DECONVOLUTION AND SELF-CALIBRATION IN RADIO ASTRONOMY USING MEM

R.K. Shevgaonkar, Indian Institute of Astrophysics, Bangalore-560 034, India.

1. INTRODUCTION

Aperture synthesis technique is the most powerful indirect imaging technique used in radio astronomy for obtaining high resolution images. The technique, instead of measuring the image directly, measures its Fourier components which form the required image after Fourier inversion. The radio images obtained using this technique are degraded mainly due to two practical reasons:

(a) Inadequate sampling of the Fourier plane (also called uvaplane in radio astronomy): In principle, to obtain the true image with infinte resolution, all the Fourier coefficients up to infinite spatial frequency must be sampled. However, due to finite size and non uniform sampling of the synthesized aperture, the synthesized image, in general, deviates from the true brightness distribution. The reason being that while Fourier inverting, the unmeasured Fourier coefficients are naturally assumed to be zero. The image thus obtained suffers from loss of resolution and introduction of artifacts. Recovery of an image from these artifacts is a deconvolution problem.

(b) Errors in the measurement of the Fourier coefficients: Apart from the fact that the aperture synthesis technique does not sample all the Fourier coefficients, the measurements are affected by systematic as well as random errors. In the presence of these errors a simple deconvolution, for obvious reasons, does not provide a good image reconstruction. Elimination of these errors is essentially a data calibration problem.

2. PRINCIPLE OF MEM

Maximum Entropy Method, after its introduction by Burg in 1967 for obtaining high resolution spectra, was adopted in image processing by Frieden in 1972. The main difficulty in the image processing is that for a given set of measurements one invariably does not find a unique image. The basic philosophy of MEM is to use the entropy of the image as a quality criterion and to obtain a physically acceptable image which is consistent with the measurements. There has been considerable discussion over the justification for MEM and the use of the word entropy in image processing (Frieden 1972, Gull and Daniell 1978, Nityananda and Narayan 1982, Cornwell and Evans 1985, Shevgaonkar 1986b). Due to the disagreement between different groups, the method, although coming into existence about two decades ago, did not receive active application. However, the recent work (Nityananda and Narayan 1982, Cornwell and Evans 1985, Shevgaonkar 1986b) suggests that the image entropy has nothing to do with the classical entropy of a system The present, widely (which has a completely different meaning). accepted understanding is that the entropy is some kind of a measure of "goodness" of the image. Since the goodness is a rather subjective criterion, the entropy can be thought as a means of providing an a priori knowledge about the image. The minimum a priori knowledge one has is that the image (considering total intensity now, rather than polarization) is positive definite. Therefore the MEM must reconstruct an image which is at least positive definite. Also for a given measurement the method must give a unique image irrespective of the user and the starting guess of the image. These constraints permit usage of only for functions like I_{PR}^{R} and R_{PR}^{R} P_{PR}^{R} P_{PR}^{R} P_{PR}^{R} P_{PR}^{R} only few functions like lnB, MB ln B, B ln(B/B_o), (B/B_o) etc. to define the entropy of an image (see Nitvananda and Naravan 1982, Shevgaonkar 1986b); where B is the image brightness and B_0 is a prior model of the image. It is noteworthy that all the functions used for defining entropy share the following common properties:

(a) The entropy function f(B) provides a non-linear restoration filter (i.e. $\partial f/\partial B \neq \text{constant}$).

(b) The function is explicitly definable for positive values of B only (positivity constraint).

(c) The function possesses a single maximum over the acceptable range of the brightness (i.e. $\partial^2 f/\partial B^2$) < 0; the uniqueness condition).

Assuming the brightness is uncorrelated from one pixel of the image to the other the entropy of an image is equal to the sum of the entropies of individual pixels. Mathematically we can write the entropy of a two dimensional image as

 $E = \int \int f[B(x,y)] dx dy$ (1)

where (x,y) are the two orthogonal image coordinates.

