[Ronny Levanda and Amir Leshem]

© PHOTODISC

Next-generation radio telescopes will be much larger, more sensitive, have a much larger observation bandwidth, and will be capable of pointing multiple beams simultaneously. Obtaining the sensitivity, resolution, and dynamic range supported by the receivers requires the development of new signal processing techniques for array and atmospheric calibration as well as new imaging techniques that are both more accurate and computationally efficient since data volumes will be much larger. This article provides an overview of existing image formation techniques and outlines some of the directions needed for information extraction from future radio telescopes. We describe the imaging process from measurement equation until deconvolution, both as a Fourier inversion problem and as an array processing estimation problem. The latter formulation enables the development of more advanced techniques based on state-of-the-art array processing. We also demonstrate the techniques on simulated and measured radio telescope data.

INTRODUCTION

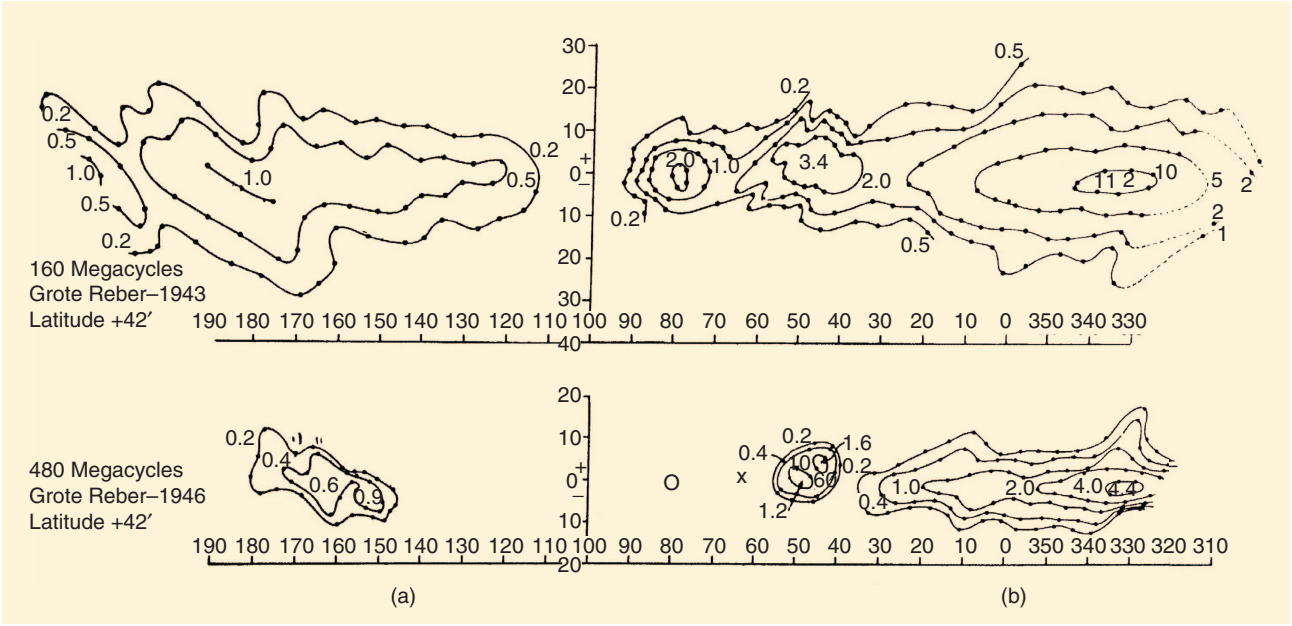
The field of radio astronomy is a relatively young field of observational astronomy and dates back to pioneering research by Jansky in the 1930s [1]. Jansky demonstrated that radio waves are emitted from the Milky Way galaxy. Inspired by his work, Reber [2] made the first radio survey of the sky using a radio telescope that he built in his backyard. Figure 1 depicts some results of his radio survey, including the strong radio emissions of Cygnus A (Cyg A) and Cassiopeia A (Cas A). In 1946, Ryle and Vonberg [3] used the Michelson interferometer to observe radio emissions from the sun at a frequency of 175 MHz. Ryle continued to construct interferometers located on rails, which allowed him to create a synthetic aperture by moving the antennas. This is the origin of modern inverse synthetic aperture radar and the active synthetic aperture radar imaging. Subsequently, the study of radio emissions from celestial sources has led to many great discoveries, such as cosmic microwave background radiation by Penzias and Wilson [4] and its anisotropy [5] and pulsars, which are rapidly rotating neutron stars, by Bell et al. [6]. Other phenomena of great interest for radio astronomers include gravitational lenses where the gravitational field of a massive object serves as a lens by bending the light wave (many of the gravitational lenses were discovered in radio frequencies, see

Synthetic Aperture Radio Telescopes

[A look at image formation techniques]

Digital Object Identifier 10.1109/MSP.2009.934719

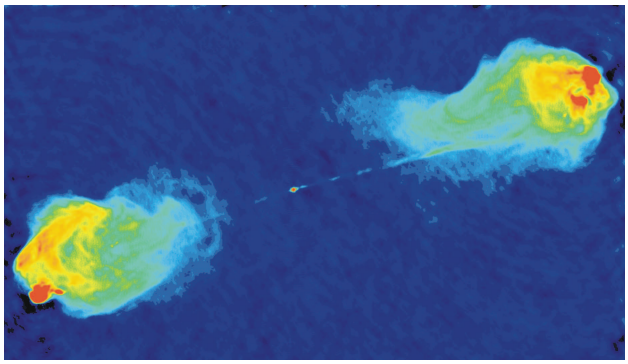
1053-5888/10/$26.00©2010IEEE



[FIG1] Reber's radio survey. We can see the Milky Way galaxy, Cygnus A, and Cassiopeia A. (Image courtesy of National Radio Astronomy Observatory (NRAO)/Associated Universities, Inc. (AUI).)

http://www.aoc.nrao.edu/smyers/class.html for more information), active galactic nuclei such as in Virgo A (also known as M87), and supernova remnants such as Cassiopeia A. Virgo A is a giant galaxy in the Virgo cluster that has jets of particles moving at relativistic speeds and emitting very strong radio waves. It is believed that the center of the Virgo A galaxy is a very massive black hole. Radio astronomy also deals with spectral lines that appear at radio frequencies such as the hydrogen spectral line that was first detected in 1951 [7]. The spectral line at 21 cm is created by a change in the energy state of neutral hydrogen. This spectral line is expected to play an important role in understanding the reionization of the universe when the first galaxies were formed. In 1962, the principle of synthesis aperture imaging using earth rotation was proposed by Ryle [9]. Ryle's idea was simple and beautiful. Instead of moving the antennas as he has been doing for about 15 years, he used the fact that the Earth rotates to generate the synthetic aperture. This quickly became the main operating mode of radio interferometers. However, imaging using Earth rotation synthesis radio telescopes is an ill-posed problem due to the irregular sub-Nyquist sampling of the Fourier domain. This subsampling results in aliasing inside the image due to the high sidelobes of the array response. To solve this problem, we need to remove the effect of the instrumental response from the image (a process known as deconvolution) to compensate for inaccuracies in the array response (known as self-calibration, but it has many similarities to blind deconvolution). It is important to understand that the improved imaging capability is a result of better equipment in conjunction with new imaging techniques. Each generation of radio telescopes involved significant hardware development effort. However, exploiting the hardware capabilities requires a constant improvement in imaging and self-calibration to match the receiver sensitivity. Figure 2 (from [8]) presents the outcome of imaging and self-calibration applied to an image of Cygnus A. It is the first discovery of the radio jets going from the center all the way to the external radio lobes. Even though Cygnus A has been observed for many years (since Reber's time) it is the image formation and self calibration algorithms that allowed the discovery of the radio jets.

Over the last 40 years, many deconvolution techniques have been developed to solve this problem. The basic idea behind a deconvolution algorithm is to exploit a priori knowledge about the image. The first algorithm (and the most popular of these techniques) is the CLEAN method proposed by Högbom [10]. The maximum entropy method algorithm (MEM) with various entropy functions was proposed in [11]–[14] and the current implementation by Cornwell and Evans [15] is the most widely used. Beyond these two techniques there are several extensions in various directions: extensions of the CLEAN algorithm to support multiresolution and wavelets as well as noncoplanar arrays and multiple



[FIG2] Cygnus A image. False color image of the radio jet and lobes in the hyperluminous radio galaxy Cygnus A. Red shows regions with the brightest radio emission, while blue shows regions of fainter emission. (Image courtesy of NRAO/AUI by Perley et al. [8].)

wavelengths (see the overview paper [16]). MEM techniques have been also extended to take into account source structure through the use of multiresolution and wavelet-based techniques [17]. Global nonnegative least squares (LS) was proposed by Briggs [18], matrix-based parametric imaging such as the LS minimum variance imaging (LS–MVI) and maximum likelihood-based techniques in [19] and [20], and sparse L_1 reconstruction in [21] and [22]. Source modeling is an important issue and various techniques to improve modeling over simple point source models by using shapelets, wavelets, and Gaussians [23] have been implemented. A more extensive overview of classical techniques and implementation issues is given in [24] and [25].

A better performance analysis of imaging and self-calibration techniques is one of the major challenges for the signal processing community. This is likely to become a more critical problem for the future generation of radio interferometers that will be built in the next two decades, such as the square kilometer array (SKA) (http://www.skatelescope.org/), the low frequency array (LOFAR) (http://www.lofar.org/p/astronomy.htm), the Allen telescope array (ATA), (see Figure 3 and http://ral.berkeley.edu/ata/), the long wavelength array (LWA) (http://lwa.unm.edu/), and the Atacama large millimeter array (ALMA) (http://www.almaobservatory.org/index.php). These radio telescopes will be composed of many stations (each station will be made up of multiple antennas that are combined using adaptive beamforming) and will have significantly increased sensitivity and bandwidth. Some of them will operate at much lower frequencies than existing radio telescopes. Improved sensitivity will therefore require a much better calibration, the capability to perform imaging with much higher dynamic range to reduce the effect of the residuals of powerful foreground sources inside and outside the field of view and better handling of noncoplanar arrays.

THE IMAGING EQUATIONS

This section reviews the basic principles of radio astronomy following Taylor et al. [25]. In radio astronomy, we observe the radio waves emitted from space. Since the source is far away, the received electromagnetic field intensity distribution can be observed only in an angular direction (no information regarding the intensity distribution in the radial direction). Defining the celestial sphere as the maximal sphere that contains no radiating sources, the observed intensity is the projection of the source intensity on the celestial sphere. For simplicity, we will deal with a quasi-monochromatic wave at frequency ν (the general case can be easily derived by a linear combination of quasi-monochromatic waves). The electric field at location $\mathbf{r}$ is given by

$$E_\nu(\mathbf{r}) = \int \epsilon_\nu(\mathbf{q}) \frac{e^{2\pi \mathcal{J} \nu |\mathbf{q}-\mathbf{r}|/c}}{|\mathbf{q}-\mathbf{r}|} \mathrm{d}S, \tag{1}$$

where $\epsilon_\nu(\mathbf{q})$ is the electric field at location $\mathbf{q}$ (on the celestial sphere), $\mathrm{d}S$ is surface area on the sphere and the integration is done over the entire sphere and c is the speed of light. For two antennas observing a distant source (receiving the electric field emitted by the source) there is a geometrical delay in one of the antennas relative to the other antenna derived from the source observation angle [see Figure 4(a)]; if the geometric delay is compensated by an electronic delay, the electric field received in one antenna should be highly correlated with the electric field received by the other antenna. The spatial coherency of the electric field for two antennas located at $\mathbf{r}_1$ and $\mathbf{r}_2$ is given by

$$V_\nu(\mathbf{r}_1, \mathbf{r}_2) = \langle \mathbf{E}_\nu(\mathbf{r}_1)\, \mathbf{E}^*_\nu(\mathbf{r}_2) \rangle, \tag{2}$$

where $\langle\rangle$ stands for the expectation value. Substituting (1) into (2) and taking into account the large distance of the source; i.e., $|\mathbf{q}-\mathbf{r}| \approx |\mathbf{q}|$ and that the electric field is spatially incoherent (i.e., $\langle \epsilon_\nu(\mathbf{q}_1)\epsilon^*_\nu(\mathbf{q}_2)\rangle = 0 \ \ \forall \mathbf{q}_1 \neq \mathbf{q}_2$) we get

$$V_\nu(\mathbf{r}_1, \mathbf{r}_2) = \int I_\nu(s) e^{-2\pi \mathcal{J} \nu \mathbf{s}(\mathbf{r}_1 - \mathbf{r}_2)/c} d\Omega, \tag{3}$$

where $I_\nu(\mathbf{s}) \equiv \langle \epsilon_\nu(\mathbf{s})^2 \rangle$ is the source intensity at direction s on the sphere ($\mathbf{s} \equiv q/|q|$), and $d\Omega = dS/|q|^2$. Representing (3) in the (u, v, w) coordinate system, for many astronomical observations (e.g., planar arrays, or small field of view imaging) we obtain

$$V_\nu(u, v) = \int\int I_\nu(l, m) e^{-2\pi \mathcal{J}(ul+vm)} \mathrm{d}l\mathrm{d}m. \tag{4}$$

