Astron. Astrophys. 121, 281–285 (1983)

ASTRONOMY AND ASTROPHYSICS

A method of stabilizing the clean algorithm

T. J. Cornwell

National Radio Astronomy Observatory* P.O. Box 0, Socorro, NM 87801, USA

Received September 9, accepted December 15, 1982

Summary. The clean algorithm, used widely in radioastronomy for the deconvolution of Fourier synthesis point spread functions from images, possesses a well-known instability concerning the estimation of the visibility function in unsampled regions of the u, v-plane. A simple and inexpensive modification to the clean algorithm which stabilizes it is described and an example of its use in the reduction of VLA data is given.

Key words: data processing – CLEAN technique

1. Indroduction

The one major drawback to the clean algorithm (e.g. Högbom, 1974) is its poor reconstruction of regions of extended emission. Among its advantages are speed (especially in the Clark implementation, see Clark 1980), relative simplicity, both of coding and of use, and suitability of use with the various selfcalibration procedures now in widespread use (e.g. Schwab, 1980; Cornwell and Wilkinson, 1981). More sophisticated image reconstruction algorithms such as the Maximum Entropy Method (Wernecke and D'Addario, 1976; Gull and Daniell, 1978; Lim and Malik, 1981) possess none of clean's advantages but do perform better in the reconstruction of extended emission since a bias towards smooth structure is explicitly included. In this paper I will describe a simple, easily coded, modification to the clean algorithm which biases it towards smoother structure and thus stabilizes the reconstruction of extended emission. Other deconvolution algorithms such as the Gerchberg-Saxton (Gerchberg, 1974) algorithm may be stabilized using the same approach.

2. The clean algorithm

Clean utilises an iterative point source subtraction technique to minimize an error term which, in the u, v-plane, can be written as:

$$Q = \sum_{i=1}^{M} w_i |V_i - \hat{V}_i|^2 , \tag{2.1}$$

where V_i is the observed, calibrated visibility at the i^{th} sample point, $\hat{V}_i$ is the predicted visibility at the i^{th} sample point, and w_i is the weight attached to i^{th} sample point, and M is the number of visibility points.

$\hat{V}$ is the Fourier transform of a model in the image plane; in the clean algorithm the model is a grid of point sources arranged at the vertices of a lattice having angular coordinates x, y. Thus,

$$\hat{V}_i = \sum_{k=1}^{N} I_k e^{j2\pi(v_i x_k + v_i y_k)} , \tag{2.2}$$

where I_k is the flux of the k^{th} component and there are N image points in all.[1]

In the image plane Q can be written:

$$Q = \sum_{k=1}^{N} \sum_{l=1}^{N} B_{kl} I_k I_l - 2 \sum_{k=1}^{N} D_k I_k + \sum_{i=1}^{M} w_i |V_i|^2 , \tag{2.3}$$

where the "dirty" image is given by:

$$D_k = \frac{\sum_{i=1}^{M} w_i \, \mathrm{Re}(V_i e^{-j2\pi(v_i x_k + v_i y_k)})}{\sum_{i=1}^{M} w_i} \tag{2.4}$$

and the "dirty" beam (or point spread function) is given by:

$$B_{kl} = \frac{\sum_{i=1}^{M} w_i \cos(2\pi(v_i(x_k - x_l) + v_i(y_k - y_l)))}{\sum_{i=1}^{M} w_i} . \tag{2.5}$$

Choosing the I_k to minimize Q we find that the solution must obey the convolution equation:

$$\sum_{l=1}^{N} B_{kl} I_l = D_k . \tag{2.6}$$

This equation, which must be obeyed in the limit of vanishing noise, simply expresses the condition that the model image when convolved with the dirty beam should yield the dirty image. In the

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

1 One unusual aspect of this formulation is that in most radio astronomical applications $N > 2M$ and so that, in contrast to more typical applications of the least squares method, Q could be reduced exactly to zero. However, this somewhat perverse formulation is helpful in describing our proposed changes to the clean algorithm

regime where $N>2M$ there are infinitely many solutions to Eq. (2.6). Clean chooses one possible solution by decomposing the source into a number of point sources spread over a finite region of the image plane (see e.g. Schwarz, 1978).[2]

3. New algorithms based upon clean

In deriving the convolution equation, which clean solves, we made no mention of the properties we expect of a image of the sky. However, clean does tend to produce images which consist of unresolved components in an otherwise blank field of view. Such a bias is built into clean, not into the convolution equation, so that if we changed the convolution equation so as to bias it towards smoother sources the clean would find smoother images. This is the approach that we will adopt. Instead of simply minimizing Q we will minimize a linear combination, J, of Q and a function, $\mathscr{H}(I)$, which is large for smooth images

$$J=Q-2\alpha\mathscr{H}(I), \tag{3.1}$$

where the variable α controls the balance between fitting the data and obtaining a smooth image.

The image which minimizes J for a given α is given by:

$$\sum_{l=1}^{N} B_{kl}I_l=D_k+\alpha\frac{\partial\mathscr{H}}{\partial I_k}. \tag{3.2}$$

In all interesting cases $\mathscr{H}(I)$ will be non-linear in I. Two different schemes arise for the solution of the resulting equations.

Scheme 1

We can alter the dirty beam so that Eq. (3.2) can be expressed as a convolution equation:

$$\sum_{l=1}^{N} B'_{kl}(I)I_l=D_k, \tag{3.3}$$

where the modified beam is:

$$B'_{kl}(I)=B_{kl}-\frac{\alpha\dfrac{\partial\mathscr{H}}{\partial I_k}\delta_{kl}}{I_k}. \tag{3.4}$$

If $\mathscr{H}(I)$ is quadratic in the I_k then $\partial\mathscr{H}/I_k$ will be linear and consequently this modified beam is not dependent on the final image (except weakly on the value of α). We can then just clean, as usual, with this modified dirty beam. If $\mathscr{H}(I)$ is not quadratic then we should use an iterative approach to find the modified beam:

1. Clean to find solution to ordinary convolution equation. Select this as the current estimate of I.
2. Calculate $\partial\mathscr{H}/\partial I_k$ from the current estimate of I and add it to the dirty beam to obtain the modified beam written in Eq. (3.4). Note that the beam will be shift-variant.
3. Solve, using clean, the modified convolution Eq. (3.3) to obtain the current estimate of I.
4. Repeat Steps 2 and 3 until convergence is attained.

The possible shift-variance of the beam means that Scheme 1 is only practical for quadratic forms of $\mathscr{H}(I)$.

2 D'Addario (1976) has noted that other decompositions, such as into a sum of Gaussians each of fixed shape, are possible and may, in some circumstances, produce superior results

Scheme 2

In all other cases, we can instead modify the dirty image to obtain a convolution equation:

$$\sum_{l=1}^{N} B_{kl}I_l=D'_k(I), \tag{3.5}$$

where the modified dirty image is:

$$D'_k(I)=D_k+\alpha\frac{\partial\mathscr{H}}{\partial I_k}. \tag{3.6}$$

The term $\partial\mathscr{H}/\partial I_k$ must be estimated iteratively. We propose the following scheme to solve for the image I:

1. Clean to find solution to ordinary convolution equation. Select this as the current estimate of I.
2. Calculate $\partial\mathscr{H}/\partial I_k$ from the current estimate of I and add it to the dirty image to obtain $D'_k(I)$.
3. Solve, using clean, the modified convolution Eq. (3.5) to obtain the current estimate of I.
4. Repeat Steps 2 and 3 until convergence is attained.

The value of Q for the final image will depend upon the value of α used. We will define the optimum value of α, α_{opt}, as that which makes Q equal to its expected value. A search method could be used to find α_{opt} but there is one argument which gives an approximate value. Multiplying Eq. (3.2) by I_l and summing over l we obtain:

$$Q+\left(\sum_{k=1}^{N}\sum_{l=1}^{N}B_{kl}I_kI_l-\sum_{i=1}^{M}w_i|V_i|^2\right)=2\alpha\sum_{k=1}^{N}I_k\frac{\partial\mathscr{H}}{\partial I_k}. \tag{3.7}$$

The difference on the left hand side is related to the discrepancy in signal-to-noise between observed and model visibilities. We will assume that for a reasonably unbiased model image this will vanish. If the expected value of Q is σ^2 per pixel then:

$$\alpha_{\text{opt}}=\frac{\sigma^2}{2\left\langle I_k\dfrac{\partial\mathscr{H}}{\partial I_k}\right\rangle}, \tag{3.7}$$

where $\langle\ldots\rangle$ denotes the average value.

The only missing ingredient is the measure of smoothness $\mathscr{H}(I)$; this we now discuss.

4. Measures of smoothness

It is helpful to sub-divide measures of smoothness according to whether they require totally positive images. Of those requiring positivity the entropy measures are simplest:

$$\mathscr{H}_1=-\sum_{k=1}^{N}I_k\ln(I_k), \tag{4.1}$$

$$\mathscr{H}_2=\sum_{k=1}^{N}\ln(I_k). \tag{4.2}$$

Note also their cousins, the $\mathscr{H}'_i$ formed by normalizing I_k with respect to the total flux:

$$I_{\text{total}}=\sum_{k=1}^{N}I_k. \tag{4.3}$$

Of those not requiring positivity the smoothness S is simplest:

$$S=-\sum_{k=1}^{N}I_k^2. \tag{4.4}$$

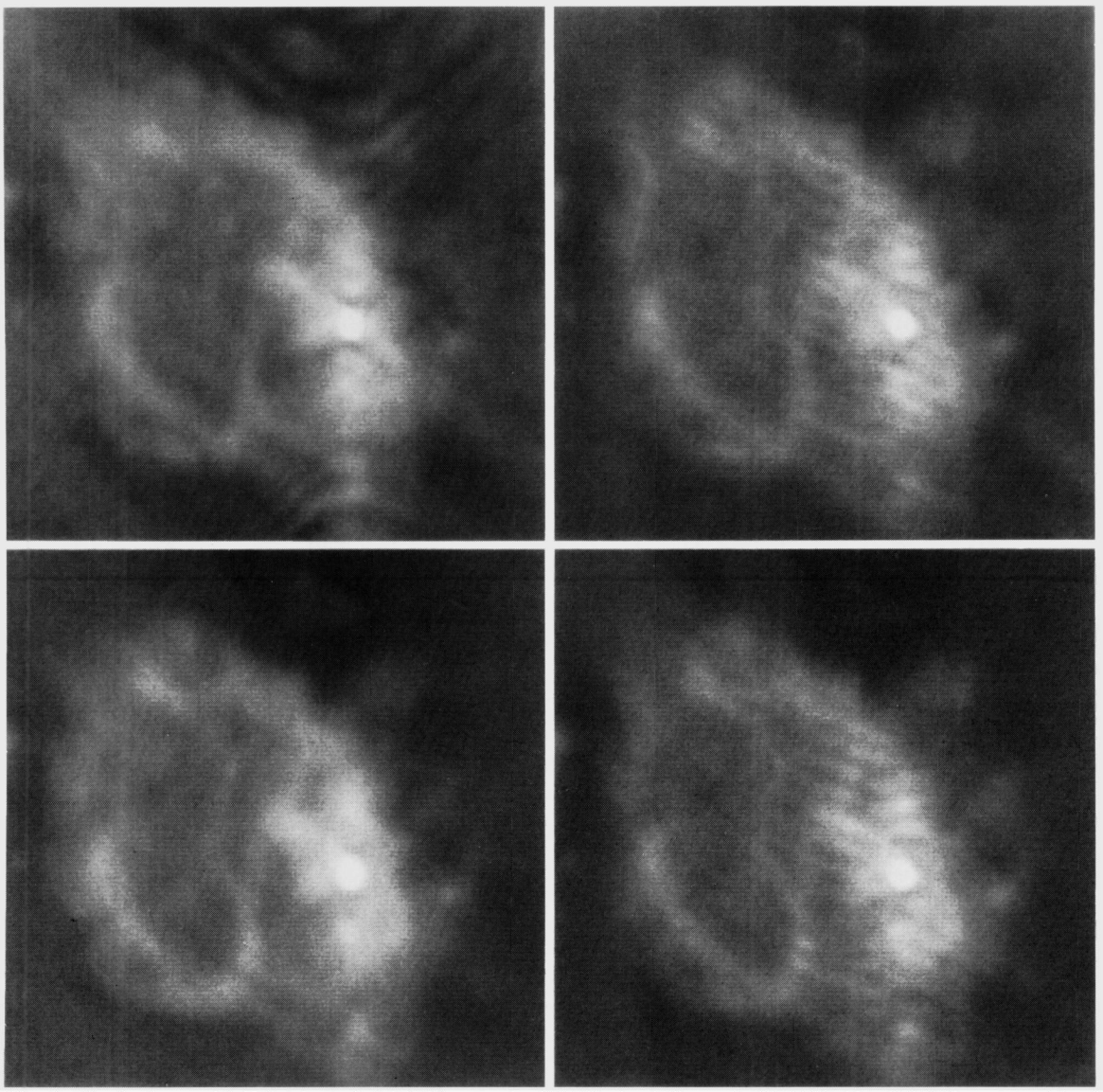

Fig. 1. Clockwise, starting from top left: *1* Dirty map of SGR-A at 20 cm. *2* CLEAN map, *3* ESC map $\sigma = 15$ mJy/beam, *4* SSC map $\alpha = 5\%$

Wernecke (1978) has formalized the meaning of smoothness preference as used here, showing that the application of any linear smoothing operator to a maximally smooth map will not decrease Q. He also demonstrates that for $\mathscr{H}_1$ and $\mathscr{H}_2$ a global minimum of J exists so that uniqueness of the image is guaranteed.

The importance of the positivity constraint is that heuristic arguments can be advanced to show that an increase in resolution is expected (Frieden, 1978). However, in the case of polarization images the positivity constraint does not apply.[3]

For our purposes, the most convenient measures of smoothness are $\mathscr{H}_2$ and S. Since the measure S is quadratic in I_k, Scheme 1 can be used. We have that:

$$\alpha_{\text{opt}} = \frac{\sigma^2}{2\langle I_k^2 \rangle}. \tag{4.5}$$

3 Ponsonby (1974) has derived an polarization entropy measure related to the $\mathscr{H}_2$ measure. This is obtained by replacing I_k in Eq. (4.2) by $I_k^2 - Q_k^2 - U_k^2 - V_k^2$ where I, Q, U, V are the Stokes parameters. Scheme 2 cluod thus be used but we would have to deal with four images

If the sidelobes are reasonably small then $\langle I_k^2 \rangle$ can be estimated from the dirty image, otherwise an initial clean image can be used. The modified beam is:

$$B'_{kl} = B_{kl} - \frac{\sigma^2 \delta_{kl}}{2\langle I_k^2 \rangle}. \tag{4.6}$$

By Parseval's theorem maximization of S is equivalent to the minimization of the power in the visibility (i.e. $|\hat{V}|^2$ at the unsampled u, v-points); thus, with no extra contraints, the solution is simply the Principal Solution (Bracewell and Roberts, 1953). We may, however, require that the convolution equation be solved by using clean; this will rule out the Principal Solution. Hereafter we will call the resulting algorithm the Smoothness Stabilised clean (SSC).

Since $\mathscr{H}_2$ is not quadratic we must use Scheme 2 (see Sect. 3) to find the solution; we find that α_{opt} is independent of the image:

$$\alpha_{\text{opt}} = \frac{\sigma^2}{2}. \tag{4.7}$$

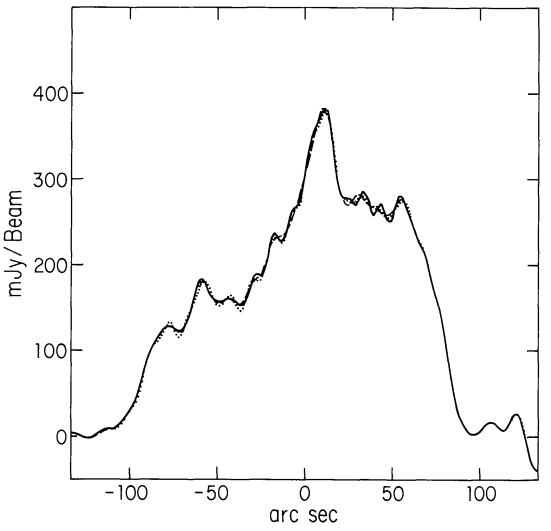

Fig. 2. Slice through the clean (solid line), SSC (dashed line), and ESC (dotted line) maps in Fig. 1. Position angle is $\sim 30°$

One problem is that clean images may possess regions of negative brightness whereas $\mathscr{H}_2$ demands positivity. Our approach has been to simply truncate below the $\sim 1\sigma$ level at Part 3 of Scheme 2; but note that other, more sophisticated, approaches, such as variable damping, have proved successful in a related problem (Lim and Malik, 1981). We will call this $\mathscr{H}_2$ maximizing algorithm, based on Scheme 2, the Entropy Stabilized clean (ESC). For completeness we give the equivalent modified dirty beam:

$$B'_{kl}(I) = B_{kl} - \frac{\sigma^2 \delta_{kl}}{2I_k^2}. \tag{4.8}$$

Two points concerning this modified dirty beam are worth noting: first, it is manifestly shift variant and secondly, it is very closely related to the modified beam for the SSC [Eq. (4.6)].

Both SSC and ESC will produce maps with varying resolution. For example, on a point source the effective beam of an SSC map will be larger than that of a CLEAN map. The degree of smearing depends upon α; too large a value will cause the resulting map to violate the ordinary convolution equation. However, both SSC and ESC are best suited to extended sources where this effect will not be important.

We have not considered the use of the normalized smoothness measures; these are easily utilized and should alleviate some problems connected with estimation of short spacing information.

5. How do these methods work?

It is clear that if spurious features are to be present in a final cleaned map then the clean algorithm must have introduced, at some stage, a small perturbation for which there is no evidence in the data. In subsequent iterations of the clean algorithm the perturbation may become amplified and appear as noticeable artifacts. In the stabilized clean the convex nature of the smoothness measures ensures that, as soon as such an error is made, a small perturbation of opposite phase is added to the dirty image and so propagates into the solution. Therefore, if $\mathscr{H}(I)$ is convex, the algorithm is stable. As a corollary, if $\mathscr{H}(I)$ is concave then the algorithm is unstable and small perturbations will be amplified. Thus, both these methods use feedback methods to stabilize the clean algorithm against small perturbations. For SSC we can go somewhat further and describe the convergence requirements: Schwarz (1978) has shown that clean will converge if the transform of the dirty beam is nowhere negative. If λ is the smallest Fourier coefficent of the dirty beam then we require that $\lambda \geqq -\alpha_{\text{opt}}$. For the SSC this condition must always be satisfied although for a smoothness minimizing algorithm α_{opt} would be less than zero and thus convergence may be impossible. One further practical point should be noted: the element of caution involved in the feedback slows both SSC and ESC considerably so that many more point source subtractions are needed to reach a given flux level.

Finally we should clarify the role of the constraints built into SSC and ESC. The constraints are:

a) Maximization of a smoothness measure (SSC only).
b) Maximization of an entropy measure (ESC only).
c) Fitting of the data.
d) Finite size of the image.
e) Point source decomposition.

These constraints have varying effects in different combinations; for example, as discussed in the previous section, under constraints a) and c) the expected map is merely the principal solution which violates constraint d). Clean relies on c), d), and e) but, as we have seen, does not necessarily produce smooth maps. In some circumstances constraint d) may be sufficient to uniquely determine the image but may be difficult to employ computationally.

6. Examples

A good example of the instabilities that can afflict clean and of their treatment using these stabilized cleans is available from VLA C-configuration observations of SGR-A made at 20 cm by Ekers et al. (1982). In Fig. 1 we show radiographs of the various maps made from this data. The dirty map shows the structure reasonably well but some beam patterns are visible. A Clark-type clean made with a loop gain of 10% and 10,000 clean components shows multiplicative fringes running through the map at p.a. $\sim 30°$. Changing the values of the loop gain, the number of clean components or the clean box size has negligible effect upon the fringes. The fringes can also be discerned in a slice through the map at the appropiate position angle (see Fig. 2). For the strongest features on this slice the amplitude of the fringe is about 5%. Examination of the Fourier transform of this clean map indicates that the fringe has a spatial frequency lying in a un-sampled region of the u, v-plane; therefore clean has not violated a data constraint but rather has poorly extrapolated the measured visibility function. An SSC map made with the same parameters as the clean map but with α set to 5% is also shown in Fig. 1 and the corresponding slice is also shown in Fig. 2. The fringe has now decreased to a much lower level and is just barely distinguishable. An ESC map with $\sigma = 15$ mJy/beam is also displayed in Fig. 1. Again the corresponding slice is shown in Fig. 2. Three passes through Scheme 2 were necessary to form this image. The fringe is rather poorly suppressed. This comparison accurately represents the typical efficacy of both ESC and SSC, the latter nearly always proving superior. The successive images calculated by ESC seem to oscillate converging slowly to some mean, the amplitude of oscillations being dependent on the value of σ used whereas SSC seems relatively insensitive to the actual value of α used. For this