The basic principle of the method is to reconstruct an image which is consistent with all the measurements and the a priori knowledge, and has a maximum entropy E. Assuming that the measurements are carried out in the Fourier plane (as is the case with indirect imaging), consistency with the measured data requires

$$\sum_{N=\sigma}^{\rho=\rho} {2 \leq N}$$
(2)

Where ρ is the measured Fourier coefficient (which we will call visibility coefficient hereafter), ρ is the true visibility coefficient, and σ is the rms of the statistical random noise of the measurement. N denotes the total number of measured data points. Equation (1) and (2) can be combined (Ables 1974, Wernecke and D'Addario 1976) to define an objective function F, which when maximized produce an image which is most consistent with the data and has maximum entropy.

Where λ is a Lagrange multiplier and essentially defines the relative weights given to the two terms in eqn.(3). To obtain the best reconstruction, a proper value λ is found by trial and error. However, a value of λ which gives equal weight to the two terms in eqn. (3) appears to be reasonable, unbiased choice.

3. DECONVOLUTION

As mentioned earlier, deconvolution is a process of recovering those visibility coefficients which were not merely modified by the instrument response but were actually never measured. Conventionally, this latter information is assumed to be hopelessly lost and therefore, in a strict mathematical sense, the deconvolution is an impossible task. However, some kind of model fitting along with a prior knowledge about the image (like positivity) makes the deconvolution possible in practice.

There are two commonly used deconvolution methods in radio astronomy:

(i) CLEAN which was introduced by Hogbom (1974) and was later mathematically justified by Schwarz (1978). This is the most routinely used deconvolution method in radio astronomy, although for extended images its performance deteriorates. This method has already been discussed in detail by A.P. Rao in this workshop. For detailed understanding of the method and its practical implication readers are referred to the above two papers and the papers by Clark (1980) and Cornwell (1983).

(ii) MEM, which was first introduced to radio astronomy by Ables (1974), has been developed to a satisfactory level over the last decade (Wernecke and D'Addario 1976, Gull and Daniell 1978, Cornwell and Evans 1985). The method can be described as follows:

Let us assume that the visibility coefficients were measured at discrete spatial locations $(m,n) \in K$; K is an arbitrary spatial region (e.g. a

radio interferometer) where total N measurements were carried out. From the Fourier transform relationship, the measured visibility coefficients can be written as

$$\rho_{mn} = \int \int B(x,y) \exp \left[i 2\pi(mx+ny) \right] dx dy + noise \qquad (4)$$

 $(m,n) \in K$ Or, inverting the Fourier integral,

$$B(\mathbf{x},\mathbf{y}) = \sum_{m \in n} \sum_{m \in n} \exp\left[-i 2\pi(m\mathbf{x}+n\mathbf{y})\right]$$
(5)

B(x,y) is the true brightness distribution and B(x,y) is the observed image obtained assuming the unmeasured visibility coefficients to be zero. B(x,y) is also called the principal solution, referred to in radio astronomy as the "dirty map".

To maximize the objective function F with respect to the visibility coefficients ρ_{mn} , we require $\partial F/\partial \rho_{mn} = 0$. Substituting for the observed brightness distribution B(x,y) in eqn.(1) and (3) and differentiating with respect to ρ_{mn} we get the gradient of the objective function as

$$g_{mn} = \frac{\partial F}{\partial \rho_{mn}} = \int \int \partial f/\partial B \exp \left[-i 2\pi (mx+ny)\right) dx dy$$

$$= \frac{2\lambda}{N} \left[\frac{\rho_{mn}}{\sigma_{mn}} \right]$$
(6)

The MEM image is an image for which the gradient g_{mn} should be equal to zero. Therefore MEM reconstruction, in otherwords reduces to a multidimensional optimization problem. Generally in any optimization problem computation of the gradient is the most tedious and time consuming part. However, in this case the gradient can be computed just through a Fourier transform as shown in eqn (6) and therefore any of the standard optimization methods can be employed for the maxima search.

Figure 1 shows a flow diagram of the MEM algorithm for deconvolving the images using the conjugate gradient search method. Note that the second box 'Add a floating constant C' is a new feature in the method discussed so far. This operation is needed to circumvent some of the practical difficulties in the implementation of the method as it is described above. Its usage requires some explanation.

It should be realized that at the beginning of the iteration the observed image generally has some negative artifacts. Since the entropy is not defined for the negative intensities, initially the entropy becomes imaginary and the algorithm can not proceed. There are two ways of taking care of this problem: (1) clip all negative intensities to a very small positive value (Cornwell and Evans 1985), (ii) Add temporarily a constant bias to the entire image such that the intensity

122



Figure 1. Flow diagram of the MEM algorithm for deconvolution.

