Polca SARA - Full polarization, direction-dependent calibration and sparse imaging for radio interferometry

Jasleen Birdi, ¹★ Audrey Repetti, ^{1,2} and Yves Wiaux ¹

- ¹ Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK
- ² Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, UK

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

New generation of radio interferometers are envisaged to produce high quality, high dynamic range Stokes images of the observed sky from the corresponding under-sampled Fourier domain measurements. In practice, these measurements are contaminated by the instrumental and atmospheric effects that are well represented by Jones matrices, and are most often varying with observation direction and time. These effects, usually unknown, act as a limiting factor in achieving the required imaging performance and thus, their calibration is crucial. To address this issue, we develop a global algorithm, named Polca SARA, aiming to perform full polarization, direction-dependent calibration and sparse imaging by employing a non-convex optimization technique. In contrast with the existing approaches, the proposed method offers global convergence guarantees and flexibility to incorporate sophisticated priors to regularize the imaging as well as the calibration problem. Thus, we adapt a polarimetric imaging specific method, enforcing the physical polarization constraint along with a sparsity prior for the sought images. We perform extensive simulation studies of the proposed algorithm. While indicating the superior performance of polarization constraint based imaging, the obtained results also highlight the importance of calibrating for direction-dependent effects as well as for off-diagonal terms (denoting polarization leakage) in the associated Jones matrices, without inclusion of which the imaging quality deteriorates.

Key words: techniques: interferometric – techniques: polarimetric – techniques: image processing – techniques: calibration

1 INTRODUCTION

In radio interferometry (RI), the acquired measurements are related to the Fourier coefficients of the sky brightness distribution under observation (Thompson et al. 2001). These sampled frequency points in the Fourier domain are dictated by the separations between the antenna pairs constituting the interferometer. A finite number of antennas within an interferometer leads to undersampling of the Fourier domain associated with the sought images. Given such measurements, the task to recover the images of interest can be performed by using a suitable imaging technique in RI. In fact, very often, the target radio sources generate polarized radiations (Ginzburg & Syrovatskii 1965). Study of these polarized emissions is crucial for various scientific goals (e.g. to determine the magnetic field distributions around the source and along the signal's propagation path (Pacholczyk 1970)), offering a way to probe the details in addition to those provided by the total intensity alone. For this purpose, Stokes parameters- I, Q, U and V provide a representation of the sky intensity distribution, where the first parameter denotes the total intensity whereas the others characterize

* E-mail: jb36@hw.ac.uk

the polarization state of the radio emissions. In particular, while the linear polarization is described by the Stokes parameters Q and U, the circular polarization is denoted by Stokes V. Conventionally speaking, although the main focus has been to develop algorithms for total intensity imaging, for the purpose of imaging these parameters, referred to as polarimetric imaging, the same techniques as developed for Stokes I imaging have been applied for imaging each of the Stokes parameters.

In practice, rather than providing an accurate representation of the true sky distribution, the RI measurements are often corrupted both by the atmospheric effects encountered by the radio emissions en route to the receiving antennas and by the antenna-based errors. These measurement-contaminating effects are represented in the form of Jones matrices (Smirnov 2011). Only in the cases when these effects are absent or known accurately, image recovery amounts to use of any RI imaging algorithm. In practical cases, where these effects are unknown, these need to be estimated along with imaging to achieve good quality reconstruction. Calibration is this process of estimation of the Jones matrices. Moreover, these time-variable calibration terms in general also exhibit direction dependency, thereby encompassing both directionindependent and -dependent effects (DIEs and DDEs, respectively).

For instance, while the antenna-based gains denote the DIEs, the antenna primary beam pattern, atmospheric phase delays etc. exhibit spatial variation within the field of view and thus constitute DDEs. Historically, only the DIE calibration has been performed (Thompson et al. 2001). However, the calibration of DDEs has become a crucial issue especially for the new generation radio telescopes including the LOw Frequency ARray (LOFAR) (http://www.lofar.org) and the upcoming Square Kilometre Array (SKA) (https://www.skatelescope.org), aiming to produce sky images at unprecedented resolution with high sensitivity. Without efficient DDE calibration, these telescopes cannot be exploited to their fullest potential. The incorporation of the DDEs in the calibration process is also essential to produce high quality images without limiting their dynamic ranges (Bhatnagar et al. 2013). Thus, the resultant problem of image recovery from the acquired corrupted data consists in estimating both the sought images and the DDE calibration terms. We now review the imaging and calibration techniques in RI, which can then be combined to solve the overall problem.

In the context of RI imaging techniques, the standard CLEAN algorithm implements a non-linear, greedy approach to perform deconvolution in an iterative manner (Högbom 1974). Working pixelwise, CLEAN implicitly assumes sparsity of the sought image, removing at each iteration, a fraction of the maximum intensity pixel convolved with the dirty beam from the computed residual image. Over the past years, many variants of this celebrated algorithm have been developed, for instance, its multi-scale version (Cornwell 2008). In addition to these CLEAN based approaches, recent developments in the field of compressive sensing (CS) applied for astronomical imaging have given birth to many imaging algorithms, particularly for radio interferometry (Wiaux et al. 2009; Carrillo et al. 2012; Garsden et al. 2015; Onose et al. 2016). Leveraging optimization framework, these methods aim to solve the underlying image recovery problem by enforcing sparsity of the image of interest in some suitable domain. As a matter of fact, these techniques have shown the potential to surpass the image reconstruction quality obtained by CLEAN based approaches (Carrillo et al. 2014; Onose et al. 2016; Pratley & Johnston-Hollitt 2016; Onose et al. 2017; Dabbech et al. 2018). While the aforementioned methods have been developed mainly for Stokes I imaging, as previously mentioned these can be extended to polarimetric imaging by following the same imaging approach for all the Stokes parameters. In the context of sparsity regularized approaches, one such extension for polarimetric imaging has been presented in Akiyama et al. (2017), promoting sparsity of each of the Stokes parameters using ℓ_1 norm combined with total variation (TV) regularization (Rudin et al. 1992). A point worth noting here is that the above mentioned approaches, whether CLEAN and its variants or the sparsity regularized method, solve for the Stokes parameters totally independently. However, these images are physically linked via polarization constraint, that is the polarized intensity cannot be greater than the total intensity. Within the CS framework, this constraint has been exploited by a recently proposed approach, namely Polarized SARA, estimating jointly the Stokes parameters (Birdi et al. 2018a,b). This approach has been shown to provide better reconstruction quality in comparison with the case when the constraint is not accounted for.

Concerning the calibration problem, it can be formulated as a non-linear least squares problem and to solve it, most of the existing solvers employ gradient-based techniques like the Levenberg-Marquardt (LM) algorithm (Moré 1978). But it suffers from its high computational cost. To overcome this issue, an efficient and fast solver named StEFCal has been proposed in Salvini & Wijnholds

(2014), which relies on an alternating direction implicit method. However, it solves only for the DIEs. In the context of DDE calibration, several methods have been proposed in the last years. In particular, the Source Peeling and Atmospheric Modeling (SPAM) method proposed in Intema et al. (2009) models the ionosphere as a phase screen, whose parameters are estimated by fitting the calibrated phases obtained from peeling (Noordam 2004) with the modelled phases, using the LM algorithm. Subsequently, this model can be used for the prediction of correction phases in any arbitrary desired direction within the field of view. To apply these corrections, SPAM relies on a facet-based approach, applying the DDE calibration solution for the facet centre, defined by a bright source or approximate centre of a cluster of closely located bright sources, to the whole facet. This scheme works under the assumption that the DDEs show smooth variation over the field of view, thereby having (approximately) constant value across a given facet. This approach has analogies with the facet calibration which consists in partitioning the sky into a number of facets and solving for the DDE solution of the facet centre using a self-calibration (SelfCal) loop, which in turn is applied to the whole facet (Van Weeren et al. 2016). Despite some similarities, these two approaches are different in the way they obtain the calibration solutions, i.e. unlike SPAM, facet calibration does not make use of global phase screen model. In contrast to these approaches solving for the calibration solutions source by source (or in terms of facets), SAGE calibration algorithm solves for the calibration solutions for all sources at once, employing an extension of the Expectation Maximization algorithm (Yatawatta et al. 2009; Kazemi et al. 2011). It therefore provides increased convergence speed in comparison with the traditional LM solvers. It is to be noted that within a given iteration, SAGE algorithm requires to define partitioning over the unknown Jones parameter space depending on the source characteristics. Another approach for DDE calibration arises from the convolutional nature of the DDEs. More precisely, the multiplication of the DDEs in the image domain can be equivalently seen as their convolution in the Fourier domain. The A-Projection algorithm makes use of the latter characteristic and offers a way to correct for those known DDEs whose associated Jones matrix are (approximately) unitary, by embedding them into the forward (degridding) and backward (gridding) operators (Bhatnagar et al. 2013). While this approach was only developed for the correction of known DDEs, it has been incorporated in the Pointing SelfCal algorithm to solve for antenna pointing errors (Bhatnagar & Cornwell 2017).

In order to estimate the sky model while having imperfect knowledge of calibration terms, any of the above mentioned calibration algorithms is generally combined with CLEAN (or its variants) for image recovery step. As such, the global algorithm fails to have any convergence guarantees. The first step in the direction of addressing this issue has been made by the joint calibration and imaging algorithm proposed in Repetti et al. (2017). This algorithm leverages non-convex optimization techniques to present a global algorithm which alternates between the estimation of the image of interest and the DDEs, while benefitting from the convergence guarantees. In addition to this, unlike the approaches requiring sky partitioning, which may not always be best achieved in an automatic manner, this technique works on the whole image all at once with minimal user intervention. However, this approach has been developed only for Stokes I imaging and calibration, without dealing with the full polarization model. As a matter of fact, in the case of full Stokes imaging and calibration, even the global algorithm consisting of combination of CLEAN based imaging and any of the previously mentioned calibration techniques, do not adopt any polarimetric imaging specific approach. These factors raise the scope for the development of a globally convergent algorithm which not only incorporates full polarization model, but also uses advanced approaches specific for full Stokes imaging to produce good quality images.

The work presented here lies within this scope and we propose Polca SARA - a full polarization, calibration and sparse imaging algorithm with proven convergence guarantees. In particular, we build our method on the approach developed in Repetti et al. (2017), and later on extended in Repetti & Wiaux (2017) and Thouvenin et al. (2018), assuming spatial and temporal smoothness of the DDEs. On the one hand, we generalize this approach to the full polarization model, developing an algorithm alternating between the estimation of the DDEs and the Stokes images. On the other hand, thanks to the underlying non-convex optimization technique, the proposed approach can deal with sophisticated priors suited to the images under consideration as well as to the DDEs. Leveraging this flexibility, we adapt Polarized SARA method, specifically designed for Stokes imaging, to be used for imaging step in the considered case.

The remainder of the article is organised as follows. Section 2 describes the full polarization model encountered in RI. The proposed approach to solve the underlying joint calibration and imaging problem is presented briefly in Section 3. A more formal and mathematical presentation of the proposed approach is provided in later sections. Particularly, the imaging problem to be solved is posed in Section 4, followed by the introduction of the calibration problem in Section 5. A blend of these problems to give the resultant joint calibration and imaging problem and the proposed algorithm to solve it are detailed in Section 6. To assess the performance of the proposed algorithm, we perform various tests which are presented in Section 7. We give the concluding remarks and discuss about possible directions of future work in Section 8.

2 FULL POLARIZATION OBSERVATION MODEL

The radio interferometric measurements, termed as visibilities, are acquired by antenna pairs. Within n_a number of total antennas, consider an antenna pair $(\alpha,\beta) \in \{1,\ldots,n_a\}^2$, at time instant $t \in \{1,\ldots,T\}$, associated with the baseline components $(u_{t,\alpha,\beta},v_{t,\alpha,\beta},w_{t,\alpha,\beta})$ expressed in observation wavelength units. Here, $u_{t,\alpha,\beta} = (u_{t,\alpha,\beta},v_{t,\alpha,\beta})$ represents the components in the plane normal to the target source direction, while $w_{t,\alpha,\beta}$ is the component in the source direction. Furthermore, let $\overline{S} = \begin{bmatrix} I & Q \\ U & V \end{bmatrix}$ encompasses the Stokes parameters of the target sky area, described in terms of the direction cosines (l,m,n) with I = (l,m) lying in the tangential plane to the celestial sphere and $n(I) = \sqrt{1-l^2-m^2}$ lying in the line of sight. Then, the measurements $\mathbf{Y}_{t,\alpha,\beta,\gamma} \in \mathbb{C}^{2\times 2}$ made by the antenna pair (α,β) at time instant t and observation frequency v is given by the radio interferometric measurement equation (RIME) (Hamaker et al. 1996; Smirnov 2011) as follows

$$\mathbf{Y}_{t,\alpha,\beta,\nu} = \int \overline{\mathbf{D}}_{t,\alpha,\nu}(l) \, \mathcal{L}(\overline{\mathbf{S}}_{t,\nu}(l)) \, \overline{\mathbf{D}}_{t,\beta,\nu}^{\dagger}(l) \mathrm{e}^{-2\mathrm{i}\pi \mathbf{u}_{t,\alpha,\beta} \cdot l} \, d^2 \mathbf{l}, \quad (1)$$

where $(\cdot)^{\dagger}$ denotes the Hermitian conjugate of its argument and $\mathcal{L}(\overline{S}_{t,\nu}(I)) = \overline{B}_{t,\nu}(I)$ is the brightness matrix formed by the application of the linear operator \mathcal{L} on the components of the Stokes matrix \overline{S} . For instance, considering linear feeds, the brightness matrix is given by $\overline{B}_{t,\nu}(I) = \begin{bmatrix} I+Q & U+\mathrm{i}V \\ U-\mathrm{i}V & I-Q \end{bmatrix}_{t,\nu}(I)$, whereas for circular trix is given by $\overline{B}_{t,\nu}(I) = \begin{bmatrix} I+Q & U+\mathrm{i}V \\ U-\mathrm{i}V & I-Q \end{bmatrix}_{t,\nu}(I)$, whereas for circular trix is given by $\overline{B}_{t,\nu}(I) = \begin{bmatrix} I+Q & U+\mathrm{i}V \\ U-\mathrm{i}V & I-Q \end{bmatrix}_{t,\nu}(I)$

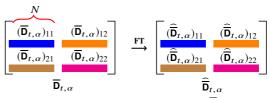


Figure 1. Schematic representation of a $2 \times 2N$ matrix, $\overline{\mathbf{D}}_{t,\alpha}$ (left) that can be seen as a 2×2 block matrix, with each block $(\overline{\mathbf{D}}_{t,\alpha})_{ii'}$, $\{i,i'\} \in \{1,2\}^2$ containing a vectorized image of dimension N. The Fourier transformation (FT) of this matrix computes 2D Fourier transform of each of the block images. For each block in matrix $\overline{\mathbf{D}}_{t,\alpha}$ on left, its corresponding FT is shown by same colored vectors in $\widehat{\overline{\mathbf{D}}}_{t,\alpha}$ on right.

lar feeds, it reads as
$$\overline{B}_{t,\nu}(I) = \begin{bmatrix} I+V & U+iQ \\ U-iQ & I-V \end{bmatrix}_{t,\nu}(I)$$
. Furthermore, in equation (1) for every antenna $\alpha \in \{1, \dots, \nu\}$ the direction.

in equation (1), for every antenna $\alpha \in \{1, \ldots, n_a\}$, the direction-independent and -dependent effects (DIEs and DDEs, respectively) at time t and frequency ν are encoded in the Jones matrix $\overline{D}_{t,\alpha,\nu}$. It is worth mentioning that equation (1) presents a general case where the Stokes images and the DIEs/DDEs can vary with respect to the observation frequency and time. However, in the current work, we consider the Stokes images without having any time or frequency dependence. In addition, for the DIEs/DDEs, we deal with their temporal dependency considered at a single observation frequency. Hence, hereafter, we take $\overline{S}_{t,\nu}(l) = \overline{S}(l)$ and drop the frequency index from the other variables.

