Astronomy & Astrophysics manuscript no. paper March 25, 2014
(DOI: will be inserted by hand later)

Complex optimisation approach to radio interferometric calibration

N. Bourbaki

Address(es) of author(s) should be given

Received ¡date¿ / Accepted ¡date¿

Abstract.

1. Complex optimisation

In this section we try to describe the shape and structure of the Jacobian over the complex scalar field, instead of what is usually done, by considering the real and imaginary separatly.

1.1. RIME formalism

Using the Radio Interferometry Measurement Equation (RIME) formalism, we can write the 4-polarisation visibility vector $\mathbf{v}_{pq}$ measured on baseline (pq), at time t and frequency ν as:

$$\mathbf{v}_{pq} = \mathrm{Vec}\left(\mathbf{V}_{(pq)t\nu}\right) \tag{1}$$

$$= \sum_d \left(\overline{\mathbf{J}^d_{qt\nu}} \otimes \mathbf{J}^d_{pt\nu}\right) \mathrm{Vec}\left(\mathbf{S}_d\right) k^d_{(pq)t\nu} \tag{2}$$

$$\text{with } k^d_{(pq)t\nu} = \exp\left(-2i\pi\left(ul + vm + w\left(\mathrm{n}-1\right)\right)\right) \tag{3}$$

$$\text{and } n = \sqrt{1-l^2-m^2} \tag{4}$$

with $[u,v,w]^T$ is the baseline vector between antennas p and q in wavelength units, and $\boldsymbol{s}_d = [l,m,n = \sqrt{1-l^2-m^2}]^T$ is a sky direction later labeled as d.

From Eq. 1, we can write the i^{th} polarisation component of $\mathbf{v}_{pq}$ as:

$$\mathrm{v}^i_{pqt\nu} = \sum_d \sum_{j=0}^{3} \left(g^d_{pt\nu,\mathbf{A}_{ij}} . \overline{g^d_{qt\nu,\mathbf{B}_{ij}}}\right) . k^d_{pqt\nu} . \mathrm{s}_{d,j} \tag{5}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 2 & 0 & 2 \\ 1 & 3 & 1 & 3 \\ 0 & 2 & 0 & 2 \\ 1 & 3 & 1 & 3 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 1 & 1 & 3 & 3 \\ 1 & 1 & 3 & 3 \end{bmatrix} \tag{6}$$

1.2. Complex derivative

In order to compute a Jacobian, we need to choose a derivative definition for complex numbers. In this section, instead of differentiating against real and imaginary parts independently, we adopt a Wirtinger differentiation point of view and consider the complex and their conjugate as being indepedent. We will see in Sec. 3 that choosing this type of differentiation turns out to be rather powerful to solve problems of the form of Eq. 1. If we write a complex number as $z = x + iy$, then the Wirtinger complex derivative operator writes as:

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \tag{7}$$

$$\text{and } \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) \tag{8}$$

where x and y are the real and imaginary parts respectively. The Wirtinger has a trivial but remarquable property that a scalar and its complex conjugate can be viewed as independent variables, and in particular:

$$\frac{\partial \overline{z}}{\partial z} = 0 \text{ and } \frac{\partial z}{\partial \overline{z}} = 0 \tag{9}$$

Considering the sky, gain, and geometry relation given in Eq. 1, according to the property of Wirtinger derivative of complex conjugate (Eq. 9), we can see that:

$$\frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial g^d_{pt\nu,\mathbf{A}_{ij}}} = \left(\overline{g^d_{qt\nu,\mathbf{B}_{ij}}} \mathrm{s}_{d,j}\right) . k^d_{pqt\nu} \tag{10}$$

$$\text{and } \frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial g^d_{qt\nu,\mathbf{B}_{ij}}} = 0 \tag{11}$$

while differentiating against the complex conjugate of those variables, we obtain:

$$\frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial\left(\overline{g^d_{pt\nu,\mathbf{A}_{ij}}}\right)} = 0 \tag{12}$$

$$\text{and } \frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial\left(\overline{g^d_{qt\nu,\mathbf{B}_{ij}}}\right)} = \left(g^d_{pt\nu,\mathbf{A}_{ij}} \mathrm{s}_{d,j}\right).k^d_{pqt\nu} \tag{13}$$

Interestinggly, we can see from Eq. 10, 11, 12 and 13 that the derivatives are always constant with respect to the differential variable.