TABLE 1 COMPARISON OF CLEAN AND MEN

Subject	CLEAN	MEM
1.Philosophy	Mathematically it is a least-squares data fitting technique	It has a firmer mathematical basis
2.Nature of Image	Good for isolated point sources but introduces coherent errors in extended sources	Good for point sources. Relatively better for extended emission compared to CLEAN
3.Quality of Reconstruction	Relatively rough image	Smooth reconstruction
4.Speed	Faster for compact sources and small fields of view but extremely slow for extended fields.	Relatively efficient for large fields
5.Prior knowledge about the image	Always assumes point source nature of the source. Does not guarantee positivity on the image	Prior information can be supplied. Reconsa tructs a positive definite image.
6.Resolution	Has essentially the same angular resolution as the principal solution	Extrapolated visibilities give enhanced resolution termed "super resolution"
7.Polarization Image	Deconvolves different component images indepent dently and therefore does not assume the images to satisfy essential conditions	Treats all the polari- zation component images simultaneously and guarantees essential conditions on the images
8.Noise treatment	Treats equally the high and the low signal to noise ratio pixels in the image	It has a provision to treat data differently having different signal to noise ratio
9.Calibration	Convolution with the clean beam of the image restores the calibration	Due to super-resolution and non-linear reconst- ruction it is difficult to get proper intensity calibration on the reconstructed image.
10.Popularity	Due to its simplicity it is very popular but only in radio astronomy	Application in variety of fields like medical sciences, crystallo ⁴⁴ graphy, X-ray astronomy etc. It is rapidly picking up in radio astronomy.

124

is positive definite in every pixel of the image. We have employed the latter choice in our algorithm and call it a floating operation. One obvious question that comes to our mind immediately is by what amount an image should be made positive definite and what is the effect on the deconvolution? An answer to this question is not very simple. We see that the choice of the floating constant 'C' is related to the slope of the extrapolated visibility function (Bhandari 1978). If 'C' is large the extrapolated visibility function drops to zero very sharply and we obtain a reconstruction which is not much superior to the principal solution. On the contrary for a very small value of C the extrapolation is much flatter and we may get an excess superresolution. The ratio of the maximum to minimum intensity in the floated image can be taken as an indirect measure of resolution enhancement and can be fixed, depending upon the requirement, to a predetermined value. Indeed the floating constant is one of the important controlling parameters of the MEM reconstruction (see Shevgaonkar 1986a, 1987; Narayan and Nityananda 1986), and its choice is left to the users. It is easy to see that at the convergence the restored image still floats by an amount equal to the last value of C. Therefore at the end this bias has to be subtracted from the image.

Other steps in the algorithm are straight forward. Using eqn (6) gradient and the conjugate gradient of the objective function are computed. Visibility coefficients are changed by a fraction 'y' of the computed conjugate gradient. Changed coefficients are Fourier inverted to obtain a new restored image. The iteration continues till a convergence is reached. The modulus of the gradient has been used as a convergence parameter. The method generally converges in about 20 iterations and therefore requires a computation time of about 40 fast Fourier transforms (FFTs).

The two methods CLEAN and MEM have been compared for their merits and demerits in Table 1.

4. CLOSURE RELATION

Assuming that the random noise is regligible, phase errors could be introduced by changes in the path lengths due to a variable atmosphere, temperature sensitive electronics etc., and the amplitude errors introduced by a front-end amplifier gain variation or an antenna pointing error. The concept of closure phase was first introduced by Jennison in 1958. He realized that the sum of the visibility phases around a closed loop of interferometers is free from all the antenna based phase errors. This can be seen very easily as follows:

Let us take three antenna elements A, B, C forming three interferometer pairs AB, BC and CA. Let ϕ_{AB} , ϕ_{BC} and ϕ_{CA} be the true visibility phases for the three interferometers respectively for any arbitrary brightness distribution. If δ_A , δ_B and δ_C are the total phase errors introduced at the three antenna elements, the observed visibility phases can be written as

It is apparent that the sum of the three observed phases

$$\phi_{ABC} = \phi_{AB} + \phi_{BC} + \phi_{CA} \tag{8}$$

is equal to the sum of the true visibility phases i.e. $(\phi_{AB} + \phi_{BC} + \phi_{CA})$. The quantity ϕ_{ABC} is free from all the antenna based systematic errors and is known as the 'closure' phase. It is clear that the closure phase is an error free quantity irrespective of the separation and orientation of the interferometer baselines. The closure relation (8) can also be written in a differential form, i.e.

$$\phi_{AB} + \phi_{BC} + \phi_{CA} = 0 \tag{9}$$

where $\phi\, 's$ are the algebraic differences between the measured and the true visibility phases.