To recover the Stokes parameters from the given measurements, we formulate the inverse problem (1) in the discrete domain. It amounts to sampling the continuous variables such that the field of view is discretized into a grid, with the vectorized form of this grid being represented by the index $n \in \{-N/2, ..., N/2 - 1\}$. Within this representation, we thus have the Stokes matrix $\overline{\mathbf{S}} \in$ $\mathbb{R}^{2\times 2N}$, and for each antenna α and time instant t, the DDEs $\overline{\mathbf{D}}_{t,\alpha} \in \mathbb{C}^{2\times 2N}$. As illustrated in Fig. 1, these matrices can also be seen as 2×2 block matrices¹, whose each block is a row vector of dimension N. In particular, for the Stokes matrix $\overline{S} = \begin{bmatrix} \overline{s}_1 & \overline{s}_2 \\ \overline{s}_3 & \overline{s}_4 \end{bmatrix}$, the elements $\overline{s}_1, \overline{s}_2, \overline{s}_3$ and \overline{s}_4 , each of size N, are respectively the discretizations of the Stokes parameters I, Q, U and V. On the other hand, for the DDEs, for each index n, we have the so called 2×2 Jones matrix $\overline{\mathbf{D}}_{t,\alpha}(n) \in \mathbb{C}^{2\times 2}$. The diagonal elements of this matrix correspond to the antenna's complex voltage pattern, whereas the off-diagonal elements account for the polarization leakage terms arising from instrumental leakage, transmission effects, etc (Bhatnagar et al. 2013). Moreover, if for every $(i,i') \in \{1,2\}^2$, $[\overline{\mathbf{D}}_{t,\alpha}]_{ii'} = \delta_{t,\alpha} \mathbf{1}_N$ with $\delta_{t,\alpha} \in \mathbb{C}$ and $\mathbf{1}_N$ being N dimensional unitary row vector, then $\overline{\mathbf{D}}_{t,\alpha}$ reduces to a DIE.

Following the introduced notations, each component of the visibility matrix $\mathbf{Y}_{t,\alpha,\beta}$, indexed by $[\cdot]_{ii'}$ with $(i,i') \in \{1,2\}^2$, at discrete spatial frequency $k_{t,\alpha,\beta}$ can be represented as

$$[\mathsf{Y}_{t,\alpha,\beta}]_{ii'} = \sum_{j,j'=1}^{2} \sum_{n=-N/2}^{N/2-1} \left[\overline{\mathbf{D}}_{t,\alpha}(n) \right]_{ij} \left[\mathcal{L}(\overline{\mathbf{S}}(n)) \right]_{jj'} \left[\overline{\mathbf{D}}_{t,\beta}^{*}(n) \right]_{j'i'}$$

$$e^{-2\mathrm{i}\pi(k_{t,\alpha,\beta})\frac{n}{N}} + [\Omega_{t,\alpha,\beta}]_{ii'}, \quad (2)$$

where $(\cdot)^*$ stands for complex conjugation of its argument and the

¹ For any such matrix $\overline{\mathbf{S}}$, $\overline{\mathbf{S}}(n)$ denotes the 2×2 matrix consisting of the n^{th} elements of each of the blocks in the parent block matrix $\overline{\mathbf{S}}$. Furthermore, for $(i,i') \in \{1,2\}^2$, $[\overline{\mathbf{S}}(n)]_{i\,i'}$ refers to the n^{th} element of the row vector block contained in the i^{th} row and i'^{th} column of the argument block matrix.

visibilities are corrupted by an additive Gaussian noise $\Omega_{t,\alpha,\beta} \in \mathbb{C}^{2\times 2}$. For $n \in \{-N/2,\ldots,N/2-1\}$, as for the continuous version (1), $\mathcal{L}(\overline{\mathbf{S}}(n)) = \overline{\mathbf{B}}(n) \in \mathbb{C}^{2\times 2}$ is the brightness matrix.

3 A BRIEF OVERVIEW OF THE PROPOSED METHOD

While a more formal description of the proposed approach is presented in the later sections, this section is designed to give an intuitive explanation of the approach. Particularly, the description in this section is targeted to facilitate readers to understand the results and conclusion sections without delving into the mathematical details provided in Sections 4-6.

Given the measurements of the form (2), the aim is to recover the images of interest. The problem of image recovery becomes more challenging when the DDEs $\overline{\mathbf{D}}$ are unknown and need to be estimated along with the images. The current work is devoted to this more challenging yet practical case, for which we propose a joint calibration and imaging algorithm to solve the underlying problem. The resultant approach is dubbed as Polca SARA.

In order to calibrate for the DDEs, we consider their smooth variation both across the field of view and in time, implying that they are spatially and temporally band-limited (Repetti et al. 2017; Thouvenin et al. 2018). Exploiting this characteristic, we instead represent the DDEs $\overline{\mathbf{D}}$ by their compact-support kernels $\overline{\mathbf{U}}$ in the spatial and temporal Fourier domain. Then, only these non-zero Fourier coefficients need to be estimated. It reduces significantly the dimension of the underlying problem. Furthermore, we note that while equation (2) is linear with respect to $\overline{\mathbf{S}}$ (when the DDEs are kept fixed), it is non-linear with respect to the DDEs. To counteract this issue, we draw on the notion of bi-linearity proposed in Salvini & Wijnholds (2014) and Repetti et al. (2017) in the context of RI calibration. It consists in introducing $\mathbf{U}_1 = \mathbf{U}_2 = \mathbf{U}$ such that problem (2) becomes bi-linear with respect to the DDEs. The resultant problem can then be written as

$$\mathbf{Y} = \Phi(\mathbf{S}, \mathbf{U}_1, \mathbf{U}_2) + \mathbf{\Omega},\tag{3}$$

where Φ is the tri-linear measurement operator, mapping the images of interest coupled with the DDE Fourier kernels \mathbf{U}_1 and \mathbf{U}_2 to the acquired visibilities \mathbf{Y} . On the imaging front, we further consider that calibration transfer has been performed and thus, the Jones matrices in equation (2) can be taken as identity as a first instance, without any directional dependency. Doing so, we first solve the problem to obtain an estimation \mathbf{S}' of the unknown images. Nevertheless, the used measurement model being devoid of DDEs, the estimated images are likely to contain artefacts. We thus instead keep a thresholded version \mathbf{S}_0 of \mathbf{S}' , considering that $\mathbf{S} = \mathbf{S}_0 + \mathbf{E}$, where \mathbf{E} is unknown and needs to be estimated.

In view of the discussion above, we propose to estimate the variables of interest by solving the following problem:

$$\underset{\mathbf{E},\mathbf{U}_1,\mathbf{U}_2}{\text{minimize}} \ h(\mathbf{E},\mathbf{U}_1,\mathbf{U}_2) + r(\mathbf{E}) + p(\mathbf{U}_1,\mathbf{U}_2), \tag{4}$$

where the data-fidelity term h is given by a least-squares criterion (equations (22)-(24)) and the regularization terms are defined both for the images (r in equation (25)) and the DDEs (p in equation (20)).

For the images of interest, we exploit the compressive sensing framework (Candès et al. 2006) and consider ℓ_1 regularization term to promote sparsity of the sought images in a suitable dictionary. More precisely, inspired by the good performance of the sparsity averaging reweighted analysis (SARA) prior for Stokes I imaging in Carrillo et al. (2013, 2014) and for full Stokes imaging in

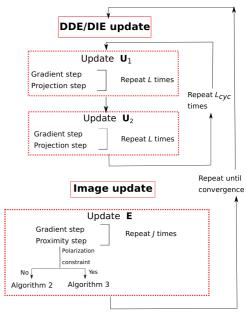


Figure 2. Block diagram illustrating the proposed Polca SARA method that alternates between the DDEs/DIEs update and the Stokes images update until convergence. For DDEs/DIEs update, $L \in \mathbb{N}$ forward-backward iterations are performed for the estimation of each of the matrices \mathbf{U}_1 and \mathbf{U}_2 . Similarly, for the image update, $J \in \mathbb{N}$ forward-backward steps are executed, wherein depending on the chosen regularization, the proximity step is either evaluated using Algorithm 2 or Algorithm 3.

Birdi et al. (2018b), we employ this prior consisting in promoting average sparsity over a concatenation of Dirac basis and first eight Daubechies wavelet bases. It is coupled with a reweighting scheme, solving sequentially the ℓ_1 regularized problems (Candès et al. 2008). Although the current work does not deal with the reweighting procedure, it can be easily incorporated. In practice, other sophisticated priors injecting additional information about the images of interest can also be taken into account in the regularization term, eg. polarization constraint in Birdi et al. (2018b). In the current work, we consider both the cases without and with the enforcement of the polarization constraint. For the DDEs, the regularization term p is defined to impose values of the Jones matrices' elements to lie within specified bounds as well as to control the distance between \mathbf{U}_1 and \mathbf{U}_2 which ideally should be equal. Additional details for the image (resp. DDEs) estimation are provided in Section 4 (resp. Section 5).

The resultant problem (4) is solved using a block-coordinate forward-backward approach (Chouzenoux et al. 2016). As the name suggests, this iterative procedure relies on a forwardbackward (FB) scheme to update each of the blocks/variables of interest (U_1, U_2, E) . The FB scheme consists in treating the smooth term with a gradient step and the non-smooth term via a proximity step. The global method is detailed in Section 6. For a better understanding, the overall algorithmic structure is depicted in Fig. 2. It has similarities with the traditional self-calibration method in the sense that it alternates between the imaging and calibration steps. Yet the two approaches are different in terms of their underlying principle and implementation. First, the imaging and calibration steps in our approach use the same optimization toolbox and hence, the global algorithm benefits from convergence guarantees. Second, to ensure this global convergence, each of these updates is obtained by performing only a finite number of iterations and not until convergence. Moreover, these updates must follow a cyclic rule, that is each variable must be updated at least once every given number of iterations. To this end, L_{cyc} global iterations are performed to update the DDEs ($\mathbf{U}_1, \mathbf{U}_2$), alternating between L FB iterations on \mathbf{U}_1 and L FB iterations on \mathbf{U}_2 . After L_{cyc} global iterations, the Stokes images \mathbf{E} are updated by executing J FB iterations.

The results obtained by implementing Polca SARA are presented in Section 7. Particularly, we compare between the following cases: on the calibration front- considering normalized DIEs (i.e. identity Jones matrices), DIE calibration, Jones matrices DDE calibration without considering off-diagonal terms, and full Jones matrices DDE calibration; and on the imaging front- in the absence and presence of the polarization constraint. While the calibration cases are chosen to judge the importance of calibrating for the off-diagonal terms as well as for the DDEs, the two imaging cases are to test the efficiency of the polarization constraint in the context of joint calibration and imaging. Furthermore, concerning these imaging cases, when the polarization constraint is not imposed, the proximity step for the update of **E** is evaluated using Algorithm 2, whereas in the case of enforcing this constraint, Algorithm 3 is employed to compute the corresponding proximity step.

A detailed mathematical formulation of the proposed approach is presented in the next sections, starting with the description of the imaging problem followed by the calibration problem and finally combining the two. On the other hand, with the intuitive description provided in this section at disposal, the reader might jump directly to the simulations and results section (Section 7) averting the mathematical details.

4 IMAGING PROBLEM

The case of having either pre-calibrated data or knowledge of DIEs/DDEs beforehand amounts to solving the problem only for image recovery. The RI literature is brimming with such methods aiming to estimate the images of interest from the underlying imaging problem. Nevertheless, most of these methods were initially proposed mainly for Stokes I imaging, including CLEAN based methods (Högbom 1974; Bhatnagar & Cornwell 2004; Cornwell 2008) and CS techniques (Wiaux et al. 2009; Carrillo et al. 2014; Garsden et al. 2015; Onose et al. 2016, 2017). In order to apply these methods for full polarization imaging, these can be extended to the estimation of each of the Stokes parameters totally independently using the same techniques as developed for Stokes I imaging. Such an extension of CS based Stokes I imaging to full polarization imaging has been presented in Akiyama et al. (2017), solving independently for the sought images. A more evolved approach aiming to jointly estimate the Stokes parameters has been very recently proposed in Birdi et al. (2018a,b). In this section, we aim to summarize this work which was proposed in the context of DIEs, and generalize it to the DDEs case, which will be useful for the understanding of the joint imaging and calibration method developed in Section 6.

In this context, a reformulation of equation (2) is more commonly used where the measurements made by an antenna pair (α, β) at time instant t are represented by a vector $y_{t,\alpha,\beta} \in \mathbb{C}^4$, whose each element $q \in \{1, \ldots, 4\}$ is given by

$$[y_{t,\alpha,\beta}]_q = \sum_{z=1}^4 \sum_{n=-N/2}^{N/2-1} [\overline{\mathbf{M}}_{t,\alpha,\beta}(n)]_{qz} [\widetilde{\mathcal{L}}(\overline{\mathbf{S}}(n))]_z$$

$$e^{-2i\pi(k_{t,\alpha,\beta})\frac{n}{N}} + [\omega_{t,\alpha,\beta}]_q. \quad (5)$$

with $\omega_{t,\alpha,\beta} \in \mathbb{C}^4$ being a realization of additive Gaussian noise,

and $\widetilde{\mathcal{L}}(\overline{\mathbf{S}}(n))$ operates on the Stokes matrix to produce the brightness matrix $\overline{\mathbf{B}}(n)$, followed by its vectorization. Furthermore, $\overline{\mathbf{M}}_{t,\alpha,\beta} = (\overline{\mathbf{D}}_{t,\alpha} \otimes \overline{\mathbf{D}}_{t,\beta}^{\dagger}) \in \mathbb{C}^{4\times 4N}$ is the Mueller matrix formed by the outer product of the Jones matrices for antennas α and β at time instant t, and in turn, this matrix can be seen as a 4×4 block matrix with each block given by N dimensional row vector. Let us note that, for the Fourier transformed Mueller matrix, denoted by $\widehat{\overline{\mathbf{M}}}_{t,\alpha,\beta}$, each block is obtained by the complex convolution of the Fourier transforms of the corresponding Jones matrices (Bhatnagar et al. 2013).

