Spatial Filtering of RF Interference in Radio Astronomy Using a Reference Antenna Array

Ahmad Mouri Sardarabadi, Alle-Jan van der Veen, and Albert-Jan Boonstra

***Abstract*—Radio astronomical observations are increasingly contaminated by RF interference. Assuming an array of telescopes, a previous technique considered spatial filtering based on projecting out the interferer array signature vector. A disadvantage is that this effectively reduces the array by one (expensive) telescope. In this paper, we consider extending the astronomical array with a reference antenna array, and develop spatial filtering algorithms for this situation. The information from the reference antennas improves the quality of the interferer signature vector estimation, hence more of the interference can be projected out. Moreover, since only the covariance data of the astronomical array has to be reconstructed, the conditioning of the problem improves as well. The algorithms are tested both on simulated and experimental data.**

***Index Terms*—Array signal processing, interference cancellation, radio astronomy, reference antenna, spatial filtering.**

I. Introduction

RADIO astronomical observations are increasingly contaminated by man-made RF interference. In bands below 2 GHz, we find TV and radio signals, mobile communication (GSM), radar, satellite communication (Iridium) and localization beacons (GPS, Glonass), etc. Although some bands are specifically reserved for astronomy, the stop-band filters of some communication systems are not always adequate. Moreover, scientifically relevant observations are not limited to these bands. Hence, there is a growing need for interference cancellation techniques.

The output of a radio telescope is usually in the form of correlations: the auto-correlation (power) of a single telescope dish, split into frequency bins and integrated over periods of 10–30 seconds or more, and/or the cross-correlations of several dishes. The astronomer uses several hours of such correlation observations to synthesize images and to create frequency-domain spectra at specific sky locations (in particular for the study of spectral emission and absorption lines). Current systems for interference cancellation mostly operate at the post-correlation level, by rejecting suspect correlation products in the time-frequency plane, or by specialized imaging algorithms. Spatial filtering at shorter time scales (pre-correlation) is not commonly applied, but would offer interesting possibilities in the first-stage suppression of continuously present wideband interference in bands that are currently avoided by astronomers. An example is the band between 174–240 MHz, which is currently being populated by Digital Audio Broadcast (DAB) transmissions but is also of interest for the LOFAR radio telescope.

Depending on the interference and the type of instrument, several kinds of RFI mitigation techniques are applicable. Overviews can be found in [2]–[6]. E.g., intermittent interference such as radar pulses can be detected using short-term Fourier transforms and the contaminated time-frequency cells omitted during long-term integration to order 10 s [2]. Similarly, during postprocessing we can suppress intermittent signals using time-frequency blanking, where detection can be based on anomalous power or higher order spectral kurtosis [7]–[9]. However, many communication signals are continuous in time. For a single-dish single-feed telescope, there are not many other options[1] than to consider an extension by a reference antenna which picks up only the interference. In this case LMS-type adaptive cancellation techniques have already been proposed by [12]–[14].

With an array of p telescope dishes (an interferometer), spatial filtering techniques are applicable as well. The desired instrument outputs in this case are $p \times p$ correlation matrices, integrated to order 10 s (more generally: the time over which astronomical array signals can be considered stationary, also taking the rotation of the earth into account). 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. The resulting long-term averages of these matrices are mostly interference-free, but they are biased because of the missing dimensions. Such a projection operation also affects the sensitivity and beam-shape of the array [15]. If the projection vector was sufficiently varying, the bias can be corrected for [16], [17]—we describe this technique in more detail later in Section III-B. For stationary interferers (e.g., TV stations or geostationary satellites), this might not work very well, and the correction has to be done during image formation [18]. A special case of a "stationary" interferer is interference entering on only a single telescope dish. The projections will simply remove that channel, and the information can never be

Manuscript received April 01, 2015; revised July 31, 2015; accepted September 08, 2015. Date of publication September 28, 2015; date of current version December 16, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Eduard Jorswieck. A preliminary version of this paper appeared in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Montreal, QC, Canada, May 2004 [1].

A. M. Sardarabadi and A.-J. van der Veen are with Department of Electrical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: a.mourisardarabadi@tudelft.nl).

A.-J. Boonstra is with ASTRON, 7990 AA Dwingeloo, The Netherlands.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2015.2483481

[1]Exceptions are techniques based on higher-order statistics [10] or estimation of outliers in variance [11].

1053-587X © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

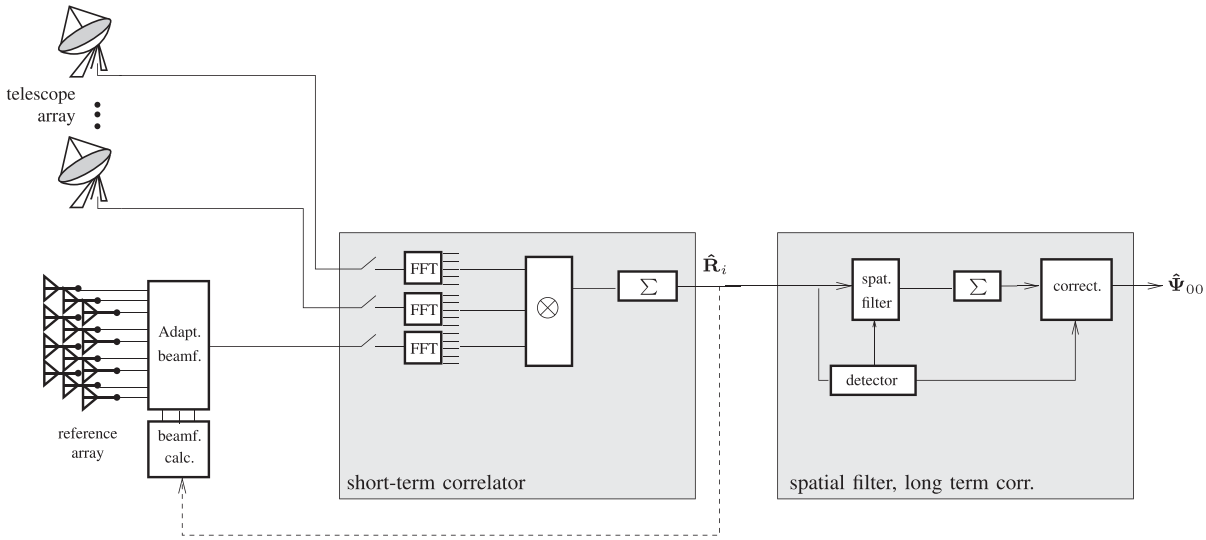

Fig. 1. Telescope array augmented with a reference phased array. A "telescope" could also be a beamformed "station" output, where a station consists of an array of antenna elements.

recovered. A third limitation is that for relatively weak interference the estimate of the signature vector will not be very accurate so that it will not be perfectly cancelled.

To improve on these aspects, we consider in this paper to extend the telescope array with one or more reference antennas. These might be simple omnidirectional antennas, located close to an interfering source (e.g., a poorly shielded computer in the observatory), or a satellite dish pointing into the direction of a geostationary satellite. Most flexibility is obtained by using a phased array which can adaptively be pointed towards the strongest interferers. In the experiment in Section VII, we have used a focal plane array that was mounted on one of the telescope dishes, pointing to zenith.

In the context of phased array telescopes consisting of stations that each form beams on the sky, such as LOFAR or SKA, the equivalent of a "telescope" is a station beam. The reference array may then be a separate array, or remaining degrees of freedom of the stations (e.g., independent beams). It may also consist of the individual station antenna outputs, if we have access to them, or a subset of these.

The generic set-up considered in this paper is shown in Fig. 1. The telescope signals (or station beams) are split into narrow sub-bands and correlated to each other over short time intervals (say to order 0.01–1 second). The reference signals are correlated along with the telescope signals as if they were additional telescopes, and spatial filtering algorithms that project out contaminated dimensions can be applied to the resulting short-term integrated covariance matrices. These matrices could also be used to adaptively beam-steer the reference array towards an interferer. The output of the spatial filter is long-term integrated (say to order 10 seconds), and formally we have to apply a correction matrix to correct for the projected dimensions. The aim of the present paper is to develop spatial filtering algorithms that act on the short-term covariance matrices.

In the literature, several papers have appeared which propose to apply some form of spatial filtering on extended arrays. Briggs *et al.* [19] consider a single dual-polarized telescope (two channels), augmented with two reference antennas. With their technique a single interferer can be cancelled; it is not immediately obvious how it can be extended to more general cases (more antennas, more interferers). Kocz *et al.* [20] propose a projection based spatial filter specifically for multi-beam receivers and show its application for detecting pulsars. Jeffs *et al.* [21], [22] propose spatial filtering algorithms along the lines of [16], [17]; we will summarize their approach in Section III-B and subsequently make extensions which may improve the performance. Hellbourg *et al.* [23] use cyclostationarity of RFI signals to improve the projection estimation while using the same projection correction technique used in [16], [17]. The improvements discussed in this paper are hence equally applicable to their technique.

The structure of the paper is as follows. In Section II, we define the data model and state the problem. In Section III, we present a number of existing spatial filtering algorithms. In Section IV, we extend on these, and also present a more generic (Maximum Likelihood) approach. Section V discusses the theoretical performance of these algorithms. Section VI shows simulation results, and Section VII shows results on experimental data.

A. Notation

Superscript ^{T} denotes matrix transpose, overbar $(\bar{\cdot})$ denotes complex conjugate, and ^{H} complex conjugate transpose, $\mathrm{vec}(\cdot)$ denotes the stacking of the columns of a matrix in a vector and $\mathrm{unvec}(\cdot)$ is the inverse operation. $\mathrm{diag}(\mathbf{a})$ creates a diagonal matrix out of a vector, $\mathrm{vectdiag}(\mathbf{M})$ creates a vector from diagonal elements of a matrix, and $\mathrm{bdiag}(\mathbf{A},\mathbf{B})$ constructs a block-diagonal matrix.

$\otimes$ denotes the Kronecker product, $\circ$ a Khatri-Rao product (column-wise Kronecker product), and $\odot$ the entrywise multiplication of two matrices of equal size. The expectation operator is $\mathrm{E}[\cdot]$, the covariance of an estimated matrix is defined as

$$\mathrm{cov}\{\hat{\mathbf{R}}\} := \mathrm{E}\left[\mathrm{vec}(\mathbf{R}')\mathrm{vec}(\mathbf{R}')^H\right]$$

where $\mathbf{R}' = \hat{\mathbf{R}} - \mathrm{E}[\hat{\mathbf{R}}]$.

For two Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$, the notation $\mathbf{A} \leq \mathbf{B}$ signifies that $\mathbf{B} - \mathbf{A}$ is positive semidefinite.

II. Problem Statement

A. Data Model

Assume we have a telescope array (primary array) with p_0 elements, and a reference array with p_1 elements.[2] The total number of elements is $p = p_0 + p_1$.

We consider complex signals $x_j(t)$ received at the antennas $j = 1, \cdots, p$ in a sufficiently narrow subband. For the interference free case the primary array output vector $\mathbf{x}_0(t)$ is modeled in complex baseband form as

$$\mathbf{x}_0(t) = \mathbf{v}_0(t) + \mathbf{n}_0(t)$$

where $\mathbf{x}_0(t) = [x_1(t), \ldots, x_{p_0}(t)]^T$ is the $p_0 \times 1$ vector of telescope signals at time t, $\mathbf{v}_0(t)$ is the received sky signal, assumed on the time scale of 10 s to be a zero-mean stationary complex Gaussian vector with covariance matrix $\mathbf{R}_{v,0}$ (the astronomical 'visibilities'), and $\mathbf{n}_0(t)$ is the $p_0 \times 1$ noise vector. In the general case for an uncalibrated antenna array where the noise on each element is independent and Gaussian, the noise covariance matrix is a diagonal matrix $\mathbf{\Sigma}_0 = \mathrm{diag}(\boldsymbol{\sigma}_0)$, where $\boldsymbol{\sigma}_0$ is a $p_0 \times 1$ vector of noise powers on each element, and if the array is calibrated this simplifies to $\sigma_0^2\mathbf{I}$. The astronomer is interested in $\mathbf{R}_{v,0}$.

If an interferer is present the primary array output vector is modeled as

$$\mathbf{x}_0(t) = \mathbf{v}_0(t) + \mathbf{a}_0(t)s(t) + \mathbf{n}_0(t)$$

where $s(t)$ is the interferer signal with spatial signature vector $\mathbf{a}_0(t)$ which is assumed stationary only over short time intervals. Without loss of generality, we can absorb the unknown amplitude of $s(t)$ into $\mathbf{a}_0(t)$ and thus set the power of $s(t)$ to 1.

Consider now that we also have a reference antenna array. The outputs of the p_1 reference antennas are stacked into a vector $\mathbf{x}_1(t)$, modeled as

$$\mathbf{x}_1(t) = \mathbf{a}_1(t)s(t) + \mathbf{n}_1(t).$$

It is assumed here that the contribution of the astronomical sources to the reference signals is negligible. The noise on the reference antenna array is assumed to be independent and Gaussian with a diagonal covariance matrix. For an uncalibrated array $\mathbf{\Sigma}_1 = \mathrm{diag}(\boldsymbol{\sigma}_1)$, and for a calibrated array it is $\mathbf{\Sigma}_1 = \sigma_1^2\mathbf{I}$.

Stacking all antenna signals in a single vector $\mathbf{x}^T = [\mathbf{x}_0^T, \mathbf{x}_1^T]^T$, and similarly for $\mathbf{v}^T = [\mathbf{v}_0^T, \mathbf{0}^T]^T$, $\mathbf{a}^T = [\mathbf{a}_0^T, \mathbf{a}_1^T]^T$, $\mathbf{n}^T = [\mathbf{n}_0^T, \mathbf{n}_1^T]^T$, we obtain

$$\mathbf{x}(t) = \mathbf{v}(t) + \mathbf{a}(t)s(t) + \mathbf{n}(t). \tag{1}$$

We make the following additional assumptions on this model:

(A1) The noise covariance matrices could be either unknown diagonal matrices or they could be known from calibration, e.g., from observations of nearby uncontaminated frequencies.

(A2) $\mathbf{R}_{v,0} \ll \mathbf{\Sigma}_0$. This is reasonable as even the strongest sky sources are about 15 dB under the noise floor.

(A3) The processing bandwidth is sufficiently narrow, so that possible multipath propagation of the interferer will add up to a single signature vector $\mathbf{a}(t)$ and the interferer is seen as a single source. For this it is at least required that the maximal propagation delay along the telescope array is small compared to the inverse bandwidth.

