Signal Processing for Radio Astronomy

van der Veen, A.-J.; Wijnholds, SJ; Sardarabadi, A.M.

DOI
10.1007/978-3-319-91734-4_9

Publication date
2019

Document Version
Final published version

Published in
Handbook of Signal Processing Systems

Citation (APA)
van der Veen, A.-J., Wijnholds, SJ., & Sardarabadi, A. M. (2019). Signal Processing for Radio Astronomy. In S. Bhattacharyya, E. Deprettere, R. Leupers, & J. Takala (Eds.), *Handbook of Signal Processing Systems* (pp. 311-360). Springer. https://doi.org/10.1007/978-3-319-91734-4_9

Signal Processing for Radio Astronomy

Alle-Jan van der Veen, Stefan J. Wijnholds, and Ahmad Mouri Sardarabadi

Abstract Radio astronomy is known for its very large telescope dishes but is currently making a transition towards the use of a large number of small antennas. For example, the Low Frequency Array, commissioned in 2010, uses about 50 stations each consisting of 96 low band antennas and 768 or 1536 high band antennas. The low-frequency receiving system for the future Square Kilometre Array is envisaged to initially consist of over 131,000 receiving elements and to be expanded later. These instruments pose interesting array signal processing challenges. To present some aspects, we start by describing how the measured correlation data is traditionally converted into an image, and translate this into an array signal processing framework. This paves the way to describe self-calibration and image reconstruction as estimation problems. Self-calibration of the instrument is required to handle instrumental effects such as the unknown, possibly direction dependent, response of the receiving elements, as well a unknown propagation conditions through the Earth's troposphere and ionosphere. Array signal processing techniques seem well suited to handle these challenges. Interestingly, image reconstruction, calibration and interference mitigation are often intertwined in radio astronomy, turning this into an area with very challenging signal processing problems.

A.-J. van der Veen (✉)
TU Delft, Faculty of EEMCS, Delft, The Netherlands
e-mail: a.j.vanderveen@tudelft.nl

S. J. Wijnholds
Netherlands Institute for Radio Astronomy (ASTRON), Dwingeloo, The Netherlands
e-mail: wijnholds@astron.nl

A. M. Sardarabadi
University of Groningen, Kapteyn Astronomical Institute, Groningen, The Netherlands
e-mail: ammsa@astro.rug.nl

S. S. Bhattacharyya et al. (eds.), *Handbook of Signal Processing Systems*,
https://doi.org/10.1007/978-3-319-91734-4_9

1 Introduction

Astronomical instruments measure cosmic particles or electromagnetic waves impinging on the Earth. Astronomers use the data generated by these instruments to study physical phenomena outside the Earth's atmosphere. In recent years, astronomy has transformed into a multi-modal science in which observations at multiple wavelengths are combined. Figure 1 provides a nice example showing the lobed structure of the famous radio source Cygnus A as observed at 240 MHz with the Low Frequency Array (LOFAR) overlaid by an X-Ray image observed by the Chandra satellite, which shows a much more compact source.

Such images are only possible if the instruments used to observe different parts of the electromagnetic spectrum provide similar resolution. Since the resolution is determined by the ratio of observed wavelength and aperture diameter, the aperture of a radio telescope has to be 5 to 6 orders of magnitude larger than that of an optical telescope to provide the same resolution. This implies that the aperture of a radio telescope should have a diameter of several hundreds of kilometers. Most current and future radio telescopes therefore exploit interferometry to synthesize a large aperture from a number of relatively small receiving elements.

An interferometer measures the correlation of the signals received by two antennas spaced at a certain distance. After a number of successful experiments in the 1950s and 1960s, two arrays of 25-m dishes were built in the 1970s: the 3 km Westerbork Synthesis Radio Telescope (WSRT, 14 dishes) in Westerbork, The Netherlands and the 36 km Very Large Array (VLA, 27 movable dishes) in Socorro, New Mexico, USA. These telescopes use Earth rotation to obtain a sequence of

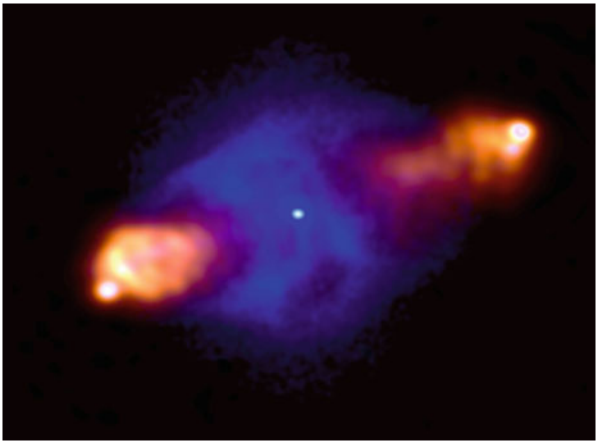

Fig. 1 Radio image of Cygnus A observed at 240 MHz with the Low Frequency Array (showing mostly the lobes left and right), overlaid over an X-Ray image of the same source observed by the Chandra satellite (the fainter central cloud) [65] (*Courtesy of Michael Wise and John McKean*)

correlations for varying antenna baselines, resulting in high-resolution images via *synthesis mapping*. A more extensive historical overview is presented in [52].

The radio astronomy community has recently commissioned a new generation of radio telescopes for low frequency observations, including the Murchison Widefield Array (MWA) [38, 53] in Western Australia and the Low Frequency Array (LOFAR) [24, 58] in Europe. These telescopes exploit phased array technology to form a large collecting area with ∼1000 to ∼50,000 receiving elements. The community is also making detailed plans for the Square Kilometre Array (SKA), a future radio telescope that should be one to two orders of magnitude more sensitive than any radio telescope built to date [18]. Even in its first phase of operation, the low-frequency receiving system of the SKA (SKA-low) is already envisaged to consist of over 131,000 receiving elements [17, 56].

The individual antennas in a phased array telescope have an extremely wide field-of-view, often the entire visible sky. This poses a number of signal processing challenges, because certain assumptions that work well for small fields-of-view (celestial sphere approximated by a plane, homogenous propagation conditions over the field-of-view), are no longer valid. Furthermore, the data volumes generated by these new instruments will be huge and will have to be reduced to manageable proportions by a real-time automated data processing pipeline. This combination of challenges led to a flurry of research activity in the area of array calibration, imaging and RFI mitigation, which are often intertwined in the astronomical data reduction.

The goal of calibration is to find the unknown instrumental, atmospheric and ionospheric disturbances. The imaging procedure should be able to apply appropriate corrections based on the outcome of the calibration process to produce a proper image of the sky. In this chapter, we review some of the array processing techniques that have been proposed for use in standard calibration and imaging pipelines, many of which are already being used in data reduction pipelines of instruments like LOFAR.

2 Notation

Matrices and vectors will be denoted by boldface upper-case and lower-case symbols, respectively. Entries of a matrix $\mathbf{A}$ are denoted by a_{ij}, and its columns by $\mathbf{a}_i$. Overbar $\overline{(\cdot)}$ denotes complex conjugation. The transpose operator is denoted by ^{T}, the complex conjugate (Hermitian) transpose by ^{H} and the Moore-Penrose pseudo-inverse by †. For matrices $\mathbf{A}$ of full column rank, i.e., $\mathbf{A}^H\mathbf{A}$ invertible, this is equal to the left inverse:

$$\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H \,. \tag{1}$$

The expectation operator is denoted by $E\{\cdot\}$.

We will multiply matrices in many different ways. Apart from the usual multiplication $\mathbf{AB}$, we will use $\mathbf{A}\odot\mathbf{B}$ to denote the Hadamard product (element-wise multiplication), and $\mathbf{A}\otimes\mathbf{B}$ to denote the Kronecker product,

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} .$$

We will also use the Khatri-Rao or column-wise Kronecker product of two matrices: let $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots]$ and $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \cdots]$, then

$$\mathbf{A}\circ\mathbf{B} = [\mathbf{a}_1\otimes\mathbf{b}_1,\ \mathbf{a}_2\otimes\mathbf{b}_2,\ \cdots] .$$

Depending on the context, $\text{diag}(\cdot)$ converts a vector to a diagonal matrix with the elements of the vector placed on the main diagonal, or converts a general matrix to a diagonal matrix by selecting its main diagonal. Further, $\text{vec}(\cdot)$ converts a matrix to a vector by stacking the columns of the matrix.

Properties of Kronecker products are listed in, e.g., [43]. We frequently use

$$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD} \tag{2}$$

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T\otimes\mathbf{A})\text{vec}(\mathbf{B}) \tag{3}$$

$$\text{vec}(\mathbf{A}\,\text{diag}(\mathbf{b})\,\mathbf{C}) = (\mathbf{C}^T\circ\mathbf{A})\mathbf{b} . \tag{4}$$

Property (3) is used to move a matrix $\mathbf{B}$ from the middle of an equation to the right of it, exploiting the linearity of the product. Property (4) is a special case of it, to be used if $\mathbf{B}$ is a diagonal matrix: in that case $\text{vec}(\mathbf{B})$ has many zero entries, and we can omit the corresponding columns of $\mathbf{C}^T\otimes\mathbf{A}$, leaving only the columns of the Khatri-Rao product $\mathbf{C}^T\circ\mathbf{A}$. A special case of (3) is

$$\text{vec}(\mathbf{a}\mathbf{a}^H) = \bar{\mathbf{a}}\otimes\mathbf{a} \tag{5}$$

which shows how a rank-1 matrix $\mathbf{a}\mathbf{a}^H$ is related to a vector with a specific "Kronecker structure".

3 Basic Concepts of Interferometry; Data Model

The concept of interferometry is illustrated in Fig. 2. An interferometer measures the spatial coherency of the incoming electromagnetic field. This is done by correlating the signals from the individual receivers with each other. The correlation of each pair of receiver outputs provides the amplitude and phase of the spatial coherence function for the *baseline* defined by the vector pointing from the first to the second

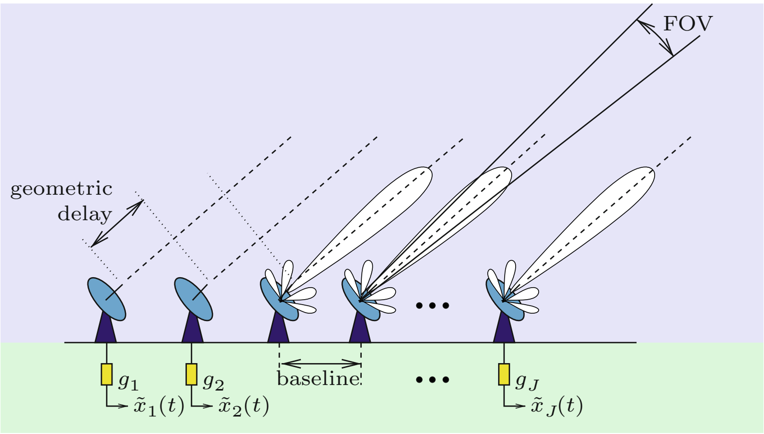

Fig. 2 Schematic overview of a radio interferometer

receiver in a pair. In radio astronomy, these correlations are called the *visibilities*. In this section, we describe the data acquisition in detail and construct a suitable data model.

3.1 Data Acquisition

Assume that there are J receiving elements. Depending on the context, a receiving element can be a telescope dish, a single antenna within a subarray (usually referred to as a *station*) or a beamformed subarray. The RF signal from the jth telescope, $\tilde{x}_j(t)$ is first moved to baseband where it is denoted by $x_j(t)$, then sampled and split into narrow subbands, e.g., of 100 kHz each, such that the narrowband condition holds. This condition states that the maximal geometrical delay across the array should be fairly representable by a phase shift of the complex baseband signal, and this property is discussed in more detail in the next subsection. The resulting signal is called $x_j(n,k)$, for the jth telescope, nth time bin, and for the subband frequency centered at RF frequency f_k. The J signals can be stacked into a $J \times 1$ vector $\mathbf{x}(n,k)$.

For each short-term integration (STI) interval m and each subband k, a covariance matrix estimate is formed by *integrating* (summing or averaging) the cross-correlation products $\mathbf{x}(n,k)\mathbf{x}^H(n,k)$ over N subsequent samples,

$$\hat{\mathbf{R}}_{m,k} = \frac{1}{N} \sum_{n=(m-1)N}^{mN-1} \mathbf{x}(n,k)\mathbf{x}^H(n,k)\,, \tag{6}$$

This processing chain is summarized in Fig. 3.

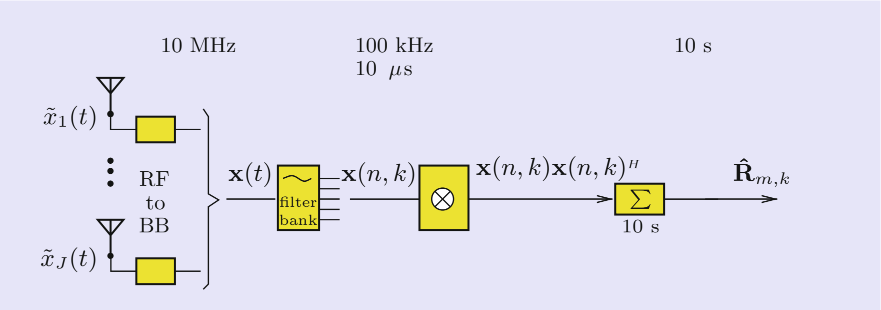

Fig. 3 The processing chain to obtain covariance data

The duration of an STI depends on the stationarity of the data, which is limited by factors like Earth rotation and the diameter of the array. For the LOFAR, a typical value for the STI is 1–10 s. A complete observation can last from a few minutes to a full night, i.e., more than 12 h. The resulting number of samples N in a snapshot observation is equal to the product of bandwidth and integration time and typically ranges from 10^3 (1 s, 1 kHz) to 10^6 (10 s, 100 kHz) in radio astronomical applications.

3.2 Complex Baseband Signal Representation

Before we can derive a data model, we need to include some more details on the RF to baseband conversion. In signal processing, signals are usually represented by their low pass equivalents, which is a suitable representation for narrowband signals in a digital communication system, and also applicable in the radio astronomy context. A complex valued bandpass signal, also called the *complex baseband signal*, with center frequency f_c may be written as

$$\tilde{s}(t) = s(t)e^{\mathrm{j}2\pi f_c t} \tag{7}$$

Suppose that the bandpass signal $\tilde{s}(t)$ is delayed by a time τ. This can be written as

$$\tilde{s}_\tau(t) := \tilde{s}(t-\tau) = s(t-\tau)e^{\mathrm{j}2\pi f_c(t-\tau)} = s(t-\tau)e^{-\mathrm{j}2\pi f_c\tau}e^{\mathrm{j}2\pi f_c t}\,.$$

The complex envelope of the delayed signal is thus $s_\tau(t) = s(t-\tau)e^{-\mathrm{j}2\pi f_c\tau}$. Let B be the bandwidth of the complex envelope (the baseband signal) and let $S(f)$ be its Fourier transform. We then have

$$s(t-\tau) = \int_{-B/2}^{B/2} S(f)e^{-\mathrm{j}2\pi f\tau}e^{\mathrm{j}2\pi ft}df \;\approx\; \int_{-B/2}^{B/2} S(f)e^{\mathrm{j}2\pi ft}df = s(t)$$

where the approximation $e^{-\mathrm{j}2\pi f\tau} \approx 1$ is valid if $|2\pi f\tau| \ll 1$ for all frequencies $|f| \leq \frac{B}{2}$. Ignoring a factor π, the resulting condition $B\tau \ll 1$ is called the narrowband condition. The quantitative interpretation of "much less than one" depends on the SNR of the received signals [67] and the sensitivity loss considered acceptable [9]. Under this condition, we have for the complex envelope $s_\tau(t)$ of the delayed bandpass signal $\tilde{s}_\tau(t)$ that

$$s_\tau(t) \approx s(t)e^{-\mathrm{j}2\pi f_c\tau} \qquad \text{for } B\tau \ll 1\,.$$

The conclusion is that, for narrowband signals, time delays smaller than the inverse bandwidth may be represented as phase shifts of the complex envelope. Phased array processing heavily depends on this step. For radio astronomy, the maximal delay τ is equal to the maximal geometric delay, which can be related to the diameter of the array. The bandwidth B is the bandwidth of each subband f_k in the RF processing chain that we discussed in the previous subsection.

3.3 Data Model

We return to the radio astronomy context. For our purposes, it is convenient to model the sky as consisting of a collection of Q spatially discrete point sources, with $s_q(n,k)$ the signal of the qth source at time sample n and frequency f_k.

The signal received at the jth antenna is a sum of delayed source signals, where the delays are geometric delays that depend on the direction under which each of the signals is observed. In the previous subsection, we saw that under the narrowband condition a delay of a narrowband signal $s(t,k)$ by τ can be represented by a phase shift:

$$s_\tau(t,k) = e^{-\mathrm{j}2\pi f_k\tau}s(t,k)$$

which takes the form of a multiplication of $s(t,k)$ by a complex number. Let $\mathbf{z}_j = [x_j, y_j, z_j]^T$ be the location of the jth antenna. Further, let $\mathbf{l}_q$ be a unit-length direction vector pointing into the direction of the qth source.

The geometrical delay τ at antenna j for a signal coming from direction $\mathbf{l}_q$ can be computed as follows. For a signal traveling directly from the origin of the coordinate system used to specify the antenna locations to antenna j, the delay is the distance from the origin to the jth antenna divided by c, the speed of light. For any other direction, the delay depends on the cosine of the angle of incidence (compared to the baseline vector) at observing time n, and is thus described by the inner product of the location vector with the direction vector, i.e., $\tau_{q,j}(n) = \mathbf{z}_j \cdot \mathbf{l}_q(n)/c$. Overall, the phase factor representing the geometric delay is

$$a_{j,q}(n,k) = e^{-\mathrm{j}2\pi f_k \tau_{q,j}(n)} = e^{-\frac{2\pi \mathrm{j} f_k}{c}\mathbf{z}_j^T \mathbf{l}_q(n)}\,. \tag{8}$$

The coordinates of source direction vectors $\mathbf{l}_q$ are expressed as[1] (ℓ, m, n), where ℓ, m and n are direction cosines and $n = \sqrt{1 - \ell^2 - m^2}$ due to the normalization. There are several conventions and details regarding coordinate systems [52], but they are not of concern for us here.

