The Generic Interferometer: II Image Solvers
AIPS++ Implementation Note 184

T.J. Cornwell, NRAO

1995, August 22

Contents

1 Introduction

In a previous memo, I described calibration and imaging for the generic interferometer (Cornwell, 1995) using the measurement equation developed by Hamaker, Bregman and Sault (1995). In this memo, I investigate how the well-known deconvolution algorithms can be adapted to the new 4-dimensional formalism to create entirely new ways of imaging.

I assume that the reader is familiar with the terminology, notation and results from the above mentioned references.

2 The Measurement Equation for a Generic Interferometer

The Generic Interferometer Measurement Equation is given by:

$$\vec{V}_{\mathrm{ij}} = \left[G_i \otimes G_j^*\right] \left[D_i \otimes D_j^*\right] \left[C_i \otimes C_j^*\right] \sum_k \left[E_i\left(\underline{\rho}_k\right) \otimes E_j^*\left(\underline{\rho}_k\right)\right] \left[P_i\left(\underline{\rho}_k\right) \otimes P_j^*\left(\underline{\rho}_k\right)\right] \left[F_i\left(\underline{\rho}_k\right) \otimes F_j^*\left(\underline{\rho}_k\right)\right] S\, \vec{\mathcal{I}}_k\, e^{-2\pi i\left(\underline{r}_i - \underline{r}_j\right)\underline{\rho}_k} \tag{1}$$

3 Image Solvers

Since the connection between estimating the sky brightness and deconvolution is becoming less and less straightforward as our approaches gain in sophistication, I will use the term *Image Solver* for what we previously called a deconvolution algorithm. An Image Solver estimates the sky brightness $\vec{\mathcal{I}}$ for a fixed set of calibration matrices. The principal algorithms that we need to accomodate are CLEAN and MEM. However, there should be flexibility sufficient to allow other algorithms. I will assume that all image solvers need:

Either a residual image, $\vec{\mathcal{I}}_k^R$

Or A measure of the misfit between a model for the sky brightness and the observed visibility data, χ^2, and the derivatives with respect to the sky brightness, $\frac{\partial\chi^2}{\partial\vec{\mathcal{I}}_k}$ and $\frac{\partial^2\chi^2}{\partial\vec{\mathcal{I}}_k^2}$.

The main point of this memo is to take this definition and demonstrate that it does indeed fit nearly all of the image solvers that we now use. This is convenient since it provides a simple interface between an image solver and the machinery for the measurement equation. A well-defined interface is useful for designing software but it is also helpful in understanding algorithms. Thus this memo should address both subjects.

As shown in the first memo:

$$\frac{\partial\chi^2}{\partial\vec{\mathcal{I}}_k} = -2\, \Re \sum_{ij} S^{*T} \left[F_i\left(\underline{\rho}_k\right) \otimes F_j^*\left(\underline{\rho}_k\right)\right]^{*T} \left[P_i\left(\underline{\rho}_k\right) \otimes P_j^*\left(\underline{\rho}_k\right)\right]^{*T} \left[E_i\left(\underline{\rho}_k\right) \otimes E_j^*\left(\underline{\rho}_k\right)\right]^{*T} \left[C_i \otimes C_j^*\right]^{*T} \left[D_i \otimes D_j^*\right]^{*T} \left[G_i \otimes G_j^*\right]^{*T} W_{\mathrm{ij}}\, \Delta\vec{V}_{\mathrm{ij}}\, e^{2\pi i\left(\underline{r}_i - \underline{r}_j\right)\underline{\rho}_k} \tag{2}$$

$$
\begin{gathered}
\frac{\partial^2\chi^2}{\partial\vec{\mathcal{I}}_k\partial\vec{\mathcal{I}}_k^T} = 2\,\Re\sum_{ij} S^{*T} \\
\left[F_i\left(\underline{\rho}_k\right)\otimes F_j^*\left(\underline{\rho}_k\right)\right]^{*T}\left[P_i\left(\underline{\rho}_k\right)\otimes P_j^*\left(\underline{\rho}_k\right)\right]^{*T}\left[E_i\left(\underline{\rho}_k\right)\otimes E_j^*\left(\underline{\rho}_k\right)\right]^{*T} \\
\left[C_i\otimes C_j^*\right]^{*T}\left[D_i\otimes D_j^*\right]^{*T}\left[G_i\otimes G_j^*\right]^{*T} \\
W_{\mathrm{ij}} \\
\left[G_i\otimes G_j^*\right]\left[D_i\otimes D_j^*\right]\left[C_i\otimes C_j^*\right] \\
\left[E_i\left(\underline{\rho}_k\right)\otimes E_j^*\left(\underline{\rho}_k\right)\right]\left[P_i\left(\underline{\rho}_k\right)\otimes P_j^*\left(\underline{\rho}_k\right)\right]\left[F_i\left(\underline{\rho}_k\right)\otimes F_j^*\left(\underline{\rho}_k\right)\right] \\
S
\end{gathered} \tag{3}
$$

