Parallel Multi-Wavelength Calibration Algorithm for Radio Astronomical Arrays[☆]

Martin Brossard[a], Mohammed Nabil El Korso[b], Marius Pesavento[c], Rémy Boyer[d], Pascal Larzabal[e], Stefan J. Wijnholds[f]

[a]*MINES ParisTech, PSL Research University, Centre for Robotics, Paris, France*
[b]*University of Paris Ouest Nanterre La Défense, IUT de Ville d'Avray, LEME EA 4416, France*
[c]*Communication Systems Group, Technische Universität, Darmstadt, Germany*
[d]*University of Paris-Sud, Laboratoire des Signaux et Systèmes (L2S), Gif-Sur-Yvette, France*
[e]*SATIE, UMR 8029, École Normale Supérieure de Cachan, Cachan, France*
[f]*Netherlands Institute for Radio Astronomy (ASTRON), P.O. Box 2, NL-7990 AA, Dwingeloo, The Netherlands*

Abstract

In order to meet the theoretically achievable imaging performance, calibration of modern radio interferometers is a mandatory challenge, especially at low frequencies. In this perspective, we propose a novel parallel iterative multi-wavelength calibration algorithm. The proposed algorithm estimates the apparent directions of the calibration sources, the direction dependent and direction independent complex gains of the array elements and their noise powers. Furthermore, the algorithm takes into account the specific variation of the aforementioned parameter values across wavelength. Numerical simulations reveal that the proposed scheme outperforms the mono-wavelength calibration scheme and approaches the derived constrained Cramér-Rao bound even with the presence of non-calibration sources at unknown directions. Finally, simulation results on LOFAR real data assess the usefulness of the proposed scheme.

Keywords: Calibration, radio astronomy, radio interferometer, sensor array, direction-of-arrival estimation, consensus optimization

1. Introduction

Advanced radio astronomical arrays, such as the existing LOw Frequency ARray (LOFAR) [1] and the future Square Kilometre Array (SKA) [2], form large sensor arrays, which are constituted of many small antenna elements. As an example, the LOFAR consists of 50 stations, mainly located in The Netherlands. Each station is a closely packed sensor array, composed of at least 96 low-band antennas (LBA, 30-90 MHz) and 48 high-band antenna (HBA, 109-240 MHz) tiles, each tile being a 4-by-4 uniform regular array of HBAs. Such interferometers offer a large aperture size and deliver large amounts of data in order to reach high performance in terms of resolution, sensitivity and survey speed [2]. Nevertheless, to achieve the theoretical optimal performance bounds, a plethora of signal processing challenges must be treated [3, 4]. This covers calibration, image synthesis and data reduction. In this paper, we focus on calibration issues by designing a computationally efficient parallel algorithm. Calibration procedures devised for such radio interferometers must estimate: i) the gain response and noise power of each antenna [5–8]; and ii) the propagation disturbances, especially the phase delays caused by the ionosphere, which scale with wavelength [9, 10].

Specifically, in this paper, we focus on the regime where the lines of sight from each antenna toward a source in the sky cross the same ionospheric layer and where the thickness of the ionosphere can be direction dependent [11], which is represented in Figure 1 and well adapted for the calibration of a LOFAR station and the future SKA stations as well as the core of these arrays. Consequently, in this regime, the ionospheric phase delays are added to the geometric delays and introduce angular-shifts for the source directions [7, 12], which are direction and wavelength dependent [9, 13]. By estimating calibrator shifts (i.e., the difference between the true calibrator directions, known from tables [14–17], and their estimated apparent directions), interpolation methods can be efficiently applied in order to obtain a phase screen model that captures the ionospheric delays over the entire Field-of-View [12]. We emphasize that in addition to the phase screen reconstruction step, the calibration usually involves the estimation of the complex direction independent (DI) gains of the antennas, their direction dependent (DD) gains toward each calibrator and their noise powers [6], for the whole available range of wavelengths.

The characteristics of the calibration sources, i.e., their true/nominal directions and their powers without the effects of the ionosphere or antenna imperfections, are a priori known from tables which is required to solve such calibration problems [7]. Based on this knowledge, state-of-
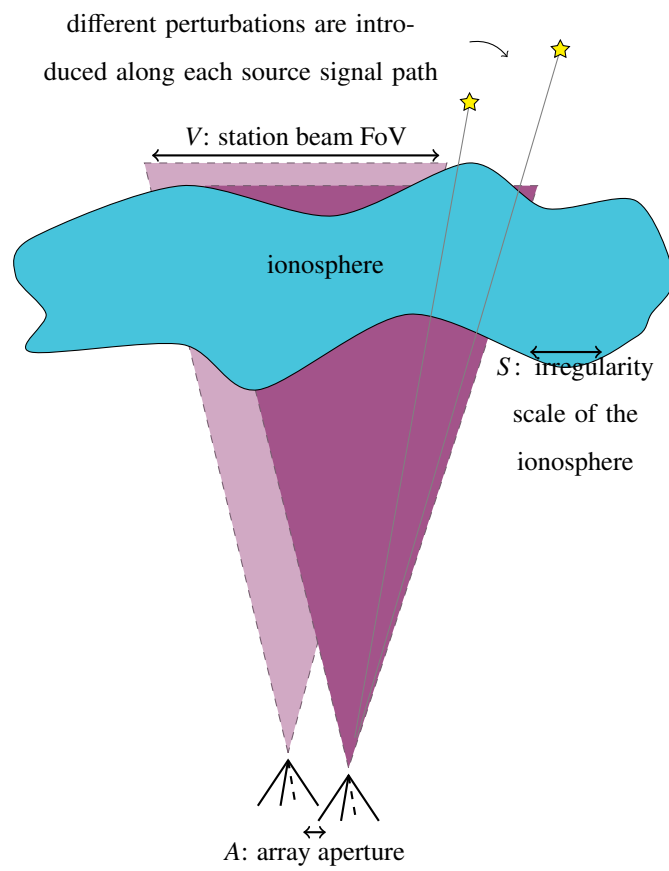
Figure 1: The so-called regime 3, which is considered in this paper, assumes that $V \gg S$ and $A \ll S$. This leads to ionospheric perturbations which are direction dependent (after [8, 11]).

the-art calibration algorithms operate mostly in an iterative manner in a mono-wavelength scenario [5–7, 18–21, 21, 22]. For instance, the (Weighted) Alternating Least Squares approach has been adapted for LOFAR station calibration [5, 6], in which closed-form expressions have been obtained for antenna gain and sensor noise power parameters. Nevertheless, such algorithms present three major limitations: i) suboptimality due to the consideration of only one wavelength at a time; ii) the assumption of a centralized processor, i.e., a single compute agent simultaneously accessing all data; and iii) the inefficiency with respect to the Direction-of-Arrival (DoA) estimation performances in the severe radio astronomical contexts.

Concerning limitation i), most current approaches calibrate a single wavelength at a time [5–7, 18–20] assuming that solutions can be combined afterwards. This is usually a suboptimal approach. For the real-time system of the Murchison Widefield Array (MWA) [23] several filtering techniques are combined to isolate the response across frequency such that ionospheric refraction and instrumental gains can be fitted across frequency on a per-source bases [24]. The MWA real-time system is highly optimized for a specific case, while we aim to provide a more general framework that can be adjusted as needed. A more general approach to multi-wavelength calibration in the context of large radio astronomical arrays is the procedure presented in [25]. That procedure treats calibration as a single, mathematically defined optimization problem that is solved using a general solving strategy. Unlike [25], we consider a structured parametric model based on the sample covariance matrix in the array processing framework based on some a priori knowledge about the physics of the system as described in the data model section. As a consequence, the problems are different, since we do not estimate the elements of the Jones matrices as in [25], but we aim to estimate the apparent directions of the calibration sources, the direction dependent and direction independent complex gains of the array elements and their noise powers using physical constraints given in Section 2.

Furthermore, regarding the limitation ii), the aforementioned state-of-the-art methods typically are designed for a centralized hardware architecture, whereas, taking the LOFAR stations as example, processing all 512 frequency bands simultaneously at a single location, if feasible, is challenging. As a solution, parallel and consensus algorithms, mostly based on the Alternating Direction of Multiple Multipliers (ADMM) [26], have recently been massively investigated in parametric estimation frameworks [27–34]. These consensus schemes can operate in various network topologies. We will consider a group of compute agents, where each agent accesses data

4

across a small bandwidth and can only communicates with a fusion center through low data rate channels. This architecture models correctly the situation for radio interferometers, where data for the full observing bandwidth is typically divided into channels and channels are grouped into subbands.

Finally, regarding the limitation iii), classical subspace methods, such as MUSIC [35], have been commonly applied in radio astronomical calibration [6]. However, these techniques are inefficient in low Signal-to-Noise-Ratio (SNR) scenarios and require knowledge of the exact number of sources in the scene. As an alternative, recent approaches, based on sparse reconstruction methods, came into focus of DoA estimation for fully calibrated arrays [36–38] as well as for partially calibrated arrays [39]. These approaches exhibit the super-resolution property, robustness and computational efficiency, without the aforementioned limitations of subspace-based methods [36]. However, most methods based on the compressive sensing framework are designed for a centralized hardware architecture and are applied in the signal time domain [20, 40]. These methods become computationally impractical with huge numbers of observations, making them unsuitable for calibration in the radio interferometer context for which we commonly access only the sample covariance matrix rather than the time signal itself [7]. This issue was also recognized recently in [22] and [41], in which the authors propose a technique for, respectively, full calibration and blind calibration of the DI gains for individual frequency channels by assuming that the observed scene is sparse. We stress that many works based of compressed sensing has been developed for image reconstruction in the radio astronomy community [42–45] as a result of the sparse nature of the interferometric sampling. Such sparse implementations produce better results on large extended objects with high angular resolution than other classical deconvolution methods. However, these works assume usually an already calibrated array. Note that the calibration and imaging steps are different. In image reconstruction, compressed sensing studies are based on sparse analysis and/or sparse synthesis using, e.g., wavelets, union of wavelet bases or synthesis IUWT (Isotropic Undecimated Wavelet Transform). In our calibration problem, compressed sensing techniques are used in order to estimate the phase shift due to the ionosphere using a Distributed Iterative Hard Thresholding method in which the grid is constructed based on array processing modeling.

In this paper, we propose an iterative algorithm, namely the Parallel Multi-wavelength Calibration Algorithm (PMCA), that performs the calibration of a radio astronomical array using

its array covariance matrix (usually referred to as a matrix of *visibilities* in radio astronomy), involving disturbances due to individual antennas and propagation effects. We assume that the sensor array has an arbitrary geometry, identical elements and is simultaneously excited by inaccurately known calibration sources and unknown weak non-calibration sources. The proposed PMCA overcomes the aforementioned limitations, by: i) reformulating the parametric model in the multi-wavelength scenario in order to exploit wavelength diversity; ii) relying on parallel and consensus algorithms; and iii) adapting the sparse reconstruction methods to the calibration of radio astronomical arrays. From the parallel calibration perspective, the PMCA successively estimates the DI antenna gains along with the DD and noise parameters for multiple subbands, where we enforce the coherence over the wavelength of the estimates based on physical and astronomical phenomena [8, 9, 13, 46].

The rest of the paper is organized as follows: in Section 2, we formulate the data model and its associated parallel multi-wavelength calibration problem. In Section 3, we present the overview of the proposed scheme and then describe its two main alternating steps. The constrained Cramér-Rao bound of the data model is derived in Section 4. Numerical simulations and real data tests, in Section 5, show the feasibility and superiority of the proposed scheme compared to mono-wavelength calibration. Finally, we give our conclusions in Section 6.

In the following, $(.)^*$, $(.)^\mathsf{T}$, $(.)^\mathsf{H}$, $(.)^\dagger$, $(.)^{\odot\alpha}$, $\mathfrak{R}(.)$, $\mathfrak{I}(.)$ and $[.]_n$ denote, respectively, conjugation, transposition, Hermitian transposition, pseudo-inverse, element-wise raising to α, real part, imaginary part and the n-th element of a vector. The expectation operator is $\mathcal{E}\{.\}$, $\otimes$ denotes the Kronecker product, $\exp(.)$ and $\odot$ represent the element-wise exponential function and multiplication (Hadamard product), respectively. The operator diag(.) converts a vector to a diagonal matrix with the vector aligned on the main diagonal, blkdiag(.) is the block-diagonal operator for matrices, whereas vecdiag(.) produces a vector from the main diagonal of its matrix argument and vec(.) converts a matrix to a vector by stacking the columns of its entry. The operators $\|.\|_0$, $\|.\|_2$ and $\|.\|_\mathsf{F}$ refer to the l_0 norm, i.e., the number of non-zero elements of its entry, the l_2 and Frobenius norms, respectively. $\mathbf{x} \succeq \mathbf{0}$ means that each element in $\mathbf{x}$ is non-negative.

2. Data Model & Problem Statement

2.1. Covariance Matrix Model

Consider an array comprised of P elements, with known locations, each referred by its Cartesian coordinates $\boldsymbol{\xi}_p = \left[x_p, y_p, z_p\right]^{\mathsf{T}}$ for $p = 1, \ldots, P$, that we stack in $\boldsymbol{\Xi} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_P]^{\mathsf{T}} \in \mathbb{R}^{P \times 3}$. This array is exposed to Q known strong calibration sources and Q^{U} unknown weak non-calibration sources. Let $\mathbf{D}^{\mathrm{K}} = \left[\mathbf{d}_1^{\mathrm{K}}, \ldots, \mathbf{d}_Q^{\mathrm{K}}\right] \in \mathbb{R}^{3 \times Q}$ and $\mathbf{D}^{\mathrm{U}} = [\mathbf{d}_1^{\mathrm{U}}, \ldots, \mathbf{d}_{Q^{\mathrm{U}}}^{\mathrm{U}}] \in \mathbb{R}^{3 \times Q^{\mathrm{U}}}$ denote the known (*true/nominal*) calibrator direction cosines and unknown non-calibrator direction cosines, respectively, in which each source direction $\mathbf{d} = [d_l, d_m, d_n]^{\mathsf{T}}$ can be uniquely described by a couple (d_l, d_m), since $d_n = \sqrt{1 - d_l^2 - d_m^2}$ [6, 9]. The ionosphere introduces an unknown angular-shift for each source direction [3, 12, 13], depending on the wavelength λ, which is related to the frequency $f = \mathrm{c}/\lambda$, with c denoting the speed of light. Consequently, we distinguish between the unknown *apparent* directions w.r.t. the calibrators, denoted by $\mathbf{D}_\lambda = [\mathbf{d}_{\lambda,1}, \ldots, \mathbf{d}_{\lambda,Q}]$, that depend on λ since the shift is wavelength dependent, and their *true/nominal* known directions $\mathbf{D}^{\mathrm{K}}$, i.e., without the propagation disturbances.

