INTRODUCTION TO INTERFERENCE MITIGATION TECHNIQUES IN RADIO ASTRONOMY

Amir Leshem and Alle-Jan van der Veen
Information Technology and Systems
Delft University of Technology
2628 CD Delft, The Netherlands
Email: leshem,allejan@cas.et.tudelft.nl

Radio-astronomical observations are increasingly corrupted by RF interference. Online detection and filtering algorithms are becoming essential. To facilitate the introduction of such techniques into radio astronomy, we formulate the astronomical problem in an array signal processing language, and give an introduction to some elementary algorithms from that field. We consider two topics in detail: interference detection by rank estimation of short-term covariance matrices, and spatial filtering by subspace estimation and projection. Finally, we discuss modifications necessary in the deconvolution step after spatial filtering, and propose a simple algebraic solution for self-calibration.

1 INTRODUCTION

Radio-astronomical observations are increasingly corrupted by RF interferers such as wireless communication and satellite navigation signals. Online detection and filtering algorithms are essential to reduce the effect of interference to an acceptable level. Until now, the most widely implemented algorithm is a single-channel total power change detector, followed by a blanking of the correlator output. Friedman [4] has implemented an improved power detection, which is based on detection of change in the power. Weber [16] proposed the use of the quantized correlation at all lags to test the presence of interference. However these are all single channel detectors which do not exploit the spatial properties of the interference. The only detector which used interferometry for on-line blanking was proposed in Kasper et al. [5] for low frequency interferometry, where a robust data censoring method based on the temporal behavior of the cross spectrum was proposed. The main drawback was the need of large number of estimated spectra (10^5) in order to obtain reliable robust estimates, and the fact that only two channels have been used. Recently Barnbaum and Bradely have proposed to use adaptive filters to excise interferer from the Green-Bank radio telescope using an LMS type algorithm [1].

The aim of the STW NOEMI project is to investigate the merits of *multichannel* detection and filtering algorithms at the Westerbork Synthesis Radio Telescope (WSRT). By combining correlation information from several sensors, we can increase the detection performance significantly, and also estimate the spatial signature of interferers. For this, it is essential to work in narrow sub-bands and detect the rank and dominant eigenvectors of each correlation matrix [9]. After detection, the interference is reduced by rejecting corrupted time-frequency slots (blanking) or projecting out certain direction vectors (spatial filtering).

The first approach is suitable for time-slotted communication signals such as in the GSM band, whereas the second will greatly enhance the performance of continuously-present interference. We describe two rank estimators: a generalized likelihood ratio test, and the minimum-description length (MDL). In a more challenging direction, adaptive excision of RFI from astronomical observations can be done by estimating spatial signature vectors of the interferers from short-term correlation matrices, followed by spatial filtering using projections. Such an approach is very promising, but the usual self-calibration process needs to be modified. For this reason, we reformulate the radio astronomical deconvolution step as a parameter estimation problem. This provides a better understanding of the CLEAN algorithm, extending the work of Schwarz [12].

The structure of the paper is as follows. After posing the astronomical measurement equations in section 2, we reformulate the model in terms of array processing matrix language in section 3. We then introduce RF interference and describe its effect on the received data. In section 5 we discuss various detection algorithms. We compare the single and multichannel detectors, for the case of a narrow-band interferer with known spatial signature vector, and then present two multichannel detectors that do not assume this

Perspectives in Radio Astronomy: Technologies for Large Antenna Arrays, A.B. Smolders and M.P. van Haarlem (eds.), Dwingeloo, April 1999.

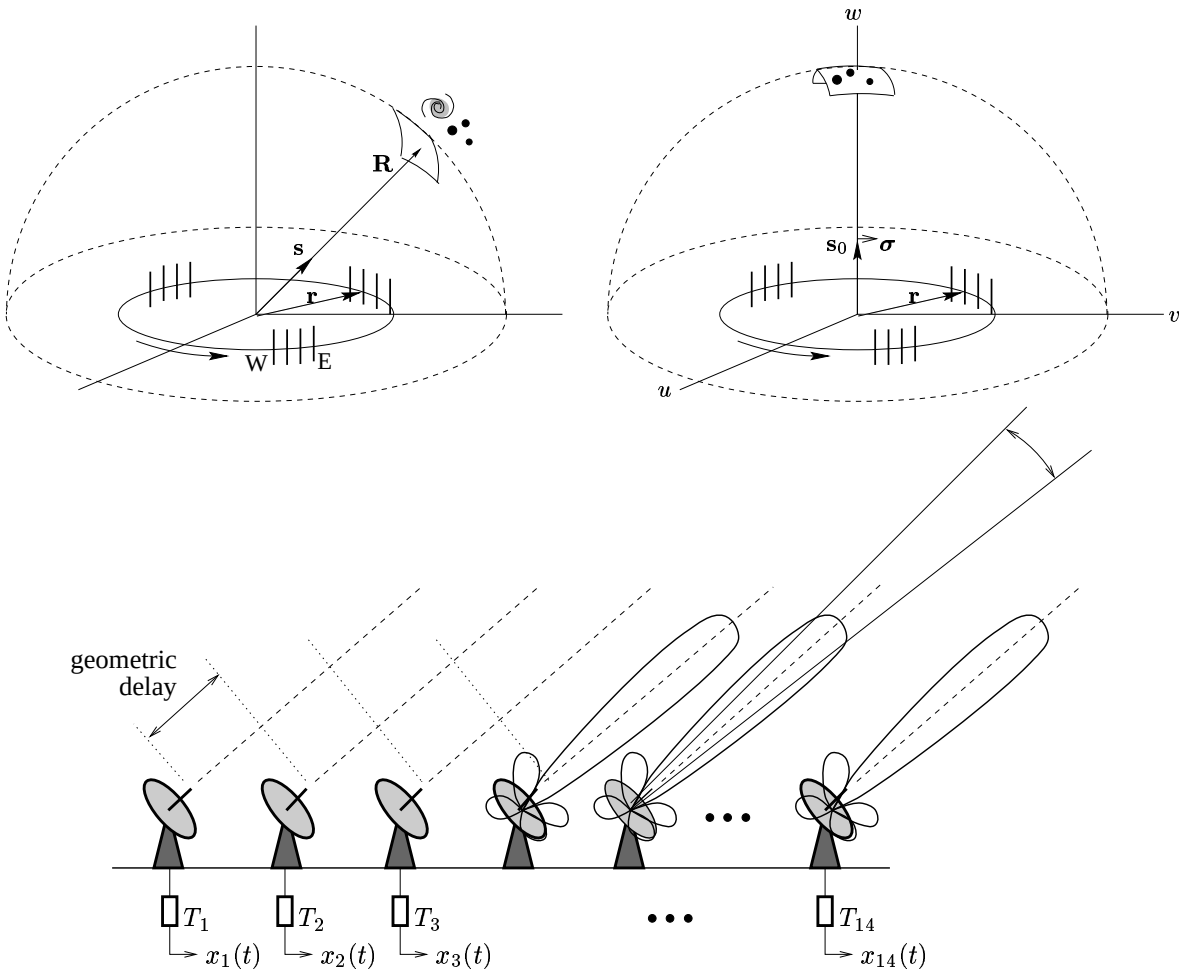

Figure 1: (a) The emitted electrical field from the celestial sphere is received by a rotating telescope array; (b) after geometrical delay compensation, the objects in the look direction appear to be at the zenith; (c) geometrical delay compensation

knowledge. We then move to spatial filtering techniques. In section 6 we formulate the basic ideas and describe a projections based approach, computational issues and other possible approaches. In section 7, we discuss the effect of spatial filtering on imaging, first in connection to Fourier-based techniques, then reformulated as a parameter estimation problem, for which we would like to find the exact maximum likelihood solution. The CLEAN and self-calibration algorithms are posed in matrix form and used to derive a closed-form solution to the calibration step. We finish by describing the effect of spatial filtering on the imaging step.

2 ASTRONOMICAL MEASUREMENT EQUATIONS

In this section we describe a simplified mathematical model for the astronomical measurement and imaging process. Our discussion follows [10]. We begin with the measurement equation, and reformulate it into a matrix form. This allows us to obtain a uniform description of various astronomical imaging operations such as deconvolution and self-calibration.

2.1 Measurement equation

The signals received from the celestial sphere may be considered as spatially incoherent wideband random noise. It is possibly polarized and perhaps contains spectral absorption or emission lines. Rather than

considering the emitted electric field $\mathcal{E}_f(\mathbf{R})$ at a location $\mathbf{R}$ on the celestial sphere, where f is a specific frequency, astronomers try to recover the *intensity* (or brightness) in the direction of unit-length vectors $\mathbf{s}$,

$$I_f(\mathbf{s}) = \langle |\mathcal{E}_f(\mathbf{s})| \rangle |\mathbf{R}|^2 \,.$$

Let $E_f(\mathbf{r})$ be the received celestial electric field at a location $\mathbf{r}$ on earth (see figure $1(a)$). The measured correlation of the electric fields between two sensors i and j with locations $\mathbf{r}_i$ and $\mathbf{r}_j$ is called a *visibility* and is (approximately) given by [10]

$$V_f(\mathbf{r}_i, \mathbf{r}_j) := \langle E_f(\mathbf{r}_i) E_f(\mathbf{r}_j)^* \rangle \;=\; \int I_f(\mathbf{s}) e^{-j2\pi f \; \mathbf{s}^T(\mathbf{r}_i - \mathbf{r}_j)/c} \; d\Omega \,.$$

Note that it is only dependent on the oriented distance $\mathbf{r}_i - \mathbf{r}_j$ between the two telescopes; this vector is called a *baseline.*

2.2 Point source model

For simplification, we may sometimes assume that the astronomical sky is a collection of d discrete point sources (maybe unresolved). This gives

$$I_f(\mathbf{s}) = \sum_k I_f(\mathbf{s}_k) \delta(\mathbf{s} - \mathbf{s}_k) \,,$$

where $\mathbf{s}_k$ is the coordinate of the k'th source, and thus

$$V_f(\mathbf{r}_i, \mathbf{r}_j) \;=\; \sum_{k=1}^{d} I_f(\mathbf{s}_k) \, e^{-j2\pi f \; \mathbf{s}_k^T(\mathbf{r}_i - \mathbf{r}_j)/c} \,. \tag{1}$$

2.3 (u, v) coordinate system

Up to this point we have worked in an arbitrary coordinate system. For earth rotation synthesis arrays, a coordinate system is often introduced as follows. We assume an array with telescopes that have a small field of view and that track a reference source location in the sky. The *geometrical delay* associated to this location is compensated by introducing a slowly time-variant delay in the first stage of the receiver (see figure $1(c)$). The result is that the reference location now appears to be located at the zenith (see figure $1(b)$). Other locations in the field of view can be written as

$$\mathbf{s} = \mathbf{s}_0 + \boldsymbol{\sigma} \,, \qquad \mathbf{s}_0 \perp \boldsymbol{\sigma} \,,$$

(valid for small $\boldsymbol{\sigma}$) and a natural coordinate system is

$$\mathbf{s}_0 = [0,\, 0,\, 1] \,, \qquad \boldsymbol{\sigma} = [\ell,\, m,\, 0] \,.$$

Similarly, for a planar array, the receiver baselines can be parameterized as

$$\mathbf{r}_i - \mathbf{r}_j = \lambda [u,\, v,\, 0] \,, \qquad \lambda = \frac{c}{2\pi f} \,.$$

Note that this coordinate system is frequency dependent.

