Image Reconstruction Methods in Radio Astronomy

C. R. Subrahmanya

Tata Institute of Fundamental Research, Indian Institute of Science Campus, Bangalore 500 012.

Abstract

An observation of a celestial source generally results in data which are insufficient for a unique reconstruction of the desired brightness profiles of the source. Two typical situations are considered here which pertain to the reconstruction of an object from the measurements of (A) an incomplete set of its Fourier components, and (B) of a diffraction limited image of the object. The inadequacy of classical restoring schemes is evident from the violation of prior knowledge, like positivity of the brightness profile. Various schemes have been developed over the last decade for obtaining a restoration, which agrees with the measurements, as well as our prior knowledge about the source. Four such schemes are reviewed and critically discussed here. *CLEAN* is an iterative subtraction of point-components from a conventional map until the residual map is no longer significant above the noise-level. In Biraud's method, the Fourier components of the object are extrapolated in steps such that the object itself is described by a positive function. A 'Maximum Entropy Method' tries to define an 'entropy' characterising the observations as well as prior knowledge and then obtains a solution which leads to a maximum of this 'entropy'. The last scheme, an 'Optimum Deconvolution Method', attempts to optimize the solution by imposing prior knowledge as constraints on a least-squares solution which is also made to satisfy a smoothness requirement of a minimum variance of its second-differences. All these methods have been found to restore the object considerably better than the classical methods even in the presence of noise. Computationally, *CLEAN* is the most attractive method and it has been routinely used in processing two-dimensional maps with as large as $\sim 10^4$ grid-points.

INTRODUCTION

In an observational science like radio astronomy, an object of interest is seldom accessible to direct measurements. Its physical properties have to be inferred from the radiation emitted by it. Unfortunately, this radiation can be detected only after it undergoes an appreciable modification in the intervening medium and/or the observing instrument. In many cases, such modifications can be approximated either by (A) a Fourier transform (FT) relation or (B) a convolution with some known function characterising the observational environment. These are quite general processes occurring in a wide variety of situations like interferometry, diffraction, crystallography, autocorrelation measurements, antenna problems, electrical filters and so on. The image-reconstruction in such cases involves an 'inverse problem' in which the mathematical analysis is expected to reverse the naturally occurring 'cause-effect' sequence. The difficulty with such a problem can be seen from the fact that even a fundamental limitation like the finiteness of the size of an antenna or of the duration of a measurement can make the given data entirely compatible with abritrary aassumption outside the range of observations. This means that an inverse problem does not possess a unique solution and is therefore mathematically 'ill-posed'. In order to define a meaningful solution it is necessary to modify the original problem in such a way that the modified, or the 'regularised' problem possesses a 'smooth' solution, which is stable against minor perturbations in the given data (see Tikhonov and Arsenin 1977 for an elaborate discussion).

Often the solution represents the brightness distribution of sources in a given field of view or the power-spectrum of the fluctuations of a certain physical property in a medium. In such cases, any physically meaningful solution to the problem should necessarily be positive. Although such a prior knowledge about the solution is generally available in many practical situations, classical methods usually fail to give solutions compatible with the prior knowledge. Hence there have recently been several attempts to evolve numerical schemes which ensure a physically meaningful solution consistent with our prior knowledge as well as being reasonably compatible with the obser vations in the presence of noise.

The purpose of this paper is to review some typical methods of Fourier inversion or deconvolution developed and used in radio astronomy during the past decade. We will also summarise the conventional approach in order to highlight its inadequacy in this context. In order to facilitate the discussion, we will assume that the given observations have been formulated in one of the following two equations:

$$(A): \; Q(u) = \int e^{-i2\pi ux} \, q_t(x) \, dx + \text{noise} \; ; \text{ or}$$

(B): $r(x') = p * q_t + \text{noise} = \int p(x'-x) q_t(x) dx + \text{noise}$,

where $Q(u)$ or $r(x')$ are the quantities which can be obtained directly from the measurements and $q_t(x)$ is the true solution which we call the 'object'. The variable u associated with x in the FT domain is a spatial frequency if x denotes a spatial coordinate, but we will refer to it simply as a 'frequency' for brevity. The asterisk (*) denotes a convolution as in the above equation.

Since the main difficulties in solving either of the problems above can be appreciated more easily in the FT domain, we take the FT of (B) which gives

$$R(u) = P(u) Q_t(u) + \text{FT (noise)},$$

where the capital letters denote FT of the functions represented by the corresponding small letters in (B). A unique solution can be defined only if all the values of $Q_t(u)$ can be estimated from the measurements, which is seldom possible in practice. For instance, for all finite apertures, $P(u) = 0$ beyond a certain critical frequency which depends on the size of the aperture and hence the measurements in (B) can never give any information on Q_t at such frequencies. Similarly, in an autocorrelation receiver, the finiteness of the duration of observations sets a limit to the frequency upto which $R(u)$ can be measured. In aperture-synthesis interferometry, there are generally an appreciable number of intermediate frequencies at which $Q(u)$ cannot be given since the corresponding baselines may not be available for observations.

