



Publication Year	2018
Acceptance in OA	2020-10-14T09:51:10Z
Title	Redundant interferometric calibration as a complex optimization problem
Authors	Grobler, T. L., BERNARDI, GIANNI, Kenyon, J. S., Parsons, A. R., Smirnov, O. M.
Publisher's version (DOI)	10.1093/mnras/sty357
Handle	http://hdl.handle.net/20.500.12386/27793
Journal	MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY
Volume	476

Redundant interferometric calibration as a complex optimization problem

T. L. Grobler,^{1,2★} G. Bernardi,^{2,3,4★} J. S. Kenyon,² A. R. Parsons^{5,6}
and O. M. Smirnov^{2,3}

¹Department of Mathematical Sciences, Computer Science Division, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

²Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa

³SKA SA, 3rd Floor, The Park, Park Road, Pinelands 7405, South Africa

⁴INAF-Istituto di Radioastronomia, via Gobetti 101, I-40129 Bologna, Italy

⁵Department of Astronomy, University of California, Berkeley, CA 94720, USA

⁶Radio Astronomy Laboratory, University of California, Berkeley, CA 94720, USA

Accepted 2018 February 7. Received 2018 February 7; in original form 2017 December 18

ABSTRACT

Observations of the redshifted 21 cm line from the epoch of reionization have recently motivated the construction of low-frequency radio arrays with highly redundant configurations. These configurations provide an alternative calibration strategy – ‘redundant calibration’ – and boost sensitivity on specific spatial scales. In this paper, we formulate calibration of redundant interferometric arrays as a complex optimization problem. We solve this optimization problem via the Levenberg–Marquardt algorithm. This calibration approach is more robust to initial conditions than current algorithms and, by leveraging an approximate matrix inversion, allows for further optimization and an efficient implementation (‘redundant STEFAL’). We also investigated using the preconditioned conjugate gradient method as an alternative to the approximate matrix inverse, but found that its computational performance is not competitive with respect to ‘redundant STEFAL’. The efficient implementation of this new algorithm is made publicly available.

Key words: instrumentation: interferometers – methods: data analysis – methods: numerical – techniques: interferometric – cosmology: observations.

1 INTRODUCTION

The quest for the redshifted 21 cm line from the epoch of reionization (EoR) is a frontier of modern observational cosmology and has motivated the construction of a series of new interferometric arrays operating at low frequencies over the last decade. EoR measurements are challenging because of the intrinsic faintness of the cosmological signal (see for instance, Furlanetto 2016; McQuinn 2016, for recent reviews) buried underneath foreground emission, which is a few orders of magnitude brighter than the EoR anywhere in the sky (e.g. Bernardi et al. 2009, 2010; Ghosh et al. 2012; Dillon et al. 2014; Parsons et al. 2014). As miscalibrated foreground emission leads to artefacts that can jeopardize the EoR signal (e.g. Grobler et al. 2014; Barry et al. 2016; Ewall-Wice et al. 2017), an exceptional interferometric calibration is required. Such calibration needs accurate knowledge of both the sky emission and the instrumental response (e.g. Smirnov 2011b). A skymodel is required for standard calibration as it would be an ill-posed problem if one tries to directly solve for all of the unknowns, i.e. the antenna gains and the uncorrupted visibilities, without first attempting to simplify the

calibration problem. When the uncorrupted visibilities, however, are predicted from a predefined skymodel, the problem simplifies and becomes solvable. In contrast, if an array is deployed in a redundant configuration, i.e. with receiving elements placed on regular grids so that multiple baselines measure the same sky brightness emission, we circumvent the need for a skymodel altogether. This is true, since redundancy leads to fewer unknowns (i.e. the number of unique *uv*-modes) and if the array is redundant enough, it reduces the number of unknowns to such an extent that the calibration problem becomes solvable without having to predict the uncorrupted visibilities from a skymodel. Redundancy, therefore, is a promising path to achieve highly accurate interferometric calibration (Noordam & De Bruyn 1982; Pearson & Readhead 1984; Wieringa 1992; Liu et al. 2010; Noorishad et al. 2012; Marthi & Chengalur 2014; Sievers 2017). Moreover, redundant configurations provide a sensitivity boost on particular spatial scales that can be tuned to be the highest signal-to-noise ratio (SNR) EoR modes (Parsons et al. 2012; Dillon & Parsons 2016). These reasons motivated the design and deployment of EoR arrays in redundant configurations like the MIT-EoR (Zheng et al. 2014), the Precision Array to Probe the Epoch of Reionization (Ali et al. 2015), and the Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017).

* E-mail: tlgrobler@sun.ac.za (TLG); giannibernardi75@gmail.com (GB)

Table 1. Notation and frequently used symbols.

$x, \mathbf{x},$ and $\mathbf{X}$	Scalar, vector, and matrix
$\mathbb{N}$	Set of natural numbers
$\mathbb{C}$	Set of complex numbers
$\mathbf{I}$	Identity matrix
$\alpha_{pq}, \phi_{pq}, \zeta_{pq}, \xi_{pq},$ and $\psi_{pq} : \mathbb{N}^2 \rightarrow \mathbb{N}$	Indexing functions
$\bar{x}$	Conjugation
$\bar{x}, \bar{\mathbf{X}}$	Element-wise conjugation
x^2	Square
$\ \mathbf{x}\ _F$	Frobenius norm
$\mathbf{X}^{-1}$	Matrix inversion
$\odot$	Hadamard product (element-wise product)
$\mathbf{X}^H$	Hermitian transpose
$\mathbf{x}^T, \mathbf{X}^T$	Transpose
$\langle \mathbf{x} \rangle_x$	Averaging over x
$\frac{\partial}{\partial z}$	Wirtinger derivative
$\hat{x}$	A new quantity derived from x
$\hat{x}$	Estimated quantity
$\tilde{\mathbf{x}}$	Augmented vector $[\mathbf{x}^T, \bar{\mathbf{x}}^T]^T$
i	$\sqrt{-1}$
$N, B, L,$ and P	Number of antennas, baselines, redundant groups, and parameters
$\mathbf{d}, \mathbf{v}, \mathbf{g}, \mathbf{y}, \mathbf{r},$ and $\mathbf{n}$	Data, predicted, antenna, true visibility, residual, and noise vector
$\mathbf{z}$	$\mathbf{z} = [\mathbf{g}^T, \mathbf{y}^T]^T$
Λ	Objective function
λ and ρ	Damping factor and $\rho = \frac{1}{1+\lambda}$
$\mathbf{J}$	Jacobian matrix
$\mathbf{H}$	$\mathbf{J}^H \mathbf{J}$ (Hessian matrix)
$\mathcal{H}$	Modified Hessian matrix
$[\mathbf{X}]_{ij}$	Element ij of matrix $\mathbf{X}$
x_{ij}	Composite antenna index
x_i	Antenna or redundant group index
x^k	Iteration number
$\mathbf{M}$	Preconditioner matrix
m	Number of non-zero entries in a matrix
κ	Spectral condition number of a matrix
γ	Sparsity factor of a matrix
Δx	Parameter update
$\exists!$	There exists a unique

In this paper, we present a new algorithm to calibrate redundant arrays based on the complex optimization formalism recently introduced by Smirnov & Tasse (2015). With respect to current algorithms, it is more robust to initial conditions, while remaining comparatively fast. We also show that given certain approximations this new algorithm reduces to the redundant calibration equivalent of the STEPCAL algorithm (Salvini & Wijnholds 2014). We investigate the speed-up that the preconditioned conjugate gradient (PCG) method provides if it is employed by the new algorithm (Liu et al. 2010). A comparison between the computational complexity of the optimized new algorithm and redundant STEPCAL is also performed.

A summary of the notation that is used within the paper is presented in Table 1. We also discuss the overset notation used within the paper in a bit more detail below (we use x as an operand proxy in this paper):

(i) $\hat{x}$ – This notation is used to denote a new scalar value that was derived from the scalar x using a proper mathematical definition.

(ii) $\bar{x}$ – This notation denotes the conjugation of its operand (the conjugation of x). For the readers' convenience, we will redefine this operator when it is first used within the paper.

(iii) $\hat{x}$ – This notation denotes that the quantity is an estimated value.

(iv) $\tilde{\mathbf{x}}$ – This notation denotes an augmented vector, i.e. $\tilde{\mathbf{x}} = [\mathbf{x}^T, \bar{\mathbf{x}}^T]^T$.

The paper is organized as follows: we review the complex calibration formalism by Smirnov & Tasse (2015) in Section 2, we extend it to the redundant case in Section 3, we present some computational complexity results in Section 4, and we present our conclusions in Section 5.

2 WIRTINGER CALIBRATION

In a radio interferometer, the true sky visibilities y_{pq} measured by a baseline formed by antenna p and q are always ‘corrupted’ by the non-ideal response of the receiver, which is often incorporated into a single, receiver-based complex number g (i.e. an antenna gain). The observed visibility d_{pq} is therefore given by (Hamaker, Bregman & Sault 1996; Sault, Hamaker & Bregman 1996; Smirnov 2011a)

$$d_{pq} = g_p \bar{g}_q y_{pq} + n_{pq}, \quad (1)$$

where $\bar{x}$ indicates complex conjugation and n_{pq} is the thermal noise component. The real and imaginary components of the thermal noise are normally distributed with a mean of zero and a standard deviation σ :

$$\sigma \propto \frac{T_{\text{sys}}}{\sqrt{\Delta\nu\tau}}, \quad (2)$$

where T_{sys} is equal to the system temperature, $\Delta\nu$ is the observational bandwidth, and τ is the integration time per visibility.