The measurement equation in (u, v) coordinates thus becomes

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

It has the form of a Fourier transformation. The function $V_f(u, v)$ is sampled at various coordinates (u, v) by first of all taking all possible sensor pairs i, j, and second by realizing that the sensor locations $\mathbf{r}_i, \mathbf{r}_j$ are actually time-varying since the earth rotates. Given sufficient samples in the (u, v) domain, the relation can be inverted to obtain an image (the 'map'), which is the topic in section 7.

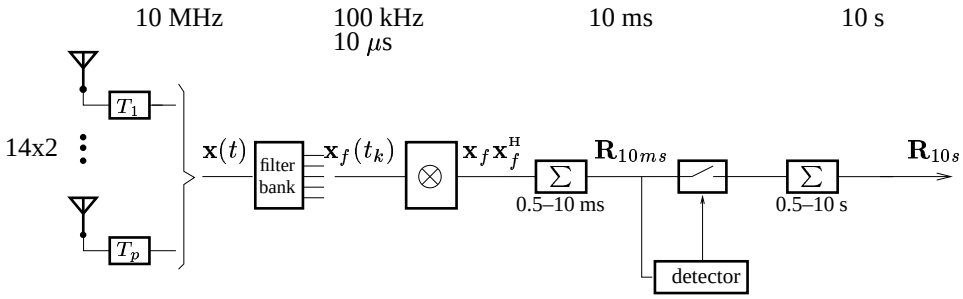

Figure 2: The astronomical correlation process

3 ARRAY SIGNAL PROCESSING FORMULATION

3.1 Obtaining the measurements

We will now describe the situation from an array signal processing point of view. The signals received by the telescopes are amplified and moved to baseband. The first stages also implement a time-varying delay for every telescope, to compensate for the geometrical delay and thus to synchronize the center of the field-of-view. Following traditional array signal processing practices, the signals at this point are called $x_i(t)$ rather than $E_f(\mathbf{r})$, and are stacked in vectors

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_p(t) \end{bmatrix},$$

where p is the number of telescopes. They are then processed by a correlation stage.

The main output of the telescope hardware is a sequence of correlation matrices, $\mathbf{R}_f(t_k)$, for a set of frequencies $\{f_k\}$ covering a 10 MHz band or so, and for a set of times $\{t_k\}$ covering 12 hours. Each correlation matrix $\mathbf{R}_f(t_k)$ is an average over 10 to 30 seconds of cross-correlations of the p telescope signals $\mathbf{x}_f(t)$, which is $\mathbf{x}(t)$ sub-band filtered at frequency f:

$$\mathbf{R}_f(t_k) \;=\; \mathrm{E}\{\mathbf{x}_f(t_k)\mathbf{x}_f(t_k)^{\mathrm{H}}\} \quad\sim\quad \hat{\mathbf{R}}_f(t_k) = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{x}_f(t_k+nT)\mathbf{x}_f(t_k+nT)^{\mathrm{H}},$$

where $NT \approx 10$ s (the superscript $^{\mathrm{H}}$ denotes a complex conjugate transpose). This is drawn schematically in figure 2 (ignoring the detection stage for the moment).[a] The matrices $\mathbf{R}_f(t_k)$ are stored for off-line spectral analysis and imaging.

Typically, each sub-band has a bandwidth in the order of 100 kHz or less. Due to the sub-band filtering, the original sampling rate of $\mathbf{x}(t)$ is reduced accordingly, resulting in T in the order of 10 μs and the number of samples N in the order of 10^5.

The connection of the correlation matrices $\mathbf{R}_f(t_k)$ to the visibilities $V_f(u,v)$ in section 2 is as follows. Each entry $r_{ij}(t_k)$ of $\mathbf{R}_f(t_k)$ is a sample of this visibility function for a specific coordinate (u,v), corresponding to the baseline vector $\mathbf{r}_i - \mathbf{r}_j$ between telescopes i and j at time t_k:

$$\mathbf{r}_i(t_k) - \mathbf{r}_j(t_k) = \lambda[u_{ik} - u_{jk},\ v_{ik} - v_{jk},\ 0]$$

$$V_f(u_{ik} - u_{jk}, v_{ik} - v_{jk}) \equiv r_{ij}(t_k)\,.$$

Note that we can obtain only a discrete set of (u,v) points. Indeed, the number of instantaneous independent baselines between p antennas is less than $\frac{1}{2}p(p-1)$. Also, using the earth rotation, the number

[a] Many telescope sites including WSRT follow actually a different scheme where the signals are first correlated at several lags and subsequently Fourier transformed. This leads to similar results.

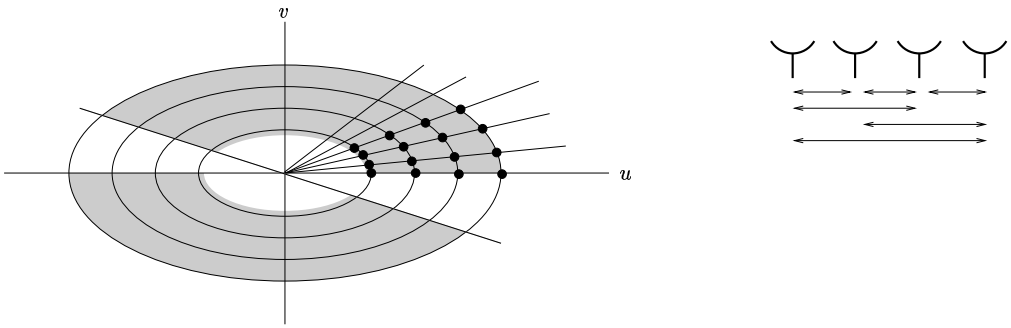

Figure 3: Sampled (u_k, v_k) points in the (u, v) plane for an East-West line array using the earth rotation

of samples $\{t_k\}$ is given by the ratio of the observation time and the covariance averaging time (e.g., 12 h/30 sec = 1440 samples). As seen in figure 3, the available sample coordinates $\{u_k, v_k\}$ give an irregular cover of the (u, v) plane.

3.2 Matrix formulation

For the discrete source model, we can now formulate our measurement equations in terms of matrices. Indeed, consider equation (1), written slightly different as

$$\begin{aligned} V_f(\mathbf{r}_i(t_k), \mathbf{r}_j(t_k)) &= \sum_{n=1}^{d} e^{-j2\pi f\, \mathbf{s}_n^T(\mathbf{r}_i - \mathbf{r}_0)/c}\, I_f(\mathbf{s}_n)\, e^{j2\pi f\, \mathbf{s}_n^T(\mathbf{r}_j - \mathbf{r}_0)/c} \\ \Leftrightarrow \quad V_f(u_{ik} - u_{jk}, v_{ik} - v_{jk}) &= \sum_{n=1}^{d} e^{-j(u_{ik}\ell_n + v_{ik}m_n)}\, I_f(\ell_n, m_n)\, e^{-j(u_{jk}\ell_n + v_{jk}m_n)}\,. \end{aligned}$$

We have introduced an arbitrary and time-varying reference point $\mathbf{r}_0(t_k)$, typically at the center of the array, so that only the phase difference with respect to that point will be of interest. The (u, v) coordinates are taken with respect to that point.

In terms of correlation matrices, this equation can be written as

$$\mathbf{R}_{k,f} \;=\; \mathbf{A}_{k,f}\mathbf{B}_f\mathbf{A}_{k,f}^{\mathrm{H}} \tag{2}$$

where

$$\mathbf{A}_{k,f} = [\mathbf{a}_{k,f}(\mathbf{s}_1), \cdots, \mathbf{a}_{k,f}(\mathbf{s}_d)]$$

$$\mathbf{a}_{k,f}(\mathbf{s}) = \begin{bmatrix} e^{-j2\pi f\, \mathbf{s}^T(\mathbf{r}_1(t_k) - \mathbf{r}_0(t_k))/c} \\ \vdots \\ e^{-j2\pi f\, \mathbf{s}(t_k)^T(\mathbf{r}_p(t_k) - \mathbf{r}_0(t_k))/c} \end{bmatrix} = \begin{bmatrix} e^{-j(u_{1k}\ell + v_{1k}m)} \\ \vdots \\ e^{-j(u_{pk}\ell + v_{pk}m)} \end{bmatrix}$$

and

$$\mathbf{B}_f = \begin{bmatrix} I_{1,f} & & \\ & \ddots & \\ & & I_{d,f} \end{bmatrix}, \qquad I_{n,f} = I_f(\mathbf{s}_n) = I_f(\ell_n, m_n)\,.$$

The vector function $\mathbf{a}_{k,f}(\mathbf{s})$ or $\mathbf{a}_{k,f}(\ell, m)$ is called the *array response vector* in array signal processing. It describes the response to a source in the direction $\mathbf{s}$ (or equivalently (ℓ, m)). As usual, the array response is frequency dependent. In this case, the response is also time-varying due to the earth rotation. Note, very importantly, that the function as shown here is completely known, since we know the locations of the telescopes very well.

3.3 Matrix formulation extended with antenna patterns

More realistically, the array response is less perfect. The telescopes are not omnidirectional: each telescope has its own beamshape (antenna pattern). Let $A_i(\mathbf{s})$ be the (complex) amplitude response of the array towards direction $\mathbf{s}$, then the array response vector $\mathbf{a}_{k,f}(\mathbf{s})$ has to be redefined as

$$\begin{bmatrix} A_1(\mathbf{s}) & & \\ & \ddots & \\ & & A_p(\mathbf{s}) \end{bmatrix} \begin{bmatrix} e^{-j2\pi f\, \mathbf{s}^T(\mathbf{r}_1-\mathbf{r}_0)/c} \\ \vdots \\ e^{-j2\pi f\, \mathbf{s}(t_k)^T(\mathbf{r}_p-\mathbf{r}_0)/c} \end{bmatrix} .$$

A significant simplification is possible if the telescopes are identical, up to unknown but direction independent gain factors. Then the $A_i(\mathbf{s})$ can be factored as $A_i(\mathbf{s}) = \gamma_i \tilde{A}(\mathbf{s})$. The function $\tilde{A}(\mathbf{s})$ is known from calibration and the same for all telescopes, and can be incorporated in the source intensity $I_f(\mathbf{s})$, i.e., the $\mathbf{B}$-factor. Thus we can leave it out of the discussion from now on. The gains γ_i may be different among the telescopes, as they also contain many other angle-independent effects such as cable losses and amplifier gains. In practice, they may also be (slowly) time-varying due to atmospheric conditions, so we will write $\gamma_{i,k}$. The model now becomes

$$\mathbf{R}_{k,f} \;=\; \mathbf{\Gamma}_k \mathbf{A}_{k,f} \mathbf{B}_f \mathbf{A}_{k,f}^{\mathrm{H}} \mathbf{\Gamma}_k^{\mathrm{H}}$$

where

$$\mathbf{\Gamma}_k = \begin{bmatrix} \gamma_{1,k} & & \\ & \ddots & \\ & & \gamma_{p,k} \end{bmatrix}$$

contains the unknown complex gains. In future equations we will drop the dependence on f.

