Astron. Astrophys. 143, 77–83 (1985)

ASTRONOMY AND ASTROPHYSICS

A simple maximum entropy deconvolution algorithm

T. J. Cornwell[1] and K. F. Evans[2]

[1] National Radio Astronomy Observatory*, PO Box 0, Socorro, NM 87801, USA
[2] Dept. of Astronomy, California Institute of Technology, Pasadena, California, USA

Received May 17, accepted July 23, 1984

Summary. A simple maximum entropy image deconvolution algorithm, now implemented in the Astronomical Image Processing System AIPS as task VM, is described. VM uses a simple Newton-Raphson approach to optimise the relative entropy of the image subject to constraints upon the rms error and total power enforced by Lagrange multipliers. Some examples of the application of VM to VLA data are given.

Key words: data processing – maximum entropy – radio astronomy

1. Introduction

In contrast to simple linear arrays, modern radio interferometric arrays such as the VLA (Thompson et al., 1980), MERLIN (Davies et al., 1980) and the proposed VLBA (NRAO staff, 1982) provide irregular, sparse coverage of the aperture plane. The resulting sidelobes in the Fourier synthesised image seriously degrade the dynamic range, usually to less than about 20 dB. The non-linear deconvolution algorithm CLEAN (Hogbom, 1974) was developed to circumvent this problem and has, until recently, performed spectacularly well. Used in conjunction with selfcalibration techniques (see, e.g., Cornwell and Wilkinson, 1984; Pearson and Readhead, 1984) CLEAN has improved the dynamic range in the best images to about 35 to 40 dB.

However, CLEAN has three major drawbacks: first, it is slow and inefficient since pixels are adjusted individually [a simple modification may improve this aspect substantially, Steer et al. (1984)]. Secondly, extended emission is reconstructed rather poorly [see, e.g., Cornwell (1983) for an example and Schwarz (1984) for a discussion of the causes]. Thirdly, in general the CLEAN image cannot be expressed in any simple mathematical form, closed or open; consequently, it is impossible to make statements about, for example, the effect on noise on a CLEANed image.

These criticism of CLEAN les us to re-consider the advantages of the Maximum Entropy Method of deconvolution. The MEM was introduced in image reconstruction some years ago (see, e.g., Frieden and Wells, 1978; Wernecke and D'Addario, 1976) but has, for a variety of reasons, not been widely accepted in radioastronomy. Early algorithms were relatively very slow and not very robust and, furthermore, claims about the propriety of the MEM based upon esoteric discussions of information theory and statistics were disquieting to many astronomers. [For those interested in the history of the attempts to justify the MEM in image reconstruction we recommend the following articles: Ables (1974), Wernecke and D'Addario (1976), Gull and Daniell (1978), Gull and Skilling (1984), and Cornwell (1984). For an overview of the maximum entropy principle in general we recommend the collection of the papers of E.T. Jaynes edited by Rosencrantz (1983).]

However, in recent years these two main objections to the MEM have been answered. Firstly, it has recently been demonstrated that the MEM in image reconstruction can be justified on purely heuristic grounds as a device for introducing a priori information (see, e.g., Nityananda and Narayan, 1983; Narayan and Nityananda, 1984, and below). Secondly, the advent of modern interferometric arrays allowing the imaging of fields spanning many resolution elements has increased the typical run time of the CLEAN algorithm by up to two orders of magnitude (although note that the Clark (1980) modification regained a factor of 2 to 5) while the typical run time of MEM algorithms, measured in terms of the number of two dimensional FFTs required, has remained approximately constant. Consequently it is now possible for MEM algorithms to be competitive with CLEAN in some applications. Although it is now in widespread use we felt that the powerful algorithm developed by Skilling and co-workers (Burch et al., 1983) is, for radio astronomical applications, made inefficient by its generality. Consequently, we were encouraged to develop a simple yet fast algorithm, which we will call VM, specifically tailored for radio interferometric applications.

The heart of our algorithm is a simple Newton-Raphson technique for optimising the relative entropy of the image subject to various constraints enforced by Lagrange multipliers. The large dimensional matrix inverse required by a second order method is approximated by taking the inverse of the diagonal elements (Cornwell, 1980). We will show that this is tantamount to neglecting the sidelobes of the point spread function, an approximation which is acceptable in the second order. Some trivial control procedures governing the Lagrange multipliers and the stopping criterion complete the algorithm. We have found that VM converges to a reasonable proximity to the true MEM image in about 10–50 iterations, where the major computational cost per iteration is two fast Fourier transforms. The total computational cost is comparable to or may be less than that of CLEAN when the image is well filled with emission. For objects which are well represented by a relatively small number of point components CLEAN is faster by an order of magnitude.

Send offprint requests to: T. J. Cornwell

* National Radio Astronomy Observatory is operated by Associated Universities Inc., under contract with the National Science Foundation

The layout of this paper is as follows: we discuss the rationale behind the MEM in Sect. 2, some properties of MEM images in Sect. 3, our algorithm in Sect. 4. Finally in Sect. 5 we give examples of its application to radio interferometric imaging, which is our main area of interest.