(A4) The interferer signature $\mathbf{a}(t)$ is stationary over short processing times (say less than 10 ms). It may or may not vary over longer periods. Note that even interferers fixed on earth will appear to move as the earth rotates and the telescopes track a source in different direction. This effect depends on the look direction and the maximal baseline length of the telescopes. (The earlier mentioned window of order 10 s over which $\mathbf{R}_{v,0}$ is stationary is derived from this as well.) The rotation of the telescopes and the associated delay compensation ('fringe stopping') introduced to keep the astronomical signals coherent, give rise to phase changes of the entries of $\mathbf{a}_0(t)$ which for long baselines are significant already over short time intervals. The amplitudes will change because the interferer is usually received via the side lobes of the telescope antenna response, which are highly non-constant and cause temporal variations for tracking dishes or beamformed stations.

The model (1) with a calibrated array was considered in [16]. The model is easily extended to multiple interfering sources, in which case we obtain

$$\begin{aligned}\mathbf{x}_0(t) &= \mathbf{v}_0(t) + \mathbf{A}_0(t)\mathbf{s}(t) + \mathbf{n}_0(t)\\ \mathbf{x}_1(t) &= \qquad\quad \mathbf{A}_1(t)\mathbf{s}(t) + \mathbf{n}_1(t)\end{aligned}$$

or equivalently

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

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

B. Covariance Model

Let be given observations $\mathbf{x}[n] := \mathbf{x}(nT_s)$, 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 $\hat{\mathbf{R}}_i$,

$$\hat{\mathbf{R}}_i = \frac{1}{M}\sum_{n=iM}^{(i+1)M-1} \mathbf{x}[n]\mathbf{x}[n]^H$$

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

Considering the $\mathbf{A}_i := \mathbf{A}(iMT_s)$ as deterministic, the expected value of each $\hat{\mathbf{R}}_i$ is denoted by $\mathbf{R}_i$, which can be written in block-partitioned form as

$$\mathbf{R}_i = \begin{bmatrix}\mathbf{R}_{00,i} & \mathbf{R}_{01,i}\\ \mathbf{R}_{10,i} & \mathbf{R}_{11,i}\end{bmatrix}$$

According to the assumptions, $\mathbf{R}_i$ has model

$$\begin{aligned}\mathbf{R}_i &= \mathbf{A}_i\mathbf{A}_i^H + \mathbf{\Psi}\\ &= \left[\begin{array}{c|c}\mathbf{R}_{v,0} + \mathbf{A}_{0,i}\mathbf{A}_{0,i}^H + \mathbf{\Sigma}_0 & \mathbf{A}_{0,i}\mathbf{A}_{1,i}^H\\ \hline \mathbf{A}_{1,i}\mathbf{A}_{0,i}^H & \mathbf{A}_{1,i}\mathbf{A}_{1,i}^H + \mathbf{\Sigma}_1\end{array}\right]\end{aligned} \tag{2}$$

[2] In subsequent notation, the subscript '0' will generally refer to the primary array and '1' to the reference array.

where $\mathbf{\Psi} := \mathbf{R}_v + \mathbf{\Sigma}$ is the interference-free covariance matrix, $\mathbf{R}_v := \mathrm{bdiag}[\mathbf{R}_{v,0}, \mathbf{0}]$ contains the astronomical visibilities, and $\mathbf{\Sigma} := \mathrm{bdiag}[\mathbf{\Sigma}_0, \mathbf{\Sigma}_1]$ is the diagonal noise covariance matrix. The objective is to estimate the interference-free covariance submatrix (long-term estimate) $\mathbf{\Psi}_{00} := \mathbf{R}_{v,0} + \mathbf{\Sigma}_0$.

III. Existing Algorithms

A. Traditional Subtraction Technique

For very high INR where the noise on the reference antennas is negligible ($\mathbf{\Sigma}_1 = \mathbf{0}$), a classical technique for interference removal using a reference antenna is based on taking the covariance of the primary antennas, $\mathbf{R}_{00,i}$, and subtracting the estimated contribution of the interferers, $\mathbf{A}_{0,i}\mathbf{A}_{0,i}^H$. In effect, the rank deficiency of the interference term

$$\mathbf{A}_i\mathbf{A}_i^H = \begin{bmatrix} \mathbf{A}_{0,i}\mathbf{A}_{0,i}^H & \mathbf{A}_{0,i}\mathbf{A}_{1,i}^H \\ \mathbf{A}_{1,i}\mathbf{A}_{0,i}^H & \mathbf{A}_{1,i}\mathbf{A}_{1,i}^H \end{bmatrix}$$

is exploited: if $q \leq p_1$ and moreover $\mathbf{A}_{1,i} : p_1 \times q$ has full column rank q, then the first p_0 columns must be linear combinations of the remaining p_1. Under these conditions, one can show that $\mathbf{A}_{1,i}^H(\mathbf{A}_{1,i}\mathbf{A}_{1,i}^H)^\dagger\mathbf{A}_{1,i}^H = \mathbf{I}_q$, where $\dagger$ denotes the Moore-Penrose pseudo-inverse, so that

$$\mathbf{A}_{0,i}\mathbf{A}_{0,i}^H = \mathbf{A}_{0,i}\mathbf{A}_{1,i}^H\left(\mathbf{A}_{1,i}\mathbf{A}_{1,i}^H\right)^\dagger\mathbf{A}_{1,i}\mathbf{A}_{0,i}^H$$

Following this observation, inspection of (2) and assuming $\mathbf{\Sigma}_1 = \mathbf{0}$ shows that a 'clean' instantaneous covariance estimate is

$$\hat{\mathbf{\Psi}}_{00,i} = \hat{\mathbf{R}}_{00,i} - \hat{\mathbf{R}}_{01,i}\hat{\mathbf{R}}_{11,i}^\dagger\hat{\mathbf{R}}_{10,i}.$$

The final 'clean' covariance estimate is obtained by averaging over N such matrices to obtain a long-term estimate

$$\hat{\mathbf{\Psi}}_{00} = \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{\Psi}}_{00,i}. \tag{3}$$

Briggs *et al.* [19] derive essentially this algorithm and several variants of it, for the special case of $q = 1$ and $p_1 = 2$. Jeffs *et al.* [21] describe the same technique as a generalization of the classical Multiple Sidelobe Canceller.

The mentioned conditions on $\mathbf{A}_{1,i}$ entail that this technique can be used for at most p_1 interferers, and only if the reference antennas are sufficiently independent so that they receive independent linear combinations of the interferers. Unlike some of the techniques to be discussed in later sections, the technique does not rely on the variation of $\mathbf{A}_i$: in principle, $\mathbf{A}_i$ can be stationary.

Note that there is no detection of the number of interferers, and the noise power is assumed to be very small. This simplifies the algorithm but might also limit its performance. With noise, the estimate is biased since $\mathbf{R}_{11,i}$ contains a noise term. Obviously, if the noise covariance is known, the noise term could be subtracted. However, if the INR of the reference array happens to be poor, the inversion of the corrected (small) term gives rise to instabilities; the resulting estimate $\hat{\mathbf{\Psi}}_{00,i}$ does not even have to be positive definite. Hence, for stability reasons, it is better not to remove the bias. This limits the performance for low INR as seen in simulations. Considering that the low INR case is important, this technique cannot be applied in many practical cases.

B. Spatial Filtering Using Projections

In [16], a spatial filtering algorithm based on projections was introduced, and subsequently analyzed in [17]. Although that algorithm did not assume the presence of reference antennas, it can also be used in our current situation. We will first discuss the case where the spatial signature of the interferers are deterministic or known, then we will generalize it to the case where it is estimated from the data.

1) Deterministic or Known Spatial Signature: Suppose that an orthogonal basis of the subspace spanned by interferer spatial signatures $\mathrm{span}(\mathbf{A}_i)$ is known. Let the basis vectors be the columns of a matrix $\mathbf{U}_i$. We can then form a spatial projection matrix $\mathbf{P}_i$,

$$\mathbf{P}_i := \mathbf{I} - \mathbf{U}_i\mathbf{U}_i^H \tag{4}$$

which is such that $\mathbf{P}_i\mathbf{A}_i = \mathbf{0}$. When this spatial filter is applied to the data covariance matrix,

$$\hat{\mathbf{Q}}_i := \mathbf{P}_i\hat{\mathbf{R}}_i\mathbf{P}_i$$

then all the energy due to the interferers will be nulled:

$$\mathrm{E}[\hat{\mathbf{Q}}_i] = \mathbf{P}_i\mathbf{\Psi}\mathbf{P}_i.$$

If we subsequently average the modified covariance matrices $\hat{\mathbf{Q}}_i$, we obtain a long-term estimate

$$\hat{\mathbf{Q}} := \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{Q}}_i = \frac{1}{N}\sum_{i=1}^{N}\mathbf{P}_i\hat{\mathbf{R}}_i\mathbf{P}_i. \tag{5}$$

$\hat{\mathbf{Q}}$ 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 $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$. This gives

$$\mathrm{vec}(\hat{\mathbf{Q}}) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_i\mathrm{vec}(\hat{\mathbf{R}}_i) \tag{6}$$

where

$$\mathbf{C}_i := \mathbf{P}_i^T \otimes \mathbf{P}_i.$$

If the interference was completely removed then

$$\mathrm{E}\left[\mathrm{vec}(\hat{\mathbf{Q}})\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_i\mathrm{vec}(\mathbf{\Psi}) = \mathbf{C}\mathrm{vec}(\mathbf{\Psi}) \tag{7}$$

where

$$\mathbf{C} := \frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_i.$$

In view of this, we can apply a correction $\mathbf{C}^{-1}$ to $\hat{\mathbf{Q}}$ to obtain the corrected estimate $\hat{\mathbf{\Psi}}$

$$\hat{\mathbf{\Psi}} := \mathrm{unvec}\left(\mathbf{C}^{-1}\mathrm{vec}(\hat{\mathbf{Q}})\right). \tag{8}$$

If the interference was completely projected out then $\hat{\mathbf{\Psi}}$ is an unbiased estimate of the covariance matrix without interference.

This algorithm was introduced in [16] and its performance was discussed in [17].

The main computational complexity is in constructing $\mathbf{C}$ and inverting it, as this is generally a very large matrix ($p^2 \times p^2$). Inversion of this would require $O(p^6)$ operations, but because the inverse is applied to only a single vector this can be reduced to $O(p^4)$ using numerical techniques.[3] Note in this respect that, for large N and sufficiently varying $\mathbf{A}_i$, $\mathbf{C}$ is usually quite close to an identity matrix, and the hope would be that the correction can be omitted or highly simplified under such conditions.

The reconstructed covariance matrix $\hat{\mathbf{\Psi}}$ is size $p \times p$. In the present case, we are only interested in the submatrix corresponding to the primary antennas. Hence, the estimate produced by the algorithm is the $p_0 \times p_0$ submatrix in the top-left corner, $\hat{\mathbf{\Psi}}_{00}$.

2) Unknown Spatial Signature, Known Noise Covariance: The spatial signatures of the interferers are generally unknown, but if the noise covariance $\mathbf{\Sigma}$ is known the interfering subspace can be estimated from an eigen-analysis of the sample covariance matrices $\hat{\mathbf{R}}_i$. If $\mathbf{\Sigma}$ is not a multiple of $\mathbf{I}$, then we first have to whiten the noise to make the noise powers on all antennas the same. This is done by working with $\mathbf{\Sigma}^{-1/2}\hat{\mathbf{R}}_i\mathbf{\Sigma}^{-1/2}$. Without interference and assuming $\mathbf{R}_v$ is negligible compared to $\mathbf{\Sigma}$, all eigenvalues of this matrix are expected to be close to 1. With q interferers, q eigenvalue become larger, and the eigenvectors corresponding to these eigenvalues are an estimate of $\mathrm{span}(\mathbf{\Sigma}^{-1/2}\mathbf{A}_i)$.

Remarks:

1) The algorithm relies on the invertibility of $\mathbf{C}$, which is constructed from projection matrices. Each projection matrix is rank deficient. Hence, $\mathbf{C}$ is invertible only if the spatial signature vectors which are projected out are sufficiently varying. In [17] it was noted that for $q = 1$, usually already 3 different projections are sufficient to guarantee that $\mathbf{C}$ is full rank.
2) The algorithm is inefficient in the sense that it first reconstructs the complete covariance matrix, then selects the submatrix of interest. Since more parameters (the complete covariance) are estimated, the performance (estimation accuracy) is reduced.
3) If the noise covariance $\mathbf{\Sigma}$ is not known, then the eigenvalue decomposition can be replaced by a more general Factor Analysis decomposition, see Section IV-B.
4) Regarding the subspace estimation, the maximum number of interferers is constrained by $q < p$. Each $\mathbf{C}_i$ has size $p^2 \times p^2$ and rank $(p-q)^2$; invertibility requires at least $N(p-q)^2 \geq p^2$ (in case the projected subspaces are completely arbitrary).

In summary, this spatial filtering algorithm does not really take advantage of the reference antennas. In the processing, it treats them like ordinary antennas. The only benefit obtained from them is that, with an improved INR, the estimate of the interference subspace will be better, so that the interference can be filtered out better. The performance is then limited by the conditioning of $\mathbf{C}$ (thus the variability of the spatial signature vectors).

[3]In comparison, image formation techniques work with correlation matrices of size $p \times p$ and have a complexity of $O(p^2)$.

