MAXIMUM ENTROPY IMAGE RESTORATION IN ASTRONOMY[1]

Ramesh Narayan

Steward Observatory, University of Arizona, Tucson, Arizona 85721

Rajaram Nityananda

Raman Research Institute, Bangalore 560080, India

1. INTRODUCTION

Imaging the two-dimensional intensity distribution of the sky has always been an important part of astronomy. This is particularly true at present, a time when aperture synthesis mapping is firmly established in radio astronomy, charge-coupled devices are revolutionizing optical imaging, and X-ray-imaging cameras are being flown in space.

Atmospheric irregularities, instrument aberrations, detector noise, and the diffraction limit all cause the observed image to deviate from the ideal one. Image restoration techniques have therefore had a long history. The field owes much to the classic papers of Bracewell & Roberts (12) and Fellgett & Linfoot (35), which focused attention on the amount and nature of information about the image present in the measurements. These ideas have become increasingly relevant with the growth of interferometry, where the data correspond to the Fourier transform of the image. The Michelson stellar interferometer (82) was an early application in optical astronomy. However, it is at radio frequencies where interferometry has proved most fruitful and where astronomers have had to face the problem

---

[1] The US Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

127

of constructing an image from incomplete knowledge of its Fourier coefficients.

An important step forward was the realization that the information contained in the measurements should be supplemented by prior knowledge about the image. A powerful example of a priori information is the positivity of intensity. In X-ray crystallography, positivity of electron density has been the cornerstone of the remarkably successful "direct methods" (71, 78), while in astronomy the early work by Biraud (10) demonstrated that resolution enhancement is possible by using the positivity constraint to supplement the data. Another successful application of a priori information is the widely used Clean algorithm of Högbom (56, 58), which exploits the fact that point sources dominate many radio images. More recently, the Maximum Entropy Method (MEM) of inference introduced by Jaynes (62, 66) has been widely used and even more widely discussed for image processing. This method, which is reviewed here, aims to obtain the most probable nonnegative image consistent with the data, based on the number of ways in which such an image could have arisen.

Early work by Frieden and collaborators (40, 42) and Ables (1) demonstrated the promise of the MEM for image processing, and this potential has been largely fulfilled in later work. Applications to real data include the image of Ganymede produced by Frieden & Swindell (43), as well as Gull & Daniell's (49) work on radio images. Applications to X-ray (34) and gamma-ray (109) data have also been made. Figure 1 shows two examples where the MEM has clearly resulted in remarkable improvement over the raw unprocessed image. With the availability of large computers and an improved understanding of the method, it is likely that use of the MEM in astronomy will continue to increase. A serious study of its properties and limitations is therefore useful, not only to the producers of MEM maps, but also to the wider community of consumers who need to examine them critically.

Understanding of the MEM has so far mostly been based on discussions of the conceptual foundations of the method, which are reviewed in Section 2. However, enough experience in the use of the MEM has accumulated to attempt a more systematic and detailed analysis of its properties, including a discussion of the role of noise. This is the subject of Section 3. Remarkably, many of the results here do not depend on the precise form of the entropy function being maximized. Practical algorithms for entropy maximization are covered in Section 4. Extensions of the MEM that have been proposed for specialized applications such as polarization data, spectral mapping, and phase refinement are discussed in Section 5. These are areas where much work remains to be done. A comparison of the MEM with other image-restoration techniques, in particular the Clean

algorithm and various positivity-enforcing schemes, is made in Section 6. Finally, Section 7 gives our view of how far the MEM has come and what the outlook is for the future.

To set the stage for our discussion of the MEM, we start in Section 1.1 below with a brief review of the image-restoration problem as it arises in single-aperture and synthesis (i.e. interferometric) measurements in astronomy. We then introduce the maximum entropy (ME) solution to this problem in Section 1.2.

1.1   *The Image-Restoration Problem*

We consider a somewhat idealized and simplified picture of the image-restoration problem. Let $I(x, y)$ be the true two-dimensional intensity distribution (we refer to this as the "object" or "true image"), where x, y are orthogonal angular coordinates in some small region of the sky. To clarify the conceptual ideas involved, we consider the case of ground-based optical or infrared imaging by a single aperture. Atmospheric "seeing" (125) causes a blurring of the image, so that the measured image $D(x, y)$ (where D stands for data), after correcting for spatial variation of the detector response ("flat-fielding"), takes the form of a convolution

$$D(x, y) = \iint P(x - x', y - y') I(x', y') \ dx' \ dy' + N(x, y). \qquad 1.$$

The function $P(x, y)$ is the point-spread function (psf) of the telescope and describes the measured image corresponding to an isolated point source. For a telescope with a circular aperture, $P(x, y)$ consists of a central peak at $x = y = 0$ with a characteristic half-width w, which measures the resolution limit imposed by the seeing disk. $N(x, y)$ is the noise in the data. Equation 1 shows that the fine details in the object are washed out in the measured image by the blurring due to the psf as well as the presence of the noise. Image-restoration techniques attempt to overcome such degradation, which is inherent to the measuring process.

The information about the object that is contained in the measurements is best appreciated in the Fourier plane. Let us represent the image as a Fourier series

$$I(x, y) = \sum_{u,v=-\infty}^{\infty} \tilde{I}(u, v) \exp \left[2\pi i(ux + vy) \right] \qquad 2.$$

with coefficients given by

$$\tilde{I}(u,v) = \tilde{I}^*(-u, -v) = \int\int_{-1/2}^{1/2} I(x, y) \exp \left[-2\pi i(ux + vy) \right] \ dx \ dy. \qquad 3.$$
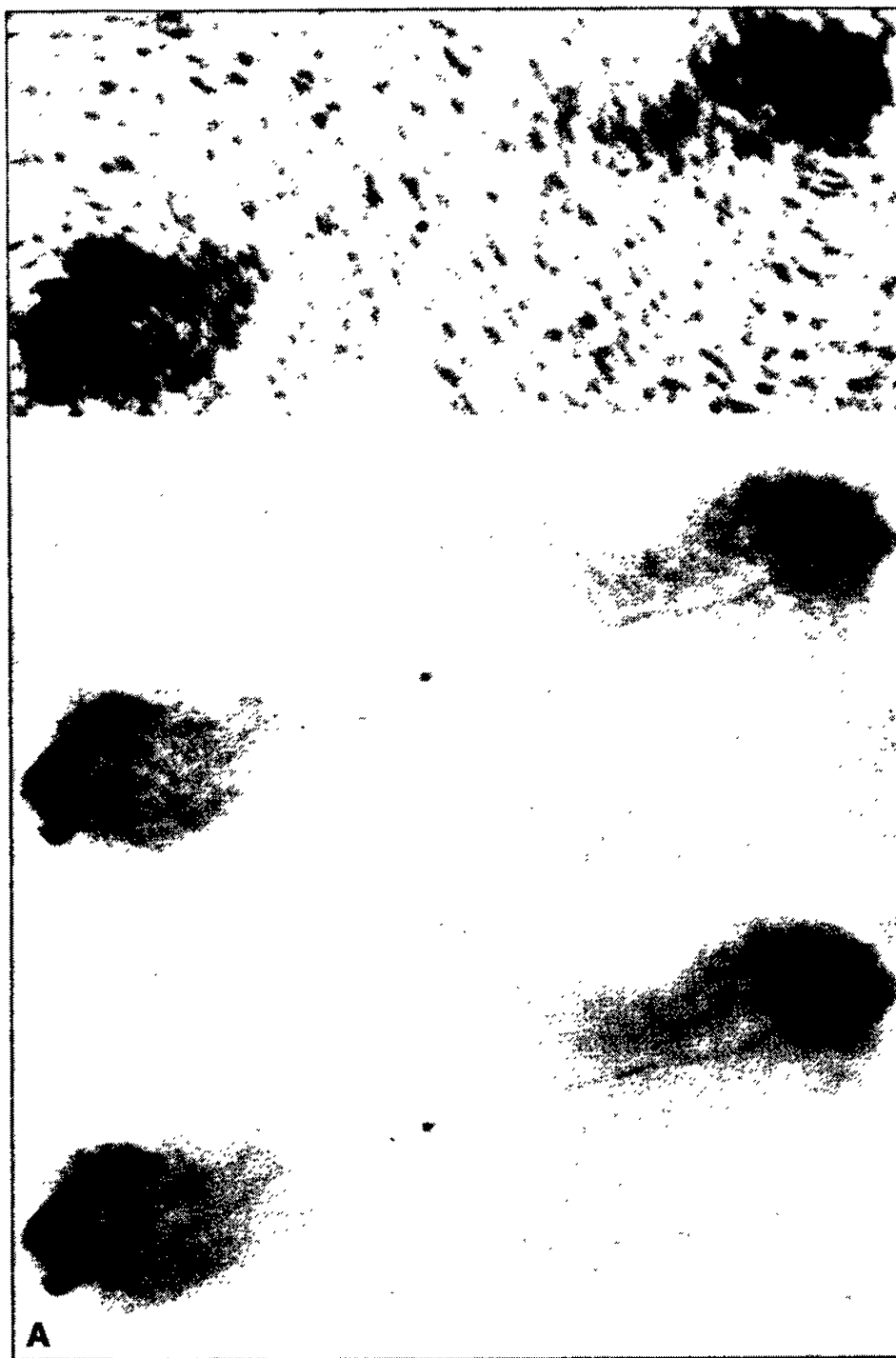
Figure 1   (*A*) An example of maximum entropy image restoration in radio astronomy. Top: An unprocessed image ("dirty map") of the radio source Cygnus A made at 6 cm with the Very Large Array of the NRAO (operated by Associated Universities, Inc., under contract with the National Science Foundation). Middle: A maximum entropy reconstruction from the same data. Bottom: A maximum entropy reconstruction after the data have been "self-calibrated" (92). Observers: R. A. Perley, J. W. Dreher. (*B*) Top: An optical photograph of the galaxy M87. Bottom: A maximum entropy reconstruction [Bryan & Skilling (18), with permission from the Royal Astronomical Society].
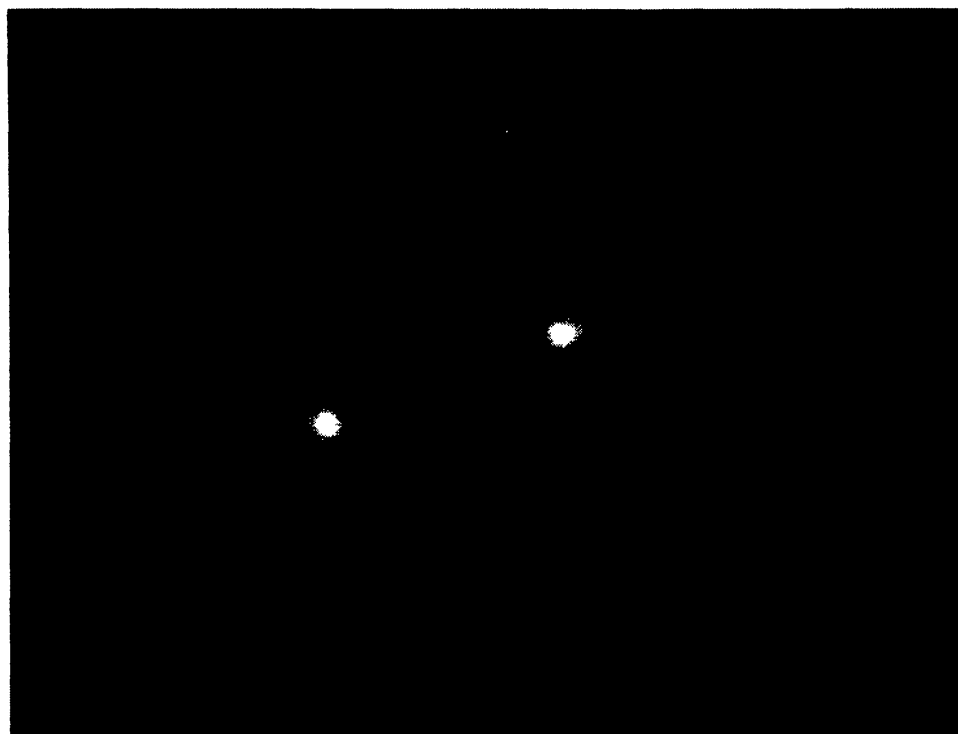
B

Figure 1 (continued)

132   NARAYAN & NITYANANDA

Equation 1 then becomes

$$\tilde{D}(u,v) = \tilde{P}(u,v)\tilde{I}(u,v) + \tilde{N}(u,v),$$ 4.

where the Fourier transforms $\tilde{D}$, $\tilde{P}$, and $\tilde{N}$ have definitions similar to that of $\tilde{I}$. We have for convenience adopted units such that the region of sky being mapped extends from $-1/2$ to $+1/2$ along x and y. The integers u and v are called spatial frequencies, since larger values correspond to a more rapidly oscillating function of the angles x and y. For the form of $P(x,y)$ described above for a single aperture, $\tilde{P}(u,v)$ consists of a central peak at $u = v = 0$ with a half-width $\tilde{w} \sim 1/\pi w$, proportional to the reciprocal of the width of the seeing disk. Thus, by Equation 4, Fourier amplitudes $\tilde{I}(u,v)$ in the true image with $(u^2+v^2)^{1/2} > \tilde{w}$ are severely reduced in the data $\tilde{D}(u,v)$ as a result of the falloff in $\tilde{P}$. Since the noise usually does not vary rapidly with u,v, this means the higher spatial frequencies in the measured image are progressively swamped by the noise, and so there is no direct way of inverting Equation 4 to obtain $\tilde{I}(u,v)$. Hence, one needs image restoration techniques to estimate the high-spatial-frequencies in the object.

X-ray imaging also is described by Equation 1, and many of the cameras flown in space [e.g. the *Einstein* telescope (46)] have point-spread functions very similar to the single-aperture case described above. However, there is an alternative design involving the use of a coded mask in front of the telescope aperture (110), where the psf consists of a number of peaks scattered over a wide area. Here, the deconvolution of Equation 1 is essential before any interpretation of the observations is possible, in contrast to the case discussed earlier, where the image is a smoothed version of the object and therefore resembles it even before processing.

