Applying Wirtinger's Jacobian to the radio interferometry calibration problem

C. Tasse[1,2], O. Smirnov[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 a dramatic gain in terms of algorithmic cost.

1. Complex optimisation

This section describes the shape and structures of the Jacobian over the complex scalar field, instead of what is usually done, by considering the real and imaginary separatly (later refered as the *Wirtinger* and *classical* Jacobians respectively).

1.1. RIME formalism

Using the Radio Interferometry Measurement Equation (RIME) formalism (for extensive discussions on the validity and limitations of the measurement equation see Hamaker et al. 1996; Smirnov 2011), the 4-polarisation visibility vector $\mathbf{v}_{pq}$ measured on baseline (pq), at time t and frequency ν can be written as

$$\begin{aligned}\mathbf{v}_{pq} =& \mathrm{Vec}\left(\mathbf{V}_{(pq)t\nu}\right) && (1)\\ =& \sum_d \left(\overline{\mathbf{J}^d_{qt\nu}} \otimes \mathbf{J}^d_{pt\nu}\right) \mathrm{Vec}\left(\mathbf{S}_d\right) k^d_{(pq)t\nu} && (2)\\ \text{with } k^d_{(pq)t\nu} =& \exp\left(-2i\pi\left(ul+vm+w\left(\mathrm{n}-1\right)\right)\right) && (3)\\ \text{and } n =& \sqrt{1-l^2-m^2} && (4)\end{aligned}$$

where $[u,v,w]^T$ 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, the i^{th} polarisation component of $\mathbf{v}_{pq}$ can be writen as

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

where

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

1.2. Wirtinger complex derivative

In order to compute a Jacobian, a derivative definition for complex numbers has to be choosen. 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 indepedent. 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

$$\begin{aligned}\frac{\partial}{\partial z} =& \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) && (7)\\ \text{and } \frac{\partial}{\partial \overline{z}} =& \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) && (8)\end{aligned}$$

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

$$\frac{\partial \overline{z}}{\partial z} = 0 \text{ and } \frac{\partial z}{\partial \overline{z}} = 0 \qquad (9)$$

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

$$\begin{aligned}\frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial g^d_{pt\nu,\mathbf{A}_{ij}}} =& \left(\overline{g^d_{qt\nu,\mathbf{B}_{ij}}}\mathrm{s}_{d,j}\right) . k^d_{pqt\nu} && (10)\\ \text{and } \frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial g^d_{qt\nu,\mathbf{B}_{ij}}} =& 0 && (11)\end{aligned}$$

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

$$\begin{aligned}\frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial \left(\overline{g^d_{pt\nu,\mathbf{A}_{ij}}}\right)} =& 0 && (12)\\ \text{and } \frac{\partial \mathrm{v}^i_{pqt\nu}}{\partial \left(\overline{g^d_{qt\nu,\mathbf{B}_{ij}}}\right)} =& \left(g^d_{pt\nu,\mathbf{A}_{ij}}\mathrm{s}_{d,j}\right) . k^d_{pqt\nu} && (13)\end{aligned}$$

Interestinggly, Eq. 10, 11, 12 and 13 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 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.

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 written as follows:

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

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

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

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

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

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

and

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

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:

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

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. Fast iterative solver using Wirtinger's framework

2.1. Levenberg-Maquardt

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

$$\widehat{\boldsymbol{g}_{i+1}} = \widehat{\boldsymbol{g}}_i + \mathbf{K}\big|^{-1}_{\widehat{\boldsymbol{g}}_i}\boldsymbol{J}\big|^H_{\widehat{\boldsymbol{g}}_i}\mathbf{C}^{-1}\left(\mathbf{v_m} - h\left(\widehat{\boldsymbol{g}}_i\right)\right) \tag{21}$$

$$\text{with } \mathbf{K}\big|_{\widehat{\boldsymbol{g}}_i} = \mathbf{H}\big|_{\widehat{\boldsymbol{g}}_i} + \lambda.\mathrm{diag}\left(\mathbf{H}\big|_{\widehat{\boldsymbol{g}}_i}\right) \tag{22}$$

$$\text{and } \mathbf{H}\big|_{\widehat{\boldsymbol{g}}_i} = \boldsymbol{J}\big|^H_{\widehat{\boldsymbol{g}}_i}\mathbf{C}^{-1}\boldsymbol{J}\big|_{\widehat{\boldsymbol{g}}_i} \tag{23}$$

where the matrix $\boldsymbol{J}$ is the Jacobian of h, 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 Jacobocian. In this paper, as mentioned in 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)

$$\boldsymbol{J} \equiv \boldsymbol{\mathcal{J}}_{\mathbf{v}} \tag{24}$$

2.2. The COHJONES algorithm

The Wirtinger $\boldsymbol{J}^H\boldsymbol{J}$ has a different structure to the classical one (Smirnov & Tasse 2014, in prep), allowing for high gain in terms of algorithmic cost, and in this section, we describe an algorithms that uses *only one* of the two independent Wirtinger variable (either z or $\overline{z}$). In short, $\boldsymbol{J}$ is defined as

$$\boldsymbol{J} \equiv \boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}} \tag{25}$$

