Understanding radio polarimetry

IV. The full-coherency analogue of scalar self-calibration: Self-alignment, dynamic range and polarimetric fidelity

J.P. Hamaker

Netherlands Foundation for Research in Astronomy, Postbus 2, 7990 AA Dwingeloo, The Netherlands

Abstract. Paper II of this series studied the calibration process in mostly qualitative terms. In developing the underlying mathematics this paper completes that analysis and extends it in several directions.

It exploits the analogy between scalar and matrix algebras to reformulate the self-calibration method in terms of 2×2 Jones and coherency matrices. The basic condition that the solutions must satisfy in either case is developed and its consequences are investigated. The fourfold nature of the matrices and the non-commutativity of their multiplication are shown to lead to a number of new effects.

In the same way that scalar selfcal leaves the brightness scale undefined, matrix selfcal gives rise to a more complicated indeterminacy. The calibration is far from complete: *self-alignment* describes more properly what is actually achieved. The true brightness is misrepresented in the image obtained by an unknown brightness-scale factor (as in scalar selfcal) and an undefined *poldistortion* of the Stokes brightness. The latter is the product of a *polrotation* of the *polvector* (Q, U, V) and a *polconversion* between unpolarized and polarized brightness. The relation of these concepts to conventional "quasi-scalar" calibration methods is demonstrated.

Like scalar selfcal, matrix self-alignment is shown to suppress spatial scattering of brightness in the image, which is a condition for attaining high *dynamic range.* Poldistortion of the brightness is an in-place transformation, but must be controlled in order to obtain *polarimetric fidelity.*

The theory is applied to reinterpret the quasi-scalar methods of polarimetry including those of Paper II, and to prove two major new assertions: (a.) An instrument calibrated on an unpolarized calibrator measures the degree of polarization correctly regardless of poldistortion; (b.) Under the usual a priori assumptions, a *heterogeneous* instrument (i.e. one with unequal feeds) can be completely calibrated without requiring a phase-difference measurement.

Key words: instrumentation: interferometers — instrumentation: polarimeters — methods: analytical — methods: observational — techniques: interferometric — techniques: polarimetric

1. Introduction

The invention of self-calibration marks a watershed in the history of radio astronomy. It turned Very Long Baseline Interferometry (VLBI) from a primitive tool to probe source structure into a full-fledged imaging technique. By adopting the selfcal technique, imaging arrays such as the Westerbork Synthesis Radio Telescope (WSRT) and the Very Large Array (VLA) suddenly acquired dynamic ranges exceeding those of their original designs by several orders of magnitude. Such performance has been the standard for newer instruments ever since.

Notwithstanding this huge success, there is good reason for dissatisfaction. Indeed, the selfcal algorithm is a scalar one and therefore fundamentally incompatible with the vector nature of the electromagnetic radiation field. Observers have nonetheless found their way through the polarization landscape, following a narrow path marked by four guide-posts:

– Weakly polarized sources;
– The availability of calibrator sources that are unpolarized;
– Mechanically pointed antennas with custom-designed feeds providing for low instrumental polarization;
– The use of arrays that are *homogeneous*, i.e. have nominally identical feeds in all antennas.

Send offprint requests to: J.P. Hamaker,
e-mail: hamaker@nfra.nl

Under these conditions, a first-order linearized approximation to the interferometer equation converts it to a set of linear equations in which the scalar selfcal algorithm can be applied, as outlined in Sect. 7 below. I shall call this the *quasi-scalar* approach. As soon as any of the guideposts is taken away, the traveler is in trouble, — and this is actually happening:

– As observers push toward higher frequencies and resolutions, they begin to encounter more strongly polarized sources. . .
– . . . and unpolarized ones become rare;
– In ad-hoc VLBI arrays such as the European VLBI Network (EVN), funding limitations dictate the use of makeshift circular feeds of inadequate quality in many stations;
– For a new generation of radio telescopes, unorthodox designs are being considered in which cost is an overriding driver. Polarization purity as well as homogeneity will very probably have to be traded in for financial economy.

The time has come to leave the quasi-scalar trail and find a less restrictive way of navigating in polarization land. A suitable type of vehicle has been proposed in Paper I of this series (Hamaker et al. 1996). Its principle is to abandon the notion of scalar electromagnetic signals and visibilities in favour of signal 2-vectors and coherency 4-vectors, and to represent their transformations by multiplications with matrices. This results in a simple, modular and complete description of the system without the need for simplifying assumptions.

In matters of calibration, Paper I looks backward, showing how the traditional methods can be described and justified in terms of the matrix/vector formalism. Paper II (Sault et al. 1996) takes a first step forward by showing how our treatment can be used to view an interferometer array as a single imaging instrument, and discussing the fundamental limitations to which such an array and its calibration are necessarily subject. Here, I zoom in on that part of the calibration process that Paper II takes for granted: My purpose is to develop a comprehensive matrix-based theory of self-calibration and find out how similar and/or different matrix selfcal is from the scalar selfcal that we know.

For this particular purpose I find that a representation of coherency and brightness in the form of *matrices* is more convenient than the vector representation of the preceding papers. However, such matrices are difficult to visualise. A third equivalent representation, that of Stokes parameters, is much more enlightening and therefore widely used. Outside its physical context, the Stokes representation is valid for any 2×2 matrix; this leads to the mathematical concept of *quaternions*: "hypercomplex" numbers composed of a scalar and a three-vector. Quaternions have multiplication rules of their own which can be used in analysing matrix products in more detail.

This proves to be a powerful tool for studying the effects represented by the matrix equations.

In its essence, the quaternion concept entails little more than a simple extension of undergraduate-level linear algebra, but it is new to radio astronomy and may take some effort to digest. Therefore, although derivations and proofs are an essential part of this paper, I have relegated them to appendices. The main text concentrates on the results and their interpretation. For some of the mathematical effects and properties that the analysis uncovers, I have chosen to introduce suitable polarimetry-specific terminology.

The layout of the paper is briefly as follows:

Section 2 establishes the basic mathematical components: coherency, Jones and brightness matrices and the Stokes brightness vector/quaternion.

Sections 3 to 5 develop the matrix form of self-calibration by exploiting the close analogies between scalar and matrix algebras. It turns out that the matrix form provides a "calibration" that is seriously incomplete: *self-alignment* describes more accurately what the algorithm actually achieves. An arbitrary *poldistortion* is left undefined; it is an in-place transformation of the brightness, composed of a *polrotation* of the *polvector* (whose components are the Stokes Q, U and V brightnesses), and a *polconversion* between the polvector and total intensity I.

Section 6 considers the elimination of poldistortion through the use of unpolarized calibrators, supplemented with prior knowledge about the feed and/or additional observations. It confirms, reinterprets amd extends the results of Paper II.

Section 7 discusses quasi-scalar calibration methods in the perspective of the matrix approach. It is shown that most of the concepts revealed by the latter also appear in one form or another in the quasi-scalar context. Recent attempts at calibrating without recourse to an unpolarized calibrator are also discussed and evaluated.

A new option that the matrix formalism offers is the use of *heterogeneous* arrays, i.e. arrays combining antennas with non-identical feeds. It is explored in Sect. 8. In such arrays, feed errors and receiver phases are coupled in such a way that constraining the former in the usual way also fixes the latter; no additional phase measurement is needed to complete the calibration.

Section 9 discusses the general problem of calibrating an observation of a completely unknown source. In this case, one depends entirely on a priori knowledge of the instrument and/or ground-based measurements. My analysis is at present inconclusive and the problem needs further study.

Section 10 makes a comparison between quasi-scalar and matrix approaches, summarises the results of the latter and speculates on its practical application.

The Appendix provides a brief summary of the mathematical background as well as proofs of the assertions in

Table 1. Analogies between scalars and 2×2 matrices, their algebraic properties and their application in interferometry. Particulars are to be found in the sections listed

Scalar form	Matrix form	Section(s)
Arbitrary scalar a	Arbitrary 2×2 matrix $\boldsymbol{A}$	
Unity $= 1$	Identity 2×2 matrix $= \mathbf{I}$	
Phase factor $\exp i\alpha$	Unitary 2×2 matrix $\boldsymbol{X}$ Unimodular unitary 2×2 matrix $\boldsymbol{Y}$	Appendix B.1
Positive real number $|a|$	Positive hermitian 2×2 matrix $\boldsymbol{G}$ Unimodular pos.-herm. 2×2 matrix $\boldsymbol{H}$	Appendix B.4
Polar representation $a = a \exp i\alpha$	Polar representation $\boldsymbol{A} = a \exp i\alpha \, \boldsymbol{H}\boldsymbol{Y}$	5.1, Appendix B.6
Complex conjugation a^*	Hermitian transposition $\boldsymbol{A}^\dagger \equiv \boldsymbol{A}^{*\mathrm{T}}$	
$d(\psi) = |ae^{i\psi} - 1|^2$ minimal for $\psi = -\arg a$	Minimum-variance theorem	Appendix C.5
Multiplication $c = ab = ba$	Multiplication $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} \neq \boldsymbol{B}\boldsymbol{A}$	
Field or voltage transfer $e'_j = g_j e_j$ $\quad g_j =$ (complex) gain	Field or voltage vector transfer $\boldsymbol{e}'_j = \boldsymbol{J}_j \boldsymbol{e}_j$ $\quad \boldsymbol{J}_j =$ (complex) Jones matrix	2.3
Visibility $e_{jk} = \,< e_j e_k^* >$	Coherency $\boldsymbol{E}_{jk} = \,< \boldsymbol{e}_j \boldsymbol{e}_k^\dagger >$	2.2
Visibility transfer $e'_{jk} = g_j e_{jk} g_k^*$	Coherency transfer $\boldsymbol{E}'_{jk} = \boldsymbol{J}_j \boldsymbol{E}_{jk} \boldsymbol{J}_k^\dagger$	2.3

the body of the paper. It also contains a few small digressions related to polarimetry that would not fit elsewhere.

1.1. Terminology and notation

Since I will be discussing scalar selfcal and its full-polarization analogue side by side, it is necessary to put a precise terminology in place.

The analogue of the scalar visibility is the *coherency*. It consists of four components and can be represented in various coordinate systems in the form of either a vector or a matrix. Each of these forms contains four scalar elements that I shall occasionally refer to as *visibilities*.

I shall use "matrix" as an antonym of "scalar", — as in "matrix selfcal". This is not strictly correct, because alternative formulations of my methods are possible that rely on other representations, e.g. using vector or tensor forms. But within the context of this paper it is the most convenient word to describe the antithesis.

The device that converts the electromagnetic field vector into a pair of voltages is called a *feed*; it consists of two *receptors* that are usually (but not necessarily, cf. Appendix B.3) sensitive to nominally opposite polarizations. In a *homogeneous* array all feeds are nominally identical; in a *heterogeneous* array they differ.

The imaging process that I consider consists of the observations proper followed by a process of self-calibration. In the latter, *models* of the instrumental errors and the source brightness distribution are developed jointly in an iterative procedure. It is assumed to have converged when the models together correctly represent the observed coherencies within the noise. The *image* is the pictorial representation of the source model; the two words are almost synonymous. The model itself may take various forms; its essential property is that it can be used to "predict" model coherency values that can be compared to those actually observed in order to estimate instrumental errors.

It is important to correctly understand the word "intensity" used as an adjective, as in "intensity calibration". I use it to say that the calibrator source is characterised by its intensity alone, i.e. it is unpolarized. It does *not* imply the exclusive calibration of receiver parameters that one might associate with intensity, e.g. voltage gains.

Vectors are denoted by bold lowercase symbols; bold uppercase represents matrices. Constant scalars and vectors are shown in roman, variables in italic font. A unit vector in the direction $\boldsymbol{x}$ will be denoted by $\mathbf{1}_{\boldsymbol{x}}$. The "dagger" superscript $\dagger$ 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.

I follow the subscript notation of Paper II: quantities in the signal domain carry a single antenna subscript j

or k; those in the coherency domain get an interferometer subscript jk. An additional subscript t will be used to indicate successive integration intervals or "time slices". The array consists of N antennas and an observation comprises M integration intervals.

2. Coherency-matrix formulation of interferometry

2.1. The scalar/matrix analogy

The algebraic properties of scalars and matrices are very similar. Every elementary property of scalars has an immediate matrix counterpart, with one very important exception, viz. that matrix multiplication is non-commutative.

The analogy extends further. There is, for example, a matrix counterpart of the polar representation $a = |a| \exp i\alpha$ of complex numbers. An overview is given in Table 1.