IV. New Algorithms

A. Improved Spatial Filter With Projections

Taking the above remarks into account, we derive an improved algorithm. Compute the projections and long-term average of the projected estimates $\hat{\mathbf{Q}}$ as before in (5). Then (7) applies:

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

Based on this, we previously set $\mathrm{vec}(\hat{\mathbf{\Psi}}) = \mathbf{C}^{-1}\mathrm{vec}(\hat{\mathbf{Q}})$, which is the solution in Least Squares sense of the covariance model error minimization problem, $\|\mathrm{vec}(\hat{\mathbf{Q}}) - \mathbf{C}\mathrm{vec}(\hat{\mathbf{\Psi}})\|^2$. Now, instead of this, partition $\mathbf{\Psi}$ as in (2) 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 $\mathbf{\Psi}_{01} = \mathbf{0}$, $\mathbf{\Psi}_{10} = \mathbf{0}$, $\mathbf{\Psi}_{11} = \mathbf{\Sigma}_1$. This corresponds to solving the reduced-size covariance model error minimization problem,

$$\hat{\mathbf{\Psi}}_{00} = \arg\min_{\mathbf{\Psi}_{00}} \left\| \mathrm{vec}(\hat{\mathbf{Q}}) - \mathbf{C}\mathrm{vec}\left(\left[\begin{array}{c|c} \mathbf{\Psi}_{00} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{\Sigma}_1 \end{array} \right] \right) \right\|^2 .$$

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

$$\left[\begin{array}{c} \mathrm{vec}(\mathbf{\Psi}_{00}) \\ \hline \boldsymbol{\sigma}_1 \\ \mathbf{0} \end{array} \right]$$

where $\boldsymbol{\sigma}_1 = \mathrm{vectdiag}(\mathbf{\Sigma}_1)$, and repartition $\mathbf{C}$ accordingly, to obtain the equivalent problem

$$\begin{aligned} \hat{\boldsymbol{\psi}}_{00} &= \arg\min_{\boldsymbol{\psi}_{00}} \left\| \mathrm{vec}(\hat{\mathbf{Q}}) - [\mathbf{C}_1 \ \ \mathbf{C}_2 \ \ \mathbf{C}_3] \left[\begin{array}{c} \boldsymbol{\psi}_{00} \\ \hline \boldsymbol{\sigma}_1 \\ \mathbf{0} \end{array} \right] \right\|^2 \\ &= \arg\min_{\boldsymbol{\psi}_{00}} \left\| \left(\mathrm{vec}(\hat{\mathbf{Q}}) - \mathbf{C}_2\boldsymbol{\sigma}_1 \right) - \mathbf{C}_1\boldsymbol{\psi}_{00} \right\|^2 \\ &= \mathbf{C}_1^{\dagger} \left(\mathrm{vec}(\hat{\mathbf{Q}}) - \mathbf{C}_2\boldsymbol{\sigma}_1 \right), \end{aligned} \tag{9}$$

where $\hat{\boldsymbol{\psi}}_{00} = \mathrm{vec}(\hat{\mathbf{\Psi}}_{00})$ and $\boldsymbol{\psi}_{00} = \mathrm{vec}(\mathbf{\Psi}_{00})$. 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.

Remarks:

1) The subspace estimation has not changed, and the maximum number of interferers is still constrained by $q < p$. Now $\mathbf{C}_1$ has size $p^2 \times p_0^2$ and invertibility requires at least $N(p-q)^2 \geq p_0^2$, in case the projected subspaces are completely arbitrary.
2) Even if the interferers are located stationary ($\mathbf{A}_i$ constant), $\mathbf{C}_1$ is expected to have full column rank and hence the improved algorithm can estimate the astronomical covariance (provided $q \leq p_1$).
3) The same advantage holds in case an interferer only contaminates one of the primary antennas ($\mathbf{a}_{i,0}$ has only one nonzero entry). Without reference antenna, the projection is always the same and cannot be corrected: the correlations corresponding to that antenna are lost. With a reference antenna, they can be recovered.

4) Asymptotically for large INR of the reference array, the algorithm is seen to behave similar to the traditional subtraction technique described in Section III-A.
5) We assumed that $\mathbf{\Sigma}_1$ is known from calibration. For an uncalibrated array we can use Factor Analysis to estimate $\mathbf{\Sigma}_1$, as described next in Section IV-B.

B. Factor Analysis

Factor Analysis (FA) [24] is a multivariate statistical technique that decomposes a covariance matrix into the sum of a low rank matrix and a diagonal matrix. Factor Analysis generalizes Principal Component Analysis (PCA), i.e., the low-rank approximation of a matrix using an Eigenvalue Decomposition (EVD). More general than PCA/EVD, FA includes an unknown diagonal matrix to represent the noise covariance, whereas in PCA/EVD this matrix is assumed to be a multiple of $\mathbf{I}$. Using FA, we can operate on an uncalibrated array.

If we ignore the astronomical correlations $\mathbf{R}_v$, then the short-term covariance matrices are modeled as in (2),

$$\mathbf{R}_i = \mathbf{A}_i \mathbf{A}_i^H + \mathbf{\Sigma}_i, \tag{10}$$

where $\mathbf{\Sigma}_i = \mathbf{\Sigma}$ is a diagonal and $\mathbf{A}_i \mathbf{A}_i^H$ is low rank. Given $\mathbf{R}_i$, FA estimation techniques obtain $\mathbf{\Sigma}_i$ and $\mathbf{A}_i' := \mathbf{A}_i \mathbf{V}_i$, where $\mathbf{V}_i$ is an unidentifiable unitary matrix. However, the subspace $\mathrm{span}(\mathbf{A}_i)$ is identified. As before, collect an orthonormal basis for this subspace in a matrix $\mathbf{U}_i$. Then we can form projections $\mathbf{P}_i$ as before in (4) and proceed with the algorithms in Section III-B or Section IV-A. From this point on, the astronomical correlations are properly taken into account.

Several estimation algorithms are listed in [25]. In general, the complexity is of $O(p^3)$ per short-term correlation matrix. Detection of the number of interferers q can be based on the generalized likelihood ratio test (GLRT), specifying a certain false-alarm ratio [26], or on more advanced model selection techniques (a summary is given in [27]) which are out of scope for this paper.

Remarks:

1) In the above approach, short-term covariance matrices are treated independently, resulting in independent estimates $\hat{\mathbf{\Sigma}}_i$. These should all be equal to the same diagonal $\mathbf{\Sigma}$, but that aspect was not taken into account. As an improvement, we can form a more accurate long-term estimate

$$\hat{\mathbf{\Sigma}} := \frac{1}{N} \sum_{i=1}^{N} \hat{\mathbf{\Sigma}}_i$$

and use this to whiten the short-term covariance matrices, $\mathbf{\Sigma}^{-1/2} \hat{\mathbf{R}}_i \mathbf{\Sigma}^{-1/2}$. We can then proceed with the algorithms in Section III-B or Section IV-A. The subspaces are estimated from the whitened model. As $\hat{\mathbf{\Sigma}}$ is more accurate than each $\hat{\mathbf{\Sigma}}_i$, this is expected to lead to more accurate results. In practical algorithms, the long-term estimate can also be based on a sliding-window estimate or on another estimate from the past, so that the covariance data does not have to be revisited.
2) Regarding identifiability of the model (10), considerations are based on comparing the number of equations to the number of parameters, while including constraints to arrive at a unique $\mathbf{A}_i$. For complex-valued data,[4] a covariance matrix estimate provides p^2 (real) equations, whereas $\mathbf{A}_i$ has $2pq$ (real) parameters, $\mathbf{\Sigma}_i$ has p parameters, and the unknown rotation matrix $\mathbf{V}_i$ is fixed using q^2 (real) constraints (explained in Section V-D). This gives a total number of "degrees of freedom" as [25]

$$s = p^2 - 2pq - p + q^2 = (p-q)^2 - p$$

and for identifiability we need the problem to be overdetermined, i.e. $s \geq 0$. This translates to $q \leq p - \sqrt{p}$ (whereas previously we only had $q < p$).

C. Direct ML Estimation Using Extended Factor Analysis

As we will show here, the data model (2) satisfies the Extended Factor Analysis (EFA) model which we introduced in [25]. This will allow us to directly find a Maximum Likelihood (ML) estimate for $\mathbf{\Psi}_{00}$.

In the EFA model, the diagonal noise covariance matrix $\mathbf{\Sigma}$ is generalized to the interference-free covariance matrix $\mathbf{\Psi}$ which has specified entries that can be nonzero. This is done using a symmetric "masking" matrix $\mathbf{M}$, consisting of entries that are 0 or 1, which constrains $\mathbf{\Psi}$ via the equation $\mathbf{\Psi} = \mathbf{M} \odot \mathbf{\Psi}$. Thus, a '1' entry in $\mathbf{M}$ corresponds to an unknown parameter in $\mathbf{\Psi}$ that needs to be estimated, whereas a '0' indicates that the corresponding entry in $\mathbf{\Psi}$ is known to be zero (this model reduces to classical FA if $\mathbf{M} = \mathbf{I}$).

The covariance model (2) is

$$\mathbf{R}_i = \mathbf{A}_i \mathbf{A}_i^H + \mathbf{\Psi} = \mathbf{A}_i \mathbf{A}_i^H + \left[\begin{array}{c|c} \mathbf{\Psi}_{00} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{\Sigma}_1 \end{array}\right]. \tag{11}$$

where we are interested in estimating the unknown square matrix $\mathbf{\Psi}_{00}$ and, for an uncalibrated array, $\mathbf{\Sigma}_1$ is unknown. Thus, the appropriate masking matrix $\mathbf{M}$ such that $\mathbf{\Psi} = \mathbf{M} \odot \mathbf{\Psi}$ is

$$\mathbf{M} = \left[\begin{array}{c|c} \mathbf{1}\mathbf{1}^T & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I} \end{array}\right].$$

If we replace $\mathbf{\Psi}$ by $\mathbf{\Psi}_i$ (i.e., for each snapshot $\mathbf{R}_i$ we estimate an independent $\mathbf{\Psi}_i$), then we can apply the algorithms for estimating $\mathbf{A}_i$ and $\mathbf{\Psi}_i$ that were introduced in [25]. In that paper, following a Maximum Likelihood approach, it was proposed to maximize

$$l(\mathbf{x}; \mathbf{A}_i, \mathbf{\Psi}_i) = M \left[\log \left| \mathbf{R}_i^{-1} \right| - \log(\pi^p) - \mathrm{tr}\left(\mathbf{R}_i^{-1} \hat{\mathbf{R}}_i \right) \right]$$

with respect to $\mathbf{A}_i$ and $\mathbf{\Psi}_i$. The maximum can be found using various optimization techniques, and in [25] a steepest descent algorithm was detailed.

Each $\hat{\mathbf{R}}_i$ will give us an estimate $\hat{\mathbf{\Psi}}_i$, and $\hat{\mathbf{\Psi}}_{00,i}$ is simply the upper left sub-block of this matrix. The long-term estimate is given by

$$\hat{\mathbf{\Psi}}_{00} = \frac{1}{N} \sum_{i=1}^{N} \hat{\mathbf{\Psi}}_{00,i}. \tag{12}$$

A necessary condition for identification is that the degree of freedom $s > 0$. Compared to FA, we see that the p parameters

[4]This generalizes the real-valued case as documented in [24], [28].

of $\mathbf{\Sigma}$ are now replaced by the $p_0^2 + p_1$ (real) parameters in $\mathbf{\Psi}$. Thus, we require

$$s = p^2 + q^2 - 2pq - \left(p_0^2 + p_1\right)$$

to be larger than 0. With $p = p_0 + p_1$ and solving for the number of reference antennas p_1 we find

$$p_1 > q - \left(p_0 - \frac{1}{2}\right) + \sqrt{q + \left(p_0 - \frac{1}{2}\right)^2}. \qquad (13)$$

Thus, if p_0 is small, we need $p_1 > q + \sqrt{q}$, and if p_0 is large, we need $p_1 > q$.

Remarks:

1) For each snapshot $2pq + p_0^2 + p_1$ unknowns are estimated. The computational complexity of the steepest descent algorithm in [25] is $O(p_0^3 + p^2 q + q^3)$ per snapshot. The term p_0^3 is caused by the inversion of $\mathbf{\Psi}_{00,i}$, which can be reduced to p_0^2 by using a power iteration, assuming the astronomical source powers are small. The complexity will then be $O(p^2 q)$ per snapshot. Note that no expensive post-processing as in (8) or (9) is needed.
2) Although each snapshot is processed with a Maximum Likelihood estimator, the overall algorithm is not maximum likelihood as $\mathbf{\Psi}_{00}$ is estimated using an average of the $\hat{\mathbf{\Psi}}_{00,i}$ in (12).
3) If not all primary antennas receive the RFI, then p_0 should be replaced by $\tilde{p}_0$ in (13), where $\tilde{p}_0$ is the number of primary antennas that receive RFI.
4) This approach assumes that the number of interferers is known. If this is not the case we can follow a similar approach as was suggested in Section IV-B.

V. Performance Analysis

The result of the algorithm is $\hat{\mathbf{\Psi}}_{00}$, an estimate of the true covariance matrix $\mathbf{\Psi}_{00}$. As a measure of accuracy, we will determine the covariance of $\hat{\mathbf{\Psi}}_{00}$ in four cases: (*i*) interference free, (*ii*) the spatial signatures $\mathbf{A}_i$ are known, (*iii*) the spatial signatures are estimated, and (*iv*) the EFA model.

The covariance of the short-term averages is defined by

$$\mathrm{cov}\{\hat{\mathbf{R}}_i\} := \mathrm{E}\left[\mathrm{vec}\left(\mathbf{R}_i'\right)\mathrm{vec}\left(\mathbf{R}_i'\right)^H\right]$$

where $\mathbf{R}_i' = \hat{\mathbf{R}}_i - \mathrm{E}[\hat{\mathbf{R}}_i]$, and

$$\mathrm{E}[\hat{\mathbf{R}}_i] = \mathbf{R}_i = \mathbf{\Psi} + \mathbf{A}_i\mathbf{A}_i^H.$$

Assuming Gaussian sources (sky signals and interference), the covariance of the short-term avarages is given by the standard result,

$$\mathrm{cov}\{\hat{\mathbf{R}}_i\} = \frac{1}{M}\overline{\mathbf{R}_i} \otimes \mathbf{R}_i \qquad (14)$$

A. Case I: Interference-Free

In the interference free case, without filtering, $\hat{\mathbf{R}}$ is given by

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{R}}_i$$

Then, from (14), the covariance of $\hat{\mathbf{R}}$ is given by

$$\mathrm{cov}\{\hat{\mathbf{R}}\} = \frac{1}{MN}\overline{\mathbf{R}} \otimes \mathbf{R}$$

where $\mathbf{R} = \mathbf{\Psi} = \mathbf{R}_v + \mathbf{\Sigma}$. Because $\mathbf{R}_v \ll \mathbf{\Sigma}$,

$$\mathrm{cov}\{\hat{\mathbf{\Psi}}\} \approx \frac{1}{MN}\mathbf{\Sigma} \otimes \mathbf{\Sigma}$$
$$\mathrm{cov}\{\hat{\mathbf{\Psi}}_{00}\} \approx \frac{1}{MN}\mathbf{\Sigma}_0 \otimes \mathbf{\Sigma}_0 \qquad (15)$$

For white noise, $\mathbf{\Sigma}_0 = \sigma_0^2\mathbf{I}$, and

$$\mathrm{cov}\{\hat{\mathbf{\Psi}}_{00}\} \approx \frac{\sigma_0^4}{MN}\mathbf{I} \qquad (16)$$

This interference-free result gives a reference performance for the estimation of $\mathbf{\Psi}_{00}$ in the case with interference.