It is not immediately obvious how to construct dirty and residual images for this measurement equation, since the Fourier inverse does not apply. Actually it is not clear just what a dirty image is supposed to be. I argued previously that a good definition of a *generalized* dirty image $\vec{\mathcal{I}}_k^D$ is:

$$
\vec{\mathcal{I}}_k^D = -\left[\frac{\partial^2\chi^2}{\partial\vec{\mathcal{I}}_k\partial\vec{\mathcal{I}}_k^T}\right]^{-1}\frac{\partial\chi^2}{\partial\vec{\mathcal{I}}_k}\,|_{\vec{\mathcal{I}}_k=0} \tag{4}
$$

Note that the first term on the RHS of this equation is the inverse of a 4 by 4 matrix. By inverting this matrix, we are correcting for the coupling of different polarizations in the interferometer. By ignoring the non-diagonal terms of the Hessian, we are ignoring the coupling between different pixels in the final image. This is reasonable since, first, the coupling is singular, and, second, it is the role of an Image Solver to correct for this coupling.

A *generalized* residual image can be similarly defined as the update direction for a given estimate of the sky brightness, $\vec{\mathcal{I}}$. Thus, as is reasonable, the residual image tends towards zero as the model image reproduces the observed visibility data. Note that since the residual image is defined in terms of the gradients and Hessian, these are the primary interface to the measurement equation. In fact, it makes good sense to think of *residual image* as an image solver of status no different from any other image solver.

In addition, there is a PSF $B(\underline{\rho},\underline{\rho}')$ associated with the generalized dirty image. Conceptually, it is calculated by propagating an appropriately centered point source, at position $\underline{\rho}'$ through the measurement equation to obtain the predicted visibility and then back into the image plane, to position $\underline{\rho}$, via the equation for the dirty image. Thus the PSF is not necessarily shift-invariant. Note also that the PSF is a (4 by 4) matrix for each point in $(\underline{\rho},\underline{\rho}')$ space (!) while the dirty and residual images are images in which each pixel is a 4-vector. This means that any image solver must estimate a 4-vector at every pixel, and that finding the PSF requires evaluating the response separately for the 4 basis vectors (*e.g.* once for each of I, Q, U, V). The Stokes parameters, I, Q, U, V are therefore coupled together both in the calibration and in the solution for the image. Note that the "normal" approach to imaging is to correct coupling between Stokes parameters in the measurements and then to deconvolve separately. Here the decoupling and deconvolution are done simultaneously. At this point, it may be a good idea to reassure the reader that conventional imaging can be acheived using this formalization by making the appropriate approximations: *e.g.* the PSF matrix is diagonal and shift-invariant, in which case one is justified in using a straightforward application of CLEAN to each of I, Q, U, V.

Let us now consider how to use *a priori* information that applies to the 4-vector $\vec{\mathcal{I}}$.

4 Projection Onto Convex Sets

Apart from the dirty image, the simplest class of Image Solvers are those that use the principle of Projection Onto Convex Sets. POCS is a simple but very general iterative algorithm. If one wants to solve a linear equation $AX = Y$ then one uses an iterative algorithm given by successive and repeated applications of various projection operators:

$$\Delta X = T_1 \left(T_2 \left(T_3 ... T_N \left(B - AY \right) ... \right) \right)$$

where the projection operators T_i impose some form of *a priori* constraint *e.g.* positivity, finite support. CLEAN is a POCS algorithm. A POCS algorithm is guaranteed to converge for convex set.

In the context of the GI measurement equation, $Y - AX$ is the residual image $\vec{\mathcal{I}}^R$. The projection operators are open to choice. There are a number of possible operators:

Finite support The finite support of the object can be represented by a projection operator, T^{supp}, that simply sets to zero all parts of estimate that are outside the region of support.

Positivity The total intensity must be positive:

$$I \geq 0 \tag{5}$$

The corresponding projection operator, T^{pos} therefore projects invalid vectors onto the plane $I = 0$.

Fractional polarization An analogy to positivity for the 4-vector $\vec{\mathcal{I}}$ is that the fractional polarization not exceed unity. Or:

$$I^2 - Q^2 - U^2 - V^2 \geq 0 \tag{6}$$

This describes a hypercone in I, Q, U, V space. For points outside the cone, projection onto the surface of the cone can be done in a number of ways. The simplest operator, T^{cone}, sets to zero all elements outside the cone. Alternatively one could boost the I component until the point lies inside the cone, T^{boost}. The most appropriate operator, T^{shrink}, is probably that which projects perpendicular to the I axis so as to shrink Q, U, V to be less than or equal to I.

A terser, and perhaps useful equation expressing this constraint is that:

$$\vec{\mathcal{I}}^T M \vec{\mathcal{I}} \geq 0 \tag{7}$$

where the matrix M is given by:

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{8}$$

One can also think of polarized radiation in terms of the coherence matrix:

$$\mathcal{B} = \frac{1}{2} \begin{pmatrix} I+V & Q+iU \\ Q-iU & I-V \end{pmatrix} \tag{9}$$

in which case the equivalent statement is that this matrix be positive semi-definite, and so the product of eigenvalues is non-negative:

$$\lambda_1 \lambda_2 \geq 0 \tag{10}$$

Pointedness The operator T^{point} returns the maximum as a delta function, suitably scaled. Used in conjuction with the finite support operator, this yields the Hogböm CLEAN algorithm. I will say more on CLEAN below.

Clumpiness The operator T^{clump} returns a clump of components greater than some fraction of the peak. This is analogous to the Steer-Dewdney-Ito variant of CLEAN.

5 The Maximum Entropy Method

Generalizing MEM is quite easy: one maximizes the entropy of $\vec{\mathcal{I}}$ subject to χ^2 taking some specified value, and possibly that the integral of $\vec{\mathcal{I}}$ be equal to some value. Nityananda and Narayan (1982) show that for an entropy measure $H()$, the entropy of polarized radiation is given the sum of the entropies for the two independent polarization states of the radiation (i.e. eigenvalues of the coherence matrix $\mathcal{B}$.)

$$H\left(\vec{\mathcal{I}}\right) = H\left(I + \sqrt{Q^2+U^2+V^2}\right) + H\left(I - \sqrt{Q^2+U^2+V^2}\right) \tag{11}$$

One then calculates the gradient of H with respect to $\vec{\mathcal{I}}$ and in combination with the gradient of χ^2, generates a search direction with which to update the current estimate of $\vec{\mathcal{I}}$ (see *Holdaway*, 1990 for an algorithm).

6 CLEAN

The original Hogböm CLEAN algorithm can be considered as a pattern recognition procedure whereby the strongest beam pattern was searched for in the residual image. The best match was obtained by cross-correlating the residual image with the PSF, which for uniform weighting can be approximated by looking for the peak in the residual image alone. In the 4-dimensional space used here, it is not clear just what is meant by the peak. One possible definition of the peak is of the length of the vector $\vec{\mathcal{I}}_k$: $I^2+Q^2+U^2+V^2$. With such a choice, the CLEAN image solver would know nothing about the physical interpretation of the components of $\vec{\mathcal{I}}$, in particular that $\vec{\mathcal{I}}^T M \vec{\mathcal{I}} \geq 0$. This is somewhat strange but is analogous to the case of CLEANing separately the dirty images for the different Stokes parameters.

To introduce some physics, one could use the maximum eigenvalue of the coherence matrix: $I + \sqrt{Q^2+U^2+V^2}$ or perhaps the determinant: $I^2 - Q^2 - U^2 - V^2$. The latter form is blind to totally polarized emission so the former choice is probably more useful. In either case, one would probably want to add an additional criterion that both eigenvalues be non-negative. Hence a reasonable prescription would be to search for a peak in $I + \sqrt{Q^2+U^2+V^2}$ for those dirty image points for which $I^2 - Q^2 - U^2 - V^2 \geq 0$.

For the GI, the PSF can vary considerably with position, and so searching for the peak could be considerably more computationally expensive.

7 Model fitting

It is common to want to represent $\vec{\mathcal{I}}$ by a set of discrete components such as Gaussians. If so, then the optimization would require gradients and the Hessian with respect to the parameters of the components rather than the pixel strengths as assumed above. Thus a model-fitter does not fit my simple definition of an Image Solver as that requires either a residual image or the gradient and inverse of the diagonal elements of the Hessian. It is important to understand the origin of this failing. It arises because the measurement equation is really an integral of complicated form that requires in most cases numerical integration. Consider, for example, a Gaussian observed by an interferometer with empirically determined primary beam patterns. In general there will not be an analytic expression for the predicted visibility and so it will be necessary to use numerical integration. The pixels in an image can then be viewed as a mechanism for numerical integration of the Measurement Equation. Hence one can use Gaussian models but I think that one is lead to numerical integration and hence the use of a pixellated image in conjunction with the chain rule. One can think of special cases that are exceptions (*e.g.* when a Gaussian is smaller in angular size than any of the size scales of the image plane calibration matrices, but in general numerical methods will be required.

8 Comments

This memo, like the previous one, covers some entirely new ground. Is this a good thing? It looks more like a research paper than a description of a well-posed problem for which software is to be written. My view is that with a new formalism come new possibilities. The old algorithms can be implemented on top of the new formalism. For example, one could clean all four elements of the coherence vector independently, as is now done. AIPS++ should provide that capability. However, the point of this memo has been to explore what new things can be done using the HBS formalism. It is clear that a lot of new possibilities for imaging will spring up. This is, of course, what we hope will occur for many different areas of AIPS++.

References

Cornwell, T.J., 1995, AIPS++ Implementation Note 183.
Hamaker, J.P., Bregman, J.D., and Sault, R.J., 1995, *Understanding radio polarimetry: I Mathematical foundations*, submitted to A&A.
Holdaway, M.A., 1990, Ph.D. Thesis, *Brandeis University.*
Noordam, J., 1995, AIPS++ Implementation Note 182.