3.4 Additive noise

In reality, most of the received signal consists of additive system noise. When this noise is zero mean, independent among the antennas (thus spatially white), and identically distributed, then asymptotically in N

$$\mathbf{R}_k = \mathbf{A}_k \mathbf{B} \mathbf{A}_k^{\mathrm{H}} + \sigma^2 \mathbf{I} \,. \tag{3}$$

Thus, the noise averages out everywhere, except on the main diagonal. Usually the noise is assumed to be Gaussian.

When we have also unknown complex gains for each antenna, then

$$\mathbf{R}_k = \mathbf{\Gamma}_k \mathbf{A}_k \mathbf{B} \mathbf{A}_k^{\mathrm{H}} \mathbf{\Gamma}_k^{\mathrm{H}} + \sigma^2 \mathbf{I} \,. \tag{4}$$

Note that this assumes that the noise is introduced *after* the gains.

In reality, the system noise is slightly different at each receiver. It is reasonable to assume that the noise is spatially white: the noise covariance matrix is diagonal. We can assume that it can be estimated using various calibration techniques; a simple diagonal scaling will then bring us back to the model (4). We further assumed that the quantization is fine, since a large dynamic range is needed to cope with strong interferers.

4 RF INTERFERENCE

RF interference usually enters the antennas through the sidelobes of the main beam. It can be stronger or weaker than the system noise. An important property is that it has a certain *directivity*, so that it does not average out in the correlation process.

Examples of harmful RFI present at WSRT are television broadcasts (Smilde station), geolocation satellites (GPS, Glonass), taxi dispatch systems, airplane communication and navigation signals, wireless communication (GSM) and satellite communication signals (Iridium). Thus, interferers may be continuous or intermittent, narrow-band or wideband, and strong or weak.

4.1 Delays of narrow-band signals

Let us digress for a moment and recall a well-known but important property of narrow-band signals, which says that a short time delay translates to a phase shift. In signal processing, narrow-band signals are usually represented by their low-pass equivalents [11]. A real-valued bandpass signal with center frequency f_c, such as received by an antenna, can be written as

$$z(t) = \text{real}\{s(t) \cdot e^{j2\pi f_c t}\}$$

where the baseband signal $s(t)$ is the complex envelope of the received signal $z(t)$. It is obtained from $z(t)$ by demodulation.

In array signal processing, we are interested in the effect of small delays on the narrow-band baseband signal $s(t)$. A delay τ on $z(t)$ results in

$$z_\tau(t) := z(t-\tau) = \text{real}\{s(t-\tau)e^{-j2\pi f_c \tau} \cdot e^{j2\pi f_c t}\}\,,$$

so that the complex envelope of the delayed signal is $s_\tau(t) = s(t-\tau)e^{-j2\pi f_c \tau}$. If the bandwidth W of $s(t)$ is sufficiently small so that $\exp(j2\pi W\tau) \approx 1$, then standard Fourier analysis yields $s(t-\tau) \approx s(t)$, so that

$$s_\tau(t) \approx s(t)e^{-j2\pi f_c \tau} \qquad \text{for } W\tau \ll 1\,.$$

The well-known conclusion is that, for narrow-band signals, time delays shorter than the inverse bandwidth amount to phase shifts of the baseband signal. This is fundamental in phased array signal processing.

4.2 Narrow-band interference model

Suppose that we have a single interferer impinging onto the telescope array. The interfering signal reaches the array with different delays τ_i for each telescopes. After demodulation to baseband, we have

$$x_i(t) = a_i\, s(t-\tau_i)\, e^{-j2\pi f_c \tau_i}\,, \qquad i = 1, \cdots, p\,.$$

Here, a_i represents the telescope gain in the direction of the interferer, including any possible attenuation of the channel. Unlike much of the array signal processing literature, the a_i are likely to be different for each telescope since the interferer is typically in the near field. This implies that it impinges on each telescope at a different angle, whereas the response of the telescopes is not omni-directional.

The model can be simplified if the narrow-band assumption holds: in that case $s(t-\tau_i) = s(t)$. Note that we have already assumed before that the signals are sub-band filtered. Let W be the bandwidth of the sub-band filters. In WSRT, the largest baseline is 3000 m, corresponding to a maximal delay of 10 μs. Hence the narrow-band assumption holds if $W \ll 100$ kHz [9]. Under this condition, we can stack the p telescope outputs from a particular sub-band filter in a vector $\mathbf{x}_f(t)$ and write

$$\mathbf{x}_f(t) = \begin{bmatrix} a_1 e^{-j2\pi f_c \tau_1} \\ \vdots \\ a_p e^{-j2\pi f_c \tau_p} \end{bmatrix} s(t) \;=:\; \mathbf{a}s(t)\,.$$

As before, $\mathbf{a}$ is an array response vector. Unlike before, it is not a simple or known function of the direction of the interferer, since we are in the near field and the sidelobes of the array are not calibrated. The vector is also called the *spatial signature* of the interfering source.

Similarly, with q interferers,

$$\mathbf{x}_f(t) = \sum_{j=1}^{q} \mathbf{a}_j s_j(t) = \mathbf{A}_s \begin{bmatrix} s_1(t) \\ \vdots \\ s_q(t) \end{bmatrix}, \qquad \mathbf{A}_s = [\mathbf{a}_1, \cdots, \mathbf{a}_q] .$$

The subscript 's' is used to distinguish $\mathbf{A}_s$ from the array response matrix of the astronomical sources. The corresponding correlation matrix and its estimate are

$$\mathbf{R}_k = \mathrm{E}\{\mathbf{x}_f(t_k)\mathbf{x}_f(t_k)^{\mathrm{H}}\} = (\mathbf{A}_s)_k(\mathbf{R}_s)_k(\mathbf{A}_s^{\mathrm{H}})_k , \qquad \hat{\mathbf{R}}_k = \frac{1}{M}\sum_{m=0}^{M-1} \mathbf{x}_f(t_k + mT)\mathbf{x}_f^{\mathrm{H}}(t_k + mT) .$$

How well the estimate fits to $\mathbf{R}_k$ depends on the stationarity of the scenario, and is open to discussion. $(\mathbf{R}_s)_k$ depends on the second-order properties of the interfering signals. The power of television signals will be stationary over long periods (order tens of seconds or better). At the other extreme, communication signals such as GSM are time slotted: time is partitioned into frames of about 5 ms and frames are partitioned into 8 slots. Each user can transmit only during its slot of 0.577 ms and then has to be silent for 7 times this period before transmitting again in the next frame. Thus, there is a short-term stationarity (over 0.577 ms), and a cyclostationarity with periods of about 5 ms.

The stationarity of $\mathbf{A}_s(t_k)$ depends on the stationarity of the location of the interferer, its distance, and the orientation of the telescopes. With multipath fading, a mobile interferer only has to move about 30 cm to create a different $\mathbf{a}$-vector, giving a stationarity in the order of 10–100 ms for a GSM user. Even for a fixed interferer such as a television station, the rotation of the telescopes will change the $\mathbf{a}$-vector within a fraction of a second, either because of multipath fading or because the interferer moves through the highly variable sidelobe pattern. After the $\mathbf{a}$-vector has changed, the interferer effectively looks like a new interferer, thus increasing the value of q.

The conclusion is that $\hat{\mathbf{R}}_k$ is a useful estimate only over short averaging periods over which the interference is stationary, say MT in the order of msec. Thus, $M \ll N$.

4.3 Overall model: astronomical signals with interference and noise

In summary, the model that we have derived is as follows:

$$\mathbf{R}_k = \mathbf{\Gamma}_k\mathbf{A}_k\mathbf{B}\mathbf{A}_k^{\mathrm{H}}\mathbf{\Gamma}_k + (\mathbf{A}_s)_k(\mathbf{R}_s)_k(\mathbf{A}_s)_k^{\mathrm{H}} + \sigma^2\mathbf{I}, \qquad k = 0, 1, \cdots .$$

Here, $\mathbf{A}_k$: $p \times d$ is the array response matrix of the d discrete sources in the sky. Its columns are known functions of the (unknown) locations of the sources. It is a (very) wide matrix: $d \gg p$, and assumed stationary over 10 s. $\mathbf{B}$: $d \times d$ is a diagonal matrix (positive real) containing the brightness of each source, and assumed time-invariant over the complete observation. $\mathbf{\Gamma}_k$ are diagonal matrices (positive real) representing unknown and slowly varying antenna gains.

$\mathbf{A}_s$: $p \times q$ is the array response matrix of the q interferers. It is likely to be unstructured. We will consider cases where $q < p$, so that $\mathbf{A}_s$ is tall. $\mathbf{R}_s$: $q \times q$ is the interference correlation matrix. $\mathbf{A}_s$ and $\mathbf{R}_s$ are usually stationary only over very short time spans.

$\sigma^2\mathbf{I}$ is the noise covariance matrix, assuming white i.i.d. noise for simplicity. The noise power σ^2 is often rather well known.

$\|\mathbf{A}_k\mathbf{B}\mathbf{A}_k^{\mathrm{H}}\|$, i.e., the observed power of the astronomical sources, is at least two orders of magnitudes smaller than σ^2, and for the purpose of detection, it can be ignored. In contrast, $\|(\mathbf{A}_s)_k(\mathbf{R}_s)_k(\mathbf{A}_s)_k^{\mathrm{H}}\|$ can be of comparable magnitude.

5 INTERFERENCE DETECTION

Ideally, the output of the correlation process are clean estimates of $\mathbf{A}_k \mathbf{B} \mathbf{A}_k^{\mathrm{H}}$, once every 10 s or so. In principle, we estimate it by

$$\hat{\mathbf{R}}^{10s}(t_k) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{x}_f(t_k + nT) \mathbf{x}_f^{\mathrm{H}}(t_k + nT) \,, \qquad NT = 10 \text{ s}. \tag{5}$$

As we have seen, these estimates are corrupted by interference and additive system noise, and unknown antenna gains. The objective of interference detection and rejection schemes is to improve the *signal to interference and noise ratio* (SINR) at the output of the integrators, i.e., at the 10 s level. Interference that is stationary at these time scales or longer can often be treated off-line. In this paper we consider *online* interference detection and excision schemes, assuming stationarity at msec time scales or less.

Many interference detection schemes exist. They differ by the amount of knowledge that we can assume on the interfering signals. E.g., if we know the signal wave form, then the optimal detector has the form of a matched filter. Extensions are possible if the waveform is known up to a few parameters such as amplitude, phase or frequency. However, usually the signal is modulated by a message and hence effectively unknown. There are two classes of detection techniques: more or less deterministic methods that exploit known properties of the signals such as modulation type or certain periodicities, and those based on statistical models with unknown parameters, leading to Generalized Likelihood Ratio Tests (GLRT), a particular example of which is power detection.