In the following, we describe the signal for one wavelength bin λ. Under the narrowband assumption, the steering vector $\mathbf{a}_\lambda(\mathbf{d})$ toward the direction $\mathbf{d}$ (at wavelength λ) is given by

$$\mathbf{a}_\lambda(d_l, d_m) = \frac{1}{\sqrt{P}} \exp\left(-\mathrm{j}\frac{2\pi}{\lambda}\boldsymbol{\Xi}\mathbf{d}\right), \tag{1}$$

that we gather for multiple directions in the steering matrix

$$\mathbf{A}_{\mathbf{D}_\lambda} = \frac{1}{\sqrt{P}} \exp\left(-\mathrm{j}\frac{2\pi}{\lambda}\boldsymbol{\Xi}\mathbf{D}_\lambda\right). \tag{2}$$

As in [6], we assume that all antennas have identical directional responses. Their DD gain responses (and propagation losses) are modeled by two diagonal matrices, $\boldsymbol{\Gamma}_\lambda \in \mathbb{C}^{Q \times Q}$ and $\boldsymbol{\Gamma}_\lambda^{\mathrm{U}} \in \mathbb{C}^{Q^{\mathrm{U}} \times Q^{\mathrm{U}}}$ toward the calibration and non-calibration sources, respectively.

The received signals from each antenna are partitioned into narrow subbands. The measurement of the sensor array collected in the subband with wavelength λ are stacked to the received signal vector

$$\mathbf{x}_\lambda(n) = \mathbf{G}_\lambda[\mathbf{A}_{\mathbf{D}_\lambda}\boldsymbol{\Gamma}_\lambda\mathbf{s}_\lambda(n) + \mathbf{A}_{\mathbf{D}_\lambda^{\mathrm{U}}}\boldsymbol{\Gamma}_\lambda^{\mathrm{U}}\mathbf{s}_\lambda^{\mathrm{U}}(n)] + \mathbf{n}_\lambda(n), \tag{3}$$

for the *n*-th observation, where we stress that each element of the right part of (3) depend on λ and $\mathbf{A}_\lambda^{\mathrm{U}}$ is the steering matrix for unknown sources, with $[\mathbf{x}_\lambda(n)]_p$ denoting the signal corresponding

to the p-th antenna, where $\mathbf{G}_\lambda = \mathrm{diag}(\mathbf{g}_\lambda) \in \mathbb{C}^{P \times P}$ models the DI antenna gains, with $[\mathbf{g}_\lambda]_p$ the DI antenna gain for the p-th antenna. $\mathbf{s}_\lambda(n) \in \mathbb{C}^Q$ and $\mathbf{s}_\lambda^{\mathrm{u}}(n) \in \mathbb{C}^{Q^{\mathrm{U}}}$ represent, respectively, the i.i.d. calibrator and non-calibrator signals, with $[\mathbf{s}_\lambda(n)]_q$ and $[\mathbf{s}_\lambda^{\mathrm{u}}(n)]_{q'}$, respectively, the signal corresponding to the q-th calibrator and q'-th non-calibrator, whereas $\mathbf{n}_\lambda(n) \sim C\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\lambda^{\mathrm{n}})$ denotes the i.i.d. noise vector, with $[\mathbf{n}_\lambda(n)]_p$ the thermal noise for the p-th antenna [7]. Assume that the sources are statistically independent from each other and let $\mathbf{\Sigma}_\lambda = \mathrm{diag}(\sigma_\lambda) \in \mathbb{R}^{Q \times Q}$, $\mathbf{\Sigma}_\lambda^{\mathrm{U}} = \mathrm{diag}(\sigma_\lambda^{\mathrm{u}})$ and $\mathbf{\Sigma}_\lambda^{\mathrm{n}} = \mathrm{diag}(\sigma_\lambda^{\mathrm{n}}) \in \mathbb{R}^{P \times P}$ be the diagonal covariance matrices for, respectively, the calibrators, non-calibration sources and sensor noises. Consequently, $\mathbf{s}_\lambda(n) \sim C\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\lambda)$, $\mathbf{s}_\lambda^{\mathrm{u}}(n) \sim C\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\lambda^{\mathrm{U}})$ and the covariance matrix $\mathbf{R}_\lambda = \mathcal{E}\{\mathbf{x}_\lambda \mathbf{x}_\lambda^{\mathrm{H}}\}$ of the observations corresponding to model (3) is given by

$$\mathbf{R}_\lambda = \mathbf{E}_{\mathbf{D}_\lambda} \mathbf{M}_\lambda \mathbf{E}_{\mathbf{D}_\lambda}^{\mathrm{H}} + \mathbf{R}_\lambda^{\mathrm{U}} + \mathbf{\Sigma}_\lambda^{\mathrm{n}}, \tag{4}$$

in which

$$\mathbf{E}_{\mathbf{D}_\lambda} = \mathbf{G}_\lambda \mathbf{A}_{\mathbf{D}_\lambda} \mathbf{\Sigma}_\lambda^{\frac{1}{2}}, \tag{5}$$

$$\mathbf{M}_\lambda = \mathbf{\Gamma}_\lambda \mathbf{\Gamma}_\lambda^{\mathrm{H}} = \mathrm{diag}(\mathbf{m}_\lambda), \tag{6}$$

and where we have defined the unknown covariance matrix for the non-calibration sources as

$$\mathbf{R}_\lambda^{\mathrm{U}} = \mathbf{G}_\lambda \mathbf{A}_{\mathbf{D}_\lambda^{\mathrm{U}}} \mathbf{\Gamma}_\lambda^{\mathrm{U}} \mathbf{\Sigma}_\lambda^{\mathrm{U}} (\mathbf{G}_\lambda \mathbf{A}_{\mathbf{D}_\lambda^{\mathrm{U}}} \mathbf{\Gamma}_\lambda^{\mathrm{U}})^{\mathrm{H}}. \tag{7}$$

In radio astronomy, sources (including the calibration sources), are typically much weaker than the antenna noise , so the covariance matrix is usually approximated by $\mathbf{R}_\lambda \approx \mathbf{\Sigma}_\lambda^{\mathrm{n}}$. Since the array consists of identical elements and mutual coupling can usually be ignored, it is commonly assumed that

$$\mathbf{\Sigma}_\lambda^{\mathrm{n}} = \mathrm{diag}(\sigma_\lambda^{\mathrm{n}}) \approx \sigma_\lambda^{\mathrm{n}} \mathbf{I}. \tag{8}$$

Such no mutual coupling is currently assumed due to antenna separation (though this may not exactly be true in a station array) and careful hardware implementations as in LOFAR [1].

In order to overcome the scaling ambiguities in the observation model (4), we consider the following commonly used assumptions in radio astronomy [6–8]: i) to resolve the phase ambiguity of $\mathbf{g}_\lambda$, we take its first element as the phase reference [47, 48]; ii) the phase information of of $\mathbf{\Gamma}_\lambda$ drops, consequently, we only estimate $\mathbf{M}_\lambda$, i.e., its absolute part; iii) $\mathbf{m}_\lambda$ shares a common scalar factor with $\mathbf{g}_\lambda$ and consequently, we assume that the directional gain towards the first calibration source is known / fixed; and iv) when solving for the calibrator directions, a common

rotation of all steering vectors can be compensated by the undirectional gain phase solution. We therefore fix the direction of the first calibration source at its known position [5].

2.2. Model Effects of the Wavelength on Antenna Gains, Source Direction Shifts and Source Powers

In the radio astronomical context, the antenna and source parameters of the covariance matrix are commonly assumed as wavelength dependent [7, 8]. Consequently, we assume smooth and/or known variations of the parameters $\mathbf{g}_\lambda$, $\boldsymbol{\Gamma}_\lambda$, $\boldsymbol{\Sigma}_\lambda$, $\boldsymbol{\Sigma}_\lambda^{\mathrm{u}}$ and $\boldsymbol{\Sigma}_\lambda^{\mathrm{n}}$ in (4) over λ, as commonly used in recent works on array calibration in radio astronomy [25, 49]. We summarize the particular behavior of the underlying parameters as follows:

- The DI gains, $\mathbf{g}_\lambda$, vary smoothly over λ. Hence, in order to impose coherence along subbands (not along different sensors), we define a set of smooth basis functions $b_{k,\lambda}$ as a function of the wavelength λ, for $k = 1, \ldots, K$. These basis functions define our coherence model, in which, for the p-th sensor, its gain $[\mathbf{g}_\lambda]_p$ is a function of wavelength λ and represented as a linear combination of basis functions, hence:

$$[\mathbf{g}_\lambda]_p = \sum_{k=1}^{K} b_{k,\lambda} \alpha_{k,p}, \forall \lambda \in \Lambda, p = 1, \ldots, P, \tag{9}$$

where $\alpha_{k,p}$ represents the linear coefficient corresponding to the k-th basis function for the p-th sensor. Common models for characterizing this behavior consist of classical polynomials of power law over λ [8, 25], e.g., we can consider the set of K-th order basis functions centered around a reference wavelength $\lambda_0 = \mathrm{c}/f_0$ that is defined by $b_{k,\lambda} = ((\lambda_0 - \lambda)/\lambda)^{k-1}$ [25] that impose smoothness w.r.t. frequency (keeping in mind that $b_{k,\lambda_0} = 1$). Let us denote

$$\mathbf{b}_\lambda = [b_{1,\lambda}, \ldots, b_{K,\lambda}]^\mathsf{T} \in \mathbb{R}^K, \tag{10}$$

representing all polynomial terms and rewrite (9) as

$$\mathbf{g}_\lambda = \left(\mathbf{b}_\lambda^\mathsf{T} \otimes \mathbf{I}\right) \boldsymbol{\alpha} = \mathbf{B}_\lambda \boldsymbol{\alpha}, \forall \lambda \in \Lambda, \tag{11}$$

where $\mathbf{B}_\lambda = \left(\mathbf{b}_\lambda^\mathsf{T} \otimes \mathbf{I}\right) \in \mathbb{R}^{P \times PK}$ and $\boldsymbol{\alpha}$ is thus the augmented vector of hidden variables defined by

$$\boldsymbol{\alpha} = [\alpha_{1,1}, \ldots, \alpha_{1,P}, \alpha_{2,1}, \ldots, \alpha_{K,P}]^\mathsf{T} \in \mathbb{C}^{PK}. \tag{12}$$

9

- For dipoles like in LOFAR, the DD gains contain a resonance frequency at which the response peaks, and a roll-off left and right that frequency. For simplicity and readability, we assume the DD gains, $\mathbf{\Gamma}_\lambda$, are inversely proportional to λ, i.e., $\mathbf{\Gamma}_\lambda \propto \lambda^{-1}$, as observed for the small wavelength range used in Section 5. Note that the proposed algorithm can be straightforwardly adapted with another given behavior (including the simplest case of a constant behavior across the wavelength range).

- As a consequence of the ionospheric delays, the directional shifts are proportional to λ^2 [9, 13, 46]. Furthermore, we assume that the calibration sources are well separated, which is common in radio astronomy [6, 7], and consider in the remainder of this paper that for every wavelength:

  A1) each apparent calibration source lies in an uncertainty sector with a predefined angular spread around its nominal location.

  A2) the displacement sectors of different calibration sources are not overlapping.

- The source powers, $\mathbf{\Sigma}_\lambda$ and $\mathbf{\Sigma}_\lambda^{\mathrm{U}}$, vary commonly with a power law with different spectral indexes. We consider the calibrator powers, $\mathbf{\Sigma}_\lambda$, to be known from tables, e.g., [14–17].

- The antenna noise covariance matrices, $\mathbf{\Sigma}_\lambda^{\mathrm{n}}$, do not exhibit a smooth behavior w.r.t. λ and noise is assumed to be independent over wavelength. Nevertheless, if any particular coherence model for the noise covariances is available, this knowledge can be incorporated in the proposed algorithm in a straightforward manner.

2.3. Joint Parameter Estimation Problem

In this subsection, we formulate the calibration problem as the estimation of the parameter vector of interest, $\mathbf{p}$, defined as

$$\mathbf{p} = \left[\mathbf{p}_{\lambda_1}^{\mathsf{T}}, \ldots, \mathbf{p}_{\lambda_J}^{\mathsf{T}} \right]^{\mathsf{T}}, \tag{13}$$

under the constraints given in Subsection 2.2, in which $\mathbf{p}_\lambda = [\mathbf{g}_\lambda^{\mathsf{T}}, \mathbf{d}_{\lambda,1}^{\mathsf{T}}, \ldots, \mathbf{d}_{\lambda,Q}^{\mathsf{T}}, \mathbf{m}_\lambda^{\mathsf{T}}, \sigma_\lambda^{\mathrm{n}\mathsf{T}}]^{\mathsf{T}}$, from J sample covariance matrices

$$\left\{ \hat{\mathbf{R}}_\lambda = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_\lambda(n) \mathbf{x}_\lambda^{\mathsf{H}}(n) \right\}_{\lambda \in \Lambda}, \tag{14}$$

where $\Lambda = \{\lambda_1, \ldots, \lambda_J\}$ represents the set of the J available wavelengths for the whole network and assume that $J \geq K$, i.e., accessing to data for at least K wavelengths, which is assumed to

be satisfied since, e.g., for the LOFAR, the signal is typically divided into 512 subbands while usually a low polynomial order is sufficient to represent the variations across wavelength (3 to 5 in our numerical simulations).

Data parallelism across wavelength is a recent field of research in radio astronomical observations, in which data are recorded as multiple channels at different wavelengths [1], thus, leading to a data which is not centralized but paralleled across the network. This network consists of: i) one fusion center, that does not access data; and ii) Z compute agents. The z-th agent, A_z, can only access data for a subset $\Lambda_z \subset \Lambda$ of $J_z \leq J$ subbands, and for each available wavelength, its associated sample covariance matrix is accessible for exactly one agent. Moreover, the agents cannot exchange information among themselves. Operation flow and signaling exchange between local processors and the central processor is given in the end of the algorithm description (c.f. Fig. ???)

Note that the estimation of the unknown matrices $\mathbf{R}_\lambda^\mathsf{U}$ represents the imaging step which is beyond the scope of the paper [6, 7, 9, 50]. Image synthesis [51–55] is usually performed as a separate procedure after the calibration and can be complemented by the proposed calibration approach. The main reason for this two-step procedure is that the calibration step is usually carried out based on a point source model (unlike the imaging step) with a known number of strong calibrators. In addition, the power of the calibrators are much higher than the power of the weakest (non-calibrator) sources. This causes a bias in the DI gain solutions that results in imaging artefacts know as *ghosts* [56, 57]. This effect can be mitigated by improving the sky model. Consequently, in the LOFAR pipeline, an alternating scheme between calibration and imaging is performed [1, 3].

3. Proposed Parallel Multi-wavelength Calibration Algorithm

3.1. Overview of the Proposed Parallel Multi-wavelength Calibration Algorithm

In this subsection, we define the main steps of our proposed algorithm, then, in the following subsections, we describe each step in detail.

