Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **00,** 1 (2018)
Advance Access publication 2018 June 8

doi:10.1093/mnras/sty1501

# PRIFIRA: General regularization using prior-conditioning for fast radio interferometric imaging⋆

Shahrzad Naghibzadeh† and Alle-Jan van der Veen

*Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, NL-2628 CD Delft, the Netherlands*

Accepted 2018 June 4. Received 2018 June 4; in original form 2017 September 14

**ABSTRACT**

Image formation in radio astronomy is a large-scale inverse problem that is inherently ill-posed. We present a general algorithmic framework based on a Bayesian-inspired regularized maximum likelihood formulation of the radio astronomical imaging problem with a focus on diffuse emission recovery from limited noisy correlation data. The algorithm is dubbed PRIor-conditioned Fast Iterative Radio Astronomy and is based on a direct embodiment of the regularization operator into the system by right preconditioning. The resulting system is then solved using an iterative method based on projections onto Krylov subspaces. We motivate the use of a beam-formed image (which includes the classical 'dirty image') as an efficient prior-conditioner. Iterative reweighting schemes generalize the algorithmic framework and can account for different regularization operators that encourage sparsity of the solution. The performance of the proposed method is evaluated based on simulated 1D and 2D array arrangements as well as actual data from the core stations of the Low Frequency Array radio telescope antenna configuration, and compared to state-of-the-art imaging techniques. We show the generality of the proposed method in terms of regularization schemes while maintaining a competitive reconstruction quality with the current reconstruction techniques. Furthermore, we show that exploiting Krylov subspace methods together with the proper noise-based stopping criteria results in a great improvement in imaging efficiency.

**Key words:** methods: numerical – methods: statistical – techniques: image processing – techniques: interferometric.

## 1 INTRODUCTION

### 1.1 The image formation problem

The advent and development of increasingly large radio interferometers such as the Low Frequency Array (LOFAR; Van Haarlem et al. 2013) and the Square Kilometer Array (Dewdney et al. 2009) has sparked renewed interest in the image formation task. The increased resolution, bandwidth, sensitivity, and sky coverage of these instruments result in many more sources, including unresolved sources and extended structures, rendering the traditional imaging algorithms based on point source detection and CLEAN iterations less effective. At the same time, image formation is expected to be the main computational bottleneck in the processing pipeline of next generation radio telescopes (Jongerius et al. 2014).

Image formation is based on solving the measurement equation, which generically has the form (Leshem & Van der Veen 2000; Wijnholds & van der Veen 2008)

$$\boldsymbol{r} = \mathbf{M}\boldsymbol{\sigma} + \boldsymbol{e},$$

where $\boldsymbol{r}$ is a vector of the measurements (antenna pair correlations or visibilities), $\mathbf{M}$ is the system matrix that models the antenna sampling pattern, $\boldsymbol{\sigma}$ is a stack of the pixels of the source brightness image, and $\boldsymbol{e}$ is an additive noise term. For high-resolution images, computing $\boldsymbol{\sigma}$ by least-squares (LS) minimization of $\|\boldsymbol{r} - \mathbf{M}\boldsymbol{\sigma}\|^2$ leads to an ill-posed problem. The solution $\hat{\boldsymbol{\sigma}} = \mathbf{M}^\dagger \boldsymbol{r}$ (where $\mathbf{M}^\dagger$ is a left inverse of $\mathbf{M}$) includes a magnified noise term $\mathbf{M}^\dagger \boldsymbol{e}$ or, if $\mathbf{M}$ is a 'wide' matrix, a unique and physically meaningful solution may not even exist.

Regularization is needed, in the form of prior knowledge, structure, or other constraints on the solution $\boldsymbol{\sigma}$. Classically, one of the options is to add a term to the cost function, e.g. $\|\boldsymbol{\sigma}\|_2^2$ (an $\ell_2$ constraint or Tikhonov regularization), $\|\boldsymbol{\sigma}\|_1$ or $\|\boldsymbol{\sigma}\|_0$ (an $\ell_1$ or $\ell_0$ constraint), a total variation constraint or a maximum entropy constraint. These options induce smoothness or sparsity of the solution. Another structural constraint is the requirement of the image pixels

⋆ Earlier versions of this paper were presented at ICASSP'17 [Naghibzadeh & van der Veen (2017b)] and CAMPSAP'17 [Naghibzadeh & van der Veen (2017a)].

† E-mail: s.naghibzadeh@tudelft.nl

to be positive. Iterative solution methods such as conjugate gradient are usually needed to compute the solution, and another form of regularization is to prematurely stop the iterations, restricting the solution to a particular data-dependent subspace. More formally, many of these regularization techniques can also be formulated in a Bayesian framework, where $\boldsymbol{\sigma}$ is modelled as a random variable, and prior knowledge on $\boldsymbol{\sigma}$ is given in the form of a prior statistical distribution $p(\boldsymbol{\sigma})$, often containing unknown parameters (e.g. scale) which can be modelled statistically as well as using hyperpriors (Tipping 2001; Wipf & Rao 2004).

It is clear that, in this generic form, the problem has been widely studied in many areas of mathematics, engineering, signal processing, and computer science (e.g. machine learning; Kaipio & Somersalo 2004). In recent years, some of these techniques are now gradually being introduced in the context of radio astronomy. Considerations in algorithm selection are (i) the accuracy (fidelity) of the resulting image, related to the definition of the optimization problem, (ii) computational complexity, related to the scalable solution of the optimization problem, and (iii) the automation and flexibility of the process regarding the selection of unknown parameters or settings such as iteration counts. For future radio telescopes, not all measurement data can be stored and image formation has to be done in an automated and quasi-real time process.

### 1.2 State-of-the-art imaging algorithms

Classical Radio Astronomical (RA) imaging algorithms are based on the CLEAN algorithm (Högbom 1974; Schwab 1984) and its multiresolution and multiscale variants (Wakker & Schwarz 1988; Cornwell 2008; Rau & Cotton 2011; Offringa & Smirnov 2017). The considered cost function is the LS objective, implicitly regularized by an $\ell_0$ constraint (Marsh & Richardson 1987) which favours maximal sparsity of the solution. The CLEAN algorithm was recently interpreted as a gradient descent method combined with a 'greedy' procedure to find the support of the image (Onose et al. 2016).

Alternatively, the problem can be regularized by posing a non-negativity constraint on the solution (Briggs 1995). The resulting Non-negative LS (NNLS) optimization can be implemented using the active set method (Sardarabadi, Leshem & van der Veen 2016) and similarly consists of two levels of iterations: (i) an outer loop to iteratively find the sparse support of the image and (ii) an inner loop in which a dimension reduced version of the LS problem is solved.

Another classical RA imaging algorithm is the maximum entropy method (MEM; Cornwell & Evans 1985). The regularization term is the entropy function $\boldsymbol{\sigma}^T \log(\boldsymbol{\sigma})$, and the problem is solved using computationally expensive non-linear optimization methods such as Newton–Raphson.

Finding the non-zero support of an image using an $\ell_0$ constraint is an NP-complete problem. Instead, this constraint may be weakened to an $\ell_1$ constraint, which still promotes sparsity of the solution, but admits a solution based on the theory of compressed sensing and convex optimization, for which efficient techniques exist. Recently, many algorithms in this direction have been proposed (Wiaux et al. 2009; McEwen & Wiaux 2011; Carrillo, McEwen & Wiaux 2012, 2014; Dabbech et al. 2014; Girard et al. 2015; Onose et al. 2016). These methods are based on a gradient descent approach. Instead of a constraint on $\|\boldsymbol{\sigma}\|_1$ (sparse image), also a more general constraint $\|\boldsymbol{\Psi}^T\boldsymbol{\sigma}\|_1$ or $\|\boldsymbol{\alpha}\|_1$ where $\boldsymbol{\sigma} = \boldsymbol{\Psi}\boldsymbol{\alpha}$ can be used, in which $\boldsymbol{\Psi}$ is an overcomplete dictionary of orthonormal bases. For example Sparsity Averaging Reweighted Analysis (SARA; Carrillo et al. 2012) employs a concatenation of wavelet dictionaries. The advantage of the methods based on convex optimization is the simplicity of imposing additional constraints on the solution, the existence of many well-developed methods with guaranteed convergence and the ability to split the work into simpler, parallelizable subproblems (Combettes & Pesquet 2011; Onose et al. 2016). The disadvantage of these algorithms is that the gradient descent steps make the algorithm convergence rather slow. Also, as remarked in Offringa & Smirnov (2017), many of these algorithms have not yet been tested on real data.

Taking another direction, the RESOLVE algorithm introduced techniques from Bayesian statistics to propose priors that regularize the solution (Junklewitz et al. 2016), aimed specifically at extended sources, and modelled these a priori using lognormal distributions. Unfortunately, the resulting method appears to be extremely slow (Junklewitz et al. 2016).

### 1.3 Results

In this paper, our interest is in developing a new method for science cases where a considerable amount of complex diffuse emissions are present such as in the studies of galactic magnetism, the epoch of reionization, and polarized imaging.

We start from a Bayesian statistical approach for regularization, but formulate a shortcut that immediately connects to a numerical method called prior-conditioning, i.e. a data-dependent Jacobi-like right preconditioner that scales the columns of $\mathbf{M}$. In this general framework, the prior conditioner can take the form of a beamformed image, such as the classical dirty image, or the Minimum Variance Distorsionless Response (MVDR) image, or any other low resolution prior image that is strictly positive on the true support of the image. This could also be determined iteratively, which gives a connection to reweighted LS solutions, often used to approximate $\ell_0$ or $\ell_1$ norm optimization by LS optimization, in particular the Focal Underdetermined System Solver (FOCUSS) algorithm (Gorodnitsky & Rao 1997) and the algorithm presented in Daubechies et al. (2010).

Next, we propose to solve the obtained regularized LS problem by a fast and efficient iterative algorithm based on the Krylov subspace-based method of LSQR (Paige & Saunders 1982). Krylov methods often exhibit a faster convergence than methods based on gradient descent (Saad 1981). Therefore, they appear to be good candidates as alternative iterative solution methods for the RA imaging problem. The stopping criterion of the LSQR algorithm is based on the norm of the residual, which provides another form of regularization, called iterative regularization or semiconvergence (Hansen 2010; Berisha & Nagy 2013). The resulting algorithm is straightforward to implement and computationally very efficient.

We compare the proposed method to classical RA imaging methods as well as methods based on convex optimization both in terms of speed and quality of the estimate. It will be seen that the proposed method is accurate and converges extremely fast (around 10 iterations).

The paper is organized as follows. In Section 2, we introduce the signal processing data model for RA imaging. In Section 3, we discuss RA imaging problem as a source power estimation problem and consider different problem formulations. In Section 4 we introduce our imaging problem formulation and generalize it based on different priors. In Section 5 we introduce the PRIor-conditioned Fast Iterative Radio Astronomy (PRIFIRA) algorithm. We compare PRIFIRA with the state-of-the art RA imaging algorithms in Section 6.

### 1.4 Notation

Matrices and vectors are denoted by boldface letters. A boldface italic letter such as $\boldsymbol{a}$ denotes a column vector, a boldface capital letter such as $\mathbf{A}$ denotes a matrix. For a matrix $\mathbf{A}$, $\boldsymbol{a}_i$ is the $i$th column of $\mathbf{A}$, and $a_{i,j}$ is its $i$, $j$th entry. $\mathbf{1}$ is a vector consisting of ones, $\mathbf{I}$ is an identity matrix of appropriate size, and $\mathbf{I}_p$ is a $p \times p$ identity matrix.

$E\{\cdot\}$ is the expectation operator, $(\cdot)^T$ is the transpose operator, $(\cdot)^*$ is the complex conjugate operator, $(\cdot)^H$ is the Hermitian transpose. $\|\boldsymbol{a}\|_p$ is the $p$-norm of a vector $\boldsymbol{a}$, defined as $\|\boldsymbol{a}\|_p^p = \sum |a_i|^p$.trace$(\cdot)$ computes the sum of the diagonal elements of a matrix. vect$(\cdot)$ stacks the columns of the argument matrix to form a vector, vectdiag$(\cdot)$ stacks the diagonal elements of the argument matrix to form a vector, diag$(\cdot)$ is a diagonal matrix with its diagonal entries from the argument vector (if the argument is a matrix diag$(\cdot)$ = diag(vectdiag$(\cdot)$)).

Let $\otimes$ denote the Kronecker product, $\circ$ the Khatri–Rao product (column-wise Kronecker product), and $\odot$ the Hadamard (element-wise) product. The following properties are used throughout the paper (for matrices and vectors with compatible dimensions):

$$
\begin{aligned}
(\mathbf{B}^T \otimes \mathbf{A})\text{vect}(\mathbf{X}) &= \text{vect}(\mathbf{AXB}) \\
(\mathbf{B} \otimes \mathbf{A})^H &= (\mathbf{B}^H \otimes \mathbf{A}^H) \\
(\mathbf{B} \otimes \mathbf{A})^{-1} &= (\mathbf{B}^{-1} \otimes \mathbf{A}^{-1}) \\
(\mathbf{B}^T \circ \mathbf{A})\boldsymbol{x} &= \text{vect}(\mathbf{A}\text{diag}(\boldsymbol{x})\mathbf{B}) \\
(\mathbf{BC} \otimes \mathbf{AD}) &= (\mathbf{B} \otimes \mathbf{A})(\mathbf{C} \otimes \mathbf{D}) \\
(\mathbf{BC} \circ \mathbf{AD}) &= (\mathbf{B} \otimes \mathbf{A})(\mathbf{C} \circ \mathbf{D}) \\
(\mathbf{B}^H\mathbf{C} \odot \mathbf{A}^H\mathbf{D}) &= (\mathbf{B} \circ \mathbf{A})^H(\mathbf{C} \circ \mathbf{D}) \\
\text{vectdiag}(\mathbf{A}^H\mathbf{XA}) &= (\mathbf{A}^* \circ \mathbf{A})^H\text{vect}(\mathbf{X}).
\end{aligned}
$$

## 2 DATA MODEL

We employ the array signal processing framework and data model for RA imaging as developed in van der Veen, Leshem & Boonstra (2005), van der Veen & Wijnholds (2013), and Wijnholds & van der Veen (2008).

Assuming a telescope array of $P$ distinct receiving elements, the baseband output signals of the array elements are sampled and split into narrow sub-bands. We assume that the narrow-band condition holds, so that propagation delays across the array can be replaced by complex phase shifts. For simplicity, we will consider only a single sub-band in this paper.

Although the sources are considered stationary, because of the earth's rotation the apparent position of the celestial sources will change with time. For this reason the data is split into short blocks or 'snapshots' of $N$ samples, where the exact value of $N$ depends on the resolution of the instrument.

The sampled signals are stacked into $P \times 1$ vectors $\boldsymbol{x}_k[n]$, where $n = 1, \ldots, N$ is the sample index, and $k = 1, \ldots, K$ denotes the snapshot index. Similarly, assuming $Q$ mutually independent source signals $s_q[n]$ impinging on the array, we stack them into $Q \times 1$ vectors $\boldsymbol{s}_k[n]$. We model the receiver noise as mutually independent zero mean Gaussian signals stacked in a $P \times 1$ vector $\boldsymbol{n}_k[n]$.

The output of the telescope array is a linear combination of the source signals and receiver noise:

$$\boldsymbol{x}_k[n] = \mathbf{A}_k\boldsymbol{s}_k[n] + \boldsymbol{n}_k[n], \tag{1}$$

where $\mathbf{A}_k = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_Q]$ of size $P \times Q$ is called the array response matrix, and $\boldsymbol{a}_q$ is its $q$th column. Ideally, entry $(p, q)$ of $\mathbf{A}_k$ follows from the geometric delay of source $q$ arriving at antenna $p$:

$$a_{p,q} = \frac{1}{\sqrt{P}} e^{-j\frac{2\pi}{\lambda}\boldsymbol{v}_p^T\boldsymbol{z}_q}, \tag{2}$$

where the scaling by $\sqrt{P}$ is such that $\|\boldsymbol{a}_q\| = 1$, $\lambda$ is the wavelength of the received signal, $\boldsymbol{v}_p$ is a $3 \times 1$ vector of the Cartesian location of the $p$th array element (at time-index $k$) with respect to a chosen origin in the field of array, and $\boldsymbol{z}_q$ contains the direction cosines of the $q$th pixel in the image plane. In practice, the position of the celestial sources are unknown. One approach is to decompose the field of view (FoV) of the telescope array into a fine grid where each grid point denotes an image pixel. In the rest of the paper we assume that $Q$ indicates the number of image pixels.

In practice, the array also suffers from antenna-dependent gains and direction-dependent gains that need to be estimated and multiply with $\mathbf{A}_k$. This estimation is done in an outer loop (the selfcal loop) and therefore, for the purpose of this paper, we can assume that $\mathbf{A}_k$ is known (although not necessarily of exactly the form 2). Nonetheless, before selfcal has converged, the data will suffer from a model mismatch.

Without loss of generality, we will from now on consider only a single snapshot $k$, and will drop the index $k$.