Similar to the closure phase a quantity called closure amplitude can also be defined, however, for at least four element arrays. If A_{AB} , A_{BC} , A_{CD} and A_{DA} are the visibility amplitudes of the four out of Six interferometers formed by four antenna elements A,B,C and D, the closure amplitude can be written as

$$T_{ABCD} = \frac{A_{AB} A_{CD}}{A_{BC} A_{DA}}$$
(10)

The observed and true closure amplitudes are also going to be identical. If we express the observed visibility amplitude

$$A_{AB} = A_{AB} (1 + A_{AB})$$
(11)

and assume that A's are small compared to unity (as is generally the case) the closure equation (10) can be simplified to

$$A_{AB} - A_{BC} + A_{CD} - A_{DA} = 0$$
 (12)

Please note the differences in the definitions of ϕ 's and A's.

An N element array measures N(N-1)/2 (=q) complex visibility coefficients whereas there are only (N-1) phase errors and N amplitude errors related to the N antenna elements. Therefore one obtains N(N-3)/2 (=r) independent closure amplitudes and (N-1)(N-2)/2 (=p)

independent closure phases which are free of all the antenna based errors. For an N element system, the closure equations (9) and (12) can be written in a matrix form as

and

$$\Pi_{\mathbf{A}} \cdot \underline{\mathbf{A}} = 0 \tag{14}$$

. . . .

Where, H is the closure phase matrix of size $p \ge q$ (p < q) and H is closure amplitude matrix of size $r \ge q$ (r < q). Both matrices are composed of elements 0 or ± 1 .

It can be seen easily that the large number of closure quantities, although free from the antenna errors, can not reconstruct the image uniquely. The most obvious thing to note is the loss of absolute position and strength of the source. A position change introduces a phase change in the visibility which depends upon the difference in the position of the interferometer elements. This phase change can be factorized into two antenna related errors and threfore cancels out in the closure phase. The loss of absolute source strength calibration is evident as the closure amplitude is only a ratio of the visibility coefficients. In any case the fact is, inspite of having large number of error-free quantities, the remaining degrees of freedom corresponding to the antenna element complex gains should still have to be balanced out by using an a priori knowledge about the distribution.

5. SELF#CALIBRATION

. .

Using closure quantities, the calibration of the data in a self consistent manner was first proposed by Readhead and Wilkinson in 1978. The approach was visibility oriented. However, Cornwell and Wilkinson (1981) later adopted the antenna based approach and this is the currently followed approach. Self-calibration is essentially a combination of CLEAN and the closure phases. The basic philosophy is to obtain a plausible model image B of the intensity distribution, the Fourier transform ρ of which, when corrected by some complex gain factors, reproduces the observed visibility coefficients within noise. By plausible model we mean a positive, confined image. The model fitting can be done (Schwab 1980) by minimizing the error term

$$\chi^{2} = \sum_{\substack{m,n \ m \neq n}}^{*} w_{mn} |\rho_{mn} - G_{m} G_{n}^{*} \hat{\rho}_{mn}|^{2}$$
(15)

where w are the weights decided from signal-to-noise ratio. G and G are the complex antenna gains for element m and n respectively, and the denotes the complex conjugate. ρ_{mn} are the observed visibility coefficients and β are the visibilities corresponding to the model B.

Equation (15) can be re-written as,

$$x^{2} = \sum_{\substack{m,n \ m \neq n}} w_{mn} |\hat{\rho}_{mn}|^{2} |X_{mn} - G_{m} G_{n}^{*}|^{2}$$
 (16)

where,

$$X_{mn} = -\frac{\rho_{mn}}{\rho_{mn}}$$
(17)

This step turns the object into a pseudo point source. The method can be applied in an iterative manner to refine the model as follows:

- i. Using all the prior knowledge we have about the object make a model image
- ii. Convert the source into a point source using model
- iii. Solve for the complex antenna gains
- iv. Correct the visibilities as

- v. Obtain a new model by Fourier inverting the corrected visibilities.
- vi. Go to (ii) unless the image is satisfactory.

