

Radio Interferometric Calibration Using The SAGE Algorithm

Sarod Yatawatta^{1,2,3}, Saleem Zaroubi¹, Ger de Bruyn^{1,2}, Leon Koopmans¹ and Jan Noordam²

Abstract—Radio Interferometry is an essential method for astronomical observations. Self-calibration techniques have increased the quality of the radio astronomical observations (and hence the science) by orders of magnitude. Recently, there is a drive towards sensor arrays built using inexpensive hardware and distributed over a wide area acting as radio interferometers. Calibration of such arrays poses new problems in terms of computational cost as well as in performance of existing calibration algorithms. We consider the application of the Space Alternating Generalized Expectation Maximization (SAGE) [1] algorithm for calibration of radio interferometric arrays. Application to real data shows that this is an improvement over existing calibration algorithms that are based on direct, deterministic non linear optimization. As presented in this paper, we can improve the computational cost as well as the quality of the calibration using this algorithm.

I. INTRODUCTION

Radio synthesis arrays have greatly benefited by self calibration techniques invented during the last 30 years. Calibration refers to estimation of errors introduced by the instrument (and also the propagation path, such as the ionosphere), and correction for such errors, before any imaging is done. At the beginning of radio astronomy, calibration was done by observing a known celestial object (called the external calibrator), in addition to the part of the sky being observed. This was improved by self-calibration, which is essentially using the observed sky itself for the calibration. Therefore, self calibration entail considering both the sky as well as the instrument as unknowns. Nevertheless, by iteratively refining the sky and the instrument model, the quality of the calibration was improved by orders of magnitude in comparison to using an external calibrator.

From a signal processing perspective, calibration is essentially the Maximum Likelihood (ML) estimation of the instrument and sky parameters using a non linear optimization technique such as the Levenberg Marquardt [2],[3] (LM) algorithm. An in depth overview of existing calibration techniques are given in [4],[5]. However, with such techniques, we have reached a limit in sensitivity that can be achieved using present radio interferometers. This is because the achievable sensitivity is limited by the receiver collecting area itself. Moreover, with finite computational cost, there is a bound in the performance of existing algorithms. Therefore, there is a drive towards building large, distributed, sensor arrays (SKA[6], LOFAR

[7]) that increase the collecting area and hence the sensitivity, enabling new scientific research. However, this also increases the parameter space that needs to be calibrated and hence, the required computational cost. A detailed analysis of calibration of LOFAR, which can be considered as a pathfinder for the next generation interferometric arrays, can be found in [15]. Although in [15], limitations (such as the Cramer Rao bound) of existing calibration techniques are defined, there is little work existing on improving the computational cost or the speed of convergence of such techniques.

This paper describes the application of the Expectation Maximization [8] (EM) algorithm for radio interferometric calibration. The EM algorithm was first presented as a solution to maximum likelihood estimation when the complete data is not observed. Since then, EM type algorithms have been widely used as an iterative method for ML estimation, with either faster convergence or reduced computational cost. The essential property of the EM algorithm is that the likelihood can only increase in each iteration. The SAGE algorithm was presented to improve the plain EM algorithm in terms of speed of convergence and has been successfully applied in diverse signal processing applications, such as medical imaging [9], communications systems [10], etc. In contrast to [15], in this paper we focus on finding algorithms that improve the computational cost and quality of calibration. Naturally, in order to improve the performance of the ML estimation, we choose the EM algorithm, and in particular the SAGE algorithm. Therein lies the novelty of this paper.

Notation: Lower case bold letters refer to column vectors (e.g. \mathbf{y}). Upper case bold letters refer to matrices (e.g. \mathbf{C}). Unless otherwise stated, all parameters are complex numbers. The matrix inverse, transpose, Hermitian transpose, and conjugation are referred to as $(\cdot)^{-1}$, $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$, respectively. The matrix Kronecker product is given by \otimes . The statistical expectation operator is given as $E\{\cdot\}$. The vectorized representation of a matrix is given by $\text{vec}(\cdot)$. The diagonal matrix consisting of only the diagonal entries of a square matrix is given by $\text{diag}(\cdot)$. The identity matrix is given by \mathbf{I} . The Kronecker delta function is given by δ_{ij} . Real and complex numbers are represented as \mathbb{R} and \mathbb{C} , respectively. Estimated parameters are denoted by a hat, $\widehat{(\cdot)}$. All logarithms are to the base e .

II. DATA MODEL

We briefly describe the data model of the radio interferometer in this section. For more information about radio interferometry, the reader is referred to [11] and for the

¹Kapteyn Astronomical Institute, University of Groningen, Groningen, The Netherlands

²ASTRON, Dwingeloo, The Netherlands

³yatawatta@astron.nl

data model in particular, [12],[13]. A more signal processing oriented description is given in [4],[15].