2. The maximum entropy method

In this section we will give two justifications for the maximum entropy approach to image reconstruction, neither of which is founded in information theory.

For simplicity in the following discussion we use a discrete representation of the images:

$$\boldsymbol{b}=(b_i|i=1,N)$$

where the images have N pixels. We should emphasize that no commitment to one dimension is implied in this notation.

The first justification is based upon the observation that we often have some a priori information concerning the object which we are to reconstruct from a data set. Suppose for concreteness that this knowledge can be expressed in the form of an image $\boldsymbol{m}$ which we expect the reconstructed image $\boldsymbol{b}$ to resemble. A reasonable reconstruction of the true image should obviously lie as near to $\boldsymbol{m}$ as the data allows. A measure of the distance between $\boldsymbol{b}$ and $\boldsymbol{m}$ is thus required. Some of the more obvious measures can be rejected immediately; for example, it is easy to show that the Euclidean distance, formed by the sum of the squares of the pixel differences, leads to linear filter when the data constraints are added. Adding supplementary non-holonomic constraints such as positivity is possible and produces a suitably non-linear filter but leads to an awkward optimisation problem. The "city-block" measure, formed by the sum of the absolute values of the pixel differences, is again awkward to optimise. The relative entropy:

$$H(\boldsymbol{b}|\boldsymbol{m})=-\Sigma_i b_i \ln(b_i/m_i)$$

has proved to be a useful measure of the difference between two probability distributions $\boldsymbol{b}$ and $\boldsymbol{m}$ (see, e.g., Shore and Johnson, 1980, for references). We should note that the negativity of the relative entropy is conventional and unimportant. Although in image reconstruction we are not directly concerned with probability distributions, we can use this as our measure of distance of $\boldsymbol{b}$, the reconstructed image, from $\boldsymbol{m}$, the image expected a priori. Of course, one may object that many other choices are possible (see, e.g., Nityananda and Narayan, 1982). However, a second, although admittedly weak, argument in favor of the relative entropy, again divorced from information theory, has been advanced by Shore and Johnson, 1980 (see Gull and Skilling, 1984 for a vivid illustration of the basic argument). Shore and Johnson showed that if one wished to construct a consistent theory of inference based upon the minimisation of some distance measure between a priori and a posteriori "degrees of belief" then the relative entropy must be used as the distance measure. Consistency in their terms means that restating the problem in an equivalent form, for example performing an intermediate grouping of possibilities to be considered, must lead to the same answer. Gull and Skilling (1984) noted that the same argument can be used for image reconstruction if one substitutes "image" for "degree of belief". The consistency requirement is then tantamount to requiring that an intermediate notional division of the image into sub-images, each with its own relative entropy, should not affect the final reconstruction. Similarly, neither should the pixel labelling be relevant. For an elaboration of these ideas the reader is referred to the original paper by Shore and Johnson (1980).

Thus to summarise our interpretation of the maximum entropy method as applied to image reconstruction: we wish to find an image $\boldsymbol{b}$ as close as is allowed by the data to an "a priori" image $\boldsymbol{m}$. A mild consistency argument dictates that the relative entropy be used as the distance measure. We do not believe that Jaynes' maximum entropy principle (see, e.g., Rosencrantz, 1983) can be directly applied to image reconstruction since the required identification of some attribute of an image with a probability distribution is not possible. A paper by one of us (Cornwell, 1984) expands upon this point (but see also the answer by Gull and Skilling, 1984).

3. Properties of images reconstructed with maximum entropy

The MEM image is obtained by maximising the relative entropy H subject to whatever data constraints apply. In the simplest and trivial case of no data we thus obtain the a priori image:

$$\boldsymbol{b}=\boldsymbol{m}.$$

Data constraints will drag the reconstructed image $\boldsymbol{b}$ away from the a priori image $\boldsymbol{m}$ but, given any freedom in obeying the constraints, for example because of noise, $\boldsymbol{b}$ will be biased towards $\boldsymbol{m}$. This has important and obvious implications for the treatment of noisy data since then the MEM image must then necessarily be biased. Of course, if $\boldsymbol{m}$ is close to the true image then this bias is unimportant. A simple procedure which makes the bias negligible, at least a low spatial frequencies, may be easily devised: we find the MEM image appropriate to a flat a priori image, then convolve this image with a broad beam to retain only the large-scale structure, and finally find the MEM image appropriate to this lower resolution a priori image. We are then implicitly using the a priori information that the true image contains little power at high spatial frequencies, an assumption which is also implicit in the smoothing with a CLEAN beam performed in the CLEAN algorithm. Alternatively, it is possible to use a lower resolution image as the a priori image. It should be clear that the a priori image allows a great deal of flexibility in that a specific model for the object to be imaged can be inserted directly in the algorithm. A further advantage is that a priori information can be expressed in terms of images rather than the properties of an algorithm. In a similar vein Frieden and Wells (1978) developed an algorithm in which knowledge of the background is used and showed that an increase in resolution results. However, the background is additive and must be chosen so that the foreground image is always positive. By contrast, the use of the a priori image is less restrictive and acts more as a position dependent scaling of the brightness.