Assuming that the signals and the receiver noise are uncorrelated and the noise on different antennas are mutually uncorrelated, the data covariance matrix of the received signals is modelled as

$$\mathbf{R} := E\{\boldsymbol{x}[n]\boldsymbol{x}^H[n]\} = \mathbf{A}\boldsymbol{\Sigma}_s\mathbf{A}^H + \boldsymbol{\Sigma}_n, \tag{3}$$

where $\boldsymbol{\Sigma}_s = \text{diag}\{\boldsymbol{\sigma}_s\}$ and $\boldsymbol{\Sigma}_n = \text{diag}\{\boldsymbol{\sigma}_n\}$ represent the covariance matrices associated with the source signals and the received noise, $\boldsymbol{\sigma} = [\sigma_{s1}^2, \sigma_{s2}^2, \ldots, \sigma_{sQ}^2]^T$ and $\boldsymbol{\sigma}_n = [\sigma_{n,1}^2, \sigma_{n,2}^2, \ldots, \sigma_{n,P}^2]^T$. We assume that the receiver noise powers $\boldsymbol{\Sigma}_n$ are known from the calibration process.

An estimate of the data covariance matrix is obtained using the available received data samples. The sample covariance matrix for a single snapshot is calculated as

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}[n]\boldsymbol{x}^H[n], \tag{4}$$

and is used as an estimate of the true covariance matrix $\mathbf{R}$.

The radio astronomical imaging process amounts to estimating the image pixel intensities $\boldsymbol{\sigma}$ based on the covariance data measured by a telescope array $\hat{\mathbf{R}}$ over the FoV of the array. To obtain a linear measurement model, we vectorize the covariance data model (3) as

$$\boldsymbol{r} = (\mathbf{A}^* \circ \mathbf{A})\boldsymbol{\sigma} + \boldsymbol{r}_n = \mathbf{M}\boldsymbol{\sigma} + \boldsymbol{r}_n, \tag{5}$$

where $\boldsymbol{r} = \text{vect}(\mathbf{R})$, $\boldsymbol{r}_n = \text{vect}(\mathbf{R}_n) = (\mathbf{I} \circ \mathbf{I})\boldsymbol{\sigma}_n$, and $\mathbf{M} = \mathbf{A}^* \circ \mathbf{A}$ is the system matrix of the linear measurement model of size $P^2 \times Q$. Based on (2), one element of $\mathbf{M}$ corresponding to the baseline between the $i$th and $j$th antenna and the $q$th pixel is computed as

$$\mathbf{M}_{ij,q} = a_{iq}^* a_{jq} = \frac{1}{P} e^{j\frac{2\pi}{\lambda}(\boldsymbol{v}_i - \boldsymbol{v}_j)^T\boldsymbol{z}_q}. \tag{6}$$

This expression is modified in the presence of calibration parameters (antenna-dependent gains and direction-dependent gains) which we assume to be known at this point.

Similarly, we vectorize the covariance measurement matrix as

$$\hat{\boldsymbol{r}} = \text{vect}(\hat{\mathbf{R}}) \tag{7}$$

and compensate $\hat{\boldsymbol{r}}$ for the (known) receiver noise powers,

$$\tilde{\boldsymbol{r}} = \hat{\boldsymbol{r}} - \boldsymbol{r}_n, \qquad \tilde{\mathbf{R}} = \hat{\mathbf{R}} - \mathbf{R}_n. \tag{8}$$

This results in a linear measurement equation for estimating $\boldsymbol{\sigma}$ based on the measured $\tilde{\boldsymbol{r}}$:

$$\tilde{\boldsymbol{r}} = \mathbf{M}\boldsymbol{\sigma} + \boldsymbol{e}, \tag{9}$$

where $\boldsymbol{e}$ represents the error due to the finite sample modelling of the covariance data. For a large number of samples $N$ we can assume that $\boldsymbol{e}$ is distributed according to a zero-mean complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \mathbf{C}_{\boldsymbol{e}})$ where $\mathbf{C}_{\boldsymbol{e}} = \frac{1}{N}(\mathbf{R}^T \otimes \mathbf{R})$ (Ottersten, Stoica & Roy 1998; Sardarabadi, Leshem & van der Veen 2016), which will be estimated from $\hat{\mathbf{R}}$.

Incidentally, we remark that many recent papers on radio astronomy image formation start from (9) and model the covariance of $\boldsymbol{e}$ as spatially white, $\mathbf{C}_{\boldsymbol{e}} \propto \mathbf{I}$. However, this is correct only under two assumptions, i.e. (i) the additive noise $\boldsymbol{n}$ is much stronger than the astronomical signals $\boldsymbol{s}$, and (ii) the additive noise is spatially white, $\mathbf{R}_{\boldsymbol{n}} = \sigma_n^2\mathbf{I}$. This requires a whitening operation after calibration of the antenna-dependent gain parameters. These assumptions are usually considered valid in radio astronomy practice.

We also remark that in actual instruments, the autocorrelations of the data are often not formed (or at least not used for imaging) because they are considered to be too much contaminated. In that case, $\tilde{\boldsymbol{r}}$ is not computed from (8), but rather by omitting the autocorrelation terms from $\hat{\boldsymbol{r}}$ (they correspond to the non-zero entries of $\boldsymbol{r}_{\boldsymbol{n}}$). Equation (9) holds but some rows of $\mathbf{M}$ have been dropped. Unfortunately, with this missing data we lose the estimate for $\mathbf{C}_{\boldsymbol{e}}$.[1]

## 3 RADIO ASTRONOMICAL IMAGING PROBLEM FORMULATION

Estimating $\boldsymbol{\sigma}$ from $\tilde{\boldsymbol{r}}$ depends on the properties of the matrix $\mathbf{M}$. Due to the physical constraints on the measurement process, $\mathbf{M}$ is ill-conditioned and in some cases where the requested resolution (number of pixels) $Q$ is very large it may become wide. Therefore, the RA imaging problem is often ill-posed and in some cases underdetermined. Additional prior information or constraints on $\boldsymbol{\sigma}$ are needed to obtain a unique and physically meaningful solution. Generally, this is done by imposing different statistical assumptions on the noise and image (Kay 1993; Bertero & Boccacci 1998).

### 3.1 Beamforming-based estimation

An initial estimate of the image can be obtained via beamforming. In this case, the $i$th pixel of the image is estimated as

$$\hat{\boldsymbol{\sigma}}_i = \boldsymbol{w}_i^H \tilde{\mathbf{R}} \boldsymbol{w}_i = \boldsymbol{w}_i^H (\hat{\mathbf{R}} - \mathbf{R}_{\boldsymbol{n}})\boldsymbol{w}_i, \quad i = 1, \ldots, Q, \tag{10}$$

where $\boldsymbol{w}_i$ is a spatially dependent beamformer (a spatial filter). We consider two common beamforming approaches: Matched Filtering (MF) and MVDR beamforming (Krim & Viberg 1996). The image estimate obtained by the MF beamformer is obtained by setting $\boldsymbol{w}_i = \boldsymbol{a}_i$, so that

$$\hat{\boldsymbol{\sigma}}_{\mathrm{MF},i} = \boldsymbol{a}_i^H(\hat{\mathbf{R}} - \mathbf{R}_{\boldsymbol{n}})\boldsymbol{a}_i \qquad \Leftrightarrow \qquad \hat{\boldsymbol{\sigma}}_{\mathrm{MF}} = \mathbf{M}^H\tilde{\mathbf{r}}. \tag{11}$$

This estimate is known as the 'dirty image' in the radio astronomy community. The expected value of this image is

$$\boldsymbol{\sigma}_{\mathrm{MF},i} = \boldsymbol{a}_i^H(\mathbf{R} - \mathbf{R}_{\boldsymbol{n}})\boldsymbol{a}_i, \quad i = 1, \ldots, Q.$$

Similarly, the MVDR beamformer is defined as (van der Veen & Wijnholds 2013)[2]

$$\boldsymbol{w}_i = \frac{\mathbf{R}^{-1}\boldsymbol{a}_i}{\boldsymbol{a}_i^H\mathbf{R}^{-1}\boldsymbol{a}_i}, \quad i = 1, \ldots, Q, \tag{12}$$

leading to the MVDR dirty image

$$\boldsymbol{\sigma}_{\mathrm{MVDR},i} = \frac{\boldsymbol{a}_i^H\mathbf{R}^{-1}\tilde{\mathbf{R}}\mathbf{R}^{-1}\boldsymbol{a}_i}{(\boldsymbol{a}_i^H\mathbf{R}^{-1}\boldsymbol{a}_i)^2} = \frac{1}{\boldsymbol{a}_i^H\mathbf{R}^{-1}\boldsymbol{a}_i} - \frac{\boldsymbol{a}_i^H\mathbf{R}^{-1}\mathbf{R}_{\boldsymbol{n}}\mathbf{R}^{-1}\boldsymbol{a}_i}{(\boldsymbol{a}_i^H\mathbf{R}^{-1}\boldsymbol{a}_i)^2}. \tag{13}$$

In this expression, the first term is the 'classical' MVDR solution, while the second term is a correction for the unwanted contribution of the noise covariance to this image. Since we do not have access to the true covariance matrix, we use the sample covariance matrix $\hat{\mathbf{R}}$ instead to obtain the beam-formed images.[3]

It is shown by Sardarabadi et al. (2016) that, if the corrections by $\mathbf{R}_{\boldsymbol{n}}$ are ignored in (11) and (13), then

$$\mathbf{0} \le \boldsymbol{\sigma}_{\mathrm{true}} \le \boldsymbol{\sigma}_{\mathrm{MVDR}} \le \boldsymbol{\sigma}_{\mathrm{MF}}. \tag{14}$$

Without ignoring $\mathbf{R}_{\boldsymbol{n}}$, we can prove that the same result holds at least if $\mathbf{R}_{\boldsymbol{n}} = \sigma_n^2\mathbf{I}$ (see Appendix A). This indicates that the MVDR dirty image is always closer to the true image than the MF beamformer.

### 3.2 LS estimation

The most straightforward formulation of the source intensity estimation problem is via LS. In this problem formulation, no statistical assumptions are made about the sources, only the available measurements are fitted to the model in an LS sense. Due to the absence of probabilistic assumptions on $\boldsymbol{\sigma}$, claims about statistical optimality of the solution and its statistical performance cannot be made (Kay 1993).

The LS RA imaging problem can be stated as

$$\hat{\boldsymbol{\sigma}} = \arg\min_{\boldsymbol{\sigma}} \|\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma}\|_2^2. \tag{15}$$

The solution to (15) satisfies the normal equations

$$\mathbf{M}^H\mathbf{M}\hat{\boldsymbol{\sigma}} = \mathbf{M}^H\tilde{\boldsymbol{r}}, \tag{16}$$

where the left-hand side shows the convolution of the image pixels with the beampattern of the array via $\mathbf{M}^H\mathbf{M}$, and the right-hand side $\hat{\boldsymbol{\sigma}}_{\mathrm{MF}} = \mathbf{M}^H\tilde{\boldsymbol{r}}$ is recognized as the MF dirty image which is the same as the image obtained by matched filtering the data.

[1] In the signal processing community, measuring the autocorrelations is considered essential since without the autocorrelations, $\hat{\mathbf{R}}$ would not constitute sufficient statistics for the collected data $\{x_k[n]\}$, i.e. information is lost. If these autocorrelations are deemed to be 'more noisy' then that should be represented in the data model. Estimates of the variance of the measurements are equivalent information and usually available in radio astronomy via the natural weights (Briggs 1995) or System Equivalent Flux Density (SEFD; Van Haarlem et al. 2013). We believe that in any case the autocorrelations will have been used in the calibration and subsequent whitening of the system noise, so that we arrive at the implicit assumptions where $\mathbf{R}_{\boldsymbol{n}} = \sigma_n^2\mathbf{I}$ and astronomical signals much weaker than the noise.

[2] Actually, a correct derivation based on minimization of (10) subject to $\boldsymbol{w}_i^H\boldsymbol{a}_i = 1$ would give a result where $\mathbf{R}^{-1}$ is replaced by $(\mathbf{R} - \mathbf{R}_{\boldsymbol{n}})^{-1}$ in (12), but this inverse is not numerically stable if $\mathbf{R}$ is replaced by its estimate $\hat{\mathbf{R}}$.

[3] If the autocorrelations are not available, then $\tilde{\mathbf{R}}$ represents the sample covariance matrix with the main diagonal replaced by zero. Moreover, $\mathbf{R}^{-1}$ cannot be formed. Under the earlier mentioned assumptions (white noise, weak signals), the factors $\mathbf{R}^{-1}$ in the nominator and denominator cancel each other and the MVDR reduces to the matched filter beamformer in (11).

In RA imaging, the columns of $\mathbf{M}$ corresponding to neighbouring pixels are nearly parallel, making $\mathbf{M}^H\mathbf{M}$ poorly conditioned and the problem ill-posed. For large $Q$, $\mathbf{M}^H\mathbf{M}$ is not even invertible and a unique solution cannot be obtained without regularizing assumptions.

### 3.3 Maximum likelihood estimation

Equation (9) shows the linear measurement model for RA imaging. For such a model, Maximum Likelihood Estimation (MLE) results in an efficient estimator that is also a Minimum Variance Unbiased (MVU) estimator (Kay 1993). If $\boldsymbol{\sigma}$ is considered deterministic (i.e. a parameter vector without associated stochastic model), the likelihood function for (9) with complex Gaussian noise $\boldsymbol{e}$ is

$$p(\tilde{\boldsymbol{r}}|\boldsymbol{\sigma}) = \frac{1}{\pi^{P^2}\det(\mathbf{C}_e)}\exp[-(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})^H\mathbf{C}_e^{-1}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})], \tag{17}$$

where, as mentioned before, $\mathbf{C}_e = \frac{1}{N}(\mathbf{R}^T \otimes \mathbf{R})$; $\det(\cdot)$ denotes the determinant of the matrix. Maximizing the likelihood function is equivalent to minimizing the cost function

$$\mathrm{J}(\boldsymbol{\sigma}) = (\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})^H\mathbf{C}_e^{-1}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma}), \tag{18}$$

which results in the weighted LS (WLS) formulation of the imaging problem as shown in Wijnholds & van der Veen (2008) and van der Veen & Wijnholds (2013).

$$\hat{\boldsymbol{\sigma}} = \arg\min_{\boldsymbol{\sigma}} \|\boldsymbol{\Gamma}(\tilde{r} - \mathbf{M}\boldsymbol{\sigma})\|_2^2, \tag{19}$$

where $\mathbf{C}_e^{-1} = \boldsymbol{\Gamma}^H\boldsymbol{\Gamma}$. The corresponding normal equations are

$$\mathbf{M}^H\mathbf{C}_e^{-1}\mathbf{M}\hat{\boldsymbol{\sigma}} = \mathbf{M}^H\mathbf{C}_e^{-1}\tilde{\boldsymbol{r}},$$

and the WLS (or MLE) solution is given by[4]

$$\hat{\boldsymbol{\sigma}} = (\mathbf{M}^H\mathbf{C}_e^{-1}\mathbf{M})^{-1}\mathbf{M}^H\mathbf{C}_{/bme}^{-1}\tilde{\boldsymbol{r}}. \tag{20}$$

As before, the inversion problem is ill-posed. A regularization term prevents overfitting of the model to the data and penalizes the solution based on the available additional information about the image. We can write the general regularized WLS RA imaging problem as

$$\hat{\boldsymbol{\sigma}} = \arg\min_{\boldsymbol{\sigma}} \|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau\mathcal{R}(\boldsymbol{\sigma}), \tag{21}$$

where $\tau$ is a regularization parameter and $\mathcal{R}(\cdot)$ denotes the regularization operator. Alternatively, we can rewrite the problem formulation (21) as

$$\min_{\boldsymbol{\sigma}} \mathcal{R}(\boldsymbol{\sigma}) \quad \text{subject to} \quad \|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 \leq \epsilon. \tag{22}$$

There is a data-dependent one-to-one correspondence between $\tau$ in (21) and $\epsilon$ in (22) for which the two optimization problems have the same solution. Therefore, they can be used interchangeably. Both $\epsilon$ and $\tau$ are related to the level of noise in the data. As discussed in the introduction, many choices for $\mathcal{R}(\cdot)$ are possible, e.g. $\|\boldsymbol{\sigma}\|_2^2$ or $\|\boldsymbol{\sigma}\|_1$ or $\|\boldsymbol{\Psi}^T\boldsymbol{\sigma}\|_1$.