We consider the radio frequency sky to be composed of discrete sources, far away from the earth such that the approaching radiation from each one of them appears to be plane waves. Let the plane wave from the i -th source be decomposed to two orthogonal polarization directions $\mathbf{u}_i = [u_{xi} \ u_{yi}]^T$. The interferometric array consists of N receiving elements or stations. At the p -th station, this plane wave causes an induced voltage, which is dependent on the beam attenuation as well as the radio frequency receiver chain attenuation. Normally, each station has dual polarized feeds. So the induced voltages at the x and y feeds, $\tilde{\mathbf{v}}_{pi} = [v_{xpi} \ v_{yqi}]^T$ due to source i are given as in (1).

$$\tilde{\mathbf{v}}_{pi} = \mathbf{J}_{pi} \mathbf{u}_i \quad (1)$$

In (1), the complex interaction of the approaching radiation with the station beam shape as well as the remaining signal path is represented by the 2 by 2 Jones matrix \mathbf{J}_{pi} . If there are K such sources, the total signal will be a superposition of K such signals as in (1). Moreover, the receiver noise $\boldsymbol{\nu} = [\nu_x \ \nu_y]^T$ is also added to this signal.

The signal of the p -th station is correlated with the signals of the other $N - 1$ receivers at the correlator. Before this is done, each signal is given a delay correction depending on the direction on the sky being observed and also depending on the absolute position of the receiver on the earth.

After correlation, the correlated signal of the p -th station and the q -th station (named as the *visibilities* [12]), $\mathbf{V}_{pq} = E\{\mathbf{v}_p \mathbf{v}_q^H\}$ is given by (2).

$$\mathbf{V}_{pq} = \sum_{i=1}^K \mathbf{J}_{pi}(\boldsymbol{\theta}) \mathbf{C}_i \mathbf{J}_{qi}^H(\boldsymbol{\theta}) + \mathbf{N}, \quad p, q \in [1, \dots, N] \quad (2)$$

This is the full matrix measurement equation, developed in [12]. In (2), $\mathbf{J}_{pi}(\boldsymbol{\theta})$ and $\mathbf{J}_{qi}(\boldsymbol{\theta})$ are the Jones matrices describing the electromagnetic and electronic interaction of the plane wave of source i with stations p and q , respectively. The parameter vector $\boldsymbol{\theta} \in \mathbb{C}^P$ describes the unknown instrument model. The 2 by 2 noise matrix is given as \mathbf{N} . We can only arrive at (2) because the radiation emitted by the sources in the sky are uncorrelated. The *coherency* [12], \mathbf{C}_i describes the intrinsic polarized radiation of the i -th source.

The instrumental properties (such as the beamshape, low noise amplifier gain, system frequency response etc.) and the path properties (such as tropospheric and ionospheric distortion etc.) are described by the Jones matrices $\mathbf{J}_{pi}(\boldsymbol{\theta})$ and $\mathbf{J}_{qi}(\boldsymbol{\theta})$ in (2). Calibration is essentially finding the parameters $\boldsymbol{\theta}$ (P complex valued parameters or $2P$ real values parameters). There are numerous ways to parametrize these unknowns using the parameter set $\boldsymbol{\theta}$. For a more specialized treatment of the parametrization as well as factoring the Jones matrices, the reader is referred to [14].

To a lesser extent, the source information \mathbf{C}_i is also unknown. However, at the initial stage, we can use prior information about the source or sky properties obtained by previous observations.

Note that in (2), the noise matrix $\mathbf{N} = \mathbf{0}$ for $p \neq q$ if the noise at each receiver is uncorrelated. However in practice this does not hold for the following reasons:

- The integration time at the correlator has to be finite in order not to decorrelate the signal from the sources (or the sky). Hence there is always some receiver noise appearing even in the cross correlations, $p \neq q$.
- The assumption that the sky is composed of a set of discrete sources is valid only up to a certain intensity level. There is low level diffused radiation from the sky, which appear especially in short baseline visibilities.
- We select only K brightest sources in our data model. However, the multitude of fainter sources that are ignored in (2) contribute to the noise.