Besides the phase factor $a_{q,j}(n,k)$, the received signals are also affected by the direction dependent response of the receiving element $b_j(\mathbf{l}, n, k)$ and the direction independent complex valued receiver path gain $g_j(n,k)$. The function $b_j(\mathbf{l}, n, k)$ is referred to as the *primary beam* to distinguish it from the array beam and the *point spread function* or *dirty beam* that results from beamforming over a full synthesis observation (more about this later). The general shape of the primary beam is known from (electromagnetic) modelling during the design of the telescope. If that model is not sufficiently accurate, it needs to be calibrated. Together with the tropospheric and ionospheric propagation conditions, the primary beam determines the direction dependent gain $g^d_{j,q}(n,k)$ of the jth receiving element. The signal $x_j(n,k)$ received by the jth receiving element can thus be described by

$$x_j(n,k) = g_j(n,k) \sum_{q=1}^{Q} g^d_{j,q}(n,k) a_{j,q}(n,k) s_q(n,k) + n_j(n,k), \tag{9}$$

where $n_j(n,k)$ denotes the additive noise in the jth receive path.

We can stack the phase factors $a_{j,q}(n,k)$ into an *array response vector* for each source as

$$\mathbf{a}_q(n,k) = \left[a_{1,q}(n,k), \cdots, a_{J,q}(n,k)\right]^T. \tag{10}$$

In a similar way, we can stack the direction independent gains $g_j(n,k)$ into a vector $\mathbf{g}(n,k)$, stack the direction dependent gains $g^d_{j,q}(n,k)$ into a vector for each source $\mathbf{g}^d_q(n,k)$ and stack the additive noise signals in a vector $\mathbf{n}(n,k)$. With these conventions, we can formulate the data model for the *array signal vector* as

$$\mathbf{x}(n,k) = \mathbf{g}(n,k) \odot \sum_{q=1}^{Q} \mathbf{g}^d_q(n,k) \odot \mathbf{a}_q(n,k) s_q(n,k) + \mathbf{n}(n,k). \tag{11}$$

For convenience of notation, we introduce the gain matrix

---

[1] With abuse of notation, as m, n are not related to the time variables used earlier.

$$\mathbf{G}(n,k) = \left[\mathbf{g}(n,k) \odot \mathbf{g}_q^d(n,k), \cdots, \mathbf{g}(n,k) \odot \mathbf{g}_Q^d(n,k)\right].$$

As we will see in Sect. 5, this gain matrix may have a specific structure depending on a priori knowledge about the direction independent gains and the direction dependent gains. This structure can then be exploited during calibration. We can also stack the array response vectors into an array response matrix $\mathbf{A}(n,k) = \left[\mathbf{a}_1(n,k), \cdots \mathbf{a}_Q(n,k)\right]^T$. These conventions allow us to write Eq. (11) as

$$\mathbf{x}(n,k) = (\mathbf{G}(n,k) \odot \mathbf{A}(n,k))\, \mathbf{s}(n,k) + \mathbf{n}(n,k), \tag{12}$$

where $\mathbf{s}(n,k) = \left[s_1(n,k), \cdots s_Q(n,k)\right]^T$.

For convenience of notation, we will in future usually drop the dependence on the frequency f_k (index k) from the notation. Previously, in (6), we defined correlation estimates $\hat{\mathbf{R}}_m$ as the output of the data acquisition process, where the time index m corresponds to the mth STI interval, such that $(m-1)N \leq n \leq mN$. Due to Earth rotation, the vectors $\mathbf{a}_q(n)$ change slowly with time, but we assume that within an STI it can be considered constant and can be represented, with some abuse of notation, by $\mathbf{a}_q(m)$. In that case, $\mathbf{x}(n)$ is wide sense stationary over the STI, and a single STI covariance matrix is defined as

$$\mathbf{R}_m = E\{\mathbf{x}(n)\, \mathbf{x}^H(n)\}, \qquad m = \left\lceil \frac{n}{N} \right\rceil \tag{13}$$

where $\mathbf{R}_m$ has size $J \times J$. Each element of $\mathbf{R}_m$ represents the interferometric correlation along the baseline vector between the two corresponding receiving elements. It is estimated by STI sample covariance matrices $\hat{\mathbf{R}}_m$ defined in (6), and our stationarity assumptions imply $E\{\hat{\mathbf{R}}_m\} = \mathbf{R}_m$.

We will model the source signals $s_q(n,k)$ and the noise signals $n_j(n,k)$ as zero mean white Gaussian random processes sampled at the Nyquist rate. We will also assume that the source signals and noise signals are mutually uncorrelated. With these assumptions, we find, by substituting Eq. (12) into Eq. (13), that

$$\begin{aligned}
\mathbf{R}_m &= E\left\{(\mathbf{G}_m \odot \mathbf{A}_m \mathbf{s}(n) + \mathbf{n}(n))\, (\mathbf{G}_m \odot \mathbf{A}_m \mathbf{s}(n) + \mathbf{n}(n))^H\right\} \\
&= (\mathbf{G}_m \odot \mathbf{A}_m)\, E\left\{\mathbf{s}(n)\mathbf{s}^H(n)\right\} (\mathbf{G}_m \odot \mathbf{A}_m)^H + E\left\{\mathbf{n}(n)\mathbf{n}^H(n)\right\} \\
&= (\mathbf{G}_m \odot \mathbf{A}_m)\, \boldsymbol{\Sigma}_s\, (\mathbf{G}_m \odot \mathbf{A}_m)^H + \boldsymbol{\Sigma}_n,
\end{aligned} \tag{14}$$

where $\boldsymbol{\Sigma}_s = \operatorname{diag}(\boldsymbol{\sigma}_s)$ with $\boldsymbol{\sigma}_s = [\sigma_1^2, \cdots, \sigma_Q^2]^T$ is the source covariance matrix and $\boldsymbol{\Sigma}_n = \operatorname{diag}(\boldsymbol{\sigma}_n)$ with $\boldsymbol{\sigma}_n = [\sigma_{n,1}^2, \cdots, \sigma_{n,J}^2]^T$ is the noise covariance matrix. In radio astronomy, the covariance data model described in Eq. (14) is usually referred to as the *measurement equation*.

3.4 Radio Interferometric Imaging Concepts

Under ideal circumstances, the array response matrix $\mathbf{A}_m$ is not perturbed by the gain matrix $\mathbf{G}_m$, i.e., we have $\mathbf{G}_m = \mathbf{1}\mathbf{1}^H$ where $\mathbf{1}$ denotes a vector of ones of appropriate size. The columns of $\mathbf{A}_m$ are given by Eq. (8). Its entries represent the phase shifts due to the geometrical delays associated with the array and source geometry. By adding the gain matrix $\mathbf{G}_m$, we can introduce directional disturbances due to non-isotropic antennas, unequal antenna gains and disturbances due to ionospheric effects.

Assuming ideal conditions and ignoring the additive noise, a single element of the array covariance matrix, usually referred to as a *visibility*, can be written as

$$(\mathbf{R}_m)_{ij} = \sum_{q=1}^{Q} a_{i,q}\overline{a_{j,q}}\sigma_q^2 = \sum_{q=1}^{Q} I\left(\mathbf{l}_q\right) e^{-\mathrm{j}\frac{2\pi}{\lambda}\left(\mathbf{z}_i(m)-\mathbf{z}_j(m)\right)^T\mathbf{l}_q}. \tag{15}$$

where $I(\mathbf{l}_q) = \sigma_q^2$ is the brightness (power) in direction $\mathbf{l}_q$. The function $I(\mathbf{l})$ is the brightness image (or *map*) of interest: it is this function that is shown when we refer to a radio-astronomical image like Fig. 1. It is a function of the direction vector $\mathbf{l}$: this is a 3D vector, but due to its normalization it depends on only two parameters. We could e.g., show $I(\cdot)$ as function of the direction cosines (ℓ, m), or of the corresponding angles.

For our discrete point-source model, the brightness image is

$$I(\mathbf{l}) = \sum_{q=1}^{Q} \sigma_q^2\, \delta(\mathbf{l} - \mathbf{l}_q) \tag{16}$$

where $\delta(\cdot)$ is a Kronecker delta, and the direction vector $\mathbf{l}$ is mapped to the location of "pixels" in the image (various transformations are possible). Only the pixels $\mathbf{l}_q$ are nonzero, and have value equal to the source variance σ_q^2.

The vector $\mathbf{z}_i(m) - \mathbf{z}_j(m)$ is the *baseline*: the (normalized) vector pointing from telescope i to telescope j. In radio astronomy, it is usually expressed in coordinates denoted by $\mathbf{u}_{ij} = (u, v, w)$ and normalized by the wavenumber, i.e., $\mathbf{u}_{ij}(m) = (2\pi/\lambda)(\mathbf{z}_i(m) - \mathbf{z}_j(m))$. The objective in telescope design is often to have as many different baselines as possible. In that case the entries of $\mathbf{R}_m$ are different and non-redundant. As the Earth turns, the baselines also turn, thus giving rise to new baseline directions. We will see later that the set of baselines during an observation determines the spatial sampling function by which the incoming wave field is sampled, with important implications on the quality of the resulting image.

Equation (15) describes the relation between the visibility model and the desired image, and it has the form of a Fourier transform; it is known in radio astronomy as the Van Cittert-Zernike theorem [49, 52]. Image formation (*map making*) is essentially the inversion of this relation. Unfortunately, we have only a finite set of observations, therefore we can only obtain a *dirty image*: if we apply the inverse Fourier transformation to the measured correlation data, we obtain

$$\hat{I}_D(\mathbf{l}) := \sum_{i,j,m} \left(\hat{\mathbf{R}}_m\right)_{ij} e^{\mathrm{j}\mathbf{u}_{ij}^T(m)\mathbf{l}_q} \tag{17}$$

In terms of the measurement data model (15), the "expected value" of the image is obtained by replacing $\hat{\mathbf{R}}_m$ by $\mathbf{R}_m$, or

$$\begin{aligned} I_D(\mathbf{l}) &:= \sum_{i,j,m} (\mathbf{R}_m)_{ij}\, e^{\mathrm{j}\mathbf{u}_{ij}^T(m)\mathbf{l}} \\ &= \sum_{i,j,m} \sum_q \sigma_q^2 e^{\mathrm{j}\mathbf{u}_{ij}^T(m)(\mathbf{l}-\mathbf{l}_q)} \\ &= \sum_q I(\mathbf{l}_q) B(\mathbf{l}-\mathbf{l}_q) \\ &= I(\mathbf{l}) * B(\mathbf{l}), \end{aligned} \tag{18}$$

where the *dirty beam* is given by

$$B(\mathbf{l}) := \sum_{i,j,m} e^{\mathrm{j}\mathbf{u}_{ij}^T(m)\mathbf{l}}. \tag{19}$$

The dirty image $I_D(\mathbf{l})$ is the desired "true" image $I(\mathbf{l})$ convolved with the dirty beam $B(\mathbf{l})$: every point source excites a beam $B(\mathbf{l}-\mathbf{l}_q)$ centered at its location $\mathbf{l}_q$. The effect of this is that the true image gets blurred, thus limiting its resolution. Note that $B(\mathbf{l})$ is a known function: it only depends on the locations of the telescopes, or rather the set of telescope baselines $\mathbf{u}_{ij}(m) = (2\pi/\lambda)(\mathbf{z}_i(m) - \mathbf{z}_j(m))$.

Note that Eq. (17) has the form of a Fourier transform, although it has been defined on (u, v, w) samples that are non-uniformly spaced. To be able to use the computationally efficient fast Fourier transform (FFT), astronomy software first applies a *gridding* operation that interpolates and resamples the visibilities onto a regular grid, after which the FFT can be used to obtain the dirty image [49, 52]. This essentially implements a non-uniform FFT as used in other science communities [19].

As an example, the antenna configuration for the six stations forming the core of the LOFAR and the resulting single-STI dirty beam is shown in Fig. 4. The dirty beam has heavy sidelobes as high as −10 dB. A resulting dirty image (in dB scale) is shown in Fig. 5. In this image, we see the complete sky, in (ℓ, m) coordinates, where the reference direction is pointing towards zenith. The strong visible sources are Cassiopeia A and Cygnus A, also visible is the Milky Way. The image was obtained by averaging 259 STIs, each consisting of 1 s data in a single frequency channel of 195 kHz wide at a central frequency of 58.9 MHz.

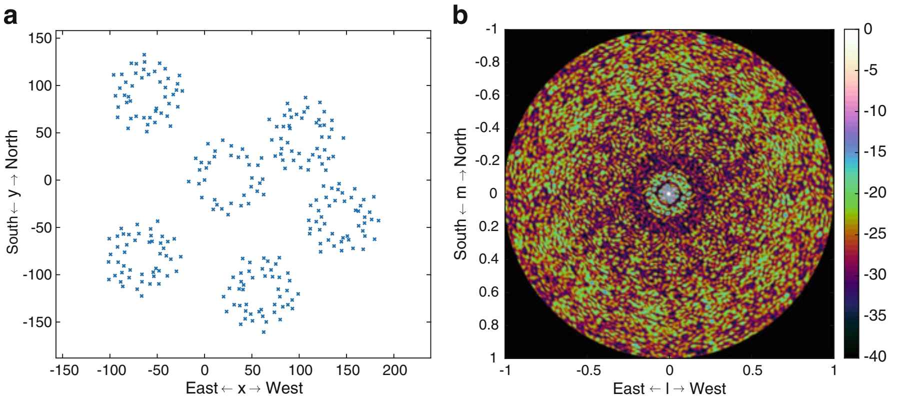

Fig. 4 (**a**) Coordinates of the antennas in the LOFAR Superterp, which defines the spatial sampling function, and (**b**) the resulting *dirty beam* in dB scale

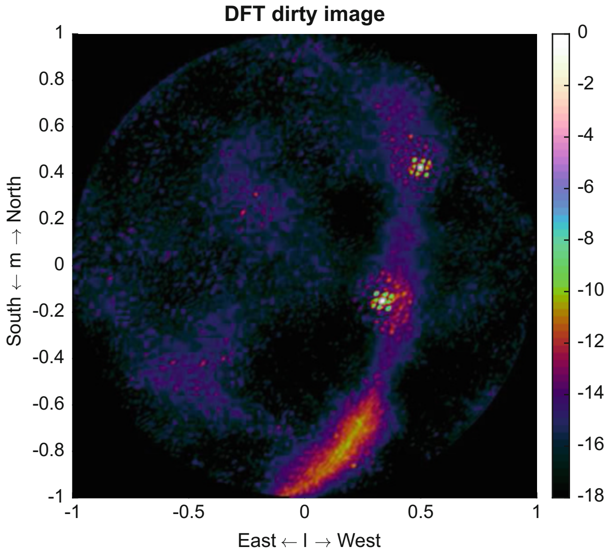

Fig. 5 Dirty image following (18), using LOFAR Superterp data

The dirty beam is essentially a non-ideal point spread function due to finite and non-uniform spatial sampling: we only have a limited set of baselines. The dirty beam usually has a main lobe centered at $\mathbf{l} = \mathbf{0}$, and many side lobes. If we would have a large number of telescopes positioned in a uniform rectangular grid, the dirty beam would be a 2-D sinc-function (similar to a boxcar taper in time-domain sampling theory). The resulting beam size is inversely proportional to the aperture (diameter) of the array. This determines the *resolution* in the dirty image. The sidelobes of the beam give rise to confusion between sources: it is unclear whether a small peak in the image is caused by the main lobe of a weak source, or the sidelobe of a strong source. Therefore, attempts are made to design the array such that the sidelobes are low. It is also possible to introduce weighting coefficients ("tapers") in (18) to obtain an acceptable beamshape.

Another aspect is the summation over m (STI intervals) in (19), where the rotation of the Earth is used to obtain essentially many more antenna baselines. This procedure is referred to as *Earth rotation synthesis* as more (u, v, w) sampling points are obtained over time. The effect of this is that the sidelobes tend to get averaged out, to some extent. Many images are also formed by averaging over a small number of frequency bins (assuming the σ_q^2 are constant over these frequency bins), which enters into the equations in exactly the same way: Replace $\mathbf{z}_i(m)$ by $\mathbf{z}_i(m, k)$ and also sum over the frequency index k.

4 Image Reconstruction

The goal of image reconstruction is to obtain an estimate of the true image $I(\mathbf{l})$. Many approaches to this problem have been proposed, which can be divided into two classes. The first is a non-parametric approach that starts from the dirty image. Since the dirty image is the convolution of the true image by the dirty beam, this reduces the image reconstruction problem to a *deconvolution* problem. Deconvolution is the process of recovering $I(\mathbf{l})$ from $I_D(\mathbf{l})$ using knowledge of the dirty beam and thus to obtain the high-resolution "clean" image. A standard algorithm for doing this is CLEAN [27] and variants; however, many other algorithms are possible, depending on the underlying model assumptions and on a trade-off between accuracy and numerical complexity.

The second class of approaches is to consider image reconstruction as an estimation problem in which an unknown set of parameters describing $I(\mathbf{l})$ need to be extracted from the measured visibilities collected in the measured array covariance matrices $\hat{\mathbf{R}}_m$. This "model matching" approach is discussed in more detail in Sect. 4.4.

After a telescope has been designed and built, algorithms for image formation are the most important topic for signal processing. Careful techniques can increase the dynamic range (ratio between powers of the strongest and the weakest features in the image) by several orders of magnitude. However, the numerical complexity is often large, and high-resolution images require dedicated hardware solutions and sometimes even supercomputers. In this section, we will describe some of the algorithms. Additional overviews are available in [13, 14, 33, 36], as well as in the books [4, 52].

4.1 Constructing Dirty Images

4.1.1 Beamforming Formulation

Previously (Eq. (17)), we formulated the dirty image as the inverse Fourier transform of the measured correlations. Here, we will interpret this process as beamforming. Once we have this formulation, we may derive many other dirty images via beamforming techniques. For simplicity of notation, we assume from now on that only a single STI snapshot is used in the imaging, hence we also drop the time index m from the equations. The results can easily be extended.

The imaging process transforms the covariances of the received signals to an image of the source structure within the field-of-view of the receivers. In array processing terms, it can be described as follows [33]. Assume a data model as in (12) with all gain factors equal to unity, and recall the definition of the array response vector $\mathbf{a}(\mathbf{l})$ in (8) and (10) (using yet another change of notation to emphasize now that $\mathbf{a}$ is a function of the source direction $\mathbf{l}$). There are J antennas. To determine the power of a signal arriving from a particular direction $\mathbf{l}$, a weight vector

$$\mathbf{w}(\mathbf{l}) = \frac{1}{J}\mathbf{a}(\mathbf{l}) = \frac{1}{J}e^{-\mathrm{j}\frac{2\pi}{\lambda}\mathbf{Z}^T\mathbf{l}}, \tag{20}$$

where $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_J]$, is applied to the array signal vector $\mathbf{x}(n)$. The operation $y(n) = \mathbf{w}^H\mathbf{x}(n)$ is generally called beamforming. The choice $\mathbf{w} = \mathbf{a}$ precisely compensates the geometric phase delays so that the antenna signals are added in-phase. This can be regarded as a spatially matched filter, or *conjugate field match*. The (often omitted) scaling by $1/J$ ensures the correct scaling of the output power. Indeed, the output power of a beamformer is, generally,

$$E\{|y|^2\} = \mathbf{w}^H E\{\mathbf{x}\mathbf{x}^H\}\mathbf{w} = \mathbf{w}^H\mathbf{R}\mathbf{w}\,.$$

For a data model consisting of a single source with power σ^2 arriving from direction $\mathbf{a}(\mathbf{l})$, i.e., $\mathbf{x}(n) = \mathbf{a}(\mathbf{l})s(n)$, we have, with $\mathbf{w} = \frac{1}{J}\mathbf{a}(\mathbf{l})$,

$$E\{|y|^2\} = \mathbf{w}^H(\mathbf{a}\sigma^2\mathbf{a}^H)\mathbf{w} = \sigma^2\frac{\mathbf{a}^H\mathbf{a}}{J}\frac{\mathbf{a}^H\mathbf{a}}{J} = \sigma^2\,. \tag{21}$$

Thus, the matched beamformer corrects precisely the signal delays (phase shifts) present in $\mathbf{a}(\mathbf{l})$, when $\mathbf{w}$ matches $\mathbf{a}(\mathbf{l})$, i.e. the beamformer is pointed into the same direction as the source. If the beamformer is pointed into other directions, the response is usually much smaller.