Deconvolution method CLEAN or MEM (Sanroma and Estallela 1984) is used to obtain a new model at every iteration. For detailed understanding of this technique a review article by Pearson and Readhead (1984) could be referred.

Since its invention, the Self*calibration has been quite routinely used in radio astronomy to obtain high dynamic range images. However, the method has certain drawbacks.

- i) It is sensitive to the choice of the model to some extent.
- ii) Since the key module of the method is CLEAN, the method faces difficulties for extended images.
- iii) For large fields of view closure quantities can be considered as an alternative to the conventional self-calibration.

6. MEM FOR DATA CALIBRATION.

Since the brightness distribution B is real, the visibility function ρ_{mn} [=A_{mn} exp(i ϕ_{mn})] must be Hermitian giving

$$^{A}mn = A_{-m-n} \text{ and } \phi_{mn} = -\phi_{-m-n}$$
(19)

Just to avoid complications, for the time being let us assume that the noise is negligible compared to the visibility amplitudes over all baselines. Using Hermitian property (19) in eqn.(5), the brightness distribution can be rewritten as

$$B(x,y) = \rho_{00} + \sum_{m,n>0} \{2A_{mn} \cos[\phi_{mn} - 2\pi(mx+ny)]\}$$
(20)

Substituting eqn.(20) in (3) and differentiating the entropy with respect to amplitude and phase we get the gradient for the entropy with respect to A_{mn} and ϕ_{mn} respectively as

$$g_{A_{mn}} = 2 \sigma_{mn} \cos(\alpha_{mn} - \phi_{mn})$$
(21)

$$g_{\phi_{mn}} = 2A_{mn}\sigma_{mn}\sin(\alpha_{mn}-\phi_{mn}) \qquad (22)$$

where we have defined

$$\sigma_{mn} e^{i\alpha_{mn}} = FT \begin{bmatrix} -\frac{\partial f}{\partial B} \end{bmatrix}$$
(23)

where FT denotes Fourier transform.

0.17

Equating eqns (21) and (22) to zero we can get the maximum entropy solution. However the difficulty is that since these equations have harmonic functions in them, one may obtain multiple solutions. By making use of the closure quantities the choice of solutions may be restricted and in most of the cases a unique solution is obtained. It should be noted that the maximization of the entropy is only a desirable condition whereas the closure relations must be strictly satisfied at every stage of iteration. The visibility phase and amplitude can be changed in the direction of the entropy gradient only as much as the closure constraints allow.

Since we are interested only in the relative changes in the visibility phase and amplitude, a differential form of closure relations (13) and (14) is adequate for this purpose. It is important to note that the differential closure equations (13) and (14) do not require any observational quantity and can be thought as characteristic closure equations of an N element array. Since there are more unknowns than the equations, the equations (13) and (14) define a surface in a q dimensional space. Any vector ($\underline{\phi}$, A) lying on these surfaces satisfies the closure constraints. For the maximization of the entropy we choose, however, that vector which is closest to the entropy gradient vector. The scalar distance between the gradient vector ${\bf g}_{\varphi}$ and ${\bf g}_A$ and $~{}_{\bot}\underline{A}$ can be written as

$$S\phi = |g_{\phi} - \Delta \phi| = \{ \sum_{m,n} (g_{\phi_{mn}} - \Delta \phi_{mn})^2 \}^{1/2}$$
(24)
$$SA = |g_A - \Delta A| = \{ \sum_{m,n} (g_{A_{mn}} - \Delta A_{mn})^2 \}^{1/2}$$
(25)