In principle, we can say that man-made interference is expected to be statistically different from the astronomical sources. Although this is a very attractive feature, it is not easy to use these properties for detection or excision, since the long averaging periods and the central limit theorem tend to jointly Gaussianize the interferers. However for strong narrow-band interferers these methods are expected to give improved suppression at an additional computational power [7].

Another distinction between interferers and astronomical signals is their spatial signature vectors. Astronomical signals enter through the main lobe of the telescopes and have a very structured (parametrically known) array response, which is used for imaging. The interferers usually enter through the sidelobes and are in the near field, leading to unstructured $\mathbf{a}$-vectors. Also, their location relative to the array is not correlated with the motion of the earth. It might even remain fixed relative to the array during the complete observation period (e.g., TV transmitters). Since the array tracks a fixed region in the sky, the directional vector of the interference is typically time varying.

From all possibilities, we consider here two schemes:

– *Multichannel interference detection and excision.* The interference is detected at short time scales (ms), and contaminated samples are removed from the averaging process in (5). This will work well if the interference is concentrated in frequency and time, as e.g., in the GSM system.

– *Spatial filtering.* This more ambitious scheme is also suitable for continuously present interference such as TV stations. After detection, we estimate the spatial signature of the interferer and project out that dimension or otherwise subtract the signal coming from that direction.

For the purpose of power detection schemes, it is sufficient to look at (short-term) correlation matrices based on measurement data in a window of length MT:

$$\hat{\mathbf{R}}_k = \frac{1}{M} \sum_{m=0}^{M-1} \mathbf{x}_f(t_k + mT) \mathbf{x}_f^{\mathrm{H}}(t_k + mT) \,, \qquad MT \approx 10 \text{ ms}.$$

If an interferer is detected in this analysis window, it is discarded, otherwise the data is accepted and the correlation matrix is used in the formation of a clean estimate of $\hat{\mathbf{R}}^{10s}(t_k)$. See figure 2. Obviously, many variations are possible, such as sliding window techniques, or discarding neighbors of contaminated samples as well (perhaps both in time and frequency).

In this section we propose sub-band detection methods and analyze their performance. Spatial filtering is discussed in section 6.

5.1 Single channel spectral detector

Detection theory is based on hypothesis testing. We test $\mathcal{H}_0$: there is no interference, versus $\mathcal{H}_1$: there is at least one interferer in this band. The implementation of this test depends on the model that we pose for the interferer. We will first discuss some particularly simple cases which will allow analysis.

Thus let us consider the single-channel case first. We assume that there is at most a single interferer, where the interfering signal is i.i.d. Gaussian noise with unknown power σ_s^2. The background noise is white Gaussian with known power σ^2.

Without interferer, the observed data samples x_m are complex normal distributed, with zero mean and variance σ^2. With an interferer, this distribution is still complex normal, but with variance $\sigma_s^2 + \sigma^2$. Thus, we test the hypothesis

$$\begin{array}{lll} \mathcal{H}_0: & x_m & \sim \mathcal{CN}(0,\, \sigma^2) \\ \mathcal{H}_1: & x_m & \sim \mathcal{CN}(0,\, \sigma_s^2 + \sigma^2)\,, \qquad m = 0, \cdots, M-1\,. \end{array}$$

We assume that we have available M samples $\{x_m\}$.

This is a rather standard problem in detection theory (cf. [6] for an introduction). A Neyman-Pearson detector selects $\mathcal{H}_1$ if the likelihood ratio,

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)}\,,$$

exceeds a threshold. It is known that this leads to an optimal probability of detection, given a certain probability of false alarm (detecting an interferer when there is none). In our case, the NP detector will compare the total received power to a threshold γ, deciding $\mathcal{H}_1$ if the test statistic

$$T(\mathbf{x}) := \frac{1}{\sigma^2} \sum_{m=0}^{M-1} |x_m|^2 \quad > \quad \gamma\,.$$

Under the above assumptions we can obtain closed form expressions for the probability of false alarm and the probability of detection. For this, recall that the sum of squares of M real i.i.d. zero-mean unit-variance Gaussian random variables has a χ^2 distribution with M degrees of freedom. Since we have complex samples, $T(\mathbf{x})$ is the sum of $2M$ real variables. Under $\mathcal{H}_0$, these have a variance $\frac{1}{2}$, hence the probability of false alarm is given by

$$P_{FA} \;:=\; P\left\{T(x) > \gamma\;;\; \mathcal{H}_0\right\} \quad = \quad Q_{\chi^2_{2M}}(2\gamma)$$

where $Q_{\chi^2_{2M}}(\gamma)$ is the tail probability of a χ^2 random variable with $2M$ degrees of freedom. It has a closed-form expression (cf. [6]):

$$Q_{\chi^2_{2M}}(2\gamma) \;=\; e^{-\gamma} \sum_{k=0}^{M-1} \frac{\gamma^k}{k!}\,.$$

Its inverse is known in terms of the inverse Gamma-function, and allows to select γ to obtain a desired level of false alarm. Similarly, the probability of detection of an interference at this threshold γ is given by

$$\begin{array}{rcl} P_D & := & P\{T(\mathbf{x}) > \gamma\;;\; \mathcal{H}_1\} \\ & = & P\left\{ \dfrac{2}{\sigma^2 + \sigma_s^2} \displaystyle\sum_{m=1}^{M} |x_m|^2 > \dfrac{2\gamma}{1 + \sigma_s^2/\sigma^2}\;;\; \mathcal{H}_1 \right\} \\ & = & Q_{\chi^2_{2M}}(\frac{2\gamma}{1+\mathrm{INR}}) \end{array} \tag{6}$$

where $\mathrm{INR} = \frac{\sigma_s^2}{\sigma^2}$ is the interference-to-noise ratio.

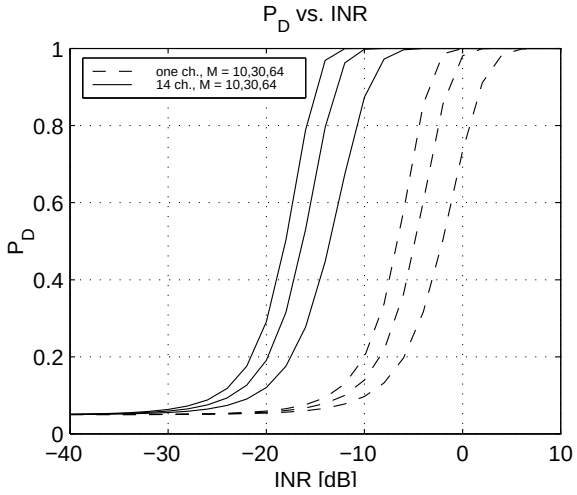

Figure 4: P_D vs. INR, for $M = 10,\ 30,\ 64$, and false alarm rate $P_{FA} = 5\%$

5.2 Multichannel detector with known spatial signature

A significant performance improvement is possible with a multichannel detector. To illustrate this, we assume again the simple case with at most a single narrow-band Gaussian interferer, with *known* spatial signature vector $\mathbf{a}$ in white Gaussian noise. Without interference, the data vectors $\mathbf{x}_m$ are complex normal distributed with zero mean and covariance matrix $\sigma^2\mathbf{I}$. With a single interferer, the covariance matrix becomes $\sigma_s^2\mathbf{a}\mathbf{a}^H + \sigma^2\mathbf{I}$. Thus,

$$\begin{aligned} \mathcal{H}_0: &\quad \mathbf{x}_m \sim \mathcal{CN}(0, \sigma^2\mathbf{I}) \\ \mathcal{H}_1: &\quad \mathbf{x}_m \sim \mathcal{CN}(0, \sigma_s^2\mathbf{a}\mathbf{a}^H + \sigma^2\mathbf{I}), \qquad m = 0, \cdots, M-1\,. \end{aligned}$$

The Neyman-Pearson detector considers the estimated data covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{M}\sum_{m=0}^{M-1} \mathbf{x}_m\mathbf{x}_m^{\mathrm{H}}$$

and is given by

$$T(\mathbf{X}) \quad := \quad \frac{1}{\sigma^2/M}\,\frac{\mathbf{a}^{\mathrm{H}}\hat{\mathbf{R}}\mathbf{a}}{\mathbf{a}^{\mathrm{H}}\mathbf{a}} \qquad \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \quad \gamma\,.$$

This test is recognized as a matched spatial filter detector; essentially we compare the received energy in the direction $\mathbf{a}$ of the interferer to σ^2. Taking the same threshold as in the single channel case will provide the same false alarm probability as before:

$$P_{FA} \;=\; P\{T(\mathbf{X}) > \gamma;\ \mathcal{H}_0\} \;=\; Q_{\chi^2_{2M}}(2\gamma)\,.$$

However, the probability of detection is now given by

$$P_D = P\{T(\mathbf{X}) > \gamma;\ \mathcal{H}_1\} \;=\; Q_{\chi^2_{2M}}\left(\frac{2\gamma}{1 + p\,\mathrm{INR}}\right).$$

Figure 4 presents the probabilities of detection as a function of interference to noise ratio for a single and for $p = 14$ channels. We have selected a threshold such that $P_{FA} = 5\%$, which means that without interference, we will throw away 5% of the data. We can clearly see that the probability of detection is greatly improved by moving to the multichannel case. The improvement is equal to the array gain, $10\log(p) = 11.5$ dB.

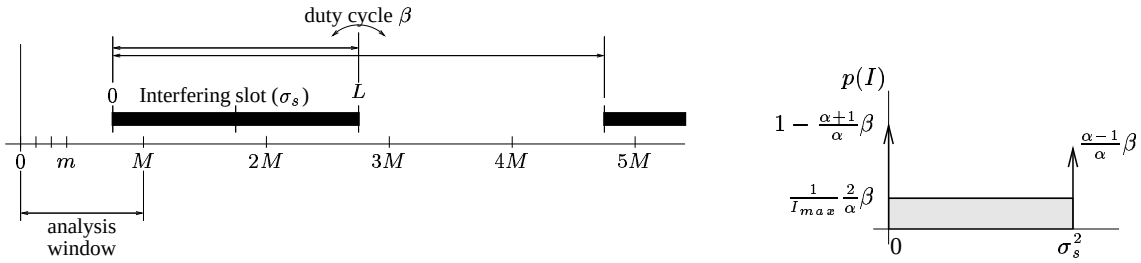

Figure 5: (a) Interferer with slot length $L = \alpha M$ samples, power σ_s^2 per on-sample, and duty cycle β. (b) Corresponding probability density of interference power in a single analysis window.