Using the beamformer to scan over all pixels $\mathbf{l}$ in an image, we can create an image via beamforming as

$$\hat{I}_{BF}(\mathbf{l}) = \mathbf{w}(\mathbf{l})^H\hat{\mathbf{R}}\mathbf{w}(\mathbf{l}) \tag{22}$$

and the corresponding model for this image is

$$I_{BF}(\mathbf{l}) = \mathbf{w}(\mathbf{l})^H \mathbf{R}\mathbf{w}(\mathbf{l})\,. \tag{23}$$

The matched filter corresponds to weights $\mathbf{w}(\mathbf{l})$ defined as in (20). Except for a factor J^2, the image $I_{BF}(\mathbf{l})$ is identical to the dirty image $I_D(\mathbf{l})$ defined in (18) for this choice! Indeed, starting from (18), we can write

$$I_D(\mathbf{l}) = \sum_{i,j} R_{ij}\, e^{\mathrm{j}\mathbf{u}_{ij}^T\mathbf{l}} = \sum_{i,j} \overline{a_i}(\mathbf{l}) R_{ij} a_j(\mathbf{l}) = \mathbf{a}(\mathbf{l})^H \mathbf{R}\mathbf{a}(\mathbf{l})$$

which is the beamforming image obtained using $\mathbf{w}(\mathbf{l}) = \mathbf{a}(\mathbf{l})$. The response to a single source at the origin is

$$\begin{aligned} B(\mathbf{l}) &= \mathbf{a}(\mathbf{l})^H \mathbf{a}(\mathbf{0})\mathbf{a}(\mathbf{0})^H \mathbf{a}(\mathbf{l}) \\ &= \mathbf{a}(\mathbf{l})^H \mathbf{1}\mathbf{1}^H \mathbf{a}(\mathbf{l}) \\ &= \mathbf{1}^H [\mathbf{a}(\mathbf{l})\mathbf{a}(\mathbf{l})^H]\mathbf{1} \\ &= \sum_{i,j} e^{\mathrm{j}\mathbf{u}_{ij}^T\mathbf{l}} \end{aligned}$$

which is the dirty beam defined in (19), now written in beamforming notation. It typically has a spike at $\mathbf{l} = \mathbf{0}$, and many sidelobes, depending on the spatial sampling function. We have already seen that these sidelobes limit the resolution, as they can be confused with (or mask) other sources.

So far, we looked at the response to a source, but ignored the effect of the noise on an image. In the beamforming formulation, the response to a data set which only consists of noise, or $\mathbf{R} = \mathbf{\Sigma}_n$ is

$$I_n(\mathbf{l}) = \mathbf{w}(\mathbf{l})^H \mathbf{\Sigma}_n \mathbf{w}(\mathbf{l})\,.$$

Suppose that the noise is spatially white, $\mathbf{\Sigma}_n = \sigma_n^2\mathbf{I}$, and that we use the matched beamformer (20), we obtain

$$I_n(\mathbf{l}) = \sigma_n^2 \frac{\mathbf{a}(\mathbf{l})^H}{J}\frac{\mathbf{a}(\mathbf{l})}{J} = \sigma_n^2 \frac{\|\mathbf{a}(\mathbf{l})\|^2}{J^2} = \frac{\sigma_n^2}{J}, \tag{24}$$

since all entries of $\mathbf{a}(\mathbf{l})$ have unit magnitude. As this is a constant, the image will be "flat". For a general data set, the responses to the sources and to the noise will be added. Comparing (21)–(24), we see that the noise is suppressed by a factor J compared to a point source signal coming from a specific direction. This is the *array gain*. If we use multiple STIs and/or frequencies f_k, the array gain can be larger than J.

4.1.2 Constructing Dirty Images by Adaptive Beamforming

Now that we have made the connection of the dirty image to beamforming, we can apply a range of other beamforming techniques instead of the matched filter, such as the class of spatially adaptive beamformers. In fact, these can be considered as 2D spatial-domain versions of (now classical) spectrum estimation techniques for estimating the power spectral density of a random process (viz. [26]), and the general idea is that we can obtain a higher resolution if the sidelobes generated by strong sources are made small.

As an example, the "minimum variance distortionless response" (MVDR) beamformer is defined such that the response towards the direction of interest $\mathbf{l}$ is unity, but signals from other directions are suppressed as much as possible, i.e.,

$$\mathbf{w}(\mathbf{l}) = \arg\min_{\mathbf{w}} \ \mathbf{w}^H\mathbf{R}\mathbf{w}\,, \qquad \text{such that} \quad \mathbf{w}^H\mathbf{a}(\mathbf{l}) = 1.$$

This problem can be solved in various ways. For example, after making a transformation $\mathbf{w}' := \mathbf{R}^{1/2}\mathbf{w}$, $\mathbf{a}' := \mathbf{R}^{-1/2}\mathbf{a}$, the problem becomes

$$\mathbf{w}'(\mathbf{l}) = \arg\min_{\mathbf{w}'} \ \|\mathbf{w}'\|^2\,, \qquad \text{such that} \quad \mathbf{w}'^H\mathbf{a}'(\mathbf{l}) = 1.$$

To minimize the norm of $\mathbf{w}'$, it should be aligned to $\mathbf{a}'$, i.e., $\mathbf{w}' = \alpha\mathbf{a}'$, and the solution is $\mathbf{w}' = \mathbf{a}'/(\mathbf{a}'^H\mathbf{a}')$. In terms of the original variables, the solution is then

$$\mathbf{w}(\mathbf{l}) = \frac{\mathbf{R}^{-1}\mathbf{a}(\mathbf{l})}{\mathbf{a}(\mathbf{l})^H\mathbf{R}^{-1}\mathbf{a}(\mathbf{l})}, \tag{25}$$

and the resulting MVDR dirty image can thus be described as

$$I_{MVDR}(\mathbf{l}) = \mathbf{w}(\mathbf{l})^H\mathbf{R}\mathbf{w}(\mathbf{l}) = \frac{1}{\mathbf{a}(\mathbf{l})^H\mathbf{R}^{-1}\mathbf{a}(\mathbf{l})}\,. \tag{26}$$

For a point-source model, this image will have a high resolution: two sources that are closely spaced will be resolved. The corresponding beam responses to different sources will in general be different: the beamshape is spatially varying. While we may represent $I_{MVDR}(\mathbf{l})$ as a convolution of the true image with a dirty beam, this is now a spatially varying convolution (viz. the convolution in a linear time-varying system). Deconvolution is still possible but has to take this into account.

Another consequence of the use of an adaptive beamformer is that the output noise power is not spatially uniform. Consider the data model $\mathbf{R} = \mathbf{A}\boldsymbol{\Sigma}_s\mathbf{A}^H + \boldsymbol{\Sigma}_n$, where $\boldsymbol{\Sigma}_n = \sigma_n^2\mathbf{I}$ is the noise covariance matrix, then at the output of the beamformer the noise power is, using (25),

$$I_n(\mathbf{l}) = \mathbf{w}(\mathbf{l})^H\mathbf{R}_n\mathbf{w}(\mathbf{l}) = \frac{\mathbf{a}(\mathbf{l})^H\mathbf{R}^{-1}(\sigma_n^2\mathbf{I})\mathbf{R}^{-1}\mathbf{a}(\mathbf{l})}{[\mathbf{a}(\mathbf{l})^H\mathbf{R}^{-1}\mathbf{a}(\mathbf{l})]^2} = \sigma_n^2\frac{\mathbf{a}(\mathbf{l})^H\mathbf{R}^{-2}\mathbf{a}(\mathbf{l})}{[\mathbf{a}(\mathbf{l})^H\mathbf{R}^{-1}\mathbf{a}(\mathbf{l})]^2}\,.$$

Thus, the output noise power is direction dependent.

As a remedy to this, a related beamformer which satisfies the constraint $\mathbf{w}(\mathbf{l})^H \mathbf{w}(\mathbf{l}) = 1$ (and therefore has spatially uniform output noise) is obtained by using a different scaling of the MVDR beamformer:

$$\mathbf{w}(\mathbf{l}) = \mu \mathbf{R}^{-1}\mathbf{a}(\mathbf{l}) , \qquad \mu = \frac{1}{[\mathbf{a}(\mathbf{l})^H \mathbf{R}^{-2}\mathbf{a}(\mathbf{l})]^{1/2}} .$$

This beamformer is known as the "Adapted Angular Response" (AAR) [8]. The resulting image is

$$I_{AAR}(\mathbf{l}) = \mathbf{w}(\mathbf{l})^H \mathbf{R}\mathbf{w}(\mathbf{l}) = \frac{\mathbf{a}(\mathbf{l})^H \mathbf{R}^{-1}\mathbf{a}(\mathbf{l})}{\mathbf{a}(\mathbf{l})^H \mathbf{R}^{-2}\mathbf{a}(\mathbf{l})} .$$

It has a high resolution and suppresses sidelobe interference under the white noise constraint.

Example MVDR and AAR dirty images using the same LOFAR stations as before are shown in Fig. 6. Comparing to Fig. 5, we observe that, as predicted, the sidelobe suppression in the MVDR and AAR dirty images is much better than the original matched beamformer dirty image. The images have a higher contrast and it appears that some additional point sources emerge as the result of lower sidelobe levels. This is especially true for the AAR dirty image.

4.2 Deconvolution

Having obtained a dirty image, we then attempt to recover the true image via *deconvolution*: inverting the effect of the (known) dirty beam.

4.2.1 The CLEAN Algorithm

A popular method for deconvolution is the CLEAN algorithm [27]. It was proposed for the classical, matched beamformer dirty image $I_D(\mathbf{l})$ defined in (17). From $I_D(\mathbf{l})$ and the known dirty beam $B(\mathbf{l})$, the desired image $I(\mathbf{l})$ is obtained via a sequential Least Squares fitting method. The algorithm is based on the assumption that the sky is mostly empty, and consists of a set of discrete point sources. The brightest source is estimated first, its contribution is subtracted from the dirty image, then the next brightest source is subtracted, etc.

The algorithm further uses the fact that $B(\mathbf{l})$ has its peak at the origin. Inside the loop, a candidate location $\mathbf{l}_q$ is selected as the location of the largest peak in $I_D(\mathbf{l})$, the corresponding power $\hat{\sigma}_q^2$ is estimated, and subsequently a small multiple

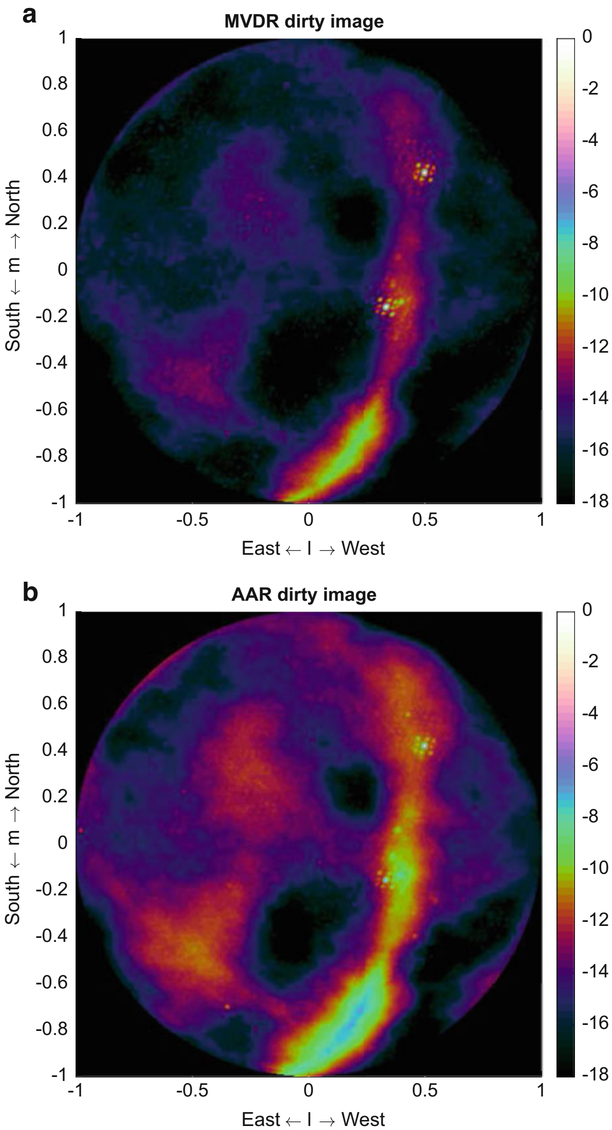

Fig. 6 Dirty images corresponding to the (**a**) MVDR and (**b**) AAR beamformers

of $\hat{\sigma}_q^2 B(\mathbf{l}-\mathbf{l}_q)$ is subtracted from $I_D(\mathbf{l})$. The objective is to minimize the residual, until it converges to the noise level:

```
$q = 0$
while $I_D(\mathbf{l})$ is not noise-like:
    $q = q + 1$
    $\mathbf{l}_q = \arg\max_{\mathbf{l}} I_D(\mathbf{l})$
    $\hat{\sigma}_q^2 = I_D(\mathbf{l}_q)/B(\mathbf{0})$
    $I_D(\mathbf{l}) := I_D(\mathbf{l}) - \gamma\hat{\sigma}_q^2 B(\mathbf{l}-\mathbf{l}_q)\,, \quad \forall\mathbf{l}$
$I_{clean}(\mathbf{l}) = I_D(\mathbf{l}) + \sum_q \gamma\hat{\sigma}_q^2 B_{synth}(\mathbf{l}-\mathbf{l}_q), \quad \forall\mathbf{l}\,.$
```

The scaling parameter $\gamma \leq 1$ is called the loop gain; for accurate convergence it should be small because the estimated location of the peak is at a grid point, whereas the true location of the peak may be in between grid points. $B_{synth}(\mathbf{l})$ is a "synthetic beam", usually a Gaussian bell-shape with about the same beam width as the main lobe of the dirty beam; it is introduced to mask the otherwise high artificial resolution of the image.

In current imaging systems, instead of the subtractions on the dirty image, it is considered more accurate to do the subtractions on the sample covariance matrix $\hat{\mathbf{R}}$ instead,

$$\hat{\mathbf{R}} := \hat{\mathbf{R}} - \gamma\hat{\sigma}_q^2\mathbf{a}(\mathbf{l}_q)\mathbf{a}(\mathbf{l}_q)^H$$

and then to recompute the dirty image. Computing a dirty image is the most expensive step in this loop, therefore usually a number of peaks are estimated from the dirty image together, the covariance is updated for this ensemble, and then the residual image is recomputed.

4.2.2 CLEAN Using Other Dirty Images

Instead of the matched beamformer dirty image $I_D(\mathbf{l})$, we can use other beamformed dirty images in the CLEAN loop, for example the MVDR dirty image. Due to its high resolution, the location of sources is better estimated than using the original dirty image (and the location estimate can be further improved by searching for the true peak on a smaller grid in the vicinity of the location of the maximum). A second modification to the CLEAN loop is also helpful: suppose that the location of the brightest source is $\mathbf{l}_q$, then the corresponding power α_q should be estimated by minimizing the residual $\|\mathbf{R} - \alpha\mathbf{a}(\mathbf{l}_q)\mathbf{a}(\mathbf{l}_q)^H\|^2$. This can be done in closed form: using (5) we find

$$\|\mathbf{R} - \alpha\mathbf{a}(\mathbf{l}_q)\mathbf{a}(\mathbf{l}_q)^H\| = \|\mathrm{vec}(\mathbf{R}) - \alpha[\bar{\mathbf{a}}(\mathbf{l}_q) \otimes \mathbf{a}(\mathbf{l}_q)]\|\,.$$

The optimal least squares solution for α is, using (1), (3) and (2) in turn,

$$\begin{aligned}\alpha_q &= [\bar{\mathbf{a}}(\mathbf{l}_q) \otimes \mathbf{a}(\mathbf{l}_q)]^\dagger \mathrm{vec}(\mathbf{R}) \\ &= \frac{[\bar{\mathbf{a}}(\mathbf{l}_q) \otimes \mathbf{a}(\mathbf{l}_q)]^H \mathrm{vec}(\mathbf{R})}{[\bar{\mathbf{a}}(\mathbf{l}_q) \otimes \mathbf{a}(\mathbf{l}_q)]^H [\bar{\mathbf{a}}(\mathbf{l}_q) \otimes \mathbf{a}(\mathbf{l}_q)]} \\ &= \frac{\mathbf{a}(\mathbf{l}_q)^H \mathbf{R} \mathbf{a}(\mathbf{l}_q)}{[\mathbf{a}(\mathbf{l}_q)^H \mathbf{a}(\mathbf{l}_q)]^2} \\ &= \frac{\mathbf{a}(\mathbf{l}_q)^H \mathbf{R} \mathbf{a}(\mathbf{l}_q)}{J^2},\end{aligned}$$

which is the power estimate of the matched filter. In the CLEAN loop, $\mathbf{R}$ should be replaced by its estimate $\hat{\mathbf{R}}$ minus the estimated components until q, and also a constraint that α_q is to be positive should be included. This method was proposed in [3].

Using the AAR dirty image in the CLEAN loop is also possible, and the resulting CLEANed image was called LS-MVI in [3].

4.3 Matrix Formulations

Because our data model is linear, it is beneficial to represent the covariance model and all subsequent operations on it in a linear algebra framework. In this more abstract formulation, details are hidden and it becomes easier to recognize the connection of image formation to standard formulations and more generic approaches, such as matrix inversion and parametric estimation techniques.