It is well established that a statistically efficient estimator can be obtained via the Maximum Likelihood method. However, from a computational viewpoint, its exact evaluation appears to be intractable in the radio astronomical context [6]. With a large number of samples, statistically

11

Algorithm 1: Parallel Multi-wavelength Calibration Algorithm

Input: $\{\hat{\mathbf{R}}_\lambda, \mathbf{m}_\lambda^{[0]}\}_{\lambda \in \Lambda}, \mathbf{D}^\kappa, , \eta_\mathbf{p}$;

Init: set $i = 0, \{\mathbf{g}_\lambda = \mathbf{g}_\lambda^{[0]}\}_{\lambda \in \Lambda}, \{\mathbf{D}_\lambda = \mathbf{D}^\kappa, \mathbf{m}_\lambda = \mathbf{m}_\lambda^{[0]}, \mathbf{\Omega}_\lambda = \mathbf{1}_{P \times P}\}_{\lambda \in \Lambda}$;

repeat

1     $i = i + 1$;

2   Estimate paralelelly $\{\mathbf{g}_\lambda^{[i]}\}_{\lambda \in \Lambda}$ with *Algorithm 1.2*;

3   Estimate paralelelly $\{\mathbf{D}_\lambda^{[i]}, \mathbf{m}_\lambda^{[i]}, \sigma_\lambda^{\mathrm{n}[i]}\}_{\lambda \in \Lambda}$ with *Algorithm 1.3*;

4   Update locally $\{\mathbf{\Omega}_\lambda^{[i]} = \left(\sigma_\lambda^{\mathrm{n}[i]} \sigma_\lambda^{\mathrm{n}[i]\mathsf{T}}\right)^{\odot - \frac{1}{2}}\}_{\lambda \in \Lambda}$;

until $\left\|\mathbf{p}^{[i-1]} - \mathbf{p}^{[i]}\right\|_2 \leq \left\|\mathbf{p}^{[i]}\right\|_2 \eta_\mathbf{p}$;

Output: $\hat{\mathbf{p}} = [\mathbf{p}_{\lambda_1}^{[i]\mathsf{T}}, \ldots, \mathbf{p}_{\lambda_J}^{[i]\mathsf{T}}]^\mathsf{T}$;

---

efficient estimators can be devised using the Weighting Least Squares approach. In this context, we define, for each $\lambda \in \Lambda$, the local cost function [5] to be minimized as

$$\kappa_\lambda(\mathbf{p}_\lambda) = \left\|\mathbf{W}_\lambda^{-\frac{1}{2}} \left(\mathbf{R}_\lambda(\mathbf{p}_\lambda) - \hat{\mathbf{R}}_\lambda\right) \mathbf{W}_\lambda^{-\frac{1}{2}}\right\|_\mathsf{F}^2, \tag{15}$$

in which

$$\mathbf{R}_\lambda(\mathbf{p}_\lambda) = \mathbf{E}_{\mathbf{D}_\lambda} \mathbf{M}_\lambda \mathbf{E}_{\mathbf{D}_\lambda}^\mathsf{H} + \mathbf{\Sigma}_\lambda^\mathrm{n} \tag{16}$$

denotes the covariance matrix when the contribution of the (weakest) non-calibrators is absorbed in the noise, and $\mathbf{W}_\lambda$ is the weighting matrix. The optimal weighting matrix for Gaussian noise is the true covariance matrix of the observation [58], which is generally unknown. Due to the low SNR regime in which radio astronomical arrays usually operate, it is often assumed that the noise power per receiving element is much larger than the source powers [5, 18], we therefore consider $\mathbf{W}_\lambda = \mathbf{I}$ in our alternating algorithm as an initialization and refine it as $\mathbf{W}_\lambda = \mathbf{\Sigma}_\lambda^\mathrm{n}$ once we obtain an estimate of $\mathbf{\Sigma}_\lambda^\mathrm{n}$. Since $\mathbf{\Sigma}_\lambda^\mathrm{n}$ is diagonal, we rewrite the local cost function (15), i.e., the cost function associated with the wavelength λ, as

$$\kappa_\lambda(\mathbf{p}_\lambda) = \left\|\left(\mathbf{R}_\lambda(\mathbf{p}_\lambda) - \hat{\mathbf{R}}_\lambda\right) \odot \mathbf{\Omega}_\lambda\right\|_\mathsf{F}^2, \text{ with} \tag{17}$$

$$\mathbf{\Omega}_\lambda = \left(\sigma_\lambda^\mathrm{n} \sigma_\lambda^{\mathrm{n}\mathsf{T}}\right)^{\odot - \frac{1}{2}}. \tag{18}$$

Finally, with (13), we define the cost function for the entire network as

$$\kappa(\mathbf{p}) = \sum_{\lambda \in \Lambda} \kappa_\lambda(\mathbf{p}_\lambda). \tag{19}$$

Our aim is to estimate $\mathbf{p}$ by minimizing $\kappa(\mathbf{p})$ in an alternating and parallel manner. Note that the overall problem is non-convex, so we can not claim finding the global minimum when the algorithm converge (neverthless, in practice we might have good initialisation that converges to the global minimum as shown in the real data simulation). We first estimate locally in each agent the parameter vector $\{\mathbf{g}_\lambda\}_{\lambda \in \Lambda}$, with the remaining parameters in $\mathbf{p}$ fixed as described in Subsection 3.2, by reformulating the problem as a consensus problem, in which the updates of α provide coherence among wavelength. In a second step, we estimate the variables $\left\{\mathbf{m}_\lambda, \mathbf{d}_{\lambda,1}, \ldots, \mathbf{d}_{\lambda,Q}, \sigma_\lambda^n\right\}_{\lambda \in \Lambda}$ for fixed $\{\mathbf{g}_\lambda\}_{\lambda \in \Lambda}$, by using a sparse representation approach as described in Subsection 3.3. Finally, we update the weighting matrices $\{\boldsymbol{\Omega}_\lambda\}_{\lambda \in \Lambda}$. During these procedures, the amount of information that needs to be exchanged between the fusion center and the compute agents is much less than the volume of data being calibrated, making this scheme computationally feasible. The overall procedure, referred to as Parallel Multi-wavelength Calibration Algorithm (PMCA), is presented in *Algorithm 1*. The algorithm is carefully initialized with the true/nominal calibrator parameters and an initial guess for the antenna gains, e.g. from precedent calibration, or by default by the unit sensor gain, $\mathbf{g}_\lambda^{[0]} = \mathbf{1}$. The stopping criterion $\eta_\mathbf{p}$ has to be sufficiently small to assure convergence. In the following subsections, we detail the two major alternating optimization steps of the proposed PMCA.

3.2. *Direction Independent Antenna Gain Estimation (*Algorithm 1.2*)*

In this subsection, we describe *Algorithm 1.2* of the PMCA. As shown in *Algorithm 1.2*, this optimization step is performed w.r.t. the DI gain parameters $\{\mathbf{g}_\lambda\}_{\lambda \in \Lambda}$, while the remaining parameters $\left\{\mathbf{m}_\lambda, \mathbf{d}_{\lambda,1}, \ldots, \mathbf{d}_{\lambda,Q}, \sigma_\lambda^n\right\}_{\lambda \in \Lambda}$ of $\mathbf{p}$ are fixed. During this step, each agent calibrates the sensor gains $\mathbf{g}_\lambda$ using the data available locally and *Algorithm 1.2.2*. Then, the agents transfers their parameter estimates to the centralized location. At the fusion center, smoothness of the parameters across wavelength is enforced (line 3 of *Algorithm 1.2*), using the coherence parameter vector α, that is passed to each agent to provide coherent processing across the whole wavelength range, and thus improving the local calibration procedure.

At this point, we distinguish between centralized and parallel based estimation of α. Specifically:

- Centralized calibration leads to a direct estimation scheme of α from the data, in which we substitute in the minimization of (19) the sensor gains $\{\mathbf{g}_\lambda\}_{\lambda \in \Lambda}$ by α according to (11).

Algorithm 1.2: estimation of $\{\mathbf{g}_\lambda\}_{\lambda\in\Lambda}$

Input: $\{\hat{\mathbf{R}}_\lambda\}_{\lambda\in\Lambda}, \mathbf{p}^{[i-1]}, \eta_{\mathbf{g}}$;

Init: set $t = 0, \{\mathbf{g}_\lambda^{[t]} = \mathbf{g}_\lambda^{[i-1]}\}_{\lambda\in\Lambda}, \{\mathbf{R}_\lambda^\kappa = \mathbf{A}_{\mathbf{D}_\lambda^{[i-1]}}\mathbf{M}_\lambda^{[i-1]}\mathbf{A}_{\mathbf{D}_\lambda^{[i-1]}}^{\mathsf{H}}\}_{\lambda\in\Lambda}$;

repeat

1    $t = t + 1$;

2   Estimate locally $\{\mathbf{g}_\lambda^{[t]}\}_{\lambda\in\Lambda}$ with *Algorithm 1.2.2* ;

3   Estimate $\boldsymbol{\alpha}^{[t]}$ at the fusion center with (36);

4   Update locally the Lagrange multipliers $\{\mathbf{y}_\lambda^{[t]}\}_{\lambda\in\Lambda}$ with (24);

until $\sum_{\lambda\in\Lambda} \left\|\mathbf{g}_\lambda^{[t-1]} - \mathbf{g}_\lambda^{[t]}\right\|_2 \leq \sum_{\lambda\in\Lambda} \left\|\mathbf{g}_\lambda^{[t]}\right\|_2 \eta_{\mathbf{g}}$;

5 Deduce locally $\{\hat{\mathbf{g}}_\lambda\}_{\lambda\in\Lambda}$ from $\boldsymbol{\alpha}^{[t]}$ with (11);

Output: $\{\hat{\mathbf{g}}_\lambda\}_{\lambda\in\Lambda}$;

However, this requires access to all data by minimizing (19) w.r.t. $\boldsymbol{\alpha}$, which is computationally unfeasible due to the restriction resulting from the large data volumes.

- Let us recall that Z computational agents are disposed on a network (see, Subsection 2.3), where the z-th agent, A_z, accesses data for wavelengths $\lambda \in \Lambda_z \subset \Lambda$. In this perspective, we propose a parallel calibration scheme, in which the sensor gains corresponding to each wavelength λ are estimated locally and consensus is enforced among agents by imposing constraint (11).

With this network setup, we formulate the parallel calibration procedure as

$$\underset{\boldsymbol{\alpha},\{\mathbf{g}_\lambda\}_{\lambda\in\Lambda}}{\text{minimize}} \sum_{\lambda\in\Lambda} \tilde{\kappa}_\lambda^{[i]}(\mathbf{g}_\lambda)$$

$$\text{subject to } \mathbf{g}_\lambda = \mathbf{B}_\lambda\boldsymbol{\alpha}, \forall\lambda \in \Lambda, \tag{20}$$

in which $\tilde{\kappa}_\lambda^{[i]}(\mathbf{g}_\lambda) = \kappa_\lambda(\mathbf{p}_\lambda|\mathbf{m}_\lambda^{[i]}, \mathbf{D}_\lambda^{[i]}, \sigma_\lambda^{\mathrm{n}[i]})$, i is the [i]-th iteration of *Algorithm 1*, and where the cost function consists of a sum of independent cost functions, one for each subband, that are coupled through the coherence constraints which however are independent across sensors. A common way to solve (20) is to consider the problem as a consensus optimization problem [26]

Algorithm 1.2.2: local estimation of $\{\mathbf{g}_\lambda\}_{\lambda\in\Lambda_z}$

Input: $\{\hat{\mathbf{R}}_\lambda, \mathbf{R}_\lambda^\kappa, \mathbf{g}_\lambda^{[t-1]}, \mathbf{y}_\lambda^{[t-1]}, \mathbf{\Omega}_\lambda^{[i]}\}_{\lambda\in\Lambda_z}, \alpha^{[t-1]}, \eta_\mathbf{g}$;

Init: set $\{t_\lambda = 0, \mathbf{g}_\lambda^{[t_\lambda]} = \mathbf{g}_\lambda^{[t-1]}\}_{\lambda\in\Lambda_z}$;

foreach $\lambda \in \Lambda_z$ **do**

    repeat

1        $t_\lambda = t_\lambda + 1$;

        for $p = 1, \ldots, P$ **do**

2            $\hat{\mathbf{r}}_\lambda^p = \mathcal{S}_p\left(\hat{\mathbf{R}}_\lambda\right)$;

3            $\mathbf{z} = \mathcal{S}_p\left(\mathbf{R}_\lambda^\kappa \mathbf{G}_\lambda^*\right)$;

4            $\mathbf{z}_\omega = \mathbf{z} \odot \mathcal{S}_p\left(\mathbf{\Omega}_\lambda^{[i]}\right)$;

5            $[\mathbf{g}_\lambda^{[t_\lambda]}]_p = \dfrac{2\mathbf{z}_\omega^\mathsf{H}(\hat{\mathbf{r}}_\lambda^p \odot \omega) + \rho[\mathbf{B}_\lambda\alpha^{[t-1]}]_p - [\mathbf{y}_\lambda^{[t-1]}]_p}{2(\mathbf{z}_\omega^\mathsf{H}\mathbf{z}_\omega) + \rho}$;

    until $\left\|\mathbf{g}_\lambda^{[t_\lambda-1]} - \mathbf{g}_\lambda^{[t_\lambda]}\right\|_2 \le \left\|\mathbf{g}_\lambda^{[t_\lambda]}\right\|_2 \eta_\mathbf{g}$;

Output: $\{\hat{\mathbf{g}}_\lambda = \mathbf{g}_\lambda^{[t_\lambda]}\}_{\lambda\in\Lambda_z}$;

---

and consequently to use the augmented Lagrangian, given by [59]

$$L^{[i]}\left(\{\mathbf{g}_\lambda\}_{\lambda\in\Lambda}, \alpha, \{\mathbf{y}_\lambda\}_{\lambda\in\Lambda}\right) = \sum_{\lambda\in\Lambda} \tilde{\kappa}_\lambda^{[i]}(\mathbf{g}_\lambda) + \Re\left\{\mathbf{y}_\lambda^\mathsf{H}(\mathbf{g}_\lambda - \mathbf{B}_\lambda\alpha)\right\} + \frac{\rho}{2}\|\mathbf{g}_\lambda - \mathbf{B}_\lambda\alpha\|_2^2$$

$$= \Re\left\{\sum_{\lambda\in\Lambda} L_\lambda^{[i]}(\mathbf{g}_\lambda, \alpha, \mathbf{y}_\lambda)\right\}, \tag{21}$$

where $\{\mathbf{y}_\lambda\}_{\lambda\in\Lambda}$ are the J Lagrange multipliers and ρ is the regularization term, chosen by the practitioner. In order to solve (20), we resort to the consensus ADMM [26]. Let t denote the local iteration counter of *Algorithm 1.2*, the updates for the [t]-th iteration are given by

$$\mathbf{g}_\lambda^{[t]} = \arg\min_{\mathbf{g}_\lambda} \Re\left\{L_\lambda^{[i]}\left(\mathbf{g}_\lambda, \alpha^{[t-1]}, \mathbf{y}_\lambda^{[t-1]}\right)\right\}, \lambda \in \Lambda, \tag{22}$$

$$\alpha^{[t]} = \arg\min_{\alpha} \Re\left\{\sum_{\lambda\in\Lambda} L_\lambda^{[i]}\left(\mathbf{g}_\lambda^{[t]}, \alpha, \mathbf{y}_\lambda^{[t-1]}\right)\right\}, \tag{23}$$

$$\mathbf{y}_\lambda^{[t]} = \mathbf{y}_\lambda^{[t-1]} + \rho\left(\mathbf{g}_\lambda^{[t]} - \mathbf{B}_\lambda\alpha^{[t]}\right), \lambda \in \Lambda, \tag{24}$$

that correspond to, respectively, the sensor gain update, the smoothness parameter update and the Lagrange multiplier update. The minimization of (22) is the computationally most expensive step and is performed locally by each agent. Similarly, (24) is solved locally, whereas (23) is solved at the fusion center. Procedures for obtaining (22) and (23) are detailed in the following.

15

3.2.1. Minimization of (22)

Toward the minimization of (22), we follow an iterative optimization approach based on [18, 30], that we adapt to our parallel minimization. Let us assume that $\mathbf{g}_\lambda$ and $\mathbf{g}_\lambda^*$ are two independent variables. We then regard $\mathbf{g}_\lambda^*$ as fixed and minimize $L_\lambda^{[i]}\left(\mathbf{g}_\lambda, \mathbf{g}_\lambda^*, \alpha^{[t]}, \mathbf{y}_\lambda^{[t]}\right)$ w.r.t. $\mathbf{g}_\lambda$ only, and without considering the diagonal elements in the cost functions in (20) that contain the unknown noise variances $\sigma_\lambda^{\mathrm{n}}$. In this case, the local cost function becomes separable w.r.t. the elements of $\mathbf{g}_\lambda$, hence,

$$\tilde{\kappa}_\lambda^{[i]}(\mathbf{g}_\lambda) = \sum_{p=1}^{P} \tilde{\kappa}_\lambda^{p[i]}([\mathbf{g}_\lambda]_p), \tag{25}$$

where $\tilde{\kappa}_\lambda^{p[i]}([\mathbf{g}_\lambda]_p)$ corresponds to the cost function for the p-th row of $\mathbf{R}_\lambda$, which depends only on $[\mathbf{g}_\lambda]_p$ since the remaining parameters are considered as fixed in this step. Let us define the operator $\mathcal{S}_p(.)$, that converts the p-th row of a matrix to a vector and removes the p-th element of this selected vector. Further, define the vector $\hat{\mathbf{r}}_\lambda^p = \mathcal{S}_p\left(\hat{\mathbf{R}}_\lambda\right)$ and the weighting vector $\omega = \mathcal{S}_p\left(\mathbf{\Omega}_\lambda^{[i]}\right)$. We can thus write $\tilde{\kappa}_\lambda^{p[i]}([\mathbf{g}_\lambda]_p)$ in (25) as

$$\tilde{\kappa}_\lambda^{p[i]}([\mathbf{g}_\lambda]_p) = \left\|\left(\hat{\mathbf{r}}_\lambda^p - \mathbf{z}[\mathbf{g}_\lambda]_p\right) \odot \omega\right\|_2^2, \tag{26}$$

in which $\mathbf{z} = \mathcal{S}_p\left(\mathbf{R}_\lambda^{\mathrm{\kappa}} \mathbf{G}_\lambda^*\right)$ and where

$$\mathbf{R}_\lambda^{\mathrm{\kappa}} = \mathbf{A}_{\mathbf{D}_\lambda} \mathbf{M}_\lambda \mathbf{A}_{\mathbf{D}_\lambda}^{\mathsf{H}} \tag{27}$$

represents the estimated calibrator sky model. Then, we decompose the augmented Lagrangian in (22) w.r.t. the elements of $\mathbf{g}_\lambda$ as

$$L_\lambda^{[i]}\left(\mathbf{g}_\lambda, \alpha^{[t]}, \mathbf{y}_\lambda^{[t]}\right) = \sum_{p=1}^{P} L_\lambda^{p[i]}\left([\mathbf{g}_\lambda]_p, \alpha^{[t]}, \mathbf{y}_\lambda^{[t]}\right), \tag{28}$$

in which

$$L_\lambda^{p[i]}\left([\mathbf{g}_\lambda]_p, \alpha^{[t]}, \mathbf{y}_\lambda^{[t]}\right) = \tilde{\kappa}_\lambda^{p[i]}\left([\mathbf{g}_\lambda]_p\right) + \tag{29}$$
$$[\mathbf{y}_\lambda^{*[t]}]_p \left([\mathbf{g}_\lambda]_p - \left[\mathbf{B}_\lambda \alpha^{[t]}\right]_p\right) + \frac{\rho}{2} \left\|[\mathbf{g}_\lambda]_p - \left[\mathbf{B}_\lambda \alpha^{[t]}\right]_p\right\|_2^2.$$

By using standard inversion techniques [25, Section 4], we calculate the minimizer of (29) as

$$[\hat{\mathbf{g}}_\lambda^{[t+1]}]_p = \frac{2\mathbf{z}_\omega^{\mathsf{H}}\left(\hat{\mathbf{r}}_\lambda^p \odot \omega\right) + \rho\left[\mathbf{B}_\lambda \alpha^{[t]}\right]_p - [\mathbf{y}_\lambda^{[t]}]_p}{2\left(\mathbf{z}_\omega^{\mathsf{H}} \mathbf{z}_\omega\right) + \rho}, \tag{30}$$

where $\mathbf{z}_\omega = \mathbf{z} \odot \mathcal{S}_p\left(\mathbf{\Omega}_\lambda^{[i]}\right)$. Then, we directly update $[\mathbf{g}_\lambda^{*[t+1]}]_p = \left([\hat{\mathbf{g}}_\lambda^{[t+1]}]_p\right)^*$ and proceed in the same manner with the remaining parameters in $\mathbf{g}_\lambda$. This procedure is summarized in *Algorithm 1.2.2* and is repeated until convergence.

3.2.2. Minimization of (23)

After gathering the estimates $\{\mathbf{g}_\lambda^{[t+1]}\}_{\lambda \in \Lambda}$, the fusion center can obtain a closed-form expression of $\boldsymbol{\alpha}$, and then, its estimated value, $\hat{\boldsymbol{\alpha}}^{[t+1]}$, is sent to all agents in the network. Specifically,

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}^{[t+1]} &= \arg\min_{\boldsymbol{\alpha}} \mathfrak{R} \left\{ \sum_{\lambda \in \Lambda} L_\lambda^{[i]}\left(\mathbf{g}_\lambda^{[t+1]}, \boldsymbol{\alpha}, \mathbf{y}_\lambda^{[t]}\right) \right\} \\
&= \arg\min_{\boldsymbol{\alpha}} \mathfrak{R} \left\{ \sum_{\lambda \in \Lambda} \mathbf{y}_\lambda^{[t]\mathsf{H}}\left(\mathbf{g}_\lambda^{[t+1]} - \mathbf{B}_\lambda \boldsymbol{\alpha}\right) + \frac{\rho}{2} \left\|\mathbf{g}_\lambda^{[t+1]} - \mathbf{B}_\lambda \boldsymbol{\alpha}\right\|_2^2 \right\},
\end{aligned}
\tag{31}
$$

which leads, after some calculus, to

$$
\hat{\boldsymbol{\alpha}}^{[t+1]} = \left(\sum_{\lambda \in \Lambda} \rho \mathbf{B}_\lambda^\mathsf{T} \mathbf{B}_\lambda\right)^\dagger \left(\sum_{\lambda \in \Lambda} \mathbf{B}_\lambda^\mathsf{T}\left(\mathbf{y}_\lambda^{[t]} + \rho \mathbf{g}_\lambda^{[t+1]}\right)\right).
\tag{32}
$$

Expression (32) can be simplified by means of (11), as

$$
\hat{\boldsymbol{\alpha}}^{[t+1]} = \frac{1}{\rho} \left(\left(\sum_{\lambda \in \Lambda} \mathbf{b}_\lambda \mathbf{b}_\lambda^\mathsf{T}\right)^\dagger \otimes \mathbf{I}\right) \left(\sum_{\lambda \in \Lambda} \mathbf{b}_\lambda \otimes \left(\mathbf{y}_\lambda^{[t]} + \rho \mathbf{g}_\lambda^{[t+1]}\right)\right).
\tag{33}
$$

with $J \geq K$ (see, Subsection 2.2). Finally, denoting

$$
\mathbf{T} = \frac{1}{\rho} \left(\sum_{\lambda \in \Lambda} \mathbf{b}_\lambda \mathbf{b}_\lambda^\mathsf{T}\right)^\dagger,
\tag{34}
$$

$$
\mathbf{U}^{[t+1]} = \sum_{\lambda \in \Lambda} \left(\mathbf{y}_\lambda^{[t]} + \rho \mathbf{g}_\lambda^{[t+1]}\right) \mathbf{b}_\lambda^\mathsf{T}
\tag{35}
$$

and by use of the Kronecker product property $\mathrm{vec}(\mathbf{ABC}) = \left(\mathbf{C}^\mathsf{T} \otimes \mathbf{A}\right) \mathrm{vec}(\mathbf{B})$, (23) is reduced to the following compact analytical expression,

$$
\hat{\boldsymbol{\alpha}}^{[t+1]} = \mathrm{vec}\left(\mathbf{I}\mathbf{U}^{[t+1]}\mathbf{T}^\mathsf{T}\right) = \mathrm{vec}\left(\mathbf{U}^{[t+1]}\mathbf{T}^\mathsf{T}\right).
\tag{36}
$$

3.3. Direction Dependent Parameter and Noise Power Estimation (Algorithm 1.3)

In this subsection, we describe *Algorithm 1.3* of the PMCA dedicated to the estimation of the DD parameters and noise powers, i.e., $\left\{\mathbf{m}_\lambda, \mathbf{d}_{\lambda,1}, \ldots, \mathbf{d}_{\lambda,Q}, \sigma_\lambda^\mathrm{n}\right\}_{\lambda \in \Lambda}$ for fixed $\{\mathbf{g}_\lambda\}_{\lambda \in \Lambda}$, which is based mainly on a sparse representation framework with a parallel implementation.

17

By taking into account the assumptions A1 and A2 of Subsection 2.2, each of following dictionaries is obtained from sampling the uncertainty sector corresponding to a particular calibrator. Let us recall that according to our model, the shifts of the directions are proportional to λ^2 [13] (see, Subsection 2.2). Consequently, we impose the same behavior w.r.t. the wavelength in the grid resolutions, i.e., $\Delta l_\lambda^q \propto \lambda^2, \Delta m_\lambda^q \propto \lambda^2$, by scaling them around λ_0 as

$$\Delta l_\lambda^q = \left(\frac{\lambda}{\lambda_0}\right)^2 \Delta l_{\lambda_0}^q, \tag{37}$$

$$\Delta m_\lambda^q = \left(\frac{\lambda}{\lambda_0}\right)^2 \Delta m_{\lambda_0}^q. \tag{38}$$

Thus, we define JQ dictionaries of steering vectors, $\tilde{\mathbf{A}}_{q,\lambda} \in \mathbb{C}^{P \times N_q}$, for $q = 1, \ldots, Q, \lambda \in \Lambda$, which contain N_q steering vectors, centered around the true/nominal direction of the q-th calibrator, namely $\mathbf{d}_q^\kappa$, with resolution $\left(\Delta l_\lambda^q, \Delta m_\lambda^q\right)$ and $N_q \gg 1$. These dictionary steering matrices are gathered in

$$\tilde{\mathbf{A}}_\lambda = \left[\tilde{\mathbf{A}}_{1,\lambda}, \ldots, \tilde{\mathbf{A}}_{Q,\lambda}\right] \in \mathbb{C}^{P \times N_g}, \lambda \in \Lambda, \tag{39}$$

with $N_g = \sum_{q=1}^Q N_q$ denoting the total number of directions on the grid.

We define then J vectors, $\{\tilde{\mathbf{m}}_\lambda\}_{\lambda \in \Lambda}$, as

$$\tilde{\mathbf{m}}_\lambda = \left[\tilde{\mathbf{m}}_{1,\lambda}^\mathsf{T}, \ldots, \tilde{\mathbf{m}}_{Q,\lambda}^\mathsf{T}\right]^\mathsf{T} \in \mathbb{R}^{N_g}, \lambda \in \Lambda, \tag{40}$$

which contain the squared DD gains towards all calibrators, where $\tilde{\mathbf{m}}_{q,\lambda}$ is the sparse vector associated with $\tilde{\mathbf{A}}_{q,\lambda}$. Due to the previous assumption of non-overlapping displacement sectors, each $\tilde{\mathbf{m}}_{q,\lambda}$ is exactly 1-sparse, i.e., $\left\|\tilde{\mathbf{m}}_{q,\lambda}\right\|_0 = 1$, for $q = 1, \ldots, Q, \forall \lambda \in \Lambda$. Since the shift resolution in the dictionaries is made proportional to λ^2, the support of $\tilde{\mathbf{m}}_{q,\lambda}$ is independent of λ. To proceed, we exploit that $\Gamma_\lambda \propto \lambda^{-1}$ (see, Subsection 2.2) in order to estimate a unique sparse vector for all wavelengths, namely $\tilde{\mathbf{m}}$. More precisely, under this assumption, we define $\tilde{\mathbf{m}} = \left[\tilde{\mathbf{m}}_1^\mathsf{T}, \ldots, \tilde{\mathbf{m}}_Q^\mathsf{T}\right]^\mathsf{T}$ as

$$\tilde{\mathbf{m}}_\lambda = \left(\frac{\lambda_0}{\lambda}\right)^2 \tilde{\mathbf{m}}, \forall \lambda \in \Lambda, \tag{41}$$

which can be also adjusted for other existing models of Γ_λ.

Using (16), the approximate covariance model can be rewritten as

$$\mathbf{R}_\lambda = \tilde{\mathbf{E}}_\lambda \tilde{\mathbf{M}} \tilde{\mathbf{E}}_\lambda^\mathsf{H} + \Sigma_\lambda^\mathrm{n}, \tag{42}$$

18

Algorithm 1.3: estimation of $\{\mathbf{D}_\lambda, \mathbf{m}_\lambda, \sigma_\lambda^n\}_{\lambda \in \Lambda}$

Input: $\{\hat{\mathbf{R}}_\lambda, \mathbf{g}_\lambda^{[i]}\}_{\lambda \in \Lambda}, \mathbf{p}^{[i-1]}, \eta_{\tilde{\mathbf{m}}}$;

Init: set $k = 0$, $\{\mathbf{g}_\lambda^{[k]} = \mathbf{g}_\lambda^{[i]}, \mathbf{M}_\lambda^{[k]} = \mathbf{M}_\lambda^{[i-1]}\}_{\lambda \in \Lambda}$, $\{\mathbf{D}_\lambda^{[k]} = \mathbf{D}_\lambda^{[i-1]}, \sigma_\lambda^{n[k]} = \sigma_\lambda^{n[i-1]}\}_{\lambda \in \Lambda}$;

repeat

1     $k = k + 1$;

    for $q = 1, \ldots, Q$ **do**

       foreach $\mathsf{A}_z, z = 1, \ldots, Z$ **do**

          foreach $\lambda \in \Lambda_z$ **do**

2   Calculate locally $\underline{\hat{\mathbf{r}}}_\lambda^q = \underline{\hat{\mathbf{r}}}_\lambda - \sum_{q'=1}^{q-1} \underline{\mathbf{V}}_\lambda^{q'} \tilde{\mathbf{m}}_{q'}^{[k]} - \sum_{q'=q+1}^{Q} \underline{\mathbf{V}}_\lambda^{q'} \tilde{\mathbf{m}}_{q'}^{[k-1]}$;

3   Send $\sum_{\lambda \in \Lambda_z} \underline{\mathbf{V}}_\lambda^{q\mathsf{T}} \underline{\hat{\mathbf{r}}}_\lambda^q$ to the fusion center;

4   Hard thresholding by the fusion center $\tilde{\mathbf{m}}_q^{[k]} = \mathcal{H}_1 \left(\tilde{\mathbf{m}}_q^{[k-1]} + \tau_q^{[k]} \sum_{\lambda \in \Lambda} \underline{\mathbf{V}}_\lambda^{q\mathsf{T}} \underline{\hat{\mathbf{r}}}_\lambda^q \right)$;