### 3.4 Bayesian estimation

An alternative approach to regularization is the Bayesian framework. Both the noise term and the image are assumed random variables, and a prior distribution $p(\boldsymbol{\sigma})$ is posed. The Maximum A Posteriori estimator is defined as (Kay 1993)

$$\begin{aligned}\hat{\boldsymbol{\sigma}} &= \arg\max_{\boldsymbol{\sigma}} p(\boldsymbol{\sigma}|\tilde{\boldsymbol{r}}) = \arg\max_{\boldsymbol{\sigma}} \frac{p(\tilde{\boldsymbol{r}}|\boldsymbol{\sigma})p(\boldsymbol{\sigma})}{\int p(\tilde{\boldsymbol{r}}|\boldsymbol{\sigma})p(\boldsymbol{\sigma})d\boldsymbol{\sigma}} \\ &= \arg\max_{\boldsymbol{\sigma}} p(\tilde{\boldsymbol{r}}|\boldsymbol{\sigma})p(\boldsymbol{\sigma}).\end{aligned} \tag{23}$$

Here, $p(\boldsymbol{\sigma}|\tilde{\boldsymbol{r}})$ denotes the posterior probability density function of the image given the observation, and Bayes' rule is used to replace it by $p(\tilde{r}|\boldsymbol{\sigma})p(\boldsymbol{\sigma})$, which is a product of the likelihood of the observation given an image with the prior probability of that image. The likelihood is given in (17). Assuming for simplicity that the prior for the image is also distributed according to a Gaussian distribution, with mean $\boldsymbol{\mu}_\sigma$ and covariance $\mathbf{C}_\sigma$, then $\boldsymbol{\sigma} \sim \mathcal{N}(\boldsymbol{\mu}_\sigma, \mathbf{C}_\sigma)$, or

$$p(\boldsymbol{\sigma}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma)^T\mathbf{C}_\sigma^{-1}(\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma)\right]. \tag{24}$$

The log of the posterior likelihood is then

$$\begin{aligned}\log p(\boldsymbol{\sigma}|\tilde{r}) \propto &-(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})^H\mathbf{C}_e^{-1}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma}) \\ &- \frac{1}{2}(\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma)^T\mathbf{C}_\sigma^{-1}(\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma).\end{aligned} \tag{25}$$

If we define the Cholesky factorization of the inverse image covariance matrix as

$$\mathbf{C}_\sigma^{-1} = \mathbf{L}^T\mathbf{L}, \tag{26}$$

we can equivalently write this as

$$\log p(\boldsymbol{\sigma}|\tilde{\boldsymbol{r}}) \propto -\|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 - \frac{1}{2}\|\mathbf{L}(\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma)\|_2^2. \tag{27}$$

Therefore, maximizing the posterior likelihood is equivalent to solving the minimization problem

$$\hat{\boldsymbol{\sigma}} = \arg\min_{\boldsymbol{\sigma}} \|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau\|\mathbf{L}(\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma)\|_2^2, \tag{28}$$

where $\tau = \frac{1}{2}$. This is also known as ridge regression and a specific case of (21), with the advantage that there is some insight in the role of $\mathbf{L}$. For example if we have accurate prior knowledge, then $\mathbf{C}_\sigma$ is small and $\mathbf{L}$ is large, and the solution $\hat{\boldsymbol{\sigma}}$ will be close to $\boldsymbol{\mu}_\sigma$. If instead of a Gaussian prior we assume a Laplace distribution for $\boldsymbol{\sigma}$,

$$p(\sigma_i) = \frac{1}{b_i}\exp\left(-\frac{|\sigma_i - \mu_i|}{b_i}\right),$$

we obtain an $\ell_1$ constraint (or LASSO). The Laplace distribution is more concentrated around zero and has long tails, which models images that are mostly zero with occasional outliers, explaining why $\ell_1$ constraints lead to sparse solutions. Similarly, a lognormal density prior will lead to constraints that generate a maximum-entropy solution (Kaipio & Somersalo 2004), and such a prior was used in RESOLVE (Junklewitz et al. 2016). Thus, the Bayesian framework is a general method to derive constrained optimization problems.

Returning to the Gaussian prior, we can rewrite (28) as

$$\hat{\boldsymbol{\sigma}} = \arg\min_{\boldsymbol{\sigma}} \left\| \begin{bmatrix} \boldsymbol{\Gamma}\mathbf{M} \\ \sqrt{\tau}\mathbf{L} \end{bmatrix} \boldsymbol{\sigma} - \begin{bmatrix} \boldsymbol{\Gamma}\tilde{\boldsymbol{r}} \\ \sqrt{\tau}\mathbf{L}\boldsymbol{\mu}_\sigma \end{bmatrix} \right\|_2^2.$$

The corresponding normal equations are

$$(\mathbf{M}^H\mathbf{C}_e^{-1}\mathbf{M} + \tau\mathbf{C}_\sigma^{-1})\boldsymbol{\sigma} = \mathbf{M}^H\mathbf{C}_e^{-1}\tilde{\boldsymbol{r}} + \tau\mathbf{C}_\sigma^{-1}\boldsymbol{\mu}_\sigma, \tag{29}$$

and the solution is

$$\hat{\boldsymbol{\sigma}} = (\mathbf{M}^H\mathbf{C}_e^{-1}\mathbf{M} + \tau\mathbf{C}_\sigma^{-1})^{-1}(\mathbf{M}^H\mathbf{C}_e^{-1}\tilde{\boldsymbol{r}} + \tau\mathbf{C}_\sigma^{-1}\boldsymbol{\mu}_\sigma).$$

[4]The weighting by $\mathbf{C}_e$ is omitted if the autocorrelations are not known, or if we may assume that the noise is white and much stronger than the sources.

For the specific case where $\boldsymbol{\mu}_\sigma = \mathbf{0}$, and assuming white processes $\mathbf{C}_e = \nu^2\mathbf{I}$ and $\mathbf{C}_\sigma = \eta^2\mathbf{I}$, equation (29) can be written as

$$(\mathbf{M}^H\mathbf{M} + \tau\mathbf{I})\boldsymbol{\sigma} = \mathbf{M}^H\tilde{\boldsymbol{r}}, \quad \tau = \frac{1}{2}\left(\frac{\nu}{\eta}\right)^2, \tag{30}$$

which is recognized as a Tikhonov regularized LS problem (Bertero & Boccacci 1998). Thus, these standard regularization methods are all included in the Bayesian framework.

The main question in the Bayesian framework is the selection of a suitable prior. For example we can select a Gaussian prior where $\boldsymbol{\mu}_\sigma$ is the currently best known estimate for the image (the current sky map), with $\mathbf{C}_\sigma$ related to the accuracy of that knowledge. As it is hard to quantify this, $\mathbf{C}_\sigma$ could be modelled as a diagonal matrix, with the unknown variances on the diagonal modelled in turn as statistical parameters, for which a distribution (with unknown parameters called hyperparameters) has to be proposed. The estimation of these hyperpriors from the data is known as Sparse Bayesian Learning (Tipping 2001) and in the context of our problem has been worked out by Wipf & Rao (2004). The RESOLVE method (Junklewitz et al. 2016) follows a similar approach. Unfortunately, the computational complexity is reported to be rather high.

Since the prior in this framework is data-dependent, the question at this point is whether it would be possible to use a (perhaps less optimal) data-dependent prior that is easier to estimate.

## 4 PROPOSED SOLUTION METHOD

### 4.1 Problem reformulation

We focus on the Tikhonov regularized WLS problem formulation and will use $\boldsymbol{\mu}_\sigma = \mathbf{0}$ and restrict $\mathbf{L}$ to be diagonal. Our aim is thus to propose a suitable $\mathbf{L}$. Since $\mathbf{C}_\sigma^{-1} = \mathbf{L}^H\mathbf{L}$, the diagonal entries of $\mathbf{L}$ model the *precision* of our prior knowledge, and a large entry of $\mathbf{L}$ will result in a dark pixel (since $\boldsymbol{\mu}_\sigma = \mathbf{0}$), whereas a small entry of $\mathbf{L}$ will make that pixel to be determined by the data.

With change of variables $\boldsymbol{\alpha} = \mathbf{L}\boldsymbol{\sigma}$, we can rewrite the objective function (21) in terms of $\boldsymbol{\alpha}$ as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\mathbf{L}^{-1}\boldsymbol{\alpha})\|_2^2 + \tau\|\boldsymbol{\alpha}\|_2^2. \tag{31}$$

The image can be recovered from $\hat{\boldsymbol{\alpha}}$ by the linear transform $\hat{\boldsymbol{\sigma}} = \mathbf{L}^{-1}\hat{\boldsymbol{\alpha}}$.

Equation (31) is equivalent to the solution of

$$(\mathbf{L}^{-H}\mathbf{M}^H\mathbf{C}_e^{-1}\mathbf{M}\mathbf{L}^{-1} + \tau\mathbf{I})\boldsymbol{\alpha} = \mathbf{L}^{-H}\mathbf{M}^H\mathbf{C}_e^{-1}\tilde{\boldsymbol{r}}. \tag{32}$$

With the choice of $\mathbf{C}_\sigma^{-1} = \mathbf{L}^H\mathbf{L}$ and $\mathbf{C}_e^{-1} = \boldsymbol{\Gamma}^H\boldsymbol{\Gamma}$ and change of variables $\bar{\mathbf{M}} = \boldsymbol{\Gamma}\mathbf{M}\mathbf{L}^{-1}$ and $\bar{\boldsymbol{r}} = \boldsymbol{\Gamma}\tilde{\boldsymbol{r}}$ we can rewrite this as

$$(\bar{\mathbf{M}}^H\bar{\mathbf{M}} + \tau\mathbf{I})\boldsymbol{\alpha} = \bar{\mathbf{M}}^H\bar{\boldsymbol{r}}. \tag{33}$$

Such a scaling of the columns of $\mathbf{M}$ by a matrix $\mathbf{L}^{-1}$ related to the prior distribution is known as prior-conditioning (Calvetti & Somersalo 2005), as it is similar in shape to the preconditioning that is sometimes done in iterative solvers to improve convergence. The difference is that preconditioning only involves $\mathbf{M}$ whereas prior-conditioning is not just based on $\mathbf{M}$ but on the interaction of $\mathbf{M}$ with the data $\tilde{\boldsymbol{r}}$.

To obtain a prior, data-dependent, estimate of $\mathbf{C}_\sigma$, the idea is to compute from the data an unbiased estimate for the image, using minimal assumptions, i.e. we consider $\boldsymbol{\sigma}$ deterministic. The variance of this estimate can then be used as estimate for $\mathbf{C}_\sigma$.

The best possible estimate under this assumption is the MLE estimate, in this case equal to the WLS estimate

$$\hat{\boldsymbol{\sigma}}_{\text{MLE}} = (\mathbf{M}^H\mathbf{C}_e^{-1}\mathbf{M})^{-1}\mathbf{M}^H\mathbf{C}_e^{-1}\tilde{\boldsymbol{r}}. \tag{34}$$

It is known that this estimator is an efficient MVU estimator (Kay 1993) with covariance

$$\mathbf{C}_{\hat{\sigma}} = (\mathbf{M}^H\mathbf{C}_e^{-1}\mathbf{M})^{-1}, \tag{35}$$

where $\mathbf{C}_e = \frac{1}{N}(\mathbf{R}^T \otimes \mathbf{R})$. Therefore

$$\begin{aligned}\mathbf{M}^H\mathbf{C}_e^{-1}\mathbf{M} &= N(\mathbf{A}^* \circ \mathbf{A})^H(\mathbf{R}^{-T} \otimes \mathbf{R}^{-1})(\mathbf{A}^* \circ \mathbf{A}) \\ &= N(\mathbf{A}^T\mathbf{R}^{-T}\mathbf{A}^*) \odot (\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A}).\end{aligned} \tag{36}$$

If we denote the variance of $\hat{\boldsymbol{\sigma}}$ as $\text{Var}(\hat{\boldsymbol{\sigma}})$, it consists of the diagonal elements of $\mathbf{C}_{\hat{\sigma}}$, i.e.

$$\text{Var}(\hat{\boldsymbol{\sigma}}) = \text{diag}(\mathbf{C}_{\hat{\sigma}}). \tag{37}$$

Based on equation (36), the $i$th diagonal element of $\mathbf{M}^H\mathbf{C}_e^{-1}\mathbf{M}$ can be computed as $N(\boldsymbol{a}_i^H\mathbf{R}^{-1}\boldsymbol{a}_i)^2$. Although equation (36) shows that the estimated pixel intensities are correlated, we ignore that and set $\mathbf{C}_{\hat{\sigma}} \approx \text{diag}(\text{Var}(\hat{\boldsymbol{\sigma}}))$ where

$$\text{Var}(\hat{\sigma}_i) = \frac{1}{N(\boldsymbol{a}_i^H\mathbf{R}^{-1}\boldsymbol{a}_i)^2}, \qquad i = 1, 2, \ldots, Q, \tag{38}$$

with $\text{Var}(\hat{\sigma}_i)$ denoting the variance of the $i$th pixel estimate. Comparing (13) and (38) we conclude that (if $\mathbf{R}_n$ is ignored in 13)

$$\mathbf{C}_{\hat{\sigma}} \approx \text{diag}(\text{Var}(\hat{\boldsymbol{\sigma}})) = \frac{1}{N}\text{diag}(\boldsymbol{\sigma}_{\text{MVDR}})^2. \tag{39}$$

Since the true data covariance matrix is not available, we will use the sample covariance matrix $\hat{\mathbf{R}}$, and obtain the estimated MVDR image $\hat{\boldsymbol{\sigma}}_{\text{MVDR}}$. This leads to the choice to set

$$\mathbf{L}^{-1} = \text{diag}(\hat{\boldsymbol{\sigma}}_{\text{MVDR}}) \tag{40}$$

as regularizing operator (A factor $\sqrt{N}$ is absorbed in $\tau$.).

While this choice is obviously a shortcut from a truly Bayesian approach (e.g. the mean value of the initial image is ignored and only the variance is taken into account), we will show in the simulations that this simple idea is very effective in obtaining regularized solutions. Moreover, it is computationally not very involved as it amounts to constructing a beam-formed image (similar to computing the classical dirty image), followed by solving (33). We propose to use Krylov subspace iterations to do this efficiently (Section 5.1).

### 4.2 Discussion and generalizations

Before we develop an efficient algorithm for finding the solution of the problem stated in Section 4.1, we discuss some of the properties of the problem and address potential generalizations of the framework.

(1) RA images contain substantial black background of radio-quiet zones. The working principle of greedy algorithms such as CLEAN and NNLS is to first obtain the support of the image, also called the active set, and to solve only for the elements of the image in the active set. Therefore, as shown by Marsh & Richardson (1987), these methods solve the regularized LS or MLE problem (21) with $\mathcal{R}(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}\|_0$,

$$\hat{\boldsymbol{\sigma}} = \arg\min_{\boldsymbol{\sigma}} \|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau\|\boldsymbol{\sigma}\|_0, \tag{41}$$

with the addition of a non-negativity constraint for NNLS. Minimizing the $\ell_0$ norm produces satisfactory results both in terms of the support of the image and the intensity estimates if the underlying image is sufficiently sparse and only consists of scattered point sources.

In line with our problem formulation, the $\ell_0$ constraint can be translated into a right preconditioner. If we assume for the moment the knowledge of the true $\boldsymbol{\sigma}$, denoted as $\boldsymbol{\sigma}_{\text{true}}$, we can define a

diagonal matrix $\mathbf{D}$ as

$$[\mathbf{D}]_{i,i} = \begin{cases} 1, & \text{if } [\boldsymbol{\sigma}_{\text{true}}]_i > 0 \\ 0, & \text{if } [\boldsymbol{\sigma}_{\text{true}}]_i = 0. \end{cases} \tag{42}$$

Therefore, in terms of the LS formulation, we need to solve the problem

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\mathbf{D}\boldsymbol{\alpha})\|_2^2 + \tau\|\boldsymbol{\alpha}\|_2^2, \tag{43}$$

where the image estimate is found by the transform $\hat{\boldsymbol{\sigma}} = \mathbf{D}\hat{\boldsymbol{\alpha}}$. Thus, $\hat{\boldsymbol{\sigma}}$ will be zero where $\mathbf{D}_{i,i}$ is zero, and $\mathbf{D}$ would be the optimal prior conditioner. In reality we do not know $\boldsymbol{\sigma}_{\text{true}}$. Finding the active set in greedy algorithms is done iteratively through outer iterations. This increases the cost of the algorithms substantially. Clearly, problem (43) is connected to problem (31) considered in Section 4.1 via $\mathbf{D} = \mathbf{L}^{-1}$. In this context, our use of a beam-formed image $\mathbf{D} = \text{diag}(\hat{\boldsymbol{\sigma}}_{\text{MVDR}})$ or $\mathbf{D} = \text{diag}(\hat{\boldsymbol{\sigma}}_{\text{MF}})$ can be interpreted as a surrogate for this.

(2) Low resolution initial estimates of the image can be obtained via MF or MVDR beamforming. Previously, we suggested in Naghibzadeh, Sardarabadi & van der Veen (2016b) and Naghibzadeh & van der Veen (2017b) to use the MF dirty image $\text{diag}(\hat{\boldsymbol{\sigma}}_{\text{MF}})$ for regularization purposes. Moreover, we showed the relation to Bayesian estimation when applying MVDR-based right prior-conditioning weights in Naghibzadeh & van der Veen (2017a). If the noise is lower or comparable to the signal, we have the relation (14),

$$\mathbf{0} \leq \boldsymbol{\sigma}_{\text{true}} \leq \boldsymbol{\sigma}_{\text{MVDR}} \leq \boldsymbol{\sigma}_{\text{MF}}. \tag{44}$$

Therefore, the prior variance $\mathbf{C}_{\boldsymbol{\sigma}}$ based on the MF dirty image is higher than when the MVDR dirty image is used, and the latter provides a better start. The correction by $\mathbf{R}_{\boldsymbol{n}}$ introduced in (13) moves the MVDR image even further towards the true image. It is also important to note from this relation that the true image is black wherever the initial image is black. This way the initial estimate provides a rough estimate for the true support of the image.