4.3.1 Matrix Formulation of the Data Model

Let us start again from the data model given by Eq. (12) assuming an ideal situation, in which all gain factors are unity. For simplicity, we consider only a single frequency bin and STI interval, but all results can be generalized straightforwardly. The model for the signals arriving at the antenna array is thus

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{n}(n)$$

and the covariance of $\mathbf{x}$ is (viz. (14))

$$\mathbf{R} = \mathbf{A}\boldsymbol{\Sigma}_s\mathbf{A}^H + \boldsymbol{\Sigma}_n\,.$$

We have available a sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_n \mathbf{x}(n)\mathbf{x}(n)^H$$

which serves as the input data for the imaging step. Let us now vectorize this data model by defining

$$\hat{\mathbf{r}} = \mathrm{vec}(\hat{\mathbf{R}}) \,, \qquad \mathbf{r} = \mathrm{vec}(\mathbf{R})$$

where $\mathbf{r}$ has the data model (using (4))

$$\mathbf{r} = (\bar{\mathbf{A}} \circ \mathbf{A})\boldsymbol{\sigma}_s + \mathrm{vec}(\boldsymbol{\Sigma}_n) \,.$$

If $\boldsymbol{\Sigma}_n$ is diagonal, we can write $\mathrm{vec}(\boldsymbol{\Sigma}_n) = (\bar{\mathbf{I}} \circ \mathbf{I})\boldsymbol{\sigma}_n$, where $\boldsymbol{\sigma}_n$ is a vector containing the diagonal entries of $\boldsymbol{\Sigma}_n$. Define $\mathbf{M}_s = \bar{\mathbf{A}} \circ \mathbf{A}$ and $\mathbf{M}_n = \mathbf{I} \circ \mathbf{I}$. Then

$$\mathbf{r} = \mathbf{M}_s\boldsymbol{\sigma}_s + \mathbf{M}_n\boldsymbol{\sigma}_n = [\mathbf{M}_s \;\; \mathbf{M}_n] \begin{bmatrix} \boldsymbol{\sigma}_s \\ \boldsymbol{\sigma}_n \end{bmatrix} = \mathbf{M}\boldsymbol{\sigma} \,. \tag{27}$$

In this formulation, several modifications can be introduced. E.g., a non-diagonal noise covariance matrix $\boldsymbol{\Sigma}_n$ will lead to a more general $\mathbf{M}_n$, while if $\boldsymbol{\Sigma}_n = \sigma_n^2\mathbf{I}$, we have $\mathbf{M}_n = \mathrm{vec}(\mathbf{I})$ and $\boldsymbol{\sigma}_n = \sigma_n^2$. Some other options are discussed in [47]. Also, if we have already an estimate of $\boldsymbol{\sigma}_n$, we can subtract it and write the model as

$$\mathbf{r}' := \mathbf{r} - \mathbf{M}_n\boldsymbol{\sigma}_n = \mathbf{M}_s\boldsymbol{\sigma}_s \tag{28}$$

The available measurements $\hat{\mathbf{r}}$ should be modified in the same way. This model is similar to (27), with the advantage that the number of unknown parameters in $\boldsymbol{\sigma}$ is smaller.

We can further write

$$\hat{\mathbf{r}} \;=\; \mathbf{r} + \mathbf{w} \;=\; \mathbf{M}\boldsymbol{\sigma} + \mathbf{w} \,, \tag{29}$$

where $\hat{\mathbf{r}}$ is the available "measurement data", $\mathbf{r}$ is its mean (expected value), and $\mathbf{w}$ is additive noise due to finite samples. It is not hard to derive that (for Gaussian signals) the covariance of this noise is [47]

$$\mathbf{C}_w = E(\hat{\mathbf{r}} - \mathbf{r})(\hat{\mathbf{r}} - \mathbf{r})^H = \frac{1}{N}(\bar{\mathbf{R}} \otimes \mathbf{R})$$

where N is the number of samples on which $\hat{\mathbf{R}}$ is based. We have thus written our original data model on $\mathbf{x}$ as a similar data model on $\hat{\mathbf{r}}$. Many estimation techniques from the literature that are usually applied to data models for $\mathbf{x}$ can be applied to the data model for $\mathbf{r}$. Furthermore, it is straightforward to extend this vectorized formulation to include multiple snapshots over time and frequency to increase the amount of measurement data and thus to improve the imaging result: Simply stack the covariance data in $\hat{\mathbf{r}}$ and include the model structure in $\mathbf{M}$; note that $\boldsymbol{\sigma}$ remains unchanged. Similarly, assuming a diagonal noise covariance matrix, astronomers often drop the autocorrelation terms (diagonal of $\hat{\mathbf{R}}$), rather than attempting to do the subtraction in (28); this corresponds to dropping rows in $\mathbf{M}$ and corresponding rows in $\mathbf{M}_s$, and leads to a model similar to (28) but without the autocorrelation terms.

The unknown parameters in the data model are, first of all, the powers $\boldsymbol{\sigma}$. These appear linear in the model. Regarding the positions of the sources, we can consider two cases:

1. We consider a point source model with a "small" number of sources. In that case, $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{M} = \mathbf{M}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is some parameterization of the unknown locations of the sources (the position vectors $\mathbf{l}_q$ for each source). These enter in a nonlinear way into the model $\mathbf{M}(\boldsymbol{\theta})$. The image $I(\mathbf{l})$ is constructed following (16), usually convolved with a synthetic beam $B_{synth}(\mathbf{l})$ to make the image look nicer. The resulting estimation techniques are very much related to *direction of arrival* (DOA) estimation in array signal processing, with a rich literature.
2. Alternatively, we consider a model where, for each pixel in the image, we assume a corresponding point source: the source positions $\mathbf{l}_q$ directly correspond to the pixels in the image. This can lead to a large number of sources. With the locations of the pixels predetermined, $\mathbf{M}$ is a priori known and not a function of $\boldsymbol{\theta}$, but $\mathbf{M}$ will have many columns (one for each pixel-source). The image $I(\mathbf{l})$ has a one-to-one relation to the source power vector $\boldsymbol{\sigma}_s$, we can thus regard $\boldsymbol{\sigma}_s$ as the image in this case.

We need to pose several requirements on $\mathbf{M}$ or $\mathbf{M}(\boldsymbol{\theta})$ to ensure identifiability. First of all, in the first case we must have $\mathbf{M}(\boldsymbol{\theta}) = \mathbf{M}(\boldsymbol{\theta}') \;\rightarrow\; \boldsymbol{\theta} = \boldsymbol{\theta}'$, otherwise we cannot uniquely find $\boldsymbol{\theta}$ from $\mathbf{M}$. Furthermore, for both cases we will require that $\mathbf{M}$ is a tall matrix (more rows than columns) and has full column rank, so that it has a left inverse (this will allow to estimate $\boldsymbol{\sigma}$). This puts a limit on the number of sources in the image (number of columns of $\mathbf{M}$) in relation to the number of observations (rows). If more snapshots (STIs) and/or multiple frequencies are available, as is the case in practice, then $\mathbf{M}$ will become taller, and more sources can be estimated thus increasing the resolution. If $\mathbf{M}$ is not tall, then there are some ways to generalize this using prior knowledge on the image, e.g. via the context of compressive sampling where we can have $\mathbf{M}$ wide as long as $\boldsymbol{\sigma}$ is sparse [59], which we will briefly discuss in Sect. 4.5.5.

For the moment, we will continue with the second formulation: one source per pixel, fewer pixels than available correlation data.

4.3.2 Matrix Formulation of Imaging via Beamforming

Let us now again interpret the "beamforming image" (22) as a linear transformation on the covariance data $\hat{\mathbf{r}}$. We can stack all image values $I(\mathbf{l})$ over all pixels $\mathbf{l}_q$ into a single vector $\mathbf{i}$, and similarly, we can collect the weights $\mathbf{w}(\mathbf{l})$ over all pixels into a single matrix $\mathbf{W} = [\mathbf{w}(\mathbf{l}_1), \cdots, \mathbf{w}(\mathbf{l}_Q)]$. From (3), we know that $\mathbf{w}^H\mathbf{R}\mathbf{w} = (\overline{\mathbf{w}} \otimes \mathbf{w})^H \text{vec}(\hat{\mathbf{R}})$, so that we can write

$$\hat{\mathbf{i}}_{BF} = (\overline{\mathbf{W}} \circ \mathbf{W})^H \hat{\mathbf{r}}\,. \tag{30}$$

We saw before that the dirty image is obtained if we use the matched filter. In this case, we have $\mathbf{W} = \frac{1}{J}\mathbf{A}$, where $\mathbf{A}$ contains the array response vectors $\mathbf{a}(\mathbf{l})$ for every pixel $\mathbf{l}_q$ of interest. In this case, the image is

$$\hat{\mathbf{i}}_D = \frac{1}{J^2}(\bar{\mathbf{A}} \circ \mathbf{A})^H \hat{\mathbf{r}} = \frac{1}{J^2}\mathbf{M}_s^H \hat{\mathbf{r}}\,. \tag{31}$$

The expected value of the image is obtained by using $\mathbf{r} = \mathbf{M}\boldsymbol{\sigma}$:

$$\mathbf{i}_D = \frac{1}{J^2}\mathbf{M}_s^H \mathbf{M}\boldsymbol{\sigma} = \frac{1}{J^2}(\mathbf{M}_s^H \mathbf{M}_s)\boldsymbol{\sigma}_s + \frac{1}{J^2}(\mathbf{M}_s^H \mathbf{M}_n)\boldsymbol{\sigma}_n\,.$$

The quality or "performance" of the image, or how close $\hat{\mathbf{i}}_D$ is to $\mathbf{i}_D$, is related to its covariance,

$$\text{cov}(\hat{\mathbf{i}}_D) = E\{(\hat{\mathbf{i}}_D - \mathbf{i}_D)(\hat{\mathbf{i}}_D - \mathbf{i}_D)^H\} = \frac{1}{J^4}\mathbf{M}_s^H \mathbf{C}_w \mathbf{M}_s$$

where $\mathbf{C}_w = \frac{1}{N}(\bar{\mathbf{R}} \otimes \mathbf{R})$ is the covariance of the noise on the covariance data. Since usually the astronomical sources are much weaker than the noise (often at least by a factor 100), we can approximate $\mathbf{R} \approx \boldsymbol{\Sigma}_n$. If the noise is spatially white, $\boldsymbol{\Sigma}_n = \sigma_n^2\mathbf{I}$, we obtain for the covariance of $\hat{\mathbf{i}}_D$

$$\text{cov}(\hat{\mathbf{i}}_D) \approx \frac{\sigma_n^4}{J^4 N}\mathbf{M}_s^H \mathbf{M}_s\,.$$

The variance in the image is given by the diagonal of this expression. From this and the structure of $\mathbf{M}_s = (\bar{\mathbf{A}} \circ \mathbf{A})$ and the structure of $\mathbf{A}$, we can see that the variance on each pixel in the dirty image is constant, $\sigma_n^4/(J^2 N)$, but that the noise on the image is correlated, possibly leading to visible structures in the image. This is a general phenomenon. Similar equations can be derived for the MVDR image and the AAR image.

4.4 Parametric Image Estimation

In Sect. 4.2, we discussed various deconvolution algorithms based on the CLEAN algorithm. This algorithm uses a successive approximation of the dirty image using a point source model. Alternatively, we take a model-based approach. The imaging problem is formulated as a parametric estimation problem where certain parameters (source locations, powers, noise variance) are unknown and need to be estimated. Although we start from a Maximum Likelihood formulation, we will quickly arrive at a more feasible Least Squares approach. The discussion was presented in [45] and follows to some extent [47], which is a general array processing approach to a very similar problem and can be read for further details.

4.4.1 Weighted Least Squares Imaging

The image formation problem can be formulated as a maximum likelihood (ML) estimation problem, and solving this problem should provide a statistically efficient estimate of the parameters. Since all signals are assumed to be i.i.d. Gaussian signals, the derivation is standard and the ML estimates are obtained by minimizing the negative log-likelihood function [47]

$$\{\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\theta}}\} = \arg\min_{\boldsymbol{\sigma},\boldsymbol{\theta}} \ \ln|\mathbf{R}(\boldsymbol{\sigma},\boldsymbol{\theta})| + \mathrm{tr}\left(\mathbf{R}^{-1}(\boldsymbol{\sigma},\boldsymbol{\theta})\hat{\mathbf{R}}\right) \tag{32}$$

where $|\cdot|$ denotes the determinant. $\mathbf{R}(\boldsymbol{\sigma},\boldsymbol{\theta})$ is the model, i.e., $\mathrm{vec}(\mathbf{R}(\boldsymbol{\sigma},\boldsymbol{\theta})) = \mathbf{r} = \mathbf{M}(\boldsymbol{\theta})\boldsymbol{\sigma}$, where $\boldsymbol{\theta}$ parameterizes the source locations, and $\boldsymbol{\sigma}$ their intensities.

We will first consider the overparameterized case, where $\boldsymbol{\theta}$ is a (known) list of all pixel coordinates in the image, and each pixel corresponds to a source. In this case, $\mathbf{M}$ is a priori known, the model is linear, and the ML problem reduces to a Weighted Least Squares (WLS) problem to match $\hat{\mathbf{r}}$ to the model $\mathbf{r}$:

$$\hat{\boldsymbol{\sigma}} = \arg\min_{\boldsymbol{\sigma}} \|\mathbf{C}_w^{-1/2}(\hat{\mathbf{r}} - \mathbf{r})\|_2^2 = \arg\min_{\boldsymbol{\sigma}} \ (\hat{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})^H \mathbf{C}_w^{-1}(\hat{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}) \tag{33}$$

where we fit the "data" $\hat{\mathbf{r}}$ to the model $\mathbf{r} = \mathbf{M}\boldsymbol{\sigma}$. The correct weighting is the inverse of the covariance of the residual, $\mathbf{w} = \hat{\mathbf{r}} - \mathbf{r}$, i.e., the noise covariance matrix $\mathbf{C}_w = \frac{1}{N}(\bar{\mathbf{R}} \otimes \mathbf{R})$. For this, we may also use the estimate $\hat{\mathbf{C}}_w$ obtained by using $\hat{\mathbf{R}}$ instead of $\mathbf{R}$. Using the assumption that the astronomical sources are much weaker than the noise we could contemplate to use $\mathbf{R} \approx \boldsymbol{\Sigma}_n$ for the weighting. If the noise is spatially white, $\boldsymbol{\Sigma}_n = \sigma_n^2\mathbf{I}$, the weighting can then even be omitted.

The solution of (33) is obtained by applying the pseudo-inverse,

$$\hat{\boldsymbol{\sigma}} = [\mathbf{C}_w^{-1/2}\mathbf{M}]^\dagger \mathbf{C}_w^{-1/2}\hat{\mathbf{r}} = (\mathbf{M}^H\mathbf{C}_w^{-1}\mathbf{M})^{-1}\mathbf{M}^H\mathbf{C}_w^{-1}\hat{\mathbf{r}} =: \mathbf{M}_d^{-1}\hat{\boldsymbol{\sigma}}_d \tag{34}$$

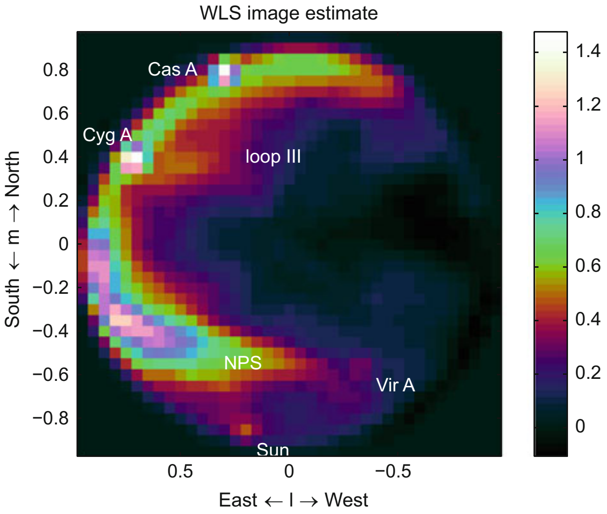

Fig. 7 Image corresponding to the WLS formulation (34)

where

$$\mathbf{M}_d := \mathbf{M}^H \mathbf{C}_w^{-1} \mathbf{M}\,, \qquad \hat{\boldsymbol{\sigma}}_d := \mathbf{M}^H \mathbf{C}_w^{-1} \hat{\mathbf{r}}\,.$$

Here, we can consider the term $\hat{\boldsymbol{\sigma}}_d = \mathbf{M}^H \mathbf{C}_w^{-1} \hat{\mathbf{r}}$ as a "dirty image": it is comparable to (31), although we have introduced a weighting by $\mathbf{C}_w^{-1}$ and estimate the noise covariance parameters $\boldsymbol{\sigma}_n$ as well as the source powers in $\boldsymbol{\sigma}_s$ (the actual image). The factor $1/J^2$ in (31) can be seen as a crude approximation of $\mathbf{M}_d^{-1}$.

Figure 7 shows an example WLS image for a single LOFAR station. The image was obtained by deconvolving the dirty image from 25 STIs, each consisting of 10 s data in 25 frequency channels of 156 kHz wide taken from the band 45–67 MHz, avoiding the locally present radio interference. As this shows data from a single LOFAR station, with a relatively small maximal baseline (65 m), the resolution is limited and certainly not representative of the capabilities of the full LOFAR array. The resolution (number of pixels) in this image is kept limited (about 1000) for reasons discussed below.

The term $\mathbf{M}_d^{-1} = (\mathbf{M}^H \mathbf{C}_w^{-1} \mathbf{M})^{-1}$ is a deconvolution operation. This inversion can only be carried out if the deconvolution matrix $\mathbf{M}_d = \mathbf{M}^H \mathbf{C}_w^{-1} \mathbf{M}$ is not rank deficient. This requires at least that $\mathbf{M}$ is a tall matrix ("less pixels than observations" in case we take one source per pixel). Thus, high resolution WLS imaging is only possible if a limited number of sources is present. The condition number of $\mathbf{M}_d$, i.e., the ratio of the largest to the smallest eigenvalue of $\mathbf{M}_d$, gives important information on our ability to compute its inverse: LS theory tells us that the noise on $\hat{\boldsymbol{\sigma}}_d$ could, in the worst case, be magnified by this factor. The optimal (smallest) condition number of any matrix is 1, which is achieved if $\mathbf{M}_d$ is a scaling of the identity matrix, or if the columns of $\mathbf{C}_w^{-1/2}\mathbf{M}$ are all orthogonal to each other. If the size of $\mathbf{M}$ becomes less tall, then the condition number of $\mathbf{M}_d$ becomes larger (worse), and once it is a wide matrix, $\mathbf{M}$ is singular and the condition number will be infinite. Thus, we have a trade-off between the resolution (number of pixels in the image) and the noise enhancement.

