

ROBUST RADIO INTERFEROMETRIC CALIBRATION

Sarod Yatawatta

Sanaz Kazemi

ASTRON,
The Netherlands Institute for Radio Astronomy
Dwingeloo, The Netherlands
Email: yatawatta@astron.nl

Kapteyn Astronomical Institute
University of Groningen
Groningen, The Netherlands
Email: kazemi@astro.rug.nl

ABSTRACT

Existing calibration algorithms in radio interferometry implicitly assume the noise to be Gaussian. However, outliers in the data due to interference or due to errors in the sky model would have adverse effects on processing based on a Gaussian noise model. Most of the shortcomings of calibration such as the loss in flux or coherence, and the appearance of spurious sources could be attributed to the deviations of the underlying noise model. In this paper, we demonstrate our previous proposal to improve the robustness of calibration by using a noise model based on the Student's t distribution. Unlike Gaussian noise model based calibration, traditional nonlinear least squares minimization would not directly extend to a case when we have a Student's t noise model. Therefore, we use the Expectation-Conditional Maximization Either (ECME) algorithm for calibration. We give simulation results to show the robustness of the proposed calibration method as opposed to traditional Gaussian noise model based calibration, especially in preserving the flux of weaker sources that are not included in the calibration model.

Index Terms— Calibration, Interferometry: Radio interferometry

1. INTRODUCTION

The square kilometre array (SKA) aims to surpass all existing radio interferometers in terms of sensitivity and spatial resolution, especially at low radio frequencies (50 MHz to 1 GHz). This implies building an instrument with hundreds of receivers that will collect data volumes that are greater by orders of magnitude. In order to reach the scientific potential, this data needs to be processed in many aspects. First, the systematic errors in the data (mainly due to ionosphere, troposphere, and the receiver beam shape) needs to be estimated and corrected for. Secondly, the effect of all bright celestial radio sources needs to be subtracted from the data so that the weak signals in the data (wherein the scientific interest lies) are suitable for further scientific study.

We call both these tasks together as calibration. These two tasks are more challenging at low radio frequencies due to two reasons. First, the celestial radio sources are significantly brighter at these frequencies. Secondly, the receiver beam shapes are much wider, thus many more sources are seen. Therefore, calibration at low radio frequencies is computationally demanding, even for conventional radio interferometers. This task is even more daunting for SKA, due to the fact of having many more receivers.

Fortunately, array signal processing methods have enabled us to make progress in tackling the aforementioned problems as work like [1],[2],[3],[4] and [5] show. In all such methods, calibration is con-

sidered as a maximum likelihood (ML) estimation problem. Crucial for such methods is an initial model of the sky, which is iteratively updated in self-calibration. An accurate sky model is difficult to obtain a priori, especially when the interferometric array has wide beam shapes and many bright sources are in the field of view. As studied by [6], [7], and [8], incomplete sky models give rise to various problems in calibration including flux loss of unmodeled sources and the appearance of spurious sources. In this paper, we consider the errors in the initial sky model to be outliers in the data. Therefore, as assumed in existing work, ML estimation under a Gaussian noise model would not perform satisfactorily due to outliers in the data. In particular, we are interested in minimizing the flux loss of unmodeled sources (and noise suppression) due to having an incomplete sky model. This will be our sole criterion in measuring the 'robustness' throughout this paper.

The novelty of the work presented in this paper (relation to prior work) is mainly to demonstrate our previous work [9, 10] in deviating from Gaussian noise model based calibration, as done in existing work [1],[2],[3],[5], with more simulations. As presented in our previous work [9, 10], we proposed to use Student's t distribution [11] based noise model to increase to robustness of calibration. The first attempt in increasing robustness [12] used an l_1 norm based cost function (equivalent to having a Laplacian noise model [13]). We choose the expectation-conditional maximization either (ECME) algorithm [14] to find the ML estimate with Student's t noise model and unlike [12], this gives a computationally affordable calibration algorithm.