Another cause of bias is the inflexibility of the logarithm term, which constrains $\boldsymbol{b}$ and $\boldsymbol{m}$ to be positive (or both negative if we change the minus sign). Thus, for example, if $\boldsymbol{m}$ is totally positive then so is $\boldsymbol{b}$ and therefore, given zero mean noise, $\boldsymbol{b}$ is biased. Although this "one-sidedness" constraint has a useful role to play in increasing the resolution [see, e.g., Frieden (1976) for a simple explanation of this point] it also forces us to acknowledge the inevitable presence of noise in any data which we may collect. In concrete terms we cannot simply add data points, e.g., noisy Fourier coefficients of $\boldsymbol{b}$, as separate constraints each with its own Lagrange multiplier since the resulting optimisation problem will almost certainly have no solution. This point was first emphasized by Ables (1974) who suggested requiring only that the global χ^2 of

the misfit between the observed data points and those predicted from $\boldsymbol{b}$ be equal to the expected value. For large numbers of data points this somewhat less stringent constraint will nearly always be consistent with a "one-sided" image $\boldsymbol{b}$. However, as emphasized above, such extra latitude will result in a small amount of bias in the final image. Other noise tolerant constraints, such as forcing the residuals to have the expected normal distribution (Bryan and Skilling, 1980), can be used to reduce the bias but we believe that choosing $\boldsymbol{m}$ appropriately is more efficient and usually reduces the bias to an acceptable level.

In order to discuss further the properties of MEM images we must be rather more specific about the constraints. Restricting our attention to an interferometer, our observed data are samples of the visibility function. These can be expressed in the form:

$$V_k = \Sigma_i t_i \exp(2\pi j \boldsymbol{u}_k \cdot \boldsymbol{x}_i) + n_k ,$$

where $\boldsymbol{n}$ represents noise, $\boldsymbol{t}$ is the true image, $\boldsymbol{u}_k$ is the position vector of the k'th sample in the u, v plane and $\boldsymbol{x}_i$ is the position vector in the image plane of the i'th pixel. The corresponding data predicted from $\boldsymbol{b}$ are:

$$V_k' = \Sigma_i b_i \exp(2\pi j \boldsymbol{u}_k \cdot \boldsymbol{x}_i) .$$

Forming the χ^2 for $\boldsymbol{n}$:

$$\chi^2 = \Sigma_k w_k |V_k - V_k'|^2 ,$$

where w_k is the relevant weighting factor for the k'th visibility point.

Maximising the relative entropy subject to the constraint that χ^2 be equal to the expected value we find that the solution is:

$$b_i = m_i \exp[-\alpha(\partial\chi^2/\partial b_i)] ,$$

where α is a Lagrange undetermined multiplier and in this particular case:

$$\partial\chi^2/\partial b_i = 2\,\mathrm{Re}[\Sigma_k w_k (V_k' - V_k) \exp(-j2\pi \boldsymbol{u}_k \cdot \boldsymbol{x}_i)] .$$

Since this represents an inverse Fourier transform of the error in prediction of the true visibility the MEM image must be biased. In the image plane:

$$\partial\chi^2/\partial b_i = 2(\Sigma_j p_{i,j} b_j - d_i) ,$$

where $\boldsymbol{d}$ is the Fourier transform of V and p is the point spread function, given by the Fourier transform of the weights.

$$p_{i,j} = \mathrm{Re}\{\Sigma_k w_k \exp[-j2\pi \boldsymbol{u}_k \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j)]\} .$$

Constraining only the global χ^2 leads to a significant systematic bias in the MEM image since to maximise the relative entropy it is advantageous to arrange the errors to be much smaller than the noise on all data points except the total power which is biased by about $K \cdot \sigma$ where K is the number of independent visibility points. The resulting image fits the data far too well and is thus too noisy. Furthermore, since positivity is the main constraint effective in eliminating sidelobes these sidelobes are poorly suppressed. For these reasons when possible we add an extra constraint that the total power of the MEM image be that expected:

$$F = \Sigma_i b_i = F_{\text{observed}} .$$

Using Lagrange's method of undetermined multipliers we can now restate the optimisation problem as: maximise the objective function:

$$J = H - \alpha \cdot \chi^2 - \beta \cdot F ,$$

where α and β are chosen so that χ^2 and F are equal to the expected values. In the next section we will discuss possible methods of solving this problem.

4. An algorithm to optimise the relative entropy

The optimisation problem consists in finding an image such that the gradient of J, ∇J, is zero. The gradient of J is to be evaluated with respect to the pixel values b_i i.e. $(\nabla J)_i = \partial J/\partial b_i$. The implicit solution has the form:

$$b_i = m_i \exp(-\alpha(\partial\chi^2/\partial b_i) - \beta) .$$

Skilling and Gull (1984) have reported that utilization of this solution in an iterative substitution algorithm leads to slow convergence and instabilities. Thus the direct optimisation of J is to be preferred. Although this optimisation is conceptually simple, in practice the large dimensionality of the space over which J is defined restricts the feasible set of algorithms. For example, schemes involving explicit second order methods such as the various variable metric optimisation algorithms (e.g., Adby and Dempster, 1974) are impractical. However, the steepness of the relative entropy near zero brightness renders unattractive any simple first order approach based purely upon the gradients of H, χ^2, and F since the low brightness parts of the trial image dominate the gradient of the objective function. In order that the gradient of J can be weighted down near zero brightness we require an approximation to a second order method. In the exact Newton-Raphson method the step to the next trial image is given by:

$$\Delta\boldsymbol{b} = (-\nabla\nabla J)^{-1} \cdot \nabla J ,$$

where

$$\nabla J = \nabla H - \alpha \nabla\chi^2 - \beta \mathbf{1}$$

and

$$\nabla\nabla J = \nabla\nabla H - 2\alpha p .$$

Since $\nabla\nabla H$ is purely diagonal the non-diagonal elements of this matrix are due entirely to the point spread function p. This suggests one simple and viable approximation: neglect the non-diagonal elements of the Hessian of J in the standard Newton-Raphson approach. Since the main purpose of the use of $(-\nabla\nabla J)^{-1}$ is to weight down the effect of weak points upon $\Delta\boldsymbol{b}$ we may, for our purposes, neglect the sidelobes of p by using in place of p a suitably scaled identity matrix:

$$\nabla\nabla J = \nabla\nabla H - 2\alpha q I$$

where q is a scaling parameter required to convert Jy(beam area) to Jy/pixel. The exact value of q is relatively unimportant but it should represent the power in the main lobe of the point spread function. We then have that:

$$(-\nabla\nabla J)^{-1}_{i,i} \sim 1/(1/b_i + 2\alpha q)$$

$$(-\nabla\nabla J)^{-1}_{i,j} \sim 0 \quad \text{if} \quad i \neq j .$$

This also provides a useful metric for the $\boldsymbol{b}$-space: the lengths of all vectors are measured in this metric. For example, the vector scalar product between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is:

$$\|\boldsymbol{XY}\| = \Sigma_{i,j} X_i (-\nabla\nabla J)^{-1}_{i,j} Y_j$$

which is, under the above approximation:

$$\|\boldsymbol{XY}\| \sim \Sigma_i X_i Y_i / (1/b_i + 2\alpha q) .$$

Thus, in calculating norms the effect of weak points is weighted down linearly by the brightness. This serves to tame the steep gradient of ∇H near zero brightness and allows the step to the new image to be of reasonable length. In contrast to the metric used by Burch et al. (1983) this form does not over-emphasize the brighter points in the image. In fact, as iteration continues we increase α and so the metric tends towards the ideal case of a flat metric.

Since the length of $\Delta \boldsymbol{b}$ is dependent upon the approximations in $\nabla\nabla J$ it is advantageous and relatively inexpensive to perform a linear search for the zero of ∇J along $\Delta \boldsymbol{b}$. This involves calculating the residuals for two images and hence two convolutions are required. Fortunately, the residuals of the optimum interpolated image resulting from the linear search can themselves be interpolated. Therefore the net cost per iteration is dominated by one two-dimensional convolution, which can be efficiently calculated by FFT techniques.

Judging nearness to the true MEM image is difficult and somewhat arbitrary. We use the criterion that the norm of ∇J should be much less than the norm of a unit vector. Typically, we require:

$$\|\nabla J \cdot \nabla J\| < \varepsilon \|\mathbf{1} \cdot \mathbf{1}\| ,$$

where ε is of order 0.01 or less.

The Lagrange multipliers α and β must be adjusted as the iteration proceeds so that the final constraints are met. A simple Taylor series expansion of χ^2 and F allows us to calculate that, to first order, the required changes at any stage are:

$$\Delta\alpha = -\Delta\chi^2 / \|\nabla\chi^2 \cdot \nabla\chi^2\|$$

$$\Delta\beta = -\Delta F / \|\nabla F \cdot \nabla F\| .$$

Since changing α and β affects the gradient of J we limit the changes so that an analog to the above stopping criterion is always obeyed. For example, if α is to be changed by $\Delta\alpha$ then the condition becomes:

$$\|(\nabla J - \Delta\alpha \cdot \nabla\chi^2) \cdot (\nabla J - \Delta\alpha \cdot \nabla\chi^2)\| < \varepsilon \|\mathbf{1} \cdot \mathbf{1}\| .$$

Thus we obtain a quadratic equation for the maximum allowed change in α which can be solved for the upper and lower bounds. A sequence of maximum entropy images are thus obtained as iteration continues. The ratio $\|\nabla J \cdot \nabla J\| / \|\mathbf{1} \cdot \mathbf{1}\|$ can also act as a diagnostic. If the deconvolution has no solution or is ill-conditioned then this number increases to of order unity or greater as iteration proceeds.

All that remains is to protect against negative values of brightness; this we do with a simple clip, decreasing the minimum brightness allowed by an order of magnitude at each iteration. Usually, this protection is only necessary for the first few iterations.

This algorithm is coded in the Astronomical Image Processing System (AIPS) as task VM. Since within AIPS it is more convenient and efficient to work the dirty image and dirty beam rather than the visibility data, we approximate χ^2 by E/q where E, the misfit in the image plane, is given by:

$$E = \Sigma_i (\Sigma_j p_{i,j} \cdot b_j - d_i)^2 .$$

It is impossible to calculate χ^2 directly since the gridding process in the formation of the dirty map and beam introduces small but significant errors into the weighting terms w_i. This shortcut of using E/q is unimportant because χ^2 is only used directly in changing α.