Using equation (5), the overall measurement model can be written as

$$\mathbf{y} = \Phi(\overline{\mathbf{S}}) + \omega,\tag{6}$$

where $y \in \mathbb{C}^{4M}$ is the resultant measurement vector corrupted by the noise vector $\omega \in \mathbb{C}^{4M}$. The measurement operator Φ mapping the images of interest to the acquired visibilities is modelled as

$$\Phi(\overline{S}) = GF\widetilde{Z}\widetilde{\mathcal{L}}(\overline{S}), \tag{7}$$

that can be explained as follows. The operation $\widetilde{\mathcal{L}}(\overline{S})$ generates the brightness vector $\boldsymbol{b} \in \mathbb{C}^{4N}$ that needs to be Fourier transformed at the sampled spatial frequencies. To evaluate these Fourier transforms in a computationally efficient manner, we make use of the non-uniform fast Fourier Transform that relies on interpolation of the Fourier coefficients from discrete to continuous domain (Fessler & Sutton 2003). In this context, the zero-padding matrix $\tilde{\mathbf{Z}} \in \mathbb{C}^{y4N\times 4N}$ oversamples the images in \boldsymbol{b} , by factor γ in each dimension and also accounts for the scale factors to pre-compensate for the interpolation convolution kernels. A fast Fourier transform operator $\mathbf{F} \in \mathbb{C}^{\gamma 4N \times \gamma 4N}$ is then applied to compute the 2D Fourier transform of each of these oversampled images. To interpolate these discrete Fourier coefficients to the continuous frequency points, the compact support convolution kernels (the so called degridding kernels) are embedded in the matrix $\mathbf{G} \in \mathbb{C}^{4M \times \gamma 4N}$. More precisely, the matrix \mathbf{G} is modelled to take into account the degridding kernels as well as the components of the Fourier transformed Mueller matrix. In particular, each of the four rows of **G** associated with a frequency $k_{t,\alpha,\beta}$ contains the convolution of the corresponding elements of $\overline{\mathbf{M}}_{t,\alpha,\beta}$ centred on $k_{t,\alpha,\beta}$ with the degridding kernel. Then, the application of this matrix on the Fourier transformed over-sampled images produce the measurements. When working in the absence of the DDEs, a direct correspondence of the measurement operator in equation (7) can be noticed with the measurement operator used for Stokes I imaging (Onose et al. 2017; Pratley et al. 2017) and more recently, for full

It is to be mentioned here that most of the imaging techniques in RI literature rely on taking the Jones matrices and hence Mueller matrix as identity in equation (5). In other words, they work under the assumption of absence of any calibration errors or availability of pre-calibrated data. However, in presence of calibration errors, the imaging techniques should account for these terms within their models. Working in this direction, we can generalize the model proposed in Birdi et al. (2018b) and employ it to solve inverse problem (6). The corresponding imaging strategy leverages the CS theory. In particular, CS based methods offer a way from optimization point of view to solve the underlying inverse problem. The main idea is to obtain the estimation of the sought images by solving a minimization problem of the form (Birdi et al. 2018a,b)

Stokes imaging (Birdi et al. 2018b).

minimize
$$r(\mathbf{S})$$
 subject to $\|\Phi(\mathbf{S}) - \mathbf{y}\|_2 \le \epsilon$, (8)

where ϵ is related to the norm of the additive noise, $\|\cdot\|_2$ denotes the ℓ_2 norm of its argument vector. The term $r(\mathbf{S})$ is the regularization term, injecting prior information in the reconstruction process, whereas the data fidelity is ensured by the constraint $\|\Phi(\mathbf{S}) - y\|_2 \le \epsilon$, implying that the residual is situated within an ℓ_2 ball whose radius is determined by ϵ . In the current work, due to technical assumptions related to the proposed joint imaging and calibration algorithm (Chouzenoux et al. 2016), we propose to solve an unconstrained version of problem (8), given by

minimize
$$\bar{h}(\mathbf{S}) + r(\mathbf{S}),$$
 (9)

where $\bar{h}(\mathbf{S})$ is the data fidelity term. Using this formulation and assuming that the additive noise is i.i.d. Gaussian, $\bar{h}(\mathbf{S})$ is given by a least squares criterion

$$(\forall \mathbf{S} \in \mathbb{R}^{2 \times 2N}) \ \bar{h}(\mathbf{S}) = \frac{1}{2} \|\Phi(\mathbf{S}) - \mathbf{y}\|_2^2. \tag{10}$$

This formulation has been introduced in Repetti et al. (2017) in the particular case of joint Stokes I imaging and calibration. In order to incorporate various polarimetric imaging specific prior informations in the regularization function, Birdi et al. (2018b) proposed to use a hybrid regularization term of the form

$$r(\mathbf{S}) = g(\mathbf{S}) + r'(\mathbf{S}),\tag{11}$$

where the function $g: \mathbb{R}^{2\times 2N} \to]-\infty, +\infty]$ imposes sparsity of the underlying Stokes images and the function $r': \mathbb{R}^{2\times 2N} \to]-\infty, +\infty]$ constrains the domains of the argument images as per some physical constraints. In particular, the first term is inspired by the CS framework which proposes to exploit the sparsity of the target images in some dictionary $\Psi \in \mathbb{C}^{\hat{N} \times J}$ (Candès et al. 2006; Donoho 2006). The choice of this dictionary is dependent on the images under scrutiny. To give an intuitive idea, images which are already sparse, for instance point sources images, are well represented by dirac basis taking Ψ to be identity. Otherwise, sparsity can be imposed in some other domain, such as TV regularization for piece-wise constant images (Rudin et al. 1992; Wiaux et al. 2010; Akiyama et al. 2017), wavelet basis for smooth images (Mallat 2009), etc. In particular for astronomical images, a collection of wavelet bases is shown to be a good candidate for sparsifying dictionary Ψ (Carrillo et al. 2012, 2014), both for Stokes I imaging (Onose et al. 2016, 2017) and specifically for polarization imaging in Birdi et al. (2018a,b). Formally, for any matrix $\tilde{\mathbf{S}} \in \mathbb{R}^{N \times 4}$, the sparse representation of the sought images in a chosen dictionary Ψ is given by Ψ^{\dagger} **S**. The regularization function is then defined to impose sparsity of this term. This can be achieved by using the ℓ_0 pseudo-norm, counting the number of non-zero components of its argument (Donoho et al. 1995). A more common approach is to use ℓ_1 norm which overcomes the problem of non-convexity of the ℓ_0 norm (Chen et al. 2001). An even better approximation of the ℓ_0 norm is provided by the (re)weighted ℓ_1 norm, which tends to diminish the magnitude dependency of the ℓ_1 norm (Candès et al. 2008; Carrillo et al. 2012). The (re)weighted ℓ_1 norm is given by

$$g(\mathbf{S}) = \|\mathbf{\Psi}^{\dagger} \mathcal{R}(\mathbf{S})\|_{\mathbf{W},1} = \sum_{i=1}^{4} \sum_{j=1}^{J} \mathbf{W}_{j,i} |[\mathbf{\Psi}^{\dagger} \mathcal{R}(\mathbf{S})]_{j,i}|,$$
(12)

where the subscripts j and i in the notation $[\cdot]_{j,i}$ stand respectively for the row and column indices of the argument matrix, and $\mathcal{R}: \mathbb{R}^{2\times 2N} \to \mathbb{R}^{N\times 4}$ is the operator consisting in placing the four Stokes images contained in the matrix \mathbf{S} in four columns, whereas its adjoint $\mathcal{R}^{\dagger}: \mathbb{R}^{N\times 4} \to \mathbb{R}^{2\times 2N}$ do the contrary, i.e. storing the four images in the rows of a 2×2 block matrix. In equation (12),

 $\mathbf{W} \in \mathbb{R}^{J \times 4}_+$ is the weighting matrix. In the case when this matrix is chosen to be identity, the ℓ_1 regularization term is obtained. Another useful piece of information which can be used to regularize the problem is the polarization constraint to be satisfied by the recovered Stokes images (Birdi et al. 2018b). This constraint comes from a physical point of view, that the polarized intensity should be smaller or at most equal to the total intensity and mathematically, can be described by the following set:

$$\mathbb{P} = \{ \widetilde{\mathbf{S}} = \mathcal{R}(\mathbf{S}) \in \mathbb{R}^{N \times 4} \, \big| \, (\forall n \in \{1, \dots, N\}) \\ - \widetilde{\mathbf{S}}_{n,1} + \|\widetilde{\mathbf{S}}_{n,2:4}\|_2 \leqslant 0 \}. \quad (13)$$

As can be noticed, this constraint also imposes implicitly the positivity of the total intensity image (Stokes I). The enforcement of this constraint amounts to incorporation of the indicator function of the set \mathbb{P} in the regularization function, i.e.

$$r'(\mathbf{S}) = \iota_{\mathbb{P}}(\mathbf{S}). \tag{14}$$

The indicator function of any such set \mathbb{P} is defined as

$$\iota_{\mathbb{P}}(\mathbf{S}) = \begin{cases} 0, & \text{if } \mathbf{S} \in \mathbb{P}, \\ +\infty, & \text{otherwise.} \end{cases}$$
 (15)

5 CALIBRATION PROBLEM

In practice, the DDEs are often unknown and need to be estimated. In this section, we formulate the calibration problem to be solved. We assume that the DDEs exhibit a smooth variation both across the field of view and in time. It implies that DDEs are band-limited spatially as well as temporally. This is enforced by considering compact-support kernels of the DDEs in both spatial (Repetti & Wiaux 2017; Repetti et al. 2017) and temporal Fourier domains (Thouvenin et al. 2018). More specifically, for each antenna $\alpha \in \{1, \ldots, n_a\}$, the DDEs $(\overline{\mathbf{D}}_{t,\alpha})_{1 \leqslant t \leqslant T}$ are represented by the Fourier kernels $\overline{\mathbf{U}}_{\alpha} \in \mathbb{C}^{2 \times 2K \times P}$, where K and P are the support sizes in spatial and temporal Fourier domain, respectively, with $K \ll N, P \ll T$. Then, the task is to estimate only non-zero Fourier coefficients of the DDEs, thereby reducing the dimension of the underlying problem significantly. Analogous to the calibration inverse problem in Repetti et al. (2017), problem (2) can be reformulated as

$$\mathbf{Y}_{t,\alpha,\beta} = \mathcal{D}_{t}(\overline{\mathbf{U}}_{\alpha}) X_{t,\alpha,\beta} \left(\mathbf{F} \widetilde{\mathbf{Z}} \widetilde{\mathcal{L}}(\overline{\mathbf{S}}) \right) \mathcal{D}_{t}'(\overline{\mathbf{U}}_{\beta}) + \Omega_{t,\alpha,\beta}, \tag{16}$$

where $\mathcal{D}_t: \mathbb{C}^{2\times 2K\times P} \to \mathbb{C}^{2\times 2N}$ is the operator acting on $\overline{\mathbf{U}}_{\alpha}$ to give a sparse matrix $\mathcal{D}_t(\overline{\mathbf{U}}_{\alpha})$ containing the compact support kernels in $\widehat{\overline{\mathbf{D}}}_{t,\alpha}$, flipped and centred in zero spatial frequency. Similarly, the operator $\mathcal{D}_t': \mathbb{C}^{2\times 2K\times P} \to \mathbb{C}^{2N\times 2}$ is defined such that $\mathcal{D}_t'(\overline{\mathbf{U}}_{\beta})$ is a sparse matrix consisting of the compact support kernels in $\widehat{\overline{\mathbf{D}}}_{t,\beta}^{\dagger}$ centred in zero spatial frequency. Finally, $X_{t,\alpha,\beta}(\mathbf{F}\widetilde{\mathbf{Z}}\widetilde{\mathcal{L}}(\overline{\mathbf{S}})) \in \mathbb{C}^{2N\times 2N}$ is a 2×2 block matrix, with each block of size $N\times N$. Each row/ column of such a block consists of a shifted version of the Fourier transform of the corresponding image in reshaped brightness vector $\widetilde{\mathcal{L}}(\overline{\mathbf{S}})$, mimicking the convolution operation. Moreover, to account for the continuous sampled frequencies, these Fourier transforms are convolved with the degridding kernels centred at the associated frequency $k_{t,\alpha,\beta}$.

It can be observed that problem (16) is non-linear with respect to the compact-support kernels $(\overline{\mathbf{U}}_{\alpha})_{1 \leq \alpha \leq n_a}$. Following the approach proposed in Salvini & Wijnholds (2014) and Repetti et al. (2017), we linearize it by introducing the matrices $\overline{\mathbf{U}}_{\alpha,1}$ and $\overline{\mathbf{U}}_{\alpha,2}$

such that $\overline{\mathbf{U}}_{\alpha,1} = \overline{\mathbf{U}}_{\alpha,2} = \overline{\mathbf{U}}_{\alpha}$ for every $\alpha \in \{1,\dots,n_a\}$. Using this strategy, the problem becomes bi-linear and then the objective is to estimate both the matrices $\overline{\mathbf{U}}_1 = (\overline{\mathbf{U}}_{\alpha,1})_{1\leqslant \alpha\leqslant n_a} \in \mathbb{C}^{2\times 2K\times P\times n_a}$ and $\overline{\mathbf{U}}_2 = (\overline{\mathbf{U}}_{\alpha,2})_{1\leqslant \alpha\leqslant n_a} \in \mathbb{C}^{2\times 2K\times P\times n_a}$, where $\overline{\mathbf{U}}_1$ (resp. $\overline{\mathbf{U}}_2$) concatenates the non-zero Fourier coefficients $\overline{\mathbf{U}}_{\alpha,1}$ (resp. $\overline{\mathbf{U}}_{\alpha,2}$) for all antennas. The estimation of these matrices is achieved by solving the following minimization problem for the DDE calibration:

$$\underset{\mathbf{U}_{1},\mathbf{U}_{2}}{\text{minimize}} \quad \widetilde{h}(\mathbf{U}_{1},\mathbf{U}_{2}) + p(\mathbf{U}_{1},\mathbf{U}_{2}), \tag{17}$$

where \tilde{h} is the least-squares data fidelity term and p is the regularization function for \mathbf{U}_1 and \mathbf{U}_2 . The data fidelity term for DDE calibration reads as

$$\widetilde{h}(\mathbf{U}_1, \mathbf{U}_2) = \sum_{1 \leq \alpha \leq n_a} \frac{1}{2} \|\mathcal{G}_{\alpha, 1}(\mathbf{U}_{\alpha, 1}) - \mathbf{Y}_{\alpha}\|_F^2, \tag{18}$$

$$= \sum_{1 \leq \alpha \leq n_a} \frac{1}{2} \|\mathcal{G}_{\alpha,2}(\mathbf{U}_{\alpha,2}) - \mathbf{Y}_{\alpha}\|_F^2, \tag{19}$$

where $\|\cdot\|_F$ stands for the Frobenious norm of the matrix argument, and considering t and β taking all the values respectively in the ranges $\{1,\ldots,T\}$ and $\{1,\ldots,n_a\}$, the following hold: $\mathbf{Y}_{\alpha} = (\mathbf{Y}_{t,\alpha,\beta})_{t,\beta\neq\alpha}$ and the operator $\mathcal{G}_{\alpha,1}(\mathbf{U}_{\alpha,1})$ in equation (18) generates a concatenation of the terms of the form $\mathcal{D}_t(\mathbf{U}_{\alpha,1}) X_{t,\alpha,\beta}(\mathbf{F} \widetilde{\mathbf{Z}} \widetilde{\mathcal{L}}(\mathbf{S})) \mathcal{D}_t'(\mathbf{U}_{\beta,2})$. Similarly, $\mathcal{G}_{\alpha,2}(\mathbf{U}_{\alpha,2})$ in equation (19) consists in the operation $\mathcal{D}_t(\mathbf{U}_{\beta,1}) X_{t,\beta,\alpha}(\mathbf{F} \widetilde{\mathbf{Z}} \widetilde{\mathcal{L}}(\mathbf{S})) \mathcal{D}_t'(\mathbf{U}_{\alpha,2})$ to produce the corresponding measurements.