If we do not know the autocorrelations in the measurement data, we can use the MF estimate without the diagonals or the MVDR image obtained by diagonal loading. However, we must take care that all brightness estimates are strictly positive by adding a constant value to the image since negative weights will be completely wrong while zero weights will result in pixels that will stay black throughout the iterations and in the final solution. We note that without autocorrelation information the results will be suboptimal. Appendix B gives a brief analysis and provides additional remarks related to the proposed imaging techniques.

(3) We show that applying MF or MVDR dirty images as prior conditioners favours smooth reconstructions and is therefore more interesting for the recovery of diffuse structures and smooth features of the sky map rather than point sources. We motivate our claim for prior-conditioning with the MF. Analysis for MVDR-based prior-conditioning would be similar.

Assuming for the moment that there is no noise and error on the covariance measurements, i.e. $\boldsymbol{e} = \mathbf{0}$ and $\boldsymbol{r}_{\boldsymbol{n}} = \mathbf{0}$, based on (11) we can write the true dirty image as

$$\boldsymbol{\sigma}_{\text{MF}} = \mathbf{M}^H(\mathbf{M}\boldsymbol{\sigma}). \tag{45}$$

If we consider the image only contains a unit norm point source in the middle of the FoV, we can rewrite (45) as

$$\boldsymbol{b} = (\mathbf{M}^H\mathbf{M})\boldsymbol{e}_{\text{mid}}, \tag{46}$$

where $\boldsymbol{e}_{\text{mid}}$ is the unit vector with the element in the middle of the FoV equal to 1. In this case the dirty image is called the dirty beam, indicated by $\boldsymbol{b}$. Dirty beam is also known as Point Spread Function (PSF) and impulse response or the beam pattern of the telescope array. If we insert an arbitrary image $\boldsymbol{\sigma}$ in the FoV of the array, resulting $\boldsymbol{\sigma}_{\text{MF}}$ would be the convolution of dirty beam with the image. Dirty beam by construction acts as a low-pass filter with the main beam corresponding to the resolution of the array. Therefore, the resulting dirty image will be a low-pass filtered version of the sky map and is smooth.

When we use the dirty image as a prior, the smoothness will be preserved in the resulting image from (31). Since the extended emissions exhibit a smooth structure, in the reconstruction they will be preserved. By the same token, isolated point sources will not be imaged sharper than the resolution of the instrument and will be spread out. This spreading is similar to the post-processing applied to the CLEAN solution to restore the natural resolution of the telescope array. In our method, since the prior obeys the resolution of the array, spreading is done automatically.

(4) Next, we show that applying the regularization operator as a prior-conditioner opens the door to various regularizations. This can be applied in cases where the underlying sky map contains isolated point sources.

It is well-known that $\ell_1$ constraints result in sparse solutions. The associated regularized MLE problem is

$$\hat{\boldsymbol{\sigma}} = \arg\min_{\boldsymbol{\sigma}} \|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau\|\boldsymbol{\sigma}\|_1. \tag{47}$$

One way to solve (47) is via the Iteratively Reweighted LS method (Daubechies et al. 2010). The $\ell_1$ constraint is transformed to an $\ell_2$ constraint by

$$\begin{aligned} \|\boldsymbol{\sigma}\|_1 &= \sum_{i=1}^{Q} |\sigma_i| = \sum_{i=1}^{Q} \frac{|\sigma_i|^2}{|\sigma_i|} \\ &= \|\mathbf{W}\boldsymbol{\sigma}\|_2^2, \quad \text{where} \quad \mathbf{W} = \text{diag}(\boldsymbol{\sigma}^{-1/2}). \end{aligned} \tag{48}$$

Equation (48) suggests that $\|\boldsymbol{\sigma}\|_1$ can be computed from a properly weighted $\ell_2$-norm. Although this optimal weight is unknown, we can enter an iteration wherein, at each step, the weight is based on the solution obtained at the previous step. It is thus sufficient to solve only weighted LS problems. Specifically, we define the weight matrix at iteration $k$ as $\mathbf{W}_k = \text{diag}(\hat{\boldsymbol{\sigma}}_{k-1}^{-1/2})$ where $\hat{\boldsymbol{\sigma}}_{k-1}$ is the solution obtained at the previous iteration $k-1$. Therefore, (47) is replaced by

$$\hat{\boldsymbol{\sigma}}_k = \arg\min_{\boldsymbol{\sigma}} \|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau\|\mathbf{W}_k\boldsymbol{\sigma}\|_2^2 \tag{49}$$

which can be transformed into a right preconditioned system using the transform $\boldsymbol{\alpha} = \mathbf{W}_k\boldsymbol{\sigma}$,

$$\hat{\boldsymbol{\alpha}}_k = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\mathbf{W}_k^{-1}\boldsymbol{\alpha})\|_2^2 + \tau\|\boldsymbol{\alpha}\|_2^2. \tag{50}$$

After the estimate $\hat{\boldsymbol{\sigma}}_k$ is obtained, problem (50) is solved again with the new weights. Therefore, this method requires solving (50) multiple times where the outer iterations are indicated by $k$. Comparing to (31), we see that (50) is a prior-conditioned problem where the prior is iterated upon as more accurate images are being computed.

If we start the iteration with $\mathbf{W}_0 = \mathbf{I}$, then the first estimate $\hat{\boldsymbol{\sigma}}_1$ will be the MLE estimate (20). The next iteration will solve a right preconditioned system where the square-root of this image is the prior.

In contrast, the prior we proposed in (40) uses the MVDR image and omits the square-root. Nonetheless, it is interesting to consider what happens if we iterate this estimate in the same way as (50), but with $\mathbf{W}_k = \text{diag}(|\hat{\boldsymbol{\sigma}}_{k-1}|^{-1})$. If this converges to a fixed point,

the corresponding constraint is

$$\|\mathbf{W}\boldsymbol{\sigma}\|_2^2 = \sum_{i=1,\sigma_i\neq 0}^{Q} \frac{|\sigma_i|^2}{|\sigma_i|^2} = \|\boldsymbol{\sigma}\|_0. \tag{51}$$

This shows that iteratively minimizing (50) in this way is a surrogate for using the $\ell_0$ norm as regularizer, and will result in a very sparse image (even if the true image is not sparse). We show this effect with simulations in a 1D setting.

Iterations of this form have been proposed in the context of our problem by Gorodnitsky & Rao (1997), and are known as the FOCUSS algorithm. As mentioned in that paper, sparsity by itself does not form a sufficient constraint to obtain a unique estimate for an underdetermined problem, and the use of a low-resolution initial estimate provides the necessary additional constraint. The paper proves the quadratic convergence to a local fixed point in the neighbourhood of the initialization, and also mentions a technique to impose a positivity constraint on the solution. Unfortunately, the proposed solution method is based on the truncated Singular Value Decomposition (SVD) and is not applicable for large-scale problems.

Overall, the regularization penalty $\|\boldsymbol{\sigma}\|_0$ assumes the image is composed of a set of separate point sources. On the other hand, the $\|\boldsymbol{\sigma}\|_2$ penalty favours solutions with similar intensity levels over different pixels to minimize the overall power and is suitable for the recovery of diffuse emissions. The $\|\boldsymbol{\sigma}\|_1$ penalty is intermediate between smoothness and sparsity penalties but is not specifically designed for extended or point sources. It is known that if the $\ell_0$-constrained problem (41) contains a sufficiently sparse solution, the $\ell_1$-constrained problem (47) as a surrogate for (41) will recover it (Bruckstein, Donoho & Elad 2009). However, in cases where both resolved and unresolved sources coexist in the sky map, (47) recovers both types of sources but it is not optimal for any of them. In Appendix C we discuss some attempts to generalize the formulation such that both resolved and unresolved sources can be recovered. This is done by means of introducing overcomplete dictionaries. Appendix C shows the capability of the proposed framework for generalization.

We see that by applying outer iterations, based on the prior-conditioning formulation, we are able to also impose sparsity-promoting norms into the framework of Section 4.1. Doing so, we are able to benefit from the efficient algebraic algorithms that exist for solving the $\ell_2$ regularized problem to also recover images with sparsity priors. In the next section, we present an efficient algorithmic framework to solve the prior-conditioned problem.

## 5 THE PRIFIRA ALGORITHM

The proposed solution method from Section 4 is now further worked out into an algorithm which we call PRIFIRA.

### 5.1 Implementation using Krylov subspace methods

We solve problem (31) by an iterative method based on projections onto Krylov subspaces. Let $\bar{\mathbf{M}} = \boldsymbol{\Gamma}\mathbf{M}\mathbf{L}^{-1}$ and $\bar{\boldsymbol{r}} = \boldsymbol{\Gamma}\tilde{\boldsymbol{r}}$, then (31) is written as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\bar{r} - \bar{\mathbf{M}}\boldsymbol{\alpha}\|_2^2 + \tau\|\boldsymbol{\alpha}\|_2^2. \tag{52}$$

Define the $t$-dimensional Krylov subspace $\mathcal{K}_t$, for $t = 1, 2, \ldots$, as

$$\mathcal{K}_t(\bar{\mathbf{M}}^H\bar{\mathbf{M}}, \bar{\mathbf{M}}^H\bar{\boldsymbol{r}}) = \text{span}\{\bar{\mathbf{M}}^H\bar{\boldsymbol{r}}, (\bar{\mathbf{M}}^H\bar{\mathbf{M}})\bar{\mathbf{M}}^H\bar{\boldsymbol{r}}, \ldots, (\bar{\mathbf{M}}^H\bar{\mathbf{M}})^{t-1}\bar{\mathbf{M}}^H\bar{\boldsymbol{r}}\}. \tag{53}$$

Krylov subspace methods instead solve the problem

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\bar{\boldsymbol{r}} - \bar{\mathbf{M}}\boldsymbol{\alpha}\|_2^2 \quad \text{subject to} \quad \boldsymbol{\alpha} \in \mathcal{K}_t(\bar{\mathbf{M}}^H\bar{\mathbf{M}}, \bar{\mathbf{M}}^H\bar{\boldsymbol{r}}) \tag{54}$$

for $t = 1, 2, \ldots$. As the iteration count $t$ increases, the Krylov subspace gradually increases in dimension as well, so that the residual $\|\bar{\mathbf{r}} - \bar{\mathbf{M}}\hat{\boldsymbol{\alpha}}_t\|_2^2$ decreases while $\|\boldsymbol{\alpha}_t\|_2^2$ usually increases.

Due to the ill-posedness of the problem, $\|\boldsymbol{\alpha}_t\|_2^2$ will grow out of bound as the iteration progresses (Hanke 1995). One way to stop the iterations while the solution is still numerically stable is via the discrepancy principle (Hanke 1995; Nemirovskii 1986). In this case, the iteration is stopped at iteration $T$ once $\|\bar{\mathbf{r}} - \bar{\mathbf{M}}\boldsymbol{\alpha}_t)\|_2^2 \leq \epsilon$, which then gives an approximate solution to (52).

The restriction to the Krylov subspace before it spans the complete space provides a regularization, called semiconvergence (Hansen 2010). If the iteration is allowed to continue, then the residual converges to zero and the solution converges to the pseudo-inverse minimum norm solution (20), so that we obtain the unregularized solution. In contrast to Tikhonov regularization or truncated SVD where the regularization only depends on $\mathbf{M}$ and not on the measured data, the regularization provided by Krylov subspace methods adapts to the data via the initial vector $\bar{\boldsymbol{r}}$. While problem (52) is not exactly equivalent to (54), their solutions are considered very similar (Hansen 2005).

Krylov subspace methods are attractive because we do not need to store $\bar{\mathbf{M}}$, rather we need to provide functions that return matrix-vector products of the form $\bar{\mathbf{M}}\boldsymbol{u}$ and $\bar{\mathbf{M}}^H\boldsymbol{v}$ (Saad 1981). With the functional form of $\mathbf{M}$ as given in (6), we can implement such a subroutine. This greatly reduces storage requirements. Related details are in Section 5.3.

To solve (54) iteratively for a non-square $\bar{\mathbf{M}}$ with arbitrary rank, the LSQR method (Paige & Saunders 1982) is appropriate (Choi 2006). This is analytically equivalent to the Conjugate Gradient method applied to the normal equations, but is numerically preferred (Hanke 1995). The LSQR method is based on the Golub–Kahan (GK) bidiagonalization algorithm (Golub & Kahan 1965), also referred to as Lanczos iterations. First define

$$\beta_1 = \|\bar{r}\|, \quad \boldsymbol{u}_1 = \bar{r}/\beta_1,$$
$$\alpha_1 = \|\bar{\mathbf{M}}^H\boldsymbol{u}_1\|, \quad \boldsymbol{v}_1 = \bar{\mathbf{M}}^H\boldsymbol{u}_1/\alpha_1.$$

Using these as initialization, in the GK process, at iteration $t$, two orthonormal vectors $\boldsymbol{v}_t$ and $\boldsymbol{u}_t$ are computed as

$$\beta_t\boldsymbol{u}_t := \bar{\mathbf{M}}\boldsymbol{v}_{t-1} - \alpha_{t-1}\boldsymbol{u}_{t-1},$$
$$\alpha_t\boldsymbol{v}_t := \bar{\mathbf{M}}^H\boldsymbol{u}_t - \beta_t\boldsymbol{v}_{t-1} \tag{55}$$

where $\beta_t$ and $\alpha_t$ are chosen such that $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$ are normalized. Let

$$\mathbf{B}_t = \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \beta_3 & \ddots & \\ & & \ddots & \alpha_t \\ & & & \beta_{t+1} \end{bmatrix}. \tag{56}$$

It can then be shown that solving (54) reduces to solving the bidiagonalized LS problem

$$\min_{\boldsymbol{y}_t} \|\mathbf{B}_t\boldsymbol{y}_t - \beta_1\boldsymbol{e}_1\|_2^2, \tag{57}$$

where $\boldsymbol{e}_1$ is the unit vector with its first element equal to one. LSQR uses QR updates to obtain $\boldsymbol{y}_t$ at each iteration $t$ (Paige & Saunders 1982).

The complete algorithm to solve (54) and compute the estimate of the image $\hat{\boldsymbol{\sigma}}$ is summarized in algorithm 1.

**Algorithm 1:** PRIFIRA (based on LSQR)

**input** : $\bar{\mathbf{r}}$, $\bar{\mathbf{M}}$ (or operator function), $\mathbf{L}^{-1}$, $\epsilon$
**output**: image $\hat{\boldsymbol{\sigma}}$

Initialize: $\beta_1\mathbf{u}_1 := \bar{\mathbf{r}}$, $\alpha_1\mathbf{v}_1 := \bar{\mathbf{M}}^H\mathbf{u}_1$, $\boldsymbol{\omega}_1 := \mathbf{v}_1$, $\boldsymbol{\alpha}_1 := \mathbf{0}$, $\bar{\Phi}_1 := \beta_1$, $\bar{\rho}_1 := \alpha_1$, $t := 1$;
**while** *stopping criteria not satisfied* **do**
  $\beta_{t+1}\mathbf{u}_{t+1} := \bar{\mathbf{M}}\mathbf{v}_t - \alpha_t\mathbf{u}_t$,
  $\alpha_{t+1}\mathbf{v}_{t+1} := \bar{\mathbf{M}}^H\mathbf{u}_{t+1} - \beta_t\mathbf{v}_t$
  Construct and apply orthogonal transform:
  $\rho_t = (\bar{\rho}_t^2 + \beta_{t+1}^2)^{1/2}$
  $c_t = \bar{\rho}_t/\rho_t$, $s_t = \beta_{t+1}/\rho_t$
  $\theta_{t+1} = s_t\alpha_{t+1}$, $\bar{\rho}_{t+1} = -c_t\alpha_{t+1}$
  $\Phi_t = c_t\bar{\Phi}_t$, $\bar{\Phi}_{t+1} = s_t\bar{\Phi}_t$
  Update:
  $\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + (\Phi_t/\rho_t)\boldsymbol{\omega}_t$
  $\boldsymbol{\omega}_{t+1} = \mathbf{v}_{t+1} - (\theta_{t+1}/\rho_t)\boldsymbol{\omega}_t$
  $t := t + 1$
**end**
$\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}_t$
Transform to the image: $\hat{\boldsymbol{\sigma}} = \mathbf{L}^{-1}\hat{\boldsymbol{\alpha}}$

If we are going to apply the generalized reweighted prior-conditioning as discussed in Section 4.2, we can do so by defining an outer iteration loop around Algorithm 1 where the weights are obtained using the values of $\hat{\boldsymbol{\sigma}}$ at the previous iteration. The reweighted algorithm is summarized as Algorithm 2, where $f(\boldsymbol{\sigma})$ refers to an arbitrary function applied to $\boldsymbol{\sigma}$ which depends on the constraint as discussed in Section 4.2, and PRIFIRA$(\cdot)$ denotes Algorithm 1 with the mentioned input. The outer loop also allows for applying more constraints such as projecting the solution into the real and positive orthant but comes at a greater computation expense due to the repeated application of the LSQR algorithm. As initialization, we can choose the MVDR dirty image, the MF dirty image, or simply set $\boldsymbol{\sigma}_0 = \mathbf{1}$.