2.2. The coherency matrix

As in the previous papers of this series, I represent the electric field in cartesian coordinates by a vector

$$\boldsymbol{e} = \begin{pmatrix} e_x \\ e_y \end{pmatrix}.$$

The equivalent of the scalar visibility is the *coherency tensor*. In mathematical terms, it is a complex-valued 2-dimensional tensor of rank 2 (Landau & Lifshitz 1995). In Paper I, we represented it in the form of the *coherency vector* and the *Stokes vector*. Here I shall use yet another representation, the *coherency matrix*. It is composed of the same four elements as the coherency vector, but arranged in the form of a 2×2 matrix

$$\boldsymbol{E}_{jk} = <\boldsymbol{e}_j \, \boldsymbol{e}_k^\dagger> = \begin{pmatrix} <e_{jx}e_{kx}^*> & <e_{jx}e_{ky}^*> \\ <e_{jy}e_{kx}^*> & <e_{jy}e_{ky}^*> \end{pmatrix}. \quad (1)$$

All of these forms are entirely equivalent representations of one and the same underlying tensor. Both the form (vector, matrix or other) and the coordinate system (e.g. cartesian or circular, cf. Paper I) of the representation are a matter of convenience. I will use geometric xy coordinates throughout.

2.3. The interferometer equation

I recall from Paper I that the elements in the signal path in one antenna transform the electric field or voltage vector:

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

where $\boldsymbol{J}_j$ is 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

$$\boldsymbol{W}_{jk} = \boldsymbol{J}_j \boldsymbol{E}_{jk} \boldsymbol{J}_k^\dagger. \quad (2)$$

Perhaps superfluously, I reiterate that this is no more than another representation of the basic underlying transformation of the coherency tensor by an interferometer. The advantage over the coherency-vector representation of Paper I is that both coherencies and antenna/receiver systems are now represented by 2×2 matrices and we need only one type of multiplication operator. This leads to a complete formal analogy between scalar and matrix self-cal, which will allow us to extrapolate our knowledge of the former in trying to understand the latter.

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.

I note in passing that in the particular case of a single dish, $j = k$ and Eq. (2) reduces to

$$\boldsymbol{W} = \boldsymbol{J}\boldsymbol{E}\boldsymbol{J}^\dagger \quad (3)$$

which is known as a *congruence transformation*. We will see in Sect. 4 that this same transformation describes a "self-aligned" synthesis array.

2.4. Matrix and Stokes brightnesses

In its original form, the van Cittert-Zernike theorem (Paper I Appendix C; Thompson et al. 1986; Perley et al. 1994; Born & Wolf 1964) establishes a spatial Fourier-transform relation between a scalar visibility function $e(\boldsymbol{r})$ of baseline $\boldsymbol{r}$ and a scalar brightness function $B(\boldsymbol{l})$ of sky position $\boldsymbol{l}$.

In an observation, we measure the visibility at discrete times t; our observables are the samples

$$e_{jkt} = e(\boldsymbol{r}_{jkt}).$$

In self-calibration theory, the sampling times are assumed to coincide for all interferometers. Also discretising the brightness, we approximate the Fourier integral by a sum

$$e(\boldsymbol{r}_{jkt}) = \sum w(\boldsymbol{r}_{jkt}, \boldsymbol{l}) B(\boldsymbol{l}). \quad (4)$$

The theorem can be readily generalised to show that each element of the coherency matrix is the Fourier transform of the corresponding element of a *brightness matrix*. In the same approximation

$$\boldsymbol{E}_{jkt} = \sum_{\boldsymbol{l}} w(\boldsymbol{l}, \boldsymbol{r}_{jkt}) \, \boldsymbol{B}(\boldsymbol{l})$$
$$j, k = 1, \ldots, N, \quad t = 1, \ldots, M. \quad (5)$$

The four elements of the brightness matrix correspond to those of the coherency matrix. A more enlightening representation is provided by the *Stokes brightness* (I, Q, U, V). It is another function of $\boldsymbol{l}$, defined by the transformation

$$\boldsymbol{B} = \begin{pmatrix} I+Q & U-iV \\ U+iV & I-Q \end{pmatrix} = I\,\mathbf{I} + Q\,\mathbf{Q} + U\,\mathbf{U} + V\,\mathbf{V} \quad (6)$$

where

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\mathbf{U} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{7}$$

The matrix constants $\mathbf{I}$, $\mathbf{Q}$, $\mathbf{U}$ and $\mathbf{V}$ are known in physics as the *Pauli (spin) matrices.*

The Stokes parameter I is the total brightness or *intensity.* For (Q, U, V) a proper name is the "polarized-brightness vector"; more conveniently, I shall call it the *polvector.* The dichotomy between I and the other Stokes parameters can be understood as a consequence of the Pauli matrix $\mathbf{I}$ being the identity matrix. The domain of the polvector is closely related to that of the Poincaré sphere (Born & Wolf 1964; Cornbleet 1976; Simmons & Guttman 1970). It is convenient to introduce a shorthand for Eq. (6):

$$\boldsymbol{B}(\boldsymbol{l}) = [\, I(\boldsymbol{l}) + \boldsymbol{p}(\boldsymbol{l}) \,] \tag{8}$$

where $\boldsymbol{p} \equiv (Q, U, V)$ is the polvector.

2.5. Quaternions

The transformation Eq. (6) or Eq. (8) does not depend on $\boldsymbol{B}$ being a brightness matrix. It can be applied to an arbitrary 2×2 matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = [\, a + \boldsymbol{a} \,]. \tag{9}$$

The entity in square brackets is known as a *quaternion.* Quaternions were invented and named by Hamilton in the middle of the nineteenth century in a mathematical quest for generalisations of the concept of complex numbers. Physicists of the time ignored them in favour of the vector algebra that was developed at the same time (Hestenes 1986). In the analysis to be presented here they prove to be extremely useful, because they can be added and multiplied in exactly the same way that matrices can: in mathematical terms the "quaternion group is isomorphous with the group of 2×2 matrices". (Korn & Korn 1961).

In the same way that the Stokes vector is preferable because of its physical content, the quaternion form of equations such as the interferometer equation Eq. (2) can be analysed in a more meaningful way than the corresponding matrix equations. The analysis is an essential part of this paper, but I have chosen to present it in an appendix. In the main text, I concentrate on the results and their physical interpretation.

The notation for quaternions is not standardised. The form Eq. (9) is an ad-hoc choice of my own.

3. Scalar self-calibration

The instability of our instruments limits the possibility of *external* calibration against cosmic or man-made standards. The high dynamic ranges that are now the norm depend entirely on *self*-calibration or *selfcal* (Thompson et al. 1986; Perley et al. 1994).

So far, selfcal has been known only in scalar form. Although it is almost universally applied, we have no complete and compelling theory to describe it. In practice it shows a strong propensity to converge to a unique solution, — provided it is given enough data. Yet we have no formal proof of this uniqueness.

My aim now is to study matrix selfcal by approaching it as a generalisation of the scalar variant. In doing so, the best one may expect is to reach an analogously incomplete understanding. As we shall see, this is enough for making interesting inferences.

I begin by reviewing scalar selfcal. I ignore the effect of noise, except to note that it introduces an element of uncertainty into the entire process that may in unfavourable cases subvert the apparent uniqueness of our solution. As a rule this does not appear to happen in practice.

3.1. Scalar self-calibration

Scalar selfcal works on the basis of two assumptions:

- All instrumental effects are *antenna-based*, i.e. the correlator is error-free. Thus our observed visibility is given by

  $$v(\boldsymbol{r}_{jkt}) = g_{jt}\, e(\boldsymbol{r}_{jkt})\, g_{kt}{}^*.$$

- The sky is "relatively empty": The source brightness is nonzero only in a minor fraction of the observed field, the source's *support.* In practice, it turns out that the support need not be known a priori, but can be found and successively refined by inspection of provisional "dirty" images.

Given a set of observations, selfcal seeks to find antenna gains g'_{jt} and visibilities $e'(\boldsymbol{r}_{jkt})$ that are consistent with them:

$$v(\boldsymbol{r}_{jkt}) = g'_{jt}\, e'(\boldsymbol{r}_{jkt})\, g'_{kt}{}^*. \tag{10}$$

Obviously, one solution consists in the *true* gains and visibilities. In addition, Eq. (10) is satisfied by the combination

$$g'_{jt} = g_{jt}\, x_{jt}^{-1}, \qquad e'(\boldsymbol{r}_{jkt}) = x_{jt}\, e(\boldsymbol{r}_{jkt})\, x_{kt}{}^*, \\ j, k = 1, \ldots, N \tag{11}$$

for any conceivable set of multipliers x_{jt}. For each of these, the visibilities $e'(\boldsymbol{r}_{jkt})$ in turn correspond to a source model B' according to Eq. (4):

$$x_{jt}\, e(\boldsymbol{r}_{jkt})\, x_{kt}^* = \sum_{\boldsymbol{l}} w(\boldsymbol{l}, \boldsymbol{r}_{jkt})\, B'(\boldsymbol{l}). \tag{12}$$

If the source support is limited as assumed, the sum contains only a limited number L of terms for which $B'(\boldsymbol{l})$ can differ from zero. For a properly conditioned observation, the number of visibility samples (of order $MN(N-1)/2$) is much greater than that of unknowns: MN values of the x_{jt} plus L values of $B'(\boldsymbol{l})$.

The system is now overdetermined, but we have already seen that it admits at least one solution. It is not unique, however. Indeed, if all the x'_{jt} equal one value x, Eq. (12) can be rewritten as

$$x \, e(\boldsymbol{r}_{jkt}) \, x^* = \sum_{\boldsymbol{l}} w(\boldsymbol{l}, \boldsymbol{r}_{jkt}) \, x \, B(\boldsymbol{l}) \, x^*$$

which defines a brightness solution

$$B'(\boldsymbol{l}) = x \, B(\boldsymbol{l}) \, x^*. \tag{13}$$

Obviously, B' is confined to the support of B. Actually it is an exact but scaled replica.

Other solutions are unlikely to exist. If we should allow the x_{jt} to take independent values, this results in scattering of brightness away from the source to other parts of the image. It is reasonable to conjecture that it is impossible for any "wild" combination of x_{jt} values to produce a false brightness image that nonetheless vanishes everywhere outside the source support. Practical experience of two decades supports this conjecture, — but I repeat that a formal proof is lacking and the solution may not always be robust against the effect of noise.

The above argument pinpoints the support limitation as the agent that makes selfcal work. This idea does not seem to have been systematically exploited before, but Leppänen et al. (1995) advance it in discussing the construction of the polarized part for a source model whose total intensity is already available (cf. Sect. 7.2).

3.2. Calibration versus alignment

Equation (14) does not represent a complete calibration: out of the infinite number of solutions that mutually differ by their positive scale factors xx^*, the selfcal procedure arbitrarily selects one. This non-uniqueness is fundamental. On the basis of selfcal alone, we have no way of knowing what value x has. We must fix the brightness scale afterwards by other means.

What selfcal does achieve is to reduce all the errors x_{jt} in the individual visibility measurements to a single value x: it *lines up* the measurements, forcing them all to conform to one common scale factor. As a result, extremely high dynamic ranges can be attained even though the absolute brightness scale is unknown.

Strictly, the calibration is incomplete and we ought to replace "self-calibration" by the more precise term "self-*alignment*". The distinction is a bit academic here, but will become crucially important when we explore matrix selfcal.

It is not immediately clear from the present discussion that the absolute sky position is also lost in self-calibration. To establish this, one must consider the properties of the Fourier-transform relation Eq. (12). The effect is not directly relevant here, but it should not be forgotten.

4. Matrix self-alignment

4.1. Self-alignment

The argument of the preceding section carries over in its entirety to the matrix domain. We may follow through the same steps, simply replacing all scalar gains by Jones matrices and visibilities by coherency matrices. We must now solve the matrix equivalent of Eq. (10):

$$\boldsymbol{V}(\boldsymbol{r}_{jkt}) = \boldsymbol{J}'_{jt} \, \boldsymbol{E}'(\boldsymbol{r}_{jkt}) \, \boldsymbol{J}'_{kt}{}^{\dagger}.$$

The solution is the analogue of Eq. (13)

$$\boldsymbol{B}'(\boldsymbol{l}) = \boldsymbol{X} \, \boldsymbol{B}(\boldsymbol{l}) \, \boldsymbol{X}^{\dagger}. \tag{14}$$

This equation is known as a *congruence transformation*. I shall follow mathematicians in referring to Eq. (14) as "the congruence transformation $\boldsymbol{X}$", i.e. using the name of $\boldsymbol{X}$ as a synonym for the transformation that it effects: The source model $\boldsymbol{B}'$ is related to the true source $\boldsymbol{B}$ by an unknown congruence transformation $\boldsymbol{X}$.

