The Generic Interferometer: I Overview of Calibration and Imaging
AIPS++ Implementation Note 183

T.J. Cornwell, NRAO

1995, August 22

Contents

1 **Introduction** 1

2 **The Measurement Equation for a Generic Interferometer** 2

3 **Calibration of the visibility** 5

4 **The Generalized Dirty Image** 7

5 **Image and Calibration Solvers** 10

6 **Open and unresolved questions** 11

7 **Summary** 11

1 Introduction

In a beautiful paper, Hamaker, Bregman and Sault (1995) described a new formalism for understanding and analysing the measurement of polarized radiation by radio-interferometers. Although the HBS paper was restricted to a point source, Bregman later developed the formalism for extended objects. Noordam (1995) pointed out that this leads to a measurement equation for a Generic Interferometer (GI), that provides a means of unifying the polarization calibration and imaging schemes for many different synthesis arrays.

In previous notes, I described mathematical approaches to imaging and calibration based upon a number of simple versions of the measurement equation (Cornwell, 1992a, 1992b). In this document, I will revisit the analysis performed in those two documents in the light of the new measurement equation.

I will avoid statements about design of software based upon these equations. Instead, analysis and design will be deferred to a later document. Hence I will try to separate concepts that are distinct, a discipline that will be useful later during the design phase.

The purpose of this document is to consider in more detail than HBS how calibration and imaging is to be performed for the Generic Interferometer. The formalism is very general and the equations seem somewhat different from those we have seen previously in interferometry. I will show that in fact straightforward generalizations of existing methods of calibration and imaging are possible. In calibration, one can derive update equations that resemble those usually seen. In imaging, the usual concepts of dirty image, residual image, and PSF can be similarly extended. I will show that the general formulas thus derived can be simplified to special cases to yield familiar results.

2 The Measurement Equation for a Generic Interferometer

I will restate the HBS measurement equation as described by Noordam but with a few additions and simplifications to clarify the physical content of the equation. In addition, I wish to correct what I believe to be mistakes and omissions in his formulation.

A Generic Interferometer measures cross correlations between two channels per feed, leading to a four-term cross-correlation that we write as a vector, $\vec{V}$, subscripted by the pair of feeds i, j:

$$\vec{V}_{\mathrm{ij}} = \begin{pmatrix} V_{\mathrm{pp}} \\ V_{\mathrm{pq}} \\ V_{\mathrm{qp}} \\ V_{\mathrm{qq}} \end{pmatrix}_{ij} \tag{1}$$

This basic measureable can be decomposed as follows. The measured cross-correlation $\vec{V}$ is given by:

$$\vec{V}_{\mathrm{ij}} = \left[G_i \otimes G_j^*\right] \left[D_i \otimes D_j^*\right] \left[C_i \otimes C_j^*\right] \vec{\mathcal{V}}_{\mathrm{ij}} \tag{2}$$

where G_i, D_i, C_i are 2 by 2 matrices representing specific feed-based effects. The operator $\otimes$ represents a direct matrix product yielding in this case 4 by 4 matrices. G_i represents the complex gain of the i'th feed, D_i the leakage, and C_i is a fixed matrix representing the nominal feed configuration. Noordam gives examples of these terms. The terms $\left[G_i \otimes G_j^*\right]$, $\left[D_i \otimes D_j^*\right]$, and $\left[C_i \otimes C_j^*\right]$ are 4 by 4 matrices, and thus the formulation is 4-dimensional. Physically the matrices G_i, D_i, C_i represent coupling between the two polarization states for each feed and so the terms $\left[G_i \otimes G_j^*\right]$, $\left[D_i \otimes D_j^*\right]$, $\left[C_i \otimes C_j^*\right]$ represent coupling between the four polarization correlation states. In appendix A, I give a catalog of various forms for the calibration matrices.

A hint on reading these and succeeding equations: to get the more usual equations, just convert matrices to scalars, ignore the transposes, and convert the direct matrix product to a simple product. The equations derived below should then look quite familiar.