For the numerical schemes it is useful to describe the observations and the solution in terms of finite samples. We represent the data by $Q_m = Q(u_m)$ and $r_m = r(x'_m)$ in (A) and (B) respectively and assume that there are M values given. The solution, the restored image of the object, will be denoted by $q(x)$ and it is assumed that there is some knowledge of the range of x, e.g., the field of view of observations, so that the infinite limits of integration can be replaced by finite limits for numerical evaluations. It is also assumed that n samples $q_k = q(x_k)$ can adequately represent the solution $q(x)$ throughout its range. For simplicity, we will also assume that the x_k are uniformly spaced at an interval Δx. The term 'data bandwidth' is used in this paper to mean the set of u_m (or the range of u) over which measurements are feasible. Thus this includes and is restricted to the region over which $P(u) \neq 0$ in problem (B) and the values u_m at which Q_m are given in problem (A).

The algorithms in the methods discussed below will be explicitly given only for the one-dimensional problems (A) and (B), but a brief mention will be made of any particular difficulty anticipated when a straight forward generalisation of any of these is attempted for a two dimensional problem.

There is one class of problems where a great deal is known about the object and the unknown part can be reduced to a few parameters. For instance, our prior knowledge may enable us to regard the object as a superposition of a few Gaussian components whose positions, widths and amplitudes are the unknown parameters. In such cases, the prior knowledge is sufficient to make the inverse problem 'well-posed', since these parameters can be obtained by a least-squares fit, which generally has a stable solution under such circumstances. Such a procedure, called 'model-fitting', has been used in radio astronomy in a few simple problems or when the observational data are too meagre to aim at a general solution. However, these methods will not be discussed here any further.

CONVENTIONAL APPROACH

The classical approach to the inverse problems of the types discussed here is based on the linear filter theory (cf. Bracewell and Roberts 1954). The first step in such a method is to define a 'window' or a tapering function $W(u)$ confined strictly to the data bandwidth, i.e., $W(u) = 0$ if a measurement at u exists (problem A) or if $P(u) = 0$ (problem B). Then the solution $q_L(x)$ is obtained by first computing its FT, $Q_L(u)$, which is defined as :

$$Q_L(u) = \begin{cases} Q(u)\, W(u) & ; \quad (A) \\ R(u)\, W(u)/P(u) & ; \quad (B) \end{cases}$$

which can be calculated for all values of u since the values outside the data bandwidth have been defined to be zero through $W(u)$. A familiar example is the 'principal solution' which corresponds to setting the nonzero values of $W(u)$ to unity.

In the absence of noise, a restoration obtained by a Fourier transformation of Q_L is equivalent to recovering $q_L = q_t * w$, where $Q(u)$ is the FT of $w(*)$ and the asterisk (*) denotes convolution. Thus w(x) may be called an equivalent 'beam' for the restoration of the object. If the data bandwidth is rectangular, the principal solution has an equivalent beam of the form $\sin x/x$, notorious for its its unending series of ripples (sidelobes) and, in particular, the first minima on either side of origin with an amplitude -25% of the central maximum. Apart from introducing unphysical negative values into the solution, the existence of sidelobes over a wide range of x is quite annoying in the interpretation of a restored object. The sidelobe level can be restricted by a suitable choice of tapering function for $W(u)$, but this invariably leads to a loss of resolution.

From the standpoint of current philosophy of data analysis (c. f. Jaynes 1968), a classical solution is unsatisfactory on three counts: (a) it is biased since it involves arbitrary assumptions on the unavailable data e.g., $Q_L(u) = 0$ outside the data bandwidth; (b) it violates our prior knowledge about it, e.g., by including negative values (sidelobes) even when it represents the brightness distribution of an object which must be positive everywhere; and (c) the equivalent beam does not take into account the existence of noise. It may be recalled here that one is not justified in introducing the effects of noise as a perturbation on the solution to an inverse problem since the solution is unstable against small changes in the data (Turchin et al. 1970).

Linear methods are still used because of their computational ease and the availability of a detailed understanding of these methods including their limitations in any given context. Sometimes, the noise is also considered within the framework of linear filter theory by making $W(u)$ depend on the signal-to-noise ratio, as in a Weiner filter. However, this still suffers from the other drawbacks mentioned above.