Like Eq. (13) for scalar selfcal and with the same proviso, this is a basic relation that any matrix self-alignment solution must satisfy. And as for the scalar case, the appearance of this indeterminacy is fundamental and unavoidable. I call it the *poldistortion*.

Although this result is formally the same as for scalar selfcal, it is worth some extra thought. We see that all the, probably time-varying, errors in the observation have given way to a *single* poldistortion representing a set of unknown errors that is *constant* over the observation. A similar effect occurs, e.g., in quasi-scalar interferometry, where scalar selfcal on the "parallel" channels takes away the temporal gain/phase variations and leaves a single, constant XY or LR phase difference in their place. In matrix self-alignment, the time-varying errors to be converted into unknown constants include not only phases and gains, but also any other variations, e.g. in feed parameters (mainly the "leakage" or "D" terms in the quasi-scalar jargon), in parallactic angle (if one were not to correct for it beforehand) and in ionospheric Faraday rotation. Moreover, all this is true not only for a single observation contiguous in time, but also for a set of observations spaced over a time interval in which the source does not change.

A schematic of the combined self-alignment and poldistortion elimination procedure is shown in Fig. 1. As in scalar selfcal, we may first correct for the errors that we know of. Note that our argument does not require this; however, the actual selfcal *algorithm* (Thompson et al. 1986; Perley et al. 1994) starts with an image, to be made from the raw observations, that must be good enough to extract a reasonable initial source model from it. If such a model can be obtained otherwise, e.g. from prior knowledge about the source, the initial corrections may just as well be omitted: they will automatically be subsumed in the corrections to be derived in self-alignment.

Whether or not we apply the prior corrections is likely to affect the poldistortion in the final solution, *but not our ignorance about its value.* No matter how we arrive at a self-aligned image, we *must* assume that it contains an unknown poldistortion and undertake to eliminate it. This problem will be discussed below.

As an aside, I observe that Fourier transforming Eq. (3) for a single dish, one obtains an equation of the same form as Eq. (14): the self-aligned array is equivalent to a single dish with unkown Jones matrix. This result was derived in Paper II by a more qualitative argument.

4.2. Scattering dynamic range and polarimetric fidelity

The solutions Eqs. (13) and (14) of the selfcal/self-alignment problem represent *in-place* transformations of the brightness: a simple uniform scaling in the scalar case, a uniform poldistortion in the matrix case. Their common characteristic is that they do not *scatter* radiation out of any point of the source into any other position in the image. In the scalar case, the scaling being the same for the entire image means that our image is a faithful (although scaled) replica of the source: *dynamic range* is conserved, - and it is this feature that makes selfcal such an important and valuable tool.

In the matrix case, image fidelity and dynamic range are no longer synonymous. Self-alignment suppresses spatial scattering in the image; the residual scattering defines to what extent weak structures remain recognisable in the presence of strong features elsewhere in the source. The concept of dynamic range is appropriate to describe this effect, but its proper quantitative definition is not as obvious as in the scalar case.

On top of any residual effect of scattering comes the poldistortion that is *independent* of it. Even if we were to produce a truly scatter-free image it would still misrepresent the source in an unknown way: as a complement to dynamic range we must consider the question of the *polarimetric fidelity* of the image.

We have no need for a formal definition of either dynamic range or polarimetric fidelity. What matters is that we are dealing with two quite different and mutually independent types of error in a self-aligned image.

4.3. Example: Faraday rotation

The concepts developed above and the benefits of the matrix approach can be neatly illustrated on the example of ionospheric Faraday rotation. This rotation changes the observed position angle of linear polarization. Over a synthesis observation, it varies and this results in scattering of linearly polarized brightness in the final image.

Corrections for the effect are based on external data: Ionosphere models and ground- and satellite-based measurements (Thompson et al. 1986). Often the
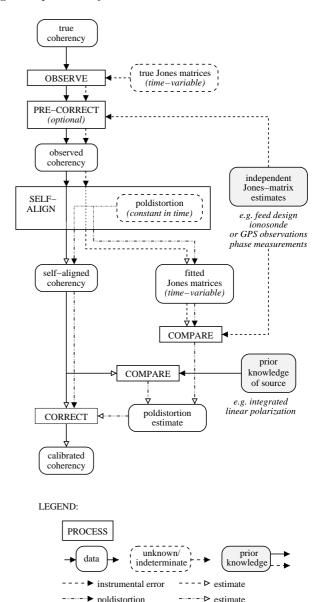

Fig. 1. Flow diagram of the self-alignment and poldistortion calibration procedure. The crux of the matter resides in the self-alignment: it effects a split between the true coherencies and the time-variable part of the observing errors. The price to be paid is the insertion of an unknown constant error artefact, the poldistortion, in the estimated coherencies and its inverse in the estimated Jones matrices. To eliminate it, we must compare these estimates with a priori instrumental and astronomical information

results leave much to be desired. In some cases, the external correction can be improved upon by noting the apparent rotation of linear polarization during the observation (A.G. de Bruyn, private communication). One might call this *Faraday self-alignment.* In a similar vein, Sakurai & Spangler (1994) used a linearly polarized source to monitor temporal fine structure in the Faraday rotation. Obviously, one can only measure and eliminate *variations* in the rotation; to find the true

position angle of linear polarization one must determine its zero point by other means.

In the matrix approach, this adjustment of Faraday rotation is no longer a distinct operation: it is subsumed in the overall self-alignment process. A scatter-free image results directly. The absolute rotation and position angle of linear polarization remain undetermined: this is now recognised as part of the poldistortion that is the unavoidable by-product of self-alignment. Obtaining a good dynamic range and correct rendition of the polarization appear as two distinct problems that must be independently addressed.

5. The poldistortion

5.1. Polrotation and polconversion

An arbitrary square matrix $\boldsymbol{X}$ can be subjected to a *polar decomposition* (Appendix B.6)

$$\boldsymbol{X} = x\boldsymbol{H}\boldsymbol{Y} = x\boldsymbol{Y}\boldsymbol{H}'$$

where

- x is a complex constant;
- $\boldsymbol{Y}$ is a *unimodular unitary* matrix (i.e. $\boldsymbol{Y}\boldsymbol{Y}^\dagger = \mathbf{I}$ and $\det \boldsymbol{Y} = 1$);
- $\boldsymbol{H}$ is a *unimodular positive hermitian* matrix (i.e. $\boldsymbol{H} = \boldsymbol{H}^\dagger$, $\det \boldsymbol{H} = 1$ and $\mathrm{Tr}\ \boldsymbol{H} > 0$); so is $\boldsymbol{H}'$.

Applying the polar decomposition to Eq. (14) we get

$$\boldsymbol{B}' = xx^* \boldsymbol{H}\ (\boldsymbol{Y}\boldsymbol{B}\boldsymbol{Y}^\dagger)\ \boldsymbol{H}^\dagger. \tag{15}$$

The positive scaling factor is the same as in scalar selfcal and I shall further ignore it. Apart from it, $\boldsymbol{B}'$ is derived from $\boldsymbol{B}$ through a succession of two specific transformations.

The first is the unitary transformation $\boldsymbol{Y}$. Its effect is to leave the intensity unchanged (naturally, since $\boldsymbol{Y}\mathbf{I}\boldsymbol{Y}^\dagger = \mathbf{I}$) and to *rotate the polvector* in its three-dimensional space (Appendix C.1). In quaternion notation

$$[\,I' + \boldsymbol{p}'\,] = [\,I + \boldsymbol{R}\boldsymbol{p}\,]$$

where $\boldsymbol{R}$ is a rotation operator. I call $\boldsymbol{Y}$ the *polrotation* (transformation). The rotation and $\boldsymbol{Y}$ can be characterised by a vector in polvector space, the *Gibbs vector* (Appendix B.1)

$$\mathbf{1}_{\boldsymbol{y}} \sin \eta,$$

where $\boldsymbol{y}$ is the direction of the rotation axis and 2η the rotation angle. In Appendix C.1 I derive the relation between the Gibbs vector and $\boldsymbol{Y}$. Obviously, any multiple of the Gibbs vector is an *eigenvector* of the transformation: $\boldsymbol{R}\mathbf{1}_{\boldsymbol{y}} = \mathbf{1}_{\boldsymbol{y}}$.

Like the polrotation, the positive hermitian transformation $\boldsymbol{H}$ is characterised by a vector

$$\mathbf{1}_{\boldsymbol{h}} \sinh \gamma$$

that may also be called a Gibbs vector (Appendix C.4). The transformation exchanges brightness between the intensity and that component of the polvector that is parallel to $\boldsymbol{h}$; any perpendicular component is an eigenvector. I call $\boldsymbol{H}$ the *polconversion*. There is no mutual conversion between polvector components: The transformation is *rotation-free*.

We may summarise the above by stating that the self-aligned source model $\boldsymbol{B}'$ is related to the true brighness $\boldsymbol{B}$ by a transformation that is the product of a polrotation, a polconversion and a positive scale factor.

The poldistortion $\boldsymbol{X}$ is far more complicated than a simple scale error. Polrotation and polconversion each contain three unknown parameters (the cartesian components of their Gibbs vectors) which, together with the scale factor, make a total of seven. This is the same number that Paper II arrived at through a more elementary analysis.

$\boldsymbol{H}$ and $\boldsymbol{Y}$ being unimodular

$$\det \boldsymbol{B} \equiv \mathbf{I}^2 - \boldsymbol{p}^2 \tag{16}$$

is, apart from the scaling of the entire image, an invariant of the poldistortion transformation.

5.2. Controlling the poldistortion

With self-alignment only the first half of the calibration job is done. To eliminate the poldistortion, we must bring other information to bear on our problem. It may take the form of either prior knowledge or additional measurements; both may be used either to constrain the self-alignment algorithm so as to (partly) suppress the poldistortion, or to remove the poldistortion $\boldsymbol{X}$ afterwards.

For example, if we have reason to believe that our source field is unpolarized, we can impose this condition on our image and source model, as in Sect. 6.1 below. Alternatively, we may allow self-alignment to produce a polarized image and determine and remove the polconversion afterwards, on the basis that in the proper image certain sources must be unpolarized.

Apart from the image $\boldsymbol{B}'$, self-alignment also yields estimates $\boldsymbol{J}'_j = \boldsymbol{J}_j\boldsymbol{X}^{-1}$ of the antenna Jones matrices. Comparing these with what we know about the true values gives us another handle on $\boldsymbol{X}$.

6. Polrotation and intensity self-alignment

In this section I briefly review Paper II from our newly acquired viewpoint, providing proofs for its assertions and extending them. A further extension, to the case of heterogenous arrays, will follow later (Sect. 8). Here I will show how the use of an intensity calibrator suppresses polconversion and how the remaining polrotation can then be

determined. The more difficult problem of eliminating pol-distortion without recourse to an intensity calibrator will be taken up in Sects. 7.2 and 9.

6.1. Intensity self-alignment

The usual first step is to observe an unpolarized source and self-align it. Because the source has intensity only, I call this *intensity self-alignment*.

The brightness matrix $\boldsymbol{B}$ reduces to a positive multiple of the identity matrix $\mathbf{I}$, the scale factor being the *intensity*. Imposing the same form on our solution, we find that

$$B'(\boldsymbol{l})\mathbf{I} = \boldsymbol{X} \; B(\boldsymbol{l})\mathbf{I} \; \boldsymbol{X}^{\dagger}. \tag{17}$$

A solution $\boldsymbol{B}'$ can only exist if $\boldsymbol{X}$ is a multiple $c\boldsymbol{Y}$ of a matrix $\boldsymbol{Y}$ for which

$$\boldsymbol{Y}\boldsymbol{Y}^{\dagger} = \mathbf{I},$$

i.e. $\boldsymbol{Y}$ is unitary. c is the scale factor discussed before and we may ignore it.

What we see, then, is that the poldistortion is reduced to a polrotation, $\boldsymbol{Y}$. Our knowledge that the source is unpolarized has completely suppressed the polconversion factor $\boldsymbol{H}$ of Eq. (15). One of the central results of Paper II is that the remaining error (which we have now recognised as a polrotation) is characterised by three unknown real parameters. I have already confirmed this assertion in Sect. 5.1.