Considering the number of visibilities B (i.e. baselines) measured by an array of N elements $B = \frac{N^2-N}{2}$, equation (1) can be expressed in the following vector form:

$$\mathbf{d} = \mathbf{v} + \mathbf{n}, \quad (3)$$

where

$$[\mathbf{d}]_{\alpha_{pq}} = d_{pq}, \quad [\mathbf{v}]_{\alpha_{pq}} = v_{pq} = g_p y_{pq} \bar{g}_q, \quad [\mathbf{n}]_{\alpha_{pq}} = n_{pq}, \quad (4)$$

and

$$\alpha_{pq} = \begin{cases} (q-p) + (p-1)(N-\frac{1}{2}p) & \text{if } p < q \\ 0 & \text{otherwise} \end{cases}. \quad (5)$$

The function α_{pq} therefore maps composite antenna indices to unique single indices, i.e.

$$\{\alpha_{12}, \alpha_{13}, \dots, \alpha_{N-1N}\} = \{1, 2, \dots, B\}. \quad (6)$$

The vectors in equation (4) are column vectors of size B (i.e. $p < q$). It is important to point out here that all of the mathematical definitions in this paper assume that we are only considering composite indices belonging to the set $\{rs | r < s\}$. For the sake of mathematical completeness, however, α_{pq} is defined for composite indices that do not belong to this set. Also note that, while the indexing function α_{pq} does not attain the value zero, in practice the zero indexing value does play a very important role in some of the definitions used within this paper (see for example equation A10).

Radio interferometric calibration aims to determine the best estimate of $\mathbf{g} = [g_1, g_2, \dots, g_N]^T$ in order to correct the data and,

following equation (3), can be formulated as a non-linear least-squares optimization problem:

$$\min_{\mathbf{g}} \Lambda(\mathbf{g}) = \min_{\mathbf{g}} \|\mathbf{r}\|_{\mathbb{F}}^2 = \min_{\mathbf{g}} \|\mathbf{d} - \mathbf{v}(\mathbf{g})\|_{\mathbb{F}}^2, \quad (7)$$

where Λ is the objective function, $\mathbf{r}$ is the residual vector, and $\|\mathbf{x}\|_{\mathbb{F}}$ denotes the Frobenius norm. In standard interferometric calibration, y_{pq} is assumed to be known at some level, for instance through the observation of previously known calibration sources.

Non-linear least-squares problems are generally solved by using gradient-based minimization algorithms [i.e. Gauss–Newton (GN) or Levenberg–Marquardt (LM)] that require the model $[\mathbf{v}]$ in equation (7)] to be differentiable towards each parameter. When the least-squares problem is complex, it becomes less straightforward to apply these gradient-based minimization methods, as many complex functions are not differentiable if the classic notion of differentiation is used, i.e. $\frac{\partial \bar{z}}{\partial z}$ does not exist if $z \in \mathbb{C}$.

In order to circumvent the differentiability conundrum associated with complex least-squares problems, standard interferometric calibration divides the complex optimization problem into its real and imaginary parts and solves for the real and imaginary parts of the unknown model parameters separately. Smirnov & Tasse (2015) showed, however, that this approach is not needed if complex calculus (Wirtinger 1927) is adopted. The Wirtinger derivatives are defined as

$$\frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right), \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right), \quad (8)$$

which lead to the following relations:

$$\frac{\partial z}{\partial z} = 1, \quad \frac{\partial \bar{z}}{\partial z} = 0, \quad \frac{\partial z}{\partial \bar{z}} = 0, \quad \frac{\partial \bar{z}}{\partial \bar{z}} = 1. \quad (9)$$

If the gradient operator is defined using equation (8), the model $\mathbf{v}$ now becomes analytic in both $\mathbf{g}$ and $\bar{\mathbf{g}}$, and equation (9) can be used to derive the complex variants of the real-valued GN and LM algorithms. In the complex GN and LM algorithms, complex parameters and their conjugates are treated as separate variables.

Assuming that y_{pq} is known, equation (7) is recast as (Smirnov & Tasse 2015)

$$\min_{\check{\mathbf{g}}} \Lambda(\check{\mathbf{g}}) = \min_{\check{\mathbf{g}}} \|\check{\mathbf{r}}\|_{\mathbb{F}}^2 = \min_{\check{\mathbf{g}}} \|\check{\mathbf{d}} - \check{\mathbf{v}}(\check{\mathbf{g}})\|_{\mathbb{F}}^2, \quad (10)$$

where $\check{\mathbf{r}} = [\mathbf{r}^T, \bar{\mathbf{r}}^T]^T$, $\check{\mathbf{d}} = [\mathbf{d}^T, \bar{\mathbf{d}}^T]^T$, $\check{\mathbf{v}} = [\mathbf{v}^T, \bar{\mathbf{v}}^T]^T$, and $\check{\mathbf{g}} = [\mathbf{g}^T, \bar{\mathbf{g}}^T]^T$.

The complex GN update is therefore defined as

$$\Delta \check{\mathbf{g}} = (\mathbf{J}^H \mathbf{J})^{-1} \mathbf{J}^H \check{\mathbf{r}}, \quad (11)$$

with

$$\mathbf{J} = \begin{bmatrix} \mathcal{J}_1 & \mathcal{J}_2 \\ \bar{\mathcal{J}}_2 & \bar{\mathcal{J}}_1 \end{bmatrix}, \quad (12)$$

and

$$[\mathcal{J}_1]_{\alpha_{pq}, i} = \frac{\partial v_{pq}}{\partial g_i}, \quad [\mathcal{J}_2]_{\alpha_{pq}, i} = \frac{\partial v_{pq}}{\partial \bar{g}_i}. \quad (13)$$

The matrix $\mathbf{J}$ is generally referred to as the Jacobian matrix.

The complex LM update is very similar, the major difference being the introduction of a damping parameter, λ :

$$\Delta \check{\mathbf{g}} = (\mathbf{J}^H \mathbf{J} + \lambda \mathbf{D})^{-1} \mathbf{J}^H \check{\mathbf{r}}, \quad (14)$$

where $\mathbf{D} = \mathbf{I} \odot \mathbf{J}^H \mathbf{J}$.

¹ A matrix composed of first-order partial derivatives.

Table 2. The dimensions of the Jacobian matrices, and their respective sub-matrices, defined in Sections 2 and 3.

Matrix	Dimension
Wirtinger calibration	
$\mathbf{J}$	$2B \times 2N$
$\mathcal{J}_1$	$B \times N$
$\mathcal{J}_2$	$B \times N$
Redundant Wirtinger calibration	
$\mathbf{J}$	$2B \times P$
$\mathcal{J}_1$	$B \times (N + L)$
$\mathcal{J}_2$	$B \times (N + L)$

In this paper, we use single subscript indices (e.g. i) to refer to a specific antenna (or as will become apparent later a specific redundant group) and composite subscript indices (e.g. pq) to refer to a specific baseline. If these indices appear in the definition of matrix elements, then their allowed ranges are determined by the dimension of the matrix that is being defined (see Table 2). Furthermore, the identity matrix is denoted by $\mathbf{I}$. Moreover, the Hadamard² product and the Hermitian transpose are denoted by $\odot$ and $\mathbf{X}^H$, respectively (Liu & Trenkler 2008). Note the use of the Wirtinger derivatives in equation (13). We will refer to $\mathbf{J}^H \mathbf{J}$ as the Hessian³ matrix $\mathbf{H}$ and to $\mathbf{J}^H \mathbf{J} + \lambda \mathbf{D}$ as the modified Hessian matrix $\mathcal{H}$ throughout this paper (Madsen, Nielsen & Tingleff 2004).

Equation (11) or (14) can now be used iteratively to update the parameter vector $\check{\mathbf{g}}$:

$$\check{\mathbf{g}}_{k+1} = \check{\mathbf{g}}_k + \Delta \check{\mathbf{g}}_k, \quad (15)$$

until convergence is reached.

In the case of the GN algorithm, the parameter update step simplifies and becomes (Smirnov & Tasse 2015)

$$\check{\mathbf{g}}_{k+1} = (\mathbf{J}^H \mathbf{J})^{-1} \mathbf{J}^H \check{\mathbf{d}} + \frac{1}{2} \check{\mathbf{g}}_k. \quad (16)$$

Smirnov & Tasse (2015) realized that the diagonal entries of the Hessian matrix $\mathbf{H}$ are much more significant than its off-diagonal entries, i.e. $\mathbf{H}$ is nearly diagonal. By approximating $\mathbf{H}$ by its diagonal and substituting the approximate Hessian matrix, the LM parameter update step becomes (Smirnov & Tasse 2015)

$$\begin{aligned} \check{\mathbf{g}}_{k+1} &\approx \frac{1}{1 + \lambda} \tilde{\mathbf{H}}^{-1} \mathbf{J}^H \check{\mathbf{d}} + \frac{\lambda}{1 + \lambda} \check{\mathbf{g}}_k, \\ &= \rho \tilde{\mathbf{H}}^{-1} \mathbf{J}^H \check{\mathbf{d}} + (1 - \rho) \check{\mathbf{g}}_k, \end{aligned} \quad (17)$$

where $\rho = \frac{1}{1 + \lambda}$. Note that equations (16) and (17) are not dependent on $\check{\mathbf{r}}$.

