

Performance Analysis of Spatial Filtering of RF Interference in Radio Astronomy

Sebastiaan van der Tol and Alle-Jan van der Veen, *Fellow, IEEE*

Abstract—Radio astronomical observations are increasingly contaminated by man-made RF interference (RFI). If these signals are continuously present, then they cannot be removed by the usual techniques of detection and blanking. We have previously proposed a spatial filtering technique, where the impact of the interferer is projected out from the estimated covariance data. Assuming that the spatial signature of the interferer is time-varying, several such estimates can be combined to recover the missing dimensions. In this paper, we give a detailed performance analysis of this algorithm. It is shown that the spatial filter introduces a small increase in variance of the estimates (because of the loss in information) and that the algorithm is unbiased in case the true spatial signatures of the interferers are known but that there may be a bias in case the signatures are estimated from the same data. Some of the bias may be removed, and moreover, the bias only affects the auto-correlations, whereas the astronomical information is mostly in the cross-correlations.

Index Terms—Algorithm performance, eigenvalues, interference cancellation, radio astronomy, RF interference, 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 stopband 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). Most interference cancellation today is done at the post-correlation level by rejecting

suspect correlation products or by specialized imaging algorithms, but these techniques have their limits and interference rejection at shorter time scales is needed.

Depending on the interference and the type of instrument, several kinds of radio frequency interference (RFI) mitigation techniques are applicable. Overviews can be found in [2] and [3], 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 [2]. However, many communication signals are continuous in time. For a single-dish telescope, there are not many other options¹ than to consider an extension by a reference antenna that picks up only the interference, so that adaptive cancellation techniques based on output power minimization can be implemented [5]–[7].

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, e.g., 10 s (in practice, this can range from a fraction of a second up to several minutes, and only the nonredundant entries are computed and retained). 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 a correction is needed to take into account that some dimensions were underrepresented. This algorithm was introduced in [8]—we describe it in more detail in Section II.

The goal of the present paper is to give a detailed performance analysis of this algorithm. It has to be demonstrated that the information the astronomer is looking for is unaffected by the spatial filter—a delicate affair because this information is at least 15 dB below the noise level and typically much more.

In Section III, we derive the bias and standard deviation of the final covariance estimate. We then analyze the effect of the interferer stationarity on the performance of the filter (Section IV). The bias is studied in more detail in Section V, and the theoretical performance equations are verified with simulations. We also show that much of the bias can be corrected. Conclusions are in Section VI.

Notation: An overbar $\bar{\cdot}$ denotes complex conjugate, superscript T denotes matrix transpose, and H denotes complex conjugate transpose. $\text{vec}(\cdot)$ denotes the stacking of the columns of a matrix in a vector, \otimes the Kronecker product, \circ the column-wise Kronecker product (Khatri–Rao product), and \odot the entrywise multiplication of two matrices of equal size. \mathbf{I} is the identity

Manuscript received August 18, 2003; revised April 2, 2004. This work was supported in part by the NOEMI project under Contract STW DEL77-476 and by VICI-SPCOM under Contract STW DTC.5893. Parts of this paper were published in [1]. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Constantinos B. Papadias.

The authors are with the Department of Electrical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: svdt@cas.et.tudelft.nl; allejan@cas.et.tudelft.nl).

Digital Object Identifier 10.1109/TSP.2004.842177

¹An exception is a technique based on higher order statistics [4].

matrix, and $\mathbf{1}$ is a vector with all ones. The covariance of an estimated matrix is defined as

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

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

II. DATA MODEL AND ALGORITHM SUMMARY

A. Received Data Model

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

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

where $\mathbf{x}(t) = [x_1(t), \dots, x_p(t)]^T$ is the $p \times 1$ vector of telescope signals at time t , $\mathbf{v}(t)$ is the received sky signal possibly due to many astronomical sources, assumed on the time scale of 10 s to be a stationary Gaussian vector with covariance matrix \mathbf{R}_v (the astronomical “visibilities”), and $\mathbf{n}(t)$ is the $p \times 1$ noise vector with independent identically distributed Gaussian entries and covariance matrix $\sigma^2 \mathbf{I}$ (this implies accurate calibration). The astronomer is interested in the nonredundant off-diagonal entries of \mathbf{R}_v .

If an interferer is present the output vector is modeled as

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

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

We make the following additional assumptions on this model:

- A1) The noise variance σ^2 is known from calibration, e.g., from observations of nearby uncontaminated frequencies.
- A2) $\mathbf{R}_v \ll \sigma^2 \mathbf{I}$, so that $\mathbf{R}_v + \sigma^2 \mathbf{I} \approx \sigma^2 \mathbf{I}$. This is reasonable as even the strongest sky sources are about 15 dB under the noise floor. This condition will be made more precise in (6).
- A3) The processing bandwidth is sufficiently narrow, meaning that the maximal propagation delay of a signal across the telescope array is small compared to the inverse bandwidth so that this delay can be represented by a phase shift of the signal. For a maximal baseline D , the condition on the maximal bandwidth is $\Delta f \ll c/(\pi D)$, where c is the speed of light, e.g., for a maximal baseline of $D = 3000$ m, $\Delta f \ll 32$ kHz. Without this assumption, it does not make sense to define an interferer signature vector $\mathbf{a}(t)$. If the assumption is not satisfied, as for many existing telescopes, a form of subband processing has to be implemented.
- A4) The interferer signature $\mathbf{a}(t)$ is stationary over short processing times (say 10 ms).

- A5) $\mathbf{a}(t)$ is varying over periods longer than the short-term integration time. Note that even interferers fixed on earth will appear to move as the earth rotates and the telescopes track in the opposite direction. This effect is worked out in Section IV-B.

This was the model considered in [8]. The model is easily extended to multiple narrowband interfering sources, in which case, we obtain

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

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

B. Covariance Model

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

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

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

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

$$\mathbf{R}_k = \Psi + \mathbf{A}_k \mathbf{A}_k^H = \mathbf{R}_v + \sigma^2 \mathbf{I} + \mathbf{A}_k \mathbf{A}_k^H \quad (1)$$

where $\Psi = \mathbf{R}_v + \sigma^2 \mathbf{I}$ is the interference-free covariance matrix, and σ^2 is assumed to be known from calibration. The objective is to estimate the interference-free covariance Ψ .

C. Spatial Filtering Using Projections

In [8], a spatial filtering algorithm based on projections was introduced. In summary, it is as follows.

Momentarily suppose that an orthogonal basis \mathbf{U}_k of the subspace spanned by interferer spatial signatures $\text{span}(\mathbf{A}_k)$ is known. We can then form a spatial projection matrix \mathbf{P}_k

$$\mathbf{P}_k := \mathbf{I} - \mathbf{U}_k \mathbf{U}_k^H \quad (2)$$

which is such that $\mathbf{P}_k \mathbf{A}_k = 0$. When this spatial filter is applied to the data covariance matrix, all the energy due to the interferer will be nulled. Indeed, let

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

then

$$\text{E}[\hat{\mathbf{Q}}_k] = \mathbf{P}_k \Psi \mathbf{P}_k.$$

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

$$\hat{\mathbf{Q}} := \frac{1}{N} \sum_{k=1}^N \hat{\mathbf{Q}}_k = \frac{1}{N} \sum_{k=1}^N \mathbf{P}_k \hat{\mathbf{R}}_k \mathbf{P}_k. \quad (3)$$

$\hat{\mathbf{Q}}$ is an estimate of Ψ , 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 $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$. This gives

$$\text{vec}(\hat{\mathbf{Q}}) = \frac{1}{N} \sum_{k=1}^N \mathbf{C}_k \text{vec}(\hat{\mathbf{R}}_k) \quad (4)$$

where

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

If the interference was completely removed, then

$$\mathbb{E} [\text{vec}(\hat{\mathbf{Q}})] = \frac{1}{N} \sum_{k=1}^N \mathbf{C}_k \text{vec}(\Psi) = \mathbf{C} \text{vec}(\Psi) \quad (5)$$

where

$$\mathbf{C} := \frac{1}{N} \sum_{k=1}^N \mathbf{C}_k.$$

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

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

If the interference was completely projected out, then $\hat{\Psi}$ is an unbiased estimate of the covariance matrix without interference. This was the algorithm introduced in [8].

In practice, $\text{span}(\mathbf{A}_k)$ is not known, and it has to be estimated. As usual in array processing, this is done by computing an eigenvalue decomposition of the sample covariance matrix $\hat{\mathbf{R}}_k$ and letting $\hat{\mathbf{U}}_k$ in (2) contain the dominant eigenvectors. The underlying assumption is that, without interference, $\hat{\mathbf{R}}_k \approx \sigma^2 \mathbf{I}$, so that eigenvalues significantly larger than σ^2 indicate the presence of interferers, and at the same time, the corresponding eigenvectors span the subspace to be projected out. For this to work, it is essential that the astronomical contribution \mathbf{R}_v is small compared with the noise $\sigma^2 \mathbf{I}$ [viz. assumption A2)]; otherwise, it would disturb the eigenvalues and the eigenvector directions. This assumption can be made more precise by requiring that the entries of \mathbf{R}_v are much smaller than the standard deviation of the entries of $\hat{\mathbf{R}}_k$ due to finite sample noise, or

$$|(\mathbf{R}_v)_{ij}| \ll \frac{\sigma^2}{\sqrt{M}}. \quad (6)$$

For a known signal-to-noise ratio (SNR), this translates into a limit on the short-term integration length M , e.g., the strongest astronomical source has SNR ≈ -15 dB, for which we obtain $M \ll 1000$. Usually, the value of M is limited to a lower value by the (non-)stationarity of the interferer.

A detection metric for the presence of interferers, and their number can be derived from statistical principles (e.g., the classical sequential hypothesis tests [9]–[11], the AIC [12], or Rismanen's MDL test [13]; see [14, ch.7.8] for additional references). A test which is simpler to implement is that which puts a threshold on the eigenvalues

$$\gamma = \sigma^2 \left(1 + \frac{\sqrt{p}}{\sqrt{M}} \right)^2 \quad (7)$$

(this equation is based on an asymptotic formula for the largest singular value of a $p \times M$ white Gaussian noise matrix by Edelman [15]). The number of eigenvalues larger than γ indicates the number of interferers. The eigenvalue threshold test is used here because it is easier to relate it to M and to the interference-to-noise ratio (INR), which gives a useful insight. The test does not have a known false-alarm rate, but for the threshold in (7), it is about 2% over a wide range of p and M .

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 that are projected out are sufficiently varying. In [1], it was noted that for $q = 1$, usually, three different projections are already sufficient to guarantee that \mathbf{C} is full rank.

III. SIGNAL COVARIANCE ESTIMATE ERROR ANALYSIS

The result of the algorithm is $\hat{\Psi}$, which is an estimate of the true covariance matrix Ψ . As a measure of accuracy, we will determine the covariance of $\hat{\Psi}$ in three cases: *i*) interference free, *ii*) the spatial signatures \mathbf{A}_k are known, and *iii*) the spatial signatures are estimated from the data.

The covariance of the short-term averages is defined by

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

where

$$\begin{aligned} \mathbf{R}'_k &= \hat{\mathbf{R}}_k - \mathbb{E}[\hat{\mathbf{R}}_k] \\ \mathbb{E}[\hat{\mathbf{R}}_k] &= \mathbf{R}_k = \Psi + \mathbf{A}_k \mathbf{A}_k^H. \end{aligned}$$

Assuming Gaussian sources (sky signals, interference and noise), the covariance of the short-term averages is given by the standard result (e.g., [16, eq. (A4)])

$$\text{cov}\{\hat{\mathbf{R}}_k\} = \frac{1}{M} \bar{\mathbf{R}}_k \otimes \mathbf{R}_k. \quad (8)$$

A. Case I: Interference-Free

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

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{k=1}^N \hat{\mathbf{R}}_k.$$

Then, from (8), the covariance of $\hat{\mathbf{R}}$ is given by

$$\text{cov}\{\hat{\mathbf{R}}\} = \frac{1}{MN} \bar{\mathbf{R}} \otimes \mathbf{R}$$

where $\mathbf{R} = \Psi = \mathbf{R}_v + \sigma^2 \mathbf{I}$. Because $\mathbf{R}_v \ll \sigma^2 \mathbf{I}$

$$\text{cov}\{\hat{\Psi}\} \approx \frac{\sigma^4}{MN} \mathbf{I}. \quad (9)$$

Hence, all entries of Ψ are disturbed by uncorrelated noise with variance $\sigma^4/(MN)$. This interference-free result gives a reference performance for the estimation of Ψ .