The vectorized form of (2), $\mathbf{v}_{pq} = \text{vec}(\mathbf{V}_{pq})$ can be written as in (3) where $\mathbf{n}_{pq} = \text{vec}(\mathbf{N})$.

$$\mathbf{v}_{pq} = \sum_{i=1}^K \mathbf{J}_{qi}^*(\boldsymbol{\theta}) \otimes \mathbf{J}_{pi}(\boldsymbol{\theta}) \text{vec}(\mathbf{C}_i) + \mathbf{n}_{pq} \quad (3)$$

Ignoring the autocorrelations where $p = q$, stacking up all cross correlations as $\mathbf{y} = [\mathbf{v}_{12}^T \ \mathbf{v}_{13}^T \ \dots \ \mathbf{v}_{(N-1)N}^T]^T$, $\mathbf{y} \in \mathbb{C}^M$, we get (4).

$$\mathbf{y} = \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n} \quad (4)$$

The size of \mathbf{y} , M , in (4) is at most $2N(N - 1)$ provided all cross correlations are used. Typically, the number of parameters in $\boldsymbol{\theta}$, P , is proportional to KN . So for large enough N and small enough K , we have enough constraints to estimate $\boldsymbol{\theta}$. The non-linear functions $\mathbf{s}_i(\boldsymbol{\theta})$ correspond to the contribution of each source to the observation. In previous formulations of the same problem, [15],[16], the noise \mathbf{n} has been ignored because only the cross correlations are used. However, we stress that in our formulation, we consider \mathbf{n} to be a Gaussian random variable with zero mean and covariance $\boldsymbol{\Pi}$ ($M \times M$ matrix), i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Pi})$.

Note that (4) is a superposition of K non linear signals, with unknown parameters, which is exactly the problem considered in [17]. The present calibration schemes estimate $\boldsymbol{\theta}$ as the least squared error estimate, typically using a gradient based optimization algorithm like LM algorithm.

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta})\|^2 \quad (5)$$

If the cost function is $\phi(\boldsymbol{\theta}) = \|\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta})\|^2$, at the k -th iteration, we estimate

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - (\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \phi(\boldsymbol{\theta}) + \lambda \mathbf{H})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta})|_{\boldsymbol{\theta}^k} \quad (6)$$

In (6), $\nabla_{\boldsymbol{\theta}}$ is the gradient with respect to $\boldsymbol{\theta}$ and λ is a regularization parameter. The matrix $\mathbf{H} = \text{diag}(\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \phi(\boldsymbol{\theta}))$ is the diagonal of the Hessian matrix. Given suitable initial values, (6) should converge to the global optimum. However, (6) suffers from the same set of problems faced with any non linear optimization problem, i.e., convergence to local

minima, slow convergence and heavy computational cost. In the next section we shall investigate the application of the EM algorithm to overcome some of these problems.

III. THE EM AND SAGE ALGORITHMS

In this section, we first proceed to apply the EM algorithm to (4), in a similar way as done in [17]. We do this before applying the SAGE algorithm to clarify the presentation.

A. EM Algorithm

The key step in applying the EM algorithm is to define a complete data set \mathbf{x} from the observed data \mathbf{y} . The obvious choice would be to associate each source with a complete data space $\tilde{\mathbf{x}} = [\tilde{\mathbf{x}}_1^T \tilde{\mathbf{x}}_2^T \dots \tilde{\mathbf{x}}_K^T]^T$, with each component as in (7), such that $\mathbf{y} = \sum_{i=1}^K \tilde{\mathbf{x}}_i$.

$$\tilde{\mathbf{x}}_i = \mathbf{s}_i(\boldsymbol{\theta}_i) + \tilde{\mathbf{n}}_i \quad (7)$$

Note that in (7), we have assumed the contribution of the i -th source depends only on a subset of parameters $\boldsymbol{\theta}_i$, not the full set of parameters $\boldsymbol{\theta}$. In other words, we partition the parameter space to K components as $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T \boldsymbol{\theta}_2^T \dots \boldsymbol{\theta}_K^T]^T$. This is justified because each source is at a unique direction on the sky. Even though the signal path for a given station is common for all sources, the different directions and the rotation of the sky makes this assumption justifiable. The noise contribution $\tilde{\mathbf{n}}_i$ is such that the total noise is decomposed into K noise sources.

$$\mathbf{n} = \sum_{i=1}^K \tilde{\mathbf{n}}_i, \quad E\{\tilde{\mathbf{n}}_i \tilde{\mathbf{n}}_j^H\} = \beta_i \delta_{ij} \mathbf{I}, \quad \sum_{i=1}^K \beta_i = 1 \quad (8)$$

The β_i s form an affine combination and we are free to choose them. Typically, we can associate stronger sources with lower noise, hence low β_i . Given the complete data \mathbf{x} , we get the observed data as in (9) where \mathbf{G} is a block matrix with K identity matrices.

$$\mathbf{y} = [\mathbf{I} \ \mathbf{I} \ \dots \ \mathbf{I}] \mathbf{x} = \mathbf{G} \mathbf{x} \quad (9)$$

Having this setup, it is rather straightforward to apply the EM algorithm to our problem as in [17].