On the other hand, the regularization term p in problem (17) is given by

$$p(\mathbf{U}_1, \mathbf{U}_2) = \gamma \|\mathbf{U}_1 - \mathbf{U}_2\|_E^2 + \iota_{\mathbb{D}}(\mathbf{U}_1) + \iota_{\mathbb{D}}(\mathbf{U}_2), \tag{20}$$

where $\gamma > 0$ is the regularization parameter and the first term in equation (20) controls the distance between the matrices U_1 and \mathbf{U}_2 , thereby imposing the constraint that these two matrices should be equal. The set \mathbb{D} is defined to constrain the values of the Fourier coefficients of the DDEs to lie within the specified bounds. In particular, \mathbb{D} is defined such that for each $t \in \{1, ..., T\}$, the Fourier kernels stored in diagonal terms of $\widehat{\mathbf{D}}_{t,lpha}$ have the central coefficients $([\widehat{\mathbf{D}}_{t,\alpha}(0)]_{11})$ and $[\widehat{\mathbf{D}}_{t,\alpha}(0)]_{22}$ belonging to an ℓ_{∞} complex ball centred in 1 with radius $\theta_1 > 0$, whereas the rest of the coefficients belong to an ℓ_{∞} complex ball centred in 0 with radius θ_1 . On the other hand, for the Fourier kernels in the off-diagonal terms, the central and the other coefficients are assumed to be contained in ℓ_{∞} balls centred in 0 with radius θ_2 and θ_3 , respectively. Intuitively, this can be understood as follows. Since we work in the scenario after the calibration transfer is performed, the zero spatial frequency coefficient of the DDEs (i.e. the DIEs) encoded in the diagonal terms of the Jones matrices are normalized to 1, thus lying in a complex neighbourhood of 1 + i0. Moreover, while the central coefficient in the spatial Fourier domain represents the mean gain, the higher order spatial frequencies characterize the gain variations across the field of view with respect to this mean gain. Therefore, these coefficients have smaller values in comparison with the central coefficient. Lastly, concerning the off-diagonal terms which encompass the polarization leakage, their values are usually much smaller than the diagonal terms.

6 PROPOSED CALIBRATION & IMAGING APPROACH

The current work deals with a practical case when neither the Stokes images nor the DDEs associated with the antennas are known. In the particular case when the sky is considered to be unpolarized, only Stokes I imaging needs to be performed and the 2×2 Jones matrices are often replaced by a scalar value. In this scenario, Repetti et al. (2017) recently proposed a method to perform joint calibration and imaging. Motivated by the good performance obtained by this flexible method with proven convergence guarantees, we extend this approach for full polarization model and propose a joint calibration and imaging algorithm to solve the problem under scrutiny. This leads to considering a global minimization problem aiming to solve for the Stokes parameters S and the calibration matrices U_1 and U_2 . This non-convex problem can be cast by combining the minimization problems (9) and calibration (17) which were proposed in the earlier sections solely for Stokes imaging and DDE calibration, respectively.

Given the non-convexity of the underlying minimization problem, choice of initialization is crucial. In this context, we exploit the fact that the calibration transfer has been performed, to obtain an initial estimate of the Stokes images, as suggested in Repetti & Wiaux (2017) (although for Stokes *I* imaging only). Such first imaging step consists in considering the Jones matrices as identity and solving the associated minimization problem (9).

Let \mathbf{S}' be the Stokes parameters estimated by solving problem (9). Since these are obtained ignoring the DDEs, in general, these may contain artefacts. Therefore, we instead use a thresholded version of \mathbf{S}' , denoted by \mathbf{S}_0 , which contains only the high amplitude coefficients of \mathbf{S}' . With this first approximation of the images at hand, the original unknown image can be seen as a sum of \mathbf{S}_0 and $\mathbf{E} \in \mathbb{R}^{2\times 2N}$, where the latter is unknown and need to be estimated.

Finally, we propose to define the estimates $(\mathbf{E}, \mathbf{U}_1, \mathbf{U}_2)$ as solutions to the following global, non-convex minimization problem:

$$\underset{\mathbf{E}, \mathbf{U}_1, \mathbf{U}_2}{\text{minimize}} \quad h(\mathbf{E}, \mathbf{U}_1, \mathbf{U}_2) + r(\mathbf{E}) + p(\mathbf{U}_1, \mathbf{U}_2), \tag{21}$$

where h is the least squares data fidelity term associated with the data model, r and p are the regularization terms for the image and the DDEs, respectively. In particular, with $\mathbf{S} = \mathbf{S}_0 + \mathbf{E}$, the data fidelity term is given by the least squares criterion as proposed in equations (10), (18) and (19), i.e.

$$h(\mathbf{E}, \mathbf{U}_1, \mathbf{U}_2) = \|\Phi(\mathbf{S}_0 + \mathbf{E}) - \mathbf{y}\|_2^2$$
 (22)

$$= \sum_{1 \leqslant \alpha \leqslant n_{\alpha}} \frac{1}{2} \| \mathcal{G}_{\alpha,1}(\mathbf{U}_{\alpha,1}) - \mathbf{Y}_{\alpha} \|_{F}^{2}, \tag{23}$$

$$= \sum_{1 \leqslant \alpha \leqslant n_{\alpha}} \frac{1}{2} \|\mathcal{G}_{\alpha,2}(\mathbf{U}_{\alpha,2}) - \mathbf{Y}_{\alpha}\|_{F}^{2}, \tag{24}$$

where operator Φ in equation (22) is formed using fixed values of $(\mathbf{U}_1,\mathbf{U}_2)$. Similarly, $\mathcal{G}_{\alpha,1}(\mathbf{U}_{\alpha,1})$ in equation (23) (and $\mathcal{G}_{\alpha,2}(\mathbf{U}_{\alpha,2})$ in equation (24), resp.) is determined by fixed $(\mathbf{U}_{\beta,2},\mathbf{E})$ ($(\mathbf{U}_{\beta,1},\mathbf{E})$, resp.) with $\beta \in \{1,\dots,n_a\}$ and $\beta \neq \alpha$. While estimating the Stokes parameters, equation (22) is employed as the data fidelity term. In particular, keeping $(\mathbf{U}_1,\mathbf{U}_2)$ fixed, the convexity of this term with respect to \mathbf{E} can be noticed. In the same manner, equation (23) ((24), resp.) is chosen while updating \mathbf{U}_1 (\mathbf{U}_2 , resp.) which is convex with respect to \mathbf{U}_1 (\mathbf{U}_2 , resp.) keeping the other two variables fixed. It is then straightforward to see that the non-convex function h is in fact convex for each of the variables while fixing the others.

Concerning the regularization terms, the function $p(\mathbf{U}_1, \mathbf{U}_2)$

is given from equation (20), whereas the function $r(\mathbf{E})$ for the images is associated with the priors introduced in equations (11), (12) and (14). In particular, one can choose whether to take the polarization constraint into account or not. In the former case, the regularization term boils down to

$$r(\mathbf{E}) = g(\mathbf{E}) + \iota_{\mathbb{P}}(\mathbf{S}_0 + \mathbf{E}) + \iota_{\mathbb{K}}(\mathbf{E}), \tag{25}$$

where $g(\mathbf{E}) = \sum_{i=1}^4 \eta_i \| (\mathbf{\Psi}^\dagger \mathcal{R}(\mathbf{S}_0 + \mathbf{E}))_{:,i} \|_1$ denotes the SARA prior considering $\mathbf{\Psi}$ to be a concatenation of Dirac basis and first eight Daubechies wavelets (Carrillo et al. 2014; Birdi et al. 2018b). Here, we have taken the weighting matrix to be equal to identity, i.e. without adapting the reweighting scheme and, for every $i \in \{1, 2, 3, 4\}, \eta_i > 0$ is the regularization parameter. The set \mathbb{K} is chosen to take into account the errors that might appear on the estimated non-zero coefficients of S_0 . Formally, it is defined as

$$\mathbb{K} = \left\{ \mathbf{E} \in \mathbb{R}^{2 \times 2N} | (\forall n \in \mathbb{S}_0) \ \mathsf{E}(n) \in [-\vartheta \, \mathsf{S}_0(n), \vartheta \, \mathsf{S}_0(n)] \right\}, \tag{26}$$

where \mathbb{S}_0 is the support of \mathbf{S}_0 . The parameter $\vartheta \in [0,1]$ is chosen according to the error percentage considered on S_0 .

On the other hand, in the absence of the polarization constraint, the positivity of Stokes I image needs to be imposed explicitly, that can be accounted for by replacing the set K in equation (25) by the set \mathbb{K}' , defined as

$$\mathbb{K}' = \left\{ \mathbf{E} \in \mathbb{R}^{2 \times 2N} \left| (\forall n \in \mathbb{S}_0) \right| \mathcal{E}(n) \in [-\vartheta \, \mathcal{S}_0(n), \vartheta \, \mathcal{S}_0(n)], \right. \\ \left. (\forall n \notin \mathbb{S}_0) \left[\mathcal{E}(n) \right]_{1,1} \geqslant 0 \right\}. \quad (27)$$

In such a case, the function r reads as

$$r(\mathbf{E}) = g(\mathbf{E}) + \iota_{\mathbb{K}'}(\mathbf{E}). \tag{28}$$

Proposed algorithm

In order to solve problem (21), we observe that it has a blockvariable structure with U_1, U_2 and E being the three blocks constituting the problem. On top of it, although the global problem is non-convex, it is convex with respect to each of these blocks. Leveraging this block-variable structure, we propose to use an iterative algorithm based on a block-coordinate forward-backward approach (Chouzenoux et al. 2016) to solve problem (21). It consists in alternating between the estimation of the DDEs and the Stokes images. In turn, for each of these estimations, FB iterations are employed, i.e. the smooth terms are dealt by their gradient, whereas the non-smooth terms are managed by their proximity operators. Formally, given a proper, lower-semicontinuous function $f: \mathbb{R}^{2 \times 2N} \to]-\infty, +\infty]$, its proximity operator (Moreau 1965) at $\mathbf{S} \in \mathbb{R}^{2 \times 2N}$ is defined as

$$\operatorname{prox}_{f}(\mathbf{S}) = \underset{\mathbf{V} \in \mathbb{R}^{2 \times 2N}}{\operatorname{argmin}} f(\mathbf{V}) + \frac{1}{2} \|\mathbf{V} - \mathbf{S}\|_{F}^{2}. \tag{29}$$

In the particular case of f being an indicator function of a set \mathbb{E} , the proximity operator reduces to commonly known projection operator $\mathcal{P}_{\mathbb{E}}$, defined as

$$\mathcal{P}_{\mathbb{E}}(\mathbf{S}) = \underset{\mathbf{V} \in \mathbb{B}}{\operatorname{argmin}} \|\mathbf{V} - \mathbf{S}\|_{F}^{2}. \tag{30}$$

In light of the discussion above, we present the proposed algorithm as Algorithm 1. It consists of a global loop and inner iteration loops. At each iteration of the global loop, indexed by $i \in \mathbb{N}$ (step 2), we choose either to update the DDEs or the image following an essentially cyclic rule, that is each of the variables must be updated at least once within a given finite number of iterations. For every i^{th} iteration, this is taken care by the choice of number of inner loop

iterations $L^{(i)} \in \mathbb{N}$ and $J^{(i)} \in \mathbb{N}$ to update the DDEs and the image, respectively. To be more precise, in the former case, each of the calibration matrices $\mathbf{U}_{1}^{(i)}$ (step 5) and $\mathbf{U}_{2}^{(i)}$ (step 9) are updated by performing $L^{(i)}$ number of FB iterations in the inner loop, using the images estimated at the previous iterate. When the images are chosen to be updated in the global loop, the updated DDEs from the previous iterate are used to estimate the image in step 15, executing $J^{(i)}$ FB iterations. The overall algorithm can then be understood by splitting it into two parts: Calibration and Imaging. These two parts are explained in what follows.

Algorithm 1 Joint calibration and imaging algorithm

- 1: **Initialization**: $\mathbf{E}^{(0)} \in \mathbb{R}^{2 \times 2N}$, $(\mathbf{U}_1^{(0)}, \mathbf{U}_2^{(0)}) \in (\mathbb{C}^{2 \times 2K \times P \times n_a})^2$. Let, for every $i \in \mathbb{N}$, $(L^{(i)}, J^{(i)}) \in \mathbb{N}^2$
- 2: **for** $i = 0, 1, \ldots$ **do** Choose to update either the DDEs or the images.

If the DDEs are updated:

$$\begin{array}{lll} \text{3:} & \left(\ \mathbf{U}_{1}^{(i,0)}, \mathbf{U}_{2}^{(i,0)} \right) = \left(\mathbf{U}_{1}^{(i)}, \mathbf{U}_{2}^{(i)} \right) \\ \text{4:} & \quad \text{for } \ell = 0, \dots, L^{(i)} - 1 \ \text{do} \\ & \quad \mathbf{U}_{1}^{(i,\ell+1)} = \mathcal{P}_{\mathbb{D}} \bigg(\mathbf{U}_{1}^{(i,\ell)} - \ \Gamma_{1}^{(i)} \cdot \nabla_{\mathbf{U}_{1}} h(\mathbf{E}^{(i)}, \mathbf{U}_{1}^{(i,\ell)}, \mathbf{U}_{2}^{(i)}, \mathbf{U}_{2}^{(i)}) \\ \text{5:} & \quad -\Gamma_{1}^{(i)} \cdot \gamma \left(\mathbf{U}_{1}^{(i,\ell)} - \mathbf{U}_{2}^{(i)} \right) \bigg) \\ \text{6:} & \quad \text{end for} \\ \text{7:} & \quad \mathbf{U}_{1}^{(i+1)} = \mathbf{U}_{1}^{(i,L^{(i)})} \end{array}$$

7:
$$\mathbf{U}_{1}^{(i+1)} = \mathbf{U}_{1}^{(i,L^{(i)})}$$

8: **for**
$$\ell = 0, ..., L^{(i)} - 1$$
 do

9:
$$\begin{aligned} \mathbf{U}_{2}^{(i,\ell+1)} &= \mathcal{P}_{\mathbb{D}} \bigg(\mathbf{U}_{2}^{(i,\ell)} - \Gamma_{2}^{(i)} \cdot \nabla_{\mathbf{U}_{2}} h \big(\mathbf{E}^{(i)}, \mathbf{U}_{1}^{(i+1)}, \mathbf{U}_{2}^{(i,\ell)} \big) \\ &- \Gamma_{2}^{(i)} \cdot \gamma \big(\mathbf{U}_{2}^{(i,\ell)} - \mathbf{U}_{1}^{(i+1)} \big) \bigg) \end{aligned}$$

10: end for

11:
$$\mathbf{U}_2^{(i+1)} = \mathbf{U}_2^{(i,L^{(i)})}$$

$$\mathbf{c}(i+1) = \mathbf{c}(i)$$

If the Stokes images are updated:

13:
$$\mathbf{E}^{(i,0)} = \mathbf{E}^{(i)}$$

14: **for**
$$j = 0, ..., J^{(i)} - 1$$
 do

15:
$$\mathbf{E}^{(i,j+1)} = \operatorname{prox}_{\sigma^{(i)}r} \left(\mathbf{E}^{(i,j)} - \sigma^{(i)} \nabla_{\mathbf{E}} h \left(\mathbf{E}^{(i,j)}, \mathbf{U}_{1}^{(i+1)}, \mathbf{U}_{2}^{(i+1)} \right) \right)$$

16:

17:
$$\mathbf{E}^{(i+1)} = \mathbf{E}^{(i,J^{(i)})}$$

18:
$$(\mathbf{U}_1^{(i+1)}, \mathbf{U}_2^{(i+1)}) = (\mathbf{U}_1^{(i)}, \mathbf{U}_2^{(i)}).$$

19: **end for**

Calibration: It comprises of the estimation of the matrices U_1 and \mathbf{U}_2 . In this case, while the data fidelity term h is differentiable, the regularization term p consist of both smooth and non-smooth terms. Thus, in each ℓ^{th} FB iteration to estimate either of these matrices, the gradient step for the differentiable terms is coupled with the proximity operator of the non-smooth term, as shown in steps 5 and 9. In particular, for the gradient step, for every $q \in \{1, 2\}$, the step size $\Gamma_q^{(i)} \in \mathbb{R}^{2 \times 2K \times P \times n_a}$ is chosen as

$$\Gamma_q^{(i)} = \left(\zeta_{q,\alpha}^{(i)} \, \mathbf{1}_{2\times 2K\times P}\right)_{1\leqslant \alpha\leqslant n_a},\tag{31}$$

where $\mathbf{1}_R$ is a matrix of ones of dimension R, and $\zeta_{q,\alpha}^{(i)}$ is given by $0 < \zeta_{q,\alpha}^{(i)} < 1/(\gamma + \upsilon_{q,\alpha}^{(i)})$, with $\upsilon_{q,\alpha}^{(i)}$ denoting the Lipschitz constant of the partial derivative of h with respect to $\mathbf{U}_{q,\alpha}^{(i)}$.

