SIGNAL PROCESSING FOR RADIO ASTRONOMICAL ARRAYS

Alle-Jan van der Veen[1], *Amir Leshem*[2] *and Albert-Jan Boonstra*[3]

[1] Delft Univ. of Technology, Dept. Electrical Eng./DIMES, Mekelweg 4, 2628 CD Delft, The Netherlands
[2] Bar-Ilan University, Dept. Electrical Eng., Tel-Aviv, Israel
[3] ASTRON, P.O Box 2, 7990 AA Dwingeloo, The Netherlands

Radio astronomy forms an interesting application area for array signal processing techniques. Current synthesis imaging telescopes consist of a small number of identical dishes, which track a fixed patch in the sky and produce estimates of the time-varying spatial covariance matrix. The observations are distorted by RFI, e.g., radio, TV, radar and satellite signals. We describe some of the tools that array signal processing offers to filter out the interference, based on eigenvalue decompositions and factor analysis, a more general technique applicable to partially calibrated arrays. We consider spatial filtering techniques using projections, and discuss how a reference antenna pointed at the interferer can improve the performance. We also consider image formation and its relation to beamforming. Finally, we briefly discuss some future radio telescopes, which will consist of distributed phased arrays with 10,000s to 100,000s of elements.

1. INTRODUCTION

The future of radio astronomical discoveries depends on achieving better resolution and sensitivity while maintaining immunity to terrestrial interference which is rapidly growing. The last two demands are obviously contradicting as improved sensitivity implies receiving more interfering signals. RFI detection and removal is now an important topic in radio astronomy. The most promising track here is to switch to massive phased array technology, where we will gain both in terms of resolution and sensitivity while increasing the flexibility to filter out interference. The international efforts in this direction are coordinated under the framework of the Square Kilometer Array programme (SKA). The first example of a phased array radio telescope is LOFAR (13,000 elements) which is currently under construction in The Netherlands.

The principle of interferometry has been used in radio astronomy since 1946 when Ryle and Vonberg constructed a radio interferometer using dipole antenna arrays [9]. In 1962 the principle of aperture synthesis using earth rotation was proposed [10], and applied in the five kilometer Cambridge radio telescope, the 3 km Westerbork Synthesis Radio Telescope (WSRT) in The Netherlands (figure 1) and the well-known 36 km Very Large Array (VLA) in the USA.

In interferometric radio astronomy the signals from various telescopes are usually split into narrow frequency bins (say 50 kHz), and correlated over 1–100 milliseconds to yield short-term correlation matrices. These are then integrated over longer periods of typically 10–60 seconds to yield long-term correlation matrices, which are stored onto tape and constitute the output of the telescope interferometer. The long-term correlation matrices contain contributions from the astronomical sources in the pointing direction through the main lobe of the telescope, from interferers in the near and far field through the side lobes, and from spatially white receiver noise. The astronomical signals usually have a signal-to-noise ratio (SNR) of -20 dB or less, and hence they are too weak to be detected over short integration periods. Harmful interference may range from -70 dB up to $+50$ dB with respect to the instantaneous system noise level.

Figure 1. Westerbork Synthesis Radio Telescope (14 dishes)

In this paper, we present a signal processing data model (section 2) and subsequently give an overview of several problems in radio astronomy where array signal processing can make a contribution, namely calibration using factor analysis (section 3), interference removal using spatial filtering (section 4), and image formation (section 5). We also have a brief look at future radio telescope designs, in particular LOFAR (section 6).

*This work was supported in part by the STW under DEL77-4476, DTC.5893. This overview paper accompanies an invited lecture at IEEE SAM'04 and covers material from several prior publications, including [1–8].

Notation Superscript $^{\mathrm{T}}$ denotes matrix transpose, $^{\mathrm{H}}$ denotes complex conjugate transpose, vec$(\cdot)$ denotes the stacking of the columns of a matrix in a vector, $\otimes$ the Kronecker product. $\mathbf{I}$ is the identity matrix, and $\mathbf{1}$ is a vector with all ones.

2. DATA MODEL

2.1. Received data model

Assume we have a telescope array with p elements. We consider the signals $x_i(t)$ received at the antennas $i = 1, \cdots, p$ in a sufficiently narrow subband. For the interference free case the array output vector $\mathbf{x}(t)$ is modeled in complex baseband form as

$$\mathbf{x}(t) = \mathbf{v}(t) + \mathbf{n}(t)$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_p(t)]^{\mathrm{T}}$ is the $p\times 1$ vector of telescope signals at time t, $\mathbf{v}(t)$ is the received sky signal possibly due to many astronomical sources, assumed on the time scale of 10 s to be a stationary Gaussian vector with covariance matrix $\mathbf{R}_v$ (the astronomical 'visibilities'), and $\mathbf{n}(t)$ is the $p\times 1$ Gaussian noise vector with covariance matrix $\mathbf{R}_n$. Usually independent identically distributed noise is assumed, for which $\mathbf{R}_n = \sigma^2\mathbf{I}$, but this implies accurate calibration as discussed in section 3. The astronomer is interested in the non-redundant off-diagonal entries of $\mathbf{R}_v$.

If an interferer is present the output vector is modeled as

$$\mathbf{x}(t) = \mathbf{v}(t) + \mathbf{a}(t)s(t) + \mathbf{n}(t)$$

where $s(t)$ is the interferer signal with spatial signature vector $\mathbf{a}(t)$ which is assumed stationary only over short time intervals. Without loss of generality, we can absorb the unknown amplitude of $s(t)$ into $\mathbf{a}(t)$ and thus set the power of $s(t)$ to 1. The model is easily extended to multiple narrowband interfering sources, in which case we obtain

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

where $\mathbf{A} : p\times q$ has q columns corresponding to q interferers, and $\mathbf{s}(t)$ is a vector with q entries.

We assume that the processing bandwidth is sufficiently narrow, meaning that the maximal propagation delay of a signal across the telescope array is small compared to the inverse bandwidth, so that this delay can be represented by a phase shift of the signal. If the assumption is not satisfied, as for many existing telescopes, a form of subband processing has to be implemented.

2.2. Covariance model

Suppose that we have obtained observations $\mathbf{x}[m] := \mathbf{x}(mT_s)$, where T_s is the sampling period. We assume that $\mathbf{A}(t)$ is stationary at least over intervals of MT_s, and construct short-term covariance estimates $\hat{\mathbf{R}}_k$,

$$\hat{\mathbf{R}}_k = \frac{1}{M}\sum_{m=kM+1}^{(k+1)M} \mathbf{x}[m]\mathbf{x}[m]^{\mathrm{H}}$$

where M is the number of samples per short-term average. Several filtering algorithms in this paper are based on applying operations to each $\hat{\mathbf{R}}_k$ to remove the interference, followed by further averaging over the resulting matrices to obtain a long-term average.

Considering the $\mathbf{A}_k := \mathbf{A}(kMT_s)$ as deterministic, the expected value of each $\hat{\mathbf{R}}_k$ is denoted by $\mathbf{R}_k$. According to the assumptions, $\mathbf{R}_k$ has model

$$\mathbf{R}_k = \mathbf{\Psi} + \mathbf{A}_k\mathbf{A}_k^{\mathrm{H}} = \mathbf{R}_v + \mathbf{R}_n + \mathbf{A}_k\mathbf{A}_k^{\mathrm{H}} \qquad (1)$$

where $\mathbf{\Psi} = \mathbf{R}_v + \mathbf{R}_n$ is the interference-free covariance matrix.

3. FACTOR ANALYSIS

3.1. The "Factor Analysis" Decomposition

Factor analysis is a statistical technique with origins in psychometrics and biometrics [11, 12]. It assumes a collection of data $\mathbf{X} = [\mathbf{x}(1), \cdots, \mathbf{x}(N)]$ with covariance

$$\mathbf{R} = \mathrm{E}\{\mathbf{x}(k)\mathbf{x}(k)^{\mathrm{H}}\} \;=\; \mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{D}$$

where $\mathbf{R} : p\times p$ Hermitian, $\mathbf{A} : p\times q$ and $\mathbf{D} : p\times p$ diagonal. The objective of factor analysis is, for given $\mathbf{R}$, to identify $\mathbf{A}$ and $\mathbf{D}$, as well as the factor dimension q. We can furthermore set

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}_0\mathbf{U}^{\mathrm{H}} + \mathbf{D} = [\mathbf{U}_s \quad \mathbf{U}_n]\begin{bmatrix}\mathbf{\Lambda}_s & \\ & 0\end{bmatrix}\begin{bmatrix}\mathbf{U}_s^{\mathrm{H}} \\ \mathbf{U}_n^{\mathrm{H}}\end{bmatrix} + \mathbf{D}$$

Thus, the "Factor Analysis" Decomposition (FAD) can be viewed as a generalization of the eigenvalue decomposition. This decomposition is relevant in case the noise covariance is unknown but diagonal, $\mathbf{R}_n = \mathbf{D}$, which corresponds to the noise being uncorrelated among the sensors. In contrast, the usual eigenvalue decomposition for estimating $\mathbf{U}_s$ is only valid if the noise powers are equal among sensors ($\mathbf{R}_n = \sigma^2\mathbf{I}$), which is generally true only after accurate calibration and noise whitening.