1.3. Complex Jacobian

In this section, we describe the structure of the Wirtinger Jacobian $\boldsymbol{\mathcal{J}}_{\mathbf{v}}$ using the results of Sec. 1.2. First, we consider the visibility vector $\mathbf{v}_{pq}$ for all given time frequency blocks, and writing the antenna, polarisation, and direction dependent gain vector as $\boldsymbol{g}$. Its size is $4n_an_d$, and have for i^{th} component $i = j + 4 \times d + 4 \times a \times n_d$ the gain of antenna a in direction d for polarisation j, where n_a, and n_d are the number of antenna and directions.

We construct the corresponding Wirtinger Jacobian $\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}_{\boldsymbol{W}}}$ of size $(4n_tn_\nu) \times (8n_an_d)$ (n_t and n_ν are the number of time and frequency points) as follows:

$$\boldsymbol{d}\mathbf{v}_{pq} = \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}_{\boldsymbol{W}}}\ \boldsymbol{dg}_{\boldsymbol{W}} \tag{14}$$

$$= \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}}\ \boldsymbol{dg} + \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\overline{\boldsymbol{g}}}\ \boldsymbol{d\overline{g}} \tag{15}$$

$$\text{where } \boldsymbol{dg}_{\boldsymbol{W}} = \begin{bmatrix} \boldsymbol{dg} \\ \boldsymbol{d\overline{g}} \end{bmatrix} \tag{16}$$

$$\text{and } \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}_{\boldsymbol{W}}} = \begin{bmatrix} \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}} & \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\overline{\boldsymbol{g}}} \end{bmatrix} \tag{17}$$

Each cell of $\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}_{\boldsymbol{W}}}$, $\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}}$ and $\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\overline{\boldsymbol{g}}}$ can be written using Eq. 10, 11, 12 and 13. Specifically, line corresponds to a single i-polarisation measurement at $(t\nu)$ for the (pq) baseline, and a column $j + 4d + a.4n_d$ to a gain for polarisation j, antenna a and direction d ($g^d_{a,j}$). The matrix $\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}_{\boldsymbol{W}}}$ can be described as follows:

$$\left[\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}}\right]_{t\nu,i} = \begin{cases} \left(\overline{g^d_{qt\nu,\mathbf{B}_{ij}}} \mathrm{s}_{d,j}\right).k^d_{pqt\nu} \text{ for } a = p \\ 0 \text{ otherwise} \end{cases} \tag{18}$$

and

$$\left[\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\overline{\boldsymbol{g}}}\right]_{t\nu,i} = \begin{cases} \left(g^d_{pt\nu,\mathbf{A}_{ij}} \mathrm{s}_{d,j}\right).k^d_{pqt\nu} \text{ for } a = q \\ 0 \text{ otherwise} \end{cases} \tag{19}$$

We can see that non-zero columns are the ones corresponding to all direction and all polarisations for antenna p. The Jacobian for all baselines is written in a similar way, by superposing the $\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}_{\boldsymbol{W}}}$ for all (pq) pairs as follows:

$$\boldsymbol{\mathcal{J}}_{\mathbf{v}} = \begin{bmatrix} \vdots \\ \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}_{\boldsymbol{W}}} \\ \vdots \end{bmatrix} \tag{20}$$

which have size $[(4n_{bl}n_tn_\nu) \times (8n_an_d)]$, where n_{bl} is the number of baselines and is typically $n_{bl} = n_a(n_a - 1)/2$. Although it has large dimensions, $\boldsymbol{\mathcal{J}}_{\mathbf{v}}$ is sparse.