The rest of the paper is organized as follows. In section 2, we give an overview of the data model. Next in section 3 we give an overview of the proposed calibration based on Student's t noise model. We give results to show its robustness in section 4 before drawing our conclusions in section 5.

Notation: Matrices and vectors are denoted by bold upper and lower case letters as \mathbf{J} and \mathbf{v} , respectively. The transpose and the Hermitian transpose are given by $(\cdot)^T$ and $(\cdot)^H$, respectively. The matrix Frobenius norm is given by $\|\cdot\|$. The set of real and complex numbers are denoted by \mathbb{R} and \mathbb{C} , respectively. The identity matrix is given by \mathbf{I} . The matrix trace operator is given by $\text{trace}(\cdot)$.

2. RADIO INTERFEROMETRIC DATA MODEL

The overview of radio interferometry in this section is brief and for more information about radio interferometry, the reader is referred to [15], and [16] for the data model in particular. 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. The interferometric array consists of R

receiving elements with dual polarized feeds. 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.

Consider the correlation of signals at the p -th receiver and the q -th receiver, with proper signal delay. After correlation, the correlated signal of the p -th station and the q -th station (named as the *visibilities*), $\mathbf{V}_{pq} \in (\mathbb{C}^{2 \times 2})$ is given by

$$\mathbf{V}_{pq} = \sum_{i=1}^K \mathbf{J}_{pi} \mathbf{C}_{pqi} \mathbf{J}_{qi}^H + \sum_{i'=1}^{K'} \mathbf{J}_{pi'} \mathbf{C}_{pqi'} \mathbf{J}_{qi'}^H + \mathbf{N}_{pq}. \quad (1)$$

In (1), \mathbf{J}_{pi} and \mathbf{J}_{qi} are the Jones matrices describing errors along the direction of source i , at station p and q , respectively. The matrices represent the effects of the propagation medium, the beam shape and the receiver. There are K known sources (that are in the sky model) and K' unknown sources, and the total signal is the superposition of $K + K'$ such signals as in (1). The noise matrix is given as $\mathbf{N}_{pq} \in (\mathbb{C}^{2 \times 2})$. The contribution from the i -th source on baseline pq is given by the coherency matrix $\mathbf{C}_{pqi} \in (\mathbb{C}^{2 \times 2})$. For a linearly polarized source along the i -th direction, with Stokes parameters $I_{pqi}, Q_{pqi}, U_{pqi}, V_{pqi}$, we have

$$\mathbf{C}_{pqi} = e^{j\phi_{pqi}} \begin{bmatrix} I_{pqi} + Q_{pqi} & U_{pqi} + jV_{pqi} \\ U_{pqi} - jV_{pqi} & I_{pqi} - Q_{pqi} \end{bmatrix} \quad (2)$$

where ϕ_{pqi} is the Fourier phase component that depends on the direction in the sky as well as the separation of stations p and q . For baseline coordinates u_{pq}, v_{pq}, w_{pq} and direction cosines of the i -th direction l_i, m_i, n_i , we have $\phi_{pqi} = -2\pi(u_{pq}l_i + v_{pq}m_i + w_{pq}n_i - 1)$. The noise matrix \mathbf{N}_{pq} is assumed to have elements with zero mean, complex Gaussian entries with equal variance in real and imaginary parts. Moreover, in (1), we have split the contribution from the sky into two parts: K sources that are known to us and K' sources that are unknown. Generally, the bright sources are always known but there are infinitely many faint sources that are too weak to be detected and too numerous to be included in the sky model. Therefore, almost always K' is much larger than K .