Furthermore, since in this case the non-smooth term is the indicator function of the set \mathbb{D} , as explained previously, this reduces to performing projection on this set, $\mathcal{P}_{\mathbb{D}}$ which basically ensures that the values of the estimated DDEs Fourier coefficients lie within the bounds specified earlier.

Imaging: This step updates the Stokes images while using the DDEs estimates from the previous iterate. As shown in step 15, it involves computing the gradient of the data fidelity term, followed by the proximity operator of the regularization function r. In this case, the step size $\sigma^{(i)}$ for the gradient step is chosen such that it satisfies

$$0 < \sigma^{(i)} < 1/\|\Phi\|_2,\tag{32}$$

where $\|\Phi\|_2$ computes the spectral norm of Φ , and Φ is generated using the updated values $(\mathbf{U}_1^{(i+1)}, \mathbf{U}_2^{(i+1)})$.

Regarding the proximity step, the regularization term $r(\mathbf{E})$ is a hybrid term incorporating a mixture of prior information. Particularly, based on the choice of inclusion or exclusion of the polarization constraint, the computation of the proximity operator differs. For comparison purposes, here we consider both the cases.

Regularization without polarization constraint: When we work in the absence of the polarization constraint, the regularization term $r(\mathbf{E})$ is given by equation (28). Then, the proximity operator of r evaluated at any point $\mathbf{R} \in \mathbb{R}^{2 \times 2N}$ amounts to

$$\operatorname{prox}_{\sigma^{(i)}r}(\mathbf{R}) = \underset{\widetilde{\mathbf{E}}}{\operatorname{argmin}} \ \sigma^{(i)}g(\widetilde{\mathbf{E}}) + \iota_{\mathbb{K}'}(\widetilde{\mathbf{E}}) + \frac{1}{2}\|\widetilde{\mathbf{E}} - \mathbf{R}\|_F^2, \tag{33}$$

which does not have an explicit formulation and thus requires sub-iterations for its computation. One such possibility is to employ dual forward-backward algorithm (Combettes & Pesquet 2011; Combettes et al. 2011), as presented in Algorithm 2.

In particular, step 3 of Algorithm 2 performs the projection onto the set \mathbb{K}' . It is followed by the computation of the proximity operator of function g in step 5, which in the current case of g being the ℓ_1 norm, corresponds to the *soft-thresholding operator* (Chaux et al. 2007). This is evaluated as

$$\operatorname{prox}_{\boldsymbol{\varepsilon}^{(i)}g}\left(\mu^{-1}\mathbf{H}^{(k)}\right) = \widetilde{\mathbf{R}},\tag{34}$$

where $\varepsilon^{(i)} = \mu^{-1}\sigma^{(i)}$ and the soft-thresholding operation is a component-wise operation such that $\widetilde{\mathbf{R}} \in \mathbb{R}^{J \times 4}$ is defined for every $q \in \{1, 2, 3, 4\}$ and $j \in \{1, \dots, J\}$ as

$$\widetilde{\mathbf{R}}_{j,q} = \begin{cases} -\mu^{-1}\mathsf{H}_{j,q}^{(k)} + \varepsilon^{(i)}\eta_q\,, & \text{if } \mu^{-1}\mathsf{H}_{j,q}^{(k)} < -\varepsilon^{(i)}\eta_q\,, \\ 0, & \text{if } -\varepsilon^{(i)}\eta_q \leqslant \mu^{-1}\mathsf{H}_{j,q}^{(k)} \leqslant \varepsilon^{(i)}\eta_q\,, \\ \mu^{-1}\mathsf{H}_{j,q}^{(k)} - \varepsilon^{(i)}\eta_q\,, & \text{otherwise.} \end{cases} \tag{35}$$

In essence, this operation sets the elements smaller than a threshold to zero, whereas shrinks all the other elements by this value. Therefore, it promotes sparsity of its argument variable.

Regularization with polarization constraint: In the case when the polarization constraint is to be enforced, the regularization term is given from equation (25) and the associated proximity operator

Algorithm 2 Dual Forward-Backward algorithm to compute (33)

1: **Initialization:** Let
$$\widetilde{\mathbf{P}}^{(0)} \in \mathbb{R}^{J \times 4}$$
, $\bar{\epsilon} \in]0$, min $\{1, 1/\|\mathbf{\Psi}^{\dagger}\|^2\}[$, and $\mu \in [\bar{\epsilon}, 2/\|\mathbf{\Psi}^{\dagger}\|^2 - \bar{\epsilon}]$

2: **for** $k = 0, 1, \dots$

3:
$$\mathbf{V}^{(k)} = \mathcal{P}_{\mathbb{K}'} \left(\mathbf{R} - \mathcal{R}^{\dagger} \left(\mathbf{\Psi} \widetilde{\mathbf{P}}^{(k)} \right) \right)$$

4:
$$\mathbf{H}^{(k)} = \widetilde{\mathbf{P}}^{(k)} + \mu \mathbf{\Psi}^{\dagger} \mathcal{R} (\mathbf{V}^{(k)})$$

5:
$$\widetilde{\mathbf{P}}^{(k+1)} = \mathbf{H}^{(k)} - \mu \operatorname{prox}_{\mu^{-1}\sigma^{(i)}g} (\mu^{-1}\mathbf{H}^{(k)})$$

6: end for

7: **Return:** $\widetilde{\mathbf{E}} = \lim_{k} \mathbf{V}^{(k)}$

at a point R is given by

$$\operatorname{prox}_{\sigma^{(i)}r}(\mathbf{R}) = \underset{\widetilde{\mathbf{E}}}{\operatorname{argmin}} \ \sigma^{(i)}g(\widetilde{\mathbf{E}}) + \iota_{\mathbb{K}}(\widetilde{\mathbf{E}}) + \iota_{\mathbb{P}}(\widetilde{\mathbf{E}} + \mathbf{S}_0) + \frac{1}{2} \|\widetilde{\mathbf{E}} - \mathbf{R}\|_F^2. \tag{36}$$

The evaluation of this operator requires projection onto the set \mathbb{P} , that does not have a closed form solution. To circumvent this difficulty, we have recently proposed a method, named Polarized SARA (Birdi et al. 2018a,b), that enforces this constraint leveraging the epigraphical projection techniques (Chierchia et al. 2015). In this context, with the introduction of an auxiliary variable $\mathbf{Z} \in \mathbb{R}^{N \times 2}$, the polarization constraint set is splitted into simpler, easily manageable constraint sets, thereby performing the projection onto these sets. Formally, it corresponds to the following reformulation of problem (36):

$$\operatorname{prox}_{\sigma^{(i)}r}(\mathbf{R}) = \underset{\widetilde{\mathbf{E}}}{\operatorname{argmin}} \ \sigma^{(i)}g(\widetilde{\mathbf{E}}) + \iota_{\mathbb{K}}(\widetilde{\mathbf{E}}) + \frac{1}{2} \|\widetilde{\mathbf{E}} - \mathbf{R}\|_{F}^{2}$$
 (37a)

subject to $(\forall n \in \{1, ..., N\})$

$$h_1(\mathcal{R}(\widetilde{\mathbf{E}} + \mathbf{S}_0)_{n,1}) = -(\widetilde{\mathbf{E}} + \mathbf{S}_0)_{n,1} \leqslant \mathbf{Z}_{n,1},\tag{37b}$$

$$\begin{cases} h_2(\mathcal{R}(\widetilde{\mathbf{E}} + \mathbf{S}_0)_{n,2:4}) = \|\mathcal{R}(\widetilde{\mathbf{E}} + \mathbf{S}_0)_{n,2:4}\|_2 \leqslant \mathsf{Z}_{n,2}, \end{cases}$$
(37c)

$$\mathsf{Z}_{n,1} + \mathsf{Z}_{n,2} \leqslant 0. \tag{37d}$$

In order to impose the constraints (37b)- (37d), we make use of the indicator functions of the corresponding sets. In particular, the constraints (37b) and (37c) are enforced using indicator functions of the epigraphs² \mathbb{E}_1 , \mathbb{E}_2 respectively of the functions h_1 and h_2 . On the other hand, the set $\mathbb{V} = \{\mathbf{Z} \in \mathbb{R}^{N \times 2} \mid (\forall n \in \{1, \dots, N\}) \ \mathsf{Z}_{n,1} + \mathsf{Z}_{n,2} \leqslant 0\}$ is used to impose the constraint (37d). For further details, we refer the reader to Birdi et al. (2018b).

Doing so, the proximity operator (37a) results into

$$\operatorname{prox}_{\sigma^{(i)}r}(\mathbf{R}) = \underset{\widetilde{\mathbf{E}}, \mathbf{Z}}{\operatorname{argmin}} \ \sigma^{(i)}g(\widetilde{\mathbf{E}}) + \iota_{\mathbb{K}}(\widetilde{\mathbf{E}}) + \frac{1}{2} \|\widetilde{\mathbf{E}} - \mathbf{R}\|_{F}^{2} + \iota_{\mathbb{V}}(\mathbf{Z})$$
$$+ \iota_{\mathbb{E}_{1}} \left(\mathcal{R}(\widetilde{\mathbf{E}} + \mathbf{S}_{0})_{:,1}, \mathbf{Z}_{:,1} \right)$$
$$+ \iota_{\mathbb{E}_{2}} \left(\mathcal{R}(\widetilde{\mathbf{E}} + \mathbf{S}_{0})_{:,2:4}, \mathbf{Z}_{:,2} \right). \tag{38}$$

For the computation of this proximity operator, we employ Polarized SARA method which is based on primal-dual forward-backward algorithm (Condat 2013; Vũ 2013; Pesquet & Repetti 2015). Particularly, leveraging the flexibility and parallelizability

² The epigraph of a proper, lower semi-continuous function $\psi \colon \mathbb{R}^N \to]-\infty, +\infty]$ corresponds to epi $\psi = \{(z, \kappa) \in \mathbb{R}^N \times \mathbb{R} \, | \, \psi(z) \leqslant \kappa \}$.

Algorithm 3 Primal-dual algorithm to solve problem (38)

1: **Initialization:**
$$\widetilde{\mathbf{E}}^{(0)} \in \mathbb{R}^{2\times 2N}, \mathbf{Z}^{(0)} \in \mathbb{R}^{N\times 2}, \mathbf{A}^{(0)} \in \mathbb{R}^{J\times 4},$$

$$\mathbf{C}^{(0)} \in \mathbb{R}^{N\times 4}, \mathbf{B}^{(0)} \in \mathbb{R}^{N\times 2}, (\rho_1, \rho_2, \tau) \in \mathbb{R}^3_+ \text{ such that } \rho_1 \|\mathbf{\Psi}^{\dagger}\|_2^2 + \rho_2 < \tau^{-1}$$

2: **repeat for** k = 0, 1, ...

Primal updates

3:
$$\overline{\widetilde{\mathbf{E}}^{(k+1)}} = \mathcal{P}_{\mathbb{K}} \left(\widetilde{\mathbf{E}}^{(k)} - \tau \left(\widetilde{\mathbf{E}}^{(k)} - \mathbf{R} \right) - \tau \mathcal{R}^{\dagger} \left(\mathbf{\Psi} \mathbf{A}^{(k)} + \mathbf{C}^{(k)} \right) \right)$$

4:
$$\mathbf{Z}^{(k+1)} = \mathcal{P}_{\mathbb{V}} \left(\mathbf{Z}^{(k)} - \tau \mathbf{B}^{(k)} \right)$$

Dual updates

Promoting sparsity:

5:
$$\widetilde{\mathbf{A}}^{(k)} = \mathbf{A}^{(k)} + \rho_1 \mathbf{\Psi}^{\dagger} \mathcal{R} \left(2 \, \widetilde{\mathbf{E}}^{(k+1)} + \mathbf{S}_0 - \widetilde{\mathbf{E}}^{(k)} \right)$$

6:
$$\mathbf{A}^{(k+1)} = \widetilde{\mathbf{A}}^{(k)} - \rho_1 \operatorname{prox}_{\rho_1^{-1}\sigma^{(i)}g} \left(\rho_1^{-1} \widetilde{\mathbf{A}}^{(k)} \right)$$

Performing epigraphical projection:

7:
$$\widetilde{\mathbf{C}}^{(k)} = \mathbf{C}^{(k)} + \rho_2 \mathcal{R} \left(2 \widetilde{\mathbf{E}}^{(k+1)} + \mathbf{S}_0 - \widetilde{\mathbf{E}}^{(k)} \right)$$

8:
$$\widetilde{\mathbf{B}}^{(k)} = \mathbf{B}^{(k)} + \rho_2 \left(2 \mathbf{Z}^{(k+1)} - \mathbf{Z}^{(k)} \right)$$

9:
$$\begin{bmatrix} \mathbf{C}_{:,1}^{(k+1)} \\ \mathbf{B}_{:,1}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{C}}_{:,1}^{(k)} \\ \widetilde{\mathbf{B}}_{:,1}^{(k)} \end{bmatrix} - \rho_2 \mathcal{P}_{\mathbb{E}_1} \left(\frac{1}{\rho_2} \begin{bmatrix} \widetilde{\mathbf{C}}_{:,1}^{(k)} \\ \widetilde{\mathbf{B}}_{:,1}^{(k)} \end{bmatrix} \right)$$

$$\begin{bmatrix} \mathbf{C}^{(k+1)} \end{bmatrix} \quad \begin{bmatrix} \widetilde{\mathbf{C}}^{(k)} \end{bmatrix} \quad (\begin{bmatrix} \widetilde{\mathbf{C}}^{(k)} \end{bmatrix}$$

10:
$$\begin{bmatrix} \mathbf{C}_{:,2:4}^{(k+1)} \\ \mathbf{B}_{:,2}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{C}}_{:,2:4}^{(k)} \\ \widetilde{\mathbf{B}}_{:,2}^{(k)} \end{bmatrix} - \rho_2 \mathcal{P}_{\mathbb{E}_2} \left(\frac{1}{\rho_2} \begin{bmatrix} \widetilde{\mathbf{C}}_{:,2:4}^{(k)} \\ \widetilde{\mathbf{B}}_{:,2}^{(k)} \end{bmatrix} \right)$$

11: until convergence

offered by the primal-dual algorithms, we can easily adapt Polarized SARA method to solve our underlying problem (38). For the sake of completeness, we also present this adapted version in Algorithm 3. It comprises of solving for the variables of interest, the primal variables, along with the associated auxiliary variables, the dual variables. In this context, the update of the primal variables is similar to the forward-backward strategy, wherein the gradient of the differentiable terms is followed by their proximity step. For the current case, it implies projection onto the set \mathbb{K} (step $\overline{\mathbf{3}}$) and \mathbb{V} (step 4) for the update of variables **E** and **Z**, respectively. It is to be noted that these updates incorporate an additive term based on the corresponding dual variables. More precisely, the update of $\widetilde{\mathbf{E}}$ involves the variables $\mathbf{A} \in \mathbb{R}^{J \times 4}$ and $\mathbf{C} \in \mathbb{R}^{N \times 4}$ related to the sparsity prior and the epigraphical constraints, respectively. Similarly, step 4 comprises of the variable $\mathbf{B} \in \mathbb{R}^{N \times 2}$ associated with the epigraphical constraints. These dual variables are in turn updated by the computation of their associated proximity operators, that is the soft-thresholding operator for A (step 6) using the definition in equations (34) and (35), and the projections onto the sets \mathbb{E}_1 and \mathbb{E}_2 for the variables C and B. A detailed description of these updates is provided in Birdi et al. (2018b).