The definition of $\mathbf{M}_d$ shows that it is not data dependent, and it can be precomputed for a given telescope configuration and observation interval. It is thus possible to explore this trade-off beforehand. To avoid numerical instabilities (noise enhancement), we would usually compute a regularized inverse or pseudo-inverse of this matrix, e.g., by first computing the eigenvalue decomposition

$$\mathbf{M}_d = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$$

where $\mathbf{U}$ contains the (orthonormal) eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues, sorted from large to small. Given a threshold ϵ on the eigenvalues, we can define $\tilde{\boldsymbol{\Lambda}}$ to be a diagonal matrix containing only the eigenvalues larger than ϵ, and $\tilde{\mathbf{U}}$ a matrix containing the corresponding eigenvectors. The ϵ-threshold pseudo-inverse is then given by

$$\mathbf{M}_d^{\dagger} := \tilde{\mathbf{U}}\tilde{\boldsymbol{\Lambda}}^{-1}\tilde{\mathbf{U}}^H$$

and the resulting image is

$$\boldsymbol{\sigma} = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Lambda}}^{-1}\tilde{\mathbf{U}}^H\boldsymbol{\sigma}_d\,. \tag{35}$$

This can be called the “Karhunen-Loève” image, as the rank reduction is related to the Karhunen-Loève transform (KLT). It corresponds to selecting an optimal (Least Squares) set of basis vectors on which to project a certain data set, here $\boldsymbol{\sigma}_d$.

An example KLT image is shown in Fig. 8. In this image, the number of pixels is much larger than before in Fig. 7 (about 9000), but the rank of the matrix $\mathbf{M}_d$ is truncated at 1/200 times the largest eigenvalue, leaving about 1300 out of 9000 image components. The result is not quite satisfactory: the truncation to a reduced basis results in annoying ripple artefacts in the image.

Computing the eigenvalue decomposition for large matrices is complex. A computationally simpler alternative is to compute a regularized inverse of $\mathbf{M}_d$, i.e., to take the inverse of $\mathbf{M}_d + \epsilon\mathbf{I}$. This should yield similar (although not identical) results.

If we use the alternative sky model where we assume a point source model with a “small” number of sources ($\mathbf{M} = \mathbf{M}(\boldsymbol{\theta})$), then the conditioning of $\mathbf{M}_d$, and thus the performance of the deconvolution, is directly related to the number of sources and their spatial distribution.

The performance of the method is assessed by looking at the covariance of the resulting image (plus noise parameters) $\hat{\boldsymbol{\sigma}}$ in (34). It is given by

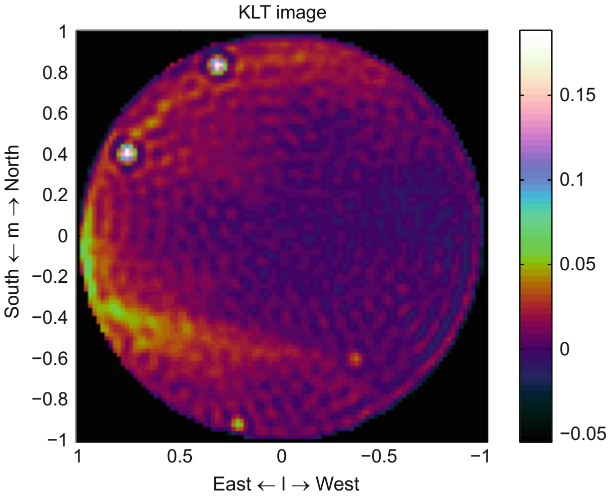

Fig. 8 Image corresponding to the KLT solution (35)

$$\begin{aligned}\mathbf{C}_{\boldsymbol{\sigma}} &= (\mathbf{M}^H\mathbf{C}_w^{-1}\mathbf{M})^{-1}\mathbf{M}^H\mathbf{C}_w^{-1}(\mathbf{C}_w)\mathbf{C}_w^{-1}\mathbf{M}(\mathbf{M}^H\mathbf{C}_w^{-1}\mathbf{M})^{-1} \\ &= (\mathbf{M}^H\mathbf{C}_w^{-1}\mathbf{M})^{-1} \quad = \quad \mathbf{M}_d^{-1}\,.\end{aligned}$$

This again shows that the performance of the imaging method follows directly from the conditioning of the deconvolution matrix $\mathbf{M}_d$. If $\mathbf{M}_d$ is sufficiently well conditioned, the noise on the image is limited, otherwise it may be large. The formulation also shows that the pixels in the image are correlated ($\mathbf{M}_d$ is in general not diagonal), as we obtained before for the dirty image.

Similarly, if we use the pseudo-inverse $\mathbf{M}_d^{\dagger} = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Lambda}}^{-1}\tilde{\mathbf{U}}^H$ for the deconvolution, then we obtain $\mathbf{C}_{\boldsymbol{\sigma}} = \mathbf{M}_d^{\dagger}$. In this case, the noise enhancement depends on the chosen threshold ϵ. Also, the rank of $\mathbf{C}_{\boldsymbol{\sigma}}$ depends on this threshold, and since it is not full rank, the number of independent components (sources) in the image is smaller than the number of shown pixels: the rank reduction defines a form of interpolation.

Using a rank truncation for radio astronomy imaging was already suggested in [10]. Unfortunately, if the number of pixels is large, this technique by itself is not sufficient to obtain good images, e.g., the resulting pixels may not all be positive, which is unplausible for an intensity image. Thus, the overparameterized case requires additional constraints; some options are discussed in Sects. 4.5.4 and 4.5.5.

4.4.2 Estimating the Position of the Sources

Let us now consider the use of the alternative formulation, where we write $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{M} = \mathbf{M}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ captures the positions of the limited number of sources in the image. In this case, we have to estimate both $\boldsymbol{\sigma}$ and $\boldsymbol{\theta}$. If we start again from the ML formulation (32), it does not seem feasible to solve this minimization problem in closed form. However, we can again resort to the WLS covariance matching problem and solve instead

$$\begin{aligned}\{\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\theta}}\} &= \arg\min_{\boldsymbol{\sigma},\boldsymbol{\theta}} \|\mathbf{C}_w^{-1/2}[\hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\sigma}, \boldsymbol{\theta})]\|^2 \\ &= \arg\min_{\boldsymbol{\sigma},\boldsymbol{\theta}} [\hat{\mathbf{r}} - \mathbf{M}(\boldsymbol{\theta})\boldsymbol{\sigma}]^H \mathbf{C}_w^{-1}[(\hat{\mathbf{r}} - \mathbf{M}(\boldsymbol{\theta})\boldsymbol{\sigma}]\,. \end{aligned} \tag{36}$$

It is known that the resulting estimates are, for a large number of samples, equivalent to ML estimates and therefore asymptotically efficient [47].

The WLS problem is separable: suppose that the optimal $\boldsymbol{\theta}$ is known, so that $\mathbf{M} = \mathbf{M}(\boldsymbol{\theta})$ is known, then the corresponding $\boldsymbol{\sigma}$ will satisfy the solution which we found earlier:

$$\hat{\boldsymbol{\sigma}} = (\mathbf{M}^H \mathbf{C}_w^{-1}\mathbf{M})^{-1}\mathbf{M}^H \mathbf{C}_w^{-1}\hat{\mathbf{r}}\,.$$

Substituting this solution back into the problem, we obtain

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\min_{\boldsymbol{\theta}}\ \hat{\mathbf{r}}^H[\mathbf{I} - \mathbf{M}(\boldsymbol{\theta})(\mathbf{M}(\boldsymbol{\theta})^H\mathbf{C}_w^{-1}\mathbf{M}(\boldsymbol{\theta}))^{-1}\mathbf{M}(\boldsymbol{\theta})^H\mathbf{C}_w^{-1}]^H\ \cdot \\ &\qquad \cdot\ \mathbf{C}_w^{-1}\cdot[\mathbf{I} - \mathbf{M}(\boldsymbol{\theta})(\mathbf{M}(\boldsymbol{\theta})^H\mathbf{C}_w^{-1}\mathbf{M}(\boldsymbol{\theta}))^{-1}\mathbf{M}(\boldsymbol{\theta})^H\mathbf{C}_w^{-1}]\hat{\mathbf{r}} \\ &= \arg\min_{\boldsymbol{\theta}}\ \hat{\mathbf{r}}^H\mathbf{C}_w^{-1/2}(\mathbf{I} - \boldsymbol{\Pi}(\boldsymbol{\theta}))\mathbf{C}_w^{-1/2}\hat{\mathbf{r}} \\ &= \arg\max_{\boldsymbol{\theta}}\ \hat{\mathbf{r}}^H\mathbf{C}_w^{-1/2}\boldsymbol{\Pi}(\boldsymbol{\theta})\mathbf{C}_w^{-1/2}\hat{\mathbf{r}}\end{aligned}$$

where $\boldsymbol{\Pi}(\boldsymbol{\theta}) = \mathbf{C}_w^{-1/2}\mathbf{M}(\boldsymbol{\theta})\big(\mathbf{M}(\boldsymbol{\theta})^H\mathbf{C}_w^{-1}\mathbf{M}(\boldsymbol{\theta})\big)^{-1}\mathbf{M}(\boldsymbol{\theta})^H\mathbf{C}_w^{-1/2}$.

$\boldsymbol{\Pi}(\boldsymbol{\theta})$ is an orthogonal projection: $\boldsymbol{\Pi}^2 = \boldsymbol{\Pi},\ \boldsymbol{\Pi}^H = \boldsymbol{\Pi}$. The projection is onto the column span of $\mathbf{M}'(\boldsymbol{\theta}) := \mathbf{C}_w^{-1/2}\mathbf{M}(\boldsymbol{\theta})$. The estimation of the source positions $\boldsymbol{\theta}$ is nonlinear. It could be obtained iteratively using a Newton iteration (cf. [47]). The sources can also be estimated sequentially [47], which provides an alternative to the CLEAN algorithm.

4.4.3 Preconditioned WLS

WLS imaging can be improved using preconditioning, and this has an interesting relation to the adaptive beamforming techniques discussed earlier. From this point forward we assume that an estimate of the noise has been subtracted from the images as in (28) such that $\mathbf{M} = \mathbf{M}_s$ and $\boldsymbol{\sigma} = \boldsymbol{\sigma}_s$.

If $\mathbf{M}$ has full column rank then $\mathbf{H}_{\text{LS}} := \mathbf{M}^H\mathbf{M}$ and $\mathbf{H}_{\text{WLS}} := \mathbf{M}^H\mathbf{C}_w^{-1}\mathbf{M}$ are non-singular and there exists a unique solution to LS and WLS. For example the solution to the LS imaging becomes

$$\boldsymbol{\sigma} = \mathbf{H}_{\text{LS}}^{-1}\hat{\boldsymbol{\sigma}}_{\text{D}} \tag{37}$$

where $\hat{\boldsymbol{\sigma}}_{\text{D}} = \mathbf{M}^H\hat{\mathbf{r}}$ is the estimated dirty image. Unfortunately, if the number of pixels is large then $\mathbf{H}_{\text{LS}}$ and $\mathbf{H}_{\text{WLS}}$ become ill-conditioned or even singular. Generally, we need to improve the conditioning of the deconvolution matrices and to find appropriate regularizations.

One way to improve the conditioning of a matrix is by applying a preconditioner. The most widely used and simplest preconditioner is the Jacobi preconditioner [1] which, for any matrix $\mathbf{M}$, is given by $[\text{diag}(\mathbf{M})]^{-1}$. Let $\mathbf{D}_{\text{WLS}} = \text{diag}(\mathbf{H}_{\text{WLS}})$, then by applying this preconditioner to $\mathbf{H}_{\text{WLS}}$ we obtain

$$[\mathbf{D}_{\text{WLS}}^{-1}\mathbf{H}_{\text{WLS}}]\boldsymbol{\sigma} = \mathbf{D}_{\text{WLS}}^{-1}\hat{\boldsymbol{\sigma}}_{\text{WLS}} \tag{38}$$

where $\hat{\boldsymbol{\sigma}}_{\text{WLS}} = \mathbf{M}^H\mathbf{C}_w^{-1}\hat{\mathbf{r}}$. We take a closer look at $\mathbf{D}_{\text{WLS}}^{-1}\hat{\boldsymbol{\sigma}}_{\text{WLS}}$. For a single STI

$$\begin{aligned}\mathbf{H}_{\text{WLS}} &= (\bar{\mathbf{A}} \circ \mathbf{A})^H(\hat{\mathbf{R}}^{-T} \otimes \hat{\mathbf{R}}^{-1})(\bar{\mathbf{A}} \circ \mathbf{A}) \\ &= (\mathbf{A}^T\hat{\mathbf{R}}^{-T}\bar{\mathbf{A}}) \odot (\mathbf{A}^H\hat{\mathbf{R}}^{-1}\mathbf{A})\end{aligned}$$

and

$$\mathbf{D}_{\text{WLS}}^{-1} = \begin{bmatrix} \frac{1}{(\mathbf{a}_1^H\hat{\mathbf{R}}^{-1}\mathbf{a}_1)^2} & & \\ & \ddots & \\ & & \frac{1}{(\mathbf{a}_Q^H\hat{\mathbf{R}}^{-1}\mathbf{a}_Q)^2} \end{bmatrix}, \tag{39}$$

where we have assumed that $\mathbf{a}_i$ is normalized by a factor $1/\sqrt{J}$ such that $\mathbf{a}_i^H\mathbf{a}_i = 1$. This means that

$$\begin{aligned}\mathbf{D}_{\text{WLS}}^{-1}\hat{\boldsymbol{\sigma}}_{\text{WLS}} &= \mathbf{D}_{\text{WLS}}^{-1}\left((\hat{\mathbf{R}}^{-T} \otimes \hat{\mathbf{R}}_1^{-1})(\bar{\mathbf{A}} \circ \mathbf{A})\right)^H\hat{\mathbf{r}} \\ &= (\hat{\mathbf{R}}^{-T}\bar{\mathbf{A}}\mathbf{D}_{\text{WLS}}^{-1/2} \circ \hat{\mathbf{R}}^{-1}\mathbf{A}\mathbf{D}_{\text{WLS}}^{-1/2})^H\hat{\mathbf{r}}\end{aligned}$$

which is equivalent to a dirty image that is obtained by applying a beamformer of the form

$$\mathbf{w}_i = \frac{1}{\mathbf{a}_i^H\hat{\mathbf{R}}^{-1}\mathbf{a}_i}\hat{\mathbf{R}}^{-1}\mathbf{a}_i \tag{40}$$

to both sides of $\hat{\mathbf{R}}$ and stacking the results, $\hat{\sigma}_i = \mathbf{w}_i^H \hat{\mathbf{R}} \mathbf{w}_i$, of each pixel into a vector. This beamformer is the MVDR beamformer which we have introduced before! This shows that the Preconditioned WLS (PWLS) image (motivated from its connection to the maximum likelihood) is expected to exhibit the features of high-resolution beamforming associated with the MVDR. The PWLS was introduced in [45].

4.5 Constraints on the Image

Another approach to improve the conditioning of a problem is to introduce appropriate constraints on the solution. Typically, image formation algorithms exploit external information regarding the image in order to regularize the ill-posed problem. For example maximum entropy techniques [21] impose a smoothness condition on the image while the CLEAN algorithm [27] exploits a point source model wherein most of the image is empty, and this has recently been connected to sparse optimization techniques [59].

4.5.1 Non-negativity Constraint

A lower bound on the image is almost trivial: each pixel in the image represents the intensity at a certain direction, hence is non-negative. This is physically plausible, and to some extent already covered by CLEAN [41]. It is an explicit condition in a Non-Negative Least Squares (NNLS) formulation [10], which searches for a Least Squares fit while requiring that the solution $\boldsymbol{\sigma}$ has all entries $\sigma_i \geq 0$:

$$\begin{aligned} &\min_{\boldsymbol{\sigma}} \; \|\hat{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}\|^2 \\ &\text{subject to } \mathbf{0} \leq \boldsymbol{\sigma} \end{aligned} \tag{41}$$

4.5.2 Dirty Image as Upper Bound

A second constraint follows if we also know an upper bound $\boldsymbol{\gamma}$ such that $\boldsymbol{\sigma} \leq \boldsymbol{\gamma}$, which will bound the pixel intensities from above. We will propose several choices for $\boldsymbol{\gamma}$.

By closer inspection of the ith pixel of the matched beamformer dirty image $\hat{\boldsymbol{\sigma}}_{\mathrm{D}}$, we note that its expected value is given by

$$\sigma_{\mathrm{D},i} = \mathbf{a}_i^H \mathbf{R} \mathbf{a}_i \,.$$

Using normalization $\mathbf{a}_i^H \mathbf{a}_i = 1$, we obtain

$$\sigma_{\mathrm{D},i} = \sigma_i + \mathbf{a}_i^H \mathbf{R}_r \mathbf{a}_i \,, \tag{42}$$

where

$$\mathbf{R}_r = \sum_{j \neq i} \sigma_j \mathbf{a}_j \mathbf{a}_j^H + \mathbf{R}_\mathbf{n} \tag{43}$$

is the contribution of all other sources and the noise. Note that $\mathbf{R}_r$ is positive-(semi)definite. Thus, (42) implies $\sigma_{\mathrm{D},i} \geq \sigma_i$ which means that the expected value of the matched beamformer dirty image forms an upper bound for the desired image, or

$$\boldsymbol{\sigma} \leq \boldsymbol{\sigma}_\mathrm{D} . \tag{44}$$

We can extend this concept to a more general beamformer $\mathbf{w}_i$. The output power of this beamformer, in the direction of the ith pixel, becomes

$$\sigma_{\mathbf{w},i} = \mathbf{w}_i^H \mathbf{R} \mathbf{w}_i = \sigma_i \mathbf{w}_i^H \mathbf{a}_i \mathbf{a}_i^H \mathbf{w}_i + \mathbf{w}_i^H \mathbf{R}_r \mathbf{w}_i . \tag{45}$$

If we require that

$$\mathbf{w}_i^H \mathbf{a}_i = 1 \tag{46}$$

we have

$$\sigma_{\mathbf{w},i} = \sigma_i + \mathbf{w}_i^H \mathbf{R}_r \mathbf{w}_i . \tag{47}$$

As before, the fact that $\mathbf{R}_r$ is positive definite implies that

$$\sigma_i \leq \sigma_{\mathbf{w},i} . \tag{48}$$

We can easily verify that the matched filter weights $\mathbf{w}_{\mathrm{D},i}$ as given in (20) satisfy (46) and, hence, that the resulting dirty image $\sigma_{\mathrm{D},i}$ is a specific upper bound.

4.5.3 Tightest Upper Bound

The next question is: What is the tightest upper bound for σ_i that we can construct using linear beamforming?