The visibility is the Fourier transform of the source intensity; therefore the inverse relation holds

$$I_\nu(l, m) = \int\int V_\nu(u, v) e^{2\pi \mathcal{J}(ul+vm)} \mathrm{d}u\mathrm{d}v. \tag{5}$$

When the coplanar approximation does not hold, (4) takes the more complicated form

$$V_\nu(u, v, w) = \int\int \frac{1}{n} I_\nu(l, m) e^{-2\pi \mathcal{J}[ul+vm+w(n-1)]} \mathrm{d}l\mathrm{d}m, \tag{6}$$

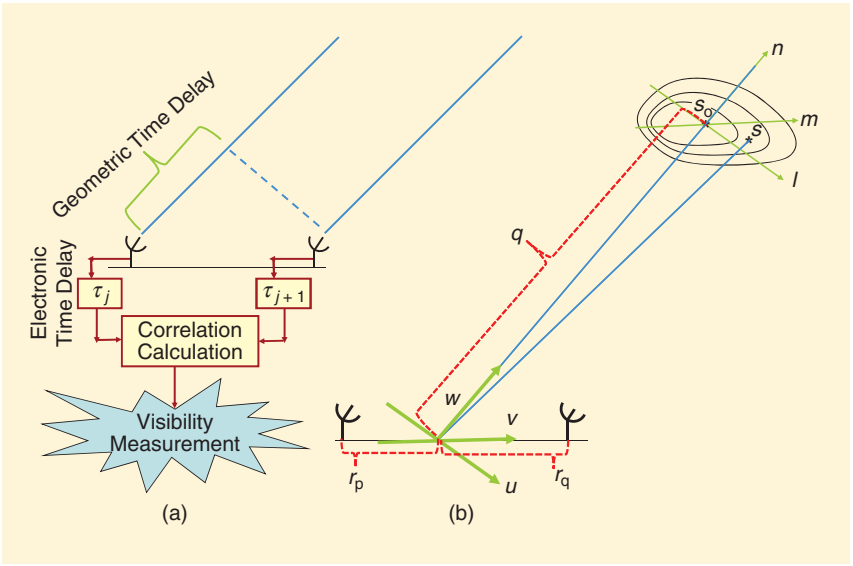
where

$$n \equiv \sqrt{1 - \ell^2 - m^2}. \tag{7}$$

For a source with visibility measurements covering the entire (u, v) domain, the source image is perfectly computed by the Fourier inversion of the visibility. In practice, only a small part of the (u, v) domain is measured by sampling the

[FIG3] The Allen telescope array. (Image courtesy of Seth Shostak and the SETI Institute.)

existing antenna pair baselines as they change with the Earth's rotation relative to the (u, v) coordinates (at time t_k two antennas p and q measure a single visibility point in the (u, v) domain at (u_{pq}^k, v_{pq}^k)) [see Figure 4(b)]. This set of samples is known as the (u, v) coverage of the radio telescope. This coverage is determined by many factors such as the configuration in which the individual receptors (telescopes or dipole) are placed on the ground, the minimal and maximal distance between antenna pairs, the time difference between consecutive measurements, and the total measurement time and bandwidth. An example of the (u, v) coverage for a simulated radio telescope (east-west array with 14 antennas logarithmically spaced from λ to 200λ, observation time of 12 hours) is shown in Figure 5. The sampled points in the (u, v) plane are a collection of ellipses. The sampling effect on the resulting image is shown in Figure 6(a) and (c). Figure 6(a) depicts an image of visibility data measured over a dense and uniform grid in the (u, v) plane (all grid points in the (u, v) plane were sampled). Figure 6(c) presents the same data with a more realistic (u, v) sampling. The image with the partial (and more realistic) measurement set is blurred, distorted and noisy. Let $S(u, v)$ be the sampling function ($S(u, v) = 1$ for each measured (u, v) pair and $S(u, v) = 0$ otherwise). We obtain that the inverse direct Fourier transform of the measured visibility, known as the dirty image I_D, is given by



[FIG4] Measurement setting. (a) The visibility is the measurement of spatial correlation between the electric field of antenna pairs. The geometric delay of the wave that propagates from the source to the two antennas is compensated for by an electronic delay. (b) A distant source is observed by an antenna pair. The baseline connecting the two antennas is the origin of the (*u*, *v*, *w*) coordinate system. The *w* axis points from the baseline toward the source reference point. (*u*, *v*) are perpendicular to *w* and selected according to the Earth's orientation. (*l*, *m*, *n*) is a unit vector in the (*u*, *v*, *w*) system pointing toward a specific location in the source (at the source reference point S_o, $l = 0$, $m = 0$) and $n = \sqrt{1 - (l^2 + m^2)}$.

$$I_{D,\nu}(l, m) = \int\int V_\nu(u, v) S(u, v) e^{2\pi \mathcal{J}(ul + vm)} du\, dv. \quad (8)$$

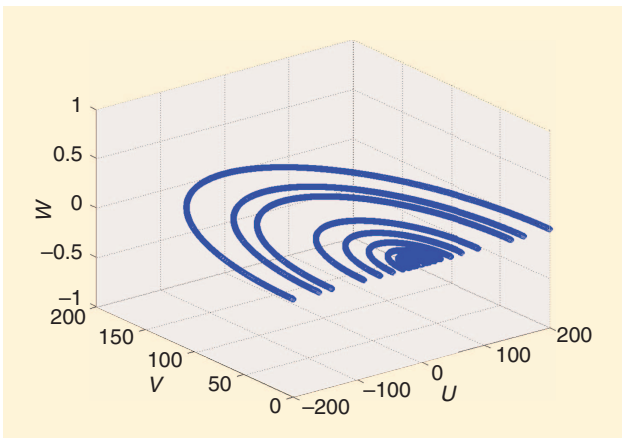
The instrument point spread function (also known as the dirty beam) is defined by

$$B(l, m) \equiv \int\int S(u, v) e^{2\pi\, \mathcal{J}(ul + vm)} du\, dv. \quad (9)$$

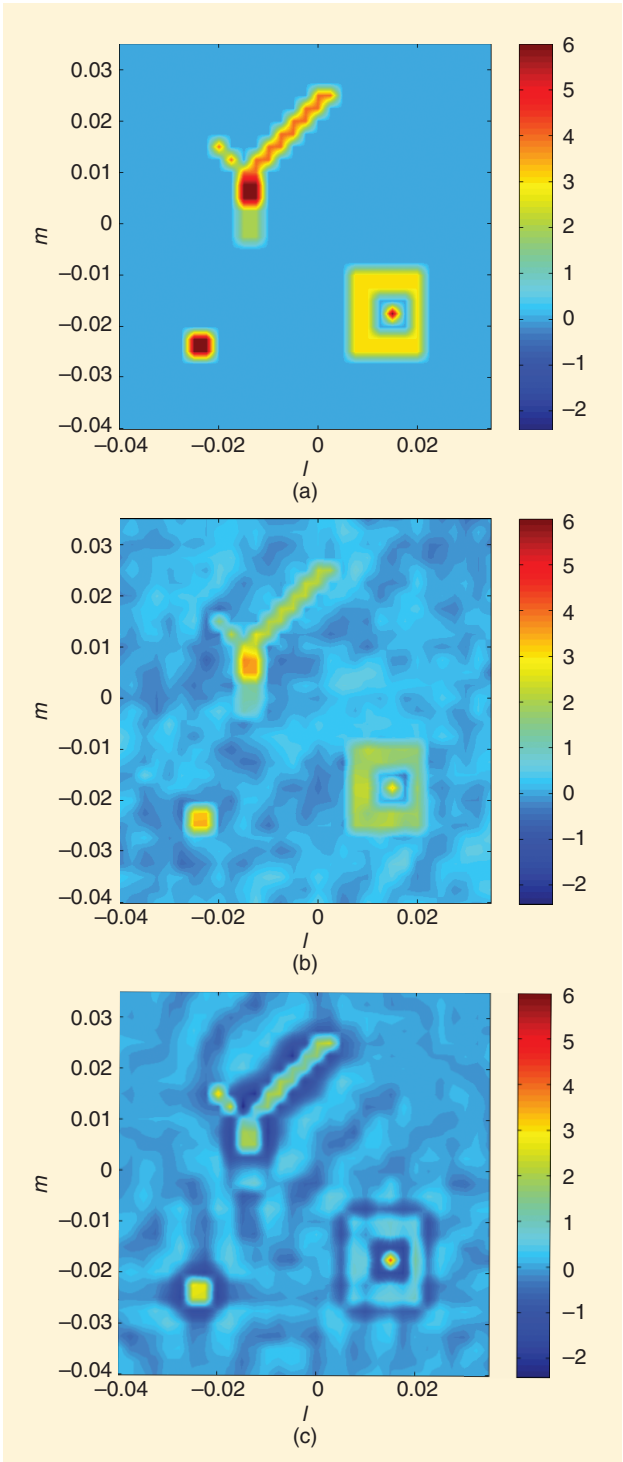
By the convolution theorem, the dirty image is the convolution of the true source intensity (5) and the dirty beam (9)

$$I_{D,\nu} = I_\nu * B. \quad (10)$$

This is the reason why image reconstruction algorithms in radio astronomy are often referred to as deconvolution algorithms, since direct synthesis produces $I_{D,\nu}$, but we want to obtain I_ν by deconvolution with respect to B. The dirty image can be calculated from the measured visibility data according to (8), or by using a fast Fourier transform (FFT) to reduce the calculation time and memory resources. To use the FFT, the visibility data must lie on a rectangular equally spaced grid. This procedure of resampling the measured visibilities on a regular grid is called gridding. The weighting is done by convolving the visibilities with a smooth kernel (this procedure is also called convolutional gridding). Choice of the gridding kernel is important and follows from standard interpolation theory. An illustration of the gridding effect for a rectangular kernel is shown in Figure 6(a) and (b). Both images were generated using simulated visibility data with complete (u, v) coverage. In Figure 6(a), the visibility



[FIG5] The (*u*, *v*) coverage of a simulated east-west radio telescope with 14 antennas logarithmically spaced with baselines up to 200λ. Observations are made every six minutes for a duration of 12 hours. From each antenna pair we get an ellipse in the (*u*, *v*) domain. Note: *u* and *v* are in λ units.



[FIG6] Illustration of sampling and gridding effects. (a) Fourier transform of visibility data on a perfect grid (visibility measurements location are on the center of the grid cells) with complete (*u, v*) coverage. (b) Fourier transform of visibility data with off-grid points (the visibility measurements location were chosen randomly) and complete (*u, v*) coverage, demonstrates gridding effect. Features in the image are blurred and distorted. (c) Fourier transform of visibility data with perfect grid and incomplete (*u, v*) coverage (of the radio telescope described in Figure 5), demonstrates the sampling effect.

data were taken on a perfect grid (all visibility measurements were located on the center of a grid cell). In Figure 6(b), the location of the visibility measurements was chosen randomly within the cells in the (u, v) plane. This results in a blurred and distorted image. For more details on gridding and tapering, the reader is referred to [24] and [25].