Nowadays, it is often witnessed in many fields that the scope of experimental techniques has been highly restricted by the available resources and there is a pressing need to explore every feasible means to get the best out of the available data. Also, the availability of high speed digital computers has now made the exploration of new methods of restoration a far simpler task than in the past. There have thus been several new restoring methods developed over the past decade or so in several different fields. The following sections are devoted to a brief description of such restoring schemes being used in radio astronomy. However, since it is not possible to scan the literature fully within the scope of this review, we confine ourselves to a few typical examples and for a picture of the current state of art, we refer the interested reader to the proceedings of a recent symposium (Schooneveld 1979).

'CLEAN'

It is generally found that a restoration seeking consistency with prior knowledge involves much more computational effort than a conventional method. For two-dimensional problems, this can sometimes become prohibitive in spite of the advantages of prior knowledge. Our first example of the new methods is an attractive compromise which incorporates prior knowledge only partially but still provides remarkably better restoration than a conventional method without much sacrifice of computational simplicity in many cases. It is an iterative algorithm called '*CLEAN*' which aims at decomposing the object into a series of point-components (delta-functions) in the field of view. In this method, a Fourier inversion problem is also treated effectively as a deconvolution problem. In the following description, we will assume that the field of view (x_1, x_n) has been divided into n equidistant points x_k, and that the point-components are all intended to be situated at these points.

As remarked earlier, a classical solution $q_L(x)$ is characterised by an effective beam $w(x)$ whose artefacts like sidelobes are quite annoying. In many problems, the actual source, i.e., the nonzero values of $q_t(x)$, can be safely presumed to be confined to a few isolated sections in (x_1, x_n) distributed in an unknown fashion but in such a way that their total extent is a small fraction of the field of view. One can then regard the object as being equivalent to a superposition of point sources, say $k << n$ in number, whose amplitudes and locations in the field of view are to be determined. This is the central theme of *CLEAN*, first suggested by Hogbom (1974). It is an iterative algorithm for picking the grid-points x_k in such a way that a superposition of beams $w(x)$ centred at these selected grid-points will be equivalent to the classical solution. This method has mostly been used in mapping radio sources from their Fourier components obtained from interferometric measurements, where the jargons 'dirty map' and 'dirty beam' are popular for $q_L(x)$ and $w(x)$ respectively. For a detailed mathematical analysis of the method, we refer the reader to Schwarz (1978).

The various steps involved in *CLEAN* can be summarised as follows :

(a) Obtain a dirty map $q_L(x)$ as the set of its values $q_L(x_k)$, $k=1$, n, using a convenient classical method and let $w(x)$ be its equivalent 'dirty beam'. Without loss of generality, we assume a normalisation $w(0)=1$.

(b) Locate the point, say x_i, where the dirty map $q_L(x)$ has the highest absolute value, and let

$$q_{max} = q_L(x_i).$$

(c) The corresponding 'component' c_1 is now equal to a delta function at x_i of amplitude $g \cdot q_{max}$, i.e. $c_1 = g \cdot q_{max}\ \delta(x - x_i)$ where g is an emiprical parameter(~ 1) called the 'loop gain'. Now subtract the contribution of this point-source from q_L to obtain the 'residual map' q_r, i. e.,

$$q_r(x_k) = q_L(x_k) - q_{max}\, w(x_k - x_i),\ k=1, n$$

(d) Treat q as the updated dirty map and repeat the last two steps to get the next component c_2 and the corresponding residual map, and then the next component c_3, and so on, each time updating the residual map and dirty map in the same manner.

(e) The iteration cycle (b)-(d) is terminated by specifying suitable criteria like

(i) the absolute value of the amplitude of the component being less than a specified value;

or

(ii) number of iterations exceeding a specified limit.

The resulting *CLEANed* map is then the sum of all the components. For better persentation, one often convolves this map with a 'clean beam' usually a Gaussian, which does not possess any undesirable artefacts like sidelobes. Although a fixed value of g, say 0.75, can work satisfactorily in most circumstances, the convergence of iterations can be accelerated by choosing it more judiciously and, if necessary, varying it during the course of iterations by examining the residual map at intermediate stages. It can be shown that in most practical cases, the convergence is guaranteed for any choice of g in the range $0 < g < 2$ (Schwarz 1978). One of the virtues of *CLEAN* is that it naturally allows for an interactive program where one can display the residual map at regular intervals; and literally supervise the *CLEANing* process. Also, it is not necessary that the x_k inspected for the determination of the components (the 'search area') should cover the entire field of view. In most applications, the users have been able to restrict the search area to a small fraction of the total field of view and process routinely, with moderate computing efforts, two-dimensional maps with fields of view chosen on a grid matrix with as many as 512 x 512 points (n=262144).