Interestingly enough, if $\lambda = 0$ we obtain the odd parameter update step of STEFAL⁴, and if $\lambda = 1$ (which corresponds to $\rho = \frac{1}{2}$), we obtain the even parameter update step of STEFAL⁵ (Mitchell et al. 2008; Salvini & Wijnholds 2014). In the STEFAL algorithm, the measurement equation (equation 1) is linearized by assuming that the gains are known, but that their conjugates are not. Under this assumption, the system of equations become linear and the conjugates of the gains can be obtained in a straightforward manner. Starting from the latest value of the gain conjugates, an updated estimate of

² Element-wise product.

³ A square matrix composed of second-order partial derivatives.

⁴ $k \in \{0, 2, \dots\}$.

⁵ $k \in \{1, 3, \dots\}$.

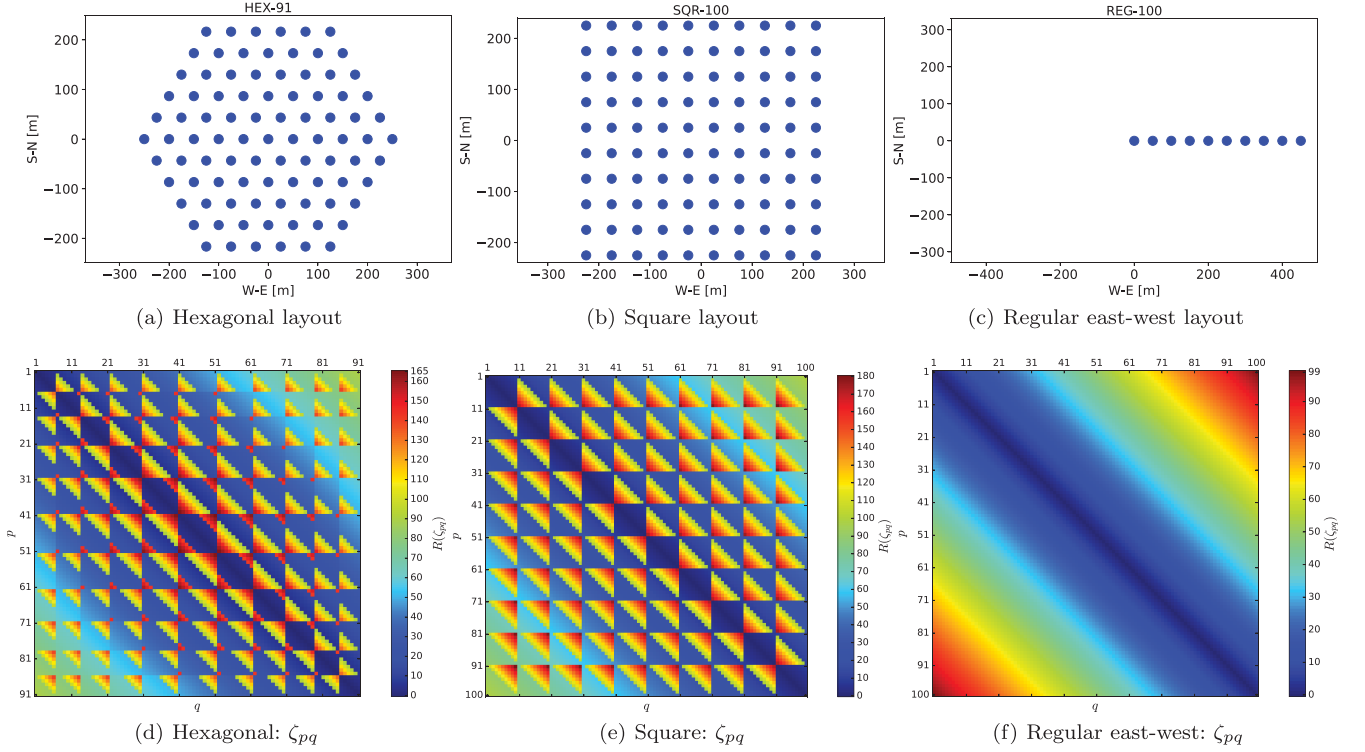


Figure 1. Three different redundant antenna layouts: hexagonal (top left), square (top middle), and regular east–west (top right) with their associated symmetric redundancy geometry functions ζ_{pq} (bottom panels). We used 91 antennas to construct the hexagonal layout, while 100 antennas were used in the square and east–west layouts. In the case of the east–west layout, we only plot the positions of the first 10 antennas. The maximal numbers of redundant baseline groups L that can be formed for the hexagonal, square, and east–west layouts are 165, 180, and 99, respectively. The analytic expressions of L for a hexagonal, square, and east–west layout are $L = 2N - \frac{1}{2}\sqrt{12N - 3} - \frac{1}{2}$, $L = 2N - 2\sqrt{N}$, and $L = N - 1$, respectively.

the gains can be obtained iteratively until convergence is reached. Alternating between solving and fixing different sets of parameters (which is exactly what STEFAL does) is referred to as the alternating direction implicit (ADI) method. The STEFAL algorithm reduces the computational complexity of calibration from $O(N^3)$ to $O(N^2)$.

3 REDUNDANT WIRTINGER CALIBRATION

Interferometric baselines are redundant when they sample the exact same visibilities in the uv -plane, i.e. if baseline pq and rs are redundant, then $y_{pq} = y_{rs}$. A redundant array layout allows us to solve for the unknown observed visibilities themselves in addition to the antenna gains (see equation 4). This is true, since in the case of a redundant layout, equation (4) is already an overdetermined system even before having predicted visibilities from a pre-existing skymodel.

It is convenient to group redundant visibilities together and label each group using a single index rather than using their antenna pairs as in equation (1). We introduce a function ϕ that maps the antenna pair associated with a specific baseline to its corresponding redundant baseline group, i.e. if baseline pq and rs are redundant, then $\phi_{pq} = \phi_{rs}$ (implying that they belong to the same group). To be exact, ϕ maps the composite index pq to its group index only if $pq \in \{rs|r \leq s\}$. If $pq \notin \{rs|r \leq s\}$, then the composite index pq is mapped to zero. The function ϕ_{pq} is, therefore, not symmetric. Equation (1) can be rewritten for a redundant array as

$$d_{pq} = g_p \bar{g}_q y_{\phi_{pq}} + n_{pq}, \quad (18)$$

with the same vector form as equation (3) if

$$\begin{aligned} [d]_{\alpha_{pq}} &= d_{pq}, \quad [v]_{\alpha_{pq}} = v_{pq} = g_p y_{\phi_{pq}} \bar{g}_q, \\ [n]_{\alpha_{pq}} &= n_{pq}, \end{aligned} \quad (19)$$

where the vectors in equation (19) are column vectors of size B (i.e. $p < q$).

We also introduce the following symmetric variant of ϕ_{pq} :

$$\zeta_{pq} = \begin{cases} \phi_{pq} & \text{if } p \leq q \\ \phi_{qp} & \text{if } p > q \end{cases}, \quad (20)$$

and we will refer to ζ_{pq} as the symmetric geometric function. It is possible to construct a simple analytic expression for ζ_{pq} for an east–west regular array, i.e. $\zeta_{pq} = |q - p|$. It becomes, however, increasingly difficult to construct analytic expressions of ζ_{pq} for more complicated array layouts. The empirically constructed symmetric geometry functions of three different redundant layouts are displayed in Fig. 1. We denote the range of ζ_{pq} with $\mathcal{R}(\zeta_{pq})$. The maximal element that ζ_{pq} can ascertain is denoted by L and can be interpreted as the greatest number of unique redundant baseline groups that can be formed for a given array layout.

We can now formulate redundant calibration as a least-squares problem:

$$\min_z \Lambda(z) = \min_z \|r\|_F^2 = \min_z \|d - v(z)\|_F^2, \quad (21)$$

where

$$\begin{aligned} \mathbf{g} &= [g_1, \dots, g_N]^T, \quad \mathbf{y} = [y_1, \dots, y_L]^T, \\ \mathbf{z} &= [\mathbf{g}^T, \mathbf{y}^T]^T. \end{aligned} \quad (22)$$

The number of model parameters to be solved for is now $P = 2(N + L)$, since redundant calibration is a complex problem. Note that equation (21) is only solvable (i.e. the array is redundant enough) if $L + N \leq B$.

In the literature, equation (21) is solved by splitting the problem into its real and imaginary parts. The real and imaginary parts of the unknown parameters are then solved for separately (Wieringa 1992; Liu et al. 2010; Zheng et al. 2014). Currently, the above is achieved by using the real-valued GN algorithm (Kurien et al. 2016). We, instead, intend to formulate the redundant calibration problem using Wirtinger calculus and recast equation (21) as

$$\min_{\mathbf{z}} \|\check{\mathbf{r}}\| = \min_{\mathbf{z}} \|\check{\mathbf{d}} - \check{\mathbf{v}}(\mathbf{z})\|, \quad (23)$$

where $\mathbf{z} = [\mathbf{z}^T, \bar{\mathbf{z}}^T]^T$.

We derive the complex Jacobian associated with equation (23) to be

$$\mathbf{J} = \begin{bmatrix} \mathcal{J}_1 & \mathcal{J}_2 \\ \bar{\mathcal{J}}_2 & \bar{\mathcal{J}}_1 \end{bmatrix}, \quad (24)$$

where

$$[\mathcal{J}_1]_{\alpha pq, i} = \begin{cases} \frac{\partial v_{pq}}{\partial g_i} & \text{if } i \leq N \\ \frac{\partial v_{pq}}{\partial \bar{y}_{i-N}} & \text{otherwise} \end{cases}, \quad (25)$$

and

$$[\mathcal{J}_2]_{\alpha pq, i} = \begin{cases} \frac{\partial v_{pq}}{\partial \bar{g}_i} & \text{if } i \leq N \\ \frac{\partial v_{pq}}{\partial \bar{y}_{i-N}} & \text{otherwise} \end{cases}. \quad (26)$$

Algorithm 1 Constructing $\mathcal{J}_1$

```

1:  $p \leftarrow 1$ 
2: while  $p \leq N$  do
3:    $q \leftarrow p + 1$ 
4:   while  $q \leq N$  do
5:      $\alpha \leftarrow (q - p) + (p - 1) \left(N - \frac{1}{2}\right)$ 
6:      $i \leftarrow 1$ 
7:     while  $i \leq (N + L)$  do
8:       if  $i \leq N$  then
9:          $\mathcal{J}_1[\alpha, i] \leftarrow \frac{\partial v_{pq}}{\partial g_i}$ 
10:      else
11:         $\mathcal{J}_1[\alpha, i] \leftarrow \frac{\partial v_{pq}}{\partial \bar{y}_{i-N}}$ 
12:       $i \leftarrow i + 1$ 
13:     $q \leftarrow q + 1$ 
14:   $p \leftarrow p + 1$ 

```

Note that we employ the same subscript indexing notation that we used in Section 2 in equations (25) and (26). To further aid the reader in understanding this indexing notation, we refer him/her to Fig. 2 (also see Algorithm 1), which depicts a flow diagram of the matrix construction procedure with which $\mathcal{J}_1$ can be constructed. The range of values the indices in equations (25) and (26) can attain should also be clear to the reader after having inspected Fig. 2 and Table 2.

We can now calculate the GN and LM updates to be

$$\Delta \mathbf{z} = (\mathbf{J}^H \mathbf{J})^{-1} \mathbf{J}^H \check{\mathbf{r}} \quad (27)$$

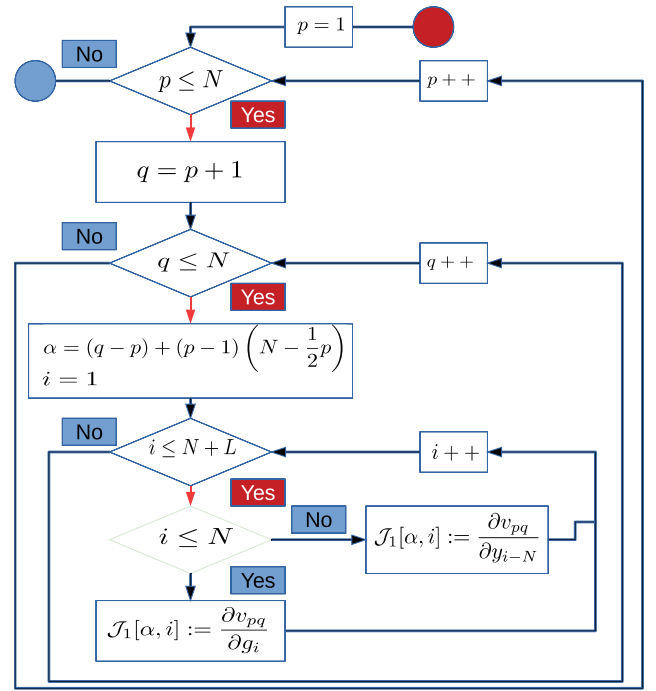


Figure 2. A flow chart representing the procedure one would follow to partially construct the Jacobian matrix (i.e. $\mathcal{J}_1$) associated with Wirtinger redundant calibration. The red circle represents the start of the flow diagram. The blue circle represents the end of the diagram. Blue diamonds denote loop conditionals, while green diamonds denote simple conditionals. The diagram elements following the red arrows just below a loop conditional element all form part of the main body of the loop that has the conditional statement of the aforementioned loop conditional element in its definition. The pseudo-code associated with this flow diagram is given in Algorithm 1.

and

$$\Delta \mathbf{z} = (\mathbf{J}^H \mathbf{J} + \lambda \mathbf{D})^{-1} \mathbf{J}^H \check{\mathbf{r}}, \quad (28)$$

respectively. As in Section 2, equation (27) can be used to iteratively update our parameter vector:

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \Delta \mathbf{z}_k. \quad (29)$$

The analytic expressions for $\mathbf{J}$, $\mathbf{H}$, and $\mathbf{J}^H \check{\mathbf{r}}$ can be found in Appendix A. Appendix A also contains two useful identities involving $\mathbf{J}$ and $\mathbf{H}$.

In the case of the GN algorithm, we can simplify equation (29) even further. Replacing $\check{\mathbf{r}}$ with $\check{\mathbf{d}} - \check{\mathbf{v}}$ in equation (27) results in

$$\Delta \mathbf{z} = (\mathbf{J}^H \mathbf{J})^{-1} \mathbf{J}^H (\check{\mathbf{d}} - \check{\mathbf{v}}). \quad (30)$$

If we substitute the first identity of equation (A20) into equation (30) and we simplify the result, we obtain

$$\Delta \mathbf{z} = (\mathbf{J}^H \mathbf{J})^{-1} \mathbf{J}^H \check{\mathbf{d}} - \frac{1}{3} \check{\mathbf{z}}. \quad (31)$$

If we use the above-simplified update in equation (29), it reduces to

$$\mathbf{z}_{k+1} = (\mathbf{J}^H \mathbf{J})^{-1} \mathbf{J}^H \check{\mathbf{d}} + \frac{2}{3} \check{\mathbf{z}}_k. \quad (32)$$

Equation (32) is the redundant equivalent of equation (16), and it shows us that in the case of redundant calibration, we can calculate the GN parameter update step without calculating the residual.

Fig. 3 shows that the Hessian matrix $\mathbf{H}$ is nearly diagonal and sparse for both the regular east-west and hexagonal layouts we

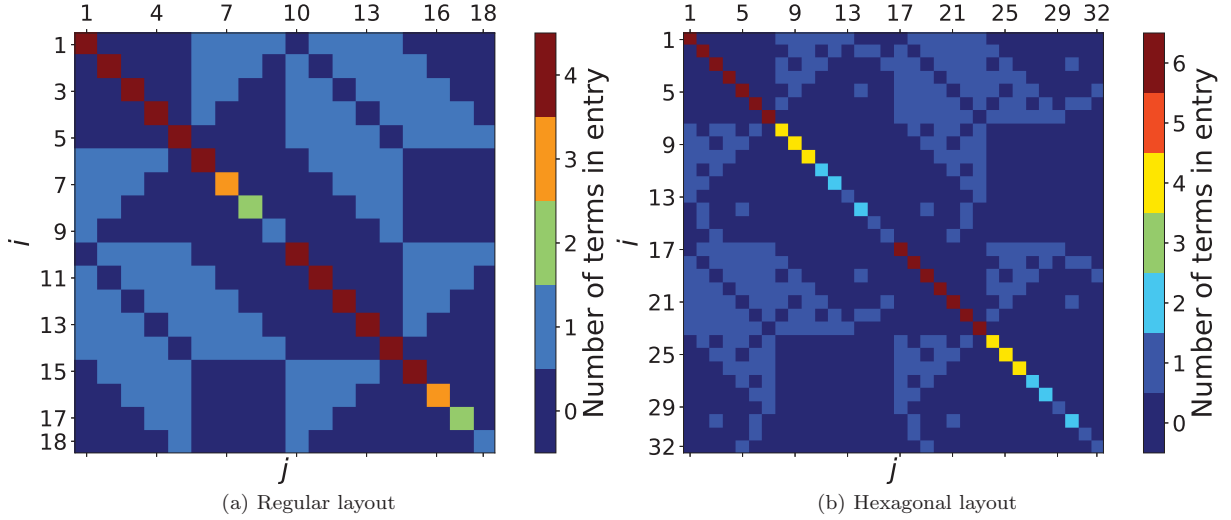


Figure 3. The number of analytic terms out of which the entries of the Hessian $\mathbf{H}$ consist for two different geometric layouts, namely a regular east–west grid with $N = 5$ (left-hand panel) and a hexagonal grid with $N = 7$. The diagonal entries of these two Hessians are clearly more significant than their off-diagonal entries. Moreover, these two Hessians also contain many zero entries. Note that the locations of the zero entries are dependent on the geometry of the array layout.