We can translate the problem of finding the tightest upper bound to the following optimization question:

$$\begin{aligned} \sigma_{\mathrm{opt},i} &= \min_{\mathbf{w}_i} \mathbf{w}_i^H \mathbf{R} \mathbf{w}_i \\ &\text{s.t. } \mathbf{w}_i^H \mathbf{a}_i = 1 \end{aligned} \tag{49}$$

where $\sigma_{\mathrm{opt},i}$ would be this tightest upper bound. This optimization problem is exactly the same as the one used in Sect. 4.1.2 to obtain the MVDR beamformer. Hence

$$\mathbf{w}_i = \frac{1}{\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i} \mathbf{R}^{-1} \mathbf{a}_i .$$

This means that for a single STI the MVDR image is the tightest upper bound that can be constructed using beamformers.

Note that $\mathbf{w}_{\mathrm{D},i}$ also satisfies the constraint in (46), i.e. $\mathbf{w}_{\mathrm{D},i}^H \mathbf{a}_i = \mathbf{a}_i^H \mathbf{a}_i = 1$, but does not necessary minimize the output power $\mathbf{w}_i^H \mathbf{R} \mathbf{w}_i$, therefore the MVDR dirty image is smaller than the matched beamformer dirty image: $\boldsymbol{\sigma}_{\mathrm{MVDR}} \le \boldsymbol{\sigma}_{\mathrm{D}}$. This relation also holds if $\mathbf{R}$ is replaced by the sample covariance $\hat{\mathbf{R}}$.

For multiple snapshots the tightest bound can be obtained by taking the minimum of the individual MVDR estimates [44]. The bound becomes

$$\sigma_{\mathrm{opt},i} = \min_m \frac{1}{\mathbf{a}_{m,i} \mathbf{R}_m^{-1} \mathbf{a}_{m,i}} .$$

One problem with using this result in practice is that $\sigma_{\mathrm{opt},i}$ depends on a single snapshot. Actual dirty images are based on the sample covariance matrix $\hat{\mathbf{R}}$ and hence they are random variables. If we use a sample covariance matrix $\hat{\mathbf{R}}$ instead of the true covariance matrix $\mathbf{R}$, this bound would be too noisy without any averaging. Hence we would like to find a beamformer that exhibits the same averaging behavior as the matched beamformer while being as tight as possible. Sardarabadi [44] shows that a modified multi-snapshot MVDR image can be defined as

$$\sigma_{\mathrm{MVDR},i} = \frac{1}{\frac{1}{M} \sum_m \mathbf{a}_{m,i}^H \mathbf{R}_m^{-1} \mathbf{a}_{m,i}} , \tag{50}$$

which satisfies $\sigma_i \le \sigma_{\mathrm{MVDR},i} \le \sigma_{\mathrm{D},i}$ and produces a very tight bound.

4.5.4 Constrained WLS Imaging

Now that we have lower and upper bounds on the image, we can use these as constraints in the LS imaging problem to provide a regularization. The resulting constrained LS (CLS) imaging problem is

$$\begin{aligned} &\min_{\boldsymbol{\sigma}} \; \|\hat{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}\|^2 \\ &\text{s.t. } \mathbf{0} \le \boldsymbol{\sigma} \le \boldsymbol{\gamma} \end{aligned} \tag{51}$$

where $\boldsymbol{\gamma}$ can be chosen either as $\boldsymbol{\gamma} = \boldsymbol{\sigma}_{\mathrm{D}}$ for the matched beamformer dirty image or $\boldsymbol{\gamma} = \boldsymbol{\sigma}_{\mathrm{MVDR}}$ for the MVDR dirty image.

The extension to constrained WLS leads to the problem formulation

$$\begin{aligned} &\min_{\boldsymbol{\sigma}} \ \|\mathbf{C}_w^{-1/2}\left(\hat{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}\right)\|^2 \\ &\text{s.t. } \mathbf{0} \le \boldsymbol{\sigma} \le \boldsymbol{\gamma}\,. \end{aligned} \tag{52}$$

It is also recommended to include a preconditioner which, as was shown in Sect. 4.4.3, relates the WLS to the MVDR dirty image. However, because of the inequality constraints, (52) does not have a closed form solution and it is solved by an iterative algorithm. In order to have the relation between the WLS and MVDR dirty image during the iterations we introduce a change of variables of the form $\check{\boldsymbol{\sigma}} = \mathbf{D}\boldsymbol{\sigma}$, where $\check{\boldsymbol{\sigma}}$ is the new variable for the preconditioned problem and the diagonal matrix $\mathbf{D}$ is given in (39). The resulting constrained preconditioned WLS (CPWLS) optimization problem is

$$\begin{aligned} \check{\boldsymbol{\sigma}} = \arg\min_{\check{\boldsymbol{\sigma}}} \ &\|\mathbf{C}_w^{-1/2}\left(\hat{\mathbf{r}} - \mathbf{M}\mathbf{D}^{-1}\check{\boldsymbol{\sigma}}\right)\|^2 \\ &\text{s.t. } \mathbf{0} \le \check{\boldsymbol{\sigma}} \le \mathbf{D}\boldsymbol{\gamma} \end{aligned} \tag{53}$$

and the final image is found by setting $\boldsymbol{\sigma} = \mathbf{D}^{-1}\check{\boldsymbol{\sigma}}$. Here we used that $\mathbf{D}$ is a positive diagonal matrix so that the transformation to an upper bound for $\check{\boldsymbol{\sigma}}$ is correct. As mentioned, the dirty image that follows from the (unconstrained) Weighted Least Squares part of the problem is given by the MVDR image $\hat{\boldsymbol{\sigma}}_{\text{MVDR}}$.

These problems are convex and their solutions can be found using various numerical optimization techniques such as the active set method, as discussed in more detail in [45]. Some experimental results using non-negative constraints are shown in [23, 37, 51].

4.5.5 Imaging Using Sparse Reconstruction Techniques

Compressive sampling/sensing (CS) is a "new" topic, currently drawing wide attention. It is connected to random or non-uniform sampling, and as such, it has been used in radio astronomy for a long time. In the CS community, the recovery of full information from undersampled data is the central problem, and to regularize this problem, the main idea has been to exploit the sparsity of the solution: the number of nonzero entries in the solution is supposed to be small. This is measured by the ℓ_0-norm: $\|\boldsymbol{\sigma}\|_0$ is the number of nonzero entries in $\boldsymbol{\sigma}$. Optimizing using this norm is difficult, and therefore as a surrogate, the ℓ_1-norm is used.

To introduce this, let us start from the Least Squares formulation, and consider the KLT regularization. This constrains the solution image to lie on a basis determined by the dominant column span of $\mathbf{M}$ (possibly giving rise to artefacts). It is straightforward to show that this regularization is connected to adding a regularization term

$$\min_{\boldsymbol{\sigma}} \|\hat{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}\|_2^2 + \lambda\|\boldsymbol{\sigma}\|_2$$

where λ is related to the truncation threshold used in the KLT. The used norm on $\boldsymbol{\sigma}$ is ℓ_2, the sum of squares, or the total "energy" of the image.

An alternative to this is to use a regularization term $\|\boldsymbol{\sigma}\|_1$ based on the ℓ_1 norm of $\boldsymbol{\sigma}$, or the sum of absolute values [35, 59]. The resulting problem is

$$\min_{\boldsymbol{\sigma}} \|\hat{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}\|_2^2 + \lambda\|\boldsymbol{\sigma}\|_1$$

An alternative formulation of this problem is

$$\min_{\boldsymbol{\sigma}} \|\boldsymbol{\sigma}\|_1 \quad \text{subject to} \quad \|\hat{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}\|_2^2 \leq \epsilon$$

where ϵ is threshold on the residual noise. Like for KLT, the results depend on the chosen noise threshold ϵ (or regularization parameter λ).

Minimizing the ℓ_1-norm is known to promote the sparsity of the solution vector. The implied sparsity assumption in the model poses that the sky is mostly empty. Although it has already long been suspected that CLEAN is related to ℓ_1-optimization [41] (in fact, it is now recognized as a Matching Pursuit algorithm [39]), CS theory states the general conditions under which this assumption is likely to recover the true image [35, 59]. Extensions are needed in case of extended emissions [37]. As images may consist of sources with different source structures, different sources may be best represented, i.e., best compressible, by different bases. This is the basic idea behind the Sparsity Averaging Reweighted Analysis (SARA) algorithm, which aims to find the sparsest representation using an overdetermined dictionary composed of multiple complete bases [11, 12].

4.5.6 Comparison of Regularization Techniques

In this section, we discussed a number of constraints to regularize the ill-posed inverse imaging problem: non-negativity, upper bound, and sparsity of the image. This can be combined into a single problem,

$$\begin{aligned} \min_{\check{\boldsymbol{\sigma}}} \; & \|\mathbf{C}_w^{-1/2}\left(\hat{\mathbf{r}} - \mathbf{M}\mathbf{D}^{-1}\check{\boldsymbol{\sigma}}\right)\|^2 + \lambda\|\mathbf{D}^{-1}\boldsymbol{\sigma}\| \\ & \text{s.t. } \mathbf{0} \leq \check{\boldsymbol{\sigma}} \leq \mathbf{D}\boldsymbol{\gamma} \end{aligned} \tag{54}$$

where $\mathbf{D}$ is an optional preconditioner, the resulting image is $\boldsymbol{\sigma} = \mathbf{D}^{-1}\check{\boldsymbol{\sigma}}$, and the norm is either ℓ_1 or ℓ_2. Many variations on this problem are possible. Taken by itself, the non-negativity constraint is already known to be a strong constraint for regularization. It can even be shown that, when certain conditions are satisfied, the non-negativity constraint alone already promotes a sparse solution [20]. In cases where there is a combination of sparse and extended structures in the image, an ℓ_2 regularization might be more appropriate.

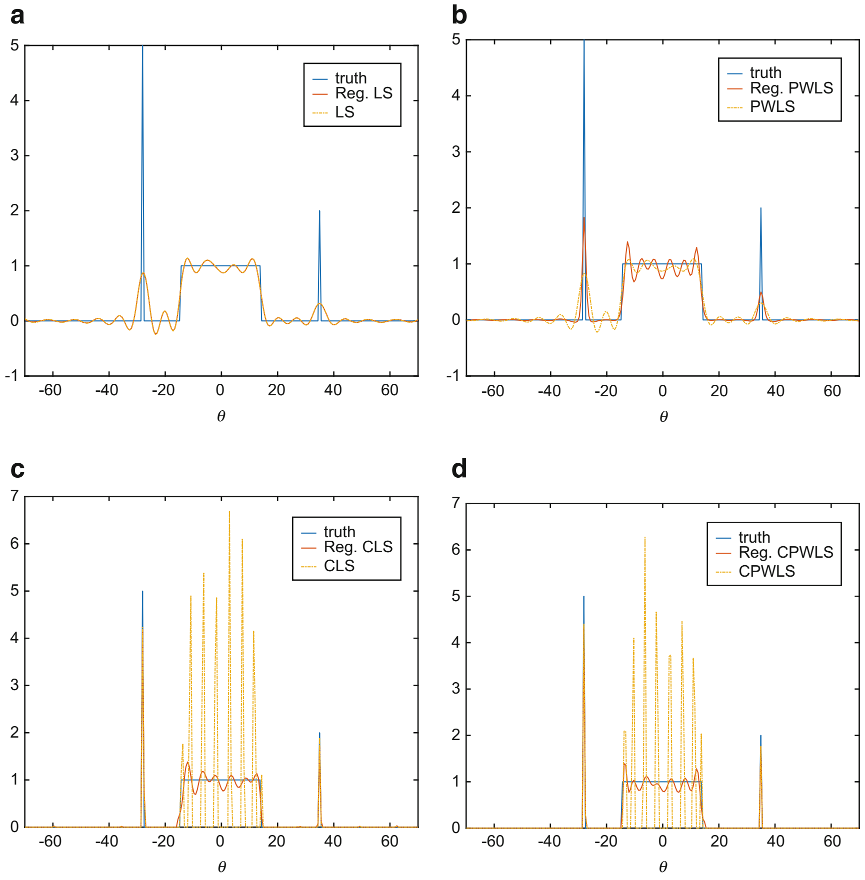

Fig. 9 Solutions for different algorithms with and without regularization; (**a**) Unconstrained LS. (**b**) Unconstrained PWLS. (**c**) Constrained LS. (**d**) Constrained PWLS

To illustrate the effects of regularization, constraints, and preconditioning, we consider a 1D "image" reconstruction example. A uniform linear array (ULA) with 20 receivers is simulated. The array is exposed to two point sources with magnitudes 5 and 2 and an extended rectangular source with a magnitude of 1. Because it is a ULA, $\text{rank}(\mathbf{M}) = 2J - 1 = 39$, while the number of pixels is $Q = 245$. This shows that $\mathbf{H}_{\text{LS}} = \mathbf{M}^H\mathbf{M}$ is singular. We use ℓ_2-norm regularization with a regularization coefficient $\lambda = 1/\sqrt{N}$ where $N = 1000$ is the number of samples in a single STI.

Figure 9 shows the result of the various estimation techniques with and without bound constraints and regularization. Figure 9a shows the result of standard LS with and without regularization, Fig. 9b shows similar results for unconstrained Preconditioned WLS, Fig. 9c incorporates the bound constraints for the LS problem, and Fig. 9d shows the results for CPWLS.

The figures show the following:

- Both standard LS and PWLS are unable to recover the point sources and suffer from high sidelobe levels. The regularization does not seem to affect the LS solution while it improves the sidelobe behavior in the PWLS solution at the cost of less accurate estimates for the extended structure.
- Both Constrained LS and Constrained PWLS without regularization attempt to model the extended structure using a series of point sources. This is the consequence of the non-negativity constraint which tends to promote sparsity.
- For CLS and CPWLS an ℓ_2-norm regularization helps with the recovery of the extended structure. The value of $\lambda = 1/\sqrt{N}$ seems to be a good balance for both extended and point sources.

5 Calibration

5.1 Non-ideal Measurements

In the previous section we showed that there are many options to make an image from radio interferometric measurements. However, we assumed that these measurements were done under ideal circumstances, such that the gain matrix in our data model given by (14) only contains ones. In practice, there are several effects that make matters more complicated causing $\mathbf{G} \neq \mathbf{1}\mathbf{1}^H$ (where we omitted the STI index m for convenience of notation as we will initially consider calibration on a per-STI basis). These effects need to be estimated and corrected for in a process called *calibration*. For this, some reference information is needed. In this section, we will assume that the locations and powers of Q reference sources are known, where Q can be small (order 1 to 10) or large (up to a complete image). In practice, calibration is an integral part of the imaging step, and not a separate phase as we will see in Sect. 6. The model given by (14) is not identifiable in its generality unless we make some assumptions on the structure of $\mathbf{G}$ (in the form of a suitable parameterization) and describe how it varies with time and frequency, e.g., in the form of (stochastic) models for these variations.

The effects captured by the gain matrix $\mathbf{G}$ can be subdivided in instrumental effects and propagation effects. We start by describing a few basic effects as understanding those will help to establish a suitable representation of the gain matrix.

5.1.1 Instrumental Effects

The instrumental effects consist of the directional response of the receiving elements (antennas) and the direction-independent electronic gains and phases of the receivers.

The directional response or *primary beam* of the receiving elements in the array can be described by a function $b_j(\mathbf{l})$, where we have assumed that this function is constant over the time and frequency span of the STI. It is generally assumed that the primary beam is equal for all elements in the array. With Q point sources, we will collect the resulting samples of the primary beam into a vector $\mathbf{b} = [b(\mathbf{l}_1), \cdots, b(\mathbf{l}_Q)]^T$. These coefficients are seen as gains that (squared) will multiply the source powers σ_q^2. The general shape of the primary beam $b(\mathbf{l})$ is known from electromagnetic modeling during the design of the telescope. If this is not sufficiently accurate, then it has to be calibrated, which is typically done off-line in the lab.

Next, each receiver element in the array is connected to a receiver chain (low-noise amplifier, bandpass filter, down-modulator), and initially the direction-independent electronic gains and phases of each receiver chain are unknown and have to be estimated. They are generally different from element to element. We thus have an unknown vector $\mathbf{g}$ (size $J \times 1$) with complex entries that each multiply the output signal of each telescope. As the direction independent gains are identical for all Q sources while the direction dependent response is identical for all elements, the gain matrix can be factored as $\mathbf{G} = \mathbf{g}\mathbf{b}^H$. By introducing the diagonal matrices $\boldsymbol{\Gamma} = \text{diag}(\mathbf{g})$ and $\mathbf{B} = \text{diag}(\mathbf{b})$, we can write $\mathbf{G} \odot \mathbf{A} = \boldsymbol{\Gamma}\mathbf{A}\mathbf{B}$.

Also the noise powers of each element are unknown and generally unequal to each other. We will still assume that the noise is independent from element to element. We can thus model the noise covariance matrix by an (unknown) diagonal $\boldsymbol{\Sigma}_n$.

For instrumental calibration, we can thus reformulate our data model in (14) to

$$\mathbf{R} = (\boldsymbol{\Gamma}\mathbf{A}\mathbf{B})\boldsymbol{\Sigma}_s(\mathbf{B}^H\mathbf{A}^H\boldsymbol{\Gamma}^H) + \boldsymbol{\Sigma}_n \tag{55}$$

Usually, $\boldsymbol{\Gamma}$ and $\mathbf{B}$ are considered to vary only slowly with time m and frequency k.

5.1.2 Propagation Effects

Ionospheric and tropospheric turbulence cause time-varying refraction and diffraction, which has a profound effect on the propagation of radio waves. In the simplest case, the ionosphere is modeled as a thin layer at some height (say 100 km) above the Earth, causing delays that can be represented as phase shifts. At the low frequencies used for LOFAR, this effect is more pronounced. Generally it is first assumed that the ionosphere is "constant" over about 10 km and about 10 s. A better model is to model the ionospheric delay as a "wedge", a linear function of the distance between piercing points (the intersection of the direction vectors $\mathbf{l}_q$ with the ionospheric phase screen). As illustrated in Fig. 10, this modifies the geometric delays, leading to a shift in the apparent position of the sources. For larger distances, higher-order functions are needed to model the spatial behaviour of the ionosphere, and if left uncorrected, the resulting image distortions are comparable to the distortions one sees when looking at lights at the bottom of a swimming pool.