During calibration, we only estimate the Jones matrices \mathbf{J}_{pi} for $p \in [1, R]$ and $i \in [1, K]$, in other words, we estimate the errors along the known bright sources. Due to our ignorance of the K' unknown sources, the effective noise during calibration becomes

$$\mathbf{N}'_{pq} = \sum_{i'=1}^{K'} \mathbf{J}_{pi'} \mathbf{C}_{pqi'} \mathbf{J}_{qi'}^H + \mathbf{N}_{pq} \quad (3)$$

and our assumption regarding the noise being complex circular Gaussian breaks down, depending on the properties of the signals of the weak sources. The prime motivation of this paper is to address this problem of the possible non-Gaussianity of the noise due to an error in the sky model. A similar situation could arise even for calibration along one direction (or direction independent calibration), when $K = 1$, if there is an error in the source model, for instance in the shape of the source.

The vectorized form of (1), $\mathbf{v}_{pq} = \text{vec}(\mathbf{V}_{pq})$ can be written as

$$\mathbf{v}_{pq} = \sum_{i=1}^K \mathbf{J}_{qi}^* \otimes \mathbf{J}_{pi} \text{vec}(\mathbf{C}_{pqi}) + \sum_{i'=1}^{K'} \mathbf{J}_{qi'}^* \otimes \mathbf{J}_{pi'} \text{vec}(\mathbf{C}_{pqi'}) + \mathbf{n}_{pq} \quad (4)$$

where $\mathbf{n}_{pq} = \text{vec}(\mathbf{N}_{pq})$. Depending on the time and frequency interval within which calibration solutions are obtained, we can stack

up all cross correlations within that interval as

$$\mathbf{d} = [\text{real}(\mathbf{v}_{12}^T) \text{ imag}(\mathbf{v}_{12}^T) \text{ real}(\mathbf{v}_{13}^T) \dots \text{ imag}(\mathbf{v}_{(R-1)R}^T)]^T \quad (5)$$

where \mathbf{d} is a vector of size $N \times 1$ of real data points. Thereafter, we have the data model

$$\mathbf{d} = \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta}) + \sum_{i'=1}^{K'} \mathbf{s}_{i'} + \mathbf{n} \quad (6)$$

where $\boldsymbol{\theta}$ is the real parameter vector (size $M \times 1$) that is estimated by calibration. The contribution of the i -th known source on all data points is given by $\mathbf{s}_i(\boldsymbol{\theta})$ (size $N \times 1$) and the unknown contribution from the i' -th unknown source is given by $\mathbf{s}_{i'}$ (size $N \times 1$). The noise vector based on a Gaussian noise model is given by \mathbf{n} (size $N \times 1$). The parameters $\boldsymbol{\theta}$ are the elements of \mathbf{J}_{pi} -s, with real and imaginary parts considered separately.

The ML estimate of $\boldsymbol{\theta}$ under a zero mean, white Gaussian noise is obtained by minimizing the least squares cost

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{d} - \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta})\|^2 \quad (7)$$

as done in current calibration approaches ([1],[2],[3],[4],[5]). However, due to the unmodeled sources, the effective noise is actually

$$\mathbf{n}' = \sum_{i'=1}^{K'} \mathbf{s}_{i'} + \mathbf{n} \quad (8)$$

even when \mathbf{n} is assumed to be Gaussian.

In order to overcome this problem, we consider the possibility of the noise model not being Gaussian during calibration. Ideally, we should estimate the exact distribution of \mathbf{n}' but in practice this is not feasible, mainly due to non-stationarity. Thus we need to choose a fixed distribution and we select a noise model based on Student's t distribution. Our motivation for this is described in [10] and existing work (e.g. [17], [18], [13]) taking a similar approach also justifies this choice.

