Deconvolution Tutorial

Tim Cornwell & Alan Bridle
National Radio Astronomy Observatory

Text last updated: 04 November 1996
HTML version available at:
`http://www.cv.nrao.edu/~abridle/deconvol/deconvol.html`

Contents

1 Purpose **2**

2 Introduction **2**

3 Solutions of the convolution equation **3**
3.1 The "principal solution" and "invisible distributions" 4
3.2 Problems with the principal solution 4

4 The 'CLEAN' algorithm **5**
4.1 The Högbom algorithm 5
4.2 The Clark algorithm 6
4.3 The Cotton–Schwab algorithm 6
4.4 Other algorithms related to 'CLEAN' 7
4.5 Practical details and problems of 'CLEAN' 7
4.6 The use of boxes: finite support 8
4.7 Number of iterations and the 'CLEAN' loop gain 9
4.8 The problem of short spacings 9
4.9 The 'CLEAN' beam 10
4.10 Use of *a priori* models in 'CLEAN' 11
4.11 Non-uniqueness of 'CLEAN' images 11
4.12 Instabilities 11

5 The Maximum Entropy Method (MEM) **12**
5.1 The default image (prior distribution) 13
5.2 Total flux density 13
5.3 Varying resolution 13
5.4 Bias 13
5.5 Point sources in extended emission 14

6 Comparing 'CLEAN' and MEM **14**

7 Algebraic deconvolution **15**
7.1 Singular value decomposition 15
7.2 Non-negative least-squares 16

8 Other methods, including hybrids **16**

9 References **17**

1 Purpose

The 'dirty' image formed directly by Fourier transforming visibility data from practical interferometric arrays has defects imposed by limited sampling of the u,v plane.

This tutorial describes how such visibility data can be processed to reduce these defects using deconvolution algorithms. It is based on lectures originally given in the NRAO Summer Schools on *Synthesis Imaging*.

2 Introduction

An interferometric array samples the complex visibility function $V(u,v)$ of the sky at points in the u,v plane. Under approximations that are valid for a sufficiently small sources in an otherwise empty sky, the visibility function $V(u,v)$ is related to the angular distribution of the source intensity $I(l,m)$ (multiplied by the primary beam of the array elements) through a two-dimensional Fourier transform:

$$V(u,v) = \int\int_S I(l,m)e^{-2\pi i(ul+vm)}\,dl\,dm\,, \tag{1}$$

where S denotes integration over the whole sky.

Practical arrays provide only a finite number of noisy samples of the visibility function $V(u,v)$, so $I(l,m)$ cannot be recovered directly. Instead, $I(l,m)$ must be estimated either from a model with a finite number of parameters, or from a non-parametric approach.

For radio astronomical imaging, a convenient (and sometimes realistic) model of the source intensity is a 2-d grid of δ-functions whose strengths are proportional to the intensity. The model can be thought of as a 'bed of nails' with strengths $\widehat{I}(p\Delta l, q\Delta m)$, where Δl and Δm are the element separations on a grid in two orthogonal sky coordinates. The visibility $\widehat{V}$ predicted by this model is given by

$$\widehat{V}(u,v) = \sum_{p=1}^{N_l}\sum_{q=1}^{N_m} \widehat{I}(p\Delta l, q\Delta m)e^{-2\pi i(pu\Delta l+qu\Delta m)}\,. \tag{2}$$

For simplicity we notate the discrete model $\widehat{I}(p\Delta l, q\Delta m)$ as $\widehat{I}_{p,q}$. Assuming reasonably uniform sampling of a region of the u,v plane, one can expect to estimate source features with widths ranging from $\mathcal{O}(1/\max(u,v))$ up to $\mathcal{O}(1/\min(u,v))$. The grid spacings, Δl and Δm, and the number of pixels on each axis, N_l and N_m, must be chosen so that all these scales can be represented. In terms of the range of u,v points sampled, the requirements are:

1. $\Delta l \leq \frac{1}{2u_{\max}}$,
2. $\Delta m \leq \frac{1}{2v_{\max}}$,
3. $N_l\Delta l \geq \frac{1}{u_{\min}}$, and

4. $N_m \Delta m \geq \frac{1}{v_{\min}}$.

The model has $N_l N_m$ free parameters: the flux densities $\widehat{I}_{p,q}$ in each cell. The measurements constrain the model such that at the sampled u,v points

$$V(u_r, v_r) = \widehat{V}(u_r, v_r) + \epsilon(u_r, v_r)\,, \tag{3}$$

where $\epsilon(u_r, v_r)$ is a complex, normally-distributed random error due to receiver noise, and r indexes the samples.

At points in the u,v plane where no sample was taken, the transform of the model can have any value without conflicting with the data. One can think of Equation 3 as a multiplicative relation

$$V(u, v) = W(u, v)\big(\widehat{V}(u, v) + \epsilon(u, v)\big)\,, \tag{4}$$

where $W(u, v)$ is a weighted sampling function which is non-zero only where we have samples in the u,v plane,

$$W(u, v) = \sum_r W_r \delta(u - u_r, v - v_r)\,. \tag{5}$$

By the convolution theorem, this corresponds to a convolution relation in the image plane:

$$I^D_{p,q} = \sum_{p',q'} B_{p-p',q-q'} \widehat{I}_{p',q'} + E_{p,q}\,, \tag{6}$$

where

$$I^D_{p,q} = \sum_r W(u_r, v_r)\,\mathrm{Re}\left(V(u_r, v_r) e^{2\pi i (p u_r \Delta l + q v_r \Delta m)}\right) \tag{7}$$

and

$$B_{p,q} = \sum_r W(u_r, v_r)\,\mathrm{Re}\left(e^{2\pi i (p u_r \Delta l + q v_r \Delta m)}\right)\,. \tag{8}$$

$E_{p,q}$ in Equation 6 is the noise image obtained by replacing V in Equation 7 by $\epsilon(u_r, v_r)$. Note that the $B_{p,q}$ given by Equation 8 is the point spread function (beam) that is synthesized after all weighting has been applied (and after gridding and grid correction if an FFT was used; to keep the notation concise, the gridding and grid correction are not explicitly included). The Hermitian nature of the visibility has been used in this rearrangement.