It is clear that for an arbitrary Hermitian matrix $\mathbf{R}$, this factorization can exist only for $q = p$, in which case we can set $\mathbf{D} = 0$, or any other value. Hence, for a noise-perturbed matrix, we wish to detect the smallest q which gives a "reasonable fit", and we will assume that q is sufficiently small so that unique decompositions exist. Furthermore, we can not estimate $\mathbf{A}$ uniquely, since $\mathbf{A}$ can be replaced by $\mathbf{A}\mathbf{V}$ for an arbitrary unitary matrix $\mathbf{V}$. Thus, we can only estimate $\mathrm{ran}(\mathbf{A}) = \mathrm{ran}(\mathbf{U}_s)$, as well as the "signal eigenvalues" $\mathbf{\Lambda}_s$. If the eigenvalues are not repeated and we sort them in descending order, then $\mathbf{U}_s$ and $\mathbf{\Lambda}_s$ can be uniquely determined. There are other ways to constrain $\mathbf{A}$ to be a unique factor, e.g. by taking it to be a lower-triangular rectangular Cholesky factor with positive real diagonal entries.

We consider the following two problems:

- Detection: given $\mathbf{X}$, estimate q.
- Identification: given $\mathbf{X}$ and q, estimate $\mathbf{D}$, $\mathbf{\Lambda}_s$ and $\mathbf{U}_s$.

3.2. Detection

The detection problem is given by a collection of hypotheses

$$\begin{array}{l}\mathcal{H}_q : \mathbf{x}(k) \;\sim\; \mathcal{CN}(0, \mathbf{R}_q) \\ \mathcal{H}' : \mathbf{x}(k) \;\sim\; \mathcal{CN}(0, \mathbf{R}'), \qquad q = 1, 2, \cdots\end{array} \qquad (2)$$

where $\mathcal{CN}(0, \mathbf{R})$ denotes the zero-mean complex normal distribution with covariance $\mathbf{R}$, $\mathbf{R}_q$ is the covariance matrix of the model with q interferers,

$$\mathbf{R}_q = \mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{D}, \quad \text{where} \quad \mathbf{A}:\ p\times q, \quad \mathbf{D}\ \text{diagonal}$$

and $\mathcal{H}'$ corresponds to a default hypothesis of an arbitrary (unstructured) positive definite matrix $\mathbf{R}'$.

The Generalized Likelihood Ratio Test (GLRT) detector for this problem tests $\mathcal{H}_q$ versus $\mathcal{H}'$, where the unknown parameters are replaced by maximum likelihood estimates under each of the hypotheses. Under $\mathcal{H}_q$, the likelihood function is given by

$$L(\mathbf{X}|\mathcal{H}_q) \;\equiv\; L(\mathbf{X}|\mathbf{R}_q) \;=\; \left(\frac{1}{|\mathbf{R}_q|}e^{-\mathrm{tr}\left(\mathbf{R}_q^{-1}\hat{\mathbf{R}}\right)}\right)^N,$$

where $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)]$ is the data and $\hat{\mathbf{R}} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}(k)\mathbf{x}(k)^H$ is the sample covariance matrix, $|\cdot|$ denotes the determinant and $\mathrm{tr}(\cdot)$ the trace operator.

The ML estimate of $\mathbf{R}_q$ is found by maximizing $L(\mathbf{X}|\mathbf{R}_q)$ over the parameters of the model $\mathbf{R}_q = \mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{D}$, or equivalently the log-likelihood function

$$\mathcal{L}(\mathbf{X}|\mathbf{R}_q) \;=\; N\left(-\ln|\mathbf{R}_q| - \mathrm{tr}(\mathbf{R}_q^{-1}\hat{\mathbf{R}})\right).$$

Denote the estimate by $\hat{\mathbf{R}}_q = \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathrm{H}} + \hat{\mathbf{D}}$. Under $\mathcal{H}'$ we obtain that the ML estimate of $\mathbf{R}'$ is given by $\hat{\mathbf{R}}$, the sample covariance matrix. The log-likelihood GLRT test statistic is thus given by

$$\ln\frac{L(\mathbf{X}|\mathcal{H}_q)}{L(\mathbf{X}|\mathcal{H}')} \;=\; -N\big(\,\mathrm{tr}(\hat{\mathbf{R}}_q^{-1}\hat{\mathbf{R}}) - \ln|\hat{\mathbf{R}}_q^{-1}\hat{\mathbf{R}}| - p\,\big).$$

A further result is that the ML estimate of $\hat{\mathbf{R}}_q$ is such that $\mathrm{tr}(\hat{\mathbf{R}}_q^{-1}\hat{\mathbf{R}}) = p$ so that we can base the test on

$$T_q(\mathbf{X}) \;:=\; N\ln|\hat{\mathbf{R}}_q^{-1}\hat{\mathbf{R}}|. \tag{3}$$

A threshold for the test can be found if we generalize the results in [11, 12] to complex data:

Lemma 3.1 *If $\mathcal{H}_q$ is true and N is moderately large (say $N-q \geq 50$), then $2T_q(\mathbf{X})$ has approximately a χ^2_ν distribution with $\nu = (p-q)^2 - p$ degrees of freedom.*

This provides a threshold for a test of $\mathcal{H}_q$ versus $\mathcal{H}'$ corresponding to a desired probability of "false alarm" (here the probability of rejecting $\mathcal{H}_q$ when it is true). The test replaces the more familiar eigenvalue test on the rank of $\hat{\mathbf{R}}$ in the case of white noise, $\mathbf{D} = \sigma^2\mathbf{I}$.

3.3. Estimation of the FAD

Assume that the factor rank q is known. Given $\hat{\mathbf{R}} = \frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{H}}$, and a sufficiently small q, we wish to estimate $\hat{\mathbf{A}}$ and $\hat{\mathbf{D}}$ such that $\hat{\mathbf{R}} \approx \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathrm{H}} + \hat{\mathbf{D}}$. There are several approaches for this.

An ML estimate of the factors $\mathbf{A} : p\times q$ and $\mathbf{D}$ is dependent on the choice of q. The largest permissible value of q is that for which the number of degrees of freedom $\nu = (p-q)^2 - p \geq 0$, or $q \leq p - \sqrt{p}$. For larger q, there is no identifiability of $\mathbf{A}$ and $\mathbf{D}$: any sample covariance matrix $\hat{\mathbf{R}}$ can be fitted. Even for smaller q, $\mathbf{A}$ can be identified only up to a $q\times q$ unitary transformation at the right, i.e., we can identify $\mathrm{span}(\mathbf{A})$. Luckily, this is sufficient for many applications.

For $q > 0$, there is no closed form solution to the estimation of the factors $\mathbf{A}$ and $\mathbf{D}$ in the ML estimation of $\hat{\mathbf{R}}_q = \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathrm{H}} + \hat{\mathbf{D}}$. There are several approaches for obtaining an estimate:

- Suppose that the optimal ML-estimate $\hat{\mathbf{D}}$ has been found. We can then whiten $\hat{\mathbf{R}}$ to $\tilde{\mathbf{R}} = \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{R}}\hat{\mathbf{D}}^{-1/2}$, and similarly the model, giving $\tilde{\mathbf{R}}_q = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^{\mathrm{H}} + \mathbf{I}$. Note that $|\mathbf{R}_q^{-1}\hat{\mathbf{R}}| = |\tilde{\mathbf{R}}_q^{-1}\tilde{\mathbf{R}}|$, which is the usual problem for white noise, solved via an eigenvalue decomposition of $\tilde{\mathbf{R}}$. This is equivalent to solving $\min\|\tilde{\mathbf{R}} - (\tilde{\mathbf{A}}\tilde{\mathbf{A}}^{\mathrm{H}} + \mathbf{I})\|_F^2$. Since $\hat{\mathbf{D}}$ is not known, this leads to an iteration where $\tilde{\mathbf{A}}$ is plugged back, $\hat{\mathbf{D}}$ is estimated, etc.

  A related technique is alternating least squares, where we alternatingly minimize $\|\hat{\mathbf{R}} - (\mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{D})\|_F^2$ over $\mathbf{A}$ keeping $\mathbf{D}$ fixed, and over $\mathbf{D}$ keeping $\mathbf{A}$ fixed. Both iterative techniques tend to converge very slowly but may be used for fine-tuning.

- Gauss-Newton iterations on the original (determinant) cost function, or on the (weighted) least squares cost. This requires an accurate starting point.

- Ad-hoc techniques for solving the least squares problem, possibly followed by a Gauss-Newton iteration. These techniques try to modify the diagonal of $\hat{\mathbf{R}}$ such that the modified matrix is low-rank q, hence can be factored as $\mathbf{A}\mathbf{A}^{\mathrm{H}}$. For this we can exploit the fact that submatrices away from the main diagonal with at least $q+1$ columns and rows have rank q, hence one column can be written as a linear combination of q others. The same linear combination can be used to estimate the main diagonal of $\mathbf{A}\mathbf{A}^{\mathrm{H}}$. See figure 2. This works for $q < \frac{1}{2}p$. The case $q = 1$ was studied in more detail in [6].

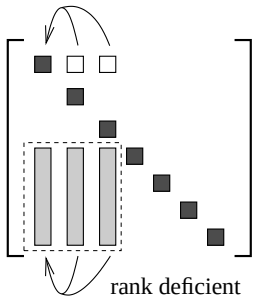

Figure 2. Column ratio factor estimation

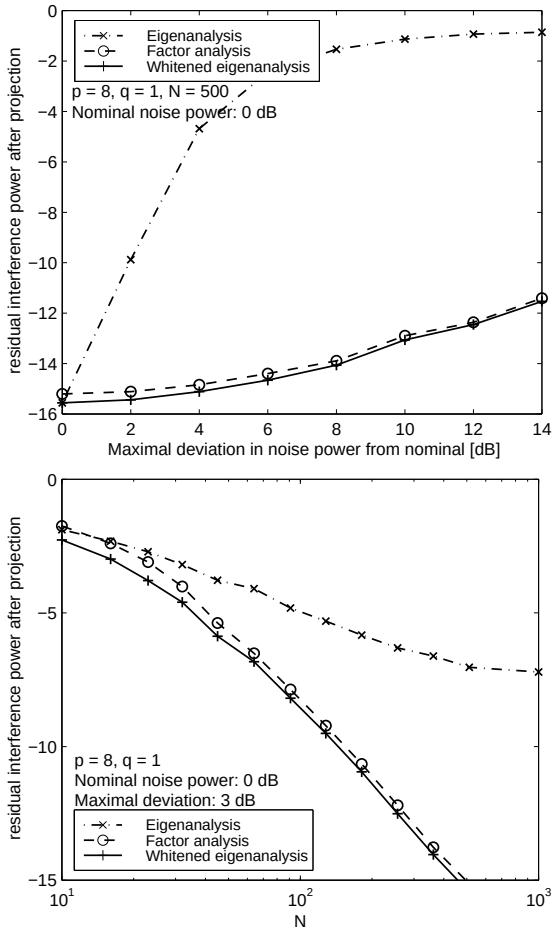

Figure 3. Residual interference power after projections. $\mathbf{U}_A$ estimated from (i) eigenvalue decomposition, (ii) Factor Analysis, and (iii) eigendecomposition after whitening by $\mathbf{D}^{-1/2}$, assuming true $\mathbf{D}$ is known.

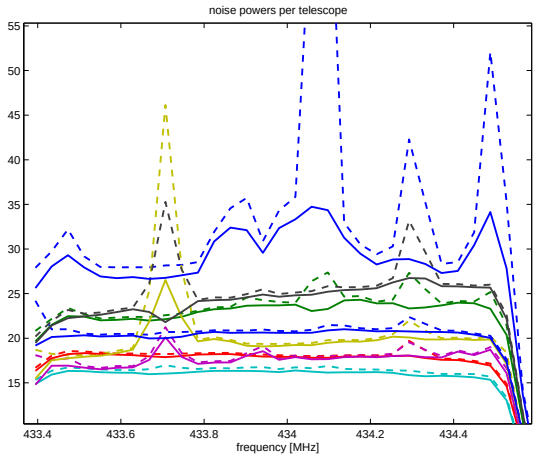

Figure 4. Amateur broadcast interference, both continuous and intermittent, recorded at the WSRT: estimated noise powers (dashed: averaged diagonal entries of the $\hat{\mathbf{R}}_k$; solid: filtered using factor analysis).

3.4. Application to radio astronomy

We mention two applications of the factor analysis decomposition.

1. *Signal subspace estimation in the presence of uncorrelated noise [4]*, e.g., for the purpose of spatial filtering of interference. An example is shown in figure 3. Here, the data model is $\mathbf{R} = \mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{D} + \mathbf{R}_\nu$, where $\mathbf{A}$ corresponds to the interfering signals, $\mathbf{D}$ is the diagonal noise covariance matrix, and $\mathbf{R}_\nu \ll \mathbf{D}$ is the sky covariance. Using factor analysis, the number of interferers q is detected, and a basis $\mathbf{U}_A \sim \mathrm{ran}(\mathbf{A})$ is estimated, subsequently a projection $\mathbf{P}_A^\perp = \mathbf{U}_A^\perp \mathbf{U}_A^{\perp H}$ is applied to $\mathbf{R}$ to cancel the interference:
$$\tilde{\mathbf{R}} = \mathbf{P}_A^\perp \hat{\mathbf{R}} \mathbf{P}_A^\perp$$
The figure shows $\|\tilde{\mathbf{R}} - \mathbf{P}_A^\perp(\mathbf{D}+\mathbf{R}_\nu)\mathbf{P}_A^\perp\|_F$. Clearly, the solution using eigenvalue decompositions is not suitable if the noise covariance is not a multiple of the identity matrix.
2. *Gain calibration [6]*. Initially the antenna gains and noise powers of the telescopes are unknown. To estimate them, a common procedure is to point the telescopes at a strong sky source and make an observation. This produces a rank-1 factor model $\mathbf{R} = \mathbf{g}\sigma_s^2\mathbf{g}^{\mathrm{H}} + \mathbf{D}$, where σ_s^2 is the source power (assumed to be known from tables), $\mathbf{g}$ is the antenna gain vector, and $\mathbf{D}$ is a diagonal matrix containing the noise powers of each antenna. These can be estimated using rank-1 factor analysis.

As an example for the noise power estimation, we consider an uncalibrated data set which includes time continuous and intermittent interference observed at the WSRT. The data set is a $p = 8$-channel recording of a 1.25 MHz-wide band at 434 MHz containing signals from the astronomical source 3C48 (white noise signal) contaminated by narrow-band amateur radio broadcasts. The data was partitioned into 32 frequency bins (each processed separately), the short term averaging period was 10 ms ($M = 781$ samples), and the number of time intervals was $N = 1000$.

For each short-term correlation matrix, the rank q_k and the factors are determined using factor analysis. This gives a sequence of $\mathbf{D}_k(f)$-matrices. These are subsequently averaged (median filtering) to obtain the noise power estimate at a single frequency. The process is repeated for all 32 frequency bins. The results are shown in figure 4: the dashed lines are the unfiltered averages of the original diagonals of $\mathbf{R}_k(f)$, the solid lines are the filtered averages. Although the factor analysis did not remove all interference (in particular interference entering on only a single telescope), the graphs give sufficient information to derive smooth noise calibration functions. The data set is subsequently whitened using these calibration parameters.

4. SPATIAL FILTERING

Interference cancellation is becoming increasingly important in radio astronomy. Depending on the interference and the type of instrument, several kinds of RFI mitigation techniques are applicable [2, 13]. For intermittent interference, the most effective techniques are based on detection and "blanking": omitting the contaminated samples from the covariance estimate, using a single sensor [14, 15] or multiple sensors [2]. For continually present interference and an array of p telescope dishes, spatial filtering is possible. The desired instrument outputs in this case are $p \times p$ correlation matrices, integrated to several seconds (eg, 10 s). Based on short-term correlation matrices (integration to e.g., 10 ms) and narrow subband processing, the array signature vector of an interferer can be estimated and subsequently projected out [5]—we describe this technique below.

An interesting option is to utilize a reference antenna which picks up only the interference, so that adaptive cancellation techniques can be implemented [16, 17]. Spatial filtering on extended arrays was first considered by Briggs et al. [18] for a single dual-polarized telescope (two channels) and two reference antennas. Jeffs et al. [19, 20] propose spatial filtering algorithms along the lines of [5].

4.1. Spatial filtering using projections

Suppose that an orthogonal basis $\mathbf{U}_k$ of the subspace spanned by interferer spatial signatures $\mathrm{span}(\mathbf{A}_k)$ is known. We can then form a spatial projection matrix $\mathbf{P}_k := \mathbf{I} - \mathbf{U}_k\mathbf{U}_k^{\mathrm{H}}$ which is such that $\mathbf{P}_k\mathbf{A}_k = 0$. When this spatial filter is applied to the data covariance matrix all the energy due to the interferer will be nulled: let
$$\hat{\mathbf{Q}}_k := \mathbf{P}_k\hat{\mathbf{R}}_k\mathbf{P}_k$$
then
$$\mathrm{E}\{\hat{\mathbf{Q}}_k\} = \mathbf{P}_k\mathbf{\Psi}\mathbf{P}_k$$
where $\mathbf{\Psi} := \mathbf{R}_\nu + \sigma^2\mathbf{I}$ is the interference-free covariance matrix. When we subsequently average the modified covariance matrices $\hat{\mathbf{Q}}_k$, we obtain a long-term estimate
$$\hat{\mathbf{Q}} := \frac{1}{N}\sum_{k=1}^{N}\hat{\mathbf{Q}}_k = \frac{1}{N}\sum_{k=1}^{N}\mathbf{P}_k\hat{\mathbf{R}}_k\mathbf{P}_k\,. \tag{4}$$
$\hat{\mathbf{Q}}$ is an estimate of $\mathbf{\Psi}$, but it is biased due to the projection. To correct for this we first write the two-sided multiplication as a single-sided multiplication employing the matrix identity $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$. This gives
$$\mathrm{vec}(\hat{\mathbf{Q}}) = \frac{1}{N}\sum_{k=1}^{N}\mathbf{C}_k\mathrm{vec}(\hat{\mathbf{R}}_k) \quad \text{where } \mathbf{C}_k := \mathbf{P}_k^{\mathrm{T}} \otimes \mathbf{P}_k\,. \tag{5}$$
If the interference was completely removed then
$$\mathrm{E}\{\mathrm{vec}(\hat{\mathbf{Q}})\} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{C}_k\mathrm{vec}(\mathbf{\Psi}) = \mathbf{C}\mathrm{vec}(\mathbf{\Psi})\,; \quad \mathbf{C} := \frac{1}{N}\sum_{k=1}^{N}\mathbf{C}_k\,. \tag{6}$$

In view of this, we can apply a correction $\mathbf{C}^{-1}$ to $\hat{\mathbf{Q}}$ and define

$$\hat{\mathbf{\Psi}} := \mathrm{vec}^{-1}(\mathbf{C}^{-1}\mathrm{vec}(\hat{\mathbf{Q}})).$$

If the interference was completely projected out then $\hat{\mathbf{\Psi}}$ is an unbiased estimate of the covariance matrix without interference. A detailed analysis of this algorithm will appear in [8]. The main conclusion is that the variance of the estimate of $\mathbf{\Psi}$ is equal to $(1/N)\mathbf{C}^{-1}\sigma^4$, whereas for "clean" data it would be $(1/N)\sigma^4$. For interferers which are sufficiently moving, $\mathbf{C}^{-1}$ is well conditioned and the penalty is comparable to a loss in number of samples. Even for stationary interferers, $\mathbf{C}^{-1}$ might be well conditioned due to the motion of the telescopes, but it depends on the integration length and the location of the sky source which is being tracked. Cases where an interferer enters only on a single telescope always lead to a singular $\mathbf{C}$ and cannot be resolved by this algorithm.

4.2. Spatial filtering with an extended array

If the telescope array is extended with one or more reference antennas, we can follow the same procedure. Let p_0 be the number of primary antennas, and p be the total number of antennas. The data covariance matrix can be partitioned accordingly as

$$\mathbf{R}_k = \begin{bmatrix} \mathbf{R}_{00,k} & \mathbf{R}_{01,k} \\ \mathbf{R}_{10,k} & \mathbf{R}_{11,k} \end{bmatrix}.$$

where $\mathbf{R}_k$ has model

$$\begin{aligned} \mathbf{R}_k &= \mathbf{\Psi} + \mathbf{A}_k\mathbf{A}_k^{\mathrm{H}} = \mathbf{R}_v + \mathbf{\Sigma} + \mathbf{A}_k\mathbf{A}_k^{\mathrm{H}} \\ &= \left[\begin{array}{c|c} \mathbf{R}_{v,0} + \mathbf{A}_{0,k}\mathbf{A}_{0,k}^{\mathrm{H}} + \sigma_0^2\mathbf{I} & \mathbf{A}_{0,k}\mathbf{A}_{1,k}^{\mathrm{H}} \\ \hline \mathbf{A}_{1,k}\mathbf{A}_{0,k}^{\mathrm{H}} & \mathbf{A}_{1,k}\mathbf{A}_{1,k}^{\mathrm{H}} + \sigma_1^2\mathbf{I} \end{array}\right] \end{aligned} \quad (7)$$

$\mathbf{\Psi}$ is the interference-free covariance matrix, and $\mathbf{\Sigma} := \mathrm{diag}[\sigma_0^2\mathbf{I}, \sigma_1^2\mathbf{I}]$ is the diagonal noise covariance matrix (assumed known). The objective is to estimate the interference-free covariance submatrix $\mathbf{\Psi}_{00} := \mathbf{R}_{v,0} + \sigma_0^2\mathbf{I}$.

Following the preceding algorithm applied to $\mathbf{R}_k$, the reconstructed covariance matrix is size $p \times p$, and we can simply select the $p_0 \times p_0$ submatrix in the top-left corner, $\hat{\mathbf{\Psi}}_{00}$ [19, 20]. An improved algorithm would not reconstruct the other blocks of $\hat{\mathbf{\Psi}}$ [7]. Indeed, let the projected estimates $\hat{\mathbf{Q}}$ be as before in (4). Then (6) applies:

$$\mathrm{E}\{\mathrm{vec}(\hat{\mathbf{Q}})\} = \mathbf{C}\mathrm{vec}(\mathbf{\Psi}).$$

Partition $\mathbf{\Psi}$ as in (7) into 4 submatrices. Since we are only interested in recovering $\mathbf{\Psi}_{00}$, the other submatrices in $\hat{\mathbf{\Psi}}$ are replaced by their expected values, respectively $\mathbf{\Psi}_{01} = \mathbf{0}$, $\mathbf{\Psi}_{10} = \mathbf{0}$, $\mathbf{\Psi}_{11} = \sigma_1^2\mathbf{I}$. This corresponds to solving the reduced-size covariance model error minimization problem,

$$\hat{\mathbf{\Psi}}_{00} = \underset{\mathbf{\Psi}_{00}}{\arg\min} \; \| \mathrm{vec}(\hat{\mathbf{Q}}) - \mathbf{C}\mathrm{vec}\Big(\left[\begin{array}{c|c} \mathbf{\Psi}_{00} & \mathbf{0} \\ \hline \mathbf{0} & \sigma_1^2\mathbf{I} \end{array}\right] \Big) \|^2 .$$

The solution of this problem reduces to a standard LS problem after separating the knowns from the unknowns. Thus, rearrange the entries of $\mathrm{vec}(\mathbf{\Psi})$ into

$$\begin{bmatrix} \mathrm{vec}(\mathbf{\Psi}_{00}) \\ \hline \sigma_1^2\mathbf{1} \\ \hline \mathbf{0} \end{bmatrix}$$

where $\mathbf{1}$ indicates a vector with all entries equal to 1, and repartition $\mathbf{C}$ accordingly, to obtain the equivalent problem

$$\begin{aligned} \mathrm{vec}(\hat{\mathbf{\Psi}}_{00}) &= \underset{\mathbf{\Psi}_{00}}{\arg\min} \; \| \mathrm{vec}(\hat{\mathbf{Q}}) - [\mathbf{C}_1 \;\; \mathbf{C}_2 \;\; \mathbf{C}_3] \begin{bmatrix} \mathrm{vec}(\mathbf{\Psi}_{00}) \\ \hline \sigma_1^2\mathbf{1} \\ \hline \mathbf{0} \end{bmatrix} \|^2 \\ &= \underset{\mathbf{\Psi}_{00}}{\arg\min} \; \| (\mathrm{vec}(\hat{\mathbf{Q}}) - \sigma_1^2\mathbf{C}_2\mathbf{1}) - \mathbf{C}_1\mathrm{vec}(\mathbf{\Psi}_{00}) \|^2 \\ &= \mathbf{C}_1^{\dagger}(\mathrm{vec}(\hat{\mathbf{Q}}) - \sigma_1^2\mathbf{C}_2\mathbf{1}). \end{aligned}$$

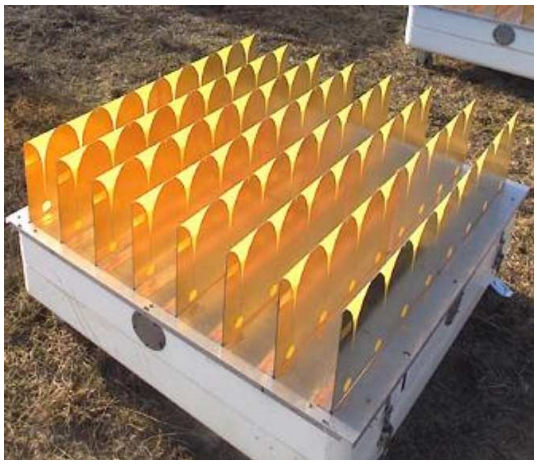

Figure 5. The "THEA platform" reference phased array (ASTRON)

(If σ_1^2 is unknown, then it can be estimated using a straightforward modification.) The advantage compared to the preceding algorithm is that $\mathbf{C}_1$ is a tall matrix, and better conditioned than $\mathbf{C}$. This improves the performance of the algorithm in cases where $\mathbf{C}$ is ill-conditioned, e.g., for stationary interferers, or an interferer entering on only a single telescope. Asymptotically for large INR of the reference array, the algorithm is seen to behave similar to the traditional subtraction technique.