The arbitrary rotation that $\boldsymbol{Y}$ represents can be factored into the product of three mutually orthogonal rotations (Korn & Korn 1961). Choosing for these the base rotations of Appendix C.2 we have:

$$\boldsymbol{Y} = \boldsymbol{Y}_{\mathbf{q}}(\phi)\,\boldsymbol{Y}_{\mathbf{u}}(\epsilon)\,\boldsymbol{Y}_{\mathbf{v}}(\theta) \tag{18}$$

$$= \begin{pmatrix} \exp i\phi & 0 \\ 0 & \exp -i\phi \end{pmatrix} \begin{pmatrix} \cos\epsilon & i\sin\epsilon \\ i\sin\epsilon & \cos\epsilon \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

This is Eq. (27) of Paper II, revised to represent our new insight into what it actually means: The order of the factors reflects the physics of the instrument as discussed in Paper I and the variables have been renamed accordingly. The last term, which represents the first element in the signal path, corresponds to an unknown geometric rotation θ of the zero point from which the orientation of the feeds is measured; likewise, the middle term represents an unknown zero-point shift ϵ of the ellipticity scale. The leading term is the unknown phase difference 2ϕ between the two subsets (X and Y or L and R) of receiver channels. (For circular feeds, the equation takes a different form, see Appendix C.3).

I add a new proposition. Rotating the polvector does not change its length: this means that *an intensity-aligned instrument correctly measures the degree of polarization*, even though the orientation of the polvector cannot be determined. This provides a simple and powerful way to detect strongly polarized sources in generally weakly polarized fields, e.g. in surveys.

6.2. Constraining the polrotation

To eliminate the polrotation $\boldsymbol{Y}$, Paper II proposes to observe calibrators whose polarization is known. It states that *two* such observations are necessary. We can now supply the proof that was missing.

Let the Stokes brightness of the first source be

$$\boldsymbol{B} = [\, I + \boldsymbol{p}\,].$$

We require the observed brightness $\boldsymbol{B}'$ to be the same. This limits the possible range of polrotations $\boldsymbol{Y}$ to those that conserve $\boldsymbol{p}$, i.e. those unitary transformations that have $\boldsymbol{p}$ as their rotation axis; but it leaves the rotation angle free. To constrain it as well, we do indeed need a second polarized source whose polvector is not collinear with $\boldsymbol{p}$.

6.3. Imposing the nominal feed characteristics

For well designed and constructed feeds, the orientations and ellipticities are quite accurately known. We may compare this prior knowledge with the polrotated Jones matrices $\boldsymbol{J}_j\boldsymbol{X}^{-1}$ to estimate $\boldsymbol{X}$. It is reasonable to propose that the feed errors are randomly distributed with a zero average. In this way the zero-point offsets, $\boldsymbol{Y}_{\mathbf{u}}(\epsilon)$ and $\boldsymbol{Y}_{\mathbf{v}}(\theta)$ in Eq. (18), can be eliminated. This method is discussed in Paper II and I shall return to it and to the elimination of the $\boldsymbol{Y}_{\mathbf{q}}(\phi)$ term in Sect. 8.2.

6.4. Calibrating polarized source fields

Elimination of poldistortion in an unpolarized field is in itself not particularly useful but, provided the instrument is stable enough, we may transfer the result to an observation of an unknown field. In this case, instrumental stability limits the accuracy of the final calibration, no matter what the theoretical potential is of the calibration methods employed.

A better option is to sidestep the problem of drifts by using supposedly unpolarized reference sources in the observed field itself. The reference can be a strong foreground source (component). At the opposite end, most fields contain many cosmic background sources whose polarization, if present at all, must be zero on the average. By analysing the source model in either case, one should be able to estimate the polconversion and correct for it.

7. Quasi-scalar methods

It is interesting to see how the concepts revealed by the above analysis appear in the usual quasi-scalar methods of polarization calibration. Unlike in the rest of this paper, *Stokes parameters in this section represent visibilities.*

It is assumed that the feed errors and degree of polarization are both small; authors typically state a few

percent as the upper limit. It is further required that all antennas have the same type of feed. In the interferometer equation Eq. (2), second-order products of order 10^{-3} are dropped and errors at a few times this level are accepted.

The linearised equations can be found in Paper II (and in almost every paper on polarimetric calibration, e.g. Thompson et al. (1986), Perley et al. (1994) and the papers quoted in Sect. 7.2). I show them here in an elementary form for feeds with left (L) and right (R) circularly-polarized receptors:

$$
\begin{aligned}
I'_{jkt} \;+\; V'_{jkt} &= g^{\mathrm{R}}_{jt} g^{\mathrm{R}*}_{kt}(I_{jkt} + V_{jkt}) \\
I'_{jkt} \;-\; V'_{jkt} &= g^{\mathrm{L}}_{jt} g^{\mathrm{L}*}_{kt}(I_{jkt} - V_{jkt}) \\
Q'_{jkt} \;+\; iU'_{jkt} &= g^{\mathrm{R}}_{jt} g^{\mathrm{L}*}_{kt}(Q_{jkt} + iU_{jkt} + D_{jk}I) \\
Q'_{jkt} \;-\; iU'_{jkt} &= g^{\mathrm{L}}_{jt} g^{\mathrm{R}*}_{kt}(Q_{jkt} - iU_{jkt} + D_{kj}I).
\end{aligned}
\tag{19}
$$

(For linearly polarized feeds, Q, U and V must be cyclically interchanged in these equations and in the following discussion, cf. Papers I and II.)

7.1. Intensity calibration

The starting point is again an intensity calibration. The procedure is discussed at length in Paper II. The case where such calibrators are not available has been taken up only recently (cf. Sect. 7.2). In Paper II a point source is assumed, but this is not necessary.

Usually V' is assumed to be zero. Scalar selfcal is applied to the first two lines of Eq. (19) to obtain a good I' image and complex R and L antenna-channel gains, each with an unknown phase.

In the second pair of equations, the difference between these phases enters through the g and g^* factors; its effect is a polrotation in the Q, U plane of polvector space. The trailing terms in the equations describe the *leakage* from I to Q and U: a polconversion. As a consequence of the linearisation, polrotation and polconversion appear in a sum rather than a product.

For an unpolarized source, the leakage terms can be measured per interferometer because Q and U are zero. The Q, U-plane polrotation can only be eliminated by a measurement of some sort. Several approaches, each with their own uncertainties, are discussed in Paper II.

As with the matrix technique, unpolarized foreground or background sources can be used as in-field calibrators. An example is the observation by Wardle et al. (1998) of weak circular polarization in a quasar whose strong core is assumed to be unpolarized.

7.2. Calibration without an intensity calibrator

A quite serious problem arises if there are no unpolarized sources that can serve as a reference. This situation is actually occuring in VLBI: observed fields are so small

that they donot contain any background sources to speak of, and sources strong enough to serve as calibrators (either in the target field or elsewhere) tend to be relatively strongly polarized at VLBI resolutions, particularly at the higher frequencies. For these reasons, VLBI observers (Cotton 1993; Roberts et al. 1994; Leppänen et al. 1995) have pioneered quasi-scalar polarization calibration without the use of an intensity calibrator.

The basic idea in all these papers is to exploit the effect of parallactic-angle variations in alt-az antennas to distinguish the leakage terms in Eq. (19) from the true source visibilities. In discussing this method, I assume that the parallactic-angle rotation is corrected for beforehand: thus Eq. (19) refers to the visibilities in sky coordinates and the leakage terms include the inverse of the rotation.

Cotton (1983) mentions the method without giving details.

Roberts et al. (1994) consider two types of calibrator: either an unpolarized one that may be resolved (i.e. my case of an intensity calibrator) or an unresolved one that may be polarized. In the latter case, the true Q and U visibilities in Eq. (19) are constants. The paper illustrates the difference between them and the rotating leakage terms graphically for a couple of interferometers. The leakage terms and the visibilities are well separable by a model fit, provided the parallactic angles cover a sufficiently broad range.

Leppänen et al. (1995) use the empty-sky principle of Sects. 3.1 and 4 to separate true linear source polarization from leakage effects: the true polarized source brightness is necessarily confined[1] to the support of the total intensity, and this provides a strong constraint that the rotated leakage terms have difficulty satisfying. A polarized-source model obeying the constraint is used to estimate the leakage terms and the procedure can be iterated.

The authors introduce the concept of a "leakage beam" to characterise the transfer of brightness from total to polarized brightness. Its value at the origin represents the polconversion, its sidelobes a polconverted spatial scattering: due to the fragmented character of the quasi-scalar method, polconversion and scattering are not neatly separated and the latter is not completely suppressed. In a simulation Leppänen et al. (1995) find the leakage beam to peak at a few .1% close to the origin, in accord with the accuracy expected under quasi-scalar assumptions.

Unlike Roberts et al. (1994), these authors make no assumptions on the source. In the more comprehensive perspective of matrix theory I doubt that the information they do use is sufficient to suppress polconversion and obtain a unique solution, — although it may be that the quasi-scalar assumptions constrain it well enough for astrophysically acceptable image fidelity.

---

[1] In Paper I we discussed situations in interferometry where this is not the case. These are unlikely to occur in VLBI.

8. Heterogeneous arrays

Homogenoeus arrays, i.e. having all identical feeds, have been a natural first choice for obvious engineering reasons and are also *required* by the quasi-scalar approach. Having removed the latter restriction, we can now take a fresh look.

8.1. Coupling of receiver phases and feed errors

Consider the case of an intensity calibrator and perfect feeds represented by known unitary Jones matrices $\boldsymbol{F}_j$. (This assumption is justified in Appendix B.3. The argument becomes slightly more complicated for imperfect feeds, but leads to the same conclusion.) After intensity self-alignment and imposition of our prior knowledge about the feeds, the only error remaining is the receiver phase-difference term of Eq. (18), so the interferometer equation becomes

$$\boldsymbol{Y}_\mathbf{q}(\phi_j)\boldsymbol{F}_j\,\boldsymbol{F}_k^\dagger\,\boldsymbol{Y}_\mathbf{q}(-\phi_k) = \boldsymbol{F}_j\,\boldsymbol{F}_k^\dagger$$

in which both $\boldsymbol{Y}_\mathbf{q}$ terms are diagonal matrices. This equation must hold for all interferometers, which is possible in only two ways[2]:

- For a homogeneous array, $\boldsymbol{F}_j = \boldsymbol{F}_k = \boldsymbol{F}$ and $\boldsymbol{F}\boldsymbol{F}^\dagger = \mathbf{I}$. This is the conventional case discussed in Paper II and Sect. 6.3. The one thing that our derivation adds is that it pertains generally to homogeneous arrays with any feed type;
- For a heterogeneous array with $\phi_j = \phi_k = 0$.

The latter solution is new: in a heterogeneous array with perfectly known feeds, this knowledge alone suffices to calibrate all receiver phases *without an additional measurement*. In reality we may at best assume our feed descriptions to be correct on the average, just as for the homogeneous array. The solutions that we find for the individual feed characteristics as well as for the receiver phases will then be imperfect, and perhaps more sensitive to errors in our a priori assumptions and to system noise than in the homogeneous case; this is a matter that needs further study.

The possibility of a priori phase alignment is surprising at first sight. Yet there are several good intuitive reasons to accept it:

Most fundamentally, the nominal receptor characteristics in the heterogeneous case are arbitrarily distributed. There is no natural way to split them into two disjoint groups between which an asymmetry such as the phase difference might arise. The only way out is for the difference not to exist.

---

[2] There are combinations $\boldsymbol{F}_j$ and $\boldsymbol{F}_k^\dagger$ whose product equals $\boldsymbol{Y}_\mathbf{q}(\xi)$, producing another type of solution for a single interferometer. Such solutions are not possible, however, for a triangle of interferometers as required for self-alignment.

Another viewpoint is that the homogeneous case is, in physical terms, a *degenerate* one. The degeneracy results in a decoupling of the feed and phase characteristics which are normally coupled. Mathematically, this decoupling is represented by the product $\boldsymbol{F}_j\boldsymbol{F}_k^\dagger$ reducing to the value $\mathbf{I}$.

A third viewpoint is that a set of identical feeds defines a preferred direction in polvector space, viz. that of the polvectors to which the two receptors are matched (e.g. the V axis for circular feeds). This results in an asymmetry in the characteristics of the instrument that is directly related to the phase-difference problem. "Randomised" feed characteristics destroy this preferential direction: the phase difference evaporates and all polarizations can be measured equally well, — or equally poorly.

An array having altaz mounts is heterogeneous in a sense, since in the course of an observation its feed orientations relative to the source assume a range of values. Yet at any instant it is homogeneous and consequently its long-term heterogeneity does not help in removing the phase error and associated polrotation. As we have seen, it may help in controlling the polconversion (Sect. 7.2) when we have no intensity calibrators.