6.2 Convergence properties

The convergence properties of the proposed joint calibration and imaging algorithm for full polarization model (Algorithm 1) can be deduced from Chouzenoux et al. (2016). Particularly, if a finite

number of iterations are performed to update each of the variables, that is for every i^{th} iteration, choosing $L^{(i)}$ and $J^{(i)}$ to be finite in Algorithm 1, and if the

- blocks $(\mathbf{E}^{(i)}, \mathbf{U}_1^{(i)}, \mathbf{U}_2^{(i)})$ in Algorithm 1 are estimated at least once in every finite number of given iterations,
- step sizes $\Gamma_1^{(i)}$ and $\Gamma_2^{(i)}$ for the gradient steps while updating \mathbf{U}_1 and \mathbf{U}_2 , respectively are chosen as per equation (31), and
- step size $\sigma^{(i)}$ for the update of **E** is chosen as per equation (32).

then Algorithm 1 is guaranteed to converge to a critical point $(\mathbf{E}', \mathbf{U}'_1, \mathbf{U}'_2)$ of the underlying objective function given in equation (21). Additionally, iteration-by-iteration, the value of this objective function decreases.

6.3 Computational complexity

The computational complexity of Algorithm 1 can be analyzed from its two sub-parts: Calibration and Imaging. In the former case, the update of matrix \mathbf{U}_1 in steps 4 - 6 (and \mathbf{U}_2 in steps 8 - 10) is performed in parallel for n_a antennas, using forward-backward (FB) iterations. In this case, the gradient evaluation of the data fidelity term in equation (23) (and in (24)) is required. It involves the degridding operation while generating the matrix $X_{t,\beta,\alpha}$ ($\mathbf{F}\widetilde{\mathbf{Z}}\widetilde{\mathcal{L}}(\mathbf{S}_0 + \mathbf{E})$) for all $t \in \{1,\ldots,T\}$ and $\beta \in \{1,\ldots,n_a\}$, with $\beta \neq \alpha$, that turns out to be the most expensive step. However, it is to be noted that for each global calibration iteration, this matrix needs to be computed only once before the inner FB iterations (both for \mathbf{U}_1 and \mathbf{U}_2 updates).

Regarding the imaging step, the FB strategy presents the computationally most demanding part of the algorithm. In particular, in step 15, the gradient computation (forward step) of the data fidelity term (equation (22)) consists in performing the application of the operator Φ , followed by its adjoint operator. Furthermore, the evaluation of the proximity operator (backward step) involves sub-iterations, performed either by Algorithm 2 or Algorithm 3. While this iterative process adds to the computational cost, the heaviest steps within either of the algorithms is the application of the wavelet transform operator Ψ consisting of 8 wavelet bases and Dirac basis to impose sparsity. This can in turn be implemented in a parallel fashion, for each sparsity basis as well as for each of the underlying images.

7 SIMULATIONS AND RESULTS

In this section, we describe the various simulations conducted using a MATLAB implementation of the proposed algorithm to assess its performance.

To simulate the data sets, we consider VLA antenna's A configuration. It consists of $n_a=27$ antennas with each antenna pair acquiring measurements at T=200 snapshots, considering linear feeds. While performing the non-uniform Fourier transform at these sampled frequencies, we make use of the Kaiser-Bessel kernels of size 5×5 for interpolation (Fessler & Sutton 2003). Furthermore, we perform tests on two sets of model images of size $N=256\times 256$. The two sets correspond to the images of the radio galaxies Cygnus A and Hydra A, shown respectively in first and second columns of Fig. 3. For each set, Stokes I, Q and U images V=0 due to negligible circular polarization) are displayed from top to bottom.

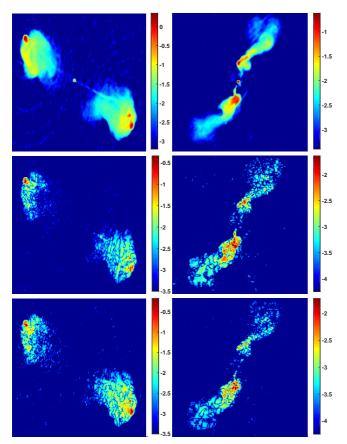


Figure 3. Cygnus A (first column) and Hydra A (second column) ground truth images used for performing simulations. In both the columns, rowwise from top to bottom the following images are shown (all in log scale): Stokes I, Stokes Q and Stokes U.

Regarding the DDE Fourier kernels, for every antenna $\alpha \in$ $\{1,\ldots,n_a\}$ and at each time instant $t\in\{1,\ldots,T\}$, these kernels are generated randomly in the Fourier domain, with a spatial Fourier support of $K = 5 \times 5$ and a temporal Fourier support of size $P = 3 \times 3$. Furthermore, the values of the Fourier coefficients are chosen as per the discussion in Section 5. More precisely, for the diagonal terms in the Jones matrix, the central coefficients that correspond to the Fourier coefficients of the DIEs have their real and imaginary parts lying in the interval $[1-\theta_1, 1+\theta_1]$ and $[-\theta_1, \theta_1]$, respectively, whereas the other frequency coefficients (related to the DDEs Fourier coefficients) belong to $[-\theta_1, \theta_1]$ with $\theta_1 = 5 \times 10^{-2}$. On the other hand, the central and all the rest coefficients in the offdiagonal terms are considered belonging to $[-\theta_2, \theta_2]$ and $[-\theta_3, \theta_3]$, respectively. Here we choose $(\theta_2, \theta_3) = (5 \times 10^{-2}, 5 \times 10^{-4})$. The chosen values are in line with the VLA characteristics, wherein the leakage terms (i.e. the off-diagonal terms) are $\sim 10^{-2}$ lower in peak amplitude compared to the diagonal terms, representing a leakage of around 5% of the diagonal terms (Bhatnagar et al. 2013).

7.1 Comparisons performed

For an assessment of the proposed algorithm, we perform an extensive study on a varied number of cases based on choices made both on calibration and imaging fronts. From the calibration viewpoint, we compare between different generations of calibration schemes, i.e. 1GC, 2GC and 3GC. Particularly, we compare the results obtained by DDE calibration with those of considering only DIEs. Furthermore, to determine the importance of off-diagonal terms in

the Jones matrices, we run tests with and without calibrating for these terms. On the other hand, from the imaging perspective, it is interesting to analyze the performance of different regularizations for the Stokes parameters. Thus, we consider the cases with and without enforcement of the polarization constraint. A blend of these approaches leads to the following list of tests:

- (1) Imaging with normalized DIEs (1GC): Working under the assumption of calibration transfer, the first step is to consider the Jones matrices as identity, without any directional dependency. It is important to mention here that this is the usual case considered by imaging algorithms in RI, either ignoring the calibration effects or relying on the data provided after the calibration transfer. In such a scenario, only imaging step is performed to obtain an estimation of the Stokes images **S**. Furthermore, in this imaging problem, we can consider two different regularizations and hence two sub-cases: Imaging with normalized DIEs (1.a) without and (1.b) with the enforcement of the polarization constraint.
- (2) Joint DIE calibration and imaging (2GC): It involves using the proposed algorithm for calibration and imaging. While calibrating, only the DIEs, i.e. considering support size K=1 for Fourier kernels, are solved for. A thresholded version of the images estimated from (1.a) and (1.b) are used respectively to initialize the problems while working in the absence and presence of the polarization constraint. This corresponds to two subcases: DIE calibration and imaging (2.a) without and (2.b) with the polarization constraint.
- (3) Joint DDE calibration and imaging (3GC-polarized): This approach follows the proposed algorithm for calibrating the DDEs in conjunction with the imaging step. Similar to (2), the images are initialized from thresholded versions of (1.a) and (1.b), resulting into four cases:
- (3.a) DDE calibration for the Jones matrices, excluding the off-diagonal terms & imaging without the polarization constraint: It corresponds to the case when the calibration steps are applied only to update the diagonal terms of the Jones matrices, without accounting for the polarization constraint in the imaging step.
- (3.b) DDE calibration for the Jones matrices, excluding the offdiagonal terms & imaging with the polarization constraint: Here, the calibration strategy is the same as (3.a), whereas the polarization constraint is considered in the imaging step.
- (3.c) DDE calibration for the whole Jones matrix, including the off-diagonal terms & imaging without the polarization constraint: It consists in calibrating for the full Jones matrix, but without enforcing the polarization constraint in the imaging step.
- (3.d) DDE calibration for the whole Jones matrix, including the off-diagonal terms & imaging with the polarization constraint: This approach comprises of full Jones matrix calibration and imposing the polarization constraint in the imaging step.

7.2 Simulation settings

As previously mentioned, one of the key points for convergence of the proposed algorithm is to update the variables at least once in every finite number of given iterations. To ensure this, within the global loop in Algorithm 1, we perform $L_{\rm cyc}=2$ global iterations for the DDEs (i.e. steps 3 to 12 are iterated twice) followed by an update of the Stokes images (steps 13 to 18). Nevertheless, after a certain number of global loops, it may happen that the DDE updates

	without polarization constraint				with polarization constraint			
	Normalized	DIE	DDE calibration	DDE calibration	Normalized	DIE	DDE calibration	DDE calibration
	DIEs	calibration	w/o off-	of full Jones	DIEs	calibration	w/o off-	of full Jones
			diagonal terms	matrix			diagonal terms	matrix
Stokes I	19.4	21.4	34	34.7	18.4	20.2	33.6	37.4
Stokes Q	18.6	16.9	20.8	21.2	16.3	15.2	23.4	24.9
Stokes U	16.9	18	19.9	22.5	15.6	16.7	21.3	25.2

(a) SNR (in dB) values for Cygnus A

	without polarization constraint				with polarization constraint			
	Normalized	DIE	DDE calibration	DDE calibration	Normalized	DIE	DDE calibration	DDE calibration
	DIEs	calibration	w/o off-	of full Jones	DIEs	calibration	w/o off-	of full Jones
			diagonal terms	matrix			diagonal terms	matrix
Stokes I	1.04	1.73	10.8	12.8	1.07	1.82	9.64	41.4
Stokes Q	3.31	3.43	4.79	4.94	2.59	2.16	6.79	9.61
Stokes U	2	2.14	2.62	3.38	0.55	0.87	1.39	19.8

(b) Dynamic range (DR) values for Cygnus A in units of 10⁴

	without polarization constraint				with polarization constraint			
	Normalized	DIE	DDE calibration	DDE calibration	Normalized	DIE	DDE calibration	DDE calibration
	DIEs	calibration	w/o off-	of full Jones	DIEs	calibration	w/o off-	of full Jones
			diagonal terms	matrix			diagonal terms	matrix
Stokes I	18.3	20.6	24.9	25	18.3	20.7	31.2	33.2
Stokes Q	17.8	18.6	11.3	11.4	17.9	18.8	19.7	20.7
Stokes U	10.3	11.5	8.09	9.07	10.9	12.2	12.6	19.8

(c) SNR (in dB) values for Hydra A

	without polarization constraint				with polarization constraint			
	Normalized	DIE	DDE calibration	DDE calibration	Normalized	DIE	DDE calibration	DDE calibration
	DIEs	calibration	w/o off-	of full Jones	DIEs	calibration	w/o off-	of full Jones
			diagonal terms	matrix			diagonal terms	matrix
Stokes I	0.82	1.54	2.35	2.45	0.82	1.49	6.43	15.9
Stokes Q	2.01	2.21	0.23	0.23	0.68	0.93	1.01	1.19
Stokes U	2.44	2.53	0.21	0.28	0.31	0.5	0.48	2.7

(d) Dynamic range (DR) values for Hydra A in units of 10⁴

Table 1. Cygnus A (top two tables) and Hydra A (bottom two tables) results: SNR and Dynamic range values for different considered cases. For each of the reconstructed Stokes image, the shown values are the mean computed over 5 performed simulations for the cases with following considerations- for calibration step: known normalized DIEs, DIE calibration, no off-diagonal Jones terms calibration, and full Jones matrix DDE calibration; and for imaging step: without and with the enforcement of the polarization constraint.

stabilize in fewer than $L_{\rm cyc}$ iterations. Thus, to avoid unnecessary computation, we define a stopping criterion for the DDE updates as the relative variation between the consecutive estimates of DDEs to be less than a threshold $\varepsilon_{\rm U}$, that is

$$\max_{q \in \{1,2\}} \left(\|\mathbf{U}_{q}^{(i+1)} - \mathbf{U}_{q}^{(i)}\|_{F} / \|\mathbf{U}_{q}^{(i)}\|_{F} \right) \leqslant \varepsilon_{\mathsf{U}}. \tag{39}$$

where we set $\varepsilon_U = 10^{-5}$. In other words, if this criterion is met, then no more DDE iterations are performed, and the algorithm leaps to the imaging step using the estimated DDEs. The estimation of the Stokes images adopts a similar strategy wherein iterations in the inner imaging loop are stopped and the algorithm resorts to the calibration step once the following stopping criterion is satisfied:

$$\max_{q \in \{1,2,3\}} \left(\| \mathbf{E}_{:,q}^{(i+1)} - \mathbf{E}_{:,q}^{(i)} \|_2 / \| \mathbf{E}_{:,q}^{(i)} \|_2 \right) \leqslant \varepsilon_{\mathsf{E}}. \tag{40}$$

with $\varepsilon_{\mathsf{E}} = 10^{-5}$. Furthermore, in order to ensure stopping of the global algorithm at convergence, we define a stopping criterion as the relative variation between the values of the objective function at consecutive iterates to be less than a threshold ε_0 , that is

$$\|\varphi^{(i+1)} - \varphi^{(i)}\|_2 / \|\varphi^{(i)}\|_2 \leqslant \varepsilon_0,$$
 (41)

where $\varphi^{(i)}$ denotes the objective function value for the i^{th} iteration computed using the updated values $(\mathbf{E}^{(i+1)}, \mathbf{U}_1^{(i+1)}, \mathbf{U}_2^{(i+1)})$ and ε_0

is fixed to 2×10^{-2} . Regarding the choice of number of iterations $L^{(i)}$ and $J^{(i)}$ for calibration and imaging inner loops, respectively, we choose $L^{(i)} = 5$ when DDE updates need to be performed and $J^{(i)} = 100$ when imaging step is carried out. Furthermore, for each of the above mentioned cases, we run 5 simulations varying the DIEs/DDEs and noise realizations.