The vector $\vec{\mathcal{V}}$ represents the visibility that would be measured in the absence of the visibility-domain calibration effects. [1]

$$\vec{\mathcal{V}}_{\mathrm{ij}} = \begin{pmatrix} \mathcal{V}_{\mathrm{pp}} \\ \mathcal{V}_{\mathrm{pq}} \\ \mathcal{V}_{\mathrm{qp}} \\ \mathcal{V}_{\mathrm{qq}} \end{pmatrix}_{ij} \tag{3}$$

This includes only image-domain calibration effects. Thus, the feed-based effects must now be a function of direction, $\underline{\rho}$.

$$\vec{\mathcal{V}}_{\mathrm{ij}} = \sum_k \left[E_i\left(\underline{\rho}_k\right) \otimes E_j^*\left(\underline{\rho}_k\right)\right] \left[P_i\left(\underline{\rho}_k\right) \otimes P_j^*\left(\underline{\rho}_k\right)\right] \left[F_i\left(\underline{\rho}_k\right) \otimes F_j^*\left(\underline{\rho}_k\right)\right]$$
$$S\, \vec{\mathcal{I}}_k\, e^{-2\pi i \left(\underline{r}_i - \underline{r}_j\right) \underline{\rho}_k} \tag{4}$$

where $E_i\left(\underline{\rho}_k\right)$ represents the feed voltage receptivity pattern of the i'th feed (Noordam used B for this term), $P_i\left(\underline{\rho}_k\right)$ parallactic angle rotation, and $F_i\left(\underline{\rho}_k\right)$ atmospheric terms including tropospheric and ionospheric phase (Noordam used this term for Faraday rotation alone; here I note that Faraday rotation can also be accomodated in F). The term $\vec{\mathcal{I}}$ represents the polarized sky brightness. In the Stokes representation, this is given by:

$$\vec{\mathcal{I}} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} \tag{5}$$

The matrix S converts from the representation used in $\vec{\mathcal{I}}$ to that most naturally used in describing the interferometer. HBS use linear polarization as the canonical representation, for which the S-matrix is:

$$S_{\mathrm{linear}} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & i \\ 0 & 0 & 1 & -i \\ 1 & -1 & 0 & 0 \end{pmatrix} \tag{6}$$

[1] In Noordam's note, $\vec{\mathcal{V}}$ is called Σ and is mistakenly called a matrix. It actually is a visibility vector.

It is worth emphasizing that this is an arbitary choice and that one could, instead, use circular polarization, in which case the S-matrix would be:

$$S_{\mathrm{circular}} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & i & 0 \\ 0 & 1 & -i & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \tag{7}$$

It may seem that C_i and S are redundant since one can choose S so that C_i is a unit matrix. However, I follow HBS and choose S to be the canonical S^{lin}. Consequently, C_i must be allowed to vary to suit the actual measurement scheme of the feed. This flexibility is required for a system of interferometers in which a mixture of linear and circular polarization is measured. Note that the calibration matrices will then inevitably be more complicated than would be the case if a single, natural representation was used.

To emphasize the true glory of the full measurement equation, here I give the whole expression:

$$\begin{aligned} \vec{V}_{\mathrm{ij}} = & \left[G_i \otimes G_j^*\right] \left[D_i \otimes D_j^*\right] \left[C_i \otimes C_j^*\right] \\ & \textstyle\sum_k \left[E_i\left(\underline{\rho}_k\right) \otimes E_j^*\left(\underline{\rho}_k\right)\right] \left[P_i\left(\underline{\rho}_k\right) \otimes P_j^*\left(\underline{\rho}_k\right)\right] \left[F_i\left(\underline{\rho}_k\right) \otimes F_j^*\left(\underline{\rho}_k\right)\right] \\ & S\, \vec{\mathcal{I}}_k\, e^{-2\pi i \left(\underline{r}_i - \underline{r}_j\right) \underline{\rho}_k} \end{aligned} \tag{8}$$

It is worth making a number of comments about this very general form of the measurement equation:

Ordering of terms The ordering of terms follows the signal path, reading right to left in the equation.

Linearity The ME is linear in the sky brightness $\vec{\mathcal{I}}$. It is almost always non-linear in various calibration parameters. This will have implications for solvers, as I discuss below.

Time and Frequency The formalism as presented here ignores indexing by time and frequency. Both of these are trivial to add but obscure the notation. Averaging over both time and frequency is then accomodated easily.