B. Case II: Interference With Known Spatial Signatures

Suppose the subspace spanned by the spatial signatures $\mathbf{A}_i$ of the interferers is deterministic and perfectly known. In that case the interference will be completely removed and the algorithms are unbiased by design. We first consider the algorithm in Section III-B. The estimate of the full-size $\mathbf{\Psi}$ is given by

$$\mathrm{vec}(\hat{\mathbf{\Psi}}) = \mathbf{C}^{-1}\mathrm{vec}(\hat{\mathbf{Q}}), \quad \mathrm{vec}(\hat{\mathbf{Q}}) = \frac{1}{N}\sum \mathbf{C}_i\mathrm{vec}(\hat{\mathbf{R}}_i) \qquad (17)$$

where $\mathbf{C} = \frac{1}{N}\sum \mathbf{C}_i$, $\mathbf{C}_i = \mathbf{P}_i^T \otimes \mathbf{P}_i$ and $\mathbf{P}_i$ is the projection onto the orthogonal complement of $\mathbf{A}_i$. (Note that $\mathbf{C}$ and $\mathbf{C}_i$ are Hermitian.) Thus the covariance of the long term estimate is

$$\mathrm{cov}\{\hat{\mathbf{\Psi}}\} = \mathbf{C}^{-1}\mathrm{cov}\{\hat{\mathbf{Q}}\}\mathbf{C}^{-1}$$

where

$$\mathrm{cov}\{\hat{\mathbf{Q}}\} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\mathbf{C}_i\mathrm{E}\left[\mathrm{vec}\left(\mathbf{R}_i'\right)\mathrm{vec}\left(\mathbf{R}_\ell'\right)^H\right]\mathbf{C}_\ell \qquad (18)$$

and $\mathbf{R}_i' = \hat{\mathbf{R}}_i - \mathbf{R}_i$. The estimation errors $\mathbf{R}_i'$ and $\mathbf{R}_\ell'$ are uncorrelated for $i \neq \ell$, and $\mathrm{E}(\hat{\mathbf{R}}_i) = \mathbf{R}_i = \mathbf{\Psi} + \mathbf{A}_i\mathbf{A}_i^H$. Thus,

$$\begin{aligned}\mathrm{cov}\{\hat{\mathbf{Q}}\} &= \frac{1}{N^2}\sum_{i=1}^{N}\mathbf{C}_i\mathrm{E}\left[\mathrm{vec}\left(\mathbf{R}_i'\right)\mathrm{vec}\left(\mathbf{R}_i'\right)^H\right]\mathbf{C}_i \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\mathbf{C}_i\mathrm{cov}\{\hat{\mathbf{R}}_i\}\mathbf{C}_i \\ &= \frac{1}{MN^2}\sum_{i=1}^{N}\mathbf{C}_i\left(\mathbf{R}_i^T \otimes \mathbf{R}_i\right)\mathbf{C}_i\end{aligned}$$

Multiplication by $\mathbf{C}_i$ projects out the contribution of the interferers $(\mathbf{A}_i)$, so that

$$\begin{aligned}\mathrm{cov}\{\hat{\mathbf{Q}}\} &= \frac{1}{MN^2}\sum_{i=1}^{N}\mathbf{C}_i(\mathbf{\Psi}^T \otimes \mathbf{\Psi})\mathbf{C}_i \\ &\approx \frac{1}{MN^2}\sum_{i=1}^{N}\mathbf{C}_i(\mathbf{\Sigma} \otimes \mathbf{\Sigma})\mathbf{C}_i \\ &= \frac{1}{MN^2}\sum_{i=1}^{N}\left(\mathbf{P}_i^T\mathbf{\Sigma}\mathbf{P}_i^T\right) \otimes (\mathbf{P}_i\mathbf{\Sigma}\mathbf{P}_i)\end{aligned}$$

If we assume $\mathbf{\Sigma} = \sigma^2\mathbf{I}$ (for simplicity of analysis), then

$$\text{cov}\{\hat{\mathbf{Q}}\} \approx \frac{\sigma^4}{MN^2}\sum_{i=1}^{N}\left(\mathbf{P}_i^T \otimes \mathbf{P}_i\right) = \frac{\sigma^4}{MN}\mathbf{C}$$

Thus,

$$\text{cov}\{\hat{\mathbf{\Psi}}\} = \mathbf{C}^{-1}\text{cov}\{\hat{\mathbf{Q}}\}\mathbf{C}^{-1} \approx \frac{\sigma^4}{MN}\mathbf{C}^{-1} \tag{19}$$

The final estimate $\hat{\mathbf{\Psi}}_{00}$ is a submatrix of $\hat{\mathbf{\Psi}}$. Its performance is a submatrix of $\text{cov}\{\hat{\mathbf{\Psi}}\}$. Compared to (15), this indicates that (a submatrix of) $\mathbf{C}^{-1}$ determines the relative performance of the spatial filtering algorithm of Section III-B. The conditioning of $\mathbf{C}^{-1}$ depends on the variability of $\mathbf{A}_i$, the spatial signatures of the interferer. For large N and sufficiently varying $\mathbf{A}_i$, $\mathbf{C} \to \mathbf{I}$ and the performance is expected to be similar to the interference-free case.

The algorithm in Section IV-A is slightly different:

$$\text{vec}(\hat{\mathbf{\Psi}}_{00}) = \mathbf{C}_1^{\dagger}\left(\text{vec}(\hat{\mathbf{Q}}) - \sigma_1^2\mathbf{C}_2\mathbf{1}\right)$$

Thus, (with $\mathbf{\Sigma} = \sigma^2\mathbf{I}$ for simplicity of analysis)

$$\text{cov}\{\hat{\mathbf{\Psi}}_{00}\} = \mathbf{C}_1^{\dagger}\text{cov}\{\hat{\mathbf{Q}}\}\mathbf{C}_1^{\dagger H} \approx \frac{\sigma^4}{MN}\mathbf{C}_1^{\dagger}\mathbf{C}\mathbf{C}_1^{\dagger H} \tag{20}$$

It is known that for any tall matrix $\mathbf{J}$ for which $\mathbf{CJ}$ is full column rank

$$(\mathbf{CJ})^{\dagger}\mathbf{C}(\mathbf{CJ})^{\dagger H} \quad \leq \quad \mathbf{J}^H\mathbf{C}^{-1}\mathbf{J}$$

(cf. [29, lemma 3.1]). Choosing $\mathbf{J}$ a selection matrix such that $\mathbf{C}_1 = \mathbf{CJ}$, it can be deduced that the algorithm of Section IV-A is always more efficient than the algorithm of Section III-B.

C. Case III: The Variance of $\hat{\mathbf{R}}$ for Interference With Deterministic Spatial Signatures

In practice the spatial signatures $\mathbf{A}_i$ are unknown and their column span will be estimated. Because of the estimation error, the projection is not perfect and there will be a residual which might affect the performance.

Again we first study the algorithm in Section III-B. As before, let $\mathbf{P}_i$ be a projection such that $\mathbf{P}_i\mathbf{A}_i = 0$, and denote the estimate of $\mathbf{P}_i$ by $\hat{\mathbf{P}}_i$. The estimate of the full size $\mathbf{\Psi}$ is then

$$\text{vec}(\hat{\mathbf{\Psi}}) = \hat{\mathbf{C}}^{-1}\text{vec}(\hat{\mathbf{Q}}), \quad \text{vec}(\hat{\mathbf{Q}}) = \frac{1}{N}\sum\hat{\mathbf{C}}_i\text{vec}(\hat{\mathbf{R}}_i) \tag{21}$$

where $\hat{\mathbf{C}} = \frac{1}{N}\sum\hat{\mathbf{C}}_i$, $\hat{\mathbf{C}}_i = \hat{\mathbf{P}}_i^T \otimes \hat{\mathbf{P}}_i$. Define the errors

$$\mathbf{R}_i' = \hat{\mathbf{R}}_i - \mathbf{R}_i$$

and likewise for $\mathbf{Q}'$, $\mathbf{C}'$ etc.

Equation (7) is not true because $\hat{\mathbf{C}}$ and $\hat{\mathbf{R}}_i$ are not independent. The algorithm is not unbiased anymore, but it can be shown that the bias of $\hat{\mathbf{\Psi}}$ is $O(M^{-1})$, whereas the standard deviation is $O(M^{-1/2})$ as we show next.

We employ a first-order perturbation analysis. From (21),

$$\text{vec}(\mathbf{\Psi}') \approx \left[\mathbf{C}^{-1}\right]'\text{vec}(\mathbf{Q}) + \mathbf{C}^{-1}\text{vec}(\mathbf{Q}') \tag{22}$$

where we can write

$$\left(\mathbf{C}^{-1}\right)' \approx -\mathbf{C}^{-1}\mathbf{C}'\mathbf{C}^{-1} \tag{23}$$

and

$$\begin{aligned}\text{vec}(\mathbf{Q}') &= \frac{1}{N}\sum_{i=1}^{N}\text{vec}\left(\mathbf{Q}_i'\right)\\ &\approx \frac{1}{N}\sum\mathbf{C}_i'\text{vec}(\mathbf{R}_i) + \mathbf{C}_i\text{vec}\left(\mathbf{R}_i'\right).\end{aligned}$$

We next show that $\frac{1}{N}\sum\mathbf{C}_i'\text{vec}(\mathbf{R}_i) \approx \frac{1}{N}\sum\mathbf{C}_i'\text{vec}(\mathbf{\Psi}) = \mathbf{C}'\text{vec}(\mathbf{\Psi})$. Indeed,

$$\text{vec}(\mathbf{R}_i) = \text{vec}(\hat{\mathbf{\Psi}}) + (\bar{\mathbf{A}}_i \circ \mathbf{A}_i)\mathbf{1}$$

and

$$\mathbf{C}_i' \approx \left(\bar{\mathbf{P}}_i' \otimes \mathbf{P}_i\right) + \left(\bar{\mathbf{P}}_i \otimes \mathbf{P}_i'\right)$$

so that

$$\begin{aligned}\mathbf{C}_i'\text{vec}(\mathbf{R}_i) \approx\ & \mathbf{C}_i'\text{vec}(\mathbf{\Psi})\\ &+ \left(\bar{\mathbf{P}}_i' \otimes \mathbf{P}_i\right)(\bar{\mathbf{A}}_i \circ \mathbf{A}_i)\mathbf{1}\\ &+ \left(\bar{\mathbf{P}}_i \otimes \mathbf{P}_i'\right)(\bar{\mathbf{A}}_i \circ \mathbf{A}_i)\mathbf{1}\end{aligned}$$

Since $(\bar{\mathbf{P}}_i' \otimes \mathbf{P}_i)(\bar{\mathbf{A}}_i \circ \mathbf{A}_i) = (\bar{\mathbf{P}}_i'\bar{\mathbf{A}}_i) \circ (\mathbf{P}_i\mathbf{A}_i) = \mathbf{0}$, and likewise for the third term, only the first term remains. It follows that

$$\text{vec}(\mathbf{Q}') \approx \mathbf{C}'\text{vec}(\mathbf{\Psi}) + \frac{1}{N}\sum\mathbf{C}_i\text{vec}\left(\mathbf{R}_i'\right).$$

Inserting in (22) and using (23) gives

$$\begin{aligned}\text{vec}(\mathbf{\Psi}') \approx\ & -\mathbf{C}^{-1}\mathbf{C}'\mathbf{C}^{-1}\text{vec}(\mathbf{Q})\\ &+\mathbf{C}^{-1}\left(\mathbf{C}'\text{vec}(\mathbf{\Psi}) + \frac{1}{N}\sum\mathbf{C}_i\text{vec}\left(\mathbf{R}_i'\right)\right).\end{aligned}$$

Since $\mathbf{C}^{-1}\text{vec}(\mathbf{Q}) = \text{vec}(\mathbf{\Psi})$, the first two terms cancel, and we remain with

$$\text{vec}(\mathbf{\Psi}') \approx \frac{1}{N}\mathbf{C}^{-1}\sum\mathbf{C}_i\text{vec}\left(\mathbf{R}_i'\right)$$

The rest of the derivation is the same as before in (17), hence the performance is in first order unchanged even if the projections are estimated. The same holds for the algorithm of Section IV-A.

We showed that for a single long-term estimate $\hat{\mathbf{\Psi}}_{00}$, the bias is dominated by the standard deviation of the estimate. In combining multiple long-term estimates (as is done in imaging), the variance reduces and a concern may be that the bias affects the final results. In this respect, it is important to note that the bias is only present on the *diagonal* elements of $\hat{\mathbf{\Psi}}_{00}$ [17], and that this does not impact current imaging algorithms. This is discussed in Section V-E.

D. Case IV: The CRB of $\mathbf{\Psi}_{00}$ and Asymptotic Statistics of EFA

Some general results from multivariate statistical analysis are as follows. Suppose that $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}_0$. Then the asymptotic distribution of $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is $\mathcal{N}(0, \mathbf{\Gamma})$ where $\mathbf{\Gamma}$ is the inverse of the Fisher information matrix $\mathbf{F}$. This is the Cramér-Rao lower bound (CRB) for an unbiased estimator.

In Appendix A we derive the CRB for a single short-term estimate $\hat{\mathbf{\Psi}}_{00,i}$, i.e., the covariance of $\mathrm{vec}(\hat{\mathbf{\Psi}}_{00,i} - \mathbf{\Psi}_{00,i})$, and we denote it by $\mathbf{\Gamma}_{00,i}$.

As a maximum likelihood technique, the EFA algorithm applied to a single short-term estimate is unbiased and will asymptotically reach the CRB, hence its asymptotic performance is given by $\mathbf{\Gamma}_{00,i}$. The long-term estimate $\hat{\mathbf{\Psi}}_{00}$ in EFA is obtained by simply averaging the short-term estimates (assuming the estimates are independent), so that its performance is given by the covariance matrix

$$\mathbf{\Gamma}_{00,efa} = 1/N^2 \sum_i \mathbf{\Gamma}_{00,i}. \tag{24}$$

Using simulations we will show that this asymptotic performance of EFA is achieved for a moderate number of samples and/or INR. The subtraction algorithm described in Section III-A also estimates the long-term estimate $\mathbf{\Psi}_{00}$ by averaging short-term estimates, so that a lower bound on its performance is given by this asymptotic performance of EFA. However, we will show by simulations that because of the bias in this algorithm, the bound is not reached.