B. Case II: Interference With Known Spatial Signatures

Suppose the subspace spanned by the spatial signatures \mathbf{A}_k of the interferers is deterministic and perfectly known. In that case, the interference will be completely removed, and the algorithms are unbiased by design. According to the algorithm in Section II-C, the estimate of Ψ is given by

$$\text{vec}(\hat{\Psi}) = \mathbf{C}^{-1} \text{vec}(\hat{\mathbf{Q}}), \quad \text{vec}(\hat{\mathbf{Q}}) = \frac{1}{N} \sum_{k=1}^N \mathbf{C}_k \text{vec}(\hat{\mathbf{R}}_k) \quad (10)$$

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

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

where

$$\text{cov}\{\hat{\mathbf{Q}}\} = \frac{1}{N^2} \sum_{k=1}^N \sum_{\ell=1}^N \mathbf{C}_k \mathbf{E} \left[\text{vec}(\mathbf{R}'_k) \text{vec}(\mathbf{R}'_\ell)^H \right] \mathbf{C}_\ell \quad (11)$$

and $\mathbf{R}'_k := \hat{\mathbf{R}}_k - \mathbf{R}_k$. The estimation errors \mathbf{R}'_k and \mathbf{R}'_ℓ are uncorrelated for $k \neq \ell$, and $\mathbf{E}(\hat{\mathbf{R}}_k) = \mathbf{R}_k = \Psi + \mathbf{A}_k \mathbf{A}_k^H$; hence

$$\begin{aligned} \text{cov}\{\hat{\mathbf{Q}}\} &= \frac{1}{N^2} \sum_{k=1}^N \mathbf{C}_k \mathbf{E} \left[\text{vec}(\mathbf{R}'_k) \text{vec}(\mathbf{R}'_k)^H \right] \mathbf{C}_k \\ &= \frac{1}{MN^2} \sum_{k=1}^N \mathbf{C}_k (\bar{\mathbf{R}}_k \otimes \mathbf{R}_k) \mathbf{C}_k. \end{aligned}$$

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

$$\begin{aligned} \text{cov}\{\hat{\mathbf{Q}}\} &= \frac{1}{MN^2} \sum_{k=1}^N \mathbf{C}_k (\bar{\Psi} \otimes \Psi) \mathbf{C}_k \\ &\approx \frac{1}{MN^2} \sum_{k=1}^N \mathbf{C}_k (\sigma^2 \mathbf{I} \otimes \sigma^2 \mathbf{I}) \mathbf{C}_k \\ &= \frac{\sigma^4}{MN^2} \sum_{k=1}^N \mathbf{P}_k^T \otimes \mathbf{P}_k \\ &= \frac{\sigma^4}{MN} \mathbf{C}. \end{aligned}$$

Thus

$$\text{cov}\{\hat{\Psi}\} = \mathbf{C}^{-1} \text{cov}\{\hat{\mathbf{Q}}\} \mathbf{C}^{-1} \approx \frac{\sigma^4}{MN} \mathbf{C}^{-1}. \quad (12)$$

Compared to (9), this indicates that \mathbf{C}^{-1} determines the relative performance of the spatial filtering algorithm of Section II-C.

In turn, \mathbf{C}^{-1} depends on the variability of \mathbf{A}_k , which are the spatial signatures of the interferer. If \mathbf{A}_k is sufficiently varying and the number of antennas p is sufficiently large, then for large N , $\mathbf{C} \rightarrow \mathbf{I}$, and the performance is similar as in the interference-free case [see also (16) later in the paper]. In practice, a moderate N is sufficient. Even for stationary located interferers, \mathbf{A}_k will change because of earth rotation. A more detailed discussion on the conditioning of \mathbf{C} in that case is found in Section IV-B.

C. Case III: Interference With Unknown Deterministic Spatial Signatures

In practice, the spatial signatures \mathbf{A}_k are unknown, and their column span will be estimated from eigendecompositions of the $\hat{\mathbf{R}}_k$. Because of the estimation error, the projection is not perfect, and there will be a residual that might affect the performance. In previous work [1], we have presented a first-order perturbation analysis, which said that the performance is the same as in (12): It is in first order unchanged, even if the projections are estimated. Obviously, this cannot be entirely true. A more detailed second-order analysis presented in this subsection reveals that, with projections estimated from an eigenvalue decomposition, a *bias* is introduced on the main diagonal of $\hat{\Psi}$, which in certain cases can be significant compared with the standard deviation. We also show that for reasonably large M , the covariance of the estimate is as derived before in (12), i.e., not affected by the interference subspace estimation.

In this section, we assume for simplicity that the noise power has been normalized to $\sigma^2 = 1$. The main results of the second-order perturbation analysis are summarized in the following theorem, and proofs are in the Appendices.

Theorem 1: Let $\hat{\Psi} = \Psi + \Psi' + \Psi'' + \dots$ be a Taylor series expansion of $\hat{\Psi}$. Assume that $\mathbf{R}_v \ll \mathbf{I}$. Let $\lambda_{i,k}$ be the i th eigenvalue of \mathbf{R}_k (sorted in descending order) and q_k the number of dimensions that are projected out in the k th interval. Assume that the largest q_k eigenvalues are distinct. Then

$$\begin{aligned} \Psi' &= O\left(\frac{1}{\sqrt{MN}}\right), \quad \mathbf{E}[\Psi'] = 0 \\ \Psi'' &= O\left(\frac{1}{(M\sqrt{N})}\right) \\ \mathbf{E}[\text{vec}(\Psi'')] &= -\mathbf{C}^{-1} \frac{1}{N} \sum_{k=1}^N \frac{1}{M} \\ &\quad \times \left(\sum_{i=1}^{q_k} \frac{\lambda_{i,k}}{\lambda_{i,k} - 1} \right) \text{vec}(\mathbf{P}_k). \quad (13) \end{aligned}$$

The bias on $\hat{\Psi}$ is given by $\mathbf{E}[\Psi'']$. If the eigenvalues of \mathbf{R}_k are stationary (independent of k), the bias is

$$\mathbf{E}[\Psi''] = -\frac{1}{M} \left(\sum_{i=1}^q \frac{\lambda_i}{\lambda_i - 1} \right) \mathbf{I}. \quad (14)$$

For sufficiently large M ($M^2 \gg M$)

$$\text{cov}\{\hat{\Psi}\} \approx \mathbf{E} \left[\text{vec}(\Psi') \text{vec}(\Psi')^H \right] \approx \frac{1}{MN} \mathbf{C}^{-1}.$$

Proof: See Appendix B. \square

Clearly, (13) and (14) indicate that there is a bias term. Under the assumption that the eigenvalues are stationary and the sky signal powers are much below the noise power, (14) also shows that *only* the diagonal entries of $\hat{\Psi}$ are biased, even if the interferer is not uniformly distributed over space (in fact, even if \mathbf{A}_k is almost stationary). This is acceptable for most astronomy applications (e.g., imaging, spectral analysis). The bias is studied in more detail in Section V.

The variance of the estimates is the same as we obtained before in (12). The effect of the projections is described entirely by \mathbf{C}^{-1} . The properties of \mathbf{C}^{-1} are analyzed in the next section.

IV. EXPECTED VALUE OF \mathbf{C}^{-1}

From the covariance expression (12), it is seen that \mathbf{C}^{-1} determines the penalty on the estimation performance due to spatial filtering. The main diagonal of \mathbf{C}^{-1} contains the factors by which the variance of Ψ is multiplied compared with the interference-free case. \mathbf{C} only depends on the projected directions, i.e., on the collection of \mathbf{a}_k , the spatial signatures of the interferer over time. We consider a single interferer and two extreme cases: *I*) the \mathbf{a}_k are independent random vectors, with normally distributed entries, and *II*) the \mathbf{a}_k are the spatial signatures of an interferer fixed to earth.

A. Case I: Interferer With Normally Distributed Spatial Signatures

If a ground-based interferer without direct line-of-sight is moving, then due to fast fading effects, it is reasonable to model the corresponding spatial signature vectors \mathbf{a}_k ($k = 1, \dots, N$) as a collection of i.i.d. random vectors with directions uniformly distributed over the p -dimensional unit sphere.² From this model, we can determine $\mathbb{E}[\mathbf{C}_k]$. For $N \rightarrow \infty$, \mathbf{C} converges to $\mathbb{E}[\mathbf{C}_k]$; hence, \mathbf{C}^{-1} converges to $(\mathbb{E}[\mathbf{C}_k])^{-1}$.

Thus, assuming $\mathbf{a}_k \sim \mathcal{CN}(0, \mathbf{I})$ and i.i.d. for different k , let

$$\mathbf{u}_k = \frac{\mathbf{a}_k}{\|\mathbf{a}_k\|}.$$

Then, \mathbf{u}_k is uniformly distributed over the unit sphere. Note that

$$\begin{aligned} \mathbb{E}[\mathbf{C}] &= \mathbb{E}[\mathbf{C}_k] \\ &= \mathbb{E}[\mathbf{P}_k^T \otimes \mathbf{P}_k] \\ &= \mathbb{E}\left[(\mathbf{I} - \mathbf{u}_k \mathbf{u}_k^H)^T \otimes (\mathbf{I} - \mathbf{u}_k \mathbf{u}_k^H)\right] \\ &= \mathbb{E}[\mathbf{I} \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{u}_k \mathbf{u}_k^H - \bar{\mathbf{u}}_k \bar{\mathbf{u}}_k^H \otimes \mathbf{I} + \bar{\mathbf{u}}_k \bar{\mathbf{u}}_k^H \otimes \mathbf{u}_k \mathbf{u}_k^H]. \end{aligned}$$

To evaluate this expression, we need to establish the second- and fourth-order moments of \mathbf{u}_k . Denote by u_i the i -th entry of \mathbf{u}_k . Due to symmetry, $\mathbb{E}[u_i \bar{u}_j]$ is zero, except when $i = j$. Let $y = u_i \bar{u}_i$; then, y has a beta distribution with parameters $(1, p-1)$ [17, p. 487] for which

$$\mathbb{E}[y] = \frac{1}{p}, \quad \text{var } y = \mathbb{E}\left[|y - \mathbb{E}[y]|^2\right] = \frac{p-1}{p^2(p+1)}.$$

²In other cases, this model may represent a workable simplification which allows for a meaningful analysis.

It follows that

$$\begin{aligned} \mathbb{E}[u_i \bar{u}_i] &= \frac{1}{p} \\ \mathbb{E}[u_i \bar{u}_i u_i \bar{u}_i] &= \frac{2}{p(p+1)} \end{aligned}$$

and subsequently

$$\begin{aligned} \mathbb{E}[u_i \bar{u}_i u_j \bar{u}_j] &= \frac{1}{p-1} \mathbb{E}[u_i \bar{u}_i (1 - u_i \bar{u}_i)] \\ &= \frac{1}{p-1} \left(\frac{1}{p} - \frac{2}{p(p+1)} \right) \\ &= \frac{1}{p(p+1)} \quad (i \neq j) \\ \mathbb{E}[u_i \bar{u}_j u_j \bar{u}_i] &= \frac{1}{p(p+1)} \quad (i \neq j). \end{aligned}$$

All other cases of $\mathbb{E}[u_i \bar{u}_j u_k \bar{u}_l]$ are 0 due to symmetry. In summary

$$\begin{aligned} \mathbb{E}[\mathbf{C}] &= \mathbf{I} - \frac{2}{p} \mathbf{I} + \frac{1}{p(p+1)} (\mathbf{I} + \text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})^T) \\ &= \frac{p^2 - p - 1}{p(p+1)} \left(\mathbf{I} + \frac{1}{p^2 - p - 1} \text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})^T \right). \end{aligned} \quad (15)$$

The inverse is given by

$$(\mathbb{E}[\mathbf{C}])^{-1} = \frac{p(p+1)}{p^2 - p - 1} \left(\mathbf{I} - \frac{1}{p^2 - 1} \text{vec}(\mathbf{I}) \text{vec}(\mathbf{I})^T \right). \quad (16)$$

For large N , \mathbf{C}^{-1} converges to $(\mathbb{E}[\mathbf{C}])^{-1}$. The variance of the entries of $\hat{\Psi}$ is determined by the main diagonal of \mathbf{C}^{-1} . In particular

$$\begin{aligned} \text{var}\{\hat{\Psi}_{ii}\} &\rightarrow \frac{\sigma^4}{MN} \left(1 + \frac{2p}{p^2 - p - 1} \right) + O\left(\frac{1}{p^3}\right) \\ \text{var}\{\hat{\Psi}_{ij}\} &\rightarrow \frac{\sigma^4}{MN} \left(1 + \frac{2p+1}{p^2 - p - 1} \right) \quad (i \neq j) \end{aligned} \quad (17)$$

e.g., with $p = 8$ antennas, the variance of the diagonal entries of $\hat{\Psi}$ is multiplied by approximately 1.29, and the variance of the off-diagonal entries is multiplied by 1.31, or an increase of about 1.1 dB. Alternatively, we can say that if one dimension is always projected out, then 31% more samples are needed to compensate for the loss of information.

