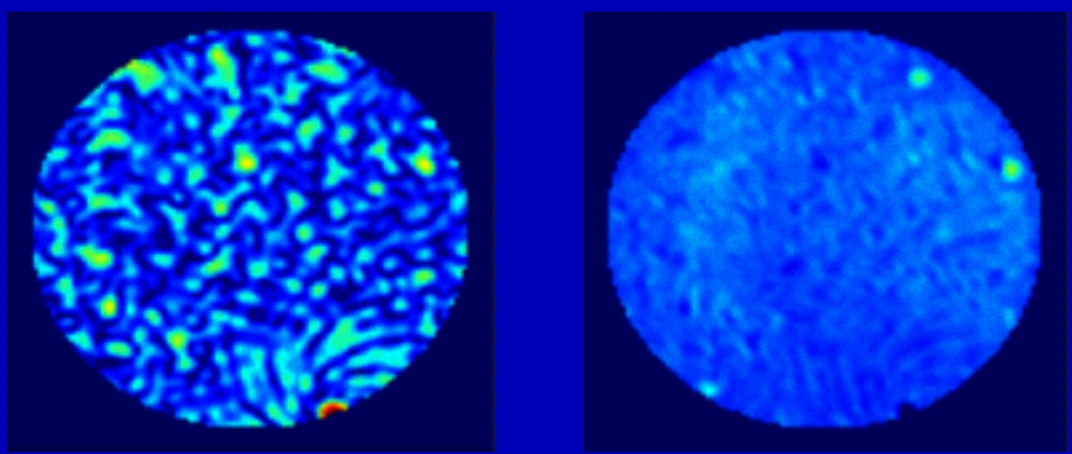
Radio Frequency Interference Mitigation in Radio Astronomy



A.J. Boonstra

Radio Frequency Interference Mitigation in Radio Astronomy

Albert-Jan Boonstra

Radio Frequency Interference Mitigation in Radio Astronomy

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 14 juni 2005 om 10.30 uur

door Albert-Jan BOONSTRA

ingenieur in de technische natuurkunde

geboren te Nijmegen

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr.ir. A.J. van der Veen

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof.dr.ir. A.J. van der Veen	Technische Universiteit Delft, promotor
Prof.dr.ir. P.M. Dewilde	Technische Universiteit Delft
Prof.dr.ir. M.H.G. Verhaegen	Technische Universiteit Delft
Prof.dr. A.G. de Bruyn	Rijksuniversiteit Groningen
Prof.dr. B.D. Jeffs	Brigham Young University Provo Utah, USA
Ir. A. van Ardenne	ASTRON, Dwingeloo
Dr. A. Leshem	Bar-Ilan University, Israël

The work described in this thesis was carried out in the context of the STW NOEMI project (DEL-77-4476, DTC.5893), and was also supported by ASTRON.

Published by ASTRON

ISBN: 90-805434-3-8

Aan Marieke, Elise, Franka, Noortje, Evelien, en Ilse,
de volgende generatie.
Als je goed om je heen kijkt zie je overal patronen.

Contents

Chapter 1

Introduction

1.1 Radio astronomy and the interference challenge

Radio astronomy is a passive service[1] and is equipped to observe extremely weak signals from outer space. Technological advances have had great influence on the sensitivity of radio telescope systems. Since the first detection of radio emissions from outer space in 1933 [76], and the development of the first radio telescope [120] and radio interferometer [128], the telescope (continuum) sensitivities have increased by five orders of magnitude [26, 30]. This corresponds to a factor of ten sensitivity increase per decade.

The sensitivity of current state-of-the-art telescopes is over ten orders of magnitude higher than in most communications systems [71]. This high sensitivity is required because radio astronomical signals are very weak, typically 40 to over 100 dB weaker than signals from most other services. Radio astronomy reaches this high sensitivity as astronomical observations usually have a duration of hours to several days as compared to only microseconds to seconds for communications systems. Also, the receiving areas of the antennas in radio astronomy are usually several orders of magnitude larger than in radio communications systems. In addition, radio telescope receivers are often equipped with cryogenically cooled receivers. This leads to very low receiver noise powers, often as low as ten times the cosmic background noise levels.

For the coming two decades, the aim is to build radio telescope systems which are one to two orders of magnitude more sensitive than the current systems. Examples are the Low Frequency Array (LOFAR) [28, 160], currently under construction in the Netherlands, and the Square Kilometer Array (SKA) [2, 161], currently in a concept study phase.

[1] Passive service: a service [75, 38] not involved in any man-made radio transmission but only concerned with the reception of naturally occurring radio waves.

Recent technological advances in the fields of electronics and communications systems have led to a vast increase in communication applications and systems. These include mobile cellular telephone systems (e.g. GSM), digital radio (e.g. DAB) and digital television (e.g. DVB), short-range devices (e.g. Bluetooth, WiFi), ultra-wideband radar and communications (UWB), and satellite systems (e.g. GPS, GLONASS). The demand for radio spectrum has increased dramatically, leading to scarcity in many frequency bands and in some cases to congestion.

From a technical point of view [139], one might argue that for active radio communication services[2], the current spectrum information transmission capacity is not exploited to its limit [131]. Indeed, monitoring information shows that not the whole allocated spectrum is occupied with transmissions all the time [36, 7, 129, 50, 12], as many transmission systems operate only intermittently. New technologies and radio communication approaches, such as described in [139], may reduce the spectrum scarcity for most of the radio communication services. One of those techniques is applying spread-spectrum[3] digitally-coded modulation schemes [176]. Another technique is "software radio" [122], which enables communication systems to transmit and receive across a broad range of frequencies as the signal processing is done in software. Because of the economic potential of these systems, it is likely that they will be developed further, and will be accompanied by (inter)national agreements. However, from the point of view of the passive services these trends do not in general lead to an improved spectrum use; often the opposite is true.

Because of the denser active use of the spectrum, and because of higher telescope sensitivities, radio astronomy is increasingly hampered by interference from other spectrum users. Figures 1.1 and 1.2 show how man-made radio signals influence radio astronomy observations. Figure 1.1 shows auto- and cross-correlation spectra (left) of an observation of OH emission lines contaminated with interference from the GLONASS satellite positioning system. The correlations are between telescopes of the WSRT array. The figure also shows an example of an astronomical pulsating star signal, a pulsar (right-hand figure). In this particular case, a stationary interferer would reduce the signal to noise ratio, but it would not influence the observed pulse shape. This illustrates that interference influences observations from different observational modes in different ways. Figure 1.2 shows astronomical images just outside the $25.55 - 25.67$ MHz band allocated to radio astronomy. The images were obtained with the LOFAR Initial Test Station (ITS), which is a LOFAR demonstration telescope located in the Netherlands. The figure shows all-sky ("fish-eye") observations, showing the astronomical sources Cas.A, Vir.A, and the North Polar Spur. The right-hand figure shows the same observation at an adjacent frequency channel,

[2] Active communication service: a service [75] which is based on both actively transmitting and receiving signals

[3] Spread spectrum: a communication technique in which the transmitted signal bandwidth is considerably larger than the frequency content of the original signal.

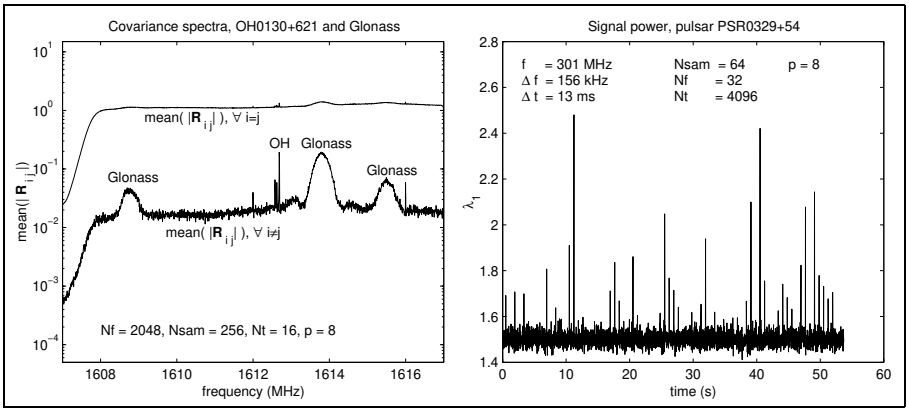


Figure 1.1. Examples of astronomical observations in the spectral domain (left) showing astronomical OH emission lines contaminated with GLONASS transmissions, and in the temporal domain (right) of a pulsar (spinning/pulsating neutron star) with rotation period 0.715 s. Both datasets were obtained with the WSRT and the NOEMI data recording system (cf. chapter 4).

occupied by a terrestrial transmitter. The spatial sidelobes of the transmitter (point spread function) obscure the astronomical sources. In this case, spatial filtering can reduce the transmitter signals to levels below the Cas.A flux level, as is described in [17] [170]. For useful astronomical science, an interferer with this power should be suppressed many more orders of magnitude. A comparison of the transmitter flux levels versus the LOFAR telescope sensitivity is given in chapter 10, together with some initial interference mitigation results.

Interference affects radio astronomy in several ways. Before going into more detail in possible interference mitigation approaches in section 1.2, a few general aspects and challenges in relation to the changes in spectrum use are briefly discussed.

First it should be noted that protection criteria exist for radio astronomy; see for instance [69, 72]. These norms cover both single dish telescope protection and aperture array protection criteria. Although relatively narrow bands are allocated to the radio astronomy service (RAST), such as the 21 cm band for neutral hydrogen, radio astronomy increasingly observes in bands in which there is no radio astronomy allocation. The main reason for this is that the cosmic radio signals are not limited to specific bands, but occur over the entire spectrum. The neutral hydrogen emissions of far-away galaxies, for example, are Doppler shifted to lower frequencies. These emissions are even observed at frequencies far below 1 GHz. A second reason is that for continuum observations (as opposed to narrow-band spectral line observations), the sensitivity can be increased by using large bandwidths, thus enabling radio astronomy to observe

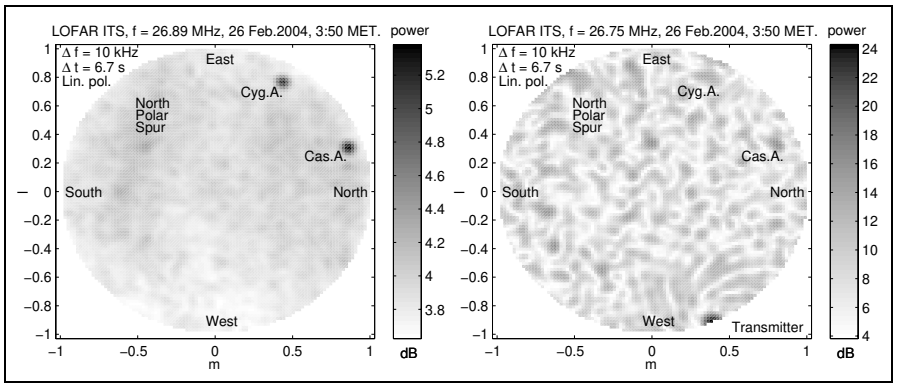


Figure 1.2. Example of radio astronomical images: observation of the northern sky with a test station (ITS) of the Low Frequency Array (LOFAR) in The Netherlands, without interference (left) and with a transmitter at the horizon (right). The ITS station is a phased array with 60 antenna elements, configured in a 200 m diameter five-armed spiral (cf. chapter 10).

very weak and distant astronomical sources. Since a considerable fraction of the bands below 2 GHz are only occupied with intermittent transmissions, radio astronomy is able to observe in some of these bands outside the ones allocated to radio astronomy, even in densely populated regions such as the Netherlands.

The increasing demand for spectrum in general leads to fewer time-frequency slots unoccupied with transmitters and interference (out of band and spurious emissions [75]). Although new modulation schemes in principle have higher information transmission capacity than the traditional ones (e.g. AM, FM), spectrum monitoring observations indicate that for several bands the spectrum occupancy increases, rather than decreases. In the Netherlands this can be seen, for example, in monitoring observations at the Westerbork Synthesis Radio Telescope (WSRT), in the LOFAR monitoring data [12], and in observations with the monitoring network (VMN) of the Dutch spectrum management agency Agentschap Telecom (AT) [3]. This means that spectrum sharing between active and passive users becomes increasingly difficult.

A consequence of the increasing spectrum demand is that in several bands the aggregate power of transmitters increases in time. This may lead to linearity problems in the analog part of telescope receivers: increasing intermodulation product levels (a mixture of harmonics appearing in the band of interest), and an increase in the noise level. These problems can often, but not always, be reduced. However, increasing the linearity of an already well-designed receiver system almost always leads to an increase in system noise.

Finally, digitally modulated wide-band and ultra wide-band (UWB) systems [127] are often designed to have many transmitters simultaneously active in the same band. Such an aggregate of transmitters does not have a specific spatial signature, and therefore it will tend to behave like spatially white background

noise. Due to this lack of specific spatial signatures, it is very difficult to suppress these signals in astronomical observations, and it will lead to an increase in the system noise. Single UWB devices may have a spectral power density which is undetectable because it is below the background radio noise [70]. However, an aggregate of many devices may hamper radio astronomy. Also, UWB pulses may hamper non-imaging radio astronomy such as pulsar research [86,157], and cosmic ray airshower research [67]. A related (EMC) issue, also potentially dangerous for radio astronomy, is an aggregate of high speed (> 1 MHz) data transmissions over conducting wires, such as power line communications (PLC) [39]. There are very practical alternatives to these systems which avoid EMC problems, but this issue is not only a technical one.

There clearly is a challenge for radio astronomy to mitigate the increasing levels of interference. Due to the technical advances, part of the spectrum occupied with interference and transmitters can be recovered with interference mitigation approaches. There are, however, limits in effectiveness and cost. Still, considering the trends mentioned, it seems reasonable to assume that the new generation of telescopes such as LOFAR , ALMA, and SKA, can deliver one to two orders of magnitude more sensitivity than the current systems.

1.2 Current RFI mitigation approaches

In order to mitigate interference, it must have characteristics which are in some ways different from the cosmic signals to be detected. If the interference cannot be distinguished, then obviously it can not be mitigated. Fortunately, there cannot be an interferer which is in all possible domains simultaneously indistinguishable from cosmic sources. The effectiveness of mitigation is limited by the estimation and detection accuracies of the signals involved. Different astronomical observing modes may require different interference mitigation techniques and approaches. Examples of these modes are spectral line observations, polarisation measurements, synthesis imaging, and pulsar research.

There are many ways to define categories for interference, such as narrowband or wideband, fixed or moving sources, categories based on statistical properties (e.g. spatial and temporal coherence) or based on modulation type [58], distinctions based on the amount of a-priori information of the transmitter or on differences in spatial properties or polarisation, categories based on field strength, power [44], and temporal-spectral occupancy, and categories of overlapping signal parameter domains [46].

The way interference is perceived by a radio telescope system obviously also depends on the telescope system itself and on its configuration. Interference mitigation methods are therefore categorised, sometimes in terms of telescope subsystems [9], and sometimes in terms of signal characteristics. Examples of telescope-based mitigation categories are analog versus digital methods, pre- versus post-correlation methods [8], single telescope versus array methods, adaptive versus non-adaptive, real-time versus off-line, aperture array versus focal

plane array. Interference mitigation methods can of course also be applied in multiple domains simultaneously. Depending on the interference and the type of instrument, several kinds of RFI mitigation techniques are applicable. Overviews of applicable methods, techniques and approaches can be found for example in [98, 53, 151, 88, 154, 40].

It is difficult to define a complete set of independent and non-overlapping interference mitigation categories into which all of the current methods fit in a straightforward way. In this section, the following, more or less ad hoc, categories are chosen: linearity in the analog (receiver) domain, fundamental signal parameters (time, frequency, location, direction, polarisation), and combinations of the fundamental parameters.

Linearity in the analog domain

The interference problem in the analog (receiver) domain is maintaining linearity. Strong interfering signals may saturate the receiver and give a decrease in sensitivity and a nonlinear response, so that sums of two signals will produce intermodulation products. Issues here are the avoidance of intermodulation products and achieving low receiver noise characteristics, even in the proximity of spatially and spectrally strong transmitters. This can be achieved by careful receiver system design in which the receiver mixing scheme is matched to the spectrum environment and the required bandwidths [155, 135, 107]. In addition, the (active) antenna [34], and receiver components such as the low noise amplifiers (LNA), need careful design [172, 173]. Here, an optimum must be found between linearity and noise characteristics, usually two contradictory requirements. Although intermodulation products in the band of interest can to some extent be filtered spatially in the digital domain [17], it is much better to suppress them at the source. Alternative receiver design studies were carried out, using for example the feed-forward concept [171]. This study confirmed that increased linearity to accommodate high power input signals increases the system noise. Receiver and LNA design must meet the technical requirements, but as the new generation of telescopes will consist of one to two orders of magnitude more receivers than current systems, cost becomes a critical issue [165] as well.

Suppressing nearby strong transmitter signals is traditionally done with conventional filters [121]. A drawback of those filters for radio astronomy is that in-band loss close to the stopband is high. In practice this means that spectrum bands spectrally and spatially near very strong transmitters such as TV stations and FM stations cannot be used by radio astronomy. A solution is using high temperature superconductor (HTSC) filters [158, 178], which have a very low passband loss. A drawback is their cost and the need for cryogenics. In the mid- or long term, miniature cryocooler systems [31] may be developed which could potentially be produced at low cost.

After passing the analog receiver parts, the received radio signals are digitised using analog to digital converters (ADC's). Only two-level (one bit), three-level, or four-level (two-bit) digitisations are traditionally used in radio

astronomical correlators [146, 24, 61]. This works well for Gaussian signals, but if there are non-Gaussian signals present as well, for instance from radio transmitters, then the digitisation process is distorted [83]. The LOFAR telescope stations in the north-eastern part of the Netherlands, for example, would require at most 12- to 14- bit ADC's [27, 13]. The required number of bits depends on the system architecture, and on the choice of the telescope site as the spectrum occupancy is site dependent.

Time and frequency domain approaches

Radio frequency transmitters and interferers which are intermittent, such as Time Division Multiple Access (TDMA) wireless communication (e.g. GSM), and airplane radar (DME) can be detected and removed from contaminated radio astronomical data without completely losing astronomical information. Traditionally, excision, also known as flagging or blanking, is applied in radio astronomy to post correlation data. The integration is usually of the order of seconds or minutes. In many cases, the time scales of transmission slot lengths during which the transmitter or interferer is active is much less than the one-minute or one-second level, and online (real time) detection and filtering algorithms are essential to reduce the effect of interference to an acceptable level. The most widely implemented algorithm is a single-channel total power change detector [82, 164, 51, 115, 49], followed by a blanking of the correlator output. The threshold level is usually determined either experimentally, or is based on the χ^2 distribution.

Detectors based on a-priori information of the transmitter have been proposed, for example detectors based on cyclostationarity properties [163, 118]. Neural network theory is also applied for detection purposes [137, 90]. A detector based on a probability density function (pdf) analysis was proposed [52], as well as a detector based on wavelet decompositions [106]. All these detectors are single-channel (i.e. non-array) detectors.

In case the interferer bandwidth is much smaller than the astronomical feature of interest, the interferer can be removed by a spectral notch filter. If the interferer is outside the band of interest, the challenge is to design (digital) spectral filters with low spectral sidelobes [1]. If the transmitter or interferer is within the band of interest it can often be mitigated [43], provided that there is a-priori knowledge of the modulation and coding scheme.

A disadvantage of detectors based on a-priori information is that they are usually more complicated and require more processing resources. An advantage is that these detectors are potentially more sensitive, which may yield more effective mitigation.

Approaches based on direction

The single-channel detectors described so far do not exploit the spatial properties of the interference. A detector that considered combining multiple telescopes for improved detection and blanking was proposed for low-frequency interferometry [80]; in this study a robust data-censoring method based on the

temporal behaviour of the cross spectrum was proposed, which requires a large number of estimated spectra to obtain robust estimates, and only two channels are used.

If an interferer is continuously present in time, spatial filtering is an approach which can be used in telescope arrays. Extensive literature exists on beamforming and spatial filtering, for example in the context of communications systems and radar array signal processing [149, 88, 168]. In radio astronomy, research on and application of these techniques started only around 1998.

Some spatial filtering research efforts are directed at subtraction approaches[4] using two to four reference antennas, employing for example LMS-type adaptive cancellation techniques [6] [29]. Although effective in interferer reduction capabilities, these methods do not fully exploit the subspace structure[5] of antenna arrays. In astronomical SKA-related phased-array studies, however, RF beamforming and subspace-based interference nulling were investigated and applied to radio astronomical demonstrators [64, 136, 43, 45]. In the digital domain, on the other hand, subspace-based interference mitigation studies were very limited [100, 147].

Wideband beamforming and nulling can be achieved in the analog domain by using time delays or by a combination of time delays and phase rotators [136,47]. Wideband beamforming and spatial filtering can be implemented digitally in many ways, for example as explained in [40].

Post correlation spatial filtering can be applied both to the correlation data and to the image plane data [94]. In [151] the relation between array beamforming , spatial filtering, and clean [130] is described.

Approaches based on location and direction

A first measure to prevent interference is to separate telescopes from transmitters and (other) interfering sources. In practice this means placing telescopes at remote places, or restricting transmitters to certain areas (coordination zones). Additional methods are shielding the telescope with screens [74], which requires propagation effects to be taken into account [73], or shielding radiating equipment, which is an EMC problem.

Arrival time differences between telescopes in an array can be used to distinguish between cosmic signals and interference. Examples are secondary radio emissions from high-energy cosmic particles entering the earth atmosphere [67], or the search for extraterrestials [144].

Approaches based on polarisation

So far, not much research has been carried out on polarisation-based interference mitigation. It is difficult to define useful mitigation methods because

[4]Methods involving the subtraction of signals from reference antennas are here considered as spatial filtering, as both methods are mathematically nearly identical.

[5]Subspace: the array output covariance matrix contains structure which can be used for calibration and spatial filtering purposes. This structure can be estimated using eigenvalue or factor analysis approaches [151].

the observed interference polarisation properties may vary rapidly in time. The telescope sidelobe structure influences the observed polarisation properties of the interference; moreover, the sidelobes change in time because the telescope rotates mechanically, or, for phased arrays, electronically. Also, because of limited transmitter/interferer polarisation information and because of diffraction and multipath effects, it is difficult to separate the interferer properties from the sidelobe properties. Moreover, it is already difficult to do polarisation work at all in radio telescopes because of system instabilities and finite calibration accuracies, and the fact that most astronomical sources are either unpolarised or weakly polarised.

1.3 Scope of this thesis

The focus of this thesis is on interference mitigation techniques for synthesis imaging, mainly exploiting the spatial signature of the interferers. In radio astronomy, image formation is based on the conversion (mostly Fourier transformation) of observed covariance matrices into sky maps [113,146]. The covariance matrices contain more structure than is traditionally used in radio astronomy. Often optimisations are not carried out on the complete covariance matrix but iteratively on an interferometer basis [130, 66]. These methods yield good results, but new challenges such as higher sensitivity for the new generation of telescopes LOFAR [160] and SKA [161] utilising interference mitigation make it worthwhile to investigate the underlying data models more closely.

A full-array covariance matrix and dual polarisation data model for radio astronomy was presented in 1996 [63]. In this model, instrumental effects (e.g. antenna dipole orientation) and non-instrumental effects (e.g. Faraday rotation [113, 124], that is, rotation of the polarisation plane) are taken into account in a straightforward and elegant way. Also, analysis of the data model is mainly based on interferometer correlations (two by two or four by four matrices).

This thesis proposes new data models and uses them for interference mitigation purposes. The data models described are partly complementary, and partly identical to the formalism defined in [63]. Modern array signal processing techniques drawn from other fields are introduced here for use in interference mitigation. These techniques are, for example, detection and estimation theory from communications and statistical signal processing [167,81,82,104,149], and subspace techniques such as eigen analysis [68], and factor analysis [91,105] from econometrics [60] and psychometrics. The theoretical performance of the interference mitigation methods studied is investigated and verified experimentally using the Westerbork Synthesis Radio Telescope (WSRT).

Until recently, multichannel interference mitigation techniques were not studied in radio astronomy. This thesis therefore focuses on multichannel detection and the excision of interference, and on spatial filtering with and without reference antennas. An advantage of these methods is that they are blind, in the sense that no a-priori information of the interferer is required for them to be

effective. This means that they are relatively simple and that they will work for a large class of interfering signals. Obviously their performance may be inferior to methods in which a-priori information is used. The main advantage of arrays is that they have more degrees of freedom than single dishes, and thus in principle have more interference mitigation capabilities.

The performance of the interference mitigation algorithms is dependent on the model and on system parameters such as integration time, but also on the estimation accuracies of the model parameters such as telescope gain. This thesis therefore also focuses on the single and dual polarisation estimation of the complex telescope gains. The influence of bandwidth and the extendedness of sources will be considered as well. Finally, implications of this thesis work for future telescopes will be discussed.

1.4 Thesis layout and summary of the main results

Data model and subspace analysis

In chapter 2 discrete source models will be derived using the approach outlined in [93]. In addition, a dual polarisation model, a multipath model, an intermodulation model and a multiple phased-array covariance model will be derived. These models describe many spatio-temporal properties relevant for (narrow-band) radio astronomy signal processing in a compact way. Chapter 3 will describe the basic signal processing tools. The influence of channel bandwidth and source extendedness on the subspace structure will be estimated. The experimental set-up for the different experiments will be described in chapter 4.

Detection and excision

It will be shown in chapter 5 (see also [98]) that by subband processing, many narrow-band techniques available in array signal processing and detection theory can be successfully applied to radio-astronomical observations contaminated with intermittent transmitter signals and interference. The benefits of multichannel spatio-spectral detection and excision, both theoretical and experimental, will be demonstrated. Finally, the theoretical limits for the maximum attenuation numbers will be derived.

Spatial filtering

In chapter 6 [119, 11] spatial projection and subtraction filters will be analysed and applied to experimental data. It will be shown that the effectiveness of the projection filter is limited by estimation accuracies. The advantage of projection filters over subtraction filters is that only the spatial signature of the interferer is needed, not an estimate of the interferer power. An advantage of a subtraction filter is that it is a relatively simple filter, but it requires more knowledge of the system noise power and the interference power. A property of mitigation filters in general is that there will always be residual distortions

of the astronomical signal. For the projection filter a distortion correction matrix exists [93]. This correction matrix reduces the distortions and requires the short-term stationarity and mid-term non-stationarity of the transmitter involved. This requirement was verified experimentally at the WSRT for different types of transmitters. In chapter 7 spatial filtering with reference antennas will be analysed and verified experimentally. It will be shown [152] that by using a-priori knowledge of the system an improved filter can be constructed.

Gain estimation

For a single polarisation array which observes one dominant source, be it either an interferer or of astronomical origin, it will be shown in chapter 8 (see also [21]) that weighted least-square gain estimators are asymptotically efficient. Closed-form estimators, which are asymptotically efficient under certain conditions, will be derived as well. The estimation accuracies are compared to the Cramer-Rao Lower bound, a technique previously unknown in radio astronomy, and it will be shown that the gain phase and gain magnitude estimation are independent.

Dual polarisation gain estimation

In chapter 9 and in [20, 19] a full dual polarised array model will be proposed, and factor analysis approaches are applied to estimate the model components. The performance is studied by simulation. A closed-form solution is found for the least squares minimisation of the model error and it turned out that at least three sky sources with different polarisation states are needed to find the telescope complex-gain factors. This three-source requirement follows straightforwardly from the model solutions, a result which is difficult to achieve otherwise.

Implications for future generation telescopes

An interference mitigation strategy for the LOFAR telescope will be given in chapter 10 [10, 17]. The interference power levels observed in spectrum monitoring data will be linked to LOFAR sensitivity levels. It will be shown that interference mitigation techniques and spatial dilution effects will reduce the interference to levels below the integrated noise levels under certain conditions and for moderate transmitter interference powers. The spatial dilution due to snapshot averaging as described in chapter 10 and in [10] is, in a sense, the two-dimensional equivalent of interference attenuation due to fringe rotation in synthesis arrays. As both SKA and LOFAR are many-element aperture synthesis arrays, the results derived for LOFAR will most likely also be applicable to SKA.

Outreach, dissemination

The KIvI Telecommunication Section Best Thesis Award 2002 was granted to this thesis work. A patent "Calibration method, device, and computer programme", based on this thesis work, was filed [14]. The patent is based on

single- and dual-polarisation array calibration methods. The single polarisation method is based on a fast, closed form, and accurate array-gain estimation procedure, a weighted logarithmic least square method, described in chapter 8. The polarised-array gain estimation is based on a least squares minimisation, involving a dual-polarisation gain model and a source model, described in chapter 9. The minimisation can be carried out, for example, by a rank two factor analysis approach. Finally, the work done for this thesis resulted in the following publications:

International journal papers

- A.J. Boonstra and S. van der Tol. Spatial filtering of interfering signals at the initial LOFAR phased array test station. *Radio Science, special section*, accepted for publication, 2005.
- A.J. van der Veen, A. Leshem, and A.J. Boonstra. Array signal processing for radio astronomy. *Experimental Astronomy, special issue*, accepted for publication, 2005.
- S.J. Wijnholds, J.D. Bregman, and A.J. Boonstra. Sky noise limited snapshot imaging in the presence of RFI with the LOFAR Initial Test Station. *Experimental Astronomy, special issue*, accepted for publication, 2005.
- A.J. Boonstra and A.J. van der Veen. Gain calibration methods for radio telescope arrays. *IEEE Transactions on Signal Processing*, 51(1):25–38, January 2003.
- J. Raza, A.J. Boonstra, and A.J. van der Veen. Spatial filtering of RF interference in radio astronomy. *IEEE Signal Processing Letters*, 9(2):64–67, February 2002.
- A. Leshem, A.J. van der Veen, and A.J. Boonstra. Multichannel interference mitigation techniques in radio astronomy. *The Astrophysical Journal Supplement Series*, 131(1):355–373, November 2000.

Refereed international conference papers

- A.J. Boonstra et al. Calibration, sensitivity and RFI mitigation requirements for LOFAR. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, USA, 2005.
- A.J. van der Veen, A. Leshem, and A.J. Boonstra. Signal processing for radio astronomical arrays. *IEEE Sensor Array and Multichannel Signal Procesing workshop (SAM)*, Barcelona, Spain, July 2004.
- A.J. van der Veen and A.J. Boonstra. Spatial filtering of RF interference in radio astronomy using a reference antenna. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.
- A.J. Boonstra and A.J. van der Veen. Dual-polarization gain estimation for radio telescope arrays. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, April 2003.

- A.J. Boonstra and A.J. van der Veen. Gain estimation for polarized radio telescope arrays. *Proc. International Union of Radio Science (URSI)*, 27^{th} General Assembly, August 2002.
- S. van der Tol, and A.J. van der Veen, and A.J. Boonstra. "Mitigation of continuous interference in radio astronomy using spatial filtering", In URSI General Assembly, Maastricht (NL), August 2002.
- A.J. Boonstra, and A.J. van der Veen, and J. Raza. Spatial filtering of continuous interference in radio astronomy, In IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pp. 2933-2936, Orlando, Florida, USA, May 2002.
- A.J. Boonstra and A.J. van der Veen. Gain decomposition methods in sensor array systems. *11th IEEE Workshop on Statistical. Signal Processing.* Singapore, August, 2001.
- A.J. Boonstra, A. Leshem, A.J. Van der Veen, A. Kokkeler, and G. Schoonderbeek, The effect of blanking of TDMA interference on radio-astronomical observations: experimental results, In IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pp. 3546-3549, Istanbul, Turkey, June 2000.
- A. Leshem, A-J. van der Veen, A. Kokkeler, A-J. Boonstra and G. Schoonderbeek, "Blanking of TDMA interference and its effect on radio-astronomical correlation measurements: Experimental results", In Proceedings of IEEE-PRORISC workshop, November 1999.

Other publications and contributions

- A.J. van der Veen, A.J. Boonstra, and S. van der Tol. Interference reduction in radio astronomy, ICT congress, The Hague, The Netherlands. September 2002.
- A.J. Boonstra. RFI mitigation strategies, SKA Workshop, Groningen, The Netherlands, August 2002.
- A.J. van der Veen, A.J. Boonstra, A. Leshem, J. Raza, and R. Calders, Exploiting the spatial signature of communications signals received at the WSRT, In IUCAF RFI Mitgation Workshop, MPIfR, Bonn, Germany, March 2001.
- A.J. Boonstra. Interference mitigation strategies for radio astronomy: RFI research areas for SKA, In SKA Workshop Technology Pathways to the Square Kilometre Array. Jodrell Bank, UK, August 2000.
- A.J. Boonstra. LOFAR RFI mitigation strategy, Technical Report LOFAR ASTRON Doc. 5, Dwingeloo, The Netherlands, October 2002 (http://www.lofar.org).

Patent

- A.J. Boonstra and A.J. van der Veen, "Calibration method, device, and computer program", ASTRON patent WO2004017090, 26 February 2002.

Award

- KIvI Telecommunication Section Best Thesis Award 2002, A.J. Boonstra, Eindhoven, April, 2002.

1.5 Notation and mathematical functions

The notation which is used throughout this thesis will be described in this section. Unless stated otherwise or when evident from the context, the numbers, vectors, and matrices are complex.

General notation

$\mathbf{a}$	Boldface letters denote vectors
$\mathbf{a}_i$	The i^{th} element of vector $\mathbf{a}$, scalar
a, A	Lowercase and uppercase letters which are not boldfaced denote scalars
$\mathbf{A}$	Boldface capital letters denote matrices, the $m \times n$ matrix $\mathbf{A}$ consists of elements $\mathbf{A}_{ij}$, or $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1n} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2n} \\ \vdots & & \ddots & \vdots \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \cdots & \mathbf{A}_{mn} \end{bmatrix}$
$\widetilde{(\cdot)}$	Polarisation matrix or vector

Matrix definitions

$\mathbf{I}$	Identity matrix, the dimension follows from the context in which it is used.
$\mathbf{I}_p$	The $p \times p$ Identity matrix
$\mathbf{I}_{st}, \mathbf{Q}_{st}, \mathbf{U}_{st}, \mathbf{V}_{st}$	Pauli (spin) matrices, $\mathbf{I}_{st} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{Q}_{st} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ $\mathbf{U}_{st} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\mathbf{V}_{st} = \begin{bmatrix} 0 & -\jmath \\ \jmath & 0 \end{bmatrix}$
$\mathbf{e}_i$	The i-th unit vector
$\mathbf{1}$	Vector containing ones

Linear algebra standard functions

$(\cdot)^H$	Conjugate transpose (Hermitian) operator
$\underline{(\cdot)}^t$	Transpose operator
$\overline{(\cdot)}$	Complex conjugate operator
$(\cdot)^{-1}$	Matrix inverse, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$
$(\cdot)^\dagger$	Pseudo inverse (Moore-Penrose inverse) of a full rank matrix, for a $M \times N$ matrix $\mathbf{A}$ it is defined by $\mathbf{A}^\dagger = \left[\mathbf{A}^H\mathbf{A}\right]^{-1}\mathbf{A}^H$, for $M \geq N$, and by $\mathbf{A}^\dagger = \mathbf{A}^H\left[\mathbf{A}\mathbf{A}^H\right]^{-1}$, for $M \leq N$
tr	Trace operator, $\mathrm{tr}(\mathbf{A}) = \sum_i [\mathbf{A}]_{ii}$, let λ_i be the i^{th} eigenvalue of $\mathbf{A}$, then $\mathrm{tr}(\mathbf{A}) = \sum_i \lambda_i$
$\|\cdot\|_F$	Frobenius matrix norm. Let $\|\mathbf{A}\|_F = \sqrt{\sum_{ij} \|\mathbf{A}_{ij}\|^2}$, let λ_i be the i^{th} eigenvalue of $\mathbf{A}$, then $\|\mathbf{A}\|_F^2 = \sum_i \lambda_i^2$
$\|\mathbf{A}\|$	Determinant. Let λ_i be the i^{th} eigenvalue of $\mathbf{A}$, then $\|\mathbf{A}\| = \prod_i \lambda_i$
$\|\mathbf{a}\|$	Vector norm, $\|\mathbf{a}\| = \sqrt{\mathbf{a}^H\mathbf{a}}$

Stacking and selection operators

$[\mathbf{a}_1, \cdots, \mathbf{a}_q]$	Stacking of length p vectors $\mathbf{a}_i$ in a $p \times q$ matrix
$\mathbf{A}_{ij}$	The ij^{th} element of matrix $\mathbf{A}$, $\mathbf{A}_{ij}$ can be either a scalar or a submatrix
$[\mathbf{A}_{ij}]$	$[\mathbf{A}_{ij}]$ denotes a matrix $\mathbf{A}$ which is composed of elements (either submatrices or single components) $\mathbf{A}_{ij}$: $[\mathbf{A}_{ij}] \equiv \mathbf{A}$
vec	Stacking of the columns of a matrix $\mathbf{A}$ into a vector $\mathbf{b}$ is denoted by $\mathbf{b} = \text{vec}(\mathbf{A})$
vec$'$	Vectorisation of a matrix, omitting the diagonal entries, $\text{vec}'(\mathbf{A}) = \mathbf{J}_{sel}\text{vec}(\mathbf{A})$
unvec	Reverse operation of vec
diag	The diag operator converts a vector into a diagonal matrix, with the vector placed on the main diagonal: $\mathbf{A} = \text{diag}(\mathbf{a})$. When operated on a matrix, it will set its off-diagonal elements to zero: $\text{diag}(\mathbf{A}) = \mathbf{A} \odot \mathbf{I}$, where $\odot$ denotes element-wise matrix multiplication
vecdiag	Returns the main diagonal of a matrix stacked into a vector: $\text{vecdiag}(\mathbf{A}) = (\mathbf{A} \odot \mathbf{I})\mathbf{1}$, where $\odot$ denotes element-wise matrix multiplication
$\mathbf{I}_c$	Complement of $\mathbf{I}$: $\mathbf{I}_c = \mathbf{1}\mathbf{1}^t - \mathbf{I}$
$\mathbf{I}_s$	A $p \times (p-1)$ selection matrix, defined as the identity matrix with its first column removed.
$\mathbf{J}_{sel}$	Selection matrix with dimensions $((p^2-p) \times p)$, defined as the $p \times p$ identity matrix with the $[1, (p+1)+1, 2(p+1)+1, \cdots, p^2]$ columns removed

Random variable functions and notation

$\widehat{(\cdot)}$	Estimated value
$\langle \cdot \rangle$	Time average of a random vector or a random matrix
$\mathcal{E}\{\cdot\}$	Expected value of a random vector a random matrix
$\sim$	Connects a (multivariate) random variable to a certain probability distribution
$\text{var}(\widehat{\mathbf{A}})$	Element-wise variance of the random variable or sample matrix $\widehat{\mathbf{A}}$: $\text{var}(\widehat{\mathbf{A}}) \equiv \mathcal{E}\{(\widehat{\mathbf{A}} - \mathcal{E}\{\widehat{\mathbf{A}}\}) \odot \overline{(\widehat{\mathbf{A}} - \mathcal{E}\{\widehat{\mathbf{A}}\})}\}$
$\text{std}(\widehat{\mathbf{A}})$	Element-wise standard deviation of the random variable or sample matrix $\widehat{\mathbf{A}}$ with elements a_{ij}: $\text{std}(\hat{a}_{ij}) \equiv \sqrt{\mathcal{E}\{\ (\hat{a}_{ij} - \mathcal{E}\{\hat{a}_{ij}\})\ \overline{(\hat{a}_{ij} - \mathcal{E}\{\hat{a}_{ij}\})}\ \}}$
$\text{cov}(\widehat{\mathbf{A}})$	Covariance of the random variable or sample matrix $\widehat{\mathbf{A}}$, with $\mathcal{E}\{\widehat{\mathbf{A}}\} = \mathbf{A}$: $\text{cov}\widehat{\mathbf{A}} \equiv \mathcal{E}\{\ [\text{vec}(\widehat{\mathbf{A}}) - \text{vec}(\mathbf{A})]\ [\text{vec}(\widehat{\mathbf{A}}) - \text{vec}(\mathbf{A})]^H\ \}$

Matrix products and division operators

$\otimes$	Denotes the Kronecker matrix product, defined as $\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B} & \mathbf{A}_{12}\mathbf{B} & \cdots \\ \mathbf{A}_{21}\mathbf{B} & \mathbf{A}_{22}\mathbf{B} & \\ \vdots & & \ddots \end{bmatrix}$
$\odot$	Element-wise matrix multiplication (Hadamard product), $[\mathbf{A} \odot \mathbf{B}]_{ij} = \mathrm{A_{ij}B_{ij}}$
$\ominus$	Element-wise matrix division, $[\mathbf{A} \ominus \mathbf{B}]_{ij} = \mathbf{A}_{ij}\mathbf{B}_{ij}^{-1}$
$\circ$	Khatri-Rao product, which is a column-wise Kronecker product, $\mathbf{A} \circ \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \mathbf{a}_2 \otimes \mathbf{b}_2 \cdots]$
$\mathbf{a}^H\mathbf{b}$	Vector inner product, $\mathbf{a}^H\mathbf{b} = \sum_{i=1}^{N} a_i^* b_i$
$\mathbf{a} \times \mathbf{b}$, $\mathbf{a}\mathbf{b}^H$	Vector outer product, $\mathbf{a} \times \mathbf{b} \equiv \mathbf{a}\mathbf{b}^H \equiv \begin{bmatrix} a_1\bar{b}_1 & \cdots & a_1\bar{b}_p \\ \vdots & \ddots & \vdots \\ a_p\bar{b}_1 & \cdots & a_p\bar{b}_p \end{bmatrix}$

Standard mathematical functions

$e^{\mathbf{a}}$	$e^{\mathbf{a}} \equiv [e^{\mathbf{a}_1}, \cdots, e^{\mathbf{a}_p}]^t$, with p the number of elements in $\mathbf{a}$
ln	Natural, e-based, logarithm
$\ln(\mathbf{A})$	Elementwise logarithm, $\ln(\mathbf{A}) \equiv [\ln(\mathbf{A}_{ij})]$. This is a non-standard definition, used in this thesis. The standard definition of a function f of $\mathbf{A}$ is: $f(\mathbf{A}) = \mathbf{U}\,\mathrm{diag}(f(\lambda_1), \cdots, f(\lambda_p))\,\mathbf{U}^{-1}$, where λ_i is the i^th eigenvalue of $\mathbf{A}$; $\mathbf{U}$ is a unitary matrix.
$^{10}\log$	The 10-based logarithm
$^{2}\log$	The 2-based logarithm
$\equiv$	By definition equal to
sinc	Sinc function, $\mathrm{sinc}(x) \equiv \frac{\sin(x)}{x}$
$\mathcal{O}$	Order function
Re	Real part of complex entities
Im	Imaginary part of complex entities
$\jmath$	Complex number, $\jmath \equiv \sqrt{-1}$
δ_{ij}	Kronecker Dirac delta function
Eigenvalue decomposition	Any correlation or covariance matrix $\mathbf{R}$ of the random process $\mathbf{x}$, with $\mathbf{R} = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}$ is Hermitian and positive definite. The matrix $\mathbf{R}$ can be decomposed in terms of eigenvalues λ_i and eigenvectors $\mathbf{u}_i$. Let $\mathbf{\Lambda} \equiv \mathrm{diag}(\lambda_1 \cdots \lambda_p)$, and $\mathbf{U} \equiv [\mathbf{u}_1, \cdots, \mathbf{u}_p]$, then $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$. Eigenvectors corresponding to distinct eigenvalues are orthonormal.
Singular value decomposition (SVD)	Any matrix $\mathbf{A}$ can be decomposed in terms of unitary matrices $\mathbf{U}, \mathbf{V}$, and a diagonal matrix $\mathbf{\Sigma}$ with positive real entries: $\mathbf{A} \equiv \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. The entries σ_i of $\mathbf{\Sigma}$ are the singular values of $\mathbf{A}$ and are usually sorted such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$. The columns of $\mathbf{U}$ and the columns of $\mathbf{V}$ are called left and right singular vectors. The left singular vectors and right singular vectors form two orthonormal sets.

Chapter 2

Data model

In this chapter[1], discrete-source data models for synthesis imaging signal processing based on the work of Leshem and van der Veen [93] and Hamaker et al [63] will be presented . Their array vector and array covariance matrix models will be extended with a full-array polarisation model, and with multipath and intermodulation models. The purpose of the models is to simplify the analysis of the effectiveness of RFI mitigation techniques, as will be discussed in the following chapters. Existing and new models will be presented in a uniform manner, so that similarities between the models will become apparent. Finally, the conditions under which the models are valid will be described. The main focus of the interference mitigation research in this thesis is on interference mitigation in the image Fourier transform plane, not in the image domain itself.