Mapping at radio frequencies using aperture synthesis (38) is somewhat different from the above. This is a case of "indirect imaging" (96), where the data consist directly of the Fourier transform $\tilde{I}(u,v)$ of the object (also called the visibilities), measured interferometrically by correlating the signals received by a pair of telescopes. The indices u and v are proportional to the spatial separation in wavelengths of the two telescopes and are hence also called baseline components. Owing to practical limitations, one obtains measurements only over a finite set of baselines (u,v), which we denote by K (standing for "known"). Thus, all spatial frequencies beyond a certain maximum spacing are unmeasured. In addition, gaps in the data may also occur at lower frequencies, often near the origin in Fourier space. We denote the measured and unmeasured frequencies by $(u,v) \in K$ and $(u,v) \notin K$, and thus the situation is described by Equation 4 with

$$\tilde{P}(u,v) \begin{cases} = 1, & (u,v) \in K, \\ = 0, & (u,v) \notin K. \end{cases}$$ 5.

© Annual Reviews Inc. • Provided by the NASA Astrophysics Data System

Because of the abrupt transition in the magnitude of $\tilde{P}(u,v)$ between measured and unmeasured spatial frequencies, the psf $P(x,y)$ (called the "dirty beam" in radio astronomy) now consists of secondary peaks, or "sidelobes," of appreciable amplitude surrounding a central peak of finite width. If the unmeasured $\tilde{D}(u,v)$ are set to zero, the measurements alone (restricted to $u,v \in K$) can be Fourier inverted to generate a map $D(x,y)$. This image is called the "principal solution" or, in radio astronomy, the "dirty map." The sidelobes in the psf cause oscillations, or "ripple," in the principal solution around strong point sources and sharp edges. At a deeper level, these artifacts are a consequence of setting unmeasured Fourier coefficients to zero, which is an arbitrary procedure (1). They make it extremely difficult to see weak features in the map, and so image restoration gets elevated from a luxury to a virtual necessity.

Summarizing the various cases discussed above, we note that the goal of image processing is to extract from the measured data an approximation to the true image with reduced ripple and possibly improved resolution, making due allowance for noise. This involves some form of *extrapolation* in the Fourier domain. This is certainly true in aperture synthesis, where the high-frequency data are missing altogether. (In this case one also needs interpolation to fill in the missing shorter spacings.) But it is also true in the single-aperture case discussed earlier, where the algorithm should give less weight to high-spatial-frequency data that are badly corrupted by noise and instead give more importance to extrapolated values obtained from better data at lower frequencies.

The classical method of reducing the effect of sidelobes and high-frequency noise is to convolve $D(x,y)$ with a suitable smoothing function, such as a Gaussian. This is equivalent to multiplying $\tilde{D}$ by a "filter," a process that unfortunately further attenuates higher spatial frequencies. Ables (1) has pointed out the irony of going to great expense to build large single apertures or arrays, only to damp out the costliest measurements to insignificance! Also, such a linear operation cannot generate nonzero values at unmeasured spatial frequencies, nor can it sensibly extrapolate into noise-contaminated regions of the uv plane. We thus conclude that successful image restoration requires *nonlinear* methods (41, 117).

Several good reviews are already available, describing a variety of linear and nonlinear algorithms for astronomical image restoration (5, 11, 41, 113, 119). A common feature of all nonlinear algorithms is that they generate nonzero values for unmeasured (or very noisy) data instead of making them zero as in linear methods. Such a procedure is sensible because these methods use a priori information, such as positivity of intensity or finite spatial extent of the source of interest. A natural effect of this extrapolation into unmeasured regions of data space is increased resolution. The actual amount of "superresolution" obtained will clearly

（左側縦書き）1986ARA&A..24..127N

depend on how strongly the a priori information constrains the un-measured data. However, a modest degree of superresolution (not more than a factor of two) will often be possible and need not be treated with suspicion. We should distinguish clearly between superresolution based on a priori information and that based on analyticity of the data. The latter requires infinite signal to noise (5, 117) and is only of academic interest.

A note on notation. In parts of Sections 2 and 4, the discussion requires only the general structure of the relation between the object, the psf, the data, and the noise; thus, we write

$$D_r = \sum_{i=1}^{n} P_{ri}I_i + N_r, \qquad r = 1, 2, \ldots, m. \qquad 6.$$

Here I_i is the intensity in the ith pixel, $P_{ri}I_i$ is the contribution it makes to the measurement D_r, and N_r is additive noise. In Section 3, however, we use the more concrete Fourier relation of Equation 4, which is a particular case of Equation 6.

1.2   Maximum Entropy Image Restoration

The MEM is a particular nonlinear image restoration scheme that is being increasingly applied in a number of fields. Here, the entropy of an image $I(x, y)$ is defined to be

$$S = \iint f[I(x, y)] \, dx \, dy, \qquad 7.$$

where f is a suitably selected function. This expression for the entropy is maximized subject to the constraints imposed by the data. Without the ME condition, the inversion of Equations 1 or 4 is *ill posed*, since the data can be satisfied by an infinity of maps that differ from each other at unmeasured spatial frequencies by arbitrary amounts, and at measured frequencies by amounts consistent with the noise. The condition that the entropy be a maximum selects one among these many possible solutions and therefore *regularizes* the problem (115).

Two forms have been proposed for the entropy function f, namely

$$f_1(I) = \ln I, \qquad f_2(I) = -I \ln I. \qquad 8.$$

The standard arguments in favor of these functions are discussed in Sections 2.3 and 2.4. Here we mention two important properties possessed by both forms. Firstly, they do not permit negative values of I, thus automatically imposing the positivity constraint on the image (see Section 6.2 for more on positivity). Secondly, when the only available measurement is the total flux, both forms have maximum entropy for a uniform image

with a constant intensity. This suggests that the MEM is maximally non-committal regarding unmeasured data (1), and that it produces images that are as featureless as possible (93). The entropy forms in Equation 8 are sometimes defined in terms of intensity proportions $I_i/\Sigma\, I_i$. This introduces additive and multiplicative constants (e.g. see Equation 15), which are of no consequence in the maximization of entropy.

The more practical-minded reader could at this point go directly to Section 3, which describes general properties of ME images. Perhaps this will whet his appetite for the discussion of the foundations of the MEM given in Section 2.

2. ORIGINS AND MOTIVATIONS OF THE MAXIMUM ENTROPY METHOD

The concept of entropy in thermodynamics has intrigued generations of physicists, chemists, and biologists. Its generalization has played an important role in the theory of probability, communication, and statistical inference. The application to images is in a sense a particular case of the last of these. This section reviews some of this background material very briefly and attempts to guide the reader through the literature on the problem of assigning a functional form to the entropy of an image.

2.1 *The Concept of Entropy*

As introduced by Boltzmann into statistical mechanics, the entropy S is a measure of the number of microscopic ways that a given macroscopic state can be realized. For W alternatives, each with probability $p_W = 1/W$, the entropy is

$$S = \ln W = -\ln p_W.$$

When the alternatives, labeled by i, have probabilities p_i, this equation is generalized to the following average:

$$S = -\overline{\ln p_i} = -\sum_i p_i \ln p_i. \qquad 9.$$

The choice of the logarithm in these equations ensures the desirable property that the entropy behaves additively when one multiplies the probabilities for two independent systems. Thus

$$-\sum_i \sum_j p_i q_j \ln (p_i q_j) = -\sum_i p_i \ln p_i - \sum_j q_j \ln q_j,$$

where we use $\Sigma p_i = \Sigma q_j = 1$. Indeed, this property of additivity, combined with a few more reasonable axioms, leads uniquely (73) to Equation 9.

The classic papers by Shannon (101, 102) introduced the idea that the information gained in a measurement depends on the number of possible outcomes out of which one is realized. Equation 9 applied to messages is the basis for the mathematical theory of communication. The direct generalization for a continuous variable with probability density $p(X)$ gives

$$S = - \int p(X) \ln [p(X)] \, dX.$$

However, if we change to a new variable Y with density $q(Y)$ given by $p \, dX = q \, dY$, the entropy would read

$$S = - \int q(Y) \ln [q(Y)/J(Y)] \, dY,$$

where $J(Y)$ ($= dX/dY$ in one dimension) is the Jacobian of the transformation from Y to X. We thus see that in general the entropy must be written as

$$S = - \int p(X) \ln [p(X)/p_0(X)] \, dX, \qquad 10.$$

where the function $p_0(X)$, known as the "prior," can be regarded as containing information about the coordinate system used. The analog of $p_0(X)$ for a discrete variable would be a degeneracy factor g_i associated with the state i. The probability p_i has to be distributed over g_i substates, and adding the contributions to the entropy gives

$$S = - \sum_i p_i \ln (p_i/g_i). \qquad 11.$$

An important property of entropy is that for a uniform prior, it is maximum when $p_i = \text{constant} = 1/(\text{number of states})$ or $p(X) = \text{constant}$. This is intuitively reasonable, since a flat probability function represents a state of maximum ignorance or minimum information. More precisely, the maximum is attained when p_i and $p(X)$ are proportional to their respective priors g_i and $p_0(X)$. The choice of prior in a problem is therefore rather important. In favorable cases, a prior may be suggested by symmetry (64), but in general, different priors (which express different counting rules) are all equally legitimate from the point of view of pure probability theory. Only further experimental or theoretical input can single out a particular prior. For example, in statistical mechanics, equal weights are assigned to equal numbers of energy levels or equal volumes in phase space. The ultimate justification for this has to be sought in the dynamics of a many-particle system [see (63), however].

2.2 The Bayesian Strategy

Out of all images that satisfy the measurements, it is natural to select the one that maximizes a given measure of entropy. A parallel procedure is used in statistical mechanics, where, given the total energy and number of particles of a gas, entropy maximization is used to deduce a distribution that is uniform in space and Maxwellian in velocity. We can picture a complete collection of images corresponding to all possible intensity distributions. On this, the measurements act as a filter, restricting our attention to those images that satisfy the data to within the noise. Among these, the one that could have arisen in the maximum number of ways (which depends on our counting rule) is a natural choice. Such an approach to statistical inference was suggested as early as 1763 by Bayes (8) and has been championed in recent times by Jaynes (67). The basic idea is contained in the following relation between conditional probabilities:

$$P(A\,|\,B) = P(A)P(B\,|\,A)/P(B). \tag{12.}$$

In our application, A is the unknown image (represented by the set of intensities I_i), and the data D_r (represented collectively by B) are a noisy convolution of the image with the psf P_{ri} (Equation 6). In Equation 12 we are interested in evaluating the conditional probability $P(A\,|\,B)$ of an image A given the measurements B. The quantities on the right-hand side have the following meanings. $P(A)$ is the a priori distribution of images based on our counting rule and is related to the entropy S of the image by

$$P(A) \equiv P(I_i) \propto \exp[S(I_i)]. \tag{13.}$$

$P(B\,|\,A)$ is the conditional probability of obtaining the measurements B for a given image A, given by

$$P(B\,|\,A) \propto \prod_r \exp(-N_r^2/2\sigma_r^2) = \prod_r \exp\left[-\left(\sum_i P_{ri}I_i - D_r\right)^2 /2\sigma_r^2 \right]$$

for the simple case of uncorrelated Gaussian noise N_r with variance σ_r^2. The denominator $P(B)$ is independent of A and serves merely to ensure the correct normalization. Given the data, the most probable image is obtained by maximizing $P(A\,|\,B)$, or equivalently by maximizing the product of the previous two equations. Taking the logarithm, we thus need to maximize

$$\ln P(A|B) = S(I_i) - \sum_r \left(\sum_i P_{ri}I_i - D_r\right)^2 /2\sigma_r^2, \tag{14.}$$

which is a linear combination of two terms: (a) the entropy of the image, and (b) a quantity corresponding to χ^2 in statistics measuring the dis-

crepancy between the data and the predictions of the model. Although we have chosen Gaussian noise for simplicity, the Bayesian approach is general and has been applied to the Poisson noise case in optical (44) and gamma-ray (109) astronomy as well.

Thus, the application of the MEM to image reconstruction fits very naturally into the Bayesian framework. While the underlying Equation 12 is not controversial, the choice of an a priori distribution $P(A)$ has been debated in statistics (72) for a long time, and the same is true of $P(I_i)$ in image restoration.

2.3   Choice of the Entropy Function

In his pioneering work on estimating a power spectrum $I(\omega)$ from its first n autocorrelations, Burg (20) used the expression $\int \ln [I(\omega)]\, d\omega$ for the entropy. To motivate this, one can loosely think of $I(\omega)$ as the absolute square of a Fourier coefficient $E(\omega)$ of a sufficiently long stretch of the time series. For a Gaussian stationary random signal, the individual Fourier coefficients are independent Gaussian variables. Then the entropy associated with each can be shown to be proportional to ln (variance), i.e. to $\ln [I(\omega)]$. (For instance, the entropy associated with the Maxwellian velocity distribution at temperature T is proportional to $\ln T$.) In the case of aperture synthesis in radio astronomy, the electric field distribution on the aperture plane is a two-dimensional stationary Gaussian random function, whose power spectrum is the sky intensity distribution $I(x, y)$. From this point of view, the entropy expression

$$S_1 = \int\!\!\int \ln [I(x,y)]\, dx\, dy$$

is applicable to images (1, 93, 122). Another derivation (74) is based on the thermodynamic entropy associated with the radiation field that constitutes $I(x, y)$. In the limit of a large number of photons per mode ($n \gg 1$), which is valid in the radio domain, the Bose formula (79) for the entropy, $(1+n) \ln (1+n) - n \ln n$, reduces to $\ln n$.

In contrast, the early work of Frieden (40) on image reconstruction was based on the entropy form

$$S_2 = -\int\!\!\int I(x,y) \ln [I(x,y)]\, dx\, dy.$$

In one motivation for this form, the image is built up from equal "luminance elements" (65) placed at random over the field—in Gull & Daniell's (49) picturesque analogy, as if monkeys were given building blocks. For a map with intensity I_i in the ith pixel, we normalize the intensity by the

total flux $D_0 = \Sigma I_i$ and define the fractional intensity $f_i = I_i/D_0$. The usual combinatorial arguments used in statistical mechanics (79) then show that the number of ways W that an image can be built up is given by $W = \exp(S_2)$, where

$$S_2 \approx -N \sum_i f_i \ln f_i = -(N/D_0) \sum_i I_i \ln I_i + \text{constant}. \qquad 15.$$

(Henceforth we disregard additive constants in the entropy, since they do not affect the maximization condition.) Another approach (74) considers the thermodynamic entropy in the limit of a small number $n \ll 1$ of photons per mode, which is usually valid at optical and higher frequencies. This reduces to the form $-n \ln n$. It has been pointed out (50) that the strictly thermodynamic approach gives different answers in the radio and optical domains, a result that is undesirable. Yet another approach (108) considers the entropy associated with the probability distribution for the arrival of the next photon [see (27) for a recent critical discussion of this approach].