B. Case II: Stationary Interferers

For \mathbf{C} to be invertible, the spatial signatures \mathbf{a}_k need to be sufficiently variable. For stationary interferers (not moving, no multipath), the only source of variability is the geometric delay compensation. This is a delay placed between each telescope and the correlator to correct for the different path lengths of the astronomical signal to each of the telescopes: After the correction, signals arriving from the look direction will arrive in-phase and will be processed coherently. The geometric delay $\tau_g(t)$ depends on the locations of the telescopes and the direction of the observed field in the sky. Due to the daily rotation of the earth, the stars are moving along the sky, and hence, the geometric delay is time varying.

For practical reasons, the compensating delay is usually inserted only after frequency downconversion. In this case, the downconversion of an RF signal with carrier frequency f and delay τ_g introduces a phase shift $2\pi f\tau_g$. For coherent processing of the sky signals, it is therefore necessary to apply an additional time-varying phase correction to each telescope signal after downconversion, which is called *fringe stopping* [18], [19]. This correction will make interferers fixed on earth appear to have a telescope-dependent time-varying phase-shift.

For a uniform linear East–West array of telescopes, the geometrical delay is $\tau_g(t) = D_\lambda/f \cdot \cos \delta \sin h(t)$, where D_λ is the longest baseline length in wavelengths, δ is the declination of the sky source ($\delta = 90^\circ$ for a source at the celestial pole), and $h(t) = \omega_e t$ is the “hour angle” (azimuth) of the source ($\omega_e \approx 7.29115 \cdot 10^{-5}$ rad/s is the earth rotation speed based on a sidereal day) [19]. For time scales of a few seconds, the required phase correction can be approximated by a linear-phase progression $2\pi f\tau_g \approx 2\pi f_F t$, where f_F is called the “fringe frequency” [19]

$$f_F = \omega_e D_\lambda \cos \delta \cos h.$$

The effect of the time-varying fringe correction on the spatial signature of an interferer can thus be modeled as

$$\mathbf{a}(t) = \begin{bmatrix} a_1 \\ a_2 e^{\frac{j\phi}{(p-1)}} \\ a_3 e^{\frac{j2\phi}{(p-1)}} \\ \vdots \\ a_p e^{\frac{j(p-1)\phi}{(p-1)}} \end{bmatrix}, \quad \phi(t) = 2\pi f_F t \quad (18)$$

where $[a_1, \dots, a_p]^T$ is the spatial signature vector of the interferer without fringe correction.

For a stationary interferer and sufficiently large N , \mathbf{C}^{-1} depends mostly on the number of antennas p and the total fringe rotation ϕ_{tot} over the longest baseline and over the complete integration period of duration $T_{\text{int}} = MNT_s$

$$\phi_{\text{tot}} = 2\pi T_{\text{int}} f_F = 2\pi T_{\text{int}} \omega_e D_\lambda \cos \delta \cos h. \quad (19)$$

Simulations [see Fig. 1(a) and (b)] show how the asymptotic value of $\max(\text{diag}(\mathbf{C}^{-1}))$ depends on ϕ_{tot} and p . In Fig. 1(a), the initial interferer signature vector is $[a_1, \dots, a_p] = [1, \dots, 1]$. It is seen that a reasonably good performance is already obtained once the total fringe rotation is larger than about two cycles (other simulations indicated that this result is independent of the number of samples per fringe cycle). The asymptotic values actually correspond to the same levels as predicted in (16) for the uniformly distributed case. In Fig. 1(b), $[a_1, \dots, a_p]$ is selected randomly, and confidence intervals for $\max(\text{diag}(\mathbf{C}^{-1}))$ are shown for $p = 8$ telescopes.

This condition on ϕ_{tot} can be translated into a division of the sky in an “observable” and an “unobservable” area (sky locations that cause insufficient fringe rotations to tolerate the projection of stationary interferers). Let α be an angle such that $\sin \alpha = \cos \delta \cos h$. In view of (19), a condition for sufficient fringe rotation corresponds to a certain minimum value of α . Translating this to values for δ and h , it can be shown that this corresponds geometrically to a band from the East horizon over the celestial pole to the West horizon, as is shown in Fig. 1(c),

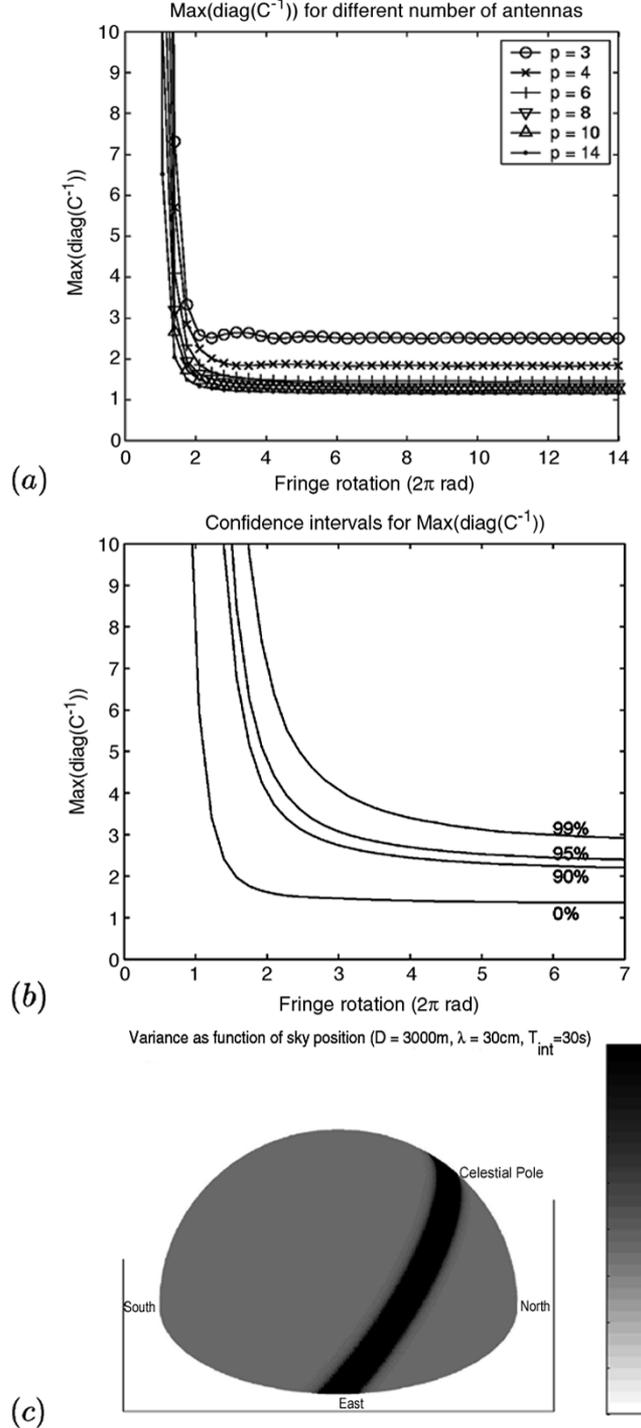


Fig. 1. (a) $\max(\text{diag}(\mathbf{C}^{-1}))$ as function of fringe rotation ϕ_{tot} . (b) Confidence intervals for a random initial signature vector ($p = 8$). (c) Unobservable area of the sky due to a fixed interferer.

which depicts the sky as seen above the observatory in a coordinate system relative to the observatory; the vertical direction corresponds to zenith, and the latitude of the observatory was assumed to be 52° . The black ring is the unobservable area.

Let ϕ_{min} be the smallest value of ϕ_{tot} that is acceptable. The width 2α of the unobservable band is given by

$$2\alpha = 2 \arcsin \frac{\phi_{\text{min}}}{2\pi T_{\text{int}} \omega_e D_\lambda} \quad (20)$$

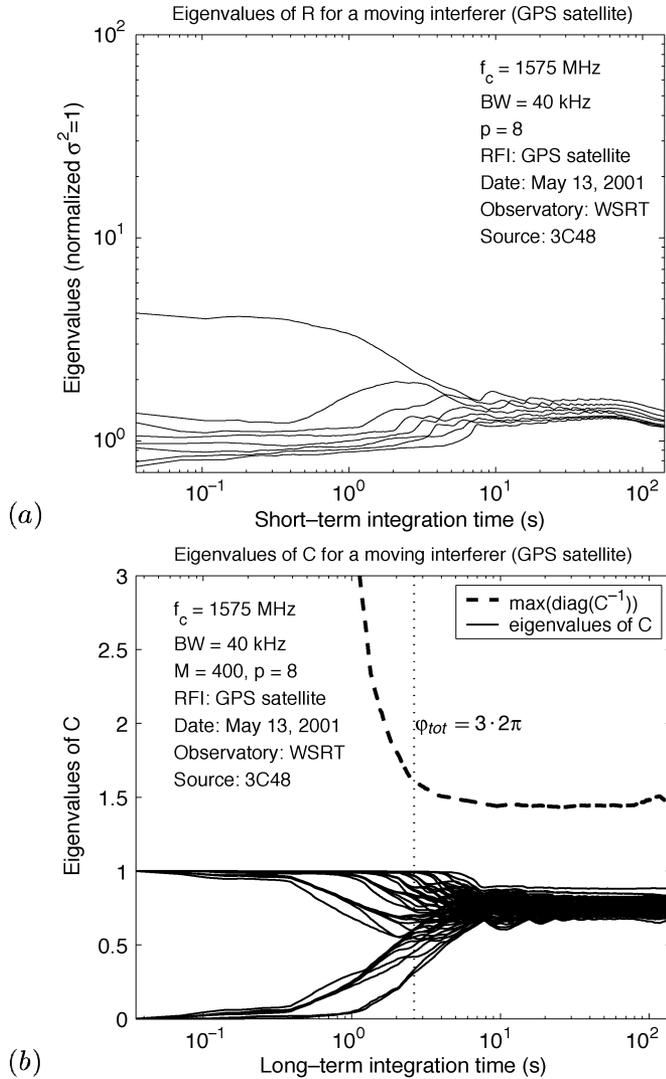


Fig. 2. Observation of 3C48 in the presence of an interfering signal from a GPS satellite. (a) Eigenvalues of $\hat{\mathbf{R}}_k$ for varying short-term integration length. (b) Eigenvalues of \mathbf{C} for varying long-term integration length.

e.g., if $\phi_{\min} = 3 \cdot 2\pi$ rad, $T_{\text{int}} = 30$ s, $\lambda = 30$ cm (corresponding to an observation frequency of about 1 GHz), and $D = 3000$ m, then the width of this band is 16° .

C. Experimental Results

The preceding results can be illustrated further using two measurement sets taken at the Westerbork Synthesis Radio Telescope (WSRT) in The Netherlands. The measurements consist of correlation products of $p = 8$ telescopes with a maximal baseline of $D = 1296$ m, split into subbands of 40 kHz and subsequently short-term integrated over $M = 400$ samples, thus corresponding to a short term integration time of $MT_s = 10$ ms. We will consider only a single subband channel. The observed source was 3C48 at declination $\delta = 33^\circ$. The data was offline calibrated to have noise power $\sigma^2 = 1$.

The first data set was measured at $f_c = 1575$ MHz and contained interference from a GPS satellite. Fig. 2 shows the results. In Fig. 2(a), the eigenvalues of the short-term data covari-

ance matrix $\hat{\mathbf{R}}_k$ are shown for increasing levels of integration (MT_s). For an integration length shorter than 0.5 s, the eigenvalues indicate a single interferer plus white noise. After 0.5 s, the interferer starts to occupy more than one spatial dimension because its motion changes the direction of the instantaneous spatial signature vector \mathbf{a}_k .

Since this clearly shows that the spatial signature is constant over the short-term integration interval of 10 ms, we subsequently do our spatial filtering at that level and vary the long-term integration length. The question now is whether the resulting \mathbf{C} is invertible. Fig. 2(b) shows the eigenvalues of \mathbf{C} for various levels of long-term integration (the short-term integration length is fixed at 10 ms). For integration lengths shorter than 0.5 s, \mathbf{C} has eigenvalues that are close to zero. After 0.5 s, these eigenvalues start to grow, and after about 3 s, \mathbf{C} is well conditioned. At the same time, the value of $\max(\text{diag}(\mathbf{C}^{-1}))$ drops to about 1.5.

The GPS satellite is in the far field and moves at a constant speed. Its spatial signatures can also be modeled by (18), i.e., by a linear phase progression, although the "fringe frequency" now does not correspond to the earth rotation. Indeed, we have verified that the data satisfies this model well and have estimated the corresponding frequency. The dotted vertical line shows where the $\phi_{\text{tot}} = 3 \cdot 2\pi$ condition is met, and it is seen that it coincides with the integration length above which \mathbf{C} is well conditioned.

The second data set, shown in Fig. 3, was measured at $f_c = 434$ MHz and contained interference from a stationary interferer, possibly an AM amateur broadcast. The stationarity of the source and the lower observing frequency result in a low fringe frequency. In this case as well, increasing the long-term integration period causes the smallest eigenvalues of \mathbf{C} to rise and $\max(\text{diag}(\mathbf{C}^{-1}))$ to drop but at a later time than in the previous case.