EXTRAPOLATION OF FT USING POSITIVITY

The first attempt at incorporating positivity in the reduction of a radio astronomical observation was made by Biraud (1969), who also demonstrated clearly the possibility of super-resolution with positivity as compared to a classicial method. Since $q(x) \geq 0$, we can write it as a, where a is a real function and hence having a FT $A(u)$ which is Hermitian. The FT of q is thus an auto-convolution of the Hermitian function A. Biraud's method is to construct a sequence of Hermitian functions iteratively converging to a specific function whose autoconvolution represents $Q(u_m)$ as closely as desired. As a measure of fitting the data, Biraud used the least-squares criterion :

$$s = \sum_m \left| Q(u_m) - G(u_m) \right|^2 = \text{minimum},$$

where G=A*A. If the observations are not all of the same accuracy, one has to use a weighted least-squares criterion by suitably weighting each term in the above sum to take care of the variation of accuracy of measurement with u_m. For simplicity, however, we will treat these weights as unity in the following paragraphs.

We will now describe the algorithm for problem A. It can be applied to problem B by using $R(u)/P(u)$ in place of $Q(u)$ and using a weighted least-squares criterion to take care of the variation of the effective signal-to-noise ratio of obsevrations according to $P(u)$. It will be assumed that a sampling interval Δu is appropriate for $Q(u)$, and hence for $A(u)$; whose autocorrelation is required to represent Q. Let $2L\Delta u$ be the highest frequency available for Q, say $= u_M$. Then the bandwidth of $A(u)$ is confined at least to $\pm L \Delta u$.

The following steps summarise Biraud's method, in which any Hermitian function (even $A_0 = 0$ will work) can be used as the initial guess A_0, provided $A(u)$ vanishes for $|u| > L\Delta u$.

(a) Assume that A can be approximated by a function A_1 bandlimited to $L\Delta u$, and perform the following cycle of iterations, steps (b1)-(b3), with the iteration count l initially assumed to be equal to 1.

(b1) Define

$A_1 = A_0 + c_1\, \delta(u) + c_2[\delta(u+l\Delta u) + \delta(u-l\Delta u)] - ic_3[\delta(u+l\Delta u) - (u-l\Delta u)]$ where the coefficients c_1, c_2 and c_3 are obtained so as to minimise

$$s_1^2 = \sum_m \left| Q(u_m) - G_1(u_m) \right|^2, \text{ with } G_1 = A_1 * A_1.$$

If $Q(0)$ is known, then s_1^2 is minimised subject to the constraint $Q(0) = G(0)$ by the method of Lagrange multipliers ; i.e., by introducing the Lagrange multiplier c_4 and determining all the c_i by solving the set of equations

$$\frac{\partial}{\partial c_i}\left[s_1^2 + c_4 \left\{ Q(0) - G(0) \right\}\right] = 0, i=1, 2, 3, 4.$$

The unconstrained case corresponds to c_4=0.

(b2) Use A_1 in place of A_0 and repeat the previous step by incrementing l by 1, until all the values of l from 1 to L are covered.

(b3) Repeat the above steps until A_1 saturates to a stage where s_1^2 no longer decreases by continuing the iterations.

(c) Now check the latest available value of s_1^2 against the desired degree of agreement with the data, i. e the expected value of the variance of noise, say, $s_1^2 \approx s_N^2$, then the latest available $G_1(u)$ is the FT of the restored object, from which q(x) is easily found by a Fourier transformation. Otherwise, one tries to increase the band width A, by replacing L by $L+1$, A_0 by A_1 and going back to step (a). The whole cycle, including bandwidth extrapolation, is repeated until the agreement with data is satisfactory, or it does not improve even by extrapolation of bandwidth.

A very clear demonstration of resolution-enhancement even in the presence of noise was given by Biraud (1969). The method is generally found to converge for arbitrary initial guesses although no proof exists for its convergence. An extensive study and evaluation of the method was the subject of Wong (1971).

MAXIMUM ENTROPY METHOD

Another approach to the inverse problems is to aim at an objective criterion enabling us to choose the most appropriate solution out of the infinitely many mathematically feasible solutions consistent with the data. The idea is to use our prior knowledge to define a suitable criterion representing a least biased solution or the 'most random' solution consistent with the observations and prior knowledge. In the spirit of information theory this can be done by using the prior knowledge to construct a prior probability distribution or, equivalently its logarithm called the 'entropy' of the object. The least biased solution is then the one with maximum entropy. Such an approach for the picture-processing problem was advocated in the pioneering works of Felgett and Linfoot (1955) and Jaynes (1968) and the latter gave examples of situations where prior probability distribution could be constructed from prior knowledge. However, a prior knowledge like positivity is too insufficient to construct any probability distribution. Hence, in practical implementations, this principle has been relaxed and replaced by a similar one which assumes a certain statistical model for the description of the object or observations and uses the logarithm of the resulting *a posteriori* probability in place of entropy. Thus, there are several restoring schemes using different models for this purpose, which are all unfortunately called 'Maximum Entropy Method (MEM)' and they maximise quite different expressions under the name 'entropy'. Essentially, there have been two major schools of thought and we denote the 'entropies' used by them by H_1 and H_2 respectively. The corresponding methods will be denoted by MEM1 and MEM2 for brevity. In our notation, the definitions are :