Since q complex measurements are coupled through p (=(N-1)(N-2)/2) closure phases and r (=N(N-3)/2) closure amplitudes, any (q-p) phases and (q-r) amplitudes can be chosen independently. For S, and S_A to be minimum with respect to these independent variables we require

$$\frac{\partial S}{\partial q_{j}} = 0; j = 1 \text{ to } (q^{-1}p)$$
 (26)
 $\partial \Delta \phi_{j}$

and

$$\partial S_A = 0$$
; k = 1 to (q=r) (26)
 $\partial \Delta A_2$

Equations (26) and (27) along with the closure equations (13) and (14) form a set of equations which can be solved uniquely to get ϕ and \underline{A} . Knowing the shift direction (ϕ , \underline{A}) an algorithm similar to the one described in section 3 can be developed. For the details of the algorithm, its application to simulated data and the treatment of noise the readers are referred to the original reference (Shevgaonkar 1986a). The method does show good promise when tested on simulated images; however, its application to real data is still awaited. The method is expected to be faster than the self-calibration especially for large fields of view and extended structures. The method as such does not require any starting model. However, a use of default image, if available, makes the convergence faster.

7. CONCLUSION

Maximmum Entropy Method has been proved to be a promising image reconstruction technique in radio astronomy and in variety of other fields. Starting from the simple deconvolution of scalar intensity images it has progressed to where it may now also be applicable to data calibration.

In recent years it has been extended to reconstruct polarization images (Nityananda and Narayan 1983, Shevgaonkar 1987) but this application to real synthesis data has yet to be made. Similar to what is mentioned above, the MEM has also been applied for phase recovery in crystallography problems (Narayan and Nityananda 1982). With the upcoming low frequency telescopes like GMRT (Giant Meter Radio Telescope), MEM may find an interesting application for reconstructing imges over non-isoplanatic fields of view (Shevgaonkar 1986c).

130

Presently, due to super-resolving nature, the MEM images are of qualitative interest only. For the quantitative analysis the common practice is still to smooth the super-resolved image to the instrument resolution. Hopefully the problem of image calibration will be well understood in the future and the MEM images will claim the right to their superiority.

REFERENCES

Ables, J.G. (1974), Astron. Astrophys. Suppl., 15, 383. Bhandari, R. (1978), Astron. Astrophys., 70, 331. Burg, J.P. (1967), "Maximum Entropy Spectral Analysis", 37th Annual Society of Exploration Geophys. Meeting, Oklahoma City. Clark, B.G. (1980), Astron. Astrophys., 89, 377. Cornwell, T.J. and Evans, K.F. (1985), Astron. Astrophys., 143, 77. Cornwell, T.J. and Wilkinson, P.N. (1981), M.N.R.A.S, 196, 1067. Cornwell, T.J. (1983), Astron. Astrophys., 121, 281. Frieden, B.R. (1972), J. Opt. Soc. Amer., 62, 511. Gull,S.F. and Daniell,G. (1978), Nature, 272, 686. Hogbom, J.A. (1974), Astron. Astrophys. Suppl., 15, 417. Jennison, R.C., (1958), M.N.R.A.S., 118, 276. Narayan, R. and Nityananda, R. (1982), Acta Crystallogra., A38, 122. Narayan, R. and Nityananda, R. (1986), Ann. Rev. Astron. Astrophys, 24, 127. Nityananda, R. and Narayan, R. (1982), J. Astrophys. Astron. 3, 419. Nityananda, R. and Narayan, R. (1983), Astron. Astrophys. 118, 194. Pearson, T.J. and Readhead, A.C.S. (1984), Ann. Rev. Astron. Astrophys., 22, 97. Readhead, A.C.S. and Wilkinson, P.N. (1978), Ap. J., 223, 25. Sanroma, M. and Estallela, R. (1984), Astron. Astrophys. 133, 299. Schwab, F. (1980), Proc. S.P.I.E., 231, 18. Schwarz, U.J. (1978), Astron. Astrophys., 65, 345. Shevgaonkar, R.K. (1985), Ph.D. Thesis, Indian Institute of Technology, Bombay. Shevgaonkar, R.K. (1986a), Astron. Astrophys., 162, 349. Shevgaonkar, R.K. (1986b), J. Astrophys. Astron., 7, 275. Shevgaonkar, R.K. (1986c), Proc. of Workshop on "Low Frequency Radio Astronomy", ed. Erickson, W.C. and Cane, H.V., p.59. Shevgaonkar, R.K. (1987), Astron. Astrophys., 176, 159. Wernecke, S.J. and D'Addario, L.R. (1976), IEEE Trans. Computers, C426, 351.