Equation 6 represents the constraint that the model $\widehat{I}_{p,q}$, when convolved with the point spread function $B_{p,q}$ (also known as the *dirty beam*) corresponding to the sampled and weighted u,v coverage, should yield $I^D_{p,q}$ (known as the *dirty image*).

The weight function $W(u, v)$ can be chosen to favor certain aspects of the data. For example, setting $W(u_r, v_r)$ to the reciprocal of the variance of the error in $V(u_r, v_r)$ optimizes the signal-to-noise ratio in the final image. Setting $W(u_r, v_r)$ to the reciprocal of some approximation of the local density of samples minimizes the sidelobe level.

3 Solutions of the convolution equation

We now consider whether the convolution equation has a unique solution.

3.1 The "principal solution" and "invisible distributions"

Consider a situation in which some spatial frequencies that are present in the source model are not sampled by the data. The fit of the model to the data is unaffected by changing the amplitudes of the sinusoids corresponding to these frequencies.

The dirty beam filters out these un-sampled spatial frequencies: if Z is an intensity distribution containing only such frequencies, then $B * Z = 0$. Thus, if I is a solution of the convolution equation, then so is $I + \alpha Z$ where α is any number. As usual, the existence of homogeneous solutions implies the non-uniqueness of any solution, *in the absence of boundary conditions.*

In interferometry, the solution in which all the un-sampled spatial frequencies have zero amplitude is called the "principal" solution. It is useful to think of the homogeneous solutions, or "invisible distributions" (Bracewell & Roberts 1954), as originating via two main shortcomings of our u,v coverage:

1. the coverage extends only out to a finite spatial-frequency limit, and
2. there are holes in the coverage.

Invisible distributions of the first sort correspond to finer detail than can be resolved. We deal with these by accepting a finite-resolution image as our final product. The most vexing problems of image construction come from finding *plausible* invisible distributions of the second sort to merge with the principal solution. To see why we need to do this, consider the shortcomings of the principal solution.

3.2 Problems with the principal solution

If the visibility data $V(u, v)$ are obtained on a regular grid, then the principal solution can easily be computed: one simply chooses the weight function W in Equation 7 to correct the bias in weight due to any vagaries of sampling. For each grid point, the visibility samples are summed with the appropriate weights, and the total weight is normalized to unity. Using such *uniform weighting*, the principal solution is the same as the dirty image: the convolution of the true brightness distribution with the dirty beam.

For most synthesis arrays now used in radio astronomy, the dirty beam has sidelobes in the range 1% to 10%. Sidelobes represent unavoidable ambiguity about the true distribution of emission in the dirty image. There are two ways to resolve this ambiguity:

1. make further observations to improve the sampling,
2. incorporate *a priori* information, e.g. about the extent of the emission on the sky, about positivity where appropriate, or about an upper bound to the degree of polarization, etc.

For example, consider uniformly weighted observations of a point source: the dirty image is the dirty beam centered on the point source position. Without *a priori* information we cannot distinguish whether the source is truly a point, or is shaped like the dirty beam. Of course we know that the Stokes parameter I must be positive, and that most radio sources are much better localized than dirty beams (they certainly do not have sidelobe patterns extending to infinity). A further unsatisfactory aspect of the principal solution, besides its implausibility, is that it changes (sometimes drastically) as more visibility data are added. A more stable estimator is obviously desirable.

The key to "successful" deconvolution is to make good use of such *a priori* information as positivity and the extent (support) of the radio source to steer the choice of invisible distributions to add while constructing the image. This process can also be thought of as specifying plausible boundary conditions while solving the deconvolution equation.

Image deconvolution in radio astronomy has so far been dominated by two algorithms, 'CLEAN' and the Maximum Entropy Method (MEM), which solve the convolution equation by placing strikingly different constraints on the invisible distributions. This tutorial focuses on these dominant approaches, and on a third that has recently shown particular promise for VLBI imaging: direct algebraic solution of the convolution equation.

4 The 'CLEAN' algorithm

The 'CLEAN' algorithm, introduced by Högbom (1974), assumes that the radio sky can be represented by a small number of point sources in an otherwise empty field of view. It uses a simple iterative procedure to find the positions and strengths of these sources. The final deconvolved 'CLEAN' image is the sum of:

1. these point-source 'CLEAN components' reconvolved ("restored") with a 'CLEAN beam' (usually Gaussian) to de-emphasize the higher spatial frequencies which are often spuriously extrapolated and,
2. (optionally, but strongly recommended) an image representing residual differences between the point-source model and the data.

We now describe some variants of the 'CLEAN' algorithm, including two that can be used on large images.

4.1 The Högbom algorithm

This algorithm proceeds as follows:

1. Find the strength and position of the peak (i.e., of the greatest absolute intensity) in the dirty image $I^D_{p,q}$.
2. Subtract from the dirty image, at the position of the peak, the dirty beam B multiplied by the peak strength and a damping factor γ (≤ 1, usually termed the *loop gain*).
3. Go to (1) unless any remaining peak is below some user-specified level. The search for peaks may be constrained to specified areas of the image, called *'CLEAN' windows*.
4. Convolve the accumulated point source model $\widehat{I}_{p,q}$ with an idealized 'CLEAN' beam (usually an elliptical Gaussian fitted to the central lobe of the dirty beam).
5. Add the residuals of the dirty image to the 'CLEAN' image.

The fifth step is sometimes omitted, but it is recommended because it can provide useful diagnostic information about the noise on the image, residual sidelobes, "bowls" near the image center (see Section 4.8 below).

4.2 The Clark algorithm

Much of the computation in 'CLEAN' consists of shifting and scaling the dirty beam. As this is essentially a convolution it may, in some circumstances, be done more efficiently with 2-d FFTs. Clark's (1980) 'CLEAN' algorithm does this, finding approximate positions and strengths of the point components using only a small patch of the dirty beam.

In detail, the Clark algorithm has two cycles, known as "minor" and "major" cycles. The *minor cycle* proceeds as follows:

1. A beam patch (a segment of the discrete representation of the beam) is selected to include the highest exterior sidelobe.
2. Points are selected from the dirty image if they have an intensity, as a fraction of the image peak, greater than the highest exterior sidelobe of the beam.
3. A Högbom 'CLEAN' is performed using the beam patch and the selected points of the dirty image. The stopping criterion for the 'CLEAN' is roughly such that any remaining points would not be selected in step (2).

The algorithm then does a *major cycle* wherein the point source model found by the minor cycle is transformed via an FFT, multiplied by the weighted sampling function (inverse transform of the beam), transformed back, and subtracted from the dirty image. Errors introduced in a minor cycle by the beam patch approximation are, to some extent, corrected in subsequent minor cycles.

4.3 The Cotton–Schwab algorithm

Cotton & Schwab (Schwab 1984b, top right corner of p. 1078!) developed a variant of the Clark algorithm whereby the major cycle subtracts 'CLEAN' components from the *un-gridded* visibility data. Aliasing noise and gridding errors can thus be removed if the inverse Fourier transform of the 'CLEAN' components to each u,v sample is accurate enough. Two routes are used for the inverse transform:

1. for small numbers of 'CLEAN' components, a 'direct Fourier transform' is performed so the accuracy is limited by the precision of the arithmetic;
2. for a large number of 'CLEAN' components, an FFT is used for efficiency, but inevitably some errors are introduced in interpolating from the grid to each u,v sample. High order Lagrangian interpolation is generally used.

A further advantage of the Cotton–Schwab algorithm, besides gridding correction, is its ability to image and 'CLEAN' many separate but proximate fields simultaneously. In the minor cycle each field is 'CLEAN'ed independently; in the major cycles, 'CLEAN' components from all fields are removed together. In calculating the residual image for each field, the full phase equation, including the w-term, can be used. Thus, the algorithm can correct images for the "non-coplanar baselines" distortion.

The Cotton–Schwab algorithm is often faster than the Clark 'CLEAN', the major exception being for data sets with a large number of visibility samples, where re-gridding many times can be prohibitively expensive. The Cotton–Schwab algorithm also allows 'CLEAN'ing with smaller guard bands around the region of interest, hence with smaller image sizes.

The algorithm has been particularly useful for 'CLEAN'ing sensitive, high-resolution images at lower radio frequencies where there may be numerous confusing sources in the primary beam. In such work, imaging a wide field of view to deconvolve the sidelobes of distant confusing sources in a Högbom or Clark 'CLEAN' could be prohibitively expensive in disk space and time. The confusing sources (once identified from a diagnostic low-resolution image) can however be handled in small sub-fields using the Cotton–Schwab approach.

The Cotton–Schwab algorithm was first implemented in the NRAO's Astronomical Image Processing System (classic AIPS) as the program 'MX' and is now the basis of the AIPS programs 'IMAGR' and 'WFCLN'.

4.4 Other algorithms related to 'CLEAN'

Several later algorithms have tried to correct deficiencies of 'CLEAN'.

Steer, Dewdney & Ito (1984) developed a variant of the Clark algorithm in which the minor cycle is replaced by simply taking all points above a sidelobe-dependent threshold, scaling them and then subtracting normally in the major cycle. This saves time compared to 'CLEAN', but the radio astronomy community has little experience with this variant of the algorithm. The algorithm's ability to handle different practical situations is therefore somewhat uncertain. It was implemented in classic AIPS as 'SDCLN'.

Segalovitz & Frieden (1978) proposed an *ad hoc* modification of the *dirty* beam to enhance the smoothness of the resulting 'CLEAN' image. Cornwell (1983) justified a similar prescription as forcing the minimization of the image power (i.e., the sum of the squares of the pixel values) and thus pushing down the extrapolated visibility function. Both approaches seem to ameliorate the striping instability (see Section 4.12 below) to which 'CLEAN' is susceptible.

4.5 Practical details and problems of 'CLEAN'

Theoretical understanding of 'CLEAN' is relatively poor even though the original algorithm is quite old. Schwarz (1978, 1979) analyzed the Högbom 'CLEAN' algorithm in detail. He notes that in the noise-free case the least-squares minimization of the difference between observed and model visibility, which 'CLEAN' performs, produces a unique answer if the number of cells in the model is not greater than the number of independent visibility measurements contributing to the dirty image and beam (*cf.* Equations 7 and 8), counting real and imaginary parts separately. This rule is unaffected by the distribution of u,v data so that, in principle, super-resolution is possible if enough visibility samples are available. In practice, however, the presence of noise and the use of the FFT algorithm to calculate the dirty image and beam corrupt our knowledge of the derivatives of the visibility function upon which super-resolution is based. Clearly, even if the FFT is not used, the presence of noise means that independence of the data must be redefined. Schwarz produced a noise analysis of the least-squares approach but it involves the inversion of a matrix of side $N_l N_m$ and so is impractical for large images; furthermore, we are really interested in 'CLEAN', not the more limited least-squares method, since 'CLEAN' will still produce a unique answer in circumstances where the least-squares method is guaranteed to fail. To date no one has succeeded in producing a noise analysis of 'CLEAN' itself. The existence of instabilities (see Section 4.12 below) in 'CLEAN' makes such an analysis highly desirable.

Schwarz also proves three conditions for the convergence of 'CLEAN':

1. The beam must be symmetric.
2. The beam must be positive definite or positive semi-definite. Thus the eigenvalues must be non-negative.
3. The dirty image must be in the *range* of the dirty beam. Roughly speaking, there must be no spatial frequencies present in the dirty image which are not also present in the dirty beam.

All three conditions are obeyed in principle for the dirty image and beam calculated by Equations 7 and 8 if the weight function W is nowhere negative. In practice, however, numerical errors, and the gridding and grid-correction process may create violations of these conditions, so 'CLEAN' will eventually diverge. 'CLEAN'ing close to the edge of a dirty image computed by an FFT is particularly risky.

Most of our understanding of 'CLEAN' comes from a combination of guessing how to apply intuition and Schwarz's analysis to real cases, and much practical experience with real and test data. We will now try to summarize the available lore about how 'CLEAN' should be used, and how it can fail.

4.6 The use of boxes: finite support

The region of the image which is searched for the peak can be limited to those areas (known as the 'CLEAN' *windows* or *boxes*) within which emission is presumed to be present. These boxes restrict the number of degrees of freedom available when fitting the data. Schwarz's work (and common sense) tells us that the number of such degrees of freedom should be minimized but that the 'CLEAN' window should include all real emission. For a simple source in an otherwise uncluttered field of view, one 'CLEAN' window will do, but multiple boxes may be needed when 'CLEAN'ing more complicated sources, or a field containing many sources. In the latter case, the presence of weak sources may be revealed only after the sidelobes of the stronger sources have been removed, so more boxes may be needed as the 'CLEAN' progresses. (Note that such *a posteriori* definition of 'CLEAN' boxes complicates any noise analysis.)

It is hard to gauge the practical implications of Schwarz's observation that the number of degrees of freedom should not exceed the number of independent constraints. In the presence of noise, u,v samples can be judged independent if the differences in visibility due to the size of structure expected are much greater than the noise. Counting visibility points in such a way, the aggregate area of the 'CLEAN' boxes in pixels should be less than twice the number of *independent* visibility samples. If the FFT is used, then the number of independent visibility samples cannot be greater than $\mathcal{O}(N_l N_m)$, so it is advisable to use 'CLEAN' boxes.

Given the uncertainty in determining the number of independent data points, and hence the number of constraints, caution dictates that boxes should always be placed tightly around the region to be 'CLEAN'ed.

4.7 Number of iterations and the 'CLEAN' loop gain

The number of 'CLEAN' subtractions N_{CL} and the loop gain γ determine how deep the 'CLEAN' goes. In particular, for a point source the residual left on the dirty image is $(1-\gamma)^{N_{CL}}$. Hence, to minimize the number of 'CLEAN' subtractions (and so to minimize the CPU time) γ should be unity; one then finds, however, that extended structure is not well represented in the corresponding 'CLEAN' image. In typical VLA applications a reasonable compromise lies in the range $0.1 \leq \gamma \leq 0.25$. (Note that this dependence of the 'CLEAN' image on the loop gain demonstrates the multiplicity of solutions to the convolution equation.) Lower loop gains may be required if the u,v coverage is poor, but the improvements in deconvolution for $\gamma \ll 0.01$ are generally minimal. If in any doubt, then it is wise to experiment (e.g., by decreasing γ and increasing N_{CL}). One exception to the use of low loop gain is in the removal of confusing sources; it is preferable to remove them with high loop gain, as their structure is usually not of interest.

The choice of the number of iterations depends upon the amount of real emission in the dirty image. One should aim at transferring all brightness greater than the noise level to 'CLEAN' components (some implementations of 'CLEAN' allow one to specify a lower intensity limit to the components instead of N_{CL}). 'CLEAN'ing deep into the noise is usually a waste of time unless you specifically wish to analyze the extended, low surface-brightness emission (but see Section 4.9 below).

Examination of the list of 'CLEAN' components, and, in particular, of the behavior of the accumulated intensity in the model, is useful in detecting divergence; sometimes the accumulated intensity diverges. As discussed above, divergence of the Högbom 'CLEAN' is always due to a computational problem. Possible culprits are the gridding process, aliasing, and finite precision arithmetic. In the case of the Clark or the Cotton–Schwab algorithms, the truncated dirty beam patch that is used in the minor cycles of these algorithms must violate Schwarz's conditions. Therefore both may be subject to instability or divergence if the minor cycle is prolonged unduly.

4.8 The problem of short spacings

Deconvolution implicitly interpolates values for un-sampled u,v spacings. In most cases 'CLEAN' does this interpolation reasonably well. However, in the case of short spacings the poor interpolation is sometimes rather more noticeable since very extended objects have much more power at the short spacings. The error is nearly always an underestimation and is manifested as a "bowl" of negative surface-brightness in which the source rests. In such a case, introducing an estimate of the zero-spacing flux density into the visibility data before forming the dirty image can help considerably. The appropriate value of this flux density would be that measured by a single element of the array. In practice, however, single array elements rarely have sufficient sensitivity or stability to provide this estimate accurately. Values estimated from surveys made with larger, more sensitive, and more directive elements are therefore frequently substituted. Choosing the weight for the zero-spacing flux density is difficult; the best estimate seems to be simply the number of unfilled cells around the origin of the gridded u,v plane. However, the results obtained are fairly insensitive to the value used *provided that the 'CLEAN' deconvolution goes deep enough.*

The 'CLEAN' windows or boxes offer a way to constrain the shape of the visibility function $V(u, v)$ near the zero spacing $u = v = 0$. For this reason, careful choice of 'CLEAN' windows may also minimize problems associated with the short spacings.

After 'CLEAN'ing, the emission should be, but is not guaranteed to be, distributed sensibly over the 'CLEAN' image. Failure of the interpolation is indicated by the presence of a "pedestal" of surface brightness within the 'CLEAN' box upon which the source rests. Such a pedestal all over the image can be caused by insufficient 'CLEAN'ing; one can experiment by simply increasing N_{CL}.

Ultimately, of course, it may be necessary to measure the appropriate short-spacing data!

4.9 The 'CLEAN' beam

The 'CLEAN' restoring beam is used to suppress high spatial frequencies that which are poorly estimated by the 'CLEAN' algorithm. There are two competing opinions on this in the radio astronomy community: some object that it is purely *ad hoc* and is undesirable—in the sense that the equivalent predicted visibilities do not then agree with those observed. Others defend it as a way of recognizing the inherent resolution limit. In practice, re-convolving with a 'CLEAN' beam seems to be necessary to produce astrophysically reasonable images.

The most common way to choose the 'CLEAN' beam is to fit an elliptical Gaussian to the main lobe of the dirty beam. This choice is a compromise between resolution and apparent image quality, however, and either larger or smaller beams may be appropriate in some cases. If one is prepared to tolerate a decrease in the apparent quality of the 'CLEAN' image, and if both the signal-to-noise ratio and the u,v coverage are good, then a smaller 'CLEAN' beam can be justified.

Various attempts have been made to improve the choice of the 'CLEAN' beam. The dirty beam, truncated outside the first zero-crossing, is appropriate in some applications since it lacks the extended wings of a Gaussian, but we emphasize that, after convolution with such a beam, just as in the case of a Gaussian clean beam, the 'CLEAN' image does not agree satisfactorily with the original visibilities. An ideal 'CLEAN' beam might be defined as a function obeying three constraints:

1. Its transform should be unity inside the sampled region of the u,v plane.
2. Its transform should tend to zero outside the sampled region as rapidly as possible.
3. Any negative sidelobes should produce effects comparable with the noise level in the 'CLEAN' image.

Constraint (1) is usually the first to be relaxed, and then only positivity of the transform is necessary. It may be that in typical applications 'CLEAN' performs so poorly that these constraints do not allow an astrophysically plausible 'CLEAN' image, however such a topic is probably worth further consideration.

An important consequence of choosing the 'CLEAN' beam poorly is that the units of the convolved 'CLEAN' components may not agree with the units of the residuals. The units of a dirty image are poorly defined, but can be called "Jy per dirty beam area" because an isolated point source of flux density S Jy will appear in the dirty image as a dirty beam shape with amplitude S Jy per dirty beam area. An extended source of total flux density S Jy will be seen in the dirty image convolved with the dirty beam, but the integral will not, in general, be S Jy. However, convolved 'CLEAN' components do have sensible units of Jy per 'CLEAN' beam, which can be converted to Jy per unit area since the equivalent area of the 'CLEAN' beam is usually well-defined. If 'CLEAN' is run to convergence, the integral of the 'CLEAN' image will often be a good estimate of the flux density of an extended object, failing only if the u,v coverage fails to sample the true peak visibility of the source adequately on the shortest spacings. If convergence is not attained, then both flux density and noise estimates obtained from the 'CLEAN' image can be significantly in error.

4.10 Use of *a priori* models in 'CLEAN'

A priori models of sources can be used to good effect in 'CLEAN'. A good example is in the 'CLEAN'ing of images of planets. In this case the visibility function of a circular disk can be subtracted from the observed visibilities before making the dirty image. 'CLEAN' is then asked only to find the small perturbations from the disk model, so the image quality and speed of convergence can both be improved.

4.11 Non-uniqueness of 'CLEAN' images

A major drawback to the use of 'CLEAN' is the way in which its answers depend upon the various control parameters: the location of 'CLEAN' boxes, the loop gain γ and the number of 'CLEAN' subtractions. By changing these one can, even for a relatively well-sampled u,v plane, produce noticeably different final images. In the absence of an error analysis of 'CLEAN', one can do nothing about this except practise vigilance and avoid interpreting any aspects of an image that are unstable to the choice of control parameters.

Part of our purpose in this tutorial is to make you aware of effects that should keep you from being over-confident in the final images produced by 'CLEAN'. In almost any astronomical application, Monte Carlo tests of 'CLEAN', and comparisons of its results with those of other deconvolution methods, are illuminating. They remain the only practical way to estimate the effects of data errors and of different 'CLEAN'ing strategies on the final image.

Eventually, you will gain experience of applying 'CLEAN' to a wide range of different images. This experience will let you guide 'CLEAN' to plausible results more quickly. The 'CLEAN' images that you then produce may not be intrinsically more reliable, but you will have calibrated your use of them for astrophysics much better!

4.12 Instabilities

One instability of 'CLEAN' is well known: its images of extended sources are sometimes modulated at spatial frequencies corresponding to un-sampled parts of the u,v plane (e.g., Cornwell 1983). Convolution with a larger 'CLEAN' beam than usual can mask this problem, especially if the un-sampled regions are in the outer parts of the u,v plane. Reducing the loop gain γ to very low values generally has little effect. Various modifications to CLEAN have been invented to try to combat this problem (see *e.g.* Cornwell (1983)), but overall the experience is that the best solution is to use another deconvolution algorithm, such as MEM.

The occurrence of the stripes is a natural consequence of the incorrect information about radio sources embodied in the 'CLEAN' algorithm. Astronomers have not found much evidence for real stripes in radio sources, so they are skeptical about stripes in 'CLEAN' images. Unfortunately the only *a priori* information built into 'CLEAN', via the use of 'CLEAN' boxes, is that astronomers prefer to see mainly blank images; there is no bias against stripes. These and other considerations motivated the development of deconvolution algorithms which incorporate extra constraints on astrophysically plausible brightness distributions or which are claimed to produce, in some way, optimal solutions to the deconvolution equation. MEM is an example of the latter.