The only historical example of a truly heterogeneous array is that of the Westerbork Synthesis Radio Telescope (WSRT) in its "crossed-dipole" configuration. Weiler (1973) analysed this system to show that, in the quasi-scalar approximation, it can be fully calibrated through observations of only one intensity calibrator. His system is too different from modern ones to relate his analysis directly to ours. Yet his work pointed to the potential of heterogeneous arrays as long as a quarter century ago and provided a major motivation for carrying the present investigation to completion.

8.2. Minimising the feed errors

The method of Sect. 6.3 for dealing with imperfect feeds is to assume that they are correct *on the average*. For a heterogeneous array, we may use the equivalent requirement that the sum of the receptor errors squared be minimised.

Following Paper I, we model the antenna Jones matrix in terms of the nominal feed $\boldsymbol{F}_j$ ($\boldsymbol{C}$ in Paper I), a feed error $\boldsymbol{D}_j$ and the receiver-gain $\boldsymbol{G}_j$ (a diagonal matrix). After intensity self-alignment we then have

$$\boldsymbol{J}_j' = \boldsymbol{G}_j\,\boldsymbol{D}_j\,\boldsymbol{F}_j\,\boldsymbol{Y} \tag{20}$$

and hence we consider values $\boldsymbol{D}_j'$, $\boldsymbol{G}_j'$ and $\boldsymbol{Y}'$ that satisfy this equation, or

$$\boldsymbol{D}_j' = \boldsymbol{G}_j'^{-1}\,\boldsymbol{J}_j'\,\boldsymbol{Y}'^{-1}\boldsymbol{F}_j^{-1}. \tag{21}$$

For an ideal system, the $\boldsymbol{D}_j$ equal $\mathbf{I}$, so we define our best guess at the polrotation $\boldsymbol{Y}'$ as the one that minimises the sum of variances (Appendix A.5)

$$\sum_j \mathrm{Var}\,(\boldsymbol{D}_j' - \mathbf{I}) = \sum_j \mathrm{Var}\,(\boldsymbol{G}_j'^{-1}\,\boldsymbol{J}_j'\,\boldsymbol{Y}'^{-1}\boldsymbol{F}_j^{-1} - \mathbf{I}),$$

given the self-aligned Jones matrices $\boldsymbol{J}'_j$ and the nominal feed matrices $\boldsymbol{F}_j$, and under the condition that the gains $\boldsymbol{G}'_j$ are diagonal and $\boldsymbol{Y}'$ is unitary.

It may not be obvious that this condition leads to a unique solution. Simulations (Appendix D) using the MATLAB (1997) programming environment show that this is indeed the case. Moreover, the system of equations becomes singular for a homogeneous array; something of this sort was to be expected because of the degeneracy discussed in Sect. 8.1.

9. The general polconversion problem

As long as we can rely on an intensity calibration of some sort to suppress polconversion, both the quasi-scalar and matrix methods provide for elimination of the polrotation. The most difficult problem arises when no intensity calibrators are available. The preceding sections provide several leads on how this problem might be approached. I now consider it further.

9.1. Use of a priori receiver characteristics

In the absence of poldistortion, prior knowledge about the instrument can be used to optimise the polrotation. An obvious question now is whether that method can take care of polconversion as well. I have not succeeded in making a proper analysis. It is informative, however, to look again at the heterogeneous array, in which we have seen that prior knowledge about the feeds alone suffices to eliminate all polrotation, after intensity self-alignment had suppressed polconversion.

In an attempt to eliminate the entire poldistortion without recourse to an intensity calibrator, I modified the least-squares minimisation method of Sect. 8.2 by removing the restriction that $\boldsymbol{X}$ be unitary. It appears that there is still a unique solution, but it is not the correct one: The algorithm transforms part of the polconversion factor $\boldsymbol{H}$ in the poldistortion $\boldsymbol{X}$ through the feeds into spurious amplitude gains in the receiver. It has the freedom to do so because the amplitude-gain matrices are also positive hermitian. To suppress this effect, we need additional prior knowledge, e.g. about those gains.

This example suffices to demonstrate that, if we want to control polconversion by applying prior instrumental knowledge, we need more of such knowledge than to control polrotation. Undoubtedly, observers will find ways to reproduce with full-fledged matrix self-alignment the polarimetric fidelity now obtainable with the linearised approximation. Whether matrix theory will alow us to progress any further remains to be seen.

9.2. Maximum entropy

A different approach might be to apply some statistical optimisation criterion representing prior assumptions on the source. It has been suggested that the maximum-entropy (ME) method can also be applied in polarimetric imaging (Ponsonby 1973; Narayan & Nityananda 1986; Sault et al. 1999). The quantity proposed for maximisation is the integral over the image of the product of the eigenvalues of the brightness matrix.

This product equals the value of the determinant (Eq. (16)), $\det \boldsymbol{B} = I^2 - \boldsymbol{p}^2$. Physically this makes some sense: indeed, maximising it is similar to maximising the unpolarized (and therefore most "disorderly") brightness $I - |\boldsymbol{p}|$.

However, we have seen that $\det \boldsymbol{B}$ is invariant under poldistortion. So the ME algorithm must be indifferent to it: out of many possible solutions it will arbitrarily select one, with an unknown poldistortion, — just as self-alignment does. For the polrotation this is obvious, because the ME criterion provides no clue as to the orientation of the polvector. What my analysis shows is that it does not provide a handle on the polconversion either.

10. Conclusion

10.1. Comparison of the quasi-scalar and matrix approaches

Sooner or later, the matrix formulation is bound to supersede the scalar one as the basis for radio interferometry and aperture synthesis: As I have argued in the Introduction, quasi-scalar theory and its first-order accounting for polarization effects are approaching the limits of their applicability. To proceed beyond these limits, one must embrace the matrix paradigm.

In Table 2 I compare the quasi-scalar and matrix methods as they stand. Due to the difference between the two approaches, the comparison is not always straightforward, but the table does bring out the important differences. In the longer term, the disadvantages of the matrix method may disappear as the practical knowledge and the ingenuity of observers are brought to bear upon it.

The linearising assumptions limit the validity of the quasi-scalar approach. It is not clear to me whether, within these limits, linearisation helps to constrain the poldistortion. Clearly, if the degree of polarization is assumed to be small in the source and found to be small in the image, polconversion must be small as well, but I have not succeeded in casting this argument in a convincing mathematical form.

10.2. Results and prospects

Use of a 2×2 matrix to represent the coherency results in an analysis whose form is an exact replica of the scalar

Table 2. Comparison of the properties of quasi-scalar intensity selfcal versus matrix self-alignment, both with inclusion of the subsequent poldistortion elimination step

Quasi-scalar	Matrix
Small-error/weak-polarization approximation.	Exact representation.
Homogeneous arrays required.	Arbitrary arrays allowed.
Fragmented stepwise self-alignment, treating successive error types in mutually isolated procedures. Coupling may be introduced through iteration.	Self-alignment includes all errors in one procedure.
Lack of overall perspective obscures view of what actually happens. Scattering and poldistortion intertwined.	Holistic perspective, individual effects clearly separated and identifiable: Alignment, dynamic range, scattering, poldistortion, polconversion, polrotation.
Faraday rotation to be externally calibrated prior to calibrating polrotation.	Faraday-rotation variations absorbed in self-alignment. One overall rotation to be calibrated externally.
Due to intermixing of various effects dynamic range cannot be clearly assessed.	Dynamic range strictly defined in self-alignment, should be comparable to that in scalar selfcal on unpolarized sources.
Second-order effects produce nasty artefacts: $(Q, U) \Rightarrow I$ "inverse leakage" shows up as interferometer-based errors in the I, V selfcal (Massi et al. 1996).	All higher-order effects are properly accounted for.
Intensity selfcal measures leakage terms absolutely per interferometer.	Leakage terms absorbed in antenna Jones matrices derived in self-alignent.
Intensity calibration suppresses polconversion through determination of leakage terms.	Intensity self-alignment suppresses polconversion directly.
Two axes of polrotation suppressed through determination of leakage terms.	Two axes of polrotation suppressed through least-squares fit of feed errors.
Homogeneous array: Phase difference representing third axis of polrotation must be measured	
Heterogeneous array: Intractable	Heterogeneous array: Phase difference does not exist as an independent term. It is coupled to the feed parameters and the feed-error fit takes care of it.

one. Since the algebra of matrices follows almost the same rules as that of scalars, many expressions of scalar interferometry remain valid when we reinterpret the variables as matrices. In this respect, the matrix representation of coherency turns out to be preferable over the vector form of Paper I. The close analogy allows us to retain most of our familiar ways of thinking and continue to reap the fruits of half a century of theoretical and instrumental developments.

There is one important exception to the conformity: matrix multiplication is non-commutative. This means that factors in a multiple product cannot be arbitrarily merged; for example, the factors Y and $Y^{\dagger}$ in Eq. (17) donot cancel, even though their product equals $\mathbf{I}$. Non-commutativity together with the fourfold content of Jones and coherency matrices gives rise to important effects that have no scalar counterpart.

The crux of these is that the matrix analogue of self-calibration fails to actually *calibrate* but only *aligns* an observation. The indeterminacy that remains is the matrix analogue of the unknown scale factor in scalar selfcal. It entails, in addition to a similar scale factor, an unknown transformation of the brightness distribution: the *poldistortion*.

Further analysis showed that the latter is the product of *a.* a *polrotation* of the Stokes *polvector* (Q, U, V) in its three-dimensional vector space; and *b.* a *polconversion* between this polvector and Stokes I, the total brightness. Both of these are *in-place* transformations: Like scalar selfcal for a scalar (i.e. unpolarized) source, matrix self-alignment should be highly effective in suppressing spatial scattering of the matrix brightness and thereby produce images with a high dynamic range.

Methods for eliminating the poldistortion follow the pattern established in quasi-scalar polarimetry:

Intensity self-alignment on unpolarized reference sources suppresses the polconversion. This has an interesting consequence that had not been recognised

before. Indeed, the remaining polrotation leaves both the total intensity and the length of the polvector invariant: although not fully calibrated, an intensity-aligned array measures the *degree* of polarization correctly. As in the quasi-scalar method, prior knowledge of the average feed characteristics can be applied to suppress two cartesian components of the polrotation. In conventional homogeneous arrays, a (phase) measurement is necessary to eliminate the remaining one.

Situations in which no unpolarized reference sources can be used are problematic in either context. In the quasi-scalar context, the variation of parallactic angle in an alt-az-mounted antenna has been advanced as a means to separate poldistortion from true source structure; it seems to provide for at least a partial solution. One may hope that it can be put to good use in matrix form as well, but this remains to be shown.

In addition to shedding a different light on polarimetric calibration, the matrix formalism allows us to consider heterogeneous arrays. I see several possible applications:

- Full freedom in the choice of feeds may become important for a new generation of arrays composed of stationary dipole elements;
- In a heterogeneous array, the functional form of the parallactic-angle effect may be different for every antenna. This may enhance its usefulness in separating source structure and polconversion;
- Most VLBI observations are made with ad-hoc combinations of antennas that were independently built. Their polarimetric quality should greatly benefit from the replacement of makeshift leaky ad-hoc feeds with the properly designed native feeds of the participating telescopes.

An interesting property of a heterogeneous array is that receiver phase is coupled to the feed parameters in such a way that aligning the latter to their nominal average has the effect of aligning the phases at the same time. No additional phase measurement is needed.

To become a practical reality, matrix-based interferometry demands an entirely new set of matrix-based computer programs. These must incorporate procedures for matrix self-alignment and the treatment of polconversion and polrotation, along with collateral ones e.g. for the extraction of polarized source models. Having endorsed Paper I as the basis of its data and processing model, AIPS++ (1998) is the obvious environment in which such software can be developed. A project is underway at NFRA using a special observation made with the Westerbork Telescope in a heterogeneous configuration.

Handicapped by a scalar foundation that is fundamentally incorrect, radio astronomy has been remarkably successful in producing meaningful results. Now at last, we can transplant our accumulated understanding and experience into a conceptual environment that does full justice to the basic vector nature of electromagnetic radiation, without sacrificing what we have learnt in the scalar domain. What remains to be seen is not *if*, but only *when* the transition will actually happen.

11. Appendix: Mathematical theory

A. Quaternion algebra

Quaternions are not part of the standard mathematical toolkit of physicists and engineers. They do find applications in the practical calculation of rotations, e.g. in computer animation, but the relevant texts concentrate on that particular application (e.g. Kuipers 1998; see also many entries on the World-Wide Web). In scientific textbooks (e.g. Cornbleet 1976; Korn & Korn 1961) one may find brief references to them but hardly anything more. Hestenes (1986) mentions them as a variant of the concepts of *bivectors* and *spinors* that have a more central place in his text, and most of the concepts that we need are to be found there in one form or another; however, his work is not very accessible as a quick reference. For lack of better, I give here a brief summary of quaternion theory in the form in which I use it.