considered. We therefore follow the approach of Smirnov & Tasse (2015) and approximate the Hessian matrix $\mathbf{H}$ with its diagonal. If we substitute $\mathbf{J}^H \mathbf{J}$ with $\tilde{\mathbf{H}} = \mathbf{H} \odot \mathbf{I}$ and replace $\tilde{\mathbf{r}}$ with $\tilde{\mathbf{d}} - \tilde{\mathbf{v}}$ in equation (27), we obtain

$$\Delta \tilde{\mathbf{z}} \approx \tilde{\mathbf{H}}^{-1} \mathbf{J}^H (\tilde{\mathbf{d}} - \tilde{\mathbf{v}}). \quad (33)$$

Utilizing the second identity in equation (A20) allows us to simplify equation (33) to

$$\Delta \tilde{\mathbf{z}} \approx \tilde{\mathbf{H}}^{-1} \mathbf{J}^H \tilde{\mathbf{d}} - \tilde{\mathbf{z}}, \quad (34)$$

which leads to

$$\tilde{\mathbf{z}}_{k+1} \approx \tilde{\mathbf{H}}^{-1} \mathbf{J}^H \tilde{\mathbf{d}}. \quad (35)$$

Using equation (28), we follow the same procedure and obtain a similar result for LM

$$\tilde{\mathbf{z}}_{k+1} \approx \frac{1}{1 + \lambda} \tilde{\mathbf{H}}^{-1} \mathbf{J}^H \tilde{\mathbf{d}} + \frac{\lambda}{1 + \lambda} \tilde{\mathbf{z}}_k, \quad (36)$$

$$= \rho \tilde{\mathbf{H}}^{-1} \mathbf{J}^H \tilde{\mathbf{d}} + (1 - \rho) \tilde{\mathbf{z}}_k. \quad (37)$$

The analytic expression of $\mathbf{J}^H \tilde{\mathbf{d}}$ will be very similar to the analytic expression of $\mathbf{J}^H \tilde{\mathbf{r}}$, the only difference being that in equation (A18) the letter r would be replaced by a d . If we substitute the analytic expression of $\mathbf{J}^H \tilde{\mathbf{d}}$ and $\tilde{\mathbf{H}}^{-1}$ (which can easily be constructed using Appendix A) into equation (37), we obtain the following two update rules:

$$g_i^{k+1} = \rho \frac{\sum_{j \neq i} g_j^k \tilde{y}_{ij}^* d_{ij}}{\sum_{j \neq i} |g_j^k|^2 |\tilde{y}_{ij}^k|^2} + (1 - \rho) g_i^k \quad (38)$$

and

$$y_i^{k+1} = \rho \frac{\sum_{rs \in \mathcal{RS}_i} \tilde{g}_r^k \tilde{g}_s^k d_{rs}}{\sum_{rs \in \mathcal{RS}_i} |\tilde{g}_r^k|^2 |\tilde{g}_s^k|^2} + (1 - \rho) y_i^k. \quad (39)$$

The index set $\mathcal{RS}_i$ and the quantity $\tilde{y}_{ij}$ are defined in equations (A14) and (A19), respectively. The computational complexity of inverting $\tilde{\mathbf{H}}$ is $O(P)$. We note that equation (38) is the gain estimator associated with STEFCAL.

Equations (38) and (39) were obtained by Marthi & Chengalur (2014) by taking the derivative of the objective function Λ relative to the elements of $\mathbf{g}$ and $\mathbf{y}$, setting the intermediate results to zero and then solving for the unknown parameters (i.e. using the gradient descent algorithm). We note that their derivation is less general. The LM algorithm has better convergence properties than gradient descent and encompasses the gradient descent algorithm as a special case. In Appendix B, we show that equations (38) and (39) can also be derived using the ADI method. For this reason, we refer to the implementation of the pseudo-LM calibration scheme derived above, i.e. equations (38) and (39), as redundant STEFCAL throughout the rest of the paper. Interestingly, Marthi & Chengalur (2014) were not the first to make use of the ADI method to perform redundant calibration; a slower alternative ADI-based calibration algorithm is presented in Wijnholds & Noorishad (2012).

The choice of the ρ parameter is somewhat unconstrained. In this paper, we chose ρ by adopting the same strategy that is used by STEFCAL and Marthi & Chengalur (2014), i.e. we chose ρ to be equal to $\frac{1}{3}$ ($\lambda = 2$). We also carried out simulations to validate this choice.

We generated a skymodel that comprised of 100 flat spectrum sources distributed over a 3° by 3° sky patch. The flux density of each source was drawn from a power-law distribution with a slope of 2 and the source position was drawn from a uniform distribution. We also made use of multiple fictitious telescope layouts each one having a hexagonal geometry (see the upper-left image of Fig. 1 for an example layout). The largest hexagonal array that we used has 217 antennas, with a minimum and maximum baseline of 20 and 320 m, respectively.

We corrupted visibilities by applying gain errors (see Tables 3 and 4) and calibrated the corrupted visibilities using redundant STEFCAL. Solutions were independently derived for each time-step and channel for five realizations.

Note that our simulations are almost ideal as we did not include a primary beam response, nor did we incorporate time and frequency smearing into our simulation.

We also did not explicitly define our noise in terms of the integration time, channel bandwidth, and T_{sys} ; we instead follow the same approach described in Liu et al. (2010) and Marthi & Chengalur

Table 3. The gain error models used in this paper. We have used the symbol x here as a proxy as it can either refer to time slots or channels. We either performed our simulations over multiple time slots and one frequency channel or one time slot and multiple frequency channels (see Table 4). Moreover, c in the leftmost column denotes the speed of light. We adopted a sinusoidal error model, similar to the sinusoidal error models used within MEQTREES (Noordam & Smirnov 2010), as well as a phase slope across frequency that mimics a real case of physical delays between different antennas. In the first model, we model a gain error with amplitude around one and an additional phase error. In the second model, we chose A in such a way that we do not unintentionally produce nonsensical gain values, i.e. a zero gain value. For the third model, the value of τ was chosen so that the number of phase wraps that occur across the observing band is restricted to a realistic number.

Number tag	1	2	3
Model	Sinusoidal: amplitude and phase	Sinusoidal: real and imaginary parts	Linear phase slope
Function	$(A + 1)e^{iP}$	$A \cos(2\pi f x + B) + 1.5A + jC \sin(2\pi f x + D)$	e^{iP}
Parameters	$A = a \cos(2\pi f x + b)$ $P = c \cos(2\pi f x + d)$ $f = 5$ $a \sim U[0.8, 0.9]$ $c \sim U[0.5, 5]$ $b, d \sim U[0, 2\pi]$	$f = 5$ $A, C \sim U[0.5, 10]$ $B, D \sim U[0, 2\pi]$	$P = \tau x$ $\tau = \frac{l}{c}$ $l \sim U[5, 50] \text{ (m)}$

Table 4. We generated results using two main setups. We either used one frequency channel and multiple time slots or one time slot and multiple frequency channels. The most important parameters used in realizing these two major setups are presented here. We chose the observational frequency band of setup 1 to coincide with the HERA array. To broaden the scope of our analysis, we chose the observational frequency of our second setup to be equal to 1.4 GHz, which is a typical observing frequency of the Westerbork Synthesis Radio Telescope.

	Setup 1	Setup 2
Num. channels	1024	1
ν -range	100–200 MHz	1.4 GHz
Num. time slots	1	50

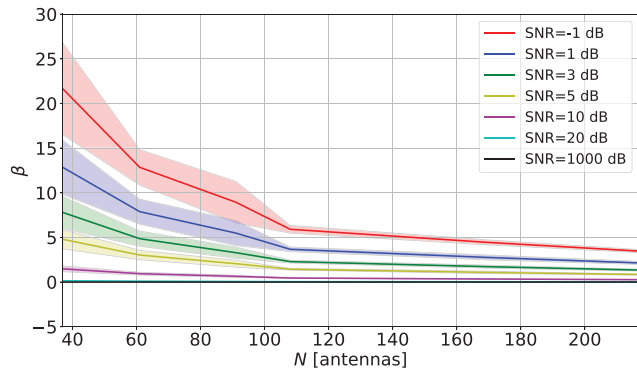


Figure 4. We plot the percentage error β between the simulated visibilities and the visibilities solved for by redundant STEFCAL for different SNR values as a function of the number of antennas (N) in the array.

(2014) and make use of the definition of SNR to introduce noise into our visibilities. We use the following definition of SNR (Liu et al. 2010; Marthi & Chengalur 2014):

$$\text{SNR} = 10 \log \left(\frac{\langle \mathbf{v} \odot \bar{\mathbf{v}} \rangle_{v,t,pq}}{\langle \mathbf{n} \odot \bar{\mathbf{n}} \rangle_{v,t,pq}} \right), \quad (40)$$

where $\langle \mathbf{x} \rangle_{v,t,pq}$ denotes averaging over frequency, time, and baseline. It should be pointed out, however, that by producing visibilities with different SNR values, we are effectively either changing our integration time or our channel bandwidth or both (assuming a single T_{sys} value for our instrument).

Fig. 4 shows the simulation results as a function of SNR and number of antennas. The accuracy of our solutions is quantified

through the percentage error

$$\beta = \frac{\|\mathbf{v} - \hat{\mathbf{v}}\|_F^2}{\|\mathbf{v}\|_F^2}, \quad (41)$$

where $\hat{\mathbf{v}}$ is the redundant STEFCAL parameter estimate.

The error magnitude follows the expected behaviour, i.e. it decreases as a function of SNR and number of antennas N . Interestingly, it reduces to a few per cent when $N > 120$ for essentially any choice of SNR.

4 PCG METHOD

Liu et al. (2010) suggested that the execution speed of redundant calibration could be reduced using the conjugate gradient (CG) method (Hestenes & Stiefel 1952), which would be computationally advantageous, since the Hessian matrix associated with redundant calibration (see Fig. 3) is sparse (Reid 1971). In this section, we study the computational complexity of the CG method when it is used to invert the modified Hessian matrix (see equation 28), in particular when preconditioning is used (i.e. the PCG method). Interestingly, empirical tests suggest that the unmodified Hessian itself is singular. It is therefore important to mention that the CG method can pseudo-invert the unmodified Hessian as well, i.e. the CG method can be directly applied to equation (27), because the Hessian is a positive semi-definite Hermitian matrix and the vector $\mathbf{J}^H \tilde{\mathbf{r}}$ is an element of its column range (Lu & Chen 2018).

The computational complexity of the CG method is

$$O(\sqrt{\kappa m}), \quad (42)$$

where m denotes the number of non-zero entries and κ denotes the spectral condition number of the matrix that is to be inverted. The spectral condition number κ of the matrix $\mathbf{A}$ is defined as

$$\kappa(\mathbf{A}) = \frac{\ell_{\max}}{\ell_{\min}}, \quad (43)$$

where $\ell_{\max}$ and $\ell_{\min}$ denote the largest and the smallest eigenvalue of $\mathbf{A}$, respectively.

Preconditioning is a technique used to improve the spectral condition number of a matrix. Let us consider a generic system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad (44)$$

and a positive-definite Hermitian matrix $\mathbf{M}$ so that

$$\mathbf{M}^{-1}\mathbf{A}\mathbf{x} = \mathbf{M}^{-1}\mathbf{b}. \quad (45)$$

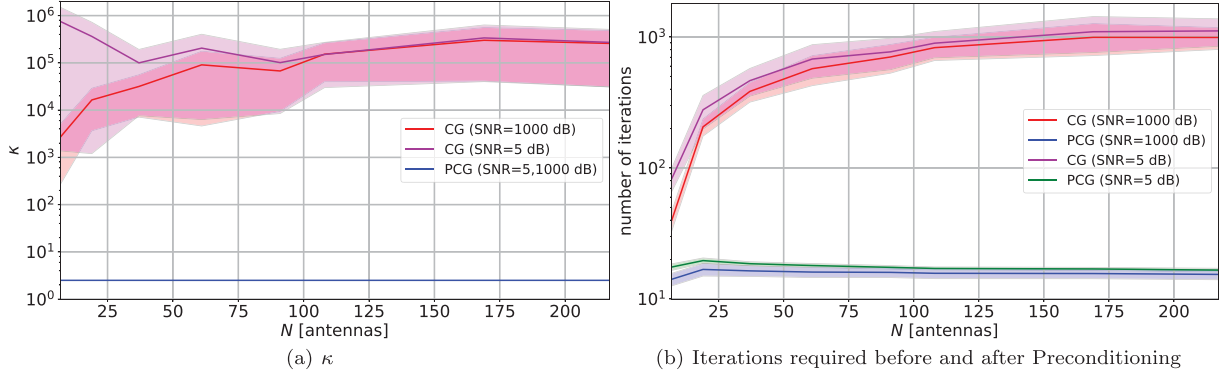


Figure 5. The graphs presented here were generated using the simulations discussed in Section 3. Left: spectral condition number κ of the modified Hessian $\mathcal{H}$ as a function of N , before (magenta and red curves) and after (blue curve) preconditioning, at different SNR values. Right: number of major iterations required by the CG method to invert $\mathcal{H}$ as a function of the number of antennas N in the array, before (magenta and red curves) and after preconditioning (blue and green curves), at different SNR values. Both plots show that the Jacobian preconditioner definitely speeds up the CG method.

The matrix $\mathbf{M}$ is a good preconditioner if

$$\kappa(\mathbf{M}^{-1}\mathbf{A}) \ll \kappa(\mathbf{A}), \quad (46)$$

i.e. if it lowers the condition number of a matrix. If $\mathbf{A}$ is a nearly diagonal matrix, the Jacobian preconditioner is a natural choice of $\mathbf{M}$ and can be computed as

$$\mathbf{M} = \mathbf{A} \odot \mathbf{I}. \quad (47)$$

In order to quantify the effectiveness of the CG method in redundant calibration, we investigate the spectral condition number and the sparsity of the modified Hessian $\mathcal{H}$ (i.e. $\lambda = 2$).

We generated simulated (i.e. corrupted) visibilities according to models described in Table 3. We used the complex LM algorithm described in Section 3 to calibrate the corrupted visibilities. To invert the modified Hessian, we used the CG method, with and without a Jacobian preconditioner. Fig. 5(a) shows us that preconditioning reduces the condition number of $\mathcal{H}$ to a small constant value and therefore effectively eliminates it from equation (42), i.e. equation (42) reduces to

$$O(m). \quad (48)$$

This result is confirmed by Fig. 5(b). Fig. 5(b) shows us that the number of major iterations needed by the PCG method to invert $\mathcal{H}$ is independent of the number of antennas in the array and that it is much less than the dimension of $\mathcal{H}$.

Let us now shift our attention towards the remaining factor m . To get a useful idea of the computational complexity of CG, we must relate m and P . This can be achieved by utilizing the modified Hessian's measure of sparsity γ :

$$\gamma = \left(1 - \frac{m}{P^2}\right), \quad (49)$$

where m is the number of non-zero entries in $\mathcal{H}$ and P^2 is total number of matrix elements.

For a regular east–west geometric configuration, the sparsity and the asymptotic sparsity of $\mathcal{H}$ can be derived analytically:

$$\gamma = \frac{5N^2 - 7N + 3}{8N^2 - 8N + 2} \quad \gamma_\infty = \lim_{N \rightarrow \infty} \gamma = \frac{5}{8}. \quad (50)$$

For more complicated array layouts, however, there is no straightforward analytical solution and we empirically determined the sparsity ratios for three different geometric layouts as a function of the number antennas in the array (see Fig. 6a).

It now follows that

$$P^c = m = (1 - \gamma)P^2, \quad (51)$$

which leads to

$$c = \log_p(1 - \gamma) + 2 \quad c_\infty = \lim_{N \rightarrow \infty} c = 2. \quad (52)$$

The computational complexity is therefore asymptotically bounded by $O(P^2)$, although it converges very slowly to its asymptotic value, and in general is equal to $O(P^c)$ (with $c < 2$). In the case of a hexagonal geometric layout with $N < 200$, we have that $c \sim 1.7$ (see Fig. 6b).

We are now finally able to compare the computational complexity of redundant STEFCAL and the PCG method. Redundant STEFCAL is computationally inexpensive as it just needs to invert a diagonal matrix; however, the PCG inversion is accurate and, therefore, may require fewer iterations, ultimately leading to a faster convergence. We computed the theoretical number of redundant STEFCAL iterations Δk in excess of the number of LM (implementing PCG) iterations required for LM to outperform redundant STEFCAL

$$(k + \Delta k)P > kP^c, \quad \Delta k > k(P^{c-1} - 1). \quad (53)$$

We then compared Δk with the empirically obtained average excess number of iterations. The results are displayed in Fig. 7, which shows that redundant STEFCAL outperforms the PCG method. We note that, in this comparison, we have only taken into account the cost of inverting $\mathcal{H}$ and ignored the cost of preconditioning.

5 CONCLUSION

In this paper, we have formulated the calibration of redundant interferometric arrays using the complex optimization formalism (Smirnov & Tasse 2015). We derived the associated complex Jacobian and the GN and LM parameter update steps. We also showed that the LM parameter update step can be simplified to obtain an ADI type of algorithm (Marthi & Chengalur 2014; Salvini & Wijnholds 2014). Our code implementation of this algorithm (redundant STEFCAL) is publicly available at <https://github.com/Trienko/heracommissioning/blob/master/code/stef.py>. We note that, in its current implementation, redundant STEFCAL does not solve for the degeneracies inherent to redundant calibration (Zheng et al. 2014; Kurien et al. 2016; Dillon et al. 2017), which will be the subject of future work. Compared to current redundant calibration

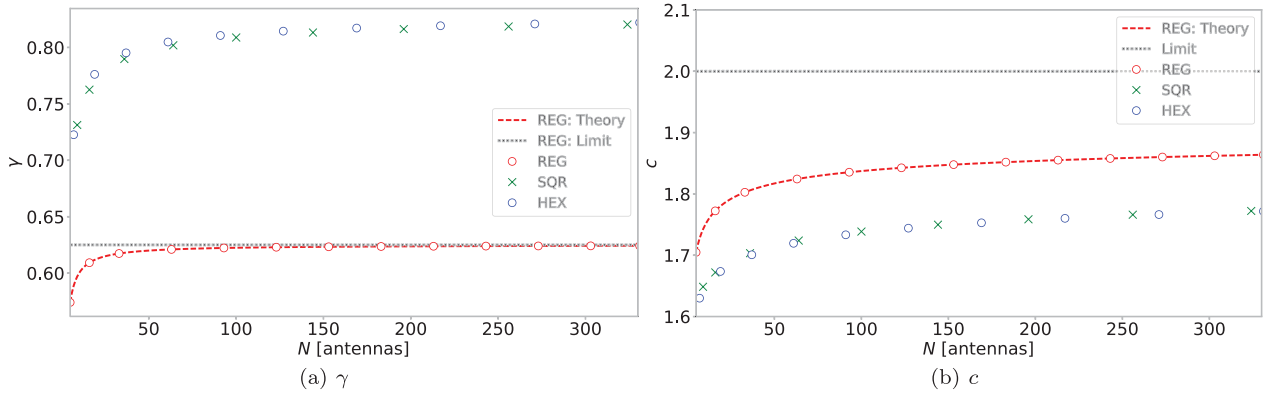


Figure 6. Left: the sparsity ratio γ of the modified Hessian $\mathcal{H}$ as a function of the number of antennas N for a hexagonal (blue circles), square (green crosses), and regular east–west (red circles) array geometry. The red dashed and black dotted lines show the analytical expression for γ and its limit for the regular east–west grid case (see the text for details). Right: the order of the computational cost c for inverting $\mathcal{H}$ as a function of N for different array geometries (same colour scheme as in the left-hand panel). The red dashed and black dotted lines are the analytical expression of c and its limit in the east–west regular grid case.

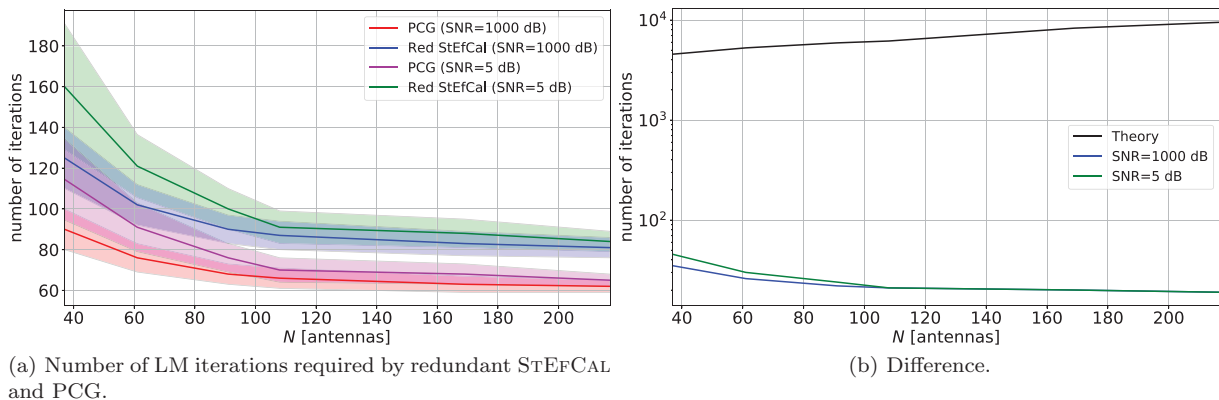


Figure 7. The graphs presented here were generated using the simulations discussed in Section 3. Left: number of LM iterations required by redundant STEFCAL (green and blue curves) and the PCG (magenta and red curves) methods to converge to a parameter error tolerance of 10^{-6} whilst using different SNR values as a function of the number of antennas N in the array. Right: average number of LM iterations (difference between the redundant STEFCAL and PCG curves in the left-hand panel) saved (green and blue curve) by computing the full inverse with the PCG method whilst using different SNR values. The black curve is Δk , the theoretical number of redundant STEFCAL iterations that are needed in excess of the number of LM iterations used for the LM algorithm to outperform redundant STEFCAL. For the black curve, we assumed $c = 1.7$ and that k could be approximated with the magenta curve plotted in the left-hand panel.

algorithms, redundant STEFCAL is more robust to initial conditions and allows for easier parallelization.

We investigated the computational cost of redundant STEFCAL and compared it with the performance of the PCG method (as suggested by Liu et al. 2010). We found that, although the PCG method greatly improves the speed of redundant calibration, it still significantly underperforms when compared to redundant STEFCAL.

The characteristics of redundant STEFCAL make it an appealing calibration algorithm for large redundant arrays like HERA (DeBoer et al. 2017), CHIME (Bandura et al. 2014), HIRAX (Newburgh et al. 2016), or even hybrid arrays like the MWA (Tingay et al. 2013) in its updated phase.

6 FUTURE OUTLOOK

We plan to apply redundant STEFCAL to HERA data in the near future. The reason for applying redundant STEFCAL to real data is twofold. First, we wish to validate the theoretical computational complexity of redundant STEFCAL (derived earlier in the paper). Secondly, we would like to test whether it could be used to calibrate HERA in

near-realtime. We are also interested in parallelizing our current implementation and to see how much can be gained by doing so. We would also like to conduct a perturbation analysis, similar to the one conducted by Liu et al. (2010), to estimate the error that is introduced into our estimated gains and visibilities when the array contains baselines that are not perfectly redundant. We are also interested in quantifying the error that is introduced into the estimated gains and visibilities because of differing primary beam patterns between array elements (Noorishad et al. 2012).

ACKNOWLEDGEMENTS

This work is based upon research supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation. GB acknowledges support from the Royal Society and the Newton Fund under grant NA150184. This work is also based on the research supported in part by the National Research Foundation of South Africa (grant No. 103424).

We would like to thank the anonymous reviewer for their useful comments, which greatly improved the quality of the paper.

REFERENCES

- Ali Z. S. et al., 2015, *ApJ*, 809, 61
 Bandura K. et al., 2014, *Proc. SPIE*, 9145, 914522
 Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, *MNRAS*, 461, 3135
 Bernardi G. et al., 2009, *A&A*, 500, 965
 Bernardi G. et al., 2010, *A&A*, 522, A67
 DeBoer D. R. et al., 2017, *PASP*, 129, 045001
 Dillon J. S., Parsons A. R., 2016, *ApJ*, 826, 181
 Dillon J. S. et al., 2014, *Phys. Rev. D*, 89, 023002
 Dillon J. S. et al., 2017, preprint ([arXiv:1712.07212](https://arxiv.org/abs/1712.07212))
 Ewall-Wice A., Dillon J. S., Liu A., Hewitt J., 2017, *MNRAS*, 470, 1849
 Furlanetto S. R., 2016, in Mesinger A., ed., *Astrophysics and Space Science Library*, Vol. 423, Understanding the Epoch of Cosmic Reionization: Challenges and Progress. Springer-Verlag, Berlin, p. 247
 Ghosh A., Prasad J., Bharadwaj S., Ali S. S., Chengalur J. N., 2012, *MNRAS*, 426, 3295
 Grobler T. L., Nunhokee C. D., Smirnov O. M., van Zyl A. J., de Bruyn A. G., 2014, *MNRAS*, 439, 4030
 Hamaker J. P., Bregman J. D., Sault R. J., 1996, *A&AS*, 117, 137
 Hestenes M. R., Stiefel E., 1952, *J. Res. Natl. Bur. Stand.*, 49, 409
 Kurien B. G., Tarokh V., Rachlin Y., Shah V. N., Ashcom J. B., 2016, *MNRAS*, 461, 3585
 Liu S., Trenkler G., 2008, *Int. J. Inf. Syst. Sci*, 4, 160
 Liu A., Tegmark M., Morrison S., Lutemirski A., Zaldarriaga M., 2010, *MNRAS*, 408, 1029
 Lu Z., Chen X., 2018, *Math. Oper. Res.*, 43, 275
 McQuinn M., 2016, *ARA&A*, 54, 313
 Madsen K., Nielsen H. B., Tingleff O., 2004, *Methods for Non-Linear Least Squares Problems*, 2nd edn. Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Lyngby
 Marthi V. R., Chengalur J., 2014, *MNRAS*, 437, 524
 Mitchell D. A., Greenhill L. J., Wayth R. B., Sault R. J., Lonsdale C. J., Cappallo R. J., Morales M. F., Ord S. M., 2008, *IEEE J. Sel. Top. Signal Process.*, 2, 707
 Newburgh L. B. et al., 2016, *Proc. SPIE*, 9906, 99065X
 Noordam J., De Bruyn A., 1982, *Nature*, 299, 597
 Noordam J. E., Smirnov O. M., 2010, *A&A*, 524, A61
 Noorishad P., Wijnholds S. J., van Ardenne A., van der Hulst J., 2012, *A&A*, 545, A108
 Parsons A., Pober J., McQuinn M., Jacobs D., Aguirre J., 2012, *ApJ*, 753, 81
 Parsons A. R. et al., 2014, *ApJ*, 788, 106
 Pearson T., Readhead A., 1984, *ARA&A*, 22, 97
 Reid J. K. ed., 1971, *Proc. Conf. on Large Sparse Sets of Linear Equations*, On the Method of Conjugate Gradients for the Solution of Large Sparse Systems of Linear Equations. Academic Press, New York, p. 231
 Salvini S., Wijnholds S. J., 2014, *A&A*, 571, A97
 Sault R. J., Hamaker J. P., Bregman J. D., 1996, *A&AS*, 117, 149
 Sievers J., 2017, preprint ([arXiv:1701.01860](https://arxiv.org/abs/1701.01860))
 Smirnov O. M., 2011a, *A&A*, 527, A106
 Smirnov O. M., 2011b, *A&A*, 527, A108
 Smirnov O., Tasse C., 2015, *MNRAS*, 449, 2668
 Tingay S. J. et al., 2013, *PASA*, 30, e007
 Wieringa M. H., 1992, *Exp. Astron.*, 2, 203
 Wijnholds S. J., Noorishad P., 2012, in Bartleson K., ed., *Signal Processing Conference (EUSIPCO)*, Statistically Optimal Self-Calibration of Regular Imaging Arrays. IEEE, Piscataway, NJ, p. 1304
 Wirtinger W., 1927, *Math. Ann.*, 97, 357
 Zheng H. et al., 2014, *MNRAS*, 445, 1084

APPENDIX A: DERIVATION OF REDUNDANT WIRTINGER CALIBRATION

If we apply the definition in equation (24) to equation (23), we obtain the following analytic result:

$$\mathbf{J} = \begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \mathbf{N}^H & \mathbf{M} \end{bmatrix}, \quad (\text{A1})$$

where

$$\mathbf{M} = \begin{bmatrix} \mathbf{O} & \mathbf{P} \end{bmatrix} \quad (\text{A2})$$

and

$$\mathbf{N} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \end{bmatrix}. \quad (\text{A3})$$

Moreover,

$$[\mathbf{O}]_{\alpha pq, j} = \begin{cases} y_{\phi pq} \bar{g}_q & \text{if } p = j \\ 0 & \text{otherwise} \end{cases}, \quad (\text{A4})$$

$$[\mathbf{P}]_{\alpha pq, j} = \begin{cases} g_p \bar{g}_q & \text{if } \phi_{pq} = j \\ 0 & \text{otherwise} \end{cases}, \quad (\text{A5})$$

and

$$[\mathbf{Q}]_{\alpha pq, j} = \begin{cases} g_p y_{\phi pq} & \text{if } q = j \\ 0 & \text{otherwise} \end{cases}. \quad (\text{A6})$$

We use $\mathbf{0}$ to denote an all-zero matrix. It is now trivial to compute the Hessian $\mathbf{H}$ by using equation (A1). If we substitute equation (A1) into $\mathbf{J}^H \mathbf{J}$, we obtain

$$\mathbf{H} = \mathbf{J}^H \mathbf{J} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{A} \end{bmatrix}, \quad (\text{A7})$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{D}^H & \mathbf{E} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{F} & \mathbf{G} \\ \mathbf{G}^H & \mathbf{0} \end{bmatrix}, \quad (\text{A8})$$

$$[\mathbf{C}]_{ij} = \begin{cases} \sum_{k \neq i} |g_k|^2 |y_{\xi_{ik}}|^2 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}, \quad (\text{A9})$$

$$[\mathbf{D}]_{ij} = \begin{cases} g_i \bar{y}_j |g_{\psi_{ij}}|^2 & \text{if } \psi_{ij} \neq 0 \\ 0 & \text{otherwise} \end{cases}, \quad (\text{A10})$$

$$[\mathbf{E}]_{ij} = \begin{cases} \sum_{rs \in \mathcal{RS}_i} |g_r|^2 |g_s|^2 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}, \quad (\text{A11})$$

$$[\mathbf{F}]_{ij} = \begin{cases} g_i g_j |y_{\xi_{ij}}|^2 & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}, \quad (\text{A12})$$

and

$$[\mathbf{G}]_{ij} = \begin{cases} g_i y_j |g_{\xi_{ij}}|^2 & \text{if } \xi_{ij} \neq 0 \\ 0 & \text{otherwise} \end{cases}. \quad (\text{A13})$$

Moreover,

$$\mathcal{RS}_i = \{rs \in \mathbb{N}^2 | (\phi_{rs} = i)\}, \quad (\text{A14})$$

Table A1. The dimensions of the matrices defined in Appendix A.

Matrix	Dimension
J	$2B \times P$
M	$B \times (N + L)$
N	$B \times (N + L)$
O	$B \times N$
P	$B \times L$
Q	$B \times N$
0	$B \times L$
H	$P \times P$
A	$(N + L) \times (N + L)$
B	$(N + L) \times (N + L)$
C	$N \times N$
D	$N \times L$
E	$L \times L$
F	$N \times N$
G	$N \times L$
0	$L \times L$

$$\xi_{ij} = \begin{cases} p & \text{if } \exists! p \in \mathbb{N}s.t.(\phi_{pi} = j) \\ 0 & \text{otherwise} \end{cases}, \quad (\text{A15})$$

and

$$\psi_{ij} = \begin{cases} q & \text{if } \exists! q \in \mathbb{N}s.t.(\phi_{iq} = j) \\ 0 & \text{otherwise} \end{cases}. \quad (\text{A16})$$

Furthermore, substituting equation (A1) into $\mathbf{J}^H \tilde{\mathbf{r}}$ results in

$$\mathbf{J}^H \tilde{\mathbf{r}} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \bar{\mathbf{a}} \\ \bar{\mathbf{b}} \end{bmatrix}, \quad (\text{A17})$$

where

$$[\mathbf{a}]_i = \sum_{k \neq i} g_k \tilde{y}_{ik} r_{ik}, \quad [\mathbf{b}]_i = \sum_{rs \in \mathcal{RS}_i} \bar{g}_r g_s r_{rs}, \quad (\text{A18})$$

and

$$\tilde{y}_{ik} = \begin{cases} \bar{y}_{\zeta_{ik}} & \text{if } k > i \\ y_{\zeta_{ik}} & \text{otherwise} \end{cases}. \quad (\text{A19})$$

Additionally, $\mathbf{a}$ and $\mathbf{b}$ are both column vectors. The lengths of $\mathbf{a}$ and $\mathbf{b}$ are N and L , respectively. The dimensions of the matrices we defined in this section are presented in Table A1.

Furthermore,

$$\frac{1}{3} \mathbf{J} \tilde{\mathbf{z}} = \tilde{\mathbf{v}}, \quad \mathbf{J}^H \tilde{\mathbf{v}} = (\mathbf{I} \odot \mathbf{H}) \tilde{\mathbf{z}}. \quad (\text{A20})$$

The above identities can be trivially established by mechanically showing that the left-hand side of each expression in equation (A20) is equal to its right-hand side.

APPENDIX B: ADI

The basic skymodel-based STEPCAL update step is equal to the leftmost term in equation (38) (barring ρ ; Salvini & Wijnholds 2014). Assume without any loss of generality that the array is in an east-west regular grid. Furthermore, assume that $\mathbf{d}$ (see equation 3) has been reordered and that the result of this reordering is the following:

$$\tilde{\mathbf{d}} = [d_{12}, \dots, d_{N-1,N}, d_{13}, \dots, d_{N-2,N}, \dots, d_{1N}]^T. \quad (\text{B1})$$

The vector $\tilde{\mathbf{n}}$ should be interpreted in a similar manner.

Equation (3) can now be rewritten as

$$\tilde{\mathbf{d}} = \mathbf{J} \mathbf{y} + \tilde{\mathbf{n}}, \quad (\text{B2})$$

if we assume that $\mathbf{g}$ and its conjugate are known vectors. In equation (B2),

$$\mathbf{J} = \begin{bmatrix} g_1 \bar{g}_2 & 0 & \dots & 0 \\ g_2 \bar{g}_3 & 0 & \dots & 0 \\ \vdots & 0 & \dots & 0 \\ g_{N-1} \bar{g}_N & 0 & \dots & 0 \\ 0 & g_1 \bar{g}_3 & \dots & 0 \\ 0 & \vdots & \dots & 0 \\ 0 & g_{N-2} \bar{g}_N & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & g_1 \bar{g}_N \end{bmatrix}. \quad (\text{B3})$$

We can now estimate $\mathbf{y}$ with

$$\mathbf{y} = (\mathbf{J}^H \mathbf{J})^{-1} \mathbf{J}^H \tilde{\mathbf{d}}, \quad (\text{B4})$$

where

$$[\mathbf{J}^H \mathbf{J}]_{ij} = \begin{cases} \sum_{rs \in \mathcal{RS}_i} |g_r|^2 |g_s|^2 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}, \quad (\text{B5})$$

and

$$[\mathbf{J}^H \tilde{\mathbf{d}}]_i = \sum_{rs \in \mathcal{RS}_i} \bar{g}_r g_s d_{rs}. \quad (\text{B6})$$

Substituting equations (B5) and (B6) into equation (B4) and simplifying the result leads to the leftmost term in equation (39) (if we bar ρ and consider only the i th entry of $\mathbf{y}$).

This paper has been typeset from a $\text{\LaTeX}$ file prepared by the author.