2.1 Introduction

The aperture synthesis principle

In conventional optical or single-dish radio telescope systems, an image of the sky is made by concentrating the incident electromagnetic waves on a focal plane, where the images are captured by the placement of sensitive elements. In optics, for example, charge coupled devices (CCDs) are placed at the focus, while in radio astronomy a (multi-antenna) receiver is used and an image is obtained by scanning the telescope over a region in the sky. In radio aperture synthesis, however [113, 146, 124], sky images are not obtained by measuring on a focal plane, but by measuring on a more or less arbitrary and usually flat aperture plane. At this aperture plane, a spatial electromagnetic interference pattern is present, caused by the emitting cosmic sources. This pattern can be estimated by measuring the spatial time correlations of the electromagnetic field using radio interferometers [124, 87]. Radio interferometers are formed by combining or correlating signals from pairs of telescopes. Figure 2.1 shows a schematic

[1]Parts of this chapter were published in [98, 19, 20]

picture of the radio astronomical observation process. It shows observing telescopes, the output signals of which are filtered and down-converted. The resulting baseband signal voltages are subsequently digitised and correlated. The pair-wise correlation output depends on the relative telescope positions, which are also known as baseline vectors or baselines. A sky image can be formed by inverse Fourier transforming the correlated signals, or using techniques related to Fourier transforms.

In radio astronomy, the spatial time correlations are known as coherencies or visibilities. Figure 2.2 shows an example of observed coherency data, obtained with the LOFAR test station ITS. The figure shows observed coherency amplitudes (upper) and coherency phases (lower). The data are presented in a (u, v) coordinate system, which consists of baseline direction cosines (cf. section 2.2.2). The spatial antenna configuration and corresponding baseline configuration of ITS is shown in figure 10.7. Sky images are reconstructed by inverse Fourier transforming the observed coherencies , or by techniques closely related to the inverse Fourier transform [124]. The sky image, corresponding to the coherency data mentioned above, is shown in figure 1.2. The astronomical sources Cas.A (supernova remnant) and Cyg.A (radio galaxy) are clearly visible near the horizon. In the aperture plane in figure 2.2, these sources are visible as phase gradients showing approximately 20 times 2π phase rotations. In the figure, these are visible as parallel line structures with about 20 lines each, which are aligned from the upper left to the lower right.

In radio astronomy, the relation between cosmic source brightness distribution and the observed spatial coherencies is also known as the *interferometer equation* or the *visibility function* [113, 146, 124]. In summary, aperture synthesis is an imaging technique which solves the cosmic source brightness distribution, using measured values of the spatial correlations for several interferometer telescope distances and orientations. The main advantage of aperture synthesis interferometers over single dishes is that very large aperture areas can be synthesised using relatively small telescopes. Imaging techniques based on the electromagnetic field coherence function are also used in other fields such as optical and holographic interferometry [123, 23, 79].

New challenges in radio aperture synthesis

Recent advances in electronics and related areas have led to a large increase in wireless communications applications. This has resulted in a denser spectrum occupation and new (wideband) transmitter modulation schemes, and, as a consequence, radio astronomical telescope systems are faced with an increasing radio interference problem: it is becoming increasingly difficult to keep the bands allocated to radio astronomy free from interference. On the other hand, advances in electronics and computer technology also allow more sophisticated data processing, which means that at least part of the interference effects could be reduced using interference mitigation techniques. It is a challenge to develop effective RFI mitigation techniques with low signal distortions in the astronomical observations which can be implemented at a reasonable cost. Another

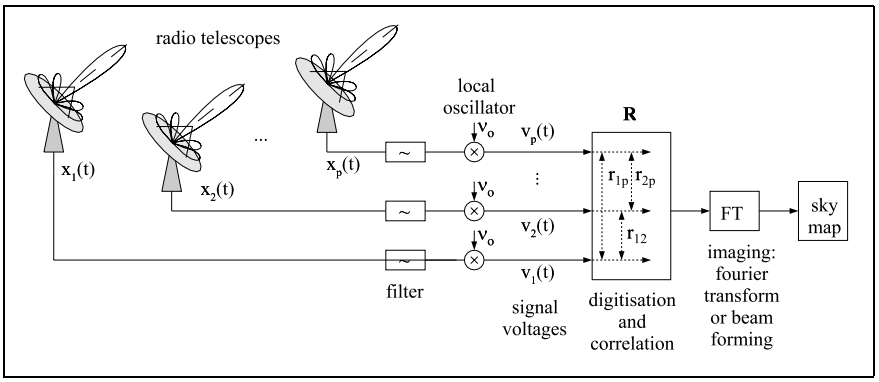


Figure 2.1. A radio telescope array of parabolic dishes for measuring the spatial coherency function. The number of telescopes is p; the telescope output signal for telescope i in the time domain is indicated by $x_i(t)$. After filtering and down conversion [124], the telescope signals are correlated pair-wise and stacked in a matrix $\mathbf{R}$. A sky image can be formed by inverse Fourier transforming the correlated signals, or using techniques related to Fourier transforms.

challenge in future aperture synthesis systems is the scale and complexity. Several large-scale aperture synthesis telescopes haven been proposed or planned (for example LOFAR, ATA, SKA), and these telescopes will have aperture areas which are one to two orders of magnitude larger than current systems. Some of these systems combine phased arrays with aperture synthesis (for example LOFAR). Challenges for these large scale systems include calibration, imaging, data processing, and RFI mitigation techniques.

Signal formalism

The mathematical formalism in the traditional radio astronomical calibration and imaging process is based on interferometer relations, that is, on pairwise telescope output correlations. In principle, all observed telescope correlations can be stacked in a (hermitian) square matrix, also known as the correlation or the covariance matrix. Data models based on such a "full" covariance matrix, in which instrumental effects are modelled as matrix multiplications, have an internal structure which can be utilised using linear algebra techniques. This internal structure is used, for example, to estimate the telescope-based calibration errors. Because the unknowns are telescope-based and not interferometer based, the data model can be simplified and telescope gain solutions were readily found (self-calibration procedure) [146]. Although the calibration methods applied so far were successful, there is structure in the full covariance matrix model which is not (yet) utilised. The challenges mentioned in the previous section may therefore be handled better if this structure in the aperture synthesis full covariance model is taken into account.

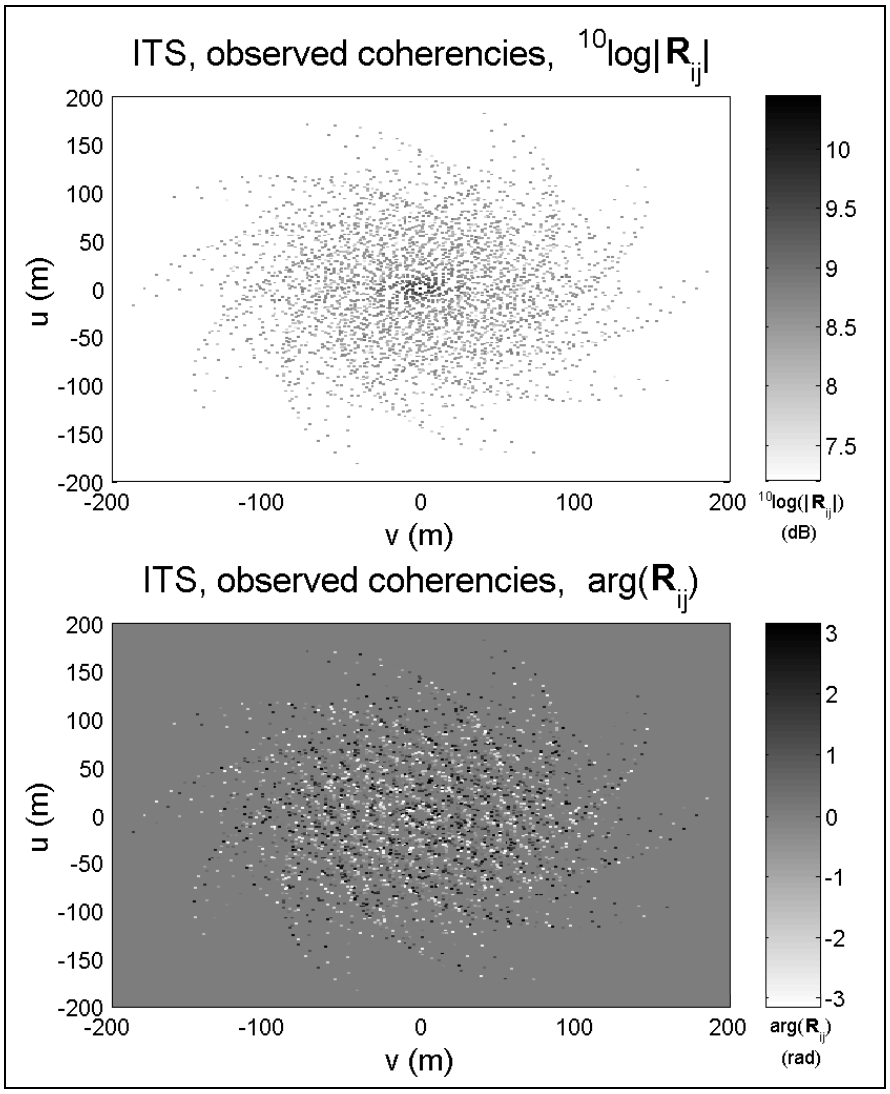


Figure 2.2. Observed telescope antenna covariance data $\mathbf{R}_{ij}$ or spatial coherencies , integrated over 6.7 s. The data was obtained with the LOFAR ITS test station, and shown are the covariance amplitude (upper) and phase (lower).

An example of a model which utilises more structure is the radio astronomical polarimetric data model which was recently proposed [63]. The aim of this matrix model is to form a coherent framework combining the polarisation concepts of Stokes parameters and Wolf coherency matrices with the Jones and

Mueller matrix calculus from optics and radio interferometry based on multiplying correlators. In this framework all telescope-based and interferometer-based effects (such as telescope gain, ionospheric rotation of the polarisation, interferometer errors) are straightforwardly modelled as multiplicative matrix operations. The advantage of this model is that it supports a better understanding of radio interferometry, especially the polarisation properties. It also potentially enhances the calibration accuracies, as system and instrumental errors can be modelled more accurately.

In the polarimetric data processing model, the analysis is usually applied to interferometer equations of dimension 2×2 or 4×4, rather than to the analysis of the full covariance matrix. Also, the polarimetric model and the current scalar models do not take interference source modelling into consideration. For these reasons, a new single-polarisation data model [93, 98], and a new dual-polarisation data model [19, 20] were proposed. The new single polarisation data model is a (hermitian) matrix-based discrete source model, and it allows interference source modelling. In several forms this data model is in use in many other areas such as statistical signal processing, communications, acoustics, econometrics, and psychometrics. A wide variety of methods and algorithms is available for detection and parameter estimation, many of which can be used in radio astronomy signal processing. A polarised version of the data model was also proposed; it is a straightforward extension of the polarimetric model in [63] in the sense that it considers the full covariance matrix rather than submatrices of dimension 2×2 or 4×4.

Chapter outline

The purpose of this chapter is to present a new data model which allows interference modelling and which, just as the polarimetric model [63], incorporates straightforward modelling of system and instrumental effects. Following the introduction, section 2.2 will briefly describe the classical aperture synthesis approach in radio astronomy, and the assumptions used in the modelling. The next section, section 2.3, will describe the proposed new single-polarisation, matrix-based, discrete sources model for the observed covariance matrix. The polarised version of the new model will be described in section 2.4, which will be followed by section 2.5 on alternative interference models. Although it is only partially used in this thesis, the covariance matrix structure of a synthesis array based on phased-array telescope stations will be described in section 2.6. The chapter will end with concluding remarks in section 2.7.

2.2 Aperture synthesis

In this section, a brief description will be given of the relation between the observed spatial coherencies and the cosmic source distribution, and also of the aperture synthesis problem: how to estimate the cosmic source distribution from the observed spatial coherencies. Following [113], relatively simple expressions

for the spatial coherence function and visibility function can be derived under certain simplifying assumptions.

2.2.1 Spatial coherence and interferometry

This section will briefly describe the relation between the observed coherencies and the cosmic source distribution. It follows the introduction of [113]. The purpose is to connect the discrete source matrix-based formalism to the conventional continuous interferometer relations. Many details of the interferometer relation derivations are omitted as these are not relevant to the description of the matrix based formalism. More details can be found in [113, 146, 87, 124].

The starting point is the definition of an arbitrary orthonormal coordinate system in which the location of the cosmic sources is given by $\mathbf{r}_s$ and where the locations within the aperture array are given by the vector $\mathbf{r}_i$. For the cosmic source signals it is assumed that the aperture far-field condition holds, which implies that no information can be obtained about the cosmic source distribution in the direction of the line of sight. As a consequence, the cosmic source distribution may be described in terms of a source distribution located at the celestial sphere at a fixed distance $\mathbf{r}_s$ without any loss of generality. It is also assumed that the space within the celestial sphere is vacuum. Let t denote time, let $\mathbf{s}$ be the source direction vector, and define the time varying electric field at a frequency ν at the cosmic source location $\mathbf{r}_s$ as $E(\mathbf{r}_s, t)$ and at an earth-bound aperture location $\mathbf{r}_i$ by $E(\mathbf{r}_i, t)$. Define the cosmic source direction vector $\mathbf{s}$ by $\mathbf{s} = \frac{\mathbf{r}_s}{|\mathbf{r}_s|}$, let dS be an infinitesimal area on the celestial sphere, $dS = |\mathbf{r}_s|^2 d\Omega$, where $d\Omega$ is the solid angle corresponding to dS, and define the source intensity or brightness by $I_B(\mathbf{s})$. Further let c denote the speed of light. A schematic drawing of the electric fields at the mentioned locations is shown in figure 2.3.

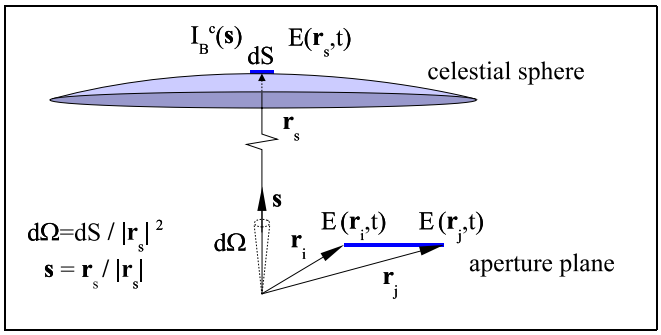


Figure 2.3. Electromagnetic field distribution at the aperture plane and at the celestial sphere source locations

As it is assumed that the source signals are stationary over the period of ob-

servation, the stochastical properties of the signals can be estimated by (spatial) correlation of the electric fields at the aperture plane. The electric field correlation between the aperture locations $\mathbf{r}_i$ and $\mathbf{r}_j$, or spatial coherence function $V(\mathbf{r}_i, \mathbf{r}_j)$, is defined by:

$$V(\mathbf{r}_i, \mathbf{r}_j) = \mathcal{E}\{E(\mathbf{r}_i, t)\overline{E(\mathbf{r}_j, t)}\} \tag{2.1}$$

where $\mathcal{E}\{\}$ denotes the expected value or ensemble average.

Under the above assumptions, the linearity properties of the Maxwell equations imply that the electromagnetic field at a point $\mathbf{r}_i$ can be described as the sum or integral of the contributions of all emitting sources in the sky, taking into account geometrical path effects. This is closely related to the Huygens-Fresnel principle in optics [79], which states that the propagation of waves through an unobscured medium or free space can be described in terms of oscillating virtual point sources[2] located at the wavefronts, or in this case the celestial sphere. These concepts are related to the Van Cittert-Zernike theorem [124] [174], which describes the relation between the scalar (sky) source intensity distribution and the electric field coherencies at the (earth-bound) aperture plane. The coherency function can be described in terms of the sky source distribution $I_B^c(\mathbf{s})$, which represents the intensities of the astronomical sources in the direction $\mathbf{s}$ [113, 146, 124]:

$$V(\mathbf{r}_i, \mathbf{r}_j) = \int_{\substack{celestial \\ sphere}} I_B^c(\mathbf{s}) e^{-2\pi\jmath\nu c^{-1}\mathbf{s}^t(\mathbf{r}_i - \mathbf{r}_j)} d\Omega \tag{2.2}$$

The subscript B in I_B^c refers to the brightness distribution of the celestial sources [113]. The superscript c refers to the fact that a continuous sky source distribution is considered, as opposed to a discrete source distribution, as will be described in section 2.3.1. In radio astronomy, the spatial coherency function V is measured by means of radio interferometers, which are pairs of telescopes the outputs of which are cross-correlated. Figure 2.1 shows an interferometer array with p telescopes. The individual telescope outputs x_i are voltages which correspond to measured electromagnetic field strength amplitudes. The telescopes have a certain effective receiving area, dependent on the direction of the incident waves, which means that a multiplicative telescope (frequency dependent) antenna gain $\mathcal{A}_i(\mathbf{s})$, also known as amplitude response or primary beam, should be introduced to the spatial coherence function to obtain a formula for measured coherencies [113]:

$$V(\mathbf{r}_i, \mathbf{r}_j) = \int_{\substack{celestial \\ sphere}} \mathcal{A}_i(\mathbf{s})\overline{\mathcal{A}_j(\mathbf{s})} I_B^c(\mathbf{s}) e^{-2\pi\jmath\nu c^{-1}\mathbf{s}^t(\mathbf{r}_i - \mathbf{r}_j)} d\Omega \tag{2.3}$$

[2] A point source is defined as a source of electromagnetic radiation which has infinitesimal spatial dimensions

The quantity $\mathbf{r}_i - \mathbf{r}_j$ is the distance between the two locations at which the electric fields are measured and is called the (telescope) baseline $\mathbf{b}_{ij}$: $\mathbf{b}_{ij} = \mathbf{r}_i - \mathbf{r}_j$.

The coherencies $V(\mathbf{r}_i, \mathbf{r}_j)$ are given in terms of $V^2m^{-2}Hz^{-1}$ (or V^2m^{-2}) whereas practical correlators have as inputs the measured voltages at a certain signal cable impedance, usually 50Ω. The relation between the coherencies $V(\mathbf{r}_i, \mathbf{r}_j)$ and the observed correlations $R(\mathbf{r}_i, \mathbf{r}_j)$ is a telescope dependent scaling factor G_i, which is constant over the sky:

$$R(\mathbf{r}_i, \mathbf{r}_j) = G_i \overline{G_j} V(\mathbf{r}_i, \mathbf{r}_j) \tag{2.4}$$

The scaling factor G_i is due to unknown variations in electronic instrument response, it is slowly time-varying and must be estimated in a calibration step. The model now becomes:

$$R(\mathbf{r}_i, \mathbf{r}_j) = G_i \overline{G_j} \int_{\substack{celestial \\ sphere}} \mathcal{A}_i(\mathbf{s}) \overline{\mathcal{A}_j(\mathbf{s})} I_B^c(\mathbf{s}) e^{-2\pi j \nu c^{-1} \mathbf{s}^t \mathbf{b}_{ij}} d\Omega \tag{2.5}$$

Equation (2.5) is the interferometer equation, which relates the sky source distribution to the observed correlations, using a multiplying correlator. Solving this equation for $\mathbf{I}_B^c(\mathbf{s})$ for measured values of $R(\mathbf{r}_i, \mathbf{r}_j)$ (for several baselines $\mathbf{d}_{ij}$) is aperture synthesis. This sky image reconstruction process involves inverse Fourier transforms or related techniques. Usually, the phases of the correlation function $R(\mathbf{r}_i, \mathbf{r}_j)$ are expressed relative to the phase centre of the field of view.

2.2.2 Aperture synthesis and imaging

The interferometer equation (2.5) can be expressed in right-hand orthonormal coordinate systems, both for the aperture plane (u, v, w), and for the source plane (l, m, n). Following [146], the $\mathbf{w}$ and $\mathbf{n}$ axes are defined parallel to the centre of the field of view (direction of view), as is shown in figure 2.4. These coordinate systems are often used in radio astronomy. Assume that an earth rotation aperture synthesis telescope tracks a certain point in the sky (centre of the field of view) at a direction $\mathbf{s}_o$. Other parts of the observable field of view can be associated with the vector $\mathbf{s}$. Both vectors are unit length vectors. Define $\boldsymbol{\sigma}$ by $\mathbf{s} = \mathbf{s}_o + \boldsymbol{\sigma}$. In general, in (l, m, n) coordinates:

$$\mathbf{s}_o = (0, 0, 1)^t, \qquad \mathbf{s} = (l_s, m_s, n_s)^t, \qquad \boldsymbol{\sigma} = (l_s, m_s, n_s - 1)^t \tag{2.6}$$

where $n_s = \sqrt{1 - (l_s^2 + m_s^2)}$. The telescope distance vector or baseline $\mathbf{b}_{ij}$ is defined as $\mathbf{b}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and can be expressed in (u, v, w) coordinates as:

$$\mathbf{b}_{ij} = \lambda(u_{ij}, v_{ij}, w_{ij})^t \tag{2.7}$$

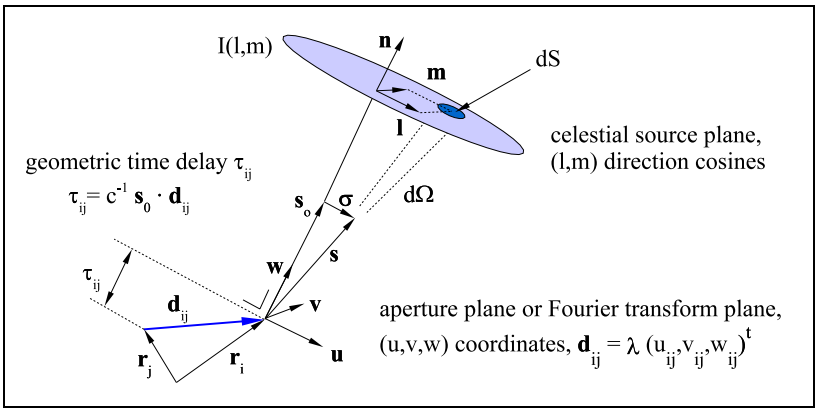


Figure 2.4. Coordinate systems in the celestial sphere with $\{\mathbf{l}, \mathbf{m}, \mathbf{n}\}$ unit vectors, and at the aperture plane with $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ unit vectors. The telescopes i and j are located at $\mathbf{r}_i$

Using the new bases, the visibility function (2.3) can be expressed as [146, 113]:

$$V(u_{ij}, v_{ij}, w_{ij}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{A}_i(l,m)\overline{\mathcal{A}_j(l,m)} I_B^c(l,m) \cdot e^{-2\pi\jmath(u_{ij}l + v_{ij}m + w_{ij}\sqrt{1-(l^2+m^2)})} \frac{dldm}{\sqrt{1-(l^2+m^2)}} \tag{2.8}$$

So far, no assumption has been made limiting this expression to small fields of view, and solving this equation in yields a (2π Sr) map of the entire visible celestial sphere. This relation between visibility and sky brightness distribution is not a Fourier transform relationship, but by assuming a small field of view ($l^2 + m^2 \ll 1$), and dropping the i, j indices, it can be written as

$$V(u,v) = e^{-2\pi\jmath w} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{A}_i(l,m)\overline{\mathcal{A}_j(l,m)} I_B^c(l,m) e^{-2\pi\jmath(ul+vm)} dldm \tag{2.9}$$

The factor $e^{-2\pi\jmath w}$ is the geometric delay factor associated with the reference location $\mathbf{s}_o$; the reference direction $\mathbf{s}_o$ is also known as the phase tracking centre. As the geometry of a telescope is known, the geometric delay can be compensated for, leading to the following Fourier relation:

$$V(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{A}_i(l,m)\overline{\mathcal{A}_j(l,m)} I_B^c(l,m) e^{-2\pi\jmath(ul+vm)} dldm \tag{2.10}$$

The beamshape $\mathcal{A}_i(l,m)$ need not be identical for all telescopes. Source position-independent complex gain factors, such as varying atmospheric conditions, and to a certain extent the telescope aperture efficiencies, can be accounted for by adding a complex gain factor to the beamshape function. As the

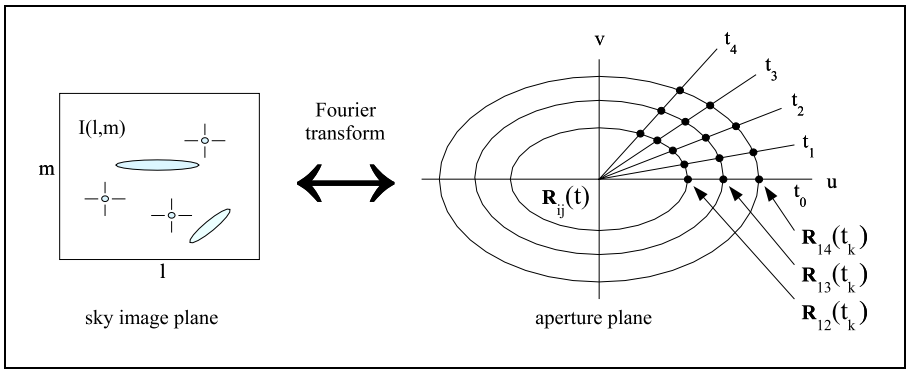


Figure 2.5. Fourier transform relation between distribution of astronomical sources in the sky (left) and electromagnetic interference pattern caused by celestial sources at the aperture plane (right). The aperture plane interference pattern is sampled at the uv plane by the correlations $\mathbf{R}_{ij}(t)$. Note that here the term interference is not related to RFI .

visibility function is a Fourier transform of the sky source brightness distribution, the brightness distribution can be recovered by inverse Fourier transforming the observed visibilities . Aperture synthesis arrays usually increase their number of aperture plane sampling points by making use of the fact that the interferometer coordinates change with respect to the source due to the rotation of the earth. This principle is called earth rotation synthesis. Figure 2.5 shows a sky source intensity image (left figure) in (l, m) coordinates, and a corresponding aperture plane (right figure). For an East-West linear array, this plane is at times t_k sampled in straight lines. Earth rotation has the effect that the telescope baseline orientation with respect to the sky changes with time, thus gradually filling the (synthesised) aperture with coherency sampling points. For an East-West array, the (u, v) coordinates can be expressed in terms of h and δ coordinates:

$$u_{ij} = \frac{|\mathbf{b}_{ij}|}{\lambda}\cos(h), \qquad v_{ij} = \frac{|\mathbf{b}_{ij}|}{\lambda}\sin(h)\cos(\delta) \tag{2.11}$$

The rotation angle around the earth axis is hour angle, or h, and is -90^o in the East, 0^o at the meridian, and 90^o in the West; δ is the declination, which is the angle with respect to the ecliptic plane, positive in the northern direction.

Aperture synthesis: classical inverse Fourier imaging

The inverse Fourier transform of equation (2.10) yields

$$\mathcal{A}_i(l,m)\overline{\mathcal{A}_j(l,m)}I_B^c(l,m) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} V(u,v)\; e^{2\pi\jmath(ul+vm)}dudv \tag{2.12}$$

However, in practical situations, where not the entire aperture plane is or

can be sampled, only discrete or quasi-continuous measurements are available. The measured visibilities $v_o(u,v)$ can be expressed as the theoretical visibilities $V(u,v)$ multiplied by a sampling function $S_{sa}(u,v)$ which is 1 except at places in the uv plane where there are no sample points available; at those places $S_{sa}(u,v)=0$. The Fourier transform of $V_{sa}(u,v)=V(u,v)S_{sa}(u,v)$ is $I_D(l,m)$, which is known as the dirty map,

$$I_D(l,m)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}V(u,v)S_{sa}(u,v)\ e^{2\pi\jmath(ul+vm)}dudv \tag{2.13}$$

But, the Fourier transform of a product of $V(u,v)$ and $S_{sa}(u,v)$ is a convolution in the transform plane. Let $S^F_{ss}(l,m)$ be the Fourier transform of $S_{sa}(u,v)$, then:

$$I_D(l,m)=\left(\mathcal{A}_i(l,m)\overline{\mathcal{A}_j(l,m)}I^c_B(l,m)\right)*S^F_{ss}(l,m) \tag{2.14}$$

Here $*$ denotes the convolution operation. The function $S^F_{sa}(l,m)$ is called the dirty beam, and can be modified by applying weights to the sampling function $S_{sa}(u,v)$, for example in order to lower the sidelobe level. Given measured $V_{sa}(u,v)$ data, and a known sampling and weighting function $S_{sa}(u,v)$, the dirty beam and the dirty map can be calculated. The desired function $\mathcal{A}_i(l,m)\overline{\mathcal{A}_j(l,m)}I^c_B(l,m)$ can be obtained via a deconvolution process. A well-known algorithm based on a point source model is CLEAN [66] [130]. This algorithm contains an iterative loop in which the location and intensity of the strongest source are found in the dirty map (correlation with dirty beam), after which the source is subtracted. In each iteration one source location is found and (after convolving with the dirty beam, and scaled with the estimated intensity) subtracted. The aim is to minimise the residue until it converges to the sky map noise level.

The introduction above only describes some of the basic principles of synthesis imaging; more information on synthesis imaging can be found in [113, 146, 124, 117, 130, 93]

2.3 Single polarisation, discrete source formalism

So far, the aperture synthesis and imaging methods were described for spatially continuous celestial sources. Expressions for the coherencies , visibilities and covariances, were given in the form of integrals. In the next section, a discrete source model will be described, giving expressions in terms of summations. These are suitable for compact matrix expressions, as is common in other fields, such as signal processing for communications.

2.3.1 Discrete source model

The sky source distribution can be approximated by a dense population of resolved or unresolved point sources. Let q_s be the number of point sources,

and let $I_B(\mathbf{s})$ represent the discrete-source brightness distribution. Then the relation between the continuous and discrete-source distributions is given by:

$$I_B^c(\mathbf{s}) = \sum_{\ell=1}^{q_s} I_B(\mathbf{s})\delta(\mathbf{s}-\mathbf{s}_\ell) \tag{2.15}$$

where the delta function $\delta(\mathbf{s}-\mathbf{s}_\ell)$ is defined in such a way that the integral of $\delta(\mathbf{s}-\mathbf{s}_\ell)$ over the unit sphere is 1. Define $V_{ij} = V(r_i, r_j)$. Inserting this expression in equation (2.3) yields the following discrete visibility function:

$$V_{ij} = \sum_{\ell=1}^{q_s} \mathcal{A}_i(\mathbf{s}_\ell)\overline{\mathcal{A}_j(\mathbf{s}_\ell)} I_B(\mathbf{s}_\ell) e^{-2\pi\jmath\nu c^{-1}\mathbf{s}_\ell^t \mathbf{b}_{ij}} \tag{2.16}$$

Let $\mathbf{b}_{io}$ and $\mathbf{b}_{jo}$ be defined by $\mathbf{b}_{ij} = (\mathbf{r}_i - \mathbf{r}_0) - (\mathbf{r}_j - \mathbf{r}_0) = \mathbf{b}_{i0} - \mathbf{b}_{j0}$, where the index 0 refers to a fixed (arbitrary) reference point $\mathbf{r}_0$. The discrete visibility function can then be rewritten as:

$$V_{ij} = \sum_{\ell=1}^{q_s} \left(e^{-2\pi\jmath\nu c^{-1}\mathbf{s}_\ell^t \mathbf{b}_{i0}} \mathcal{A}_i(\mathbf{s}_\ell)\right) I_B(\mathbf{s}_\ell) \left(\overline{\mathcal{A}_j(\mathbf{s}_\ell)} e^{2\pi\jmath\nu c^{-1}\mathbf{s}_\ell^t \mathbf{b}_{j0}}\right) \tag{2.17}$$

This equation describes the visibility response for source $\mathbf{s}_k$ at a baseline $\mathbf{b}_{ij}$, or in other words, it describes the spatial electric field coherency which can be measured by means of an interferometer with a certain baseline. For p telescopes, $p \times p$ correlation products or visibilities can be formed. These correlation products, and the components of the visibility function of which it is consists, can be conveniently stacked in matrices. The advantage of the matrix formalism is that a wide range of signal processing algorithms, originating from other fields, becomes available to radio astronomy signal processing applications.

Define the $p \times p$ visibility or coherency matrix $\mathbf{V}$ in which all observed V_{ij} visibility points ($i = 1, \cdots, p$, $j = 1, \cdots, p$) are stacked as

$$\mathbf{V} = \begin{pmatrix} V_{11} & \cdots & V_{1p} \\ \vdots & \ddots & \vdots \\ V_{p1} & \cdots & V_{pp} \end{pmatrix} \tag{2.18}$$

Define $\mathcal{R}$ by

$$\mathcal{R} = (\mathbf{b}_{1o}, \cdots, \mathbf{b}_{po})^t \tag{2.19}$$

and let the geometric delay vector be $\mathbf{a}^d(\mathbf{s}_\ell)$, and the telescope beam response of the ℓ^{th} source be $\mathbf{a}^b(\mathbf{s}_\ell)$, then

$$\mathbf{a}^d(\mathbf{s}_\ell) = e^{-2\pi\jmath\nu c^{-1}\mathcal{R}\mathbf{s}_\ell}, \qquad \mathbf{a}^b(\mathbf{s}_\ell) = [\mathcal{A}_1(\mathbf{s}_\ell), \cdots, \mathcal{A}_p(\mathbf{s}_\ell)]^t \tag{2.20}$$

The geometric delay vector is time-dependent, but known; the telescope main beam gain $\mathcal{A}_i$ is direction-dependent, and is assumed to be constant in time and is now taken equal for all telescopes: $\mathcal{A}_i(\mathbf{s}_\ell) = \mathcal{A}(\mathbf{s}_\ell)$ and hence $\mathbf{a}^b(\mathbf{s}_\ell) = \mathcal{A}(\mathbf{s}_\ell)\mathbf{1}$.

The array response vector $\mathbf{a}(\mathbf{s}_\ell)$ is defined by the element-wise multiplication of the vectors defined previously: $\mathbf{a}(\mathbf{s}_\ell) = \mathbf{a}^d(\mathbf{s}_\ell) \odot \mathbf{a}^b(\mathbf{s}_\ell)$. Let $B_\ell = I_B(\mathbf{s}_\ell)$, then the $p \times p$ visibility matrix $\mathbf{V}$ can be expressed as

$$\mathbf{V} = \sum_{\ell=1}^{q_s} \mathbf{a}(\mathbf{s}_\ell)\, B_\ell\, \mathbf{a}(\mathbf{s}_\ell)^H \tag{2.21}$$

Define the diagonal gain matrix $\mathbf{G}$ by

$$\mathbf{G} = \begin{pmatrix} G_1 & & 0 \\ & \ddots & \\ 0 & & G_p \end{pmatrix} \tag{2.22}$$

Because $\mathbf{R} = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}$, $\mathbf{V}_{ij} = \mathcal{E}\{E(\mathbf{r}_i)\overline{E(\mathbf{r}_j)}\}$, and $x(\mathbf{r}_i) = G_i E(\mathbf{r}_i)$, the relation between $\mathbf{R}$ and $\mathbf{V}$ is:

$$\mathbf{R} = \mathbf{G}\mathbf{V}\mathbf{G}^H \tag{2.23}$$

and thus

$$\boxed{\mathbf{R} = \textstyle\sum_{\ell=1}^{q_s} \mathbf{G}\, \mathbf{a}(\mathbf{s}_\ell)\, B_\ell\, \mathbf{a}(\mathbf{s}_\ell)^H\, \mathbf{G}^H} \tag{2.24}$$

This can be written more compactly. Let the $p \times q_s$ array response matrix $\mathbf{A}$be defined by

$$\mathbf{A} = (\mathbf{a}(\mathbf{s}_1), \cdots, \mathbf{a}(\mathbf{s}_{q_s})) \tag{2.25}$$

and let the diagonal $q_s \times q_s$ brightness distribution matrix $\mathbf{B}$ be defined by

$$\mathbf{B} = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_{q_s} \end{pmatrix} \tag{2.26}$$

Straightforward verification shows that the visibility function (2.17) and the covariance matrix can be written in compact matrix forms as:

$$\mathbf{V} = \mathbf{A}\mathbf{B}\mathbf{A}^H \tag{2.27}$$

$$\mathbf{R} = \mathbf{G}\, \mathbf{A}\mathbf{B}\mathbf{A}^H\, \mathbf{G}^H \tag{2.28}$$

Equations (2.24) and (2.28) are discrete source versions of equation (2.5); equations (2.21) and (2.27) are discrete source versions of equation (2.3). The complex gain matrix $\mathbf{G}$ usually slowly varies (> 1 minute), and the array response matrix $\mathbf{A}$ varies with time due to source tracking. Most sources do not time vary with time, so the brightness distribution $\mathbf{B}$ is usually time-invariant. Hence equation (2.28) is more accurately written as

$$\boxed{\mathbf{R}(t) = \mathbf{G}\, \mathbf{A}(t)\mathbf{B}\mathbf{A}(t)^H\, \mathbf{G}^H} \tag{2.29}$$

Following [63], a-priori knowledge of some of the telescope system parameters and propagation paths can be included in the modelling in a straightforward

way. The gain and array response vectors may be composed partly of known and partly of unknown parameters. Assuming the number of unknowns is not too large, they can be found. Detailed analysis of this topic is not a goal of this thesis, but it will become an important issue for telescopes such as LOFAR [16] and SKA .

Discrete source imaging approaches

Using a parametric point-source model, the image deconvolution problem in the noiseless case can be interpreted as a direction-of-arrival (DOA) estimation problem. Assuming the gains $\mathbf{G}$ are constant and known from calibration, they can be omitted from the model without loss of generality. Writing the time dependence of the model and model estimates as a discrete time index k and assuming K measurement snapshots, a least squares covariance fitting approach to the DOA problem is written as

$$[\{\hat{\mathbf{s}_\ell}\}, \widehat{\mathbf{B}}] = \arg \min_{\{\mathbf{s}_\ell\}, \mathbf{B}} \sum_{k=1}^{K} \| \widehat{\mathbf{R}}_k - \mathbf{A}_k \, \mathbf{B} \mathbf{A}_k^H \|_F^2 \tag{2.30}$$

where $\{\mathbf{s}_\ell\} \equiv \{\mathbf{s}_1, \cdots, \mathbf{s}_{q_s}\}$, and where $\mathbf{A}_k = (\mathbf{a}_k(\mathbf{s}_1), \cdots, \mathbf{a}_k(\mathbf{s}_{q_s}))$. In the estimation of $\{\mathbf{s}_\ell\}$ and $\mathbf{B}$, $\widehat{\mathbf{B}}$ is constrained to be diagonal with positive entries. This model is similar to the model used for DOA estimation in array processing. Note that there are many more sources than the dimension of each covariance matrix $\widehat{\mathbf{R}}_k$. In this notation, the image formation can be formulated as follows. If we write $I_D(\mathbf{s}) \equiv I_D(l, m)$ and $\mathbf{a}_k(\mathbf{s}) \equiv \mathbf{a}_k(l, m)$, we can rewrite the dirty image (2.13) as [151]

$$I_D(\mathbf{s}) = \sum_{k=1}^{K} \mathbf{a}_k(\mathbf{s})^H \, \mathbf{R}_k \, \mathbf{a}_k(\mathbf{s}) \tag{2.31}$$

With system noise, $\widehat{\mathbf{R}}$ is replaced by $\widehat{\mathbf{R}} - \widehat{\mathbf{D}}$, where $\widehat{\mathbf{D}}$ is a diagonal covariance noise matrix; see for instance section 2.3.2 for details. The iterative beam removing in CLEAN can now be posed as an iterative LS fitting between the sky model and the observed visibility [130]. Finding the brightest point $\mathbf{s}_0$ in the image is equivalent to trying to find a point source using classical Fourier beamforming, that is:

$$\hat{\mathbf{s}}_0 = \arg \max_{\mathbf{s}} \sum_{k=1}^{K} \mathbf{a}_k(\mathbf{s})^H \, \widehat{\mathbf{R}}_k \, \mathbf{a}_k(\mathbf{s}) \tag{2.32}$$

Thus, the CLEAN algorithm can be regarded as a generalised classical sequential beamformer, where the brightest points are found one by one, and subsequently removed from $\mathbf{R}_k$ until the LS cost function (2.30) is minimised. An immediate consequence is that the estimated source locations will be biased, a well-known fact in array processing. If the sources are well separated the bias is negligible compared to the standard deviation, otherwise it might be significant. This explains the limited performance of CLEAN in imaging extended structures (see e.g. [113]).

An alternative solution to the deconvolution problem is to solve equation (2.30) in closed form in the following way. Let $\hat{\mathbf{v}}_k \equiv \text{vec}(\widehat{\mathbf{R}}_k)$, and $\mathbf{b} \equiv \text{vecdiag}(\mathbf{B})$. Applying equation (B.22) to $\mathbf{A}_k \mathbf{B} \mathbf{A}_k^H$ yields the following expression

$$\hat{\mathbf{v}}_k = [\overline{\mathbf{A}_k} \circ \mathbf{A}_k]\mathbf{b} \tag{2.33}$$

Stacking all time sample data on top of each other yields:

$$\begin{bmatrix} \hat{\mathbf{v}}_1 \\ \vdots \\ \hat{\mathbf{v}}_K \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{A}_1} \circ \mathbf{A}_1 \\ \vdots \\ \overline{\mathbf{A}_K} \circ \mathbf{A}_K \end{bmatrix} \mathbf{b} \tag{2.34}$$

Define

$$\mathbf{A}_+ = \begin{bmatrix} \overline{\mathbf{A}_1} \circ \mathbf{A}_1 \\ \vdots \\ \overline{\mathbf{A}_K} \circ \mathbf{A}_K \end{bmatrix} \tag{2.35}$$

and $\hat{\mathbf{v}}_+ = [\hat{\mathbf{v}}_1, \cdots, \hat{\mathbf{v}}_K]^t$. Inserting these in the previous equations and solving for $\mathbf{b}$ yields:

$$\mathbf{b} = \mathbf{A}_+^\dagger \mathbf{v}_+ = (\mathbf{A}_+^H \mathbf{A}_+)^{-1} \mathbf{A}_+^H \hat{\mathbf{v}}_+ \tag{2.36}$$

The factor $\mathbf{A}_+^H \mathbf{v}_+$ corresponds to the dirty image, and the matrix $(\mathbf{A}_+^H \mathbf{A}_+)^{-1}$ to the deconvolution step. This approach will lead to a least squares solution, but will in practice be computationally very expensive, as the matrices to be inverted are extremely large: $\mathcal{O}(N_q \times N_q)$ with N_q the number of pixels in the image. The matrix inversion requires $\mathcal{O}(N_q^3)$ operations. An approach such as CLEAN may lead to less optimal solutions, but the computational requirements are much lower: for each pixel (brightest point) the beam has to be subtracted from all pixels, requiring subtraction operations on $\mathcal{O}(N_q^2)$ pixels.

2.3.2 Additive interference and noise

In this section a model is described for (additive) interfering signals and noise. Assume that the telescope signals x_i are composed of q_s astronomical source signals, q_r interfers, and noise. As before, the telescope output signals x_i are stacked in a vector $\mathbf{x} = (x_1, x_2, \cdots, x_p)^t$. Let $\mathbf{x}_\ell^s$ be the telescope array output signal corresponding to the ℓ^{th} astronomical source in the direction $\mathbf{s}_\ell$, let $\mathbf{x}_k^r$ be the telescope array output signal corresponding to the k^{th} interferer in the direction $\mathbf{s}_k^r$, and let $\mathbf{x}^n$ be the noise vector, then the resulting array output signal can be expressed as

$$\mathbf{x} = \sum_{\ell=1}^{q_s} \mathbf{x}_\ell^s + \sum_{k=1}^{q_r} \mathbf{x}_k^r + \mathbf{x}^n \tag{2.37}$$

The noise $\mathbf{x}^n$ is independent identically distributed (i.i.d.) Gaussian noise, so it is uncorrelated between the array elements, or, in other words, spatially white

at the aperture plane. The astronomical source signals are assumed to be identically distributed Gaussian noise signals. The sources are assumed to be independent, or in other words spatially white at the celestial sphere. Finally, the interferers are assumed to be quasi stationary, Gaussian noise signals. The interferers may either be mutually correlated (multipath) or not. Figure 2.6 schematically shows the telescope set-up with the three types of input signals. A matrix formulation for the source contributions is already derived, so the next

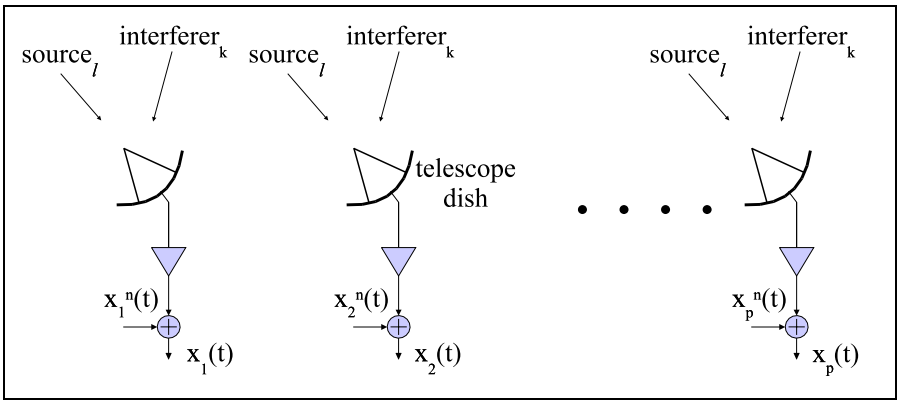


Figure 2.6. Telescope array output signal vector $\mathbf{x} = (x_1, \cdots, x_p)^t$ contains contributions from sky sources, from interference, and includes system noise.

step is to derive a matrix expression for the interference contributions and noise.