Choice of a reasonable value for q will help convergence. The most obvious choice would seem to be the sum of the point spread function but, unfortunately, this is zero in the application to radio interferometry. We have found that usually the sum of either the equivalent Gaussian fitted to the main lobe or the main lobe itself is satisfactory. From the equation for the step length can see that if q is too high then the step length is underestimated. Therefore, we can use the result of the linear search in ∇J to adjust q so that the next step taken will be of the correct length.

Setting q to zero yields a metric similar to that used by Burch et al. (1983). We find that this metric drastically over-estimates the step length and that an arbitrary clip of the step is necessary. By contrast, in our algorithm such clipping of the step length is rarely needed. This "zero-q" metric also over-emphasizes the importance of the bright points and hence leads to spuriously strong peaks. Dulk et al. (1984) report that the Burch et al. (1983) algorithm appears to over-resolve peaks of components as is expected from use of the "zero-q" metric.

With careful coding the main cost of the VM algorithm is two fast Fourier transforms per iteration. The number of iterations typically required lies in the range 10–50 and so the total computational cost is of order 20–100 FFTs. For comparison we note that for objects spanning many resolution elements the Clark CLEAN algorithm (Clark, 1980) can take in excess of 500 FFTs. At the other extreme, for simple high dynamic range images CLEAN can be faster than VM by about two orders of magnitude.

In summary, the main elements of our algorithm are:

1. Optimisation of the relative entropy subject to constraints upon the rms error and total flux enforced by Lagrange multipliers.

2. A Newton-Raphson approach to optimising the objective function. In calculating the inverse Hessian non-diagonal elements are neglected. We have found this approximation to be perfectly adequate if the number of pixels per beam is restricted ($q < 50$). If not, then convergence can be very slow.

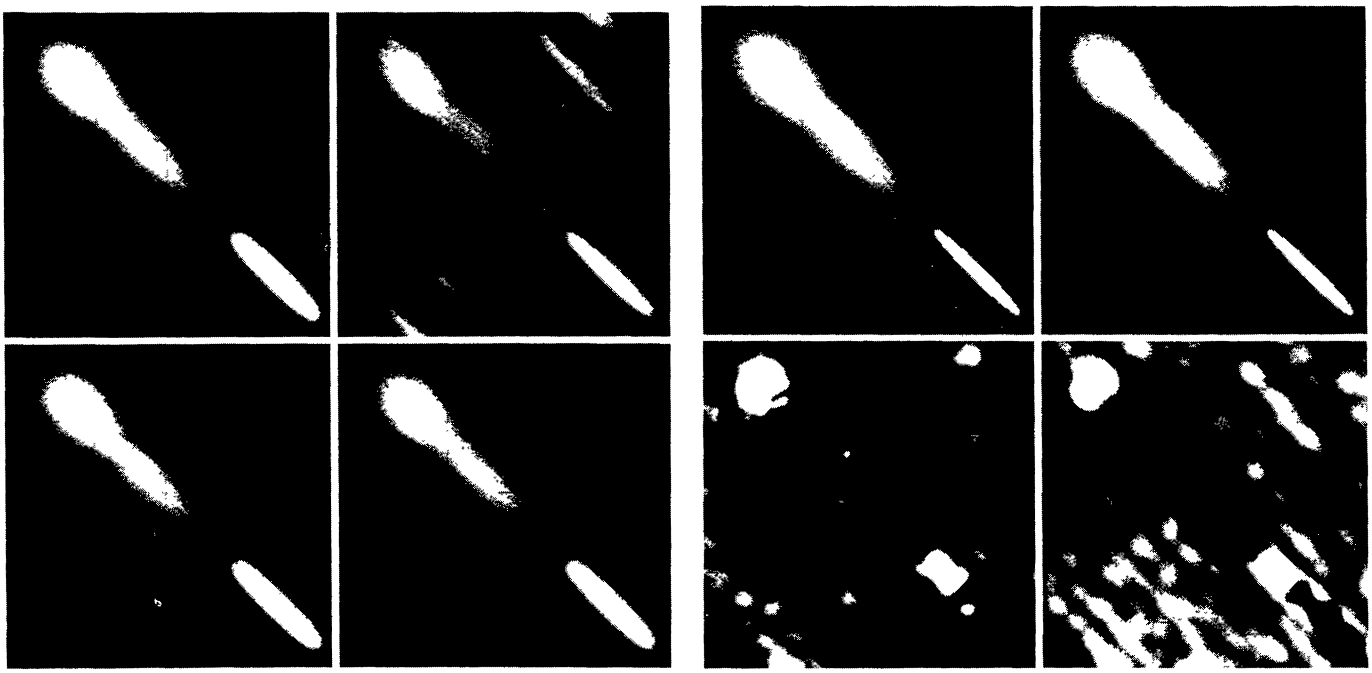
3. Iterative adjustment of the Lagrange multipliers so that the resulting gradient is small compared to the size of the unit vector.