The two distinct entropy expressions S_1 and S_2 do not contradict each other, since they correspond to different probability distributions that one can associate with the same image. The basic question, as posed by Wernecke (120), is "Entropy of what?" This has not yet been conclusively answered, although for image restoration the form S_2 seems to have more supporters at present.

2.4 *Axiomatic Approaches to the Inference Problem*

Since the Bayesian strategy is impeded by our ignorance of the true entropy function, many authors have attempted to derive the ME principle from other considerations. As early as 1957, Jaynes (62) suggested that any function other than S_2 may lead to logical contradictions. More recently (51, 105, 106, 116), formal arguments have been advanced to prove that, given expectation values of various observables $\langle A \rangle = \Sigma A_i p_i$, the only consistent way to determine the underlying probability distribution is to maximize $-\Sigma p_i \ln p_i$. The essential point is that when we deal with measurements on two independent systems with states labeled by X and Y, there are two approaches that must give the same probability for the joint system: (*a*) Maximize the entropy associated with the probability distributions $p_1(X)$ and $p_2(Y)$ separately (each constrained by its own data) and form the product, which by construction will have a factorized form. (*b*) Construct a joint distribution $p(X, Y)$ by maximizing the entropy while simultaneously satisfying the measurements on X as well as Y. Only S_2 gives a factorized solution by method (*b*) (51) and could therefore be singled out as the only consistent choice.

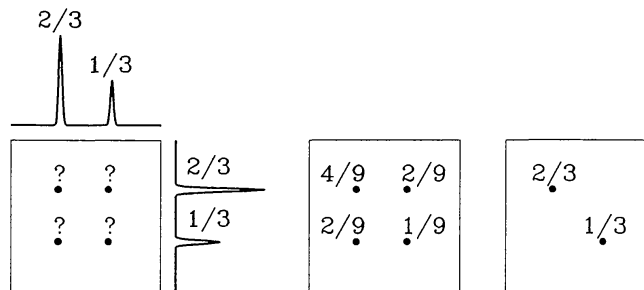There are three points to keep in mind concerning the axiomatic approach.

Figure 2 (*Left*) Projections on two perpendicular axes of an unknown image. (*Middle*) A reconstruction from these projections by the maximum entropy method with $-I \ln I$, showing the factorized form with four sources on a perfect rectangle. (*Right*) An astronomically more plausible model with only two sources fitting the same data.

1. It is not obvious that a brightness distribution can be identified directly with a probability distribution. Hence, a factorized solution is not necessarily the best choice in image restoration. In fact, Figure 2 gives an example showing that a factorized image would be most undesirable in a simple example with two sources and two perpendicular scans.

2. For continuous probability distributions, the form of the entropy is given by Equation 10 with a general prior distribution $p_0(X)$, and even in the discrete case a degeneracy factor g_i is possible (Equation 11). With this added freedom, there is a wide variety of solutions (48).

3. The a priori information that the two systems are independent should be included by maximizing the entropy over distributions of the factorized form $p_1(X)p_2(Y)$. When this is done, many more expressions, such as $p^{-\alpha}$ and $\ln p$, also lead to consistent results (70).

2.5 *Entropy Maximization and A Priori Information*

Jaynes (62) and Burg (20) have noted that the MEM gives solutions that are positive, as required for probabilities, power spectra, and images. Högbom (57) pointed out that functions like $\ln I$ and $-I \ln I$ have two vital properties: (*a*) infinite slope at $I = 0$, which enforces positivity; and (*b*) a negative second derivative, which discourages ripple. This is illustrated in Figure 3. Also, Subrahmanya (114) has obtained successful reconstructions from lunar occultations by maximizing a "penalty function" that discourages negative values and rapid variations. All this suggests that in favorable cases, very general a priori information like positivity and absence of ripple may be used to constrain the solution rather tightly, and that the detailed form of the entropy/penalty function may not be of much consequence so long as it satisfies the above general properties. Tests with different entropy forms such as S_1 and S_2 with model

data do indeed confirm that the results are similar (55, 88, 121). More detailed reasons for this are given in Section 3.1.

In implementing the penalty function approach, one maximizes a linear combination of the χ^2 statistic and the penalty function in order to satisfy both the data and the a priori information. This is rather similar to what is done in the Bayesian approach (Equation 14) if we identify the penalty function with the logarithm of the a priori distribution. Experimenting with different penalty functions can thus be regarded as equivalent to exploring the consequences of different a priori distributions, which is regarded (68) as an important part of the Bayesian framework.

3. GENERAL PROPERTIES OF MAXIMUM ENTROPY IMAGES

Having decided to maximize a suitable entropy function, one would like to know what kind of images to expect. The standard statements that the image is "maximally noncommittal" or "as featureless as the data allow" are too general to be useful. Fortunately, as we show in this section, the actual condition for the entropy to be a maximum allows a rather detailed analysis of the general properties of ME images (just as Newton's laws give a better feeling for the motion of a particle than does the principle of least action). The material that follows is unavoidably long and technical, since we feel that the nature of ME images deserves a detailed and self-contained treatment. Hopefully, the following preview will help readers locate the main results. The form of the ME image is described by Equations 18 and 19 below and is illustrated in Figure 4, on the basis of which many properties of the restoration become rather easy to understand. These are listed in Section 3.1.1 under points (1)–(10) for the conceptually
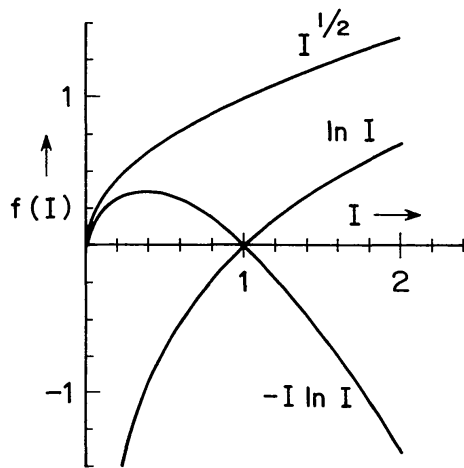

Figure 3 Graphs of the functions ln I and $-I \ln I$, showing the infinite slope at $I = 0$ and negative second derivative that make them suitable as penalty functions for image restoration (57). The function $I^{1/2}$, which shares these properties and represents an intermediate case, is also shown.

1986ARA&A..24..127N

142 NARAYAN & NITYANANDA

simplest case, namely noise-free interferometric data. Section 3.1.2 is devoted to the effect of noise and the important question of the residuals, while Section 3.1.3 deals with the single-aperture case. The properties (1)–(10) retain their validity even in these more general situations, while additional features are listed under points (11)–(15). Section 3.2 describes various suggestions that have been made to overcome the defects of the MEM. The different choices of entropy function are compared in Section 3.3, and from this comparison the unique properties of $-I \ln I$ are revealed.

3.1 Form of the ME Image

3.1.1 NOISE-FREE INTERFEROMETRY In order to introduce the ideas one at a time, we start with the simplest situation here and generalize in later subsections. Consider the problem of reconstructing $I(x,y)$ from noise-free interferometric measurements $\tilde{D}(u,v)$ over the "known" spatial frequencies $(u,v) \in K$. Each data constraint is of the form (Equations 4 and 5 with $\tilde{N} = 0$)

$$\tilde{I}(u,v) = \tilde{D}(u,v), \qquad (u,v) \in K, \qquad\qquad 16.$$

where $\tilde{I}(u,v)$ is related to the reconstructed image through Equation 3. Introducing Lagrange multipliers $\lambda(u,v)$ for the data constraints (31), we then have to maximize a linear combination of the entropy (Equation 7) and the constraints, i.e.

$$S' = \iint f(I) \, dx \, dy + \sum_{u,v \in K} \lambda(u,v) \left\{ \iint I \exp\left[-2\pi i(ux+vy)\right] \, dx \, dy \right.$$

$$\left. - \tilde{D}(u,v) \right\}. \qquad\qquad 17.$$

If we differentiate with respect to $I(x,y)$, this gives

$$f'[I(x,y)] = - \sum_{u,v \in K} \lambda(u,v) \exp\left[-2\pi i(ux+vy)\right] \equiv J(x,y), \qquad 18.$$

where f' is the derivative of f. (The symbol $\equiv$ denotes a definition.) When this solution is substituted in the constraints (Equation 16), we obtain a set of nonlinear equations for the $\lambda(u,v)$. For our purposes, the crucial observation is that the function $J(x,y)$ in Equation 18 is "band limited," i.e. it has only a finite set of Fourier frequencies [in this case the set $(u,v) \in K$ corresponding to the measurements]. The ME image I can now be found by formally inverting the function f' in Equation 18, i.e.

$$I(x,y) = f'^{-1}[J(x,y)] \equiv g[J(x,y)]. \qquad\qquad 19.$$

© Annual Reviews Inc. • Provided by the NASA Astrophysics Data System

For the two standard forms of entropy ($f_1 = \ln I$, $f_2 = -I \ln I$), we have

$$g_1(J) = 1/J, \qquad g_2(J) = \exp(-1-J).$$

Thus, *the image is a nonlinear function* (*reciprocal or exponential*) *of the band-limited function* $J(x, y)$. (The exponential function is responsible for the factorization property of ME images constructed with the $-I \ln I$ entropy mentioned in Section 2.4 and Figure 2.)

The transformation between the image $I(x, y)$ and the associated band-limited function $J(x, y)$ is illustrated in Figure 4, which can be used (88) to bring out the following properties of ME images:

1. Since there are no Fourier components in $J(x, y)$ corresponding to the unmeasured spatial frequencies ($u, v \notin K$), these can be generated in the restored image only by the *nonlinearity* of $g(J)$, or equivalently of $f'(I)$. Thus, for Fourier extrapolation and interpolation, the slope $f''(I)$ of the "transfer function" in Figure 4 should vary significantly over the range of values of I in the restoration.

2. If the data correspond to a map $I(x, y)$ with a large constant baseline, we have $I_{max} - I_{min} \ll I_{min}$. We then operate on a small linear section of the curve $I(J)$, and there will be no generation of unmeasured spatial frequencies (9), i.e. we get back to the principal solution. As a corollary, if we add a constant C to the map before processing [equivalent to increasing $\tilde{I}(0, 0)$ by C or to doing ME with $f(I+C)$] and subtract it out at the end, the resulting image will depend strongly on C.

3. The two functions f_1 and f_2 of Equation 8 have the common features

$$f''(I) < 0, \qquad f'''(I) > 0. \qquad\qquad 20.$$

Therefore, the slope of the curve $f'(I)$ in Figure 4 has a small magnitude at large I. The peaks in the image are thus sharper than those in a band-
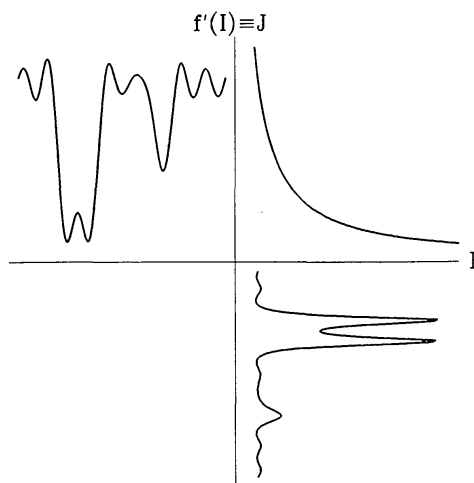

Figure 4 The relationship (Equations 18 and 19) between a one-dimensional $I(x)$ and $J(x)$ for the entropy form $f(I) = \ln I$. The band-limited "input signal" $J(x)$ at top left (x increasing to the right) is modified by the "transfer function" $I = g(J)$ at top right to produce the maximum entropy image $I(x)$ at bottom right (x increasing downward). Note the flattening of the baseline and the sharpening of peaks in I. These effects would occur for all choices of $f(I)$ that satisfy Equation 20, including the forms $-I \ln I$ and $I^{1/2}$ discussed in the text.

limited function, i.e. we have "superresolution." The reason for this is, of course, the Fourier extrapolation caused by the nonlinearity. Note that the sign of $f(I)$ or $f'(I)$ is immaterial for this argument.

4. Since the slope has a large magnitude at lower values of I, the oscillations present in the baseline of I are weaker than those in a band-limited function, i.e. we have ripple suppression. This implies Fourier extrapolation and, more importantly, interpolation when there are gaps in the data at low spatial frequencies.

These properties of the MEM are illustrated in Figure 5, which shows a model source that is complex enough to resemble real cases. A comparison of region A in the true map, the principal solution, and the ME reconstructions brings out the basic peak-sharpening tendency of the MEM. Similarly, region B illustrates the baseline-flattening effect. These two fundamental properties of the MEM were clearly stated in Frieden's 1972 paper (40).

5. The two entropy functions (Equation 8) give similar results when the degree of nonlinearity, measured by the parameter

$$R = f''(I_{min})/f''(I_{max}),\qquad\qquad 21.$$

has similar values (88). In fact, one could consider other functions sharing the same general properties as those given in Equation 20. In Table 1, the three functions $\ln I$, $I^{1/2}$, and $-I\ln I$ are compared, and it is clear that $I^{1/2}$ represents an intermediate case between the usual $\ln I$ and $-I\ln I$ entropies. This is also seen directly in Figure 5 from the reconstructed images produced by the three forms of $f(I)$. The success of $I^{1/2}$, which has no information-theoretic backing (no logarithms!), is a strong point in favor of the penalty function interpretation mentioned in Section 2.5. This similarity of reconstructions made with different choices of $f(I)$ was first noted by Wernecke (121; see also 55, 88).