4.3. Experiment

A reference signal is useful only if it has a better SNR than the primary antennas. Therefore, an omnidirectional antenna is not good enough. To be versatile, we have tested the preceding technique on a reference signal obtained from the beamformed output of a wideband phased array of 64 elements, shown in figure 5. This system has a bandwidth of 600-1700 MHz, a baseband of 20 MHz, two digital beamforming outputs, and it is part of an envisioned "Thousand Elements Array" (THEA), developed by ASTRON. The reference signal is correlated along with the telescope signals as if it was an additional telescope, and spatial filtering algorithms can be applied to the resulting short-term integrated covariance matrices.

The test data is an observation of the strong astronomical source 3C48 contaminated by Afristar satellite signals. The primary array consists of $p_0 = 6$ of the 14 telescope dishes of the WSRT. The reference signals are $p_1 = 2$ beamforming outputs of the THEA system. One beam was pointed approximately to the satellite, the other was used for scanning. We recorded 65 kSamples at 20 MS/s, and processed these offline. After short-term windowed Fourier transforms, the data was split into 64 frequency bins, correlated, and averaged over 32 samples to obtain 16 short-term covariance matrices.

The resulting auto- and crosscorrelation spectra after filtering are shown in figure 6. The autocorrelation spectra are almost flat, and close to 1 (the whitened noise power). The cross-correlation spectra show that the spatial filtering with reference antenna has done much better to remove the interference than the case without reference antenna. The residual correlation of about 4% is known to be the SNR of the astronomical source. The lines are noisy due to the finite sample effect; the predicted standard deviation (based on number of samples averaged) are indicated for a few frequencies.

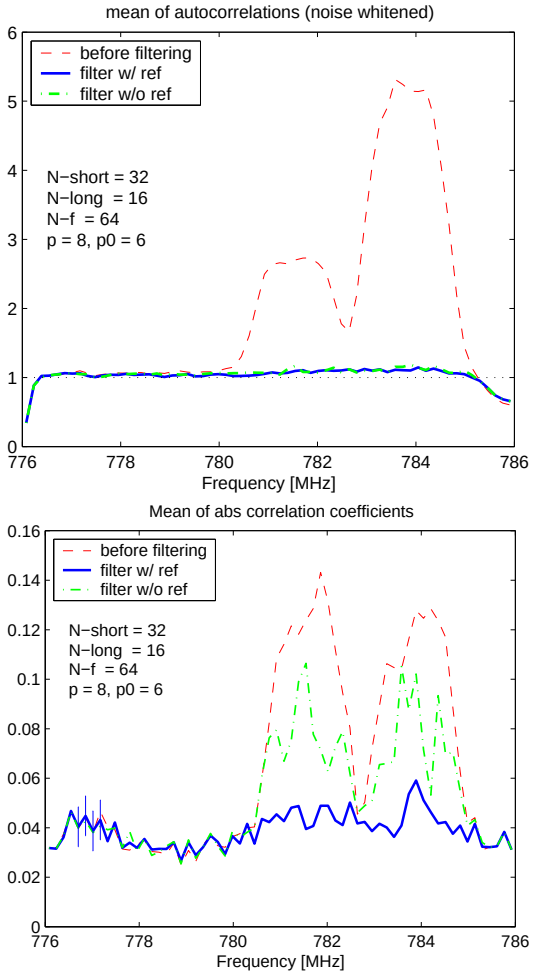

Figure 6. (a) Averaged autocorrelation spectrum before and after filtering, (b) Averaged cross-correlation spectrum

5. IMAGING

As described in more detail in [1], image formation is also a fruitful area for array signal processing techniques. Astronomers try to estimate the intensity (brightness) $I_f(\mathbf{s})$ of the sky as a function of the location $\mathbf{s}$ and frequency f. They do this by measuring the correlation (called the "visibility" V_f) between identical sensors i and j with locations $\mathbf{r}_i$ and $\mathbf{r}_j$, corresponding to a baseline $\mathbf{r}_i - \mathbf{r}_j$. Let (ℓ, m) denote normalized coordinates of the sky source $(-1 \le \ell, m \le 1)$, and (u, v, w) the baseline vector of the antenna pair measured in wavelengths. Assuming a planar array, w can be removed from the equations via geometrical delay compensation. Under certain approximations, the "measurement equation" is given by [21]

$$V_f(u,v) \;=\; \iint I_f(\ell,m)\, e^{-j2\pi(u\ell+vm)}\, d\ell dm\,. \qquad (8)$$

It has the form of a Fourier transformation.

The function $V_f(u,v)$ is sampled at various coordinates $(u_{ij}(t), v_{ij}(t))$ by first of all taking all possible sensor pairs i, j or baselines $\mathbf{r}_i - \mathbf{r}_j$, and second by realizing that the sensor locations $\mathbf{r}_i, \mathbf{r}_j$ are actually time-varying since the earth rotates. Given a sufficient number of samples in the (u,v) domain, the relation can be inverted to obtain an image (the 'map').

5.1. Matrix formulation

Assume that the sky consists of a large number (d) point sources. Equation (8) can then be written slightly differently as

$$\begin{aligned} &V(u_{ij}(t), v_{ij}(t)) = \\ &= \sum_{l=1}^{d} e^{-j2\pi(u_{i0}(t)\ell_l + v_{i0}(t)m_l)} \cdot I(\ell_l, m_l) \cdot e^{j2\pi(u_{j0}(t)\ell_l + v_{j0}(t)m_l)}\,. \end{aligned}$$

where (u_{i0}, v_{i0}) are coordinates of the ith antenna with respect to a common reference point. The connection to our previous framework is obtained by collecting the visibilities into correlation matrices $\mathbf{R}$, where $\mathbf{R}_{ij}(t) = V(u_{ij}(t), v_{ij}(t))$. The above equation can then be written as

$$\mathbf{R}_k \;=\; \mathbf{A}_k \mathbf{B} \mathbf{A}_k^{\mathrm{H}}$$

where $\mathbf{R}_k \equiv \mathbf{R}(t_k)$, $\mathbf{A}_k = [\mathbf{a}_k(\ell_1, m_1), \ldots, \mathbf{a}_k(\ell_d, m_d)]$, and

$$\mathbf{a}_k(\ell,m) = \begin{bmatrix} e^{-j2\pi(u_{10}(t_k)\ell + v_{10}(t_k)m)} \\ \vdots \\ e^{-j2\pi(u_{p0}(t_k)\ell + v_{p0}(t_k)m)} \end{bmatrix} \qquad (9)$$

$$\mathbf{B} = \begin{bmatrix} I(\ell_1,m_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & I(\ell_d,m_d) \end{bmatrix}$$

where $\mathbf{a}_k(\ell, m)$ is recognized as the array response vector. As usual, the array response is frequency dependent. The response is also slowly time-varying due to the earth rotation.

5.2. Classical inverse Fourier imaging

The relation between sky brightness $I(\ell, m)$ and visibilities $V(u,v)$ (where u, v are taken at frequency f) is given by the measurement equation (8). We have measured V on a discrete set of baselines $\{(u_i, v_i)\}$. The "dirty image" (a lumpy image obtained via direct Fourier inversion possibly modified with some weights c_i) is defined by

$$I_D(\ell,m) \;:=\; \sum_i c_i\, V(u_i,v_i)\, e^{j2\pi(u_i\ell + v_i m)} \qquad (10)$$

It is equal to the 2D convolution of the true image I with a point spread function known as the "dirty beam":

$$\begin{aligned} &I_D(\ell,m) = \\ &= \textstyle\sum_i c_i\, V(u_i,v_i)\, e^{j2\pi(u_i\ell + v_i m)} \\ &= \textstyle\sum_i c_i \left[\iint I(\ell',m')\, e^{-j2\pi(u_i\ell' + v_i m')}\, d\ell'\, dm'\right] e^{j2\pi(u_i\ell + v_i m)} \\ &= \iint I(\ell',m') \left[\textstyle\sum_i c_i\, e^{j2\pi(u_i(\ell-\ell') + v_i(m-m'))}\right] d\ell'\, dm' \\ &= \iint I(\ell',m')\, B_0(\ell-\ell', m-m')\, d\ell'\, dm' \end{aligned}$$

or

$$I_D \;=\; I * B_0\,, \qquad B_0(\ell,m) \;:=\; \sum_i c_i\, e^{j2\pi(u_i\ell + v_i m)}$$

B_0 is the dirty beam, centered at the origin. The weights $\{c_i\}$ are arbitrary coefficients designed to obtain an acceptable beam-shape, with low side lobes, in spite of the irregular sampling.