No new physics There is no new physics in this formulation. The equations are totally equivalent to those derived earlier (see *e.g.* Schwab, 1984). However, I expect that the formulation will suggest new forms of calibration and imaging.

Use of these equations In general, one will want to determine one or both of the calibration parameters and the sky brightness. To determine the calibration, one will fix the sky brightness model, and solve for the calibration parameters, perhaps by a least squares approach. The determine the sky brightness, one will want to fix the calibration parameters and solve for the sky brightness, using a deconvolution algorithm. One could alternate between calibration and sky brightness estimation, as is done in most self-calibration procedures, or one could do a joint solution by some very powerful least squares type algorithm.

Analytical forms for $\vec{\mathcal{I}}$ In the calibration cycle, one will have to perform an integration over $\underline{\rho}$. This will in general be quite difficult and usually, the integration will have to be performed numerically. The "art" of programming this equation will principally be in finding quick methods of numerical integration using, for example, FFTs.

Non-invertibility The ME is clearly not invertible for either the sky brightness or the calibration matrices. This is really a truism since it represents only one sample.

Self-cal assumption The fundamental assumption yielding the direct products is that all these instrumental effects factorize per feed. It is hard to think of any exceptions to this, apart from problems with the correlation process itself. Note that failure of closure due to different bandpasses at the different feeds is actually accomodated by the summation over frequency.

Non-isoplanatism The location of the atmospheric phase terms is actually ambiguous in most cases. Physically it belongs with the F term. In isoplanatic conditions, it cannot be distinguished easily from the electronic term G, and so it is usually written as belonging in G. However, if the atmosphere is non-isoplanatic then the atmospheric phase must be included in the F-term. F then actually decomposes a follows:

$$F_i\left(\underline{\rho}_k\right) = F_i\left(\underline{\rho}_k\right)^{troposphere} F_i\left(\underline{\rho}_k\right)^{ionosphere} \tag{9}$$

Assumption of perfect correlation Note that the correlation process itself introduces errors that must be corrected, such as the van Vleck correction. These we ignore for the moment.

Completeness of information It is assumed that the full 4-dimensional visibility is measured. In many cases, this is not so and, for example, only the parallel hands will be correlated. Full correction is not then possible but one expects that good approximations will be obtained in many cases simply by inserting zeroes in the appropriate 4-vectors. A similar but distinct difficulty arises in those interferometers that measure different hands at different times. Such wrinkles can be dealt with straightforwardly but further discussion is deferred to a specifications document.

3 Calibration of the visibility

Calibration involves solving for the possibly unknown feed-based matrices such as G_i, D_i from calibration observations[2]. Imaging means estimating $\vec{\mathcal{I}}$ from data corrected for the calibration matrices. Note that parametrization of the free variables is probably wise. For calibration, this means that, for example, G_i is diagonal with parameters a_p, a_q, ϕ_p and ϕ_q describing the amplitude and phase of the gain in the two polarization channels. For imaging, the parameters could be the pixels of an image or the parameters of a number of Gaussian components.

If the image parameters are known for some subset of the measurements, then one can solve for the calibration parameters and thus calibrate. Then from another subset one can solve for image parameters of another object.

Suppose that we have a number of measurements $\vec{V}$ of the visibility function, and that furthermore we have estimates of the visibility function $\vec{V}$ that would be measured in the absence of visibility-based calibration effects. We can then estimate the calibration matrices by a least squares fit whereby the matrices are adjusted to fit the data in a least-squares sense. Define an error norm χ^2:

$$\chi^2 = \sum_{ij} \Delta\vec{V}_{\mathrm{ij}}^{*T} W_{ij} \Delta\vec{V}_{\mathrm{ij}} \tag{10}$$

where the residual is given by:

$$\Delta\vec{V} = \vec{V} - \widehat{\vec{V}} \tag{11}$$

and $\widehat{\vec{V}}$ denotes an estimate:

$$\widehat{\vec{V}} = \left[G_i \otimes G_j^*\right]\left[D_i \otimes D_j^*\right]\left[C_i \otimes C_j^*\right]\widehat{\vec{\mathcal{V}}}_{\mathrm{ij}} \tag{12}$$

and W_{ij} is a (4 by 4) weight matrix. For natural weighting, W will usually be diagonal with elements given by the inverse of the corresponding variance. However, it is more strictly the inverse of the covariance matrix of the errors. For uniform weighting, the usual correction for the local density of samples will be applied.