5.3 Single TDMA interferer with known spatial signature

Let us now consider a TDMA signal: an interferer which is periodically active in a fraction β of the time (see figure 5). Here, $0 < \beta < 1$ is known as the duty cycle of the periodic signal. Assume that the interferer is present in the selected frequency band and that the duration of the slot in which the interferer is active is equal to αM samples $\mathbf{x}_m$, where we take $\alpha > 1$. Let as before σ_s^2 denote the power of a single sample of the interferer when it is present.

Since the interfering slots need not be synchronized to the analysis window, a single interfering slot will give rise to two analysis windows in which the interferer is partially present, and possibly one or more analysis windows in which the interferer is present in all the samples. Since the interferer is time-slotted with duty cycle β, there will also be windows that contain no interference.

The corresponding probability density $p(I)$ of having a certain average interference power I per sample in an arbitrary analysis window of length M can be computed in closed form, as

$$p(I) = \begin{cases} (1 - \dfrac{\alpha+1}{\alpha}\beta)\,\delta(I)\,, & I = 0 \\ \dfrac{1}{I_{max}}\dfrac{2}{\alpha}\beta\,, & 0 < I < I_{max} \\ \dfrac{\alpha-1}{\alpha}\beta\,\delta(I - I_{max})\,, & I = I_{max}\,. \end{cases}$$

It is plotted in figure 5, where the vertical arrows indicate the unit impulse function. For example, for an interferer of strength σ_s^2 per sample when it is on, the maximal average interference power per sample is obviously σ_s^2, when all samples are contaminated. The probability of this is $(\alpha-1)/\alpha\,\beta$. Power densities less than σ_s^2 occur with a uniform distribution for analysis windows that are only partly corrupted, at the edges of the interference slot.

We can define

– the average interference power per sample before detection:

$$I_{eff} \;=\; \int I\,p(I)\,dI \;=\; \beta\,\sigma_s^2\,,$$

– the average interference power per sample after detection and blanking:

$$I_{res} \;=\; \int I\,(1 - P_D(I))\,p(I)\,dI\,,$$

– the fraction of number of samples kept after detection and blanking:

$$n_{res} \;=\; \int (1 - P_D(I))\,p(I)\,dI\,.$$

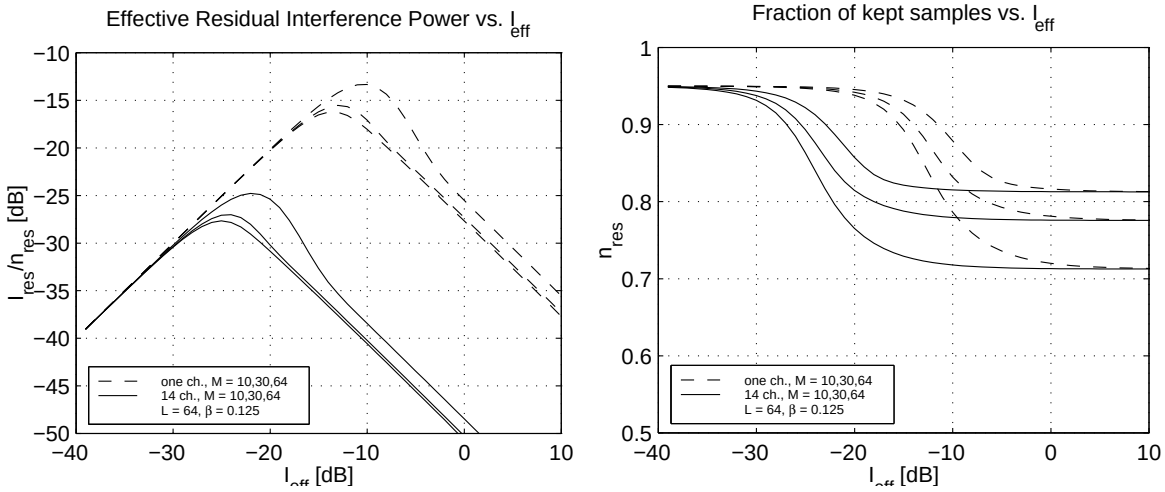

Figure 6: (a) Effective residual INR after blanking versus effective INR at the input; (b) fraction of remaining samples after blanking

Figure 6 shows the dependence of the residual INR as a function of M (the number of samples in an analysis block), for an interferer of length $L = 64$ sub-band samples, a duty cycle $\beta = 1/8$, and a false alarm rate of 5%. Obviously, very weak interference is not detected, and in that case we throw away 5% of the data due to the false alarm rate. High interference powers are easily detected, and almost all contaminated analysis windows will be detected and blanked. Only the tails of an interfering slot might be missed, so that there is still some interference remaining after detection. The worst case occurs for interference that is not strong enough to be detected all the time, but not weak enough to be harmless.

Several other interesting facts can be seen in these figures. The most important is the large performance gain in the multichannel approach, as compared to a single channel. As seen in figure 4, the effect of using an array is to shift the graphs of probability of detection to the left by the array gain, e.g., for the 14-channel detector the graph is shifted by 11.5 dB. Hence, we require 11.5 dB less interference power in order to detect it. However, the effective gain is given by the vertical distance between the graphs: this shows the amount of interference suppression for a given interference power. In figure 6 the suppression can be approximately 21 dB larger than that of the single antenna case.

A second interesting phenomenon is the fact that the interference suppression is almost the same for a large range of analysis windows M. Thus, we would take this window rather small, so that the the residual number of samples is larger. This effect is mainly due to the fact that the case of partial blocks with weaker power is less frequent as the analysis block becomes shorter.

5.4 Eigenvalue analysis

So far, we have looked at the detection problem from a rather idealistic viewpoint: at most 1 interferer, and a known spatial signature. The reason was that for this case, we could derive optimal detectors with closed-form expressions for the performance.

Our goal is the detection of the presence of an interferer from observed correlation data. As a start, let us first consider the (asymptotic) equation

$$\mathbf{R} = \mathbf{A}_s \mathbf{R}_s \mathbf{A}_s^{\mathrm{H}}$$

where $\mathbf{A}_s$ has size $p \times q$ and $\mathbf{R}_s$ has size $q \times q$. For a low number of interferers q, this brings us to familiar grounds in array signal processing, as it admits analysis by subspace-based techniques. We give a brief introduction here.

If $q < p$, then the rank of $\mathbf{R}$ is q. Thus, we can estimate the number of narrow-band interferers from a

rank analysis. This is also seen from an eigenvalue analysis: let

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{H}}$$

be an eigenvalue decomposition of $\mathbf{R}$, where the $p \times p$ matrix $\mathbf{U}$ is unitary ($\mathbf{U}\mathbf{U}^{\mathrm{H}} = \mathbf{I}$, $\mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}$) and contains the eigenvectors, and the $p \times p$ diagonal matrix $\mathbf{\Lambda}$ contains the corresponding eigenvalues ($\lambda_i \geq 0$). Since the rank is q, there are only q nonzero eigenvalues. We can collect these in a $q \times q$ diagonal matrix $\mathbf{\Lambda}_s$, and the corresponding eigenvectors in a $p \times q$ matrix $\mathbf{U}_s$, so that

$$\mathbf{R} = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^{\mathrm{H}} \,. \tag{7}$$

The remaining $p-q$ eigenvectors from $\mathbf{U}$ can be collected in a matrix $\mathbf{U}_n$, and they are orthogonal to $\mathbf{U}_s$ since $\mathbf{U} = [\mathbf{U}_s \;\; \mathbf{U}_n]$ is unitary. The subspace spanned by the columns of $\mathbf{U}_s$ is called the *signal subspace*, the orthogonal complement spanned by the columns of $\mathbf{U}_n$ is known as the *noise subspace* (although this is in fact a misnomer, since the noise is not confined to this space).

In the presence of white noise,

$$\mathbf{R} \;=\; \mathbf{A}_s\mathbf{R}_s\mathbf{A}_s^{\mathrm{H}} + \sigma^2\mathbf{I}_p \,.$$

In this case, $\mathbf{R}$ is full rank: its rank is always p. However, we can still detect the number of interferers by looking at the eigenvalues of $\mathbf{R}$. Indeed, the eigenvalue decomposition is derived as (expressed in terms of the previous decomposition (7))

$$\begin{aligned}
\mathbf{R} &= \mathbf{A}_s\mathbf{R}_s\mathbf{A}_s^{\mathrm{H}} + \sigma^2\mathbf{I}_p \\
&= \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^{\mathrm{H}} + \sigma^2[\mathbf{U}_s \;\; \mathbf{U}_n][\mathbf{U}_s \;\; \mathbf{U}_n]^{\mathrm{H}} \\
&= [\mathbf{U}_s \;\; \mathbf{U}_n] \left[\begin{array}{c|c} \mathbf{\Lambda}_s + \sigma^2\mathbf{I}_q & 0 \\ \hline 0 & \sigma^2\mathbf{I}_{p-q} \end{array}\right] \left[\begin{array}{c} \mathbf{U}_s^{\mathrm{H}} \\ \mathbf{U}_n^{\mathrm{H}} \end{array}\right]
\end{aligned} \tag{8}$$

hence $\mathbf{R}$ has $p-q$ eigenvalues equal to σ^2, and q that are larger than σ^2. Thus, we can detect q by comparing the eigenvalues of $\mathbf{R}$ to a threshold defined by σ^2.

But the eigenvalue decomposition shows more. Indeed, the columns of $\mathbf{U}_s$ span the same subspace as the columns of $\mathbf{A}_s$. This is clear in the noise-free case (7), but the decomposition (8) shows that the eigenvectors contained in $\mathbf{U}_s$ and $\mathbf{U}_n$ respectively are the same as in the noise-free case. Thus,

$$\mathrm{span}(\mathbf{U}_s) = \mathrm{span}(\mathbf{A}_s) \,, \qquad \mathbf{U}_n^{\mathrm{H}}\mathbf{A} = 0 \,. \tag{9}$$

Although we cannot directly identify each individual column of $\mathbf{A}$, we can at least determine its subspace from the observed covariance matrix. This can be used to filter out the interference — such spatial filtering algorithms are discussed in section 6. Note that it is crucial that the noise is spatially white. For colored noise, an extension (whitening) is possible but we have to know the coloring.

To see how this matches with the previous section, let us we specialize to the case of a single interferer, with spatial signature $\mathbf{a}$. It follows that the dominant eigenvector $\mathbf{u}_1 = \alpha\mathbf{a}$ (for some scaling α), so that

$$\frac{\mathbf{a}^{\mathrm{H}}\mathbf{R}\mathbf{a}}{\mathbf{a}^{\mathrm{H}}\mathbf{a}} = \frac{\mathbf{u}_1^{\mathrm{H}}\mathbf{R}\mathbf{u}_1}{\mathbf{u}_1^{\mathrm{H}}\mathbf{u}_1} = \lambda_1 \,.$$

Thus, the test statistic of the previous section reduces to testing the dominant eigenvalue of $\mathbf{R}$. With an interferer, $\lambda_1 = \sigma_s^2 + \sigma^2$, without interference it is $\lambda_1 = \sigma^2$. In the current case (asymptotic, $M \to \infty$), the threshold can be very sharp: just above σ^2. With finite M, however, the dominant eigenvector of the estimated covariance matrix $\hat{\mathbf{R}}$ will only be approximately equal to $\mathbf{a}$, the dominant eigenvector will only be approximately equal to $\sigma_s^2 + \sigma^2$, and the threshold has to be relaxed.