For quantitative comparison of the results obtained by different tests, we use signal-to-noise ratio (SNR) as a metric. Given an original image $\overline{s} \in \mathbb{R}^N$, the SNR of the reconstructed image s' with respect to \overline{s} is defined as

$$SNR = 20 \log_{10} \left(\frac{\|\overline{\mathbf{s}}\|_2}{\|\nu' \mathbf{s}' - \overline{\mathbf{s}}\|_2} \right), \tag{42}$$

where $\upsilon' = \underset{\upsilon>0}{\operatorname{argmin}} \|\upsilon s' - \overline{s}\|_2^2$ accounts for the ambiguity problem in the underlying blind deconvolution problem. Furthermore, the values of the regularization parameters $\eta = [\eta_1, \eta_2, \eta_3]$ and γ are chosen to maximize the SNR.

We also report the dynamic range (DR) obtained for the reconstructed images. For every image contained in the estimated Stokes matrix $\mathbf{S}' = \mathbf{S}_0 + \mathbf{E}'$, with $(i,j) \in \{1,2\}^2$ it is defined as follows:

$$DR_{ij} = \frac{\sqrt{N} \|\Phi\|_{2}^{2}}{\|\left[\Phi^{\dagger}(y - \Phi(S'))\right]_{ij}\|_{2}} \max_{n} [S'(n)]_{ij}.$$
 (43)

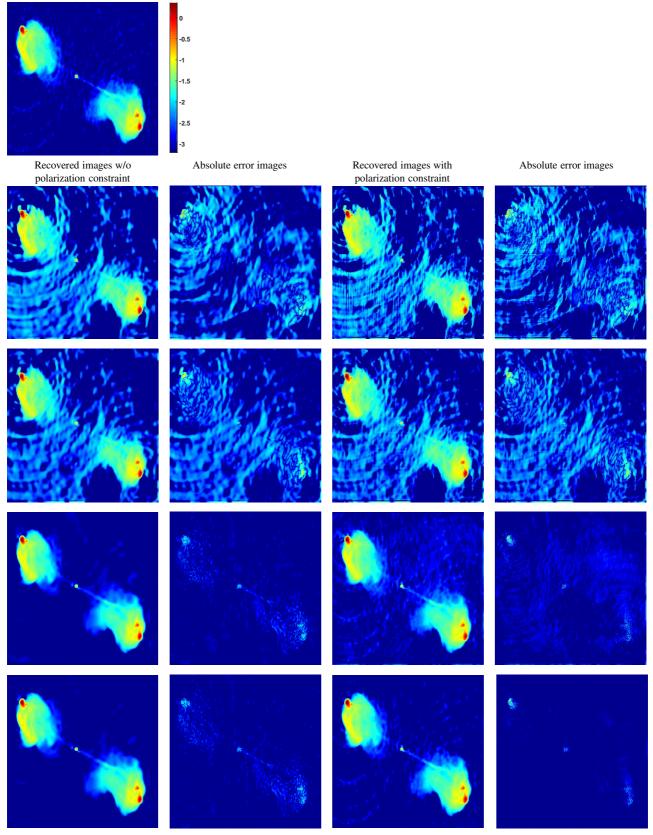


Figure 4. Cygnus A Stokes *I* true image in first row and reconstructed images (best ones over 5 performed simulations for each case) in other rows for the cases: Imaging with normalized DIEs (second row), Joint DIE calibration and imaging (third row), Joint DDE calibration and imaging excluding the off-diagonal terms (fourth row), and considering full Jones matrix (fifth row). In each case, column-wise recovered images followed by their corresponding error images are displayed when imaging is performed without polarization constraint (first two columns) and with polarization constraint (last two columns). All the images are shown in log scale, with the same color range corresponding to the colorbar given in first row.

14 J. Birdi et al.

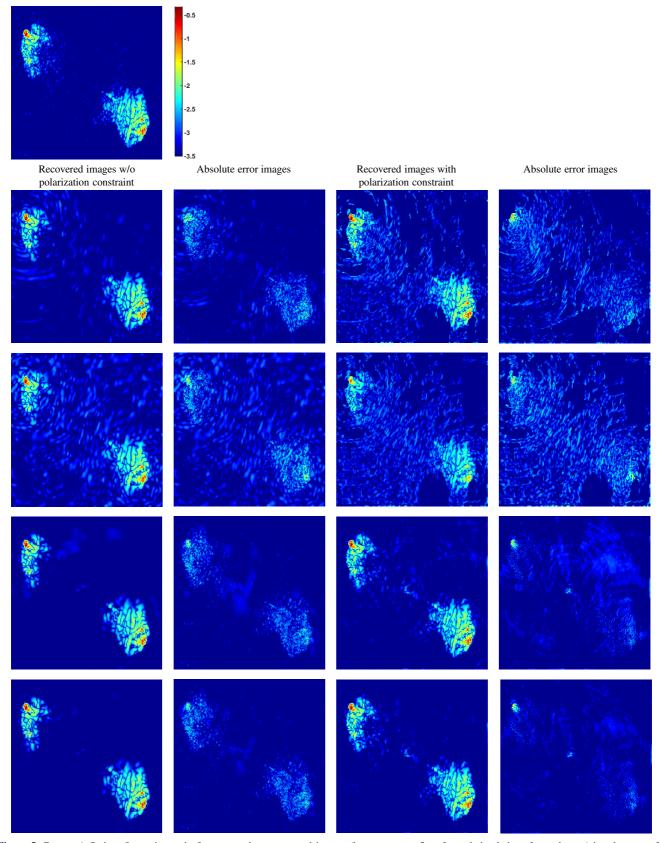


Figure 5. Cygnus A Stokes Q true image in first row and reconstructed images (best ones over 5 performed simulations for each case) in other rows for the cases: Imaging with normalized DIEs (second row), Joint DIE calibration and imaging (third row), Joint DDE calibration and imaging excluding the off-diagonal terms (fourth row), and considering full Jones matrix (fifth row). In each case, column-wise recovered images followed by their corresponding error images are displayed when imaging is performed without polarization constraint (first two columns) and with polarization constraint (last two columns). All the images are shown in log scale, with the same color range corresponding to the colorbar given in first row.

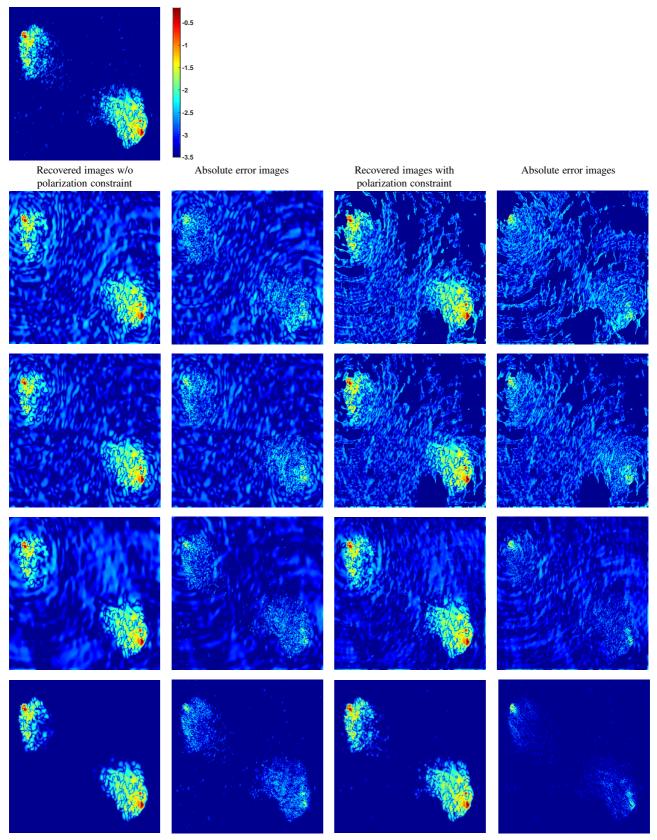


Figure 6. Cygnus A Stokes U true image in first row and reconstructed images (best ones over 5 performed simulations for each case) in other rows for the cases: Imaging with normalized DIEs (second row), Joint DIE calibration and imaging (third row), Joint DDE calibration and imaging excluding the off-diagonal terms (fourth row), and considering full Jones matrix (fifth row). In each case, column-wise recovered images followed by their corresponding error images are displayed when imaging is performed without polarization constraint (first two columns) and with polarization constraint (last two columns). All the images are shown in log scale, with the same color range corresponding to the colorbar given in first row.

7.3 Results and analysis

We now present the results obtained by conducting tests for the various cases mentioned before. In terms of quantitative comparison between the different cases performed, Table 1 provides the obtained SNR and dynamic range values for both the sets of images. In each case, the mean values evaluated over the 5 performed simulations are shown. From the calibration front, it can be noticed that joint DDE calibration and imaging leads to a significant improvement in the reconstruction quality, both in terms of SNR and DR, in comparison with either of the cases when only imaging step is performed considering normalized DIEs or jointly calibrating for DIEs and imaging. In particular, the case of normalized DIEs considers Jones matrices to be identity, whereas the DIE calibration scheme also accounts for the DIEs in the off-diagonal Jones terms and is thus further affected by the estimation of these terms. In fact, these terms lead to flux leakage from one Stokes parameter to others and if not perfectly calibrated, can lead to error propagation. It particularly affects the low amplitude Stokes Q and U images. This can be seen from the smaller SNR values of Stokes Q image than that obtained by the case of normalized DIEs (Table 1(a)). Another observation that can be made for these two cases is regarding the higher DR values obtained for Stokes Q and U images than that of Stokes I image. This non-physical feature can be attributed to the usage of incomplete measurement model (i.e. without the incorporation of the DDEs), severely affecting the underlying non-convex minimization problem giving inaccurate solutions. The aforementioned observations for these cases further indicate the necessity of taking direction-dependent terms into account and performing their calibration along with imaging for a better reconstruction quality. Second, comparing the results obtained with and without calibrating for the off-diagonal terms in the Jones matrices shows better SNR and DR values are achieved in the former case, thereby demonstrating the significance of calibrating for the off-diagonal Jones terms to recover high quality, high dynamic range images.

On the other hand, on the imaging front, it is interesting to note that when the image recovery is performed either using only normalized DIEs or calibrating for DIEs, i.e. having inexact knowledge of the Jones matrices and neglecting the DDEs, regularizing the underlying problem with the polarization constraint may not be very effective. Indeed, in such a case, not enforcing the constraint can produce slightly better results, as seen for Cygnus A images. Nevertheless, when the DDE calibration scheme is incorporated, the result analysis highlights the superior performance of this constraint, recovering images with much higher SNR and DR compared to the case where it is not imposed. Especially in the case of model images for Hydra A, the enforcement of the polarization constraint leads to an appreciable improvement of $\sim 9-10$ dBs in SNR of Stokes Q and U images (Table 1(c)) and around one order of magnitude in DR (Table 1(d)). It should be noted that Hydra A images have lesser amplitude than the first set of images (Cygnus A). This is true specifically for Stokes Q and U images which are around one order of magnitude lower in amplitude than the corresponding Stokes I image and hence, difficult to be recovered. In this respect, the obtained results emphasize the crucial role played by suitably chosen regularization prior in enhancing the reconstruction quality. In fact, if the problem is not well regularized (i.e. the polarization constraint is not enforced), the solution of the underlying non-convex joint DDE calibration and imaging problem may stuck in a local minimum leading to poor reconstruction quality. The good performance of the polarization constraint to recover these images shows suitability of this prior for full polarization imaging. This validates the findings of Birdi et al. (2018b) and further extends them to the case of joint calibration and imaging.

The above listed observations are further supported by the visual inspection of the recovered images and the associated absolute error images, shown in Figs. 4-9. While Figs. 4, 5 and 6 display the Cygnus A images for Stokes I, Q and U, respectively, the corresponding images for Hydra A are shown in Figs. 7, 8 and 9. The error images are obtained by computing the absolute difference between the ground truth and reconstructed images. In each of these figures, the recovered images (first and third columns) followed by their absolute error images (second and fourth columns) are shown. The shown images correspond to the reconstructions (and the associated error images) obtained when imaging is performed with: (second row) normalized DIEs, (1.a) without polarization constraint and (1.b) with polarization constraint, and (third row) DIE calibration (2.a) without and (2.b) with the enforcement of the polarization constraint. Similarly, the reconstructed images (and the associated error images) for the case of joint DDE calibration and imaging, excluding the off-diagonal terms and performing full Jones matrix calibration are presented respectively in third and fourth rows, both for (3.a) (and (3.c)) without polarization constraint and (3.b) (and (3.d)) with polarization constraint. It can be observed that joint DDE calibration and imaging offers remarkable advantage over considering only DIEs, mitigating the artefacts occurring because of the calibration errors. Moreover, calibration of the off-diagonal terms is crucial for high dynamic range imaging by diminishing the diffused calibration artefacts in the background. In addition to these remarks, the quality of reconstruction is promoted by considering the polarization constraint in the reconstruction process. While the enforcement of this constraint yields lesser residual in the error images, a careful examination of the recovered images also indicates that this prior is able to produce highly resolved images with finer details as opposed to the case when the constraint is not taken into account.

8 CONCLUSIONS

We have presented a joint calibration and imaging technique taking into account the full polarization model for radio interferometry. The proposed technique, dubbed Polca SARA, unifies the estimation of the DDEs for the full Jones matrix and the Stokes images of interest within a global algorithmic structure, exploiting the same optimization framework for both calibration and imaging. In particular, it solves the underlying non-convex minimization problem employing a block-coordinate forward-backward algorithm, thereby following a forward-backward scheme for estimation of each of the variables. The MATLAB code of the proposed method will soon be made available on GitHub (https://baspgroup.github.io/Polca-SARA/).

While our approach is shipped with convergence guarantees, it can also be adapted to incorporate suitable regularization priors for the variables under consideration. Thanks to this flexibility, for the imaging step in the global algorithm, we have employed an approach specifically developed for full Stokes imaging enforcing the physical polarization constraint (Birdi et al. 2018b). These features offered by our method are in contrast with the existing calibration and imaging algorithms in RI which (i) do not benefit from global convergence and (ii) use in fact Stokes *I* imaging based techniques even for polarimetric imaging. The significance of the latter remark is further highlighted by the results obtained from various numerical simulations performed, achieving better SNR and higher dynamic range with the enforcement of the polarization constraint.