E Step: We find the conditional mean of $\tilde{\mathbf{x}}_i = E\{\tilde{\mathbf{x}}_i | \mathbf{y}, \boldsymbol{\theta}^k\}$. Taking into account that \mathbf{y} and \mathbf{x} are jointly Gaussian, we get

$$\hat{\tilde{\mathbf{x}}}_i = \mathbf{s}_i(\boldsymbol{\theta}_i^k) + \beta_i (\mathbf{y} - \sum_{l=1}^K \mathbf{s}_l(\boldsymbol{\theta}_l^k)) \quad (10)$$

M Step: For the $k+1$ -th iteration, we find $\boldsymbol{\theta}_i^{k+1}$ that minimizes the cost $\phi_i(\boldsymbol{\theta}_i^{k+1}) = \|\hat{\tilde{\mathbf{x}}}_i - \mathbf{s}_i(\boldsymbol{\theta}_i^{k+1})\|^2$, given by:

$$\boldsymbol{\theta}_i^{k+1} = \boldsymbol{\theta}_i^k - (\nabla_{\boldsymbol{\theta}_i} \nabla_{\boldsymbol{\theta}_i}^T \phi_i(\boldsymbol{\theta}_i) + \lambda \mathbf{H}_i)^{-1} \nabla_{\boldsymbol{\theta}_i} \phi_i(\boldsymbol{\theta}_i) |_{\boldsymbol{\theta}_i^k} \quad (11)$$

where $\mathbf{H}_i = \text{diag}(\nabla_{\boldsymbol{\theta}_i} \nabla_{\boldsymbol{\theta}_i}^T \phi_i(\boldsymbol{\theta}_i))$. We repeat the above two steps starting from iteration $k=1$ until convergence or an upper limit has reached. At each iteration, we update each source, so i goes from 1 to K .

B. SAGE Algorithm

Next, we investigate the application of the SAGE algorithm to our problem. As before we need to find a complete data space (or a hidden data space as defined in [1]). Similar to [1], we select the hidden data space \mathbf{x}^S as in (12).

$$\mathbf{x}^S = \mathbf{s}_i(\boldsymbol{\theta}_i) + \mathbf{n} \quad (12)$$

This gives the observed data \mathbf{y} as in (13).

$$\mathbf{y} = \mathbf{x}^S + \sum_{l=1, l \neq i}^K \mathbf{s}_l(\boldsymbol{\theta}_l) \quad (13)$$

Note that in (12) and (13), we have selected the index set $[1, S]$ to be the i -th source. Moreover, we have associated all the noise to \mathbf{x}^S , unlike in the classic EM algorithm. Once again, we arrive at the following EM scheme:

SAGE E Step: We find the conditional mean of $\mathbf{x}^S = E\{\mathbf{x}^S | \mathbf{y}, \boldsymbol{\theta}^k\}$.

$$\hat{\mathbf{x}}^S = \mathbf{s}_i(\boldsymbol{\theta}_i^k) + (\mathbf{y} - \sum_{l=1}^K \mathbf{s}_l(\boldsymbol{\theta}_l^k)) = \mathbf{y} - \sum_{l=1, l \neq i}^K \mathbf{s}_l(\boldsymbol{\theta}_l^k) \quad (14)$$

SAGE M Step: For the $k+1$ -th iteration, $\boldsymbol{\theta}_i^{k+1}$ that minimizes the cost $\phi_S(\boldsymbol{\theta}_i^{k+1}) = \|\hat{\mathbf{x}}^S - \mathbf{s}_i(\boldsymbol{\theta}_i^{k+1})\|^2$. This is similar to (11). As before, we iterate from $k=1$ to an upper limit. At each iteration, we change the index set S to update all or some sources.

Note that instead of partitioning per source, we could also perform the partitioning to include more than one source. This would be better if some sources are closer in the sky and hence share some parameters.

C. Computational Cost

The computational cost of direct estimation using (6) and the EM algorithmic approach can be compared as follows. Solving (6) (without calculating the inverse) involves the solution of a linear system of order KN . So the computational cost of the direct approach is $\mathcal{O}((KN)^2)$. On the other hand, the computational cost of solving (11), K times, is $K\mathcal{O}(N^2)$. Thus, we gain a factor K by using the EM algorithm. Furthermore, we can increase this gain if the convergence of the EM approach is faster (fewer iterations).