Specializing to a point source model, $I(\ell,m) = \sum_l I_l\, \delta(\ell - \ell_l, m - m_l)$ where I_l is the intensity of the source at location (ℓ_l, m_l), gives

$$V(u,v) \;=\; \sum_l I_l\, e^{-j2\pi(u\ell_l + v m_l)}$$

$$I_D(\ell,m) \;=\; \sum_l I_l\, B_0(\ell-\ell_l, m-m_l)$$

Thus, every point source excites the dirty beam centered at its location (ℓ_l, m_l).

From the dirty image I_D and the known dirty beam B_0, the desired image I is obtained via a deconvolution process. A popular method for doing this is the CLEAN algorithm [22]. The algorithm assumes that B_0 has its peak at the origin, and consists of a loop in which a candidate location (ℓ_l, m_l) is selected as the largest peak in I_D, and subsequently a small multiple of $B_0(\ell-\ell_l, m-m_l)$ is subtracted from I_D. The objective is to minimize the residual, until it converges to the noise level.

5.3. Inverse Fourier imaging after projections

If we take projections or any other linear combination $[c_{ij}]$ of the visibilities $\{V(u_i, v_i)\}$ during measurements we have instead available

$$Z(u_i, v_i) \;=\; \sum_j c_{ij}\, V(u_j, v_j)$$

E.g., in section 4 we took projections of the short-term covariance matrices and averaged the results. In that case, the coefficients c_{ij} correspond to $\mathbf{C}$ in equation (6), and $Z(u_i, v_i)$ corresponds to $\hat{\mathbf{Q}}$ in equation (5). This further assumes that we did not make the correction by $\mathbf{C}^{-1}$ as suggested in that section, perhaps because it is ill conditioned.

Suppose we compute the dirty image in the same way as before, but now from Z,

$$\begin{aligned} I_D(\ell, m) &:= \textstyle\sum_i Z(u_i, v_i)\; e^{j2\pi(u_i\ell + v_i m)} \\ &= \textstyle\sum_i \sum_j c_{ij}\, V(u_j, v_j)\; e^{j2\pi(u_i\ell + v_i m)}\,. \end{aligned}$$

Then

$$\begin{aligned} I_D(\ell, m) &= \textstyle\sum_i \sum_j c_{ij}\, V(u_j, v_j)\; e^{j2\pi(u_i\ell + v_i m)} \\ &= \textstyle\sum_i \sum_j c_{ij} \left[\iint I(\ell', m')\; e^{-j2\pi(u_j\ell' + v_j m')}\; d\ell'\, dm' \right] e^{j(u_i\ell + v_i m)} \\ &= \iint I(\ell', m') \left[\textstyle\sum_i \sum_j c_{ij}\; e^{-j2\pi(u_j\ell' + v_j m')}\; e^{j2\pi(u_i\ell + v_i m)} \right] d\ell'\, dm' \\ &= \iint I(\ell', m')\, B(\ell, m, \ell', m')\; d\ell'\, dm' \end{aligned}$$

where

$$B(\ell, m, \ell', m') \;:=\; \sum_i \sum_j c_{ij}\; e^{-j2\pi(u_j\ell' + v_j m')}\; e^{j2\pi(u_i\ell + v_i m)}\,.$$

Thus, the dirty image is again obtained via a convolution, but the dirty beam is now space-varying. $B(\ell, m, \ell', m')$ is a beam centered at (ℓ', m') and measured at (ℓ, m).

With a point source model,

$$I_D(\ell, m) \;=\; \sum_l I_l\, B(\ell, m, \ell_l, m_l) \;=\; \sum_l I_l\, B_l(\ell, m)$$

where

$$B_l(\ell, m) \;:=\; \sum_i \sum_j c_{ij}\; e^{-j2\pi(u_j\ell_l + v_j m_l)}\; e^{j2\pi(u_i\ell + v_i m)}\,.$$

Again, every point source excites a beam centered at its location (ℓ_l, m_l), but the beams may all be different: they are *space varying*. Nonetheless, they are completely known if we know the linear combinations that we took during observations. Thus, the CLEAN algorithm can readily be modified to take the varying beam shapes into account: simply replace $B_0(\ell, m)$ by $B_l(\ell, m)$ everywhere in the algorithm.

5.4. Imaging via beamforming techniques

CLEAN and sequential beamforming Using a parametric point-source model, the image deconvolution problem can be interpreted as a direction-of-arrival (DOA) estimation problem, e.g., as

$$[\{\hat{\mathbf{s}}_l\}, \hat{\mathbf{B}}] = \arg\min_{\{\mathbf{s}_l\}, \mathbf{B}} \sum_{k=1}^{K} \| \hat{\mathbf{R}}_k - \mathbf{A}_k(\{\mathbf{s}_l\})\, \mathbf{B}\, \mathbf{A}_k^{\mathrm{H}}(\{\mathbf{s}_l\}) - \sigma^2 \mathbf{I} \|_F \quad (11)$$

($\mathbf{B}$ is constrained to be diagonal with positive entries.) This is recognized as the same model as used for DOA estimation in array processing. Note however that the array is moving ($\mathbf{A}_k$ is time-dependent), and that there are many more sources than the dimension of each covariance matrix.

In this notation, the image formation in section 5.2 can be formulated as follows. If we write $I_D(\mathbf{s}) \equiv I_D(\ell, m)$ and $\mathbf{a}_k(\mathbf{s}) \equiv \mathbf{a}_k(\ell, m)$, we can rewrite the dirty image (10) as

$$\begin{aligned} I_D(\mathbf{s}) &= \sum_{i,j,k} V(u_{ij}(t_k), v_{ij}(t_k))\, e^{j2\pi(u_{i0}(t_k)\ell + v_{i0}(t_k)m)}\,. \\ &\qquad \cdot e^{-j2\pi(u_{j0}(t_k)\ell + v_{j0}(t_k)m)} \\ &= \sum_{i,j,k} (\mathbf{R}_k)_{ij} (\bar{\mathbf{a}}_k(\mathbf{s}))_i (\mathbf{a}_k(\mathbf{s}))_j \\ &= \sum_k \mathbf{a}_k^{\mathrm{H}}(\mathbf{s}) \mathbf{R}_k \mathbf{a}_k(\mathbf{s})\,. \end{aligned}$$

(We omitted the optional weighting. Also note that, with noise, we have to replace $\mathbf{R}_k$ by $\mathbf{R}_k - \sigma^2\mathbf{I}$.) The iterative beam removing in CLEAN can now be posed as an iterative LS fitting between the sky model and the observed visibility [23]. Finding the brightest point $\mathbf{s}_0$ in the image is equivalent to trying to find a point source using classical Fourier beamforming, i.e., ,

$$\hat{\mathbf{s}}_0 = \arg\max_{\mathbf{s}} \sum_{k=1}^{K} \mathbf{a}_k^{\mathrm{H}}(\mathbf{s}) \left(\mathbf{R}_k - \sigma^2 \mathbf{I} \right) \mathbf{a}_k(\mathbf{s})\,. \quad (12)$$

Thus, the CLEAN algorithm can be regarded as a generalized classical sequential beamformer, where the brightest points are found one by one, and subsequently removed from $\mathbf{R}_k$ until the LS cost function (11) is minimized. An immediate consequence is that the estimated source locations will be biased, a well known fact in array processing. When the sources are well separated the bias is negligible compared to the standard deviation, otherwise it might be significant. This gives an explanation for the poor performance of the CLEAN in imaging extended structures (see e.g., [21]).

Minimum variance beamforming approaches Once we view image formation/deconvolution as equivalent to direction-of-arrival (DOA) estimation with a moving array, we can try to adapt various other DOA estimators for handling the image formation. In particular the deflation approach used in the CLEAN algorithm can be replaced by other source parameters estimators. One approach that seems particularly relevant in this context is the Minimum-Variance Distortionless Response (MVDR) method of beamforming [24]. The major new aspect here is the fact that the array is moving and that there are more sources than sensors.