6. Around a maximum or a minimum, the band-limited function $J(x,y)$ can be Taylor expanded up to quadratic terms. For the f_1 and f_2 cases, the form of the restoration $I(x,y)$ near such a point (x_0, y_0) would then be

$$I_1(x,y) = 1/[a(x-x_0)^2+2b(x-x_0)(y-y_0)+c(y-y_0)^2+d],$$

$$\qquad\qquad 22.$$

$$I_2(x,y) = \exp[-a(x-x_0)^2-2b(x-x_0)(y-y_0)-c(y-y_0)^2-(d+1)],$$

respectively. We thus have anisotropic Lorentzian (f_1) and Gaussian (f_2) peaks (88). This argument is applicable to sharp peaks, where the function $I(x,y)$ has fallen appreciably while the change in $J(x,y)$ is still small enough to be described by the quadratic terms.

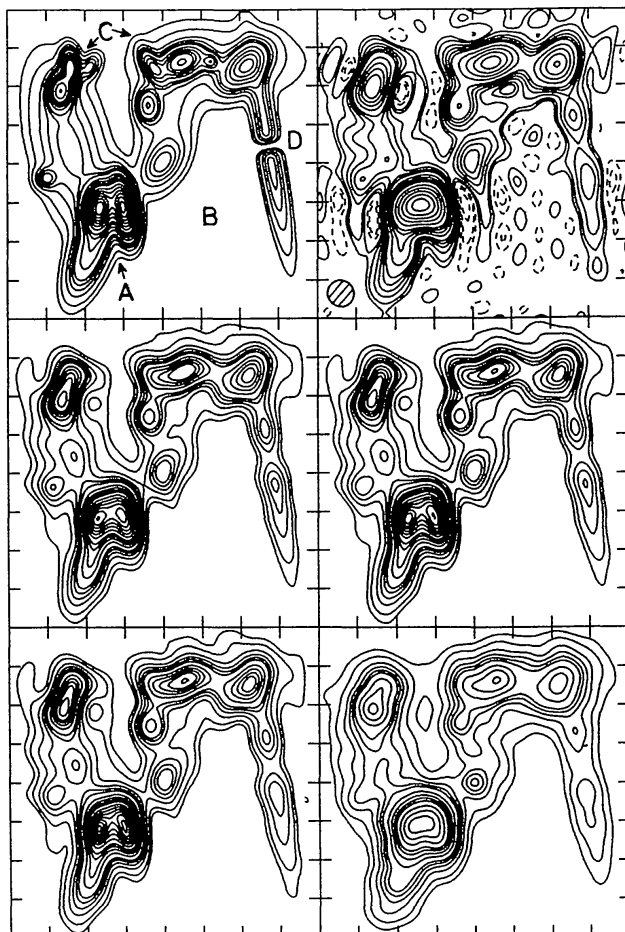7. The resolution is greatest for the highest peaks in the map and

Figure 5 (*Top left*) A model source. (*Top right*) The principal solution (dirty map) when the coverage is limited to $(u^2+v^2)^{1/2} < 8.5$. The size of the psf at half-maximum is shown by the hatched circle. (*Center*) ME restorations with $-I \ln I$ (*left*) and $I^{1/2}$ (*right*). (*Bottom left*) The ln I restoration. (*Bottom right*) A typical restoration with the Clean algorithm. The contour levels in the figures are as follows: solid contours at 1, 2, 4, 6, 8, 10, 15, 20, 25, 30, 40, 50, 60, . . . , dashed contours at -1, -2, -4, . . . , in units where the main peaks in the model have height 100. The ME restorations, which all correspond to $R = 100$ (see Equation 21), resolve the two peaks in region A of the model, remove the ripple in the baseline (region B), have reduced resolution for the lower peaks in region C, and create spurious peaks near the absorption feature in region D. Note the remarkable similarity among the three ME restorations.

Table 1 Properties of three choices of entropy function

$f(I)$	$f'(I) \equiv J$	$f'^{-1}(J) \equiv g(J)$	$f''(I)$	$f'''(I)$
$\ln I$	$1/I$	$1/J$	$-1/I^2$	$2/I^3$
$I^{1/2}$	$1/2I^{1/2}$	$1/4J^2$	$-1/4I^{3/2}$	$3/8I^{5/2}$
$-I \ln I$	$-1-\ln I$	$\exp(-1-J)$	$-1/I$	$1/I^2$

progressively decreases for lower ones (32, 76, 88), as is seen by comparing regions A and C of Figure 5. Because of this variable resolution, it is dangerous to attach too much significance to the heights of restored peaks. Integrated fluxes are likely to be more reliable.

8. Ripple suppression on a plateau is less effective than on the baseline (13, 76, 88). Thus, the sidelobes of an isolated source are better suppressed than those of a point source embedded in an extended region of emission.

9. Points (3) and (4) above show that forms of entropy that satisfy Equation 20 implicitly assume that the source of interest occurs with high values of I and the background with low values of I. If we consider the opposite case, where a smooth background continuum is *absorbed* by compact foreground sources, all such forms of the MEM will give poor results (85). This is illustrated in region D of Figure 5, where spurious peaks appear on either side of the absorption feature.

10. The condition $f''(I) < 0$ in Equation 20 ensures that the entropy is a convex function of the intensity. It can then be shown (21, 121) that the maximum of S' in Equation 17 is unique.

One can regard the functional form (Equation 18 or 19) of the MEM image as a model that the method fits to the measurements. The various items listed above are then properties of this model, which may or may not be desirable for a specific problem.

3.1.2 INTERFEROMETRY WITH NOISY DATA   Ables (1) suggested that for noisy data, Equation 16 should be replaced by a single constraint that sets the χ^2 term equal to its expected value Ω, so that the predictions of the model deviate from the data by the expected noise. We thus have the constraint

$$\chi^2 \equiv \sum_{u,v \in K} [|\tilde{I}(u,v) - \tilde{D}(u,v)|^2 / \tilde{\sigma}^2(u,v)] = \Omega \simeq m, \qquad 23.$$

where $\tilde{\sigma}^2(u,v)$ is the variance of the measurement $\tilde{D}(u,v)$. The last relation says that Ω (the expected value of χ^2) is equal to m (the number of independent data) when m is large. Using the method of Lagrange multipliers once more, we now have to maximize

$$S' = \iint f(I) \, dx \, dy - \lambda(\chi^2 - \Omega), \qquad 24.$$

which, for $f(I) = -I \ln I$, appears very similar to the quantity that had to be maximized in the Bayesian framework (Equation 14 with S given by Equation 15). However, there is a crucial difference, namely that the relative weight of the two terms in the Bayesian case is not an adjustable

parameter but instead depends on N (the number of luminance elements or building blocks involved). There is thus no guarantee that the residuals will have the right magnitude. O'Sullivan & Komesaroff (91) have pointed out that when N becomes large, the a priori distribution $P(I_i)$ of Equation 13 will peak so sharply at the flat map [$I(x,y) = $ constant] that the data will be almost entirely disregarded and the residuals will be enormous. While there have been various suggestions about a choice of N in the Bayesian framework (68), it appears that this paradox is still unsolved in principle. In practice, users of the MEM follow Ables and adjust the Lagrange multiplier to obtain the expected level of residuals. For brevity, we refer to this procedure as least-squares MEM.

Returning then to our aperture synthesis example with the inclusion of noise, we now have to maximize Equation 24. Substituting Equation 3 and differentiating with respect to an unmeasured Fourier coefficient of the image [i.e. $\tilde{I}(u,v)$, $(u,v) \notin K$], we find that

$$\iint f'(I) \exp\left[2\pi i(ux+vy)\right] \, dx \; dy \equiv \tilde{J}(-u,-v) = 0, \qquad (u,v) \notin K. \qquad 25.$$

[The measured set K is symmetric (i.e. if u,v are in K, then $-u,-v$ are also in K).] The above equation states that the Fourier coefficients $\tilde{J}(u,v)$ of $J(x,y)$ vanish except for those spatial frequencies $(u,v) \in K$ that are measured.

11. Thus, $J(x,y)$ is a band-limited function as before (87), and all the consequences of this statement discussed in Section 3.1.1 hold.

The new feature in least-squares ME is that over the measured spacings, the model predictions $\tilde{I}(u,v)$ differ from the data $\tilde{D}(u,v)$. Differentiating Equation 24 with respect to $\tilde{I}^*(u,v) = \tilde{I}(-u,-v)$, $(u,v) \in K$, we find that

$$\tilde{J}(u,v) - \lambda[\tilde{I}(u,v) - \tilde{D}(u,v)]/\tilde{\sigma}^2(u,v) = 0, \qquad (u,v) \in K. \qquad 26.$$

12. *The residuals $\tilde{I} - \tilde{D}$ between the model $\tilde{I}$ and the measurements $\tilde{D}$ are thus far from random* (18, 88). In the simple case when $\tilde{\sigma}(u,v)$ is independent of (u,v), the residuals are proportional to the Fourier coefficients of the band-limited function $J(x,y) \equiv f'[I(x,y)]$. Equivalently, in the map plane these residuals are given by $J(x,y)$. In actual practice, since the total flux in the map is usually independently constrained [see (14) below], the residuals are given by $J(x,y) - \bar{J}$, where $\bar{J}$ is the average value of J. Figure 4 thus shows that the residuals will be highly correlated with the map, being systematically negative at the peaks. In the baseline region, the residuals have a positive average and tend to smooth out fluctuations, since excursions in I and J have opposite signs.

13. Since the residuals are negative at the peaks, they transfer flux from the peaks to the background. The effect is proportional to $J(x, y)$ locally. Since peaks in I with a wide range of heights correspond to a much smaller range in J (e.g. see Figure 4), the amount of flux transferred is not very sensitive to peak height. Thus, the fractional flattening is more pronounced for the smaller peaks (32, 91).

The systematic behavior of the residuals can be called bias in the same sense that the term is used in statistics. A point to be emphasized is that one is not solving for the noise by doing least-squares MEM. In fact, it can be shown (88) that for a high signal-to-noise ratio, the ME model predictions for the visibilities $\tilde{I}$ are farther from their true values than were the original noisy data by a factor $\sim \sqrt{2}$. (The case of low signal to noise is discussed in greater detail in Section 3.1.3.) This is in contrast to usual applications of the method of least squares, where one determines a small number of parameters from a large number of measurements and the residuals really do represent the noise.

14. The total flux $\tilde{I}(0,0) = \iint I(x, y) \; dx \; dy$ has a special role. If it is included in the χ^2 term of Equation 24, then Equation 26 implies that we have a residual $\tilde{I}(0,0) - \tilde{D}(0,0)$ that is proportional to $\tilde{J}(0,0) = \bar{J}$, the average of the band-limited function $J(x, y)$. Most users of the MEM would hesitate to modify the measured value of the flux in a systematic manner, and it is usual to add a separate equality constraint for the total flux (see Equation 30). This will ensure that the average of the residuals over the map, $\tilde{I}(0,0) - \tilde{D}(0,0)$, vanishes.

The systematic behavior of the residuals in a least-squares ME image becomes clear in a geometric picture [after Bryan & Skilling (18)] illustrated in Figure 6 for the case of constant $\tilde{\sigma}(u, v) \equiv \sigma$. In the space of all maps
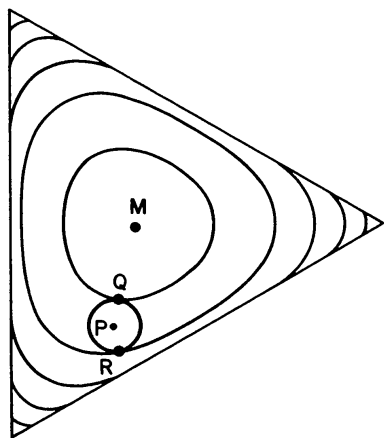

Figure 6 A geometric picture of ME restoration from noisy data [after Bryan & Skilling (18)]. The perpendicular distances of a point inside the triangle from the three sides of the triangle represent the intensity values in a three-pixel map with a given total flux. Points outside the triangle violate the positivity constraint. The circle represents maps that differ from the measurements (P) by the expected noise. Here Q and R are both stationary points of the entropy subject to this constraint, but only Q is a maximum. In certain circumstances, there may even be multiple maxima.

with a given total flux, the entropy is highest for the flat map ($I = $ constant) represented by the point M. This is enclosed by convex contours of progressively lower entropy out to the boundary, beyond which positivity is violated. Our noisy data are represented by the point P, and all maps satisfying the χ^2 constraint lie on a generalized sphere centered on P of radius $\sigma\sqrt{m}$, where m is the number of independent data. When we allow for the unmeasured data, this surface is like a generalized cylinder. The least-squares ME map consistent with the data to within the errors is the point Q where this cylinder touches a constant entropy surface. We see that the vector PQ, which is the residual, points along the direction of increasing entropy. This geometric statement is equivalent to Equation 26 and makes it clear that the residuals conspire to take one toward the totally flat map M.

3.1.3   THE SINGLE-APERTURE CASE

Consider, finally, the single-aperture case discussed at the beginning of Section 1.1 (Equations 1 and 4). Here, data are in principle available at all spatial frequencies less than the diffraction limit, but the signal-to-noise ratio reduces catastrophically beyond the seeing cutoff, which usually occurs at much smaller spatial frequencies. We now have to maximize

$$S' = \iint f(I)\, dx\, dy - \lambda\left[\sum_{u,v} |\tilde{P}(u,v)\tilde{I}(u,v) - \tilde{D}(u,v)|^2/\tilde{\sigma}^2(u,v) - \Omega\right].$$

(Equation 24 is clearly a special case of this, corresponding to the psf of Equation 5.) As before, by varying with respect to $\tilde{I}(-u, -v)$, we obtain

$$\tilde{J}(u,v) - \lambda\tilde{P}(u,v)\,[\tilde{P}(u,v)\tilde{I}(u,v) - \tilde{D}(uv)]/\tilde{\sigma}^2(u,v) = 0, \qquad\qquad 27.$$

where $\tilde{J}$ is again the Fourier transform of $J \equiv f'(I)$. [We have assumed a symmetric psf, so that $\tilde{p}(-u, -v) = \tilde{p}(u,v)$.] We expect the residuals to be of order $\tilde{\sigma}$ and hence find that

$$|\tilde{J}(u,v)| \sim \lambda\tilde{P}(u,v)/\tilde{\sigma}(u,v),$$

i.e. the spatial frequencies of $J(x, y)$ corresponding to data with poor signal-to-noise ratios (low values of $\tilde{P}/\tilde{\sigma}$) are attenuated.