2. Iterative direction-dependent solver

2.1. Linear system for calibration

We find the remarquable property that around the solution, Eq. 1 behaves like a linear system. Specifically, from Eq. 19 and 20, it is easy to check that:

$$\mathbf{v} = \frac{1}{2}\boldsymbol{\mathcal{J}}_{\mathbf{v}}\ \boldsymbol{g}_{\boldsymbol{W}} \tag{21}$$

$$= \boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}\ \boldsymbol{g} = \boldsymbol{\mathcal{J}}_{\mathbf{v},\overline{\boldsymbol{g}}}\ \overline{\boldsymbol{g}} \tag{22}$$

Both Eq. 21 and 22 form linear systems.

TO BE MODIFIED AFTER THAT

Assuming a linear operator satisfying Eq. 21 is given, we can build an iterative scheme to derive an estimate $\widehat{\boldsymbol{g}}$ of $\boldsymbol{g}$. The linear operator $\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}$ is build from the estimate $\widehat{\boldsymbol{g}}_i$ at step i. Then Eq. 21 is solved using the least-squares solution given by computing pseudo-inverse as follows:

$$\widehat{\boldsymbol{g}_{i+1}} = \left[\mathbf{A}^H\mathbf{C}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}^H\mathbf{C}^{-1}\mathbf{v} \tag{23}$$

$$\text{with } \mathbf{A} = \boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}} \tag{24}$$

where $\mathbf{C}$ is the covariance matrix of $\mathbf{v}$.

2.2. Convergence and averaging

As shown in Fig. ... the convergence of this algorithm is slow, and following Stef-the-great, instead of estimating $\boldsymbol{\mathcal{J}}_{\mathbf{v}}$ at $\widehat{\boldsymbol{g}}_i$, we build it at a modified location constructed from previous iterations, and Eq. 24 becomes:

$$\mathbf{A} = \boldsymbol{\mathcal{J}}_{\mathbf{v}}\big|_{\widetilde{\boldsymbol{g}}_i} \tag{25}$$

$$\text{with } \widetilde{\boldsymbol{g}}_i = (\widehat{\boldsymbol{g}_{i-1}} + \widehat{\boldsymbol{g}}_i)/2 \tag{26}$$

3. Connection with other existing algorithms

In this section, we investigate the connections between the complex optimisation point of view (see Sec. 1 and 2) and the existing algorithm. We show that antenna based calibration schemes are always obtainable using this framework, while general pealing based approaches are appliable under certain condition of orthogonality between Fourier kernels.

3.1. Levenberg-Maquardt and pealing

3.2. StefCal

4. Test on real data

We test the algorithm described above on a LOFAR dataset. The visibilities produced by this interferometer