THE PARAMETRIC MATRIX FORMULATION OF THE IMAGE FORMATION PROBLEM

We now describe an alternative formulation of the image formation problem. In this formulation imaging is viewed as a parameter estimation problem, where the locations and powers (and possibly polarization parameters and frequency dependence of power) are the unknown parameters. This formalism was first proposed in [26] and [19] to allow for the introduction of interference mitigation techniques in the imaging process. It was extended to noncoplanar array and polarimetric imaging in [20]. This formulation also allows the easy inclusion of space-dependent calibration parameters [27]. Assume that the observed image is a collection of D point sources, i.e.,

$$I_\nu(l, m) = \sum_{d=1}^{D} I_\nu(l, m)\delta(l - l_d)\delta(m - m_d). \tag{11}$$

Since (u, v) are the baseline coordinates (i.e., $u \equiv u_{ij}^k = u_i^k - u_j^k$ and $v \equiv v_{ij}^k = v_i^k - v_j^k$), the visibility (4) can be rewritten as

$$V_\nu(u_{ij}^k, v_{ij}^k) = \sum_{d=1}^{D} I_\nu(l_d, m_d) e^{-2\pi\mathcal{J}(u_{ij}^k l_d + v_{ij}^k m_d)}, \tag{12}$$

where k denotes the measurement time t_k. Selecting a (time varying) reference point at one of the antennas (u_0^k, v_0^k) and manipulating (12) yields

$$V_\nu(u_{ij}^k, v_{ij}^k) = \sum_{d=1}^{D} e^{-2\pi\mathcal{J}(u_{i,0}^k \; l_d + v_{i,0}^k m_d)} I_\nu(l_d, m_d). \\ e^{2\pi\mathcal{J}(u_{j,0}^k \; l_d + v_{j,0}^k \; m_d)} \tag{13}$$

We define the kth measurement correlation matrix $\mathbf{R}_k$ by

$$(\mathbf{R}_k)_{ij} \equiv V_\nu(u_{ij}^k, v_{ij}^k). \tag{14}$$

The correlation matrix is illustrated in Figure 7 for a single frequency bin. Cell R_{ij}^k of the correlation matrix is the visibility measurement at time t_k from antenna pair (i, j). The size of the correlation matrix $\mathbf{R}_k$ is $p \times p$ where p is the number of antennas in the array. The autocorrelation of each antenna is also used (the diagonal of the correlation matrix). When an observation uses more than a single frequency bin, each correlation matrix is computed using a single bin, as illustrated in Figure 8.

Let the Fourier component vector at time t_k be

$$\mathbf{a}_k(l, m) \equiv \begin{pmatrix} e^{-2\pi\mathcal{J}(u_{1,0}^k \, l + v_{1,0}{}^k m)} \\ e^{-2\pi\mathcal{J}(u_{1,0}^k \, l + v_{1,0}{}^k m)} \end{pmatrix}, \tag{15}$$

and let the Fourier component matrix at time t_k be

$$\mathbf{A}_k \equiv [\mathbf{a}_k(l_1, m_1), \ldots, \mathbf{a}_k(l_d, m_d)]. \quad (16)$$

Define the point source intensity matrix by

$$\mathbf{B} \equiv \begin{bmatrix} I(l_1, m_1) & & 0 \\ & \ddots & \\ 0 & & I(l_D, m_D) \end{bmatrix}. \quad (17)$$

Using (15)–(17), (13) can be rewritten as

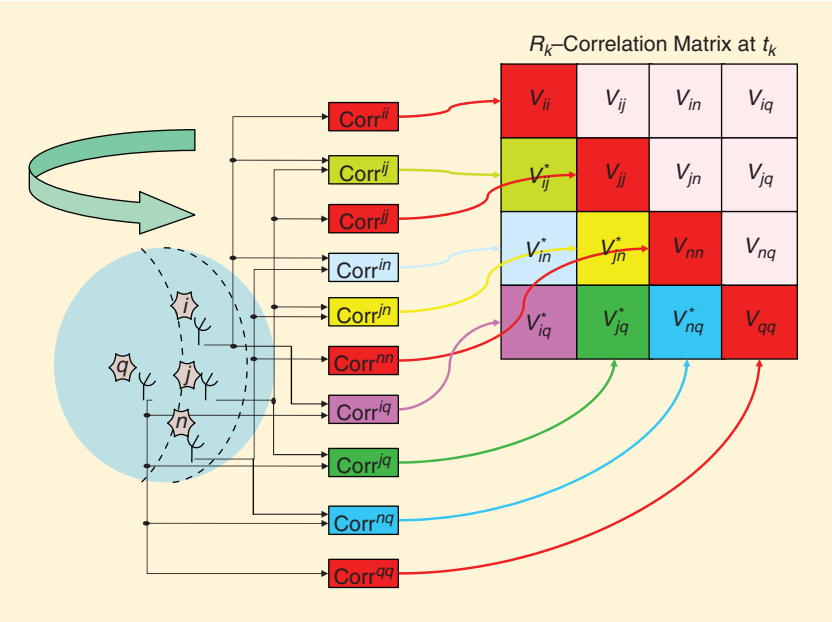
$$\mathbf{R}_k = \mathbf{A}_k \mathbf{B} \mathbf{A}_k^H. \quad (18)$$

Matrix (18) is the parametric form of the classical (4). It will allow us to consider the problem as an estimation problem, where we observe a set of measured covariance matrices which depend smoothly on the unknown source and instrument calibration parameters, as well as receiver noise. Using this formulation we can easily use well-known techniques from estimation theory (such as maximum a posteriori, ML, MVDR, and robust techniques) to solve the image formation problem. It also enables a simple extension to the noncoplanar array case as well as polarimetric imaging and multifrequency synthesis, where sources have frequency dependent (parametrically known) characteristics. The classical dirty image (8) can be rewritten as

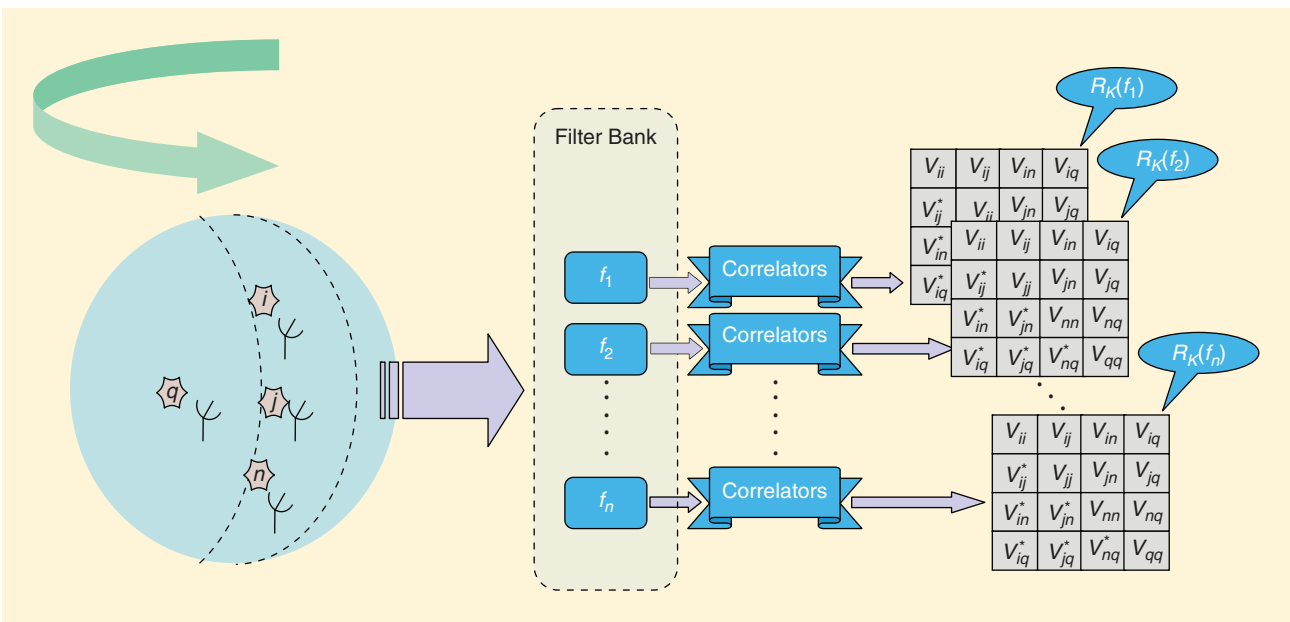
$$I_D(l, m) = \frac{1}{K} \sum_k \mathbf{a}_k^H(l, m)\, \mathbf{R}_k\, \mathbf{a}_k(l, m). \quad (19)$$

Note that this is identical to the mean power output of a classical beamformer oriented towards direction (l, m). More realistically, the antenna response varies slightly between different antennas and there is an additional noise per antenna. The antenna response can be measured prior to the observation and taken into account in the model. Since the noise in two antennas is independent, the noise correlation matrix is diagonal. Denoting by $\gamma_{i,k}$ the unknown complex gain of antenna i at observation time t_k and by σ^2 the noise variance, the correlation matrix now becomes

$$\mathbf{R}_k = \Gamma_k \mathbf{A}_k \mathbf{B} \mathbf{A}_k^H \Gamma_k^H + \sigma^2 \mathbf{I}, \quad (20)$$



[FIG7] The correlation matrix (for the simple case of a single frequency) in a specific time measurement t_k is built from the visibility measurement of antenna pairs.



[FIG8] The correlation matrices for an observation in the multifrequency case. The correlation matrices are calculated for each frequency separately (on filtered measurements from the antennas).

where

$$\Gamma_k \equiv \begin{bmatrix} \gamma_{1,k} & & 0 \\ & \ddots & \\ 0 & & \gamma_{p,k} \end{bmatrix}.$$

Estimation of the $\gamma_{i,k}$ is discussed in an article by Wijnholds et al. [28]. Typically, $\gamma_{i,k}$ varies slowly so it can be assumed to be constant over multiple times. Similarly, the noncoplanar array case is given by replacing $\mathbf{a}_k$ and B by

$$\mathbf{a}_k(l,m) \equiv \left[e^{-2\pi\mathcal{J}(u^k_{1,0}l+v^k_{1,0}m+w^k_{1,0}n)}, \ldots, e^{-2\pi\mathcal{J}(u^k_{p,0}l+v^k_{p,0}m+w^k_{1,0}n)}\right]^T \tag{21}$$

and

$$\mathbf{B} \equiv \begin{bmatrix} \dfrac{I(l_1,m_1)}{\sqrt{1-\ell_1^2-m_1^2}} & & 0 \\ & \ddots & \\ 0 & & \dfrac{I(l_D,m_D)}{\sqrt{1-\ell_D^2-m_D^2}} \end{bmatrix}. \tag{22}$$

The radio imaging problem can now be reformulated as follows: Given a set of measured covariance matrices $\hat{\mathbf{R}}_1, \ldots, \hat{\mathbf{R}}_K$ estimate the parameters $\mathbf{s}_1, \ldots, \mathbf{s}_D$, $I(\mathbf{s}_1), \ldots, I(\mathbf{s}_D)$ and the calibration matrices Γ_k:$k = 1, \ldots, K$. Note that (20) can be easily generalized to deal with direction dependent calibration parameters, polarized sources, as well as multifrequency synthesis. All that we need to change is the source and the calibration parametric model by simple adaptation of (20). The parametric approaches described in this article can be applied uniformly to all these problems. However, for simplicity we will focus on the calibrated array case.

CLASSICAL AND PARAMETRIC APPROACHES BASED ON SEQUENTIAL SOURCE REMOVAL

Many algorithms in radio astronomy are based on sequential source removal. The most commonly used is the CLEAN algorithm originally proposed by Högbom [10]. These iterative algorithms assume that the observed field is a collection of sources with simple structure. CLEAN assumes that the sources are point sources. During each iteration, a single point source is estimated and removed from the data. The reconstructed image is the collection of all point sources with their estimated power convolved with an ideal reconstruction beam (usually an elliptical Gaussian fitted to the central lobe of the dirty beam). The general structure common to all the sequential source removal algorithms is described in Table 1. The algorithms differ from each other by the exact definition of the dirty image used, the way the point source is removed from the image (either in the image domain after gridding or in the visibility domain), the intensity estimation method of the point sources and the exact modeling of the sources (such as point source, Gaussian, wavelet coefficients, and shapelets). Some versions, like the Cotton-Schwab technique, estimate multiple sources based on the same dirty image. This significantly accelerates the algorithm, since the number of Fourier transforms of the image is reduced.

We describe two sequential source removal algorithms. The first is the CLEAN algorithm, and the second is a parametric estimation-based algorithm known as LS-MVI.

THE CLEAN ALGORITHM

The CLEAN algorithm assumes that the observed field of view is composed of point sources. Since the image of a point source is given by the convolution of the point source and the dirty beam (10), CLEAN iteratively removes the brightest point source from the image until the residual image is noise-like. There are several variants of CLEAN [10], [29]–[31]. The CLEAN algorithm is implemented either in the image or in the visibility domain. In each iteration, the brightest point in the dirty image (8) is found (position and strength) and added to a point source list. A fraction of it (γ, $0 < \gamma < 1$) is removed from the dirty image. The γ parameter is called the loop gain and is usually taken to be 0.1–0.2. The iterations continue until the residual image is noise like. The subtraction can be done either in the image domain or in the visibility domain. The visibility domain CLEAN is more accurate since we are not limited to pixel resolution. The algorithm flow for ungridded visibility domain CLEAN is summarized in Table 2.