5.5 Multichannel detector with unknown spatial signature

In case we only have an estimate $\hat{\mathbf{R}}$ based on a finite amount of samples M, and the spatial signature vectors of the interference are unknown, there are no optimal results. The eigenvalue analysis suggested

that we should compare the eigenvalues to a threshold defined by σ^2. We will discuss two detectors, one for the case where σ^2 is known, and another one for which it is unknown.

If the noise power σ^2 is known, we can apply the likelihood ratio test (LRT), which leads to a method due to Box [2] for testing the null hypothesis that $\sigma^{-2}\hat{\mathbf{R}} = \mathbf{I}$ (no interference). The test statistic is given by

$$-Mp \log \prod_{i=1}^{p} \frac{\lambda_i}{\sigma^2} \quad \sim \quad \chi^2_{(p+1)(p-2)} \tag{10}$$

where λ_i is the i-th eigenvalue of $\hat{\mathbf{R}}$. This basically tests if all eigenvalues are equal to σ^2.

If also the noise power is unknown, we propose to use the Minimum Description Length (MDL) detector [15]. In this case, rather than setting a threshold based on the asymptotic distribution of the LRT, we try to find the correct model order which minimizes the description length of the data. The MDL rank estimator is given by

$$\hat{k} = \arg\min_k \mathrm{MDL}(k) \tag{11}$$

where

$$\mathrm{MDL}(k) \;=\; (p-k)M \log \frac{\frac{1}{p-k}\sum_{i=k+1}^{p} \lambda_i}{\left(\prod_{i=k+1}^{p} \lambda_i\right)^{\frac{1}{p-k}}} \;+\; \frac{1}{2}k(2p-k+1)\log M$$

and an interference is detected if $\hat{k} \neq 0$. This basically tests if the geometric mean of the eigenvalues is equal to the arithmetic mean, which is only true if all eigenvalues are equal to each other. This rank detector is simple to implement since it is independent of the varying SINR in the system.

Simulations of these detectors in a scenario with GSM interference have appeared in [8]. We are currently testing them on measured data, with good results.

6 SPATIAL FILTERING

Let us now assume that we have obtained a covariance matrix $\mathbf{R}$, which contains the rather weak covariance matrix of the astronomical sources (visibilities) $\mathbf{R}_v$, plus white noise. Suppose also that there is an interferer with power σ_s^2:

$$\mathbf{R} \;=\; \mathbf{R}_v \;+\; \sigma_s^2 \mathbf{a}\mathbf{a}^{\mathrm{H}} \;+\; \sigma^2 \mathbf{I}\,.$$

In the previous section, we considered schemes to detect the interferer from the eigenvalues of $\mathbf{R}$. After detection, we proposed to discard $\mathbf{R}$ if it is found to be contaminated, but what if the interferer is present all the time? In that case, it is more suitable to try to suppress its contribution $\sigma_s^2 \mathbf{a}\mathbf{a}^{\mathrm{H}}$. This leads to *spatial filtering* techniques.

6.1 Projecting out the interferer

A possibility is to null all energy with spatial signature $\mathbf{a}$. To this end, we can introduce the projection matrices

$$\mathbf{P}_{\mathbf{a}} = \mathbf{a}(\mathbf{a}^{\mathrm{H}}\mathbf{a})^{-1}\mathbf{a}^{\mathrm{H}}\,, \qquad \mathbf{P}_{\mathbf{a}}^{\perp} = \mathbf{I} - \mathbf{a}(\mathbf{a}^{\mathrm{H}}\mathbf{a})^{-1}\mathbf{a}^{\mathrm{H}}\,.$$

It is easily seen that $\mathbf{P}_{\mathbf{a}}^{\perp}\mathbf{a} = 0$, so that

$$\tilde{\mathbf{R}} = \mathbf{P}_{\mathbf{a}}^{\perp}\mathbf{R}\mathbf{P}_{\mathbf{a}}^{\perp} \;=\; \mathbf{P}_{\mathbf{a}}^{\perp}\mathbf{R}_v\mathbf{P}_{\mathbf{a}}^{\perp} \;+\; \sigma^2\mathbf{P}_{\mathbf{a}}^{\perp}\,. \tag{12}$$

Thus, the interference is removed. At the same time, the visibility matrix is modified by the projections, and the noise is not white anymore, since one dimension is missing. The imaging stage has to be aware of this, which is the topic of section 7. We do need to store the effective spatial filter on the 10s data.

Since all information is now confined to a $(p-1)$-dimensional space, we might as well define a smaller-size $(p-1)\times(p-1)$ covariance matrix. To this end, recall from equation (8) the eigenvalue decomposition of $\mathbf{R}$, and in particular the matrix containing an orthonormal basis of the "noise subspace" $\mathbf{U}_n$, which is the orthogonal complement of $\mathbf{a}$, with $p-1$ columns. According to (9), $\mathbf{a} \perp \mathbf{U}_n$, so that $\mathbf{P}_{\mathbf{a}}^{\perp} = \mathbf{U}_n\mathbf{U}_n^{\mathrm{H}}$. As an alternative to (12), we can define

$$\tilde{\mathbf{R}} = \mathbf{U}_n^{\mathrm{H}}\mathbf{R}\mathbf{U}_n \;=\; \mathbf{U}_n^{\mathrm{H}}\mathbf{R}_v\mathbf{U}_n \;+\; \sigma^2\mathbf{I}_{p-1}\,, \qquad \mathbf{U}_n \perp \mathbf{a}\,. \tag{13}$$

Although smaller, this matrix contains the same information as $\mathbf{P}_{\mathbf{a}}^{\perp}\mathbf{R}\mathbf{P}_{\mathbf{a}}^{\perp}$. Besides the dimension reduction, an advantage of this scheme is that the noise stays white.

These expressions can immediately be generalized to the more general case of $q < p$ interferers and unknown $\mathbf{a}$-vectors. Indeed, in this case, the projection onto the complement of the $\mathbf{A}_s$-matrix of the interference is given by

$$\mathbf{P}_{\mathbf{A}_s}^{\perp} = \mathbf{I} - \mathbf{A}_s(\mathbf{A}_s^{\mathrm{H}}\mathbf{A}_s)^{-1}\mathbf{A}_s^{\mathrm{H}} = \mathbf{U}_n\mathbf{U}_n^{\mathrm{H}}$$

and we can form $\tilde{\mathbf{R}} = \mathbf{U}_n^{\mathrm{H}}\mathbf{R}\mathbf{U}_n$ as above. Note that we do not have to know $\mathbf{A}_s$: the relevant noise subspace is estimated from the eigenvalue decomposition of $\mathbf{R}$. This hinges upon the fact that the noise covariance is white (in general: known), and the visibility matrix $\mathbf{R}_v$ is insignificant at these time scales (otherwise, it might disturb the eigenvalue decomposition).

6.2 Keeping track of projections

To enable the imaging step, it is essential to store the linear operation represented by the projections. At the same time, it might be necessary to adapt the projection several times per second, since the $\mathbf{a}$-vectors of interferers are time-varying. Hence, in the construction of the 10 s correlation average, we also have to construct the effective linear operation.

Thus consider the short-term averages, denoting for generality the linear operation by $\mathbf{L}_k$:

$$\tilde{\mathbf{R}}_k := \mathbf{L}_k\mathbf{R}_k\mathbf{L}_k^{\mathrm{H}} \;=\; \mathbf{L}_k\mathbf{R}_v\mathbf{L}_k^{\mathrm{H}} \;+\; \sigma^2\mathbf{L}_k\mathbf{L}_k^{\mathrm{H}}\,, \qquad k = 0, 1, \cdots, N-1\,.$$

By simply averaging these, the long-term average will be

$$\tilde{\mathbf{R}}^{10s} \;=\; \frac{1}{N}\sum_{k=0}^{N-1}\tilde{\mathbf{R}}_k \;=\; \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{L}_k\mathbf{R}_k\mathbf{L}_k^{\mathrm{H}}\,.$$

The $\mathbf{L}_k$ appear here at both sides of $\mathbf{R}_k$. To move them to one side, we make use of the general expression

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

where $\otimes$ denotes a Kronecker product, and $\mathrm{vec}(\cdot)$ the column-wise stacking of a matrix into a vector,

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

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots] \quad\Rightarrow\quad \mathrm{vec}(\mathbf{A}) := \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \end{bmatrix}$$

In this case, we obtain

$$\begin{aligned} \mathrm{vec}(\tilde{\mathbf{R}}^{10s}) &= \tfrac{1}{N}\textstyle\sum\left[(\bar{\mathbf{L}}_k \otimes \mathbf{L}_k)\mathrm{vec}(\mathbf{R}_k)\right] \\ &= \left[\tfrac{1}{N}\textstyle\sum\bar{\mathbf{L}}_k \otimes \mathbf{L}_k\right]\mathrm{vec}(\mathbf{R}_v) \;+\; \sigma^2\left[\tfrac{1}{N}\textstyle\sum\bar{\mathbf{L}}_k \otimes \mathbf{L}_k\right]\mathrm{vec}(\mathbf{I}_p) \\ &= \mathbf{C}\mathrm{vec}(\mathbf{R}_v) \;+\; \sigma^2\mathbf{C}\mathrm{vec}(\mathbf{I}_p) \end{aligned}$$

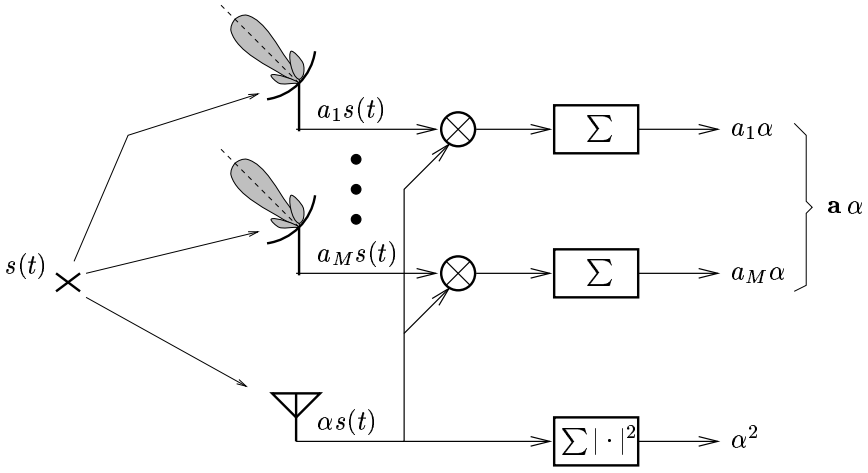

Figure 7: Estimation of $\mathbf{a}$ using a reference antenna

where

$$\mathbf{C} \;:=\; \frac{1}{N}\sum_{k=0}^{N-1} \bar{\mathbf{L}}_k \otimes \mathbf{L}_k$$

For the imaging step, we have to know how $\tilde{\mathbf{R}}^{10s}$ depends on $\mathbf{R}_v$. Thus, we have to construct and store $\mathbf{C}$ along with $\tilde{\mathbf{R}}^{10s}$. Unfortunately, it is a rather large matrix: $p^2 \times p^2$ if $\mathbf{L}_k$ doesn't contain dimension reductions. Another problem for imaging might be that the noise contribution on $\tilde{\mathbf{R}}^{10s}$ is no longer white, but determined by $\mathbf{C}$. Two possible remedies are

– Assume that the $\mathbf{a}$-vectors were sufficiently variable over the time interval. In that case, $\mathbf{C}$ is likely to be of full rank and thus invertible, and we can construct

$$\mathbf{C}^{-1}\mathrm{vec}(\tilde{\mathbf{R}}^{10s}) = \mathrm{vec}(\mathbf{R}_v) \;+\; \sigma^2\mathrm{vec}(\mathbf{I}_p)\,.$$

That is, we recover our interference-free model $\mathbf{R}_v + \sigma^2\mathbf{I}$. However, the inversion of $\mathbf{C}$ might be a formidable, and numerically dubious, task.