15. While $J(x, y)$ is thus no longer strictly band limited as in Sections 3.1.1 and 3.1.2, it still is a tapered function, with a resolution approximately equal to that of the original psf. As before, any increase in the resolution of the final map can arise only because of the nonlinearity of f' or g, and so the properties discussed earlier continue to be valid. (Ripple suppression is, however, no longer an issue, because there is no sharp boundary between measured and unmeasured frequencies.) It is satisfying that the situation

150   NARAYAN & NITYANANDA

with very noisy data ($\tilde{P}/\tilde{\sigma}$ small) goes over into the case of missing data ($\tilde{J} = 0$; see Section 3.1.2) as the signal-to-noise ratio goes to zero.

The systematic nature of the residuals in Equation 27 is again obvious because $\tilde{P}\tilde{I} - \tilde{D}$ is seen to be proportional to $\tilde{J}$, which is intimately related to the restored image. The nature of the bias in this case is best understood by going to a limit in which the image has weak structure on a strong, flat background. As discussed in point (2) of Section 3.1.1, the transformation $J = f'(I)$ then becomes linear, i.e. we have $J = \alpha - \beta I$ or $\tilde{J}(u,v) = -\beta \tilde{I}(u,v)$, $(u,v) \neq (0,0)$. (The sign of βI is appropriate for $f''(I) < 0$.) Substituting this in Equation 27, we find that

$$\tilde{I} = \frac{1}{(1 + \beta \tilde{\sigma}^2/\lambda \tilde{P}^2)} \frac{\tilde{D}}{\tilde{P}} \sim \frac{1}{[1 + (\tilde{P}_0/\tilde{P})^2/s_{max}]} \frac{\tilde{D}}{\tilde{P}}.$$

The second relation above is approximate. It is obtained by estimating λ from the χ^2 constraint on the assumption of unresolved sources and constant $\tilde{\sigma}(u,v)$. Here $\tilde{P}/\tilde{P}_0$ is the normalized psf and the signal-to-noise ratio $\tilde{P}\tilde{I}/\tilde{\sigma}$ has been assumed to attain its maximum value s_{max} near $u = v = 0$. As $\tilde{P}/\tilde{P}_0 \to 0$ (poor signal to noise), this formula gives $\tilde{I} \to 0$. Thus, the linear limit of least-squares MEM discards poor data, a very sensible procedure and one that is closely related to many other linear image-processing algorithms [e.g. the Wiener filter (6)]. However, when the signal-to-noise ratio is good ($\tilde{P}/\tilde{P}_0 \sim 1$), $\tilde{I}$ takes on a value $\tilde{D}/\tilde{P}$ (the noise-free solution) *multiplied by a factor that is always less than* 1. This is the bias we discussed earlier, seen now in a linear situation. It means that the least-squares MEM puts less structure in the reconstruction than the data warrant. This errs on the conservative side, but it is still a systematic error that we feel should be avoidable.

16. In the full nonlinear MEM with the χ^2 constraint, the data with poor signal-to-noise ratios are not set to zero as in the linear case discussed above; instead, they are essentially superseded by an extrapolation from better data. This is a desirable feature of the method. The systematic modification of good data, however, continues to be an undesirable side effect.

3.2   Controlling the MEM

The purely behavioristic analysis of the MEM given in the previous sections has brought out its strong points, namely its ability to flatten the baseline, to resolve peaks, and to supersede excessively noisy data. Some undesirable features (such as resolution increasing with peak height, inability to clean up ripple in regions of extended emission, and biased residuals) have also emerged. Practical users of the MEM are becoming

aware of its weaknesses, and various suggestions to control them have been made. These are mostly ad hoc schemes that go beyond the ME principle but often give excellent results.

Bhandari (9) suggested that the sensitivity of the ME reconstruction to an additive constant could be exploited by maximizing $f(I+C)$ with a suitable constant C. This has been used in practical implementations (55, 88). One scheme (88) is to adjust C until the nonlinearity parameter R defined in Equation 21 attains a suitable value, say 100 (see Figure 5), so that one obtains ripple suppression and useful but not excessive resolution enhancement. This approach ensures that the restoration is independent of the dc level in the map, a desirable feature (2). A more general scheme due to Frieden & Wells (44) is to subtract out a slowly varying background, which will eliminate the dependence of ripple suppression on background level (Section 3.1.1). These authors give examples to demonstrate that peak sharpening too depends on the background level, as we would expect.

We saw earlier that the bias in least-squares MEM has the effect of transferring flux from the source to the background, and therefore it behaves somewhat like the additive constant C discussed above. It has been found (13) that a realistic value for the noise can sometimes lead to excessive peak sharpening, as one expects for too low a baseline, or equivalently for too high a value of the parameter R. In such cases, one could deliberately use a larger value of σ in order to achieve the desired level of resolution [see the examples in (87)]. However, it appears more direct to simply add a constant as described above.

It is frequently the practice to specify a window around sources of interest in the field and to constrain the intensity to be zero outside this window. An important consequence is that sidelobes outside the window are completely suppressed, which helps the algorithm to concentrate on restoring the source of interest. Another side effect is that the flux transferred from the source to the background by the least-squares method has a greater effect on the baseline because it is concentrated in a smaller area. Thus, the noise parameter selected will have a greater influence on the resolution enhancement in the final map.

The presence of spurious features in MEM maps of extended regions of brightness (13, 76, 88) can be handled (51) by maximizing $-I \ln (I/I_0)$ (see Equation 10) with a "default map" $I_0(x, y)$ that includes known extended structure. The image then takes the form

$$I(x, y) = I_0(x, y) \exp\left[-1 - J(x, y)\right]. \tag{28}$$

It has been suggested (13, 48) that the map obtained with $I_0 = $ constant could itself serve as the default for further processing by the MEM. However, notice that both factors in Equation 28 would be exponentials

of band-limited functions, and the resulting map would be no different from the one obtained with constant default when taken to full convergence. (When the χ^2 constraint is included, the residuals will have some dependence on $I_0(x, y)$, but this is a weak effect.) This objection does not hold if the constant default map is smoothed in some way before being used as I_0 in the next iteration. The qualitative effect of the default approach is that the second factor in Equation 28 is only asked to represent peaks on a flat baseline and thus will not have excessive ripple in regions of extended emission. However, since this factor is finally multiplied by $I_0(x, y)$, it is not clear that a great deal is gained. The approach of Frieden & Wells (44) of *subtracting* the background appears to be superior to the default idea of *dividing* by it.

Horne (59) has successfully used the default idea in mapping accretion disks from eclipse data. Artifacts that appeared in the constant default reconstruction were suppressed by using a circularly symmetrized version as the new default. Another use of the default approach is described in Section 5.2.

A possible way of handling ripple on regions of extended emission is to first use Clean (Section 6.1) to remove obvious point sources in the principal solution (T. J. Cornwell, private communication). The remaining map could then be processed by the MEM. This is another ad hoc prescription like many of the others discussed in this section, but it has the virtue of recognizing the strengths and weaknesses of the two methods and making the best of both.

The variation of resolution with peak height is intrinsic to the method, and it would appear that nothing much can be done within a pure ME framework. Braun & Strom (13) have suggested that the final MEM restoration could be convolved with a function of similar resolution to the psf and the residuals added back, just as in Clean (see Section 6.1 for details on Clean). Thus all peaks in the final restoration will have approximately the same resolution. Also, gradients will be better represented and the noise will be more uniform. However, the map will no longer fit the data, and the discrepancy will systematically increase at larger spatial frequencies. A related scheme is to apply the convolution to the principal solution before deconvolving with the MEM (or, equivalently, to suitably taper the data). This appears to converge faster and to give better results (T. J. Cornwell, private communication).

A scheme for imposing a Gaussian distribution of residuals as an added constraint has been used (9) as a cure for bias. However, the spatial distribution still has some nonrandom features, and the method has not been used very much since. From the discussion of the linear limit of least-squares MEM in Section 3.1.3, we see that what is needed is a prescription

whereby poor data are handled by least squares, but good data are fitted exactly. It is relatively straightforward to develop such an approach for the linear case. Application to the full nonlinear algorithm is a challenging problem.

3.3   Comparison of Different Entropy Forms

It is clear from the discussion in Sections 3.1 and 3.2 that one can regard the MEM as a model that fits a function of a specified form (e.g. the exponential of a band-limited function) to the measurements as modified by the least-squares residuals. Such a view does no justice to the origins and motivations of the method discussed in Section 2 but instead focuses on its behavior in practical situations. Within this framework, one could ask whether, purely as a model, one form of entropy is better than another. We saw in Equation 22 that $-I \ln I$ has the desirable property of producing Gaussian peaks. We now show that there are other reasons why this form of entropy may be considered superior to other contenders.

We first give a semiquantitative estimate of the resolution enhancement in ME images. Consider the restoration of a two-dimensional image using the $-I \ln I$ form of entropy. Let a peak have intensity I_p at its maximum, and let I_b represent the average intensity in the baseline. Let us define $r \equiv I_p/I_b$, a sort of "dynamic range" appropriate to the peak. We have seen in Section 3.1 that the form of the image is given by $I = \exp(-1-J)$, with J having a resolution appropriate to the psf. For the peak of interest, J varies over a range $\sim \ln r$, and so the coefficients a, b, c in Equation 22 have typical values $\sim \ln r/w_0^2$, where w_0 is the width of the psf. The width of the peak in the restored image is then $w(r) \sim w_0/(\ln r)^{1/2}$, i.e. the degree of sharpening is $s \equiv w_0/w(r) \sim (\ln r)^{1/2}$. Thus the resolution increases with r, as shown earlier, but only weakly as the square root of a logarithm. [We further see that a peak of given height would be more sharpened on a low baseline than on a plateau, as was also noted in (44).] A similar analysis shows that $s \sim r^{1/4}$ for $f(I) = I^{1/2}$ and $s \sim r^{1/2}$ for $f(I) = \ln I$. These show a greater sensitivity of the resolution to peak height, and thus the $-I \ln I$ form of entropy should be preferred on this account.

We now ask whether the model that is fitted by the MEM gives a solution for all physically realizable data. Since we are dealing with a set of nonlinear equations, this question is not entirely straightforward. Let us consider the maximization of the entire family (88) of "entropy" functions

$$f(I) = I^{1-\varepsilon}/\varepsilon(1-\varepsilon). \qquad 29.$$

This includes $\ln I$ for $\varepsilon \to 1$ and $-I \ln I$ for $\varepsilon \to 0$ (apart from constants and terms linear in I, which are irrelevant and are assumed to have been eliminated). This gives a unified discussion of these two standard forms,

and it also includes cases like $I^{1/2}$, which would be allowed from the penalty function approach of Section 2.5. (One could call this a maximally noncommittal approach to the MEM!)

The convexity property $f''(I) < 0$ only guarantees the uniqueness of a maximum if one exists. However, there can be cases where there is no maximum within the domain of nonnegative images. As an illustration, consider $\varepsilon = -1$ in Equation 29, which gives the function $f(I) = -I^2/2$. In this case, $I = g(J) = -J$, and so I is itself a band-limited function, and there is no nonlinearity or extrapolation. The (unique) "MEM" solution with this function is directly the principal solution, which can go negative. The same holds for all functions of the above family with $\varepsilon < 0$, because none of these has the infinite slope at $I = 0$ that is necessary for imposing positivity (57). The function $-I \ln I$, corresponding to $\varepsilon = 0$, has a derivative going to infinity as $\ln I$, and therefore it does not suffer from the above problem. Functions for $\varepsilon > 0$ have more strongly divergent derivatives, but they suffer from a different shortcoming because they enforce positivity too strongly. This has been illustrated (88, Appendix B) for $\varepsilon = 1$, corresponding to $f_1 = \ln I$, for the case of three-dimensional imaging, as could occur in spectral line mapping (Section 5.2). Let the data correspond to a peak on a baseline. As the strength of the baseline decreases relative to the peak, the reconstructured peak narrows and develops a δ function even when the baseline is nonzero. This is a case of infinite superresolution with perfectly reasonable data. The effect, which is mathematically similar to the appearance of a δ function in the momentum distribution of a Bose gas (79), depends on dimension d and occurs for all $d > 2/\varepsilon$. Thus, it occurs at a finite dimension for every member of the family in Equation 29 with $\varepsilon > 0$. The $-I \ln I$ entropy, corresponding to $\varepsilon = 0$, is free from this problem and is once again seen to be superior to the other forms.

4. ALGORITHMS

Two factors simplify the solution of the system of nonlinear equations that express the ME image in terms of the data. Firstly, the solution is unique (20, 88) for a given entropy form $f(I)$ [satisfying $f''(I) < 0$] whenever the data are fitted exactly. (The situation with the χ^2 constraint is shown in Figure 6.) Consequently, algorithms are likely to be robust. Secondly, the measured data considered so far are linear combinations of the unknown map intensities, with a Fourier relation connecting the two spaces. Therefore, the Fast Fourier Transform (15, 25) helps to reduce the computation time considerably.

Two types of algorithm are discussed in Sections 4.1 and 4.2. One technique is to iterate using an implicit relation connecting the unknown

map intensities. Such "fixed-point" methods are often extremely fast but can be unstable. The other method is to directly maximize the chosen entropy (with the χ^2 term if least-squares MEM is being implemented) using one of the variants of the gradient method.

Clever algorithms have been developed for ME spectral analysis (20, 52), where given the first n autocorrelations of a time series, one is interested in estimating the underlying spectrum. Attempts have been made (86) to generalize these algorithms to two dimensions, but these have run into difficulties (88, Appendix C). In any case, such approaches are not always suitable for image restoration, since they require the data to be available on a rectangular array in Fourier space.

4.1   *Fixed-Point Methods*

As we have seen in Sections 3.1.1 and 3.1.2, the ME image from interferometric data has the property that the auxiliary function $J(x, y)$, which is related to the map $I(x, y)$ by $J = f'(I)$, is "band limited." This leads to the following fixed-point scheme for noise-free data (80, 88) (the case discussed in Section 3.1.1):

Step 0   Start with an initial map that is constant everywhere.

