Radio polarimetry and selfcal in the 21st century: Harnessing the power of simple matrix algebra

Johan Hamaker

Netherlands Foundation for Research in Astronomy NFRA/ASTRON

P.O. Box 2, 7990 AA Dwingeloo, Netherlands

E-mail:hamaker@astron.nl

1. INTRODUCTION

Conventional methods of dealing with polarization in interferometry are based on a nonlinear equation published in 1964. To make that equation tractable, linearising approximations are introduced which are based on the assumptions of low instrumental polarization and weakly polarized sources. Instrumental effects other than the feeds' orientations and ellipticities cannot conveniently be included in the formulas and their treatment in practical situations requires lengthy explanations. Although it does, in the end, lead to practical answers, this approach to polarimetry is no more than an ad-hoc engineering solution.

Fundamental physics demands that one considers the vector nature of the electromagnetic field from the very start and ignores the vain promise of simplicity that a scalar approach offers. The required mathematical tools are offered by the standard algebra of vectors and matrices. In terms of these, the standard interferometer equations retain their familiar form, but rather than scalars the variables now represent 2×2 matrices. This approach has been systematically developed in a series of papers: See Hamaker (2000) and references therein.

The latter paper develops a theory of *matrix self-calibration* and derives the conditions that its solutions obey. As in the scalar case, the theory provides little insight into the iteration process that is to produce the solution. The only way to address this matter is through practical examples — which are most readily created through computer simulations.

In this paper, I briefly summarise the matrix theory as it stands (secs. 2 and 3) and outline a program of computer simulations that is now underway to explore its practicality. It is hoped that some results of these simulations can be presented in the oral presentation in August.

The practical rationale for this work lies in the fact that the engineering solution is becoming untenable. On one hand, source polarizations get stronger with higher resolutions, as is already the case in VLBI and will be with the intruments now being planned. On the other hand, new types of radio telescopes such as *LOFAR* are being developed in which the antenna elements are stationary rather than pointed at the source of interest. Consequently, the paraxial conditions prevailing in observations with parabolic reflectors will no longer hold. The 'envelope beam' defined by the individual elements of a phased array is inherently strongly polarized in directions away from the optical axis.

1.1. Terminology and notation

The *signals* being considered are 2-vectors, either representing two mutually orthogonal components of the electric field or a voltage pair in the receiver system. The device that converts the field into a voltage pair is the *feed*.

The analogue of the scalar visibility is the *coherency*, a tensor that I represent in the form of a 2×2 matrix. It consists of four complex components that I shall occasionally refer to as *visibilities*.

The vectors are denoted by bold lowercase symbols; bold uppercase represents matrices. Constant scalars and vectors are shown in roman, variables in italic font. The 'dagger' superscript † stands for hermitian transposition or conjugation, i.e. transposition plus complex conjugation.

Primes are generally used to distinguish *observed* or *fitted* values from *true* ones; occasionally they will also be used to distinguish input and output of a transformation or values of one variable under different conditions.

Quantities pertaining to a single antenna carry a subscript j or k; coherencies pertaining to an interferometer get a double subscript jk. An additional index t distinguishes samples taken at different times.

Self-calibration is typically used to correct observations of strong sources in which receiver noise is relatively insignificant. Such noise introduces an element of uncertainty that must be accounted for by solving the equations in a least-squares sense rather than exactly. I ignore this aspect here.

2. COHERENCY-MATRIX FORMULATION OF INTERFEROMETRY

2.1. Signal vector and coherency matrix

The proper way to represent the electric field of radiation moving in the z direction of a cartesian coordinate frame is by a vector

$$\boldsymbol{e} = \begin{pmatrix} e_{\mathrm{x}} \\ e_{\mathrm{y}} \end{pmatrix} \tag{1}$$

The equivalent of the scalar visibility is the *coherency tensor*. It may be represented in various ways; the representation I use here is the *coherency matrix*.

$$\boldsymbol{E}_{jk} = <\boldsymbol{e}_j\, \boldsymbol{e}_k^{\dagger}> = \begin{pmatrix} <e_{j\mathrm{x}}e_{k\mathrm{x}}^{*}> & <e_{j\mathrm{x}}e_{k\mathrm{y}}^{*}> \\ <e_{j\mathrm{y}}e_{k\mathrm{x}}^{*}> & <e_{j\mathrm{y}}e_{k\mathrm{y}}^{*}> \end{pmatrix} \tag{2}$$

2.2. The interferometer equation

The elements in the signal path in one antenna linearly transform the electric field or voltage vector. The transformations may be represented by equations of the form

$$\boldsymbol{w}_j = \boldsymbol{J}_j \boldsymbol{e}_j \tag{3}$$

where $\boldsymbol{J}_j$ is called a *Jones matrix*. It is then readily seen that an interferometer with Jones matrices $\boldsymbol{J}_j$ and $\boldsymbol{J}_k$ transforms the coherency matrix according to

$$\boldsymbol{W}_{jk} = \boldsymbol{J}_j \boldsymbol{E}_{jk} \boldsymbol{J}_k^{\dagger} \tag{4}$$

Coherency and Jones matrices having the same form, it should be clear from the context which is which, just as in the scalar domain. In addition, note that Jones matrices carry the single index of an antenna whereas the coherency matrices have a double, interferometer index. This difference will remain also when we later add another index t for sampling time.

2.3. Matrix and Stokes brightnesses; intensity and polvector

The van Cittert-Zernike theorem relates sky brightness $B(\boldsymbol{l})$ and aperture visibility $E(\boldsymbol{r}_{jkt})$ by a Fourier transform. We may represent this for the present discussion as

$$B(\boldsymbol{l}) = \sum_{jkt} w(\boldsymbol{l}, \boldsymbol{r}_{jkt})\ E(\boldsymbol{r}_{jkt}) \tag{5}$$

where the $\boldsymbol{r}_{jkt}$ are the baseline vectors sampled at times t in the aperture, $\boldsymbol{l}$ is the position in the sky and the w represent values of the Fourier kernel.

The matrix equivalent is

$$\boldsymbol{B}(\boldsymbol{l}) = \sum_{jkt} w(\boldsymbol{l}, \boldsymbol{r}_{jkt})\ \boldsymbol{E}(\boldsymbol{r}_{jkt}) \tag{6}$$

where, as in (5), the Fourier kernel is scalar. This relation defines the *matrix brightness* $\boldsymbol{B}$ which I shall use below. Note that (6) represents four scalar transforms: between each brightness element of the matrix function $\boldsymbol{B}$ and the corresponding visibility element of the matrix function $\boldsymbol{E}$.

The physical content of $\boldsymbol{B}$ is more meaningfully described by the *Stokes brightness* defined by

$$\boldsymbol{B} = \begin{pmatrix} I+Q & U-iV \\ U+iV & I-Q \end{pmatrix} \tag{7}$$

where the scalar *intensity* I and the *polvector* $\boldsymbol{p} = (Q,\, U,\, V)$ are also functions of $\boldsymbol{l}$.

3. MATRIX SELF-CALIBRATION

For brevity, I present the following derivation directly in matrix form. To view it in terms with which he is more familiar, the reader may temporarily read all matrix symbols as scalars and recognise a description of the familiar scalar selfcal.

Two fundamental assumptions underlie self-calibration. One is that *the correlators are error-free*, so that (4) adequately represents the interferometers. The other is that *the sky is mostly empty*, the true sources accounting for only a small fraction of the observed field.

Under these conditions, self-calibration seeks to fit a *source-brightness model* $\boldsymbol{B}'(\boldsymbol{l})$ and *antenna Jones-matrix models* $\boldsymbol{J}'_{jt}$ to the observed visibilities $\boldsymbol{W}(\boldsymbol{r}_{jkt})$. In the coherency domain, $\boldsymbol{B}'$ is represented by its Fourier transform $\boldsymbol{E}'(\boldsymbol{r}_{jkt})$; hence the equations to be solved are

$$\boldsymbol{W}(\boldsymbol{r}_{jkt}) = \boldsymbol{J}'_{jt}\,\boldsymbol{E}'(\boldsymbol{r}_{jkt})\,\boldsymbol{J}'^{\dagger}_{kt} \tag{8}$$

Clearly, the *true* values of $\boldsymbol{E}(\boldsymbol{r}_{jkt})$ and the $\boldsymbol{J}_{jt}$ constitute a solution. Other solutions may be obtained by multiplying each $\boldsymbol{J}_{jt}$ with a matrix factor $\boldsymbol{X}^{-1}_{jt}$ and adjusting the $\boldsymbol{E}'$ accordingly; clearly there is thus a manifold infinity of such solutions.