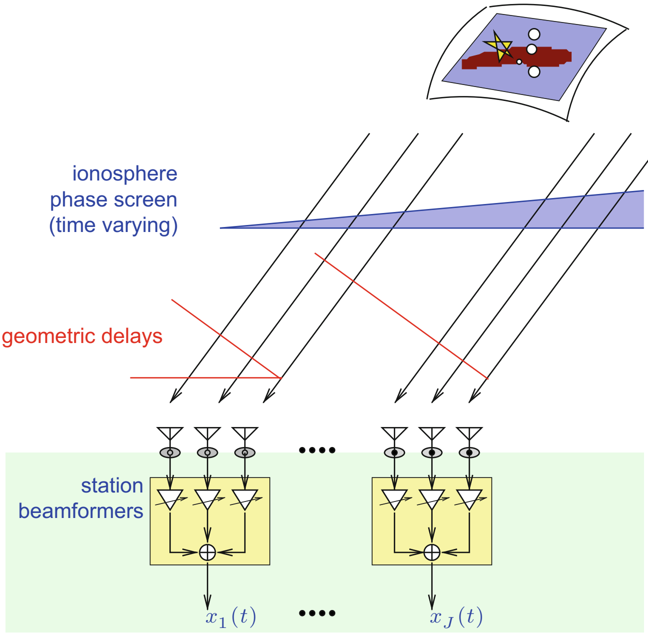

Fig. 10 A radio interferometer where stations consisting of phased array elements replace telescope dishes. The ionosphere adds phase delays to the signal paths. If the ionospheric electron density has the form of a *wedge*, it will simply shift the apparent positions of all sources

Previously, we described the array response matrix $\mathbf{A}$ as a function of the source direction vectors $\mathbf{l}_q$, and we wrote $\mathbf{A}(\boldsymbol{\theta})$ where the vector $\boldsymbol{\theta}$ was a suitable parameterization of the $\mathbf{l}_q$ (typically two direction cosines per source). If a linear model for the ionospheric disturbance is sufficient, then it is sufficient to replace $\mathbf{A}(\boldsymbol{\theta})$ by $\mathbf{A}(\boldsymbol{\theta}')$, where $\boldsymbol{\theta}'$ differs from $\boldsymbol{\theta}$ due to the shift in apparent direction of each source.

The modified data model that captures the above effects is thus

$$\mathbf{R} = (\boldsymbol{\Gamma}\mathbf{A}(\boldsymbol{\theta}')\mathbf{B})\boldsymbol{\Sigma}_s(\mathbf{B}^H\mathbf{A}(\boldsymbol{\theta}')^H\boldsymbol{\Gamma}^H) \;+\; \boldsymbol{\Sigma}_n\,. \tag{56}$$

In the next subsection, we will first describe how models of the form (55) or (56) can be identified. This step will serve as a stepping stone in the identification of a more general $\mathbf{G}$.

5.2 *Calibration Algorithms*

5.2.1 Estimating the Element Gains and Directional Responses

Let us assume a model of the form (55), where there are Q dominant calibration sources within the field of view. For these sources, we assume that their positions and source powers are known with sufficient accuracy from tables, i.e., we assume that $\mathbf{A}$ and $\mathbf{\Sigma}_s$ are known. We can then write (55) as

$$\mathbf{R} = \boldsymbol{\Gamma}\mathbf{A}\mathbf{\Sigma}\mathbf{A}^H\boldsymbol{\Gamma}^H + \mathbf{\Sigma}_n \tag{57}$$

where $\mathbf{\Sigma} = \mathbf{B}\mathbf{\Sigma}_s\mathbf{B}$ is a diagonal matrix with apparent source powers. With $\mathbf{B}$ unknown, $\mathbf{\Sigma}$ is unknown, but estimating $\mathbf{\Sigma}$ is precisely the problem we studied in Sect. 4 when we discussed imaging. Thus, once we have estimated $\mathbf{\Sigma}$ and know $\mathbf{\Sigma}_s$, we can easily estimate the directional gains $\mathbf{B}$. The problem thus reduces to estimate the diagonal matrices $\boldsymbol{\Gamma}$, $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_n$ from a model of the form (57).

For some cases, e.g., arrays where the elements are traditional telescope dishes, the field of view is quite narrow (degrees) and we may assume that there is only a single calibrator source in the observation. Then $\mathbf{\Sigma} = \sigma^2$ is a scalar and the problem reduces to

$$\mathbf{R} = \mathbf{g}\sigma^2\mathbf{g}^H + \mathbf{\Sigma}_n$$

and since $\mathbf{g}$ is unknown, we could even absorb the unknown σ in $\mathbf{g}$ (it is not separately identifiable). The structure of $\mathbf{R}$ is a rank-1 matrix $\mathbf{g}\sigma^2\mathbf{g}^H$ plus a diagonal $\mathbf{\Sigma}_n$. This is recognized as a "rank-1 factor analysis" model in multivariate analysis theory [32, 40]. Given $\mathbf{R}$, we can solve for $\mathbf{g}$ and $\mathbf{\Sigma}_n$ in several ways [6, 7, 64]. For example, any submatrix away from the diagonal is only dependent on $\mathbf{g}$ and is rank 1. This allows direct estimation of $\mathbf{g}$. This property is related to the gain and phase closure relations often used in the radio astronomy literature for calibration (in particular, these relations express that the determinant of any 2×2 submatrix away from the main diagonal will be zero, which is the same as saying that this submatrix is rank 1).

In general, there are more calibrator sources (Q) in the field of view, and we have to solve (57). A simple idea is to resort to an Alternating Least Squares approach. If $\boldsymbol{\Gamma}$ would be known, then we can correct $\mathbf{R}$ for it, so that we have precisely the same problem as we considered before, (33), and we can solve for $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_n$ using the techniques discussed in Sect. 4.4.1. Alternatively, with $\mathbf{\Sigma}$ known, we can say we know a reference model $\mathbf{R}_0 = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^H$, and the problem is to identify the element gains $\boldsymbol{\Gamma} = \text{diag}(\mathbf{g})$ from a model of the form

$$\mathbf{R} = \boldsymbol{\Gamma}\mathbf{R}_0\boldsymbol{\Gamma}^H + \mathbf{\Sigma}_n$$

or, after applying the vec(·)-operation,

$$\mathrm{vec}(\mathbf{R}) = \mathrm{diag}(\mathrm{vec}(\mathbf{R}_0))(\overline{\mathbf{g}} \otimes \mathbf{g}) + \mathrm{vec}(\boldsymbol{\Sigma}_n) .$$

This leads to the Least Squares problem

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \|\mathrm{vec}(\hat{\mathbf{R}} - \boldsymbol{\Sigma}_n) - \mathrm{diag}(\mathrm{vec}(\mathbf{R}_0))(\overline{\mathbf{g}} \otimes \mathbf{g})\|^2 .$$

This problem cannot be solved in closed form. Alternatively, we can first solve an unstructured problem: define $\mathbf{x} = \overline{\mathbf{g}} \otimes \mathbf{g}$ and solve

$$\hat{\mathbf{x}} = \mathrm{diag}(\mathrm{vec}(\mathbf{R}_0))^{-1}\mathrm{vec}(\hat{\mathbf{R}} - \boldsymbol{\Sigma}_n)$$

or equivalently, if we define $\mathbf{X} = \mathbf{g}\mathbf{g}^H$,

$$\hat{\mathbf{X}} = (\hat{\mathbf{R}} - \boldsymbol{\Sigma}_n) \oslash \mathbf{R}_0.$$

where $\oslash$ denotes an element-wise matrix division. After estimating the unstructured $\mathbf{X}$, we enforce the rank-1 structure $\mathbf{X} = \mathbf{g}\mathbf{g}^H$, via a rank-1 approximation, and find an estimate for $\mathbf{g}$. The element-wise division can lead to noise enhancement; this is remediated by only using the result as an initial estimate for a Gauss-Newton iteration [22] or by formulating a *weighted* least squares problem instead [61, 64].

With $\mathbf{g}$ known, we can again estimate $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_n$, and make an iteration. Overall we then obtain an alternating least squares solution. A more optimal solution can be found by solving the overall problem (57) as a covariance matching problem with a suitable parameterization, and the more general gradient descent algorithms (e.g., Gauss-Newton and Levenberg-Marquardt) presented in [47] lead to an asymptotically unbiased and statistically efficient solution.

For large arrays, Gauss-Newton iterations or weighted least squares approaches become computationally expensive as they scale cubicly with the number of receiving elements in the array. Several people have therefore proposed an iterative alternating direction implicit (ADI) method [25, 42, 50], which was demonstrated to have robust convergence and to be statistically efficient for typical scenarios encountered in radio astronomy in which the noise powers dominate over the source powers and are very similar for all elements in the array [50].

The resulting calibration algorithms are one step in the classical self-calibration (SelfCal) algorithm [15, 48] widely used in the radio astronomy literature, in particular for a single calibrator source. In the calibration step of SelfCal, $\mathbf{R}_0$ is a reference model, obtained from the best known map at that point in the iteration. Next, in the imaging step of SelfCal, the calibration results are used to correct the data $\hat{\mathbf{R}}$ and the next best image is constructed. This leads to a new reference model $\mathbf{R}_0$, etc.

5.2.2 Estimating the Ionospheric Perturbation

The more general calibration problem (56) follows from (55) by writing $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta}')$ where $\boldsymbol{\theta}'$ are the apparent source locations. This problem can be easily solved in quite the same way: in the alternating least squares problem we solve for $\mathbf{g}$, $\boldsymbol{\theta}'$, $\boldsymbol{\sigma}_s$ and $\boldsymbol{\sigma}_n$ in turn, keeping the other parameters fixed at their previous estimates. After that, we can relate the apparent source locations to the (known) locations of the calibrator sources $\boldsymbol{\theta}$.

The resulting phase corrections $\mathbf{A}'$ to relate $\mathbf{A}(\boldsymbol{\theta}')$ to $\mathbf{A}(\boldsymbol{\theta})$ via $\mathbf{A}(\boldsymbol{\theta}') = \mathbf{A}(\boldsymbol{\theta}) \odot \mathbf{A}'$ give us an estimate of the ionospheric phase screen in the direction of each source. These "samples" can then be interpolated to obtain a phase screen model for the entire field of view. This method is limited to the regime where the phase screen can be modeled as a linear gradient over the array. An implementation of this algorithm is called Field-Based Calibration [16].

Other techniques are based on "peeling" [42]. In this method of successive estimation and subtraction, calibration parameters are obtained for the brightest source in the field. The source is then removed from the data, and the process is repeated for the next brightest source. This leads to a collection of samples of the ionosphere, to which a model phase screen can be fitted.

5.2.3 Estimating the General Model

In the more general case (14), viz.

$$\mathbf{R} = (\mathbf{G} \odot \mathbf{A})\boldsymbol{\Sigma}_s(\mathbf{G} \odot \mathbf{A})^H + \boldsymbol{\Sigma}_n ,$$

we have an unknown full matrix $\mathbf{G}$. We assume $\mathbf{A}$ and $\boldsymbol{\Sigma}_s$ known. Since $\mathbf{A}$ element-wise multiplies $\mathbf{G}$ and $\mathbf{G}$ is unknown, we might as well omit $\mathbf{A}$ from the equations without loss of generality. For the same reason also $\boldsymbol{\Sigma}_s$ can be omitted. This leads to a problem of the form

$$\mathbf{R} = \mathbf{G}\mathbf{G}^H + \boldsymbol{\Sigma}_n ,$$

where the $J \times Q$ matrix $\mathbf{G}$ and $\boldsymbol{\Sigma}_n$ (diagonal) are unknown. This problem is known as a rank-Q factor analysis problem. Note that if the noise would be spatially white ($\boldsymbol{\Sigma}_n = \sigma_n^2\mathbf{I}$), then $\mathbf{G}$ can be solved from an eigenvalue decomposition of $\mathbf{R}$, up to a unitary factor at the right.

The more general Factor Analysis problem is a classical problem in multivariate statistics that has been studied since the 1930s [32, 40]. Currently, FA is an important and popular tool for latent variable analysis with many applications in various fields of science [2]. However, its application within the signal processing community has been surprisingly limited. The problem can be regarded as a special case of covariance matching, studied in detail in [47]. Thus, the problem can be solved using Gauss-Newton iterations. The current algorithms are robust and have a computational complexity similar to that of an eigenvalue decomposition of $\mathbf{R}$ [44].

It is important to note that $\mathbf{G}$ can be identified only up to a unitary factor $\mathbf{V}$ at the right: $\mathbf{G}' = \mathbf{G}\mathbf{V}$ would also be a solution. This factor makes the gains unidentifiable unless we introduce more structure to the problem. To make matters worse, note that this problem is used to fine-tune earlier coarser models (56). At this level of accuracy, the number of dominant sources Q is often not small anymore, and at some point $\mathbf{G}$ is not identifiable: counting number of equations and unknowns, we find that the maximum factor rank is limited by $Q < J - \sqrt{J}$.

As discussed in [46] and studied in more detail in [55], more structure needs to be introduced to be able to solve the problem. Typically, what helps is to consider the problem for a complete observation (rather than for a single snapshot $\mathbf{R}$) where we have many different frequencies f_k and time intervals m. The directional response matrix $\mathbf{A}_{m,k}$ varies with m and k in a known way, and the instrumental gains $\mathbf{g}$ and $\mathbf{b}$ are relatively constant. The remaining part of $\mathbf{G} = \mathbf{g}\mathbf{b}^H \odot \mathbf{A}'$ is due to the ionospheric perturbations, and models can be introduced to describe its fluctuation over time, frequency, and space using some low order polynomials. We can also introduce stochastic knowledge that describe a correlation of parameters over time and space.

For LOFAR, a complete calibration method that incorporates many of the above techniques was recently proposed in [28]. In general, calibration and imaging need to be considered in unison, leading to many potential directions, approaches, and solutions. Once calibration reaches the stage of full image calibration at the full resolution, we basically try to identify a highly detailed parametric model using gradient descent techniques. The computational complexity can be very high. To limit this, SAGEcal [31] clusters parameters into non-overlapping sets associated with different directions on the sky, solves the "independent" problems separately, and then combines in a parameter-fusing step. Distributed SAGEcal [66] also exploits parallelism such as continuity over time and frequency, again solving "independent" problems separately in parallel, followed by a fusion step.

6 A Typical Signal Processing Pipeline

To conclude this chapter, we discuss how calibration and imaging techniques are put together to form an imaging pipeline. We do this using a pipeline developed to guide the design of the SKA computing systems [29, 30] as an example. If the receiving elements of such a system are phased array stations, as is the case for the low-frequency system of the SKA, an end-to-end imaging pipeline consists of three stages of processing: Station Beamforming, processing in the Central Signal Processor (CSP), and the Science Data Processor (SDP). Block diagrams for each stage are shown in Figs. 11, 12 and 13.

Figure 11 shows a typical block diagram for signal processing within a phased array station. The signals from the receiving elements within a station are digitized

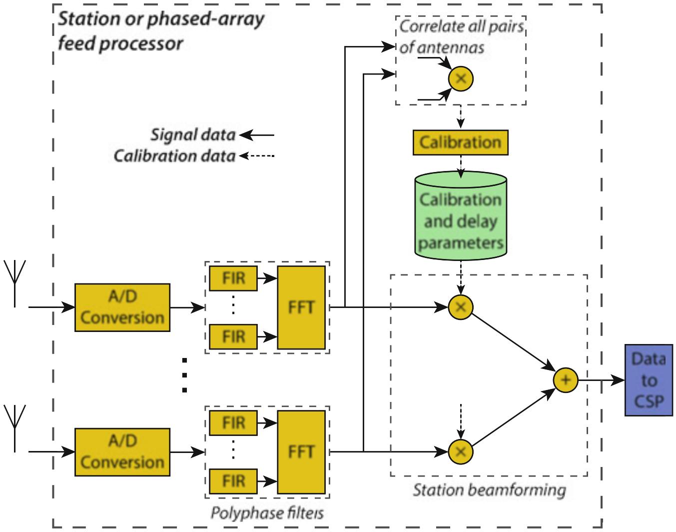

Fig. 11 Typical block diagram for signal processing within a phased array station [29, 30]

and combined into a single beamformed output, providing a well-defined beam on the sky. This is usually done by a standard delay beamformer by applying weights as described in (20). As the delays are represented by phase shifts, the signals need to be narrowband with respect to this delay. This is ensured by splitting the digitized signal of each receiver path into multiple coarse frequency channels (typically order (a few) 100 kHz wide) by a polyphase filter bank. The time series produced for each of these coarse channels can also be fed into a correlator to produce array covariance matrices for the station. These covariance matrices can be used to perform calibration. Usually, this only concerns direction independent gain calibration as described in Sect. 5.2.1. Those calibration solutions can be used to adapt the beamformer weights to correct for complex valued gain differences between receive paths. The beamformed output of each phased array station is sent to the CSP for further processing.

Figure 12 shows the block diagram for the signal processing within the CSP of the SKA. The goal of the CSP is to combine data from the receiving elements of the SKA interferometer by correlating its input signals. As the signals can be integrated after correlation, this step can significantly reduce the data volume using relatively simple operations. The input signals are either beamformed signals from phased array stations or coarsely channelized signals from reflector dishes. As the longest baselines of the SKA interferometer are much longer than the size of an individual station, much narrower frequency channels are required to satisfy the narrowband assumption discussed in Sect. 3.2. This is achieved by a second polyphase filter bank, which splits the coarse frequency channels further into fine channels (typically order 1 kHz wide). Any residual time delay across the array remaining after the coarse delay correction done by shifting time series with respect

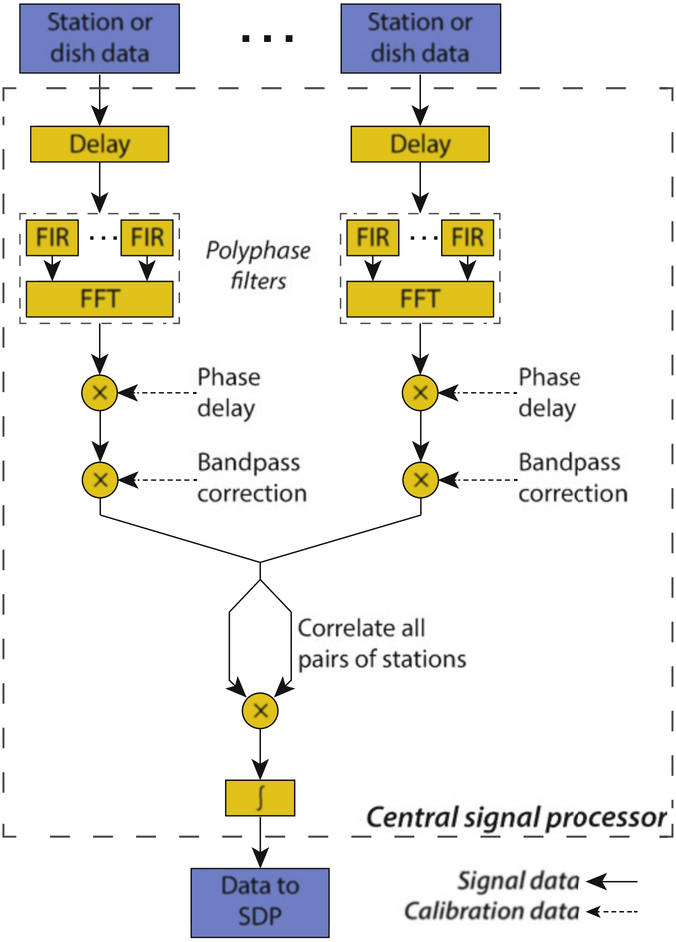

Fig. 12 Block diagram for data processing in the Central Signal Processor (CSP) of the SKA [29, 30]

to each other before the polyphase filter bank, is then corrected by applying an appropriate phase rotation. As the power received in individual frequency channels may vary significantly across frequency due to the intrinsic spectrum of most astronomical sources and the gain characteristics of the instrument, a bandpass correction is applied to equalize the power across frequency before the signals are correlated. After correlation, the data is integrated into STIs and data corrupted by radio frequency interference (RFI) is flagged before the data is transferred to the SDP.