In this case, it is seen that at $\phi_{\text{tot}} = 3 \cdot 2\pi$ (dotted vertical line), $\max(\text{diag}(\mathbf{C}^{-1}))$ has not dropped, as low as in the previous case. The reason is that one antenna was receiving most of the energy, and such an imbalance was not considered in the analysis of the previous section.³ Still, at $T_{\text{int}} = 30$ s, $\max(\text{diag}(\mathbf{C}^{-1}))$ has dropped to 2.8 and to 2.1 at $T_{\text{int}} = 60$ s. Integration times of this length are not uncommon for the antenna configuration used for this experiment. Indeed, in radio astronomy, a condition that is used to estimate the longest permissible integration time is [19]

$$T_{\text{int}} < \frac{d}{\omega_e D}$$

where d is the antenna diameter (here 25 m) and D the longest baseline length (here 1296 m); as before, $\omega_e = 7.29115 \cdot 10^{-5}$ rad/s. As a result, $T_{\text{int}} < 265$ s. For $T_{\text{int}} = 265$ s and $f_c = 434$ MHz, (20) gives an unobservable band with width 10° .

³Moreover, this measurement was done 2 hr after the fringe frequency reached its maximum value so that better results can still be achieved.

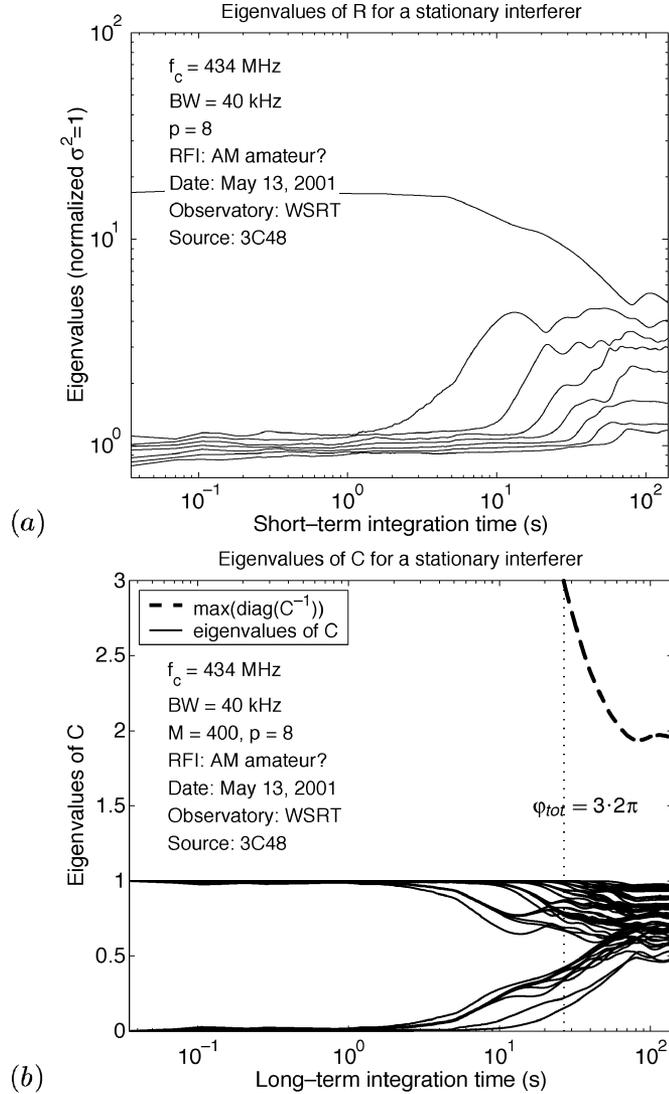


Fig. 3. Observation of 3C48 in the presence of a stationary interferer. (a) Eigenvalues of $\hat{\mathbf{R}}_k$ for varying short-term integration length. (b) Eigenvalues of $\hat{\mathbf{C}}$ for varying long-term integration length.

V. BIAS EXPRESSIONS AND CORRECTIONS

A. Bias for Low SNR

The final result of Section III was an expression for the bias of the spatial filtering algorithm (14), viz

$$\mathbb{E}[\Psi''] = -\frac{1}{M} \left(\sum_{i=1}^q \frac{\lambda_i}{\lambda_i - 1} \right) \mathbf{I} \quad (21)$$

which is valid for stationary eigenvalues but otherwise independent of the interferer statistics (a normalized noise power $\sigma^2 = 1$ is assumed). To gain some insight in this expression for the bias, we specialize to the case of a single interferer and very small SNR. In that case, $\mathbf{R}_k = \mathbf{a}_k \mathbf{a}_k^H + \mathbf{I}$. All eigenvalues are 1, except for $\lambda_{1,k} = 1 + \|\mathbf{a}_k\|^2$. If the interference power is constant over time, we can define the average interference-to-noise ratio per antenna as $\text{INR} = \|\mathbf{a}_k\|^2/p$ so that $\lambda_{1,k} = 1 + p \cdot \text{INR}$, and

$$\mathbb{E}[\Psi''] = -\frac{1}{M} \left(1 + \frac{\text{INR}^{-1}}{p} \right) \mathbf{I}. \quad (22)$$

The bias is independent of N and smallest for large INR: In that case, it converges to $-1/M$. For $N > M$, this is significant compared with the standard deviation, which is of order $1/\sqrt{MN}$. This is even more so for small values of the INR since the bias will be larger.

B. Bias for Low INR

For low INR (say zero INR), the result in (22) is obviously not valid. The reason is that the analysis assumed that the eigenvectors are unique, whereas for zero INRs, all eigenvalues of \mathbf{R}_k are the same and cannot be distinguished—and the corresponding eigenvectors are no longer unique. Hence, the analysis has to be modified for this case.

If the interference is not detectable but, nonetheless, a projection is made, it will be in the direction of the largest instantaneous eigenvector of $\hat{\mathbf{R}}_k$. The direction of that vector is random, but systematically, the component with the largest energy is removed, and this sorting effect will lead to a bias.

An estimate of this bias for the case $\hat{q}_k = 1$ is obtained as follows. Assume very low INR and SNR; hence, $\mathbf{R}_k = \mathbf{I}$, $\Psi = \mathbf{I}$. With projections computed from the largest eigenvector, we have

$$\hat{\mathbf{Q}}_k = \hat{\mathbf{P}}_k \hat{\mathbf{R}}_k \hat{\mathbf{P}}_k = \hat{\mathbf{R}}_k - \hat{\lambda}_{1,k} \hat{\mathbf{u}}_{1,k} \hat{\mathbf{u}}_{1,k}^H.$$

For large N

$$\hat{\mathbf{Q}} \rightarrow \mathbb{E}[\hat{\mathbf{Q}}_k] = \left(1 - \frac{\mathbb{E}[\hat{\lambda}_{1,k}]}{p} \right) \mathbf{I} \quad (23)$$

since $\mathbb{E}[\mathbf{u}\mathbf{u}^H] = (1/p)\mathbf{I}$ for vectors with an arbitrary direction (see Section IV-A), and the eigenvalues are statistically independent from the eigenvectors. With the asymptotic value of $\mathbb{E}[\mathbf{C}]$ from (15), we can derive

$$(\mathbb{E}[\mathbf{C}])^{-1} \text{vec}(\mathbf{I}) = \frac{p}{p-1} \text{vec}(\mathbf{I}). \quad (24)$$

Thus, for sufficiently large N

$$\begin{aligned} \text{vec}(\hat{\Psi} - \Psi) &\rightarrow \mathbb{E}[\mathbf{C}]^{-1} \text{vec}(\mathbb{E}[\hat{\mathbf{Q}}_k]) - \text{vec}(\mathbf{I}) \\ &= \left(\frac{p}{p-1} \left[1 - \frac{\mathbb{E}[\hat{\lambda}_{1,k}]}{p} \right] - 1 \right) \text{vec}(\mathbf{I}) \\ &= \left(\frac{1 - \mathbb{E}[\hat{\lambda}_{1,k}]}{p-1} \right) \text{vec}(\mathbf{I}). \end{aligned}$$

The expectation of the largest eigenvalue is approximately [15]⁴

$$\mathbb{E}[\hat{\lambda}_{1,k}] = \left(1 + \frac{\sqrt{p}}{\sqrt{M}} \right)^2. \quad (25)$$

Inserting this and taking only the dominant terms, we obtain that the bias for undetectable low INR is approximately

$$\hat{\Psi} - \Psi \rightarrow -\frac{p}{p-1} \left(\frac{2}{\sqrt{pM}} + \frac{1}{M} \right) \mathbf{I}. \quad (26)$$

This is the bias caused by doing projections when they are not needed. It is of order $1/\sqrt{pM}$ but only affects the main diagonal of $\hat{\Psi}$.

⁴It is known that this expression is an overestimation.

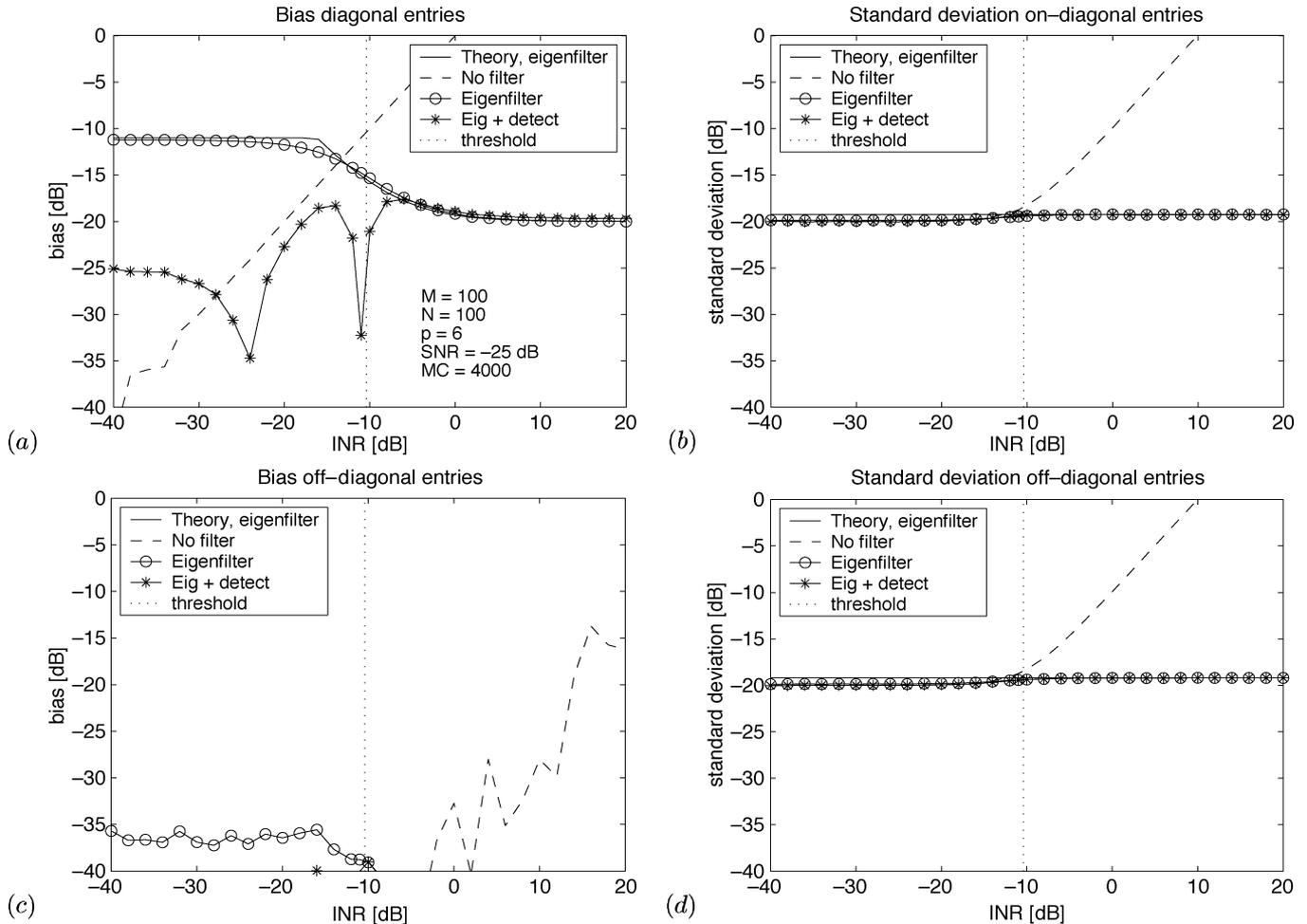


Fig. 4. (a) Bias and (b) standard deviation of the on-diagonal entries of $\hat{\Psi}$ versus input INR. (c) Bias and (d) standard deviation of the off-diagonal entries.

From simulations, it is seen that the combined bias for any INR can be estimated rather well as the minimum (in absolute value) of the two estimates (22) and (26).