In the algebraic view, the vector part of a quaternion is a generalisation of the imaginary part of a complex number; correspondingly, the scalar and vector components are real and the square of a unit vector equals -1. Hestenes' version of quaternions emphasises the geometrical viewpoint that also underlies my work; it is then more appropriate for a unit vector squared to equal $+1$, which is achieved by inserting factors $i = \sqrt{-1}$ in the Pauli-matrix definitions. Another departure from the algebraic approach is that we explicitly allow the coefficients to be complex. The reader consulting literature on quaternions should be aware of these possible differences.

Another connection that looks interesting is that with the theory of special relativity, in which four-vectors appear with three spatial and one temporal component. The discussion of the Lorentz transformation by Feynman et al. (1975) shows a close formal analogy with my poldistortions (which was recently also noted by Britton 2000), but I have found no useful inspiration in it.

I derive here the algebraic rules that quaternions must follow to make them behave in the same way as the equivalent matrices.

Once we have these rules in place we may proclaim, in mathematical language, that *the multiplicative quaternion*

Table 3. Algebraic entities and their notation as used in this paper

Entity	Vector	2×2 Matrix	Stokes vector	Quaternion
Coherency	$\boldsymbol{e}_{jk} = \boldsymbol{e}_j \otimes \boldsymbol{e}_k^*$	$\boldsymbol{E}_{jk} = \boldsymbol{e}_j \boldsymbol{e}_k^\dagger$	$\boldsymbol{s}_{jk} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & -i & i & 0 \end{pmatrix} \boldsymbol{e}_{jk}$	
General element	$\begin{pmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{pmatrix}$	$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ $= \begin{pmatrix} a + a_1 & a_2 - ia_3 \\ a_2 + ia_3 & a - a_1 \end{pmatrix}$	$\begin{pmatrix} a \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a \\ \boldsymbol{a} \end{pmatrix}$	$[a + \boldsymbol{a}],$ $\boldsymbol{a} \equiv \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}$
Scalar part			intensity a	$[a + \boldsymbol{0}] \equiv [a] \equiv a$
Vector part			polvector $\boldsymbol{a}$	$[0 + \boldsymbol{a}] \equiv [\boldsymbol{a}] \not\equiv \boldsymbol{a}$
Base for vector part		$\mathbf{Q}, \mathbf{U}, \mathbf{V},$ Eq. (7)	$\mathbf{1_q}, \mathbf{1_u}, \mathbf{1_v},$ Eq. (28)	$[\mathbf{1_q}], [\mathbf{1_u}], [\mathbf{1_v}]$
Alternate base		$\mathbf{V}, \mathbf{Q}, \mathbf{U},$ Eq. (30)	$\mathbf{1_v}, \mathbf{1_q}, \mathbf{1_u}$	$[\mathbf{1_v}], [\mathbf{1_q}], [\mathbf{1_u}]$
Unit element		$\mathbf{I}$		$[1] = 1$
Conjugation		$\boldsymbol{A}^\dagger$	$\begin{pmatrix} a^* \\ \boldsymbol{a}^* \end{pmatrix}$	$[a^* + \boldsymbol{a}^*]$
Multiplication		$\boldsymbol{AB}$		Eq. (25)
Unimodular unitary matrix		$\mathbf{Y}$		$\cos\eta + i\mathbf{1_y}\sin\eta$
Unimodular positive hermitian matrix		$\mathbf{H}$		$\cosh\gamma + \mathbf{1_h}\sinh\gamma$

group and the multiplicative group of 2×2 matrices are *isomorphous* (Korn & Korn 1961). In simpler terms, we may consider "2×2 matrix" and "quaternion" as names for the same object in two different languages, or even as synonyms. Such phrases as "the scalar and vector parts of a matrix" will then make sense.

In Table 3 I present a dictionary for translating from matrix to quaternion language.

A.1. Addition

I start from the matrix/quaternion equivalence of Sect. 2.5:

$$\boldsymbol{A} = [\, a + \boldsymbol{a}\,] \equiv [\, a_0 + (a_1, a_2, a_3)\,].$$

The use of the "+" sign is justified by the expansion that the right-hand side actually represents:

$$\boldsymbol{A} = a_0\,\mathbf{I} + a_1\,\mathbf{Q} + a_2\,\mathbf{U} + a_3\,\mathbf{V}. \tag{22}$$

The addition rules for quaternions are the obvious ones as can be shown in the same way.

A.2. Transposition and conjugation

The definition of the vector part of a quaternion does not specify whether it is a row or column vector and the same is true for the dot and cross products that we will need later.

Since the Pauli matrices are hermitian, it follows that

$$\boldsymbol{A}^\dagger = [\, a^* + \boldsymbol{a}^*\,]. \tag{23}$$

A.3. Multiplication

The following identities follow directly from the definitions Eq. (7) of the Pauli matrices:

$$\begin{aligned} \mathbf{Q}^\dagger &= \mathbf{Q}, \quad \mathbf{U}^\dagger = \mathbf{U}, \quad \mathbf{V}^\dagger = \mathbf{V} \\ \mathbf{Q}\,\mathbf{Q} &= \mathbf{U}\mathbf{U} = \mathbf{V}\mathbf{V} = \mathbf{I} \\ \mathbf{Q}\,\mathbf{U} &= -\mathbf{U}\mathbf{Q} = i\,\mathbf{V} \\ \mathbf{U}\,\mathbf{V} &= -\mathbf{V}\mathbf{U} = i\,\mathbf{Q} \\ \mathbf{V}\,\mathbf{Q} &= -\mathbf{Q}\mathbf{V} = i\,\mathbf{U}. \end{aligned} \tag{24}$$

Now consider the product

$$\boldsymbol{AB} = [\, a + \boldsymbol{a}\,][\, b + \boldsymbol{b}\,].$$

Writing out the expansions Eq. (22) for $\boldsymbol{A}$ and $\boldsymbol{B}$ and multiplying term by term using Eq. (24), one finds the *multiplication rule*

$$[a + \boldsymbol{a}][b + \boldsymbol{b}] = [(ab + \boldsymbol{a}\cdot\boldsymbol{b}) + (a\boldsymbol{b} + b\boldsymbol{a} + i\boldsymbol{a} \times \boldsymbol{b})]. \quad (25)$$

An important corrolary is that $[\mathbf{1}_{\boldsymbol{x}}]^2 = [1]$ for any unit vector $\mathbf{1}_{\boldsymbol{x}}$.

Like the multiplication of the equivalent matrices, quaternion multiplication is generally non-commutative. An exception occurs when $\boldsymbol{a}$ and $\boldsymbol{b}$ are collinear so that $\boldsymbol{a} \times \boldsymbol{b} = \mathbf{0}$. In that case I also call the matrices/quaternions $[a + \boldsymbol{a}]$ and $[b + \boldsymbol{b}]$ collinear.

Also note that the product of two real quaternions is not real unless they commute. The corresponding matrix property is that the product of two hermitian matrices is generally non-hermitian.

A.4. Scalars as quaternions

The *scalar quaternion* $[a]$ represents the 2×2 matrix $a\mathbf{I}$, and since $\mathbf{I}\mathbf{I} = \mathbf{I}$, both the addition and the multiplication rule for scalar quaternions are the same as those for scalars. We may consider scalars as a subset of the quaternions:

$$a\mathbf{I} \equiv [a] \equiv a.$$

A.5. Determinant, trace and variance

Every familiar property of a 2×2 matrix has a quaternion counterpart. Thus a quaternion has a determinant

$$\det[a + \boldsymbol{a}] = a^2 - \boldsymbol{a}^2 \quad (26)$$

and $\det \boldsymbol{A}\boldsymbol{B} = \det \boldsymbol{A} \det \boldsymbol{B}$. A *unimodular* matrix/quaternion is one whose determinant equals 1.

The trace is

$$\mathrm{Tr}\,[a + \boldsymbol{a}] = 2a$$

$\mathrm{Tr}\,\boldsymbol{A} = \mathrm{Tr}\,\boldsymbol{A}^{\mathrm{T}}$, $\mathrm{Tr}\,(\boldsymbol{A} + \boldsymbol{A}^\dagger) = 2\mathrm{Tr}\,\mathrm{Re}\,\boldsymbol{A}$, and $\mathrm{Tr}\,(\boldsymbol{A}\boldsymbol{B}) = \mathrm{Tr}\,(\boldsymbol{B}\boldsymbol{A})$. The coefficients of the Pauli matrices in Eq. (7) are given by expressions such as

$$a_0 = \tfrac{1}{2}\mathrm{Tr}\,\boldsymbol{A}\mathbf{I}; \qquad a_1 = \tfrac{1}{2}\mathrm{Tr}\,\boldsymbol{A}\mathbf{Q}; \qquad \text{etc.}$$

The trace is invariant under a unitary transformation:

$$\mathrm{Tr}\,(\boldsymbol{Y}\boldsymbol{A}\boldsymbol{Y}^\dagger) = \mathrm{Tr}\,\boldsymbol{A}.$$

I define the *variance* of a matrix as the sum of the moduli squared of its elements. It is the square of the "Frobenius norm" (Lancaster & Tismenetsky 1985) and given by

$$\mathrm{Var}\,\boldsymbol{A} = \mathrm{Tr}\,(\boldsymbol{A}\boldsymbol{A}^*) = aa^* + \boldsymbol{a} \cdot \boldsymbol{a}^* \quad (27)$$

$\mathrm{Var}\,(\boldsymbol{A} - \boldsymbol{B})$ may be used as a measure of "how different" $\boldsymbol{A}$ and $\boldsymbol{B}$ are. It is readily shown that, like the trace, the variance is also invariant under unitary transformations.

A.6. Coordinate systems

The vector parts of quaternions form a three-dimensional *quaternion-vector space*. It is convenient to choose the coordinates in this space in accordance with our definition of the Stokes quaternion. When we express the electric field vectors in geometric xy coordinates, and use the conventional definition of the Stokes vector (cf. Paper I), the definition Eq. (6) follows. The corresponding base vectors are the quaternion vectors corresponding to the Pauli matrices $\mathbf{Q}$, $\mathbf{U}$ and $\mathbf{V}$:

$$\mathbf{1_q} = (1,0,0), \quad \mathbf{1_u} = (0,1,0), \quad \mathbf{1_v} = (0,0,1). \quad (28)$$

Analogously to Eq. (24) we have

$$[\mathbf{1_q}][\mathbf{1_u}] = -[\mathbf{1_u}][\mathbf{1_q}] = i[\mathbf{1_v}] \qquad \text{etc.} \quad (29)$$

If, instead, we use circular rl coordinates to describe the electric field, this results in a cyclic permutation of the coordinate axes (Paper I) and instead of Eq. (22) we have

$$\boldsymbol{B} = a\,\mathbf{I} + a_1\,\mathbf{V} + a_2\,\mathbf{Q} + a_3\,\mathbf{U}. \quad (30)$$

This form is convenient for analysing systems with nominally circular feeds.

B. Special matrices

B.1. Unitary matrices

A 2×2 matrix $\boldsymbol{Y}$ is *unitary* if $\boldsymbol{Y}\boldsymbol{Y}^\dagger = \mathbf{I}$. To derive its quaternion form I cast the most general quaternion $[a + \boldsymbol{a}]$ in the form

$$\boldsymbol{Y} = \mathrm{e}^{i\xi}[y + \boldsymbol{x} + i\,\boldsymbol{y}], \quad y, \boldsymbol{x}, \boldsymbol{y}\ \text{real}. \quad (31)$$

I now expand $\boldsymbol{Y}\boldsymbol{Y}^\dagger$ in quaternion form and require that it equal $\mathbf{I}$:

$$\boldsymbol{Y}\boldsymbol{Y}^\dagger = [y + \boldsymbol{x} + i\,\boldsymbol{y}][y + \boldsymbol{x} - i\,\boldsymbol{y}]$$
$$= [y^2 + \boldsymbol{x}^2 + \boldsymbol{y}^2 + 2(y\boldsymbol{x} + \boldsymbol{x} \times \boldsymbol{y})] = [1].$$

Since $\boldsymbol{x} \times \boldsymbol{y}$ is perpendicular to $\boldsymbol{x}$, the vector part in the product can vanish only if $\boldsymbol{x} = \mathbf{0}$. It follows that

$$y^2 + \boldsymbol{y}^2 = 1.$$

Now, if $\boldsymbol{Y}$ is to be unimodular, ξ must be 0. Hence we may rewrite Eq. (31) as

$$\boldsymbol{Y} = [\cos\eta + i\mathbf{1}_{\boldsymbol{y}}\sin\eta] \quad (32)$$

$\boldsymbol{Y}$ is completely defined by the three real components of its *Gibbs vector* (Korn & Korn 1961) $\mathbf{1}_{\boldsymbol{y}}\sin\eta$.