Before leaving this section, we would like to point out the similarities of the VM algorithm to the enhanced CLEAN algorithm described by Steer et al. (1984). In both algorithms corrections are made to a trial image on the basis of discrepancy of the predicted dirty image from that observed. In the Steer et al. algorithm all points above a "trim" level are scaled and added to the current estimate whereas in VM, ignoring the entropy term, the metric weights down all points much less than the inverse of α before adding a correction to the current image. The effective trim level is thus decreased as α is increased to obtain agreement with the dirty image. This accidental convergence of different approaches to similar algorithms may be useful in suggesting new areas of research in image reconstruction.

Finally, it should be noted that this algorithm is effective for other forms of "entropy" e.g. $\Sigma b^{1/2}$ or $\Sigma \ln(b)$, at the same computational cost.

5. Some examples of MEM images

The VM algorithm has been used often in analyzing VLA data, most successfully when applied to observations of radio sources covering many resolution elements (>10,000) for which CLEAN is inefficient and produces poor results. It is most successfully applied to data in which the sidelobe level is small (<5% peak) and for which the total flux F and the rms noise level are both



a b

Fig. 1a and b. Tests of VM on model data: **a** Clockwise from top left, model convolved with CLEAN beam, dirty image, best VM image convolved with CLEAN beam, CLEAN image. **b** Model, best VM image, difference between VM image and model convolved with CLEAN beam, difference between CLEAN image and model convolved with CLEAN beam. The latter two images are saturated a factor 20 times lower

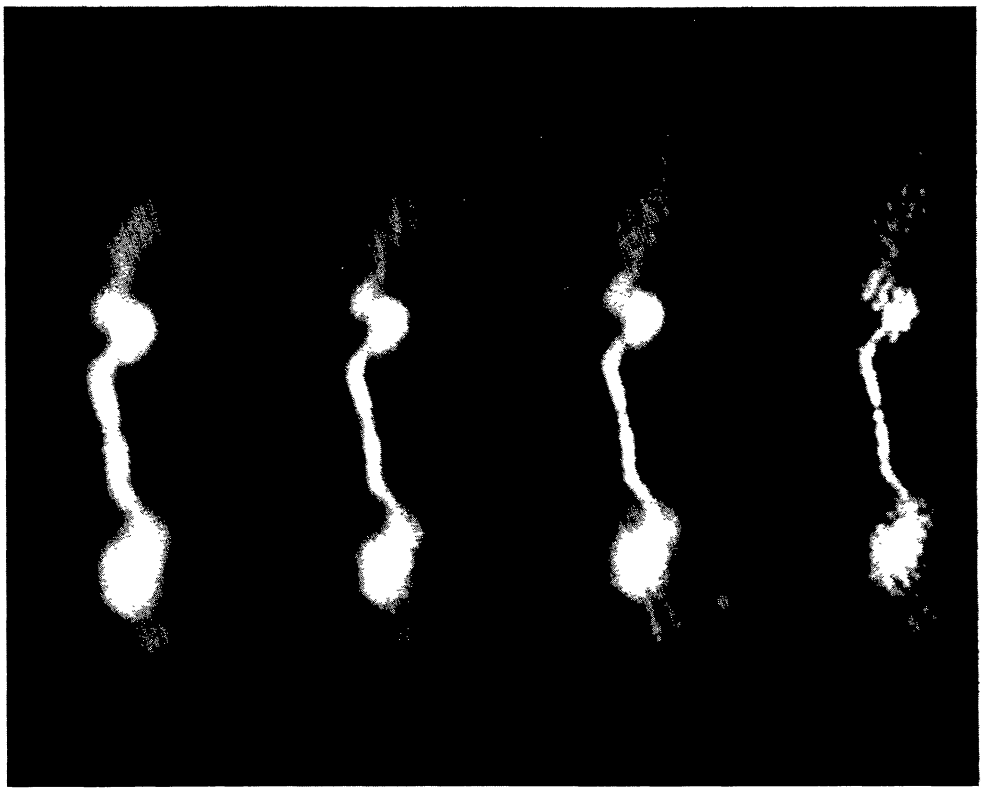


Fig. 2. Tests of superresolution (from left): CLEAN image made from C-configuration data only, VM image made from C-configuration only, CLEAN image made from B- and C-configuration data, CLEAN image made from C-configuration data only with CLEAN beam appropriate to B-configuration data