In the above, we set $\hat{q}_k = 1$ for all k . In practice, a different algorithm is used: The number of interferers is detected from the eigenvalues, and a projection is only done if the interference is seen. This avoids most of the bias in (26). However, some bias remains due to false alarm, as will be seen in the simulations.

C. Simulations

To verify the results in this section, we performed a simulation with $q = 1$ interferer of varying INR. There is one astronomical source ($\text{SNR} = -25$ dB) and $p = 6$ telescopes. The spatial signature of the interferer in each interval is a random vector with i.i.d. complex normal distributed entries, all with the same average power (INR). The statistics are computed over 4000 Monte Carlo runs.

In Fig. 4, we show the bias and standard deviation of three algorithms: *i*) a direct covariance estimate with no spatial filter, *ii*) the covariance estimate using the spatial filter with projections computed from the eigenvectors, with fixed $q = 1$, and *iii*) the spatial filter with \hat{q}_k detected from the eigenvalues using a threshold γ as in (7). A vertical “threshold” line indicates the

INR that is barely detected, i.e., where $\gamma = \lambda_{1,k} = 1 + \text{INR} \cdot p$ (assuming large M), or

$$\text{INR}(\gamma) = \frac{\gamma - 1}{p}.$$

Fig. 4(a) is the bias of the on-diagonal entries of $\hat{\Psi}$. The theoretical curve is the minimum (in absolute value) of (22) and (26). It matches very well with the experimental bias of the spatial filter algorithm without detection. Note that the bias is much larger (-10 to -20 dB) than the contribution of the astronomical signal (-25 dB) and is also larger than the standard deviation (-20 dB). The bias of the spatial filter with detection looks a bit erratic because it changes signs two times ($- + -$). For very small INR, it is equal to the value in (26) times the false alarm rate of the detector: approximately 0.031 in this simulation; hence, it is 15 dB lower than the filter without detection. For large INR, the interference is always detected, and it is the same as (22). In between, the bias is caused by the fact that the interference is present but not detected, i.e., it follows the “no filter” line.

Fig. 4(c) shows the bias of the off-diagonal entries of $\hat{\Psi}$. Theoretically, it is zero, but the eigenfilter without detection has a very small nonzero bias for small INR: It is caused by the nonzero SNR, which gives a small preferred direction to the

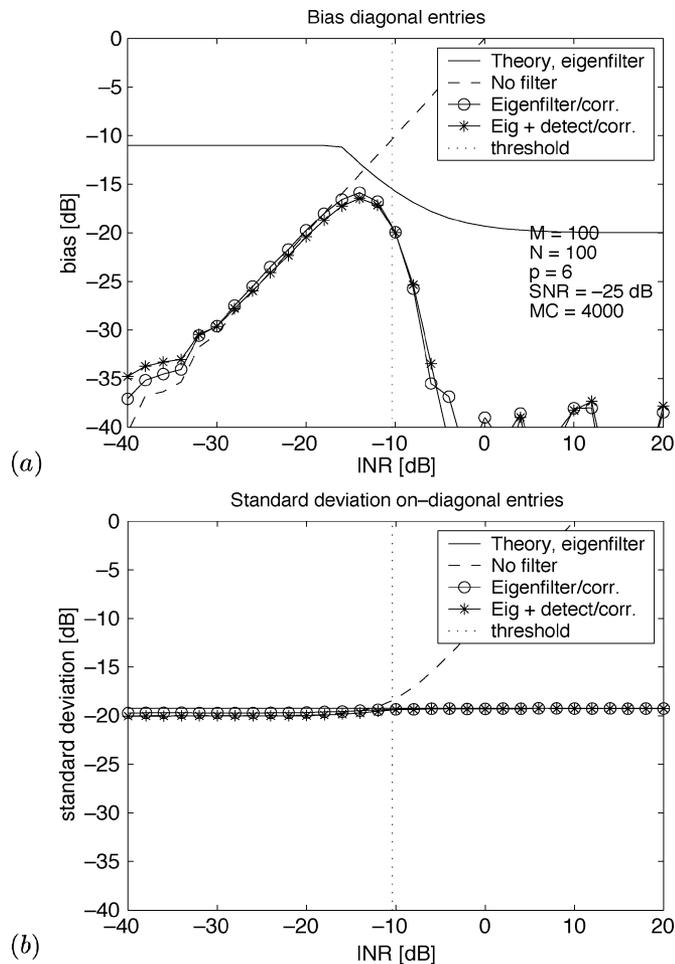


Fig. 5. (a) Bias. (b) Standard deviation of the on-diagonal entries of $\hat{\Psi}$ versus input INR after bias correction.

eigenvectors. Thus, the spatial filter tends to remove a small fraction of the astronomical signal. The filter with detection did not show a bias because it would be 15 dB lower.

In all cases, the variance of the covariance estimate is as predicted (17).

D. Bias Removal

Since the preceding expressions give quite an accurate idea about the bias, it is natural to try to remove the bias using them. A complication is that the bias is dependent on the INR, on the detected number of interferers in each interval, and on the threshold and the corresponding false alarm rates (these are also dependent on the number of interferers). We have derived some experimental techniques that were able to reduce the bias to below -35 dB for low INRs or INRs above the threshold [see (35) and (36) in Appendix C]. These results are not completely general and depend on knowledge of the false alarm rates of the detectors for various numbers of interferers.

Fig. 5 shows the results after this bias correction. Compared to Fig. 4(a), it is seen that most of the bias is removed, except in a region around the detection limit: Interference below this limit is not removed; hence, the bias line follows that of the unfiltered curve, but above the limit, the interference is quickly filtered out completely. Hence, the remaining bias directly corresponds

to the remaining (averaged) interference. Whether the bias is above the standard deviation (hence visible) depends on N ; the bias is independent of N , but the standard deviation scales as $1/\sqrt{N}$. Hence, for long integration lengths, there is always an INR window where the interference remains and will be visible after integration.

The bias is only present on the main diagonal of $\hat{\Psi}$. For many astronomical applications, this is acceptable, e.g., astronomical imaging does not use the auto-correlations. For spectral line observations, there is usually a single source of interest in the covariance data, for which the power can be recovered from the off-diagonal entries using factor analysis techniques (a form of self-calibration); see, e.g., [20]. Thus, we expect that the bias on the auto correlations will not be a major impediment, even if it is not removed.

VI. CONCLUSIONS

Spatial filtering can give a significant reduction in unwanted interference. Its effectiveness depends on the strength of the interferer: Stronger interference is more easily detected and can be better estimated and removed. The quality (covariance) of the resulting estimate is given by (12) and is almost as good as the unperturbed estimate in case the projection direction is sufficiently nonstationary. There is a small increase in variance due to the loss of information in the dimensions that were projected out.

Assuming the noise power is normalized to $\sigma^2 = 1$, a detection threshold can be defined as

$$\text{INR} = \frac{1}{p} \left(\left(1 + \frac{\sqrt{p}}{\sqrt{M}} \right)^2 - 1 \right)$$

where p is the number of antennas and M the short-term integration length. Interference with INR above this threshold is usually detected and filtered out. Below the threshold, the INR is usually too weak to be seen and will not be filtered out, leading to a bias on the diagonal entries of the covariance estimate. The most difficult interference has INR at this threshold.

In addition, for higher INR, spatial filtering can give some bias in the on-diagonal entries of the covariance estimate. This is because the projection directions and sample covariance matrices are estimated from the same data. Some of the bias can be removed, but there are two remaining effects that are hard to compensate for:

- In case of very weak/no interference, the false-alarm rate of the detector (a few percent) will lead to projections in random directions.
- Interference just below the detection threshold will not be reliably detected and will remain present in the covariance estimate. For randomly oriented interference, the effect will only be visible on the main diagonal of the final covariance estimate.

The use of reference antennas can improve the detection of weak interference and the estimation of its projection vectors, can handle interference with stationary projection directions, and is expected to generate smaller bias. Hence, this is an important direction for further research. Algorithms for this have been studied in [21], [22], and in our recent work.

APPENDIX A
PRELIMINARY RESULTS

In the proof of Theorem 1 in Appendix B, we use a few results on eigenvalue perturbation theory. They are listed in this section. The results are known—see, e.g., [16], [23]–[26]—but often derived and written in indexed form (i.e., in terms of individual matrix entries). We will write them in an integrated form using Kronecker products.

To set some notation, let \mathbf{R}_k be the true covariance matrix for the k th interval and $\hat{\mathbf{R}}_k$ be the corresponding finite sample estimate. All our estimates, eigenvalues, eigenvectors etc., are functions $\mathbf{f}(\hat{\mathbf{R}}_k)$ of $\hat{\mathbf{R}}_k$ and can be written in terms of the true function value $\mathbf{f}(\mathbf{R}_k)$ and a Taylor series expansion on the error $\mathbf{R}'_k = \hat{\mathbf{R}}_k - \mathbf{R}_k$, or $\hat{\mathbf{f}} = \mathbf{f} + \mathbf{f}' + \mathbf{f}'' + \dots$, where $\hat{\mathbf{f}} = \mathbf{f}(\hat{\mathbf{R}}_k)$, etc. We will only consider the first-order perturbation \mathbf{f}' and second-order perturbation \mathbf{f}'' .

Let $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ be an eigenvalue decomposition of \mathbf{R} , where \mathbf{U} is unitary, and $\mathbf{\Lambda}$ is diagonal with entries λ_i sorted in descending order,⁵ and define

$$\mathbf{\Sigma}_i^\dagger = (\mathbf{\Lambda} - \lambda_i \mathbf{I})^\dagger.$$

Lemma 2: For the unique eigenvectors (i.e., those corresponding to distinct eigenvalues)

$$\mathbf{u}'_i = -\mathbf{U}\mathbf{\Sigma}_i^\dagger \mathbf{U}^H \mathbf{R}' \mathbf{u}_i \quad : O\left(\frac{1}{\sqrt{M}}\right).$$

Proof: See, e.g., [24] and [25, eq. (C.2)]. □

Lemma 3: For Gaussian sources

$$\begin{aligned} \mathbb{E} [\text{vec}(\mathbf{R}') \text{vec}(\mathbf{R}')^H] &= \frac{1}{M} (\bar{\mathbf{R}} \otimes \mathbf{R}) \\ \Leftrightarrow \mathbb{E} [\bar{\mathbf{R}}' \otimes \mathbf{R}'] &= \frac{1}{M} \text{vec}(\mathbf{R}) \text{vec}(\mathbf{R})^H. \end{aligned}$$

Proof: For the first line, see, e.g., [16, eq. (A4)]. The second line is derived from the first as follows: Let \mathbf{e}_i denote the i th column of an identity matrix; then, a specific entry of $\bar{\mathbf{R}}' \otimes \mathbf{R}'$ is

$$\begin{aligned} &(\mathbf{e}_i \otimes \mathbf{e}_j)^T (\bar{\mathbf{R}}' \otimes \mathbf{R}') (\mathbf{e}_k \otimes \mathbf{e}_\ell) \\ &= \mathbf{e}_i^T \bar{\mathbf{R}}' \mathbf{e}_k \cdot \mathbf{e}_j^T \mathbf{R}' \mathbf{e}_\ell \\ &= (\mathbf{e}_\ell \otimes \mathbf{e}_j)^T \text{vec}(\mathbf{R}') \cdot \text{vec}(\mathbf{R}')^H (\mathbf{e}_k \otimes \mathbf{e}_i). \end{aligned}$$

Apply the expectation operator and use the first line to obtain

$$\begin{aligned} &\mathbb{E} [(\mathbf{e}_i \otimes \mathbf{e}_j)^T (\bar{\mathbf{R}}' \otimes \mathbf{R}') (\mathbf{e}_k \otimes \mathbf{e}_\ell)] \\ &= \frac{1}{M} (\mathbf{e}_\ell \otimes \mathbf{e}_j)^T (\bar{\mathbf{R}} \otimes \mathbf{R}) (\mathbf{e}_k \otimes \mathbf{e}_\ell) \\ &= \frac{1}{M} \mathbf{e}_\ell^H \bar{\mathbf{R}} \mathbf{e}_k \cdot \mathbf{e}_j^T \mathbf{R} \mathbf{e}_i \\ &= \frac{1}{M} (\mathbf{e}_i \otimes \mathbf{e}_j)^T \text{vec}(\mathbf{R}) \text{vec}(\mathbf{R})^H (\mathbf{e}_k \otimes \mathbf{e}_\ell). \end{aligned}$$

□

⁵The phase ambiguity on the columns of \mathbf{U} is resolved in some default manner; see [27].

The following result is a direct result of the preceding two lemmas and appears, e.g., in [16] and [25] where it is written using summations.