An illustration of the CLEAN algorithm on a simulated image is shown in Figure 9. The simulated radio telescope is the same as in Figure 5. The loop gain used is 0.2. In every

[TABLE 1] GENERIC SOURCE REMOVAL ALGORITHM FLOW.

INITIALIZATION:

- CALCULATE THE DIRTY IMAGE I_D ACCORDING TO MEASURED VISIBILITIES.
- CALCULATE THE RECONSTRUCTION BEAM B_{rec} FOR LATER USE.

WHILE STOPPING CRITERIA NOT MET:

- FIND THE BRIGHTEST LOCATION IN THE DIRTY IMAGE (l_i, m_i). THIS IS THE LOCATION OF A NEW POINT SOURCE.
- ESTIMATE THE NEW POINT SOURCE INTENSITY λ_i.
- ADD THE NEW POINT SOURCE TO THE SOURCE LIST. (WITH THE ESTIMATED INTENSITY).
- REMOVE THE NEW SOURCE RESPONSE FROM THE DATA (BOTH THE DIRTY IMAGE AND THE VISIBILITY MEASUREMENTS).

FINALIZE:

- CALCULATE THE RECONSTRUCTED IMAGE I_{rec} BY CONVOLVING THE SOURCE LIST WITH THE RECONSTRUCTION BEAM.

[TABLE 2] VISIBILITY DOMAIN CLEAN ALGORITHM FLOW.

INITIALIZATION:

- CALCULATE I_D (EQ. 8).
- $i = 0$.
- B_{rec} = Gaussian.

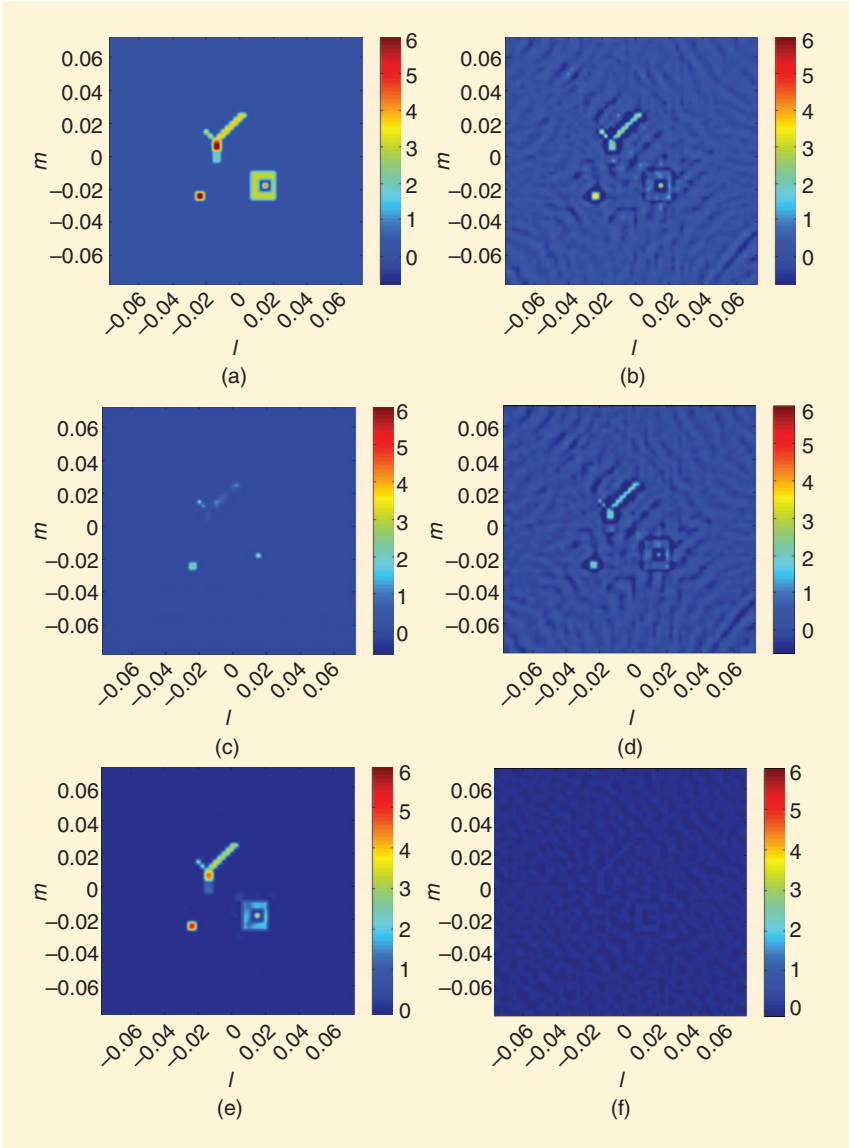
WHILE I_D IS NOT NOISE-LIKE:

- $(l_i, m_i) = \text{argmax}\, I_D(l, m)$.
- $\lambda_i = I_D(l_i, m_i)$.
- FOR ALL p, q, k:
  $V(u^k_{p,q}, v^k_{p,q}) = V(u^k_{p,q}, v^k_{p,q}) - \gamma\lambda_i e^{-2\pi j[u_{p,q}{}^k l_i + v_{p,q}{}^k m_i]}$.
- UPDATE I_D (EQ. 8).
- $i = i + 1$.

FINALIZE:

- $I_{\text{rec}} = I_D + \sum_i \gamma\lambda_i B_{\text{rec}}(l - l_i, m - m_i)$.

iteration, the strongest point source is found, added to the reconstructed image, and subtracted from the dirty image. The loop gain serves three purposes. First, it prevents (or at least reduces) the effects of over-estimation of the power due to sidelobes from other sources. Second, it allows for interpolation of sources that are located off the grid. Third, it improves performance with extended sources. However, this limits the dynamic range of the image. The effect of pixelization and choice of grid on the dynamic range of the imaging process is further discussed in [32] and [33]. Improved versions of CLEAN allow for estimation of location off the grid by using interpolation, and subtraction of the effect from the visibility rather than the dirty image. This has the positive effect of eliminating gridding accuracy effects. Acceleration of the CLEAN algorithm can be achieved by estimating multiple point sources based on a single dirty image (major cycle) as well as defining windows for the search procedure. Practically, defining windows reduces the size of the search space.



[FIG9] **CLEAN steps for a simulated image. (a) The original image and (b) the initial dirty image. The reconstructed image after 50 and 25,000 CLEAN iterations is shown in (c) and (e), respectively. The residual dirty images after 50 and 25,000 iterations are shown in (d) and (f), respectively. In general, the sources are nicely reconstructed except from the square ring extended source. All images share the same color map. All images were up-sampled by four using MATLAB basic interpolation.**

CLARK CLEAN ALGORITHM

One of the important variants of CLEAN was proposed by Clark in 1980 [30]. The main advantage of Clark's algorithm is the reduction of computational load. The algorithm is performed in two cycles: a major cycle and a minor cycle. A major cycle is constructed by selecting intensity limit value (according to a histogram of the dirty image values) and approximating a dirty beam (central patch of the true dirty beam) to be used during the subsequent minor cycles. A minor cycle consists of finding the brightest pixel in the image (i.e., a new point source) and removing a fraction of the point source response from the dirty image. In principle, the minor cycle is the same as described in the "While" loop in Table 2, when the dirty beam used is only the central patch of the full dirty beam (hence computational complexity is significantly reduced). The inaccuracies caused by working with an approximated dirty beam are corrected during the major cycle. The Clark algorithm is performed in the visibility domain instead of the image domain, yielding a multiplication instead of a convolution for calculating the point source response.

COTTON-SCHWAB ALGORITHM

Cotton and Schwab [31] developed a variant of the Clark CLEAN. Like the Clark CLEAN, in the Cotton-Schwab CLEAN, the procedure involves major and minor cycles. The main improvements over the Clark algorithm are that the Cotton-Schwab algorithm calculation is done over the ungridded visibility data, thus avoiding gridding errors, and multisource removal is done independently in each minor cycle (from different fields). The CLEAN components from all fields are removed in the major cycle. Working with the ungridded visibility measurement is done using a measurement list as described in Table 2. An element $V(u_{p,q}^k, v_{p,q}^k)$ of the measurement list is the

measured visibility by an antenna pair (p, q), corresponding to a baseline $(u_{p,q}, v_{p,q})$ measured at time k.

THE W-PROJECTION ALGORITHM

One of the main limitations of the previous technique is the case of noncoplanar arrays and large field of view. To overcome problems related to noncoplanar arrays, the W-projection algorithm has been proposed by Cornwell et al. [34].

The W-projection algorithm deals with noncoplanar arrays i.e., when the planar approximation is violated and the imaging equation is given by (6). Originally, Frater and Docherty [35] showed that a projection of visibility measurements from a constant w plane to $w = 0$ plane can be done. This corresponds to a radio telescope with antennas arranged in a plane with a single antenna outside the plane. In this case, the measured visibilities are projected onto $w = 0$ plane (real and imaginary part separately), a deconvolution is performed (such as CLEAN), and the resulting cleaned images are combined taking the constant w value into account.

In the general case (projection of any w values), the relation between $V(u, v, w)$ and $V(u, v, w = 0)$ is given by

$$V(u, v, w) = \widetilde{G}(u, v, w) * V(u, v, w = 0), \tag{23}$$

where

$$\begin{aligned} G(l, m, w) &\equiv e^{-2\pi\mathcal{J}\,[w(\sqrt{1-l^2-m^2}-1)]} \\ &\approx e^{\pi\mathcal{J}[w(l^2+m^2)]} \\ \widetilde{G}(u, v, w) &= \frac{\mathcal{J}}{w} e^{-\pi\mathcal{J}\left[\frac{u^2+v^2}{w}\right]} \end{aligned} \tag{24}$$

and $G(l, m, w)$ is the Fourier transform of $\widetilde{G}(v, u, w)$ called the W-projection function. Given a model of the sky brightness, the visibility on the $w = 0$ plane can be calculated using the two-dimensional (2-D) Fourier transform. The visibility measurement outside the $w = 0$ plane may then be calculated using the convolution function $\widetilde{G}(u, v, w)$. Note that representing the visibility as a convolution and using the FFT algorithm to compute the convolution is similar to the one-dimensional chirp z-transform algorithm. Calculating the image for a given set of visibility measurements is done using iterative algorithms since there is no inverse transform. The W-projection is a minor-major cycle algorithm that receives three-dimensional visibility measurements $V(u, v, w)$ and projects the w coordinate "out" (projection on $w = 0$ plane). The 2-D visibilities $V(u, v, w = 0)$ are used to calculate the reconstructed image in the (l, m) domain by a 2-D Fourier transform. Then a deconvolution is performed (such as CLEAN) on the resulting image. The W-projection algorithm has both high-performance and high-computational speed.

THE LS-MVI ALGORITHM

We now describe a recent approach that enables the use of modern array processing algorithms in the framework of image deconvolution. The method will be demonstrated on simulated and measured data. However, in contrast to the CLEAN algorithm it is in initial research stages and further development of the technique is an interesting research problem. The LS-MVI algorithm is a novel matrix-based sequential source removal algorithm originally proposed in [19] and further improved in [20]. It is based on the matrix-based approach to direction-of-arrival (DOA) estimation techniques. We would like to replace the vectors $\mathbf{a}_k(\mathbf{s})$ in (19) by a set of beamforming vectors $\mathbf{w}_k(s)$, $k = 1, \ldots, K$. The main goal of the LS-MVI is to eliminate interference from other points in the image when estimating the location and power of a given source. To that end, filterbank techniques such as the MVDR and its extensions have proven very effective. Minimizing the interference from sidelobes of the dirty beam while observing a point source in direction $\mathbf{s} = (l, m)$ can be formulated as a constrained beamforming problem (For simplicity we denote $\mathbf{w}_k(\mathbf{s})$ by $\mathbf{w}_k$ and assume that $\mathbf{w} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]^T$).

$$\begin{aligned} \hat{\mathbf{w}}(\mathbf{s}) \;=\; & \operatorname{argmin}_{\mathbf{w}} \sum_{k=1}^{K} \mathbf{w}_k^H \hat{\mathbf{R}}_k \mathbf{w}_k \\ & \text{subject to} \\ & \sum_{k=1}^{K} \mathbf{w}_k^H \mathbf{a}_k(\mathbf{s}) = 1. \end{aligned} \tag{25}$$