$$H_1 = \sum \ln q_k, \qquad \text{(Burg 1967)}$$

$$H_2 = -\sum q_k \ln q_k \qquad \text{(Frieden 1972)}$$

where $q = q(x)$. For details of derivation of these expressions and illustrations with specific applications, we refer the reader to Burg (1967) and Ables (1974) for MEM1 and Frieden (1972, 1975) for MEM2. Here, we will only give a brief description of the methods, by assuming the above expressions for 'entropy'. For convenience, we will also use a common notation for both problems A and B by representing either problem in the form,

$$r_m = \sum_k P_{mk}\, q_t(x_k) + \text{noise} = \sum_k P_{mk}\, q_k + D_m$$

where D_m are called the residuals and the new unknowns are q_k. The correspondence with our earlier notation is established by the following convention :

Problem A: $r_m = Q(u_m)$; $p_{mk} = \exp(-2\pi i u_m x_k) \Delta x$

Problem B: $r_m = r(x'_m)$; $p_{mk} = p(x'_m - x_k)\ \Delta x$.

For simplicity, we will first consider the noise-free case and later indicate how noise is introduced in MEM. The specification of MEM can now be written as :

$$\text{maximise } H = \begin{cases} H_1 = \sum_k \ln q_k & \text{(MEM1)} \\ H_2 = -\sum_k q_k \ln q_k & \text{(MEM2)} \end{cases}$$

subject to the M constraints $r_m = \sum_k P_{mk}\, q_k$.

This can be solved by the method of Lagrange multipliers by introducing M Largange multipliers λ_m which are obtained along with the n unknowns of the solution by solving the system of equations

$$\frac{\partial}{\partial q_i}\Big(H + \sum_m \sum_k \lambda_m p_{mk}\, q_k\Big) = \partial H/\partial q_i + \sum_m \lambda_m p_{mi} = 0,$$

$$i = 1,\ 2, \ldots\ldots\ldots,\ n$$

along with the constraint-equations mentioned above. By substituting H_1 or H_2 for H in the above equations, one can eliminate q_k as:

$$q_k = \begin{cases} -1 / \sum_m \lambda_m\, p_{mk}, & \text{(MEM1)} \\ \exp\big(-1 - \sum_m \lambda_m\, p_{mk}\big), & \text{(MEM2).} \end{cases}$$

Thus it is only necessary to solve M (nonlinear) equations for the λ_m in MEM. Standard iterative methods like Newton-Raphson method have generally been used with success for this purpose. However, there is a very special case in which the q_k can be obtained in closed form without recourse to any iterative method. This happens for MEM1 applied for the noise-free case of problem A when $Q(u_m)$ are the first M Fourier components of $q_t(x)$, i.e., $u_m = (m - 1)\, u_2$. However, since the data are not always available in this fashion, and since noise is generally not negligible, this advantage is lost in most practical problems.

The introduction of noise is not straightforward in MEM. Here again, we see two essentially different approaches. The first scheme uses a least-squares criterion seeking a minimum of $\sigma^2 = \sum_m D^2_m / M$ simultaneously with maximising the entropy. In the second scheme, noise is also described by a model, defining an 'entropy' H_N, & both H & H_N are sought to be maximised by the solution. In order to maximise two functions simultaneously, a procedure analogous to the Lagrange multiplier methods is used in the following manner:

$$\text{maximise} \quad H - \lambda\sigma^2 \quad \text{(first scheme)}$$

$$\text{or} \quad H + \lambda H_N \quad \text{(second scheme)}$$

where λ, an analogue of Lagrange multiplier, is introduced here as an empirical (positive) constant. It is a slowly-varying function of the signal-to-noise ratio. Since it is found that the solution is not sensitive to the exact choice of λ, the fact that it is empirical does not pose any problem here. For further details and illustrations to specific problems, we refer the reader to Wernecke and D'Addario (1977) and Gull and Daniell (1978) for the first scheme and to Frieden (1972) for the second. As a slight modification of the first scheme, one may assume σ^2_N to be known and introduce the constraint $\sigma^2 = \sigma^2_N$ to determine λ as the usual Lagrange multiplier (Ables 1974), but it is generally simpler to treat it as empirical from the computational point of view.

Although two different approaches have thus been suggested in literature to consider noisy data in MEM, it is not yet clear if these two schemes lead to different results in practice. In fact, it is possible to show that the second scheme (Frieden 1972) can often be expected to be equivalent to the first scheme (e.g., Gull and Daniell 1978) for all practical purposes. Since this does not seem to have been recognised in literature, we will present the argument below.