– If we take $\mathbf{L}_k = (\mathbf{U}_n)_k^{\mathrm{H}}$ as in (13), then the noise contribution on each $\tilde{\mathbf{R}}_k$ is white. We can average the $\tilde{\mathbf{R}}_k$ if they have the same dimension, i.e., $p-q$ where the number of interferers q is constant over the interval. In that case,

$$\sigma^2\frac{1}{N}\sum_{n=0}^{N-1}(\mathbf{U}_n)_k^{\mathrm{H}}(\mathbf{U}_n)_k \;=\; \sigma^2\mathbf{I}_{p-q}$$

so that the noise contribution on $\tilde{\mathbf{R}}^{10s}$ is white. Note that no inversion is necessary, but $\mathbf{C}$ now has size $(p-q)^2 \times p^2$, thus compressing the data. Simple but suboptimal extensions are possible in case q is not constant.

Since $\mathbf{C}$ is a factor p^2 larger than $\tilde{\mathbf{R}}^{10s}$, it might in fact be more efficient to store the short-term averages $\mathbf{R}_k$, along with the corresponding spatial filters $\mathbf{L}_k$. This is the case if $\mathbf{L}_k$ is to be updated at time scales of $10\ s/p^2 = 50$ ms or less.

6.3 Other spatial filtering possibilities

Without going into too much detail, we mention a few other possibilities for spatial filtering and interference cancellation. Suppose there is a single interferer,

$$\mathbf{R}_k = \mathbf{R}_v + \sigma_s^2\mathbf{a}\mathbf{a}^{\mathrm{H}} + \sigma^2\mathbf{I}\,.$$

– *Subtraction.* With an estimate of $\mathbf{a}$ and an estimate of its power, we can try to subtract it from the covariance data:

$$\tilde{\mathbf{R}}_k = \mathbf{R}_k - \hat{\sigma}_s^2 \hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathrm{H}} . \tag{14}$$

Without other knowledge, the best estimate of $\mathbf{a}$ is the dominant eigenvector, $\mathbf{u}_1$, of $\mathbf{R}_k$, and likewise the best estimate of σ_s^2 is $\lambda_1 - \sigma^2$. Since both of these are derived from $\mathbf{R}_k$, it turns out to be not too different from the projection scheme. Indeed, if we look at

$$(\mathbf{I} - \alpha \mathbf{u}_1 \mathbf{u}_1^{\mathrm{H}}) \mathbf{R}_k (\mathbf{I} - \alpha \mathbf{u}_1 \mathbf{u}_1^{\mathrm{H}}) \;=\; \mathbf{R}_k - \mathbf{u}_1 \mathbf{u}_1^{\mathrm{H}} \lambda_1 (2\alpha - \alpha^2)$$

we can make it equal to (14) by selecting α such that $\lambda_1(2\alpha - \alpha^2) = \hat{\sigma}_s^2$. The projection scheme had $\alpha = 1$.

– *Spatial whitening.* In this scheme, we try to make the interference plus noise white again. This component is equal to $\sigma_s^2 \mathbf{a}\mathbf{a}^{\mathrm{H}} + \sigma^2 \mathbf{I}$, so we pre- and postmultiply with square-root factors of it:

$$\tilde{\mathbf{R}}_k \;=\; (\hat{\sigma}_s^2 \hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathrm{H}} + \hat{\sigma}^2 \mathbf{I})^{-1/2} \mathbf{R}_k (\hat{\sigma}_s^2 \hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathrm{H}} + \hat{\sigma}^2 \mathbf{I})^{-1/2} \;=\; (\cdot)^{-1/2} \mathbf{R}_v (\cdot)^{-1/2} + \mathbf{I} .$$

– *Subtraction of a reference signal.* If we have a reference antenna that receives a 'clean' copy of the interfering signal, then we might try to subtract this reference signal from the telescope signals. There are many adaptive schemes for doing so, e.g., LMS. The general scheme is as illustrated in figure 7. In this figure, the $\mathbf{a}$-vector of the interferer is found by cross-correlating with the reference antenna. We also estimate its power. After correcting for the noise power on the reference antenna, we can reconstruct and subtract $\mathbf{a}s(t)$.

This scheme is rather similar to the original projection approach where we reduce the dimension to the noise subspace, viz. equation (13). The main difference is that, now, we reduce the dimension from $p+1$ antennas back to p antennas, so there is no loss of dimensions from the astronomy point of view. It appears that this only has advantages if the reference antenna has a better INR than the telescopes. Also, we need as many reference antennas as there are interferers.

7 IMAGING AFTER SPATIAL FILTERING

In the previous section, we discussed spatial filtering techniques. It was shown that an attractive scheme for removing the interference is by projecting it out. However, by doing so we replace the observed visibilities $V(u_i, v_i)$ in the matrix $\mathbf{R}_v$ by some (known) linear combination. In this section, we discuss the implications of this for the imaging.

7.1 Classical inverse Fourier imaging

The relation between sky brightness $I(\ell, m)$ and visibilities $V(u, v)$ (where u, v are taken at frequency f) is

$$V(u,v) = \iint I(\ell, m)\; e^{-j(u\ell + vm)}\; d\ell\; dm$$

We have measured V on a discrete set of baselines $\{(u_i, v_i)\}$. The dirty image (via direct Fourier inversion with weights w_i) is defined by

$$I_D(\ell, m) \;:=\; \sum_i \; w_i \; V(u_i, v_i)\; e^{j(u_i \ell + v_i m)}$$

It is equal to the 2D convolution of the true image with the dirty beam:

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

or

$$I_D \;=\; I * B_0\,, \qquad B_0(\ell,m) \;:=\; \sum_i\; w_i\; e^{j(u_i\ell+v_im)}$$

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

Specializing to a point source model, $I(\ell,m) = \sum_k\, I_k\, \delta(\ell-\ell_k, m-m_k)$, where I_k is the intensity of the source at location (ℓ_k, m_k), gives

$$V(u,v) \;=\; \sum_k\; I_k\; e^{-j(u\ell_k+vm_k)}$$

$$I_D(\ell,m) \;=\; \sum_k\; I_k\; B_0(\ell-\ell_k, m-m_k)$$

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

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

7.2 Inverse Fourier imaging after projections

If we take projections or any other linear combination $\mathbf{C} = [c_{ij}]$ of the visibilities $\{V(u_i,v_i)\}$ during measurements, as in section 6, we have instead available

$$\text{vec}(\tilde{\mathbf{R}}) = \mathbf{C}\text{vec}(\mathbf{R}_v) \qquad \Leftrightarrow \qquad Z(u_i,v_i) \;=\; \sum_j\; c_{ij}\; V(u_j,v_j)$$

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

$$\begin{array}{rcl} I_D(\ell,m) & := & \sum_i\; Z(u_i,v_i)\; e^{j(u_i\ell+v_im)} \\ & = & \sum_i \sum_j\; c_{ij}\; V(u_j,v_j)\; e^{j(u_i\ell+v_im)} \end{array}$$

Then

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

where

$$B(\ell,m,\ell',m') \;:=\; \sum_i \sum_j\; c_{ij}\; e^{-j(u_j\ell'+v_jm')}\; e^{j(u_i\ell+v_im)}$$

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

With a point source model

$$I_D(\ell,m) \;=\; \sum_k\; I_k\; B(\ell,m,\ell_k,m_k) \;=\; \sum_k\; I_k\; B_k(\ell,m)$$

where

$$B_k(\ell, m) \;:=\; \sum_i \sum_j \; c_{ij} \; e^{-j(u_j \ell_k + v_j m_k)} \; e^{j(u_i \ell + v_i m)}$$

Again, every point source excites a beam centered at its location (ℓ_k, m_k), but the beams may all be different: they are *space varying.* Nonetheless, they are completely known if we know the linear combinations that we took during observations.

In principle, the CLEAN algorithm is easily extended to work with space-varying beams, but at a large computational cost since the beam at each supposed source location has to be computed. However, CLEAN is strongly based on the assumption that a beam has a large peak at its center, since this is how we find candidate point source locations. In the new situation, the beam shape depends on $\mathbf{C}$, and it is not a priori clear what it will look like.

7.3 Maximum likelihood imaging

Let us consider the imaging step from a more fundamental viewpoint. In principle, the construction of the image using the observed correlation matrices and assuming the parametric model can be viewed as a parameter estimation problem. One of the most important inference methods is the maximum likelihood method. Given a parametric model for the received data, choose the parameters which maximizes the probability of obtaining the observed data. This is different than the most probable image approach [3] where no parametric model is imposed on the image, leading to maximum entropy image formation. Maximum likelihood estimators (MLEs) are known to be consistent and efficient (i.e., they provide unbiased estimators with minimum variance), under very general conditions, and thus are the natural choice for many parameter estimation problems.