**Algorithm 2:** Reweighted PRIFIRA

Initialize:
$\boldsymbol{\sigma}_0 = \mathbf{1}$, $\mathbf{W}_1^{-1} = \mathrm{diag}(f(\boldsymbol{\sigma}_0))$, $\bar{\mathbf{M}}_1 = \boldsymbol{\Gamma}\mathbf{M}\mathbf{W}_1^{-1}$
**for** $k = 1, 2, \ldots, K$ **do**
  $\boldsymbol{\sigma}_k = PRIFIRA(\bar{\mathbf{r}}, \bar{\mathbf{M}}_k, \mathbf{W}_k^{-1}, \epsilon)$
  $\mathbf{W}_{k+1}^{-1} = \mathrm{diag}(f(\boldsymbol{\sigma}_k))$,
  $\bar{\mathbf{M}}_{k+1} = \boldsymbol{\Gamma}\mathbf{M}\mathbf{W}_{k+1}^{-1}$,
**end**
$\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}_K$

### 5.2 Stopping criteria

Algorithm 1 requires an appropriate stopping rule. As mentioned, this goes back to equation (54), where we increase the iteration count $t$ until $\|\bar{\mathbf{r}} - \bar{\mathbf{M}}\boldsymbol{\alpha}_t\|_2^2 \leq \epsilon$. This is known as the discrepancy principle (Morozov 1968). The threshold $\epsilon$ on the residual norm can be set using the expected error on the data at the 'true' solution $\boldsymbol{\alpha}$,

$$E\|\bar{r} - \bar{\mathbf{M}}\boldsymbol{\alpha}\|_2^2 = E\|\boldsymbol{\Gamma}\boldsymbol{e}\|_2^2 = \mathrm{trace}(\mathrm{Cov}(\boldsymbol{\Gamma}\boldsymbol{e})), \tag{58}$$

where $\boldsymbol{\Gamma}\boldsymbol{e}$ is the whitened error on the data, and Cov$(\cdot)$ denotes the covariance. Note that, by definition of $\boldsymbol{\Gamma}$,

$$\mathrm{Cov}(\boldsymbol{\Gamma}\boldsymbol{e}) = E\{\boldsymbol{\Gamma}\boldsymbol{e}\boldsymbol{e}^H\boldsymbol{\Gamma}^H\} = \mathbf{I}, \tag{59}$$

where $\mathbf{I}$ is a $P^2 \times P^2$ identity matrix. Therefore, we set $\epsilon = P^2$, or a slightly larger value to account for finite sample noise.

If the autocorrelations of the measurements are not available, we need to resort to the unweighted LS estimator ($\boldsymbol{\Gamma} = \mathbf{I}$). The stopping criteria are based on an estimate of the noise on the visibilities.

### 5.3 Implementation details

As mentioned earlier, the system matrix $\bar{\mathbf{M}}$ is a full matrix. However it exhibits a 'data sparse' structure. Since $\mathbf{M} = \mathbf{A}^* \circ \mathbf{A}$, we can represent the system matrix $\mathbf{M}$ of dimension $P^2 \times Q$ with a lower dimensional matrix $\mathbf{A}$ of dimension $P \times Q$. In the case of $\bar{\mathbf{M}}$ we need to apply the proper right and left preconditioners to $\mathbf{A}$. Considering the Cholesky factorization

$$\hat{\mathbf{R}}^{-1} = \mathbf{B}^H\mathbf{B}, \tag{60}$$

we find $\boldsymbol{\Gamma} = N^{1/2}(\mathbf{B}^* \otimes \mathbf{B})$ and therefore $\bar{\mathbf{M}} = \bar{\mathbf{A}}^* \circ \bar{\mathbf{A}}$ with $\bar{\mathbf{A}} := N^{\frac{1}{4}}\mathbf{B}\mathbf{A}\mathbf{L}^{-\frac{1}{2}}$. If the dimensions of the imaging problem are such that we can store matrix $\bar{\mathbf{A}}$ in memory, we implement the matrix vector operations $\bar{\mathbf{M}}\boldsymbol{v}$ and $\bar{\mathbf{M}}^H\boldsymbol{u}$ as

$$\begin{aligned}\bar{\mathbf{M}}\boldsymbol{v} &= \mathrm{vect}(\bar{\mathbf{A}}\mathrm{diag}(\boldsymbol{v})\bar{\mathbf{A}}^H),\\ \bar{\mathbf{M}}^H\boldsymbol{u} &= \mathrm{vectdiag}(\bar{\mathbf{A}}^H\mathbf{U}\bar{\mathbf{A}}) = [\bar{\mathbf{a}}_i^H\mathbf{U}\bar{\mathbf{a}}_i]_{i=1}^{Q},\end{aligned} \tag{61}$$

where $\mathrm{diag}(\boldsymbol{v})$ is a diagonal matrix with the vector $\boldsymbol{v}$ on its main diagonal, vectdiag$(\cdot)$ selects the diagonal of a matrix and stores it in a vector, and $\mathbf{U}$ is a $P \times P$ matrix such that $\boldsymbol{u} = \mathrm{vect}(\mathbf{U})$. The diagonal matrices are stored in a sparse manner for memory considerations.

If the dimensions of $\mathbf{A}$ are also higher than the available physical memory, the matrix–vector multiplications can be directly implemented through the function representation of matrix $\mathbf{M}$ as denoted in equation (6), or more efficiently through the W-projection algorithm or its various implementations (Cornwell, Golap & Bhatnagar 2008).

### 5.4 Computational complexity of PRIFIRA

As can be seen from the description of Algorithm 1, the computational complexity of PRIFIRA is dominated by the two matrix–vector multiplications $\bar{\mathbf{M}}\boldsymbol{v}_t$ and $\bar{\mathbf{M}}^H\boldsymbol{u}_t$. Therefore, the implementation of these matrix–vector multiplications determines the computational complexity of the algorithm. The first operation is in fact equivalent to computing correlation data from an image (sky model) using the measurement equation, while the second corresponds to the computation of an MF dirty image from correlation data. These are standard operations in any radio astronomy imaging toolbox, and many fast algorithms (based on gridding and FFTs) have been proposed and implemented.

Assuming that no fast transform is used to obtain the matrix–vector multiplications, the complexity of computing $\bar{\mathbf{M}}\boldsymbol{v}_t$ is $\mathcal{O}(P^2Q)$, and computing $\bar{\mathbf{M}}^H\boldsymbol{u}_t$ has the same complexity. The complexity of PRIFIRA is thus $\mathcal{O}(T\ P^2Q)$ where $T$ denotes the required number of iterations until the stopping criteria are satisfied. Simulations indicate that $T$ is usually quite small (around 5 to 10). In case of the reweighted PRIFIRA, the complexity increases to $\mathcal{O}(KT\ P^2Q)$, where $K$ is the total number of reweighted outer iterations.

To compare this complexity to the existing imaging algorithms, we first note that all of them require basic operations of the form $\mathbf{M}\boldsymbol{v}$ (a forward step, computing correlation data from a sky model) and $\mathbf{M}^H\boldsymbol{u}$ (a backward step, computing a dirty image from correlation data), and often these are implemented efficiently using gridding, FFTs, and W-projections.

Existing algorithms can be classified into (i) greedy algorithms such as CLEAN and NNLS, which require an exhaustive search over all the pixels of the image to find potential sources; and (ii) compressed-sensing based algorithms, implemented using convex optimization, such as SARA implemented using Alternating Direction Method of Multipliers (ADMM). The first category requires a large number of iterations where the basic work is comparable to PRIFIRA, although significant reductions are possible by utilizing multiresolution and image partitioning techniques. NNLS also requires a number of subiterations for each iteration to solve a system of equations and is therefore more costly. These algorithms are sensitive to grid mismatch due to misalignment of the sources with the grid points.

Like PRIFIRA, compressed sensing algorithms consider the complete image at once and therefore are less sensitive to grid mismatch. CS algorithms are based on gradient descent steps which in the end boil down to matrix–vector multiplications with $\mathbf{M}$ and $\mathbf{M}^H$. Less costly subiterations and (non-linear) outer loops are used to satisfy positivity and sparsity constraints through proximity operators. While the amount of work per iteration is therefore comparable, the simulations presented in Section 6 show that the prior-conditioning used by PRIFIRA provides an order of magnitude faster convergence. A good estimate of the image is already obtained after a few iterations, resulting in major savings on the overall cost of the imaging algorithm.

## 6 SIMULATIONS AND EXPERIMENTAL RESULTS

### 6.1 Terminology

We proposed several variants of the PRIFIRA algorithm, based on the initial prior-conditioners and the optional use of reweighting iterations. We therefore indicate the right prior-conditioner as a prefix to the name of the algorithm; i.e. *X*-PRIFIRA where *X* indicates the prior-conditioner. We consider *X* as MF, MVDR, IR0 and IR1 where MF and MVDR respectively denote the matched filtered and MVDR dirty images, and IR0 and IR1 indicate the iteratively reweighted PRIFIRA resulting in $\ell_0$ and $\ell_1$ image norm minimizations respectively, as discussed in Section 4.2. For comparison and to show the effect of right preconditioners on the reconstruction quality, we also consider the LSQR algorithm which is equivalent to PRIFIRA when there is no right preconditioner applied.

We compare variants of PRIFIRA with the MATLAB implementations of some of the state-of-the-art algorithms. Among the greedy sparse reconstruction methods we use the NNLS optimization implemented using the active set algorithm as discussed in Sardarabadi et al. (2016). The CLEAN algorithm (Högbom 1974) is implemented in MATLAB with both minor cycles and occasional major cycles. MEM is implemented based on Newton–Raphson iterations. Among the compressed sensing techniques based on convex optimization we focus on $\ell_1$ norm minimization and the SARA formalism (Carrillo et al. 2012) implemented based on the ADMM (Boyd 2011). Furthermore, we compare the results with the conventional deconvolution method of Richardson–Lucy (RL; Richardson 1972).

We mention that there is no shortage of RA imaging algorithms and it is not possible to compare the proposed method with all the implementations of the present methods. Therefore, we have categorized the imaging methods and compare our algorithm with the basic implementation of the main methods. It is also worth noting that many of the imaging methods have been optimized both in software and hardware to perform faster. There are many possibilities to also optimize PRIFIRA in the future but for the current paper we focus on the most basic implementation and for a fair comparison compare it with basic implementations of the state-of-the-art algorithms.

### 6.2 One-dimensional tests

We first demonstrate the effects of prior-conditioning using a 1D test example. For this simulation, we use a non-uniform linear array with $P = 10$ elements as shown in Fig. 1(a). The conditioning of matrix $\mathbf{M}$ is shown via its singular value spectrum in Fig. 1(b). Two Gaussian sources with the same height 2 and different width positioned at direction cosines $l = -0.5$ and $l = 0.5$ are used to model resolved and unresolved sources, respectively. The sources are shown in Fig. 1(c). We discretize the line 'image' into $Q = 201$ pixels. The operating frequency is set to 80 MHz and the covariance data contains correlated noise with power 100. The correlation data $\hat{\mathbf{R}}$ is constructed from $N = 10^5$ samples.

The PSF of the antenna array is shown in Fig. 1(d) and as can be seen contains large sidelobes. We can see based on the PSF that the left Gaussian source, with a width larger than the main beam, can be considered as a resolved source and the right Gaussian, that is significantly narrower than the main beam, can be considered as an unresolved source. The MF and the MVDR dirty images are plotted in Fig. 1(e). As can be seen, due to the large noise power, the MF and MVDR images are relatively close. Fig. 1(f),(g),(h),(i), and (j), respectively, show the reconstruction results for LSQR, MF-PRIFIRA, MVDR-PRIFIRA, IR1-PRIFIRA, and IR0-PRIFIRA. The total number of iterations until the stopping criteria are achieved for LSQR, MF-PRIFIRA, MVDR-PRIFIRA, IR1-PRIFIRA, and IR0-PRIFIRA are 4, 3, 3, 60, and 60, respectively. 20 outer iterations are used for IR1- and IR0-PRIFIRA. For IR0-PRIFIRA, the non-zero coefficients in $\hat{\boldsymbol{\alpha}}_k$ converge to 1 after 20 iterations. Therefore, the solution $\hat{\boldsymbol{\sigma}}$ becomes invariant with the increase of outer iterations. We have added two more reconstructions based on IR1-PRIFIRA. Fig. 1(k) is a modified version of IR1-PRIFIRA such that the prior-conditioning weights are computed as $\mathbf{W}_k^{-1} = \mathrm{diag}(\hat{\boldsymbol{\sigma}}_{k-1}^{1/2} + \epsilon)$ with $\epsilon = 0.2$. Similar modifications are proposed in Chartrand & Yin (2008) and Daubechies et al. (2010) for stability reasons. The number of outer iterations is kept at 20 for this result. Furthermore, Fig. 1(l) shows an extreme case of IR1-PRIFIRA with 100 outer iterations.

The figure shows that the LSQR reconstructed image has many sidelobes, some of which are negative. MF-PRIFIRA and MVDR-PRIFIRA stabilize the solution such that the sidelobes disappear to a large extent with MVDR-PRIFIRA being more successful in this regard. Both MF-PRIFIRA and MVDR-PRIFIRA recover the resolved source reliably and smear out the unresolved source as was expected from the third argument in Section 4.2. IR1-PRIFIRA attempts to narrow the smearing while IR0-PRIFIRA aims for an optimally sparse and spiky solution which is not the preferred solution in cases where retrieving extended emissions are of interest. The modified version of IR1-PRIFIRA is more faithful to the recovery of the extended emission while smearing the unresolved source. In the extreme case, IR1-PRIFIRA recovers the unresolved source almost

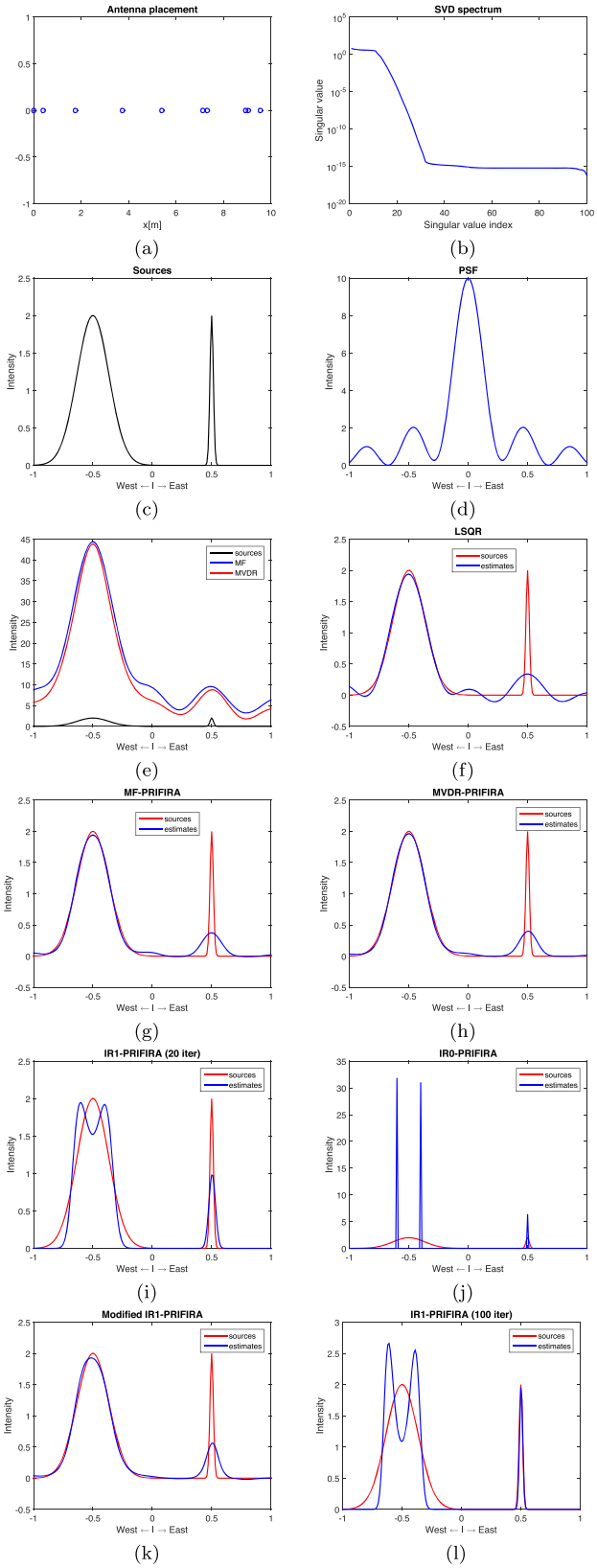


**Figure 1.** (a) Antenna positions, (b) singular value distribution of $\mathbf{M}$, (c) source distribution, (d) PSF, (e) MF and MVDR dirty images, (f) LSQR reconstruction, (g) MF-PRIFIRA reconstruction, (h) MVDR-PRIFIRA reconstruction, (i) IR1-PRIFIRA reconstruction, (j) IR0-PRIFIRA reconstruction, (k) modified IR1-PRIFIRA reconstruction, (l) IR1-PRIFIRA reconstruction after 100 outer iterations.

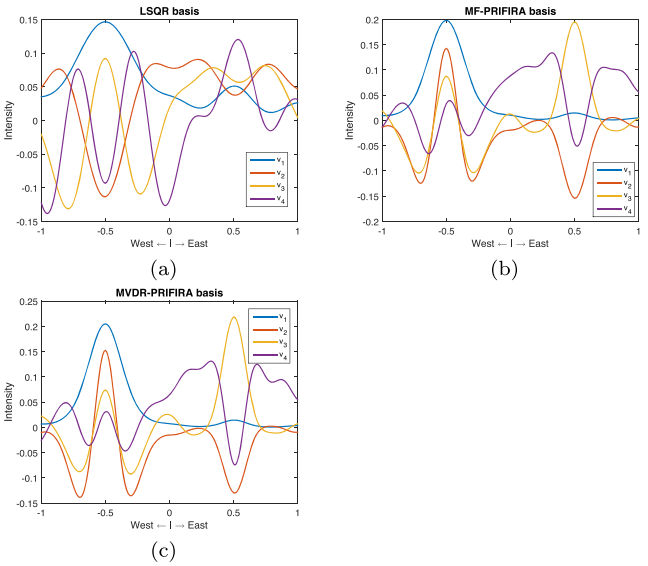


**Figure 2.** (a) LSQR basis vectors, (b) MF-PRIFIRA basis vectors, and (c) MVDR-PRIFIRA basis vectors.

perfectly while narrowing the extended emission into two peaks similar to the effect observed with the recovery of IR0-PRIFIRA. Both IR1- and IR0-PRIFIRA do not observe the natural resolution of the instrument while MF and MVDR-PRIFIRA maintain this resolution.

In Appendix B, we discuss the effect of removing the autocorrelations analytically and based on a simulation for a 1D scenario.

We now look more closely into the Krylov basis vectors produced by the various algorithms. As mentioned in Section 5, Krylov subspace-based methods restrict the solution space to the first $t$ Krylov vectors. When applying the LSQR algorithm, the Krylov vectors are reorthogonalized as Lanczos vectors indicated by $\boldsymbol{v}_t$ at iteration $t$. Therefore, the solution space is spanned by $[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_t]$. It is informative to look at the Lanczos vectors with and without the application of prior-conditioners. We show these effects for the simple 1D test case. Fig. 2 shows the first four initial Lanczos vectors. It is seen that the LSQR basis has a non-zero support where the true image is zero while MF- and MVDR-PRIFIRA bases capture the support of the image already in the initial Lanczos vectors. This indicates that the latter bases provide a good space to represent the image.

### 6.3 Tests on model images

The proposed methods are now tested on noisy simulated data using the configuration of antennas from the core stations of the LOFAR telescope array. As test image, we consider an image of the W28 supernova remnant, shown in Fig. 3(a), obtained from https://casaguides.nrao.edu/index.php. The core stations contain $P = 273$ antennas with a maximum baseline length of about 326 m as shown in Fig. 3(b). The operating frequency is chosen as 58.975 MHz and a single time snapshot is considered. The u-v coverage of the telescope array is shown in Fig. 3(c). Fig. 3(d) illustrates the PSF of the array showing the limited resolution of the array and the existence of sidelobes. To construct the sampled covariance matrix, $\mathbf{R}$ is generated from the test image, $\mathbf{R}^{1/2}$ is used to shape white Gaussian noise into data vectors $\boldsymbol{x}[n]$ that has the required covariance structure, and white Gaussian receiver noise with variance $\sigma_n^2 = 4$ is added. $N = 10^5$ data samples $\boldsymbol{x}[n]$ are used

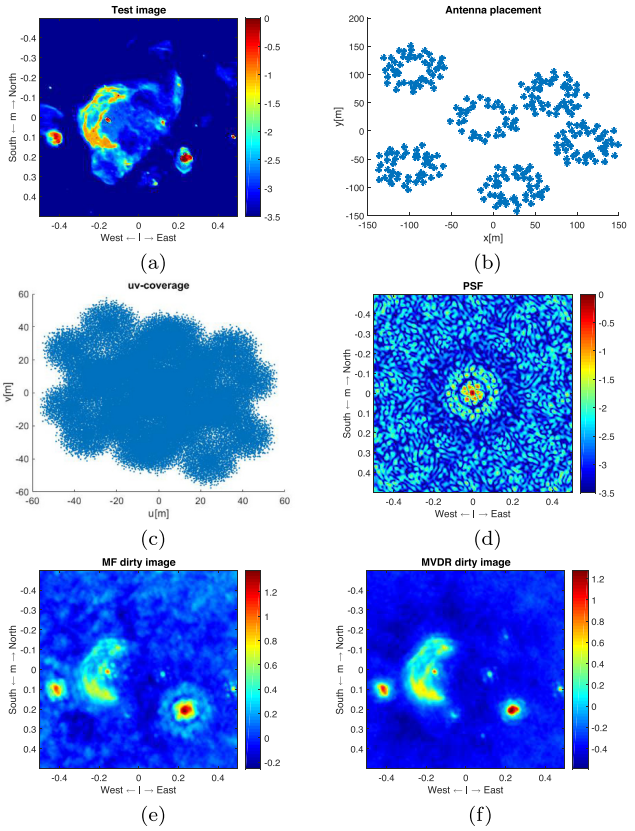


**Figure 3.** (a) Test image in dB, (b) antenna placement, (c) $u$-$v$ coverage, (d) normalized PSF in dB, (e) MF dirty image in dB, and (f) MVDR dirty image in dB.

to construct $\hat{\mathbf{R}}$. The image is discretized into $Q$= 84 681 pixels. The dirty image obtained from the matched filtered beamformer is shown in Fig. 3(e), and the MVDR dirty image is shown in Fig. 3(f). The simulations were performed in MATLAB R2014b on a computer with Intel i5-4670 CPU 3.40 GHz under 64-bit Windows 7 with an 8 GB RAM. The images are shown in logarithmic scale and for demonstration and for comparison reasons are limited to scales in the range 1–$10^{-3.5}$.

Fig. 4 compares the reconstructed images for the various imaging algorithms. Fig. 4(a) and (b), respectively, show the CLEAN and NNLS (Sardarabadi et al. 2016) reconstructions after applying post-processing with a Gaussian main beam which was fitted to the PSF. 10 major cycles of 500 minor cycles are chosen for running the CLEAN algorithm. Fig. 4(c) shows the ADMM reconstruction with an $\ell_1$ sparsity constraint. Fig. 4(d) is the reconstructed image based on the SARA formalism (Carrillo et al. 2012), implemented with ADMM. Fig. 4(e) is the reconstruction based on the Richardson–Lucy algorithm (Richardson 1972). Fig. 4(f) is the maximum entropy reconstruction (Cornwell & Evans 1985) based on the implementation (Hanke, Nagy & Vogel 2000). This method is very sensitive to the choice of the regularization parameter and the starting vector, and we chose the scaled MF dirty image as the starting vector. Figs 4(g),(h), (i), (j), and (k) show the results for LSQR, MF-PRIFIRA, MVDR-PRIFIRA, IR1-PRIFIRA, and IR0-PRIFIRA, respectively. Five outer iterations are chosen for IR0-PRIFIRA and IR1-PRIFIRA.

Qualitatively, Fig. 4 shows that CLEAN and NNLS have less resolution than the other methods due to the correction with the main

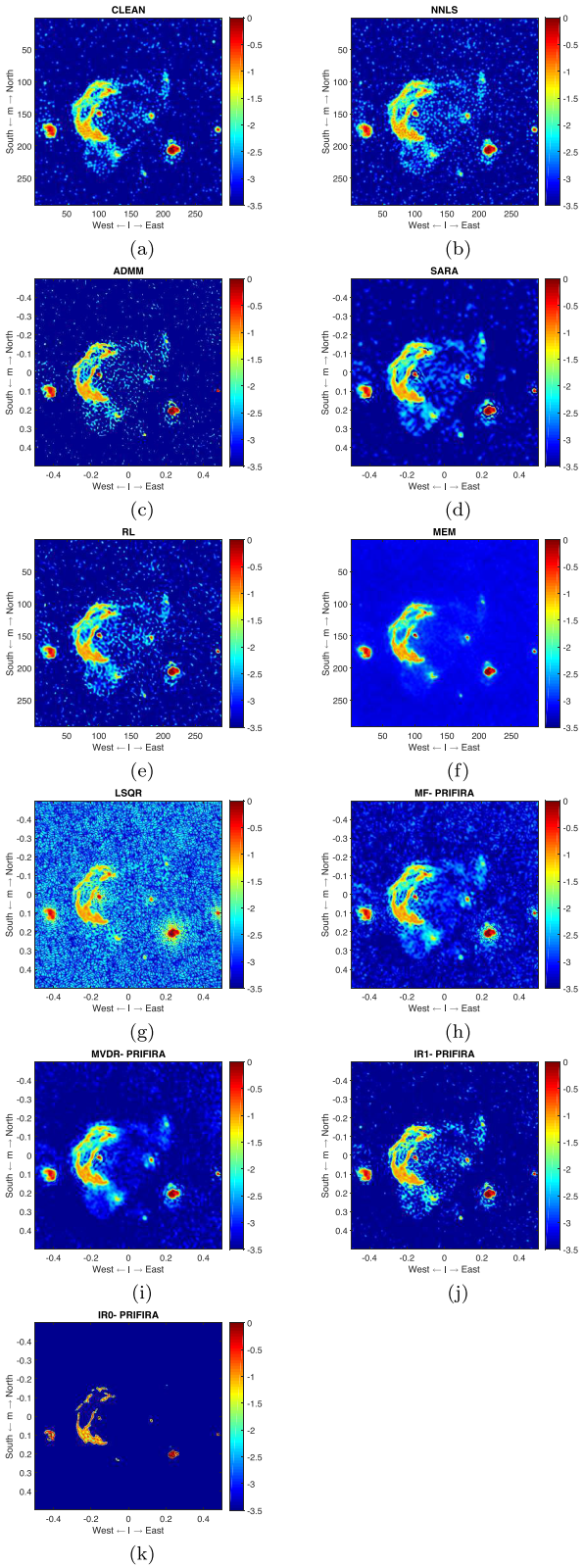


**Figure 4.** (a) CLEAN, (b) NNLS, (c) ADMM, (d) SARA, (e) RL, (f) MEM, (g) LSQR, (h) MF-PRIFIRA, (i) MVDR-PRIFIRA, (j) IR1-PRIFIRA, and (k) IR0-PRIFIRA.

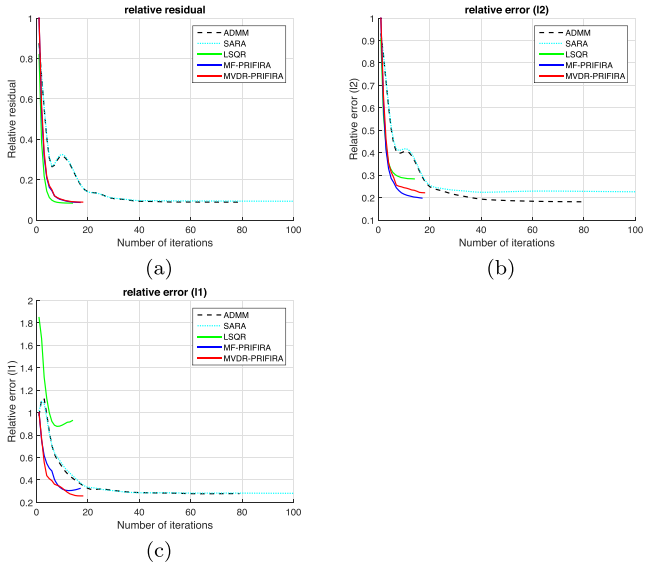


**Figure 5.** (a) Relative residual per iteration, (b) $\ell_2$ norm error, and (c) $\ell_1$ norm error.

beam. MEM results in a rather 'flat' image outside the area of significant emission. LSQR has significant remaining side lobes (ringing effects indicating insufficient regularization), and MF-PRIFIRA also shows this to a lesser extent. MVDR-PRIFIRA and IR1-PRIFIRA are comparable, although the latter results in a 'sharper' image due to the imposed sparsity. IR0-PRIFIRA has converged to a very sparse solution, indicating it is not suitable to capture extended structures. These observations are consistent with the 1-D case.

The convergence in terms of relative residual, $\ell_1$-norm error and $\ell_2$-norm error per iteration are compared in Fig. 5. The relative $\ell_i$-norm error for $i = 1, 2$ at iteration $t$, $e_{i,t}$, is defined as

$$e_{i,t} = \frac{\|\hat{\boldsymbol{\sigma}}_t - \boldsymbol{\sigma}\|_i}{\|\boldsymbol{\sigma}\|_i}, \tag{62}$$

where $\boldsymbol{\sigma}$ is the model true image and $\hat{\boldsymbol{\sigma}}_t$ is the reconstructed image at iteration $t$. We use $e_{1,t}$ as an indicator of how accurately the algorithm is capable of retrieving the source positions as well as the intensities whereas $e_{2,t}$ is mostly concerned with retrieving the correct overall intensity in the image. ADMM, LSQR, MF-PRIFIRA, and MVDR-PRIFIRA are shown in blue, black, and red graphs, respectively. For comparison reasons, LSQR, MF-PRIFIRA, and MVDR-PRIFIRA are run beyond the stopping threshold. The figure shows that methods based on LSQR exhibit a substantially faster convergence than steepest descent-based ADMM while maintaining comparable reconstruction quality.

The performance of the imaging algorithms is summarized in Table 1, which shows the number of iterations, reconstruction time, and error norm for the considered algorithms. The table shows that methods based on LSQR and PRIFIRA with one iteration level (i.e. no outer iterations), namely MF- and MVDR-PRIFIRA, exhibit greatly reduced number of iterations and reconstruction time. We can see that SARA, ADMM, and RL exhibit good reconstruction qualities but are considerably slower than the PRIFIRA-based methods. Among the PRIFIRA-based methods IR1-PRIFIRA, MF-PRIFIRA, and MVDR-PRIFIRA exhibit the best reconstruction quality.

**Table 1.** Performance summary.

| | # iterations | Reconstruction time (s) | $\|\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|_2$ | $\|\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|_1$ |
|---|---|---|---|---|
| CLEAN | 5000 | 69.53 | 2.86 | 116.7 |
| NNLS | 2656 | 11.88 (h) | 3.08 | 124.8 |
| ADMM | 79 | 134.61 | 1.45 | 3.01 |
| SARA | 200 | 353.66 | 1.76 | 2.85 |
| RL | 200 | 342.9 | 1.7 | 2.61 |
| MEM | 30 | 43.9 (min) | 1.8 | 94.4 |
| LSQR | 12 | 22.57 | 2.27 | 9.8 |
| MF-PRIFIRA | 15 | 35.08 | 1.57 | 3.5 |
| MVDR-PRIFIRA | 15 | 32.57 | 1.76 | 2.79 |
| IR1-PRIFIRA | 66 | 128 | 1.34 | 2.53 |
| IR0-PRIFIRA | 84 | 165.45 | 6.42 | 8.34 |

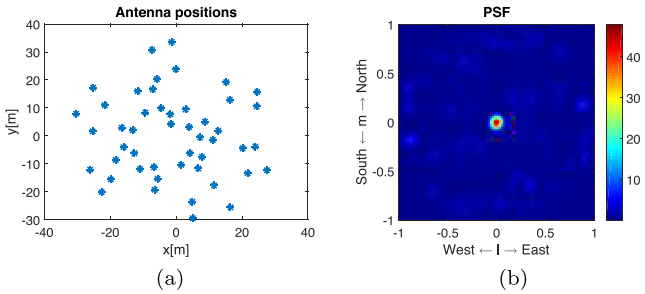


**Figure 6.** (a) LOFAR single station antenna position and (b) PSF.

### 6.4 Tests on real data

Next, we test the proposed imaging algorithm on measured correlation data from a single LOFAR station. The data set as introduced in Wijnholds & Van der Veen (2011) and Wijnholds (2010) is used. The station consists of an array of 48 antennas as shown in Fig. 6(a) and the related full-sky PSF is shown in Fig. 6(b). An observation from a single 10 s snapshot at frequency 50.3125 MHz is considered to construct an image with $Q = 8937$ pixels. The normalized MF and MVDR images are shown in Fig. 7(a) and (b). The power of the additive noise on the antennas is unknown, and we compute an estimate of it as

$$\hat{\boldsymbol{\sigma}}_{\boldsymbol{n}} = |\hat{\mathbf{R}}^{-1}|^{-\odot 2}\text{vectdiag}(\hat{\mathbf{R}}^{-1}), \tag{63}$$

as discussed in Wijnholds (2010), where the notation $|\cdot|^{-\odot 2}$ denotes entrywise taking the absolute value, inverting, and squaring. MF and MVDR images are computed based on the noise corrected covariance data $\hat{\mathbf{R}} - \mathbf{R}_{\boldsymbol{n}}$. The reconstruction results are shown in Fig. 7. Since the ground truth of the sky map is unknown with the real data, we show the residual image after reconstruction as is customary in radio astronomical society. The residual image is computed as

$$\boldsymbol{\delta} = \mathbf{M}^H(\tilde{\boldsymbol{r}} - \mathbf{M}\hat{\boldsymbol{\sigma}}), \tag{64}$$

where $\boldsymbol{\delta}$ indicates the residual image and $\hat{\boldsymbol{\sigma}}$ is the estimated image.

Figs 7(c), (e), and (g), respectively, show the reconstruction results for LSQR, MF-PRIFIRA, and MVDR-PRIFIRA after seven iterations and Figs 7(d), (f), and (h) show the corresponding residual images. Figs 7(i) and (k), respectively, show the reconstructed images using IR1-PRIFIRA and IR0-PRIFIRA after three inner iterations and four outer iterations and Figs 7(j) and (l) are the corresponding residual images. The image scales on the residual images are cropped at $[-0.2, 0.2]$ for ease of comparison.

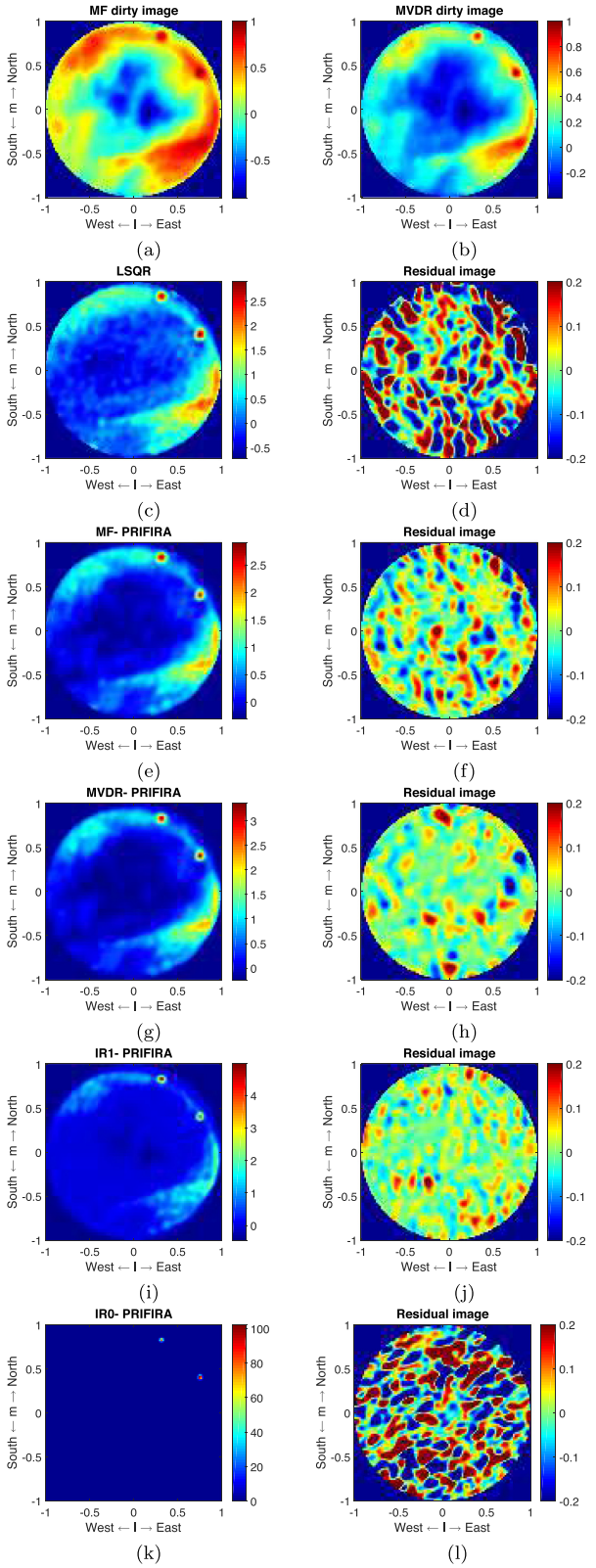


**Figure 7.** (a) MF, (b) MVDR, (c) LSQR, (d) LSQR residual image, (e) MF-PRIFIRA, (f) MF-PRIFIRA residual image, (g) MVDR-PRIFIRA, (h) MVDR-PRIFIRA residual image, (i) IR1-PRIFIRA, (j) IR1-PRIFIRA residual image, (k) IR0-PRIFIRA, and (l) IR0-PRIFIRA residual image.

Since the ground truth is not known in this case, we use the LS image generated from combining 25 frequency channels and 10 s integration per channel as discussed in Wijnholds (2010) as a reference. The bright sources are identified as Cyg A and Cas A and the presence of a Galactic loop emerging from Cyg A is identified as loop III in the Haslam survey (Wijnholds 2010). Most of the middle and west part of the image do not contain recognizable emissions. This example is interesting as the data contains the contribution from both point sources as well as extended emissions. It is worth noting that we only use data from one frequency and snapshot to test our algorithm. We can see from the LSQR reconstruction in Fig. 7(c) that there is considerable excess power in the middle and west part of the image due to the instability of the solution. MF-PRIFIRA and MVDR-PRIFIRA correct to a large extent for the faulty reconstruction in the middle and west part of the image while reducing the residual power with MVDR-PRIFIRA showing the smallest and smoothest residual image. IR1-PRIFIRA seems to capture most of the relevant emissions with a similar residual level. However, IR0-PRIFIRA only captures the point sources discarding the extended emissions as predicted and is more appropriate for images with only point sources.

## 7 CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced an algorithmic framework to efficiently solve radio astronomical imaging problems with a focus on the recovery of extended emissions. An initial image based on beamforming techniques is used to regularize the Maximum Likelihood image estimation problem by means of prior (or right) preconditioning. We further generalize the proposed framework to also handle images with sparsity priors. To achieve an efficient implementation, we have proposed the use of Krylov subspace methods. We call the algorithmic framework PRIFIRA which consists of different variants referring to the type of regularization applied.

We have compared the performance of PRIFIRA with several state-of-the-art imaging algorithms and have shown the computational savings and improvements in accuracy of the estimations. In particular, prior-conditioning using an MF or MVDR dirty image is seen to provide very fast convergence of the Krylov iterations to a solution that has comparable reconstruction quality as the state-of-the art methods at a significantly reduced computational cost.

An initial PYTHON implementation of the algorithm is being made with the goal to be included in the LOFAR imaging software and benefit from the existing fast implementations of the gridding and degridding operators.

## ACKNOWLEDGEMENTS

This work was supported in part by the NWO DRIFT project (contract 628.002.002). The authors would like to thank Ahmad Mouri Sardarabadi for discussions on an initial idea captured in Naghibzadeh et al. (2016b) which evolved into this paper, Patrick Dewilde for in-depth discussions and comments regarding the algorithm and the paper, Martin van Gijzen for discussions regarding the iteratively reweighted algorithms and Stefan Wijnholds for providing the LOFAR station data.

## REFERENCES

Berisha S., Nagy J. G., 2013, in Chellappa R., Theodoridis S., eds, Academic Press Library in Signal Processing, Vol. 4, Image, Video Processing and

Analysis, Hardware, Audio, Acoustic and Speech Processing. Academic Press, New York, p. 193
Bertero M., Boccacci P., 1998, Introduction to Inverse Problems in Imaging. Institute of Physics Publishing, London
Boyd S., 2011, in Talk at NIPS Workshop on Optimization and Machine Learning, Vol. 4, Distributed Optimization via the Alternating Direction Method of Multipliers, NV, USA.
Briggs D. S., 1995, PhD thesis, The New Mexico Institute of Mining and Technology
Bruckstein A. M., Donoho D. L., Elad M., 2009, SIAM Rev., 51, 34
Calvetti D., Somersalo E., 2005, Inverse Probl., 21, 1397
Carrillo R., McEwen J., Wiaux Y., 2012, MNRAS, 426, 1223
Carrillo R. E., McEwen J. D., Wiaux Y., 2014, MNRAS, 439, 3591
Chartrand R., Yin W., 2008, in IEEE International Conference on Acoustics, Speech and Signal Processing, 2008 (ICASSP 2008): Iteratively Reweighted Algorithms for Compressive Sensing, IEEE, Las Vegas. p. 3869.
Choi S.-C. T., 2006, PhD thesis, Stanford University
Combettes P. L., Pesquet J.-C., 2011, in Bauschke H. H., Burachik R. S., Combettes P. L. , Elser V. , Luke D. R. , Wolkowicz H., eds, Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer, New York, p. 185
Cornwell T. J., 2008, IEEE J. Sel. Top. Signal Process., 2, 793
Cornwell T. J., Evans K., 1985, A&A, 143, 77
Cornwell T. J., Golap K., Bhatnagar S., 2008, IEEE J. Sel. Top. Signal Process., 2, 647
Dabbech A., Ferrari C., Mary D., Slezak E., Smirnov O., Kenyon J., 2014, A&A, 576
Daubechies I., DeVore R., Fornasier M., Güntürk C. S., 2010, Commun. Pure Appl. Math., 63, 1
Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L., 2009, Proc. IEEE, 97, 1482
Girard J. et al., 2015, in 2nd Int. Summer School on Intelligent Signal Processing for Frontier Research and Industry. IOP Publishing, Univ. Paris-Diderot Campus, Paris, France
Golub G., Kahan W., 1965, J. Soc. Ind. Appl. Math. B, 2, 205
Gorodnitsky I. F., Rao B. D., 1997, IEEE Trans. Signal Proc., 45, 600
Hanke M., 1995, Conjugate Gradient Type Methods for Ill-posed Problems. Vol. 327, Longman Group Limited, Great Britain
Hanke M., Nagy J. G., Vogel C., 2000, Linear Algebr. Appl., 316, 223
Hansen P. C., 2005, Rank-Deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion. Vol. 4, SIAM, Philadelphia
Hansen P. C., , 2010, Discrete Inverse Problems: Insight and Algorithms. Vol. 7, SIAM, Philadelphia
Högbom J., 1974, A&AS, 15, 417
Jensen T. K., Hansen P. C., 2007, BIT Numer. Math., 47, 103
Jongerius R., Wijnholds S., Nijboer R., Corporaal H., 2014, IEEE Comput., 479, 48
Junklewitz H., Bell M., Selig M., Enßlin T., 2016, A&A, 586, A76
Kaipio J. P., Somersalo E., 2004, Statistical and Computational Inverse Problems. Applied Mathematical Sciences Vol. 160, Springer, New York
Kay S. M., 1993, Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory. Prentice Hall, Upper Saddle River, NJ
Krim H., Viberg M., 1996, IEEE Signal Process. Mag., 13, 67
Lawson C. L., Hanson R. J., 1974, Solving Least Squares Problems. Vol. 161, SIAM, Philadelphia
Leshem A., Van der Veen A.-J., 2000, IEEE Trans. Inf. Theory, 46, 1730
Marsh K., Richardson J., 1987, A&A, 182, 174
McEwen J., Wiaux Y., 2011, MNRAS, 413, 1318
Morozov V. A., 1968, Zh. vychisl. mat. mat. fiz, 8, 295
Naghibzadeh S., van der Veen A.-J., 2017a, in 2017 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, Curacao, Netherlands. p. 173
Naghibzadeh S., van der Veen A.-J., 2017b, in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, New Orleans. p. 3385
Naghibzadeh S., Sardarabadi A. M., van der Veen A.-J., 2016a, in Sensor Array and Multichannel Signal Processing Workshop (SAM), 2016 IEEE, IEEE, Rio de Janerio, p. 1
Naghibzadeh S., Sardarabadi A. M., van der Veen A.-J., 2016b, in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Shanghai, China, p. 3316
Nemirovskii A. S., 1986, USSR Comput. Math. Math. Phys., 26, 7
Offringa A. R., Smirnov O., 2017, MNRAS, 471, 301
Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J.-P., Pesquet J.-C., Wiaux Y., 2016, MNRAS, 462, 4314
Ottersten B., Stoica P., Roy R., 1998, Digit. Signal Process., 8, 185
Paige C. C., Saunders M. A., 1982, ACM Trans. Math. Softw., 8, 43
Rau U., Cotton T. J., 2011, A&A, 532, A71
Richardson W. H., 1972, J. Opt. Soc. Am., 62, 55
Saad Y., 1981, Math. Comput., 37, 105
Sardarabadi A. M., Leshem A., van der Veen A.-J., 2016, A&A, 588, A95
Schwab F., 1984, AJ, 89, 1076
Tipping M. E., 2001, J. Mach. Learn. Res., 1, 211
van der Veen A.-J., Wijnholds S. J., 2013, in Bhattacharyya S. S., Deprettere E. F., Leupers R., Takala J., eds, Handbook of Signal Processing Systems. Springer, New York, p. 421
van der Veen A.-J., Leshem A., Boonstra A.-J., 2005, in Hall P. J., ed., The Square Kilometre Array: An Engineering Perspective. Springer, Dordrecht, p. 231
Van Haarlem M. et al., 2013, A&A, 556, A2
Wakker B., Schwarz U., 1988, A&A, 200, 312
Wiaux Y., Jacques L., Puy G., Scaife A., Vandergheynst P., 2009, MNRAS, 395, 1733
Wijnholds S. J., 2010, Fish-Eye Observing with Phased Array Radio Telescopes. TU Delft, Delft University of Technology
Wijnholds S. J., van der Veen A.-J., 2008, IEEE J. Sel. Top. Signal Process., 2, 613
Wijnholds S. J., Van der Veen A.-J., 2011, in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Prague, Czech Republic, p. 2704
Wipf D. P., Rao B. D., 2004, IEEE Trans. Signal Proc., 52, 2153

## APPENDIX A: PROOF OF (14)

It is known that (Sardarabadi et al. 2016)

$$\boldsymbol{a}_i^H \mathbf{R} \boldsymbol{a}_i \geq \frac{1}{\boldsymbol{a}_i^H \mathbf{R}^{-1} \boldsymbol{a}_i}.$$

Therefore, it is sufficient to prove that

$$\boldsymbol{a}_i^H \mathbf{R}_n \boldsymbol{a}_i \leq \frac{\boldsymbol{a}_i^H \mathbf{R}^{-1} \mathbf{R}_n \mathbf{R}^{-1} \boldsymbol{a}_i}{(\boldsymbol{a}_i^H \mathbf{R}^{-1} \boldsymbol{a}_i)^2}.$$

This is the same as

$$\boldsymbol{a}_i^H \mathbf{R}^{-1} \boldsymbol{a}_i \cdot \boldsymbol{a}_i^H \mathbf{R}_n \boldsymbol{a}_i \cdot \boldsymbol{a}_i^H \mathbf{R}^{-1} \boldsymbol{a}_i \leq \boldsymbol{a}_i^H \mathbf{R}^{-1} \mathbf{R}_n \mathbf{R}^{-1} \boldsymbol{a}_i.$$

For this to hold, it is sufficient if

$$\boldsymbol{a}_i \boldsymbol{a}_i^H \mathbf{R}_n \boldsymbol{a}_i \boldsymbol{a}_i^H \leq \mathbf{R}_n.$$

While this is not true in general, the relation holds if $\boldsymbol{a}_i$ is an eigenvector of $\mathbf{R}_n$, which includes special cases such as $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$, this relation holds.

## APPENDIX B: EFFECT OF MISSING AUTOCORRELATIONS

Traditionally in radio astronomy, the autocorrelations are not measured or are discarded for image formation, as they are considered inaccurate due to the addition of large noise power terms. We briefly discuss the effect of missing autocorrelations on the proposed method.

If the autocorrelations are not available, we need to change the data model accordingly. We redefine $\tilde{\boldsymbol{r}}$ as

$$\tilde{\mathbf{R}} = \hat{\mathbf{R}} - \text{diag}(\hat{\mathbf{R}}), \qquad \tilde{\boldsymbol{r}} = \text{vect}(\tilde{\mathbf{R}}), \tag{B1}$$

which includes 'zero' entries in place of the missing autocorrelations. It is straightforward to derive that $\tilde{\boldsymbol{r}}$ is related to $\hat{\boldsymbol{r}}$ as

$$\tilde{\boldsymbol{r}} = \boldsymbol{\Pi}\hat{\boldsymbol{r}}, \tag{B2}$$

where

$$\boldsymbol{\Pi} = \mathbf{I}_{P^2} - (\mathbf{I}_p \circ \mathbf{I}_p)(\mathbf{I}_p \circ \mathbf{I}_p)^H \tag{B3}$$

is an orthogonal projection matrix that projects out the diagonal entries from $\hat{\boldsymbol{r}}$. The resulting data model is

$$\tilde{\boldsymbol{r}} = \tilde{\mathbf{M}}\boldsymbol{\sigma} + \tilde{\boldsymbol{e}},$$

where $\tilde{\mathbf{M}} = \boldsymbol{\Pi}\mathbf{M}$, and $\tilde{\boldsymbol{e}} = \boldsymbol{\Pi}\boldsymbol{e}$ is the finite sample noise, modelled as complex Gaussian with zero mean and variance

$$\tilde{\mathbf{C}}_e = \boldsymbol{\Pi}\mathbf{C}_e\boldsymbol{\Pi} = \frac{1}{N}\boldsymbol{\Pi}(\mathbf{R}^T \otimes \mathbf{R})\boldsymbol{\Pi}.$$

This has a number of consequences:

(i) $\tilde{\boldsymbol{r}}$ does not correspond to a positive (correlation) matrix;

(ii) A straightforward estimate of $\tilde{\mathbf{C}}_e$ is unknown as $\hat{\mathbf{R}}$ is unavailable. Moreover, $\tilde{\mathbf{C}}_e$ is not invertible. Thus, the weight matrix $\boldsymbol{\Gamma}$ in the regularized WLS problem (31) is not available, and we need to resort to the unweighted LS formulation

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\Pi}(\tilde{\boldsymbol{r}} - \mathbf{M}\mathbf{L}^{-1}\boldsymbol{\alpha})\|_2^2 + \tau\|\boldsymbol{\alpha}\|_2^2; \tag{B4}$$

(iii) The MVDR beamformer weights cannot be formed, for the same reason. We used this to form an initial image for the regularization operator $\mathbf{L}$. Instead, we should resort to the MF (classical) dirty image (omitting autocorrelation terms), $\tilde{\boldsymbol{\sigma}}_{\text{MF}} = \tilde{\mathbf{M}}^H\tilde{\mathbf{r}}$ (cf. 11) and set $\mathbf{L}^{-1} = \text{diag}(\tilde{\boldsymbol{\sigma}}_{\text{MF}})$ as a surrogate.