Lemma 4: For the unique eigenvectors of the covariance matrix of Gaussian sources

$$\begin{aligned} \mathbb{E} [\mathbf{u}'_i \mathbf{u}'_i{}^H] &= \frac{1}{M} \lambda_i \mathbf{U} \mathbf{\Sigma}_i^\dagger \mathbf{\Lambda} \mathbf{\Sigma}_i^\dagger \mathbf{U}^H \\ \mathbb{E} [\mathbf{u}'_i \mathbf{u}'_j{}^H] &= 0, \quad (j \neq i). \end{aligned}$$

Proof: Using Lemma 2

$$\begin{aligned} \mathbf{u}'_i \mathbf{u}'_j{}^H &= \mathbf{U} \mathbf{\Sigma}_i^\dagger \mathbf{U}^H \mathbf{R}' \mathbf{u}_i \mathbf{u}_j^H \mathbf{R}' \mathbf{U} \mathbf{\Sigma}_j^\dagger \mathbf{U}^H \\ &= (\mathbf{u}_i^T \otimes \mathbf{U} \mathbf{\Sigma}_i^\dagger \mathbf{U}^H) \text{vec}(\mathbf{R}') \text{vec}(\mathbf{R}')^H (\bar{\mathbf{u}}_j \otimes \mathbf{U} \mathbf{\Sigma}_j^\dagger \mathbf{U}^H). \end{aligned}$$

Using Lemma 3

$$\begin{aligned} \mathbb{E} [\mathbf{u}'_i \mathbf{u}'_j{}^H] &= \frac{1}{M} (\mathbf{u}_i^T \otimes \mathbf{U} \mathbf{\Sigma}_i^\dagger \mathbf{U}^H) (\bar{\mathbf{R}} \otimes \mathbf{R}) (\bar{\mathbf{u}}_j \otimes \mathbf{U} \mathbf{\Sigma}_j^\dagger \mathbf{U}^H) \\ &= \frac{1}{M} (\bar{\mathbf{u}}_j^H \mathbf{R} \mathbf{u}_i) \otimes (\mathbf{U} \mathbf{\Sigma}_i^\dagger \mathbf{U}^H \mathbf{R} \mathbf{U} \mathbf{\Sigma}_j^\dagger \mathbf{U}^H) \\ &= \frac{1}{M} (\bar{\mathbf{u}}_j^H \mathbf{R} \mathbf{u}_i) \otimes (\mathbf{U} \mathbf{\Sigma}_i^\dagger \mathbf{\Lambda} \mathbf{\Sigma}_j^\dagger \mathbf{U}^H). \end{aligned}$$

If $i \neq j$, then $\bar{\mathbf{u}}_j^H \mathbf{R} \mathbf{u}_i = 0$, else $\bar{\mathbf{u}}_j^H \mathbf{R} \mathbf{u}_i = \lambda_i$. □

APPENDIX B
PROOF OF THEOREM 1

In this section, for simplicity of notation, we assume that the noise power has been scaled to unity ($\sigma^2 = 1$).

We derive the perturbation on $\hat{\Psi}$ in the case that q_k interferers are present in the k th period. More precisely, the algorithm computes eigendecompositions $\hat{\mathbf{R}}_k = \hat{\mathbf{U}}_k \hat{\mathbf{\Lambda}}_k \hat{\mathbf{U}}_k^H$ and applies the projection $\hat{\mathbf{P}}_k = \mathbf{I} - \hat{\mathbf{U}}_{s,k} \hat{\mathbf{U}}_{s,k}^H$, where $\hat{\mathbf{U}}_{s,k}$ is the set of dominant eigenvectors (spanning the “signal subspace”). Then, with second-order accuracy

$$\begin{aligned} \hat{\mathbf{P}}_k &= \mathbf{I} - \hat{\mathbf{U}}_{s,k} \hat{\mathbf{U}}_{s,k}^H \\ &= \mathbf{I} - (\mathbf{U}_{s,k} + \mathbf{U}'_{s,k} + \mathbf{U}''_{s,k}) (\mathbf{U}_{s,k} + \mathbf{U}'_{s,k} + \mathbf{U}''_{s,k})^H \\ &= \mathbf{I} - \mathbf{U}_{s,k} \mathbf{U}_{s,k}^H - \mathbf{U}'_{s,k} \mathbf{U}_{s,k}^H - \mathbf{U}_{s,k} \mathbf{U}'_{s,k}{}^H \\ &\quad - \mathbf{U}'_{s,k} \mathbf{U}'_{s,k}{}^H - \mathbf{U}''_{s,k} \mathbf{U}_{s,k}^H - \mathbf{U}_{s,k} \mathbf{U}''_{s,k}{}^H. \end{aligned}$$

Let $\hat{\mathbf{P}}_k = \mathbf{P}_k + \mathbf{P}'_k + \mathbf{P}''_k + \dots$ be the Taylor series expansion of $\hat{\mathbf{P}}_k$; then, we can identify

$$\begin{aligned} \mathbf{P}_k &= \mathbf{I} - \mathbf{U}_{s,k} \mathbf{U}_{s,k}^H && : O(1) \\ \mathbf{P}'_k &= -\mathbf{U}'_{s,k} \mathbf{U}_{s,k}^H - \mathbf{U}_{s,k} \mathbf{U}'_{s,k}{}^H && : O\left(\frac{1}{\sqrt{M}}\right) \\ \mathbf{P}''_k &= -\mathbf{U}'_{s,k} \mathbf{U}'_{s,k}{}^H - \mathbf{U}''_{s,k} \mathbf{U}_{s,k}^H - \mathbf{U}_{s,k} \mathbf{U}''_{s,k}{}^H && : O\left(\frac{1}{M}\right). \end{aligned}$$

Subsequently

$$\begin{aligned} \hat{\mathbf{C}}_k &= \hat{\mathbf{P}}_k^T \otimes \hat{\mathbf{P}}_k \\ &= (\mathbf{P}_k + \mathbf{P}'_k + \mathbf{P}''_k)^T \otimes (\mathbf{P}_k + \mathbf{P}'_k + \mathbf{P}''_k) \\ &= (\mathbf{P}_k^T \otimes \mathbf{P}_k) + (\mathbf{P}'_k{}^T \otimes \mathbf{P}_k) + (\mathbf{P}_k^T \otimes \mathbf{P}'_k) \\ &\quad + (\mathbf{P}'_k{}^T \otimes \mathbf{P}'_k) + (\mathbf{P}''_k{}^T \otimes \mathbf{P}_k) + (\mathbf{P}_k^T \otimes \mathbf{P}''_k). \end{aligned}$$

Define $\hat{\mathbf{C}}_k = \mathbf{C}_k + \mathbf{C}'_k + \mathbf{C}''_k + \dots$; then, we can identify

$$\begin{aligned} \mathbf{C}_k &= (\mathbf{P}_k^T \otimes \mathbf{P}_k) && : O(1) \\ \mathbf{C}'_k &= (\mathbf{P}'_k{}^T \otimes \mathbf{P}_k) + (\mathbf{P}_k^T \otimes \mathbf{P}'_k) && : O\left(\frac{1}{\sqrt{MN}}\right) \\ \mathbf{C}''_k &= (\mathbf{P}'_k{}^T \otimes \mathbf{P}'_k) + (\mathbf{P}''_k{}^T \otimes \mathbf{P}_k) \\ &\quad + (\mathbf{P}_k^T \otimes \mathbf{P}''_k) && : O\left(\frac{1}{M}\right). \end{aligned}$$

Similarly

$$\begin{aligned} \hat{\mathbf{C}} &:= \frac{1}{N} \sum_{k=1}^N \hat{\mathbf{C}}_k =: \frac{1}{N} \sum_{k=1}^N (\mathbf{C}_k + \mathbf{C}'_k + \mathbf{C}''_k) \\ &\Rightarrow \begin{cases} \mathbf{C} = \frac{1}{N} \sum_{k=1}^N \mathbf{C}_k & : O(1) \\ \mathbf{C}' = \frac{1}{N} \sum_{k=1}^N \mathbf{C}'_k & : O\left(\frac{1}{\sqrt{MN}}\right) \\ \mathbf{C}'' = \frac{1}{N} \sum_{k=1}^N \mathbf{C}''_k & : O\left(\frac{1}{M}\right). \end{cases} \end{aligned}$$

Note that \mathbf{C}''_k has some entries on the main diagonal that are always positive. Since these do not average to 0, \mathbf{C}'' does not scale with $1/\sqrt{N}$.

Continuing, we have

$$\begin{aligned} \text{vec}(\hat{\mathbf{Q}}_k) &= \hat{\mathbf{C}}_k \text{vec}(\hat{\mathbf{R}}_k) = (\mathbf{C}_k + \mathbf{C}'_k + \mathbf{C}''_k) \text{vec}(\mathbf{R} + \mathbf{R}'_k) \\ &\Rightarrow \begin{cases} \text{vec}(\mathbf{Q}_k) = \mathbf{C}_k \text{vec}(\mathbf{R}_k) & : O(1) \\ \text{vec}(\mathbf{Q}'_k) = \mathbf{C}'_k \text{vec}(\mathbf{R}_k) \\ \quad + \mathbf{C}_k \text{vec}(\mathbf{R}'_k) & : O\left(\frac{1}{\sqrt{M}}\right) \\ \text{vec}(\mathbf{Q}''_k) = \mathbf{C}''_k \text{vec}(\mathbf{R}_k) \\ \quad + \mathbf{C}'_k \text{vec}(\mathbf{R}'_k) & : O\left(\frac{1}{M}\right) \end{cases} \\ \hat{\mathbf{Q}} &= \frac{1}{N} \sum_{k=1}^N \hat{\mathbf{Q}}_k = \frac{1}{N} \sum_{k=1}^N (\mathbf{Q}_k + \mathbf{Q}'_k + \mathbf{Q}''_k) \\ &\Rightarrow \begin{cases} \mathbf{Q} = \frac{1}{N} \sum_{k=1}^N \mathbf{Q}_k & : O(1) \\ \mathbf{Q}' = \frac{1}{N} \sum_{k=1}^N \mathbf{Q}'_k & : O\left(\frac{1}{\sqrt{MN}}\right) \\ \mathbf{Q}'' = \frac{1}{N} \sum_{k=1}^N \mathbf{Q}''_k & : O\left(\frac{1}{M}\right). \end{cases} \end{aligned}$$

As with \mathbf{C}''_k , in addition, \mathbf{Q}'' does not scale with $1/\sqrt{N}$ because certain entries of \mathbf{Q}''_k are always positive. Subsequently, we obtain

$$\begin{aligned} \text{vec}(\hat{\Psi}) &= \hat{\mathbf{C}}^{-1} \text{vec}(\hat{\mathbf{Q}}) \\ &= (\mathbf{C}^{-1} + (\mathbf{C}^{-1})' + (\mathbf{C}^{-1})'') \text{vec}(\mathbf{Q} + \mathbf{Q}' + \mathbf{Q}'') \\ &\Rightarrow \begin{cases} \text{vec}(\Psi) = \mathbf{C}^{-1} \text{vec}(\mathbf{Q}) & : O(1) \\ \text{vec}(\Psi') = (\mathbf{C}^{-1})' \text{vec}(\mathbf{Q}) + \mathbf{C}^{-1} \text{vec}(\mathbf{Q}') & : O\left(\frac{1}{\sqrt{MN}}\right) \\ \text{vec}(\Psi'') = (\mathbf{C}^{-1})'' \text{vec}(\mathbf{Q}) + (\mathbf{C}^{-1})' \text{vec}(\mathbf{Q}') \\ \quad + \mathbf{C}^{-1} \text{vec}(\mathbf{Q}'') & : O\left(\frac{1}{M} + \frac{1}{(MN)} + \frac{1}{M}\right). \end{cases} \end{aligned} \quad (27)$$

By expansion of $\hat{\mathbf{C}}^{-1}$, it can be shown that

$$\begin{aligned} \hat{\mathbf{C}}^{-1} &= \mathbf{C}^{-1} - \mathbf{C}^{-1}(\mathbf{C}' + \mathbf{C}'')\mathbf{C}^{-1} + \mathbf{C}^{-1}(\mathbf{C}' + \mathbf{C}'') \\ &\quad \times \mathbf{C}^{-1}(\mathbf{C}' + \mathbf{C}'')\mathbf{C}^{-1} + \dots \\ &\Rightarrow \begin{cases} (\mathbf{C}^{-1})' = -\mathbf{C}^{-1}\mathbf{C}'\mathbf{C}^{-1} & : O\left(\frac{1}{\sqrt{MN}}\right) \\ (\mathbf{C}^{-1})'' = -\mathbf{C}^{-1}\mathbf{C}''\mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{C}'\mathbf{C}^{-1} \\ \quad \times \mathbf{C}'\mathbf{C}^{-1} & : O\left(\frac{1}{M}\right) + O\left(\frac{1}{(MN)}\right). \end{cases} \end{aligned} \quad (28)$$