In case the interferers are in the far field of the telescope array, and are stationary for the duration of the observation, and assuming that there is no multipath, then the matrix derivations are completely analogous to the astronomical sources case. However, many interferers occur in the near field of large radio telescopes such as the WSRT and the VLA . In this case, assuming that the narrowband condition is valid, all the formulas in the previous derivations still hold, the only difference being that the array response vector for interferers $\mathbf{a}^r$ is written differently (because of the near field condition), and that the array response vector is usually unknown.

Assume that there is an interferer with the index k, with signal $y_k^r(t)$. Because the narrowband condition holds, the telescope output signal $\mathbf{x}_k^r(t)$ can be written in terms of (unknown) gains a_i^r. Let the array response vector $\mathbf{a}_k^r$ be defined by

$$\mathbf{a}_k^r = \begin{pmatrix} a_1^r e^{2\pi\jmath\frac{1}{\lambda}(|\mathbf{r}_k^r - \mathbf{r}_1|)} \\ \vdots \\ a_p^r e^{2\pi\jmath\frac{1}{\lambda}(|\mathbf{r}_k^r - \mathbf{r}_p|)} \end{pmatrix} \tag{2.38}$$

then, dropping the time dependence from $y_k^r(t)$, this yields $\mathbf{x}_k^r = \mathbf{a}_k^r y_k^r$. Assume

there are q_r interferers, then $\mathbf{x}^r = \sum_k \mathbf{x}_k^r$, and

$$\mathbf{x}^r = \sum_{k=1}^{q_r} \mathbf{a}_k^r y_k^r \tag{2.39}$$

The diagonal interferer power matrix $\mathbf{B}^r$ and the array response matrix for interferers $\mathbf{A}^r$ are defined analogously to the astronomical source definitions. The diagonal gain matrix $\mathbf{G}$ for the interferers is identical to the gain matrix for the astronomical sources, as it contains the telescope electronic gain contributions, which are source-independent. Taking into account only the interferer contribution, the resulting covariance matrix is given by: $\mathbf{G}\mathbf{A}^r\mathbf{B}^r(\mathbf{A}^r)^H\mathbf{G}^H$.

Receiver noise, originating from, for example, amplifiers, is usually spatially uncorrelated, so that it appears only on the main diagonal of the covariance matrix: $\mathcal{E}\{x_i^n(x_j^n)^H\} = 0 \ \forall \ (i \neq j)$. Let the channel receiver noise power standard deviations be σ_{ni}, and stacked in a vector $\boldsymbol{\sigma_n} = (\sigma_{n1}, \cdots, \sigma_{np})^t$. The receiver noise covariance matrix $\mathbf{D} = \mathcal{E}\{\mathbf{x}^n(\mathbf{x}^n)^H\}$ is given by $\mathbf{D} = (\text{diag}(\boldsymbol{\sigma_n}))^2$.

Finally, interferometer correlator errors may occur if the correlation process is somehow affected. For example, inaccuracies in determining the threshold values for ADC levels will generate correlator offsets which are interferometer based and are not modelled easily. Correlator errors can be represented as an additive term in the covariance matrix. However, if we disregard correlator errors and instrumental crosstalk effects, we have:

$$\mathbf{R} \ = \mathbf{G}(\mathbf{A}\mathbf{B}\mathbf{A}^H + \mathbf{A}^r\mathbf{B}^r\mathbf{A}^{rH})\mathbf{G}^H + \mathbf{D} \tag{2.40}$$

This model also assumes that there is no correlation between the astronomical sources, the interferers and the channel/system noise. As before, an alternative expression for this equation is:

$$\mathbf{R} = \mathbf{G}(\sum_{\ell=1}^{q_s} \mathbf{a}(\mathbf{s}_\ell)B_\ell\mathbf{a}(\mathbf{s}_\ell)^H + \sum_{k=1}^{q_r} \mathbf{a}^r(\mathbf{s}_k^r)B_k^r\mathbf{a}^r(\mathbf{s}_k^r)^H)\mathbf{G}^H + \mathbf{D} \tag{2.41}$$

For map-making, the goal is to estimate $\mathbf{B}$, given that a sample estimate $\widehat{\mathbf{R}}$ of $\mathbf{R}$ is available. The gains $\mathbf{G}$ and some terms in $\mathbf{A}$ are initially unknown but can be estimated from $\widehat{\mathbf{R}}$ using calibration schemes, such as selfcal techniques in which the telescope based calibration errors/unknowns are found by comparing measured data with expected data based on sky models [169], and by the CLEAN process. A relatively simple calibration scheme in which one dominant source was observed will be described and thoroughly analysed in chapter 8.

The telescope noise powers can be calibrated, for example, by comparing the off-diagonal elements of the normalised covariance matrix, that is, a matrix containing correlation coefficients [167]

$$\mathbf{R}_{cc} = (\text{diag}\mathbf{R})^{-\frac{1}{2}}\mathbf{R}((\text{diag}\mathbf{R})^{-\frac{1}{2}})^H, \tag{2.42}$$

with the predicted values based on the sky model. For one dominant source, this scenario is part of the calibration scheme described in chapter 8.

Applying the discrete source model to interference mitigation tests involving an eigenvalue decomposition of $\widehat{\mathbf{R}}$ often requires that the noise matrix takes the form $\mathbf{D} = \sigma_n^2\mathbf{I}$ or that the noise matrix is estimated and subtracted from $\widehat{\mathbf{R}}$ before applying the eigenvalue decomposition technique. The reason for this is that the subspace estimation results are changed if the diagonal noise matrix entries are non-identical, as will be explained in the next chapter.

One approach if $\mathbf{D} \neq \sigma_n^2\mathbf{I}$ would be to apply eigenvalue decomposition to the normalised covariance matrix $\mathbf{R}_{cc}$, and then to reverse the normalisation. A second approach, which has already been mentioned, is to estimate and subtract $\mathbf{D}$. A third method is to pre-whiten $\widehat{\mathbf{R}}$ with estimates of $\widehat{\mathbf{D}}$ in the following way. Define the pre-whitened matrix $\mathbf{R}_w$ by $\mathbf{R}_w = \widehat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{R}(\widehat{\mathbf{D}}^{-\frac{1}{2}})^H$, then:

$$\mathbf{R}_w = \widehat{\mathbf{D}}^{-\frac{1}{2}}\,\mathbf{G}\,(\mathbf{A}\mathbf{B}\mathbf{A}^H + \mathbf{A}_r\mathbf{R}_r\mathbf{A}_r^H)\,\mathbf{G}^H\,(\widehat{\mathbf{D}}^{-\frac{1}{2}})^H + \mathbf{I} \tag{2.43}$$

The noise matrix is now the identity matrix, which is advantageous for many interference mitigation algorithms, as these often assume i.d.d. noise. More on whitening in relation to RFI mitigation and subspace analysis can be found in the next chapter.

2.4 Polarisation formalism

2.4.1 Discrete source model

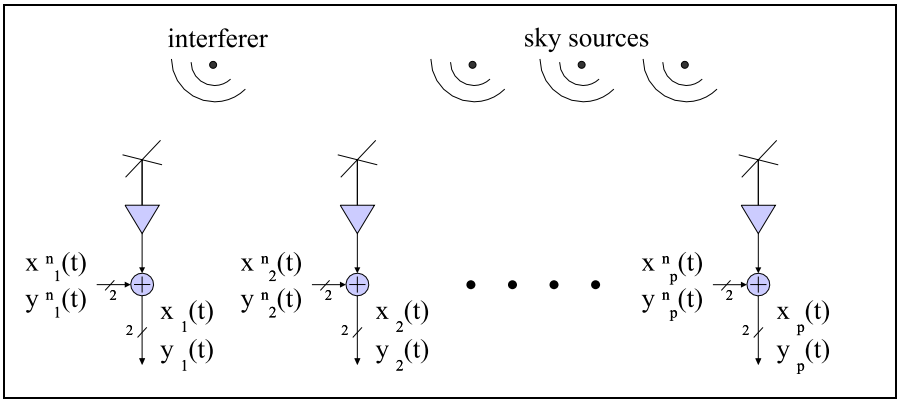


Figure 2.7. Dual polarisation telescope array output signal vectors contain contributions from sky sources, from interference, and from system noise.

In this section the scalar description of the coherencies is extended to a representation in which the electric field polarisation is taken into account. It follows the work of Hamaker et al [63] [62], which is extended in the sense that a full covariance matrix representation is derived. At this point it is assumed

that there is no interference. The electric field at the location $\mathbf{r}_i$ of an antenna element can be described by two linear polarisation components, stacked in a 2×1 vector: $\mathbf{e}_i = (E_x(\mathbf{r}_i),\ E_y(\mathbf{r}_i))^t$. The correlation between two different telescopes i and j is a 2×2 interferometer coherency matrix defined by $\widetilde{\mathbf{V}}_{ij} = \mathcal{E}\{\mathbf{e}_i \mathbf{e}_j^H\}$. If there are p telescopes, each with two polarisations, then the $2p$ observed electric fields can similarly be stacked in one vector: $\mathbf{e} = \left(\mathbf{e}_1^t, \cdots, \mathbf{e}_p^t\right)^t$. The $2p \times 2p$ Hermitian coherency matrix $\widetilde{\mathbf{V}}$ is defined by $\widetilde{\mathbf{V}} = \mathcal{E}\{\mathbf{e}\mathbf{e}^H\}$ which can be written in terms of interferometer coherency matrices $\widetilde{\mathbf{V}}_{ij}$ as

$$\widetilde{\mathbf{V}}_{ij} = \mathcal{E}\{\mathbf{e}_i \mathbf{e}_j^H\} = \begin{pmatrix} \mathcal{E}\{E_x(\mathbf{r}_i)\overline{E_x(\mathbf{r}_j)}\} & \mathcal{E}\{E_x(\mathbf{r}_i)\overline{E_y(\mathbf{r}_j)}\} \\ \mathcal{E}\{E_y(\mathbf{r}_i)\overline{E_x(\mathbf{r}_j)}\} & \mathcal{E}\{E_y(\mathbf{r}_i)\overline{E_y(\mathbf{r}_j)}\} \end{pmatrix} \tag{2.44}$$

and

$$\widetilde{\mathbf{V}} = \mathcal{E}\{\mathbf{e}\mathbf{e}^H\} = \begin{pmatrix} \widetilde{\mathbf{V}}_{11} & \cdots & \widetilde{\mathbf{V}}_{1p} \\ \vdots & \ddots & \vdots \\ \widetilde{\mathbf{V}}_{p1} & \cdots & \widetilde{\mathbf{V}}_{pp} \end{pmatrix} \tag{2.45}$$

Note that each $\widetilde{\mathbf{V}}_{ij}$ is non-Hermitian for $i \neq j$, but that $\widetilde{\mathbf{V}}_{ij} = \widetilde{\mathbf{V}}_{ji}^H$ so that $\widetilde{\mathbf{V}}$ is Hermitian. $\widetilde{\mathbf{V}}$ is dependent on frequency and time, but for the analysis it is assumed that the narrow-band assumption is valid.

Suppose that the ℓ^{th} source has sky brightness $\widetilde{\mathbf{B}}_\ell$, which is a 2×2 matrix specifying the source flux polarisation components or Stokes parameters [62]. Let $\widetilde{\mathbf{A}}_{i\ell}$ be the source dependent polarisation multiplication matrix. This is the array response matrix, or in the Hamaker formalism [63], a Jones matrix or a combination of Jones matrices. Furthermore let the source brightness be defined in terms of Stokes parameters :

$$\widetilde{\mathbf{B}}_\ell = I_s^\ell \mathbf{I}_s + Q_s^\ell \mathbf{Q}_s + U_s^\ell \mathbf{U}_s + V_s^\ell \mathbf{V}_s \tag{2.46}$$

with I_s^ℓ, Q_s^ℓ, U_s^ℓ, and V_s^ℓ real scalars (Stokes parameters of the ℓ^{th} source), and

$$\mathbf{I}_s = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{Q}_s = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \mathbf{U}_s = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \mathbf{V}_s = \begin{pmatrix} 0 & -\imath \\ \imath & 0 \end{pmatrix} \tag{2.47}$$

which are known in physics as Pauli spin matrices. Then for each of the four $\widetilde{\mathbf{V}}_{ij}$ coherency components, for each interferometer ij, for each source ℓ, the following relation, consisting of 2×2 matrices, holds.

$$\widetilde{\mathbf{V}}_{ij\ell} = \widetilde{\mathbf{A}}_{i\ell} \widetilde{\mathbf{B}}_\ell \widetilde{\mathbf{A}}_{j\ell}^H \tag{2.48}$$

The source-dependent polarisation multiplication matrix $\widetilde{\mathbf{A}}_{i\ell}$ represents different physical and instrumental effects. All of these effects are represented by matrices, also known as Jones matrices [63], and correspond to transformations due to ionospheric polarisation rotation or Faraday rotation $\widetilde{\mathbf{A}}_\ell^f$, parallactic offsets due to the orientation of the antenna with respect to the source $\widetilde{\mathbf{A}}_\ell^p$, the

geometric delay offsets $\widetilde{\mathbf{A}}^d_{i\ell}$, and source-dependent telescope gain $\widetilde{\mathbf{A}}^g_\ell$ which is assumed to be the same for all telescopes: $\widetilde{\mathbf{A}}_{i\ell} = \widetilde{\mathbf{A}}^g_\ell \widetilde{\mathbf{A}}^d_{i\ell} \widetilde{\mathbf{A}}^p_\ell \widetilde{\mathbf{A}}^f_\ell$.

Following the derivations in the non-polarised case, the factors in $\widetilde{\mathbf{A}}_{il}$ can be expressed as

$$\begin{aligned}
\widetilde{\mathbf{A}}^f_\ell &= \begin{pmatrix} \cos(\phi_\ell) & -\sin(\phi_\ell) \\ \sin(\phi_\ell) & \cos(\phi_\ell) \end{pmatrix} \\
\widetilde{\mathbf{A}}^p_\ell &= \begin{pmatrix} \cos(\alpha_\ell) & -\sin(\alpha_\ell) \\ \sin(\alpha_\ell) & \cos(\alpha_\ell) \end{pmatrix} \\
\widetilde{\mathbf{A}}^d_{i\ell} &= e^{-2\pi\jmath\lambda^{-1}\mathbf{b}^t_{io}\mathbf{s}_\ell} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
\widetilde{\mathbf{A}}^d_\ell &= \begin{bmatrix} \mathcal{A}_x & 0 \\ 0 & \mathcal{A}_y \end{bmatrix}
\end{aligned} \tag{2.49}$$

The Faraday rotation ϕ_ℓ is assumed to be constant over the array and is initially unknown. The parallactic offset phase α_ℓ is known from the telescope geometry. The delay matrix is telescope-dependent, and is determined by the geometry as well. Finally, assuming that there is no crosstalk, the telescope gain matrix $\widetilde{\mathbf{A}}^g_\ell$ is diagonal and assumed to be the same for all telescopes. Initially it is unknown, but it can be calibrated. Using the following definitions for the polarisation brightness and polarisation array response matrices

$$\widetilde{\mathbf{B}} \equiv \begin{pmatrix} \widetilde{\mathbf{B}}_1 & & 0 \\ & \ddots & \\ 0 & & \widetilde{\mathbf{B}}_{q_s} \end{pmatrix}, \quad \widetilde{\mathbf{A}}_\ell = \begin{pmatrix} \widetilde{\mathbf{A}}_{1\ell} \\ \vdots \\ \widetilde{\mathbf{A}}_{p\ell} \end{pmatrix}, \quad \widetilde{\mathbf{A}} = (\widetilde{\mathbf{A}}_1, \cdots, \widetilde{\mathbf{A}}_{q_s}) \tag{2.50}$$

$\widetilde{\mathbf{V}}$ can be expressed as

$$\widetilde{\mathbf{V}} = \sum_{\ell=1}^{q_s} \widetilde{\mathbf{A}}_\ell \widetilde{\mathbf{B}}_\ell \widetilde{\mathbf{A}}^H_\ell = \widetilde{\mathbf{A}} \widetilde{\mathbf{B}} \widetilde{\mathbf{A}}^H \tag{2.51}$$

Instead of the field strengths, each telescope measures a voltage vector $\widetilde{\mathbf{x}}_i$, where $\widetilde{\mathbf{x}}_i = (x_{ix}, x_{ij})^t$. Their relation is given by $\widetilde{\mathbf{x}}_i = \widetilde{\mathbf{G}}_i \mathbf{e}_i$, where $\widetilde{\mathbf{G}}_i$ is a 2×2 telescope gain matrix The observed voltages of the dual polarisation output signals of the telescopes i and j are cross-correlated into covariance matrices $\widetilde{\mathbf{R}}_{ij}$, for which $\widetilde{\mathbf{R}}_{ij} = \mathcal{E}\{\widetilde{\mathbf{x}}_i \widetilde{\mathbf{x}}^H_j\} = \widetilde{\mathbf{G}}_i \widetilde{\mathbf{V}}_{ij} \widetilde{\mathbf{G}}^H_j$. The telescope output voltages $\widetilde{\mathbf{x}}_i$ are stacked into a $2p$-dimensional vector $\widetilde{\mathbf{x}} = [\widetilde{\mathbf{x}}^t_1, \cdots, \widetilde{\mathbf{x}}^t_p]^t$ and subsequently $\widetilde{\mathbf{R}}$ is defined by $\widetilde{\mathbf{R}} = \mathcal{E}\{\widetilde{\mathbf{x}} \widetilde{\mathbf{x}}^H\}$:

$$\widetilde{\mathbf{R}} = \begin{bmatrix} \widetilde{\mathbf{R}}_{11} & \cdots & \widetilde{\mathbf{R}}_{1p} \\ \vdots & \ddots & \vdots \\ \widetilde{\mathbf{R}}_{p1} & \cdots & \widetilde{\mathbf{R}}_{pp} \end{bmatrix} \tag{2.52}$$

Next, the block diagonal gain matrix is defined by

$$\widetilde{\mathbf{G}} = \text{blockdiag}(\widetilde{\mathbf{G}}_1, \cdots, \widetilde{\mathbf{G}}_p) \tag{2.53}$$

The $\widetilde{\mathbf{G}}_i$ matrices are also called Jones matrices, and in our definition they were astronomical source-independent. Ideally, the two polarisation signal paths of a telescope are the same, leading to diagonal matrices $\widetilde{\mathbf{G}}_i$. Now it can easily be seen that the $2p \times 2p$ covariance matrix $\widetilde{\mathbf{R}}$ can be expressed as

$$\widetilde{\mathbf{R}} = \widetilde{\mathbf{G}}\widetilde{\mathbf{V}}\widetilde{\mathbf{G}}^H \tag{2.54}$$

Note that these expressions are nearly the same as for the non-polarised case, and that the matrix dimensions are twice the dimensions of the non-polarised case. By inserting the expression for $\widetilde{\mathbf{V}}$ in the equation above, the polarisation covariance matrices are given by

$$\boxed{\widetilde{\mathbf{R}} = \textstyle\sum_{\ell=1}^{q_s} \widetilde{\mathbf{G}}\ \widetilde{\mathbf{A}}_\ell\ \widetilde{\mathbf{B}}_\ell\ \widetilde{\mathbf{A}}_\ell^H\ \widetilde{\mathbf{G}}^H} \tag{2.55}$$

$$\boxed{\widetilde{\mathbf{R}} = \widetilde{\mathbf{G}}\ \widetilde{\mathbf{A}}\ \widetilde{\mathbf{B}}\ \widetilde{\mathbf{A}}^H\ \widetilde{\mathbf{G}}^H} \tag{2.56}$$

2.4.2 Additive noise and interference

In practice the observations are influenced by noise. The system noise signals of each of the two polarisation channels $\widetilde{\mathbf{n}}_i \equiv [n_{ix}, n_{iy}]^t$ are stacked into a vector: $\widetilde{\mathbf{n}} \equiv \left(\widetilde{\mathbf{n}}_1^t, \cdots, \widetilde{\mathbf{n}}_p^t\right)^t$. The noise signals are uncorrelated between the telescopes, and up to a certain level also uncorrelated between the two polarisations of a telescope.

If the noise between the two polarisations of a single telescope is uncorrelated then the noise matrix $\widetilde{\mathbf{D}} \equiv \mathcal{E}\{\widetilde{\mathbf{n}}\widetilde{\mathbf{n}}^H\}$ is diagonal:

$$\widetilde{\mathbf{D}} = \text{diag}(\sigma^2_{n_1x}, \sigma^2_{n_1y}, \cdots, \sigma^2_{n_px}, \sigma^2_{n_py}) \tag{2.57}$$

where $\sigma^2_{n_ix}$ and $\sigma^2_{n_iy}$ are the channel noise variances of the x- and y- polarisation respectively. On the other hand, if the noise between the two polarisations of a single telescope is correlated then the noise matrix is blockdiagonal. Let $\widetilde{\mathbf{D}}_{ij}$ be $\mathcal{E}\{\widetilde{\mathbf{n}}_i\widetilde{\mathbf{n}}_j^H\}$, then:

$$\widetilde{\mathbf{D}} = \begin{bmatrix} \widetilde{\mathbf{D}}_{11} & & 0 \\ & \ddots & \\ 0 & & \widetilde{\mathbf{D}}_{pp} \end{bmatrix} \tag{2.58}$$

The system noise can be considered additive, so that the covariance matrix of the received data can be written as

$$\widetilde{\mathbf{R}} = \widetilde{\mathbf{G}}\,\widetilde{\mathbf{V}}\widetilde{\mathbf{G}}^H + \widetilde{\mathbf{D}} \tag{2.59}$$

If interferers are in the far field of a telescope and are stationary, then the model derivations are analogous to the astronomical source derivations. In case the interferes are in the far field of the telescope antennas (dishes), but in the near field of the array, then the model is still valid, but the gain matrices $\widetilde{\mathbf{A}}^r$, and $\widetilde{\mathbf{A}}^r_\ell$ will have to be adapted. This is a straightforward operation, and is not included here. Ignoring correlator errors, the resulting two equivalent expressions for $\widetilde{\mathbf{R}}$ are:

$$\boxed{\widetilde{\mathbf{R}} = \widetilde{\mathbf{G}} \left(\sum_{\ell=1}^{q_s} \widetilde{\mathbf{A}}_\ell \, \widetilde{\mathbf{B}}_\ell \, \widetilde{\mathbf{A}}_\ell^H + \sum_{k=1}^{q_r} \widetilde{\mathbf{A}}_k^r \, \widetilde{\mathbf{B}}_k^r \, \widetilde{\mathbf{A}}_k^{rH} \right) \widetilde{\mathbf{G}}^H + \widetilde{\mathbf{D}}} \tag{2.60}$$

$$\boxed{\widetilde{\mathbf{R}} = \widetilde{\mathbf{G}} \left(\widetilde{\mathbf{A}} \, \widetilde{\mathbf{B}} \, \widetilde{\mathbf{A}}^H \; + \; \widetilde{\mathbf{A}}^r \, \widetilde{\mathbf{B}}^r \, \widetilde{\mathbf{A}}^{rH} \right) \widetilde{\mathbf{G}}^H + \widetilde{\mathbf{D}}} \tag{2.61}$$

As in the non-polarised case, a difference between astronomical and man-made interferers is that the properties of the interferers (direction, distance, antenna sidelobe level) are usually not known.

2.5 Interference models

2.5.1 Multipath model

The interference model derived in the previous section assumed that the interferers were mutually independent. In the non-polarisation case this led to the model with the $p \times q_r$ array response matrix and the diagonal matrix $\mathbf{B}^r$ corresponding to the interferers. In the polarisation case it would lead to a $2p \times 2q_r$ array response matrix, and to the $2q_r \times 2q_r$ blockdiagonal matrix $\widetilde{\mathbf{B}}^r$. Now the question is how these models will change when there is multipath. The effect of multipath is that there are copies of the original signal which are time-delayed and attenuated. This section is limited to the non-polarised case, but the model is easily extended to polarisation. Let $\mathbf{R}^r$ be the covariance matrix of the telescope array output signals only taking into account the interferers and disregarding the astronomical sources and noise for the moment. Let the number of interferers be q_r and the number of multipaths, including the main path, for each interferer k be L_k. Then the telescope output vector with multipaths taken into account is given by:

$$\mathbf{x}^r(t) = \sum_{k=1}^{q_r} \left(\sum_{l=1}^{L_k} \mathbf{a}^r_{kl} \, y_k(t - \tau_{kl}) \right) \tag{2.62}$$

where $\mathbf{a}^r_{kl}$ is the array response vector of an interferer multipath , and $y_k(t-\tau_{kl})$ is the corresponding signal with a multipath time delay τ_{kl}. The relative path differences to all telescopes for any interferer-path combination is absorbed in $\mathbf{a}_{kl}$, so that the time delay τ_{kl} does not require a telescope index i. Let the array response vector, including multipath directions, be defined by

$$\mathbf{A}^r = [(\mathbf{a}^r_{11}, \cdots, \mathbf{a}^r_{1L_1}), \cdots, (\mathbf{a}^r_{q_r 1}, \cdots, \mathbf{a}^r_{q_r L_{q_r}})] \tag{2.63}$$

and stack the interferer signals in a vector $\mathbf{y}(t)$

$$\mathbf{y}(t) = [\,[y_1(t-\tau_{11}),\cdots,y_1(t-\tau_{1l_1})],\cdots,[y_{q_s}(t-\tau_{q_r1}),\cdots,y_{q_r}(t-\tau_{q_rL_{q_r}})]\,]^t \tag{2.64}$$

This yields

$$\mathbf{x}^r(t) = \mathbf{A}^r\mathbf{y}(t) \tag{2.65}$$

Let the covariance matrix $\mathbf{R}^r$ be defined by $\mathbf{R}^r = \mathcal{E}\{\mathbf{x}^r(t)(\mathbf{x}^r(t))^H\}$, and define the interferer multipath spatial correlation $\sigma_{kk'}^{ll'}$ for paths (k,l) and (k',l') by

$$(\sigma_{kk'}^{ll'})^2 = \mathcal{E}\{y_k(t-\tau_{kl})\overline{y_{k'}(t-\tau_{k'l'})}\} \tag{2.66}$$

Note that the correlations between different statistically independent interferers are zero, so that $\sigma_{kk'}^{ll'} = 0\ \forall\ k \neq k'$. Straightforward manipulation shows that the covariance matrix $\mathbf{R}^r$ can be written as

$$\mathbf{R}^r = \mathbf{A}^r\mathbf{B}^r\mathbf{A}^H \tag{2.67}$$

where

$$\mathbf{B}^r = \begin{pmatrix} (\sigma_{11}^{11})^2 & \cdots & (\sigma_{11}^{1L_1})^2 & & 0 & \cdots & 0 & \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots & \\ (\sigma_{11}^{L_11})^2 & \cdots & (\sigma_{11}^{L_1L_1})^2 & & 0 & \cdots & 0 & \\ & \vdots & & \ddots & & \vdots & & \\ 0 & \cdots & 0 & & (\sigma_{L_{q_r}L_{q_r}}^{11})^2 & \cdots & (\sigma_{L_{q_r}L_{q_r}}^{1L_{q_r}})^2 & \\ \vdots & \ddots & \vdots & \cdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & & (\sigma_{L_{q_r}L_{q_r}}^{L_{q_r}1})^2 & \cdots & (\sigma_{L_{q_r}L_{q_r}}^{L_{q_r}L_{q_r}})^2 & \end{pmatrix} \tag{2.68}$$

In this model it is assumed that there is no correlation between different interferers k. A main difference between this model and the model without multipaths for interferers is that in the multipath case not all off-diagonal elements of $\mathbf{R}^r$ are zero. The model can straightforwardly be extended to a polarisation model, but that is omitted here. The next step is to verify the structure and determine the rank of the model in case of full and partial correlation between the different interferers k.

Fully correlated interferers

Suppose there is only one interferer $y(t-\tau_l)$ with L multipaths with signature vector $\mathbf{a}_l^r$, and assume that all L multipath signals are fully correlated with observed signal power σ^2. Fully correlated means that the propagation delay differences τ_l between all multipaths and all telescopes are much smaller than the inverse bandwidth $(2\pi\Delta\nu)^{-1}$ of the frequency channel under consideration:

$$|\tau_l - \tau_{l'}| \ll \frac{1}{2\pi\Delta\nu} \tag{2.69}$$

If this condition holds, then the delay can be represented as phase shifts which can be absorbed in the array response vectors, and the signal (envelope) $y(t-\tau_l)$ can be replaced by $y(t)$. This criterion is further explained in chapter 3. Defining $\mathbf{a}_o^r = \sum_{l=1}^{L} \mathbf{a}_l^r$ yields

$$\mathbf{x}^r(t) = \sum_{l=1}^{L} (\mathbf{a}_l^r y(t)) = \left(\sum_{l=1}^{L} \mathbf{a}_l^r \right) y(t) = \mathbf{a}_o^r y(t) \tag{2.70}$$

for the array output signal vector, which means that in this case $\mathbf{R}^r = \mathcal{E}\{\mathbf{x}^r(t)$ $(\mathbf{x}^r(t))^H\} = \mathbf{a}_o^r(\mathbf{a}_o^r)^H\sigma^2$, so that this interference model has rank one. This can also be seen directly from the equation $\mathbf{R}^r = \mathbf{A}^r\mathbf{B}^r(\mathbf{A}^r)^H$, where $\mathbf{R}^r$ for fully correlated signals with covariance σ^2, can be written as $\mathbf{B}^r = \sigma^2\mathbf{1}\mathbf{1}^H$. The covariance $\mathbf{R}^r$ can then be written as $\mathbf{R}^r = \sigma^2(\mathbf{A}^r\mathbf{1})(\mathbf{A}^r\mathbf{1})^H$, which is a multiplication of two vectors, and thus has rank one.

Partially correlated interferers

Suppose that there is only one interferer with two multipaths : $y(t-\tau_1)$ and $y(t-\tau_2)$. The two paths correspond to array signature vectors $\mathbf{a}_1^r$ and $\mathbf{a}_2^r$ respectively, and the array output vector is given by

$$\mathbf{x}^r(t) = \mathbf{a}_1^r y(t-\tau_1) + \mathbf{a}_2^r y(t-\tau_2) \tag{2.71}$$

The covariance of $\mathbf{x}^r(t)$ is defined by $\mathbf{R}^r = \mathcal{E}\{\mathbf{x}^r(t)(\mathbf{x}^r(t))^H\}$, let $\sigma_{ij}^2 \equiv \mathcal{E}\{y(t-\tau_i)\overline{y(t-\tau_j)}\}$, and define $\mathbf{B}^r = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix}$, then:

$$\mathbf{R}^r = [\mathbf{a}_1\ \mathbf{a}_2] \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1^H \\ \mathbf{a}_2^H \end{bmatrix} \tag{2.72}$$

In this case, for partial correlation, i.e. $\sigma_{11}^2 = \sigma_{22}^2 = 1$, $|\sigma_{12}|^2 = |\sigma_{21}|^2 < 1$ the rank of $\mathbf{R}^r$ is two. In general:

$$\text{rank}(\mathbf{R}^r) = \min(p, \text{rank}(\mathbf{B}^r)) \tag{2.73}$$

This formula leads to table 2.1, in which an overview is given of the rank of q_s interferers, each with M_k multipaths and varying degrees of correlation. The second and third rows in the table correspond to a situation where the propagation delay differences are larger than the narrowband criterium requires. If the channel bandwidth is reduced in a system, then the rank of $\mathbf{R}^r$ reduces and the system rank tends towards the situation in the first row of the table. See chapter 3 for further details on the narrowband assumption.

2.5.2 Intermodulation product model

When linear devices, such as low-noise amplifiers in the first stages of radio astronomical receivers, are exposed to very high input signals, these devices

Multipath correlation	$\text{rank}(\mathbf{R}^r)$
All paths full correlation, no correlation between different transmitters	$\text{rank}(\mathbf{R}^r) = \min(p, q_r)$
All paths partial (arbitrary) correlations, no correlation between different transmitters	$\text{rank}(\mathbf{R}^r) = \min(p, \sum_{k=1}^{q_r} M_k)$
No correlation between paths or transmitters	$\text{rank}(\mathbf{R}^r) = \min(p, \sum_{k=1}^{q_r} M_k)$

Table 2.1. Rank of $\mathbf{R}^r$ for different degrees of multipath correlation, for $k = 1 \cdots q_r$ interferers, each with M_k multipaths .

usually show nonlinear behaviour. In some cases this is advantageous, for example for mixer devices [124], where a strong constant reference signal (local oscillator signal) is added to the receiving signal. The nonlinearity generates higher harmonics both for the received signal and for the local oscillator signal. Systems are usually designed such that the frequency difference between the local oscillator signal and the received signal fall within the pass-band of the system, usually at lower frequencies. The original signals and harmonics, and all other sum and difference frequencies (intermodulation products) are filtered out.

In some cases, interference which may initially fall outside the band of interest may be so strong that it generates nonlinearities in the receiving system. The resulting frequency of the sum or difference of the harmonics of two or more interferers may fold into the pass-band of the system, thereby causing in-band interference. The unwanted interference caused by this nonlinear effect is called an intermodulation product. Now the question is, what data model is suitable for such an interferer, and what properties the resulting interferer has. The effect of the higher-order intermodulation terms on the array response matrix is especially interesting.

Consider two interferers impinging on a telescope i with sky frequencies ν_1 and ν_2, and narrowband real signals $m_1(t)$ and $m_2(t)$ which are spectrally centred at the sky frequencies. As will be described in the section on narrowband assumptions in chapter 3, the geometric time-shift of the signal envelope $m_{1i}(t - \tau_{1i})$ and $m_{2i}(t - \tau_{2i})$ can be represented as a geometric phase shift of the interfering signal, assuming that the narrowband condition holds. At radio frequencies therefore, the real signal $x_i(t)$ which enters the first low-noise amplifier (LNA) of telescope i can be described as

$$x_i(t) = m_{1i}(t)\cos(2\pi\nu_1(t - \tau_{1i})) + m_{2i}(t)\cos(2\pi\nu_1(t - \tau_{2i})) \tag{2.74}$$

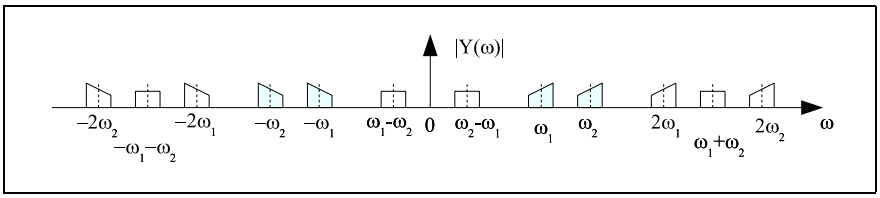


Figure 2.8. Spectrum showing two strong transmitters with angular frequencies ω_1 and ω_2. If these transmitters drive an amplifier into nonlinear operation, then higher order harmonics and difference and summing frequencies (intermodulation products) are generated. The figure shows some of the corresponding occupied frequency bands.

The terms $2\pi\nu_1\tau_{1i}$ and $2\pi\nu_2\tau_{2i}$ can also be written as phase offsets ϕ_{1i} and ϕ_{2i} respectively. Now suppose that the interferers are strong enough to drive the LNA into nonlinear behaviour. Then the output signal $y_i(t)$ of the LNA of the i^{th} telescope can be described as a power series of the input signal, with (unknown) coefficients:

$$y_i(t) = \sum_{k=0}^{N} c_k x_i(t)^k \tag{2.75}$$

Using the binomial theorem and an expansion for $cos^n(.)$ the formula for $y_i(t)$ can be expanded. In this expansion there are higher order factors of $m_{1i}(t)$ and m_{2i}^l, and harmonics of the two impinging signals. This expansion is straightforward but tedious, and is omitted here. Of all the (real) generated intermodulation products only those which fall within the band of interest are selected. Assume that there is only one such product, with orders n_1 and n_2, then $y_i(t)$ can be written as:

$$y_i(t) = \alpha_i \cos((2\pi n_1\nu_1 \pm 2\pi n_2\nu_2)t + (n_1\phi_{1i} \pm n_2\phi_{2i}))m'(t) \tag{2.76}$$

where $m'(t)$ is the higher order signal envelope product of the two interferers (non-Gaussian), and α_i is an (unknown) real scaling or gain factor. This formula was explicitly derived in [17] for a second order product. In order to obtain this expression for the entire array the following vectors are formed: $\mathbf{y} = (y_1, \cdots, y_p)^t$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p)^t$, $\boldsymbol{\phi}_1 = (\phi_{11}, \cdots, \phi 1p)^t$, and $\boldsymbol{\phi}_2 = (\phi_{21}, \cdots, \phi 2p)^t$. The array signal vector of the intermodulation product is

$$\mathbf{y}(t) = \boldsymbol{\alpha} \odot \cos[\,(2\pi n_1\nu_1 \pm 2\pi n_2\nu_2)\,t\,\mathbf{1} + (n_1\boldsymbol{\phi}_1 \pm n_2\boldsymbol{\phi}_2)\,]\,m'(t) \tag{2.77}$$

From this expression it is clear that the rank of $\mathbf{R} = \mathcal{E}\{\mathbf{y}(t)\mathbf{y}(t)^H\}$ is one. On the basis of this model, and verified experimentally using LOFAR data, in [17] it was shown that the second order intermodulation product of point sources in a sky map remain point sources. After downconversion, the array response vector takes the form

$$\mathbf{a}^r = \boldsymbol{\gamma} \odot e^{\jmath(n_1\boldsymbol{\phi}_1 \pm n_2\boldsymbol{\phi}_2)} \tag{2.78}$$

where $\boldsymbol{\gamma}$ is a scaling vector, and the phase delays $\boldsymbol{\phi}_j$ are related to the RFI source direction and array geometry by $\boldsymbol{\phi}_j = 2\pi\lambda^{-1}\mathcal{R}\mathbf{s}_j$ (as before). It is obvious (cf. the next chapter) that the earth rotation-related fringe frequency for the intermodulation product scales with the orders n_1 and n_2. This, together with the fact that the rank of the covariance matrix is one, means that intermodulation products behave as ordinary sources and can be filtered with the same methods as single in-band interferers, at least when the transmitters are spatial point sources, and stationary for the duration of the processing interval. If there are more than two interferers and/or products, the rank of the covariance matrix in general will increase.

2.6 Phased array covariance model

In this section, expressions will be derived for the covariance matrices of outputs of phased array systems. The results will be used in the chapter on spatial filtering using a phased-array reference antenna. The model described in this section will be made more general: every telescope will be considered a phased array with m_i phased-array antennas, where $m_i = 1$ for each of the WSRT telescopes. This extended model will describe the (non-polarised) next-generation telescope LOFAR , and the model will be used in the chapter on implications for next generation telescopes.

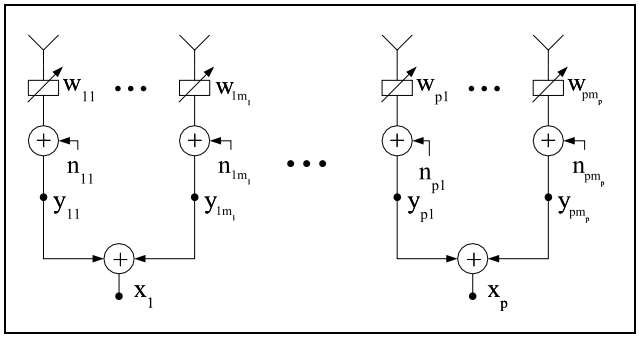


Figure 2.9. Telescope system with p telescopes, each based on subarrays consisting of m_i antenna elements. The number of antenna elements m_i in each subarray i may be different.

Subarray model.

Figure 2.9 schematically shows p telescopes, where each telescope is a subarray composed of m_i antennas, and where i is the telescope index number. Assume that the narrowband assumption holds, and that the complex signals impinging on a subarray i are multiplied by the complex adjustable weights w_{ij}, where j denotes the subarray antenna index. The adjustable weight vector $\mathbf{w}_i$ is defined by $\mathbf{w}_i = [w_{i1}, \cdots, w_{im}]^t$. Let s_k be the k^{th} impinging source,

and define $\mathbf{s} = [s_1, \cdots s_q]^t$. Further let $\mathbf{r}_{ij}$ be the location of antenna ij, then the geometric delay-phase factor a_{ijk} for the k^{th} source is defined as before: $a_{ijk} = e^{-2\pi\lambda^{-1}\mathbf{r}_{ij}^t\mathbf{s}_k}$. Let the subarray vector $\mathbf{a}_i$ for q sources be defined by $\mathbf{a}_1 = [a_{i1}, \cdots, a_{im}]^t$, and define the subarray array response matrix $\mathbf{A}_i$ by $\mathbf{A}_i = [\mathbf{a}_{i1}, \cdots, \mathbf{a}_{iq}]^t$. Furthermore, also stack the subarray antenna noise powers n_{ij} in a vector: $\mathbf{n}_i = [n_{i1}, \cdots, n_{im}]^t$. Finally, define the subarray output vector $\mathbf{y}_i$ by $\mathbf{y}_i = [y_{i1}, \cdots, y_{im}]^t$, which can also be written as $\mathbf{y}_i = \text{diag}(\mathbf{w}_i)\mathbf{A}_i\mathbf{s} + \mathbf{n}_i$. The phased array summed output x_i can be written as $x_i = \mathbf{1}_{m_i}^t\mathbf{y}_i$. Using $\mathbf{1}_{m_i}^t\text{diag}(\mathbf{w}_i) = \mathbf{w}_i^t$, this is equal to:

$$x_i = \mathbf{w}_i^t\mathbf{A}_i\mathbf{s} + \mathbf{1}_{m_i}^t\mathbf{n}_i \tag{2.79}$$

Correlation of subarrays.

The summed subarray signals defined above can be correlated with each other. Define the array output signal vector $\mathbf{x}$ by $\mathbf{x} = (x_1, \cdots x_p)^t$ as before. The overall array response matrix $\mathbf{A}$ and noise vector $\mathbf{n}$ are defined by respectively $\mathbf{A} = ((\mathbf{w}_1^t\mathbf{A}_1)^t, \cdots (\mathbf{w}_p^t\mathbf{A}_p)^t)^t$ and $\mathbf{n} = (\mathbf{1}_{m_1}^t\mathbf{n}_1, \cdots \mathbf{1}_{m_p}^t\mathbf{n}_p)^t$. Using these definitions, the output vector $\mathbf{x}$ can be expressed by

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} \tag{2.80}$$

Let $\mathbf{G}$ be the matrix describing the electronic gains g_i after the subarray summation: $\mathbf{G} = \text{diag}\,[(g_1, \cdots, g_p)^t]$. Define the source coherency matrix by $\mathbf{B}_s = \mathcal{E}\{\mathbf{s}\mathbf{s}^H\}$, and define the noise matrix $\mathbf{D}$ by $\mathbf{D} = \mathcal{E}\{\mathbf{n}\mathbf{n}^H\}$, then the subarray's output covariance matrix $\mathbf{R} = \mathbf{G}\mathcal{E}\{\mathbf{x}\mathbf{x}^H\}\mathbf{G}^H$ can be written as

$$\mathbf{R} = \mathbf{G}\mathbf{A}\mathbf{B}_s\mathbf{A}^H\mathbf{G}^H + \mathbf{G}\mathbf{D}\mathbf{G}^H \tag{2.81}$$

The elements D_{ij} of the noise matrix $\mathbf{D}$ are given by $D_{ij} = \mathbf{1}_{m_i}^t\mathcal{E}\{\mathbf{n}_i\mathbf{n}_j^H\}\mathbf{1}_{m_j}$. This means that if the system noise powers among the subarrays are not correlated, the matrix D is diagonal: $\mathbf{D} = \text{diag}(D_{11}, \cdots, D_{pp})$.

2.7 Concluding remarks

In this chapter compact, consistent, discrete source algebraic models have been presented. The models are narrow-band representations for array output vectors and covariance matrices, applied to (polarised or scalar) astronomical sources and man-made radio interference.

The models are limited to radio interferometry; an interesting future extension would be to include the correlation of telescopes which are composed of phased arrays with beamforming capabilities. Another interesting extension would be to model focal plane arrays, which are beamforming arrays at the focus of parabolic dishes. A minor extension of the model is the addition of a small reference antenna; this will be described in the chapter on spatial filtering.