In deriving the maximum likelihood estimator of the image parameters, we should take into account the Gaussianity of the astronomical signals. Contrary to the claim in [14], the corresponding MLE is *not* equivalent to parametric optimization of the CLEAN cost function. Using the discrete point source model we obtain:

$$\mathbf{R}_k \;=\; \mathbf{\Gamma}_k \mathbf{A}_k(\boldsymbol{\theta}) \, \mathbf{B} \mathbf{A}_k^{\mathrm{H}}(\boldsymbol{\theta}) \, \mathbf{\Gamma}_k^{\mathrm{H}} \;+\; \sigma^2 \mathbf{I} \tag{15}$$

where the d astronomical sources are Gaussian with covariance matrix $\mathbf{B} = \mathrm{diag}[I_1, \cdots, I_d]$ and sky coordinates $\boldsymbol{\theta} = \{\mathbf{s}_\ell\}_{\ell=1}^d$, and the noise is Gaussian with covariance $\sigma^2 \mathbf{I}$. Let $\hat{\mathbf{R}}_k$ be the sample covariance matrix during the k-th epoch, based on N_k collected samples. The likelihood of the observations at the k-th epoch given map parameters $\boldsymbol{\theta}, \mathbf{B}, \sigma^2, \mathbf{\Gamma}_k$ is then given by

$$\left(\frac{1}{\pi^p |\mathbf{R}_k|} e^{-\mathrm{tr}(\mathbf{R}_k^{-1} \hat{\mathbf{R}}_k)} \right)^{N_k}$$

Using all observation epochs we obtain that the log-likelihood function is given by (after omitting constants)

$$\mathcal{L}(\hat{\mathbf{R}}_1, \ldots, \hat{\mathbf{R}}_K | \mathbf{B}, \boldsymbol{\theta}, \{\mathbf{\Gamma}_k\}_1^K, \sigma^2) = -\sum_{k=1}^K N_k \log |\mathbf{R}_k| - \sum_{k=1}^K N_k \mathrm{tr}(\mathbf{R}_k^{-1} \hat{\mathbf{R}}_k) \tag{16}$$

where K is the number of epochs. The MLE is found by maximizing (16) over $\mathbf{B}, \boldsymbol{\theta}, \{\mathbf{\Gamma}_k\}_1^K, \sigma^2$. This is rather complicated, and efficient ways to implement it will be discussed elsewhere. A simplified model where perfect calibration is assumed has been analyzed in [13], where a generalized least squares solution has been proposed.

7.4 CLEAN and self-calibration

Since the direct optimization of (16) appears to be intractable, it is natural to resort to least-squares model fitting. In the case of the CLEAN algorithm we just try to fit in the least squares sense a parametric model

to equations (15), but assuming perfect calibration, i.e., $\mathbf{\Gamma}_k = \mathbf{I}$. The LS cost function is given by:

$$[\hat{\boldsymbol{\theta}} = \{\hat{\mathbf{s}}_\ell\}, \hat{\mathbf{B}}] = \min_{\boldsymbol{\theta},\mathbf{B}} \sum_{k=1}^{K} \parallel \hat{\mathbf{R}}_k - \mathbf{A}_k(\boldsymbol{\theta})\, \mathbf{B} \mathbf{A}_k^{\mathrm{H}}(\boldsymbol{\theta}) \, - \sigma^2 \mathbf{I} \parallel_F^2 \tag{17}$$

($\mathbf{B}$ is constrained to be diagonal with positive entries.) It is interesting to observe that if one ignores the Gaussianity of the astronomical signals, instead introducing nuisance parameters for each of the astronomical signal samples, one ends up with a likelihood function equivalent to the above CLEAN cost function. In a variety of simpler cases in array signal processing, this is known to have a degraded performance compared to (16).

With unknown array gains, equation (17) generalizes to

$$\left[\hat{\boldsymbol{\theta}} = \{\hat{\mathbf{s}}_\ell\}, \hat{\mathbf{B}}, \{\hat{\mathbf{\Gamma}}_k\}\right] = \min_{\boldsymbol{\theta},\mathbf{B},\{\mathbf{\Gamma}_k\}} \sum_{k=1}^{K} \parallel \hat{\mathbf{R}}_k - \mathbf{\Gamma}_k \mathbf{A}_k(\boldsymbol{\theta})\, \mathbf{B} \mathbf{A}_k^{\mathrm{H}}(\boldsymbol{\theta})\, \mathbf{\Gamma}_k^{\mathrm{H}} - \sigma^2 \mathbf{I} \parallel_F^2 \tag{18}$$

The solution can be obtained by the "self-cal" algorithm: an alternating least squares algorithm which solves iteratively for the parameters $\mathbf{B}$, $\{\mathbf{s}_\ell\}$ by a CLEAN step (with fixed gains), and the gain parameters $\{\mathbf{\Gamma}_k\}$ by a calibration step (with fixed source parameters $\mathbf{B}, \boldsymbol{\theta}$).

It has not been noted before in the literature that the latter step admits a direct algebraic solution. Indeed, to minimize (18) with fixed $\{\mathbf{A}_k\}$ and $\mathbf{B}$, we we can minimize separately for each k the related expression

$$\parallel \hat{\mathbf{R}}_k - \mathbf{\Gamma}_k \mathbf{A}_k \mathbf{B} \mathbf{A}_k^{\mathrm{H}} \mathbf{\Gamma}_k^{\mathrm{H}} - \sigma^2 \mathbf{I} \parallel_F^2$$

Let $\mathbf{g}^{(k)}$ be the vector of diagonal elements of $\mathbf{\Gamma}_k$. We can define for each k a $(p \times p)$ matrix $\mathbf{X}^{(k)}$ with entries

$$\mathbf{X}_{ij}^{(k)} = \frac{(\hat{\mathbf{R}}_k - \sigma^2 \mathbf{I})_{ij}}{(\mathbf{A}_k \mathbf{B} \mathbf{A}_k^{\mathrm{H}})_{ij}}$$

and fit $\mathbf{g}^{(k)}$ with entries $g_i^{(k)}$ such that

$$\mathbf{X}_{ij}^{(k)} = g_i^{(k)} \bar{g}_j^{(k)} \tag{19}$$

In the usual self-calibration algorithm, this equation is solved iteratively using phase closure relations. Instead, we can resort to the algebraic structure of the problem and obtain a closed form solution. In matrix form, (19) is

$$\mathbf{X}^{(k)} = \mathbf{g}^{(k)} \mathbf{g}^{(k)^{\mathrm{H}}} \tag{20}$$

This asks for the best hermitian rank-one approximation to the matrix $\mathbf{X}^{(k)}$, which is known to be given by

$$\hat{\mathbf{g}}^{(k)} = \sqrt{\lambda_1} \mathbf{v}_1$$

where λ_1 is the largest eigenvalue of $\mathbf{X}^{(k)}$ and $\mathbf{v}_1$ is its corresponding eigenvector. This also shows that for each sky model the calibration step has (w.p. 1) a unique solution that minimizes (18).

A satisfactory consequence is the convergence of the self-calibration iteration. This follows from the iterative least squares interpretation of the complete self-calibration procedure, together with the fact that the minimization of (18) along any coordinate always reduces its value.

7.5 Effect of spatial filtering

Now that we have a matrix formulation for the self-calibration algorithm it will be easy to see the effect of spatial filtering on the imaging process. The measurement equation now becomes

$$\mathbf{R}_k = \mathbf{A}_k \mathbf{B} \mathbf{A}_k^{\mathrm{H}} + (\mathbf{A}_s)_k (\mathbf{R}_s)_k (\mathbf{A}_s)_k^{\mathrm{H}} + \sigma^2 \mathbf{I} \tag{21}$$

The interference is removed by linear operations $\mathbf{L}_k$, for example projections, acting on the left and right of $\mathbf{R}_k$. This modifies the least squares optimization problem to

$$\left[\hat{\boldsymbol{\theta}} = \{\hat{\mathbf{s}}_l\}, \hat{\mathbf{B}}, \{\hat{\boldsymbol{\Gamma}}_k\}\right] = \min_{\boldsymbol{\theta}, \mathbf{B}, \{\boldsymbol{\Gamma}_k\}} \sum_{k=1}^{K} \| \mathbf{L}_k \left(\hat{\mathbf{R}}_k - \boldsymbol{\Gamma}_k \mathbf{A}_k(\boldsymbol{\theta}) \, \mathbf{B} \mathbf{A}_k^{\mathrm{H}}(\boldsymbol{\theta}) \, \boldsymbol{\Gamma}_k^{\mathrm{H}} - \sigma^2 \mathbf{I} \right) \mathbf{L}_k^{\mathrm{H}} \|_F^2 \, . \tag{22}$$

The cost function is similar to (18) and thus its minimization does not pose stronger computational demands. The main difference is that the initialization of the multidimensional search will become more complicated because the Fourier relation between the measured data and the sky brightness now leads to a deconvolution problem with space-varying beams (section 7.2).

8 Conclusions

We considered various aspects of multichannel interference suppression for radio-astronomy. It was shown that sub-band processing will be necessary in order to exploit narrow-band techniques. We have analyzed a simple case demonstrating the great improvements that are possible by multichannel detection and blanking of intermittent interference. We have also discussed spatial filtering techniques and their effect on the image formation process. The main conclusion is that spatial filtering is very interesting but requires to be taken into account in the imaging step.

During the writing of this paper, we have applied the multichannel detection and blanking algorithm to measured GSM signals, resulting in great improvements of the spectral estimates. This will be reported elsewhere.

Acknowledgements

Amir Leshem was supported by the NOEMI project of the STW under contract no. DEL77-4476. We would like to thank our project partners at NFRA, especially A. van Ardenne, A.J. Boonstra, P. Friedman, A. Kokkeler, J. Noordam, and G. Schoonderbeek, for the very useful collaboration.

[1] C. Barnbaum and R.F. Bradely. A new approach to interference excision in radio astronomy: Real time adaptive filtering. *The astronomical journal*, 115:2598–2614, 1998.

[2] G.E.P. Box. A general distribution theory for a class of likelihood criteria. *Biometrika*, 36:317–346, 1949.

[3] B.R. Frieden. Restoring with maximum likelihood and maximum entropy. *Journal of the optical society of America*, 62:511–518, 1972.

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

[5] B.L. Kasper, F.S. Chute, and D. Routledge. Excising terrestrial radio interference in low frequency radio astronomy. *Monthly notices of the Royal Astronomical Society*, 199:345–354, 1982.

[6] S.M. Kay. *Fundumentals of statistical signal processing: Detection theory*. PTR, Prentice Hall, 1998.

[7] A. Leshem and A.J. van der Veen. Detection and filtering in radio-astronomy. Technical Report NOEMI-98-01, Delft University of Technology, 1998.

[8] A. Leshem and A.J. van der Veen. The effect of blanking of TDMA signals on radio-astronomical correlation measurements. In *Proc. IEEE workshop on Higher Order Statistics 99*, pages 25–29, June 1999.

[9] A. Leshem, A.J. van der Veen, and E. Deprettere. Detection and blanking of GSM signals in radio-astronomical observations. In *Proc. IEEE SPAWC99*, May 1999.

[10] R.A. Perley, F. Schwab, and A.H. Bridle, editors. *Synthesis imaging in radio astronomy*. Astronomical society of the pacific, 1989.

[11] J.G. Proakis. *Digital communications.* McGraw-Hill, 3rd edition, 1995.

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

[13] J. Sheinvald, M. Wax, and A.J. Weiss. Localization of multiple sources using moving arrays. *IEEE trans. on Signal Processing*, 46:2736–2743, 1998.

[14] S.M. Tan. An analysis of the properties of CLEAN and smoothness stabilized CLEAN — some warnings. *Month. Not. R. Astr. Soc.*, 220:971–1001, 1986.

[15] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust., Speech, Signal Proc.*, 33(2):387–392, April 1985.

[16] R. Weber, C. Faye, F. Biraud, and J. Dansou. Spectral detector for interference time blanking using quantized correlator. *Astronomy and Astrophysics Supplement Series*, 126(1):161–167, November 1997.