The solution is given by

$$\hat{\mathbf{w}}_k(\mathbf{s}) = \beta(\mathbf{s}) \hat{\mathbf{R}}_k^{-1} \mathbf{a}_k(\mathbf{s}), \tag{26}$$

where $\beta(\mathbf{s}) = 1 \,/\, \sum_{k=1}^{K} \mathbf{a}_k^H(\mathbf{s})\, \hat{\mathbf{R}}_k^{-1} \mathbf{a}_k(\mathbf{s})$, $\mathbf{a}_k(\mathbf{s})$ is given in (15) and $\hat{\mathbf{R}}_k$ is the covariance matrix measured at time t_k. The vectors $\hat{\mathbf{w}}(\mathbf{s})$ have different magnitudes for different values of $\mathbf{s}$. This is undesirable since it generates noise related spatial features. Therefore, the adapted angular response (AAR) solution normalizes the norm of $\mathbf{w}$ to 1. The resulting solution is given by

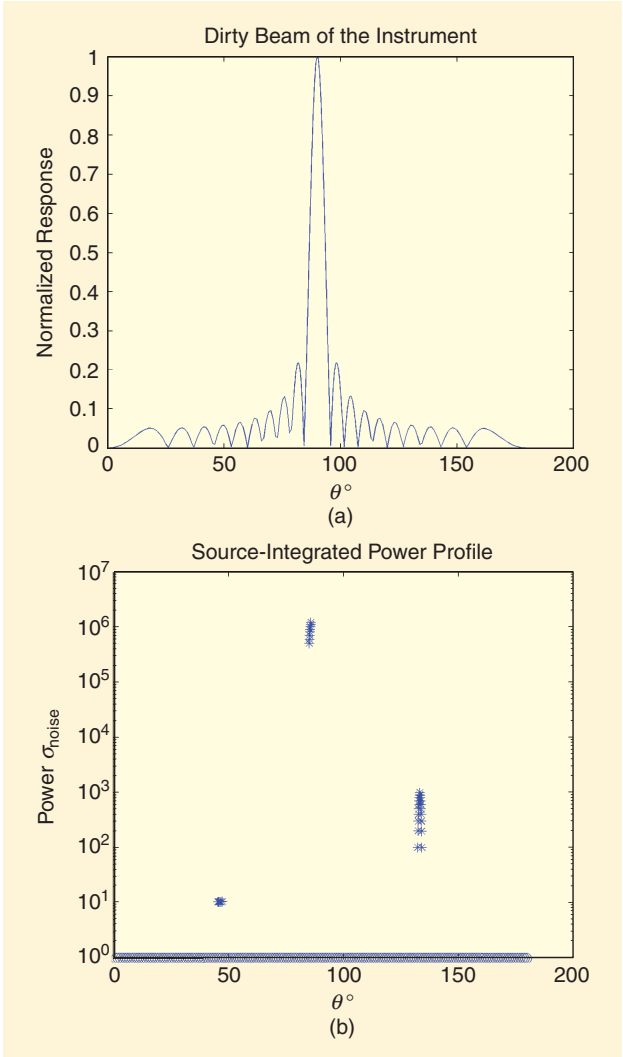
$$I_D^{\mathrm{AAR}}(l, m) \;\equiv\; \frac{\sum_{k=1}^{K} \mathbf{a}_k(l, m)^H \hat{\mathbf{R}}_k^{-1} \mathbf{a}_k(l, m)}{\sum_{k} \mathbf{a}_k(l, m)^H \hat{\mathbf{R}}_k^{-2} \mathbf{a}_k(l, m)}. \tag{27}$$

This modified dirty image replaces the classical dirty image in the LS-MVI deconvolution process (a simple 1-D example of the AAR robustness is shown in "Example of AAR Algorithm Robustness").

The intensity estimation used by the LS-MVI algorithm is a LS estimation of a point source at location (l, m) and given by the following equation:

EXAMPLE OF AAR ALGORITHM ROBUSTNESS

For simplicity, a 1-D simulated example demonstrates AAR robustness. The simulated sparse array response is displayed in Figure 10(a). There are three simulated sources [see Figure 10(b)], two relatively bright extended sources, and a faint source. The brightests source intensity is 10^6 times the RMSE of the noise after integration and the faint source power is only ten times the noise RMSE. The faint source is well below the sidelobes level in the classic dirty image [see Figure 11(a)].



[FIG10] **Simulated case: (a) the sparse array response and (b) sources' integrated power.**

$$\alpha = \arg\min_{\alpha} \sum_{k} \|\hat{\mathrm{R}}_k - \alpha \mathrm{a}_k(l,m)\mathrm{a}_k^H(l,m)\|_F^2 \quad \text{subject to } \alpha \geq 0. \tag{28}$$

This estimate of the source power has been independently used in ASP-CLEAN [36]. The closed form solution of (28) is given by

$$\alpha = \max\left\{\frac{\boldsymbol{h}^H \boldsymbol{r}}{\boldsymbol{h}^H \boldsymbol{h}}, 0\right\}, \tag{29}$$

where

$$h \equiv [\mathrm{vec}(\boldsymbol{a}_1(l,m)\boldsymbol{a}_1^H(l,m))^T, \ \ldots, \mathrm{vec}(\boldsymbol{a}_K(l,m)\boldsymbol{a}_K^H(l,m))^T]^T$$

and $\boldsymbol{r} \equiv [\mathrm{vec}(\hat{\boldsymbol{R}}_1)^T, \ \ldots, \mathrm{vec}(\hat{\boldsymbol{R}}_K)^T]^T$ are obtained by stacking the array response and the measured covariance matrices respectively.

The intensity estimation can be improved by adding the semidefinite constraint

[TABLE 3] LS-MVI ALGORITHM FLOW.

INITIALIZATION:

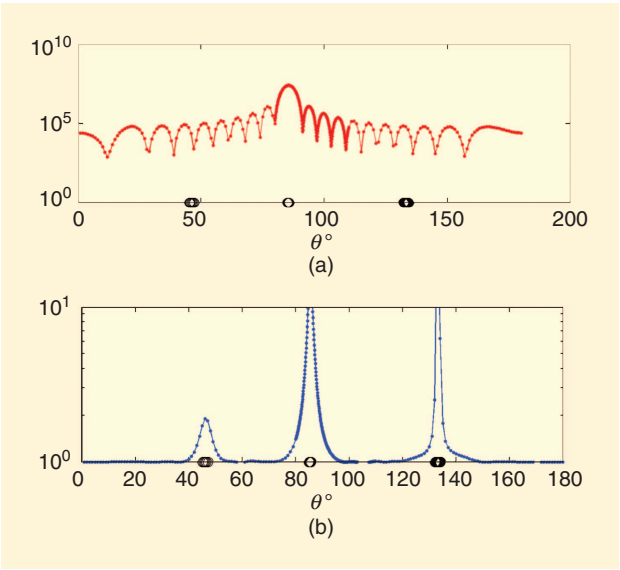
- $\boldsymbol{R}_k^0 = \boldsymbol{R}_k$, $\forall k = 1, \ \ldots, K$
- CALCULATE I_D^{AAR} USING EQ. (27)
- $i = 0$
- $B_{\mathrm{rec}} = Gaussian$

WHILE I_D IS NOT NOISE-LIKE:

- $(l_i, m_i) = \mathrm{argmax}\, I_D^{\mathrm{AAR}}(l, m)$
- ESTIMATE α_i ACCORDING TO EQ.(29)
- OPTIONALLY IMPROVE α_i ESTIMATION ACCORDING TO EQ. (30)
- $R_k^{i+1} = R_k^i - \gamma\alpha_i a_k(l_i, m_i)\, a_k^H(l_i, m_i)$, $\forall k = 1 \ldots K$
- CALCULATE I_D^{AAR} USING R_k^{i+1} USING EQ. (27)
- $i = i + 1$

FINALIZE:

- $I_{\mathrm{rec}} = \sum_i \gamma\alpha_i B_{\mathrm{rec}}(l - l_i, m - m_i)$



[FIG11] **Dirty images: (a) the classic dirty image and (b) the AAR dirty image.**

$$\hat{\mathrm{R}}_k - \sigma^2 \mathrm{I} - \alpha \mathrm{a}_k(l,m)\mathrm{a}_k^H(l,m) \succcurlyeq 0. \tag{30}$$

The intensity estimation is bounded between the solution (29) and 0. Hence, a better intensity estimation can be achieved using a simple bisection. A summary of the LS-MVI algorithm is given in Table 3. Another improvement that has low computational complexity is to use a joint LS estimate of all previously estimated sources. Assuming that we have collected L components the estimator is given by

$$\boldsymbol{\alpha} = \mathrm{argmin}_{\boldsymbol{\alpha}} \sum_{k=1}^{K} \| \mathrm{r}_k - \sum_{i=1}^{L} \alpha_i \mathrm{q}_{ki} \|^2 \quad \text{s.t.} \alpha_i \geq 0 \ \text{for all} \ i, \tag{31}$$

where $\alpha = [\alpha_1, \ \ldots, \alpha_L]$, $\mathrm{r}_k = vec(\mathrm{R}_k)$ and $\mathrm{q}_{ki} = \mathrm{vec}(\mathrm{a}_k(l_i, m_i)\, \mathrm{a}_k^H(l_i, m_i))$. Similarly to the CLEAN algorithm this improvement can be implemented only at major cycles, after several sources have been estimated.

There are two main differences between LS-MVI and CLEAN. First, the LS-MVI uses a different type of dirty image and

second, the LS-MVI performs a more sophisticated intensity estimation than CLEAN. The dirty image used by the LS-MVI is I_D^{AAR} given in (27). The main advantage of the AAR dirty image over simple MVDR is the isotropic noise response that prevents the formation of spatially varying noise related artifacts. In [20], further extensions for enforcing semidefinite constraints in a Cotton-Schwab type of iteration are also presented. It should also be noted that there is no need to compute the complete dirty image to find the maximum and optimization techniques can do this much faster, especially if the user can provide windows similar to CLEAN windows currently used by radio astronomers. Like CLEAN, the LS-MVI should be implemented in the visibility domain to eliminate gridding effects.

GLOBAL OPTIMIZATION-BASED TECHNIQUES

We now turn to a second family of solutions to the image formation problem. These solutions are based on optimizing a global property of the image subject to goodness of fit to the data. They vary from LS-based techniques to maximum entropy and ℓ_1-based reconstruction.

LINEAR DECONVOLUTION

Computationally, the simplest way to solve the image formation problem is through linear inversion. There are two main approaches in this area: The well-known LS technique and linear minimum mean square error (LMMSE). Such techniques can work well when the (u, v) coverage is good and the inversion is well conditioned. Furthermore, linear inversion can work independently of the complexity of the source structure. However, linear techniques can result in significant noise enhancement in ill-posed problems. For a fully sampled visibility domain, these techniques can provide a first approximation to the image. To overcome this problem, one can use a constrained LS, also known as nonnegative LS (NNLS), first proposed for radio synthesis imaging by Briggs [18]. The idea is that the image is positive. Putting these constraints into the deconvolution yields a computationally expensive, though feasible, algorithm. An excellent overview of the implementation of the NNLS can be found in [18].

MAXIMUM ENTROPY IMAGE RECONSTRUCTION

The maximum entropy image formation technique is one of the two most popular deconvolution techniques in radio astronomy (together with CLEAN). The maximum entropy principle was first proposed by Jaynes [37]. A good overview of the philosophy behind the idea can be found in [38]. Since then, it has been used in a wide spectrum of imaging problems. The basic idea behind MEM is that out of all the images that are consistent with the measured data where the noise distribution does not satisfy the positivity demand, i.e., the sky brightness is a positive function, consider only those that satisfy the positivity demand. From these select the one that is most likely to have been created randomly. This idea was also proposed by [11] for optical images and applied to radio astronomical imaging in [12]. Other approaches based on differential entropy have also been suggested [13], [14]. An extensive collection of papers discussing these different methods and aspects of maximum entropy can be found in a number of papers in [39]. An overview of various maximum entropy techniques and the use of the various options for choosing the entropy measure is provided by [40]. Interestingly, in that paper, a closed-form solution is given for the noiseless case, but not for the general case.

The approach in [12] begins with a prior image and iterates between maximizing the entropy function and updating the χ^2 fit to the data. The computation of the image based on a prior image is done analytically, but at each step the model visibilities are updated, through a 2-D Fourier transform. This type of algorithm is known as a fixed-point algorithm, since it is based on iterating a function until it converges to a fixed point.