Both astronomical sources and interferer sources have been assumed to be at least quasi-stationary. The astronomical sources have been assumed to be mutually uncorrelated. In some limited cases, the astronomical sources may be correlated, and therefore the model should be adapted for those cases. The coherent RFI model (RFI model with fully correlated multipath) could be used as a starting point.

The models presented are narrow-band, but for the new generation of radio telescopes a broad-band approach could be fruitful as well. Broad-band modelling and broadband RFI mitigation are more complicated than narrowband work, but will be worthwhile.

Chapter 3

Data analysis

The topic of this chapter is the data analysis of experimental data. The chapter will start with an overview of assumptions on which the models described in chapter 2 are based. It will be followed by a description of the data acquisition and subsequent data processing steps, starting from the reception of radio waves at the antenna elements to digitisation of the telescope baseband signals up to the creation of covariance matrices. In the section after that, a short description will be given of the covariance matrix subspace estimation techniques used in this thesis. These techniques will be described in some detail as they will be used in chapters 5, 6 and 7 for interference mitigation. The chapter concludes with a description of the influence of the bandwidth and of the spatial extendedness of a source on the covariance matrix subspace structure. As this structure is used explicitly for interference mitigation, the influence of deviations from the narrow-band criterion and from the point-source assumption on the subspace structure needs to estimated.

3.1 Experimental considerations

3.1.1 Assumptions

Statistical independence and stationarity

Although some cosmic source signals are non-stationary, such as pulsars, Jovian radio emissions or supernovae, most cosmic signals are stationary due to the underlying stochastic physical processes. The analysis in this thesis will be limited to the class of signals which are stationary for the duration of an astronomical observation. It is assumed that the cosmic source signals are temporally independent and identically distributed (i.d.d.) Gaussian random wide-band signals, which may contain narrow-band or wide-band emission and absorption lines. The cosmic source signals are also assumed to be mutually independent (temporally and spatially white). The processes are also assumed to be ergodic, so that ensemble parameters can be estimated by their corresponding time av-

erages [112]. For interfering signals quasi stationarity is assumed, which means that the signal statistical parameters may vary over long time intervals, whereas over short observation intervals (usually of order 10 ms) the statistical parameters are constant.

Narrowband condition

Cosmic signals usually span the entire spectral band under consideration or at least several frequency bins in the case of spectral line studies. For convenient modelling it is assumed that the narrowband condition holds and that both the cosmic source signals and the interference signals can be described as a sum of narrowband or quasi-monochromatic signals. Several narrowband conditions can be defined, of which two are used throughout this thesis. The first condition states that the source signals are band-limited (either by their nature or by sub-band filtering) to a frequency range $\Delta\nu$ centred at a frequency ν_c, with the property $\Delta\nu \ll \nu_c$ [116]. In such a case, a complex signal $y(t)$ can be described as a product of a complex band-limited baseband signal $m(t)$ and a monochromatic signal $e^{2\pi\jmath\nu_c t}$ with frequency ν_c :

$$y(t) = m(t)e^{2\pi\jmath\nu_c t} \tag{3.1}$$

The signal $m(t)$ is also known as the complex envelope of the real-valued bandpass signal $z(t) = \mathrm{Re}(y(t))$. The second narrowband assumption [167] [68] is an extension of the first, and it is said to hold when the propagation delay differences $\tau_{ij} = \tau_i - \tau_j$ between the array elements i and j are small relative to the inverse bandwidth $\Delta\nu^{-1}$of the signal $y(t)$:

$$\Delta\nu \ll (2\pi\tau_{ij})^{-1} \tag{3.2}$$

When this inequality holds, time delays across the telescope array can be approximated by phase shifts of the baseband signal, as will be explained next.

The signal $z(t) = \mathrm{Re}\{m(t)e^{2\pi\jmath\nu_c t}\}$ as was defined above can also be written as

$$z(t) = \mathrm{Re}\{m(t)\}\cos(2\pi\nu_c t) - \mathrm{Im}\{m(t)\}\sin(2\pi\nu_c t) \tag{3.3}$$

The real and imaginary parts of $m(t)$ are called the in-phase and quadrature components of $z(t)$ and are obtained in practical systems by multiplying the received signals with $\cos(2\pi\nu_c t)$ and $\sin(2\pi\nu_c t)$ followed by low-pass filtering. Now suppose that the bandpass signal $z(t)$ is delayed in time by an amount τ. This can be written as

$$z(t-\tau) = \mathrm{Re}\{m(t-\tau)e^{-2\pi\jmath\nu_c\tau}e^{2\pi\jmath\nu_c t}\} \tag{3.4}$$

which implies that the complex envelope $m_\tau(t)$ of the delayed signal $z(t-\tau)$ is $m_\tau(t) = m(t-\tau)e^{-2\pi\jmath\nu_c\tau}$. Let $\mathcal{M}(\nu)$ be the Fourier transform of the complex envelope $m(t)$, and let $\Delta\nu$ be the bandwidth of the complex envelope. The complex envelope signals can then be expressed as inverse Fourier transforms,

using the time-shift theorem for the time-delayed version

$$m(t) = \int_{-\frac{1}{2}\Delta\nu}^{\frac{1}{2}\Delta\nu} \mathcal{M}(\nu) e^{2\pi \jmath \nu t} d\nu \tag{3.5}$$

$$m(t-\tau) = \int_{-\frac{1}{2}\Delta\nu}^{\frac{1}{2}\Delta\nu} \mathcal{M}(\nu) e^{-2\pi \jmath \nu \tau} e^{2\pi \jmath \nu t} d\nu \tag{3.6}$$

If condition (3.2) is satisfied then both integrals are approximately the same and $m(t) \approx m(t-\tau)$, hence

$$m_\tau(t) \approx m(t) e^{-2\pi \jmath \nu_c \tau} \tag{3.7}$$

which means that if the narrowband condition holds, a time delay of the band-pass signal $z(t)$ can be represented by a phase shift $e^{2\pi \jmath \nu_c t}$. Note that this criterion is independent of frequency.

Suppose that the telescope gains $\mathcal{A}$ are the same for all telescopes and assume further that the telescope signal delay offsets τ_i are determined by a source in direction $\mathbf{s}_\ell$ and a telescope array geometry matrix $\mathcal{R}$ (cf. equation (2.19). Then stacking all telescope signals $y_i(t)$ in a vector $\mathbf{y}(t) = [y_1(t), \cdots, y_p(t)]^t$, and realising that $c[\tau_1, \cdots, \tau_p]^t = \mathcal{R}\mathbf{s}_\ell$ yields

$$\mathbf{y}(t) = \mathcal{A} e^{-2\pi \nu_c c^{-1} \mathcal{R}\mathbf{s}_\ell} \, m(t) \, e^{2\pi \nu_c t} \tag{3.8}$$

This equation shows how the sky-signal time delays are converted to complex phase offsets for narrowband time signals. The first and second factor in this equation are consistent with the definitions in equation (2.20).

As an example the narrow-band condition for the WSRT at 21 cm wavelength is estimated. The WSRT has a maximum baseline length b_{max} of 3 km, and therefore has a maximum geometric time delay of $\tau = b_{max}/c = 10\ \mu s$, where c is the speed of light. This corresponds to the narrow-band condition $\Delta\nu \ll 16\ kHz$.

Near-field and far-field assumptions

It is assumed that the individual-telescope far-field condition for astronomical sources holds [87, 124]. This means that the sources are at a distance at which an increase in the source distance does not lead to a change in the (apparent) telescope beam shape. In other words, the emitted spherical wavefronts at the telescope antenna can be considered flat: the relative phases at different locations within the antenna aperture should be much less than 2π. Let D_{nt} be the telescope aperture diameter, r be the distance between telescope and source, and let λ be the signal wavelength, then the telescope far field criterion is given by [124]:

$$r > \frac{2D_{nt}^2}{\lambda} \tag{3.9}$$

For example, a cosmic source emitting at 1 GHz and entering via the telescope mainbeam can be considered in the far field of a single WSRT telescope dish

(diameter 25 m) at a distance $d > 4$ km. This condition is valid and useful only for the cosmic sources. Interferers are different in the sense that they mostly enter the telescope via the sidelobes, not the main beam. In that case, the signal will reach the antenna via a sidelobe of the antenna element itself and possibly via multiple reflections at the dish. The signals will not arrive within 2π phase differences whatever the distance.

An additional aperture array far-field requirement for emitting sources is that the spherical wave fronts impinging on an aperture array have curvatures corresponding to phase differences much less than 2π over the array. Let D_{ap} be the diameter of the aperture array, then straightforward geometry leads to the aperture array far-field condition:

$$r \gg \frac{D_{ap}^2}{\lambda} \tag{3.10}$$

The far field condition for the WSRT array at 1 GHz for example is $r \gg 30\,000$ km, a distance close to the geostationary orbits. Interference sources for the WSRT will in most cases be in the near-field of the aperture array. For the LOFAR ITS test station however, with its 200 m diameter and at a frequency of 30 MHz, the array far-field condition is reached when $r \gg 4$ km. This means that for the WSRT in most cases the array response vector in (2.38) must be used, and that in most cases both response vector representations (2.20) and (2.38) can be used for LOFAR ITS. The later assumes that the interferers are at distances $r \gg 4$ km.

Aperture plane sampling

In the modelling to be presented, the aperture array is in principle assumed to be a three-dimensional, non-coplanar array with an arbitrary configuration. Often the dimension is removed in the direction of the astronomical sources to be observed by compensating for the geometric path delay differences between telescopes. This can be done by modelling time delays as phase shifts if the narrowband condition holds, as is done in this thesis. Alternatively the path delays can be corrected by applying time delays, or by a combination of time delays and phase corrections, as is done, for example, in the WSRT telescope systems [117]. It is important to note that, although the WSRT is an East-West array and is used in this thesis for testing the RFI mitigation algorithms, the general models to be presented do not assume the telescope array to be a spatial linear array.

Mutual coupling and multipath effects

Methods exist to incorporate mutual coupling effects [56,103,143] in (phased) array signal processing models. In most radio telescope systems based on large parabolic dishes, however, there is no significant antenna coupling between telescopes because the relative telescope distances $\mathbf{r}_i - \mathbf{r}_j$ are large ($|\mathbf{r}_i - \mathbf{r}_j| \gg \lambda$), and the telescope antenna sidelobe levels are low, usually at levels close to those of an isotropic radiator [87, 108, 126].

In this thesis it is assumed that there are no multipaths for astronomical sources, or that they are at a level where they cannot be detected. In a worst-case situation of a flat and perfectly conductive ground plane under a telescope, the sky sources would be reflected in this "mirror". The reflected astronomical signals would enter the telescope via the sidelobes , and would be attenuated with respect to the main beam with the antenna gain: $G = \frac{4\pi}{\lambda^2} A_{eff}$, where A_{eff} is the effective telescope receiving area. For a 25-m diameter WSRT dish at 21 cm wavelengths this yields approximately $G = 50$ dB. In practice the ground plane is not flat, nor perfectly conductive, and the multipath phase changes erratically due to the telescope rotations, leading to an additional large attenuation factor. The astronomical multipath can therefore be neglected.

For interferers there may be multipaths, but the models to be presented can take this into account; see, for instance, section 2.5.

Field of view

In principle, the aperture synthesis is a wide field-of-view technique. Both the continuous and discrete source and interferer models are wide field-of-view models, up to (at least) 2π Sr solid angle. However, in signal and image processing, there may be limiting assumptions. In the Fourier imaging method, for example, an approximation is often used to simplify the calculations [124], which limits the imaging to a narrow field of view.

3.1.2 Obtaining spatial coherence data

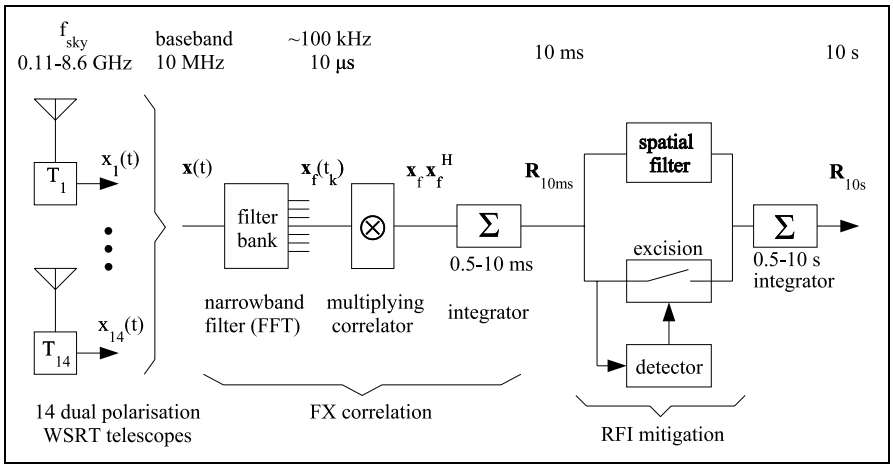


Figure 3.1. Schematic overview of the data processing steps in obtaining coherency data at the WSRT for interference mitigation tests.

Before describing the experimental set-ups used at the WSRT for obtaining

experimental data in the next chapter, this section will describe the way in which coherency data were obtained at the WSRT. It will also describe how telescope delay offset affects the array response vectors. Figure 3.1 shows a schematic picture of the data processing steps which were used to obtain coherency data at the WSRT for interference mitigation tests. The WSRT telescopes receive astronomical radio signals at high RF (sky) frequencies (110 MHz - 8.6 GHz), which are down-converted to baseband (0-10 MHz). Each of the telescope signals $x_i(t)$ is digitised, and after narrow-band filtering, the telescope signals of each of the narrow-band frequency channels $x_i(t)$ are stacked in a vector

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_p(t) \end{bmatrix} \tag{3.11}$$

were p is the number of telescopes. For each frequency channel all telescopes are cross-correlated and averaged, yielding a $p \times p$ correlation or covariance matrix $\widehat{\mathbf{R}}_{10ms}$.

$$\mathbf{R}(t) = \mathcal{E}\{\mathbf{x}(t)\mathbf{x}(t)^H\} \tag{3.12}$$

$$\widehat{\mathbf{R}}_{10ms}(t) = \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{x}(t+kT)\mathbf{x}(t+kT)^H \tag{3.13}$$

The true covariance matrix is $\mathbf{R}(t)$, its estimate at $NT = 10$ ms scales is given by $\widehat{\mathbf{R}}$. At this time scale the interferers are assumed to be stationary. The astronomical sources are assumed to be stationary throughout the entire observation (which may take several hours). After this correlation and averaging stage, there is an optional interference filtering stage. It is followed by a second integrator, averaging the (filtered) covariance matrices further to 0.5-10 s. These 10 s covariance matrix elements correspond to sampling points in the synthesised aperture plane; they contain the desired coherencies , which can be used for imaging. Usually radio astronomical observations take several hours, thus increasing the number of 10 s aperture samples, which leads to an increase in the signal-to-noise ratio in the sky map. Due to earth rotation this also leads to a denser sampling of the synthesised aperture plane. The interferers may or may not be stationary at 10 s scales. A correlator scheme in which a subband filtering stage is applied before correlation is called FX correlation. An alternative scenario would be correlation at several time lags before subband filtering (Fourier transform), and is known as XF correlation. The WSRT correlator , for example, is of the XF type. Actual receiving systems and correlators contain many signal processing details which are not discussed here; a more detailed description can be found, for example, in [113, 146, 124, 117, 87]. However, one of the details which is relevant for this thesis is fringe rotation or natural fringe which is explained next.

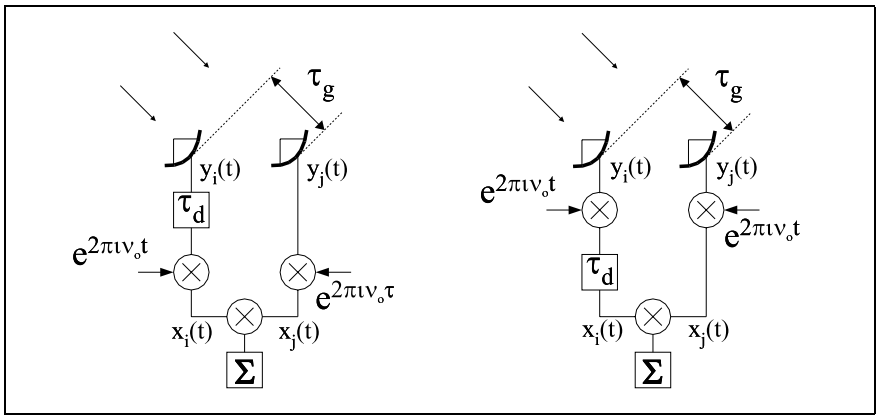


Figure 3.2. Signal down-conversion after applying geometric delay compensation (left) and before applying delay compensation(right).

Geometric delay compensation and natural fringe

Let the received signal of telescope i at sky frequency ν_c be given by $y_i(t)$ and the signal of telescope j by $y_j(t)$. Assume that the signal from an astronomical source at telescope j has a geometric delay offset τ_g. This delay needs to be compensated for in order to avoid signal decorrelation for bandwidths not satisfying the narrowband condition: the WSRT standard operational mode includes geometric delay compensation. This situation is drawn in figure 3.2, where two different scenarios are depicted for geometric delay compensation. In one scenario (left figure) the delay compensation with τ_d is applied before the frequency down-conversion stage, and in the second scenario (right) the compensation is applied after the down-conversion stage (right). For both scenarios the signals at the antenna outputs $y_i(t)$ and $y_j(t)$ of telescopes i and j can be written as

$$y_i(t) = \mathcal{A}(\mathbf{s}_o)\, m(t) e^{\jmath 2\pi\nu_c t} \tag{3.14}$$

$$y_j(t) = \mathcal{A}(\mathbf{s}_o)\, m(t-\tau_g) e^{\jmath 2\pi\nu_c (t-\tau_g)} \tag{3.15}$$

where τ_g is the geometric time delay difference between telescopes i and j, and where $\mathcal{A}(\mathbf{s}_o)$ is the telescope gain in the source direction $\mathbf{s}_o$. Let the down-conversion process be represented by multiplication with $e^{-2\pi\jmath\nu_o t}$, where ν_o is the local oscillator frequency of the mixer [124]: $x_i(t) = y_i(t)e^{-2\pi\jmath\nu_o t}$, where $x_i(t)$ is the baseband signal. If the geometric delay compensation $\tau_d = \tau_g$ of telescope i is applied before the first down-conversion stage, then it is easily seen that

$$\mathbf{R}_{ij} = \mathcal{E}\{x_i(t)\overline{x_j(t)}\} = |\mathcal{A}(\mathbf{s}_o)|^2 \mathcal{E}\{m(t)\overline{m(t)}\} \tag{3.16}$$

However, in most cases the delay is applied after (several) mixer stages. If the signal is down-converted after one mixer stage with local oscillator frequency

ν_o, then the output signals are equal to

$$x_i(t) = \mathcal{A}(\mathbf{s}_o)\, m(t-\tau_d)\, e^{\jmath 2\pi(\nu_c-\nu_o)t}\, e^{-\jmath 2\pi(\nu_c-\nu_o)\tau_d} \quad (3.17)$$

$$x_j(t) = \mathcal{A}(\mathbf{s}_o)\, m(t-\tau_g)\, e^{\jmath 2\pi(\nu_c-\nu_o)t}\, e^{-\jmath 2\pi\nu_c\tau_g} \quad (3.18)$$

and the covariance matrix elements $\mathbf{R}_{ij}(t)$ for $\tau_d = \tau_g$ is given by

$$\mathbf{R}_{ij}(t) = |\mathcal{A}(\mathbf{s}_o)|^2 \mathcal{E}\{m(t)\overline{m(t)}\}\, e^{\jmath 2\pi\nu_o\tau_d} \quad (3.19)$$

The slowly varying phase $2\pi\nu_o\tau_d$ is an instrumental effect and is called the natural fringe. In radio telescopes, this effect is usually cancelled by applying a signal with opposite sign (phase) to the local oscillator signal. In order to derive a fringe-array vector, the delayed time signal for telescope i can be expressed more generally. Let the i^{th} telescope output signal for an astronomical source be denoted by $x_i(t)$ and for an interferer by $x_i^r(t)$. Further let the interferer baseband (envelope) signal be denoted by $m^r(t)$, and the interferer geometric time delay by $\tau_{g_i}^r$. Straightforwardly applying the delay and down-conversion steps to the time signals yields, for the i^{th} telescope:

$$\begin{aligned} x_i(t) &= \mathcal{A}(\mathbf{s}_o)\, m(t-(\tau_{d_i}+\tau_{g_i}))\, e^{2\pi\jmath(\nu_c-\nu_o)t}\, e^{-2\pi\jmath(\nu_c(\tau_{d_i}+\tau_{g_i})-\nu_o\tau_{d_i})} \\ x_i^r(t) &= m^r(t-(\tau_{d_i}+\tau_{g_i^r}))\, e^{2\pi\jmath(\nu_c-\nu_o)t}\, e^{-2\pi\jmath(\nu_c(\tau_{d_i}+\tau_{g_i^r})-\nu_o\tau_{d_i})} \end{aligned} \quad (3.20)$$

The sidelobe gain $\mathcal{A}^r$for the interfering signal can be absorbed in $m^r(t)$ without loss of generality. It is further assumed that the narrow-band condition holds, which means (cf. section 3.1.1) that $m(t-(\tau_{d_i}+\tau_{g_i})) = m(t)\ \forall i$ and that $m(t-(\tau_{d_i^r}+\tau_{g_i})) = m(t)\ \forall i$. It is also assumed that $\nu_o = \nu_c$. In order to compensate for the geometric delays τ_{g_i} for all telescopes, τ_{d_i} is chosen such that $\tau_{d_i}+\tau_{g_i}$ is constant, or $\tau_0 = \tau_{d_i}+\tau_{g_i}\ \forall i$. This leads to the simplified expressions

$$\begin{aligned} x_i(t) &= \mathcal{A}(\mathbf{s}_o)\, m(t)\, e^{-2\pi\jmath\nu_o((\tau_o-\tau_{d_i}))} \\ x_i^r(t) &= m^r(t)\, e^{-2\pi\jmath\nu_o((\tau_o-\tau_{d_i})+(\tau_{g_i}^r-\tau_{g_i}))} \end{aligned} \quad (3.21)$$

The geometric delays τ_{g_i} for a source $\mathbf{s}_o$ can be expressed in terms of $\mathcal{R}\mathbf{s}_o$ by: $[\tau_{g_1},\cdots,\tau_{g_p}]^t = \mathcal{R}\mathbf{s}_o c^{-1}$. Now stack the signals $x_i(t)$ and $x_i^r(t)$ in the vectors: $\mathbf{x}(t) = [x_1(t),\cdots,x_p(t)]^t$ and $\mathbf{x}^r(t) = [x_1^r(t),\cdots,x_p^r(t)]^t$, and define $\mathbf{x}(t) = \mathbf{a}(\mathbf{s}_o)m(t)$ and $\mathbf{x}^r(t) = \mathbf{a}^r(\mathbf{s}_o)m^r(t)$. Further let the natural fringe vector $\mathbf{a}^f(\mathbf{s}_o)$ and the interferer fringe vector $\mathbf{a}^{fr}(\mathbf{s}_o)$ be defined by

$$\mathbf{a}^f(\mathbf{s}_o) = \begin{bmatrix} e^{2\pi\jmath\nu_o\tau_{d_1}} \\ \vdots \\ e^{2\pi\jmath\nu_o\tau_{d_p}} \end{bmatrix}, \quad \mathbf{a}^{fr}(\mathbf{s}_o) = \begin{bmatrix} e^{-2\pi\jmath\nu_o(\tau_{g_1}^r-\tau_{g_1})} \\ \vdots \\ e^{-2\pi\jmath\nu_o(\tau_{g_p}^r-\tau_{g_p})} \end{bmatrix} \quad (3.22)$$

Then the array response vector $\mathbf{a}(\mathbf{s}_o)$ corresponding to the direction $\mathbf{s}_o$ being "tracked", and the resulting interferer response vector $\mathbf{a}^r(\mathbf{s}_o)$ can be written as

$$\mathbf{a}(\mathbf{s}_o) = \mathcal{A}(\mathbf{s}_o)\, e^{-2\pi\jmath\nu_o\tau_0}\mathbf{1} \odot \mathbf{a}^f(\mathbf{s}_o) \quad (3.23)$$

$$\mathbf{a}^r(\mathbf{s}_o) = \boldsymbol{\gamma} \odot^{\jmath\phi} \odot e^{-2\pi\jmath\nu_o\tau_0}\mathbf{1} \odot \mathbf{a}^{fr} \odot \mathbf{a}^f(\mathbf{s}_o) \quad (3.24)$$

Fringe compensation is obtained by (element-wise) multiplying both expressions with $\overline{\mathbf{a}^f(\mathbf{s}_o)}$. This "stops" the astronomical fringe $\mathbf{a}^f(\mathbf{s}_o)$ in (3.23)and (3.24), as $\mathbf{a}^f(\mathbf{s}_o)\overline{\mathbf{a}^f(\mathbf{s}_o)} = \mathbf{1}$ but not the "interferer" fringe $\mathbf{a}^{fr}$ in (3.24). Note that for a source in the direction $\mathbf{s}_\ell$, with "fringe-stopping" carried out at the field-of-view centre $\mathbf{s}_o$, equation (3.23) takes the form

$$\mathbf{a}(\mathbf{s}_\ell) = \mathcal{A}(\mathbf{s}_\ell)\, e^{2\pi\jmath\nu_o c^{-1}\mathcal{R}(\mathbf{s}_\ell - \mathbf{s}_o)} \odot e^{-2\pi\jmath\nu_c\tau_0}\mathbf{1} \tag{3.25}$$

To illustrate the different signal processing steps involved in the process of measuring coherencies and to illustrate the different time scales at which they operate, the signal waveforms of observed television stereo sound carrier waves (TV Lingen, channel 59) superimposed onto the signal of the astronomical source 3C48 are given in figure 3.3. The observed astronomical source power is 13 dB lower than the system noise and the observed sound carrier power is about 15 dB higher than the system noise, which means that the sound carriers are dominant. Clearly visible is the beating of the two sound carrier waves. Figure (a) shows what the electric field of the signal impinging on the telescope would look like. Figure (b) shows the time behaviour of the down-converted and digitised signal, as observed with one of the WSRT telescopes. Comparison of figure (a) and (b) shows that down-conversion reduces the signal processing load by a factor 1000. [1] After baseband conversion and digitisation, the next step is Fourier transforming the data. Figures (c1) and (c2) show the resulting signal power and signal phase of one telescope for a single-frequency bin centred at one of the two sound carriers. The observed signal power in figure (c1) is roughly constant, which is to be expected as most of the modulated signal power of the audio signal falls within the 78 kHz band. The signal phase changes rapidly, as is also expected. The next signal processing step is to cross-correlate the telescope signals. Figures (d1) and (d2) show the magnitude and phase for eight interferometer pairs: one autocorrelation (baseline 0m) and seven cross-correlations with baselines up to 1008 m. The fast phase-drift in figure (c2) disappears because an interferometer measures relative phases between telescopes, not the absolute phase of the impinging wave. The phase slopes as a function of time in figure (d2) are caused by the instrumental phase correction (primary fringe correction) which was described in the previous sections. The dashed lines show the expected baseline-dependent phase slopes. The fluctuation of observed signal power and phase are caused by several effects, such as non-stationarity of the sound carriers and multipath effects. The correlation signals in figures (d1) and (d2) are the observed coherencies , which are used in the imaging process.

[1] The signal in (a) is not actually measured; it is the up-converted (frequency shifted) version of the received signal in (b).

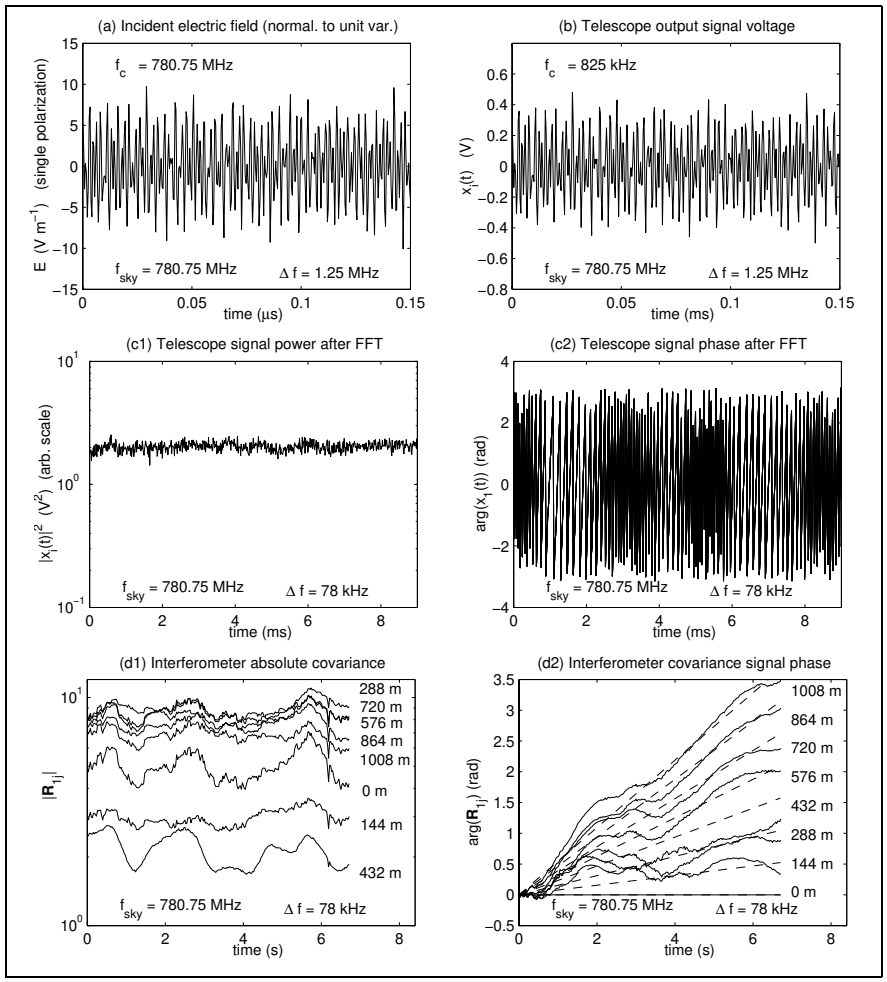


Figure 3.3. Signal waveforms of the stereo sound carrier waves of the television transmitter TV Lingen (channel 59) present in observations of the astronomical source 3C48 (telescope pointing at HA=57^o, $\delta = 33^o$) as observed with the WSRT at different stages in the signal processing chain. More information is given in section 3.1.2.

3.2 Subspace analysis

The RFI mitigation algorithms described in this thesis are based on the structure of telescope output covariance matrices. The noise, astronomical signal properties, and interferer characteristics each create different structures in the covariance matrices. This difference in structure can be used for separation (and mitigation) of the interfering signals. The following two subsections briefly de-

scribe two linear algebra approaches to characterise the eigen-structure of the covariance matrices: eigenvalue decomposition and factor analysis decomposition. The subsequent subsection describes the estimation variance of some of the eigen-structure parameters. Finally, an eigenvalue decomposition is described for a noiseless rank two model, which is subsequently used to estimate the effect of bandwidth and extendedness of sources on the covariance matrix eigen-structure. The influence of bandwidth and source extendedness on the eigen-structure is relevant as the effect of interference mitigation depends on it, as will be explained in chapters 5-7.

3.2.1 Eigenvalue decomposition

In this section a brief introduction is given of the subspace estimation by using eigenvalue decomposition; a more detailed analysis can be found for example in [88, 54]. The model under consideration is (2.40)

$$\mathbf{R} = \mathbf{G}(\mathbf{A}\mathbf{B}\mathbf{A}^H + \mathbf{A}_r\mathbf{B}_r\mathbf{A}_r^H)\mathbf{G}^H + \mathbf{D} \tag{3.26}$$

in which the diagonal matrix $\mathbf{B}$ contains the astronomical signal powers σ_i^2 and the matrix $\mathbf{B}_r$ the interferer signal powers σ_{ri}^2:

$$\begin{aligned} \mathbf{B} &= \text{diag}([\sigma_1^2, \cdots, \sigma_{q_s}^2])^t \\ \mathbf{B}_r &= \text{diag}([\sigma_{r1}^2, \cdots, \sigma_{rq_r}^2])^t \end{aligned} \tag{3.27}$$

The size of $\mathbf{B}$ is $q_s \times q_s$ assuming q_s distinct sources, and the size of $\mathbf{B}_r$ is $q_r \times q_r$ assuming q_r interferers. The size of $\mathbf{R}$ and $\mathbf{D}$ is $p \times p$ as there are p telescopes; the sizes of the array response matrices $\mathbf{A}$ and $\mathbf{A}_r$ are $p \times q$ and $p \times q_r$ respectively. Now write the model (3.26) in terms of astronomical source covariance $\mathbf{R}_s = \mathbf{G}\mathbf{A}\mathbf{B}\mathbf{A}^H\mathbf{G}^H$, and interferer covariance $\mathbf{R}_r = \mathbf{G}_r\mathbf{A}_r\mathbf{B}_r\mathbf{A}_r^H\mathbf{G}_r^H$:

$$\mathbf{R} = \mathbf{R}_s + \mathbf{R}_r + \mathbf{D} \tag{3.28}$$

The matrix $\mathbf{D}$ is a full-rank matrix, and as a consequence $\mathbf{R}$ is too, even in the absence of astronomical and interferers. In order to analyse the subspace structure due to the astronomical sources and interferers, the model without noise is considered first. The matrix $\mathbf{R}_s + \mathbf{R}_r$ is also full rank as the number of astronomical sources is much larger than the array dimension $q_s \gg p$. If there are q_r interferers with received signal powers much larger than the astronomical source signals, $\sigma_{ri}^2 \gg \sigma_j^2 \; \forall(i,j)$, then the astronomical source contribution may be neglected. Assume for the moment that the complex gains $\mathbf{G}$ are identity, $\mathbf{G} = \mathbf{I}$, or that they are calibrated and "absorbed" in $\mathbf{A}$ and $\mathbf{A}_r$. This simplifies the model in (3.26) to

$$\mathbf{R} = \mathbf{A}_r\mathbf{B}_r\mathbf{A}_r^H \tag{3.29}$$

The rank of $\mathbf{R}$ is $\min(p, q_r)$, and the number of narrowband and mutually uncorrelated interferers q_r can be estimated by means of an eigenvalue decomposition, assuming $q_r < p$. Let

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \tag{3.30}$$

be an eigenvalue decomposition of $\mathbf{R}$, where the $p \times p$ unitary matrix $\mathbf{U}$ ($\mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}$) contains the eigenvectors $\mathbf{u}_i$, $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_p]$, and the diagonal matrix $\mathbf{\Lambda}$ contains the corresponding eigenvalues λ_i, $\mathbf{\Lambda} = \mathrm{diag}([\lambda_1, \cdots, \lambda_p]^t)$. The eigenvalues are sorted in decreasing order: $\lambda_1 \geq \lambda_2 \geq, \cdots, \lambda_p \geq 0$. As the rank of $\mathbf{R}$ is q_r there are q_r nonzero eigenvalues and $p - q_r$ zero eigenvalues. The nonzero eigenvalues are stacked in a diagonal matrix $\mathbf{\Lambda}_r$, and the corresponding eigenvectors in a tall $p \times q_r$ signal subspace matrix $\mathbf{U}_r$. The remaining eigenvectors are stacked in a $p \times (p - q_r)$ "noise" subspace matrix $\mathbf{U}_n$. As $\mathbf{U} = [\mathbf{U}_r \mathbf{U}_n]$ is unitary, the two subspaces are orthogonal, $\mathbf{U}_r^H \mathbf{U}_n = \mathbf{0}$. In the noise-free case the model can be expressed in terms of subspaces as

$$\mathbf{R} = [\mathbf{U}_r \mathbf{U}_n] \begin{bmatrix} \mathbf{\Lambda}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U}_r^H \\ \mathbf{U}_n^H \end{bmatrix} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^H \tag{3.31}$$

where the zero matrices have sizes compatible with the subspace dimensions defined above.

The next step is to include noise in the model:

$$\mathbf{R} = \mathbf{A}_r \mathbf{B}_r \mathbf{A}_r^H + \mathbf{D} \tag{3.32}$$

If the noise matrix contains unequal elements on the diagonal, then it distorts the eigenvalue decomposition solution. For example, the eigenvectors obtained in the noise-free case differ from the ones obtained in the noisy case unless $\mathbf{D}$ contains identical noise powers. This is explained next.

Let $\mathbf{u}_i$ be the i^{th} eigenvalue of $\mathbf{R}_r$, then $\mathbf{R}_r\,\mathbf{u}_i = \lambda_i\,\mathbf{u}_i$. The eigenvectors of $\mathbf{R}_r$ are also eigenvectors of the identity matrix, $c\,\mathbf{I}\,\mathbf{u}_i = c\,\mathbf{u}_i$, where c is an arbitrary complex constant. Adding the two equations yields:

$$(\mathbf{R}_r + c\mathbf{I})\mathbf{u}_i = (\lambda_i + c)\mathbf{u}_i = \lambda_i'\mathbf{u}_i \tag{3.33}$$

which means that the eigenvectors in the eigenvalue decomposition of $\mathbf{R}_r$ are not changed by adding $c\mathbf{I}$. The eigenvalues are raised with a constant c. In general this is not the case for adding arbitrary diagonal matrices. If the noise is i.i.d. distributed and spatially white with covariance matrix $\sigma_n^2\,\mathbf{I}_p$, then the covariance matrix $\mathbf{R} = \mathbf{R}_r + \sigma_n^2\mathbf{I}$ can be decomposed as follows:

$$\begin{aligned} \mathbf{R} &= \mathbf{A}_r\,\mathbf{B}_r\,\mathbf{A}_r + \sigma_n^2\mathbf{I}_p \\ &= \mathbf{U}_s\,\mathbf{\Lambda}_r\,\mathbf{U}_r^H + \sigma_n^2\,(\mathbf{U}_r\mathbf{U}_r^H + \mathbf{U}_n\mathbf{U}_n^H) \\ &= [\mathbf{U}_r\mathbf{U}_n] \begin{bmatrix} \mathbf{\Lambda}_r + \sigma_n^2\mathbf{I}_{q_r} & 0 \\ 0 & \sigma_n^2\mathbf{I}_{p-q_r} \end{bmatrix} \begin{bmatrix} \mathbf{U}_r^H \\ \mathbf{U}_n^H \end{bmatrix} \\ &= \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^H \end{aligned} \tag{3.34}$$

A physical interpretation of the eigenvalue decomposition can be as follows. The eigenvectors give an orthogonal set of " directions " (spatial signatures)[2] present

[2]Here direction is not to be interpreted as the physical direction-of-incidence $\mathbf{s}_r$ of the interferer, but rather the abstract direction of a unit-norm vector in the vector space $\mathbf{C}^p$, cf. definition of $\mathbf{a}_r$ in (2.20).

in the covariance matrix, sorted in decreasing order of dominance. The eigenvalues give the power of the signal coming from the corresponding directions, or the power of the output of a beamformer matched to that direction. Indeed, let the i^{th} eigenvector be $\mathbf{u}_i$ then this output power will be

$$\mathbf{u}_i^H \mathbf{R} \mathbf{u}_i = \lambda_i. \tag{3.35}$$

The first eigenvector, $\mathbf{u}_1$ always points in the direction from which most energy comes. The second one, $\mathbf{u}_2$, points in a direction orthogonal to $\mathbf{u}_1$ from which most of the remaining energy comes, etcetera.

If there is no interference and only noise, then there is no dominant direction, and all eigenvalues are equal to the noise power. If there is a single interferer with power σ_{r1}^2 and spatial signature $\mathbf{a}_{r1}$, normalised to $||\mathbf{a}_{r1}||^2 = p$, then the covariance matrix is $\mathbf{R} = \sigma_{r1}^2 \mathbf{a}_{r1} \mathbf{a}_{r1}^H + \sigma_n^2 \mathbf{I}$. It follows from the previous that there is only one eigenvalue larger than σ_n^2. The corresponding eigenvector is $\mathbf{u}_1 = \mathbf{a}_{r1} ||\mathbf{a}_{r1}||^{-1}$, and is in the direction of $\mathbf{a}_{r1}$. The power coming from that direction is

$$\lambda_1 = \mathbf{u}_1^H \mathbf{R} \mathbf{u}_1 = p\sigma_{r1}^2 + \sigma_n^2 \tag{3.36}$$

Since there is only one interferer, the power coming from any other direction orthogonal to $\mathbf{u}_1$ is σ_n^2, the noise power.

With more than one interferer, this can be generalised. Suppose there are two interferers with powers σ_{r1}^2 and σ_{r2}^2, and spatial signatures $\mathbf{a}_{r1}$ and $\mathbf{a}_{r2}$. If the spatial signatures are orthogonal, $\mathbf{a}_{r1}^H \mathbf{a}_{r2} = 0$, then $\mathbf{u}_1$ will be in the direction of the strongest interferer, number 1, say, and $\lambda_1 = p\sigma_{r1}^2 + \sigma_n^2$ will be the corresponding power. Similarly, $\lambda_2 = p\sigma_{r2}^2 + \sigma_n^2$.

In general, the spatial signatures are not orthogonal to each other. In that case, $\mathbf{u}_1$ will point to a direction that is common to both $\mathbf{a}_1$ and $\mathbf{a}_2$, and $\mathbf{u}_2$ will point in the remaining direction orthogonal to $\mathbf{u}_1$. The power λ_1 coming from direction $\mathbf{u}_1$ will be larger than before because it combines power from both interferers, whereas λ_2 will be smaller.

Usually, the diagonal elements of the noise covariance matrix $\mathbf{D}$ are not equal, and also the elements of $\mathbf{G}$ are usually not equal. They also may vary slowly in time. This means that in general $\mathbf{U}$ and $\mathbf{\Lambda}$ found from an eigenvalue decomposition of $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ are not equal to $\mathbf{U}_r$ and $\mathbf{\Lambda}_r$ found from an eigenvalue decomposition of $\mathbf{R}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$. As the interference mitigation approaches in this thesis make use of eigenvalue decompositions, it is important to compensate for these distortions. To handle this subspace data processing problem, there are several approaches, three of which are listed below. The model $\mathbf{R} = \mathbf{G}(\mathbf{A}\mathbf{B}\mathbf{A}^H + \mathbf{A}_r \mathbf{B} \mathbf{A}_r^H)\mathbf{G}^H + \mathbf{D}$ is considered here, and it is assumed that the astronomical source signal powers are much smaller than the system noise powers.

1. No corrections for $\mathbf{D}$ *and* $\mathbf{G}$

In this case the model under consideration is $\mathbf{R} = \mathbf{G}(\mathbf{A}\mathbf{B}\mathbf{A}^H + \mathbf{A}_r \mathbf{B} \mathbf{A}_r^H)\mathbf{G}^H + \mathbf{D}$, and the matrix $\mathbf{R}$ is always full rank. In case a strong interferer exists with

$\sigma_{r1}^2 \gg \sigma_{ni}^2$ $(\forall i)$ then the largest eigenvalue and eigenvector are close to the "true" values. For moderate and weak interferers, the eigenvalue and eigenvector estimated will be suboptimal, and approaches such as described below are preferred.

2. Applying corrections for $\mathbf{G}$ *and* $\mathbf{D}$

Assuming that the astronomical source powers are known for the strongest sources, then in time-frequency regions of observed data where there is no interference, the data itself can be used for gain and noise estimation, yielding $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{D}}$. The corrected model $\mathbf{R}'$ to which the eigenvalue composition will be applied is

$$\mathbf{R}' = \widehat{\mathbf{G}}^{-1}(\mathbf{R} - \widehat{\mathbf{D}})\widehat{\mathbf{G}}^{-H} \tag{3.37}$$

The noise matrix in $\mathbf{R}'$ takes the form $\mathbf{D} - \widehat{\mathbf{D}}$ and is approximately zero, which means that the eigenvalue decomposition of $\mathbf{R}'$ is close to the "true" one $(\mathbf{ABA}^H + \mathbf{A}_r\mathbf{B}_r\mathbf{A}_r^H)$ if $\widehat{\mathbf{G}}$ is close to $\mathbf{G}$ and if $\widehat{\mathbf{D}}$ is close to $\mathbf{D}$. A disadvantage is that with estimated data, $\widehat{\mathbf{R}}'$ may no longer be positive definite.

3. Pre-whitening using an estimate for $\mathbf{D}$

Assuming that the noise matrix $\widehat{\mathbf{D}}$ is estimated by a calibration procedure, then the corrected model $\mathbf{R}'$ to which the eigenvalue composition will be applied is

$$\mathbf{R}' = \widehat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{R}\widehat{\mathbf{D}}^{-\frac{1}{2}} = \widehat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{G}(\mathbf{ABA}^H + \mathbf{A}_r\mathbf{BA}_r^H)\mathbf{G}^H\widehat{\mathbf{D}}^{-\frac{1}{2}} + \widehat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{D}\widehat{\mathbf{D}}^{-\frac{1}{2}} \tag{3.38}$$

If $\widehat{\mathbf{D}}$ is close to $\mathbf{D}$, then $\widehat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{D}\widehat{\mathbf{D}}^{-\frac{1}{2}} \approx \mathbf{I}$. In this approach $\widehat{\mathbf{D}}$ is known and the multiplication with $\widehat{\mathbf{D}}^{-\frac{1}{2}}$ can be reversed in further processing steps, but $\mathbf{G}$ is still unknown and has to be estimated at some later stage.

3.2.2 Factor analysis decomposition

Factor analysis is a statistical technique with origins in psychometrics and biometrics [91, 105]. It considers a collection of data $\mathbf{X} = [\mathbf{x}(1), \cdots, \mathbf{x}(N)]$ with covariance $\mathbf{R} = \mathcal{E}\{\mathbf{XX}^H\} = \mathbf{AA}^H + \mathbf{D}$, where $\mathbf{R}$ is a $p \times p$ hermitian matrix, $\mathbf{A}$ a $p \times q$ matrix, and $\mathbf{D}$ a diagonal matrix. The objective in factor analysis is to estimate the factor dimension q, $\mathbf{A}$, and $\mathbf{D}$, given $\mathbf{R}$. The model which factor analysis considers closely resembles the models in this thesis for astronomical correlations. Several algorithmic approaches to solve this problem exist [91,105]; in chapter 8 several algorithms for a rank-one problem are analysed in detail. The factor analysis decomposition can be viewed as an extension of the eigenvalue decomposition. Let $\mathbf{U\Lambda U}^H$ be the eigenvalue decomposition of $\mathbf{AA}^H$, with $\mathbf{U} = [\mathbf{U}_r\mathbf{U}_n]$, then

$$\mathbf{R} = \mathbf{U\Lambda U}^H + \mathbf{D} = \begin{bmatrix}\mathbf{U}_r\mathbf{U}_n\end{bmatrix}\begin{bmatrix}\mathbf{\Lambda}_r & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{U}_r^H\\ \mathbf{U}_n^H\end{bmatrix} + \mathbf{D} \tag{3.39}$$

Note that a useful factorisation only exists for $q \leq p$. Furthermore, an estimate of $\mathbf{A}$ is not unique since $\mathbf{A}$ can be replaced by $\mathbf{AQ}$ for any unitary matrix $\mathbf{Q}$ without altering the product $\mathbf{AA}^H$ as $\mathbf{AA}^H = \mathbf{AQQ}^H\mathbf{A}^H$. As a result only $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{U}_r) = q$ and $\mathbf{\Lambda}_r$ can be determined uniquely. There are ways to constrain $\mathbf{A}$ to be a unique factor [151], but the resulting factor, although consistent with the data, probably is not the "true" factor.

3.2.3 Finite sample effect

The (co)variance of the sample array covariance matrix is a crucial mathematical tool for the analysis and interpretation of interference mitigation methods. The estimation accuracies of the eigenvectors and eigenvalues are crucial as well, as subspace analysis is also used throughout this thesis. This section therefore lists a few useful formulas for these estimation (co)variances.

Covariance of the covariance matrix

Given baseband sample data stacked in a $p \times N$ matrix $\mathbf{X}$, with p the number of telescopes and N the number of samples, then the $(p \times p)$ covariance matrix sample estimate $\widehat{\mathbf{R}}$ is defined by

$$\widehat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{X}\mathbf{X}^H \tag{3.40}$$

The sample data are considered Gaussian zero mean random variables, with $\mathcal{E}\{\widehat{\mathbf{R}}\} = \mathbf{R}$ having a central complex Wishart distribution [89, 33]. In appendix C, the variance $\text{var}(\widehat{\mathbf{R}}) = \mathcal{E}\{(\widehat{\mathbf{R}} - \mathbf{R}) \odot \overline{(\widehat{\mathbf{R}} - \mathbf{R})}\}$ and covariance $\text{cov}(\widehat{\mathbf{R}}) = \mathcal{E}\{(\text{vec}(\widehat{\mathbf{R}}) - \text{vec}(\mathbf{R}))(\text{vec}(\widehat{\mathbf{R}}) - \text{vec}(\mathbf{R}))^H\}$ of the array covariance matrix $\widehat{\mathbf{R}}$ are derived, together with the weighted versions $\widehat{\mathbf{R}}_w = \mathbf{W}\widehat{\mathbf{R}}\mathbf{W}$ where $\mathbf{W}$ is derived from $\widehat{\mathbf{R}}$ or is known from a-priori knowledge of the model $\mathbf{R}$.

$$\text{cov}(\widehat{\mathbf{R}}) = \frac{1}{N}\overline{\mathbf{R}} \otimes \mathbf{R} \tag{3.41}$$

$$\text{var}(\widehat{\mathbf{R}}) = \frac{1}{N}\text{vecdiag}(\mathbf{R})(\text{vecdiag}(\mathbf{R}))^t \tag{3.42}$$

$$\text{cov}(\widehat{\mathbf{R}}_w) = (\mathbf{W} \otimes \mathbf{W})\text{cov}(\widehat{\mathbf{R}})(\mathbf{W} \otimes \mathbf{W}) \tag{3.43}$$

$$\text{var}(\widehat{\mathbf{R}}_w) = \mathbf{W}^2\text{var}(\widehat{\mathbf{R}})\mathbf{W}^2 \tag{3.44}$$

The dimension of $\text{cov}(\widehat{\mathbf{R}})$ and $\text{cov}(\widehat{\mathbf{R}}_w)$ is $p^2 \times p^2$, the dimension of $\text{var}(\widehat{\mathbf{R}})$ and $\text{var}(\widehat{\mathbf{R}}_w)$ is $p \times p$. The (co)variance of $\widehat{\mathbf{R}}_w$ above is exact for $\mathbf{W}$ fixed, but it is only approximately true if $\mathbf{W}$ is derived from $\widehat{\mathbf{R}}$.

Eigenvector covariance

Consider the eigenvalue decomposition $\mathbf{R} = \mathbf{U\Lambda U}^H$, where the eigenvector matrix $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_p]$ contains the eigenvectors $\mathbf{u}_i (i = 1 \cdots, p)$, and the

eigenvalue matrix $\mathbf{\Lambda}$ contains the corresponding eigenvalues λ_i: $\mathbf{\Lambda} = \text{diag}(\lambda_1, \cdots, \lambda_p)$. Let $\mathbf{\Psi}_i = \mathbf{\Lambda} - \lambda_i \mathbf{I}$, then the covariance of the eigenvector $\mathbf{u}_i$ is given by [141, 105]:

$$\text{cov}(\hat{\mathbf{u}}_i) = \frac{1}{N}\lambda_i \mathbf{U}\mathbf{\Psi}_i^{\dagger}\mathbf{\Lambda}\mathbf{\Psi}_i^{\dagger}\mathbf{U}^H = \frac{1}{N}\sum_{j\neq i}\frac{\lambda_i\lambda_j}{(\lambda_j-\lambda_i)^2}\mathbf{u}_j\mathbf{u}_j^H \tag{3.45}$$

$$\text{var}(\hat{\mathbf{u}}_i) = \text{vecdiag}(\text{cov}(\hat{\mathbf{u}}_i)) \tag{3.46}$$

This equation is valid for eigenvectors of $\widehat{\mathbf{R}}$ corresponding to a Gaussian model and corresponding to unequal eigenvalues.

Eigenvalue variance

Consider the variance $\text{var}(\widehat{\mathbf{R}})$ for the model given in (2.40), which is simplified by assuming that there are no interferers and that the astronomical sources are so weak they can be ignored. This leads to the noise-dominated model $\mathbf{R} = \mathbf{D}$, with $\text{var}(\widehat{\mathbf{R}}) = \text{var}(\widehat{\mathbf{D}}) = \frac{1}{N}\text{vecdiag}(\mathbf{D})\text{vecdiag}(\mathbf{D})^t = \frac{1}{N}\mathbf{D}^2$. Recall that $\mathbf{D}$ is diagonal, with eigenvalue decomposition $\mathbf{D} = \mathbf{I}\mathbf{\Lambda}\mathbf{I}^H$, which leads in the interferer-free and weak astronomical sources case to:

$$\text{var}(\hat{\mathbf{\Lambda}}) = \frac{1}{N}\mathbf{\Lambda}^2 \tag{3.47}$$

This formula is also valid in a more general case with arbitrary positive entries on the diagonal of $\mathbf{D}$, as is shown in [33].

3.2.4 Eigenvalue estimates and exact solutions

For the use in interference mitigation filters or for the evaluation and analysis of such filters it is sometimes convenient to have estimates of eigenvalues of the covariance model. The eigenvalues can be easily estimated by computer simulation in case all noise and signal contributions are known (system noise, interferer and astronomical signals) and assuming a calibrated telescope. However, simple closed-form expressions exist for the approximate values of the largest and smallest eigenvalue, if the system is fully dominated by system noise. These expressions are listed below, followed by a closed-form eigenvalue solution for a two-source model with i.i.d. noise.

Eigenvalue extremes for systems with noise

Consider a data model based on N baseband data samples in which there is only system noise with equal system power for each of the p telescopes, $\mathbf{R} = \sigma^2\mathbf{I}$. For this model, the maximum and minimum eigenvalues γ_{max} and γ_{min} of $\widehat{\mathbf{R}}$ are approximately given by [42]:

$$\lambda_{max} \approx \sigma^2\left(1+\sqrt{\frac{p}{N}}\right)^2 \tag{3.48}$$

$$\lambda_{min} \approx \sigma^2 \left(1 - \sqrt{\frac{p}{N}}\right)^2 \tag{3.49}$$

The eigenvalues λ_{max} and λ_{min} converge to the limits given above for $N \to \infty$. These expressions are useful, for example, to determine thresholds for a decision to apply an RFI mitigation algorithm. It was used, for example, for determining the detector threshold in interference excision.

Eigenvalues for a two sources model with noise

Consider a sensor array model consisting of two point sources with baseband signals $s_1(t)$ and $s_2(t)$ at (normalised) directions $\mathbf{a}_1$ and $\mathbf{a}_2$. Assume that the additive system noise $\mathbf{n}(t)$ is i.d.d. with $\mathcal{E}\{\mathbf{n}(t)\mathbf{n}(t)^H\} = \sigma_n^2\mathbf{I}$:

$$\mathbf{x}(t) = \mathbf{a}_1 s_1(t) + \mathbf{a}_2 s_2(t) + \sigma_n^2\mathbf{I} \tag{3.50}$$

Define the source power variances by $\sigma_1^2 = \mathcal{E}\{s_1(t)\overline{s_1(t)}\}$ and $\sigma_2^2 = \mathcal{E}\{s_2(t)\overline{s_2(t)}\}$, and define the inter signal correlation by $\rho_{12} = \mathcal{E}\{s_1(t)\ \overline{s_2(t)}\}$ and $\rho_{21} = \mathcal{E}\{s_2(t)\overline{s_1(t)}\}$. Using these definitions the covariance matrix $\mathbf{R} = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}$ of this system is given by

$$\mathbf{R} = \sigma_1^2\mathbf{a}_1\mathbf{a}_1^H + \sigma_2^2\mathbf{a}_2\mathbf{a}_2^H + \rho_{12}\mathbf{a}_1\mathbf{a}_2^H + \rho_{21}\mathbf{a}_2\mathbf{a}_1^H + \sigma_n^2\mathbf{I} \tag{3.51}$$

Closed form solutions for the eigenvalues of this system exist; the solutions for the two largest eigenvalues λ_1 and λ_2 are given by [68] [177]:

$$\begin{aligned}\lambda_{1,2} \quad = \quad & \sigma_n^2 + \left(\frac{1}{2}(\sigma_1^2 + \sigma_2^2) + \mathrm{Re}(\rho_{12}\,\mathbf{a}_2^H\mathbf{a}_1)\right). \\ & \left(1 \pm \sqrt{1 - 4\frac{(1 - |\mathbf{a}_1^H\mathbf{a}_2|^2)(\sigma_1^2\sigma_2^2 - |\rho_{12}|^2)}{(\sigma_1^2 + \sigma_2^2 + 2\mathrm{Re}(\rho_{12}\,\mathbf{a}_2^H\mathbf{a}_1))^2}}\right)\end{aligned} \tag{3.52}$$

The remaining eigenvalues are equal to σ_n^2. This result is useful for the estimation of several effects such as the influence of the spatial extendedness of sources and the influence of bandwidth on the subspace structure of the array covariance matrix. This tool will be helpful in determining which factors limit the interference mitigation effectiveness, as will be described in the next section.

3.3 Influence of assumption violations on subspace

Throughout this thesis, it is assumed that the narrowband approximation and the point-source approximation hold. The advantage of this assumption is simplicity. The first assumption allows time delays of impinging signals to be described by phases, which are modelled relatively easily. The advantage of the second assumption allows modelling of a continuum source distribution by discrete sources, which is also relatively easy. The question arises what the

influence of (limited) assumption violations is on interference mitigation effectiveness.

Consider, for example, a spatial interference detector (cf. chapter 5) or spatial interference filter (cf. chapter 6), based on subspace analysis in which only the largest eigenvalue and its corresponding eigenvector are used. In the noise-free case and in case there are no astronomical sources, all energy of the transmitter will be represented by the largest eigenvalue, and the corresponding eigenvector will be the true array signature vector. In case there is noise (cf. section (3.2.3)) or in case the assumptions mentioned above are violated, the eigenvalue and eigenvector distribution will change. The rank-one approach will lead to suboptimal interference filters, but a detector or filter based on a rank-two model, or more generally on a rank-q_r model, will perform better. A disadvantage of higher-rank interference filters is that those filters distort the astronomical signals more than rank-one filters. This is caused by the fact that a higher-rank spatial filter will remove a larger subspace than a rank-one filter. It is therefore relevant to investigate the influence of these assumption violations on the subspace structure, as will be discussed in the next two sections.

3.3.1 Narrowband assumption

Influence of bandwidth on covariance matrix eigenvalues

The purpose of this section is to derive an expression for the eigenvalues of the model

$$\mathbf{R} = \sigma_s^2 \mathbf{a}\mathbf{a}^H + \sigma_n^2 \mathbf{I} \tag{3.53}$$

as a function of the bandwidth used. This is the model for a single point source with source power σ_s^2, with a normalised array response vector for the source $\mathbf{a}$, and i.i.d. spatially white noise with power $\sigma_n^2\mathbf{I}$. The first part of the derivation concerns a Taylor expansion of the spatial signature vector $\mathbf{a}$ [68], while the second part differs from [68] in that no orthogonality requirements are needed, but a direct solution is given using the two-source eigenvalue model. The frequency-dependent covariance matrix $\mathbf{R}(\nu)$ of the system under consideration is given by

$$\mathbf{R}(\nu) = \sigma_s^2 \mathbf{a}(\nu)\mathbf{a}(\nu)^H + \sigma_n^2 \mathbf{I} \tag{3.54}$$

The array response vector $\mathbf{a}(\nu)$ is frequency dependent, and it is assumed that different frequency components ν_1 and ν_2 are uncorrelated: let $S(\nu)$ be the Fourier transform of $s(t)$, then $\mathcal{E}\{\overline{S(\nu_1)}S(\nu_2)\} = 0$ for $\nu_1 \neq \nu_2$. The total array covariance $\mathbf{R}$ is an integral over frequency, with ν_0 the centre frequency of the band under consideration, and $\Delta\nu$ the bandwidth:

$$\mathbf{R} = \frac{\sigma_s^2}{2\Delta\nu} \int_{\nu_o-\Delta\nu}^{\nu_o+\Delta\nu} \mathbf{a}(\nu)\mathbf{a}(\nu)^H d\nu + \sigma_n^2 \mathbf{I} \tag{3.55}$$

This expression can be evaluated using a Taylor expansion of $\mathbf{a}(\nu)$:

$$\mathbf{a}(\nu) = \sum_{n=0}^{\infty} \frac{(\nu-\nu_o)^n}{n!} \left[\frac{d^n \mathbf{a}(\nu)}{d\nu^n}\right]_{\nu=\nu_o} \tag{3.56}$$

Let $\mathbf{a}_o^{(n)} = \left[\frac{d^n \mathbf{a}(\nu)}{d\nu^n}\right]_{\nu=\nu_o}$, $\mathbf{a}_o^{(0)} = 1$, and let $\mathbf{R}_a = \mathbf{R} - \sigma_n^2\mathbf{I}$, and insert up to order two terms of the Taylor expansion in the expression for $\mathbf{R}_a$:

$$\begin{aligned}
\mathbf{R}_a &\approx \frac{\sigma_s^2}{2\Delta\nu}\int_{\nu_o-\Delta\nu}^{\nu_o+\Delta\nu}\left[\mathbf{a}_o^{(0)} + (\nu-\nu_o)\mathbf{a}_o^{(1)} + \frac{1}{2}(\nu-\nu_o)^2\mathbf{a}_o^{(2)}\right]\cdot \\
&\qquad \left[\mathbf{a}_o^{(0)} + (\nu-\nu_o)\mathbf{a}_o^{(1)} + \frac{1}{2}(\nu-\nu_o)^2\mathbf{a}_o^{(2)}\right]^H d\nu \\
&= \sigma_s^2(\mathbf{a}_o^{(0)}\mathbf{a}_o^{(0)^H}) + \frac{\sigma_s^2}{6}\Delta\nu^2\left(\mathbf{a}_o^{(0)}\mathbf{a}_o^{(2)^H} + \mathbf{a}_o^{(2)}\mathbf{a}_o^{(0)^H}\right) + \\
&\qquad \frac{\sigma_s^2}{3}\Delta\nu^2\mathbf{a}_o^{(1)}\mathbf{a}_o^{(1)^H} + \frac{\sigma_s^2}{20}\Delta\nu^4\mathbf{a}_o^{(2)}\mathbf{a}_o^{(2)^H}
\end{aligned} \tag{3.57}$$

The integral of the products of the term containing $\mathbf{a}_o^{(1)}$ with the terms containing $\mathbf{a}_o^{(0)}$ and $\mathbf{a}_o^{(2)}$ are zero because these are odd functions, integrated over a symmetric interval with respect to ν_o. Let

$$\sigma_1\mathbf{a}_1 = \sigma_s\left(\mathbf{a}_o^{(0)} + \frac{1}{6}\Delta\nu^2\mathbf{a}_o^{(2)}\right), \quad \sigma_2\mathbf{a}_2 = \sigma_s\left(\frac{1}{\sqrt{3}}\Delta\nu\mathbf{a}_o^{(1)}\right) \tag{3.58}$$

with the constraints $|\mathbf{a}_1| = |\mathbf{a}_2| = 1$. Inserting these definitions in the previous equation yields

$$\mathbf{R} = \sigma_1^2\mathbf{a}_1\mathbf{a}_1^H + \sigma_2^2\mathbf{a}_2\mathbf{a}_2^H + \sigma_n^2\mathbf{I} + \mathcal{O}(\Delta\nu^4) \tag{3.59}$$

This equation corresponds to the two-source model with $\rho_{12} = \rho_{21} = 0$, for which the eigenvalues are known in closed form. In a narrowband system only the zero$^{\text{th}}$ order expansion term would be nonzero, and $\mathbf{R}_a$ would be rank one. A system which deviates slightly from being perfectly narrowband , the higher order terms of the expansion would be nonzero, and $\mathbf{R}_a$ would be rank two. This means that the value of the second eigenvalue λ_2 (i.e. the one-but-largest eigenvalue) of $\mathbf{R}_a$ is a useful measure for the narrowness of a frequency band. The largest and second-but-largest eigenvalues (cf. section 3.2.4) are given by:

$$\lambda_{1,2} = \sigma_n^2 + \frac{1}{2}(\sigma_1^2+\sigma_2^2)\left(1 \pm \sqrt{1 - 4\frac{\sigma_1^2\sigma_2^2(1-|\mathbf{a}_1^H\mathbf{a}_2|^2)}{(\sigma_1^2+\sigma_2^2)^2}}\right) \tag{3.60}$$

As in equation 2.19, define $\mathcal{R} = (\mathbf{r}_1, \cdots, \mathbf{r}_p)^t$, and let $\mathbf{s}_\ell$ be the direction vector of the ℓ^{th} source, then the derivatives of the normalised array response vector $\mathbf{a}(\nu)$ at $\nu = \nu_o$ (cf. section 2.3.1) can be straightforwardly expressed by

$$\mathbf{a}_o^{(o)} = p^{-\frac{1}{2}}e^{-\imath\nu_o c^{-1}\mathcal{R}\mathbf{s}_\ell} \tag{3.61}$$

$$\mathbf{a}_o^{(1)} = -\imath c^{-1}\mathcal{R}\mathbf{s}_\ell \odot \mathbf{a}_o^{(o)} \tag{3.62}$$

$$\mathbf{a}_o^{(2)} = -c^{-2}\mathcal{R}\mathbf{s}_\ell \odot \mathcal{R}\mathbf{s}_\ell \odot \mathbf{a}_o^{(o)} \tag{3.63}$$

These three expressions are used for calculating the rank-two model in (3.59), given a telescope configuration $\mathcal{R}$ and a source direction $\mathbf{s}_\ell$. Initially the model $\mathbf{R} - \sigma_n^2\mathbf{I}$ is rank-one, assuming there is only one source, see equation (3.53), but the finite bandwidth transforms it into a rank-two model, see equation (3.59). The largest and second largest eigenvalues of this single source model are dependent on the bandwidth and can be calculated using (3.60). The effect of bandwidth on the second eigenvalue for a single source model for the WSRT is illustrated in the next paragraph.

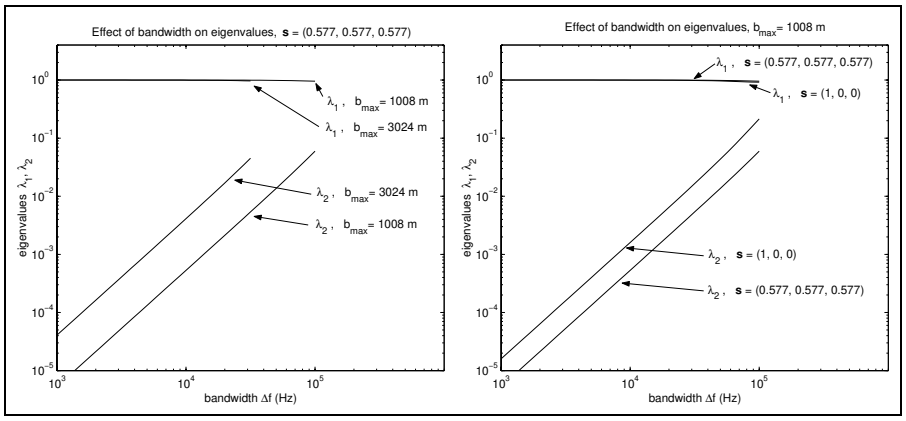


Figure 3.4. Effect of bandwidth on the two largest eigenvalues for a one-source model with noise: $\mathbf{R} = \sigma_s^2\mathbf{a}\mathbf{a}^H + \sigma_n^2\mathbf{I}$. The left-hand figure shows the eigenvalues for $p = 8$ and $(\mathbf{r}_i - \mathbf{r}_j)_{max} = 1008$ m (the configuration used for the majority of the RFI mitigation experiments), and for $p = 22$ and $(\mathbf{r}_i - \mathbf{r}_j)_{max} =$ 3 km (a full 3 km baseline WSRT configuration). The source direction used was $\mathbf{s} = 3^{-\frac{1}{2}}(1, 1, 1)^t$. The figure to the right shows the eigenvalues for $p = 8$ and $(\mathbf{r}_i - \mathbf{r}_j)_{max} = 1008$ m for two different directions $\mathbf{s} = 3^{-\frac{1}{2}}(1, 1, 1)^t$ and $\mathbf{s} = (1, 0, 0)^t$

Bandwidth requirement for the WSRT

Using the narrow bandwidth formula (3.2) for the WSRT telescopes, with 8 telescopes and a maximum baseline of 1008 m, the bandwidth $\Delta\nu$ of the frequency bins should be $\Delta\nu \ll 300$ kHz. A bandwidth of 30 to 50 kHz seems reasonable in this case. However, considering a single source impinging on the array ($(\mathbf{R} - \sigma_n^2\mathbf{I})$ should be rank one), such a bandwidth has a relatively large influence on the second eigenvalue as is demonstrated in figure 3.4, where the ratio of the second and the first (largest) eigenvalue reaches -30 to -20 dB. In figure 3.4, left, the influence of telescope configuration (maximum baseline length) is shown, the right-hand figure illustrates the effect of the pointing direction. Because of the east-west orientation of the WSRT , sources in the east or west directions give the most stringent constraints on the narrowness

of the band. Directions in the meridian plane pose no requirements to the narrowness of the band because in this case there is no time delay between the different telescopes.

For example, consider a spatial filter for interference mitigation which should be able to suppress an interferer with a certain factor. Suppose also that the filter should remove only one subspace dimension (cf. chapter 6 for details). Then the second eigenvalue in the bandwidth model above should be at least 3 dB lower than the suppression factor. If a 30 dB spatial filter is needed which suppresses only one subspace dimension, then the second eigenvalue should be $\lessapprox$ -33 dB with respect to the largest eigenvalue, which corresponds to a 10 kHz bandwidth for 1 km maximum baseline length.

3.3.2 Point source assumption

As was shown earlier in section 2.5.1, the multipaths of interferers which are not fully correlated with the main path increase the rank of the corresponding covariance matrix. If a spatial filter is based on the observed largest eigenvalue, then an increase in rank due to multipath reduces the effectiveness of the filter. Multipath in this case can be considered as a deviation from the point source assumption.

Another example where deviations from the point-source assumption may have influence is the gain estimation accuracy of a telescope using an astronomical point source in the centre of the field of view. An ideal point source would yield a rank-one system; an extended source would yield rank two or more if different sections of the extended source were not fully correlated.

The relevant question to be answered in both cases is how deviations from the point-source assumption influence the eigenvalues of the corresponding covariance matrix $\mathbf{R}$. This question will be answered by considering a single-source model with a single multipath. The formulas which will be derived are valid for the multipath case, but are only approximately valid for the extended sources case.

Single point source model with multipaths

Suppose that there is one emitting source in direction $\mathbf{a}_1$ with signal $s_1(t)$, and one multipath in direction $\mathbf{a}_2$ with signal $s_2(t)$. These two signals are either spatially correlated, partially correlated, or not correlated. The array output vector $\mathbf{x}(t)$ for the source with multipath can be written as

$$\mathbf{x}(t) = \mathbf{a}_1\, s_1(t) + \mathbf{a}_2\, s_2(t) + \mathbf{n}(t) \tag{3.64}$$

where $\mathbf{a}_i = e^{-2\pi\jmath\nu_i c^{-1}\mathcal{R}\mathbf{s}_i}$ is the spatial signature vector for the i^{th} source(cf. section 2.3.1), $s_i(t)$ is the baseband signal of source i, and $\mathbf{n}(t)$ is the noise vector. The angle between $\mathbf{s}_1$ and $\mathbf{s}_2$ is given by θ. Assume that $\mathbf{s}_1$ is fixed, and that $\mathbf{s}_2$ varies. Note that by definition $||\mathbf{s}_1|| = ||\mathbf{s}_2|| = 1$. The covariance of the array output is defined by: $\mathbf{R} = \mathcal{E}\{\mathbf{x}(t)\mathbf{x}(t)^H\}$. Let $\sigma_1^2 = \mathcal{E}\{\overline{s_1(t)}s_1(t)\}$,

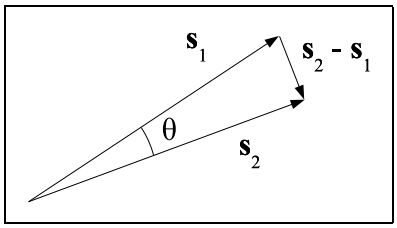


Figure 3.5. Direction vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ of a point-source with multipath.

$\sigma_2^2 = \mathcal{E}\{\overline{s_2(t)}s_2(t)\}$, and let $\rho_{12} = \mathcal{E}\{s_1(t)\overline{s_2(t)}\}$ and $\rho_{12} = \overline{\rho_{21}}$, then

$$\mathbf{R} = \sigma_1^2\mathbf{a}_1\mathbf{a}_1^H + \sigma_2^2\mathbf{a}_2\mathbf{a}_2^H + \rho_{12}\mathbf{a}_1\mathbf{a}_2^H + \rho_{21}\mathbf{a}_2\mathbf{a}_1^H + \mathbf{D} \tag{3.65}$$

where $\mathbf{D} = \mathcal{E}\{\mathbf{n}(t)\mathbf{n}(t)^H\}$. This is basically the two-source model described in section 3.3.1, so the distribution of the power of a point-source and its multipath over the two largest eigenvalues is given by equation (3.52). This is a general case which is valid for dual sources with various degrees of mutual correlation.

Now suppose the point-source and its multipath are fully correlated: $\rho_{12} = \sigma_1\sigma_2$. Inserting this in equation 3.52 yields $\lambda_1 = \sigma_n^2 + \sigma_1^2 + \sigma_2^2 + \mathrm{Re}\{\sigma_1\sigma_2\mathbf{a}_2^H\mathbf{a}_1\}$, and $\lambda_2 = 0$. When the point-source and its multipath are not fully correlated then $\lambda_2 > 0$. A spatial filter based on a rank-one analysis would then lead to a suboptimal filter (as compared to a larger-rank filter).

If the narrowband condition holds, then $\rho_{12} = \sigma_1\sigma_2$, and $\mathbf{R} - \mathbf{D}$ will be rank one. If, however, the propagation delay of the multipath is much larger than the geometric delay differences across the array, and if for this multipath the narrowband condition does not hold, then the resulting model $\mathbf{R} - \mathbf{D}$ is rank two. Figure 3.6 shows ρ_{12} for a single source with one multipath as a function of path delay. For a box car frequency taper, ρ_{12} is proportional to $\mathrm{sinc}(2\pi\Delta\nu\tau)$. At 1420 MHz, and for a 1152 m pathlength distance (between nine WSRT telescopes), ρ_{12} is $0.34\sigma_1\sigma_2$ for 30 kHz bandwidth and $0.999\sigma_1\sigma_2$ for 1 kHz bandwidth.

Figure 3.7 shows the results of a calculation of the two largest eigenvalues for a single point-source model with a single multipath, having varying degrees of mutual correlation. These calculations were carried out for the WSRT using nine fixed telescopes. The point-source is located in the $3^{-\frac{1}{2}}(1,1,1,)^t$ direction.

The upper right-hand figure shows the two largest eigenvalues as a function of point-source - multipath separation angle for $\rho_{12} = 0.33\sigma_1\sigma_2$ and $\Delta\nu = 30$ kHz, as was mentioned above. This figure shows that if a rank-one based spatial interference filter is used, with an attenuation factor of, for example, 20 dB, then this filter would be optimal for source separations less than 0.2^o at 8.6 GHz sky frequency. If in this case the multipath delay is more than the delay across the array, then the 20 dB attenuation cannot be reached unless a higher-rank filter is used.

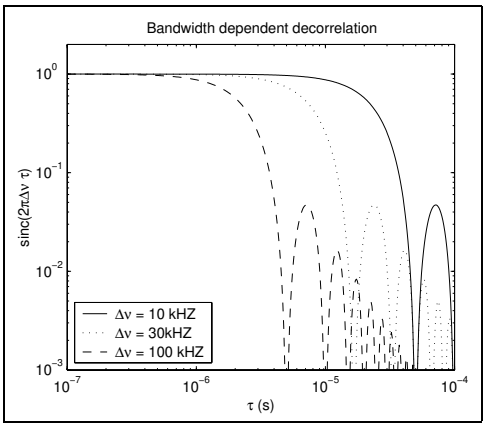


Figure 3.6. Correlation $\rho_{12}\sigma_1\sigma_2$ for a single source with one multipath as a function of path delay. For a box car frequency taper, ρ_{12} is proportional to $\text{sinc}(2\pi\Delta\nu\tau)$.

The lower left-hand figure shows that for $\rho_{12} = 0.999$ (corresponding to $\Delta\nu = 1$ kHz), a rank-one 20 dB filter puts no requirements on source-multipath separation. The upper left and lower right-hand figures show the eigenvalue distribution for $\rho_{12} = 0$ and $\rho_{12} = 0.99999$ respectively.

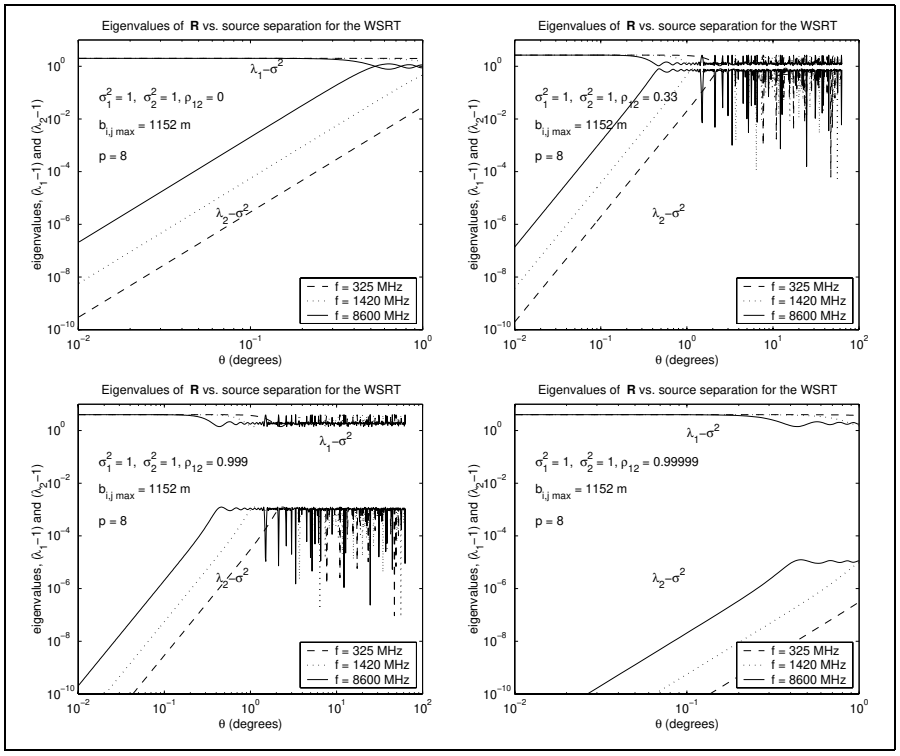


Figure 3.7. Calculation results of the two largest eigenvalues for a single point-source model with an uncorrelated multipath (upper left), and with partially correlated multipaths (remaining subfigures). The results were obtained for the WSRT configuration with nine fixed telescopes. The point-source is located in the $3^{-\frac{1}{2}}(1,1,1,)^t$ direction.

Chapter 4

Experimental set-up

In this chapter a description will be given of the experimental set-ups used for obtaining measurement data. The main part of the set-up is the WSRT radio telescope described in section 4.1. The WSRT data were recorded using a data acquisition system developed in the context of the project Nulling Osbtructing Electromagnetic Interferers (NOEMI). The general goal of this project was to investigate the merits of modern array signal processing algorithms for interference mitigation in radio interferometry. More specifically, the project aims were (1) to select viable interference rejection scenarios including performance requirements, (2) to characterise and record the signals of interest, (3) to select and analyse appropriate signal detection and rejection algorithms based on subspace techniques, (4) to develop a small-scale RFI mitigation demonstrator system, (5) to conduct field tests, and (6) to explore implications for LOFAR and/or SKA. The project started in 1999, ended in 2003, and was supported by the Dutch Technology Foundation STW under DEL-77-4476, DTC.5893. The NOEMI baseband time sample recording system will be described in section 4.2, and the NOEMI online DSP correlation system will be described in section 4.3. Finally, an offline/online recording mode for WSRT data including a phased-array reference antenna will be described in section 4.4. This reference antenna, the NOEMI-THEA tile, is a phased array antenna developed by ASTRON.

4.1 The Westerbork Synthesis Radio Telescope

The Westerbork Synthesis Radio Telescope (WSRT , figure 4.1) consists of fourteen dual polarisation, twenty-five meter diameter, parabolic dishes, and operates in the frequency bands specified in table 4.1. The telescope is configured as a linear East-West array, shown in figure 4.2, with regular 144-m distances between ten of the fourteen telescopes. Two telescopes are mounted on rail tracks close to the fixed telescopes, and two are mounted on rail tracks at approximately 1.5 km distance. All telescopes are linearly polarised, except the 13-cm

band, which is circularly polarised. After amplification with low noise amplifiers (LNAs) and filtering, the telescope signals are down converted to baseband, using a local oscillator system which is locked to a frequency reference (Hydrogen MASER). The baseband signals are digitised, and correlated and stored for further image data processing after application of a digital delay. The WSRT also has other operational modes (offline correlation in VLBI modes, pulsar time series observations), but these are not discussed here, as this thesis focuses on RFI mitigation applied to WSRT correlation data.

Figure 4.1. Westerbork Synthesis Radio Telescope (WSRT), The Netherlands, operating in selected frequency bands from 110 MHz to 8.6 GHz.

System properties

Table 4.1 lists system properties of the WSRT. The first two columns give the available receiver bands and corresponding frequency ranges. The third column gives the system noise power or system temperature T_{sys} in K, defined as [124, 87]

$$T_{sys} = \frac{P_R}{k\Delta\nu} \quad \text{(K)} \tag{4.1}$$

where P_R is the electric power (W) at the antenna terminals (after signal reception but prior to amplification), where k is the Boltzmann constant (J/K), and where $\Delta\nu$ (Hz) is the bandwidth. The system power includes instrumental noise, signal losses, ground radiation spill-over entering via the telescope

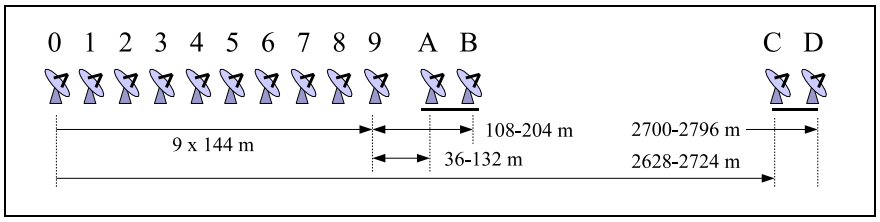


Figure 4.2. Westerbork Synthesis Radio Telescope layout: telescope positions [84]. All telescopes are equatorially mounted on an East-West line. Ten telescopes are located at regular 144 m distances. Two telescopes are mounted on rail tracks in close proximity to the fixed telescopes. The remaining two (also on rail tracks) are located at a 1.5 km distance.

Receiver	frequency (MHz)	T_{sys} (K)	η_a	Γ_{RT} (K/Jy)
LFFE	110-180			
UHF low	250-460	120-250	0.30-0.55	0.053-0.098
92 cm	310-390	125	0.59	0.105
49 cm	560-610	75	0.59	0.105
UHF high	700-1200	120-180	0.35-0.55	0.062-0.098
21/18 cm	1150-1750	27-31	0.54	0.096
13 cm	2215-2375	60	0.54	0.096
6 cm	4470-5020	65	0.48	0.085
3.6 cm	8150-8650	110	0.35	0.062

Table 4.1. WSRT telescope specifications: frequency bands, system temperatures (T_{sys}), aperture efficiencies (η_a), and telescope sensitivities (Γ_{RT}).

sidelobes, astronomical source contributions, and RFI (if present). The power received from an astronomical source antenna with an effective aperture area A_{eff} (m^2), due to an incident flux Ψ (in Jy, where $1Jy = 10^{-26}Wm^{-2}Hz^{-1}$) is given by [87] $P_\Psi = \frac{1}{2}A_{eff}\Delta\nu\Psi$. This means that the telescope sensitivity or gain Γ_{RT}, defined by $\Gamma_{RT} = \frac{T_\Psi}{\Psi}$, can be expressed as

$$\Gamma_{RT} = \frac{A_{eff}}{2k} \quad \text{(K/Jy)} \tag{4.2}$$

For the WSRT, the effective area is given by $A_{eff} = \eta_a \frac{1}{4}\pi D_{rt}^2$, where D_{rt} is the telescope dish diameter, and where η_a is the aperture efficiency. The aperture efficiency, given in the fourth column of table 4.1, is a factor corresponding to antenna losses, and to the fact that the antenna at the focus of the parabolic dish does not "illuminate" the dish area uniformly. The fifth column lists the telescope sensitivity.

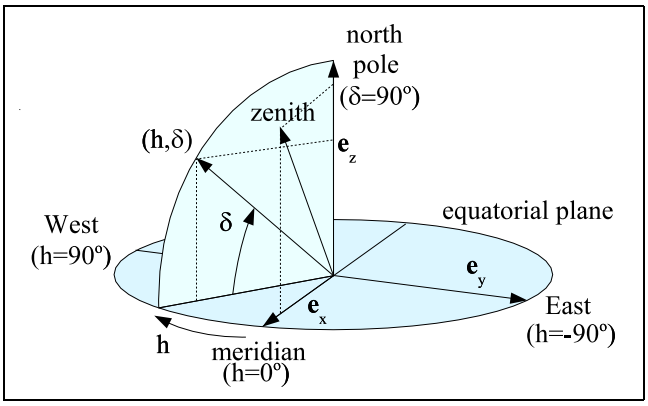


Figure 4.3. Coordinate system.

Coordinate system

The coordinate system usually used at the WSRT is the hour angle (h) and declination (δ) equatorial coordinate system shown in figure 4.3. The unit vectors $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ point in the $(\delta, h) = (0^o, 0^o)$, $(\delta, h) = (0^o, -90^o)$, and $(\delta, h) = (0^o, 90^o)$ directions respectively. An array signature vector $\mathbf{s}_o$ for the direction o can be written in carthesian coordinates $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ as:

$$\mathbf{s}_o = \begin{pmatrix} \cos(\delta_o)\cos(h_o) \\ -\cos(\delta_o)\sin(h_o) \\ \sin(\delta_o) \end{pmatrix} \tag{4.3}$$

Assuming that the first 10 telescopes with only one polarisation are used, the baselines matrix $\mathcal{R}$ can be written as $\mathcal{R} = 144\,((0,0,0)^t, (0,1,0)^t, \cdots, (0,9,0)^t)^t$ (cf. equation 2.19). In this representation telescope 0 is the baseline reference telescope. The resulting geometric delay vector $\mathbf{a}_\ell^d$ as defined in (2.20) can then easily be found by

$$\mathbf{a}_\ell^d(\mathbf{s}_\ell) = e^{-2\pi\jmath\lambda^{-1}\mathcal{R}\mathbf{s}_\ell} \tag{4.4}$$

Fringe rotation

As discussed in section 3.1.2, a geometric fringe frequency term occurs due to the application of a geometric delay correction after the mixer stage. In the literature this topic is addressed for example in [146,113]. This fringe frequency term is telescope- (baseline) dependent and needs to be compensated for. At the WSRT this is done by applying an offset frequency term which is added to the local oscillator signals. Interferers are spatially fixed with respect to the telescopes or move with different speeds and in different directions than the astronomical sources. The fringe correction term therefore introduces a changing geometric delay vector for fixed interferers. As this influences the interfering mitigation capabilities (cf. chapter 6), it is worthwhile to quantify

the fringe phase rotation magnitudes for the WSRT. The starting point is the fringe spatial signature vector (3.22) derived in section 3.1.2:

$$\mathbf{a}^f(\mathbf{s}_o) = \begin{bmatrix} e^{-2\pi\jmath\nu_o\tau_{d_1}} \\ \vdots \\ e^{-2\pi\jmath\nu_o\tau_{d_p}} \end{bmatrix} \tag{4.5}$$

As was stated in section 3.1.2, the applied geometric delay offset τ_{d_i} for telescope i is equal to $\tau_o - \tau_{g_i}$, where τ_o is an arbitrary positive offset. Recall that $(\tau_{g_1}, \cdots, \tau_{g_p})^t = \mathcal{R}\mathbf{s}_o c^{-1}$ so that

$$\mathbf{a}^f(\mathbf{s}_o) = e^{-2\pi\jmath\nu_o\tau_o} e^{-2\pi\jmath\nu_o c^{-1}\mathcal{R}\mathbf{s}_o} \tag{4.6}$$

The fringe frequency vector $\boldsymbol{\nu}_{fr}$ is defined by

$$\boldsymbol{\nu}_{fr} = \frac{\partial}{\partial t}\arg(\mathbf{a}^f(\mathbf{s}_o))) \tag{4.7}$$

Evaluating this expression yields

$$\begin{aligned} \boldsymbol{\nu}_{fr} &= -2\pi\nu_o c^{-1}\mathcal{R}\frac{\partial \mathbf{s}_o}{\partial t} \\ &= -2\pi\nu_o c^{-1}\frac{\partial\{-\cos(\delta_o)\sin(h_o)\}}{\partial t}\mathcal{R}\mathbf{1} \\ &= 2\pi\nu_o c^{-1}\cos(\delta_o)\cos(h_o)\frac{\partial h_o}{\partial t}\mathcal{R}\mathbf{1} \end{aligned} \tag{4.8}$$

The factor $\frac{\partial h_o}{\partial t}$ is the earth rotation angular frequency, and is $2\pi/(24 \times 3600)$ in radians per sidereal second.