Since $\mathbf{1}_{\boldsymbol{y}}^2 = 1$, it can be shown from the Taylor-series expansions of the cosine and sine functions that

$$\cos[\mathbf{1}_{\boldsymbol{y}}]\eta = \cos\eta, \quad \sin[\mathbf{1}_{\boldsymbol{y}}]\eta = [\mathbf{1}_{\boldsymbol{y}}]\sin\eta$$

and hence

$$\boldsymbol{Y} = \exp i[\mathbf{1}_{\boldsymbol{y}}\eta].$$

For small η, we may replace the exponential by its first-order approximation. The value of η then provides a direct measure for the deviation of $\boldsymbol{Y}$ from $\mathbf{I}$.

B.2. Complex linear polarized brightness

In the quasi-linear treatment of polarization, the linearly polarized visibility and brightness frequently appear in the form of the complex variable $Q+iU$, cf. (Eq. 19). This form is directly related to the quaternion exponential above.

In particular, for a unit vector in the $\mathbf{q}, \mathbf{u}$ plane of quaternion-vector space

$$[\mathbf{1_q}\cos\phi + \mathbf{1_u}\sin\phi] = [\mathbf{1_q}][\cos\phi - i\mathbf{1_v}\sin\phi]$$
$$= [\mathbf{1_q}]\exp -i[\mathbf{1_v}]\phi.$$

B.3. Unitary Jones matrices and perfect feeds

In Sect. 8 unitary Jones matrices were postulated. Since $\boldsymbol{YIY}^\dagger = \mathbf{I}$, a feed with such a matrix transfers all incident radiation to its output: it must be loss-free and matched at its in- and outputs. Matching implies that either receptor must absorb all the radiation that the other one does not: the receptors must be of opposite polarizations (Born & Wolf 1964; Cornbleet 1976; Thompson et al. 1986).

This is the way feeds for radio telescopes are normally designed. Note, however, that e.g. a stationary pair of crossed dipole receptors is not matched to radiation from an arbitrary direction. Designs for arrays of phased dipoles will have to take the problems ensuring into account.

B.4. Positive hermitian matrices

$\boldsymbol{H}$ is *hermitian* or *self-adjoint* if $\boldsymbol{H} = \boldsymbol{H}^\dagger$; since the Pauli matrices are hermitian, the quaternion form $\boldsymbol{H} = [h+\boldsymbol{h}]$ of a hermitian 2×2 matrix is real, cf. Eq. (23). A 2×2 matrix is positive if its eigenvalues are both positive. An equivalent condition is that both its trace and its determinant are positive.

For the matrix to be positive hermitian and unimodular

$$h > 0; \qquad h^2 - \boldsymbol{h}^2 = 1.$$

We may then write it as

$$\boldsymbol{H} = \cosh\gamma + [\mathbf{1_h}]\sinh\gamma \equiv \exp[\mathbf{1_h}]\gamma. \tag{33}$$

It is completely defined by the three real components of $\mathbf{1_h}\sinh\gamma$ which I will also call a Gibbs vector. It is readily shown from Eq. (33) that

$$\boldsymbol{H}^2 = \exp[\mathbf{1_h}]2\gamma. \tag{34}$$

B.5. Matrix square root

We will need the positive hermitian square root $\boldsymbol{H}$ of the product $\boldsymbol{MM}^\dagger$ for an arbitrary 2×2 matrix $\boldsymbol{M}$. Let

$$\boldsymbol{M} = [m + \boldsymbol{m}].$$

Then

$$\boldsymbol{A} = \boldsymbol{MM}^\dagger = [mm^* + \boldsymbol{m}\cdot\boldsymbol{m}^* + 2\mathrm{Re}\,m\boldsymbol{m}]$$

is readily shown to be positive hermitian. We now seek to find a positive hermitian matrix $\boldsymbol{H}$ such that $\boldsymbol{HH} = \boldsymbol{A}$, that is

$$h^2 + \boldsymbol{h}^2 = a; \quad 2h\boldsymbol{h} = \boldsymbol{a}. \tag{35}$$

Conceptually the simplest way to find the root is through Eqs. (33) and (34). Computationally it is more efficient to solve the quadratic equation that Eq. (35) represents. Out of four possible solutions, the positive definite one is

$$\boldsymbol{H} = \sqrt{(a + \sqrt{a^2 - \boldsymbol{a}^2})/2}$$
$$+ [\mathbf{1_a}]\sqrt{(a - \sqrt{a^2 - \boldsymbol{a}^2})/2}.$$

B.6. Polar decomposition

An arbitrary matrix $\boldsymbol{X}$ can be represented (Lancaster & Tismenetsky 1985) as the product of a unitary and a positive hermitian matrix:

$$\boldsymbol{X} = \boldsymbol{HY} \quad \text{with} \quad \boldsymbol{YY}^\dagger = \mathbf{I} \quad \text{and} \quad \boldsymbol{H}^\dagger = \boldsymbol{H}.$$

This is the matrix/quaternion analogue of the polar form of a complex scalar. Defining

$$x = \sqrt{\det \boldsymbol{X}}$$

we may rewrite the decomposition as

$$\boldsymbol{X} = x\boldsymbol{HY}$$

where $\boldsymbol{H}$ and $\boldsymbol{Y}$ are now unimodular. To find $\boldsymbol{H}$ we form the product

$$(xx*)^{-1}\boldsymbol{XX}^\dagger = \boldsymbol{HY}\,\boldsymbol{Y}^\dagger\boldsymbol{H}^\dagger = \boldsymbol{HH}$$

and find $\boldsymbol{H}$ by taking the positive hermitian square root.

C. Congruence transformations

The *congruence transformation* is defined in Eqs. (3) and (14). Substituting the polar decomposition for $\boldsymbol{X}$ we get Eq. (15):

$$\boldsymbol{B}' = xx^* \boldsymbol{H}\,(\boldsymbol{YBY}^\dagger)\,\boldsymbol{H}^\dagger.$$

Since $\boldsymbol{H}$ and $\boldsymbol{Y}$ are unimodular, $\det \boldsymbol{B}' = (xx^*)^2 \det \boldsymbol{B}$ or, apart from the scale factor

$$b'^2 - \boldsymbol{b}'^2 = b^2 - \boldsymbol{b}^2.$$

(In the main text of this paper, b and $\boldsymbol{b}$ are written as I and $\boldsymbol{p}$, respectively, to emphasize the physical interpretation of $\boldsymbol{B}$ as a brightness.)

The effect of the component unitary and positive hermitian transformations can now be analysed by replacing $\boldsymbol{B}$, $\boldsymbol{Y}$ and $\boldsymbol{H}$ with their equivalent quaternions and carrying out the multiplications:

C.1. Unitary transformations

Consider the most general unitary transformation
$$[\,b' + \boldsymbol{b}'_\| + \boldsymbol{b}'_\perp\,]$$
$$= \mathrm{e}^{i\psi}[\cos\eta + i\mathbf{1_y}\sin\eta\,]\,[\,b + \boldsymbol{b}_\| + \boldsymbol{b}_\perp\,]\,\mathrm{e}^{-i\psi}[\cos\eta - i\mathbf{1_y}\sin\eta\,]$$
where $\boldsymbol{b}_\|$ and $\boldsymbol{b}_\perp$ are the components of $\boldsymbol{b}$ that are parallel and perpendicular to $\mathbf{1_y}$.

The unitary quaternions are collinear with $[\,b + \boldsymbol{b}_\|\,]$ so
$$[\,b' + \boldsymbol{b}'_\|\,] = [\,b + \boldsymbol{b}_\|\,].$$
For $\boldsymbol{b}_\perp$ we must carry out the multiplications, and obtain
$$[\,\boldsymbol{b}'_\perp\,] = [\,\boldsymbol{b}_\perp\,\cos 2\eta - \mathbf{1_y}\times\boldsymbol{b}_\perp\,\sin 2\eta\,].$$
That is, $\boldsymbol{b}'_\perp$ is a copy of $\boldsymbol{b}_\perp$ rotated over an angle 2η.

The two results combined show that a unitary transformation $\boldsymbol{Y}$ leaves the scalar part of its input invariant and rotates its vector part around the axis $\mathbf{1_y}$ over an angle 2η; this is the polrotation effect of Sect. 5.1. The scalar and vector parts are transformed *independently*.

The vector $\mathbf{1_y}\sin\eta$ that characterises the rotation is known as the *Gibbs vector* (Korn & Korn 1961). Vectors collinear with it are invariant; they are *eigenvectors* of the rotation.

From the theory of linear transformations I take the result that the mathematical expression for a rotation is
$$\boldsymbol{b}' = \boldsymbol{R}\,\boldsymbol{b}$$
where the 3×3 transformation matrix $\boldsymbol{R}$ is *real, orthogonal* and *unitary*. Its unitarity guarantees the invariance of scalar products and in particular of real vector lengths.

The Euclidian rotations of the vector part $\boldsymbol{b}$ form a subset of the general pseudo-Euclidian rotations represented by the Lorentz transformation (Feynman et al. 1975), just as the unitary transformations form a subset of the general congruence transformations.

C.2. The fundamental unitary matrices

Any rotation in a three-dimensional Euclidian space can be represented as a succession of three rotations around mutually perpendicular axes (Korn & Korn 1961). Choosing for these axes the three base vectors of Eq. (28) we find the three respective basic unimodular unitary matrices

- The *phase-difference* transformation
$$\boldsymbol{Y_q}(\phi) = \begin{pmatrix} \exp i\phi & 0 \\ 0 & \exp -i\phi \end{pmatrix} = \exp i\,[\mathbf{1_q}]\phi. \tag{36}$$

- The *ellipticity* transformation
$$\boldsymbol{Y_u}(\epsilon) = \begin{pmatrix} \cos\epsilon & i\sin\epsilon \\ i\sin\epsilon & \cos\epsilon \end{pmatrix} = \exp i\,[\mathbf{1_u}]\epsilon. \tag{37}$$

- The (feed or xy frame) *rotation* transformation
$$\boldsymbol{Y_v}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \exp i\,[\mathbf{1_v}]\theta. \tag{38}$$

C.3. The polrotation in circular coordinates

It is of some interest to consider the form that Eq. (18) takes in circular coordinates. The feed and receiver terms $\boldsymbol{Y_u}$ and $\boldsymbol{Y_q}$ operate on the signal vector formed by the l and r voltages in the feed-receiver system and need not change. The geometric rotation term $\boldsymbol{Y_v}(\theta)$ must be transformed to the circular lr coordinate frame (Paper I) in which the radiation is measured. From Paper I we take the result that this transformation transforms Stokes V into Q, hence $\mathbf{1_v}$ into $\mathbf{1_q}$. Thus $\boldsymbol{Y_v}(\theta)$ assumes the form $\boldsymbol{Y_q}(\theta)$, and the equivalent of Eq. (18) becomes
$$\boldsymbol{Y} = \boldsymbol{Y_q}(\phi)\,\boldsymbol{Y_u}(\epsilon)\,\boldsymbol{Y_q}(\theta). \tag{39}$$
This is another generic way of factoring an arbitrary rotation (Korn & Korn 1961).

When the feed-error term $\boldsymbol{Y_u}$ is reduced to unity, the remaining two $\boldsymbol{Y_q}$ terms fuse into one. The usual practice of merging them regardless of the feed errors amounts to inverting the order of the factors in Eq. (39); this is justifiable only in the quasi-scalar approximation (cf. Sect. 7).

C.4. Positive hermitian transformations

The treatment of the unimodular positive hermitian transformation is analogous to that of the unitary one. Starting from Eq. (33) for $\boldsymbol{H}$ one finds
$$\begin{aligned} b' &= b && \cosh 2\gamma + \boldsymbol{b}_\|\cdot\mathbf{1_h} && \sinh 2\gamma \\ \boldsymbol{b}'_\| &= b\,\mathbf{1_h}\,\sinh 2\gamma + \boldsymbol{b}_\| && \cosh 2\gamma \\ \boldsymbol{b}'_\perp &= \boldsymbol{b}_\perp. \end{aligned} \tag{40}$$
The effect is in a sense complementary to that of the unitary transformation, but there is no analogous geometric interpretation. The scalar and vector parts are not transformed independently but get mixed: this is the polconversion effect of Sect. 5.1.