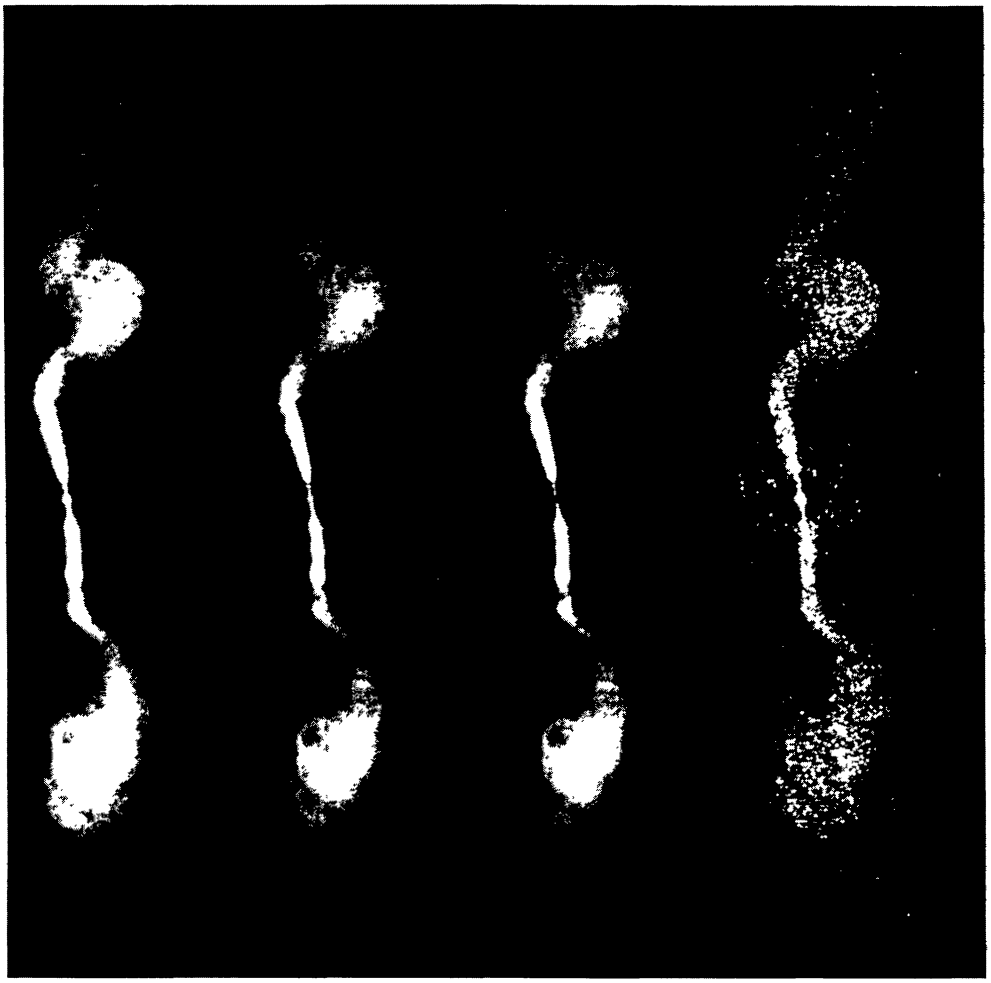


Fig. 3. Tests on very extended source (from left): CLEAN image, VM image, VM image after point source removed, difference between last two images (saturated a factor of 20 lower)

known. It works rather poorly on very high dynamic range images since the convergence is then slow. We note that these attributes are complementary to those of the CLEAN algorithm and so MEM will not replace CLEAN as the main deconvolution algorithm but rather will be used in those circumstances where CLEAN fails.

Our three examples of MEM images are designed to illustrate both the strengths and weaknesses of MEM and the VM algorithm.

First, we show the application of VM and the standard AIPS CLEAN program APCLN to an artificial data set. This simulates a VLA snapshot of a linear source somewhat resembling a jet in an extragalactic radio source. Figure 1a shows the model, the dirty image, the best CLEAN image (made using the stabilised CLEAN, see Cornwell, 1983), the best VM image, both raw and convolved with the CLEAN beam for comparison with the CLEAN image. The best VM image was made using the procedure described in Sect. 2: namely, the initial a priori image is flat, the resultant VM image is convolved with the CLEAN beam and then used as the a priori image for a final VM image. This final VM image produced in this way shows better noise suppression and less bias than the VM image made using a flat a priori image. Also, the finishing value of α is smaller since the a priori image closely resembles the final image. Figure 1b shows the errors made by CLEAN and VM in reconstructing the model (convolved with the CLEAN beam). It is interesting to see that the errors are comparable both in location (within the main brightness) and magnitude (about 5%).

Our second example shows the potential for superresolution in MEM images. The four scaled arrays of the VLA (Thompson et al., 1980) allow the testing of superresolution of real data: we make observations of a source in two adjacent configurations, the high resolution of the MEM image made from the smaller configuration data then allows comparison with a CLEAN image made from data collected with the larger configuration. Figure 2 shows the result of such a test on data on the radio source 3C449. The CLEAN image made from C-configuration data only shows far less resolution than the MEM image made from the same data. Comparison with a CLEAN image made using B-configuration data as well shows that the super resolution is reliable qualitatively. An attempt to super-resolve the C-configuration CLEAN image by using a CLEAN beam appropriate to the B-configuration results in very blobby structure both in the jets and the extended structure.

In this test the best MEM image took approximately 60 FFTs to converge whereas the CLEAN required about 10 major cycles, equivalent to about 15–20 FFTs.

Our final example concerns the application of VM to the deconvolution of an image spanning a very large number of

resolution elements. The image is of the radio source 3C449 observed in the VLA A, B, and C-configurations at a wavelength of 20 cm (Cornwell et al., in preparation). Figure 3 shows the CLEAN image, the best MEM image using a previous MEM image convolved as the a priori image, the latter with the point source removed prior to deconvolution, and the difference between the last two images. VM took about 50 FFTs to converge in each case.