Astronomical sources and transmitter identifications used

Finally, table 4.2 lists some of the astronomical sources which are used for the RFI mitigation experiments. The second column of the table lists the frequencies at which the sources were observed for RFI mitigation test purposes. Column three lists the source fluxes [5]. The sources are, from a radio-astronomical point of view, very strong sources. For the WSRT telescope, for example, the detection limit in sky maps at 1420 MHz, is about 0.24 mJy (per beam and depending on the observation mode) for a 12-hour synthesis observation for a 78 kHz channel. This is about 50 dB lower than the 3C48 and 3C286 source fluxes. Transmitters and interferers at the WSRT, on the other hand, may have observed signal powers up to several orders larger than the telescope receiver noise.

4.2 The NOEMI data recorder, time sample mode

The function of the NOEMI data recorder system, shown in figure 4.4, is to record eight telescope output baseband signals simultaneously for a short du-

Source	Frequency	Ψ (Jy)	Transmitter
3C48	781	24.9	TV Lingen
”	905	22.3	GSM uplink
”	1228	17.7	GPS
”	1420.4	15.8	-
”	1480	15.2	Afristar
”	1624	14.1	Iridium
3C286	781	19.4	TV lingen
”	1090	16.7	DME
”	1158	16.2	Aeron. radionav.
”	1481	14.4	Afristar
”	1575	14.0	GPS
OH0130+621	1612		Glonass

Table 4.2. Astronomical sources used for RFI Mitigation experiments. The source fluxes, given in the third column are from [5]. The fourth column gives some of the transmitters at or close to the bands which were investigated.

ration (of order one second to a minute), and to store the data on disk or CDROM for offline processing. The goal is to record transmitters or interference for characterisation purposes, and for the study of the effectiveness of RFI and transmitter mitigation purposes. Both blanking (excision) of TDMA and frequency hopping signals, as well as spatial filtering techniques are studied. The recorder is based on an industrial PC, with a Pentium II 300 MHz board containing 384 MB SDRAM memory on its PCI backplane. The industrial PC is equipped with 6 and 28 GB hard disks, and a CD writer for storing data. Into four PCI slots, PCI212 sampling boards are plugged, each of which has two channels. Each of the PCI212 channels is equipped with an analog to digital converter of a maximum of 20 MSamples/s and RAM for 16 MSamples of data (32 MB). This allows 0.8 seconds of sampling of eight telescope (single polarisation) output baseband signals at 20MS/s. The sampling rate can be reduced with a divider down to 312.5 kSamples/s (54 sec sampling). To obtain one sampling moment, the four boards are linked together. The internal sampler clock was free running, not locked to the WSRT MASER clock systems. Via the PCI-bus the boards can be read out and stored on disk and subsequently on CDROM.

Figure 4.5 shows the functional layout of the NOEMI recording system, and the connection with the WSRT telescope. Eight of the twenty-eight telescope baseband channels are connected to the recorder using the WSRT DLB IF system. After baseband low-pass filtering, the DLB-IF baseband output power nominally is +10 dBm/10MHz in 50 Ω coax, corresponding to $\frac{1}{\sqrt{2}}\,V_{rms}$. The ADCs range can be set in factor two steps from 2V downto 0.25 V. The measured interboard

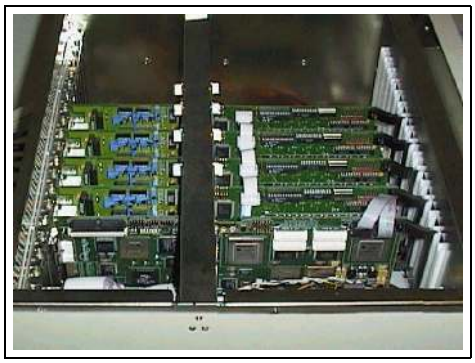


Figure 4.4. NOEMI eight channel data recorder with four dual ADC boards and with online DSP correlator board.

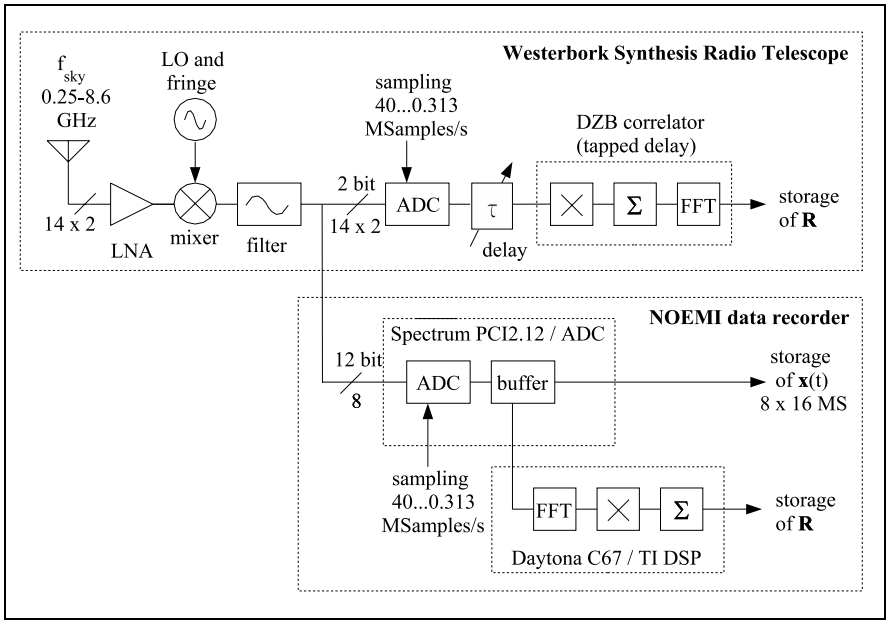


Figure 4.5. Experimental set-up consisting of the WSRT, with 14 dual polarisation telescope antenna outputs, which are amplified and downconverted to baseband. For standard WSRT imaging observations, the signals are digitised and correlated after applying a geometrical delay compensation. Eight single polarisation telescope baseband outputs are connected to the NOEMI data recorder.

crosstalk was 90 dB (power) and the onboard crosstalk 51 dB. Compared to the observed SNRs no significant problems were expected (or encountered).

4.3 NOEMI data recorder, correlator mode

Purpose

Due to limited RAM memory size, the time-sample mode of the NOEMI recorder is limited to 0.8 s of continuous data for a 10 MHz bandwidth (or 54 s for a 156 kHz bandwidth). The maximum observation duration was increased by adding a Daytona DSP board to the PC. The first reason for adding the DSP board was to demonstrate on-line blanking of intermittent RFI. The second reason was to increase the sensitivity of the system, thereby increasing the range of astronomical sources which could be observed. And finally, as the fringe frequency is low, typically 2 Hz for a 1 km baseline at 1420 MHz observing frequency, its effects on the mitigation filters could be studied for longer time stretches. The DSP board enables the recording of longer time intervals by data reduction. The reduction is reached by online correlation of the eight channels and averaging. The computed and averaged covariance matrices are stored on disk. The online correlator is used for online detection and blanking of intermittent interference, and for the recording of covariance matrices for long-term (up to half an hour) offline spatial filter processing.

System description of the online correlator

The Daytona board is a DSP board with two processing nodes. Each processing node is based on a TMS320C6701 DSP, with $4M \times 32$ bit SDRAM and $128k \times 32$ SSRAM memory. The TMS320C6701 is a floating point DSP of the Texas Instruments C600 platform, and runs at a clock speed of 166 MHz. The processor has a 32-bits interface to external SDRAM. For low-latency data exchange between the two processing nodes, the nodes share $8k \times 32$ bit dual-port RAM. The data connection between the nodes, and between the nodes and the computer PCI bus, is the Daytona local PCI bus. The data transfer between the node memory and the local PCI bus runs via a special chip, the Hurricane. The processors run at 166 MHz, which gives the processor nodes a maximum processing capacity of 1 GFLOP.

The operating system for the NOEMI system is Windows NT; software drivers for both the PCI212 and the Daytona boards have been installed. The host program, for controlling the PCI212 ADC boards and controlling the tasks for the DSPs, is written in Visual C++. The DSP code was written with the TI Code Composer Studio. Graphical visualisation of the results was done with Matlab. Calibration and detector monitoring was also done within Matlab.

Observational modes

The NOEMI DSP system uses a basic integration time of 0.5ms for the DSP system and a 10 s, 10 ms, or 0.5 ms time interval for the integrated data. It is

assumed that the 0.5 ms interval is short enough to ensure quasi-stationarity of the transmitters and RFI sources. The 10 s integrated output is the standard output for observations. The 0.5 ms output was used for system tests. The system can run continuously for hours, storing 10 s integrated covariance matrices. Due to limitations of the PCI bus speed, the effective duty cycle for most of the observations was 20%. In the online blanking mode, both the blanked data and original data were stored. In the offline mode, only the original data were stored.

4.4 NOEMI THEA tile reference antenna set-up

Purpose

The spatial filtering algorithms for RFI mitigation can be improved by incorporating directional reference antennas pointed at the interferer. The advantages of a reference antenna with high gain towards the interferer are that the interferer characteristics (signature vector) can be estimated more accurately, and that the (matrix) conditioning of spatial filters is improved. Both effects, in principle, lead to better spatial filter performance. Obviously, the reference antenna should have a higher gain towards the interferer than the telescope sidelobe gain, otherwise the noise introduced by the reference antenna will deteriorate the astronomical signals. For spatial filtering tests with a reference antenna at the WSRT , the NOEMI ASTRON THEA tile phased array reference antenna, shown in figure 4.6, was used. The tile has a directional gain toward the interferer which exceeds the WSRT sidelobe level by approximately 10 to 15 dB.

Phased array reference antenna description

The THEA tile phased array, shown in figure 4.6, is a 64-antenna element (tapered slot Vivaldi antennas) system with a two-octave wide RF bandwidth (700-1700 MHz) [133, 135]. The antenna signals are amplified with a 40 K noise temperature LNA, followed by a vector modulator for controlling the amplitudes and phases of each of the antenna elements. The 64 phase- and amplitude-adjusted signals are combined in an RF combining network, thereby creating an RF beam. The modulators and combining network are implemented two-fold, creating two independently steerable beams. The beam signals are down-converted and digitised with 40 MHz 12 bit analog to digital converters. The digital output of each beam is transmitted to the THEA back-end with a 1.2 GBit/s fibre link. For the reference antenna experiments at the WSRT site, the digitisation and digital THEA back-ends wer not used. Instead the tile down-conversion electronics was slightly adapted to match the WSRT IF output frequency ($f \approx 170$ MHz) and band direction. The parameters of the dual receiver chain are controlled with a Front-End Controller (FEC), which sets the vector-modulators and the parameters for the receiver unit. It is capable of

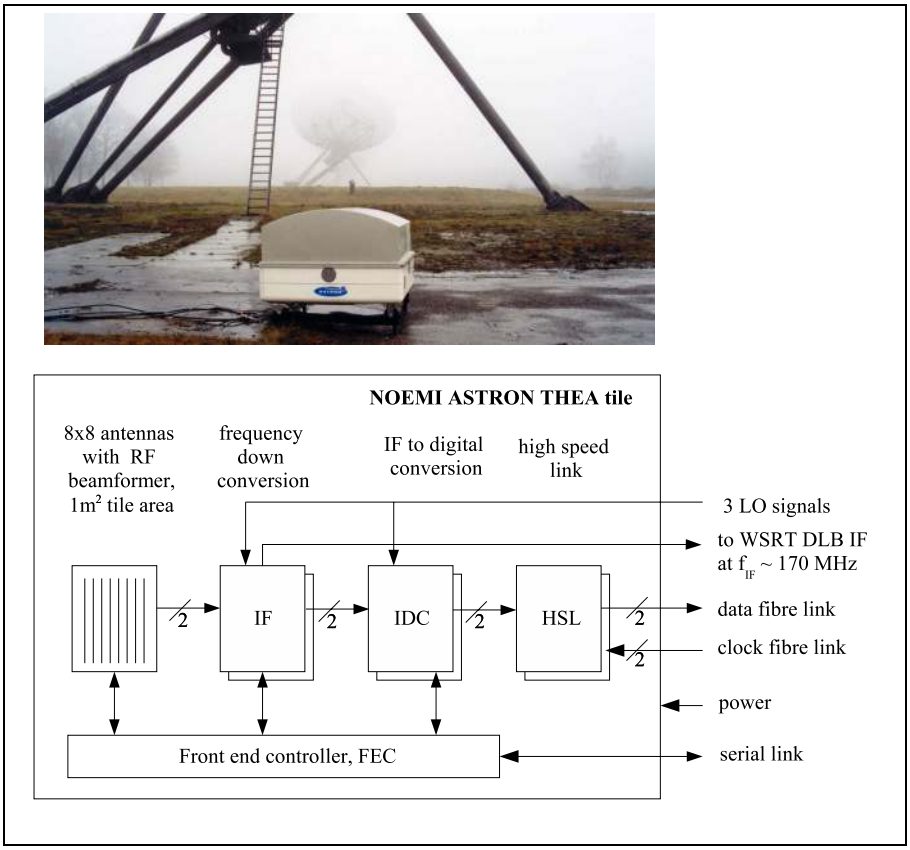


Figure 4.6. NOEMI ASTRON THEA phased-array tile at the WSRT site during initial tests (upper figure) and a schematic overview of a dual beam 64 antenna element THEA tile phased-array (lower figure).

storing 1200 pre-calculated beams, which allow fast beam steering.

Experimental configuration

Figure 4.7 shows the setup used for reference antenna experiments at the WSRT. The tile was located at a 30-m distance to telescope 6. The two tile beams, A and B, and six of the fourteen (single polarisation) telescope outputs were connected to the WSRT DLB IF system. In order to match the WSRT frequency mixing scheme, the tile (internal) IF system was slightly adapted. The DLB outputs were connected to the NOEMI data recorder as before. The tile and the NOEMI data recorder were controlled via a controller PC (NOEMI phased array controller) using serial links. One of the THEA beams (beam A, the searching beam) was used to find the direction of the interferer, the second

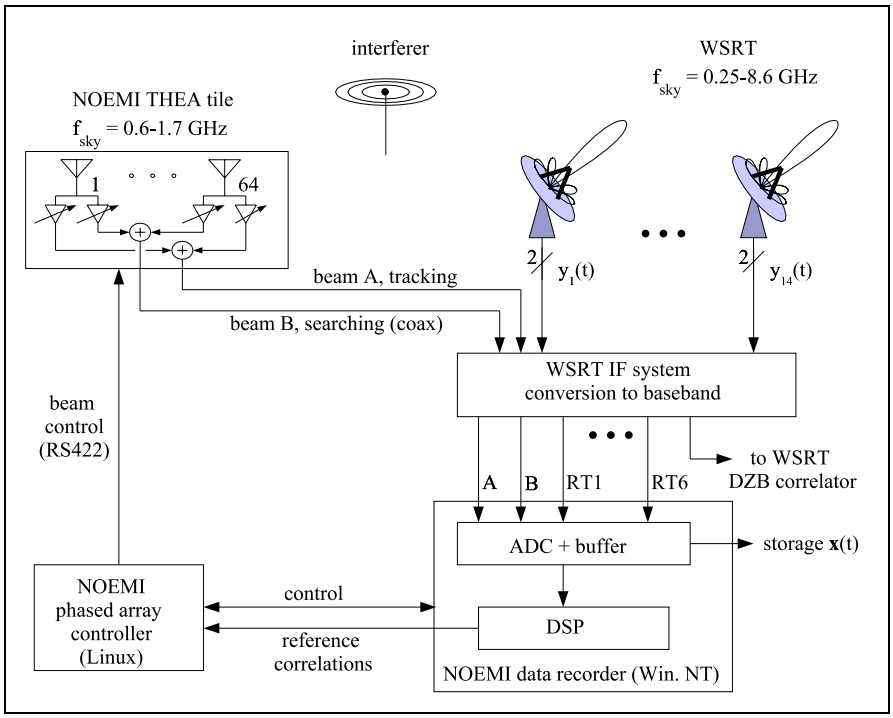


Figure 4.7. Experimental set-up used for spatial filtering experiments with a phased-array reference antenna (NOEMI THEA tile). Two array tile beams and six single polarisation WSRT telescope output signals are connected to the NOEMI data recorder. The NOEMI phased-array controller steers the array beams and synchronises the data acquisition process.

beam (beam B, tracking beam) was constantly aimed at the interferer. In the RFI searching mode of the system, the NOEMI DSP subsystem calculated the power spectra of the beams and transferred them to the control PC. Initially, a full-sky beam scan was made to find the interferer direction. During the astronomical observations, searching was done by observing eight directions close to the expected direction, and updating the RFI source tracking direction if necessary. The (relatively short) tracking and updating observations were conducted with the astronomical observations of the six telescope signals and beam A. The recorded data were used for offline spatial filter tests.

Chapter 5

Spectral-temporal blanking

5.1 Introduction

Radio frequency transmitters and interferers which are intermittent, such as Time Division Multiple Access (TDMA) wireless communication and airplane radar (DME), can be detected and removed from contaminated radio astronomical data without completely losing astronomical information. This interference detection and removal in time-frequency space is also known as excision, blanking, or flagging. Flagging in radio astronomy is traditionally done by removing time-frequency samples from post correlation data. This is an offline process and is interferometer based. The integration period of the data samples is usually of the order of seconds or minutes.

The time scale of transmission slot lengths at which the transmitter or interferer is active is in many cases much less than the one-minute or one- second level. This means that fast online detection and excision algorithms are essential to reduce the effect of interference in bands which are densely occupied with intermittent transmitters or interferers.

Examples of man-made intermittent or time-slotted signals are the European mobile telephony standard GSM (slot length ≈ 0.5 ms) or the TDMA (time-division multiple access) based mobile telephony standards IS- 54/136 in the US. Other examples are DME/SSR airplane radar (≈ 20 μs bursts), Iridium satellite transmissions (≈ 10 ms slots), and mobile communication channels where the transmission is determined by the conversation length (transmission durations of order seconds to minutes).

Existing detection and excision methods in the pre-correlation domain have a limited scope. Until recently, the most widely implemented algorithm has been a single-channel total power change detector [82], followed by a blanking of the correlator output [51]. In addition, wavelet decomposition-based detectors have been proposed [106], as well as detectors based on Cyclostationarity assumptions of the transmitter / interference [163] [118]. Also the use of quantised correlations at all correlation lags was proposed [164] to test the presence

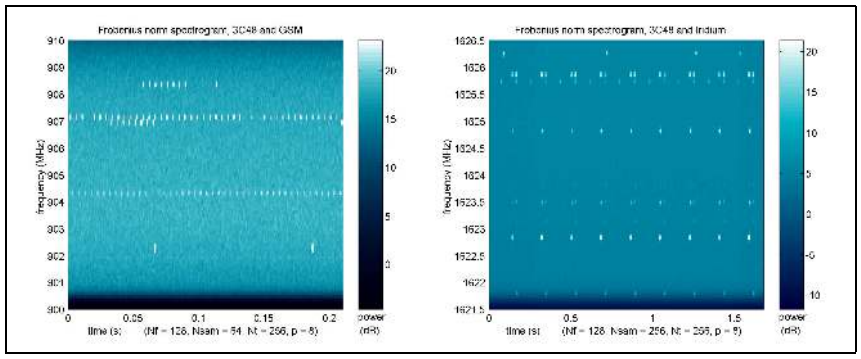


Figure 5.1. Examples of time-slotted transmissions. The left-hand figure shows observed GSM transmissions (frequency channel width 200 kHz, time slot length ≈ 0.5 ms) in an observation of 3C48 at the WSRT . The brightness in the spectrogram corresponds to the Frobenius norm of an observed eight-antenna covariance matrix. The figure to the right shows a similar spectrogram of Iridium satellite transmissions (frequency channel width 100 kHz, slot length ≈ 10 ms).

of interference. Short-pulse detection schemes were considered [115] [49] for application in wideband radio spectroscopy, neural networks were applied for detection [137], and detection based on a pdf analysis was proposed [52]. Many of these detection schemes were tested on radio telescope systems [125] [4] [51] [162]. However, these are all single-channel detectors that do not exploit the spatial properties of the interference. A detector that considered combining multiple telescopes for improved detection and blanking was proposed by [80] for low-frequency interferometry, which implied a robust data censoring method based on the temporal behaviour of the cross spectrum. This requires a large number of estimated spectra to obtain reliable robust estimates, while only two channels are used.

The aim of this chapter[1] is to introduce modern array signal processing detection and blanking techniques in the context of radio astronomy and to investigate the merits of multichannel detection and blanking algorithms at the Westerbork Synthesis Radio Telescope (WSRT). By combining cross-correlation information of a large number of sensor pairs, we can increase the detection performance significantly and also estimate the spatial signature of interferers. In essence, the approach is to compute short-term spatial correlation matrices in narrow subbands , and then to compute the eigenvalue decomposition of each of these matrices [99]. A rank-estimate based on the eigenvalues allows us to detect the number of interfering signals in each time-frequency slot, while the dominant eigenvectors give information on the spatial signature of the interferers. Alternatively, detectors based on the determinant and Frobenius norm of

[1]Parts of the results of this chapter were published in [92] [98] [15]

the correlation matrices will be discussed and applied as well.

The effectiveness of the space-time detection and blanking process will be demonstrated by applying the algorithms (offline) to data measured at the WSRT using the NOEMI data recording system. It will be shown that the galactic absorption line of 3C48, masked by a superimposed GSM transmission, was completely recovered by multichannel blanking whereas it could not be recovered by single-channel techniques.

5.2 Data model and interference detection

Many interference detection schemes exist. They differ in the amount of knowledge that we can assume about the interfering signals; if, for example, we know the signal wave form, then the optimal detector has the form of a matched filter. Extensions are possible if the wave form is known up to a few parameters such as amplitude, phase, or frequency. Usually, however, 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 the known properties of the signals such as modulation type or certain periodicities, and those methods based on statistical models with unknown parameters, leading to generalised likelihood ratio tests (GLRT). A particular example of this last is power detection.

In principle, man-made interference is expected to be statistically diffe-rent from the astronomical sources. Although this difference in statistics is very attractive for interference detection [48,163], it is a somewhat complicated technique, as it involves a-priori knowledge of the transmission modulation and coding schemes, and of the transmitted messages. For detection based on long averaging periods, the statistics-based methods are less applicable, as the averaging tends to jointly Gaussianise the interfering signals (central limit theorem).

Another distinction between interferers and astronomical signals is the difference between their spatial signature vectors. Astronomical signals enter through the main lobe of the telescopes and have a very structured (parametrically known) array response (viz. eq. (2.12, 2.30)), which is used for imaging. The interferers usually enter through the sidelobes and are either in the near or far field, leading to unstructured varying direction vectors ($\mathbf{a}^r$-vectors).

Even for fixed-location transmitters, the slow rotation of the telescopes as they track the sky will change the a^r-vector within a fraction of a second, either because of multipath fading or because the interferer moves through the highly variable sidelobe pattern. With multipath propagation, a mobile transmitter or interferer only has to move about 30 cm to create a different $\mathbf{a}^r$-vector, giving a maximum quasi-stationarity period in the order of $10 - 100$ ms for a GSM user.

Another source of nonstationarity is the fringe correction introduced at the first IF stage (cf. equation (3.22)) to compensate for the geometrical delay in synthesis telescopes. As the telescopes rotate, this introduces a time-varying phase for fixed-location interferers. This phase drift is different for each tele-

scope, and has a rate in the range of $0 - 10$ Hz for the WSRT .

For optimal multiple-antenna based interferer detection, the integration times should match the transmitted slot lengths, but not exceed the spatial quasi-stationarity time scales: order 10-100 ms for the WSRT . The chapter on spatial filtering focuses more on this stationarity issue; this chapter assumes quasi-stationary transmitters and interferers.

Before continuing with the detection schemes, the data model is revisited briefly. Ideally, the output of the correlation process produces clean estimates of the visibilities $\mathbf{R}_v$, $\mathbf{R}_v = \mathbf{ABA}^H$ (cf. ch.2), once every 10 s or so. In principle, we estimate it by correlating the array output vector: $\widehat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}\mathbf{x}_n^H$. In practice however, the estimates of $\widehat{\mathbf{R}}$ are influenced by additive noise, additive man-made transmissions, interference, and unknown antenna gains. The data model under consideration, including astronomical signals with interference and noise (cf. ch.2), is given by:

$$\mathbf{R} = \mathbf{G}\,(\mathbf{ABA}^H + \mathbf{A}^r\mathbf{B}^r\mathbf{A}^{rH})\mathbf{G}^H + \mathbf{D} \tag{5.1}$$

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 $10s$ level) by reducing the contribution of $\mathbf{A}^r\mathbf{B}^r\mathbf{A}^{rH}$ in the model above. Intermittent interference that is quasi-stationary at these time scales or longer can often be treated off-line (post-correlation flagging). In this chapter on-line interference detection and excision schemes are considered for interferers which are quasi-stationary at subsecond time scales.

5.2.1 Single-antenna temporal-spectral detection

Detection theory is based on hypothesis testing. Tested is $\mathcal{H}_0$: 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. First some particularly simple cases will be discussed which will allow analysis. Consider the single-channel case first. Assume that there is at most a single interferer, where the interfering signal is independent identically distributed (i.i.d.) Gaussian noise with unknown power σ_r^2. The background noise is i.i.d. Gaussian noise with known power σ_n^2. The astronomical signal powers are assumed to be much lower than the system noise so these signals can be ignored.

Without interferer, the observed data samples $x_n = x(t_n)$ are complex normal ($\mathcal{CN}$) distributed, with zero mean and variance σ_n^2. With an interferer, this distribution is still complex normal, but with variance $\sigma_n^2 + \sigma_r^2$. Thus, we test the hypothesis

$$\mathcal{H}_0 : \quad x_n \sim \mathcal{CN}(0, \sigma_n^2) \tag{5.2}$$

$$\mathcal{H}_1 : \quad x_n \sim \mathcal{CN}(0, \sigma_n^2 + \sigma_r^2), \quad n = (0, \cdots, N-1) \tag{5.3}$$

We assume that we have available N samples $\{x_n\}$, collected in a vector $\mathbf{x} = [x_1, \cdots, x_N]^t$. This is a rather standard problem in detection theory (cf. [82,167]

for an introduction). A Neyman-Pearson detector selects $\mathcal{H}_1$ if the likelihood ratio,

$$\mathcal{L}(\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} \tag{5.4}$$

exceeds a threshold, where $p(\mathbf{x}|\mathcal{H})$ denotes the probability density function of $\mathbf{x}$ under the hypothesis $\mathcal{H}$. 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 this case, the Neyman-Pearson detector is simplified to comparing the total received power to a threshold γ, deciding $\mathcal{H}_1$ if the test statistic

$$T(\mathbf{x}) = \frac{1}{\sigma_n^2} \sum_{n=1}^{N} |x_n|^2 > \gamma \tag{5.5}$$

The Neyman-Pearson detector can be used for intermittent interference, when the a-priori probabilities are not known and also when the costs related to the detection test outcomes are not known or specified. This makes the Neyman-Pearson generally applicable, that is, more or less independent of the intermittent interference and telescope characteristics. Under the above assumptions, closed form expressions for the probability of a false alarm and the probability of detection can be obtained. For this, recall that the sum of squares of N real i.i.d. zero-mean unit-variance Gaussian random variables has a χ^2 distribution with N degrees of freedom (dof). Since we have complex samples x_n, $T(\mathbf{x})$ is the sum-square of $2N$ real variables. Indeed, let $x_n = a_n + \imath b_n$, where a_n and b_n are random real variables with variance $\frac{1}{2}\sigma_n^2$. Define c_n with variance $\frac{1}{2}\sigma_n^2$ by

$$c_n = \begin{cases} a_n & 1 \leq n \leq N \\ b_n & N+1 \leq n \leq 2N \end{cases} \tag{5.6}$$

then the summation of the $|x_n|^2$ terms can be expressed in terms of c_n:

$$\sum_{n=1}^{N} |x_n|^2 = \sum_{n=1}^{N} |a_n + \imath b_n|^2 = \sum_{n=1}^{2N} c_n^2 \tag{5.7}$$

The probability of a false alarm is therefore given by

$$P_{FA} = P\{T(\mathbf{x}) > \gamma | \mathcal{H}_0\} = P\left\{ \frac{1}{\sigma_n^2} \sum_{n=1}^{2N} c_n^2 > \gamma \mid \mathcal{H}_0 \right\} \tag{5.8}$$

In the equation above, the term before the inequality sign, needs to have a unit variance in order to be expressable in terms of χ^2 distributions (otherwise Gamma probability density functions are needed). This is accomplished by multiplying both sides of the inequality by 2:

$$P_{FA} = P\left\{ \frac{1}{\frac{1}{2}\sigma_n^2} \sum_{n=1}^{2N} c_n^2 > 2\gamma \mid \mathcal{H}_0 \right\} = 1 - \chi^2_{cdf,2N}(2\gamma) = Q_{\chi^2_{2N}}(2\gamma) \tag{5.9}$$

where $Q_{\chi^2_{2N}}(.)$ is the tail probability of a χ^2 random variable with $2N$ degrees of freedom. It has a closed-form expression (see Kay 1998), and its inverse is known in terms of the inverse Gamma function, and allows us to select γ to obtain a desired level of false alarm. Similarly, the probability of detection of interference at this threshold γ is given by, see figure 5.2:

$$\begin{aligned} P_D &= P\{T(\mathbf{x}) > \gamma : \mathcal{H}_1\} && (5.10)\\ &= P\left\{\frac{1}{\sigma_n^2}\sum_{n=1}^{N}|x_n|^2 > \gamma \mid \mathcal{H}_1\right\} && (5.11)\end{aligned}$$

Let $INR = \sigma_r^2/\sigma_n^2$ be the interference-to-noise ratio, and normalise the term left of the inequality sign in the expression above by multiplying both sides with $\sigma_n^2/(\sigma_n^2+\sigma_r^2)$. This yields:

$$\begin{aligned} P_D &= P\left\{\frac{2}{\sigma_n^2+\sigma_r^2}\sum_{n=1}^{2N}c_n^2 > \frac{2\gamma}{1+INR} \mid \mathcal{H}_1\right\} && (5.12)\\ &= 1-\chi^2_{cdf,2N}\left(\frac{2\gamma}{1+INR}\right) = Q_{\chi^2_{2N}}\left(\frac{2\gamma}{1+INR}\right) && (5.13)\end{aligned}$$

As an illustration, figure 5.2 shows the $\chi^2_{pdf,N}$ probability density functions for the cases without interference and with interference. The probability of a false alarm and the probability of detection are given by the grey areas in the figure.

5.2.2 Multiple-antenna spatial-temporal detection

A significant performance improvement is possible with a multichannel detector. To illustrate this, we assume again the simple case with at most a single narrowband Gaussian interferer, with a known spatial signature vector $\mathbf{a}^r$ in white Gaussian noise. The source power of the interference is denoted by σ_r^2; to normalise the receiver gain we set $||\mathbf{a}^r||^2 = (\mathbf{a}^r)^H\mathbf{a}^r = p$, where p is the number of antennas. Without interference, the data vectors $\mathbf{x}_m$ are complex normal distributed with zero mean and covariance matrix $\sigma_n^2\mathbf{I}$. With a single interferer, the covariance matrix becomes $\mathbf{R} = \mathcal{E}\{\mathbf{x}_m\mathbf{x}_m^H\} = \sigma_r^2\mathbf{a}^r(\mathbf{a}^r)^H + \sigma_n^2\mathbf{I}$. Thus,

$$\begin{aligned} \mathcal{H}_0: &\quad \mathbf{x}_m \sim \mathcal{CN}(0,\sigma_n^2\mathbf{I}) && (5.14)\\ \mathcal{H}_1: &\quad \mathbf{x}_m \sim \mathcal{CN}(0,\sigma_n^2\mathbf{I}+\sigma_r^2\mathbf{a}^r(\mathbf{a}^r)^H), \quad m=(0,\cdots,N-1) && (5.15)\end{aligned}$$

The Neyman-Pearson detector based on the data matrix $\mathbf{X} = [\mathbf{x}1,\cdots,\mathbf{x}_N]$ considers the estimated data covariance matrix

$$\widehat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n^H \qquad (5.16)$$

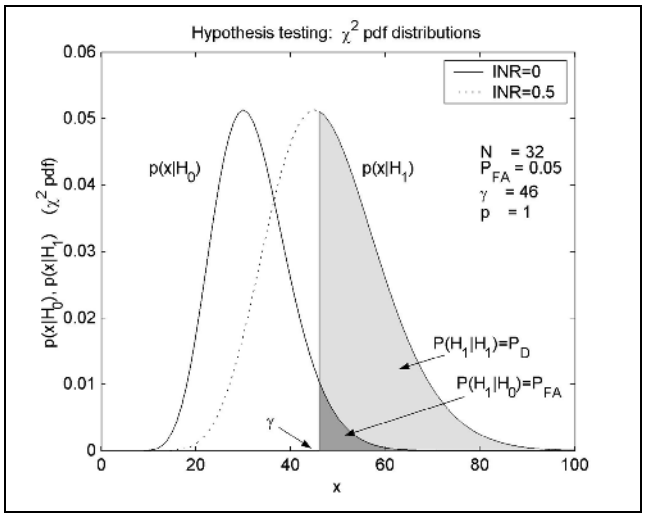


Figure 5.2. Probability density functions for the cases without interference (INR=0) and with interference (INR=0.5). Only one interferer is assumed to be present. Both the interferer and the system noise are assumed to be i.i.d. Gaussian. The probability of false alarm, assuming a threshold γ, is given by the dark grey area; the probability of detection by the light grey area.

and is given by [82]

$$T(\mathbf{X}) = \frac{N}{\sigma_n^2} \frac{(\mathbf{a}^r)^H \widehat{\mathbf{R}} \mathbf{a}^r}{(\mathbf{a}^r)^H \mathbf{a}^r} \underset{<\mathcal{H}_0}{>\mathcal{H}_1} \gamma \tag{5.17}$$

Note that $T(\mathbf{X}) \approx \frac{N}{\sigma_n^2}\lambda_1$ if $\mathbf{a}^r$ is estimated by $\mathbf{u}_1$, where λ_1 and $\mathbf{u}_1$ are the largest eigenvalue and the corresponding eigenvector of $\widehat{\mathbf{R}}$. This test is recognised as a matched spatial filter detector; essentially, we compare the received energy in the direction $\mathbf{a}^r$ of the interferer to σ_n^2. If we define y_n to be the output of the matched beamformer[2] in the direction of $\mathbf{x}_n$, $y_n = ||\mathbf{a}^r||^{-1}(\mathbf{a}^r)^H \mathbf{x}_n$, then

$$\mathcal{H}_0 : \quad y_n \sim \mathcal{CN}(0, \sigma_n^2) \tag{5.18}$$

$$\mathcal{H}_1 : \quad y_n \sim \mathcal{CN}(0, \sigma_n^2 + p\sigma_r^2), \quad n = (1, \cdots, N) \tag{5.19}$$

and it is seen that taking the same threshold as in the case of the single channel will provide the same false alarm probability as before :

$$P_{FA} = P\{T(\mathbf{x}) > \gamma | \mathcal{H}_0\} = Q_{\chi^2_{2N}}(2\gamma) \tag{5.20}$$

[2]A matched beamformer $y_n = \mathbf{w}^H \mathbf{x}_n$ for the model $\mathbf{x}_n = \mathbf{a}^r s_n + \mathbf{n}$, where s_n is the impinging signal and $\mathbf{n}$ is the noise vector, can be found by maximising $|y_n|^2$ while keeping the weight vector $\mathbf{w}$ constrained to one: $||\mathbf{w}|| = 1$ [82] [88] [104]. The weighting vector then becomes $\mathbf{w} = \mathbf{a}^r/||\mathbf{a}^r||$. Also: $\mathcal{E}\{|y_n|^2\} = \mathcal{E}\{\mathbf{w}^H \mathbf{x}_n (\mathbf{w}\mathbf{x}_n)^H\} = ((\mathbf{a}^r)^H/||\mathbf{a}^r||)\mathcal{E}\{\mathbf{x}_n \mathbf{x}_n^H\}(\mathbf{a}^r/||\mathbf{a}^r||) = \sigma_n^2$

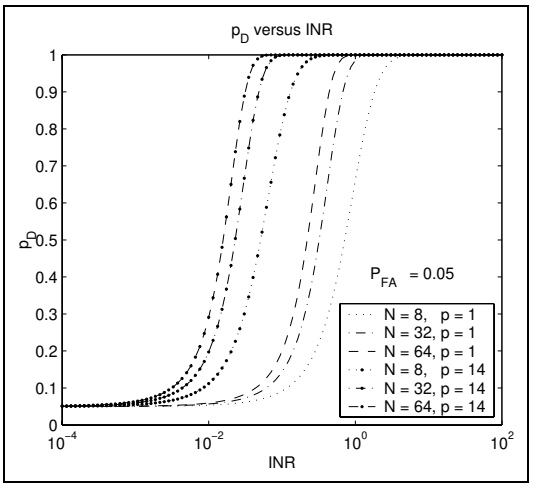


Figure 5.3. Probability of detection P_D versus INR for N = 8,32,64, and false alarm rate $P_{FA} = 0.05$. The three curves to the right correspond to the single telescope case, the curves to the left correspond to an array of 14 telescopes.

However, the probability of detection is now given by

$$P_D = P\{T(\mathbf{x}) > \gamma | \mathcal{H}_1\} \tag{5.21}$$

$$= Q_{\chi^2_{2M}}\left(\frac{2\gamma}{1+pINR}\right) \tag{5.22}$$

Figure 5.3 presents the probabilities of detection as a function of interference-to-noise ratio for a single-channel and for $p = 14$ channels. We have selected a threshold such that $P_{FA} = 0.05$, 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, p, which is for $p = 14$: 11.5 dB.

5.2.3 Residual interference after blanking

Consider a TDMA signal: an interferer which is periodically active in a fraction β of the time, see figure 5.4. 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 αN samples $\mathbf{x}_N$, where we take $\alpha > 1$. Let as before σ_r^2 denote the power of a single sample of the interferer when it is present. Since the interfering slots need not be synchronised 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

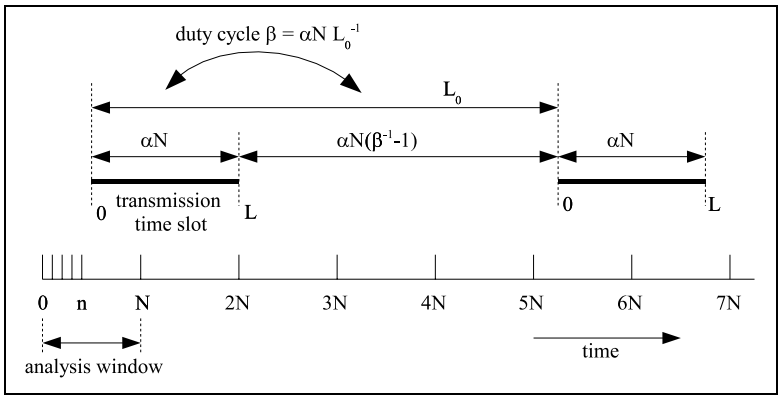


Figure 5.4. Temporal occupancy figure for a transmitter/interferer with period L_0, slot length $L = \alpha N$ samples, power σ_s^2 (when on), and duty cycle $\beta = L/L_0$

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 N can be computed in closed form as

$$p(I) = \begin{cases} (1-(\alpha+1)\alpha^{-1}\beta)\delta(I), & I = 0 \\ 2\beta(I_{max}\alpha)^{-1}, & 0 < I < I_{max} \\ (\alpha-1)\alpha^{-1}\beta\delta(I-I_{max}), & I = I_{max} \end{cases} \tag{5.23}$$

where $\delta(.)$ is the Dirac delta function. For example, for an interferer of strength σ_r^2 per sample when it is on, the maximal average interference power per sample is obviously σ_r^2, when all samples are contaminated. The probability of this is $(\alpha-1)\alpha^{-1}\beta$. Power densities less than σ_r^2 occur with a uniform distribution for analysis windows that are only partly corrupted, at the edges of the interference slot. Define the average interference power per sample before detection by I_{before}, the average interference power per sample after detection and blanking by I_{res}, and the fraction of the number of samples kept after detection and blanking by N_{res}. Finally, define the average interferer power per (kept) sample by I_{after}:

$$I_{before} = \int_0^{\sigma_r^2} Ip(I)dI = \beta\sigma_r^2 \tag{5.24}$$

$$I_{res} = \int_0^{\sigma_r^2} I(1-p(I))p(I)dI \tag{5.25}$$

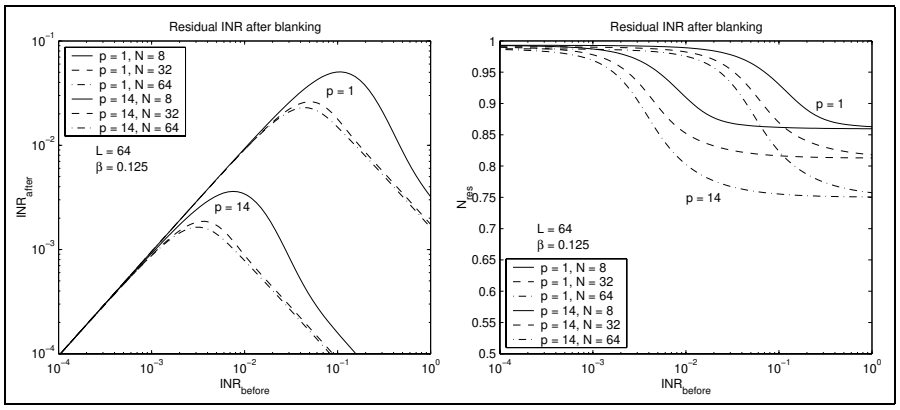


Figure 5.5. Effective INR after detection and excision versus INR before detection (left), and a fraction of remaining samples after detection and excision (right)

$$N_{res} = \int_0^{\sigma_r^2} (1 - P_D(I))p(I)dI \tag{5.26}$$

$$I_{after} = I_{res}/N_{res} \tag{5.27}$$

As the functions defined above are based on a system with unit noise variance, INR values before $INR_{before} = \sigma_r^2/\sigma_n^2$ and after detection and excisions $\mathrm{INR}_{\mathrm{after}}$ are defined, in which the noise may have a non-unity variance. Figure 5.5 shows the dependence of the residual interference-to-noise ratio as a function of N (the number of samples in an analysis block), for an interferer of length $L = 64$ subband 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 because of 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 5.5, the effect of using an array is to shift the graphs of the probability of detection to the left by the array gain; for the 14-channel detector for example, 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 5.5 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 N. Thus, we would make this window rather small, so that 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 smaller. Further study of this model appeared in [92].

The (left-hand) curves in figure 5.5 show a maximum of $\mathrm{INR}_{\mathrm{after}}$ at a certain level. For low $\mathrm{INR}_{\mathrm{before}}$, obviously, $\mathrm{INR}_{\mathrm{after}}$ is low also. For high $\mathrm{INR}_{\mathrm{before}}$, the interferer can be detected easily, and can be efficiently excised. The $\mathrm{INR}_{\mathrm{after}}$ however, cannot be zero because some of the temporal windows of the detector will have only a small overlap with the interferer "on" time slot. The interference energy in such a time slot may be below the detection threshold. The (approximate) maxima of the detection curves in figure 5.5 can be estimated as follows. Suppose the false alarm rate is set to 0.05. This yields, according to formula (5.9), a certain threshold value of γ. Now assume that at the maximum, the probability of detection is 0.5, then, using the value of γ derived above, the INR follows from formula (5.22). A straightforward combination of the formulas leads to the following expression for the (approximate) maximum residual interference after detection and excision:

$$INR_{res} \lessapprox \frac{\beta}{p}\left(\frac{\chi^{2^{inv}}_{cdf,2N}(1-P_{FA})}{\chi^{2^{inv}}_{cdf,2N}(1-P_{D_{max}})}-1\right) \tag{5.28}$$

An alternative approach to estimate the maximum residual interference after detection and excision is based on an estimate of the variance of the eigenvalues of the covariance matrix. Let $\mathbf{R}$ be a covariance matrix of a system with only i.i.d. noise, with a diagonal noise matrix $\mathbf{D} = \mathrm{diag}({\sigma_{n_1}}^2, \cdots, {\sigma_{n_p}}^2)$: $\mathbf{R} = \mathbf{D}$. The eigenvalue decomposition of $\mathbf{R}$ yields: $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, or $\mathbf{\Lambda} = \mathbf{U}^H\mathbf{D}\mathbf{U}$. The variance of $\mathbf{\Lambda}$ is then given by[3]:

$$\mathrm{var}(\mathbf{\Lambda}) = \frac{1}{N}\mathbf{\Lambda}^2 \tag{5.29}$$

Assume that the noise variance of all telescope are equal, ${\sigma_{n_i}}^2 = \sigma_n^2$ $(\forall i)$, and assume there is a single interferer with power σ_r^2, then $\lambda_1 = \sigma_n^2 + \sigma_r^2$. As a detection threshold, a value of, say, 2 times the standard deviation of the largest eigenvalue[4] for the interference-free case $(\lambda = \sigma_n^2)$can be used: $\gamma = 2\,\mathrm{std}(\lambda_1) = 2\sigma_n^2/\sqrt{N}$. When a multichannel detector, based on a largest eigenvalue λ_1 is used, the maximum residual interference after detection and excision is given

[3]From $\mathrm{cov}(\widehat{\mathbf{R}}) = \frac{1}{N}\bar{\mathbf{R}} \otimes \mathbf{R}$ it can be derived that $\mathrm{var}(\widehat{\mathbf{D}}) = \frac{1}{N}\mathrm{diag}(\mathbf{D})\mathrm{diag}(\mathbf{D})^t$, and $\mathrm{var}(\widehat{\mathbf{D}}) = \frac{1}{N}\mathbf{D}^2$. But the eigenvalues of $\mathbf{R} = \mathbf{D}$ are equal to $\mathbf{D}$ itself, or $\mathrm{var}(\mathbf{\Lambda}) = \frac{1}{N}\mathbf{\Lambda}^2$.