In MEM2, Frieden (1972) obtained an expression for the 'noise-entropy' $H_N = -\sum_m (D_m+B) \ln (D_m+B)$ by following a procedure very similar to the one he employed for deriving H_2. The constant B is an empirically chosen positive number ($\geq 2\sigma_N$) such that D_m+B may be safely assumed to be positive for the desired solution for all m. An advantageous feature of this method is that the choice of B is not critical and even an overcautious choice of a very large value of B will work effectively (Frieden 1972). Thus, without loss of generality, we can assume a choice of B such that $|D_m| \ll B$ for all m, in which case, we can use the approximation $\ln (1+D_m/B) \simeq D_m/B$. Thus

$$(D_m+B) \ln (D_m+B) = (D_m+B) [\ln (1+D_m/B). B]$$
$$\simeq (D_m + B) \; (\ln B + D_m/B)$$

$$H_N \simeq \sum_m [-B \ln B + D_m(\ln B+1) - D^2_m/B].$$

Again, without loss of generality, we can assume the noise to have a zero mean and hence require that D_m also should have a zero mean, i.e., $\sum D_m = 0$. Hence $H_N = -\sum D^2_m/B + \text{constant}$. Thus, maximising H_N is the same as minimising $\sum D^2_m$, i.e. using the least-squares criterion!

If the observations were not all of the same accuracy, one would have multiplied each term in H_N or the variance of D_m by a weight related to the relative individual accuracy of the particular measurement. With this provision, the methods used by Frieden (1972) and Gull and Daniell (1978) are indeed equivalent for all practical purposes.

For the noise-free case, there are alternate ways of of arriving at MEM1 as an unbiased restoration consistent with positivity (see e.g., Schooneveld 1979a, Komesaroff and Lerche 1979). Note that when MEM1 is used in the first scheme to take note of noise, it is strikingly similar to a purely numerical scheme called the 'logarithmic penalty function method' (Fiacco and McCormick 1970) for minimising σ^2 requiring posisivity of q_k.

Summarising, we can say that MEM has certain inherent positive features which are undisputed. The most important of these is the fact that it is a scheme which automatically leads to positivity by the very definition of 'entropy', and also that it uses the data only to the extent available without making any assumptions on the unavailable data. In all the examples cited in literature, it has consistently been found to lead to a superresolution compared to a classical method. However, it is still questionable whether the term 'entropy' has indeed the same significance in picture-processing as it has in the information theory and thus whether it is justifiable to claim the 'most likely object' to be obtained solely from the specifications of MEM.

AN 'OPTIMUM DECONVOLUTION METHOD'

Our last example of a restoring method illustrates another way of looking at the inverse problem. This is to regard the problem on hand as an optimisation problem setting four criteria as the necessary requirements of the optimum solution: (a) it should satisfy a suitable stabilising criterion to ensure that it is stable against minor perturbations in the data; (b) it should be unbiased and should lead to a reasonable agreement with the data in the presence of noise, e.g., fulfilling the least-squares criterion ; (c) it should be derived without making any assumptions on the unavialable data and using the data only to the extent given; and (d) it should be consistent with our prior knowledge about it. Clearly, since the first two criteria are qualitative in nature and not mathematically precise statements, there could conceivably be several ways of meeting these requirements. One possibility was suggested by the present author (Subrahmanya 1979) as an Optimum Deconvolution Method (ODM) in connection with the restoration of an object from its Fresnel diffraction observed by the method of lunar occultations (Hazard 1976). A highlight of this method is a simple iterative algorithm for introducing positivity as a constraint on the solution. The method will be summarised below.

The first two requirements of the solution are combined in ODM by defining a 'regularised least-squares solution' (RLS) which should minimise

$$S = \sigma^2 + \lambda \sum_k (\Delta^2 q_k)^2$$

where $\Delta^2 q_k = q_{k+1} - 2\, q_k + q_{k-1}$ are the second differences and $\lambda > 0$ is an empirical parameter similar to the one introduced in MEM for handling noisy data. Its purpose is to get a RLS which is as smooth as possible (i.e., has as low a variance of second-difference as possible) but still leads to a σ^2 not appreciably different from the minimum value, or σ^2_N, if the latter is known. The exact choice of λ is not critical to within about a factor of 5 and it is also a slowly decreasing function of signal-to-noise ratio for any given problem. Thus, for a routine application, it is enough to determine it once by trial and error for a typical signal-to-noise ratio, and redetermine it only if the signal-to-noise ratio changes drastically, say by a few factors. The use of second-differences to obtain a stable solution was originally suggested by Phillips (1962) and for a detailed discussion of this and other related schemes of 'regularisation', we refer the reader to Twomey (1965) and Tikhonov and Arsenin (1977).