In reality, the short-term estimates are not independent as the data model shows that they have $\mathbf{\Psi}$ in common, and an ideal algorithm would do a joint estimate over the entire data set of N covariance matrices. The CRB on the long-term estimate for $\mathbf{\Psi}$ is also derived in Appendix A and the corresponding bound for $\hat{\mathbf{\Psi}}_{00}$ is denoted by $\mathbf{\Gamma}_{00}$.

E. Performance for Long Term Integration (Imaging)

So far we have discussed the statistics for a single long-term estimate $\hat{\mathbf{\Psi}}_{00}$. In many astronomical applications we need to combine a large number of these estimates in order to boost the desired signals to detectable levels. One example of such application is producing a two dimensional image of the sky.

Fourier based imaging (called the 'dirty' image in astronomy, i.e., prior to further deconvolution) can be viewed as computing a weighted average of the entries of the long-term covariance estimates [18]. If K estimates are averaged, then the variance of the estimates is scaled by $1/K$. Without RFI, the individual estimates have a variance given by (16). With RFI removal using projections, the estimates have a slightly higher variance given by (19) or (20). The performance penalty corresponds to the missing data in the projected dimensions, which is natural and acceptable.

The main worry for astronomers would come from any bias that is present in the long-term covariance estimates. The following remarks can be made.

- As a maximum likelihood technique, the EFA is not biased. However, a bias can be present in case a weak interferer is present but not detected (i.e., model mismatch in the EFA). This is a natural limitation in any interference removal technique. The residual interference must be detected and removed after further averaging.
- The projection techniques have a bias, but this bias is present on the diagonal entries of the long-term covariance estimates [17]. Many imaging techniques routinely omit these diagonal entries (the auto-correlations of the antennas) because they are dominated by the system noise. Alternatively, it is possible to correct for the bias to a certain extent [17]. A second source of bias corresponds to RFI that is present but not detected (which as mentioned above is a common problem for RFI mitigation techniques).

In any case, the averaging inherent in the imaging process has a tendency to wash out any residual interference.

In the next section we use the CRB to show that the fundamental bound on the total variance of the estimated covariance is very close to the RFI free case, and we also show that the proposed algorithms are close to this bound provided that the RFI is strong enough to be detected.

VI. Simulations

We first test the performance of the algorithms in a simulation set-up. We use $p = 7$ antennas, with $p_0 = 5$ primary antennas (telescopes) and $p_1 = 2$ reference antennas. For simplicity, the array is a uniform linear array with half-wavelength spacing and the same noise power on all antennas.

The astronomical source is simulated by a source with a constant direction-of-arrival of 10° with respect to array broadside. The source has $\mathrm{SNR}_0 = -20$ dB with respect to each primary array element, and $\mathrm{SNR}_1 = -40$ dB for the reference antenna.

The interferer is simulated by a source with a randomly generated and varying complex $\mathbf{a}_i$, and varying INRs. This corresponds to a Rayleigh fading interferer. A GLRT is performed with a false alarm probability of 0.1 to detect the interfering signal.

The following algorithms are compared:

— the traditional subtraction method, Section III-A, denoted 'Trad Filter',
— the spatial filtering algorithm using projections and eigenvalue computations, Section III-B, denoted 'eig-ref',
— the improved spatial filtering algorithm with reduced-size covariance reconstruction, Section IV-A, denoted 'eig-ref-red' and for Factor Analysis version 'fa-ref',
— the version that uses Extended Factor Analysis is denoted as 'EFA', Section IV-C,
— for comparison, the spatial filtering technique without reference antenna, denoted 'eig-no-ref', the covariance estimate without RFI ('RFI free'), and the estimate obtained without any filtering ('no-filter').

Fig. 2(a) shows the relative mean-squared-error (MSE) of the primary filtered covariance estimate compared to the theoretical value $\mathbf{R}_{v,0} + \sigma_0^2\mathbf{I}$, for varying interferer powers INR_0 on the primary array. Here, we took $M = 5000$ short-term samples and $N = 2$ long-term averages, which is unrealistically small but serves to illustrate the effect of limited variability of $\mathbf{a}_i$ (only two different vectors). The interferer array gain was $\mathrm{INR}_1 - \mathrm{INR}_0$ of 5 dB. Similarly, Fig. 2(b) shows the MSE for varying INR difference and an INR_0 of 10 dB. The amplitudes and phases are varying after each short-term averaging period (i.e., the interferer is Rayleigh fading).

In Fig. 3, we consider a case where the RFI enters the primary array on only a single element. In Fig. 3(a), the INR_0 is varied, while the INR difference is 0 dB and we consider a shorter short-term integration time $M = 200$ and a longer long-term averaging time $N = 5$ than before. Similarly, Fig. 3(b) shows the MSE for varying short-term integration samples M, for an INR_0 of 10 dB.

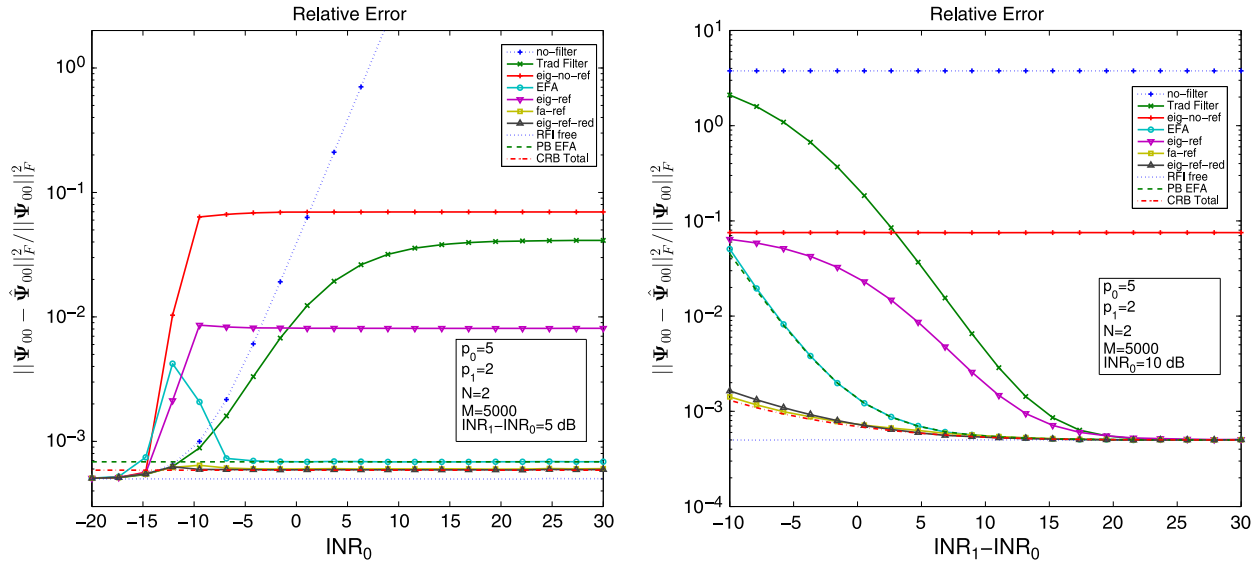

Fig. 2. Simulation with limited variability of the interferer array response vector ($N = 2$). Relative MSE (a) as function of interferer power at the primary array elements (b) as function of the interferer power difference between the reference elements and the primary elements.

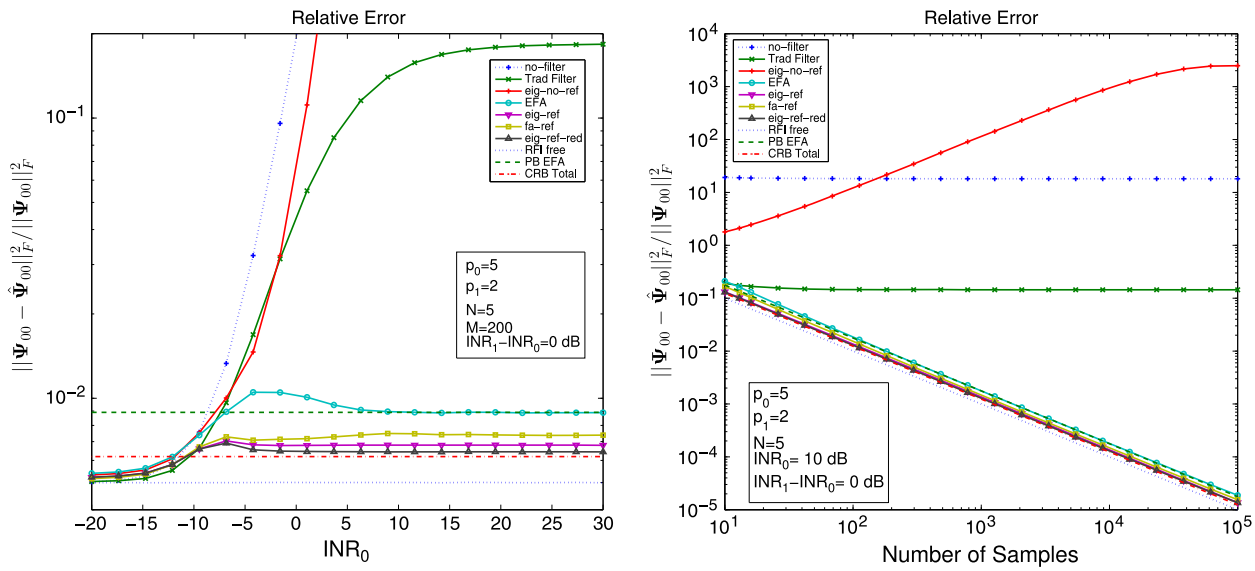

Fig. 3. (a) Simulation with the interferer entering on only a single primary antenna. Relative MSE (a) as function of interferer power at the primary elements (b) as a function of short-term integration samples M.

As reference line, we show the CRB $\mathbf{\Gamma}_{00}$ derived in (38). Because this is a matrix, we use the following relation between the MSE and the trace of the CRB:

$$\begin{aligned}\mathrm{E}\left[\left\|\mathbf{\Psi}'_{00}\right\|_F^2\right] &= \mathrm{E}\left[\mathrm{vec}\left(\mathbf{\Psi}'_{00}\right)^H \mathrm{vec}\left(\mathbf{\Psi}'_{00}\right)\right] \\ &= \mathrm{E}\left[\mathrm{tr}\left(\mathrm{vec}\left(\mathbf{\Psi}'_{00}\right)\mathrm{vec}\left(\mathbf{\Psi}'_{00}\right)^H\right)\right] \\ &\geq \mathrm{tr}(\mathbf{\Gamma}_{00}) \end{aligned} \tag{25}$$

hence $\mathrm{tr}(\mathbf{\Gamma}_{00})$ is a bound on the MSE performance of the proposed algorithms. The MSE is estimated using Monte-Carlo runs.

Observations are:

— The new algorithms that use projections with a reference antenna array (eig-ref-red and fa-ref) operate close to the CRB and have a great advantage over the spatial filtering algorithm without reference antenna (eig-no-ref) in case the $\mathbf{a}_i$-vector is not sufficiently varying [see Fig. 2(a)]. The MSE performance is flat for varying INR and INR difference, which is very desirable. Moreover, it is very close to the RFI-free case. Using FA to find the projections does not noticeably degrade the performance of the filter even though more parameters are estimated.

— The CRB is generally close to the RFI-free case. For low INR, the performance can be better than the CRB because the RFI is not detected and the CRB does not take this model mismatch into account.

— The EFA method also performs well for reasonable INR difference. It operates close to its theoretical performance bound unless the RFI is not detected. However, this bound is seen to be appreciably higher than the CRB in some simulations. This is because the EFA estimates the parameters

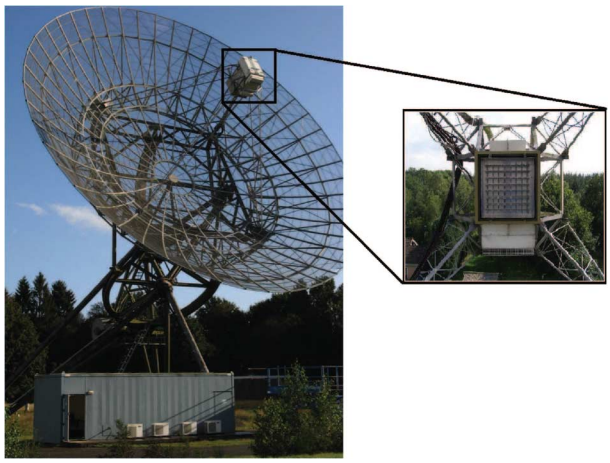

Fig. 4. Reference focal-plane array mounted on a dish.

of each short-term covariance matrix separately, whereas they have parameters in common (i.e., $\mathbf{\Psi}$). Joint processing would be better.

— The new algorithms are often better than the subtraction technique (Trad Filter). The subtraction is only accurate if the INR difference is large compared to the INR at the primary array. If the INR difference is small, or if the INR at the primary array is relatively large, then the subtraction technique fails. This is probably caused by the bias in the inverted term (power of the interferer, with added noise power). It makes the algorithm not reliable to use. As is illustrated in Fig. 3(b) the traditional subtraction does not improve with a higher number of short-term samples which indicates that this is not an efficient estimator.

— If the interferer enters only on one telescope and on the reference antenna, as in Fig. 3, then the algorithm without a reference antenna is performing poorly: it cannot reconstruct the contaminated dimension. The algorithms with reference antennas perform fine.

In summary, based on these simulations, we recommend to consider 'fa-ref' and 'eig-ref-red' (Section IV-A), which are very similar with 'fa-ref' being slightly more general. We expect that EFA can be improved by introducing joint processing.

VII. Examples on Experimental Data

A. Experiment I

To test the algorithm on actual data, we have made a short observation of the strong astronomical source 3C48 contaminated by Afristar satellite signals. The set-up follows Fig. 1. The primary array consists of $p_0 = 3$ of the 14 telescope dishes of the Westerbork Synthesis Radio Telescope (WSRT), located in The Netherlands. As reference signals we use $p_1 = 27$ of 52 elements of a focal--plane array that is mounted on another dish of WSRT which is set off-target (see Fig. 4) such that it has no dish gain towards the astronomical source nor to the interferer.