5   The fusion center communicates the non-zero element of $\tilde{\mathbf{m}}_q^{[k]}$ and its associated

         direction $\mathbf{d}_{\lambda_0,q}^{[k]}$;

   until $\left\| \tilde{\mathbf{m}}^{[k-1]} - \tilde{\mathbf{m}}^{[k]} \right\|_2 \leq \left\| \tilde{\mathbf{m}}^{[k]} \right\|_2 \eta_{\tilde{\mathbf{m}}}$;

6 Deduce locally $\{\hat{\mathbf{m}}_\lambda\}_{\lambda \in \Lambda}$ from $\tilde{\mathbf{m}}^{[k]} = \left[\tilde{\mathbf{m}}_1^{[k]\mathsf{T}}, \ldots, \tilde{\mathbf{m}}_Q^{[k]\mathsf{T}} \right]^\mathsf{T}$, with (41);

7 Deduce locally $\{\hat{\mathbf{D}}_\lambda\}_{\lambda \in \Lambda}$ from $\mathbf{D}_{\lambda_0}^{[k]} = \left[\mathbf{d}_{\lambda_0,1}^{[k]}, \ldots, \mathbf{d}_{\lambda_0,Q}^{[k]} \right]$ with (37) and (38);

8 Estimate locally $\{\sigma_\lambda^n\}_{\lambda \in \Lambda}$ with (57);

   Output: $\{\hat{\mathbf{D}}_\lambda, \hat{\mathbf{m}}_\lambda, \hat{\sigma}_\lambda^n\}_{\lambda \in \Lambda}$;

---

in which $\tilde{\mathbf{M}} = \text{diag}(\tilde{\mathbf{m}})$,

$$\tilde{\boldsymbol{\Sigma}}_\lambda = \text{blkdiag}\left(\mathbf{I}_{N_1 \times N_1} [\boldsymbol{\sigma}_\lambda]_1, \ldots, \mathbf{I}_{N_Q \times N_Q} [\boldsymbol{\sigma}_\lambda]_Q\right) \text{ and} \tag{43}$$

$$\tilde{\mathbf{E}}_\lambda = \frac{\lambda_0}{\lambda} \mathbf{G}_\lambda \tilde{\mathbf{A}}_\lambda \tilde{\boldsymbol{\Sigma}}_\lambda^{\frac{1}{2}}. \tag{44}$$

Let us then define

$$\mathbf{V}_\lambda = (\boldsymbol{\Sigma}_\lambda^n)^{-\frac{1}{2}} \tilde{\mathbf{E}}_\lambda^* \otimes (\boldsymbol{\Sigma}_\lambda^n)^{-\frac{1}{2}} \tilde{\mathbf{E}}_\lambda, \tag{45}$$

$$\mathbf{N}_\lambda = \boldsymbol{\Sigma}_\lambda^{n-\frac{1}{2}} \otimes \boldsymbol{\Sigma}_\lambda^{n-\frac{1}{2}}, \tag{46}$$

$$\hat{\mathbf{r}}_\lambda = \text{vec}\left(\hat{\mathbf{R}}_\lambda \odot \boldsymbol{\Omega}_\lambda\right), \tag{47}$$

computed with the previous estimate of $\boldsymbol{\Sigma}_\lambda^n$ and $\boldsymbol{\Omega}_\lambda$ in (18). Thus, we formulate the minimization

problem as

$$\underset{\tilde{\mathbf{m}}, \{\sigma_\lambda^n\}_{\lambda\in\Lambda}}{\text{minimize}} \sum_{\lambda\in\Lambda} \left\| \hat{\mathbf{r}}_\lambda - \mathbf{V}_\lambda\tilde{\mathbf{m}} - \mathbf{N}_\lambda\sigma_\lambda^n \right\|_2^2$$

$$\text{subject to } \tilde{\mathbf{m}} \succeq \mathbf{0}, \sigma_\lambda^n \succeq \mathbf{0}, \forall\lambda\in\Lambda \tag{48}$$

$$\left\| \tilde{\mathbf{m}}_q \right\|_0 = 1 \text{ for } q = 1,\dots,Q.$$

Since the p-th element of σ_λ^n, $[\sigma_\lambda^n]_p$, is only present in the p-th diagonal term of $\mathbf{R}_\lambda$, ignoring this term does not affect (asymptotically) the estimation of $\tilde{\mathbf{m}}$. Consequently, it is more convenient to estimate σ_λ^n after the estimation of $\tilde{\mathbf{m}}$.

For this purpose, we denote $\hat{\underline{\mathbf{r}}}_\lambda$ and $\underline{\mathbf{V}}_\lambda$, that refer, respectively, to $\hat{\mathbf{r}}_\lambda$ and $\mathbf{V}_\lambda$, where the elements corresponding to the diagonal of $\mathbf{R}_\lambda$ are discarded. We further define

$$h_\lambda(\tilde{\mathbf{m}}) = \left\| \hat{\underline{\mathbf{r}}}_\lambda - \underline{\mathbf{V}}_\lambda\tilde{\mathbf{m}} \right\|_2^2, \lambda\in\Lambda, \tag{49}$$

to obtain the solution of $\tilde{\mathbf{m}}$ in (48) after supposing $\sigma_\lambda^n > \mathbf{0}$, as

$$\hat{\tilde{\mathbf{m}}} = \arg\min_{\tilde{\mathbf{m}}} \sum_{\lambda\in\Lambda} h_\lambda(\tilde{\mathbf{m}})$$

$$\text{subject to } \tilde{\mathbf{m}} \succeq \mathbf{0}, \tag{50}$$

$$\left\| \tilde{\mathbf{m}}_q \right\|_0 = 1 \text{ for } q = 1,\dots,Q,$$

which is used in *Algorithm 1.3*.

To consider the l_0 constraints in (50), which are non-convex and NP-hard [60], we choose the Distributed Iterative Hard Thresholding method [61–63], which is based on Iterative Hard Thresholding [64]. This greedy algorithm consists of a projected gradient descend direction algorithm and offers strong theoretical guarantees that have been successfully employed in the DoA estimation context [65, 66]. Particularly, when the grid is fine and the columns of $\tilde{\mathbf{A}}_{q,\lambda}$ are strongly coherent, we can guarantee that each $\tilde{\mathbf{m}}_q$ obtained from (50) is exactly 1-sparse. Thus, using the Coordinate Descent algorithm [67] to minimize (50), we obtain an analytic solution for each sub-problem and the sparsity of the desired minimizer $\tilde{\mathbf{m}}$ reduces the computational complexity. Each step involves the hard thresholding operator $\mathcal{H}_s(.)$, that keeps the s-largest components of a vector and sets the remaining entries equal to zero, thus, it automatically satisfies both constraints of sparsity and positivity. We obtain the update for the $[k]$-th iteration as

$$\tilde{\mathbf{m}}_q^{[k]} = \mathcal{H}_1\left(\tilde{\mathbf{m}}_q^{[k-1]} + \tau_q^{[k]} \sum_{\lambda\in\Lambda} \underline{\mathbf{V}}_\lambda^{q\mathsf{T}}\left(\hat{\underline{\mathbf{r}}}_\lambda^q - \underline{\mathbf{V}}_\lambda^q\tilde{\mathbf{m}}_q^{[k-1]} \right) \right), \tag{51}$$

20

with $\underline{\mathbf{V}}_\lambda^q$ the q-th column of $\underline{\mathbf{V}}_\lambda$, where we can allow a step size $\tau_q^{[k]}$ that depends on $\tilde{\mathbf{m}}_q^{[k-1]}$ and the $[k]$-th iteration, by the use of the Normalized Iterative Hard Thresholding [68]. Let us take benefit of the 1-sparsity of $\tilde{\mathbf{m}}_q$ (see, Subsection 2.2) in order to choose $\tau_q^{[k]}$: firstly, we define its associated residual as

$$\underline{\hat{\mathbf{r}}}_\lambda^q = \underline{\hat{\mathbf{r}}}_\lambda - \sum_{\substack{q'=1 \\ q' \neq q}}^{Q} \underline{\mathbf{V}}_\lambda^{q'} \tilde{\mathbf{m}}_{q'}. \tag{52}$$

Secondly, we consider the initialization $\tilde{\mathbf{m}}_q^{[k-1]} = \mathbf{0}$ and note that the dictionary of $\tilde{\mathbf{m}}_q$ is given by

$$\underline{\mathbf{V}}^q = \left[\underline{\mathbf{V}}_{\lambda_1}^{q\mathsf{T}}, \ldots, \underline{\mathbf{V}}_{\lambda_J}^{q\mathsf{T}} \right]^\mathsf{T}, \tag{53}$$

whose columns have the same norm. Consequently, by choosing

$$\tau_q^{[k]} = \frac{1}{\left\| \underline{\mathbf{v}}^q \right\|_2^2} = \frac{1}{\sum_{\lambda \in \Lambda} \left\| \underline{\mathbf{v}}_\lambda^q \right\|_2^2}, \tag{54}$$

where $\underline{\mathbf{v}}^q$ and $\underline{\mathbf{v}}_\lambda^q$ are, respectively, any column of $\underline{\mathbf{V}}^q$ and $\underline{\mathbf{V}}_\lambda^q$, we obtain directly the step size and then solution for $\tilde{\mathbf{m}}_q$.

In the network, the z-th agent, A_z, accesses only $\{h_\lambda(.)\}_{\lambda \in \Lambda_z}$ in (50). In order to estimate $\tilde{\mathbf{m}}_q$ and then deduce its DoA and $\mathbf{m}_{q,\lambda}$, each agent A_z can calculate the values $\sum_{\lambda \in \Lambda_z} \underline{\mathbf{V}}_\lambda^{q\mathsf{T}} \underline{\hat{\mathbf{r}}}_\lambda^q$ and send them to the fusion center, which performs the thresholding step in (51). Then, the fusion center transmits only the non-zero value of $\tilde{\mathbf{m}}_q$ and its corresponding direction $\mathbf{d}_{\lambda_0,q}$. Benefiting from the positivity of $\tilde{\mathbf{m}}_q$, we are able to adopt the procedure of [61], that solves a top-K problem. Thus, the Z agents can send only a fraction of the estimates to the fusion center further reduce the communication overhead. We remark, that for compactness of the presentation, this procedure is not included in *Algorithm 1.3*, since it only improves the communication efficiency.

Afterward, the estimation of $\left\{ \sigma_\lambda^{\mathsf{n}} \right\}_{\lambda \in \Lambda}$ is performed locally, without the need of transmitting the estimated values. Firstly, note that without considering non-calibration sources, i.e., $\mathbf{R}_\lambda^{\mathsf{U}} \approx \mathbf{0}$, the estimation of $\sigma_\lambda^{\mathsf{n}}$ is given by

$$\underline{\sigma}_\lambda^{\mathsf{n}} = \mathrm{vecdiag}\left(\hat{\mathbf{R}}_\lambda - \hat{\mathbf{G}}_\lambda \hat{\mathbf{R}}_\lambda^{\mathsf{K}} \hat{\mathbf{G}}_\lambda^{\mathsf{H}} \right), \tag{55}$$

since we assume independence of $\sigma_\lambda^{\mathsf{n}}$ across wavelength. Secondly, we remove the bias introduced by the non-calibration sources as follows: we calculate the power

$$\hat{\sigma}_\lambda^r = \frac{\mathbf{a}_\lambda^{\mathsf{H}}(\mathbf{d}_{\lambda,r}) \left(\hat{\mathbf{R}}_\lambda - \hat{\mathbf{G}}_\lambda \hat{\mathbf{R}}_\lambda^{\mathsf{K}} \hat{\mathbf{G}}_\lambda^{\mathsf{H}} \right) \mathbf{a}_\lambda(\mathbf{d}_{\lambda,r})}{\left\| \mathbf{a}_\lambda(\mathbf{d}_{\lambda,r}) \right\|_2^2} \tag{56}$$

21

input: $\left\{\hat{\mathbf{R}}_\lambda\right\}_{\lambda \in \Lambda_z}$
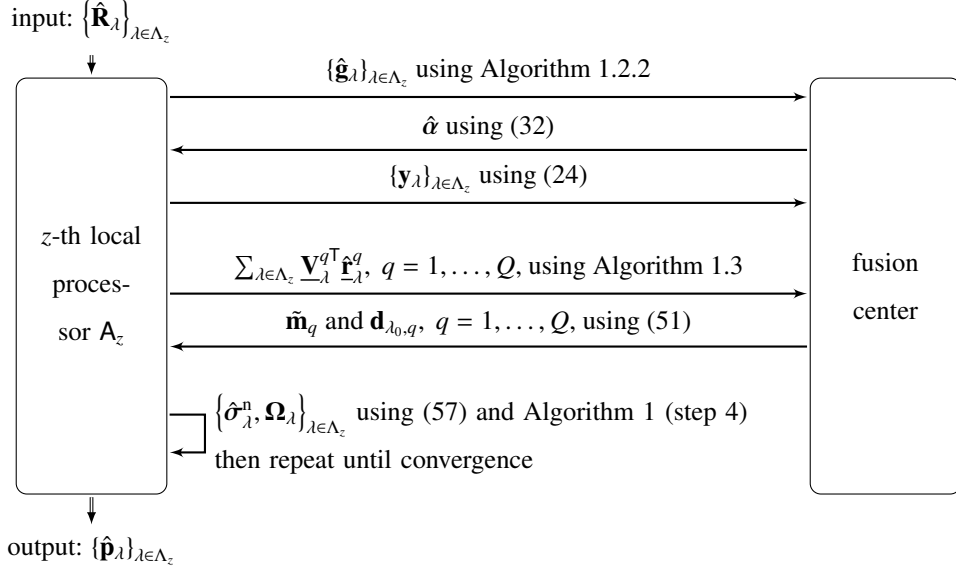

Figure 2: Operation flow and signaling exchange between a local processor A_z and the central processor. The first third arrows are performed repetitively during Algorithm 1.2 and the two middle arrows during Algorithm 1.3.

of the residual sample covariance matrix for a random direction $\mathbf{d}_{\lambda,r}$, where no source is supposed to be present. We then approximate $\mathbf{a}_\lambda^{\mathsf{H}}(\mathbf{d}_{\lambda,r})\mathbf{a}_\lambda(\mathbf{d}_{\lambda,q}) \approx 0$ for any $\mathbf{d}_{\lambda,r} \neq \mathbf{d}_{\lambda,q}$, which yields $\hat{\sigma}_\lambda^r$ as the sum of the sensor noise powers [7, 69]. By imposing $\sum_{p=1}^P \left[\sigma_\lambda^{\mathrm{n}}\right]_p = \hat{\sigma}_\lambda^r$ to the minimization of (17) w.r.t. $\sigma_\lambda^{\mathrm{n}}$, the new unbiased solution is given by

$$\hat{\sigma}_\lambda^{\mathrm{n}} = \underline{\sigma}_\lambda^{\mathrm{n}} + \frac{1}{P}\left(\hat{\sigma}_\lambda^r - \mathbf{1}_{P\times1}^{\mathsf{T}}\underline{\sigma}_\lambda^{\mathrm{n}}\right)\mathbf{1}_{P\times1}, \tag{57}$$

which concludes the description of *Algorithm 1.3*. Finally, operation flow and signaling exchange between a local processor A_z and the central processor is given in Fig 2.

4. Cramér-Rao Bound

The Cramér-Rao Bound (CRB) expresses a lower bound on the variance of the estimation error of a deterministic vector parameter for an unbiased estimator [70, 71]. In this section, after obtaining the CRB for the mono-wavelength scenario, we define the unconstrained CRB in the multi-wavelength scenario and finally take into account the dependence across wavelength (see,

Subsection 2.2) to obtain the constrained CRB that corresponds to the data model (4) and the constraints in Subsection 2.2.

Let us consider the mono-wavelength scenario and stack the unknown parameters in

$$\mathbf{p}_\lambda = [\mathfrak{R}(\mathbf{g}_\lambda^\mathsf{T}), \mathfrak{I}(\mathbf{g}_\lambda^\mathsf{T}), \mathbf{m}_\lambda^\mathsf{T}, \mathbf{m}_\lambda^{\mathsf{U}\mathsf{T}}, \mathbf{d}_{l,\lambda}^\mathsf{T}, \mathbf{d}_{l,\lambda}^{\mathsf{U}\mathsf{T}}, \mathbf{d}_{m,\lambda}^\mathsf{T}, \mathbf{d}_{m,\lambda}^{\mathsf{U}\mathsf{T}}, \sigma_\lambda^{\mathsf{n}\mathsf{T}}]^\mathsf{T}, \tag{58}$$

where $\mathbf{d}_{l,\lambda} = [d_{l,\lambda,1}, \ldots, d_{l,\lambda,Q}]^\mathsf{T}$, $\mathbf{d}_{l,\lambda}^\mathsf{U} = [d_{l,\lambda,1}^\mathsf{U}, \ldots, d_{l,\lambda,Q^\mathsf{U}}^\mathsf{U}]^\mathsf{T}$, $\mathbf{d}_{m,\lambda} = [d_{m,\lambda,1}, \ldots, d_{m,\lambda,Q}]^\mathsf{T}$, $\mathbf{d}_{m,\lambda}^\mathsf{U} = [d_{m,\lambda,1}^\mathsf{U}, \ldots, d_{m,\lambda,Q^\mathsf{U}}^\mathsf{U}]^\mathsf{T}$ and $\mathbf{\Gamma}_\lambda^\mathsf{U} \mathbf{\Gamma}_\lambda^{\mathsf{U}\mathsf{H}} = \mathrm{diag}\left(\mathbf{m}_\lambda^\mathsf{U}\right)$. We obtain the CRB, $\mathbf{C}_\lambda$, corresponding to the parameter vector in (58), after straightforward adaptations from [6, Chapter 4]. Then, for the multi-wavelength scenario, we gather the unknown parameters in a vector $\mathbf{p} = [\mathbf{p}_{\lambda_1}^\mathsf{T}, \ldots, \mathbf{p}_{\lambda_J}^\mathsf{T}]^\mathsf{T}$, suppose that the signals are i.i.d. across wavelength and ignore the constraints in Subsection 2.2 for the parameters. Consequently, we obtain the unconstrained CRB, $\tilde{\mathbf{C}}$, as

$$\tilde{\mathbf{C}} = \begin{bmatrix} \mathbf{C}_{\lambda_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{\lambda_J} \end{bmatrix}. \tag{59}$$

From $\tilde{\mathbf{C}}$, we obtain the constrained CRB [72] corresponding to the data model, $\mathbf{C}$, as

$$\mathbf{C} = \tilde{\mathbf{C}} - \tilde{\mathbf{C}}\mathbf{\Psi}^\mathsf{T}(\mathbf{\Psi}\tilde{\mathbf{C}}\mathbf{\Psi}^\mathsf{T})^{-1}\mathbf{\Psi}\tilde{\mathbf{C}}^\mathsf{T}, \tag{60}$$

where $\mathbf{\Psi}$ is the gradient matrix of the constraints in Subsection 2.2, given by

$$\mathbf{\Psi} = \left[\mathbf{\Psi}_{\mathbf{g}_\lambda}^\mathsf{T}, \mathbf{\Psi}_{\mathbf{m}_\lambda}^\mathsf{T}, \mathbf{\Psi}_{\mathbf{m}_\lambda^\mathsf{U}}^\mathsf{T}, \mathbf{\Psi}_{\mathbf{D}_\lambda}^\mathsf{T}, \mathbf{\Psi}_{\mathbf{D}_\lambda^\mathsf{U}}^\mathsf{T}\right]^\mathsf{T}, \tag{61}$$

in which the constraints on $\mathbf{g}_\lambda, \mathbf{m}_\lambda, \mathbf{m}_\lambda^\mathsf{U}, \mathbf{D}_\lambda$ and $\mathbf{D}_\lambda^\mathsf{U}, \lambda \in \Lambda$, are represented in $\mathbf{\Psi}_{\mathbf{g}_\lambda}, \mathbf{\Psi}_{\mathbf{m}_\lambda}, \mathbf{\Psi}_{\mathbf{m}_\lambda^\mathsf{U}}, \mathbf{\Psi}_{\mathbf{D}_\lambda}$ and $\mathbf{\Psi}_{\mathbf{D}_\lambda^\mathsf{U}}$, respectively. Since $\mathbf{m}_\lambda \propto \lambda^{-2}$, we have

$$\lambda_1^2\mathbf{m}_{\lambda_1} = \lambda_2^2\mathbf{m}_{\lambda_2} = \ldots = \lambda_J^2\mathbf{m}_{\lambda_J}, \tag{62}$$

leading to

$$\mathbf{\Psi}_{\mathbf{m}_\lambda} = \begin{bmatrix} \lambda_1^2\mathbf{I} & -\lambda_2^2\mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} \\ \lambda_1^2\mathbf{I} & \mathbf{0} & \ddots & & \mathbf{0} \\ \vdots & \vdots & & \ddots & \vdots \\ \lambda_1^2\mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} & -\lambda_J^2\mathbf{I} \end{bmatrix} \tag{63}$$

23

and we add zeros for the indices corresponding to the remaining parameters in $\mathbf{p}$. $\boldsymbol{\Psi}_{\mathbf{m}_\lambda^U}$ is obtained in the same way. On the other hand, in order to derive $\boldsymbol{\Psi}_{\mathbf{D}_\lambda}$, we make use of the following constraints

$$\lambda_1^{-2} d_{l,q,\lambda_1} = \lambda_2^{-2} d_{l,q,\lambda_2} = \ldots = \lambda_J^{-2} d_{l,q,\lambda_J}, \tag{64}$$

$$\lambda_1^{-2} d_{m,q,\lambda_1} = \lambda_2^{-2} d_{m,q,\lambda_2} = \ldots = \lambda_J^{-2} d_{m,q,\lambda_J}, \tag{65}$$

leading to

$$\boldsymbol{\Psi}_q^l = \begin{bmatrix} \lambda_1^{-2} & -\lambda_2^{-2} & 0 & \ldots & 0 \\ \lambda_1^{-2} & 0 & \ddots & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \lambda_1^{-2} & 0 & \ldots & 0 & -\lambda_J^{-2} \end{bmatrix} \tag{66}$$

and we add zeros for the indices corresponding to the remaining parameters in $\mathbf{p}$ and process in a same way for d_m to obtain $\boldsymbol{\Psi}_q^m$. Thus, $\boldsymbol{\Psi}_{\mathbf{D}_\lambda}$ is given by

$$\boldsymbol{\Psi}_{\mathbf{D}_\lambda} = \left[\boldsymbol{\Psi}_1^{lT}, \ldots, \boldsymbol{\Psi}_Q^{lT}, \boldsymbol{\Psi}_1^{mT}, \ldots, \boldsymbol{\Psi}_Q^{mT} \right]^T, \tag{67}$$

and we derive in the same way $\boldsymbol{\Psi}_{\mathbf{D}_\lambda^U}$. Finally, we consider the constraint (11) that reduces the degree of freedom of $\{\mathbf{g}_\lambda\}_{\lambda \in \Lambda}$ from JP to KP, i.e., we add $(J-K)P$ constraints. Let us define

$$\mathbf{g}_K = [\mathbf{g}_{\lambda_1}^T, \ldots, \mathbf{g}_{\lambda_K}^T]^T, \tag{68}$$

$$\mathbf{B}_K = \left[\mathbf{B}_{\lambda_1}^T, \ldots, \mathbf{B}_{\lambda_K}^T \right]^T. \tag{69}$$

Thus,

$$\mathbf{g}_\lambda = \mathbf{B}_\lambda \mathbf{B}_K^{-1} \mathbf{g}_K, \lambda = \lambda_{K+1}, \ldots, \lambda_J, \tag{70}$$

leading to

$$\boldsymbol{\Psi}_{\mathbf{g}_\lambda} = \begin{bmatrix} \mathbf{B}_{\lambda_{K+1}} \mathbf{B}_K^{-1} & -\mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{B}_{\lambda_{K+2}} \mathbf{B}_K^{-1} & \mathbf{0} & \ddots & & \mathbf{0} \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{B}_{\lambda_J} \mathbf{B}_K^{-1} & \mathbf{0} & \ldots & \mathbf{0} & -\mathbf{I} \end{bmatrix} \tag{71}$$

and we add zeros for the indexes corresponding to the remaining parameters in $\mathbf{p}$, which concludes our derivation of the constrained CRB.

24

5. Simulations

The proposed method is evaluated in realistic situations, with typical parameter values commonly used in radio astronomy applications [3, 6, 7]. In order to analyze the estimation of $\{\mathbf{g}_\lambda\}_{\lambda \in \Lambda}$, we first focus on *Algorithm 1.2* and *Algorithm 1.2.2* and then show results for the PMCA.
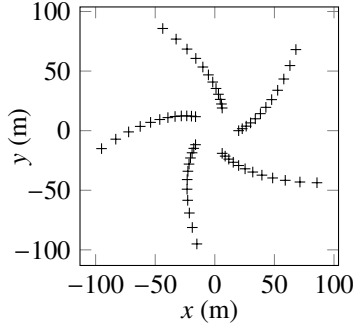

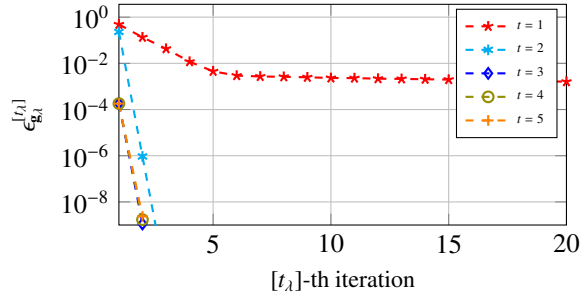Figure 3: LOFAR's Initial Test Station antenna locations [73].

Figure 4: $\mathbf{g}_\lambda$-residual, $\epsilon_{\mathbf{g}_\lambda}$, as function of the iteration number t_λ of *Algorithm 1.2.2*, for different values of the [t]-th iteration of *Algorithm 1.2*.

5.1. Data Setup

The antenna locations correspond to the LOFAR's Initial Test Station [73], with $P = 60$ antennas disposed in a five-armed spiral, as shown in Figure 3. We assume a sky model with $\lambda_0 = 10\,\text{m}$ ($f_0 = 30\,\text{MHz}$) consisting of $Q = 2$ strong calibration sources and $Q^\text{U} = 8$ weak non-calibration sources, provided from the ten strongest sources in the table of [14], such that they represent the weak contributed in the covariance matrices (the matrix $\mathbf{R}_\lambda^\text{U}$ in (7)). The total power

25
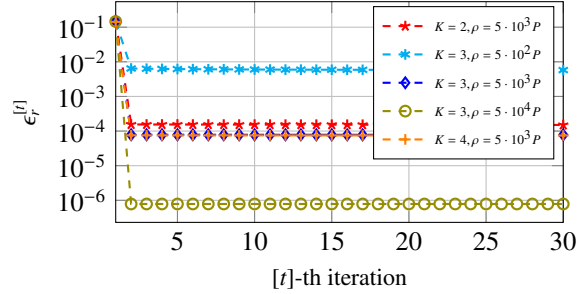
Figure 5: Primal residual, ϵ_r, as function of the iteration number t of *Algorithm 1.2*, for smoothing polynomial terms $K = 2, 3, 4$ and regularization term $\rho = 5 \cdot 10^2 P, 5 \cdot 10^3 P, 5 \cdot 10^4 P$.

of these sources is assumed to be 1% of the total antenna noise power, a typical scenario for radio interferometers [6]. We consider data taken at $J = 13$ wavelengths from 7.1 m to 16.5 m. We create $\{\mathbf{g}_\lambda\}_{\lambda \in \Lambda}$ by using a polynomial of order $K^\kappa = 3$, with $b_{\lambda,k} = ((\lambda_0 - \lambda)/\lambda)^{k-1}$, given as one realization sample from $\mathcal{CN}\left(\mathbf{1}, \left(\sigma_\mathbb{R}^\alpha + j\sigma_\mathbb{C}^\alpha\right)\mathbf{I}\right)$ with $\sigma_\mathbb{R}^\alpha = \sigma_\mathbb{C}^\alpha = 0.25$ and we consider $\mathbf{g}_\lambda^{[0]} = \mathbf{1}$ as an initialization. To initialize *Algorithm 1.2*, we consider the regularization parameter $\rho = 0$ during the first estimation of $\mathbf{g}_\lambda$, i.e., the first estimation of $\mathbf{g}_\lambda$ is done without enforcing smoothness. We generate the shifts for $\left(l_{q,\lambda_0}, m_{q,\lambda_0}\right), q = 1, \ldots, Q$ and the diagonal of $\boldsymbol{\Gamma}_{\lambda_0}$ with one realization sample from a uniform distribution centred at $\left(\left(l_q^\kappa, m_q^\kappa\right)\right)$ and $\mathbf{1}$ with a variance of $\sigma_{\lambda_0}^\mathbf{D} = 10^{-1}/\sqrt{3}$ and $\sigma_\mathbb{R}^\mathbf{\Gamma} = \sigma_\mathbb{C}^\mathbf{\Gamma} = 1/\sqrt{60}$, respectively. We initialize with $\boldsymbol{\Gamma}_\lambda = \mathbf{I}$ and data are produced via the signal model given in (3), in order to obtain the sample covariance matrices (14).

5.2. Results

5.2.1. Results for the estimation of $\{\mathbf{g}_\lambda\}_{\lambda \in \Lambda}$

We illustrate here the convergence behavior and the performance of both *Algorithm 1.2* and *Algorithm 1.2.2*. The number of observations is kept to $N = 2^{14}$ and all the stopping criteria are set to 0.03. In order to analyse the convergence, we define the $\mathbf{g}_\lambda$-residual, $\epsilon_{\mathbf{g}_\lambda}$, the primal
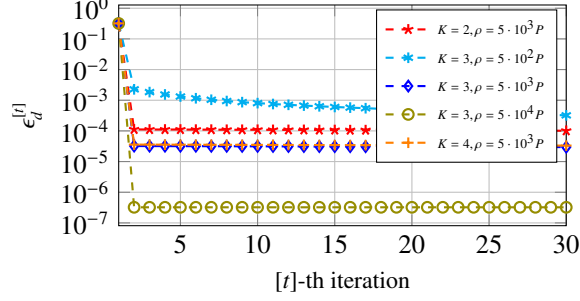
Figure 6: Dual residual, ϵ_d, as function of the iteration number t of *Algorithm 1.2*, for smoothing polynomial terms $K = 2, 3, 4$ and regularization term $\rho = 5 \cdot 10^2 P, 5 \cdot 10^3 P, 5 \cdot 10^4 P$.

residual, ϵ_r, and the dual residual, ϵ_d, as

$$\epsilon_{\mathbf{g}_\lambda}^{[t_\lambda]} = \frac{1}{J} \sum_{\lambda \in \Lambda} \frac{\left\| \mathbf{g}_\lambda^{[t_\lambda]} - \mathbf{g}_\lambda^{[t_\lambda - 1]} \right\|_2}{\left\| \mathbf{g}_\lambda^{[t_\lambda]} \right\|_2}, \tag{72}$$

$$\epsilon_r^{[t]} = \frac{1}{\sqrt{PJ}} \sum_{\lambda \in \Lambda} \left\| \mathbf{g}_\lambda^{[t]} - \mathbf{B}_\lambda \boldsymbol{\alpha}^{[t]} \right\|_2, \tag{73}$$

$$\epsilon_d^{[t]} = \frac{1}{\sqrt{PJ}} \left\| \boldsymbol{\alpha}^{[t]} - \boldsymbol{\alpha}^{[t-1]} \right\|_2. \tag{74}$$

The primal residual depicts the error between the local solution and the predicted consensus value. On the other hand, the dual residual depicts the convergence of the reference variable $\boldsymbol{\alpha}$.
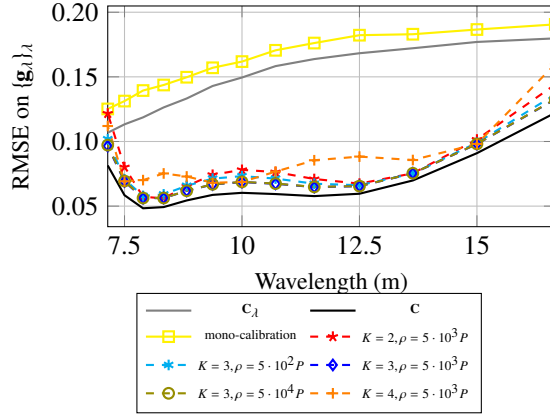
Figure 7: RMSE on $\{\mathbf{g}_\lambda\}_\lambda$ as function of wavelength and compared to the CBRs. The edge wavelengths have a higher error, particularly for $K = 2$ and $K = 4$, due to our choice of false interpolating polynomials compared to the real/true polynomial order $K^{\mathrm{K}} = 3$.

27

In Figure 4, we focus on the convergence of *Algorithm 1.2.2*. For $t > 1$, we observe that the convergence is almost immediate ($t_\lambda \approx 5$ iterations). In Figure 5 and Figure 6, we show, respectively, the primal and dual residuals, both as function of the [t]-th iteration number, when $N = 2^{14}$. We have set the regularization term $\rho = 5 \cdot 10^2 P, 5 \cdot 10^3 P, 5 \cdot 10^4 P$ and the smoothing polynomial order $K = 2, 3, 4$, with $K = 2$ underestimating the simulated polynomial order while $K = 4$ overestimating it. It is clear that as the value of ρ increases, the primal and dual residuals converge faster, for $t \approx 5$ iterations, for a sufficient value of ρ. Meanwhile, the primal and dual residuals differ slightly for different polynomial order K.

The statistical performance is then compared with mono-calibration scheme, the unconstrained-CRB $\mathbf{C}_\lambda$ and the multi-constrained-CRB, $\mathbf{C}$. In Figure 7, we plot the Root Mean Square Error (RMSE) for the estimates of $\mathbf{g}_\lambda$. Results are averaged for 500 Monte-Carlo runs, for each chosen value of K and ρ. We approach the multi-constrained-CRB for $K = K^\kappa$ and even with both $K = 2$ and $K = 4$, we significantly improve mono-calibration and be close to the CRB. Moreover, we also have errors due to polynomial interpolation, which is clearly seen at the edge wavelengths.

5.2.2. *Results for the PMCA*

We similarly analyze both convergence and performance of the proposed PMCA. During the DoA estimation, we choose initially a coarse grid, with the same resolution for each coordinate of each calibrator. We apply grid refinements [36] until we avoid off-grid mismatch.

Firstly, we concentrate on the convergence of *Algorithm 1* and *Algorithm 1.3*, respectively. For this purpose, we define the $\tilde{\mathbf{m}}$, $\sigma_\lambda^{\mathrm{n}}$ and $\mathbf{p}$-residuals, respectively, by

$$\epsilon_{\tilde{\mathbf{m}}}^{[k]} = \frac{1}{J} \sum_{\lambda \in \Lambda} \left(\frac{\left\| \mathbf{m}_\lambda^{[k]} - \mathbf{m}_\lambda^{[k-1]} \right\|_2}{\left\| \mathbf{m}_\lambda^{[k]} \right\|_2} + \frac{1}{Q} \sum_{q=1}^{Q} \left\| \mathbf{d}_{\lambda,q}^{[k]} - \mathbf{d}_{\lambda,q}^{[k-1]} \right\|_2 \right), \tag{75}$$

$$\epsilon_{\sigma_\lambda^{\mathrm{n}}}^{[i]} = \frac{1}{J} \sum_{\lambda \in \Lambda} \frac{\left\| \sigma_\lambda^{\mathrm{n}[i]} - \sigma_\lambda^{\mathrm{n}[i-1]} \right\|_2}{\left\| \sigma_\lambda^{\mathrm{n}[i]} \right\|_2}, \tag{76}$$

$$\epsilon_{\mathbf{p}}^{[i]} = \epsilon_{\mathbf{g}_\lambda}^{[i]} + \epsilon_{\tilde{\mathbf{m}}}^{[i]} + \epsilon_{\sigma_\lambda^{\mathrm{n}}}^{[i]}. \tag{77}$$

In Figure 8, the $\tilde{\mathbf{m}}$-residual for *Algorithm 1.3* decreases during the first iterations ($k \approx 5$) and stops due to alternating between close directions on the grid. In Figure 9, the previous residuals and $\mathbf{p}$-residual decline more slowly and we have to wait $i \approx 10$ iterations to assure a correct convergence.
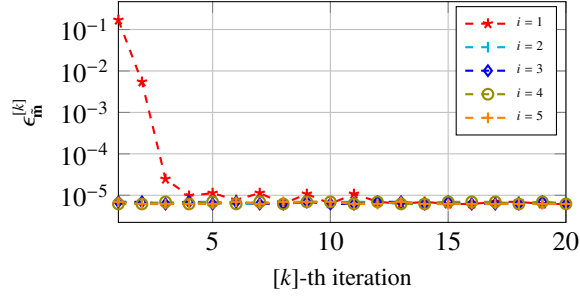
28

Figure 8: $\tilde{\mathbf{m}}$-residual, $\epsilon_{\tilde{\mathbf{m}}}$, as function of the iteration number k of *Algorithm 1.3*, for different values of the [i]-th iteration of *Algorithm 1*.

In order to investigate the statistical performance, we perform 200 Monte-Carlo runs for different sample sizes N, and setting $\rho = 5 \cdot 10^3 P$ and K to its true value. We plot the RMSE of different parameters in Figure 10 and Figure 11, as function of the number of samples N and compared to their corresponding multi-constrained-CRB. As expected, the method approaches the multi-constrained-CRB. This clearly show the good robustness of the method in low SNR scenario with a presence of non-calibration sources. This is mainly due to i) considering simultaneously multi-wavelength observations (thus, more observations) and ii) forcing smoothness constraints which can attenuate the effect of such unmodelled sources considered as outliers. However, it should be noted that this robustness has its own limit depending on the number and the power of the unmodelled sources.
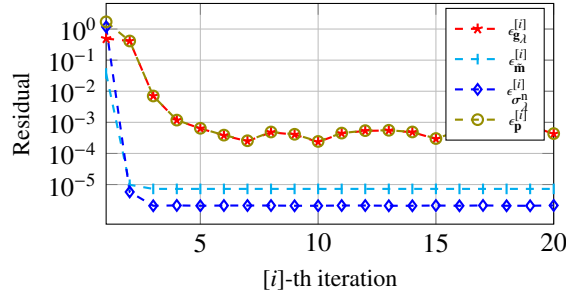
Figure 9: Variation of the residuals as function of the iteration number i of *Algorithm 1*.
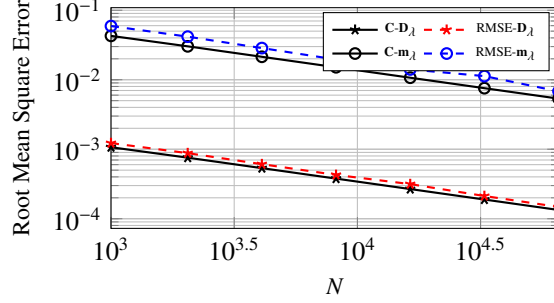
Figure 10: RMSE on the directions of the calibrators and their associated directional gains as function of number of samples N, and compared to their corresponding multi-constrained-CRB.
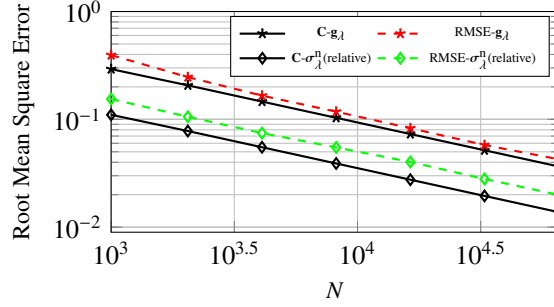

Figure 11: RMSE on the undirectional gains and antenna noise powers as function of number of samples N, and compared to their corresponding multi-constrained-CRB.

5.3. Application to LOFAR data

The LOFAR is a phased array radio telescope consisting of 50 subarrays (referred to as *stations*) of which 24 are located in a relatively small centeal area near Exloo in The Netherlands (the *core stations*), 14 are spread throughout The Netherlands (the *remote stations*) and 12 are located in several European countries (the *international stations*) [1]. For this test, we use data from a single polarization measurement with a 48-element low band antenna (LBA) array of a single core station, which have been computed between 21:01:29 UTC and 21:10:00 UTC on July 30, 2011, a night with mild ionospheric disturbances. The data consist of a set of array covariance matrices, each obtained for a different 195 kHz wide frequency channel after 1 s of integration. Figure 12 and Figure 14 show, respectively, the locations of the 48 antennas and the uncalibrated image at 49.4 Mhz, i.e., an image plotted by assuming unit sensor gain. This image clearly reveals two bright point sources, Cass A and Cyg A, but we observe also a large

30

amount of diffuse emission from the Milky Way. These covariance matrices constitute in this way a relevant test for the robustness w.r.t. unmodelled sources of our algorithm, since only the two main sources are considered as calibrators ($Q = 2$).
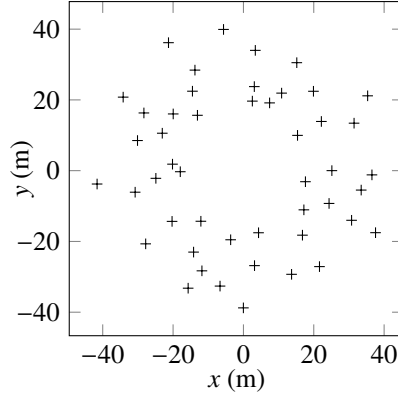

Figure 12: LOFAR's core station antenna locations.

We select data for 31 frequencies in the range 47.6-53.5 MHz, with central frequency (respectively, wavelength) $f_0 = 50.5$ MHz ($\lambda_0 = 5.9$ m) and choose a polynomial order $K = 3$. It is commonly assumed that the resolution of a radio telescope is in the order of λ/D, with D the maximum baseline [74]. This led us to define our two displacement sectors as 10×10 grid with a spacing of $0.2\lambda/D$ (much larger than possible angular-shifts), centered at the true directions obtained from tables. We plot results of the images produced from the calibrated (using the PMCA algorithm) covariance matrix for the frequency channels centered at 49.4 MHz in Figure 15, compared with the state-of-the art [5], which considers the apparent direction of calibrators as fixed w.r.t. λ.

The two calibrated images are very similar and show improved contrast between the two point sources and the noise in the background of the image compared with the uncalibrated image shown in Figure 14. This indicates that calibration has improved the images in both cases. A closer look at the two images in Figure 15 reveals that the difffuse emission on the South of the sky, which is a bright section of the Galactic plane (which can be seen by eye in the night sky as the Milky Way), appears slightly brighter in the image calibrated using PMCA. This indicates that PMCA is more robust to the presence of unmodelled sources in the image thereby reducing the likelihood that power in the unmodelled sources is fitted to the calibrator sources by the

calibration algorithm, which causes a bias in the calibration solutions.

Another indication that PMCA provides robustness to unmodelled flux is that the DI gain solutions, as shown in Figure 13 for a few representative examples, are similar to those found with the conventional approach despite the fact that the diffuse emission was not spatially filtered as is done in the conventional approach. This comparison also shows that the gain solutions produced by the conventional approach vary quite significantly from one frequency channel to the next while the PMCA provides smoothed solutions, which are physically more realistic.
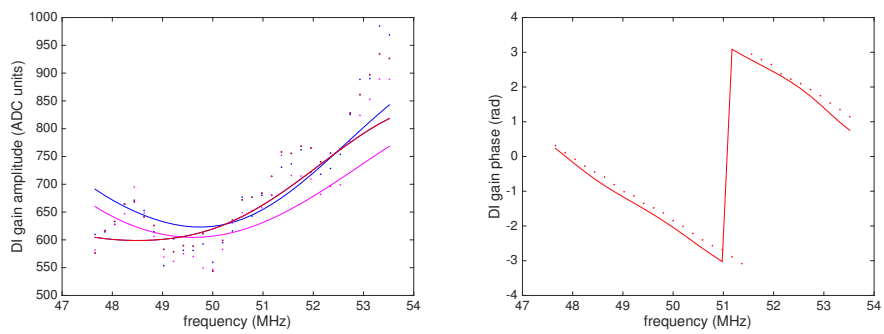
Figure 13: Comparison of DI gain amplitude estimates (left) and DI gain phase estimates (right) for a few representative examples between the solutions found using PMCA (solid lines) and the conventional per-channel calibration approach (dots).
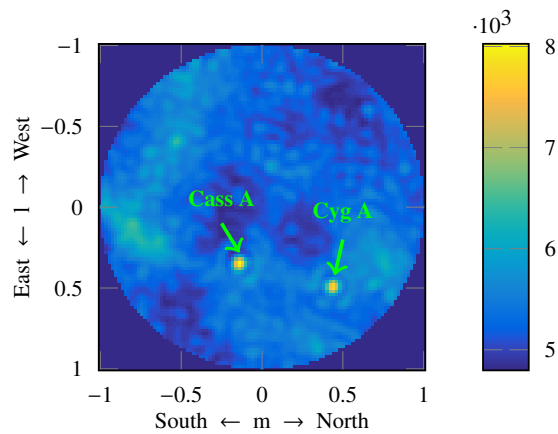

Figure 14: image of the uncalibrated covariance matrix at 49.4 MHz. The presence of diffuse emission is noticeable.
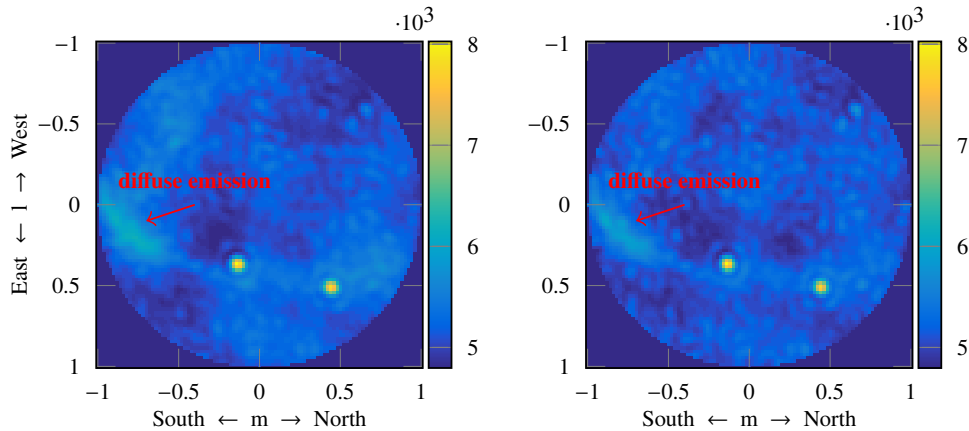
32

Figure 15: Image results of the calibrated covariance matrix at 49.4 MHz from the proposed PMCA (left), the calibrated solution of the existing method used for LOFAR [5] (right). The diffuse emission appears slightly brighter in the image calibrated using PMCA.

6. Conclusion

In this paper, we proposed a novel iterative scheme for parallel calibration of next generation radio astronomical arrays, for which direction dependent effects affect the apparent directions of the calibration sources and parameter values vary across wavelength. The proposed algorithm, named Parallel Multi-wavelength Calibration Algorithm (PMCA), iteratively estimates the complex direction independent antenna gains and their noise powers, whereas, it jointly estimates the directions of the calibrators and their associated direction dependent gain. These two main steps are, respectively, based on Alternating Direction of Multiple Multipliers and Distributed Iterative Hard Thresholding procedures. This leads to a statistically efficient and robust scheme to unmodelled sources as shown by numerical simulations and by application to actual LOFAR data.

PMCA offers a framework that is easily adaptable to variations on the scenarios presented in this paper, while maximizing the use of a priori available knowledge of the physics of the instrument and the measurement process. The parallelism in the PMCA ensures that the algorithm is scalable to large data sets, which can be paralleled over a number of nodes. The algorithm is set up in such a way that the use of data remains local, i.e., only the node on which a specific part of the data resides needs access to that data.

7. References

[1] M. P. van Haarlem, et al., LOFAR: The LOw-Frequency ARray, Astronomy & Astrophysics 556.

[2] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, T. J. L. W. Lazio, The Square Kilometre Array, Proceedings of the IEEE 97 (8) (2009) 1482–1496.

[3] S. J. Wijnholds, S. van der Tol, R. Nijboer, A.-J. van der Veen, Calibration challenges for future radio telescopes, IEEE Signal Processing Magazine 27 (2010) 30–42.

[4] S. J. Wijnholds, A. J. van der Veen, F. D. Stefani, E. L. Rosa, A. Farina, Signal processing challenges for radio astronomical arrays, in: International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 5382–5386.

[5] S. J. Wijnholds, A.-J. van der Veen, Multisource self-calibration for sensor arrays, IEEE Transactions on Signal Processing 57 (2009) 3512–3522.

[6] S. J. Wijnholds, Fish-eye observing with phased array radio telescopes, Ph.D. thesis, Delft University of Technology (2010).

[7] A.-J. van der Veen, S. J. Wijnholds, Signal processing tools for radio astronomy., in: Handbook of Signal Processing Systems, Springer, 2013, pp. 421–463.

[8] S. van der Tol, B. D. Jeffs, A. J. van der Veen, Self-calibration for the LOFAR radio astronomical array, IEEE Transactions on Signal Processing 55 (9) (2007) 4497–4510.

[9] A. R. Thompson, J. M. Moran, G. W. Swenson, Interferometry and Synthesis in Radio Astronomy; 2nd ed., Wiley-VCH, 2001.

[10] S. van der Tol, Bayesian estimation for ionospheric calibration in radio astronomy, Ph.D. thesis, Delft University of Technology (2009).

[11] C. Lonsdale, Calibration approaches, 2004 (2004).

[12] W. D. Cotton, J. J. Condon, R. A. Perley, N. Kassim, J. Lazio, A. Cohen, W. Lane, W. C. Erickson, Beyond the isoplanatic patch in the VLA Low-frequency Sky Survey, in: Ground-based Telescopes, Vol. 5489 of Proceedings of SPIE, 2004, pp. 180–189.

[13] A. S. Cohen, H. J. A. Röttgering, Probing Fine-Scale Ionospheric Structure with the Very Large Array Radio Telescope, Astronomical Journal 138 (2009) 439–447.

[14] A. S. Bennett, The revised 3C catalog of radio sources, Memoirs of the Royal Astronomical Society 68 (1962) 163–172.

[15] J. W. M. Baars, R. Genzel, I. I. K. Pauliny-Toth, A. Witzel, The absolute spectrum of CAS A - an accurate flux density scale and a set of secondary calibrators, Astronomy & Astrophysics 61 (1977) 99–106.

[16] A. E. Kimball, Ž. Ivezić, A unified catalog of radio objects detected by NVSS, FIRST, WENSS, GB6, and SDSS, The Astronomical Journal 136 (2) (2008) 684.

[17] J. J. Bryant, H. M. Johnston, J. W. Broderick, R. W. Hunstead, C. De Breuck, B. M. Gaensler, A new search for distant radio galaxies in the southern hemisphere – III. optical spectroscopy and analysis of the mrcr–sumss sample, Monthly Notices of the Royal Astronomical Society 395 (2) (2009) 1099–1120.

[18] S. Salvini, S. J. Wijnholds, Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications, Astronomy & Astrophysics 571 (2014) A97.

[19] A. M. Sardarabadi, A. J. van der Veen, Application of Krylov based methods in calibration for radio astronomy, in: Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014, pp. 153–156.

[20] S. Kazemi, P. Hurley, O. Öçal, G. Cherubini, Blind calibration for radio interferometry using convex optimization, in: Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa), 2015, pp. 164–168.

[21] V. Ollier, M. N. E. Korso, R. Boyer, P. Larzabal, M. Pesavento, Robust calibration of radio interferometers in non-Gaussian environment, IEEE Transactions on Signal Processing 65 (21) (2017) 5649–5660.

[22] M. Brossard, M. El Korso, M. Pesavento, R. Boyer, P. Larzabal, Calibration of radio interferometers using a sparse DOA estimation framework, in: 24th European Signal Processing Conference (EuSiPCo), Budapest, Hungary, 2016, pp. 265–269.

[23] S. J. Tingay, The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies, Publications of the Astronomical Society of Australia 30 (7).

[24] D. A. Mitchel, et al., Real-Time Calibration of the Murchison Widefield Array, IEEE Journal of Selected Topics in Signal Processing 2 (5) (2008) 707–717.

[25] S. Yatawatta, Distributed radio interferometric calibration, Monthly Notices of the Royal Astronomical Society 449 (2015) 4506–4514.

[26] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundation and Trends in Machine Learning 3 (1) (2011) 1–122.

[27] C. Song, S. Yoon, V. Pavlovic, Fast ADMM algorithm for distributed optimization with adaptive penalty, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16, AAAI Press, 2016, pp. 753–759.

[28] E. Wei, A. Ozdaglar, Distributed alternating direction method of multipliers, in: Conference on Decision and Control (CDC), 2012, pp. 5445–5450.

[29] F. Iutzeler, P. Bianchi, P. Ciblat, W. Hachem, Linear convergence rate for distributed optimization with the Alternating Direction Method of Multipliers, in: Conference on Decision and Control (CDC), 2014, pp. 5046–5051.

[30] T. Goldstein, B. O'Donoghue, S. Setzer, R. Baraniuk, Fast alternating direction optimization methods, SIAM Journal on Imaging Sciences 7 (3) (2014) 1588–1623.

[31] T. Chang, M. Hong, X. Wang, Multi-agent distributed optimization via inexact consensus ADMM, IEEE Transactions on Signal Processing 63 (2) (2015) 482–497.

[32] T. Erseghe, A distributed and scalable processing method based upon ADMM, IEEE Signal Processing Letters 19 (9) (2012) 563–566.

[33] W. Shi, Q. Ling, G. Wu, W. Yin, A proximal gradient algorithm for decentralized composite optimization, IEEE Transactions on Signal Processing 63 (22) (2015) 6013–6023.

[34] J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, M. Puschel, D-ADMM: A communication-efficient distributed algorithm for separable optimization, IEEE Transactions on Signal Processing 61 (10) (2013) 2718–2723.

[35] R. O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Transactions on Antennas and Propagation 34 (3) (1986) 276–280.

[36] D. Malioutov, M. Cetin, A. S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Transactions on Signal Processing 53 (8) (2005) 3010–3022.

[37] X. Wei, Y. Yuan, Q. Ling, DOA estimation using a greedy block coordinate descent algorithm, IEEE Transactions

on Signal Processing 60 (12) (2012) 6382–6394.

[38] E. T. Northardt, I. Bilik, Y. I. Abramovich, Spatial compressive sensing for Direction-of-Arrival estimation with bias mitigation via expected likelihood, IEEE Transactions on Signal Processing 61 (5) (2013) 1183–1195.

[39] C. Steffens, P. Parvazi, M. Pesavento, Direction finding and array calibration based on sparse reconstruction in partly calibrated arrays, in: IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014, pp. 21–24.

[40] C. Bilen, G. Puy, R. Gribonval, L. Daudet, Convex optimization approaches for blind sensor calibration using sparsity, IEEE Transactions on Signal Processing 62 (18) (2014) 4847–4856.

[41] S. J. Wijnholds, S. Chiarucci, Blind calibration of phased arrays using sparsity constraints on the signal model, in: 24th European Signal Processing Conference (EuSiPCo), Budapest, Hungary, 2016, pp. 271–274.

[42] F. Li, T. J. Cornwell, F. de Hoog, The application of compressive sampling to radio astronomy i: Deconvolution, arXiv preprint arXiv:1106.1711.

[43] H. Garsden, J. Girard, J.-L. Starck, S. Corbel, C. Tasse, A. Woiselle, J. McKean, A. S. Van Amesfoort, J. Anderson, I. Avruch, et al., LOFAR sparse image reconstruction, Astronomy & astrophysics 575 (2015) A90.

[44] J. D. McEwen, Y. Wiaux, Compressed sensing for wide-field radio interferometric imaging, Monthly Notices of the Royal Astronomical Society 413 (2) (2011) 1318–1332.

[45] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, P. Vandergheynst, Compressed sensing imaging techniques for radio interferometry, Monthly Notices of the Royal Astronomical Society 395 (3) (2009) 1733–1742.

[46] J. D. Bregman, System Design and Wide-field Imaging Aspects of Synthesis Arrays with Phased Array Stations: To the Next Generation of SKA System Designers, University Library Groningen, 2012.

[47] D. Astely, A. L. Swindlehurst, B. Ottersten, Spatial signature estimation for uniform linear arrays with unknown receiver gains and phases, IEEE Transactions on Signal Processing 47 (8) (1999) 2128–2138.

[48] A. J. Boonstra, A. J. van der Veen, Gain calibration methods for radio telescope arrays, IEEE Transactions on Signal Processing 51 (1) (2003) 25–38.

[49] C. Tasse, Nonlinear Kalman filters for calibration in radio interferometry, Astronomy & Astrophysics 566 (2014) A127.

[50] S. van der Tol, A.-J. van der Veen, Ionospheric calibration for the LOFAR radio telescope, in: International Symposium on Signals, Circuits and Systems (ISSCS), 2007, pp. 1–4.

[51] G. B. Taylor, C. L. Carilli, R. A. Perley, Synthesis Imaging in Radio Astronomy II, Astronomical Society of the Pacific, 1998.

[52] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, P. Vandergheynst, Compressed sensing imaging techniques for radio interferometry, Monthly Notices of the Royal Astronomical Society 395 (2009) 1733–1742.

[53] R. E. Carrillo, J. D. McEwen, Y. Wiaux, PURIFY: a new approach to radio-interferometric imaging, Monthly Notices of the Royal Astronomical Society.

[54] A. Ferrari, D. Mary, R. Flamary, C. Richard, Distributed image reconstruction for very large arrays in radio astronomy, in: IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014, pp. 389–392.

[55] H. Garsden, J. N. Girard, J. L. Starck, et. al., LOFAR sparse image reconstruction, Astronomy & Astrophysics 575 (2015) A90.

[56] T. L. Grobler, C. D. Nunhokee, O. M. Smirnov, A. J. van Zyl, A. G. de Bruyn, Calibration artefacts in radio

interferometry - I. Ghost sources in Westerbork Synthesis Radio Telescope data, Monthly Notices of the Royal Astronomical Society 439 (4) (2014) 4030–4047.

[57] S. J. Wijnholds, T. L. Grobler, O. M. Smirnov, Calibration artefacts in radio interferometry - II. Ghost patterns for irregular arrays, Monthly Notices of the Royal Astronomical Society 457 (3) (2016) 2331–2354.

[58] B. Ottersten, P. Stoica, R. Roy, Covariance matching estimation techniques for array signal processing applications, Digital Signal Processing 8 (3) (1998) 185–210.

[59] L. Li, X. Wang, G. Wang, Alternating direction method of multipliers for separable convex optimization of real functions in complex variables, Mathematical Problems in Engineering 2015.

[60] E. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math. 59 (8) (2006) 1207–1223.

[61] S. Patterson, Y. C. Eldar, I. Keidar, Distributed sparse signal recovery for sensor networks, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013, pp. 4494–4498.

[62] S. Patterson, Y. C. Eldar, I. Keidar, Distributed compressed sensing for static and time-varying networks, IEEE Transactions on Signal Processing 62 (19) (2014) 4931–4946.

[63] P. Han, R. Niu, Y. C. Eldar, Modified distributed iterative hard thresholding, in: International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 3766–3770.

[64] T. Blumensath, M. E. Davies, Iterative Hard Thresholding for compressed sensing, Applied and Computational Harmonic Analysis 27 (3) (2009) 265–274.

[65] E. Ollila, H. Kim, V. Koivunen, Robust iterative hard thresholding for compressed sensing, in: International Symposium on Communications, Control and Signal Processing (ISCCSP), 2014, pp. 226–229.

[66] E. Ollila, Multichannel sparse recovery of complex-valued signals using Huber's criterion, in: 2015 3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa), 2015, pp. 26–30.

[67] J. Friedman, T. Hastie, H. Höfling, R. Tibshirani, Pathwise coordinate optimization, Annals of Applied Statistics 1 (2) (2007) 302–332.

[68] T. Blumensath, M. E. Davies, Normalized Iterative Hard Thresholding: Guaranteed stability and performance, Selected Topics in Signal Processing, IEEE Journal of 4 (2) (2010) 298–309.

[69] A. Leshem, A. J. van der Veen, Radio-astronomical imaging in the presence of strong radio interference, IEEE Transactions on Information Theory 46 (5) (2000) 1730–1747.

[70] S. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice-Hall, Inc., 1993.

[71] P. Stoica, R. Moses, Spectral Analysis of Signals, Pearson Prentice Hall, 2005.

[72] T. L. Marzetta, A simple derivation of the constrained multiple parameter cramer-rao bound, IEEE Transactions on Signal Processing 41 (6) (1993) 2247–2249.

[73] S. J. Wijnholds, J. D. Bregman, A.-J. Boonstra, Sky noise limited snapshot imaging in the presence of RFI with LOFAR's Initial Test Station, Experimental Astronomy 17 (1) (2004) 35–42.

[74] J. D. Kraus, Radio astronomy, Cygnus-Quasar Books, 1986.