Step 1   Calculate a misfit factor ε between the Fourier coefficients $\tilde{I}(u, v)$ of the current map and the data $\tilde{D}(u, v)$:

$$\varepsilon = \sum_{u,v \in K} |\tilde{I}(u,v) - \tilde{D}(u,v)|^2 \Big/ \sum_{u,v \in K} |\tilde{D}(u,v)|^2.$$

If ε is small enough (say $< 10^{-8}$), stop the refinement. If not, go to step 2.

Step 2   Move the current $\tilde{I}(u, v)$ a fractional distance $1 - \alpha$ toward $\tilde{D}(u, v)$:

$$\tilde{I}_{\text{new}}(u, v) = (1 - \alpha)\tilde{D}(u, v) + \alpha\tilde{I}_{\text{old}}(u, v), \qquad (u, v) \in K.$$

Unmeasured coefficients $\tilde{I}(u, v)$, $(u, v) \notin K$, are left alone.

Step 3   Compute the new map $I(x, y)$ and calculate $J = f'(I)$ and its Fourier coefficients $\tilde{J}(u, v)$.

Step 4   For all $(u, v) \notin K$, move $\tilde{J}$ a fractional distance $1 - \beta$ toward the expected value of 0, i.e.

$$\tilde{J}_{\text{new}}(u, v) = \beta\tilde{J}_{\text{old}}(u, v), \qquad (u, v) \notin K.$$

Step 5   Compute the new $J(x, y)$ and calculate $I = g(J)$ and its Fourier coefficients $\tilde{I}(u, v)$. Go to step 1.

The key to a successful implementation of this algorithm is the choice

of the "damping factors" α, β. The straightforward choice $\alpha = \beta = 0$ works well when the map is not strongly constrained by positivity, i.e. when the nonlinearity parameter R (Equation 21) is small. However, in realistic cases, when the map has large empty regions, the scheme could diverge even with values of $\alpha, \beta \sim 1$. When it works, the algorithm usually converges rapidly.

A fixed point-scheme has also been developed for least-squares MEM. Using the notation of Equation 6, we need to maximize

$$S' = \sum_i f(I_i) - \mu \left[\sum_i I_i - D_0 \right] - \lambda \left\{ \sum_r \left[\left(\sum_i P_{ri} I_i - D_r \right)^2 / \sigma_r^2 \right] - \Omega \right\}, \qquad 30.$$

where a separate constraint has been introduced for the total flux, as is the usual practice (see Section 3.1.2, point 14). The last term corresponds to the χ^2 constraint. The condition $\partial S'/\partial I_i = 0$ gives an implicit relation for the I_i. This leads to an iterative algorithm (49, 124), where the map estimate $I_i^{(n+1)}$ at the $(n+1)$th iteration is given by

$$I_i^{(n+1)} = (1-\gamma)I_i^{(n)} + \gamma g \left[\mu + 2\lambda \sum_r (P_{ri}^2 I_i^{(n)} - P_{ri} D_r)/\sigma_r^2 \right].$$

As before, $g \equiv f'^{-1}$. The Lagrange parameters μ and λ need to be adjusted during the course of the computations to satisfy the corresponding constraints. Once again the choice of the damping factor $1 - \gamma$ is critical, and convergence is not assured.

Because they lack robustness, fixed-point schemes have been limited to special applications. Present-day general-purpose routines are mostly based on gradient techniques.

4.2   Gradient Methods

To maximize a nonlinear function of several variables, a standard technique is to retain up to quadratic terms in the Taylor expansion and to solve the resulting set of linear equations for the location of the maximum. This is equivalent to solving the nonlinear equation defining the maximum (e.g. Equations 25 or 27) by the Newton-Raphson method (31). Thus, to maximize S' in Equation 30, one computes both its gradient $\nabla S'$ ($\equiv \partial S'/\partial I_i$) with respect to the intensities I_i at the n pixels in the map as well as the so-called Hessian matrix made up of second derivatives, $\nabla\nabla S'$ ($\equiv \partial^2 S'/\partial I_i \partial I_j$). The shifts in the intensities are then given by

$$\Delta I_i = -\sum_j (\nabla\nabla S')_{ij}^{-1} (\nabla S')_j. \qquad 31.$$

The formal geometric statement of this equation would be that the

contravariant shift vector ΔI is related to the covariant gradient vector $\nabla S'$ by the "metric" $\nabla\nabla S'$.

The above strategy cannot be directly implemented, since the Hessian has n^2 elements, and with present-day maps (where n could be a million or more) the storage and inversion of such a matrix would be prohibitively expensive [see the algorithm in (44), however]. The conjugate gradient methods (31, 37) are, however, quite convenient and were applied to the MEM by Wernecke & D'Addario (122). Here, information about the Hessian is progressively built up as the calculation proceeds, and in each iteration the direction of the shifts in n-dimensional I space is taken to be a suitable linear combination of the current and previous gradients. It can be shown that the maximum of a quadratic function is reached in n iterations. However, most of the gain appears to occur in the first few iterations. Also, when one is well away from the quadratic region near the maximum of the function, the technique becomes ineffective after a few iterations, since the partial information about second derivatives is obtained at different points and thus quickly gets out of date. It is therefore more efficient to restart the conjugate gradient algorithm every few cycles. In practice it has been found that over a wide range of n, about 50 iterations in all are adequate to carry out the full nonlinear maximization.

A change of variables can significantly speed up the course of a gradient search by making the contours of the function being maximized more nearly spherical. For example, a change to $\ln I$ as the variables was suggested by Wernecke & D'Addario (122). The corresponding search directions in a simple steepest-ascent algorithm would be $I_i \, \partial S'/\partial I_i$. This could be regarded as a metric I_i^{-1} in Equation 31 and has the advantage that pixels with intensity close to zero do not experience large shifts.

A robust algorithm, involving use of the I_i^{-1} metric and searching in a three-dimensional subspace, has been described by Skilling & Bryan (107; see also 19) and has been widely used. Here, we illustrate the variable metric idea using the simplified approach of Cornwell & Evans (28). Equation 30 shows that

$$(\nabla\nabla S')_{ij} = f''(I_i)\delta_{ij} - 2\lambda \sum_r P_{ri}P_{rj}/\sigma_r^2.$$

Since the entropy enters only in the diagonal elements, one could neglect the off-diagonal terms that come from the psf and write

$$(\nabla\nabla S')_{ij}^{-1} \sim \delta_{ij}[f''(I_i) - \alpha\lambda]^{-1}. \qquad 32.$$

The parameter α has to be suitably selected to include the full effect of the central peak in the psf. Substituting Equation 32 into Equation 31 and comparing with Table 1, we see that the magnitude of the shifts is reduced

in those regions of the map where $I(x, y)$ is small, which thus prevents the algorithm from being unduly preoccupied with near-zero regions of the map.

Gradient schemes are also possible in data space (88). This is a useful strategy when one wishes to fit the data exactly. One then varies only the unmeasured visibilities, and so the algorithm operates throughout in the space of images that are consistent with the data. The generalization to include the χ^2 constraint is straightforward (88). However, in this approach one loses the flexibility of being able to impose windows in the restored map, which is easy with gradient schemes in I space.

The schemes described so far deal directly with the map intensities or the visibilities and apparently have to solve for a large number n of parameters (where n is the number of pixels). However, the true number of unknowns is actually the number of independent data points m, since this is the number of parameters in the band-limited function $J(x, y)$ that completely characterizes the MEM map. This could be significantly less than n in interferometric applications. Agmon et al. (3) and Alhassid et al. (4) have given an interesting variational principle that involves these parameters and the data alone. Their original scheme was for noiseless data, but it has since been extended to include the χ^2 constraint (123). For the specific case of $-I \ln I$ entropy, the maximization of S' in Equation 30 is equivalent to maximizing

$$S' = -\sum_i I_i \ln I_i - \sum_r v_r \left[\sum_i P_{ri}I_i + N_r - D_r \right] - \mu \left[\sum_i I_i - D_0 \right]$$
$$- \lambda \left[\sum_r (N_r^2/\sigma_r^2) - \Omega \right],$$

where $\Omega \approx m$ corresponds to one standard deviation. Now, S' is a maximum with respect to variations in I_i and N_r, since this is the least-squares MEM condition. In addition, we also have $\partial S'/\partial v_r = \partial S'/\partial \mu = \partial S'/\partial \lambda = 0$, which gives back the constraint equations. It is then permissible to eliminate I_i, N_r, and μ from S' using the conditions $\partial S'/\partial I_i = 0$, $\partial S'/\partial N_r = 0$ and $\Sigma I_i = D_0$. Omitting constants, we then obtain a variational principle for v_r and λ in terms of the function

$$F(v_r, \lambda) = D_0 \ln \left[\sum_i \exp\left(-\sum_r v_r P_{ri} \right) \right] + \sum_r v_r D_r + \sum_r v_r^2 \sigma_r^2/4\lambda + \lambda\Omega.$$

The ME solution corresponds to an extremum of F, and thus we have an optimization problem in only m dimensions. As a bonus, no search is required to attain the required χ^2 value. It is not immediately obvious that

this algorithm will be significantly faster than the earlier ones, since the number of iterations for convergence of the gradient algorithms in image space seems to be independent of n, the dimensionality of the problem. A real virtue of this scheme, however, is that any set of values for the Lagrange multipliers gives a valid positive map, and so there are no boundaries in this space. In contrast, gradient algorithms that work in I space need various controls to prevent negative intensities. We note that it is not straightforward to include windows in the map in this scheme.

5. EXTENSIONS OF MAXIMUM ENTROPY IMAGING

We have so far considered an image characterized by a single scalar intensity at each pixel and by data related to these intensities by a linear transformation. When we deal with polarization, multifrequency data, and interferometry with incomplete phase information, these requirements are not met. The new techniques required for these more general situations are just now being explored.

5.1 Polarization

When one includes the polarization of the radiation field, a map of the sky comprises four Stokes parameters I, Q, U, and V at each pixel (22). Here, I is the total intensity, Q and U describe the linear polarization parallel to and at 45° with respect to (arbitrary) orthogonal axes, and V describes circular polarization. The 2 × 2 correlation matrix constructed from the x and y components of the complex electric field is related to the Stokes parameters as follows:

$$I = \begin{bmatrix} E_x^* E_x & E_y^* E_x \\ E_x^* E_y & E_y^* E_y \end{bmatrix} \equiv \frac{1}{2} \begin{bmatrix} I+Q & U+iV \\ U-iV & I-Q \end{bmatrix}.$$

In interferometry, each telescope of a pair receives two polarizations, which gives a 2 × 2 matrix of correlations at each spatial frequency (u, v). There is a Fourier transform relation between the image plane and the uv plane (83).

   It is clearly incorrect in principle to process the I, Q, U, and V maps independently. Firstly, the positivity constraint is inappropriate for Q, U, and V (e.g. the two signs of V describe opposite circular polarizations). Secondly, independent processing ignores the physical constraint that the total intensity should be greater than or equal to the polarized intensity $(Q^2 + U^2 + V^2)^{1/2}$. Ponsonby (93) suggested a reasonable extension of the MEM to the polarization case. For the ln I entropy, he proposed that one

160   NARAYAN & NITYANANDA

should maximize

$$S = \iint \ln (\lambda_1 \lambda_2) \, dx \, dy,$$

where λ_1, λ_2 are the eigenvalues of the correlation matrix defined above, i.e.

$$\lambda_{1,2} = [I \pm (Q^2 + U^2 + V^2)^{1/2}]/2.$$

The generalization to an arbitrary entropy form $f(I)$ gives (89)

$$S = \iint [f(\lambda_1) + f(\lambda_2)] \, dx \, dy = \iint \text{Trace} \, [f(I)] \, dx \, dy.$$

Physically, λ_1 and λ_2 represent the intensities of the two orthogonally polarized, mutually incoherent components into which the radiation emitted from each point in the source can be resolved, and the final expression is the sum of the entropies of these two components. It should be noted that even though the entropy is written in terms of λ_1 and λ_2, the final solution does not consist of images for λ_1 and λ_2, but rather of four images for I, Q, U, and V, respectively, as required.

The algorithms discussed in Section 4 can for the most part be generalized for polarization by simply replacing the scalar quantities I, D, etc., by appropriate 2×2 matrices I, D, etc. (89). In interferometry, the reconstructed polarized ME image is now a nonlinear transform of a band-limited Hermitian 2×2 matrix (89), a natural generalization of the scalar (unpolarized) case. We expect that the general properties of reconstructions discussed in Section 3 will hold. In the scalar case, the entropy function provides a barrier at $I = 0$ that is responsible for the suppression of ripple. In the present case, we have an additional barrier at $I = (Q^2 + U^2 + V^2)^{1/2}$ that corresponds to 100% polarized radiation. Thus, *the suppression of ripple in the degree of polarization will not be effective for a weakly polarized source.* For example, we expect a strongly polarized region against an unpolarized background to be surrounded by spurious oscillations in the degree of polarization (for the same reason that absorption features are reconstructed poorly; see Section 3.1.1). Similar effects may occur at a sharp junction between two orthogonally polarized regions of the map (as is sometimes found in radio jets) if the degree of polarization is weak. These conclusions are based on the general nature of the function being maximized, but they can be confirmed by a detailed analysis of the exponential of a band-limited matrix in the $-I \ln I$ case.

Polarized MEM has been studied in a few computer simulations (85, 104) with rather strongly polarized model sources, and the results are in

general very similar to those shown in Figure 5 for the scalar case. There have been no applications to real data so far, and such applications should be of great interest.