5 The Maximum Entropy Method (MEM)

We have seen that deconvolution tries to select one answer from the many that are possible. 'CLEAN' uses a *procedure* to select a plausible image from the feasible set. Some of 'CLEAN''s problems arise just because it is procedural, so there is no simple equation describing the output image. This makes it difficult to analyze the errors (noise) in 'CLEAN'. By contrast, the Maximum Entropy Method (MEM) is not procedural: the image selected is that which fits the data, to within the noise level, and also has maximum entropy. The use of the term *entropy* has lead to some confusion about the justification for MEM. There is no consensus on this subject (e.g., Frieden 1972; Wernecke & D'Addario 1976; Gull & Daniell 1978; Jaynes 1982; Narayan & Nityananda 1984, 1986; Cornwell & Evans 1985). The authors' preferred justification defines the entropy as something which, when maximized, produces a positive image with a compressed range in pixel values. Image entropy thus defined is therefore not to be confused with a "physical entropy" (see Cornwell 1984a). The compression in pixel values forces the MEM image to be "smooth", and the positivity forces super-resolution on bright, isolated objects. There are many possible forms of this extended type of entropy, see e.g., Narayan & Nityananda 1984, but one of the best for general purpose use is:

$$\mathcal{H} = -\sum_k I_k \ \ln \frac{I_k}{M_k e} \,, \tag{9}$$

where M_k is a "default" image that incorporates *a priori* knowledge about the object. For example, a low resolution image of the object can be used to good effect as the default.

A requirement that each visibility point be fitted exactly is nearly always incompatible with the positivity of the MEM image. Consequently, data are usually incorporated in a constraint that the fit, χ^2, of the predicted visibility to that observed, be close to the expected value:

$$\chi^2 = \sum_r \frac{\left|V(u_r, v_r) - \widehat{V}(u_r, v_r)\right|^2}{\sigma^2_{V(u_r,v_r)}} \,. \tag{10}$$

Simply maximizing $\mathcal{H}$ subject to the constraint that χ^2 be equal to its expected value leads to an image which fits the long spacings much too well (better than 1σ) and the zero and short spacings very poorly. The cause of this effect is that the entropy $\mathcal{H}$ is insensitive to spatial information. It can be avoided (Cornwell & Evans 1985) by constraining the predicted zero-spacing flux density to be that provided by the user.

Algorithms for solving this maximization problem have been given by Wernecke & D'Addario (1976), by Cornwell & Evans (1985), and by Skilling & Bryan (1984). The Cornwell–Evans algorithm was coded in the NRAO's Astronomical Image Processing System (classic AIPS) as 'VTESS'. This algorithm works well for many cases and its code is in the public domain. It is generally faster than 'CLEAN' for larger images, the break-even point being around 1 million pixels.

MEM is an extremely flexible approach to deconvolution that can readily handle heterogeneous data types; this has made it particularly powerful for mosaicing.

5.1 The default image (prior distribution)

Equation 9 implies that if no data constraints exist, the MEM image is the default image, so the MEM image is always biased towards the default. A reasonable "default default" image is flat, with total flux density equal to that specified. A low-resolution image, if available, can be used as the default to good effect; this is a way to combine single-dish data with interferometric data, for example.

5.2 Total flux density

As described above, if the total flux density in the MEM image is left unspecified then the value found may be seriously biased if the signal-to-noise ratio is low. There is no way around this at the moment, except to guess a value and then adjust it to get an image that looks "reasonable"—for example, possessing a flat baseline. For bright objects, only an order-of-magnitude estimate is required to set the flux density scale. The estimated flux density is not then fitted, but is used only to set a reasonable default image. Guessing low by about an order of magnitude often works well.

5.3 Varying resolution

In the folklore, MEM is criticized for resolution that depends on the signal-to-noise ratio. But there are sound theoretical reasons to believe that this effect is common to all nonlinear algorithms that know about noise (Andrews & Hunt 1977). If you want to "fix" the resolution in MEM, the best answer is to do as is done in CLEAN: convolve the final MEM image with a Gaussian beam of appropriate width to smear out the fine scale structure, and add the residuals back in.

There are occasions when the super-resolution exhibited by MEM images is reliable, although it is not yet feasible to predict these in advance.

5.4 Bias

Another common complaint about MEM is that the answer is biased, i.e., that the ensemble average of the estimated noise is not zero. This is true; it is the price paid by any method which does not try to fit exactly to the data as 'CLEAN' does. Bias in an estimator is both common and acceptable, as it usually leads to smaller variance. Cornwell (1980) has estimated the magnitude of the bias, and has shown that it is much less than the noise for pixels having signal-to-noise ratio much greater than one. In fact, with good u,v coverage, for bright pixels the effect of noise on an MEM image is similar to that on a dirty image. The effect of bias can be substantially reduced by using a reasonable default, such as a previous MEM image smoothed with a Gaussian; then only the highest spatial frequencies are biased. The effect of bias can also be eliminated by adding back the residuals, after ensuring a similar flux scale via convolution of the MEM image with a Gaussian (as outlined above).

5.5 Point sources in extended emission

Most of MEM's power to remove sidelobes comes from the positivity constraint. Hence, if the source sits on a background level of emission, then the sidelobes will not be removed fully. The only consistently effective solutions are either (a) to remove point sources using ‘CLEAN’, whose modeling assumption is better suited to them, or (b) to smooth the dirty image prior to deconvolution. In our experience, (a) is superior.

6 Comparing ‘CLEAN’ and MEM

‘CLEAN’ has dominated deconvolution in radio astronomy since its invention, but has not been widely used in other disciplines. Its decomposition of the image into point sources is is often not appropriate for other types of image. In contrast, MEM has spread to many disciplines, probably because most of its justification is independent of the type of data to which it is applied.

The philosophy behind MEM is intriguing and may convince some of you about the objectivity of MEM (see Jaynes 1982 for an exposition of MEM from its inventor). For those of you who do not become acolytes, the practical differences between ‘CLEAN’ and MEM may be more interesting.

‘CLEAN’ is nearly always faster than MEM for small and simple images for which its approach of optimizing a small number of pixels is more efficient. For typical VLA images, the break-even point comes at around a million pixels of brightness. For large, complicated images such as those of supernova remnants at high resolution (up to 100 million pixels), ‘CLEAN’ is impossibly slow so an MEM-type deconvolution is mandatory.