To work out $\text{vec}(\Psi')$, we first show that $(1/N) \sum \mathbf{C}'_k \text{vec}(\mathbf{R}_k) = (1/N) \sum \mathbf{C}'_k \text{vec}(\Psi) = \mathbf{C}' \text{vec}(\Psi)$. Indeed

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

and

$$\mathbf{C}'_k = (\bar{\mathbf{P}}'_k \otimes \mathbf{P}_k) + (\bar{\mathbf{P}}_k \otimes \mathbf{P}'_k)$$

so that

$$\begin{aligned} \mathbf{C}'_k \text{vec}(\mathbf{R}_k) &= \mathbf{C}'_k \text{vec}(\Psi) + (\bar{\mathbf{P}}'_k \otimes \mathbf{P}_k)(\bar{\mathbf{A}}_k \circ \mathbf{A}_k)\mathbf{1} \\ &\quad + (\bar{\mathbf{P}}_k \otimes \mathbf{P}'_k)(\bar{\mathbf{A}}_k \circ \mathbf{A}_k)\mathbf{1}. \end{aligned}$$

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

It follows that

$$\text{vec}(\mathbf{Q}') = \mathbf{C}' \text{vec}(\Psi) + \frac{1}{N} \sum_{k=1}^N \mathbf{C}_k \text{vec}(\mathbf{R}'_k).$$

Inserting in (27) and using (28) gives

$$\begin{aligned} \text{vec}(\Psi') &= -\mathbf{C}^{-1}\mathbf{C}'\mathbf{C}^{-1}\text{vec}(\mathbf{Q}) \\ &\quad + \mathbf{C}^{-1} \left(\mathbf{C}' \text{vec}(\Psi) + \frac{1}{N} \sum_{k=1}^N \mathbf{C}_k \text{vec}(\mathbf{R}'_k) \right). \end{aligned}$$

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

$$\text{vec}(\Psi') = \mathbf{C}^{-1} \frac{1}{N} \sum_{k=1}^N \mathbf{C}_k \text{vec}(\mathbf{R}'_k) \quad : O\left(\frac{1}{\sqrt{MN}}\right). \quad (29)$$

The second-order perturbation on $\hat{\Psi}$ is evaluated as

$$\begin{aligned} \text{vec}(\Psi'') &= (\mathbf{C}^{-1})'' \text{vec}(\mathbf{Q}) + (\mathbf{C}^{-1})' \text{vec}(\mathbf{Q}') + \mathbf{C}^{-1} \text{vec}(\mathbf{Q}'') \\ &= - \underbrace{\mathbf{C}^{-1}\mathbf{C}''\mathbf{C}^{-1}\text{vec}(\mathbf{Q})}_{\frac{1}{M}} \\ &\quad + \underbrace{\mathbf{C}^{-1}\mathbf{C}'\mathbf{C}^{-1}\mathbf{C}'\mathbf{C}^{-1}\text{vec}(\mathbf{Q})}_{\frac{1}{(MN)}} \\ &\quad - \underbrace{\mathbf{C}^{-1}\mathbf{C}'\mathbf{C}^{-1}\text{vec}(\mathbf{Q}')}_{\frac{1}{(MN)}} + \underbrace{\mathbf{C}^{-1}\text{vec}(\mathbf{Q}'')}_{\frac{1}{M}}. \end{aligned}$$

If N is large, then the second and third terms can be neglected, and we obtain

$$\begin{aligned} \text{vec}(\Psi'') &= \mathbf{C}^{-1} \frac{1}{N} \sum_{k=1}^N -\mathbf{C}_k'' \mathbf{C}^{-1} \text{vec}(\mathbf{Q}) + \text{vec}(\mathbf{Q}_k'') \\ &= \mathbf{C}^{-1} \frac{1}{N} \sum_{k=1}^N -\mathbf{C}_k'' \text{vec}(\Psi) + \mathbf{C}_k'' \text{vec}(\mathbf{R}_k) \\ &\quad + \mathbf{C}_k' \text{vec}(\mathbf{R}_k') \\ &= \mathbf{C}^{-1} \frac{1}{N} \sum_{k=1}^N \mathbf{C}_k'' \text{vec}(\mathbf{A}_k \mathbf{A}_k^H) + \mathbf{C}_k' \text{vec}(\mathbf{R}_k'). \end{aligned} \quad (30)$$

Each of the two terms inside the summation is $O(1/M)$. Using $\mathbf{A}_k \mathbf{A}_k^H = \mathbf{U}_{s,k} (\Lambda_{s,k} - \mathbf{I}) \mathbf{U}_{s,k}^H$ (the ‘‘sky’’ contribution is ignored, and the ‘‘noise’’ eigenvalues are 1), the first term is worked out as

$$\begin{aligned} \mathbf{C}_k'' \text{vec}(\mathbf{A}_k \mathbf{A}_k^H) &= [\bar{\mathbf{P}}_k' \otimes \mathbf{P}_k' + \bar{\mathbf{P}}_k'' \otimes \mathbf{P}_k + \bar{\mathbf{P}}_k \otimes \mathbf{P}_k'] \text{vec}(\mathbf{A}_k \mathbf{A}_k^H) \\ &= \text{vec}(\mathbf{P}_k' \mathbf{A}_k \mathbf{A}_k^H \mathbf{P}_k') \\ &= \text{vec} \left[\left(\mathbf{U}_{s,k}' \mathbf{U}_{s,k}^H + \mathbf{U}_{s,k} \mathbf{U}_{s,k}'^H \right) \mathbf{U}_{s,k} (\Lambda_{s,k} - \mathbf{I}) \mathbf{U}_{s,k}^H \right. \\ &\quad \left. \cdot \left(\mathbf{U}_{s,k} \mathbf{U}_{s,k}^H + \mathbf{U}_{s,k}' \mathbf{U}_{s,k}'^H \right) \right] \\ &= \text{vec} \left[\mathbf{U}_{s,k}' (\Lambda_{s,k} - \mathbf{I}) \mathbf{U}_{s,k}'^H + \mathbf{U}_{s,k}' (\Lambda_{s,k} - \mathbf{I}) \right. \\ &\quad \times \mathbf{U}_{s,k}^H \mathbf{U}_{s,k}' \mathbf{U}_{s,k}^H + \mathbf{U}_{s,k} \mathbf{U}_{s,k}'^H \mathbf{U}_{s,k} (\Lambda_{s,k} - \mathbf{I}) \\ &\quad \times \mathbf{U}_{s,k}'^H + \mathbf{U}_{s,k} \mathbf{U}_{s,k}'^H \mathbf{U}_{s,k} (\Lambda_{s,k} - \mathbf{I}) \\ &\quad \left. \times \mathbf{U}_{s,k}^H \mathbf{U}_{s,k}' \mathbf{U}_{s,k}^H \right] \\ &= \text{vec} \left[\sum_{i=1}^{q_k} (\lambda_{i,k} - 1) \mathbf{u}_{i,k}' \mathbf{u}_{i,k}'^H + \sum_{i=1}^{q_k} \sum_{j=1}^{q_k} (\lambda_{i,k} - 1) \right. \\ &\quad \times \mathbf{u}_{i,k}' \mathbf{u}_{i,k}^H \mathbf{u}_{j,k}' \mathbf{u}_{j,k}^H + \sum_{i=1}^{q_k} \sum_{j=1}^{q_k} (\lambda_{j,k} - 1) \mathbf{u}_{i,k} \mathbf{u}_{i,k}'^H \\ &\quad \times \mathbf{u}_{j,k} \mathbf{u}_{j,k}'^H + \sum_{i=1}^{q_k} \sum_{j=1}^{q_k} \sum_{\ell=1}^{q_k} (\lambda_{j,k} - 1) \mathbf{u}_{i,k} \mathbf{u}_{i,k}'^H \\ &\quad \left. \times \mathbf{u}_{j,k} \mathbf{u}_{j,k}^H \mathbf{u}_{\ell,k}' \mathbf{u}_{\ell,k}^H \right]. \end{aligned}$$

The expected value of this term is worked out as (the tedious but straightforward derivation uses Lemmas 2–4)

$$\mathbb{E} [\mathbf{C}_k'' \text{vec}(\mathbf{A}_k \mathbf{A}_k^H)] = \frac{1}{M} \sum_{i=1}^{q_k} \frac{\lambda_{i,k}}{\lambda_{i,k} - 1} \cdot \text{vec}(\mathbf{P}_k). \quad (31)$$

The second term in (30) is worked out as

$$\begin{aligned} \mathbf{C}_k' \text{vec}(\mathbf{R}_k') &= \text{vec} [\mathbf{P}_k' \mathbf{R}_k' \mathbf{P}_k + \mathbf{P}_k \mathbf{R}_k' \mathbf{P}_k'] \\ &= \text{vec} \left[- \left(\mathbf{U}_{s,k}' \mathbf{U}_{s,k}^H \mathbf{R}_k' - \mathbf{U}_{s,k} \mathbf{U}_{s,k}'^H \mathbf{R}_k' \right) \mathbf{P}_k \right. \\ &\quad \left. - \mathbf{P}_k \left(\mathbf{R}_k' \mathbf{U}_{s,k}' \mathbf{U}_{s,k}^H - \mathbf{R}_k' \mathbf{U}_{s,k} \mathbf{U}_{s,k}'^H \right) \right]. \end{aligned}$$

Using a similar derivation as before, this is worked out as

$$\mathbb{E} [\mathbf{C}_k' \text{vec}(\mathbf{R}_k')] = -\frac{2}{M} \sum_{i=1}^{q_k} \frac{\lambda_{i,k}}{\lambda_{i,k} - 1} \cdot \text{vec}(\mathbf{P}_k).$$

Altogether, we obtain

$$\mathbb{E} [\text{vec}(\Psi'')] = -\mathbf{C}^{-1} \frac{1}{N} \sum_{k=1}^N \frac{1}{M} \left(\sum_{i=1}^{q_k} \frac{\lambda_{i,k}}{\lambda_{i,k} - 1} \right) \text{vec}(\mathbf{P}_k)$$

which is the result stated in the theorem. If the eigenvalues are stationary (independent of k), this can be further simplified as

$$\mathbb{E}[\Psi''] = -\frac{1}{M} \left(\sum_{i=1}^q \frac{\lambda_i}{\lambda_i - 1} \right) \mathbf{I} \quad (32)$$

since, by definition of \mathbf{C}

$$\begin{aligned} \frac{1}{N} \mathbf{C}^{-1} \sum_{k=1}^N \text{vec}(\mathbf{P}_k) &= \frac{1}{N} \mathbf{C}^{-1} \sum_{k=1}^N \text{vec}(\mathbf{P}_k \mathbf{I} \mathbf{P}_k) \\ &= \frac{1}{N} \mathbf{C}^{-1} \sum_{k=1}^N (\bar{\mathbf{P}}_k \otimes \mathbf{P}_k) \text{vec}(\mathbf{I}) \\ &= \text{vec}(\mathbf{I}). \end{aligned}$$

Ψ' is $O(1/\sqrt{MN})$ with expected value 0, and Ψ'' is $O(1/(M\sqrt{N}))$ with nonzero expected value. Thus

$$\mathbb{E}[\hat{\Psi}] = \Psi + \mathbb{E}[\Psi''].$$

The bias is given by $\mathbb{E}[\Psi'']$. The covariance of the estimate of Ψ is given by

$$\text{cov}\{\hat{\Psi}\} = \mathbb{E} \left[\text{vec}(\Psi' + \Psi'' - \mathbb{E}[\Psi'']) \text{vec}(\Psi' + \Psi'' - \mathbb{E}[\Psi''])^H \right].$$

In the evaluation of $\text{cov}\{\hat{\Psi}\}$, for large M , we can ignore Ψ'' , and the covariance is equal to the case where the \mathbf{A}_k are known:

$$\text{cov}\{\hat{\Psi}\} \approx \mathbb{E} \left[\text{vec}(\Psi') \text{vec}(\Psi')^H \right] \approx \frac{1}{MN} \mathbf{C}^{-1}$$

(note that we scaled $\sigma = 1$). This performance depends only on the total number of samples. A more detailed analysis shows that this result is even valid for small M (the covariance will be larger by terms of $O(1/(M^2N))$, which can be ignored for $M > 10$).

APPENDIX C BIAS REMOVAL EQUATIONS

We try to remove the bias by adding correction terms to $\hat{\mathbf{Q}}_k$. The complication here is that, even if we know quite precisely the bias as function of INR, in practice, the INR is unknown. Hence, we need to make corrections that are independent of the INR. The idea is to use the eigenvalues $\hat{\lambda}_{i,k}$ to correct the corresponding $\hat{\mathbf{Q}}_k$ rather than the final $\hat{\Psi}$.