Under the usual assumptions in radio astronomy (noise much stronger than the sources, noise powers have been whitened), it can be argued that the difference between WLS and LS is small, and also the difference between MF and MVDR is small. Alternatively, we can apply diagonal loading and replace $\tilde{\mathbf{R}}$ by $\tilde{\mathbf{R}} + \eta\mathbf{I}$, where $\eta$ is a noise variance estimate.

More important is the fact that $\tilde{r}$ does not correspond to a positive matrix. The resulting MF dirty image $\tilde{\boldsymbol{\sigma}}_{\text{MF}}$ does not have to be positive and sources can have negative sidelobes. Similarly, the PSF, or dirty beam, is defined as $\boldsymbol{b} = \mathbf{M}^H\mathbf{1}$, and becomes $\tilde{\boldsymbol{b}} = \tilde{\mathbf{M}}^H\mathbf{1}$. Since $\mathbf{M} = \mathbf{A}^* \circ \mathbf{A}$, and assuming normalized array response vectors $\|\boldsymbol{a}_i\| = 1$, it can be shown that $\tilde{\boldsymbol{b}} = \boldsymbol{b} - \mathbf{1}$. Thus, also the modified PSF can have negative sidelobes, although it is straightforward to correct this.

The negative sidelobes in $\tilde{\boldsymbol{\sigma}}_{\text{MF}}$ makes this unsuitable to be used as weight in (B4). Some entries in this vector may be close to zero, causing the resulting solution to have a black pixel at that location. Negative values should be avoided by shifting up all the pixels by (at least) the smallest negative value of the sidelobes. If we assume the entries of $\mathbf{A}$ to contain only phases, as in (2), then all entries have equal magnitude, and it is straightforward to show that the difference between the original MF image and the MF image without autocorrelations is a constant, equal to the total neglected power. (This is essentially because the MF dirty beam is spatially invariant.) Thus, to correct the MF dirty image we only have to estimate a single shift common to all the pixels. For MVDR, the PSF is spatially variant and we cannot use a single common shift to obtain the MVDR image where autocorrelations are available.

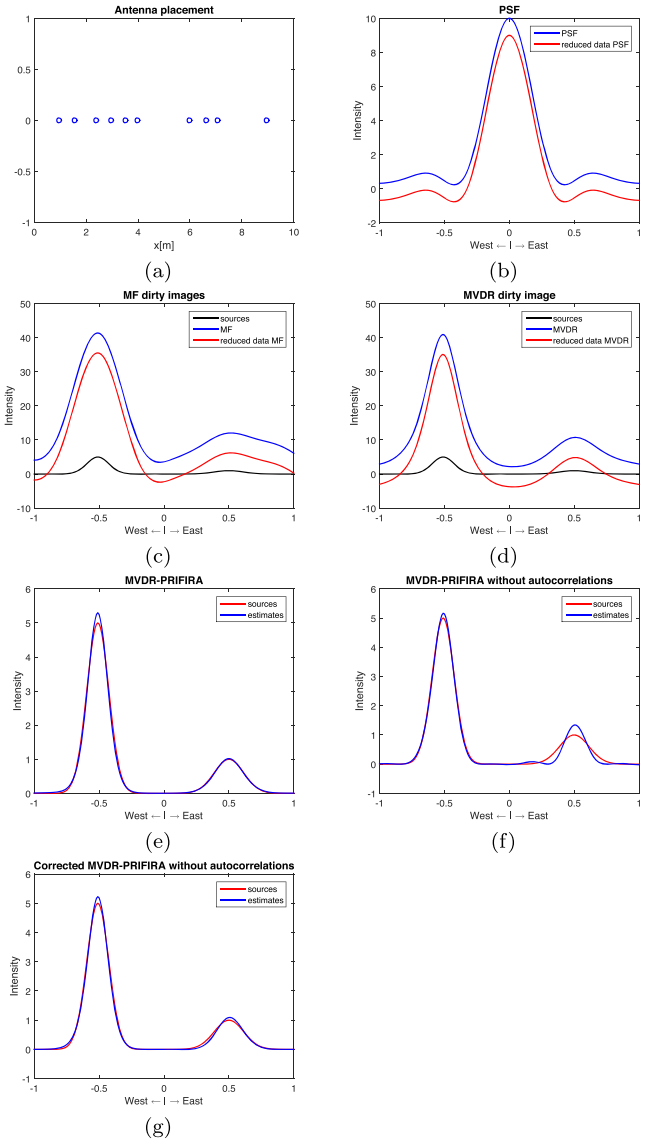


**Figure B1.** Effect of missing autocorrelations: (a) Antenna positions, (b) PSF, (c) MF dirty image, (d) MVDR dirty image, (e) MVDR-PRIFIRA, (f) MVDR-PRIFIRA without autocorrelations, and (g) MVDR-PRIFIRA without autocorrelations with correction to make sidelobes positive.

Discarding autocorrelations results in the PSF, MF, and MVDR image to have negative sidelobes. Since MF and MVDR are applied as weights to the columns of the $\mathbf{M}$, any zero value in the weights will enforce zero values in the estimated coefficients and eventually in the solution. The upper bound property of MF and MVDR ensures that none of the non-zero image pixels will not be set to zero. However, when autocorrelations are missing we should avoid zero values by shifting up all the pixels in the MF and MVDR image by the smallest negative value of the sidelobes.

In Fig. B1, we illustrate with a 1D simulation the effect of dropping the autocorrelations on the PSF, MF, and MVDR image, and on the reconstructed MVDR-PRIFIRA image (MF-PRIFIRA would give similar results). We use a similar setting as for the 1D example presented in Section 6 but with a different antenna placing to better show the effect of the negative sidelobes. We choose for this experiment two Gaussian sources of heights 5 and 1 centred at direction cosines $l = -0.5$ and $l = 0.5$, respectively, from left to right.

Fig. B1(a), (b), (c), and (d), respectively, show the antenna placement, PSF, MF dirty image, and MVDR dirty image. Lines related to the setting where we have access to the complete correlation matrix are shown in blue, while the red lines represent the case where the autocorrelations are not available. Fig. B1(e) and (f) represent MVDR-PRIFIRA, and MVDR-PRIFIRA when we do not have the autocorrelations. Fig. B1(g) is when we make the MVDR image strictly positive by adding the smallest negative sidelobe to the image and apply it as the right preconditioner in MVDR-PRIFIRA. As can be seen, this correction improves the estimation performance, although the performance of the algorithm is still suboptimal compared to the case where we have full information.

## APPENDIX C: MULTIDICTIONARY GENERALIZATION

In this appendix we include some early efforts on extending the algorithmic framework presented in Section 4 to cases where point sources and extended emissions coexist in the sky map. We explain the idea here but further developments are deferred to a future publication. The initial idea for this section is based on one of our early works published in Naghibzadeh, Sardarabadi & van der Veen (2016a).

We assume that we can model the sky map via an overcomplete dictionary $\boldsymbol{\Psi}$ as $\boldsymbol{\sigma} = \boldsymbol{\Psi}\boldsymbol{\eta}$ such that the coefficient $\boldsymbol{\eta}$ is sparse. We generalize the $\ell_0$-norm regularization problem (41) such that we have

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \|\boldsymbol{\Gamma}(\tilde{\boldsymbol{r}} - \mathbf{M}\boldsymbol{\Psi}\boldsymbol{\eta})\|_2^2 + \tau\|\boldsymbol{\eta}\|_0. \tag{C1}$$

We choose a simple dictionary $\boldsymbol{\Psi} = [\mathbf{I}, \mathbf{D}]$ with size $Q \times 2Q$ composed of (i) $\mathbf{I}$, the identity matrix, pixel basis, to model the point sources and (ii) $\mathbf{D}$, the Gaussian clean beam basis matrix. $\mathbf{D}$ consists of all the shifts of the clean beam centred on all the pixel locations and is used as a simple basis to model the extended emissions. We apply IR0-PRIFIRA to obtain an estimate $\hat{\boldsymbol{\eta}}$. The image estimate $\hat{\boldsymbol{\sigma}}$ can be obtained as $\hat{\boldsymbol{\sigma}} = \boldsymbol{\Psi}\hat{\boldsymbol{\eta}}$. We note that we need to normalize $\mathbf{D}$ such that the norm of the image is preserved during this transform.

We show the effect of applying the overcomplete dictionary on IR0-PRIFIRA based on a 1D test with a point source with size 4 and a Gaussian source with height 2, modelling an extended emission as shown in Fig. C1(a). We take as before $P = 10$ antennas with a random linear placement as shown in Fig. C1(b). The number of pixels in the image is $Q = 201$ as with the previous simulations. Gaussian receiver noise with variance $\sigma_n^2 = 100$ is added to the measurements. The beam pattern and the clean beam used in defining the dictionary are superimposed in Fig. C1(c). The MF and MVDR dirty images are shown in Fig. C1(d). The estimated basis coefficients $\hat{\boldsymbol{\eta}}$ and the reconstructed image based on the multidictionary version of IR0-PRIFIRA superimposed with the image are shown in Fig. C1(e) and (f), respectively. For this estimate 20 outer iterations of IR0-PRIFIRA are performed. These simulation results show that the original image is recovered faithfully.

This indicates that we can modify IR0-PRIFIRA by modelling the sky map based on an overcomplete dictionary such that the dictionary coefficients are sufficiently sparse. Doing so, we are able to obtain a generalization of the proposed framework such that images containing both extended emissions and point sources can be reconstructed sufficiently well. Further investigation on optimal basis designs is outside the scope of this work and is deferred to a future work.

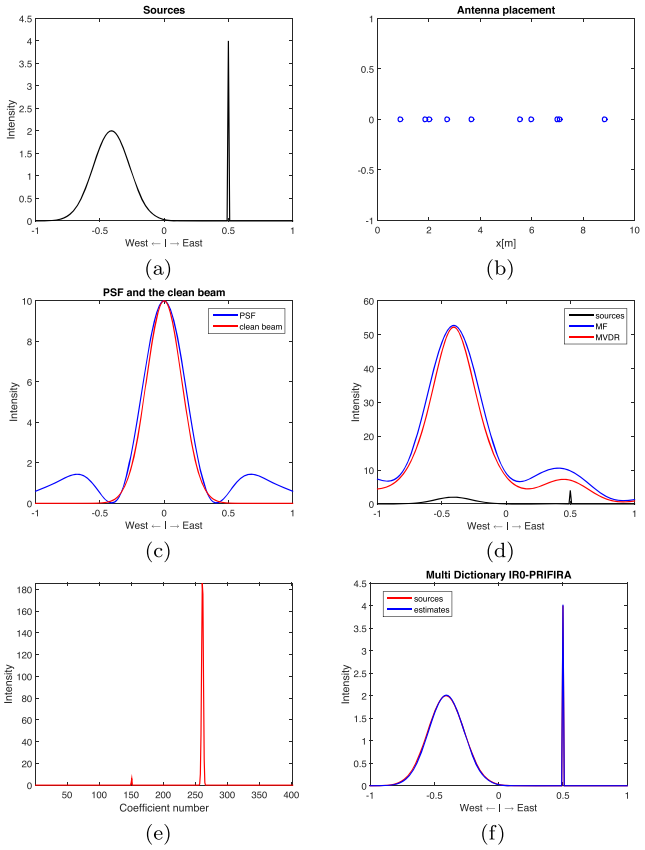


**Figure C1.** (a) Sources, (b) antenna positions, (c) PSF and the clean beam, (d) MF and MVDR dirty images, (e) estimated basis coefficients, and (f) multidictionary IR0-PRIFIRA estimates.

## APPENDIX D: COMBINED REGULARIZING EFFECT OF PRIOR-CONDITIONING AND EARLY STOPPING

In this appendix we discuss the combined regularizing effect of the prior-conditioning and early stopping on the solution of Problem (54). The SVD is a powerful tool for the analysis of LS problems. It is well-known that the minimum norm solution of a linear LS problem is obtained via the pseudo-inverse (Lawson & Hanson 1974). For example we consider Problem (20). If the SVD of $\boldsymbol{\Gamma}\mathbf{M}$ can be stated as $\boldsymbol{\Gamma}\mathbf{M} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^H$ where $\mathbf{U}$ and $\mathbf{V}$ contain the left and right singular vectors, respectively, and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the singular values of $\boldsymbol{\Gamma}\mathbf{M}$, the minimum norm solution can be expressed as

$$\hat{\boldsymbol{\sigma}} = (\boldsymbol{\Gamma}\mathbf{M})^{\dagger}\tilde{\boldsymbol{r}} = \mathbf{V}\boldsymbol{\Lambda}^{\dagger}\mathbf{U}^H\tilde{\boldsymbol{r}}. \tag{D1}$$

Unfortunately, for ill-posed problems the pseudo-inverse solution is unstable due to the noise amplification by the inversion of small singular values (Hansen 2010).

Applying regularization stabilizes the solution. The solution of many regularized LS problems can be stated in the form of a filtered SVD (Hansen 2005). If regularization is applied on (20), the solution in terms of filtered SVD can be stated as

$$\hat{\boldsymbol{\sigma}} = \mathbf{V}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\dagger}\mathbf{U}^H\tilde{\boldsymbol{r}}, \tag{D2}$$

where $\boldsymbol{\Phi}$ is a diagonal matrix containing the regularization filter factors and is dependent on the type of regularization applied. The purpose of the regularizing filter factors is to filter out the effect

of small singular values in $\mathbf{\Lambda}$ that cause noise amplification and instability of the estimated solution when inversion is performed.

We present the solution of (54) in terms of filtered SVD to show the regularizing effect of the iteration count and the prior-conditioner. Assuming in this case the SVD of $\bar{\mathbf{M}} = \mathbf{\Gamma M L}^{-1}$ is given as $\bar{\mathbf{U}}\bar{\mathbf{\Lambda}}\bar{\mathbf{V}}^H$, starting from (53) and following the approach from Jensen & Hansen (2007), Hansen (2005) we obtain:

$$K_t(\bar{\mathbf{M}}^H\bar{\mathbf{M}}, \bar{\mathbf{M}}^H\bar{\boldsymbol{r}}) = \text{span}\{\bar{\mathbf{V}}\bar{\mathbf{\Lambda}}\bar{\mathbf{U}}^H\bar{\boldsymbol{r}}, \bar{\mathbf{V}}\bar{\mathbf{\Lambda}}^3\bar{\mathbf{U}}^H\bar{\boldsymbol{r}}, \ldots, \bar{\mathbf{V}}\bar{\mathbf{\Lambda}}^{2t-1}\bar{\mathbf{U}}^H\bar{\boldsymbol{r}}\}. \tag{D3}$$

Since the solution is a linear combination of vectors, the filtered SVD solution of $\hat{\boldsymbol{\sigma}}$ can be stated as

$$\hat{\boldsymbol{\sigma}} = \mathbf{L}^{-1}\bar{\mathbf{V}}\bar{\mathbf{\Phi}}_t\mathbf{\Lambda}^\dagger\bar{\mathbf{U}}^H\tilde{\boldsymbol{r}}, \tag{D4}$$

where $\bar{\mathbf{\Phi}}_t$ is a diagonal matrix of the form $\bar{\mathbf{\Phi}}_t = \mathcal{P}_t(\bar{\mathbf{\Lambda}}^2)\bar{\mathbf{\Lambda}}^2$ where $\mathcal{P}_t$ indicates a polynomial of degree smaller than $t-1$. This polynomial is shown to be dominated by large singular values in the initial iterations. As the iteration continues, more singular values are recovered and the effect of small singular values becomes prominent. Therefore, choosing the right stopping iteration, $T$, limits the influence of the small singular values and therefore stabilizes the solution (Hansen 2010).

We can conclude from equation (D4) that the sky map obtained using PRIFIRA is in the form of a regularized LS solution. In this solution, both the prior-conditioning weights and the iteration count contribute to the regularization filter factors for filtering out the small singular values and thus stabilizing the inversion.

This paper has been typeset from a TEX/LATEX file prepared by the author.