UNIVERSITY OF CAMBRIDGE

MPhil in Data Intensive Science

Feb. 2024

Astronomy in the SKA era: SKA-low mini project

Dr. Quentin Gueuning (qdg20) and Dr. Daniel Molnar (dm611)

The coursework will be submitted via a GitLab repository which we will create for you. You should place all your code and your report in this repository. The report should be in PDF format in a folder called "report". You will be provided access to the repository until 11:59pm on Thursday the 28th of March. After this you will lose access which will constitute submission of your work. The weighting of marks for each question is indicated in brackets, up to a maximum of 100 marks.

The code associated with the coursework should be written in Python or C++ and follow best software development practice as defined by the Research Computing module.

This should include:

- *Writing clear readable code that is compliant with a common style guide and uses suitable build management tools.*
- *Providing appropriate documentation that is compatible with auto-documentation tools like Doxygen/Sphinx.*
- *The project must be well structured (sensible folder structure, README.md, licence etc..) following standard best practice.*
- *Appropriate and robust unit tests for automatic validation of your code.*
- *Uses appropriate version control best practice, including branching for development and testing, and commit hooks to protect "main".*
- *Appropriate containerisation to ensure portability of the project to other computers and operating systems.*

1 Gain calibration of a SKA-low station

(a) Introduction

Calibration is an essential part of radio astronomy telescopes and their receiver systems and ensures that known instrumental effects can be taken into account, i.e. calibrated out when recovering signals received by the antenna. The calibration of the aperture array stations within the Square Kilometer Array (SKA) telescope is a critical aspect in achieving optimal performances, i.e. dynamic range and sensitivity. This mini project focuses on implementing an algorithm for the retrieval of gain solutions for a single SKA-low station, which comprises 256 antennas.

The SKA1-low instrument is designed with 512 stations, each featuring 256 elements and covering the frequency range of 50-350 MHz. The analog chain associated with log-periodic antennas includes low-noise amplifiers and other analog components connected via coaxial cables or optical fibers to remote analog-to-digital converters. A typical RF system is illustrated in Fig. 1. Digital station beamforming is an integral part of the SKA1-low telescope and relies on a precise calibration of the analog chain. The propagation of the signal through the analog chain is simply modeled as a series of linear transformations applied to the input signal, and can thus be accounted for by a complex-valued gain per antenna. A key calibration technique, known as 'self-calibration,' utilizes a model of sky temperatures and embedded element patterns (EEPs) to determine analog gains. The term 'self' indicates that the system is calibrated internally without relying on external devices. Most algorithms implementing self-calibration formulate the problem as a quadratic system of equations, using the correlation matrix containing all pairs of cross-correlations, or "visibilities", between the antenna voltages.

Figure 1: 64-element SKA-low station at the Mullard Radio Astronomy Observatory (MRAO) in Cambridge, UK.

This mini project, drawing references from lectures 5-6 of the module, aims to guide you in implementing an algorithm capable of retrieving gain solutions for a single SKA-low station. It is recommended to revisit definitions from lecture 5, particularly those related to embedded element patterns (EEPs), station beam, beamforming, and station gain calibration.

(b) **Background**

In this exercise, we assume that the unknown complex-valued gains for feed X and Y of any antenna i are identical, i.e., $G_{ix} = G_{iy} = G_i$. The voltage at the input of the Analog-to-Digital converter associated with antenna i at a given frequency f is given by

$$v_{i,p} = G_i \, \boldsymbol{F}_{i,p}(\theta, \phi) \cdot \boldsymbol{E}(\theta, \phi) \tag{1}$$

where θ and ϕ are the zenith angle and the azimuth angle, respectively, and $\boldsymbol{E} = E_\theta \hat{\boldsymbol{\theta}} + E_\phi \hat{\boldsymbol{\phi}}$ is an incident electric field with spherical components E_θ, E_ϕ, and $\boldsymbol{F}_{i,p}$ is the EEP associated with feed ports $p = \{X, Y\}$ of antenna i. Spherical coordinates used here are shown in Fig. 2.

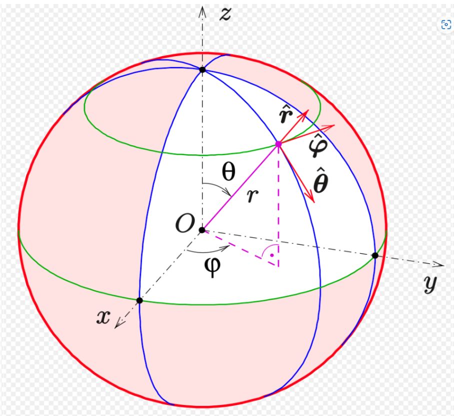

Figure 2: Spherical system of coordinates

The measured visibility $R_{ij,p}$ represents a time cross-correlation between measured voltages $v_{i,p}$ and $v_{j,p}$ and can be modeled as $R_{ij,p} = G_i G_j^* M_{ij,p}$ with the model visibility $M_{ij,p}$ expressed as

$$M_{ij,p} = \iint \Big(\boldsymbol{F}_{i,p}(\theta, \phi) \cdot \boldsymbol{F}^*_{j,p}(\theta, \phi) \Big) T_b(\theta, \phi) e^{-j\boldsymbol{k}\cdot(\boldsymbol{r}_i - \boldsymbol{r}_j)} \sin\theta \mathrm{d}\theta \mathrm{d}\phi \tag{2}$$

where * means complex conjugate, T_b is the sky brightness temperature, $\boldsymbol{r}_i$ is the position of antenna i, and $\boldsymbol{k} = k \sin\theta \cos\phi \hat{\boldsymbol{x}} + k \sin\theta \sin\phi \hat{\boldsymbol{y}} + k \cos\theta \hat{\boldsymbol{z}}$ is the wavevector with the wavenumber k. The distribution of diffuse emission used in this project is illustrated in Fig. 3 in Jansky units.

In self-calibration algorithms, we aim to fit measured visibilities $R_{ij,p}$ with the model $M_{ij,p}$ that incorporates known EEPs and sky models to retrieve the

(TURN OVER

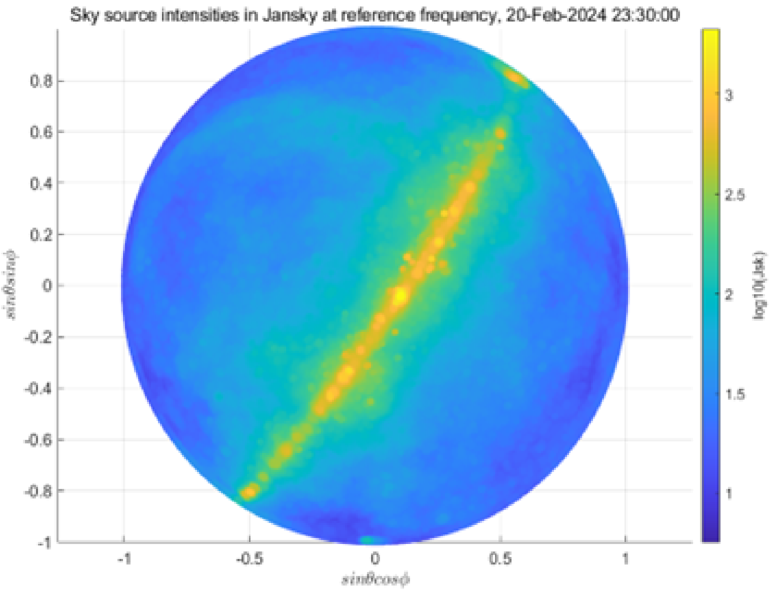

Figure 3: Diffuse emission intensities at 100 MHz, 23 february 2024, 23.30 UTC0

unknown gains G_i. By summing up the measured visibilities $R_{ij} = R_{ij,X} + R_{ij,Y}$ and the modelled ones $M_{ij} = M_{ij,X} + M_{ij,Y}$ and by grouping them into a covariance matrix $\boldsymbol{R}$ and a model matrix $\boldsymbol{M}$, we can re-write the calibration equations in matrix form as:

$$\boldsymbol{R} = \boldsymbol{G}\boldsymbol{M}\boldsymbol{G}^H \tag{3}$$

where H denotes the Hermitian transpose, and $\boldsymbol{G}$ is a diagonal matrix with the ith diagonal element equal to G_i. In practice, the autocorrelations R_{ii} are strongly contaminated by the system noise and are thus set to zero. The solution to this system of equations can be expressed as an optimization problem, with

$$\hat{\boldsymbol{G}} = \text{argmin}_{G_i} ||\boldsymbol{R} - \boldsymbol{G}\boldsymbol{M}\boldsymbol{G}^H||_F \tag{4}$$

where $\hat{\boldsymbol{G}}$ is an estimator of the gain solutions, and F denotes the Frobenius norm.

(c) **Questions**

1. Characterize the system of equations (4) and the properties of the matrix $\boldsymbol{M}$. What happens to the residual error when you multiply all gains by the same phase factor $e^{j\phi}$? (5 points)
2. Using the provided function "compute_EEPs," plot the power of the EEPs and the average element pattern (AEP) in dBVolts units ($dBV = 20\log_{10} V$) in the plane $\phi = 0, \theta = [-90, 90]deg$. Comment on variability. (5 points)

3. An algorithm known as "StEFCal" solves the system in (4) and estimates the per-antenna complex-valued gain iteratively. Implement the algorithm as described in Fig. 4. Why does the algorithm take, each even iteration i, the

Algorithm 1 Algorithm StEFCal

Initiate $\mathbf{G}^{[0]}$; $\mathbf{G}^{[0]} = \mathbf{I}$ is adequate in most cases
for $i = 1, 2, \cdots, i_{\max}$ **do**
  for $p = 1, 2, \cdots, P$ **do**
    $\mathbf{z} \leftarrow \mathbf{G}^{[i-1]} \cdot \mathbf{M}_{:,p} \equiv \mathbf{g}^{[i-1]} \odot \mathbf{M}_{:,p}$
    $g_p \leftarrow (\widehat{\mathbf{R}}_{:,p}^{H} \cdot \mathbf{z}) / (\mathbf{z}^H \cdot \mathbf{z})$
  end for
  if $\mathrm{mod}_2(i) = 0$ **then**
    if $\|\mathbf{g}^{[i]} - \mathbf{g}^{[i-1]}\|_F / \|\mathbf{g}^{[i]}\|_F \leq \tau$ **then**
      Convergence reached
    else
      $\mathbf{G}^{[i]} \leftarrow (\mathbf{G}^{[i]} + \mathbf{G}^{[i-1]})/2$
    end if
  end if
end for

Figure 4: Picture taken from S. Salvini and S.J. Wijnholds, *Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications*, https://arxiv.org/pdf/1410.2101.pdf

average between the current gain solution and the gain solution of the previous odd iteration ? Comment on convergence rate. (30 points). Optimize the algorithm for the fewest iterations (10 points).

4. Two model matrices $\boldsymbol{M}$ are provided, $\boldsymbol{M}_{AEP}$ and $\boldsymbol{M}_{EEPs}$. In $\boldsymbol{M}_{AEP}$, we used only the AEP to compute the sky integral (2) while $\boldsymbol{M}_{EEPs}$ incorporates all EEPs. Using the exact solution $\boldsymbol{G}$ provided, plot the absolute error of the gain solutions as a function of iteration number, as well as the absolute error in their amplitude and phase separately. What is the minimum error achieved using both model matrices? (20 points)

5. Using your most accurate gain solutions (or using the provided estimations $\hat{\boldsymbol{G}}_{AEP}$ and $\hat{\boldsymbol{G}}_{EEPS}$ if you haven't been able to complete questions 3 and 4), calibrate out the unknown gains G_i from the received voltages and then beamform all 256 voltages to zenith. Plot the power (in dBV) of the beamformed voltages in the plane $\phi = 0$ using the three different gain solutions: provided true solutions, and those obtained using $\boldsymbol{M}_{AEP}$ and $\boldsymbol{M}_{EEPs}$. Comment on achieved reconstruction errors. (20 points)

6. Plot your most accurate calibrated station beam, steering at direction $(\theta, \phi) = (40, 80)$, in a "3D" plot in sine-cosine coordinates. (10 points)

END OF PAPER