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

Intuitively, the idea relies on that the RIME (Eq. 1) has the remarquable property to behaves like a linear system around the solution. Specifically, from Eq. 18 and 19, it is easy to check that:

$$\mathbf{v} \approx \frac{1}{2}\left(\boldsymbol{\mathcal{J}}_{\mathbf{v}}\big|_{\boldsymbol{g}_{\boldsymbol{W}}}\right)\ \boldsymbol{g}_{\boldsymbol{W}} \tag{26}$$

while

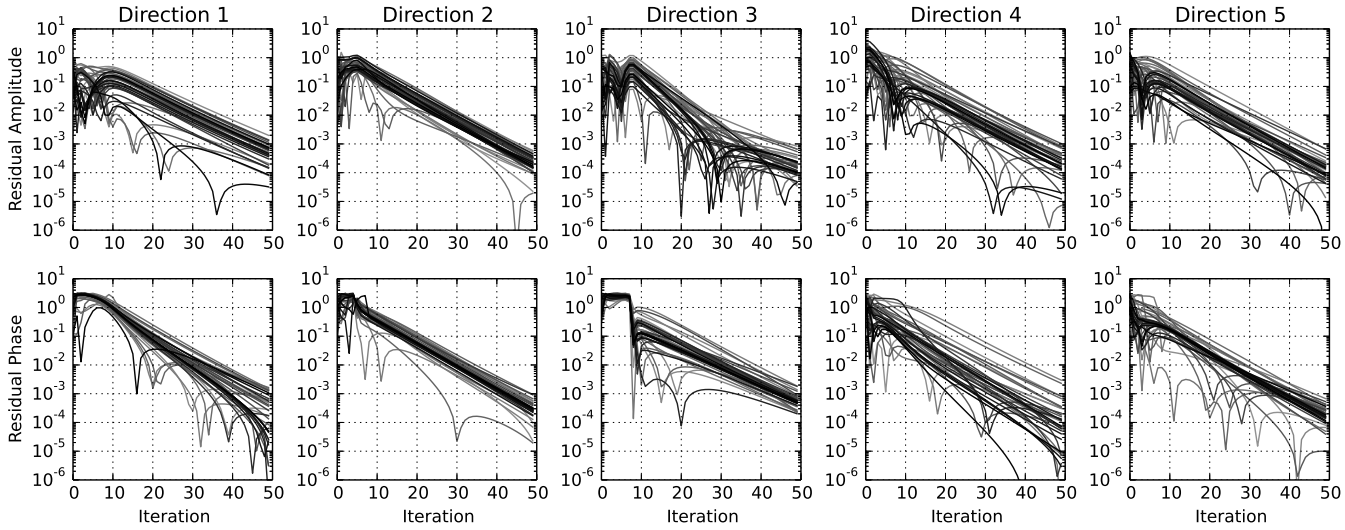

Fig. 1. This plot shows the amplitude (top panels) and phase (bottom panels) of the difference beween the estimated gains and the true (random) gains in the different directions.

$$\mathbf{v} \approx \left(\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}\big|_{\overline{\boldsymbol{g}}}\right)\ \boldsymbol{g} \tag{27}$$

$$\text{and } \mathbf{v} \approx \left(\boldsymbol{\mathcal{J}}_{\mathbf{v},\overline{\boldsymbol{g}}}\big|_{\boldsymbol{g}}\right)\ \overline{\boldsymbol{g}} \tag{28}$$

Both Eq. 26 and 27 form linear systems. Furthermore it is shown in Smirnov & Tasse (2014, in prep) that the principal blocks of $\boldsymbol{\mathcal{J}}^H_{\mathbf{v}}\boldsymbol{\mathcal{J}}_{\mathbf{v}}$ 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.

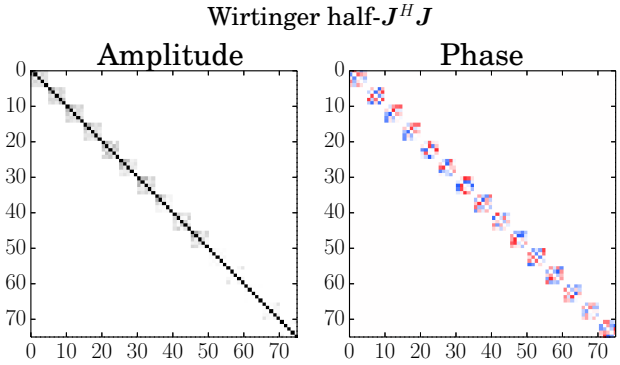

Fig. 2. This figure shows amplitude (left panel) and phase (right panel) of the block-diagonal matrix $\boldsymbol{\mathcal{J}}^H_{\mathbf{v},\boldsymbol{g}}\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}$ for the dataset described in the text. Each block corresponds to the different directions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

The structure of $\boldsymbol{\mathcal{J}}^H_{\mathbf{v},\boldsymbol{g}}\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}$ is shown in Fig. 2 for the dataset decribed 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 $\boldsymbol{\mathcal{J}}^H_{\mathbf{v}}\boldsymbol{\mathcal{J}}_{\mathbf{v}}$ matrix inversion cost is $\mathcal{O}\left(n_d^3 n_a\right)$ 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 $\mathbf{K}^{-1}\boldsymbol{J}$ is n_a^2 times cheaper.