We recorded 13.4 seconds of data with 80 MS/s, and processed these offline. Using short-term windowed Fourier transforms, the data was first split into 8192 frequency bins (from which we used 1537), and subsequently correlated and averaged over $M = 4048$ samples to obtain $N = 64$ short-term covariance matrices.

Fig. 5(a) shows the autocorrelation and crosscorrelations on the primary antennas and Fig. 5(b) shows the autocorrelation of 6 reference antennas. The interference is clearly seen in the spectrum. The interference consists of a lower and higher frequency part. The low frequency part is stronger on the reference antenna and the higher part stronger on the primary antenna. However, because of a relatively high number of reference antennas the total INR, as we will see, is high enough for the algorithms to be effective.

Because no calibration step has been performed we use a generalized likelihood ratio test (GLRT) [26] to detect if each frequency bin is contaminated with RFI and then we use FA to estimate the noise powers and the signal spatial signature. The result of whitening the spectrum with the estimated result of FA is shown in Fig. 6(a).

The resulting auto- and crosscorrelation spectra after filtering are shown in Fig. 7. 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 removed the RFI within the sensitivity of the telescope. Also it shows the power of using FA and EFA at this stage in the processing chain, as they do not require the array to be calibrated.

B. Experiment II

In a second experiment, we use raw data from the LOFAR station RS409 in HBA mode 5 (100–200 MHz), acquired via the transient buffer board. Data from the 46 (out of 48) x-polarization receiving elements are sampled with a frequency of 200 MHz and correlated. Samples are then divided into 1024 subbands with the help of tapering and an FFT. From these samples we form $N = 4$ covariance matrices with an integration time of 19 ms ($M = 1862$) for each subband. No calibration was done on the resulting covariance matrices.

The LOFAR HBA has a hierarchy of antennas, where a single receiving element output is the result of analog beamforming on 16 antennas (4×4) in a tile. During the measurements the analog beamformers were tracking the strong astronomical source Cyg A.

The received spectrum is shown in Fig. 8. Above 174 MHz, the spectrum is heavily contaminated by wideband DAB transmissions.

We have used 6 of the 46 receiving elements as reference array for our filtering techniques and the rest as primary array. Because we do not have dedicated reference antennas and that the data is already beamformed the assumption that the source is too weak at each short integration time (19 ms) is not completely valid. Also the assumption that the sky sources are much weaker on the reference antennas is not valid in this case because the reference array elements are also following Cyg A. Finally, we have the same exposure to the RFI on the secondary array as we have on the primary so there is no additional RFI gain for the secondary array.

To illustrate the performance of the filtering technique we produce snapshot images of the sky (i.e., images based on a single covariance matrix). For an uncontaminated image, we have chosen subband 250 at 175.59 MHz, see Fig. 9(a), while for RFI-contaminated data we take subband 247 at

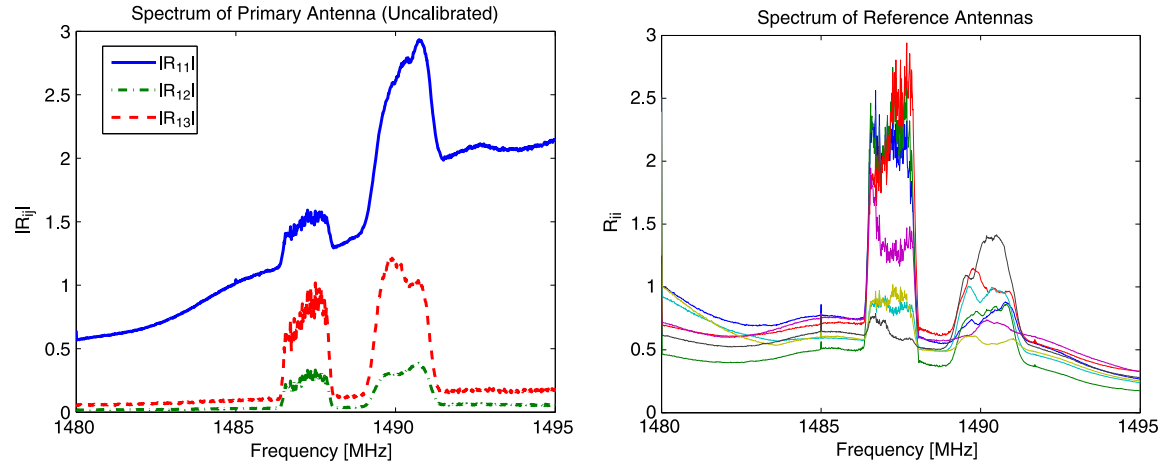

Fig. 5. Observed spectrum from (a) the primary telescopes and (b) 6 of the reference antennas.

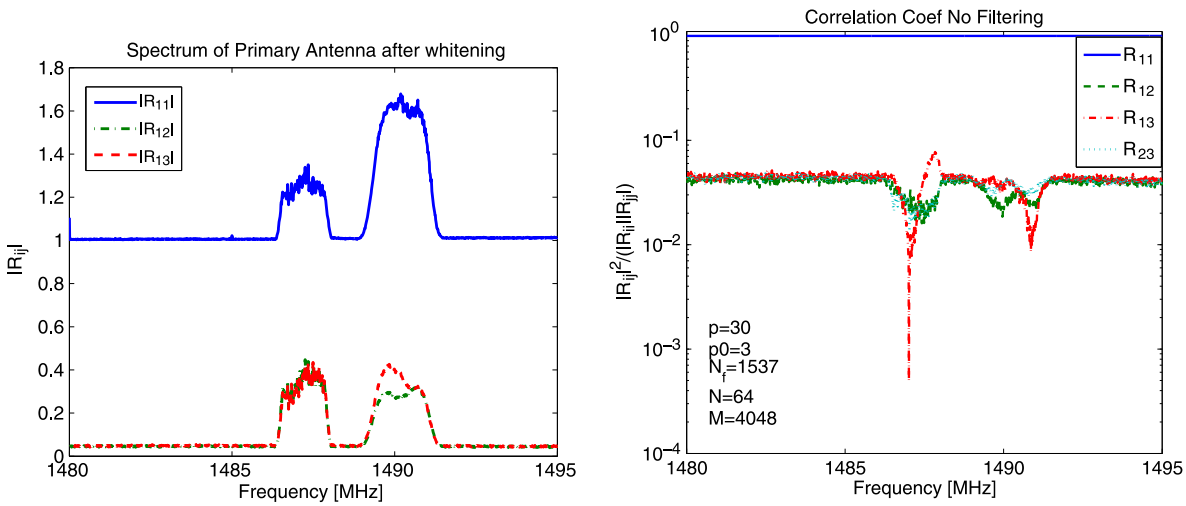

Fig. 6. (a) Spectrum of primary antenna after whitening. (b) Average normalized correlation coefficients without filtering.

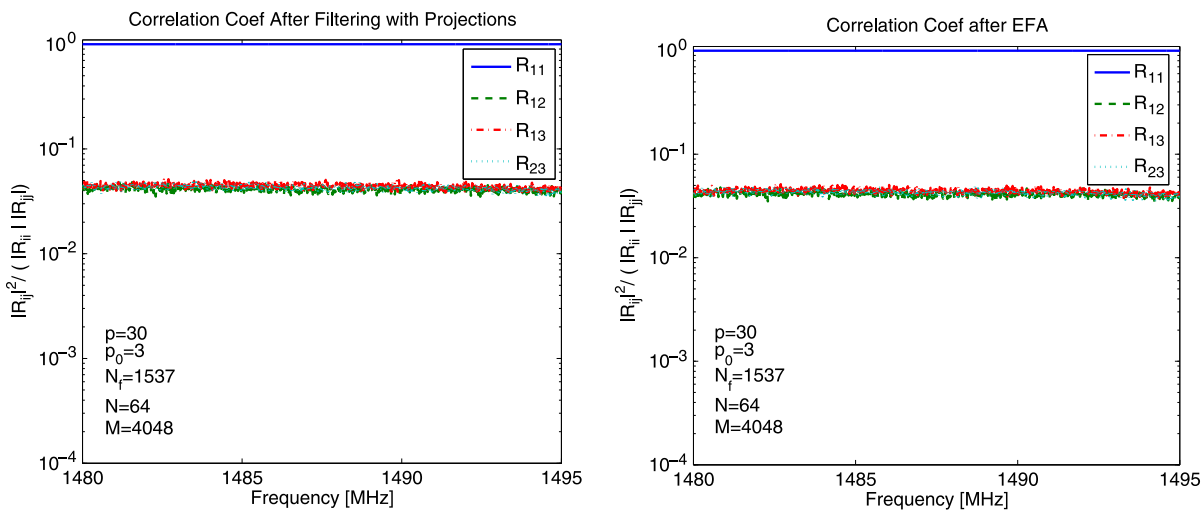

Fig. 7. Averaged normalized correlation coefficients (a) after filtering using method in Section IV-A and (b) after using EFA.

175.88 MHz, see Fig. 9(b). These two subbands have been chosen because they are close to each other (in frequency) and we expect that the astronomical images for these bands would be similar. Subband 247 is heavily contaminated and has a 10 dB flux increase on the auto-correlations and a 20 dB increase on the cross-correlations.

The repeated source visible in Fig. 9(a) is Cyg A; the repetition is due to the spatial aliasing which occurs at these frequen-

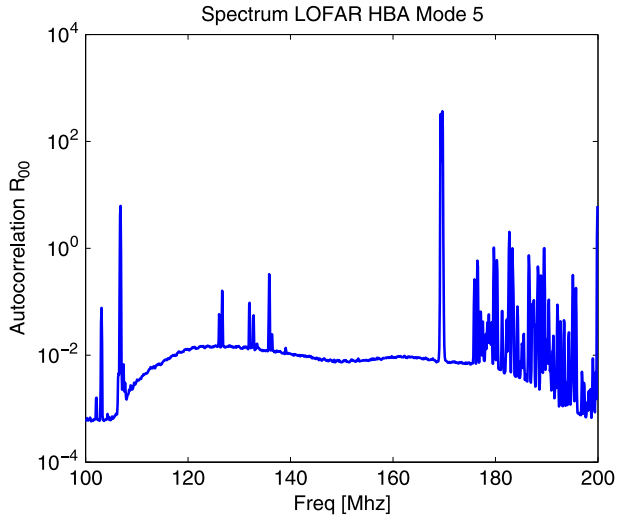

Fig. 8. Spectrum received at a LOFAR HBA station.

cies (the tiles are separated by more than half a wavelength). The contaminated image in Fig. 9(b) shows no trace of Cyg A; note the different amplitude scale which has been increased by a factor 100.

Fig. 10(a) shows the image after filtering the RFI using the algorithm with FA and projections ('fa-ref') as presented in Section IV-A, and Fig. 10(b) shows the image after using EFA (Section IV-C). Both images are nearly identical, and very similar to the clean image in Fig. 9(a)).

Remarks:

— These data show an example where the contaminated portion of the spectrum is broad, limiting the applicability of time-frequency blanking (post-processing).
— Both filtering techniques appear robust against the modeling errors implicit in this experiment setup.
— The resulting covariance estimates produces snapshot images comparable to an RFI-free channel.

Unfortunately the available data collection system at LOFAR did not allow us to create images with longer integration times.

VIII. Conclusions

Spatial filtering algorithms for removing RFI on covariance matrix estimates using reference antennas have been proposed, applicable to both calibrated and uncalibrated arrays. For the uncalibrated case, Factor Analysis is used to estimate the interference subspace. An algorithm to estimate RFI-free covariance matrices directly using Extended Factor Analysis (EFA) has also been presented. The statistical performance of the proposed algorithms has been evaluated and the CRB for the entire dataset is presented.

These algorithms generalize previously proposed spatial filtering algorithms that did not use a reference array. Simulations show that using a reference array is beneficial even if the reference antennas receive less interference power than the primary antennas. Another advantage of a reference array is that the algorithms are applicable even if the interference enters on only a single primary antenna, which was not the case for the previously proposed projection algorithm. The algorithms for uncalibrated arrays based on FA and EFA have also been tested on experimental data from astronomical instruments to illustrate their applicability and performance in real-world scenarios; the results are very encouraging.

A disadvantage of the projection techniques is that they require a computationally unattractive matrix inversion which is needed to correct the covariance estimates for the missing (projected) dimensions. The EFA technique is a direct technique which averages maximum likelihood estimates of cleaned short-term covariance matrices; it is computationally more attractive. Unfortunately, the simulations indicate that the EFA method has a lower performance for low INR and/or low number of samples. The key of the problem is that EFA processes short-term covariance matrices independently and does not exploit that they have a common term $\mathbf{\Psi}$. The solution is in joint processing of these matrices, which will be addressed in future work.

Appendix A
Derivation of the Fisher Information Matrix and CRB

In this section we derive the CRBs, $\mathbf{\Gamma}_{00,i}$ and $\mathbf{\Gamma}_{00}$ discussed in Section V-D.

For normally distributed data with covariance matrix $\mathbf{R}$, the Fisher information matrix is given by Bangs' formula [30]

$$\mathbf{F} = M\mathbf{J}^H(\mathbf{R}^{-T} \otimes \mathbf{R}^{-1})\mathbf{J}$$

where

$$\mathbf{J} = \frac{\partial \mathrm{vec}(\mathbf{R})}{\partial \boldsymbol{\theta}^T}.$$

Suppose now that $\mathbf{F}$ is singular, then for identifiability we need to pose additional constraints on $\boldsymbol{\theta}$, say $\mathbf{h}(\boldsymbol{\theta}_0) = \mathbf{0}$, where $\mathbf{h}(\boldsymbol{\theta})$ is a vector of functions. Let the Jacobian of $\mathbf{h}(\boldsymbol{\theta})$ be given by

$$\mathbf{H}(\boldsymbol{\theta}) := \frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}.$$

The constrained CRB, $\mathbf{\Gamma}$, is then given by [31]

$$\mathbf{\Gamma} = \mathbf{U}\left(\mathbf{U}^H\mathbf{F}(\boldsymbol{\theta}_0)\mathbf{U}\right)^{-1}\mathbf{U}^H \tag{26}$$

where $\mathbf{U}$ is a semi-unitary matrix, and the columns of $\mathbf{U}$ form an orthonormal basis for the null-space of $\mathbf{H}$ such that $\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{U}(\boldsymbol{\theta}_0) = \mathbf{0}$. The constraints should be chosen such that $\mathbf{U}^H\mathbf{F}(\boldsymbol{\theta}_0)\mathbf{U}$ is invertible.

We will now apply these results to our situation. First we consider only a single snapshot $\mathbf{R}_i$ with model $\mathbf{R}_i = \mathbf{A}_i\mathbf{A}_i + \mathbf{\Psi}$, with $\mathbf{\Psi} = \mathrm{bdiag}(\mathbf{\Psi}_{00}, \mathbf{\Sigma}_1)$ as given by (11).

The unconstrained Fisher information $\mathbf{F}_i$ is singular, because the model is invariant with respect to a multiplication of the matrix $\mathbf{A}_i$ with a unitary matrix at the right, including a phase change for each column. For identifiability, we need to pose q^2 constraints on the matrix $\mathbf{A}_i$. Without loss of generality we choose $\mathbf{A}_i^H\mathbf{A}_i$ to be diagonal (which poses $q(q-1)$ real-valued constraints), and $\mathrm{diag}(\mathbf{A}_i^T\mathbf{A}_i)$ to be real (which poses another q constraints).

In order to find the Fisher information we will parametrize $\mathrm{vec}(\mathbf{\Psi})$ as

$$\mathrm{vec}(\mathbf{\Psi}) = \mathbf{S}_U\boldsymbol{\psi} + \mathbf{S}_L\bar{\boldsymbol{\psi}} + (\mathbf{I}_p \circ \mathbf{I}_p)\mathbf{d},$$

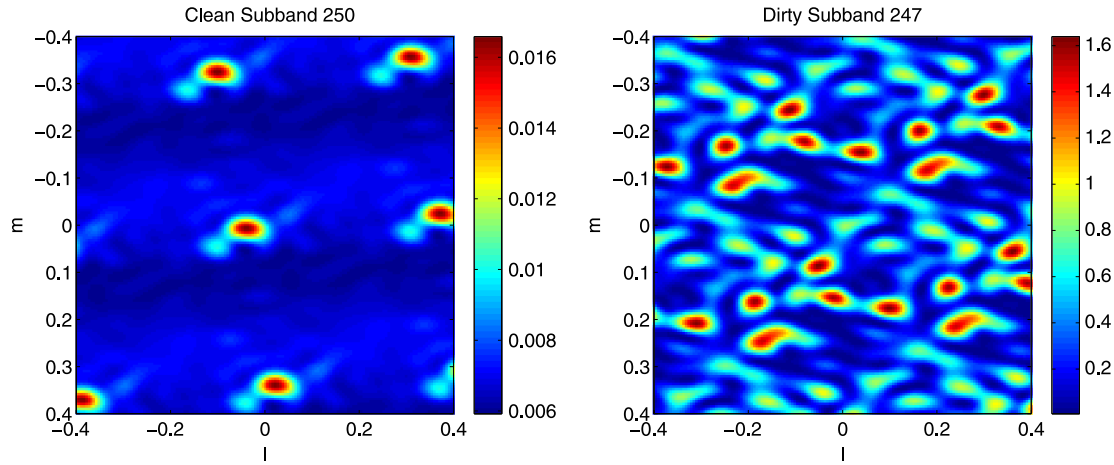

Fig. 9. (a) Clean subband 250 and (b) Contaminated subband 247.

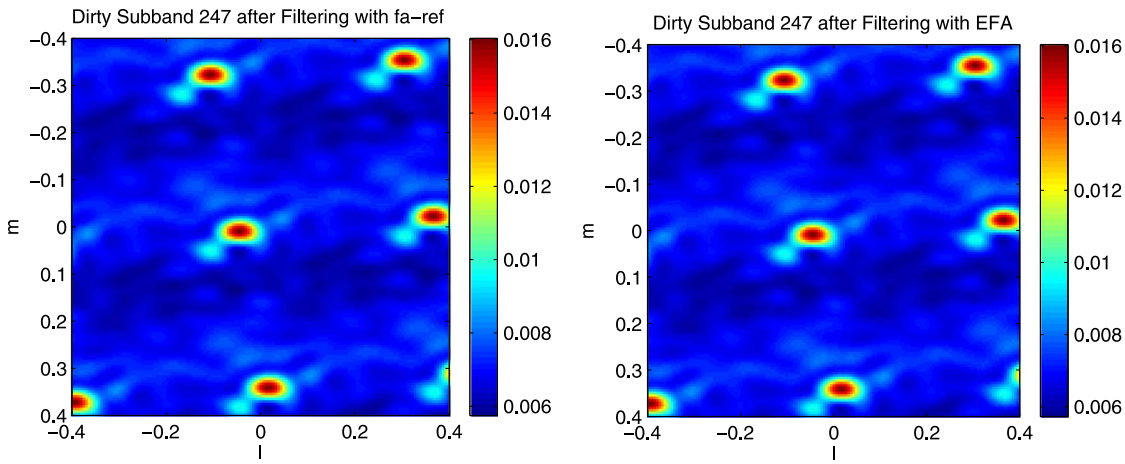

Fig. 10. (a) Result of filtering using 'fa-ref' and (b) result of filtering using EFA.

where $\mathbf{S}_L$ and $\mathbf{S}_U$ are suitable selection matrices, the entries of the strictly upper-triangular part of $\mathbf{\Psi}_{00}$ are stacked into the vector $\boldsymbol{\psi}$, its diagonal entries into $\mathbf{d}_0$, and $\mathbf{d} = [\mathbf{d}_0^T, \boldsymbol{\sigma}_1^T]^T$. The unknown complex parameters are stacked into a vector $\boldsymbol{\theta}_i$,

$$\boldsymbol{\theta}_i = \begin{bmatrix} \boldsymbol{\theta}_{\mathbf{A}_i} \\ \boldsymbol{\theta}_{\mathbf{\Psi}} \end{bmatrix} \tag{27}$$

where

$$\boldsymbol{\theta}_{\mathbf{A}_i} = \begin{bmatrix} \mathrm{vec}(\mathbf{A}_i) \\ \mathrm{vec}(\bar{\mathbf{A}}_i) \end{bmatrix}, \quad \boldsymbol{\theta}_{\mathbf{\Psi}} = \begin{bmatrix} \boldsymbol{\psi} \\ \bar{\boldsymbol{\psi}} \\ \mathbf{d} \end{bmatrix}. \tag{28}$$

Using this parametrization we can partition the Fisher information matrix as

$$\mathbf{F}_i = M \begin{bmatrix} \mathbf{F}_{\mathbf{A}_i\mathbf{A}_i} & \mathbf{F}_{\mathbf{A}_i\mathbf{\Psi}} \\ \mathbf{F}_{\mathbf{A}_i\mathbf{\Psi}}^H & \mathbf{F}_{i,\mathbf{\Psi}\mathbf{\Psi}} \end{bmatrix}.$$

To find these submatrices, we partition the corresponding Jacobian $\mathbf{J}_i$, conforming to the partitioning of $\boldsymbol{\theta}_i$, such that $\mathbf{J}_i = [\mathbf{J}_{\mathbf{A}_i}, \mathbf{J}_{\bar{\mathbf{A}}_i}, \mathbf{J}_{\boldsymbol{\psi}}, \mathbf{J}_{\bar{\boldsymbol{\psi}}}, \mathbf{J}_{\mathbf{d}}]$. Using Wirtinger derivatives we find

$$\begin{aligned} \mathbf{J}_{\mathbf{A}_i} &= (\bar{\mathbf{A}}_i \otimes \mathbf{I}_p) \\ \mathbf{J}_{\bar{\mathbf{A}}_i} &= (\mathbf{I}_p \otimes \mathbf{A}_i)\mathbf{K}_{p,q} \\ \mathbf{J}_{\boldsymbol{\psi}} &= \mathbf{S}_U \\ \mathbf{J}_{\bar{\boldsymbol{\psi}}} &= \mathbf{S}_L \\ \mathbf{J}_{\mathbf{d}} &= (\mathbf{I}_p \circ \mathbf{I}_p) \end{aligned}$$

where $\mathbf{K}_{p,q}$ is a permutation matrix such that $\mathrm{vec}(\mathbf{X}^T) = \mathbf{K}_{p,q}\mathrm{vec}(\mathbf{X})$ for $\mathbf{X}$ a $p \times q$ matrix. We also used the relation

$$\begin{aligned} \mathrm{vec}(\mathbf{R}_i) &= (\bar{\mathbf{A}}_i \otimes \mathbf{I}_p)\mathrm{vec}(\mathbf{A}_i) + \mathrm{vec}(\mathbf{\Psi}) \\ &= (\mathbf{I}_p \otimes \mathbf{A}_i)\mathbf{K}_{p,q}\mathrm{vec}(\bar{\mathbf{A}}_i) + \mathrm{vec}(\mathbf{\Psi}). \end{aligned}$$

To write the constraints on $\mathbf{A}_i$ as a function $\mathbf{h}(\boldsymbol{\theta}_i) = \mathbf{0}$, let $\mathbf{E}_1 = (\mathbf{I}_q \circ \mathbf{I}_q)^T$ and let $\mathbf{E}_2$ be a complementary set of rows such that $[\mathbf{E}_1^T, \mathbf{E}_2^T]^T$ is a permutation matrix. Then $\mathbf{E}_1\mathrm{vec}(\mathbf{A}_i^T\mathbf{A}_i)$ selects the diagonal elements of $\mathbf{A}_i^T\mathbf{A}_i$, and $\mathbf{E}_2\mathrm{vec}(\mathbf{A}_i^H\mathbf{A}_i)$ selects the off-diagonal entries of $\mathbf{A}_i^H\mathbf{A}_i$. The constraint function $\mathbf{h} = [\mathbf{h}_1^T, \mathbf{h}_2^T]^T$ is then

$$\mathbf{h}(\boldsymbol{\theta}_i) = \begin{bmatrix} \mathbf{E}_1\mathrm{vec}\left(\mathbf{A}_i^T\mathbf{A}_i - \mathbf{A}_i^H\bar{\mathbf{A}}_i\right) \\ \mathbf{E}_2\mathrm{vec}(\mathbf{A}_i^H\mathbf{A}_i) \end{bmatrix} = \mathbf{0}$$

and its Jacobian $\mathbf{H}(\boldsymbol{\theta}_i)$ is

$$\mathbf{H}(\boldsymbol{\theta}_i) = \begin{bmatrix} \mathbf{H}_{\mathbf{A}_i} & \mathbf{H}_{\bar{\mathbf{A}}_i} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

where

$$\mathbf{H}_{\mathbf{A}_i} = \begin{bmatrix} \mathbf{E}_1 \left[\left(\mathbf{I}_q \otimes \mathbf{A}_i^T \right) + \left(\mathbf{A}^T \otimes \mathbf{I}_q \right) \mathbf{K}_{p,q} \right] \\ \mathbf{E}_2 \left(\mathbf{I}_q \otimes \mathbf{A}_i^H \right) \end{bmatrix},$$
$$\mathbf{H}_{\bar{\mathbf{A}}_i} = \begin{bmatrix} -\mathbf{E}_1 \left[\left(\mathbf{I}_q \otimes \mathbf{A}_i^H \right) + \left(\mathbf{A}_i^H \otimes \mathbf{I}_q \right) \mathbf{K}_{p,q} \right] \\ \mathbf{E}_2 \left(\mathbf{A}_i^T \otimes \mathbf{I}_p \right) \mathbf{K}_{p,q}. \end{bmatrix}$$

and the trailing zeros correspond to derivatives of $\mathbf{h}(\boldsymbol{\theta}_i)$ with respect to $\boldsymbol{\psi}$, $\bar{\boldsymbol{\psi}}$ and $\mathbf{d}$.

Using QR or SVD on $\mathbf{H}(\boldsymbol{\theta}_i)$, we can find a basis $\mathbf{U}_i$ for the null-space of $\mathbf{H}(\boldsymbol{\theta}_i)$, and calculate the Constrained CRB $\boldsymbol{\Gamma}_i$ for a single snapshot.

As showned above, the constraint function $\mathbf{h}(\boldsymbol{\theta}_i)$ depends only on $\mathbf{A}_i$ and its derivatives with respect to $\boldsymbol{\theta}_{\boldsymbol{\Psi}}$ are zero. This allows us to partition $\mathbf{U}_i$ as

$$\mathbf{U}_i = \begin{bmatrix} \mathbf{U}_{\mathbf{A}_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Using this partitioning and (26) we have

$$\begin{aligned} \boldsymbol{\Gamma}_i &= \begin{bmatrix} \boldsymbol{\Gamma}_{\mathbf{A}_i\mathbf{A}_i} & \boldsymbol{\Gamma}_{\mathbf{A}_i\boldsymbol{\Psi}} \\ \boldsymbol{\Gamma}_{\mathbf{A}_i\boldsymbol{\Psi}}^H & \boldsymbol{\Gamma}_{i,\boldsymbol{\Psi}\boldsymbol{\Psi}} \end{bmatrix} \\ &= \frac{1}{M} \mathbf{U}_i \begin{bmatrix} \mathbf{U}_{\mathbf{A}_i}^H \mathbf{F}_{\mathbf{A}_i\mathbf{A}_i} \mathbf{U}_{\mathbf{A}_i} & \mathbf{U}_{\mathbf{A}_i}^H \mathbf{F}_{\mathbf{A}_i\boldsymbol{\Psi}} \\ \mathbf{F}_{\mathbf{A}_i\boldsymbol{\Psi}}^H \mathbf{U}_{\mathbf{A}_i} & \mathbf{F}_{i,\boldsymbol{\Psi}\boldsymbol{\Psi}} \end{bmatrix}^{-1} \mathbf{U}_i^H \end{aligned} \quad (29)$$

and hence

$$\begin{aligned} \boldsymbol{\Gamma}_{i,\boldsymbol{\Psi}\boldsymbol{\Psi}} = \frac{1}{M} \big[&\mathbf{F}_{i,\boldsymbol{\Psi}\boldsymbol{\Psi}} \\ &- \mathbf{F}_{\mathbf{A}_i\boldsymbol{\Psi}}^H \mathbf{U}_{\mathbf{A}_i} \left(\mathbf{U}_{\mathbf{A}_i}^H \mathbf{F}_{\mathbf{A}_i\mathbf{A}_i} \mathbf{U}_{\mathbf{A}_i} \right)^{-1} \mathbf{U}_{\mathbf{A}_i}^H \mathbf{F}_{\mathbf{A}_i\boldsymbol{\Psi}} \Big]^{-1}. \end{aligned} \quad (30)$$

The CRB for $\hat{\boldsymbol{\Psi}}_{00}$ is the sub-matrix of $\boldsymbol{\Gamma}_{i,\boldsymbol{\Psi}\boldsymbol{\Psi}}$ with respect to $\boldsymbol{\psi}$ and $\mathbf{d}_0$, which we will denote by $\boldsymbol{\Gamma}_{00,i}$.

Now that we have the Fisher Information and CRB for each snapshot we can use this to find the CRB of the entire dataset. Because the time samples are independent, the log-likelihood of the entire dataset then becomes

$$l(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N) = \sum_{i=1}^{N} l_i(\mathbf{X}_i) \quad (31)$$

where $\mathbf{X}_i$ is a $p \times M$ data matrix for each snapshot. To find the gradients of this log-likelihood, we first take derivatives with respect to $\boldsymbol{\theta}_{\mathbf{A}_i}$. If we assume that all $\mathbf{A}_i$ are independent, so that $\partial l_j(\mathbf{X}_j)/\partial \mathbf{A}_i = \mathbf{0}$ for $j \neq i$, we find

$$\frac{\partial l(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)}{\partial \boldsymbol{\theta}_{\mathbf{A}_i}} = \frac{\partial l_i(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_{\mathbf{A}_i}} \quad (32)$$

which is the same as for estimating the ML separately. However, for $\boldsymbol{\Psi}$ we have

$$\frac{\partial l(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)}{\partial \boldsymbol{\theta}_{\boldsymbol{\Psi}}} = \sum_{i=1}^{N} \frac{\partial l_i(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_{\boldsymbol{\Psi}}}. \quad (33)$$

Now we can write the Fisher information for the entire dataset as

$$\mathbf{F}_{total} = M \begin{bmatrix} \mathbf{F}_{\mathbf{A}_1\mathbf{A}_1} & \mathbf{0} & \cdots & \mathbf{F}_{\mathbf{A}_1\boldsymbol{\Psi}} \\ \mathbf{0} & \mathbf{F}_{\mathbf{A}_2\mathbf{A}_2} & \cdots & \mathbf{F}_{\mathbf{A}_2\boldsymbol{\Psi}} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{F}_{\mathbf{A}_1\boldsymbol{\Psi}}^H & \mathbf{F}_{\mathbf{A}_2\boldsymbol{\Psi}}^H & \cdots & \sum_{i=1}^{N} \mathbf{F}_{i,\boldsymbol{\Psi}\boldsymbol{\Psi}} \end{bmatrix}. \quad (34)$$

where the submatrices follow from the previous results as

$$\mathbf{F}_{\mathbf{A}_i\mathbf{A}_i} = \begin{bmatrix} \mathbf{J}_{\mathbf{A}_i}^H \\ \mathbf{J}_{\bar{\mathbf{A}}_i}^H \end{bmatrix} \left(\mathbf{R}_i^{-T} \otimes \mathbf{R}_i^{-1} \right) \begin{bmatrix} \mathbf{J}_{\mathbf{A}_i} & \mathbf{J}_{\bar{\mathbf{A}}_i} \end{bmatrix} \quad (35)$$

$$\mathbf{F}_{\mathbf{A}_i\boldsymbol{\Psi}} = \begin{bmatrix} \mathbf{J}_{\mathbf{A}_i}^H \\ \mathbf{J}_{\bar{\mathbf{A}}_i}^H \end{bmatrix} \left(\mathbf{R}_i^{-T} \otimes \mathbf{R}_i^{-1} \right) \mathbf{J}_{\boldsymbol{\Psi}} \quad (36)$$

$$\mathbf{F}_{\boldsymbol{\Psi}\boldsymbol{\Psi}} = \mathbf{J}_{\boldsymbol{\Psi}}^H \left[\sum_{i=1}^{N} \left(\mathbf{R}_i^{-T} \otimes \mathbf{R}_i^{-1} \right) \right] \mathbf{J}_{\boldsymbol{\Psi}} \quad (37)$$

where

$$\mathbf{J}_{\boldsymbol{\Psi}} = \begin{bmatrix} \mathbf{S}_U & \mathbf{S}_L & (\mathbf{I}_p \circ \mathbf{I}_p) \end{bmatrix}.$$

Because the constraint matrix is only a function of $\boldsymbol{\theta}_{\mathbf{A}_i}$, the constraint matrix for the entire dataset becomes

$$\mathbf{H}_{total} = \begin{bmatrix} \mathbf{H}_{\mathbf{A}_1} & \mathbf{0} & \cdots & \mathbf{0} & & & \\ \mathbf{0} & \mathbf{H}_{\mathbf{A}_2} & \cdots & \vdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & & & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}_{\mathbf{A}_N} & & & \end{bmatrix}$$

which is very sparse. We can use QR decomposition algorithms to find a unitary basis, $\mathbf{U}_{total}$ for its null space efficiently.

The constraint CRB can now be found, same as before, using $\mathbf{F}_{total}$ and $\mathbf{U}_{total}$. Using the matrix inversion lemma on the final result we obtain the expression for $\boldsymbol{\Gamma}_{\boldsymbol{\Psi}\boldsymbol{\Psi}}$ as

$$\begin{aligned} \boldsymbol{\Gamma}_{\boldsymbol{\Psi}\boldsymbol{\Psi}} = \frac{1}{M} \Bigg[&\sum_i \mathbf{F}_{i,\boldsymbol{\Psi}\boldsymbol{\Psi}} \\ &- \sum_i \mathbf{F}_{\mathbf{A}_i\boldsymbol{\Psi}}^H \mathbf{U}_{\mathbf{A}_i} \left(\mathbf{U}_{\mathbf{A}_i}^H \mathbf{F}_{\mathbf{A}_i\mathbf{A}_i} \mathbf{U}_{\mathbf{A}_i} \right)^{-1} \mathbf{U}_{\mathbf{A}_i}^H \mathbf{F}_{\mathbf{A}_i\boldsymbol{\Psi}} \Bigg]^{-1}. \end{aligned} \quad (38)$$

The CRB for $\boldsymbol{\Psi}_{00}$ is the corresponding submatrix of $\boldsymbol{\Gamma}_{\boldsymbol{\Psi}\boldsymbol{\Psi}}$ and it is denoted by $\boldsymbol{\Gamma}_{00}$.

ACKNOWLEDGMENT

The help of W. van Cappellen (ASTRON) in collecting the WSRT data using the Digestif focal plane array, and of S. Wijnholds and M. Norden (both ASTRON) in collecting the LOFAR HBA data is gratefully acknowledged.

REFERENCES

[1] A. van der Veen and A. Boonstra, "Spatial filtering of RF interference in radio astronomy using a reference antenna," in *Proc. IEEE ICASSP*, Montreal, QC, Candada, May 2004, pp. II.189–II.193.

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

[3] P. Fridman and W. Baan, "RFI mitigation methods in radio astronomy," *Astron. Astrophys.*, vol. 378, pp. 327–344, 2001.

[4] A. Boonstra, "Radio frequency interference mitigation in radio astronomy," Ph.D. dissertation, EEMCS Dept., TU Delft, Delft, The Netherlands, Jun. 2005, 90-805434-3-8.

[5] A. van der Veen, A. Leshem, and A. Boonstra, "Array signal processing for radio astronomy," in *The Square Kilometre Array: An Engineering Perspective*, P. Hall, Ed. Dordrecht, The Netherlands: Springer, 2005, pp. 231–249, 1-4020-3797-x.
[6] A. Boonstra and S. van der Tol, "Spatial filtering of interfering signals at the initial low frequency array (Lofar) phased array test station," *Radio Sci.*, vol. 40, no. RS5S09, 2005, 10.1029/2004RS003135.
[7] B. Guner, J. Johnson, and N. Niamsuwan, "Time and frequency blanking for radio-frequency interference mitigation in microwave radiometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3672–3679, 2007.
[8] A. Offringa, A. de Bruyn, M. Biehl, S. Zaroubi, G. Bernardi, and V. Pandey, "Post-correlation radio frequency interference classification methods," *Month. Notices Roy. Astron. Soc.*, vol. 405, no. 1, pp. 155–167, 2010.
[9] G. Nita, D. Gary, Z. Liu, G. Hurford, and S. White, "Radio frequency interference excision using spectral-domain statistics," *Publications Astron. Soc. Pac.*, vol. 119, no. 857, pp. 805–827, 2007.
[10] P. Fridman, "RFI excision using a higher order statistics analysis of the power spectrum," *Astron. Astrophys.*, vol. 368, pp. 369–376, 2001.
[11] P. Fridman, "Statistically stable estimates of variance in radio-astronomy observations as tools for radio-frequency interference mitigation," *Astron. J.*, vol. 135, no. 5, p. 1810, 2008.
[12] C. Barnbaum and R. Bradley, "A new approach to interference excision in radio astronomy: Real time adaptive cancellation," *Astron. J.*, vol. 115, pp. 2598–2614, 1998.
[13] S. Ellingson, J. Bunton, and J. Bell, H. R. Butcher, Ed., "Cancellation of GLONASS signals from radio astronomy data," *Proc. SPIE*, vol. 4015, pp. 400–407, Mar. 2000.
[14] L. Li and B. Jeffs, "Analysis of adaptive array algorithm performance for satellite interference cancellation in radio astronomy," in *URSI Gen. Assem.*, Maastricht, The Netherlands, Aug. 2002.
[15] C. Hansen, K. Warnick, B. Jeffs, J. Fisher, and R. Bradley, "Interference mitigation using a focal plane array," *Radio Sci.*, vol. 40, no. 5, 2005.
[16] J. Raza, A. Boonstra, and A. van der Veen, "Spatial filtering of RF interference in radio astronomy," *IEEE Signal Process. Lett.*, vol. 9, no. 2, pp. 64–67, Mar. 2002.
[17] S. van der Tol and A.-J. van der Veen, "Performance analysis of spatial filtering of RF interference in radio astronomy," *IEEE Trans. Signal Process.*, vol. 53, pp. 896–910, Mar. 2005.
[18] A. Leshem and A.-J. van der Veen, "Radio-astronomical imaging in the presence of strong radio-intrerference," *IEEE Trans. Inf. Theory*, pp. 1730–1747, Aug. 2000.
[19] F. Briggs, J. Bell, and M. Kesteven, "Removing radio interference from contaminated astronomical spectra using an independent reference signal and closure relations," *Astron. J.*, vol. 120, pp. 3351–3361, 2000.
[20] J. Kocz, F. Briggs, and J. Reynolds, "Radio frequency interference removal through the application of spatial filtering techniques on the Parkes multibeam receiver," *Astron. J.*, vol. 140, no. 6, p. 2086, 2010.
[21] B. Jeffs, K. Warnick, and L. Li, "Improved interference cancellation in synthesis array radio astronomy using auxiliary antennas," presented at the IEEE ICASSP, Hong Kong, Apr. 2003.
[22] B. Jeffs, L. Li, and K. Warnick, "Auxiliary antenna-assisted interference mitigation for radio astronomy arrays," *IEEE Trans. Signal Process.*, vol. 53, pp. 439–451, Feb. 2005.
[23] G. Hellbourg, R. Weber, C. Capdessus, and A.-J. Boonstra, "Cyclostationary approaches for spatial RFI mitigation in radio astronomy," *Comptes Rendus Physique*, vol. 13, no. 1, pp. 71–79, 2012.
[24] T. Anderson, *An Introduction to Multivariate Statistics*, ser. Wiley Publications in Statistics. New York, NY, USA: Wiley, 1958.
[25] A. M. Sardarabadi and A.-J. van der Veen, "Subspace estimation using factor analysis," *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, pp. 477–480, 2012.
[26] A. Leshem and A.-J. van der Veen, "Multichannel detection of Gaussian signals with uncalibrated receivers," *IEEE Signal Process. Lett.*, vol. 8, no. 4, pp. 120–122, 2001.
[27] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, pp. 36–47, Jul. 2004.
[28] K. Mardia, J. Kent, and J. Bibby, *Multivariate Analysis*. New York, NY, USA: Academic, 1979.
[29] B. Porat, *Digital processing of random signals-Theory and methods*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1994.
[30] W. Bangs, "Array processing with generalized beamformers," Ph.D. dissertation, Yale Univ., New Haven, CT, USA, 1971.
[31] A. Jagannatham and B. Rao, "Cramer-Rao lower bound for constrained complex parameters," *IEEE Signal Process. Lett.*, vol. 11, Nov. 2004.

Ahmad Mouri Sardarabadi was born in Iran in 1985. He received the M.Sc. degree (*cum laude*) from TU Delft, The Netherlands, in 2011. He is currently a Ph.D. candidate at TU Delft, working on the NWO "Signal Processing for Radio Astronomy" project with Alle-Jan van der Veen. His research interests are signal processing and multivariate analysis, with applications to radio astronomy.

Alle-Jan van der Veen was born in The Netherlands in 1966. He received the Ph.D. degree (*cum laude*) from TU Delft, The Netherlands, in 1993. Throughout 1994, he was a postdoctoral scholar at Stanford University, Stanford, CA. He was Chairman of the IEEE SPS Signal Processing for Communications Technical Committee from 2002 to 2004, Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS From 2002 to 2005, and Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2008. At present, he is a Full Professor in Signal Processing at TU Delft. His research interests are in the general area of system theory applied to signal processing, and in particular algebraic methods for array signal processing, with applications to wireless communications and radio astronomy.

Albert-Jan Boonstra was born in The Netherlands in 1961. He received the M.Sc. degree in applied physics from Groningen University, the Netherlands, in 1987. The following four years he was with the Laboratory for Space Research Groningen, where he conducted optical and cryogenic infrared tests on the short wavelength spectrometer (SWS) for the infrared space observatory satellite (ISO). In 1992, he joined ASTRON, the Netherlands Institute for Radio Astronomy, initially at the Radio Observatory Westerbork, where he coordinated the telescope upgrades in the nineties. He is currently with the ASTRON R&D Department in Dwingeloo where he leads the ASTRON-IBM Dome project. His research interests lie in the area of signal processing, specifically radio frequency interference mitigation by digital filtering. In 1999–2004 he has also been with Delft University of Technology pursuing his Ph.D., which he received in 2005.