The maximum entropy solution is given by solving the following Lagrangian optimization problem [12]:

$$I^{\mathrm{MEM}} = \arg\max_I - \sum_{l,m} I(l,m)\log\frac{I(l,m)}{F(l,m)} - \frac{\lambda}{2}\chi^2(V), \quad (32)$$

where

$$\chi^2(V) = \sum_{(u,v)\in A}\frac{1}{\sigma^2}|\hat{V}(u,v) - V(u,v)|^2, \quad (33)$$

$V(u,v)$ are the model-based visibilities, λ is a Lagrange multiplier for the constraint that $V(u,v)$ should match the measured visibilities $\hat{V}(u,v)$, A is the (u,v) coverage of the radio telescope, and $F(l,m)$ is a reference image. Taking the derivative with respect to $I(l,m)$ we obtain that the solution is given by

$$I(l,m) = \exp(-1 + \log F(l,m) + \lambda\Delta(l,m)), \quad (34)$$

where

$$\Delta(l,m) = \sum_{(u,v)\in A}\frac{1}{\sigma^2}Re\left((\hat{V}(u,v) - V(u,v))e^{\frac{\mathcal{J}\,2\pi(ul+vm)}{N}}\right).$$

The basic MEM now proceeds by choosing an initial image model (typically a flat image or a low-resolution image) computing the model-based visibilities $V(u,v)$ on a grid A. Using these visibilities a new model image is computed by (34). New visibilities are computed from the new model and the process is iterated until convergence.

While it is known that for the maximum entropy, this approach usually converges, the convergence can be slow [40]. Improved methods based on the Newton method and the conjugate gradient technique were put forward by [15], [41], and [42]. These methods perform direct optimization of the entropy function subject to the χ^2 constraint. Generalization of the maximum entropy using wavelets and multiresolution techniques have also been proposed (see e.g., [17] and [43]).

COMPRESSED SENSING AND SPARSE RECONSTRUCTION TECHNIQUES

Recently there has been growing interest in using ℓ_1-based cost functions for deconvolution (see [21], [22], and the unpublished notes by Ludwig Schwardt, which can be found at https://safe.nrao.edu/wiki/pub/Software/Callm09Program/calim2009_ludwig.pdf).

This renewed interest in ℓ_1 comes from recent results related to compressed sampling using Fourier bases. It is worth noting that as early as 1987, Marsh and Richardson [44] proved that the CLEAN algorithm can be regarded as an ℓ_1 minimization for a single-point source image. However, ℓ_1 is not the only criterion. Recovery of noisy and blurred images using total variation (TV) optimization for smooth images was discussed by Dobson and Santosa [45]. Chen et al. [46] dealt with ℓ_1 minimization of an image basis to achieve image sparseness using linear programming. Feuer and Nemirovski [47] and Elad and Bruckstein [48] established sufficient and necessary conditions for replacing ℓ_0 optimization (computing the sparsest solution with high computational complexity) by linear programming when searching for the unique sparse representation. Rudelson and Vershynim [49] proved the best known guarantees for exact reconstruction of a sparse signal from its Fourier measurements.

Radio astronomical image reconstruction is done based on the visibility measurement in the (u, v) domain. Reconstruction of the source image $I(l, m)$ is equivalent to estimating the missing visibility points. The missing $V(u, v)$ measurements together with the image itself are estimated by minimizing a cost function $\|I(l, m)\|_{\ell_1}$ in the (l, m) domain using the constraints of image positivity and the measured visibility data. Note that since $I(l, m)$ is a positive quantity we have

$$\|I(l, m)\|_{\ell_1} = \sum_{l=1}^{N} \sum_{m=1}^{N} I(l, m) \tag{35}$$

that allows us to use linear programming. To solve the reconstruction problem fast, we represent the problem as a linear programming problem with real variables. To that end let $\langle \cdot, \cdot \rangle$ be a one-to-one pairing function mapping $\{0, \ldots, N-1\} \times \{0, \ldots, N-1\}$ onto $\{0, \ldots, N^2-1\}$. Let $\mathbf{F}$ be an $N^2 \times N^2$ matrix whose elements satisfy

$$\mathbf{F}_{\langle l,m\rangle,\langle u,v\rangle} = e^{\frac{-2\pi\mathcal{I}}{N}(ul+vm)}. \tag{36}$$

Let $\xi = \text{vec}(\mathbf{V})$ and let $\mathbf{t} = \text{vec}(\mathbf{I})$. We have

$$\xi = \mathbf{F}\mathbf{t}. \tag{37}$$

Note that $\mathbf{t}$ is a real vector since the visibility measurements satisfy $V(u, v) = \overline{V}(-u, -v)$. To make the problem real, we define $\mathbf{F}_R = \text{Re}(\mathbf{F})$, $\mathbf{F}_I = \text{Im}(\mathbf{F})$ and variables $\xi_R = \text{Re}(\xi)$, $\xi_I = \text{Im}(\xi)$. (37) now becomes

$$\begin{aligned} \xi_R &= \mathbf{F}_R\mathbf{t} \\ \xi_I &= \mathbf{F}_I\mathbf{t}. \end{aligned} \tag{38}$$

For the measured locations (u_i, v_i) we have

$$\begin{aligned} \xi_R(\langle u_i, v_i\rangle) &= \text{Re}(\hat{V}(u_i, v_i)) \quad i = 1, \ldots, M \\ \xi_I(\langle u_i, v_i\rangle) &= \text{Im}(\hat{V}(u_i, v_i)) \quad i = 1, \ldots, M, \end{aligned} \tag{39}$$

where M is the number of given measurements in the (u, v) domain. The linear programming problem is described in Table 4 (for more details the reader is referred to [21]).

In [22], a joint ℓ_1 and ℓ_2 is also discussed. This makes it possible to include prior knowledge on the noise power. Using the total variation is also a possibility that leads to ℓ_1 optimization.

[TABLE 4] ℓ_1 OPTIMIZATION USING LINEAR PROGRAMMING.

$\min_t \sum_{i=1}^{N^2} t_i$
SUBJECT TO
$\xi_R(\langle u_i, v_i\rangle) = Re(\hat{V}(u_i, v_i))$
$\xi_I(\langle u_i, v_i\rangle) = Im(\hat{V}(u_i, v_i))$
$0 \le \mathbf{t}$

Note that using total variation and maximum entropy are related since both functionals impose smoothness on the image.

SELF-CALIBRATION AND ROBUST MVDR FOR SYNTHETIC APERTURE ARRAYS

We now turn to the case where the array response is not completely known, but we have some statistical knowledge of the error, e.g., we know the covariance matrix of the array response error at each epoch (measurement time). Typically this covariance will be time invariant or will have slow temporal variation. In this case, we extend the robust dirty image as described in [50] to the synthetic aperture array case. This generalization follows the analysis in [20]. Since the positive definite constraint on the residual covariance matrices is important in our application, we extended the robust Capon estimator of [51]. To that end, assume that at each epoch we have an uncertainty ellipsoid describing the uncertainty of the array response (as well as unknown atmospheric attenuation). This is described by

$$(\mathbf{a}_k(\mathbf{s}) - \bar{\mathbf{a}}_k(\mathbf{s}))^H \mathbf{C}_k (\mathbf{a}_k(\mathbf{s}) - \bar{\mathbf{a}}_k(\mathbf{s})) \le 1, \tag{40}$$

where $\bar{\mathbf{a}}_k(\mathbf{s})$ is the nominal value of the array response towards the point $\mathbf{s}$ and $\mathbf{C}_k$ are the covariance matrices of the uncertainty in the calibration parameters at time k. Generalizing the LS-MVI we would like to solve the following problem:

$$\begin{aligned} &[\hat{\rho}, \hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_k] = \arg\max_{\rho, \mathbf{a}_1, \ldots, \mathbf{a}_k} \rho \\ &\text{subject to} \\ &\hat{\mathbf{R}}_k - \sigma^2\mathbf{I} - \rho\mathbf{a}_k\mathbf{a}_k^H \succcurlyeq \mathbf{0} \qquad k = 1, \ldots, K \\ &(\mathbf{a}_k(\mathbf{s}) - \bar{\mathbf{a}}_k(\mathbf{s}))^H \mathbf{C}_k (\mathbf{a}_k(\mathbf{s}) - \bar{\mathbf{a}}_k(\mathbf{s})) \le 1 \quad k = 1, \ldots, K. \end{aligned} \tag{41}$$

Let $\tau = 1/\rho$. The problem (41) is equivalent to the following problem

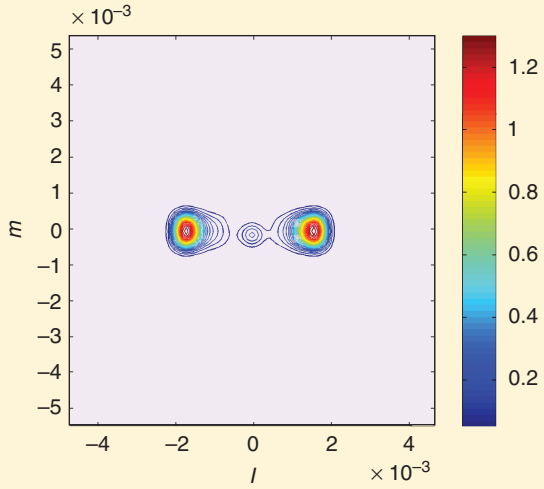
$$[\hat{\tau}, \hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_k] = \arg\min_{\tau, \mathbf{a}_1, \ldots, \mathbf{a}_k} \tau$$

subject to

$$\begin{bmatrix} \hat{\mathbf{R}}_k - \sigma^2\mathbf{I} & \mathbf{a}_k \\ \mathbf{a}_k^H & \tau \end{bmatrix} \succcurlyeq \mathbf{0} \qquad k = 1, \ldots, K$$

$$\begin{bmatrix} \mathbf{C}_k & (\mathbf{a}_k(\mathbf{s}) - \bar{\mathbf{a}}_k(\mathbf{s})) \\ (\mathbf{a}_k(\mathbf{s}) - \bar{\mathbf{a}}_k(\mathbf{s}))^H & 1 \end{bmatrix} \succcurlyeq \mathbf{0} \qquad k = 1, \ldots, K. \tag{42}$$

This problem is once again a semidefinite programming problem that can be solved efficiently via interior point techniques [52]. We can now replace the MVDR estimator by this robust version. Interestingly, we obtain estimates of the corrected array response $\hat{\mathbf{a}}(\mathbf{s}_k)$. Using the model we obtain for each k



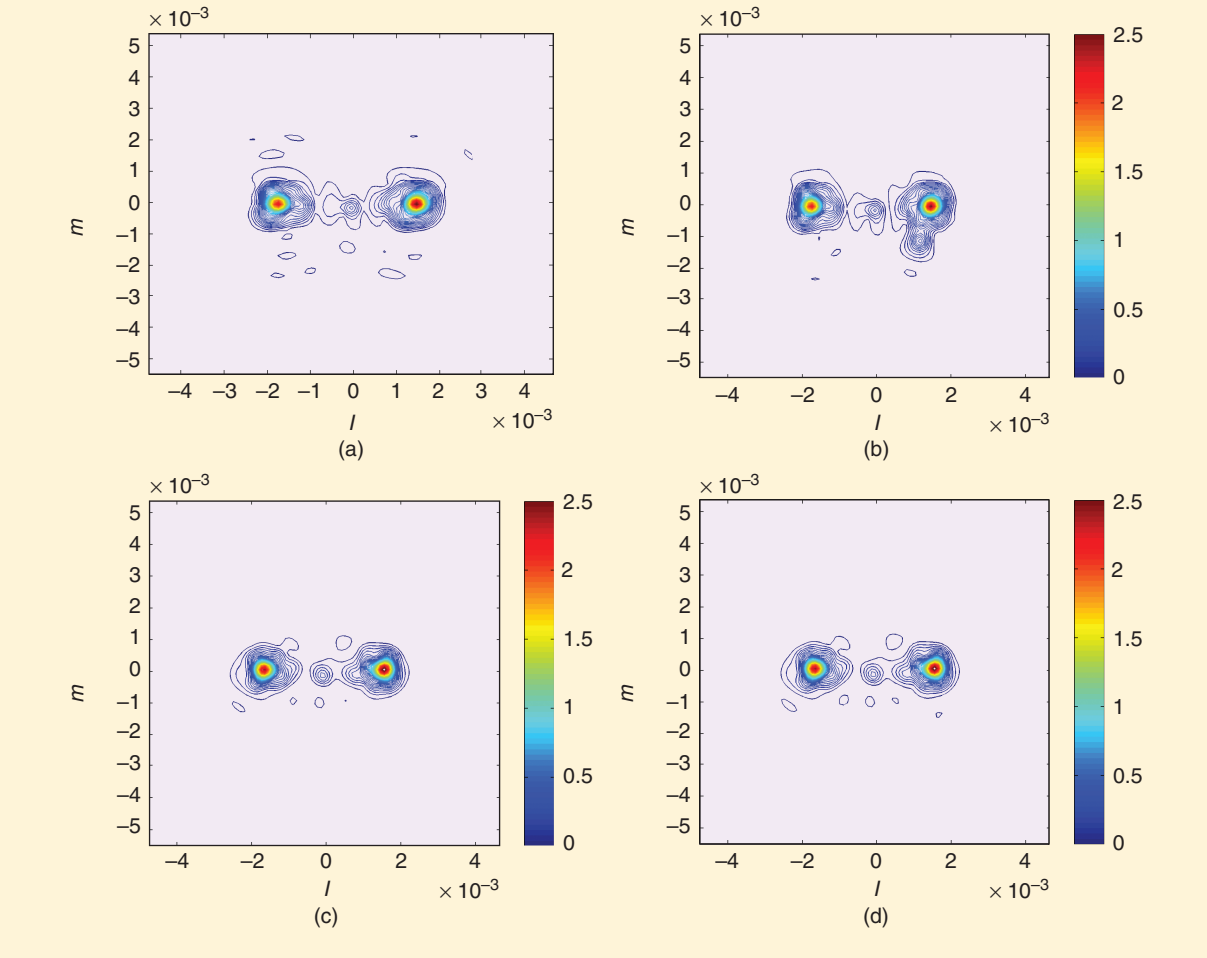
[FIG12] **Original extended source image.**