[4]This more or less arbitrary factor 2 can be related to the probability of a false alarm as before.

by

$$INR_{res} \lessapprox \frac{2\beta\sigma_n^2}{p\sqrt{N}} \tag{5.30}$$

The factor p in the denominator is a result of the array gain. Obviously, this formula can be refined by considering unequal noise variances, but this refinement will not be pursued here. The two results of the formulas (5.28) and (5.30) are shown in figure 5.6 (right), and are compared to the theoretical $INR_{after} - INR_{before}$ curves calculated using equations (5.24) to (5.27). The estimated maxima of the curves (left) match reasonably well with the predicted ones, and can be considered as upper bounds. Note that the outcome of equation (5.28) depends on the particular choice of $P_{D_{max}}$.

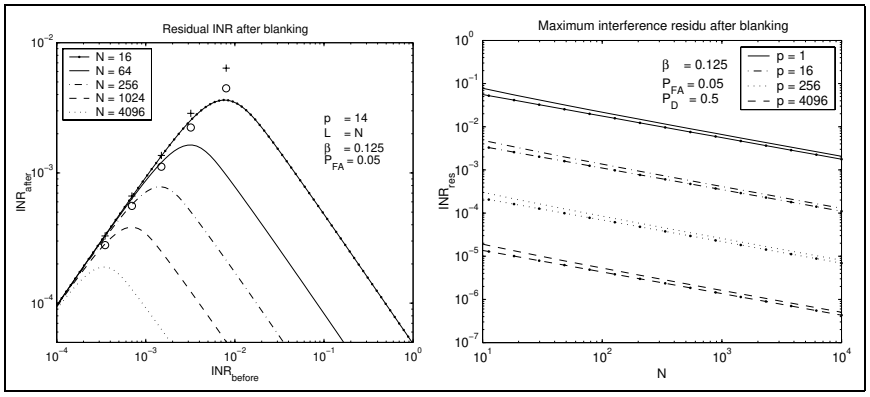


Figure 5.6. Effective INR after detection and excision versus INR before detection (left), and upper limit for the residual interference after detection and excision as a function of the number of samples (right). There are four groups of two curves. Each group corresponds to a specific number of antenna elements (p). The upper curves of the groups correspond to a detector based on formula (5.28); the lower curves correspond to a detector based on formula (5.30). The plusses in the left-hand figure correspond to the maxima according to equation (5.28), the circles correspond to equation (5.30)

5.2.4 Multiple-antenna detection scenarios

Several detectors, described in this section, will be used for excision experiments. These detectors differ in the way the observed covariance matrices $\widehat{\mathbf{R}}$ are processed.

Known spatial signature

If a telescope array is calibrated, and if the location or direction of an interferer is known, then the spatial signature of the interferer is known as well. In that case, the matched spatial filter detector as in (5.17) can be used with

known $\mathbf{a}^r$.

Unknown spatial signature, eigenvalue test

If the interferer spatial signature is not known, the test statistic in (5.17) takes the form of a matched spatial filter

$$T(\mathbf{X}) = \frac{N}{\sigma_n^2}\hat{\lambda}_1 \underset{<}{\overset{>}{}}^{\mathcal{H}_1}_{\mathcal{H}_0} \gamma \tag{5.31}$$

as $\hat{\mathbf{a}}^r\widehat{\mathbf{R}}(\hat{\mathbf{a}}^r)^H = \hat{\lambda}_1$ for $\hat{\mathbf{a}}^r = \hat{\mathbf{u}}_1$, where $\hat{\lambda}_1$ and $\hat{\mathbf{u}}_1$ are the largest eigenvalue and corresponding eigenvector of $\widehat{\mathbf{R}}$. The threshold γ can be defined as is done in sections 5.2.1 and 5.2.2. A simple alternative for the threshold is the use of the maximum eigenvalue estimate as defined in (3.48) and [42]

$$\gamma = \sigma^2\left(1 + \sqrt{\frac{p}{N}}\right)^2 \tag{5.32}$$

The threshold γ can be given a small offset to change the probability of a false alarm. The offset value can be found by simulation.

If there is more than one interferer, a threshold test for the eigenvalues can be used as well. A complication is that in this case the number of interferers must be estimated, as well as thresholds for each of the eigenvalues. The sum of all eigenvalues could be used for detection as well. The complication here is also in finding suitable threshold levels. Alternative approaches will be described next.

Unknown spatial signature, GLRT detector

In case we only have an estimate $\widehat{\mathbf{R}}$ based on a finite number of samples N 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_n^2: without interference, all eigenvalues are asymptotically equal to σ_n^2. If the noise power σ_n^2 is known, we can apply the (generalised) likelihood ratio test (GLRT), which leads to a method for testing the null hypothesis that $\sigma_n^{-2}\widehat{\mathbf{R}} = \mathbf{I}$ (no interference) [25]. The GLRT leads to a test statistic given by [98, 97]

$$T(\mathbf{X}) = -Np\log\prod_{i=1}^{p}\frac{\hat{\lambda_i}}{\sigma_n^2} = -Np\log\frac{|\widehat{\mathbf{R}}|}{\sigma_n^2} \tag{5.33}$$

where $\hat{\lambda_i}$ is the i^{th} eigenvalue of $\widehat{\mathbf{R}}$ and we detect an interferer if $T(\mathbf{X}) > \gamma$. This basically tests whether all eigenvalues are equal to σ_n^2 with a certain confidence. In the no-interference case, one can show that

$$T(\mathbf{X}) \sim \chi^2_{(p+1)(p-2)} \tag{5.34}$$

This makes it possible to select the value of γ to achieve a desired false alarm rate.

If the noise power is unknown as well, the Minimum Description Length (MDL) detector [159] can be used. In this case, rather than setting a threshold based on the asymptotic distribution of the LRT, the correct model order which minimises the description length of the data is estimated [98]. This detector is simple to implement, but a disadvantage is that the false alarm rate is not known and not fixed. This detector will not be further pursued in this thesis.

Ad hoc detectors

The eigenvalue test and the GLRT detector described above require eigenvalue computations, which need $\mathcal{O}(p^3)$ operations. An ad-hoc detector which has a lower computational load, $\mathcal{O}(p^2)$ operations, is the Frobenius norm detector. This detector is applied to pre-whitened covariance matrices $\widehat{\mathbf{R}}_w$ with test statistic

$$T(\mathbf{X}) = ||\widehat{\mathbf{R}}_w||_F^2 \underset{<}{\overset{>}{\mathcal{H}_0}}{}^{\mathcal{H}_1} \gamma \tag{5.35}$$

Note that $||\widehat{\mathbf{R}}||_F^2 = \sum \lambda_i^2$. In [97] it is shown that this detector is equivalent to the GLRT detector for weak interferer signals, up to the third order.

5.3 Experimental results

To test the blanking algorithms, the WSRT antennas were attached to the NOEMI multichannel data recorder which can collect baseband data at 0.313 to 20 MHz sampling rates. The system was used both in the sampling and the off-line processing mode, and in the on-line real-time processing mode. In this way, a variety of actual intermittent transmitter and interfering signals were processed.

In the following sections, the performance of the blanking algorithms will be demonstrated using experimental data. First, two examples will be shown of the eigenstructure of observed datasets with Iridium and GSM tranmsmitters. Next the estimates will be shown of the probability of detection using single and multichannel detectors, using observational data of GSM mobile phone and Iridium satellite TDMA transmissions. Thirdly, an experiment will be shown in which observed GSM transmissions were added (offline) to observational data of the galactic absorption of the astronomical source 3C48. The detection and blanking results were quite good, as it was possible to recover the 3C48 absorption line which was completely masked by the GSM interference. In addition, several detectors were applied to observations of the source 3C48, which were affected by Iridium TDMA transmissions. An illustration will be shown of the eigenstructure of an astronomical intermittent source: a radio pulsar. Finally, a result will be shown of the online demo blanking system with which online detection and blanking measurements were carried out.

5.3.1 Eigenstructure examples

The covariance matrix eigenvalue structure can be nicely illustrated on data collected at the WSRT . From the observed data a sequence of short-term (ms scale) cross-spectral matrices was computed, and an eigenvalue decomposition was applied. Figure 5.7, left, shows an observation of 3C48 with Iridium satellite

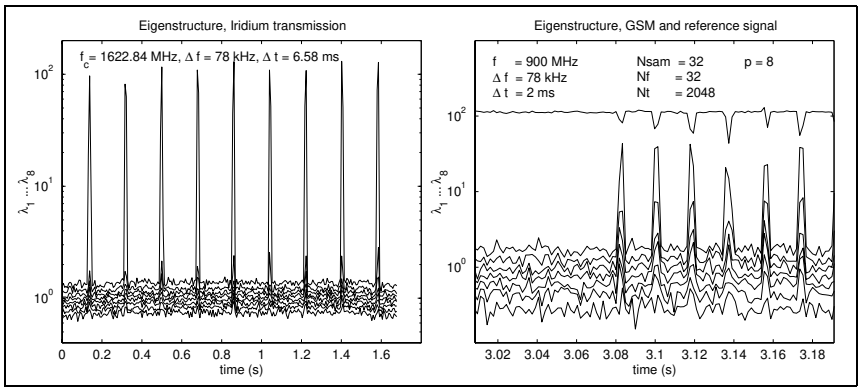


Figure 5.7. Single Iridium TDMA satellite transmitter (left) and multiple interferers: GSM TDMA transmissions and a strong continuous source (right).

transmissions. It is an example of the time evolution of the eigenvalues of the covariance matrices. There is only one dominant interferer, as can be seen from the fact that the largest eigenvalue is 20 dB larger than the remaining $p-1$ eigenvalues. This means that either there are no other interferers or that the interferers have signal powers less than $\approx 1\%$ of the observed Iridium transmitted power. The 1% rise of the $p-1$ eigenvalues can also be explained by other causes. The finite bandwidth effect (cf. section 3.3.1) for the used 78 kHz bandwidth yields a second eigenvalue which lies approximately 20 dB below the largest eigenvalue. Moreover, the dataset also contains signals from the astronomical source 3C48, which has a power of about 5% of the channel noise.

The figure to the right shows an observation of a weak GSM uplink transmission, together with a strong continuous narrowband continuous wave (CW) signal that leaked in from a local oscillator of the WSRT receiver system. The largest eigenvalue is due to the CW signal and is always present. The GSM interference is intermittent: at time intervals where it is present the number of large eigenvalues increases to two. The remaining eigenvalues are slightly above the noise floor. The reasons for this effect are the same as in the case of the Iridium transmissions.

Note that it is crucial that the noise is spatially white. For coloured noise, an extension (whitening) is possible but we have to know the colouring.

5.3.2 Comparison of single-channel and multiple-channel detectors: influence on p_D

In this section, single-channel and multichannel detectors will be applied to astronomical datasets which are affected by intermittent transmissions or interference.

The first experiment is to verify the derived theoretical $p(D_1|H_1)$ curves experimentally, using datasets of 3C48 with GSM and Iridium transmissions. The detection probability estimation approach was to count the number of detected transmissions, while gradually increasing the noise level. This was done adding spatially white noise to the GSM and Iridium datasets during offline processing. Figure 5.8 shows the autocorrelations of eight WSRT telescopes for both datasets. The communications transmissions enter the WSRT telescopes via sidelobes. The figure shows that the sidelobe gains of the telescopes differ by a factor of about 10. The fast time variability of the GSM signals are caused by multipath fading close to the source. The 0.5-second time variability of the GSM and Iridium transmissions are caused by changes in the sidelobe level due to telescope and source motion, and due to multipath reflections close to the telescopes. The eigenvalue decomposition of both datasets, shown in figure 5.7, indicates that the dominant eigenvalues are $\approx$ 10 to 20 dB above the noise floor, which means that nearly all interfering power is concentrated in the largest eigenvalue.

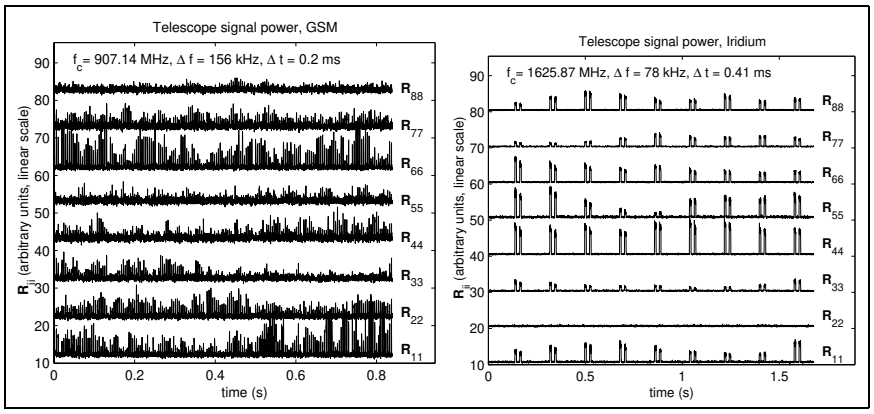


Figure 5.8. Autocorrelations of 3C48 observations with TDMA GSM transmissions (0.4 ms duration bursts) for eight WSRT telescopes (left), and with Iridium satellite transmissions (9 ms duration bursts), also for eight WSRT telescopes (right).

Before applying detectors to the datasets, the observations were pre-whitened. The whitening was done by using the diagonal of the covariance matrix, dis-

carding the sections where there were GSM transmissions present:

$$\mathbf{R}_w = (\mathbf{R} \odot \mathbf{I})^{-\frac{1}{2}} \, \mathbf{R} \, (\mathbf{R} \odot \mathbf{I})^{-\frac{1}{2}} \tag{5.36}$$

As before, the covariance matrix $\mathbf{R}$ is composed of an interference power matrix $\mathbf{B}_r$, interference array response $\mathbf{A}_r$, astronomical source power matrix $\mathbf{B}_s$, source array response $\mathbf{A}_s$, and noise power matrix $\mathbf{D}$

$$\mathbf{R} = \mathbf{A}_r \mathbf{B}_r (\mathbf{A}_r)^H + \mathbf{A}_s \mathbf{B}_s \mathbf{A}_s^H + \mathbf{D} \tag{5.37}$$

Without loss of generality it is assumed that the electronic gains $\mathbf{G}$ are "absorbed" in the spatial signature matrices $\mathbf{A}_r$ and $\mathbf{A}_s$. As the flux of 3C48 is small compared to the channel noise, $\mathbf{R} \odot \mathbf{I}$ is a good estimate of $\mathbf{D}$ in time-frequency bins without interference, and as a result:

$$\mathbf{R}_w \approx \mathbf{D}^{-\frac{1}{2}} \mathbf{A}_r \mathbf{B}_r \mathbf{A}_r^H \mathbf{D}^{-\frac{1}{2}} + \mathbf{I} \tag{5.38}$$

The INR is defined, as before, by the ratio of the transmitter power and the noise power, but due to the different sidelobe levels and multipath effects, the received transmitter power varies from telescope to telescope by about 10 dB. The INR is therefore defined by $INR \equiv \hat{\lambda_1}/p$. Here, $\hat{\lambda}_1$, the largest eigenvalue estimate, is obtained from an eigenvalue decomposition of $\widehat{\mathbf{R}}_w$.

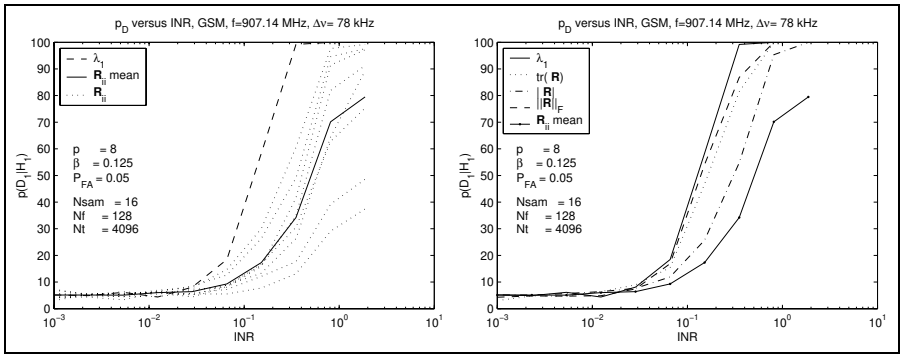


Figure 5.9. Estimated $p(D_1|H_1)$ curves for observed GSM transmissions. The figure to the left shows single-channel detectors based on the telescope autocorrelations $\mathbf{R}_{ii}$ $(i = 1 \cdots 8)$, and on a multichannel detector based on the largest eigenvalue λ_1. The figure to the right shows the estimated $p(D_1|H_1)$ curves for four multichannel detectors based on respectively λ_1, tr($\mathbf{R}$), $|\mathbf{R}|$, and $||\mathbf{R}||_F$. The figure also shows the mean of the single channel results, $(\mathbf{R}_{ii})_{mean}$. The detectors were configured such that $p_{FA} = 0.05$.

The single and multichannel detectors were applied to $\widehat{\mathbf{R}}_w$. Figure 5.9 shows the estimated $p(D_1|H_1)$ curves as a function of INR for the GSM dataset. The

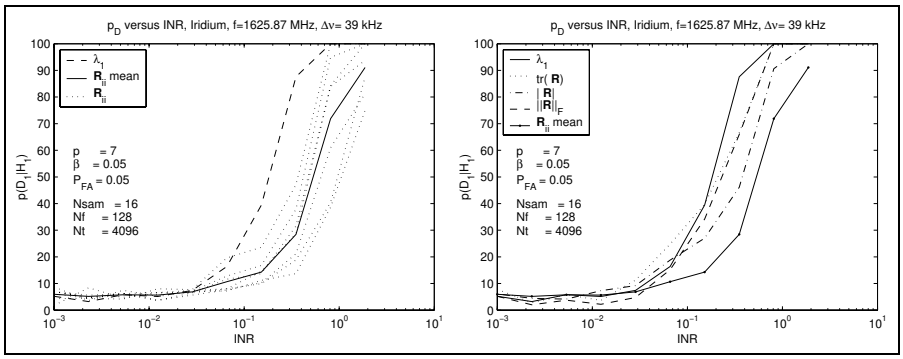


Figure 5.10. Estimated $p(D_1|H_1)$ curves for observed Iridium satellite transmissions. The applied processing and results are similar to the GSM observations shown in figure 5.9.

left-hand figure shows a detector based on the largest eigenvalue for a false alarm rate of 0.05. The single channel detectors based on the $(\mathbf{R}_w)_{ii}$ estimates are also shown in the figure. Depicted by $(\mathbf{R}_{ii})_{mean}$ is the mean of the single-detector curves. For $p = 1$ and for $p = 8$ (assuming $N = 16$ samples and a false alarm rate of 0.05) the INR corresponding to a theoretical estimated $p(D_1|H_1) = 0.5$ is 0.048 and 0.06 respectively. For $p = 1$ this is close to the experimental value in figure 5.9. For $p = 8$ this number is about a factor 2 lower than the theoretical value. This is probably caused by the fact that the sidelobe gain of the WSRT telescopes, for the observatons under consideration, differ by more than 10 dB. However, the figures clearly show the advantage of the multichannel detection scheme over the single-channel approach.

Figure 5.9 (right) shows similar results as the left-hand figure, but now detectors based on $\mathrm{tr}(\mathbf{R}) = \sum_i \lambda_i$, $|\mathbf{R}| = \prod_i \lambda_i$, and $||\mathbf{R}||_F^2 = \sum_i \lambda_i^2$ are added. All multichannel detectors, except the determinant detector, have comparable performance. This means that if computational efficiency is an issue for certain applications, the detector based on the Frobenius norm is preferred. Calculation of the Frobenius norm requires multiplicative operations in the order of p^2; full-rank eigenstructure decompositions require multiplicative operations in the order of p^3. The determinant detector, which for low INR is equal to the Frobenius detector up to third order [97], apparently does not perform as well as the Frobenius norm detector for higher INR values.

Figure 5.10 shows $p(D_1|H_1)$ the results of observations with Iridium transmissions. The performance of the detectors in this case is similar to the detector performance in the observations with GSM transmissions.

The GSM transmissions in the datasets discussed before showed fast, order 0.1 s, fluctuations (fast fading). This suggests that these fluctuations are caused by a cluster of scattering objects close to a moving transmitter giving rise to many multipaths . This can be verified by determining the distribution of the

magnitude of the observed complex samples. If there are many reflections on the scattering objects nearby a moving source, then it can be shown [153] that the distribution of the amplitude $|x|$ of observed complex samples is Rayleigh distributed [82, 176]:

$$p_{|x|}(|x|) = \frac{2|x|}{\sigma_r^2} e^{-\frac{|x|^2}{2\sigma_r^2}}. \tag{5.39}$$

The reason for this is that the multipath reflections can be described as signals with random phases, and that therefore the sum of many reflections can be represented as Gaussian noise (the amplitude of Gaussian complex noise signals is Rayleigh distributed). However, if there is a dominant path, then the amplitude distribution is a Rice distribution:

$$p_{|x|}(|x|) = \frac{|x|}{\sigma_r^2} e^{-(\frac{|x|^2+a^2)}{2\sigma_r^2}} I_0(a|x|/\sigma^2). \tag{5.40}$$

where a is defined by $a = \mathcal{E}\{|x|\}$, and I_0 is the modified Bessel function of the first kind of order 0. Figure 5.11 shows the amplitude of a baseband signal of telescope 6 for a dataset wit GSM transmissions. The figure also shows a mask which is used for separation of the GSM bursts from time slots which do not contain GSM transmissions. Figure 5.12 shows the distribution of the

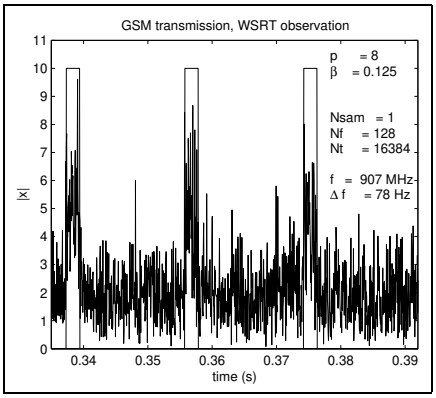


Figure 5.11. Absolute values of the baseband signal x_6 of telescope 6 for a narrow frequency band of 78 kHz width. A mask is defined, depicted by rectangular bars.

amplitude of the observed baseband signal of telescope 6, for time-slots without GSM transmissions (left) and for time-slots with GSM transmissions (right). In the situation in which there is no transmission the distribution is a Rayleigh distribution, as is expected for Gaussian signals. If there is a GSM transmitter present the distribution appears to be a Rice distribution, which means that there exists a dominant transmission path. The solid curves in the figure are the theoretical Rayleigh and Rice distributions. The Iridium dataset was processed

in a similar way as the GSM dataset. For Iridium, the distributions are similar to the GSM dataset, indicating that in this case (as expected) there also is a dominant path.

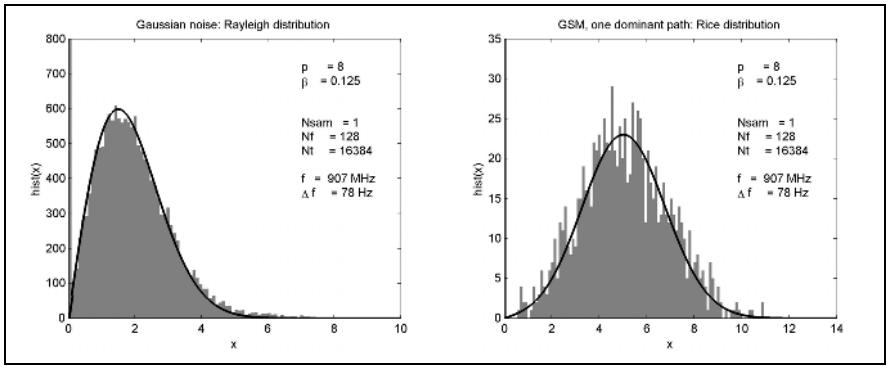


Figure 5.12. Distribution of the amplitude of the observed baseband signal of telescope 6, for time-slots without GSM transmissions (left) and for time-slots with GSM transmissions (right).

5.3.3 Application of multichannel detectors: further examples

In order to demonstrate the performance of the subband detection and blanking method, as described earlier in this chapter, an observation containing GSM transmissions at 905 MHz was superimposed on an observation of 3C48 at 1420 MHz, both having the same bandwidth. The 3C48 dataset shows a galactic spectral absorption line which was masked by the added GSM data. Although a bit artificial, the excellent linearity of the WSRT system implies that had a GSM signal been transmitted with a carrier frequency of 1420 MHz, then the measured data would be the superposition of the two signals plus system noise. The overlay allows us to verify the blanking performance for various mixtures of signal-to-interference power, since the clean data are now available as a reference, while the theoretical coherency is also well-known. As described before, the detection of an interferer in a specific time-frequency cell is based on the eigenvalues of the corresponding correlation matrix of the resulting mixture. In this scheme, if one or more eigenvalues are above a threshold, then an interferer is detected and that data block is omitted. However, to avoid the selection of the threshold based on a desired false alarm rate, a choice was made to simply throw away the worst 30 percent of the data according to the value of the detector.

Figure 5.13 shows the coherency functions over all baselines for a particular mixture of signals and interference: scaling the GSM data by 0.1 and the clean

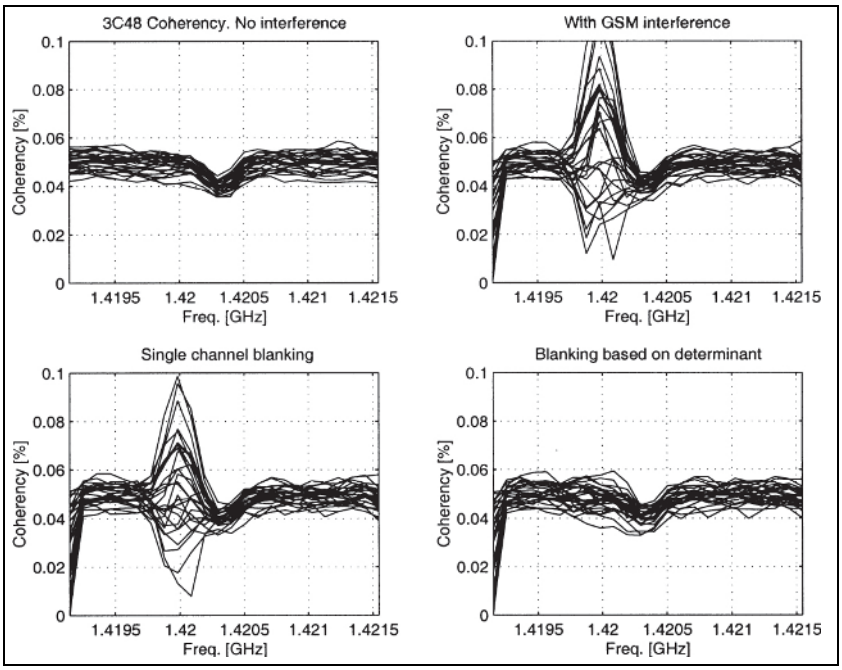


Figure 5.13. Magnitude of the coherency functions of 3C48 mixed with GSM interference: clean data (upper left), 3C48 data mixed with GSM data (upper right), effect of detection and excisison with a single channel detector (lower left), and effect of detection and excision with a mutichannel detector (lower right).

3C48 data by 0.9. It is seen that (top left) the clean 3C48 spectrum shows the absorption line[5], which is (top right) completely masked when GSM interference is added. After blanking, (bottom right) the absorption line is almost perfectly recovered. For comparison we also included (bottom left) the results of blanking based on single-channel power detection from channel 2 only, without the subband decomposition. The failure of this common way of single-channel detection is clearly seen. The reason is that the GSM signal was rather weak, so that for single-channel wideband processing the probability of detection was quite low, even for a false alarm rate of up to 30%.

To verify the phase behaviour of the coherency the unwrapped phase as a function of frequency was computed. Note that the geometrical delay compensation and fringe corrections were not included in the recording. As a result of

[5]The source flux of 3C48 at 1420.4 MHz is 15.9 Jy, yielding a correlation coefficient of approximately 0.055. The source 3C48 is an extragalactic spectral continuum source. Its radio waves, at 1420.4 MHz, are absorbed by the hydrogen gas of the Milky Way, which is visible as an absorption dip.

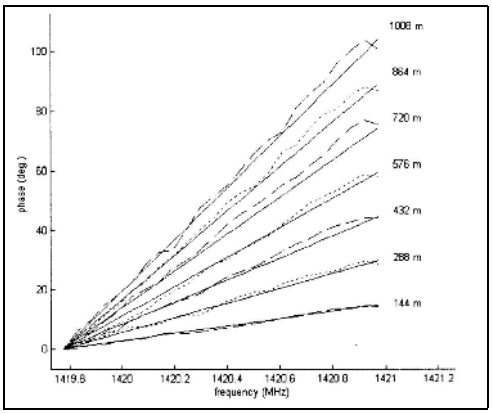


Figure 5.14. 3C48 coherency phase function versus frequency, for various baselines, predicted (solid lines) and observed (broken and dotted lines).

narrowband processing, the delay offset of one channel with respect to another shows up as a frequency-dependent phase shift (the fringe). The direction of the sources 3C48 and all baseline vectors is known, so the coherency phases were easily computed. Figure 5.14 shows the observed phase differences averaged over all identical baselines, and the computed phase, both as a function of frequency and baseline length. It is seen that the correspondence is very good. Note that for the shorter baselines we have more realisations so that their correspondence is better.

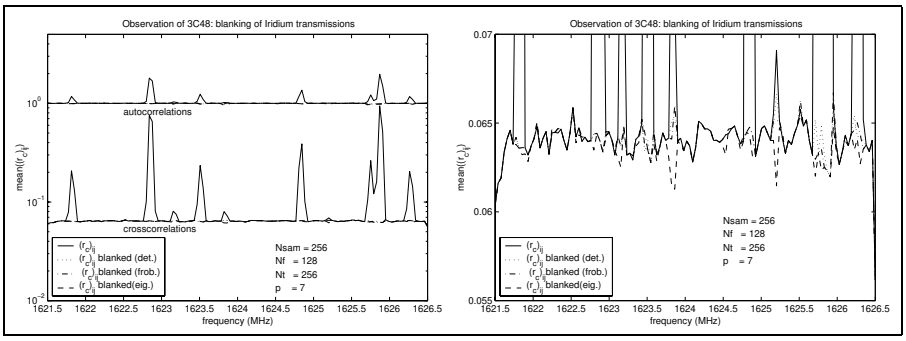


Figure 5.15. 3C48 observations affected by Iridium TDMA transmissions, before and after detection and excision. Both the auto (r_{cii}) and cross-correlation coefficients ($r_{cij},\ i \neq j$) are shown. The figure to the right is a zoomed-in version of the left figure.

Another example of the excision of communication transmissions in radio astronomy data is the excision of Iridium TDMA signals in the 3C48 contin-

uum source, shown in figure 5.15. Normalised auto- and cross-correlations are plotted in solid lines[6]. The observed TDMA time slot length is $L \approx 9$ ms, the observed duty cycle is $\beta = 0.05$. The processing time resolution, that is, the observation slot length the detection is based on, is 6.8 ms. As before, both single-channel detectors and multichannel detectors (based on the Frobenius norm , the largest eigenvalue and the determinant) are used. The filtered (excised) spectra are shown as discontinuous and dotted curves. The multichannel detector attenuates the transmission down to the noise level. To verify this, the noise variance is calculated using the derivations in chapter 3. The normalised covariance matrix and its variance are defined by

$$\mathbf{R}_w = \mathbf{W}\,\mathbf{R}\,\mathbf{W}, \qquad \mathbf{W} = (\mathbf{R} \odot \mathbf{I})^{-\frac{1}{2}} \tag{5.41}$$

$$\mathrm{var}\{\widehat{\mathbf{R}_w}\} = \mathbf{W}^2\,\mathrm{var}\{\widehat{\mathbf{R}}\}\,\mathbf{W}^2 = \frac{1}{N}(\mathbf{W}\mathrm{vecdiag}(\overline{\mathbf{R}}))(\mathbf{W}\mathrm{vecdiag}(\mathbf{R}))^t \tag{5.42}$$

In interferometer terms, the normalised covariance or correlation coefficient for interferometer ij is given by

$$r_{cij} = \frac{\mathcal{E}\{x_i\overline{x_j}\}}{\mathcal{E}\{x_i\overline{x_i}\}\mathcal{E}\{x_j\overline{x_j}\}} = \frac{\sigma_s^2}{\sqrt{(\sigma_{n_i}^2 + \sigma_s^2)(\sigma_{n_j}^2 + \sigma_s^2)}} \tag{5.43}$$

Assuming that $\sigma_s^2 \ll \sigma_{n_i}^2$ and $\sigma_s^2 \ll \sigma_{n_j}^2$, which is approximately true for the Iridium transmissions under consideration, then $r_{cij} \approx \frac{\sigma_s^2}{\sigma_{n\,i}\sigma_{n\,j}}$, and

$$\mathrm{var}\{r_{cij}\} = \frac{1}{N} r_{cii} r_{cjj} \approx \frac{1}{N}\frac{\sigma_s^4}{\sigma_{n_i}^2\sigma_{n_j}^2} \tag{5.44}$$

Assuming that the noise variances of the telescopes are equal, or $\sigma_{n_i}^2 = \sigma_n^2 \ \ (\forall i)$[7], the variance of the cross-correlations averaged over the $(\frac{1}{2}p(p-1))$ different entries in the normalised covariance matrix is given by:

$$\mathrm{mean}\{\mathrm{var}\{r_{cij}\}\} \approx \frac{1}{N\frac{1}{2}p(p-1)}\frac{\sigma_s^4}{\sigma_n^4} \tag{5.45}$$

For the settings used in figure 5.15 ($N = 256$, $p = 7$, $\sigma_s^2/\sigma_n^2 = 0.06$) , the mean of the standard deviation is 0.00087, which fits in with the observed value. The upper limit of the residual interference after detection and excision is approximately given by formula 5.30:

$$INR_{res} \lessapprox \frac{2\beta\sigma_n^2}{p\sqrt{(N)}} \tag{5.46}$$

[6]The source flux of the spectral continuum source 3C48 at 1624.4 MHz is 14.1 Jy, yielding a correlation coefficient of approximately 0.06.

[7]The noise variance differences between the WSRT telescopes are typically of the order of 15%.

For the settings used ($N = 256$, $p = 7$, $\sigma_n^2 = 1$, $\beta = 0.05$), this yields $INR_{res} \lessapprox 0.00089$. Not surprisingly, this is approximately equal to the mean standard deviation estimated above. The conclusion is that the transmitter signals are suppressed at least down to the estimated level as given by formula 5.30.

Radio astronomical signals, for instance from pulsar sources [146, 124], can be intermittent. An example of a pulsar, PSR0329+54, as observed with the NOEMI recorder at the WSRT , is shown in figure 5.16. The pulsar pulse period is 0.7145 s [145], the (FWHMP) pulse length is 5.9 ms, and its average flux at 301 MHz is 2.18 Jy. The figure shows the eigenvalues of the observation after a whitening operation. Note that the maximum and minimum eigenvalues outside the pulsar-off time slots correspond to the predictions of formula 3.48 [42][8]: $\widehat{\lambda_1} \approx 1.56$, and $\widehat{\lambda_p} \approx 0.56$. The pulsar signal is dominantly present in the largest eigenvalue, as expected. The reason is that the pulsar source is in power a dominant point-source which, in a noise-free situation, would give a rank one covariance matrix. Pulsar signal powers are known to be highly variable; this is clearly visible in the figure. Also here, some of the pulsar power "leaks" into the remaining $p - 1$ eigenvalues for reasons discussed earlier, namely finite bandwidth and sorting effects.

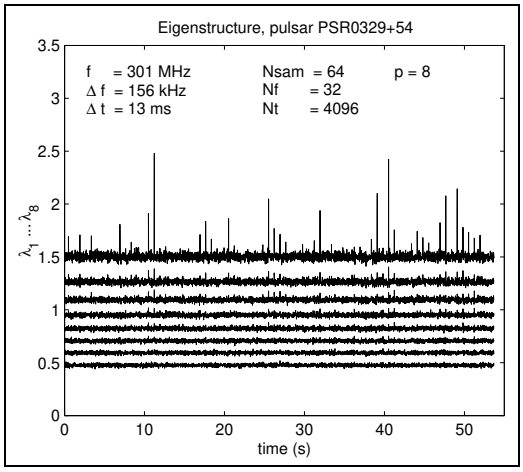


Figure 5.16. Eigenstructure time dependence for an astronomical intermittent source: the pulsar PSR 0329+54. The pulsar period is 0.715 s; the average pulse width is approximately 5.6 ms.

If the blanking of interference is necessary in observations with intermittent astronomical sources, then in most cases the detection and excision scenario can still be used, as the majority of those sources is weak with respect to system noise. However, an additional criterium for the detector that should be used is

[8]Here the number of observed real samples $N = 2 \times N_{sam}$, where N_{sam} is the number of complex samples.

that the detector threshold is higher than the expected maximum astronomical signal. Assuming that, for the duration of the on-period, the astronomical signal can be (approximately) described as a white Gaussian noise process, then the corresponding detector threshold can be derived in the same way as was done for the interference case.

5.3.4 Online blanking demonstration

In addition to the offline interference detection and excision experiments, online experiments were also carried out using the NOEMI online DSP system connected to the WSRT. The goal of these experiments was to demonstrate realtime operation of a multichannel detection and excision system at the WSRT. The operational mode of the online experiments was to alternately observe for a short period, in order to estimate optimal detector thresholds, and to perform the actual observation with the new detector threshold. Regular updates of the detector thresholds are necessary to account for gain and noise power drifts and fluctuation, which are present at radio telescopes such as the WSRT. As TDMA/FDMA and radar pulse signals were observed, each with duty cycles $\beta < 20\%$, the thresholds could easily be calculated from data from which 20% of the samples with the strongest power were removed. Figure 5.17 shows an example of an online blanking experiment at the WSRT of airplane radar at 1080 MHz. Both the original data and the blanked data were, simultaneously, stored on disk. The observational duty cycle, that is, the sustained processing capacity, was 20 %, mainly due to limitations of the PC PCI bus. The interfering power was removed, at least down to the noise level.

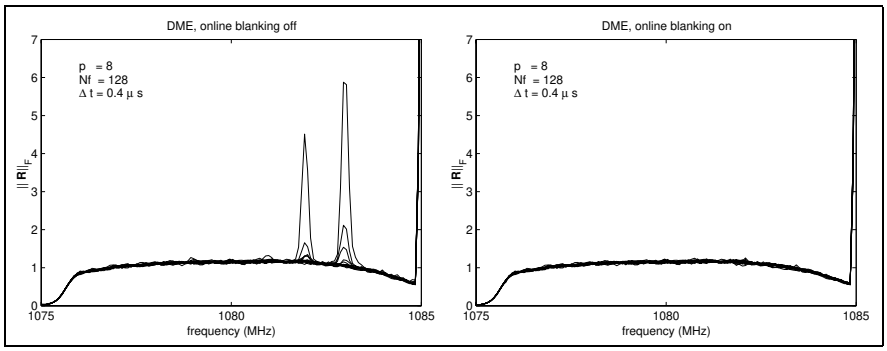


Figure 5.17. Online detection and excision of airplane radar DME: original spectrum (left) and spectrum with radar bursts removed (right).

5.4 Conclusions and further research

In this chapter the theoretical performance of single-channel and multichannel interference detectors were considered and applied to experimental radio astronomical data. It was shown that by subband processing, many narrowband techniques available in array signal processing and detection theory can be successfully applied to radio astronomical observations contaminated with intermittent transmitter signals and interference. The benefits of multichannel spatiospectral detection and excision were demonstrated, both theoretically and experimentally. Finally, the theoretical limits for the maximum attenuation numbers were derived.

The detection theory (Neyman-Pearson detector) used in this chapter was based on the assumption that the a-priori probabilities (P(H1) and P(H0)) are not known, and that there are no costs assigned to the different possible outcomes of the hypothesis tests. This makes the detection methods generally applicable. However, if for transmitters and interference the a-priori probabilities are known, and (or) if costs for the different outcomes of the hypothesis tests are defined, then also several alternative hypothesis testing criteria can be used. It is likely that by following some of these alternative scenarios the effect of the remaining interference on the astronomical data can be further reduced. It is also expected that the improvement over the method described in this chapter is not dramatic, but it may be worthwhile.

In the analysis presented in this chapter, the focus was on single-frequency channel processing. In practice the transmitter or interferer bandwidth may be (much) different from the analysis bandwidth. If this is the case, then, obviously, a great improvement of the detection and excision performance can be achieved by combining the frequency bins prior to detection.

When excising interference, one must keep track of the time-frequency slots excised because they influence the sensitivity of the observations. The remaining data (i.e. the un-excised data) must be multiplied with a scaling factor in order to be comparable to frequency bins which are not affected by excision. If this scaling is not applied, then the observed spectral lines will be distorted. Note that this scaling procedure is interferometer -based.

Chapter 6

Spatial filtering

6.1 Introduction

This chapter will focus on the efficacy of multichannel spatial filtering for the removal of continually present man-made interference or radio transmissions, such as TV signals, radio broadcasts, or communications and position reference satellite systems. Examples of these transmissions are shown in figure 6.1. Spatial filtering can be applied to interferometric radio telescope arrays such as the Westerbork Synthesis Radio Telescope (WSRT) in the Netherlands, the Very Large Array (VLA) in the USA, or future massive phased array telescopes, such as the Low Frequency Array (LOFAR), currently under design in the Netherlands, and the Square Kilometer Array (SKA).

Extensive literature exists concerning spatial filtering and nulling in beamforming arrays, for example in the context of communications systems and radar array signal processing [149] [88] [168]. In radio astronomy, research on and application of these techniques started only fairly recently, around 1998. Most of the efforts are directed at subtraction approaches using a few ($\approx$ two to four) antennas [6] [29]. Although effective in interferer reduction capabilities, these methods do not fully exploit the subspace structure of the antenna arrays. In astronomical SKA-related phased-array studies (i.e. pre-correlation), however, beamforming and subspace-based nulling were investigated and applied to radio astronomical demonstrators [64] [136] [43] [45]. Until recently [100] [147], post-correlation interference mitigation studies using the subspace structure of the full array covariance matrix , were very limited. The aim of the research presented in this chapter therefore is to focus on post-correlation spatial filtering techniques, based on a subspace structure analysis of the telescope output signal covariance matrices[1].