5.2 Spectral Imaging

In certain applications, a region of sky is imaged at several frequencies in order to obtain spectral information as a function of position. In the simplest case, the source has a smooth spectrum and measurements are made at a few widely separated frequencies v. One then attempts to reconstruct a map of spectral index α [where $I(v) \propto v^{-\alpha}$]. Note that in general, the coverage of the uv plane could be different at the different frequencies. Because of the various nonuniformities in ME restorations discussed in Section 3, it is dangerous to estimate α by taking ratios between independently restored images. Clearly the problem arises because the a priori information that the frequency dependence of the map is smooth has not been put in. One way to do this using defaults has been suggested by Gull & Skilling (51). Denoting the maps at two frequencies by I_{Ai} and I_{Bi}, each can be normalized as before by its total flux:

$$f_{Ai} = I_{Ai} \bigg/ \sum_i I_{Ai}, \qquad f_{Bi} = I_{Bi} \bigg/ \sum_i I_{Bi}.$$

A common default image for the two frequencies can then be obtained by averaging

$$f_{0i} = (f_{Ai} + f_{Bi})/2. \qquad 33.$$

This particular way of constructing the default ensures that multiplication of the A or B data by a constant only affects the scale of the final maps and not their shapes. An alternative choice is discussed in (51). The function to be maximized is now

$$S = -\sum_i f_{Ai} \ln (f_{Ai}/f_{0i}) - \sum_i f_{Bi} \ln (f_{Bi}/f_{0i}).$$

Clearly, this will encourage the A and B images to resemble each other (apart from their overall scale) to the extent allowed by the data.

An apparently natural implementation of this scheme would be to average the MEM images for A and B from one iteration to obtain the new default f_{0i} for the next iteration (starting with f_{0i} from the constant default A and B images in the very first round). However, the two images f_{Ai} and f_{Bi} must have the form given in Equation 28, and Equation 33 requires two such functions to add up to a constant over the whole map, which is generally impossible. What is therefore required is that one must explicitly substitute Equation 33 into S and vary both the numerator and denomi-

nator within the logarithms. The general arguments used to derive the entropy form (Section 2), as well as the experience with algorithms (Section 4), do not apply to this kind of floating default. Tests of this multifrequency technique should therefore prove very interesting.

Another application is spectral line mapping with a large number of frequency channels, where one has a "data cube." The obvious approach again would be to make an independent image in each channel, but this is dangerous when the lines are weak compared with the continuum (33), since small errors in the restoration in adjacent channels can badly corrupt the line information. A possible solution is to subtract the continuum before ME restoration, but this will not work when absorption lines are present, since these will not be consistent with the positivity constraint. Since the frequency information is also usually obtained through auto-correlation data, this problem should really be considered to be a case of three-dimensional restoration, with one frequency and two spatial dimensions. Depending on the details of the source structure, the image would consist of peaks, lines, or sheets in this space. The ln I form of entropy has strange artifacts in three-dimensional image restoration (Section 3) and should probably be avoided.

5.3   *Determination and Refinement of Visibility Phases*

In aperture synthesis, telescope-based errors arising from fluctuations in the atmosphere or from the electronics can corrupt the quality of the data, especially the visibility phases. In such cases, the recovery of phase information during the process of image reconstruction is particularly important because phases are known to be much more important than amplitudes (90, 94). Since n telescopes give correlations on $n(n-1)/2$ baselines but have only n errors, the corrupted visibilities do have partial phase information in the form of closure phases (69). (The sum of the phases on three baselines that form a triangle is called a closure phase.) Powerful "self-calibration" algorithms have been developed (29, 92) that make use of closure information to correct for telescope errors. Presently, these algorithms are mostly based on the Clean technique. Attempts to directly substitute the MEM for Clean in the standard self-calibration packages have not been very successful and have sometimes led to an unphysical restoration where all the visibilities are given zero phase (112). A more successful strategy is to clip the ME image at each iteration, leaving only the bright features (112). Thus one is effectively using the positivity-enforcing and peak-sharpening properties of the MEM. A more direct approach that seems to work quite well is to maximize the entropy using the closure phases as additional constraints (92, 97, 103).

With adequate closure information, it seems that the entropy maximum

is usually unique, barring a trivial ambiguity as to the position of the source. (There is no formal proof of uniqueness, only empirical evidence from simulations.) When no phase information at all is available, as in the case of Michelson interferometry (30), intensity interferometry (16), primitive speckle interferometry (7, 77), or earlier VLBI experiments, the uniqueness of the MEM reconstruction is not guaranteed. Some arguments suggest that when the positivity constraint is included and the spatial extent of the image is strongly confined, restorations from pure amplitude data are nearly always unique (17, 60). However, nontrivial ambiguities might still creep in when the effect of noise in the data is included (47). A few astronomical applications of the MEM to the problem of reconstruction without phases have been made (49). The method has been pursued more actively in other fields such as crystallography (14, 84), where it shows good promise.

6.   COMPARISON WITH OTHER NONLINEAR METHODS

6.1   *Clean*

The most widely used nonlinear restoration scheme in radio astronomy is Clean, introduced in 1974 by Högbom (56). Here the sky is modeled by a collection of point sources. At each iteration, the maximum I_{max} in the current map is located and a point source of strength gI_{max}/P_{max} is fitted at that location, where P_{max} is the height of the central maximum in the psf and g is an adjustable parameter ("loop gain") typically chosen in the range 0–1. A copy of the psf corresponding to this point source is now subtracted from the current map, and the difference map is taken as the current map for the next iteration. The algorithm is stopped when the difference map has been reduced to the noise level. In the final step, the set of point sources is convolved with a "clean beam" having a Gaussian or similar smooth shape of width comparable to the psf in order to get an image with the same resolution as the original data, and then the residual difference map is added. Figure 5 shows a Clean restoration of the model discussed in Section 3. Note the efficient removal of ripple but the rather modest resolution enhancement. The initial applications did indeed involve unresolved radio sources surrounded by large empty regions, in close agreement with the Clean model, and the method was remarkably success-ful. As discussed in Section 3, the MEM can also represent sharp peaks on a flat background. Numerical simulations comparing the performance of the two techniques have been carried out (32, 88), and they agree more than perhaps one might expect. We make a general comparison of the two methods below.

The major advantage of Clean over the MEM is its simplicity and intuitive appeal. A major disadvantage is its sequential nature, which is an obstacle to a deeper understanding of its properties. Schwarz (99) has shown that the position and strength of each point source identified by Clean can be understood as a least-squares fit to the current difference map. However, there is no least-squares interpretation of the entire collection of point sources. The MEM, on the other hand, is a clearly formulated variational problem that may ultimately prove easier to understand.

With bigger telescope arrays coming into operation, radio sources are routinely being resolved and do not conform to a point source model. Both Clean and the MEM do cope with this complication, but both have some difficulties. Clean images of extended sources often have "stripes" arising from the sequential nature of the algorithm (100). Modifications of the basic Clean algorithm to deal with such problems have been proposed and implemented (26, 58, 111) with some success, but the simplicity of the original scheme is lost. On balance, the MEM seems to be better equipped to handle extended sources. The basic MEM has problems with sidelobes when point sources are superposed on extended structure. Some suggestions to overcome this defect were discussed in Section 3.2.

For a map with n pixels, Clean requires of order n^2 operations, since the number of point sources fitted as well as the number of operations per point source are proportional to n. Recent modifications have made Clean more computer efficient (23, 24), but they do not alter the n^2 scaling. Experience has shown that the number of iterations required in the MEM (~ 50) is essentially independent of n, and since each iteration consists mainly of a few fast Fourier transforms, the computational load goes as $n \ln n$. Consequently, it appears that nonsequential schemes like the MEM will ultimately win if the present trend toward larger images continues.

It should be noted that Clean as presently implemented does not fit the data because of the final step of convolving the point source model with a "clean" beam. This step, which effectively multiplies the data by the transform of the clean beam in the uv plane, is necessary to avoid extrapolating the data too far. It would be interesting if a modification of Clean could be found that produced an image fitting the data. Present implementations of least-squares MEM also do not fit the data, but the discrepancy is only of the order of the noise as against the much larger discrepancy with Clean.

Another point to be noted is that Clean does not assume positivity of the image and might therefore be more suitable than the MEM for spectral imaging, where emission and absorption features can occur simultaneously (Section 5.2).

6.2 Positivity-Enforcing Algorithms

The Gerchberg-Saxton algorithm (45) is one of the simplest and best known among nonlinear algorithms that enforce positivity of the reconstructed image. This is an iterative scheme that repeatedly transforms between the image and data planes, enforcing image positivity (and possibly restricted spatial extent) in one plane and agreement with the data in the other. (The practical implementation of the algorithm is very similar to the iterative scheme described in Section 4.1.) Since the algorithm enforces nothing more than positivity (and image diameter where known), it makes fewer assumptions than the MEM. However, the method will find the "closest" positive map to the initial guess that is consistent with the data, and it is therefore sensitive to the initial conditions. Global maximization schemes like the MEM have the virtue that they produce the same final image regardless of the initial guess. Applications of the Gerchberg-Saxton algorithm in astronomy have been suggested for the phase problem in speckle interferometry (36) and VLBI (39).

One of the earliest applications of positivity in astronomy was by Biraud (10), who demonstrated significant resolution enhancement of a one-dimensional image in the presence of noise. The idea is to write the positive image in the form $I(x) \equiv [i(x)]^2$ so that, by the convolution theorem, the Fourier transforms $\tilde{I}_j$ and $\tilde{i}_j$ of $I(x)$ and $i(x)$ satisfy the following relation:

$$\tilde{I}_j = \sum_{k=-\infty}^{\infty} \tilde{i}_k \tilde{i}^*_{k-j}. \qquad\qquad 34.$$

Iterative schemes have been developed (10, 98) to solve for the $\tilde{i}_j$, given data on some of the $\tilde{I}_j$. The key result is that even when the data are band limited, one is usually forced to extend the band of the restoration in order to satisfy Equation 34, and this leads to superresolution. We are not aware of applications to two-dimensional image restoration.

In an interesting application of the decomposition given in Equation 34, Komesaroff & Lerche (75) considered a one-dimensional problem where measurements on $\tilde{I}_j$ are available only for $|j| \leq n$. They find (75; see also 76) that the positivity constraint forces $\tilde{I}_{n+1}$ to lie within a circular region in the complex plane. The radius of this circle is clearly a measure of how stringent the positivity constraint is. When $\tilde{I}_{n+1}$ is given a value corresponding to the center of its circle, the constraint on $\tilde{I}_{n+2}$ is another circle of the same radius, and so on. The natural choice of the centers of all the successive circles leads to an extrapolation of the $\tilde{I}_j$ that is identical to that obtained by maximizing the ln I form of entropy. Unfortunately, there seems to be no simple generalization of this beautiful result to two or more dimensions, perhaps because there is no longer a natural sequence in which to arrange the data.

Two other nonlinear positivity-enforcing restoration schemes deserve mention. Jansson et al. (61) have developed an iterative scheme that forces the restored intensity to lie between user-specified minimum and maximum values. It is a simple nonlinear modification of a previously known linear algorithm. Another iterative scheme based on positivity was suggested independently by Richardson (95) and Lucy (81). In this algorithm, the "likelihood" of the observed data increases with each iteration (81). Both methods have been remarkably successful in restoring optical images of planets (53, 54). Application to the M87 jet of Figure 1b (53) suggests that the methods are competitive with the MEM. Both schemes are fast and easy to implement, with the Jansson et al. method being somewhat superior in resolution enhancement (probably because the upper bound on the restored intensity is a powerful additional constraint, particularly in the case of planets). Further work on these algorithms seems to be worthwhile.

7. SUMMARY AND OUTLOOK

There is no doubt that ME image reconstruction has come to stay in astronomy. It is able to compete with other techniques based on positivity and gives results comparable to Clean for maps dominated by point sources. At the same time, it can be extended to handle more complex situations using the default map idea or one of the other suggestions discussed in Section 3.2. Moreover, being based on a variational principle, the MEM is easier to analyze than other methods.

The thrust of much ME research has so far been toward clarifying the conceptual and statistical foundations of the method, issues that we discussed in Section 2. The form of entropy to be maximized has been a thorny question in the past, but the majority opinion seems to be converging on $-I \ln (I/I_0)$ with a suitable default $I_0(x, y)$. However, examples like the one shown in Figure 2 and the failure of the MEM with absorption features (Section 3.1.1; see also 85) indicate that the method should not be used as a black box. This is because the basic qualities of "good" images are in general complex and may be hard to quantify in some situations. Even though most astronomers would agree on identifying an obvious artifact (like the four sources on an exact rectangle in Figure 2), an automatic and machine-implementable procedure to do this is an unrealized goal of research into pattern recognition and artificial intelligence.

Regardless of one's confidence in the foundations of the MEM, it is important to understand the properties of ME images. This is an area where a fair amount of empirical knowledge has accumulated, and we have tried to systematize this information in Section 3. We identified the peak-sharpening and baseline-flattening tendency of the MEM and

discussed the sensitivity of the results to the background level. These led to suggestions for improving the MEM such as using an additive constant or subtracting a suitable background or combining the MEM and Clean. Many of these extensions would be regarded by purists as going beyond the MEM proper, since they end up with various knobs that the user can tune depending on his application. However, a close analysis reveals that most current image restoration schemes do indeed contain user-selectable parameters.

A somewhat unwelcome feature of the least-squares modification of the MEM is the biased nature of the residuals between the data and the reconstruction (Section 3). We find it disturbing that the method modifies good data in such a manner as to actually increase the noise. On the other hand, the method treats poor data in a reasonable way. An interesting area for future research is to develop a variant of the MEM that retains its least-squares character with poor data without systematically modifying good data. The understanding gained from the linear limit of the MEM (Section 3.1.3) may be useful for this.

The various schemes of Section 3.2 for controlling the MEM need to be extended to polarized maps (Section 5.1), which have a richer variety of possible structures. Combinations of emission and absorption features (as encountered in work with the 21-cm line) are likely to stretch the MEM or any other reconstruction technique to its limits. The same is true of visibilities with phase errors. These and other topics discussed in Section 5 are virtually unexplored and promise to be interesting areas of future research.

As presently formulated, the entropy of an image depends only on the frequency with which various intensity values occur and not on their spatial distribution. The a priori knowledge or belief that adjacent pixels tend to be correlated will undoubtedly be incorporated in the future. A beginning has already been made (114) by including derivatives of the intensity into the function being maximized. This acts to discourage ripple independently of its height from the baseline. It is possible that encouraging such correlations between neighboring pixels, especially along the frequency axis, will help in the spectral imaging problem (Section 5.2).