It is well known for the scalar case that multiplying true source visibilities by random factors will produce an image in which brightness is scattered all over the place; the same is true in the matrix case. One way to constrain the image to a limited number of patches in an otherwise empty sky is by requiring all the $\boldsymbol{X}_{jt}$ to have the same value $\boldsymbol{X}$; then

$$\boldsymbol{E}' = \boldsymbol{X}\;\boldsymbol{E}\;\boldsymbol{X}^{\dagger} \qquad \text{and hence} \qquad \boldsymbol{B}' = \boldsymbol{X}\;\boldsymbol{B}\;\boldsymbol{X}^{\dagger} \tag{9}$$

I conjecture that these are the only solutions that fullfill the empty-sky requirement, but want to emphasise that a proof of this is lacking (as for scalar selfcal).

The *poldistortion* transformation (9) has the form of a *congruence transformation.* The two conditions imposed on the selfcal solution are insufficient to prevent it from appearing in our solution and provide us no clue as to its value.

It can be shown that the poldistortion transformation is a product of three elementary transformations:

- A positive scale factor (as in scalar selfcal).
- A *polrotation* $\boldsymbol{Y}$ (a *unitary* matrix) that leaves the intensity unchanged and *rotates the polvector* $\boldsymbol{p}$ in its three-dimensional space.
- A *polconversion* $\boldsymbol{H}$ (a *positive hermitian matrix*) that exchanges power between intensity $\mathbf{I}$ and the polvector $\boldsymbol{p}$.

3.1. Self-alignment and post-calibration; dynamic range and polarimetric fidelity

The solution being non-unique, self-*calibration* is actually a misnomer. Self-*alignment* describes more accurately what the procedure actually achieves. The solution (9) represents an *in-place* transformation of the brightness: It does not *scatter* radiation out of any point of the source into any other position in the image. Hence, like scalar selfcal, self-alignment conserves *dynamic range.*

To obtain a polarimetrically correct image, one must eliminate the poldistortion $\boldsymbol{X}$ in a process of *post-calibration.* To estimate $\boldsymbol{X}$, one confronts the models $\boldsymbol{J}'_{jt}$ and $\boldsymbol{B}'(\boldsymbol{l})$ with whatever prior knowledge one can get hold of, both of instrumental and of astrophysical nature.

3.2. Relation to 'classical' radio polarimetry

Naturally, the effects and their calibration discussed above must occur in any polarimetric measurement. For example, in 'classical' radio polarimetry polconversion and polrotation enter in the familiar forms of *leakage* and *phase-zero difference.* Their calibration involves the use of prior knowledge about the feeds (average errors should be near-zero) and about some of the sources (use of unpolarized calibrators, circular polarization is very weak).

Matrix post-calibration must rely on the same additional information. In a sense, it provides nothing fundamentally new. Unlike the linearised equations, however, those of the matrix theory are exact and this allows them to show a clear distinction between effects that appear as a confused mixture in the former equations.

4. MATRIX SELF-CALIBRATION IN PRACTICE?

AIPS++ has in 1995 endorsed an earlier form of the theory of Sec. 2 as one of its foundations, but it has until now ignored the subsequent theoretical development of matrix selfcal. Thus, the present implementation does not provide for an empirical test nor for investigating practical procedures. Many implementation questions remain to be answered, such as

- Is iterative self-calibration as apt to converge in the matrix domain as we know it to be in the scalar case?
- Is matrix selfcal as robust as its scalar peer? Theoretically, one may argue either way: Being more complicated, matrices may lead to more finicky algorithms *or* the greater connectivity of matrix equations could make them more robust.
- In the presence of poorly known Faraday rotation, how does one obtain an initial brightness image of linear polarization that can serve as a seed for the selfcal loop?
- In which ways can prior knowledge be injected into the procedure? Should post-calibration be included in each iteration or performed only as the final step?
- Are there tricks that speed up convergence and/or operations that should be avoided because they slow the process down?
- What is the effect of the statistical bias in the solutions that the noise brings about? (By the way, this question has hardly been touched upon for the scalar case as well.)

To investigate these matters, simulations are currently being developed in a MATLAB environment. At the time of this writing, first results are being obtained but it is too early to draw conclusions other than that the algorithms are in place and working. In the presentation in August, the theory described here will be supplemented by what the simulations should by then have revealed.

REFERENCE

Hamaker, J.P. "Understanding Radio Polarimetry, IV: "The full-coherency analogue of scalar self-calibration: Self-alignment, dynamic range and polarimetric fidelity". *Astron. and Astrophys. Suppl.*, **143**, 515-534, May I 2000.