It can be seen that the RLS automatically meets one more requirement of the desired solution–the available data are used in the form given without any need for assuming anything about the unavailable data. As for prior knowledge, we summarise below the scheme adopted for introducing positivity and refer the reader to a more detailed paper (Subrahmanya 1980) for incorporating other types of prior knowledge and also for a critical evaluation of the method. Intuitively, one can see that an attempt to minimise the negative artefacts of the RLS will also control the positive part of the suprious ripples because of the least-squares criterion, In ODM, this is achieved by minimising, in each iteration, a weighted sum of squares of those values of the solution which were found to be negative in the previous iterations. These weights are defined in the method to be proportional to the degree of constraint-violation (square of the negative value obtained) in the previous iterations. In the actual algorithm used, the Jth iteration can be interpreted as requiring a minimum of

$$S_J = S + \sum_k \lambda_{k,J}\, q^2_k / \Omega$$

with respect to the unknowns q_k. Denoting the solution in the Jth iteration by $q_{k,J}$' we can write the new terms in the above expression as $\lambda_{k,0} = 0$ (initial solution)

$$\lambda_{k,+1J} - \lambda_{k,J} = \begin{cases} 0 \ , & q_{k,J} \geq 0 \\ q^2_{k,J}, & q_{k,J} < 0; \end{cases}$$

$$\Omega = \begin{cases} \sum_k \lambda_{k,0}, & J = 1 \\ \sum_k \lambda_{k,l}, & J \geq 2. \end{cases}$$

The purpose of Ω is only to introduce a scale factor to ensure that the added terms do not dominate in the definition of S_J. The choice given above was found to lead to a rapid convergence of iterations, mostly within 3 to 5 iterations, in the lunar occultation problem. Good results were also found from a simpler choice $\Omega = n\sigma^2_N$. In order to minimise S_J in the above iterations, one has to solve the system of equations,

$$\partial S_J / \partial q_k = 0, \ k = 1, 2, \ \ldots, n,$$

which are all linear in the unknown q_k since S_J is quadratic. For further details and computational simplifications, we refer to Subrahmanya (1980). Here we only note that the method has successfully been used for routine reductions of the observations of several hundred lunar occultations. In general, it has been found that it is possible to obtain an enhanced resolution with ODM, sometimes by more than a factor of two, over a classical method for this problem.

CONCLUSION

The methods described in the preceding sections illustrate the different possible approaches to an inverse problem. *CLEAN* is essentially a deconvolution problem and it converts any other type of problem to problem B before beginning the iterations. On the other hand, Biraud's method treats the problem as an FT extrapolation situation (problem A) before proceeding with the iterations. Neither of these methods introduce any empirical parameters. In both of these methods, the limitations to the restoration are approached in a natural manner by terminating the iterations as soon as the data and prior knowledge are satisfied reasonably by the estimated parameters — 'point-sources' in *CLEAN* and Fourier components in Biraud's method.

The other two methods — MEM and ODM— work with the problem as given without any need for mapping the data or the solution obtained from the iterations into a different domain by a Fourier transformation. Both these methods involve empirical parameters, denoted here by λ, the choice of which, however, is not found to pose any problem in practice. Unlike the other two methods, the question of the highest resolution which can be reliably obtained from these methods cannot be answered in a straightforward manner. One possible scheme would be to start with sampling the restored solution at an interval just sufficient for the classical resolution-limit, and then to repeat the restoration by choosing finer and finer sampling intervals. This will allow for the possibility of enhancing the resolution gradually and the process can be terminted as soon as there is no appreciable change in the degree of consistency with the data (σ^2) or and with prior knowledge. But this is not a very practical scheme for large problems (large n) since the methods are already time-consuming and it is too expensive to apply them several times to find out the limiting resolution.

At present it is a difficult task to discuss quantitatively the absolute limit to the resolution upto which a *reliable* restoration is possible in a given situation. Presumably, this is also the limit to which there is no appreciable difference between the various possible restorations which can satisfy both the observations and our prior knowledge. However, this opinion is not generally shared by the strong proponents of MEM who feel that the 'entropy' could still be exploited to surpass this limiting resolution. We refer to two specific papers in the current literature in order to illustrate the controversy prevailing on this issue. In a paper by Kikuchi and Soffer (1977), one finds an attempt to define the correct 'entropy' as a function of some measurable parameters of the source as well as the observing conditions. On the other hand, Kermisch (1977) gives a numerical example of simulated data which, when used in MEM1 and MEM2, gave distinct restorations which are appreciably different from each other as well as the true source although both are equally plausible in a given physical problem. The present situation may be summarized by saying that an improvement in resolution by about a factor of two compared to a classical method seems generally feasible by the methods incorporating positivity for data with signal-to-noise ratios as low as 5 or so.