D. Comparison With "Peeling"

As described in [15] in detail, *Peeling* is the conversion of the K source model in (4) to a series of single source calibration problems, exploiting the temporal diversity due to the rotation of the earth. The steps taken in *Peeling* can be briefly given as follows:

- Out of sources, $i \in [1, K]$, select the strongest source (say $i = q$).
- Optionally, subtract the contributions of the remaining sources from (4), using an approximate a priori instrument model.
- Multiply (4) by a diagonal matrix \mathbf{F} such that the q -th source is at the phase center. The matrix \mathbf{F} is computed

for given time and the absolute position of the q -th source in the sky.

- Provided that the remaining sources are weak enough, over a finite time interval, the contribution of those sources in (4) are averaged out, while the contribution of the q -th source remains constant.
- Ignore the contribution from the other sources $i \in [1, K], i \neq q$ in (4) and solve for the parameters of the q -th source. Subtract this from (4). Now, we have $K - 1$ sources left and we repeat the whole procedure.

As seen from above, the application of the proposed algorithm does not rely on the weaker sources being averaged out. Moreover, the proposed algorithm does not require an a priori instrument model. When we have equally strong sources, *Peeling* might not work satisfactorily compared to the proposed algorithm.

IV. STATISTICAL MODEL ORDER SELECTION

So far in our analysis, we have assumed the number of sources, K , in (2), is known a priori. To some extent, this is true, given prior observational data and receiver noise characteristics. It is straightforward to find K sources that has sufficient SNR given the aforementioned information. However, in situations where the receiver antenna beamshape varies with time due to earth rotation (as in LOFAR), this is hard to predict (e.g. some sources might go close to, or, even below, the horizon). In this situation without a priori knowledge, we could use information theoretic criteria to find the optimal K for a given observation. In this section, we describe the use of Akaike's Information Criterion (AIC) [18] for this purpose. There are also alternative criteria, see for instance [19] for more information.

From (4), the likelihood of \mathbf{y} is given by (15).

$$f(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{\pi^M |\boldsymbol{\Pi}|} \exp\left(-(\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta}))^H \boldsymbol{\Pi}^{-1} (\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta}))\right) \quad (15)$$

Assuming the noise to be white, $\boldsymbol{\Pi} = \sigma^2 \mathbf{I}$, we get the simplified log-likelihood $L(\boldsymbol{\theta})$ as in (16).

$$\begin{aligned} L(\boldsymbol{\theta}) &= \log f(\mathbf{y}|\boldsymbol{\theta}) \\ &= -M \log \pi - M \log \sigma^2 \\ &\quad - \frac{1}{\sigma^2} (\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta}))^H (\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta})) \end{aligned} \quad (16)$$

The maximum likelihood estimate for the noise variance σ^2 (given $\boldsymbol{\theta}$) is given by (17).

$$\widehat{\sigma^2} = \frac{1}{M} (\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\widehat{\boldsymbol{\theta}}))^H (\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\widehat{\boldsymbol{\theta}})) \quad (17)$$

Using (17) in (16), we arrive at (18).

$$\begin{aligned} L(\widehat{\boldsymbol{\theta}}) &= -M \log \pi - M \\ &\quad - M \log \left(\frac{1}{M} (\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\widehat{\boldsymbol{\theta}}))^H (\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\widehat{\boldsymbol{\theta}})) \right) \end{aligned} \quad (18)$$

Using (18), we get the Akaike's Information Criterion as (19). We select K that gives the minimum value for (19).

$$AIC(K) = -2L(\widehat{\boldsymbol{\theta}}) + 2(2P) \quad (19)$$

V. NUMERICAL EXAMPLE

We consider the calibration of some data obtained by the LOFAR test core station (CS1). This has $N = 16$ dipoles (with dual polarization) acting as an interferometric array. Since each station is a single dipole, there is no beamforming and thus the whole sky (hemisphere) is observed. In this setting, the two brightest sources are Cassiopeia A (CasA) and Cygnus A (CygA), with intensities about 20000 Jy each at 50 MHz. The observation lasts for 24 hours. The correlator integration time is 30 sec. During the observation, the positions of the sources (azimuth and elevation) vary as shown on Fig. (1).

Since both CasA and CygA are equally bright, traditional algorithms such as peeling [15] will not work satisfactorily, as described in section III-D. So we have a model with $K = 2$ in (4). Note that as seen on Fig. (1), CygA goes very close to the horizon at one point. Around this time, the contribution from CygA is almost negligible due to the attenuation by the dipole beam. So, instead of using $K = 2$, we should be using $K = 1$. However, in this example, we only consider the data where both CasA and CygA are high in elevation (about 18 hours). Future work will address using e.g., (19) to determine this.