We have not found it necessary or even possible to single out one particular image restoration scheme from the many that are available. Ultimately, in our view, any image consistent with the data and free from obvious artifacts must be taken seriously. The goal of restoration techniques is to produce such images. When the results are relatively independent of the method, we gain confidence in them. However, if we obtain significantly different restorations, the places where they agree should tell us what we can really believe, and the differences should indicate

what extra data are needed. Such interaction with observations would be a most constructive role for image reconstruction.

ACKNOWLEDGMENTS

Literature Cited

1. Ables, J. G. 1974. *Astron. Astrophys. Suppl.* 15: 383–93
2. Ables, J. G. 1984. See Ref. 96, pp. 321–22
3. Agmon, N., Alhassid, Y., Levine, R. D. 1979. *J. Comput. Phys.* 30: 250–58
4. Alhassid, Y., Agmon, N., Levine, R. D. 1978. *Chem. Phys. Lett.* 53: 22–26
5. Andrews, H. C., Hunt, B. R. 1977. *Digital Image Restoration.* Englewood Cliffs, NJ: Prentice-Hall. 238 pp.
6. Arp, H., Lorre, J. 1976. *Ap. J.* 210: 58–64
7. Bates, R. H. T. 1982. *Phys. Rep.* 90: 203–97
8. Bayes, T. 1763. *Phil. Trans. R. Soc. London* 53: 330–418
9. Bhandari, R. 1978. *Astron. Astrophys.* 70: 331–33
10. Biraud, Y. 1969. *Astron. Astrophys.* 1: 124–27
11. Bracewell, R. N. 1979. *Ann. Rev. Astron. Astrophys.* 17: 113–34
12. Bracewell, R. N., Roberts, J. A. 1954. *Aust. J. Phys.* 7: 615–40
13. Braun, R., Strom, R. G. 1986. *Astron. Astrophys. Suppl.* 63: 345–401
14. Bricogne, G. 1984. *Acta Crystallogr. A* 40: 410–45
15. Brigham, E. O. 1974. *The Fast Fourier Transform.* Englewood Cliffs, NJ: Prentice-Hall. 252 pp.
16. Brown, R. H. 1974. *The Intensity Interferometer.* London: Taylor & Francis. 184 pp.
17. Bruck, Yu. M., Sodin, L. G. 1979. *Opt. Commun.* 30: 304–8
18. Bryan, R. K., Skilling, J. 1980. *MNRAS* 191: 69–79
19. Burch, S. F., Gull, S. F., Skilling, J. 1983. *Comput. Vision, Graphics, Image Process.* 23: 113–28
20. Burg, J. P. 1967. *Ann. Meet. Int. Soc. Explor. Geophys.* Reprinted in *Modern Spectral Analysis,* 1978, ed. D. G. Childers, pp. 34–41. New York: IEEE Press
21. Burg, J. P. 1975. *Maximum entropy spectral analysis.* PhD thesis. Dep. Geophys., Stanford Univ., Stanford, Calif.
22. Chandrasekhar, S. 1960. *Radiative Transfer.* New York: Dover. 393 pp.
23. Chen, C. F., Frater, R. H. 1984. See Ref. 96, pp. 425–30
24. Clark, B. G. 1980. *Astron. Astrophys.* 89: 377–78
25. Cooley, J. W., Tukey, J. W. 1965. *Math. Comput.* 19: 297–301
26. Cornwell, T. J. 1983. *Astron. Astrophys.* 121: 281–85
27. Cornwell, T. J. 1984. See Ref. 96, pp. 291–96
28. Cornwell, T. J., Evans, K. F. 1985. *Astron. Astrophys.* 143: 77–83
29. Cornwell, T. J., Wilkinson, P. N. 1984. See Ref. 96, pp. 207–19
30. Davis, J. 1984. See Ref. 96, pp. 125–41
31. Dixon, L. C. W. 1972. *Nonlinear Optimisation.* London: English Univ. Press. 214 pp.
32. Dulk, G. A., McLean, D. J., Manchester, R. N., Ostry, D. I., Rogers, P. G. 1984. See Ref. 96, pp. 355–61
33. Ekers, R. D., van Gorkom, J. H. 1984. See Ref. 96, pp. 21–32

34. Fabian, A. C., Willingale, R., Pye, J. P., Murray, S. S., Fabbiano, G. 1980. *MNRAS* 193: 175–88
35. Fellgett, P. B., Linfoot, E. H. 1954. *Proc. R. Soc. London Ser. A* 247: 369–407
36. Fienup, J. R. 1984. See Ref. 96, pp. 99–109
37. Fletcher, R., Reeves, C. M. 1964. *Comput. J.* 7: 149–54
38. Fomalont, E. B. 1979. See Ref. 118, pp. 3–18
39. Fort, D. N., Yee, H. K. C. 1974. *Astron. Astrophys.* 50: 19–22
40. Frieden, B. R. 1972. *J. Opt. Soc. Am.* 62: 511–18
41. Frieden, B. R. 1975. In *Picture Processing and Digital Filtering*, ed. T. S. Huang, pp. 177–248. Berlin: Springer-Verlag
42. Frieden, B. R., Burke, J. J. 1972. *J. Opt. Soc. Am.* 62: 1202–10
43. Frieden, B. R., Swindell, W. 1976. *Science* 191: 1237–41
44. Frieden, B. R., Wells, D. C. 1978. *J. Opt. Soc. Am.* 68: 93–103
45. Gerchberg, R. W., Saxton, W. O. 1972. *Optik* 35: 237–46
46. Giacconi, R., Branduardi, G., Briel, U., Epstein, A., Fabricant, D., et al. 1979. *Ap. J.* 230: 540–50
47. Greenaway, A. H., Walker, J. G., Coombs, J. A. G. 1984. See Ref. 96, pp. 111–17
48. Gull, S. F. 1984. See Ref. 96, p. 321
49. Gull, S. F., Daniell, G. J. 1978. *Nature* 272: 686–90
50. Gull, S. F., Daniell, G. J. 1979. See Ref. 118, pp. 219–25
51. Gull, S. F., Skilling, J. 1984. See Ref. 96, pp. 267–79
52. Haykin, S. 1979. *Non-Linear Methods of Spectral Analysis*. Berlin: Springer-Verlag. 247 pp.
53. Heasley, J. N. 1984. *Publ. Astron. Soc. Pac.* 96: 767–72
54. Heasley, J. N., Pilcher, C. B., Howell, R. P., Caldwell, J. J. 1984. *Icarus* 57: 432–42
55. Heffernan, P. B., Bates, R. H. T. 1982. *Optik* 60: 129–42
56. Högbom, J. A. 1974. *Astron. Astrophys. Suppl.* 15: 417–26
57. Högbom, J. A. 1979. See Ref. 118, pp. 237–39
58. Högbom, J. A. 1984. See Ref. 96, pp. 247–54
59. Horne, K. D. 1982. *Eclipse mapping of accretion disks in cataclysmic binaries*. PhD thesis. Dep. Astron., Calif. Inst. Technol., Pasadena
60. Huiser, A. M. J., van Toorn, P. 1980.
61. Jansson, P. A., Hunt, R. H., Plyler, E. K. 1970. *J. Opt. Soc. Am.* 60: 596–99
62. Jaynes, E. T. 1957. *Phys. Rev.* 106: 620–30
63. Jaynes, E. T. 1957. *Phys. Rev.* 108: 171–90
64. Jaynes, E. T. 1968. *IEEE Trans. Syst. Sci. Cybern.* 4: 227–41
65. Jaynes, E. T. 1982. *Proc. IEEE* 70: 939
66. Jaynes, E. T. 1983. *Papers on Probability, Statistics and Statistical Physics*, ed. R. D. Rosenkrantz. Dordrecht: Reidel. 434 pp.
67. Jaynes, E. T. 1984. *Proc. Workshop on Bayesian/Maximum Entropy Methods*, ed. J. H. Justice. Calgary: Univ. Calgary
68. Jaynes, E. T. 1984. *Proc. Workshop on Bayesian/Maximum Entropy Methods*, ed. J. H. Justice. Calgary: Univ. Calgary
69. Jennison, R. C. 1958. *MNRAS* 118: 276–84
70. Karbelkar, S. N. 1986. *Pramana* 26: 301
71. Karle, J., Hauptman, H. 1950. *Acta Crystallogr.* 3: 181–87
72. Kendall, M., Stuart, A. 1977. *The Advanced Theory of Statistics*, Vol. 1. London: Griffin. 472 pp.
73. Khinchin, A. I. 1957. *Mathematical Foundations of Information Theory*. New York: Dover. 120 pp.
74. Kikuchi, R., Soffer, B. H. 1977. *J. Opt. Soc. Am.* 67: 1656–65
75. Komesaroff, M. M., Lerche, I. 1979. See Ref. 118, pp. 241–47
76. Komesaroff, M. M., Narayan, R., Nityananda, R. 1981. *Astron. Astrophys.* 93: 269–81
77. Labeyrie, A. 1978. *Ann. Rev. Astron. Astrophys.* 16: 77–102
78. Ladd, M. F. C., Palmer, R. A. 1980. *Theory and Practice of Direct Methods in Crystallography*. New York: Plenum. 421 pp.
79. Landau, L. D., Lifshitz, E. M. 1958. *Statistical Physics*. Oxford: Pergamon. 484 pp.
80. Lim, J. S., Malik, N. A. 1981. *IEEE Trans. Acoust. Speech Signal Process.* 29: 401–13
81. Lucy, L. B. 1974. *Astron. J.* 79: 745–54
82. Michelson, A. A., Pease, F. G. 1921. *Ap. J.* 53: 249–59
83. Morris, D., Radhakrishnan, V., Seielstad, G. C. 1964. *Ap. J.* 139: 551–59
84. Narayan, R., Nityananda, R. 1981. *Curr. Sci.* 50: 168–70

Opt. Lett. 5: 499–501

85. Narayan, R., Nityananda, R. 1984. See Ref. 96, pp. 281–90
86. Newman, W. I. 1977. *Astron. Astrophys.* 54: 369–80
87. Newman, W. I. 1977. *IEEE Trans. Inf. Theory* 23: 89–93
88. Nityananda, R., Narayan, R. 1982. *J. Astrophys. Astron.* 3: 419–50
89. Nityananda, R., Narayan, R. 1983. *Astron. Astrophys.* 118: 194–96
90. Oppenheim, A. V., Lim, J. S. 1981. *Proc. IEEE* 69: 529–41
91. O'Sullivan, J. D., Komesaroff, M. M. 1984. See Ref. 96, pp. 297–302
92. Pearson, T. J., Readhead, A. C. S. 1984. *Ann. Rev. Astron. Astrophys.* 22: 97–130
93. Ponsonby, J. E. B. 1973. *MNRAS* 163: 369–80
94. Ramachandran, G. N., Srinivasan, R. 1970. *Fourier Methods in Crystallography.* New York: Wiley. 259 pp.
95. Richardson, W. H. 1972. *J. Opt. Soc. Am.* 62: 55–59
96. Roberts, J. A., ed. 1984. *Proc. Int. Symp. Indirect Imaging, Sydney, Aust., 1983.* Cambridge: Cambridge Univ. Press. 439 pp.
97. Sanroma, M., Estalella, R. 1983. *Proc. Int. Conf. VLBI Technol., Toulouse, Fr., 1982,* pp. 391–407. Toulouse: Cent. Natl. Étud. Spat. 488 pp.
98. Schell, A. C. 1965. *Radio Electron. Eng.* 29: 21
99. Schwarz, U. J. 1978. *Astron. Astrophys.* 65: 345–56
100. Schwarz, U. J. 1984. See Ref. 96, pp. 255–60
101. Shannon, C. E. 1948. *Bell System Tech. J.* 27: 379–423
102. Shannon, C. E. 1948. *Bell System Tech. J.* 27: 623–56
103. Shevgaonkar, R. K. 1986. *Astron. Astrophys.* In press
104. Shevgaonkar, R. K. 1986. *Astron. Astrophys.* In press
105. Shore, J. E., Johnson, R. W. 1980. *IEEE Trans. Inf. Theory* 26: 26–37
106. Shore, J. E., Johnson, R. W. 1983. *IEEE Trans. Inf. Theory* 29: 942–43
107. Skilling, J., Bryan, R. K. 1984. *MNRAS* 211: 111–24
108. Skilling, J., Gull, S. F. 1984. *SIAM Proc. Am. Math. Soc.* 14: 167
109. Skilling, J., Strong, A. W., Bennett, K. 1979. *MNRAS* 187: 145–52
110. Skinner, G. K. 1984. *Nucl. Instrum. Methods Phys. Res.* 221: 33–40
111. Steer, D. G., Dewdney, P. E., Ito, M. R. 1984. *Astron. Astrophys.* 137: 159–65
112. Steer, D. G., Ito, M. R., Dewdney, P. E. 1984. See Ref. 96, pp. 347–53
113. Subrahmanya, C. R. 1980. *Bull. Astron. Soc. India* 8: 5–13
114. Subrahmanya, C. R. 1980. *Astron. Astrophys.* 89: 132–39
115. Tikhonov, A. N., Arsenin, B. Y. 1977. *Solution of Ill-Posed Problems.* New York: Wiley. 258 pp.
116. Tikochinsky, Y., Tishby, N. Z., Levine, R. D. 1984. *Phys. Rev. Lett.* 52: 1357–60
117. Van Schooneveld, C. 1979. See Ref. 118, pp. 197–218
118. Van Schooneveld, C., ed. 1979. *Proc. IAU Colloq. 49, Image Formation From Coherence Functions in Astronomy, Groningen, Neth., 1978.* Dordrecht: Reidel. 340 pp.
119. Wells, D. C. 1980. *Proc. Soc. Photo-Opt. Instrum. Eng.* 264: 148–56
120. Wernecke, S. J. 1976. *Maximum entropy techniques for image reconstruction from interferometer measurements and projections.* PhD thesis. Dep. Electr. Eng., Stanford Univ., Stanford, Calif.
121. Wernecke, S. J. 1977. *Radio Sci.* 12: 831–44
122. Wernecke, S. J., D'Addario, L. R. 1977. *IEEE Trans. Comput.* 26: 351–64
123. Wilczek, R., Drapatz, S. 1985. *Astron. Astrophys.* 142: 9–12
124. Willingale, R. 1981. *MNRAS* 194: 359–64
125. Woolf, N. J. 1982. *Ann. Rev. Astron. Astrophys.* 20: 367–98