3. ROBUST CALIBRATION

In this section, we summarize our previous proposal [10] of the utilization of robust Student's t noise model for calibration. The univariate Student's t distribution ([17], [18]) can be described as follows. Let X be a random variable with a normal distribution $\mathcal{N}(\varepsilon, \sigma^2/\gamma)$ where γ is also a random variable. Then the conditional distribution of X is

$$p(x|\varepsilon, \sigma^2, \gamma) = \frac{1}{(\sigma/\sqrt{\gamma})\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\varepsilon}{\sigma/\sqrt{\gamma}}\right)^2\right). \quad (9)$$

Consider γ to have a Gamma distribution, $\gamma \sim \text{Gamma}(\frac{\nu}{2}, \frac{\nu}{2})$, where $\nu \in [0, \infty]$ is the number of degrees of freedom. The density function of γ can be given as

$$p(\gamma|\nu) = \frac{1}{\Gamma(\frac{\nu}{2})} \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \gamma^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu\gamma}{2}\right). \quad (10)$$

Therefore, the marginal distribution of X is [18]

$$p(x; \varepsilon, \sigma^2, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{(\pi\nu)^{1/2} \Gamma(\frac{\nu}{2}) \sigma} \left(1 + \frac{1}{\nu} \left(\frac{x-\varepsilon}{\sigma}\right)^2\right)^{-\frac{1}{2}(\nu+1)} \quad (11)$$

and this is the probability density function which defines the Student's t distribution. For low values of the number of degrees of freedom ν , Student's t distribution has a higher tail compared to a Gaussian. The asymptotic limit of Student's t distribution is Gaussian as $\nu \rightarrow \infty$.

As discussed in [10], we consider the increase in the noise variance due to the unmodeled sources in (8) as the effect of γ in (9). Therefore, we consider the noise vector \mathbf{n}' to have independent, identically distributed entries, with the distribution given by (11) with $\varepsilon = 0$ and $\sigma = \rho = 1$. We rewrite our data model (6) as

$$\mathbf{d} = \sum_{i=1}^K \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n}' = \mathbf{f}(\boldsymbol{\theta}) + \mathbf{n}' \quad (12)$$

where the unknowns are $\boldsymbol{\theta}$ and ν characterizing the noise \mathbf{n}' . The i -th element of the vector \mathbf{d} (denoted by \mathbf{d}_i) in (12) is assumed to have a distribution as (11) with $\sigma = 1$ and $\mu_i = \mathbf{f}_i(\boldsymbol{\theta})$, where $\mathbf{f}_i(\boldsymbol{\theta})$ is the i -th element of the vector function $\mathbf{f}(\boldsymbol{\theta})$. The likelihood function is

$$l(\boldsymbol{\theta}, \nu | \mathbf{d}) = \prod_{i=1}^N \frac{\Gamma(\frac{\nu+1}{2})}{(\pi\nu)^{1/2} \Gamma(\frac{\nu}{2})} \left(1 + \frac{(\mathbf{d}_i - \mathbf{f}_i(\boldsymbol{\theta}))^2}{\nu} \right)^{-\frac{1}{2}(\nu+1)} \quad (13)$$

and the log-likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta}, \nu | \mathbf{d}) &= N \log \Gamma\left(\frac{\nu+1}{2}\right) - N \log \Gamma\left(\frac{\nu}{2}\right) - \frac{N}{2} \log(\pi\nu) \\ &- \frac{(\nu+1)}{2} \sum_{i=1}^N \log \left(1 + \frac{(\mathbf{d}_i - \mathbf{f}_i(\boldsymbol{\theta}))^2}{\nu} \right). \end{aligned} \quad (14)$$

Since the noise is not Gaussian, minimizing a least squares cost function (or maximizing the likelihood) will not give us the ML estimate. In addition, we have an extra parameter, ν , which is the number of degrees of freedom. Hence, we use the expectation-conditional maximization either algorithm (ECME [14],[19]) to solve this problem. The ECME algorithm is an extension of the EM algorithm for t distribution presented by [17].