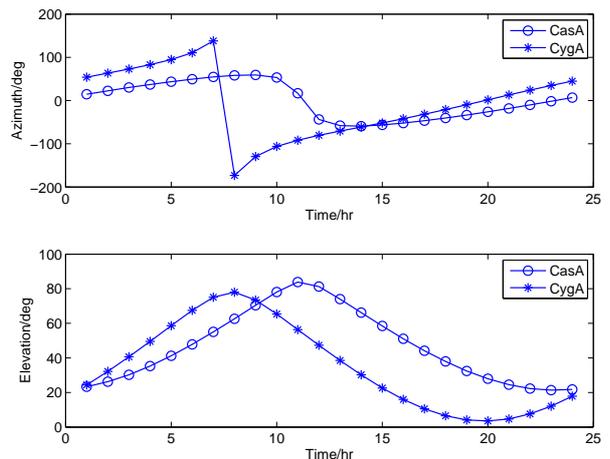


Fig. 1. The positions of CasA and CygA on the sky in azimuth and elevation, for a geographic latitude of 53° .

The parametrization of the Jones matrices are done as follows: We consider each entry of the 2 by 2 matrix to be a parameter. For $K = 2$ and $N = 16$, there are $2 \times 16 \times 4 = 128$ parameters.

$$\begin{aligned} \mathbf{J}_p(\boldsymbol{\theta}_i) &= \begin{bmatrix} J_{p11i} & J_{p12i} \\ J_{p21i} & J_{p22i} \end{bmatrix}, \\ \boldsymbol{\theta}_i &= [\text{vec}(\mathbf{J}_p(\boldsymbol{\theta}_i))^T, \dots]^T \quad \forall p \in [1 \dots, N] \end{aligned} \quad (20)$$

The Jones matrices are initialized such that the diagonal entries each have a (real) value 0.0001 and the off diagonal entries to be zero.

We consider the estimation of the parameters using (5) (Normal Algorithm) and using the SAGE algorithm. For the normal algorithm, we use 12 and 24 LM iterations to estimate 128 parameters using (5). For the SAGE algorithm, we alternate between estimating parameters for CasA and CygA. In each iteration, we use 3 LM iterations for the M step (11). We use 4 EM iterations, keeping the number of LM iterations at 12. Yet, the SAGE algorithm is computationally less expensive than the normal algorithm with 12 iterations as noted in section III-C.

Once we have estimated θ , we make images of the residual, i.e. $\mathbf{y} - \sum_{i=1}^K \mathbf{s}_i(\hat{\theta})$. We also correct the residual using the estimated θ [20]. We have given the images made by the normal algorithm and the SAGE algorithm on Fig. 2. These images show an area around CasA and CygA, respectively. Perfect subtraction should leave no residual from both these sources. However, we see that there is about 1% (of the original value) peak residual left by using the normal algorithm with 12 iterations. On the other hand, the SAGE algorithm and the normal algorithm with 24 iterations reduce this residual to 0.1% level. Moreover, fainter, known sources can also be seen on both images. Closer scrutiny reveals that the remaining sources are fainter in the results obtained using the SAGE algorithm. This is due to over subtraction of the fluxes of the remaining sources and in fact, we could reduce the number of SAGE iterations, to overcome this effect. Future work will address determining the correct number of iterations to avoid over subtraction.

In order to have a quantitative handle on the results, we have also calculated the root mean square (rms) value of the residual on these images. For an image with $L_1 \times L_2$ pixels, the rms value, η can be defined as in (21). In (21), the value at pixel i, j is given by z_{ij} .

$$\eta = \sqrt{\frac{1}{L_1 L_2} \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} z_{ij}^2} \quad (21)$$

We have evaluated (21) on images centered around CasA and CygA, with 64 by 64 pixels in size, for about 30 different frequencies around 50 MHz. The results can be seen on Fig. 3 for both CasA and CygA. It is clearly seen that for the same number of iterations, the SAGE algorithm performs better. The normal algorithm needs about more iterations to have the same performance as the SAGE algorithm.

VI. CONCLUSIONS

We have presented the application of the SAGE algorithm for calibration of radio interferometric arrays. This is an improvement over the generally used direct optimization methods in performance as well as in computing cost, as seen from the results. We have only given some initial results of using the SAGE algorithm on real data. One of the fundamental assumptions made in this paper was that the noise is white and Gaussian. Future work will address exact characterization of the noise and adaptation of the SAGE algorithm especially

when the noise is non Gaussian. Moreover, future work will address situations where we have more than 2 strong sources.

VII. ACKNOWLEDGMENT

The first author would like to thank ASTRON for kind hospitality and NOVA for support. This work was also supported by LOFAR and SNN. LOFAR is being funded by the European Union, European Regional Development Fund, and by "Samenwerkingsverband Noord-Nederland", EZ/KOMPAS. We acknowledge Ronald Nijboer for reviewing an earlier version of this paper.