A block diagram for an imaging pipeline within the SDP is shown in Fig. 13. After some pre-processing, consisting of demixing, integration and initial calibration, a self-calibration and imaging cycle is started.

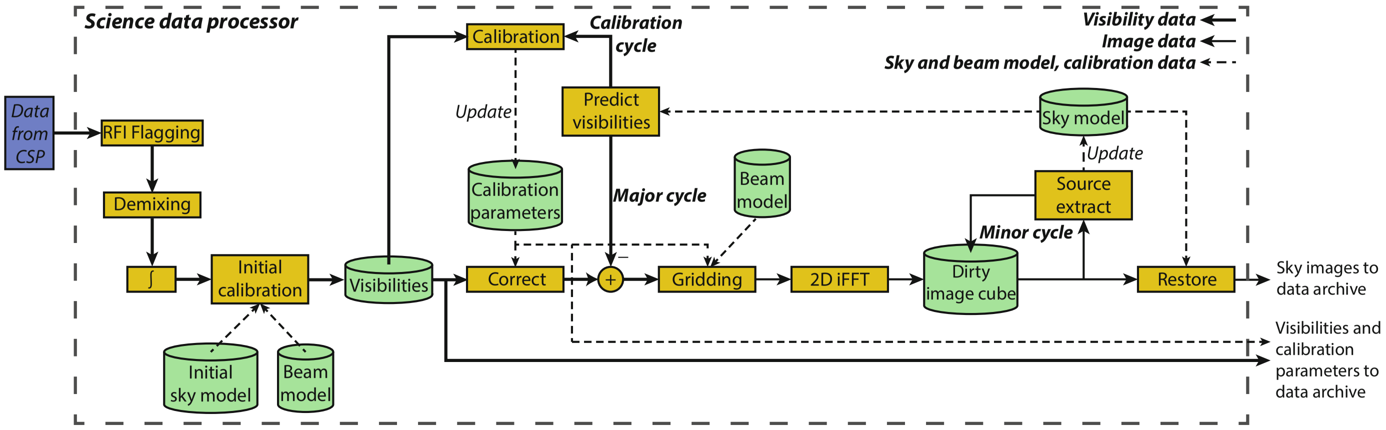

Fig. 13 Block diagram for the imaging pipeline in the Science Data Processor of the SKA [29, 30]

We first discuss the pre-processing steps. A few exceptionally bright astronomical radio sources, like Cas A and Cyg A, are so bright that their signature can be detected in the data even in observations on fields that are at a considerable distance from these sources. This is mitigated by applying phase rotation (effectively applying beamforming weights to the visibilities without adding them together) towards these sources, estimating and subtracting their response, and undoing the phase rotation again. This process is called *demixing*. After demixing, further integration is possible, which reduces the computational burden in further stages of the pipeline. Initial calibration usually consists of direction independent calibration of the complex valued gains of the individual receive paths in the interferometer array. The algorithms used here are very similar to those exploited in the station calibration mentioned before.

After initial calibration, the self-calibration and imaging cycle is entered, which is the main part of the SDP imaging pipeline. It starts by computing the residual visibilities obtained after subtracting the best available model for the visibilities based on the current best knowledge of calibration parameters and sky model from the measured visibilities. A dirty image is made from the residual visibilities. The required operations (17) are essentially a Fourier transform, but on non-uniformly sampled data. To be able to use the fast Fourier transform (required because this step is the most expensive in the entire processing pipeline), the residual visibilities are *gridded* onto a uniform grid, after which the inverse FFT is applied. Other computationally efficient implementations for non-uniform fast Fourier transforms may be considered. As this processing step is similar in many other image formation instruments (e.g., geophysics [19] and MRI), the available literature is rich.

Iterative algorithms such as CLEAN are used to find and subtract new sources in the residual image. This is referred to as the *minor cycle*. The new source components are added to the sky model, which is then used in the next iteration of the self-calibration and imaging cycle, the *major cycle*. Once this process has converged sufficiently, the sky model (deconvolved image) is added to the residual image, which should ideally only contain noise at this stage. That result is then presented as the final image. Since the major cycle is very expensive, the usual approach is to detect thousands of sources in each minor cycle, and to run the major cycle less than 10 times.

7 Concluding Remarks and Further Reading

In this chapter, we presented a signal processing viewpoint on radio astronomy. We showed how, with the right translations, the "measurement equations" are connected to covariance matrix data models used in the phased array signal processing literature. In this presentation, the resulting data models are very compact and clean, in the sense that the most straightforward covariance data models, widely studied in the signal processing literature as theoretical models, already seem valid. This is because far field assumptions clearly hold, and the propagation channels are very simple (no multipath), in contrast to other array processing applications such as seismology, synthetic aperture radar, or biomedical tomography.

However, this does not mean that radio astronomy is a "simple" application: data volumes are massive, and the requirements on resolution and accuracy are mind-boggling. Current telescopes, developed in the 1970s, start with signals sampled at 1–2 bits accuracy (because anyway the signals are mostly noise), and after data reduction and map making routinely end up with images with a dynamic range of 10^5.

So far, radio astronomy has done very well without explicit connection to the array signal processing literature. However, we expect that, by making this connection, a wealth of new insights and access to "new" algorithms can be obtained. This will be beneficial, and possibly essential, for the development of new instruments like LOFAR and SKA.

For further reading we suggest, first of all, the classical radio astronomy textbooks, e.g., by Thompson et al. [52] and by Perley et al. [49]. The August 2009 issue of the *Proceedings of the IEEE* was devoted to the presentation of new instruments. The January 2010 issue of *IEEE Signal Processing Magazine* gave a signal processing perspective. For general insights into imaging and deconvolution, we suggest Blahut [4].

Challenges for signal processing lie in (1) imaging, (2) calibration, (3) interference suppression. These problems are really intertwined. It is interesting to note that, especially for calibration and interference suppression, factor analysis is an essential tool. Our contributions in these areas have appeared in [3, 6, 33, 34, 36, 55, 57, 62–64] and are summarized in the PhD theses [5, 44, 54, 60], which should provide ample details for further reading.

References

1. Barrett, R., Berry, M., Chan, T.F., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., der Vorst, H.V.: Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition. SIAM, Philadelphia, PA (1994)
2. Bartholomew, D.J., Knott, M., Moustaki, I.: Latent Variable Models and Factor Analysis: A Unified Approach. John Wiley and Sons (2011)
3. Ben-David, C., Leshem, A.: Parametric high resolution techniques for radio astronomical imaging. IEEE Journal of Selected Topics in Signal Processing **2**(5), 670–684 (2008)
4. Blahut, R.E.: Theory of remote image formation. Cambridge University Press (2004). ISBN 0521553733
5. Boonstra, A.J.: Radio frequency interference mitigation in radio astronomy. Ph.D. thesis, TU Delft, Dept. EEMCS (2005). ISBN 90-805434-3-8
6. Boonstra, A.J., van der Veen, A.J.: Gain calibration methods for radio telescope arrays. IEEE Transactions on Signal Processing **51**(1), 25–38 (2003)
7. Boonstra, A.J., Wijnholds, S.J., van der Tol, S., Jeffs, B.: Calibration, sensitivity and RFI mitigation requirements for LOFAR. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Philadelphia (Penn.), USA (2005)
8. Borgiotti, G.B., Kaplan, L.J.: Superresolution of uncorrelated interference sources by using adaptive array techniques. IEEE Transactions on Antennas and Propagation **27**, 842–845 (1979)

9. Bridle, A.H., Schwab, F.R.: Bandwidth and Time-Average Smearing. In: G.B. Taylor, C.L. Carilli, R.A. Perley (eds.) Synthesis Imaging in Radio Astronomy II, *Astronomical Society of the Pacific Conference Series*, vol. 180, chap. 18, pp. 371–382. Astronomical Society of the Pacific (1999)
10. Briggs, D.S.: High fidelity deconvolution of moderately resolved sources. Ph.D. thesis, New Mexico Inst. of Mining and Technology, Socorro (NM) (1995)
11. Carrillo, R.E., McEwen, J.D., Wiaux, Y.: Sparsity averaging reweighted analysis (SARA): a novel algorithm for radio-interferometric imaging. Monthly Notices of the Royal Astronomical Society **426**(2), 1223–1234 (2012)
12. Carrillo, R.E., McEwen, J.D., Wiaux, Y.: PURIFY: a new approach to radio-interferometric imaging. Monthly Notices of the Royal Astronomical Society **439**(4), 3591–3604 (2014)
13. Cornwell, T., Braun, R., Brigss, D.S.: Deconvolution. In: G.B. Taylor, C.L. Carilli, R.A. Perley (eds.) Synthesis Imaging in Radio Astronomy II, *Astronomical Society of the Pacific Conference Series*, vol. 180, pp. 151–170. Astronomical Society of the Pacific (1999)
14. Cornwell, T.J.: Multiscale CLEAN deconvolution of radio synthesis images. IEEE Journal of Selected Topics in Signal Processing **2**(5), 793–801 (2008)
15. Cornwell, T.J., Wilkinson, P.N.: A new method for making maps with unstable radio interferometers. Monthly Notices of the Royal Astronomical Society **196**, 1067–1086 (1981)
16. Cotton, W.D., et al.: Beyond the isoplanatic patch in the VLA Low-frequency Sky Survey. In: Proceedings of the SPIE, vol. 5489, pp. 180–189. Glasgow (2004)
17. Dewdney, P.E., Braun, R.: SKA1-low configuration coordinates - complete set. Tech. Rep. SKA-TEL-SKO-0000422, SKA Office, Manchester (UK) (2016)
18. Dewdney, P.E., Hall, P.J., Schilizzi, R.T., Lazio, T.J.L.W.: The square kilometre array. Proceedings of the IEEE **97**(8), 1482–1496 (2009)
19. Duijndam, A.J.W., Schonewille, M.A.: Nonuniform fast Fourier transform. Geophysics **64**(2), 539–551 (1999)
20. Foucart, S., Koslicki, D.: Sparse recovery by means of nonnegative least squares. IEEE Signal Processing Letters **21**(4), 498–502 (2014)
21. Frieden, B.: Restoring with maximum likelihood and maximum entropy. Journal of the Optical Society of America **62**, 511–518 (1972)
22. Fuhrmann, D.R.: Estimation of sensor gain and phase. IEEE Transactions on Signal Processing **42**(1), 77–87 (1994)
23. Garsden, H., et al.: LOFAR sparse image reconstruction. Astronomy & Astrophysics **575**(A90), 1–18 (2015)
24. van Haarlem, M.P., et al.: LOFAR: The low frequency array. Astronomy & Astrophysics **556**(A2), 1–53 (2013)
25. Hamaker, J.P.: Understanding radio polarimetry - iv. the full-coherency analogue of scalar self-calibration: Self-alignment, dynamic range and polarimetric fidelity. Astronomy & Astrophysics Supplement **143**(3), 515–534 (2000)
26. Hayes, M.H.: Statistical Digital Signal Processing and Modeling. John Wiley and Sons (1996)
27. Hogbom, J.A.: Aperture synthesis with non-regular distribution of interferometer baselines. Astronomy and Astrophysics Suppl. **15**, 417–426 (1974)
28. Intema, H.T., et al.: Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme. I. Method description and first results. Astronomy & Astrophysics **501**(3), 1185–1205 (2009)
29. Jongerius, R.: Exascale computer system design: The square kilometre array. Ph.D. thesis, Eindhoven University of Technology (2016). ISBN 978-90-386-4136-2
30. Jongerius, R., Wijnholds, S., Nijboer, R., Corporaal, H.: An end-to-end computing model for the square kilometre array. IEEE Computer **47**(9), 48–54 (2014)
31. Kazemi, S., Yatawatta, S., Zaroubi, S., Lampropoulos, P., de Bruyn, A.G., Koopmans, L.V.E., Noordam, J.: Radio interferometric calibration using the sage algorithm. Monthly Notices of the Royal Astronomical Society **414**(2), 1656 (2011)
32. Lawley, D.N., Maxwell, A.E.: Factor Analysis as a Statistical Method. Butterworth & Co, London (1971)

33. Leshem, A., van der Veen, A.J.: Radio-astronomical imaging in the presence of strong radio interference. IEEE Transactions on Information Theory **46**(5), 1730–1747 (2000)
34. Leshem, A., van der Veen A. J., Boonstra, A.J.: Multichannel interference mitigation technique in radio astronomy. Astrophysical Journal Supplements **131**(1), 355–374 (2000)
35. Levanda, R., Leshem, A.: Radio astronomical image formation using sparse reconstruction techniques. In: IEEE 25th convention of Elec. Electron. Eng. in Israel (IEEEI 2008), pp. 716–720 (2008)
36. Levanda, R., Leshem, A.: Synthetic aperture radio telescopes. IEEE Signal Processing Magazine **27**(1), 14–29 (2010)
37. Li, F., Cornwell, T.J., de Hoog, F.: The application of compressive sampling to radio astronomy; I deconvolution. Astronomy and Astrophysics **528**(A31), 1–10 (2011)
38. Lonsdale, C., et al.: The Murchison Widefield Array: Design overview. Proceedings of the IEEE **97**(8), 1497–1506 (2009)
39. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing **41**(12), 3397–3415 (1993)
40. Mardia, K.V., Kent, J.T., Bibby, J.M.: Multivariate Analysis. Academic Press, New York (1979)
41. Marsh, K.A., Richardson, J.M.: The objective function implicit in the CLEAN algorithm. Astronomy and Astrophysics **182**(1), 174–178 (1987)
42. Mitchell, D.A., et al.: Real-time calibration of the Murchison Widefield Array. IEEE Journal of Selected Topics in Signal Processing **2**(5), 707–717 (2008)
43. Moon, T.K., Stirling, W.C.: Mathematical Methods and Algorithms for Signal Processing. Prentice Hall (2000). ISBN 0201361868
44. Mouri Sardarabadi, A.: Covariance matching techniques for radio astronomy calibration and imaging. Ph.D. thesis, TU Delft, Dept. EEMCS (2016)
45. Mouri Sardarabadi, A., Leshem, A., van der Veen, A.J.: Radio astronomical image formation using constrained least squares and Krylov subspaces. Astronomy & Astrophysics **588**, A95 (2016)
46. Noordam, J.E.: Generalized self-calibration for LOFAR. In: XXVIIth General Assembly of the International Union of Radio Science (URSI). Maastricht (The Netherlands) (2002)
47. Ottersten, B., Stoica, P., Roy, R.: Covariance matching estimation techniques for array signal processing applications. Digital Signal Processing, A Review Journal **8**, 185–210 (1998)
48. Pearson, T.J., Readhead, A.C.S.: Image formation by self-calibration in radio astronomy. Annual Review of Astronomy and Astrophysics **22**, 97–130 (1984)
49. Perley, R.A., Schwab, F.R., Bridle, A.H.: Synthesis Imaging in Radio Astronomy, *Astronomical Society of the Pacific Conference Series*, vol. 6. BookCrafters Inc. (1994)
50. Salvini, S., Wijnholds, S.J.: Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications. Astronomy & Astrophysics **571**(A97), 1–14 (2014)
51. Schwardt, L.C.: Compressed sensing imaging with the KAT-7 array. In: International Conference on Electromagnetics in Advanced Applications (ICEAA), pp. 690–693 (2012)
52. Thompson, A.R., Moran, J.M., Swenson, G.W.: Interferometry and Synthesis in Radio Astronomy, 2nd edn. Wiley, New York (2001)
53. Tingay, S.J., et al.: The Murchison widefield array: The square kilometre array precursor at low radio frequencies. Publications of the Astronomical Society of Australia **30**(7) (2013)
54. van der Tol, S.: Bayesian estimation for ionospheric calibration in radio astronomy. Ph.D. thesis, TU Delft, Dept. EEMCS (2009)
55. van der Tol, S., Jeffs, B.D., van der Veen, A.J.: Self-calibration for the LOFAR radio astronomical array. IEEE Transactions on Signal Processing **55**(9), 4497–4510 (2007)
56. Turner, W.: SKA phase 1 system requirements specification. Tech. Rep. SKA-TEL-SKO-0000008, SKA Office, Manchester (UK) (2016)
57. van der Veen, A.J., Leshem, A., Boonstra, A.J.: Array signal processing for radio astronomy. Experimental Astronomy **17**(1–3), 231–249 (2004)
58. de Vos, M., Gunst, A., Nijboer, R.: The LOFAR telescope: System architecture and signal processing. Proceedings of the IEEE **97**(8), 1431–1437 (2009)

59. Wiaux, Y., Jacques, L., Puy, G., Scaife, A.M.M., Vandergheynst, P.: Compressed sensing imaging techniques for radio interferometry. Monthly Notices of the Royal Astronomical Society **395**, 1733–1742 (2009)
60. Wijnholds, S.J.: Fish-eye observing with phased array radio telescopes. Ph.D. thesis, TU Delft, Dept. EEMCS (2010). ISBN 978-90-9025180-6
61. Wijnholds, S.J., Boosntra, A.J.: A multisource calibration method for phased array telescopes. In: Fourth IEEE Workshop on Sensor Array and Multi-channel Processing (SAM). Waltham (Mass.), USA (2006)
62. Wijnholds, S.J., van der Tol, S., Nijboer, R., van der Veen, A.J.: Calibration challenges for the next generation of radio telescopes. IEEE Signal Processing Magazine **27**(1), 32–42 (2010)
63. Wijnholds, S.J., van der Veen, A.J.: Fundamental imaging limits of radio telescope arrays. IEEE Journal of Selected Topics in Signal Processing **2**(5), 613–623 (2008)
64. Wijnholds, S.J., van der Veen, A.J.: Multisource self-calibration for sensor arrays. IEEE Transactions on Signal Processing **57**(9), 3512–3522 (2009)
65. Wise, M.W., Rafferty, D.A., McKean, J.P.: Feedback at the working surface: A joint X-ray and low-frequency radio spectral study of the Cocoon Shock in Cygnus A. In: 13th Meeting of the American Astronomical Society's High Energy Astrophysics Division (HEAD), pp. 88–89 (2013)
66. Yatawatta, S.: Distributed radio interferometric calibration. Monthly Notices of the Royal Astronomical Society **449**(4), 4506 (2015)
67. Zatman, M.: How narrow is narrowband. IEE Proc. Radar, Sonar and Navig. **145**(2), 85–91 (1998)