Instead of working with the dirty image $I_D(\mathbf{s}) = \sum_k \mathbf{a}_k^{\mathrm{H}}(\mathbf{s}) \mathbf{R}_k \mathbf{a}_k(\mathbf{s})$, the basis for high-resolution beamforming techniques is to look at more general "pseudo-spectra"

$$I_D'(\mathbf{s}) \;:=\; \sum_k \mathbf{w}_k^{\mathrm{H}}(\mathbf{s}) \mathbf{R}_k \mathbf{w}_k(\mathbf{s}) \quad (13)$$

Here, $\mathbf{w}_k(\mathbf{s})$ is the beamformer pointing towards direction $\mathbf{s}$, and $I_D'(\mathbf{s})$ is the output energy of that beamformer. Previously we had $\mathbf{w}_k(\mathbf{s}) = \mathbf{a}_k(\mathbf{s})$; the objective is to construct beamformers that provide better separation of close sources.

A generalized MVDR follows by defining the problem as follows: At each time instance k we would like to generate a weight vector $\mathbf{w}_k$ which minimizes the output power at time k subject to the constraint that we have a fixed response towards the look direction $\mathbf{s}$ of the array, i.e.,

$$\hat{\mathbf{w}}_k(\mathbf{s}) \;=\; \arg\min_{\mathbf{w}_k} \mathbf{w}_k^{\mathrm{H}} \hat{\mathbf{R}}_k \mathbf{w}_k \qquad \text{such that} \qquad \mathbf{w}_k^{\mathrm{H}} \mathbf{a}_k(\mathbf{s}) = 1\,.$$

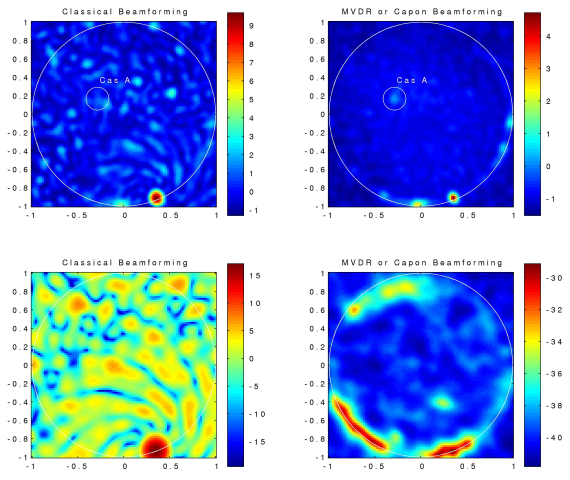

Figure 7. Examples of "dirty images" from the LOFAR test station: *(top row)* classical beamforming and MVDR image of Cassiopeia A along with interference at the horizon, *(bottom row)* similar images at a different frequency where interference is more dominant (Cas-A is not expected to be visible at this frequency due to ionosphere occlusion).

The solution to this problem is

$$\hat{\mathbf{w}}_k = \beta_k \hat{\mathbf{R}}_k^{-1} \mathbf{a}_k(\mathbf{s}), \qquad \text{where} \quad \beta_k = \frac{1}{\mathbf{a}_k^{\mathrm{H}}(\mathbf{s}) \hat{\mathbf{R}}_k^{-1} \mathbf{a}_k(\mathbf{s})}.$$

Inserting in (13) shows that the overall spectral estimator is given by

$$I'_D(\mathbf{s}) = \sum_{k=1}^{K} \frac{1}{\mathbf{a}_k^{\mathrm{H}}(\mathbf{s}) \hat{\mathbf{R}}_k^{-1} \mathbf{a}_k(\mathbf{s})}. \tag{14}$$

and the locations of the strongest sources are given by the maxima of $I'_D(\mathbf{s})$. It is known that the MVDR has improved resolution compared to the classical beamformer which is the basis for the CLEAN algorithm. The algorithm is readily extended to handle the "space-varying" beamshapes that occur after spatial filtering. It is also possible to use more advanced forms of beamforming, e.g., "robust Capon beamforming" (RCB) [25].

Figure 7 illustrates this by comparing a dirty image produced in the classical way to the dirty image corresponding to (14). The measurement data is a 1-minute "snapshot" ($K = 1$) collected from a 60-element test station for the LOFAR telescope discussed in section 6. Since this is a two-dimensional array, it does not depend on earth rotation to enable imaging. Due to the limited integration time, the sky sources are not yet observed and only interference shows up, which is visible at the horizon. All other features are due to the sidelobes of the dirty beam. However, using MVDR imaging it is possible to recover Cassiopeia A, which is the strongest extra-solar radio source in the sky. Figure 8 shows some additional detail at the location of the interferer; the Robust Capon Beamforming technique is seen to give superior resolution.

6. FUTURE RADIO TELESCOPES

Technological advances in the last decade have created interesting possibilities for large distributed interferometric radio telescopes with very large receiving areas and a sensitivity which is one to

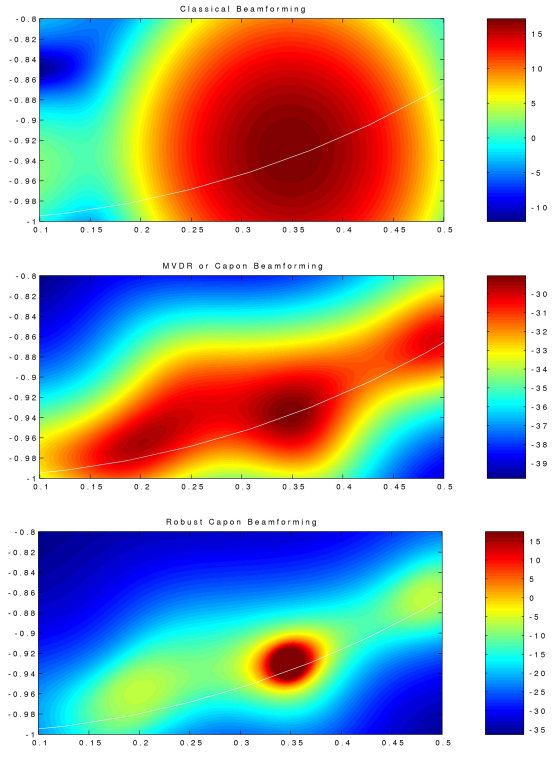

Figure 8. Zoom of figure 7 (bottom row) at the location of the dominant interferer: (a) Classical beamforming, (b) MVDR, (c) Robust Capon Beamforming.

two orders of magnitude better than the current generation. Two proposed instruments in this context are LOFAR [26, 27] and SKA [28, 29]; the previously mentioned THEA system is a smallscale step-up in the design of SKA. Prominent among the challenges of designing and building these telescopes (apart from the costs) are the mitigation of radio interference, the calibration of the system, and the sheer signal processing complexity.

LOFAR (Low Frequency Array) is a next generation radio telescope which is currently under design in an international consortium formed by ASTRON (Dwingeloo, The Netherlands), MIT Haystack Observatory (Boston area), and the US Naval Research Laboratory (Washington), as well as in a Dutch national consortium with many partners, headed by ASTRON. The goal of LOFAR is to enable radio astronomical observations in the 20–240 MHz band with an unprecedented high resolution and sensitivity. This portion of the spectrum corresponds to very high redshifts, and can be used to image the most distant and oldest radio galaxies and quasars in the universe. These can hardly be studied with existing instruments, and since their signals are so faint, an order of magnitude increase in collecting area and baseline lengths is needed. This is not practical anymore with the traditional mechanically steerable dishes.

The preliminary LOFAR design calls for an instrument consisting of about 13,000 'simple' omni-directional antennas, grouped in about 50 stations spread in spirals over an area with a diameter of about 360 km, as well as in a more densely populated central core. The 200 antennas in each remote station are used as a phased array and are combined in such a way that a beam is formed into a

Figure 9. (a) LOFAR distribution of the stations; (b) LOFAR Initial Test Station under construction by ASTRON in snow-covered Borger-Odoorn.

desired look-direction. The same is done for the 3,000 elements in the core. The resulting output of each beamformer is similar to the output of a telescope dish pointing into the same direction, but is obtained without the use of any moving parts. For large collecting areas, this results in significant cost savings; moreover, by duplicating the electronics it is simple to form multiple beams in different directions at the same time. The preliminary LOFAR design defines 8 simultaneous beams, each spanning a (non-contiguous) bandwidth of 4 MHz in the range 20–240 MHz. At this moment, a test station is under construction in Borger-Odoorn (The Netherlands), see figure 9(b), and preliminary datasets using 60 elements are becoming available.