REFERENCES

- [1] J.A. Fessler and A.O. Hero, "Space alternating generalized expectation maximization algorithm," *IEEE Trans. on Sig. Proc.*, vol. 42, no. 10, pp. 2664-2677, Oct. 1994.
- [2] K. Levenberg, "A method for the solution of certain non linear problems using least squares," *The Quarterly Jnl. of App. Math.*, vol. 2, pp. 164-168, 1944.
- [3] D. Marquardt, "An algorithm for least squares estimation of nonlinear parameters," *SIAM Jnl. of App. Math.*, vol. 11, pp. 431-441, 1963.
- [4] A.J. Boonstra and A.J. van der Veen, "Gain calibration methods for radio telescope arrays," *IEEE Trans. on Sig. Proc.*, vol. 51, no. 1, pp. 25-38, Jan. 2003.
- [5] A.J. van der Veen, A. Leshem, and A.J. Boonstra, "Array signal processing for radio astronomy," *Experimental Astronomy*, vol. 17, no. 1-3, pp. 231-249, Jun. 2004.
- [6] <http://www.skatelescope.org/>
- [7] <http://www.lofar.org/>
- [8] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Jnl. Royal Stat. Soc. Series B*, vol. 39, no. 1, pp. 1-38, 1977.
- [9] M. Krzywinski, V. Sossi, and T.J. Ruth, "Comparison of FORE, OSEM and SAGE algorithms to 3DRP in 3D PET using phantom and human subject data," *IEEE Trans. on Nuclear Science*, vol. 45, no. 4-2, pp. 1114-1120, Aug. 1999.
- [10] A. Logothetis and C. Carelmalin, "SAGE algorithms for multipath detection and parameter estimation in asynchronous CDMA systems," *IEEE Trans. on Sig. Proc.*, vol. 48, no. 11, pp. 3162-3174, Nov. 2000.
- [11] A.R. Thompson, J.M. Moran and G.W. Swenson Jr., "Interferometry and synthesis in radio astronomy, 2nd ed." *John Wiley and Sons, New York* 1998.
- [12] J.P. Hamaker, J.D. Bregman and R.J. Sault, "Understanding radio polarimetry, paper I," *Astronomy and Astrophysics Supp.*, vol. 117, no. 137, pp. 96-109, Jan. 1996.
- [13] J.P. Hamaker, "Understanding radio polarimetry, paper V," *Astronomy and Astrophysics Supp.*, vol. 456, pp. 395-404, Jan. 2006.
- [14] J. Noordam, "The measurement equation of a generic radio telescope, AIPS++ implementation note 185," ASTRON, Tech. Report, 1996
- [15] S. van der Tol, B. Jeffs, and A.J. van der Veen, "Self calibration for the LOFAR radio astronomical array," *IEEE Trans. Sig. Proc.*, vol. 55, no. 9, pp. 4497-4510, Sep. 2007.
- [16] B. Jeffs, S. van der Tol, and A.J. van der Veen, "Direction dependent self calibration of large distributed sensor arrays," in *Proc. IEEE ICASSP* pp. 1069-1072, 2006
- [17] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust. Speech and Sig. Proc.* vol. 34, no. 4, pp. 477-489, Apr. 1988
- [18] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. Second Int. Symp. on Information Theory*, pp. 267-281, Budapest, 1973.
- [19] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Sig. Proc.* vol. ASSP-32, pp. 387-392, 1985.
- [20] S. Yatawatta, "LOFAR beamshapes and their use in calibration and imaging," ASTRON, Tech. Report, 2007.