Choosing the calibration matrices that minimize χ^2 is a non-linear least squares problem. To solve it, we will need the first and second derivatives of χ^2 with respect to the calibration matrices. Since this gets quite complicated, I've deferred complete exposition to another memo. For the moment, I say only that to check these formulas, I generated the scalar update equation for the case where only the gain matrix need be calculated. I do indeed obtain the update formula used in the usual feed-based gain solution algorithms.

[2]Note that the matrix C_i denotes the nominal feed configuration and thus does not need to be estimated

This is all good as far as it goes, but we probably will prefer to parametrize the calibration matrices by a small number of parameters. We will then require derivatives with respect to these parameters, instead of the elements of the matrices directly. By applying the chain rule, this is quite straightforward, if tedious.

Having solved for the calibration matrices, we can now correct for them:

$$\vec{\mathcal{V}}_{\mathrm{ij}} = \left[C_i \otimes C_j^*\right]^{-1} \left[D_i \otimes D_j^*\right]^{-1} \left[G_i \otimes G_j^*\right]^{-1} \vec{V}_{\mathrm{ij}} \tag{13}$$

The other calibration matrices, $E_i\left(\underline{\rho}_k\right)$, $P_i\left(\underline{\rho}_k\right)$, $F_i\left(\underline{\rho}_k\right)$, cannot be corrected in the visibility domain. Instead, it is necessary to make an image. This I describe next.

4 The Generalized Dirty Image

Imaging from the measurement equation requires some careful analysis. We have to accept that there is no inverse to the measurement equation. The *dirty* image is simply one means of estimating the sky brightness from the measurements. It is convenient because it is linear in the measurements. However, giving it special status obscures the underlying simplicity of imaging.

The dirty image is known to be useful in imaging from simple interferometric observations. Here I will shown how to generalize it to this more complicated measurement equation.

We vary the estimate of the sky brightness $\vec{\mathcal{I}}_k$ to minimize the function χ^2. For this we need to evaluate $\frac{\partial \chi^2}{\partial \vec{\mathcal{I}}_k}$. Old hands at this sort of stuff will recognize that it looks like:

$$\begin{aligned}\frac{\partial \chi^2}{\partial \vec{\mathcal{I}}_k} = \quad & -2\, \Re \textstyle\sum_{ij} S^{*T} \\ & \left[F_i\left(\underline{\rho}_k\right) \otimes F_j^*\left(\underline{\rho}_k\right)\right]^{*T} \left[P_i\left(\underline{\rho}_k\right) \otimes P_j^*\left(\underline{\rho}_k\right)\right]^{*T} \left[E_i\left(\underline{\rho}_k\right) \otimes E_j^*\left(\underline{\rho}_k\right)\right]^{*T} \\ & \left[C_i \otimes C_j^*\right]^{*T} \left[D_i \otimes D_j^*\right]^{*T} \left[G_i \otimes G_j^*\right]^{*T} \\ & W_{\mathrm{ij}}\, \Delta\vec{V}_{\mathrm{ij}}\, e^{2\pi i\left(\underline{r}_i - \underline{r}_j\right)\underline{\rho}_k}\end{aligned} \tag{14}$$

It is still a long way from this to the dirty image. To make a start, set the initial model to be zero. $\Delta\vec{V}$ then becomes $\vec{V}$. We can read the equation from the inside out as:

- Apply the transposed complex conjugate of the gain $\left[G_i \otimes G_j^*\right]$,
- Apply the transposed complex conjugate of the D-terms $\left[D_i \otimes D_j^*\right]$,

- Apply the transposed complex conjugate of the feed configuration $\left[C_i \otimes C_j^*\right]$,
- Apply the transposed complex conjugate of the primary beam $\left[E_i\left(\underline{\rho}_k\right) \otimes E_j^*\left(\underline{\rho}_k\right)\right]$,
- Apply the transposed complex conjugate of the parallactic angle $\left[P_i\left(\underline{\rho}_k\right) \otimes P_j^*\left(\underline{\rho}_k\right)\right]$,
- Apply the transposed complex conjugate of the atmospheric phase term $\left[F_i\left(\underline{\rho}_k\right) \otimes F_j^*\left(\underline{\rho}_k\right)\right]$,
- Apply the transposed complex conjugate of the conversion from Stokes representation,
- Apply the inverse Fourier transform for this sample,
- Sum over all visibility samples with complex weights, *i.e.* the Fourier Transform for one sample.

The presence of all the transposes is a clue that we are still not quite there. For unitary matrices, the transpose of the complex conjugate is the inverse, and so we can recognize the correction of the various effects. However, there remains a normalization problem. Furthermore, for non-unitary matrices, we have also to correct for the amplitude of the various matrices. These points are connected to the fact that the gradient is not a good search direction for a least squares method. It is better to pre-multiply by the inverse of the Hessian. Remember that the Hessian has elements: $\frac{\partial^2\chi^2}{\partial\vec{\mathcal{I}}_k\partial\vec{\mathcal{I}}_l^T}$ Since, in general, this is not practicable we make do with the inverse of the diagonal elements: $\frac{\partial^2\chi^2}{\partial\vec{\mathcal{I}}_k\partial\vec{\mathcal{I}}_k^T}$ With a bit more heavy lifting, we get that it must be:

$$\begin{aligned}
\frac{\partial^2\chi^2}{\partial\vec{\mathcal{I}}_k\partial\vec{\mathcal{I}}_k^T} = \; & 2\,\Re\sum\nolimits_{ij} S^{*T} \\
& \left[F_i\left(\underline{\rho}_k\right) \otimes F_j^*\left(\underline{\rho}_k\right)\right]^{*T} \left[P_i\left(\underline{\rho}_k\right) \otimes P_j^*\left(\underline{\rho}_k\right)\right]^{*T} \left[E_i\left(\underline{\rho}_k\right) \otimes E_j^*\left(\underline{\rho}_k\right)\right]^{*T} \\
& \left[C_i \otimes C_j^*\right]^{*T} \left[D_i \otimes D_j^*\right]^{*T} \left[G_i \otimes G_j^*\right]^{*T} \\
& W_{\mathrm{ij}} \\
& \left[G_i \otimes G_j^*\right] \left[D_i \otimes D_j^*\right] \left[C_i \otimes C_j^*\right] \\
& \left[E_i\left(\underline{\rho}_k\right) \otimes E_j^*\left(\underline{\rho}_k\right)\right] \left[P_i\left(\underline{\rho}_k\right) \otimes P_j^*\left(\underline{\rho}_k\right)\right] \left[F_i\left(\underline{\rho}_k\right) \otimes F_j^*\left(\underline{\rho}_k\right)\right] \\
& S
\end{aligned} \tag{15}$$

An appropriate generalization of the dirty image can be defined as the initial update direction for a least squares fit. Applying the Newton-Raphson approach with the approximation that only the diagonal terms are retained, we therefore have that:

$$\vec{\mathcal{I}}_k^D = -\left[\frac{\partial^2\chi^2}{\partial\vec{\mathcal{I}}_k\partial\vec{\mathcal{I}}_k^T}\right]^{-1}\frac{\partial\chi^2}{\partial\vec{\mathcal{I}}_k}\,|_{\vec{\mathcal{I}}_k=0} \tag{16}$$

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 a deconvolution algorithm or Image Solver (see below) to correct for this coupling.

A residual image can be similarly defined as the update direction for a given estimate of the sky brightness, $\vec{\mathcal{I}}$.

Again, we should check that these equations give the right answers for known, simple cases. Let us consider imaging of **Stokes parameter I**. The generalized dirty image, as defined in equation 16, is then simply the familiar dirty image:

$$I_k^D = \frac{\sum_{ij} W_{\rm ij}\,\Re\Big(V_{\rm ij}\; e^{2\pi i\left(\underline{r}_i-\underline{r}_j\right)\underline{\rho}_k}\Big)}{\sum_{ij} W_{\rm ij}} \tag{17}$$

where I, W and V are now all scalars.