There is no interaction between vector components in different directions: The transformation is said to be *rotation-free*.

C.5. Minimal-variance theorem

Of all complex numbers $z = a\,\mathrm{e}^{i\phi}$, $z = a$ has the smallest distance squared $|(z-1)|^2$ to unity. I shall now prove an analogous property for 2×2 matrices:

Theorem: Let $\boldsymbol{H}$ be a given positive hermitian and $\boldsymbol{Y}$ an arbitrary unitary matrix. Of all products $\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{H}$, $\boldsymbol{X} = \boldsymbol{H}$ minimises the quantity
$$F_{\boldsymbol{H}}(\boldsymbol{Y}) = \mathrm{Var}\,(\boldsymbol{Y}\boldsymbol{H} - \mathbf{I}).$$

Using the definition of Var , the commutation and transposition invariance of Tr and the hermiticity of $\boldsymbol{H}$, convert the equation to
$$\begin{aligned} F_{\boldsymbol{H}}(\boldsymbol{Y}) &= \mathrm{Tr}\,(\boldsymbol{Y}\boldsymbol{H} - \mathbf{I})(\boldsymbol{H}\boldsymbol{Y}^\dagger - \mathbf{I}) \\ &= \mathrm{Tr}\,(\boldsymbol{H}\boldsymbol{H} + \mathbf{I} - 2\mathrm{Re}\,\boldsymbol{Y}\boldsymbol{H}). \end{aligned} \tag{41}$$

To find the minimum, consider the term

$$\text{Tr Re } \boldsymbol{YH} = \text{Tr Re } \exp i\psi [\cos\eta + i\mathbf{1}_y \sin\eta] [h+\boldsymbol{h}] \\ = h\cos\psi\cos\eta - \mathbf{1}_y \cdot \boldsymbol{h} \sin\psi\sin\eta. \tag{42}$$

Differentiation with respect to ψ and η gives the equations

$$h\sin\psi\cos\eta + \mathbf{1}_y \cdot \boldsymbol{h}\cos\psi\sin\eta = 0$$

$$h\cos\psi\sin\eta + \mathbf{1}_y \cdot \boldsymbol{h}\sin\psi\cos\eta = 0$$

whose obvious solutions are $\cos\psi = \cos\eta = 0$ and $\sin\psi = \sin\eta = 0$.

Since $\boldsymbol{H}$ is positive hermitian,

$$h > |\boldsymbol{h}| \geq \mathbf{1}_y \cdot \boldsymbol{h}$$

so a. there are no other solutions and b. the former one, which is equivalent to $\boldsymbol{Y} = \mathbf{I}$, maximises Eq. (42) and hence minises $F_{\boldsymbol{H}}(\boldsymbol{Y})$. Moreover this is true for all directions $\mathbf{1}_y$.

D. Matrix solution techniques

The methods discussed in this paper require nonlinear minimisation of the variance of a function of several matrices wrt one of them. One way to handle it is by writing out all equations and their derivatives in terms of the real and imaginary parts of all matrix elements and then applying standard nonlinear solution methods (Press et al. 1989). This is the approach used in AIPS++ (1998); it is cumbersome and the resultant code is complex and difficult to verify (T. Cornwell, private communication).

In the methods to be described below, entire matrices are the atomic variables. Thus, full advantage is taken of the conceptual efficiency of matrix algebra, which in turn reflects in simple solution algorithms that are very easily coded. The algorithms explicitly exploit the structure of the equations; for this reason they may well be more efficient than the general-purpose approach. They also lend themselves to quick experiments in an environment such as AIPS++ (1998) in which matrix operations can be coded directly.

D.1. Differentiation

The problem is that of finding the matrix $\boldsymbol{V}$ that minimises the variance of some matrix function $\boldsymbol{M}$ of $\boldsymbol{V}$. Attacking this problem in the conventional way requires the definition of the derivative of Var $\boldsymbol{M}(\boldsymbol{V})$ with respect to $\boldsymbol{V}$. An equivalent approach that requires no new definitions is to consider the differentials themselves rather than their quotient.

For a variation $\delta\boldsymbol{M}$, the corresponding variation in Var $\boldsymbol{M}$ is (cf. Eq. 41)

$$\delta\text{Var } \boldsymbol{M} = \text{Tr } (\boldsymbol{M}\,\delta\boldsymbol{M}^\dagger + \delta\boldsymbol{M}\,\boldsymbol{M}^\dagger) \\ = 2\,\text{Tr Re } (\boldsymbol{M}\,\delta\boldsymbol{M}^\dagger). \tag{43}$$

In the applications of interest, $\boldsymbol{M}\,\delta\boldsymbol{M}^\dagger$ is a sum of (products of) several other matrices, one of which is $\delta\boldsymbol{V}$. In each

of these products, we may cyclically permute the factors to move $\delta\boldsymbol{V}$ to the trailing position (cf. Appendix C.5). Thus we convert each product term in Eq. (43) to the form

$$\text{Tr Re } (\boldsymbol{Z}\,\delta\boldsymbol{V}) = 0. \tag{44}$$

If $\boldsymbol{V}$ is constrained, e.g. to being diagonal or unitary, corresponding constraints are to be imposed upon $\delta\boldsymbol{V}$.

If, for any permitted variation $\delta\boldsymbol{V}$, the variation $i\,\delta\boldsymbol{V}$ is also allowed, Eq. (44) can be simplified by omitting the Re operator. If, moreover, $\delta\boldsymbol{V}$ is completely free, Eq. (44) implies that $\boldsymbol{Z}$ itself is $\mathbf{0}$.

D.2. Self-alignment decomposition

I show a simple least-squares self-aligment algorithm as an example. Given a set of observed coherencies $\boldsymbol{W}_{jk}$ and a source model $\boldsymbol{E}'_{jk}$, we seek to fit values $\boldsymbol{J}'_j$ that minimise the noise power at the interferometer inputs:

$$S = \sum_{jk} \text{Var } (\boldsymbol{J}'^{-1}_j \boldsymbol{W}_{jk} \boldsymbol{J}'^{-1\dagger}_k - \boldsymbol{E}'_{jk}).$$

For a change $\delta\boldsymbol{J}'^{-1}_j$ we have, from Eq. (43)

$$\delta S = 2\text{Tr Re } \big(\sum_k \ (\boldsymbol{J}'^{-1}_j W'_{jk} \boldsymbol{J}'^{-1\dagger}_k - \boldsymbol{E}_{jk}) \\ \boldsymbol{J}'^{-1}_k \boldsymbol{W}'^{\dagger}_{jk} \, \delta(\boldsymbol{J}'^{-1\dagger}_j) \big) = 0.$$

Since $\delta(\boldsymbol{J}'^{-1\dagger}_j)$ is arbitrary, it follows that

$$\boldsymbol{J}'^{-1}_j \sum_k \boldsymbol{W}'_{jk} \boldsymbol{J}'^{-1\dagger}_k \ \boldsymbol{J}'^{-1}_k \boldsymbol{W}'^{\dagger}_{jk} = \sum_k \boldsymbol{E}_{jk} \boldsymbol{J}'^{-1}_k \boldsymbol{W}'^{\dagger}_{jk}.$$

Given a set of estimates $\boldsymbol{J}'^{-1}_k$, this equation provides the basis for an iterative algorithm by producing a new estimate for $\boldsymbol{J}'^{-1}_j$.

Note the similarity of the first three factors on the lefthand side to $\sum \boldsymbol{E}'_{jk}$. In the same way as dimension comparisons in physics, this similarity provides a partial check on the correctness of an equation.

This method is easily generalised to a more proper χ^2 form for the case where the four polarisation channels in each interferometer carry the same noise level. This is probably an adequate assumption in most practical cases.

D.3. Feed-error minimisation

Section 8.2 poses the problem of minimising

$$S = \sum_j \text{Var } (\boldsymbol{D}'_j - \mathbf{I}), \tag{45}$$

where

$$\boldsymbol{D}'_j = \boldsymbol{G}'^{-1}_j \boldsymbol{J}'_j \boldsymbol{Y}'^{-1} \boldsymbol{F}^{-1}_j$$

$\boldsymbol{G}'_j$ are unknown diagonal gain matrices and $\boldsymbol{Y}'$ is the unknown unitary polrotation matrix (which is not necessarily unimodular). Taking differentials

$$\delta S = 2 \sum_j \text{Tr Re } \big((\boldsymbol{D}'_j - \mathbf{I})\,\delta\boldsymbol{D}'^{\dagger}_j \big).$$

G'^{-1}_j is found given the current value of Y' in the way indicated above. The constraint on $\delta G'^{-1}_j$ is that it be diagonal. The result is

$$(G'^{-1}_j)_{ll} = (J_j \, J'^{\dagger}_j)_{ll} \, / \, (J'_j \, J'^{\dagger}_j)_{ll} \,, \quad l = 1, 2.$$

To solve for Y' given the current values of the G'^{-1}_j, we begin by applying unitary transformations F^{-1}_j to the summands in Eq. (45) to obtain (cf. Appendix A.5)

$$S = \sum_j \mathrm{Var}\, (F^{-1}_j \, G'^{-1}_j \, J'_j \, Y'^{-1} - I).$$

We may now minimise S by invoking the minimum-variance theorem of Appendix C.5, in combination with the fact that

$$\sum_j \mathrm{Var}\, (Z_j Y'^{-1} - I)$$

and

$$\mathrm{Var}\, \big(\sum_j (Z_j Y'^{-1} - I)\big)$$

are minimal for the same value of Y'.

References

AIPS++ 1998, Software under development by a consortium of observatories, for the processing of (primarily radio) astronomical data. See the website at NRAO: www.nrao.edu

Born M., Wolf E., 1964, Principles of Optics. Pergamon Press

Britton M.C., 2000, ApJ, April 1 (in press)

Cornbleet S., 1976, Microwave Optics. Academic Press

Cotton W.D., 1993, AJ 106, 1241

Feynman R.P., Leighton R.B., Sands M., 1975, The Feynman Lectures on Physics, Vol. 1. Addison-Wesley

Hamaker J.P., Bregman J.D., Sault R.J., 1996, A&AS 117, 137 (Paper I)

Hestenes D., 1986, New Foundations of Classical Mechanics.

Kluwer Acad. Publ.

Korn G.A., Korn T.M., 1961, Mathematical Handbook for Scientists and Engineers. McGraw Hill

Kuipers J.B., 1998, Quaternions and Rotation Sequences - A Primer with Applications to Orbits, Aersospace and Virtual Reality. Princeton Univ. Press

Lancaster P., Tismenetsky M., 1985, The Theory of Matrices. Academic Press

Landau L.D., Lifshitz E.M., 1995, The Classical Theory of Fields (Course of Theoretical Physics Vol. 2). Butterworth/Heinemann

Leppänen K.J., Zensus J.A., Diamond P.J., 1995, AJ 110, 2479

Massi M., Comoretto G, Rioja M., Tofani G., 1996, A&AS 116, 167

MATLAB, 1997, High-Performance Numeric Computation and Visualisation Software, version 5. The MathWorks, Inc, Natick, Mass, U.S.A.; http://www.mathworks.com

Narayan R., Nintyananda R., 1986, ARA&A 24, 124

Perley R.A., Schwab F.R., Bridle A.H., 1994, Synthesis Imaging in Radio Astronomy, a Collection of Lectures from the Third NRAO Synthesis Summer School, ASP Conf. Ser. 6

Press W.H., Flannery B.P., Teukolsky S.A., Vetterling W.A., 1989, Numerical Recipes – The Art of Scientific Programming. Cambridge Univ. Press

Ponsonby J.E.B., 1973, MNRAS 163, 269

Roberts D.H., Wardle J.F.C., Brown L.F., 1994, ApJ 427, 718

Sakurai T., Spangler S.R., 1994, Radio Sci. 29, 635

Sault R.J., Hamaker J.P., Bregman J.D., 1996, A&AS 117, 149 (Paper II)

Sault R.J., Bock D.C.-J., Duncan A.R., 1999, A&AS (in press)

Simmons J.W., Guttman M.J., 1970, States, waves and photons: a modern introduction to light. Addison-Wesley

Thompson A.R., Moran J.M., Swenson G.W. Jr., 1986, Interferometry and Synthesis in Radio Astronomy. John Wiley & Sons, New York

Wardle J.F.C., Homan D.C., Ojha R., Roberts D.H., 1998, Nat 395, 457

Weiler K.W., 1973, A&A 26, 404