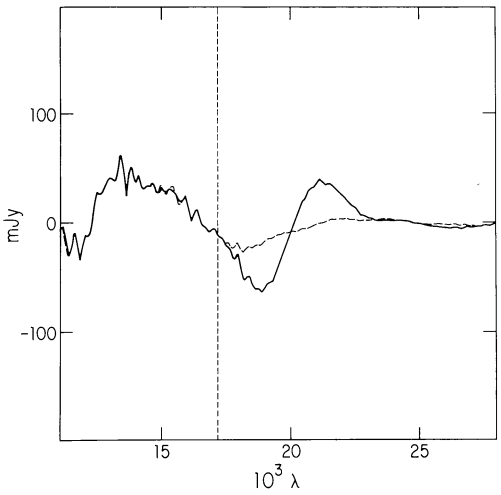

Fig. 3. Slices through the imaginary part of the Fourier transform of the clean (solid line) and SSC (dashed line) maps. The position angle is as in Fig. 2. The vertical dashed line indicates the edge of the sampled u, v-plane at this position angle

example, increasing the value of α to 100% only changes the peak flux in the map by about 4% and has no noticeable effect on the quality of the resultant SSC map. For both ESC and SSC, the peak flux density and the total flux decrease and the smoothness increases as α is increased.

A graphic demonstration of the effect of SSC is to be seen in Fig. 3 where, for both CLEAN and SSC maps, slices through the imaginary part of the Fourier transform of the map are shown, the position angle being that of the stripes. The edge of the sampled u, v-plane is indicated by the vertical line. It may be seen that the extrapolation performed by SSC decays to zero faster than that due to the ordinary CLEAN; this is in accordance with the minimization of the dispersion of the image performed by SSC since Parseval's Theorem says that, equivalently, the dispersion in the extrapolated visibilities is minimized.

7. Discussion

We have shown that the clean algorithm can be stabilized so that the multiplicative fringes sometimes found in the clean map of extended regions of emission can be avoided. Two quite different algorithms were described, one based on iterative modification of the dirty map used in clean and the other based upon single pass modification of the dirty beam. The latter, here called Smoothness Stabilized clean, seems to be superior in speed, simplicity, insensitivity to the control parameter α, and applicability to the mapping of polarization distributions. The lack of widespread experience with SSC suggests that initial results should be regarded with some scepticism. However, to encourage experimentation we emphasize that a SSC algorithm can be obtained by a trivial modification of a clean algorithm.

Other deconvolution algorithms can also be improved in the same way; for example, the Gerchberg-Saxton algorithm (Gerchberg, 1974) could be easily modified to include maximization of the smoothness. Finally, it may be argued that the improvements to CLEAN suggested here are undesirable in that one more degree of freedom is added to the algorithm; however in view of the multiplicity of degrees of freedom in the conventional algorithm this consideration may not be important.

Acknowledgements. I am grateful to Ron Ekers and Larry D'Addario for a number of useful discussions.

References

Bracewell, R.N., Roberts, J.A.: 1954, *Australian J. Phys.* **7**, 615
Clark, B.G.: 1980, *Astron. Astrophys.* **89**, 377
Cornwell, T.J., Wilkinson, P.N.: 1981, *Monthly Notices Roy. Astron. Soc.* **196**, 1067
D'Addario, L.R.: 1976, (unpublished notes)
Ekers, R.D., Goss, W.M., Schwarz, U.J.: 1982 (in preparation)
Frieden, B.R.: 1978, *J. Opt. Soc. Am.* **62**, 511
Gerchberg, R.W.: 1974, *Optica Acta* **21**, 709
Gull. S.F., Daniell, G.J.: 1978, *Nature* **272**, 686
Högbom, J.: 1974, *Astron. J. Suppl.* **15**, 417
Lim, J.S., Malik, N.A.: 1981, *Proc. IEEE, ASSP-29*, 401
Ponsonby, J.E.B.: 1974, *Monthly Notices Roy. Astron. Soc.* **163**, 369
Schwab, F.R.: 1980, *Proc. SPIE* **231**, 18
Schwarz, U.J.: 1978, *Astron. Astrophys.* **65**, 345
Wernecke, S.J.: 1978, *Radio Science* **12**, 831
Wernecke, S.J., D'Addario, L.R.: 1976, *Proc. IEEE, C-26*, 251–364