Another interesting approximation is that, assuming $\mathrm{diag}\left(\mathbf{H}\right) \approx \mathbf{H}$ and injecting 27 into 21, we find:

$$\widehat{\boldsymbol{g}_{i+1}} = \lambda\left(\lambda+1\right)^{-1}\widehat{\boldsymbol{g}}_i + \left(\lambda+1\right)^{-1}\mathbf{H}^{-1}_{\widehat{\boldsymbol{g}}_i}\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}\big|^H_{\widehat{\boldsymbol{g}}_i}\mathbf{C}^{-1}\mathbf{v_m} \tag{29}$$

$$\text{with } \mathbf{H}_{\boldsymbol{g}} = \boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}\big|^H_{\widehat{\boldsymbol{g}}_i}\mathbf{C}^{-1}\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}\big|_{\overline{\widehat{\boldsymbol{g}}_i}} \tag{30}$$

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

3. Testing cohjoneson simulated data

3.1. Constant gains

A visibility dataset is simulated assuming the Low Frequency Array (LOFAR) antenna layout. The phase center is located at $(\alpha,\delta) = (14^h11^m20.5^s, +52^o12'10.0'')$, observting frequency is 50 MHz time bins are 10 sec wide, and using a sky model containing five sources, distributed in a cross. The gains applied to the antenna p are constant through time, and are taken at random along a normal distribution $g_p \sim \mathcal{N}\left(0,1\right) + i\mathcal{N}\left(0,1\right)$.

The matrix corresponding matrix $\boldsymbol{\mathcal{J}}^H_{\mathbf{v},\boldsymbol{g}}\boldsymbol{\mathcal{J}}_{\mathbf{v},\boldsymbol{g}}$ is shown in Fig. 2. The calibration solution convergence are shown in Fig. 1.

3.2. Variable gains

In order to simulate a more realistic dataset, we use a 100 sources skymodel. The gains are simulated assuim-

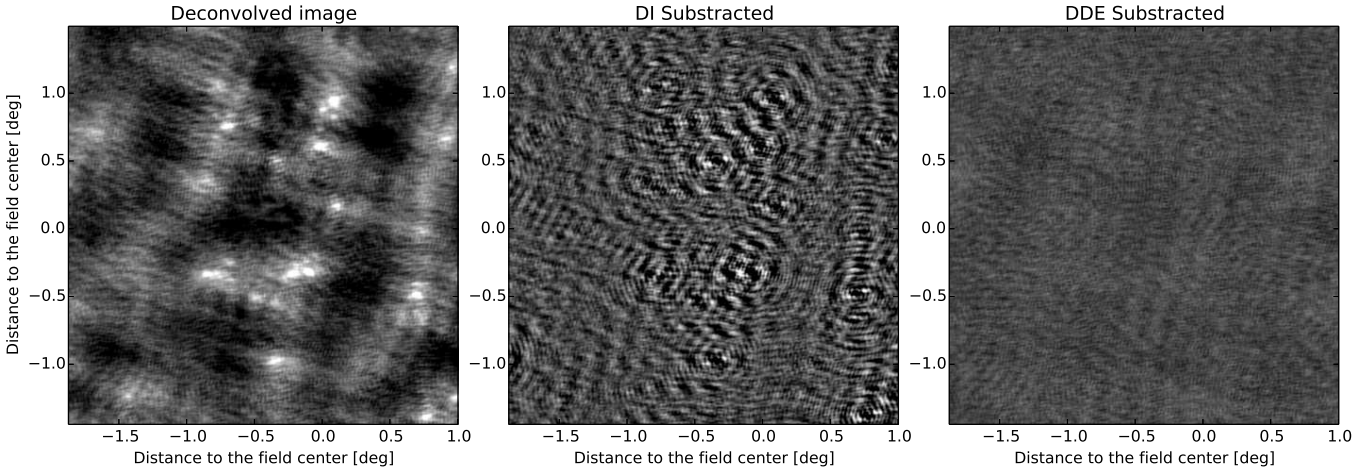

Fig. 3. This figure shows compares the image (left), the residuals data after simple skymodel substraction (center), and the residuals data after substracting the sky model corrupted by the direction-dependent solution (right).

ing 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. 3 shows the residuals the residuals as computed by substracting the model data in the visibility domain, and the model data affected by DDEs. Clearly, COHJONES gives a significant improvement over direct substraction.

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 (COHJONES) has been presented.

This framework, and its connection with existing algorithms will be further discussed in Smirnov & Tasse (2014, in prep.).

References

Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, A&AS, 117, 137
Smirnov, O. & Tasse, C. 2014, ArXiv e-prints
Smirnov, O. M. 2011, A&A, 527, A106
van der Tol, S. 2009, PhD thesis, TU Delft