Note that prior knowledge is sought to be introduced rigorously in all the methods discussed above except *CLEAN*, which is the most popular method being used in radio astronomy. It is an attractive compromise between a full implementation of prior knowledge and the computational simplicity of a classical method. It has been used routinely over years in two-dimensional problems in order to map radio sources from their interferometric observations, where the number of unknowns can be as large as $\sim 10^4$. On the other hand, except for a few isolated attempts with MEM, the other methods have generally been restricted to one-dimensional problems in view of the heavy computing requirements. However, this situation is likely to improve in the near future in view of the rapid developments taking place in the field of image-reconstruction and the increasing availability of faster computers and array-processors.

ACKNOWLEDGEMENT

I thank Dr. A. P. Rao and Dr. V. K. Kapahi for their comments on the manuscript.

REFERENCES

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

Biraud, Y. 1969, *Astr. Astrophys.*, **1**, 124.

Bracewell, R. N. and Roberts, J. A. 1954, *Austr. J. Phys.*, **7**, 615.

Burg, J. P., 1967, *Maximum Entropy Spectral Analysis*, Stanford Univ., Geophysics Dept., (Paper presented at the 37th Annual Society of Exploration Geophysics Meeting, Oklahama City).

Fellgett, P. B. and Linfoot, E. H. 1955, *Phil. Trans.*

Royal Soc. (*London*), **A 247**, 369.

Fiacco, A. V. and McCormick, G. P. 1970, *Nonlinear Programming*, Wiley.

Frieden, B. R. 1972, *J. Opt. Soc. America*, **62**, 511.

Frieden, B. R. 1975, in *Picture Processing and Digital.*

Filtering, Ed. Huang, T. S., Springer-Verlag, 176.

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

Hazard, C. 1976, in *Methods of Experimental Physics*, **Vol. 12C**, Ed. Meeks, M. L., Ch. 4.6.

Hogbom, J. A. 1974, *Astr. Astrophys. Suppl.*, **15**, 61.

Jaynes, E. T. 1968, *IEEE Trans. Syst. Sci. Cyber. SSC*-4, 227.

Kermisch, D. 1977, *J. Opt. Soc. America*, **67**, 1154.

Kikuchi, R. and Soffer, B. H. 1977, *J. Opt. Soc. America*, 67, 1656.

Komesaroff, M. M. and Lerche, I. 1979, in Schooneveld, (1979), p. 241.

Phillips, D. L. 1965, *J. Assoc. Comp. Machy.*, **9**, 84.

Schooneveld, C. van, 1979, (Ed.), *Image Formation from Coherence Functions in Astronomy,* D. Reidel Publishing Co.

Schooneveld, C. van 1979a, in Schooneveld (1979), p. 197.

Schwarz, U. J. 1978, *Astron. Astrophys.,* **65**, 345.

Subrahmanya, C. R. 1979, in Schooneveld (1979), p. 287.

Subrahmanya, C. R. 1980, *Astron. Astrophys.,* (in press)

Tikhonov, A. N. and Arsenin, V. Y. 1977, *Solutions of Ill-posed problems,* (New York, Winston/Wiley).

Twomey, S. 1965, *J. Franklin Inst.,* 179, 95.

Turchin, V. F., Kozlov, V. P. and Mahevich, M. S. 1970, *Sov. Phys. Uspekhi,* **13,** 681.

Wernecke, S. J. and d' Addario, L. R. 1977, *IEEE Trans. Computers,* **C26,** 351.

Wong, H. F. 1971, A Study and Evaluation of Biraud's Deconvolution Algorithm (Masters thesis), Electrical Engg. Dept., Queen's Univ., Kingston, Ontario, Canada.

WE HEAR THAT

Dr. Yash Pal, Director of the Space Application Centre of the Indian Space Research Organization and presently President of the Indian Physics Association has been awarded the Sixth Marconi International Fellowship, along with a grant of U. S. $ 25,000 for his work on Satellite Instructional Television Experiment (SITE) in India in conjunction with the NASA's ATS-6 Satellite.

Professor Jayant V. Narlikar, Tata Institute of Fundamental Research, has been awarded the Bhatnagar Award in Physics for 1978.

Dr. K. S. Krishna Swamy, Tata Institute of Fundamental Research has been selected for the Shri Hari Om Ashram Prerit Dr. Vikram Sarabhai Research Award in field of Planetary and Space Sciences for 1979.

ANNOUNCEMENT

The Eleventh International Conference on Solid State Track Detectors will be held between September 7-12 1981 at the University of Bristol, England. For further details, contact Professor P. H. Fowler, H. H. Wills Physics Laboratory, Tyndall Avenue, Birstol BS8 1TL, England.