The VM image made with the point source *not* removed has reconstructed the source structure reasonably well except in the neighborhood of the point source where sidelobes are still present. Removal of the point source prior to deconvolution considerably improves the image. Such hybrid approaches combine the best features of both CLEAN and MEM but require intervention to split the work.

6. Summary

We have described a simple algorithm, now implemented in AIPS, for performing a maximum entropy deconvolution of radio interferometric data. Tests show that for objects covering many resolution elements the performance of this algorithm can match or exceed that of the Clark CLEAN algorithm (Clark, 1980). A priori information about the image can be easily introduced in the form of an image which the reconstructed image will resemble as closely as the data allows. Tests of the new algorithm on real and simulated data indicate that some useful superresolution, by about a factor of 2 or 3, is possible.

Both the ease of coding VM and its presence in the AIPS package allow experimentation with Maximum Entropy Methods of deconvolution. We therefore wish to encourage others to experiment so that a wide base of practical knowledge about MEM and its application to astronomical imaging may be built up. That the CLEAN algorithm was only accepted after a similar process indicates the importance of such testing. We also feel that this may spur the development of new and improved deconvolution algorithms. Finally, we note that the analysis of MEM performed by Nityananda and Narayan (1982) indicates the usefulness of a pragmatic and non-information theoretic approach to MEM which may be fruitful in suggesting further avenues of research.

Acknowledgements. We thank the following people for many stimulating conversations on the theory and practice of image deconvolution: Ron Ekers, Martin Brown, Ramesh Narayan, Rajaram Nityananda, and Steve Gull. The AIPS group, and in particular Bill Cotton, provided the base of software upon which VM was built.

References

Ables, J.G.: 1974, *Astron. Astrophys. Suppl.* **15**, 383
Adby, P.R., Dempster, M.: 1974, Introduction to Optimisation Methods, Chapman and Hall, London
Bryan, R.K., Skilling, J.: 1980, *Monthly Notices Roy. Astron. Soc.* **191**, 69
Burch, S.F., Gull, S.F., Skilling, J.: 1983, *Computer Vision, Graphics and Image Processing* **23**, 113
Clark, B.G.: 1980, *Astron. Astrophys.* **89**, 377
Cornwell, T.J.: 1980, Ph.D. Thesis, University of Manchester
Cornwell, T.J.: 1983, *Astron. Astrophys.* **121**, 281
Cornwell, T.J.: 1984, in *Indirect Imaging, Proc. IAU/URSI Symp.*, ed. J.A. Roberts, Cambridge Univ. Press, p. 291
Cornwell, T.J., Wilkinson, P.N.: 1984, *Indirect Imaging, Proc. IAU/URSI Symp.*, ed. J.A. Roberts, Cambridge Univ. Press, p. 207
Davies, J.G., Anderson, B., Morison, I.: 1980, *Nature* **288**, 64
Dulk, G.A., McLean, D.J., Manchester, R.N., Ostry, D.I., Rogers, P.E.: 1984, in *Indirect Imaging, Proc. IAU/URSI Symp.*, ed. J.A. Roberts, Cambridge Univ. Press, p. 305
Frieden, B.R.: 1976, in *Proc. Conf. Image Analysis and Evaluation*, Toronto, Canada, p. 261
Frieden, B.R., Wells, D.C.: 1978, *J. Opt. Soc. America* **68**, 93
Gull, S.F., Daniell, G.J.: 1978, *Nature* **272**, 686
Gull, S.F., Skilling, J.: 1984, in *Indirect Imaging, Proc. IAU/URSI Symp.*, ed. J.A. Roberts, Cambridge Univ. Press, p. 267
Hogbom, J.: 1974, *Astron. Astrophys. Suppl.* **15**, 417
Narayan, R., Nityananda, R.: 1984, in *Indirect Imaging, Proc. IAU/URSI Symp.*, ed. J.A. Roberts, Cambridge Univ. Press, p. 281
National Research Astronomy Observatory Staff: 1982, The Very Long Baseline Array Radio Telescope
Nityananda, R., Narayan, R.: 1982, *J. Astrophys. Astron.* **3**, 419
Pearson, T.J., Readhead, A.C.S.: 1984, *Ann. Rev. Astron. Astrophys.*
Rosencrantz, R.D. (ed.): 1983, E.T. Jaynes: Papers on probability, statistics, and statistical physics, D. Reidel
Schwab, F.R.: 1980, *SPIE* **231**, 18
Schwarz, U.J.: 1978, *Astron. Astrophys.* **65**, 345
Shore, J.E., Johnson, R.W.: 1980, *IEEE* IT-**26**, 26
Skilling, J., Gull, S.F.: 1984, (preprint)
Steer, D.G., Dewdney, P.E., Ito, M.R.: 1984, *Astron. Astrophys.* (in press)
Thompson, A.R., Clark, B.G., Wade, C.M., Napier, P.J.: 1980, *Astrophys. J. Suppl.* **44**, 151
Wernecke, S.J., D'Addario, L.R.: 1976, *IEEE* C-**26**, 351