The auxiliary variables in the ECME algorithm are the weights w_i (N values) and a scalar λ . All these are initialized to 1 at the beginning. The expectation step in the ECME algorithm involves the conditional estimation of hidden variables γ_i (or the weights w_i) as

$$w_i \leftarrow E\{\gamma_i | \mathbf{d}_i, \boldsymbol{\theta}, \nu\} = \lambda \frac{\nu+1}{\nu + (\mathbf{d}_i - \mathbf{f}_i(\boldsymbol{\theta}))^2} \quad (15)$$

and the update of the scalar λ

$$\lambda \leftarrow \frac{1}{N} \sum_{i=1}^N w_i. \quad (16)$$

The maximization step involves finding the value for ν that is a solution for

$$\begin{aligned} \Psi\left(\frac{\nu+1}{2}\right) - \log\left(\frac{\nu+1}{2}\right) - \Psi(\nu/2) \\ + \log(\nu/2) + \frac{1}{N} \sum_{i=1}^N (\log(w_i) - w_i) + 1 = 0 \end{aligned} \quad (17)$$

where $\Psi(x) = \frac{d}{dx} \log(\Gamma(x))$ is the digamma function. Once w_i is known, \mathbf{d}_i has a normal distribution with variance determined by

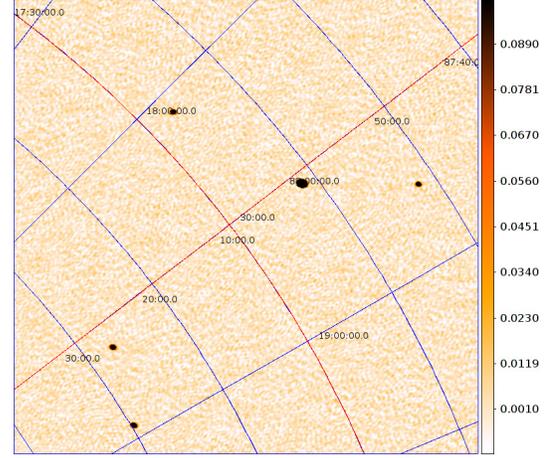


Fig. 1. An area of size 1×1 degrees in the simulated sky, without any corruptions. There are a few weak sources and one strong source in this part of the image.

w_i . Therefore, in the maximization step of the ECME algorithm, we minimize the weighted least squares cost function

$$l(\boldsymbol{\theta} | \nu) = \sum_{i=1}^N w_i (\mathbf{d}_i - \mathbf{f}_i(\boldsymbol{\theta}))^2 \quad (18)$$

to obtain a solution for $\boldsymbol{\theta}$.

The minimization of (18) can be done using well known minimization algorithms such as the Levenberg-Marquardt (LM [20],[21]) algorithm. We give a detailed description of this in [10]. Moreover, in our previous work [5], we have presented the space alternating generalized expectation maximization (SAGE, [22]) algorithm as an efficient and accurate method to solve (7), when the noise model is Gaussian. We have also adopted this to solve (12) with a Student's t noise model and the details can be found in [10].

4. SIMULATIONS

We simulate an interferometer (similar to LOFAR <http://www.lofar.org>) with $R = 47$ stations, observing at 150 MHz. The field of view of the observation is about 12 degrees in diameter. The sources in the sky are assumed to have a power law distribution in their intensities, uniformly distributed over the field of view. In Fig. 1, we show a small area of the true sky (without any errors) where we see bright sources as well as weak sources.

During a 6 hour interferometric observation, where each data sample has 10 s duration, we simulate errors along the brightest sources, and an image with such errors is shown in Fig. 2. In order to recover the true sky, the contribution of the strong sources need to be estimated and subtracted from the data. Once the data is calibrated, we get the image in Fig. 3, where we see the weak sources again.

In our sky, there are 200 sources with 100 sources having flux below 0.7 Jy (flux unit) and the rest between 0.7 and 40 Jy. We vary the number of sources selected for calibration (therefore the number of directions in the sky model) by varying the flux cutoff. The remainder of the sources act as outliers to the data. We introduce errors to the selected sources and add Gaussian noise (20 dB). After calibration, we measure the recovered flux of the 100 sources below



Fig. 2. Image of the sky (same area as Fig. 1) with corruptions. Due to the corruptions and also due to incomplete Fourier sampling, the sidelobe patterns generated by the strong sources completely overwhelm the image and the weak sources are completely obscured.

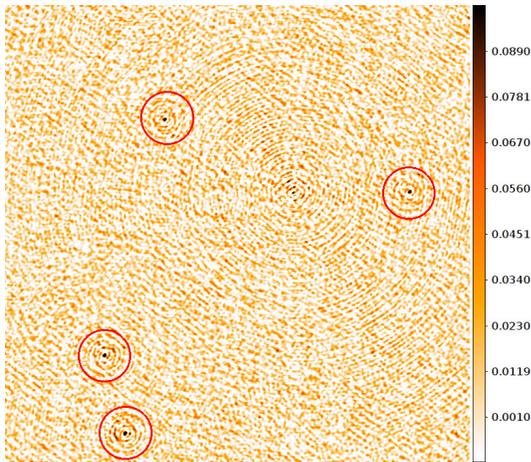


Fig. 3. Calibrated image after estimating the errors along the strong sources and subtracting them from the data. The weak sources (within red circles) are barely visible.

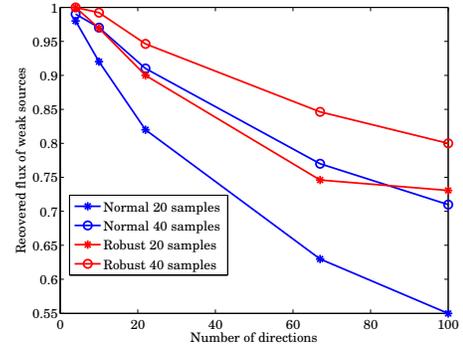


Fig. 4. Performance of robust calibration compared with normal calibration. The number of directions calibrated is increased from 4 to 100. The recovered flux of the weak sources is shown as a ratio to the original flux.

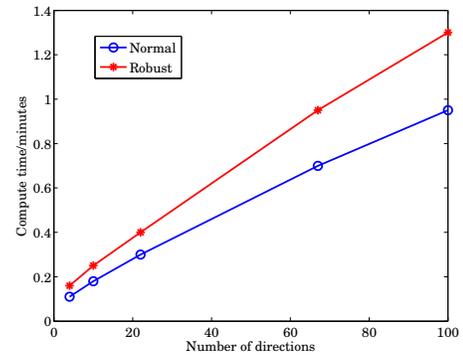


Fig. 5. Computational time required by traditional calibration and robust calibration. Robust calibration uses about 40% more time.

0.7 Jy. In Fig. 4, we have shown the recovered flux as a ratio to original flux after calibration for varying number of directions.

We make several observations from Fig. 4. First, we see that as we increase the number of directions calibrated, the recovered flux of the weak sources (that are not part of the model) decreases. Moreover, we see that robust calibration always has a higher value for recovered flux. In Fig. 4, the calibration is performed per every 20 data samples or per every 40 data samples. As the number of data samples are increased (and also the number of constraints), the flux loss is lowered. Finally, in Fig. 5, we present the computational time spent by normal calibration and robust calibration, for a single calibration run. Both implementations are optimized [23] to use hardware acceleration. We see that robust calibration takes about 40% more computing time.

5. CONCLUSIONS

We have presented increasing the robustness of radio interferometric calibration using a Student's t noise model instead of a Gaussian. This results in an increase in preservation of the flux of weaker background sources that are not included in the data model. Future work would address derivation of analytical bounds for the performance of robust calibration and improving the computational cost [24]. The source code for the software is available at <http://sourceforge.net/projects/sagecal/>.

6. REFERENCES

- [1] A.J. Boonstra and A.J. van der Veen, "Gain calibration methods for radio telescope arrays," *IEEE Trans. Sig. Proc.*, vol. 51, no. 1, pp. 25–38, Jan. 2003.
- [2] 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, 2004.
- [3] 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, Sept. 2007.
- [4] S.J. Wijnholds and A.J. van der Veen, "Multisource self-calibration for sensor arrays," *IEEE Trans. Sig. Proc.*, vol. 57, no. 9, pp. 3512–3532, May 2009.
- [5] S. Kazemi, S. Yatawatta, S. Zaroubi, P. Labropoulos, A.G. de Bruyn, L. Koopmans, and J. Noordam, "Radio interferometric calibration using the SAGE algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 414, no. 2, pp. 1656–1666, 2011.
- [6] T. Cornwell and E. B. Fomalont, "Self-Calibration," in *Synthesis Imaging in Radio Astronomy II*, G. B. Taylor, C. L. Carilli, and R. A. Perley, Eds., 1999, vol. 180 of *Astronomical Society of the Pacific Conference Series*, p. 187.
- [7] I. Martí-Vidal, E. Ros, M.A. Perez-Torres, J.C. Guirado, S. Jiménez-Monferrer, and J.M. Marcaide, "Coherence loss in phase-referenced VLBI observations," *Astronomy and Astrophysics*, vol. 515, pp. A53, 2010.
- [8] I. Martí-Vidal and J.M. Marcaide, "Spurious source generation in mapping from noisy phase-self-calibrated data," *Astronomy and Astrophysics*, vol. 480, pp. 289–295, 2008.
- [9] S. Yatawatta, S. Kazemi, and S. Zaroubi, "Robust radio interferometric calibration using the t-distribution," in *IEEE International Symposium on Signal Processing and Information Technology, Ho Chi Minh City, Vietnam*, Dec. 2012.
- [10] S. Kazemi and S. Yatawatta, "Robust radio interferometric calibration using the t-distribution," *Monthly Notices of the Royal Astronomical Society*, vol. 435, pp. 597–605, Oct. 2013.
- [11] W. (Student) Gosset, "The probable error of mean," *Biometrika*, vol. 6, pp. 1–25, 1908.
- [12] F.R. Schwab, "Robust solution for antenna gains," *Very Large Array (VLA) Scientific Memorandum 136*, pp. 1–20, 1982.
- [13] A Aravkin, M. Friedlander, and T. van Leeuwen, "Robust inversion via semistochastic dimensionality reduction," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 5425–5428.
- [14] Chuanhai Liu and D.B. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, , no. 5, pp. 19–39, 1995.
- [15] A.R. Thompson, J.M. Moran, and G.W. Swenson, *Interferometry and synthesis in radio astronomy (3rd ed.)*, Wiley Interscience, New York, 2001.
- [16] 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, 1996.
- [17] K.L. Lange, R.J.A. Little, and J.M.G. Tylor, "Robust statistical modeling using the t distribution," *J. of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, Dec. 1989.
- [18] Anna Bartkowiak, "Should normal distribution be normal? the student's t alternative," in *6th Int. Conf. on Computer Information Systems and Industrial Management Applications (CISIM'07)*, 2007, pp. 3–8.
- [19] Shaojun Li, Hong Wang, and Tianyou Chai, "A t-distribution based particle filter for target tracking," in *American Control Conference*, 2006, p. 6.
- [20] K. Levenberg, "A method for the solution of certain non linear problems using least squares," *The Quarterly J. of App. Math.*, vol. 2, pp. 164–168, 1944.
- [21] D. Marquardt, "An algorithm for least squares estimation of nonlinear parameters," *SIAM J. of App. Math.*, vol. 11, pp. 431–441, 1963.
- [22] 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.
- [23] S. Yatawatta, S. Kazemi, and S. Zaroubi, "GPU accelerated nonlinear optimization in radio interferometric calibration," in *Proceedings of Innovative Parallel Computing (InPar'12)*, 2012.
- [24] S. Yatawatta, "Radio interferometric calibration using a riemannian manifold," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 3866–3870.