For higher INRs, we can set, based on (13)

$$\hat{\mathbf{Q}}_k = \hat{\mathbf{P}}_k \hat{\mathbf{R}}_k \hat{\mathbf{P}}_k + \frac{1}{M} \sum_{i=1}^{q_k} \frac{\hat{\lambda}_{i,k}}{\hat{\lambda}_{i,k} - 1} \cdot \hat{\mathbf{P}}_k. \quad (33)$$

The added term will increase $\hat{\mathbf{Q}}$ by a term

$$\frac{1}{N} \sum_{k=1}^N \frac{1}{M} \left(\sum_{i=1}^{q_k} \frac{\hat{\lambda}_{i,k}}{\hat{\lambda}_{i,k} - 1} \right) \hat{\mathbf{P}}_k$$

and the premultiplication by $\hat{\mathbf{C}}^{-1}$ will give the desired result. Here, q_k is the number of interferers in interval k or the number of detected interferers in case of eigenvalue filtering with detection. This correction removes most of the bias for higher INRs.

For very low INRs, we need a different correction. In that case, we can set

$$\hat{\mathbf{Q}}_k = \hat{\mathbf{P}}_k \hat{\mathbf{R}}_k \hat{\mathbf{P}}_k + \frac{1}{p} (\hat{\lambda}_{1,k} - 1) \mathbf{I}. \quad (34)$$

Using (23), we obtain $\mathbb{E}[\hat{\mathbf{Q}}] = (1 - (1/p))\mathbf{I} = (p - 1/p)\mathbf{I}$. Premultiplication of $\text{vec}(\hat{\mathbf{Q}})$ by \mathbf{C}^{-1} [cf. (24)] gives the required result: $\mathbb{E}[\hat{\Psi}] = \mathbf{I}$.

For growing INRs, this term would grow linearly and be too large. Hence, even if we always project out one dimension, a correct bias removal still requires detection of the presence of interference. If it is detected, we correct using (33); else (34).

A further refinement is needed to take into account the false alarm rate. For very low INR, there are P_{FA} cases where the bias was removed according to (33) instead of (34). The following correction algorithms take this into account as well.

Bias Removal, Spatial Filtering of One Dimension

Consider the spatial filtering algorithm, where always, $q = 1$ dimension is projected out (independent of the actual INR). To correct for the bias, assume that a detector is available, which, in the k th interval, detects \hat{q}_k and has false-alarm rate P_{FA} . Set

$$\begin{aligned} \text{if } \hat{q}_k = 0 : \hat{\mathbf{Q}}_k &= \hat{\mathbf{P}}_k \hat{\mathbf{R}}_k \hat{\mathbf{P}}_k + \frac{1}{p} (\hat{\lambda}_{1,k} - 1) \cdot \mathbf{I} + \frac{P_{\text{FA}}}{1 - P_{\text{FA}}} \\ &\times \left[\frac{1}{p} (\mathbb{E}[\hat{\lambda}_1] - 1) - \frac{1}{M} \frac{\mathbb{E}[\hat{\lambda}_1]}{\mathbb{E}[\hat{\lambda}_1] - 1} \frac{p - 1}{p} \right] \mathbf{I} \\ \text{if } \hat{q}_k = 1 : \hat{\mathbf{Q}}_k &= \hat{\mathbf{P}}_k \hat{\mathbf{R}}_k \hat{\mathbf{P}}_k + \frac{1}{M} \frac{\hat{\lambda}_1}{\hat{\lambda}_1 - 1} \cdot \hat{\mathbf{P}}_k \end{aligned} \quad (35)$$

then this removes the bias for low and high INR up to second order terms. Indeed, for high INR, the interference is always detected, and the correction is as in (33). For low INR, we have $\hat{q} = 0$ in $(1 - P_{\text{FA}})$ number of cases and $\hat{q}_k = 1$ in P_{FA} number of cases. The expected value of $\hat{\mathbf{Q}}$ is then

$$\begin{aligned} \mathbb{E}[\hat{\mathbf{Q}}] &= \mathbb{E}[\hat{\mathbf{P}}_k \hat{\mathbf{R}}_k \hat{\mathbf{P}}_k] + (1 - P_{\text{FA}}) \frac{1}{p} (\mathbb{E}[\hat{\lambda}_1] - 1) \cdot \mathbf{I} \\ &+ P_{\text{FA}} \left[\frac{1}{p} (\mathbb{E}[\hat{\lambda}_1] - 1) - \frac{1}{M} \frac{\mathbb{E}[\hat{\lambda}_1]}{\mathbb{E}[\hat{\lambda}_1] - 1} \frac{p - 1}{p} \right] \mathbf{I} \\ &+ P_{\text{FA}} \frac{1}{M} \frac{\mathbb{E}[\hat{\lambda}_1]}{\mathbb{E}[\hat{\lambda}_1] - 1} \cdot \mathbb{E}[\hat{\mathbf{P}}_k] \\ &= \mathbb{E}[\hat{\mathbf{P}}_k \hat{\mathbf{R}}_k \hat{\mathbf{P}}_k] + \frac{1}{p} (\mathbb{E}[\hat{\lambda}_1] - 1) \cdot \mathbf{I} \end{aligned}$$

since $\mathbb{E}[\hat{\mathbf{P}}_k] = (1 - (1/p))\mathbf{I}$ for arbitrary 1-D projections. This gives us back the situation in (34). For $\mathbb{E}[\hat{\lambda}_1]$, we can take the expression in (25).

Bias Removal, Spatial Filtering With Detection

Consider the spatial filtering algorithm where the number of interferers q is first detected (\hat{q}_k) and the corresponding eigenvectors projected out. Only if an interferer is detected, a correction using (33) is needed. If the interference is not detected, then no correction is needed, except to correct for the bias due to false alarm.

Let $P_{\text{FA}}(q)$ be the false alarm rate of detecting $q + 1$ interferers when there are only q , and let $\hat{\mathbf{P}}_k(q)$ be the projection matrix that projects out the first q eigenvectors. Without further motivation, we mention the following algorithm:

$$\begin{aligned} \text{if } \hat{q}_k = 0 : \hat{\mathbf{Q}}_k &= \hat{\mathbf{R}}_k + \frac{P_{\text{FA}}(0)}{1 - P_{\text{FA}}(0)} \\ &\times \left[\frac{1}{p} (\mathbb{E}[\hat{\lambda}_1] - 1) - \frac{1}{M} \frac{\mathbb{E}[\hat{\lambda}_1]}{\mathbb{E}[\hat{\lambda}_1] - 1} \frac{p - 1}{p} \right] \mathbf{I} \\ \text{if } \hat{q}_k > 0 : \hat{\mathbf{Q}}_k &= \hat{\mathbf{P}}_k(\hat{q}_k) \hat{\mathbf{R}}_k \hat{\mathbf{P}}_k(\hat{q}_k) + \frac{1}{M} \sum_{i=1}^{\hat{q}_k} \frac{\hat{\lambda}_{i,k}}{\hat{\lambda}_{i,k} - 1} \\ &\cdot \hat{\mathbf{P}}_k(\hat{q}_k) + P_{\text{FA}}(\hat{q}_k) \frac{1}{p - \hat{q}_k} \\ &\times \left(\left(1 + \frac{\sqrt{p - \hat{q}_k}}{\sqrt{M}} \right)^2 - 1 \right) \hat{\mathbf{P}}_k(\hat{q}_k - 1). \end{aligned} \quad (36)$$

The algorithm tries to take into account that for $q = 1$ interferer and high INR (hence always detected), it might happen that a second eigenvalue is larger than the threshold ($\hat{q} = 2$, with probability $P_{\text{FA}}(1)$).

REFERENCES

- [1] S. van der Tol, A. J. van der Veen, and A. J. Boonstra, "Mitigation of continuous interference in radio astronomy using spatial filtering," in *Proc. URSI General Assembly*, Maastricht, The Netherlands, Aug. 2002.
- [2] A. Leshem, A.-J. van der Veen, and A.-J. Boonstra, "Multichannel interference mitigation techniques in radio astronomy," *Astrophys. J. Supplement Series*, vol. 131, pp. 355–373, Nov. 2000.
- [3] P. A. Fridman and W. A. Baan, "RFI mitigation methods in radio astronomy," *Astron. Astrophys.*, vol. 378, pp. 327–344, 2001.
- [4] P. A. Fridman, "RFI excision using a higher order statistics analysis of the power spectrum," *Astron. Astrophys.*, vol. 368, pp. 369–376, 2001.
- [5] C. Barnbaum and R. F. Bradley, "A new approach to interference excision in radio astronomy: real time adaptive filtering," *Astron. J.*, vol. 115, pp. 2598–2614, 1998.
- [6] S. W. Ellingson, J. D. Bunton, and J. F. Bell, "Removal of the GLONASS C/A signal from OH spectral line observations using a parametric modeling technique," *Astrophys. J. Supplement*, vol. 135, no. 1, pp. 87–93, 2001.
- [7] L. Li and B. D. Jeffs, "Analysis of adaptive array algorithm performance for satellite interference cancellation in radio astronomy," in *Proc. URSI General Assembly*, Aug. 2002.
- [8] J. Raza, A. J. Boonstra, and A. J. van der Veen, "Spatial filtering of RF interference in radio astronomy," *IEEE Signal Process. Lett.*, vol. 9, no. 2, pp. 64–67, Feb. 2002.
- [9] M. S. Bartlett, "Tests of significance in factor analysis," *British J. Psych. (Statist. Sect.)*, vol. 3, pp. 77–85, 1950.
- [10] —, "A note on the multiplying factors for various χ^2 approximations," *J. R. Statist. Soc.*, ser. B, vol. 16, pp. 296–298, 1954.
- [11] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Stat.*, vol. 34, pp. 122–148, 1963.
- [12] H. Akaike, "Information theory and an extension of the Maximum Likelihood principle," in *Proc. 2nd Int. Symp. Inf. Theory*, 1973, pp. 267–281.
- [13] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 1190–1196, Aug. 1989.
- [14] H. L. van Trees, *Optimum Array Processing (Part IV of Detection, Estimation, and Modulation Theory)*. New York: Wiley-Interscience, 2002.

- [15] A. Edelman, "The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type," *Lin. Alg. Appl.*, vol. 159, pp. 55–80, 1991.
- [16] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 331–341, Apr. 1986 [Corrections in ASSP-34 (6) 1986].
- [17] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. New York: Academic, 1979.
- [18] R. A. Perley, F. Schwab, and A. H. Bridle, Eds., *Synthesis Imaging in Radio Astronomy (A Collection of Lectures From the Third NRAO Synthesis Imaging Summer School)*. San Francisco, CA: Astronomical Soc. Pacific, 1989, vol. 6.
- [19] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radioastronomy*, Second ed. New York: Wiley, 2001.
- [20] A. J. Boonstra and A. J. van der Veen, "Gain calibration methods for radio telescope arrays," *IEEE Trans. Signal Process.*, vol. 51, pp. 25–38, Jan. 2003.
- [21] B. D. Jeffs, K. Warnick, and L. Li, "Improved interference cancellation in synthesis array radio astronomy using auxiliary antennas," in *Proc. IEEE ICASSP*, Hong Kong, Apr. 2003.
- [22] B. D. Jeffs, L. Li, and K. Warnick, "Auxiliary antenna assisted interference cancellation for radio astronomy arrays," *IEEE Trans. Signal Processing*, vol. 53, pp. 439–451, Feb. 2005.
- [23] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. Oxford, U.K.: Clarendon, 1965.
- [24] T. W. Anderson, *An Introduction to Multivariate Analysis*. New York: Wiley, 1984.
- [25] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Process.*, vol. 40, pp. 590–600, Mar. 1992.
- [26] N. Yuen and B. Friedlander, "Asymptotic performance analysis of ESPRIT, higher-order ESPRIT, and virtual ESPRIT algorithms," *IEEE Trans. Signal Process.*, vol. 44, pp. 2537–2550, Oct. 1996.
- [27] B. Friedlander and A. J. Weiss, "On the second-order statistics of the eigenvectors of sample covariance matrices," *IEEE Trans. Signal Processing*, vol. 46, pp. 3136–3139, Nov. 1998.



Sebastiaan van der Tol was born in The Netherlands in 1977. He received the M.Sc. degree in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 2004, where, since January 2004, he has been a research assistant and has been pursuing the Ph.D. degree in electrical engineering.

His current research interests include array signal processing and interference mitigation techniques for large phased-array radio telescopes.



Alle-Jan van der Veen (S'87–M'94–SM'02–F'05) was born in The Netherlands in 1966. He graduated (cum laude) from Department of Electrical Engineering, Delft University of Technology, Delft, The Netherlands, in 1988, from which he received the Ph.D. degree (cum laude) in 1993.

Throughout 1994, he was a postdoctoral scholar at Stanford University, Stanford, CA. At present, he is a Full Professor in the Signal Processing Group of DIMES, Delft University of Technology. His research interests are in the general area of system

theory applied to signal processing and, in particular, to algebraic methods for array signal processing.

Dr. van der Veen received the 1994 and the 1997 IEEE SPS Young Author paper award and was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1998 to 2001. He was chairman of the IEEE SPS SPCOM Technical Committee from 2002 to 2004 and is an Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS.