Applying Wirtinger derivatives to the radio interferometry calibration problem

C. Tasse[1,2]

[1] GEPI, Observatoire de Paris, CNRS, Université Paris Diderot, 5 place Jules Janssen, 92190 Meudon, France
[2] Department of Physics & Electronics, Rhodes University, PO Box 94, Grahamstown, 6140, South Africa

Abstract. This paper presents a fast algorithm for full-polarisation, direction dependent calibration in radio interferometry. It is based on Wirtinger's approach to complex differentiation. Compared to the classical case, the Jacobian appearing in the Levenberg-Maquardt iterative scheme presents a sparser structure, allowing for an algorithmic cost gain corresponding to the number of antenna squared. This improvement is significant and varies between $\sim 10^3$ for VLA or LOFAR to $\sim 10^{6-7}$ for the SKA.

1. Complex optimisation

Given a set of cross correlations between antenna voltages (the visibility set) in a given time-frequency domain and a model of the sky brightness, the calibration step in radio interferometry consists in estimating a set of direction-dependent Jones matrices per antenna. This is usually done using a chi-square minimisation technique such as the Levenberg-Maquardt. In this paper, I use the alternative Wirtinger's definition of complex derivative, instead of what is usually done by considering the real and imaginary (later referred as the *Wirtinger* and *classical* approaches respectively).

1.1. RIME formalism

Using the Radio Interferometry Measurement Equation (RIME) formalism (Hamaker et al. 1996), the 4-polarisation visibility vector $\mathbf{v}_{pq}$ measured on baseline (pq), at time t and frequency ν can be written as

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

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

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

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

where $\mathbf{J}^d_{pt\nu}$ is the Jones matrix of antenna p in direction d, $\mathbf{S}_d$ is the four-polarization sky matrix, $\otimes$ is the Kronecker product, $[u,v,w]^T$ 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. In the following, the Jones matrix $\mathbf{J}^d_{pt\nu}$ is represented by scalars $g^d_{pt\nu,k}$ as

$$\mathbf{J}^d_{pt\nu} = \begin{bmatrix} g^d_{pt\nu,0} & g^d_{pt\nu,2} \\ g^d_{pt\nu,1} & g^d_{pt\nu,3} \end{bmatrix} \quad (5)$$

and the set of Jones matrices is represented by a gain vector $\boldsymbol{g}$ containing the scalars $g^d_{pt\nu,k}$ for all directions, antennas and polarisations.

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

$$\mathrm{v}^i_{pqt\nu} = h^i_{pq}(\boldsymbol{g}) \quad (6)$$

$$= \sum_d \sum_{j=0}^{3} \left(g^d_{pt\nu,\mathbf{A}_{ij}} . \overline{g^d_{qt\nu,\mathbf{B}_{ij}}}\right) . k^d_{pqt\nu} . \mathrm{s}_{d,j} \quad (7)$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 2 & 0 & 2 \\ 1 & 3 & 1 & 3 \\ 0 & 2 & 0 & 2 \\ 1 & 3 & 1 & 3 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 1 & 1 & 3 & 3 \\ 1 & 1 & 3 & 3 \end{bmatrix} \quad (8)$$

1.2. Wirtinger complex derivative

In order to compute a Jacobian, a derivative definition for complex numbers has to be chosen. Instead of differentiating against real and imaginary parts independently, one can adopt a Wirtinger differentiation point of view and consider the complex and their conjugate as being independent. Choosing this type of differentiation turns out to be rather powerful to solve problems of the form of Eq. 1 (see Sec. 2). If a complex number is written as $z = x + iy$, the Wirtinger complex derivative operator becomes

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \quad (9)$$

$$\text{and } \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) \quad (10)$$

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

$$\frac{\partial \overline{z}}{\partial z} = 0 \text{ and } \frac{\partial z}{\partial \overline{z}} = 0 \quad (11)$$

Considering the sky, gain, and geometry relation given in Eq. 1, according to the property of Wirtinger derivative of complex conjugate (Eq. 11)

$$\frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial g^d_{pt\nu,\mathbf{A}_{ij}}} = \left(\overline{g^d_{qt\nu,\mathbf{B}_{ij}}}\mathrm{s}_{d,j}\right).k^d_{pqt\nu} \quad (12)$$

$$\text{and } \frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial g^d_{qt\nu,\mathbf{B}_{ij}}} = 0 \quad (13)$$

while differentiating against the complex conjugate of those variables, one obtain

$$\frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial \left(\overline{g^d_{pt\nu,\mathbf{A}_{ij}}}\right)} = 0 \quad (14)$$

$$\text{and } \frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial \left(\overline{g^d_{qt\nu,\mathbf{B}_{ij}}}\right)} = \left(g^d_{pt\nu,\mathbf{A}_{ij}}\mathrm{s}_{d,j}\right).k^d_{pqt\nu} \quad (15)$$

Interestingly, Eq. 12, 13, 14 and 15 show that the derivatives are always constant with respect to the differential variable.

1.3. Wirtinger Jacobian

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

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

$$\boldsymbol{d}\mathbf{v}_{pq} = \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}_{\boldsymbol{W}}}\ \boldsymbol{dg}_{\boldsymbol{W}} \quad (16)$$

$$= \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}}\ \boldsymbol{dg} + \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\overline{\boldsymbol{g}}}\ \boldsymbol{d\overline{g}} \quad (17)$$

$$\text{where } \boldsymbol{dg}_{\boldsymbol{W}} = \begin{bmatrix}\boldsymbol{dg}\\ \boldsymbol{d\overline{g}}\end{bmatrix} \quad (18)$$

$$\text{and } \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}_{\boldsymbol{W}}} = \begin{bmatrix}\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}} & \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\overline{\boldsymbol{g}}}\end{bmatrix} \quad (19)$$

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. 12, 13, 14 and 15. 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

$$\left[\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}}\right]_{t\nu,i} = \begin{cases}\left(\overline{g^d_{qt\nu,\mathbf{B}_{ij}}}\mathrm{s}_{d,j}\right).k^d_{pqt\nu} \text{ for } a = p\\ 0 \text{ otherwise}\end{cases} \quad (20)$$

and

$$\left[\boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\overline{\boldsymbol{g}}}\right]_{t\nu,i} = \begin{cases}\left(g^d_{pt\nu,\mathbf{A}_{ij}}\mathrm{s}_{d,j}\right).k^d_{pqt\nu} \text{ for } a = q\\ 0 \text{ otherwise}\end{cases} \quad (21)$$

One 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:

$$\boldsymbol{\mathcal{J}}_{\mathbf{v}} = \begin{bmatrix}\vdots\\ \boldsymbol{\mathcal{J}}_{\mathbf{v}_{pq},\boldsymbol{g}_{\boldsymbol{W}}}\\ \vdots\end{bmatrix} \quad (22)$$

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. Fast iterative solver using Wirtinger's framework

This section describes a Levenberg-Maquardt based calibration algorithm, using Wirtinger derivative to define the Jacobian (the COHJONES algorithm, see Sec. 2.2). This problem is addressed in greater detail in Smirnov & Tasse (in preparation).

2.1. Levenberg-Maquardt

In the following h is the non-linear operator mapping the 4-polarisations, direction-dependent gain vector $\boldsymbol{g}$ to the visibility vector $\mathbf{v}$ containing all baselines, time and frequency data

$$\mathbf{v} = h(\boldsymbol{g}) = \begin{bmatrix}\vdots\\ h_{pq}(\boldsymbol{g})\\ \vdots\end{bmatrix} \quad (23)$$

where h_{pq} is defined in Eq. 6.

The direction-dependent Jones matrices appearing in the measurement equation (Eq. 1) can then be estimated using a chi-square minimisation technique such as the Levenberg-Maquardt. Specifically, the gain vector $\boldsymbol{g}$ can be iteratively estimated from the measured visibilities $\mathbf{v_m}$ as

$$\widehat{\boldsymbol{g}_{i+1}} = \widehat{\boldsymbol{g}}_i + \mathbf{K}\big|^{-1}_{\widehat{\boldsymbol{g}}_i}\boldsymbol{J}\big|^{H}_{\widehat{\boldsymbol{g}}_i}\mathbf{C}^{-1}\left(\mathbf{v_m} - h\left(\widehat{\boldsymbol{g}}_i\right)\right) \quad (24)$$

$$\text{with } \mathbf{K}\big|_{\widehat{\boldsymbol{g}}_i} = \mathbf{H}\big|_{\widehat{\boldsymbol{g}}_i} + \lambda.\mathrm{diag}\left(\mathbf{H}\big|_{\widehat{\boldsymbol{g}}_i}\right) \quad (25)$$

$$\text{and } \mathbf{H}\big|_{\widehat{\boldsymbol{g}}_i} = \boldsymbol{J}\big|^{H}_{\widehat{\boldsymbol{g}}_i}\mathbf{C}^{-1}\boldsymbol{J}\big|_{\widehat{\boldsymbol{g}}_i} \quad (26)$$

where the matrix $\boldsymbol{J}\big|_{\widehat{\boldsymbol{g}}_i}$ is the Jacobian of h estimated at $\widehat{\boldsymbol{g}}_i$, and $\mathbf{C}$ is the covariance matrix of $\mathbf{v}$.

The Levenberg-Maquardt algorithm can be equivalently applied by using the Wirtinger Jacobian or the classical Jacobian. In this paper, as mentioned in Sec. 1, instead of computing the Jacobian using the real and imaginary parts of gains as independent variables, we use the Wirtinger derivative definition (see Sec. 1.3)

$$\boldsymbol{J} := \boldsymbol{\mathcal{J}}_{\mathbf{v}} \quad (27)$$

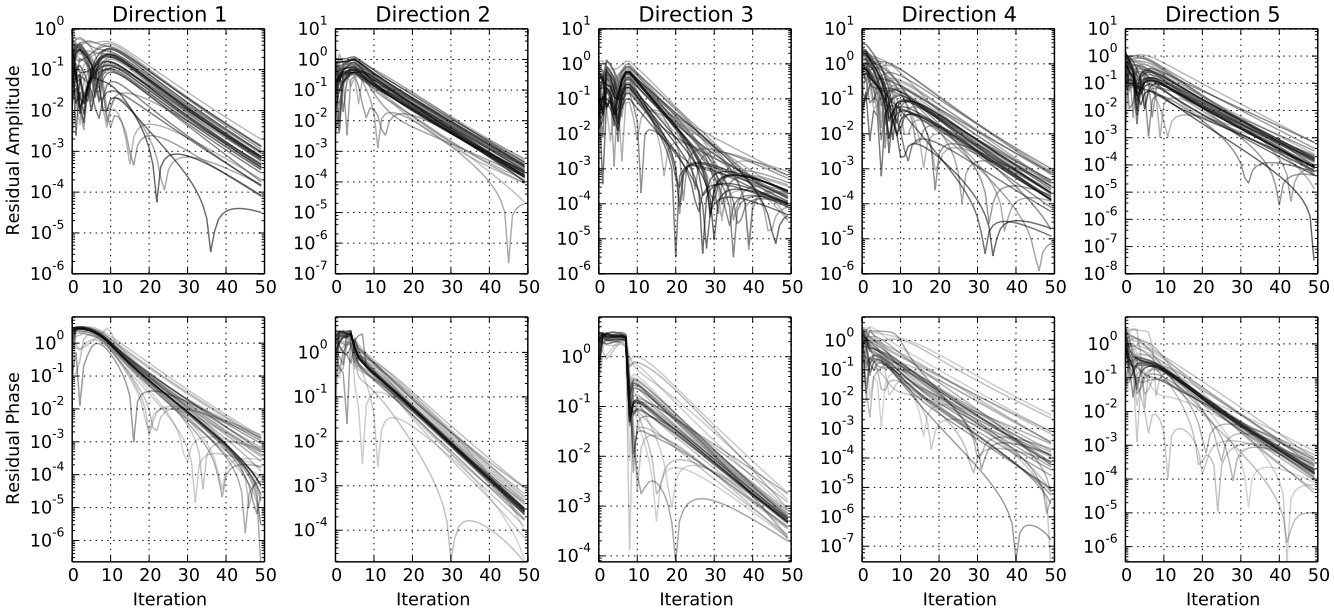

Fig. 1. This plot shows the amplitude (top panels) and phase (bottom panels) of the difference between the estimated gains and the true (random) gains in the different directions for the different antenna (shaded greys full lines).

2.2. The COHJONES *algorithm*

The Wirtinger $\boldsymbol{J}^H\boldsymbol{J}$ has a different structure to the classical one. In this section, we describe an algorithms that uses *only one* of the two independent Wirtinger variable (either z or $\overline{z}$). In short, in this section $\boldsymbol{J}$ is reduced to

$$\boldsymbol{J} := \boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}} \tag{28}$$

which means only half of the Jacobian $\boldsymbol{\mathcal{J}}_{\mathbf{v}}$ is used, and a single principal block in the matrix $\boldsymbol{\mathcal{J}}_{\mathbf{v}}^H\boldsymbol{\mathcal{J}}_{\mathbf{v}}$ is constructed (see Fig. 2). The algorithm is referred to *Complex Half-Jacobian Optimization for N-directional EStimation* (COHJONES).

Intuitively, the idea relies on that the RIME (Eq. 1) has the remarkable property to behaves like a linear system around the solution. Specifically, from Eq. 20 and 21, it is easy to check that:

$$\mathbf{v} = \frac{1}{2}\left(\boldsymbol{\mathcal{J}}_{\mathbf{v}}\big|_{\boldsymbol{g}_W}\right)\ \boldsymbol{g}_W \tag{29}$$

while

$$\mathbf{v} = h\left(\boldsymbol{g}\right) = \left(\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}\big|_{\overline{\boldsymbol{g}}}\right)\ \boldsymbol{g} \tag{30}$$

$$\text{and } \mathbf{v} = \left(\boldsymbol{\mathcal{J}}_{\mathbf{v},\overline{\boldsymbol{g}}}\big|_{\boldsymbol{g}}\right)\ \overline{\boldsymbol{g}} \tag{31}$$

Both Eq. 29 and 30 form linear systems. Furthermore it is shown in Smirnov & Tasse (in preparation) that the principal blocks of $\boldsymbol{\mathcal{J}}_{\mathbf{v}}^H\boldsymbol{\mathcal{J}}_{\mathbf{v}}$ corresponding to the $(\boldsymbol{g},\overline{\boldsymbol{g}})$ and $(\overline{\boldsymbol{g}},\boldsymbol{g})$ cross terms have a smaller amplitude than the $(\boldsymbol{g},\boldsymbol{g})$ and $(\overline{\boldsymbol{g}},\overline{\boldsymbol{g}})$ blocks.

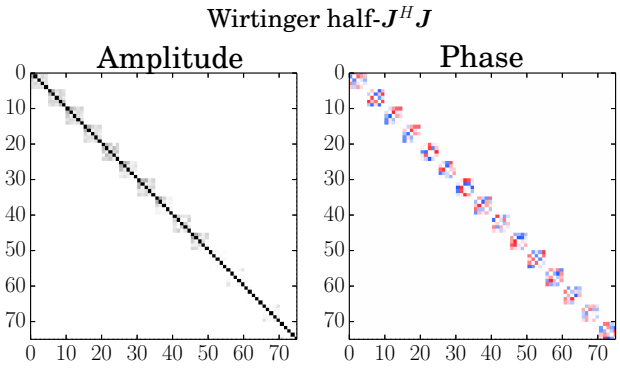

Fig. 2. This figure shows amplitude (left panel) and phase (right panel) of the block-diagonal matrix $\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}^H\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}$ for the dataset described in the text. Each block corresponds to the different directions for each specific antenna. Its block structure make it easily invertible.

The structure of $\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}^H\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}$ is shown in Fig. 2 for the dataset described in Sec. 3.1. This matrix is block diagonal, essentially because $\partial\overline{g}/\partial g = 0$ in Wirtinger's framework. This allows for dramatic improvement in algorithmic cost, as the $\boldsymbol{\mathcal{J}}_{\mathbf{v}}^H\boldsymbol{\mathcal{J}}_{\mathbf{v}}$ matrix inversion cost is $\mathcal{O}\left(n_d^3 n_a^3\right)$ instead of being $\mathcal{O}\left(n_d^3 n_a^3\right)$ corresponding to a net gain of n_a^2. In the same manner, the matrix product $\mathbf{K}^{-1}\boldsymbol{J}^H$ as well as the computation of $\boldsymbol{\mathcal{J}}_{\mathbf{v}}^H\boldsymbol{\mathcal{J}}_{\mathbf{v}}$ are n_a^2 times cheaper.

Another interesting property is that, assuming $\text{diag}\left(\mathbf{H}\right) \approx \mathbf{H}$ and injecting Eq. 30 into Eq. 24, we find:

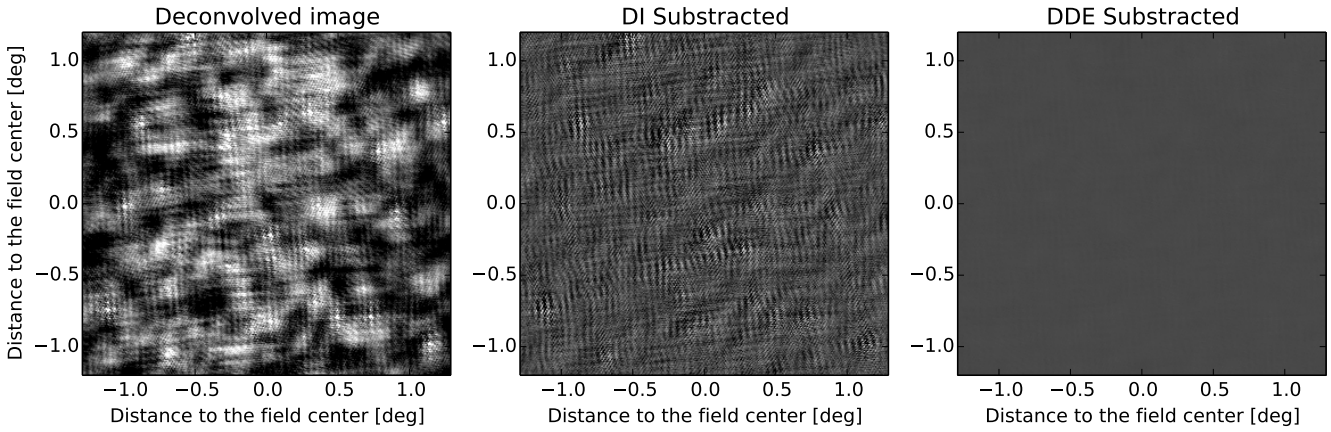

Fig. 3. This figure shows the comparison between the deconvoled image (left), the residuals data after simple skymodel subtractions (center), and the residuals data after subtracting the sky model corrupted by the direction-dependent COHJONES estimated solution (right). The color scale is the same in all panels. In this simulation, COHJONES reduces the residual data level by a factor of ~ 30.

$$\widehat{\boldsymbol{g}_{i+1}} = \lambda\left(\lambda+1\right)^{-1}\widehat{\boldsymbol{g}}_i + \left(\lambda+1\right)^{-1}\mathbf{H}_{\widehat{\boldsymbol{g}}_i}^{-1}\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}\big|_{\widehat{\boldsymbol{g}}_i}^{H}\mathbf{C}^{-1}\mathbf{v_m} \tag{32}$$

$$\text{with } \mathbf{H}_{\boldsymbol{g}} = \boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}\big|_{\widehat{\boldsymbol{g}}_i}^{H}\mathbf{C}^{-1}\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}\big|_{\widehat{\boldsymbol{g}}_i} \tag{33}$$

meaning the predict step does not have to be computed along the iteration.

3. Tests on simulated data

In this section, COHJONES is tested on simulated data for scalar Jones matrices only. In Sec. 3.1, the Jones matrices are constant in time. In that case only the convergence of COHJONES can be studied. In Sec. 3.2, the Jones matrices vary in time.

3.1. Time-constant Jones matrices

For this test, a visibility dataset is simulated assuming the Low Frequency Array (LOFAR) antenna layout. The phase center is located at $(\alpha,\delta) = (14^h11^m20.5^s, +52^o12'10.0'')$, the observing frequency is set 50 MHz, and time bins are 10 sec wide. To generate the visibilities, we use a sky model containing five sources, distributed in a cross, and separated by a degree. The gains applied to the antenna p in direction d are constant through time, and are taken at random along a normal distribution $g_p \sim \mathcal{N}(0,1) + i\mathcal{N}(0,1)$. The data vector is then built from all baselines, and a 20 minutes time chunk.

The corresponding matrix $\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}^{H}\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}$ is shown in Fig. 2. It is block diagonal, each block having size $n_d \times n_d$. The calibration solution convergence are shown in Fig. 1. It is important to note that the problems becomes better conditioned as the blocks of $\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}^{H}\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}$ become more diagonal. In that case COHJONES converges faster, and this happens (i) when more data are taken into account in the construction of the data vector or (ii) if the directions are put further away from each other.

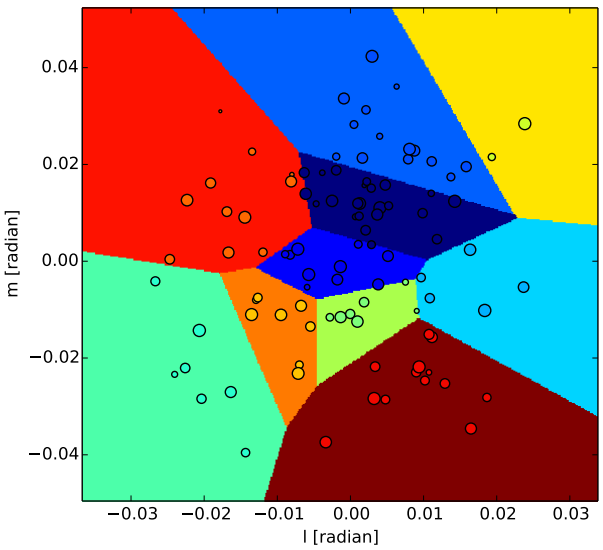

Fig. 4. In order to conduct direction-dependent calibration, the sources of the sky model are clustered using a Voronoi tessellation algorithm.

3.2. Variable gains

In order to simulate a more realistic dataset, we use a 100 sources sky model which flux density is randomly distributed (uniform distribution). Noise is added to the visibilities at the 1% level of the total flux. The scalar Jones matrices are simulated assuming an ionospheric model consisting of a purely scalar, direction-dependent phase (an infinitesimally thin layer at a height of 100 km). The

total electron content (TEC) values at a set of sample points are generated using Karhunen-Loeve decomposition (the spatial correlation is given by Kolmogorov turbulence, see van der Tol 2009). The sources are clustered in 10 directions using Voronoi tesselation (Fig. 4).

Fig. 3 shows the residuals data computed by subtracting the model data in the visibility domain, and the model data affected by DDEs. The residual data standard deviation reduces by a factor ~ 30 after COHJONES has been applied.

4. Conclusion

This paper has presented a Levenberg-Maquardt based algorithm that uses the Wirtinger's framework for complex derivative. The Jacobian harbors a different structure that is sparser than in the classical case. Based on this, a new optimisation algorithm has been presented (COHJONES).

This framework, and its connection with existing algorithms will be further discussed in Smirnov & Tasse (in preparation).

References

Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, A&AS, 117, 137
Smirnov, O. & Tasse, C. in preparation, MNRAS
van der Tol, S. 2009, PhD thesis, TU Delft