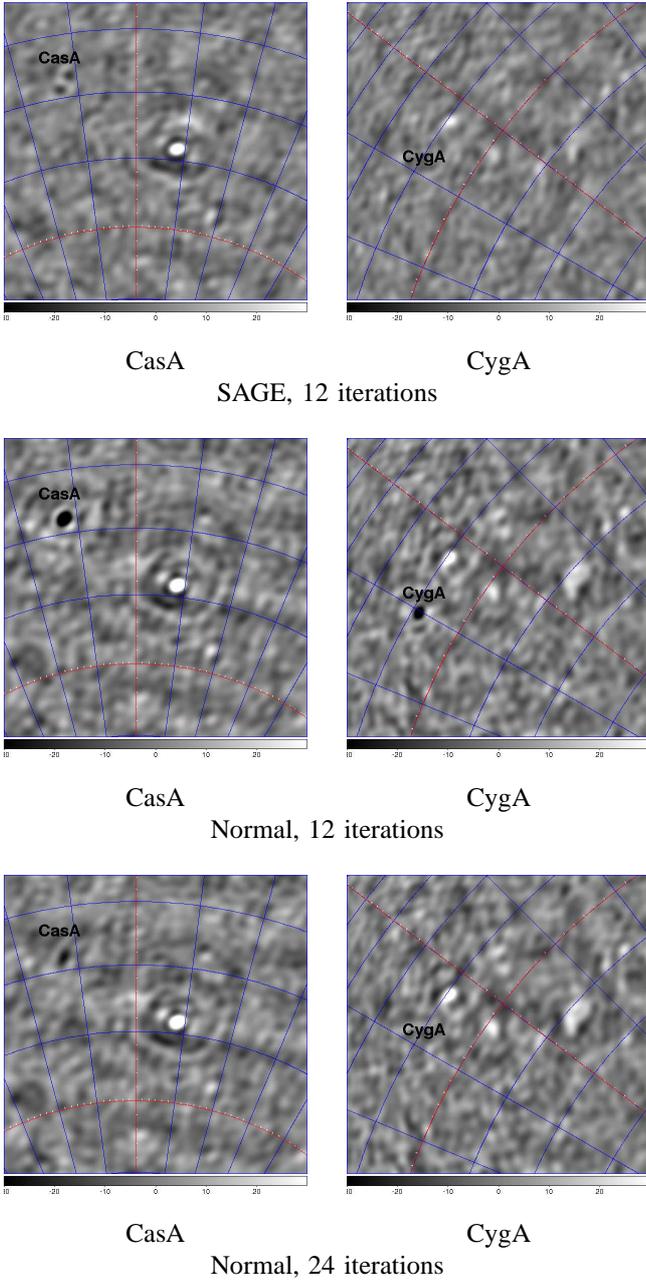


Fig. 2. Images around CasA (left column) and CygA (right column) after applying the SAGE algorithm (first row), Normal algorithm, 12 iterations (second row) and Normal algorithm, 24 iterations (bottom row). The residual of CasA is seen at top left on the images in the left column. The residual of CygA is seen at center left on the images in the right column. The grid lines correspond to sky coordinates: right ascension and declination.

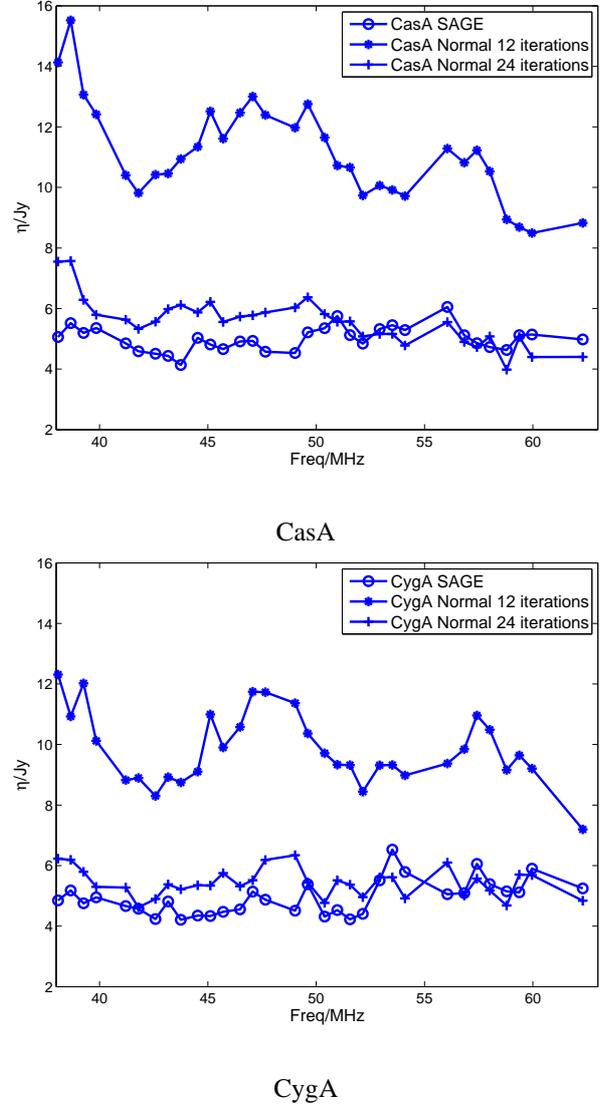


Fig. 3. Pixel rms values around CasA and CygA using the normal calibration algorithm and the SAGE calibration algorithm. For equal number of iterations, the normal algorithm has higher residual compared to the SAGE algorithm. With higher number of iterations, the normal algorithm gives comparable performance.