$$\mathbf{a}_k(\mathbf{s}) = \Gamma_k \bar{\mathbf{a}}_k(\mathbf{s}). \tag{43}$$

Hence, the self-calibration coefficients can be estimated using LS fitting

$$\hat{\Gamma}_k = \arg\min_{\gamma_1,\ldots,\gamma_p} \sum_{\ell=1}^{L} \|\hat{\mathbf{a}}_k(\mathbf{s}_\ell) - \Gamma_k \bar{\mathbf{a}}_k(\mathbf{s}_\ell)\|^2, \tag{44}$$

where $\Gamma_k = \text{diag}\{\gamma_{k,1}, \ldots, \gamma_{k,p}\}$. Of course, when the self-calibration parameters vary slowly we can combine the estimation over multiple epochs. This might prove instrumental in calibration of LOFAR type arrays, where the calibration coefficients vary across the sky. Since the computational complexity of the self-calibration semidefinite programming is higher than that of the MVDR dirty image, it is too complicated to solve this problem for each source in the image. Hence, it should be used in a way similar to the external self-calibration cycle [53] where this problem is solved using a nominal source locations model. The advantage over ordinary self-calibration is that beyond the reevaluation of the calibration parameters, we obtain better estimates of the source powers, without



[FIG13] **Reconstructed images of the CLEAN and LS-MVI algorithms: (a) CLEAN reconstructed image after 100 iterations, (b) CLEAN reconstructed image after 120 iterations, (c) the LS-MVI reconstructed image after 100 iterations, and (d) the LS-MVI reconstructed image after 300 iterations.**

significant increase in the complexity. Another interesting alternative is to use the doubly constrained robust Capon beamformer that combines a norm constraint as in the AAR dirty image with robust Capon beamforming [54].

EXAMPLES AND COMPARISONS

In this section, we describe three examples of the various algorithms, including a simulated example of an extended source, an example from the LOFAR test station, and an example of Abell 2256 observed by the VLA (initial calibration was conducted by Tracy Clarke).

SIMULATED EXTENDED SOURCE

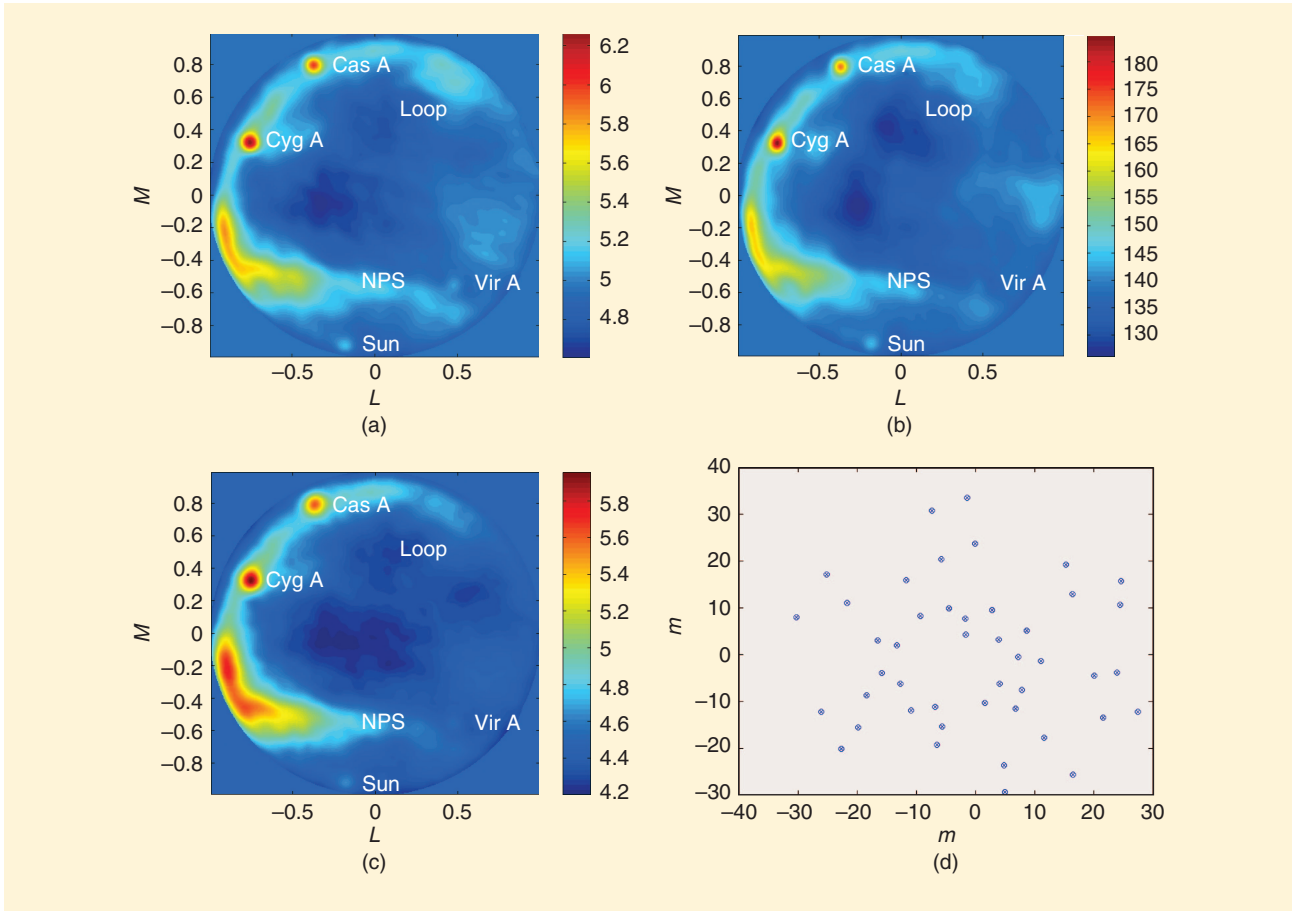
An extended source (Figure 12) was simulated using an east-west array containing ten antennas logarithmically spaced up to 1,000 λ. CLEAN deconvloution results are depicted in Figure 13(a) and (b). After 100 CLEAN iterations, the center of the source is partially reconstructed with distortion. After additional 20 iterations an artifact is generated (below the strong point on the right). This divergence can often occur in CLEAN when applying it to extended sources. The LS-MVI results are presented in Figure 13(c) and (d). After 100 iterations, the center of the source is reconstructed and after 200 additional iterations, the center of the source is stable. The reason for this is the fact that CLEAN overestimate the power due to the high sidelobes level. Further analysis of this example is given in [20].

LOFAR TEST STATION DATA

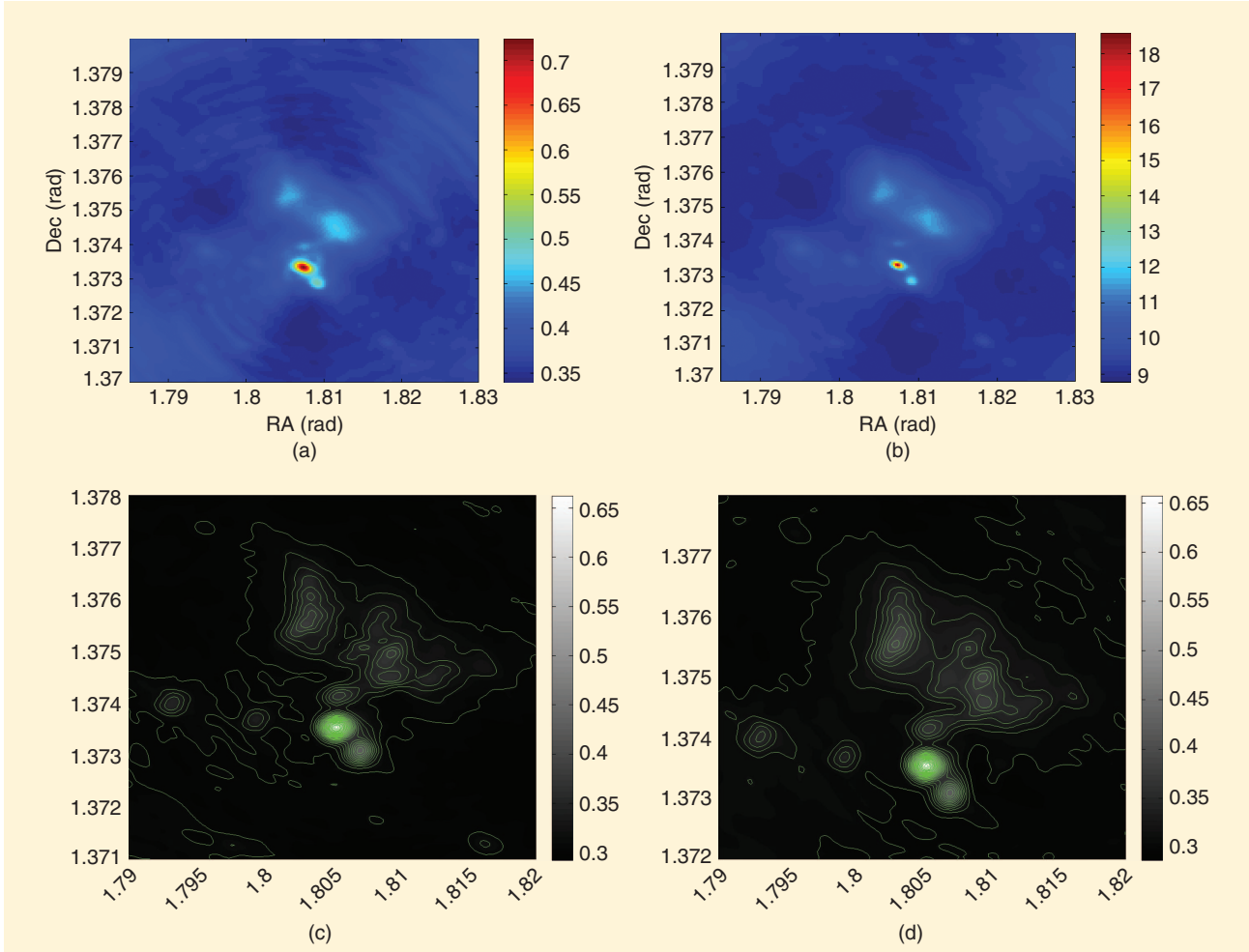
The LOFAR test station data were recorded using 25 frequency bands of 156 kHz using 45 antennas [array geometry is given in Figure 14(d)]. The data were calibrated by S. Wijnholds. The AAR dirty image and the classic dirty image are given in Figure 14(a) and (b), respectively. Since the LOFAR station benefits from a good (u, v) domain coverage, the two dirty images are similar. The reconstructed image using the LS-MVI algorithm is displayed in Figure 14(c); the spurious emission on the right side of the image was removed.

ABELL 2256

The last example used VLA data of Abell 2256 [55]. Abell 2256 is a merging of two (or three) large clusters of more than 500 galaxies. It exhibits strong radio emissions and is one of the strongest X-ray emitters. The data measured by Clarke and Ensslin [56] contain a single frequency band around 1369 MHz. The data were processed using both CLEAN and LS-MVI algorithms for 30 iterations. This is the



[FIG14] LOFAR station images: (a) classic dirty image, (b) AAR dirty image, (c) reconstructed image using AAR-based cleaning, and (d) array geometry.