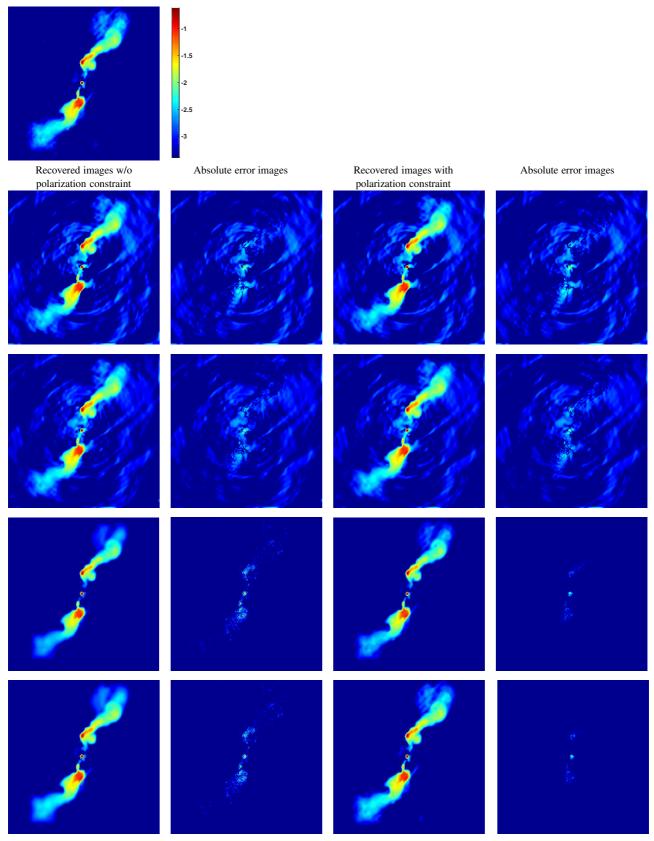


Figure 7. Hydra A Stokes *I* true image in first row and reconstructed images (best ones over 5 performed simulations for each case) in other rows for the cases: Imaging with normalized DIEs (second row), Joint DIE calibration and imaging (third row), Joint DDE calibration and imaging excluding the off-diagonal terms (fourth row), and considering full Jones matrix (fifth row). In each case, column-wise recovered images followed by their corresponding error images are displayed when imaging is performed without polarization constraint (first two columns) and with polarization constraint (last two columns). All the images are shown in log scale, with the same color range corresponding to the colorbar given in first row.

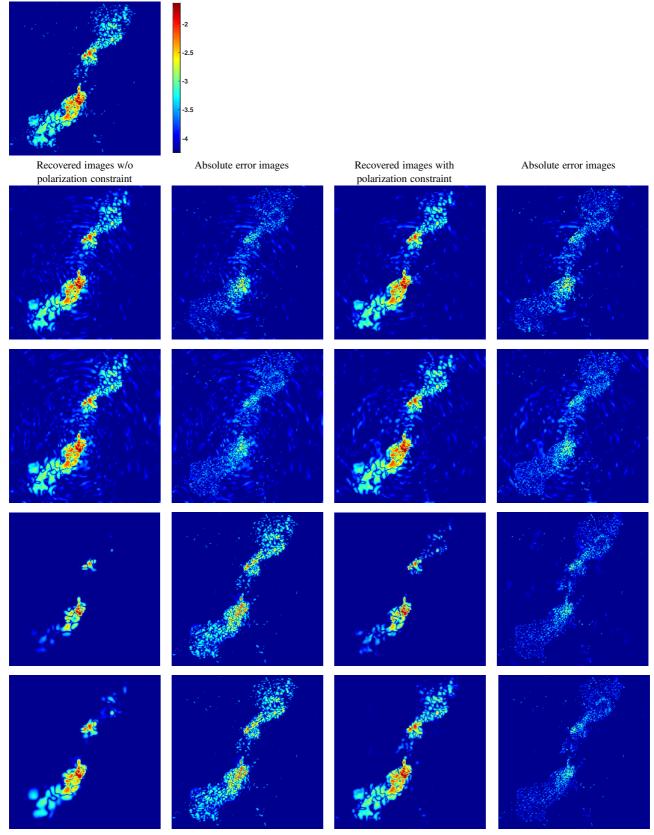


Figure 8. Hydra A Stokes Q true image in first row and reconstructed images (best ones over 5 performed simulations for each case) in other rows for the cases: Imaging with normalized DIEs (second row), Joint DIE calibration and imaging (third row), Joint DDE calibration and imaging excluding the off-diagonal terms (fourth row), and considering full Jones matrix (fifth row). In each case, column-wise recovered images followed by their corresponding error images are displayed when imaging is performed without polarization constraint (first two columns) and with polarization constraint (last two columns). All the images are shown in log scale, with the same color range corresponding to the colorbar given in first row.

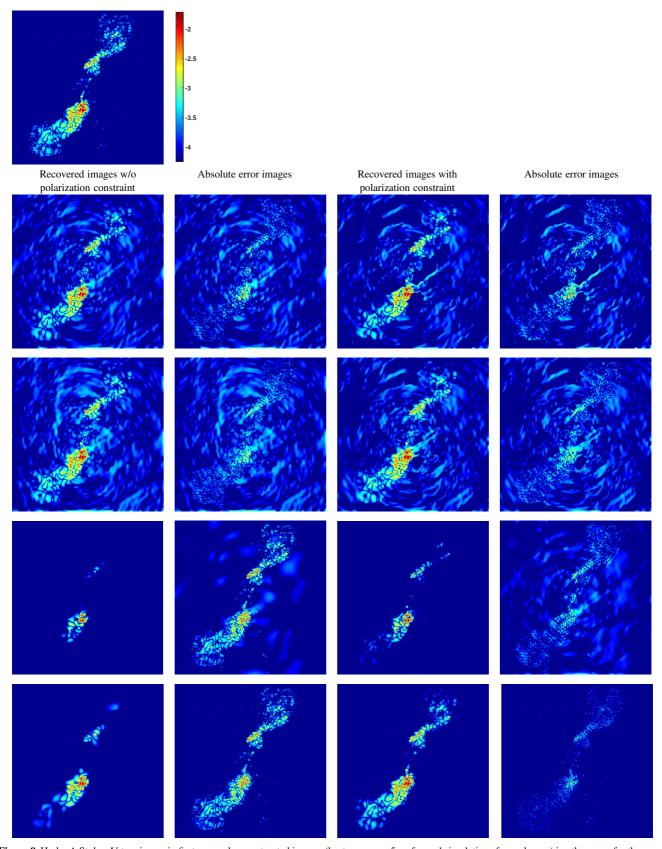


Figure 9. Hydra A Stokes U true image in first row and reconstructed images (best ones over 5 performed simulations for each case) in other rows for the cases: Imaging with normalized DIEs (second row), Joint DIE calibration and imaging (third row), Joint DDE calibration and imaging excluding the off-diagonal terms (fourth row), and considering full Jones matrix (fifth row). In each case, column-wise recovered images followed by their corresponding error images are displayed when imaging is performed without polarization constraint (first two columns) and with polarization constraint (last two columns). All the images are shown in log scale, with the same color range corresponding to the colorbar given in first row.

Furthermore, in terms of the calibration, the results have shown the importance of calibrating for full Jones matrix, including DDEs and off-diagonal terms, to mitigate the artefacts appearing otherwise in the reconstructed images. Calibration of off-diagonal terms in Jones matrices, denoting the polarization leakage, has been shown to be particularly crucial to obtain high dynamic range images.

Although in this first presentation of the proposed algorithm we have dealt with simulated data, the good performance obtained by this approach unveil it as a promising joint calibration and imaging technique in RI. Keeping this in mind, we plan to investigate its performance on real RI data sets in near future. We further note that in practice, the polarization angle deduced from reconstructed Stokes Q and U images can suffer from a unitary rotational ambiguity in the measurement model (Hamaker 2000; Smirnov 2011). A way to tackle this problem could be to add in the minimization problem some prior information about this rotation, that can be obtained from external calibration.

Another possible direction of work is to extend the proposed framework to account for hyperspectral data. While on the calibration part it would involve incorporation of DIEs/DDEs variation with observation frequency (Dabbech et al. 2019), the imaging part would need to be extended for wide-band polarimetric imaging, leveraging the adaptability offered by the currently developed approach.

ACKNOWLEDGEMENTS

We would like to thank R. Perley from NRAO for providing the two sets of images used for simulation studies. This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC, grant EP/M011089/1).

References

Akiyama K., et al., 2017, ApJ, 838, 1

Bhatnagar S., Cornwell T. J., 2004, A&A, 426, 747

Bhatnagar S., Cornwell T., 2017, The Astronomical Journal, 154, 197

Bhatnagar S., Rau U., Golap K., 2013, ApJ, 770, 91

Birdi J., Repetti A., Wiaux Y., 2018a, in 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM). IEEE, pp 465–469

Birdi J., Repetti A., Wiaux Y., 2018b, MNRAS, 478, 4442

Candès E., Romberg J., Tao T., 2006, IEEE Trans. Inf. Theory, 52, 489

Candès E., Wakin M., Boyd S., 2008, J. Fourier Anal. Appl., 14, 877

Carrillo R., McEwen J., Wiaux Y., 2012, MNRAS, 426, 1223

Carrillo R., McEwen J., Van De Ville D., Thiran J.-P., Wiaux Y., 2013, IEEE Signal Process. Lett., 20, 591

Carrillo R., McEwen J., Wiaux Y., 2014, MNRAS, 439, 3591

Chaux C., Combettes P. L., Pesquet J.-C., Wajs V. R., 2007, Inverse Problems, 23, 1495

Chen S. S., Donoho D. L., Saunders M. A., 2001, SIAM Rev., 43, 129

Chierchia G., Pustelnik N., Pesquet J.-C., Pesquet-Popescu B., 2015, Sig. Image Vid. Process., 9, 1737

Chouzenoux E., Pesquet J.-C., Repetti A., 2016, J. Global Optim., pp 1–29Combettes P. L., Pesquet J.-C., 2011, Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer, New York

Combettes P. L., Duñg D., Vũ B. C., 2011, J. Math. Anal. Appl., 380, 680 Condat L., 2013, J. Opt. Theory Appl., 158, 460

Cornwell T. J., 2008, IEEE J. Sel. Top. Sig. Process., 2, 793

Dabbech A., Onose A., Abdulaziz A., Perley R. A., Smirnov O. M., Wiaux Y., 2018, MNRAS, 476, 2853

Dabbech A., Repetti A., Wiaux Y., 2019, in International BASP Frontiers workshop 2019.

Donoho D. L., 2006, IEEE Trans. Inf. Theory, 52, 1289

Donoho D., Johnstone I., Kerkyacharian G., Picard D., 1995, J. R. Stat. Soc. Ser. B, 57, 301

Fessler J. A., Sutton B. P., 2003, IEEE Trans. Signal Process., 51,560 Garsden H., et al., 2015, A&A, 575, A90

Ginzburg V., Syrovatskii S., 1965, Ann. Rev. Astron. Astrophys., 3, 297 Hamaker J. P., 2000, A&A Supp.Ser., 143, 515

Hamaker J., Bregman J., Sault R., 1996, A&A Supp.Ser., 117, 137 Högbom J. A., 1974, A&A, 15, 417

Intema H., Van der Tol S., Cotton W., Cohen A., Van Bemmel I., Röttgering H., 2009, A&A, 501, 1185

Kazemi S., Yatawatta S., Zaroubi S., Lampropoulos P., De Bruyn A., Koopmans L., Noordam J., 2011, MNRAS, 414, 1656

Mallat S. G., 2009, A wavelet tour of signal processing : the Sparse way., 2nd edn. Academic Press, Burlington, MA

Moré J., 1978, The Levenberg-Marquardt algorithm: implementation and theory. Springer, Berlin, Heidelberg

Moreau J.-J., 1965, Bull. Soc. Math. France, 93, 273

Noordam J. E., 2004, in Ground-Based Telescopes. pp 817-826

Onose A., Carrillo R., Repetti A., McEwen J., Thiran J., Pesquet J.-C., Wiaux Y., 2016, MNRAS, 462, 4314

Onose A., Dabbech A., Wiaux Y., 2017, MNRAS, 469, 938

Pacholczyk A., 1970, Radio astrophysics: Nonthermal processes in galactic and extragalactic sources. Freeman, San Francisco

Pesquet J.-C., Repetti A., 2015, J. Nonlinear Convex Anal., 16, 2453

Pratley L., Johnston-Hollitt M., 2016, MNRAS, 462, 3483

Pratley L., McEwen J. D., d'Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2017, MNRAS, 473, 1038

Repetti A., Wiaux Y., 2017, in Wavelets and Sparsity XVII 2017. International Society for Optics and Photonics, p. 103941W

Repetti A., Birdi J., Dabbech A., Wiaux Y., 2017, MNRAS, 470, 3981

Rudin L., Osher S., Fatemi E., 1992, Phys. D: Nonlinear Phenomena, 60, 259

Salvini S., Wijnholds S., 2014, A&A, 571, A97

Smirnov O., 2011, A&A, 527, A107

Thompson A., Moran J., Swenson G., 2001, Interferometry and Synthesis in Radio Astronomy. Wiley-Interscience, New York

Thouvenin P.-A., Repetti A., Dabbech A., Wiaux Y., 2018, in 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM). IEEE, pp 475–479

Van Weeren R., et al., 2016, The Astrophysical Journal Supplement Series, 223, 2

Vũ B., 2013, Adv. Comp. Math., 38, 667

Wiaux Y., Jacques L., Puy G., Scaife A., Vandergheynst P., 2009, MNRAS, 395, 1733

Wiaux Y., Puy G., Vandergheynst P., 2010, MNRAS, 402, 2626

Yatawatta S., Zaroubi S., De Bruyn A., Koopmans L., Noordam J., 2009, in DSP Workshop and 5th IEEE SPE Workshop, 2009.. pp 150–155

APPENDIX A: TABLE OF NOTATIONS

A list of useful notations and algorithmic parameters to be configured are provided in Table A1.

This paper has been typeset from a $T_EX/I \Delta T_EX$ file prepared by the author.

Stokes matrix containing the high amplitude,
thresholded part of S obtained with 1GC
Stokes matrix containing the unknown part of S
DDEs for antenna α and time instant t
Compact-support Fourier kernels corresponding
to $\overline{D}_{t,\alpha}$
Concatenation of matrices $(\overline{\mathbf{U}}_{\alpha,1})_{1 \leq \alpha \leq n_a}$
$(\text{and } (\overline{\mathbf{U}}_{\alpha,2})_{1\leqslant \alpha\leqslant n_{\alpha}}))$
Support sizes in spatial and temporal
Fourier domains of DDEs, respectively
Number of inner-loop FB iterations for DDEs and
images updates, respectively
Regularization parameters for the images
Regularization parameter for the DDEs
Stopping criterion for the DDEs and images
update, respectively
Global stopping criterion for Algorithm 1

 $\begin{tabular}{ll} \textbf{Table A1.} Important \ variable \ notations \ and \ configurable \ algorithmic \ parameters for Polca SARA \ method \ (Algorithm 1). \end{tabular}$