The above-mentioned approximation of dividing by the inverse of the diagonal elements of the Hessian thus corresponds in this case to normalizing by the summed weights, something that is common in constructing a dirty image. This is a good approximation for a point source[3] and so the strength of an isolated point source in the generalized dirty image is correctly represented.

Let us turn now to a case for which the full vector formulation is needed: for **full polarized observations with gain errors**, the generalized dirty image is:

$$\begin{aligned}\vec{\mathcal{I}}_k^D = \;& \left[\sum_{ij}\Re\Big(S^{*T}\left[G_i\otimes G_j^*\right]^{*T} W_{\rm ij}\;\left[G_i\otimes G_j^*\right]S\Big)\right]^{-1} \\ & \sum_{ij}\Re\Big(S^{*T}\left[G_i\otimes G_j^*\right]^{*T} W_{\rm ij}\;\vec{V}_{\rm ij}\; e^{2\pi i\left(\underline{r}_i-\underline{r}_j\right)\underline{\rho}_k}\Big)\end{aligned} \tag{18}$$

This looks a little more curious and unexpected: the formalism requires that one corrects the gains only on average, not per sample as is usually done, and as arose in the previous section. Furthermore, the correction for the coupling between the different polarization channels is performed by a matrix inversion and multiplication in the image plane. A little thought tells one that this is

[3] I note that for an extended source, one might wish to use a different approximation for the inverse Hessian

actually the correct prescription so, as Jan Noordam says in his memo, our calibration procedures do benefit from some careful scrutiny.

Finally, let us look at the case of **primary beam correction**, and its big brother, mosaicing. Here the vector formulation yields:

$$\vec{\mathcal{I}}^D_k = \left[\sum_{ij} \Re\left(S^{*T}\left[E_i\left(\underline{\rho}_k\right)\otimes E^*_j\left(\underline{\rho}_k\right)\right]^{*T} W_{\mathrm{ij}}\ \left[E_i\left(\underline{\rho}_k\right)\otimes E^*_j\left(\underline{\rho}_k\right)\right] S\right)\right]^{-1} \sum_{ij} \Re\left(S^{*T}\left[E_i\left(\underline{\rho}_k\right)\otimes E^*_j\left(\underline{\rho}_k\right)\right]^{*T} W_{\mathrm{ij}}\ \vec{V}_{\mathrm{ij}}\ e^{2\pi i\left(\underline{r}_i-\underline{r}_j\right)\underline{\rho}_k}\right) \tag{19}$$

Thus again, the primary beam division is done only on average. However, unlike the case of gain errors, this result is somewhat familiar from mosaicing theory (see *e.g.* Cornwell, Holdaway, and Uson, 1993), except that it is a vector formulation. In this case, one has a linear method that corrects for an image plane effect in an intuitive way.

In summary, I have shown that the GI measurement equation can be used to define a generalized Dirty Image that seems to be quite useful in representing the sky brightness at some level of approximation. Note that the so-called image plane effects, the primary beam, the parallactic angle and the atmospheric phase effects, can be corrected without going through a full non-linear deconvolution. Another way of understanding this evidently unexpected result is that the dirty image is best viewed as the result of one form of deconvolution, special only in its linearity. This same point is made in my memo on the A-matrix formalism (Cornwell, 1992a) but is presented more explicitly here.

5 Image and Calibration Solvers

The sky brightness and the calibration matrices now appear in the formalism on an equal basis. This is in accord with our experience that one can use interferometers to image the sky or to calibrate the interferometer.

Deconvolution can now be described as a processing of solving for the sky brightness $\vec{\mathcal{I}}$. It is probably better to use a term such as *Image Solver* since the connection to classic deconvolution is getting harder and harder to follow. Image Solvers such as CLEAN, including all variants such as the Clark and Schwab-Cotton algorithms, and MEM, can be written in terms of χ^2 and the gradient terms described above. Thus the machinery for deconvolution can be separated from the machinery for the measurement equation. This is the essence of abstraction.

Solving for calibration is conceptually much more straightforward than imaging. Let us assume that we have a model for the sky brightness, $\widehat{\vec{\mathcal{I}}}$. We then find those calibration matrices that minimize χ^2. I described above, how one might imagine doing this by a simple gradient search method analogous to that using in imaging. One caution is in order: since the calibration matrices are parametrized by other free variables and since such parametrizations are likely to be strongly non-linear, it is likely that a simple approach like that advocated for imaging will be ineffective and so a more complicated algorithm may be required. Even if this is true, it is probably also true that all that is required from the measurement equation is knowledge of first and second derivatives.

6 Open and unresolved questions

1. Can the coherence matrix be used as the prime observable instead of the coherence vector? The mixed product rule cannot be applied directly but the connection should be possible. Does it have any advantages? One advantage is that the eigenvalues of the coherence matrix are the two independent states of the radiation field and so, for example, one obtains the entropy of a polarized radiation field by summing the entropies of the two eigenvalues. Presumably something similar must apply to CLEAN. This question, therefore, is most important when considering Image Solvers.

2. Can this formalism be usefully applied to single dish observations? For mosaicing, this would be convenient. One would need to add terms for total power offsets that are neglected in most synthesis applications.

3. Does this formalism lend itself naturally to the expression of fringe fitting algorithms? The answer is almost certainly yes.

7 Summary

The GI measurement equation lends itself to the following:

- Conceptually simple (but mathematically elaborate) calibration algorithms based upon derivatives of the misfit, χ^2, with respect to the calibration matrices.

- A generalized dirty image in which image plane effects can be corrected to some degree,

- Image Solvers that interface to the measurement equation solely through derivatives of the misfit, χ^2.

In all cases, the general equations can be simplified to yield familiar results, thus reassuring us of the basic correctness of this approach.

Acknowledgements

Fred Schwab provided invaluable advice and references on Matrix Calculus and calibration.

References

Cornwell, T.J., 1992a, AIPS++ Implementation Note 147.
Cornwell, T.J., 1992b, AIPS++ Implementation Note 148.
Cornwell, T.J., Holdaway, M.H., and Uson, J.M., 1993, Astron. & Astrophys. **271**, 697-713.
Hamaker, J.P., Bregman, J.D., and Sault, R.J., 1995, *Understanding radio polarimetry: I Mathematical foundations*, submitted to A&A.
Noordam, J., 1995, AIPS++ Implementation Note 182.
Schwab, F.R., 1984, *Comment on the Use of Linearly Polarized Feeds for the VLBA*, VLBA memo 337.

Appendix A: Catalog of Matrices

Here I repeat Noordam's description of the various matrices:

- G_i represents complex gain (amplitude, phase) per IF channel. It includes the effects of all the electronics after the feed (amplifiers, mixers, LO, cables etc), excluding the correlator of course.

$$G_i = \begin{pmatrix} g_{i\mathrm{p}} & 0 \\ 0 & g_{i\mathrm{q}} \end{pmatrix} \tag{20}$$

- D_i represents the signal 'leakage' between the two receptors in a feed, due to deviations from nominal. The real part corresponds roughly to a dipole alignment error, and the imaginary part to ellipticity.

$$D_i = \begin{pmatrix} 1 & d_{i\mathrm{p}} \\ -d_{i\mathrm{q}} & 1 \end{pmatrix} \tag{21}$$

  NB: Other processes may also contribute to leakage. For example, cross-talk between the two IF-channels can be described by adding a non-zero mutual coupling factor c_i to the off-diagonal terms.

- C_i represents the nominal feed configuration, which includes rotation of the dipoles w.r.t. the antenna (e.g. WSRT) , and conversion from linear to circular polarisation (e.g. VLA).

- E_i represents the feed beam response. For the important case of a dominating compact source in the centre of the field (and an axially symmetric antenna-feed combination), it is the unit matrix.

- P_i represents parallactic angle ϕ, i.e. rotation of the antenna w.r.t. the sky. For an *equatorial* antenna like used in the WSRT, it is a unit matrix. For an *alt-az* antenna, ϕ varies smoothly with Hour Angle (HA), as a function of LATitude and DEClination:

$$P_i^{alt-az} = \begin{pmatrix} \cos\phi(t) & \sin\phi(t) \\ -\sin\phi(t) & \cos\phi(t) \end{pmatrix} \tag{22}$$

- F_i represents (ionospheric) Faraday rotation.

$$F_i = \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \tag{23}$$