‘CLEAN’ images are nearly always rougher than MEM images. This may be traced to the basic iterative scheme. In ’CLEAN”, what happens to one pixel is not directly coupled to what happens to its neighbors, save by the data constraints, so there is no mechanism to introduce smoothness. MEM couples pixels by minimizing the spread in their values, so the resulting images are smooth although the entropy term does not explicitly embody spatial information.

Both MEM and ‘CLEAN’ fail on some types of structure. ‘CLEAN’ usually makes extended emission blotchy, and may introduce coherent errors such as stripes. MEM copes poorly with point sources in extended emission. Both work quite well on isolated sources with simple structure, and can produce meaningful enhancement of resolution, though MEM does slightly better in most cases. Both do poorly on mildly resolved objects, a surprising result that was first demonstrated by Briggs (1995), and that was the motivation for investigating algebraic deconvolution, see Section 7 below.

Both MEM and CLEAN can behave problematically when interpolating at the inner edge of the sampled u,v plane. MEM tends to over-estimate the intensity of the broadest-scale emission (the positivity bias), whereas ‘CLEAN’ tends to underestimate it.

Since MEM tries to separate signal and noise, it is necessary to know the noise level reasonably well. Also, as mentioned above, knowledge of the total flux density in the image helps considerably. Apart from this, MEM has no other important control parameters, although it can be helped enormously by specifying a default image. ‘CLEAN’ makes no attempt to separate out the noise, so specification of the noise level is not required. The main control parameters are the loop gain γ, and the number of iterations N_{CL}, both of which are important in determining the final deconvolution.

The default image of MEM is a powerful way to introduce *a priori* information. The effect of the default image can be easily mimicked by 'CLEAN': the default image is simply used as the starting point for the collection of 'CLEAN' components. The use of a disk model for a planet is an example of the use of a default in 'CLEAN'.

Both 'CLEAN' and MEM perform better if either any bright point sources are removed beforehand or that the dirty images are constructed such that the such point sources are exactly registered at pixels. Without this latter registration, 'CLEAN' attempts to construct a multi-component (i.e. extended) model of such sources to represent their positional offset. It is possible for the algorithm to correct itself by the use of negative 'CLEAN' components, but its attempts to do this complicate the assessment of how well 'CLEAN' is progressing. As point-source misregistration also creates difficult problems for positivity-constrained algorithms such as MEM, it is much better to choose image centers and pixel sizes to avoid it for the brightest compact features in any image.

7 Algebraic deconvolution

Since synthesis telescopes are linear devices, one might expect linear algebra to be of use in image deconvolution. Andrews & Hunt (1977) first analyzed image deconvolution problems in terms of linear algebra. In principle, one can express the deconvolution problem as a matrix equation $\mathbf{A}S = D$ where S is a vector of the (unknown) intensity distribution on the sky and D represents the observed data that constrain S via the *measurement matrix* $\mathbf{A}$.

In the image plane, D represents the pixel values in the dirty image and $\mathbf{A}$ the dirty beam pattern that relates values in S and D. The elements of S would be the strengths of the δ-function components in the 'CLEAN' representation, for example.

In the u,v plane, D represents the real and imaginary parts of the visibility samples $V(u, v)$, and $\mathbf{A}$ contains the sine and cosine terms that represent the Fourier transform relationship between S and D.

If the extent of the source brightness is poorly known then the S vector can contain many elements. The $\mathbf{A}$ matrix is then almost certainly singular, so there are either no solutions to $\mathbf{A}S = D$, or infinitely many (the "invisible distribution" problem). However, if the source extent is sufficiently small then the $\mathbf{A}$ matrix may be non-singular and a unique solution may be possible. Even if $\mathbf{A}$ is mildly singular, it may be that quite reasonable contraints on the solution S lead to an effectively unique solution. A serious practical obstacle to the use of linear algebra in the past has been the computing problem: since the size of $\mathbf{A}$ goes roughly as the square of the number of pixels, for many solution algorithms, the solution time goes roughly as the *sixth* power of the number of pixels. However, modern workstations have sufficient resources to allow linear algebra-based deconvolution of images with up to 5000-6000 pixels. Such algorithms have been investigated by Briggs (1995).

7.1 Singular value decomposition

Singular value decomposition (SVD) is a general linear-algebraic tool for dealing with singular or near-singular matrices (Noble & Daniel 1977). It is a generalization of eigenfunction analysis to systems split into two domains, such as the sky and the u,v planes. SVD determines the form of S which has minimum length by discarding the singular or near-singular terms in the formal algebraic solution. Briggs (1995) briefly discussed the use of SVD for deconvolving VLBA data; he showed that for about 3000 pixel intensities and visibilities, SVD produces an image whose quality is subjectively on a par with that of 'CLEAN', but at the cost of large memory use and of long running time (on IBM RS/6000-580 workstations).

7.2 Non-negative least-squares

Non-negative least-squares (NNLS), introduced by Lawson & Hanson (1974), also solves the basic matrix equation algebraically, but subject to the added constraint that S contains no negative elements. In principle, the algorithm has the merit that, given sufficient time, it will satisfy well-defined termination conditions, and thus requires no arbitrary cutoff parameter. This makes it a 'hands-off' algorithm whose output is not susceptible to mis-tuning by unfortunate choice of the input parameters. In practice, however, the computation time and memory usage can be impossibly large if the number of non-zero pixels exceeds about 6000-8000.

The point source model output by NNLS is again smoothed with a Gaussian beam and added to any residual emission when making the final image.

NNLS distinguishes itself on bright, compact sources that neither 'CLEAN' nor MEM can process adequately. Briggs showed that on such sources, both CLEAN and MEM produce artifacts that resemble calibration errors and that limit dynamic range. NNLS has no difficulty imaging such sources. It also has no difficulty with sharp edges, such as those of planets or of strong shocks, and can be very advantageous in producing models for self-calibration for both types of sources. Briggs (1995) showed that NNLS deconvolution can reach the thermal noise limit in VLBA images for which 'CLEAN' produces demonstrably worse solutions.

NNLS is therefore a powerful deconvolution algorithm for making high dynamic range images of compact sources for which strong finite support constraints are applicable.

8 Other methods, including hybrids

Braun & Walterbos (1985) proposed a way to address the problem of incomplete short spacing information in the absence of other shortcomings in the visibility sampling. A least-squares fit to a matched functional form is used to analytically continue the background beneath the locations of extended sources. The technique is efficient and successful for this restricted problem where the confinement constraint can be applied effectively.

Hybrid techniques try to exploit the virtues of several algorithms simultaneously while avoiding their pitfalls. For example, the awkward but common problem of deconvolving compact structure on an extended background can be tackled by 'CLEAN'ing the compact structure down to the level of the extended emission, followed by a MEM deconvolution of what remains. The component models of each method are then combined, restored, and added to the residuals.

A variant of this approach which is also effective for multi-pointing deconvolution problems consists of 'CLEAN'ing the individual pointings at the full available resolution and forming the linear combination with appropriate weighting, while using MEM to simultaneously deconvolve the data at low resolution. These results are merged by extracting the inner Fourier transform plane of the MEM result and combining it (with appropriate normalization) with the outer Fourier transform plane of the 'CLEAN' result and back-transforming. Such techniques offer considerable promise to general application, especially if their use can be streamlined.

It is ironic that, formally, more is known about the type of images generated by MEM than by 'CLEAN' (see e.g., Narayan & Nityananda 1986), since 'CLEAN' is rather more widely used. Indeed many criticisms of MEM arise because certain of its properties, such as the bias, can be analyzed. Schwarz's analysis of 'CLEAN' is incomplete in that it does not address the interesting under-determined case in which there are fewer data than pixels. We hope that someday this problem might be investigated satisfactorily.

Although deconvolution algorithms are now as important in determining the quality of images produced by a radio telescope as the receivers, correlators and other equipment, they are far less well understood. A good description is that they are poorly engineered. Only further research and development of new and existing algorithms can redress this imbalance.

9 References

Andrews, H. C. & Hunt, B. R. (1977), *Digital Image Restoration*, Prentice–Hall (Englewood Cliffs, NJ).

Bracewell, R. N. & Roberts, J. A. (1954), "Aerial smoothing in radio astronomy", *Aust. J. Phys.*, **7**, 615–640.

Braun, R. & Walterbos, R. A. M. (1985), "A solution to the short spacing problem in radio interferometry", *Astron. Astrophys.*, **143**, 307–312.

Briggs, D. S. (1995), *High Fidelity Deconvolution of Moderately Resolved Radio Sources*, Ph.D. thesis, New Mexico Inst. of Mining & Technology.

Clark, B. G. (1980), "An efficient implementation of the algorithm 'CLEAN' ", *Astron. Astrophys.*, **89**, 377–378.

Cornwell, T. J. (1980), *The Mapping of Radio Sources from Interferometer Data*, Ph. D. Thesis, University of Manchester.

Cornwell, T. J. (1983), "A simple method of stabilizing the clean algorithm", *Astron. Astrophys.*, **121**, 281–285.

Cornwell, T. J. (1984a), "Is Jaynes' maximum entropy principle applicable to image reconstruction?", in *Indirect Imaging*, J. A. Roberts, Ed., Cambridge University Press (Cambridge, England), pp. 291–296.

Cornwell, T. J. & Evans, K. F. (1985), "A simple maximum entropy deconvolution algorithm", *Astron. Astrophys.*, **143**, 77–83.

Frieden, B. R. (1972), "Restoring with maximum likelihood and maximum entropy", *J. Opt. Soc. Am.*, **62**, 511–518.

Gull, S. F. & Daniell, G. (1978), "Image reconstruction from noisy and incomplete data", *Nature*, **272**, 686–690.

Högbom, J. (1974), "Aperture synthesis with a non-regular distribution of interferometer baselines", *Astrophys. J. Suppl. Ser.*, **15**, 417–426.

Jaynes, E. T. (1982), "The rationale of maximum entropy methods", *Proc. IEEE*, **70**, 939–952.

Lawson, C. L. & Hanson, B. J. (1974), *Solving Least Squares Problems*, Prentice–Hall (Englewood Cliffs, NJ).

Narayan, R. & Nityananda, R. (1984), "Maximum entropy—flexibility versus fundamentalism", in *Indirect Imaging*, J. A. Roberts, Ed., Cambridge University Press (Cambridge, England), pp. 281–290.

Narayan, R. & Nityananda, R. (1986), "Maximum entropy image restoration in astronomy", *Ann. Rev. Astron. Astrophys.*, **24**, 127–170.

Noble, B. & Daniel, J. W. (1977), *Applied Linear Algebra*, Prentice–Hall (Englewood Cliffs, NJ).

Schwab, F. R. (1984b), "Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry", *Astron. J.*, **89**, 1076–1081.

Schwarz, U. J. (1978), "Mathematical-statistical description of the iterative beam removing technique (method CLEAN)", *Astron. Astrophys.*, **65**, 345–356.

Schwarz, U. J. (1979), "The method 'CLEAN'—use, misuse and variations", in *Image Formation from Coherence Functions in Astronomy*, C. van Schooneveld, Ed., D. Reidel (Dordrecht, Holland), pp. 261–275.

Segalovitz, A. & Frieden, B. R. (1978), "A 'CLEAN'-type deconvolution algorithm", *Astron. Astrophys.*, **70**, 335–343.

Skilling, J. & Bryan, R. K. (1984), "Maximum entropy image reconstruction: general algorithm", *Mon. Not. Roy. Astr. Soc.*, **211**, 111–124.

Steer D. G., Dewdney, P. E., & Ito, M. R. (1984), "Enhancements to the deconvolution algorithm 'CLEAN' ", *Astron. Astrophys.*, **137**, 159–165.

Wernecke, S. J. & D'Addario, L. R. (1976), "Maximum entropy image reconstruction", *IEEE Trans. Computers*, **C-26**, 351–364.