A similar example is the Square Kilometer Array (SKA). This new generation telescope should have an effective aperture area of one square kilometer, in the frequency range from 150 MHz up to 20 GHz. Just as as LOFAR, it will be a large distributed telescope with many individual elements. The telescope concept for SKA is not yet defined.

A few of the research challenges for LOFAR relevant for array signal processing are mentioned next.

Calibration. Initially the locations and frequency-dependent gains and phases of each receiver unit are unknown and need to be estimated. Additionally, the disturbance due to the propagation through the earth ionosphere has to be measured and compensated for. For LOFAR this is very much complicated by the fact that, at low frequencies, the ionosphere is not uniform and can change within minutes. The final sensitivity of the instrument is strongly dependent on how accurate this calibration is carried out.

RFI mitigation. The band 10–240 MHz contains many sources of RFI (radio frequency interference). The band 88-108 MHz is occupied by FM radio transmitters, and actually this band is given up in the LOFAR design. Apart from this, there are TV broadcasts, digital audio broadcasts, etc. LOFAR will be the first radio telescope in which RFI mitigation techniques will (necessarily) form an integral part of the system design.

There will be two levels of RFI mitigation: at the station level and at the central level. At the station level, the 200 antennas in a station can be combined to modify the beamshape such that an interferer is nulled. This will at the same time modify the reception of the sky signals, and hence the resulting beamshape has to be known at the central correlator, which is complicated by the irregular structure of the array. This form of spatial filtering can be used to null strong local interference. At the central level, the signals from each station can be combined to null any remaining interference that is received by several stations simultaneously. A correction is needed to take the disturbance of the sky signals into account [5].

Postcorrelation processing and imaging. In its simplest form, image formation consists of a spatial Fourier transform of the received correlation data—similar to seismic, synthetic aperture radar (SAR) and NMR imaging. Accurate array calibration parameters are needed to perform this step correctly.

Because the number of correlation samples (u, v data) is finite and irregularly spaced, the response of a single point source in the sky is not a point, but an irregular cloud in the initial, "dirty" image. Iterative deconvolution algorithms such as CLEAN are used to find the locations of the point sources and subtract their effect in the image. (Certain types of interfering sources can be removed in this way as well.) At each step, the residual image can be compared to a image with only point sources, and the difference (error) can be used to improve the calibration parameters into a direction that will minimize the error. This gradient search technique is the essence of the iterative calibration algorithm SELFCAL [23, 30–32], developed in the 1970s for the Westerbork telescope and subsequently used in most other synthesis telescopes.

For LOFAR, the idea is to use similar imaging algorithms. However, the complexity is orders of magnitude higher: the number of calibration parameters is in the order of a few million rather than a hundred, the number of dominant point sources to be estimated is also much larger, and the convergence of a proposed modified SELFCAL algorithm for such a large number of parameters has not been demonstrated.

7. ACKNOWLEDGEMENTS

We would like to thank S. van der Tol at TU Delft and our project partners at ASTRON for the very useful collaboration.

References

[1] A. Leshem and A.J. van der Veen, "Radio astronomical imaging in the presence of strong radio interference," *IEEE Tr. Information Th.*, vol. 46, pp. 1730–1747, Aug. 2000.

[2] A. Leshem, A.J. van der Veen, and A.J. Boonstra, "Multichannel interference mitigation techniques in radio astronomy," *Astrophysical Journal Supplements*, vol. 131, pp. 355–374, Nov. 2000.

[3] A. Leshem and A.J. van der Veen, "Multichannel detection of gaussian signals with uncalibrated receivers," *IEEE Signal Processing Letters*, vol. 8, pp. 120–122, Apr. 2001.

[4] A. Leshem and A.J. van der Veen, "Multichannel detection and spatial signature estimation with uncalibrated receivers," in *11th IEEE Workshop on Stat. Signal Proc.*, (Singapore), Aug. 2001.

[5] J. Raza, A.J. Boonstra, and A.J. van der Veen, "Spatial filtering of RF interference in radio astronomy," *IEEE Signal Processing Letters*, vol. 9, pp. 64–67, Feb. 2002.

[6] A.J. Boonstra and A.J. van der Veen, "Gain calibration methods for radio telescope arrays," *IEEE Tr. Signal Processing*, vol. 51, pp. 25–38, Jan. 2003.

[7] A.J. van der Veen, "Spatial filtering of RF interference in radio astronomy using a reference antenna," in *Proc. IEEE ICASSP*, (Montreal (Candada)), IEEE, May 2004.

[8] S. van der Tol and A.J. van der Veen, "Performance analysis of spatial filtering of rf interference in radio astronomy," *accepted for IEEE Tr. Signal Proc.*, Mar. 2004.

[9] M. Ryle, "A new radio interferometer and its application to the observation of weak stars," *Proc. Royal Society A*, vol. 211, pp. 351–375, 1952.

[10] M. Ryle, "The new Cambridge radio telescope," *Nature*, vol. 194, pp. 517–518, 1962.

[11] D.N. Lawley and A.E. Maxwell, *Factor Analysis as a Statistical Method*. Butterworth & Co, London, 1971.

[12] K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*. Academic Press, 1979.

[13] P.A. Fridman and W.A. Baan, "RFI mitigation methods in radio astronomy," *Astronomy and Astrophysics*, vol. 378, pp. 327–344, 2001.

[14] P. Friedman, "A change point detection method for elimination of industrial interference in radio astronomy receivers," in *Proc. 8th IEEE Signal Proc. Workshop on Stat. Signal Array Proc.*, pp. 264–266, 1996.

[15] R. Weber, C. Faye, F. Biraud, and J. Dansou, "Spectral detector for interference time blanking using quantized correlator," *Astronomy and Astrophysics Supp.*, vol. 126, pp. 161–167, Nov. 1997.

[16] C. Barnbaum and R.F. Bradley, "A new approach to interference excision in radio astronomy: Real time adaptive filtering," *The Astronomical Journal*, vol. 115, pp. 2598–2614, 1998.

[17] S.W. Ellingson, J.D. Bunton, and J.F. Bell, "Removal of the GLONASS C/A signal from OH spectral line observations using a parametric modelling technique," *Astrophysical Journal Supplement*, vol. 135, no. 1, pp. 87–93, 2001.

[18] F.H. Briggs, J.F. Bell, and M.J. Kesteven, "Removing radio interference from contaminated astronomical spectra using an independent reference signal and closure relations," *The Astronomical Journal*, vol. 120, pp. 3351–3361, 2000.

[19] B.D. Jeffs, K. Warnick, and Lisha Li, "Improved interference cancellation in synthesis array radio astronomy using auxiliary antennas," in *IEEE ICASSP*, (Hong Kong), Apr. 2003.

[20] B.D. Jeffs, Lisha Li, and K. Warnick, "Auxiliary antenna assisted interference cancellation for radio astronomy arrays," *accepted for IEEE Trans. Signal Processing*, Feb. 2004.

[21] R.A. Perley, F. Schwab, and A.H. Bridle, eds., *Synthesis imaging in radio astronomy (A collection of lectures from the third NRAO Synthesis Imaging Summer School)*, vol. 6. San Francisco: Astronomical Society of the Pacific, 1989. ISBN 0-937707-23-6.

[22] J.A. Hogbom, "Aperture synthesis with non-regular distribution of interferometer baselines," *Astronomy and Astrophysics Supp.*, vol. 15, pp. 417–426, 1974.

[23] U.J. Schwarz, "Mathematical-statistical description of the iterative beam removing technique (method CLEAN)," *Astronomy and Astrophysics*, vol. 65, pp. 345–356, 1978.

[24] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, pp. 1408–1418, 1969.

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

[26] "LOFAR web site." http://www.lofar.org.

[27] J.D. Bregman, "Design concepts for sky noise limited low frequency array," in *Perspectives on Radio Astronomy - Technologies for Large Antenna Arrays* (A.B. Smolders and M.P. van Haarlem, eds.), ASTRON, 1999.

[28] "SKA web site." http://www.skatelescope.org.

[29] A. van Ardenne and H. Butcher, "Square kilometer array project: Concepts and technologies," in *URSI General Assembly*, (Maastricht (NL)), Aug. 2002.

[30] J.E. Noordam and A.G. de Bruyn, "High dynamic range mapping of radio sources, with application to 3C84," *Nature*, vol. 299, pp. 597–600, 1982.

[31] T.J. Pearson and A.C.S. Readhead, "Image formation by self-calibration in radio astronomy," *Annual Rev. Astronomy and Astrophysics*, vol. 22, pp. 97–130, 1984.

[32] J. P. Hamaker, "Understanding radio polarimetry. IV. The full-coherency analogue of scalar self-calibration: Self-alignment, dynamic range and polarimetric fidelity," *Astron. Astrophys. Suppl. Ser.*, vol. 143, pp. 515–534, May 2000.