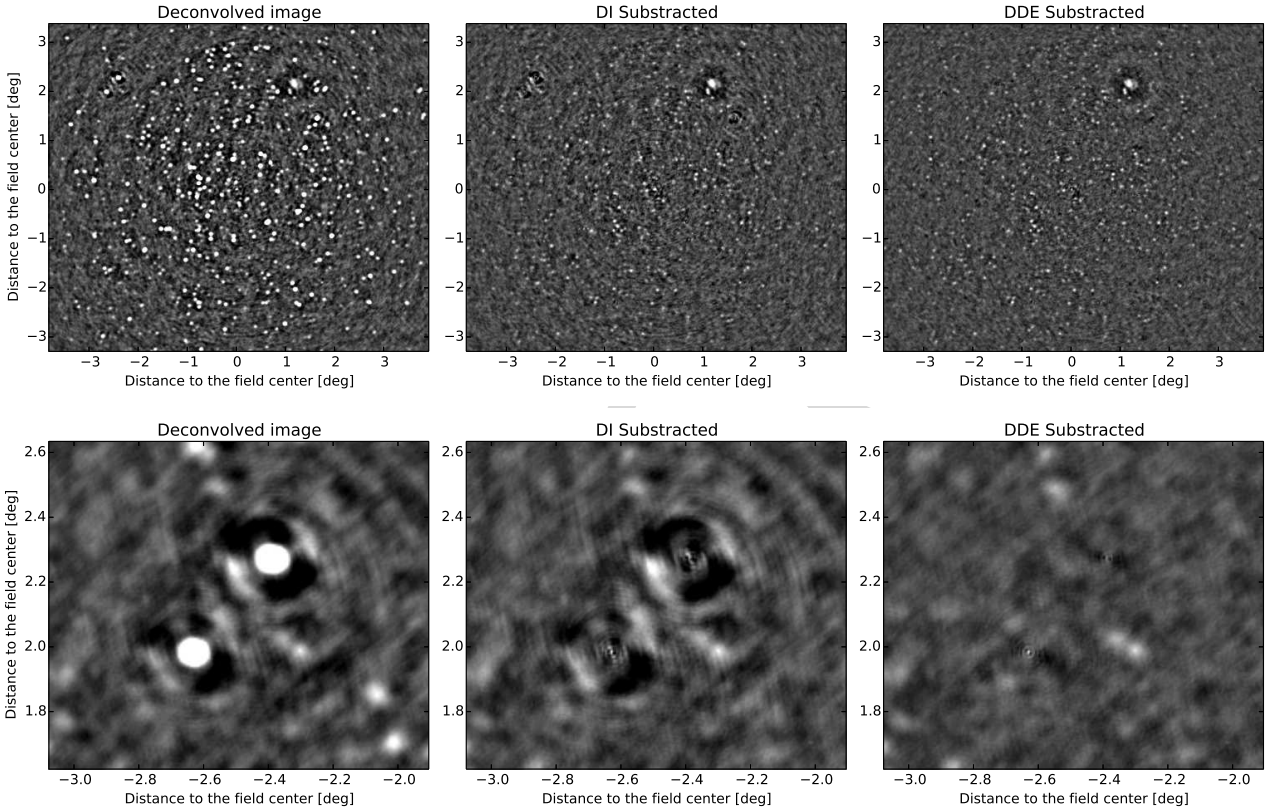

Fig. 1. This figure shows compares the image (left), the residuals data after simple skymodel substraction (center), and the residuals data after substracting the sky model corrupted by the direction-dependent solution (right).

are predominantly affected by direction dependent effects including (i) the phased beam instability and deviation from the theoritical model, (ii) ionosphere time delays shifts, and (iii) Faraday rotation.

We first calibrate the data using BBS, and in order to build a pertinent model of the field, we substract 3C295. We extract the sources using pyBDSM. The sources are the clustered in 10 directions using Voronoi tesselation (fig. 2). In Fig 1, we compare the residuals as computed by substracting the model data in the visibility domain, and the model data affected by DDEs.

References

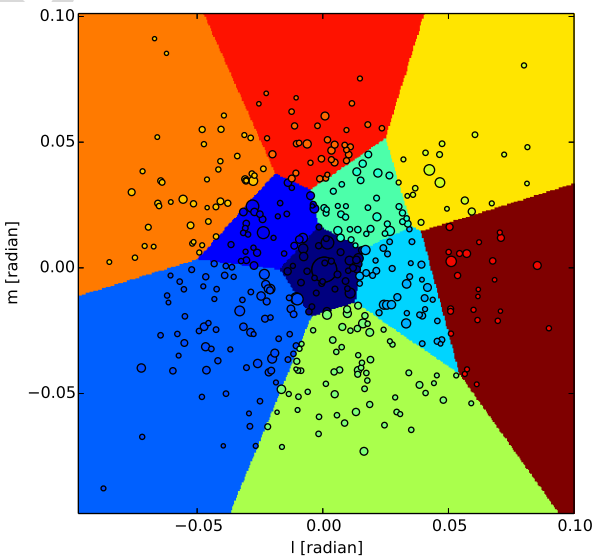

Fig. 2. In order to minimize the number of degrees of freedom, and increase the amount of signal in each direction, we cluster the sources in 10 direction using a Voronoi tesselation.