[FIG15] Abell 2256 images: (a) initial classic dirty image, (b) initial AAR dirty image, (c) classic CLEAN reconstructed image, and (d) the LS-MVI reconstructed image.

first example of an application of the LS-MVI algorithm for measured data. As such it is only a preliminary example and significant improvements can be made, e.g., in [56] data were also self-calibrated using phase data and then amplitude and phase. This is required here to achieve a deeper level of cleaning. We used a visibility domain CLEAN (updates were performed on the ungridded visibility). The initial data (dirty image) of the CLEAN are shown in Figure 15(a). The strong sidelobes structure is clearly visible as large circles in the dirty image. In contrast the initial AAR dirty image is shown in Figure 15(b). The sidelobes level is much lower and several point sources that are invisible in the classical dirty image are now visible. The reconstruction using the visibility domain CLEAN is shown in Figure 15(c). The sidelobes level is reduced and the source structure is clearly seen. The reconstruction using the LS-MVI algorithm is shown in Figure 15(d). Similar to the CLEAN, the sources structure is visible and the sidelobes level is significantly reduced. It should be emphasized that even though we have used only 30 iterations, the strong structure is consistent with that of [56] and [57].

ACKNOWLEDGMENTS

We would like to thank T. Clarke, H. Intema, H. Rottgering, and S. Wijnholds for providing the data used to demonstrate the various techniques, Seth Shostak and the SETI Institute for providing the photo of the Allen telescope array, and NRAO for permission to use VLA images. We would also like to thank the anonymous reviewers and the guest editor A-J. van der Veen for comments that significantly enhanced the presentation. Amir Leshem was partially supported by NWO-STW grants 10459 and DTC.5893 (VICI-SPCOM).

AUTHORS

Ronny Levanda (ronny.levanda@gmail.com) received her B.Sc. degree in physics and her M.Sc. degree in neural networks from Tel Aviv University, in 1995 and 2000, respectively. She is currently studying towards her Ph.D. degree at Bar-Ilan University in Israel.

Amir Leshem (leshema@eng.biu.ac.il) received the B.Sc. degree (cum laude) in mathematics and physics, the M.Sc. degree (cum laude) in mathematics, and the Ph.D. degree in mathematics, all from the Hebrew University, Jerusalem, Israel. He is one of

the founders of the School of Electrical and Computer Engineering, Bar-Ilan University, Ramat Gan, Israel, where he is currently an associate professor and head of the signal processing track. His main research interests include multichannel communication, applications of game theory to communication, array and statistical signal processing with applications to sensor arrays and networks, wireless communications, radio-astronomy, brain research, set theory, logic, and foundations of mathematics.

REFERENCES

[1] K. Jansky, "Electrical disturbances apparently of extraterrestrial origin," *Proc. IRE*, vol. 21, pp. 1387–1398, Oct. 1933.

[2] G. Reber, "Cosmic statics," *Proc. IRE*, vol. 28, pp. 68–70, Feb. 1940.

[3] M. Ryle and D. Vonberg, "Solar radiation on 175 Mc./s," *Nature*, vol. 158, pp. 339–340, Sept. 1946.

[4] A. A. Penzias and R. W. Wilson, "A measurement of excess antenna temperature at 4080 Mc/s.," *Astrophys. J.*, vol. 142, pp. 419–421, July 1965.

[5] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P.D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D.T. Wilkinson, "Structure in the COBE differential microwave radiometer first-year maps," *Astrophys. J. Lett.*, vol. 396, pp. L1–L5, Sept. 1992.

[6] A. Hewish, S. J. Bell, J. D. H. Pilkington, P. F. Scott, and R. A. Collins, "Observation of a rapidly pulsating radio source," *Nature*, vol. 217, pp. 709–713, Feb. 1968.

[7] H. Ewen and E. Purcell, "Observation of a line in the galactic radio spectrum," *Nature*, vol. 168, p. 356, Feb. 1951.

[8] R. A. Perley, J. W. Dreher, and J. J. Cowan, "The jet and filaments in Cygnus A," *Astrophys. J. Lett.*, vol. 285, pp. L35–L38, Oct. 1984.

[9] M. Ryle, "The new Cambridge radio telescope," *Nature*, vol. 194, no. 4828, pp. 517–518, 1962.

[10] J. A. Högbom, "Aperture synthesis with nonregular distribution of interefer-ometer baselines," *Astron. Astrophys. Suppl.*, vol. 15, pp. 417–426, June 1974.

[11] B. Frieden, "Restoring with maximum likelihood and maximum entropy," *J. Opt. Soc. Amer.*, vol. 62, no. 4, pp. 511–518, 1972.

[12] S. Gull and G. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, pp. 686–690, April 1978.

[13] J. Ables, "Maximum entropy spectral analysis," *AAS*, vol. 15, pp. 383–393, June 1974.

[14] S. Wernecke, "Two dimensional maximum entropy reconstruction of radio brightness," *Radio Sci.*, vol. 12, pp. 831–844, Oct. 1977.

[15] T. Cornwell and K. Evans, "A simple maximum entropy deconvolution algorithm," *Astron. Astrophys.*, vol. 143, no. 1, pp. 77–83, Feb. 1985.

[16] U. Rau, S. Bhatnagar, M. Voronkov, and T. Cornwell, "Advances in calibration and imaging techniques in radio interferometry," *Proc. IEEE*, vol. 97, pp. 1472–1481, Aug. 2009.

[17] E. Pantin and J.-L. Starck, "Deconvolution of astronomical images using the multiscale maximum entropy method," *Astron. Astrophys. Suppl.*, vol. 118, pp. 575–585, Sept. 1996.

[18] D. S. Briggs, "High fidelity deconvolution of moderately resolved sources," *Ph.D. thesis*, The New Mexico Inst. Mining and Technol., Socorro, New Mexico, 1995.

[19] A. Leshem and A. van der Veen, "Radio-astronomical imaging in the presence of strong radio interference," *IEEE Trans. Inform. Theory (Special Issue on Inform. Theoretic Imag.)*, vol. 46, pp. 1730–1747, Aug. 2000.

[20] C. Ben-David and A. Leshem, "Parametric high resolution techniques for radio astronomical imaging," *IEEE J. Select. Top. Signal Process.*, vol. 2, pp. 670–684, Oct. 2008.

[21] R. Levanda and A. Leshem, "Radio astronomical image formation using sparse reconstruction techniques," in *Proc. IEEE 25th Convention of Electrical and Electronics Engineers in Israel*, Dec. 2008, pp. 716–720.

[22] Y. Wiaux, L. Jacques, G. Puy, A. Scaife, and P. Vandergheynst, "Compressed sensing imaging techniques for radio interferometry," *Monthly Notices Royal Astron. Soc.*, vol. 395, pp. 1733–1742, May 2009.

[23] R. Reid, "Smear fitting: A new image-deconvolution method for interferometric data," *Monthly Notices Royal Astron. Soc.*, vol. 367, no. 4, pp. 1766–1780, 2006.

[24] A. Thompson, J. Moran, and G. Swenson, Eds., *Interferometry and Synthesis in Radio Astronomy*. New York: Wiley, 1986.

[25] G. Taylor, C. Carilli, and R. Perley, "Synthesis imaging in radio-astronomy," *Astron. Soc. of the Pacific*, 1999, National Radio Astronomy Observatory, Charlottesville, VA.

[26] A. Leshem, A. van der Veen, and A. J. Boonstra, "Multichannel interference mitigation techniques in radio-astronomy," *Astrophys. J. Suppl.*, vol. 131, pp. 355–373, Nov. 2000.

[27] S. van der Tol, Bayesian Estimation for Ionospheric Calibration in Radio Astronomy. Ph.D. dissertation, Delft Univ. of Technology, 2009.

[28] R. N. S. Wijnholds, S. van der Tol, and A.-J. van der Veen, "Calibration challenges for future radio telescopes," *IEEE Signal Processing Mag.*, vol. 27, no. 1, pp. 30–42, 2010.

[29] T. Cornwell, "Multiscale CLEAN deconvolution of radio synthesis images," *IEEE J. Select. Top. Signal Process.*, vol. 2, pp. 793–801, Oct. 2008.

[30] B. G. Clark, "An efficient implementation of the algorithm "clean," *Astron. Astrophys.*, vol. 89, pp. 377–378, Sept. 1980.

[31] F. R. Schwab, "Relaxing the isoplanatism assumption in self-calibration: Applications to low-frequency radio interfetomerty," *Astron. J*, vol. 89, pp. 1076–1081, July 1984.

[32] W. D. Cotton and J. M. Uson, "Pixelization and dynamic range in radio interferometry," *Astron. Astrophys.*, vol. 490, pp. 455–460, Oct. 2008.

[33] M. A. Voronkov and M. H. Wieringa, ""The Cotton-Schwab CLEAN at ultra-high dynamic range," *Exp. Astron.*, vol. 18, pp. 13–29, Apr. 2004.

[34] T. Cornwell, K. Golap, and S. Bhatnagar, "The non-coplanar baselines effect in radio interferometry: The w-projetion algorithm," *IEEE J. Select. Top. Signal Process.*, vol. 2, pp. 647–657, Oct. 2008.

[35] R. H. Frater and I. S. Docherty, "On the reduction of three dimensional interferometer measurements," *Astron. Astrophys.*, vol. 84, pp. 75–77, Apr. 1980.

[36] S. Bhatnager and T. Cornwell, "Adaptive scale sensitive deconvolution of interferometric images I. Adaptive scale pixel (asp) decomposition," *Astron. Astrophys.*, vol. 426, no. 2, pp. 747–754, 2004.

[37] E. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, pp. 620–630, May 1957.

[38] E. T. Jaynes, "On the rational of maximum-entropy methods," *Proc. IEEE*, vol. 70, pp. 939–952, Sept. 1982.

[39] J. Roberts, Ed., *Indirect Imaging*. Cambridge, U.K.: Cambridge Univ., 1984.

[40] R. Narayan and R. Nityananda, "Maximum entropy image restoration in astronomy," *Annu. Rev. Astron. Astrophys.*, vol. 24, pp. 127–170, 1986.

[41] R. Sault, "A modification of the Cornwell and Evans maximum entropy algorithm," *Astrophys. J.*, vol. 354, pp. L61–63, May 1990.

[42] J. Skilling and R. Bryan, "Maximum entropy image reconstruction – General algorithm," *Monthly Notices Royal Astron. Soc.*, vol. 211, no. 1, pp. 111–124, 1984.

[43] K. Maisinger, M. P. Hobson, and A. N. Lasenby, "Maximum-entropy image reconstruction using wavelets," *Monthly Notices Royal Astron. Soc.*, vol. 347, pp. 339–354, Jan. 2004.

[44] K. Marsh and J. Richardson, "The objective function implicit in the CLEAN algorithm," *Astron. Astrophys.*, vol. 182, pp. 174–178, Aug. 1987.

[45] D. Dobson and F. Santosa, "Recovery of blocky images from noisy and blurred data," *SIAM J. Appl. Math.*, vol. 56, no. 4, pp. 1181–1198, 1996.

[46] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *Siam J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[47] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1579–1581, June 2003.

[48] M. Elad and A. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory,* vol. 48, no. 9, pp. 2558–2567, Sept. 2002.

[49] M. Rudelson and R. Vershynim, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," Mar 2006.

[50] A. van der Veen, A. Leshem, and A. Boonstra, "Array signal processing in radio-astronomy," *Exp. Astron.*, vol. 17, pp. 231–249, June 2004.

[51] P. Stoica, Z. Wang, and J. Li, "Robust Capon beamforming," *IEEE Signal Process*. Lett., vol. 10, pp. 172–175, June 2003.

[52] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, pp. 49–95, Mar. 1996.

[53] T. Pearson and A. Readhead, "Image formation by self-calibration in radio astronomy," *Annu. Rev. Astron. Astrophys.*, vol. 22, pp. 97–130, 1984.

[54] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust Capon beamformer," *IEEE Trans. Signal Process.*, vol. 52, pp. 2407–2423, Sept. 2004.

[55] H. Rottgering, I. Snellen, G. Miley, J. P. de Jong, R. J. Hanisch, and R. Perley, "VLA observations of the rich X-ray cluster Abell 2256," *Astrophys. J.*, vol. 436, pp. 654–668, Dec. 1994.

[56] T. Clarke and T. Ensslin, "Deep 1.4 GHz very large array observations of the radio halo and relic in Abell 2256," *Astron. J.*, vol. 131, pp. 2900–2912, June 2006.

[57] A. H. Bridle and E. B. Fomalont, "Complex radio emission from the X-ray cluster Abell 2256," *Astron. Astrophys.*, vol. 52, pp. 107–113, Oct. 1976. SP