In interferometric radio astronomy the signals from various sensors (telescopes) are usually split into narrow frequency bins (say 50 kHz), and correlated

[1] Parts of the results of this chapter were published in [98, 22, 119, 148]

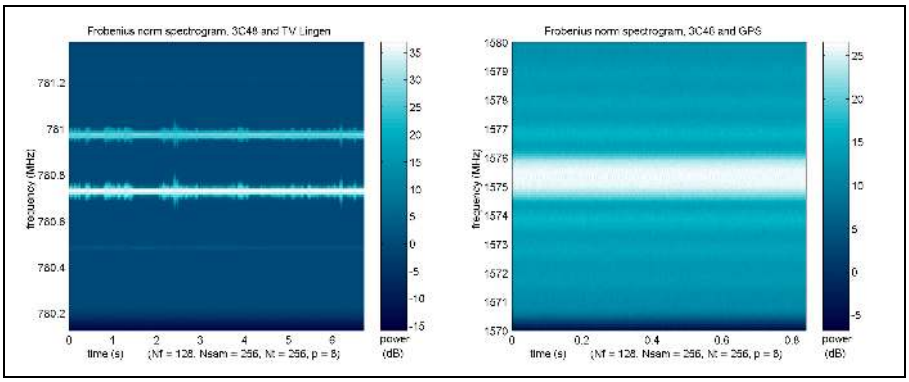


Figure 6.1. Examples of time-continuous transmissions in observation of the astronomical source 3C48. The spectrograms show stereo sound carrier waves of the analog television station TV Lingen (left) and transmissions of the GPS satellite system (right).

over 1–100 milliseconds to yield short-term correlation matrices. These are then integrated over longer periods of typically 10–60 seconds to yield long-term correlation matrices (containing the "visibilities"), which are stored onto disk or tape and constitute the output of the telescope interferometer. Images of the sky are usually constructed by fourier transforming several hours of "visbility" data.

The long-term correlation matrices contain contributions from the astronomical sources in the pointing direction through the main lobe of the telescope, from interferers in the near and far field through the side lobes, and from spatially white receiver noise. The astronomical signals usually have a signal-to-noise ratio (SNR) of -20 dB or less, and hence they are too weak to be detected over short integration periods. Harmful interference may range from -70 dB up to $+50$ dB with respect to the instantaneous system noise level.

Continually present interferers cannot be cut out in the time-frequency plane without losing astronomical information as well. In some cases, for example in continuum observations, spectral notch filters can be used. However, in many other cases, such as spectral line research, spectral filters cannot be used and other techniques are needed, such as spatial filtering. Assuming that the frequency bins are sufficiently narrow-band compared to the maximal propagation delay across the array, we can associate a spatial signature vector to each interferer, and estimate these from the short-term correlation matrices. By projecting out the corresponding dimensions, the interference is removed. The filtered correlation matrices are then integrated over 10 s.

Spatial filtering modifies the correlation matrix and therefore a correction must be applied to recover the original visibility matrix of interest to radio astronomers. The correction is possible under the assumption that the spatial signatures of interferers change sufficiently over the 10 s period, so that different

dimensions are knocked out each time. This is expected since (a) ground-based interferers are subject to multipath fading, (b) satellite and airplane interferers move, and moreover (c) the telescopes are rotating slowly while tracking a point in the sky, and continuously compensate for the changing baseline lengths by delay tracking and a phase rotation (fringe correction) in the order of a few Hz. This causes even stationary interferers (TV stations) to change in the spatial signature vector over a period of 10 s.

In the next sections, the spatial filtering algorithm is introduced and the correction that has to be applied to recover a good estimate of the desired covariance matrix is discussed. Next, the performance of the algorithm is shown both on simulated data, and on real data collected at the WSRT .

6.2 Data model

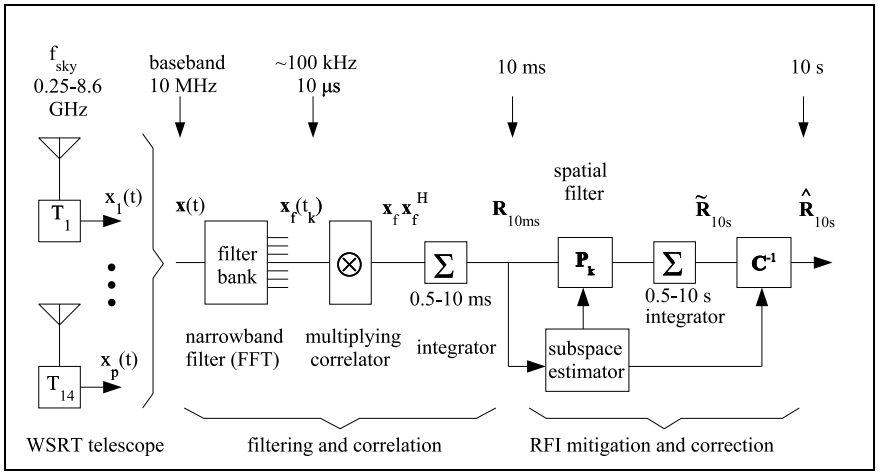


Figure 6.2. Se-tup of the correlation procedure with spatial filtering and distortion correction.

Consider a telescope array as having p elements, and consider a single frequency bin, which for simplicitys sake is assumed to have at most $q = 1$ interferer present. The array output vector $\mathbf{x}(t)$ is modelled in complex baseband form (cf. section 2.3.2) as

$$\mathbf{x}(t) = \mathbf{a}(t)s(t) + \mathbf{v}(t) + \mathbf{n}(t) \tag{6.1}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_p(t)]^t$ is the $p \times 1$ vector of output signals at time t, $s(t)$ is the interferer signal with spatial signature vector $\mathbf{a}(t)$ which is assumed stationary only over short time intervals, $\mathbf{v}(t)$ is the received sky signal, assumed a stationary Gaussian vector with covariance matrix $\mathbf{R}_v$, and $\mathbf{n}(t)$ is the $p \times 1$ noise vector with independent identically distributed Gaussian entries and

covariance matrix $\sigma^2\mathbf{I}$. We assume that σ_n^2 is known from a calibration observation, and that $\mathbf{R}_v \ll \sigma_n^2\mathbf{I}$. Given observations $\mathbf{x}_n = \mathbf{x}(nT_s)$, where T_s is the sampling period, the objective is to estimate $\mathbf{R}_v$.

Figure 6.2 schematically shows the set-up of the data acquisition and correlation process, followed by a spatial filtering and distortion correction step. The processing steps will be discussed in more detail in the next section.

6.3 Spatial filtering algorithm

Given the observations $\mathbf{x}_n$, the first step is to form short-term covariance estimates $\widehat{\mathbf{R}}_k$,

$$\widehat{\mathbf{R}}_k = \frac{1}{T}\sum_{n=kM}^{(k+1)M} \mathbf{x}_n\mathbf{x}_n^H \tag{6.2}$$

where M is the number of short-term samples to average, MT_s is in the order of 1–100 millisecond . In the usual procedure, these matrices are then further averaged to obtain a long-term (say $NMT_s = 10$ second) estimate

$$\widehat{\mathbf{R}}^{10s} = \frac{1}{N}\sum_{k=1}^{N}\widehat{\mathbf{R}}_k \tag{6.3}$$

The reason for choosing the 1–100 millisecond and 10 second values is given later in this section, and in section 6.6.

If there is only an astronomical signal and white Gaussian noise, $\widehat{\mathbf{R}}^{10s}$ is an unbiased estimate of the true covariance matrix

$$\mathbf{R}_0 = \mathbf{R}_v + \sigma_n^2\mathbf{I} \tag{6.4}$$

where $\mathbf{R}_v$ contains the astronomical 'visibilities' and σ_n^2 is the telescope noise power.

Consider now the situation where there is an interferer with zero mean, power σ_k^2 and a spatial signature vector $\mathbf{a}_k$ (normalised to unit norm), assumed constant over the short integration periods. The expected value of the short-term estimates $\widehat{\mathbf{R}}_k$ will then be

$$\mathbf{R}_k \;=\; \mathbf{R}_0 \;+\; \sigma_k^2\mathbf{a}_k\mathbf{a}_k^H \;=\; \mathbf{R}_v + \sigma_n^2\mathbf{I} \;+\; \sigma_k^2\mathbf{a}_k\mathbf{a}_k^H \tag{6.5}$$

In the construction of the long-term estimate, the interferer contribution will be $(1/N)\sum\sigma_k^2\mathbf{a}_k\mathbf{a}_k^H$. Depending on the variability of $\mathbf{a}_k$, the contribution will average out somewhat, but if σ_k is strong, its influence will be felt: the estimate of $\mathbf{R}_v$ will be biased and will also have an increased variance. It is therefore desirable to filter the interferer out.

Suppose that the spatial signature $\mathbf{a}_k$ of the interferer is known. We can then form a spatial filter $\mathbf{P}_k$,

$$\mathbf{P}_k := \mathbf{I} - \mathbf{a}_k(\mathbf{a}_k^H\mathbf{a}_k)^{-1}\mathbf{a}_k^H \tag{6.6}$$

which is such that $\mathbf{P}_k\mathbf{a}_k = 0$. Thus, when this spatial filter is applied to the data covariance matrix all the energy of the interferer will be nulled. Denoting the modified (filtered) covariance matrix by $\check{\mathbf{R}}_k$ we have

$$\check{\mathbf{R}}_k = \mathbf{P}_k\widehat{\mathbf{R}}_k\mathbf{P}_k\,, \quad \mathcal{E}\{\check{\mathbf{R}}_k\} = \mathbf{P}_k\mathbf{R}_0\mathbf{P}_k\,. \tag{6.7}$$

Note that the astronomical data are modified as well, so that we will have to apply a correction at a later stage. When the spatial signature of the interferers is unknown, it can be estimated by an eigenanalysis of the sample covariance matrix $\widehat{\mathbf{R}}_k$. The covariance matrix $\mathbf{R}_k$ can be written in terms of eigenvalues and eigenvectors as

$$\mathbf{R}_k =: \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{U}_k^H \tag{6.8}$$

where $\mathbf{U}_k$ is a unitary matrix containing the eigenvectors, and $\mathbf{\Lambda}_k$ is a diagonal matrix containing the eigenvalues. Assuming that the astronomical contribution is small, the eigenvalue decomposition (cf. section 3.2.2) can be expressed as

$$\mathbf{R}_k = [\mathbf{a}_k \mid \mathbf{U}_{nk}] \left[\begin{array}{c|c} \sigma_k^2 + \sigma_n^2 & 0\cdots 0 \\ \hline 0 & \\ \vdots & \sigma_n^2\mathbf{I}_{p-1} \\ 0 & \end{array}\right] \left[\begin{array}{c} \mathbf{a}_k^H \\ \hline \mathbf{U}_{nk}^H \end{array}\right] \tag{6.9}$$

where $\mathbf{U}_{nk}$ is the noise subspace and σ_n^2 the telescope noise variance. Thus, the principal eigenvector is equal to $\mathbf{a}_k$. This means that an estimate of $\mathbf{a}_k$ can be obtained from the eigenvalue decomposition of $\mathbf{R}_k$, and that it can be used to construct the projection matrix $\mathbf{P}_k$. As in practice the covariance matrix $\mathbf{R}_k$ may not be known, the sample covariance matrix $\widehat{\mathbf{R}}_k$ has to be used to construct $\mathbf{P}_k$. The next step is averaging the modified covariance matrices $\check{\mathbf{R}}_k$, which yields the long-term estimate

$$\check{\mathbf{R}}^{10s} := \frac{1}{N}\sum_{k=1}^{N}\check{\mathbf{R}}_k \;=\; \frac{1}{N}\sum_{k=1}^{N}\mathbf{P}_k\widehat{\mathbf{R}}_k\mathbf{P}_k \tag{6.10}$$

To recover an unbiased estimate of $\mathbf{R}_0$, assuming that the interferer has been projected out completely, a correction has to be applied to $\check{\mathbf{R}}^{10s}$. The following matrix identity is employed

$$\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B}) \tag{6.11}$$

Thus, the expected value of the distorted, filtered, covariance matrix can be expressed as

$$\begin{aligned} \mathcal{E}\{\mathrm{vec}(\check{\mathbf{R}}^{10s})\} &= \mathcal{E}\{\frac{1}{N}\sum_{k=1}^{N}(\mathbf{P}_k^T \otimes \mathbf{P}_k)\mathrm{vec}(\widehat{\mathbf{R}}_k)\} \\ &= \left\{\frac{1}{N}\sum_{k=1}^{N}(\mathbf{P}_k^T \otimes \mathbf{P}_k)\right\}\mathrm{vec}(\mathbf{R}_0) \\ &= \mathbf{C}\;\mathrm{vec}(\mathbf{R}_0) \end{aligned} \tag{6.12}$$

where $\mathbf{C} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{P}_k^T \otimes \mathbf{P}_k)$. Thus, an unbiased estimate of $\mathbf{R}_0$ can be obtained by applying the inverse of $\mathbf{C}$ to $\mathrm{vec}(\check{\mathbf{R}}^{10s})$,

$$\begin{aligned} \widehat{\mathbf{R}}_0^{10s} &= \mathrm{unvec}(\,\mathbf{C}^{-1}\mathrm{vec}(\check{\mathbf{R}}^{10s})\,) && (6.13)\\ \mathcal{E}\{\widehat{\mathbf{R}}_0^{10s}\} &= \mathrm{unvec}(\,\mathbf{C}^{-1}\mathcal{E}\{\mathrm{vec}(\check{\mathbf{R}}^{10s})\}\,) \\ &= \mathbf{R}_0 = \mathbf{R}_v + \sigma_n^2\mathbf{I} && (6.14)\end{aligned}$$

In short, to obtain the covariance matrix representing the astronomical sources, the projected short-term covariance matrices can be averaged as usual to long-term averages, but the correction matrix $\mathbf{C}$ has to be applied, which is formed in the same way by averaging $\mathbf{P}_k^T \otimes \mathbf{P}_k$. At this point, several aspects will be considered.

- The invertibility of $\mathbf{C}$ is crucial for recovering the correlation matrix of the astronomical signals. If all $\mathbf{P}_k$ are the same ($\mathbf{a}_k$ is stationary), then $\mathbf{C}$ will not be invertible. One can show that an average of only a few different $\mathbf{P}_k$ is needed to ensure invertibility. Indeed, if the $\mathbf{a}_k$ are random independent vectors, then $\mathbf{C} \to \mathbf{I}$ as $N \to \infty$. Thus $\mathbf{a}_k$ needs to be sufficiently variable over the long integration period. On the other hand, to be able to estimate $\mathbf{a}_k$ from the data, the interferer signature has to be constant over short integration periods. An analysis of the spatial stationarity properties of the interferer gives suitable information to determine the partitioning of a long integration interval into short intervals where $\mathbf{a}_k$ can be considered constant.

- The stationarity of the interferer is dependent not only on its own movement (in thecase of mobile telephones or satellites), but also on the geometrical delay compensations applied to the telescope signals to compensate for the rotation of the earth, meanwhile tracking the desired astronomical objects. Depending on the baseline length and the elevation angle of the pointing direction, the stationarity is limited to about 0.05–0.5 sec in the case of WSRT, and hence varies sufficiently over the usual integration period of 10 seconds.

- A good estimate of the interferer signature can be obtained only if it is sufficiently strong to be detected. This is dependent also on the short-term integration length, and thus limited by the period over which the interferer can be considered stationary. The accuracy of this estimate determines the amount of residual interference.

  If the principal eigenvector is always projected out, regardless of the interference power, then a bias will result if the power is weak. Thus it is a better approach to detect the presence of an interferer using a standard test for this, and to apply a projection only if the interferer has been detected.

- Another issue is that of computational expenditure. $\mathbf{C}$ will be a matrix of size $p^2 \times p^2$. Online constructing and inverting such a large matrix can be time-consuming. An alternative is to store the projection directions $\{\mathbf{a}_k\}$ and apply the corrections off-line.

The single-interferer derivations are easily generalised to multiple interferers. Suppose that there are q_r interferers with unknown $\mathbf{a}$ vectors, with $q_r < p$. As described in sections 2.3.2 and 3.2.2, the covariance matrix including multiple interferers can be described by

$$\mathbf{R} = \mathbf{R}_v + \sigma_n^2 \mathbf{I} + \mathbf{A}_r \mathbf{B}^r \mathbf{A}_r^H \tag{6.15}$$

where $\mathbf{A}_r$ and the source power matrix $\mathbf{B}_r$ are defined as in section 2.3.2. The projection matrix in this case is defined by

$$\mathbf{P}_k = \mathbf{I} - \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H \tag{6.16}$$

If the interferer directions ar not known, then the interferer subspace can be found by means of an eigenvalue analysis of $\widehat{\mathbf{R}}_k$. Also here, the astronomical source signal power must be much less than the interferer powers because otherwise they would distort the interferer subspace estimation. The derivation of the distortion correction for the multiple interferer case is identical to the single interferer case.

6.4 Alternative spatial filtering algorithms

Without going into too much detail, a few other possibilities for spatial filtering and interference cancellation will be mentioned next. Assume that the model described in the previous section, $\mathbf{R}_k = \mathbf{R}_v + \sigma_k^2 \mathbf{a}\mathbf{a}^H + \sigma_n^2 \mathbf{I}$ (equation (6.5)), holds. For this model, two alternative techniques will be discussed: the subtraction of interference and interference mitigation using a reference signal. This section will conclude with a brief comparison of pre- correlation and post-correlation filtering techniques.

Subtraction of interference

An alternative filtering method is interference subtraction. With σ_k^2 and σ_n^2 or their estimates, the contribution of the interferer can be reduced by subtracting it from the observed covariance matrix. Let $\check{\mathbf{R}}$ be the filtered covariance matrix, that is, the observed covariance matrix with the estimated interference removed by subtraction:

$$\check{\mathbf{R}}_k = \widehat{\mathbf{R}}_k - \hat{\sigma_k}^2 \hat{\mathbf{a}}_k \hat{\mathbf{a}}_k^H \tag{6.17}$$

Without further knowledge, the best estimate of $\mathbf{a}_k$ is the dominant eigenvector $\mathbf{u}_{1k}$ of $\widehat{\mathbf{R}}_k$, and likewise the best estimate of the interferer power $\hat{\sigma}_k^2$ is $\lambda_{1k} - \sigma_n^2$. Since both of these are derived from $\widehat{\mathbf{R}}_k$, it turns out not to be too different

from the projection scheme. Indeed, using a scaling parameter α_k, and using the definition $\mathbf{L}_k \equiv (\mathbf{I} - \alpha_k \mathbf{u}_{1k}\mathbf{u}_{1k}^H)$, the two methods are related as follows:

$$\begin{aligned}
\check{\mathbf{R}}_k^{\perp} &= \mathbf{L}_k \widehat{\mathbf{R}}_k \mathbf{L}_k^H && (6.18)\\
&= \widehat{\mathbf{R}}_k - \mathbf{u}_{1k}\mathbf{u}_{1k}^H \lambda_{1k}(2\alpha_k - \alpha_k^2) && (6.19)\\
\check{\mathbf{R}}_k^{-} &= \widehat{\mathbf{R}}_k - \hat{\sigma_k}^2 \mathbf{u}_{1k}\mathbf{u}_{1k}^H && (6.20)
\end{aligned}$$

where $\check{\mathbf{R}}_k^{\perp}$ is a generalisation of projection-filtered covariance matrix. It would be the equal to the projection-filtered covariance matrix for $\alpha = 1$. $\check{\mathbf{R}}_k^{-}$ the subtraction filtered matrix. Now, $\check{\mathbf{R}}_k^{\perp}$ can be made equal to $\check{\mathbf{R}}_k^{-}$ by selecting α_k such that $\lambda_{1k}(2\alpha_k - \alpha_k^2) = \sigma_k^2$. Using $\lambda_{1k} = \sigma_r^2 + \sigma_n^2$, the scaling factor α_k can be written as

$$\alpha_k = 1 \pm \sqrt{\frac{\sigma_n^2}{\sigma_n^2 + \sigma_k^2}} \tag{6.21}$$

The projection scheme had $\alpha_k = 1$. Therefore, subtraction of eigenvector components can also be represented by a two-sided linear operation on the correlation matrix, for an arbitrary $\alpha_k > 0$. Consequently, the visibility matrix $\mathbf{R}_v$ is altered, which means that the corrections mentioned in section 6.3 are in order.

For known interferer direction vectors $\mathbf{a}_k$, a distortion correction only needs to be applied if the interferer power estimate is biased, as will be shown next. Define $\check{\mathbf{R}}^{\perp} = \sum_k \check{\mathbf{R}}_k^{\perp}$ and $\check{\mathbf{R}}^{-} = \sum_k \check{\mathbf{R}}_k^{-}$ so that

$$\begin{aligned}
\mathcal{E}\{\check{\mathbf{R}}^{\perp}\} &= \frac{1}{N}\sum_{k=1}^{N} \mathbf{L}_k \mathbf{R}_0 \mathbf{L}_k + \frac{1}{N}\sum_{k=1}^{N} \sigma_k^2 \mathbf{L}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{L}_k && (6.22)\\
\mathcal{E}\{\check{\mathbf{R}}^{-}\} &= \mathbf{R}_0 + \frac{1}{N}\sum_{k=1}^{N} \left[\sigma_k^2 - \mathcal{E}\{\hat{\lambda}_{1k}(2\hat{\alpha_k} - \hat{\alpha_k}^2)\}\right] \mathbf{a}_k \mathbf{a}_k^H && (6.23)\\
&= \mathbf{R}_0 + \frac{1}{N}\sum_{k=1}^{N} \left[\sigma_k^2 - \mathcal{E}\{\hat{\sigma}_k^2\}\right] \mathbf{a}_k \mathbf{a}_k^H && (6.24)
\end{aligned}$$

where $\mathbf{R}_o$ is the interference-free covariance matrix. Recall that for the projection scheme, $\alpha_k = 1$ and $\mathbf{L}_k \mathbf{a}_k = \mathbf{0}$, hence $\mathbf{R}_0$ changes. For α_k equal to (6.21), the expected value in (6.24) is not equal to $\mathbf{R}_o$ if $\hat{\sigma}_k^2$ is biased, which means that corrections are needed.

Pre correlation versus post correlation spatial filtering

Spatial filters, both projection and subtraction filters, can be applied before correlation or after correlation. Both are similar from a mathematical point of view, assuming a stationary filter. Indeed, as before, let the time-dependent telescope array output signal vector be denoted by $\mathbf{x}$, and let this vector be modified by a pre-correlation constant filter matrix $\mathbf{P}$. The filtered array output vector $\check{\mathbf{x}}$ is then given by

$$\check{\mathbf{x}} = \mathbf{P}\mathbf{x} \tag{6.25}$$

Let $\mathbf{R} = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}$, and denote the filtered covariance matrix by $\check{\mathbf{R}}$, then

$$\check{\mathbf{R}} = \mathcal{E}\{\check{\mathbf{x}}\check{\mathbf{x}}^H\} = \mathcal{E}\{\mathbf{P}\mathbf{x}\mathbf{x}^H\mathbf{P}^H\} = \mathbf{P}\mathbf{R}\mathbf{P} \tag{6.26}$$

which means that spatial filtering at the array output level is mathematically identical to filtering after correlation, assuming the filter is stationary and data-independent. Therefore, in applying spatial filters in linear systems it does not make much sense to make a classification distinction between applying these filters before or after correlation. A more relevant distinction would be the update rate of the estimation of the spatial signature vectors. Note that in the detection of interference, the difference between pre-correlation and post-correlation is essential.

6.5 Spatial filter attenuation estimates

6.5.1 Model error simulations

In this section the results of a computer experiment will be presented. Consider a scenario in which there are $p = 8$ telescopes, a weak astronomical signal (-20 dB w.r.t. the receiver noise), and a single interferer of varying power. The received data are correlated over $M = 100$ samples, the spatial filters are applied, and the result is further averaged over $N = 100$ such matrices. The performance measure used is the Mean-Squared-Error of the resulting estimate $\widehat{\mathbf{R}}^{10s}$ compared to the desired outcome $\mathbf{R}_0 = \mathbf{R}_v + \sigma^2\mathbf{I}$, averaged over 100 Monte-Carlo experiments, that is,

$$\text{MSE} = \text{E}\left\{\| \widehat{\mathbf{R}}^{10s} - \mathbf{R}_0 \|_F\right\}$$

Figure 6.3, left, shows the MSE curves for several cases:

- Unfiltered interference: this shows what happens if nothing is done and the long-term covariance estimate is computed traditionally.
- RFI-free: the data do not contain interference and the covariance is estimated as traditionally done (no filtering).
- Perfect projection filter: assumes that the spatial signatures of the interferer are perfectly known, and the distortion correction is applied.
- Perfect projection filter, uncorrected: assumes that the spatial signatures of the interferer are perfectly known, but the distortion correction is not applied.
- Subtraction filter, λ_1: a filter based on the substraction of the interference using an estimate of the largest eigenvalue and assumes a known spatial signature $\mathbf{a}_k$: $\check{\mathbf{R}} = \widehat{\mathbf{R}} - \frac{1}{N}\sum_{k=1}^{N} \hat{\lambda}_{1k}\, \mathbf{a}_k\mathbf{a}_k^H$. The distortion correction is not applied. This curve lies above the uncorrected perfect projection filter because of uncertainties in the estimated $\hat{\lambda}_{1k}$.

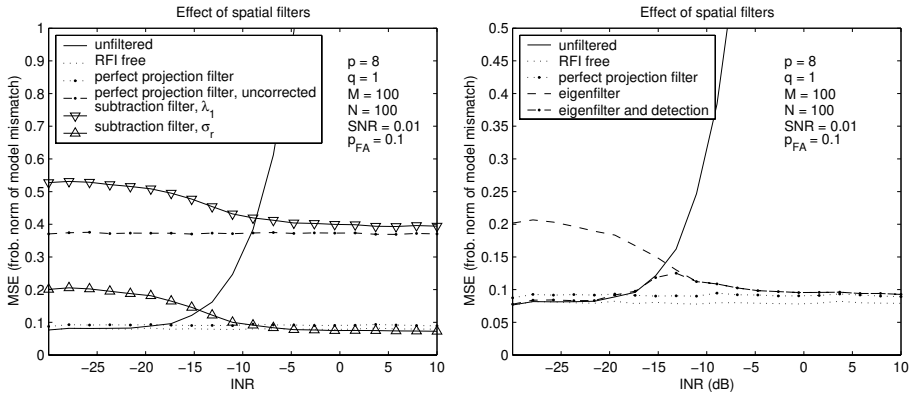


Figure 6.3. MSE for several spatial filters as function of interferer power. Subtraction filters (left) and eigenfilters (right) are compared to perfect projection filters, to applying no filters at all, and to a situation in which there is no interference.

- Subtraction filter, σ_k^2: a filter based on the substraction of the interference using an estimate of the interferer power $\hat{\sigma}_k^2 = \hat{\lambda}_1 - \hat{\sigma}_n^2$ and assumes a known spatial signature $\mathbf{a}_k$ or an estimate $\hat{\mathbf{a}}_k = \hat{\mathbf{u}}_1$: $\check{\mathbf{R}} = \widehat{\mathbf{R}} - \frac{1}{N}\sum_{k=1}^{N} \hat{\sigma_k^2}\, \mathbf{a}_k \mathbf{a}_k^H$.

The curves clearly demonstrate the effectiveness of spatial filters and the need for distortion corrections. Figure 6.3, right, shows the MSE curves for several cases:

- Unfiltered interference: this shows what happens if nothing is done and the long-term covariance estimate is computed traditionally.
- RFI-free: the data do not contain interference and the covariance is estimated as traditionally done (no filtering).
- Perfect projection filter: assumes that the spatial signatures of the interferer are perfectly known; the distortion correction is applied.
- Eigenfilter: the spatial signatures are estimated from the eigenvalue decomposition of the short-term data covariance matrices; the distortion correction is applied.
- Eigenfilter and detection: first it is checked whether the interference is observable in the data using a standard Box-test (white-noise test) on the short-term covariance estimates [105],

$$L = 2M[\mathrm{tr}(\widehat{\mathbf{R}}_k/\sigma^2) - p - \log\det(\widehat{\mathbf{R}}_k/\sigma^2)] \;\overset{>}{<}\; \gamma$$

where γ is a detection threshold. (In figure 6.3, the detection threshold was selected to obtain a false alarm probability of 0.1.) If an interferer is detected, then the spatial projection is applied as before.

Note that the scale of the right-hand figure differs from that of the left-hand figure. The results indicate that for INRs above -15 dB,[2] it is essential to apply the spatial filter. If the spatial signatures of the interferer are perfectly known, then the final estimate is almost as good as in the RFI-free case. If the spatial signatures are estimated from the data, then it is important first to detect if there is an interferer, otherwise for weak interferers the final covariance estimate is biased. In combination with detection, it is seen that the covariance estimate is very close to the interference-free result. The same is true for the subtraction filter based on $\hat{\sigma}_k^2$ estimates.

6.5.2 Interference attenuation limits

Asymptotic expressions for residual interference after applying spatial filtering can be estimated using a perturbation analysis of the eigenvalue decomposition [132]. Assume that there is only one interferer, and define the interference to noise ratios INR, before spatial filtering by

$$INR_{before} = \frac{p\,\mathrm{tr}(\mathbf{R}-\mathbf{D})}{p\,\mathrm{tr}(\mathbf{D})} = \frac{p\,\mathrm{tr}(\sigma_r^2\mathbf{a}\mathbf{a}^H)}{p\sigma_n^2} = \frac{\sigma_r^2}{\sigma_n^2} \tag{6.27}$$

where it is assumed that $\mathbf{a}$ is normalised: $\mathbf{a}^H\mathbf{a} = 1$. The INR after spatial projections is defined by

$$INR_{after} = \frac{p\,\mathrm{tr}(\mathbf{P}(\widehat{\mathbf{R}}-\widehat{\mathbf{D}})\mathbf{P})}{\mathrm{tr}(\mathbf{D})} = \frac{\sigma_r^2}{\sigma_n^2}\mathrm{tr}(\mathbf{P}\hat{\mathbf{a}}\hat{\mathbf{a}}^H\mathbf{P}) \tag{6.28}$$

As before, see for instance section 3.2.2, $\mathbf{R}$ can be decomposed in terms of eigenvalues λ_i, a noise subspace $\mathbf{U}_n$, and a signal subspace $\mathbf{U}_s$. Then $\mathbf{P}\hat{\mathbf{a}}$ is asymptotically a zero-mean Gaussian random process with variance determined by the asymptotic estimate of $\mathbf{P}\hat{\mathbf{a}}\hat{\mathbf{a}}^H\mathbf{P}$ [94] [132]:

$$\mathcal{E}\{\mathbf{P}\hat{\mathbf{a}}\hat{\mathbf{a}}^H\mathbf{P}\} = \frac{\sigma_n^2}{N}\frac{\lambda_1}{(\lambda_1-\sigma_n^2)^2}\mathbf{a}\mathbf{a}^H + o(N^{-1}) \tag{6.29}$$

For large N, the INR after spatial filtering is therefore given by:

$$\mathrm{INR_{after}} = \frac{1}{\mathrm{N}}\left(1 + \frac{1}{\mathrm{p\,INR_{before}}}\right) \tag{6.30}$$

It can be shown [94] that the residual interference is spatially white up to order N^{-2}. This means that the filtering residuals do not (up to order N^{-2} effects) add spatial features to $\widehat{\mathbf{R}}$.

[2] This bound is determined by the level at which the filtering algorithm can start to detect the presence of interference, and depends on the number of telescopes p, the number of short-term samples M, and on the selected false alarm rate.

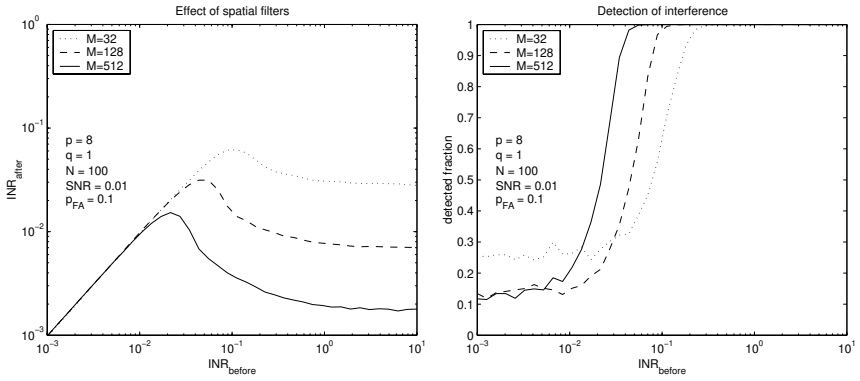


Figure 6.4. Simulation result. The left-hand figure shows the INR after applying spatial filters ($\mathrm{INR}_{\mathrm{after}}$) versus the initial INR before spatial filtering ($\mathrm{INR}_{\mathrm{before}}$). The short-term correlation length (M samples) is varied. The spatial filter is activated only when a certain threshold is exceeded. The right-hand figure shows the fraction of the data which in which interference was detected.

Figure 6.4 shows a simulation in which a spatial filter (plus distortion correction) was applied to simulated data. The spatial filter was activated only when a detector (Box test) detected interference. The false-alarm rate of the detector was set to 0.1. The left-hand figure shows that for low INR no RFI was detected, so that $\mathrm{INR}_{\mathrm{after}} = \mathrm{INR}_{\mathrm{before}}$. For high INR values, the curves decrease monotonously and converge to a limit. This part of the curve is nicely described by the asymptotic formula 6.29, as can be seen in figure 6.5, where the limiting values for high $\mathrm{INR}_{\mathrm{before}}$ are plotted together with the theoretical residual interference curve. The three limiting values lie a little below the theoretical curve because the three values were estimated for $\mathrm{INR}_{\mathrm{before}} = 10$ whereas the theoretical curve is the limiting case for $\mathrm{INR}_{\mathrm{before}} = \infty$. The curves saturate for high $\mathrm{INR}_{\mathrm{before}}$ values, apparently because the higher $\mathrm{INR}_{\mathrm{before}}$ power is fully compensated for by better estimates of the spatial signature of the interferer. The left-hand figure in 6.4 also confirms that increasing the number of samples M in the short-term time interval leads to better estimates of the spatial signature of the interferer and to more effective filters. The right-hand figure in 6.4 shows the fraction of samples in which RFI was detected. These curves imply that the occurrence of maxima in the left-hand figure are not caused by the fact that the interference is not detected by the detector. The reason that the maxima are larger than the asymptotic values for high $\mathrm{INR}_{\mathrm{before}}$ is that interference is detected but projected out (or subtracted) with very inaccurate estimates of $\mathbf{a}$.

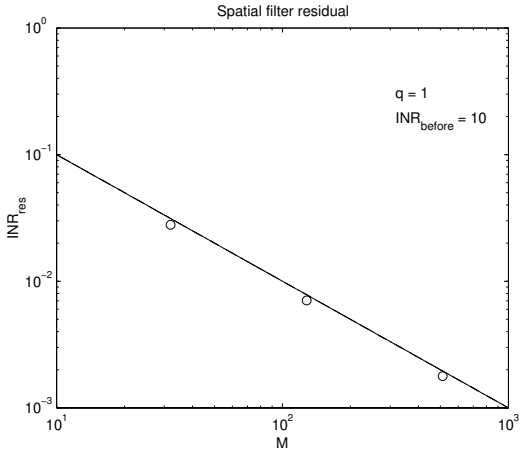


Figure 6.5. Residual INR after spatial filtering for large INR values. The curve is the theoretical limit for large INR, the circles are obtained from the simulations, shown in figure 6.4.

6.6 Correction matrix condition number estimates

Recall that for an optimal projection filter distortion correction, it is necessary that the interferer is stationary for short time intervals, and nonstationary for long intervals. Nonstationarity at long time scales ensures invertibility and a low condition number of the distortion correction matrix $\mathbf{C}$. High condition numbers need to be avoided because this increases the system noise. An important factor which makes stationary interferers nonstationary on long time scales, is the fringe frequency. The aim of this section is to find the relation between the fringe frequency, the condition number of the correction matrix, and the noise increase due to filtering and distortion correction.

Consider the case in which there is only one interferer, without multipath. In this case, at any point in time, there is only one array response vector $\mathbf{a}_r$ that corresponds to this interferer. The rank of the corresponding covariance matrix excluding noise ($\mathbf{R}-\mathbf{D} = \sigma_r^2 \mathbf{a}_r \mathbf{a}_r^H$) therefore is one. Under the influence of the fringe rotation during the integration interval, the array response vector changes. As a consequence, the rank of $\mathbf{R}-\mathbf{D}$ increases to p after a certain integration time length, and this is needed to ensure the invertibility of $\mathbf{C}$. In practically all cases, the full-rank situation is reached when the array response vector at time $t+\Delta t$ is perpendicular to $\mathbf{a}_r(t)$, or $\mathbf{a}_r(t+\Delta t)^H \mathbf{a}_r(t) = 0$. This is the somewhat arbitrary (although verified experimentally) criterion used for sufficient nonstationarity of $\mathbf{C}$. This criterion is supported by simulations in [147], and an expression in terms of phase rotations is derived next.

Suppose that the direction vector of an observed astronomical source is given by $\mathbf{s}(t)$. Because of the earth rotation, this direction vector is a function of time. Recall from section 2.3.1 that the telescope baseline coordinate matrix is given by $\mathcal{R}$. The array response vector for the astronomical source at a time t for the direction $\mathbf{s}(t)$ can then be written as

$$\mathbf{a}_s(t) = e^{-2\pi\jmath\lambda^{-1}\mathcal{R}\mathbf{s}(t)} \tag{6.31}$$

At the WSRT telescope, a (fringe) phase is added in the electronic receiver systems in order to compensate for the geometric delay-related phase rotation. One effect of this is that the phase of the field of view centre is fixed. The second effect is that an interferer at a fixed geographical location, initially with a fixed array response $\mathbf{a}_{r0}$, now will have a changing array response vector $\mathbf{a}_r(t)$:

$$\mathbf{a}_r(t) = \mathbf{a}_{r0} \odot \overline{\mathbf{a}_s(t)} \tag{6.32}$$

For an east-west linear array such as the WSRT , the array response vector can also be expressed (cf. section 4.1), in terms of the fringe frequency for the shortest baseline ν_{f0}, the baseline vector $\mathbf{b}$, and the earth rotation $\frac{dh}{dt}$. Let

$$\mathbf{b} = b_0\,[0, 1, \cdots, p-1]^t \tag{6.33}$$

$$\nu_{f0} = \frac{b_0}{\lambda}\frac{dh}{dt}\cos(\delta)\cos(h) \tag{6.34}$$

where b_0 is the smallest telescope distance (144 m). Then $\mathbf{a}_s(t)$ can be expressed by

$$\mathbf{a}_s(t) = e^{-2\pi\jmath\,\frac{\mathbf{b}}{b_0}\,\nu_f\,t} \tag{6.35}$$

Using these definitions, the inner product $\mathbf{a}_r(t+\Delta t)^H\mathbf{a}_r(t)$ is easily found by evaluating the exponential power series:

$$\mathbf{a}_r(t+\Delta t)^H\mathbf{a}_r(t) = \sum_{k=0}^{p-1} e^{2\pi\jmath k\nu_{f0}\Delta t} \tag{6.36}$$

$$= e^{\pi\jmath(p-1)\,\nu_{f0}\,\Delta t}\cdot\frac{\sin(\pi p\,\nu_{f0}\Delta t)}{\sin(\pi\nu_{f0}\Delta t)} \tag{6.37}$$

Thus the $|\mathbf{a}_r(t+\Delta t)^H\mathbf{a}_r(t)|$ expression is a sinc function with its first null at

$$\Delta t = \frac{1}{p\,\nu_{f0}} \tag{6.38}$$

Figure 6.6 (left) shows, for the WSRT with only the first eight telescopes, the value of $|\mathbf{a}_r(t+\Delta t)^H\mathbf{a}_r(t)|$ as a function of the fringe rotation phase. Because of this choice for the abscissa, the curve in the figure is independent of hour angle and declination. The first null, for p=8 telescopes, is found at $5.5\times 2\pi$ phase

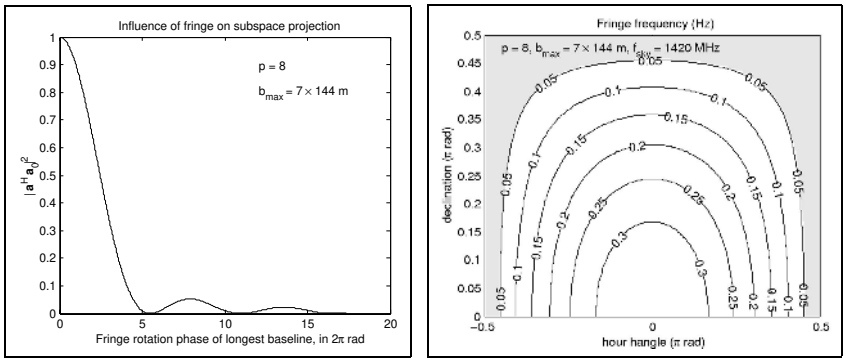


Figure 6.6. The left-hand figure shows the influence of fringe rotation on the inner product of two array response vectors, which initially are identical. One vector ($\mathbf{a}_{r0}$)remains fixed, the other($\mathbf{a}_r$) is affected by the fringe. The figure to the right shows curves with identical fringe rotation frequency for the hour angle - declination plane, for a baseline length of $(p-1) \times 144\,\mathrm{m}$,with p=8. Both figures are calculated in terms of fringe phases for the longest baseline.

rotations. Figure 6.6 (right) shows a contour plot of the fringe frequency for the WSRT for a baseline of $(p-1)b_0$ and an observation frequency of 1420 MHz. For example, at an hour angle and a declination of 45°, the fringe frequency is 0.2 Hz. This means that after 27.5 s, the phase is rotated over $5.5 \times 2\pi$ rad, as was required by the"first null" condition. Simulations [147] show that a sufficiently low condition number of $\mathbf{C}$ is obtained after 2 to 5 $\times 2\pi$ phase rotations. Apparently the "first null" criterion is slightly stronger; however, it can be used as a worst case criterion. Assume, for example, an integration time of two minutes, then formula (6.38) yields that a minimum fringe frequency of 0.05 Hz is required for an acceptable projection correction matrix. In this particular case, a projection filter in the shaded region of figure 6.6 will not work and a subtraction filter is preferable.

6.7 Measurement results

6.7.1 Application examples of projection and subtraction filters for time-continuous interference

In this section, the projection and subtraction filters which were discussed in the previous sections will be applied to observations carried out at the WSRT using the NOEMI data recorder. As the filters are based on subspace analysis, the observed covariance matrix is whitened, see for instance section 3.2.2: $\mathbf{R}_W = \mathbf{D}^{-\frac{1}{2}}\mathbf{R}\mathbf{D}^{-\frac{1}{2}}$. The whitening is done by estimating the noise matrix $\mathbf{D}$ in regions of the observed spectrum where no interference or transmitters were observed. Two methods will be applied: one method just uses the diagonal

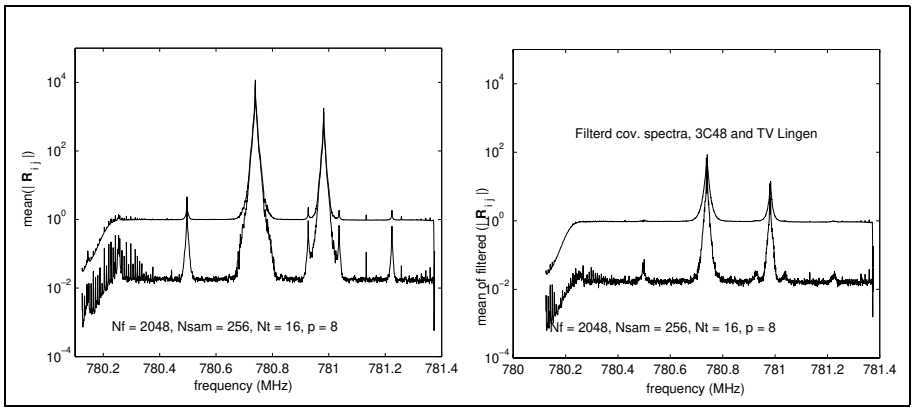


Figure 6.7. Observed autocorrelation (upper) and cross-correlation spectra (lower) of source 3C48 and TV Lingen, before (left) and after (right) applying a projection filter and distortion correction.

of $\mathbf{R}$ after averaging over time, the second estimates it by one of the ad-hoc methods (LOGLS) described in chapter 8. The LOGLS method yields better estimates for the noise matrix $\mathbf{D}$ in case there is interference, but in case there is no interference, the two methods yield the same results. In this section, the interference mitigation effectiveness is illustrated, and related to the subspace structure and observed correction matrix condition numbers.

Television sound carrier waves in an observation of 3C48

Figure 6.7 shows an observation of source 3C48 at 780.75 MHz, which is affected by television transmissions of TV Lingen. The strongest transmission in the data set occurs at 780.75 MHz, and has an INR of 10^4. The figure shows the mean of the observed auto- and cross-correlation spectra before (left) and after (right) applying a spatial projection filter and distortion correction. The two stereo sound carrier waves of the television transmitter TV Lingen are clearly suppressed. The whitened autocorrelation curves and their mean have a magnitude of approximately one (cf. equation 5.43), that is, in frequency bins which are not affected by interference. The whitened cross-correlation curve lies about 17 dB below the autocorrelation curve, this corresponds to the observed source power of 3C48 relative to the receiver noise.

Figure 6.8 shows the observed filter attenuation numbers for the projection filter (left) and for the subtraction filter (right). The expected attenuation from equation 6.30 is a factor 256, which is dominated by the number of short time scale samples N_{sam}. Recall that this attenuation limit, the "finite sample effect", is caused by estimation errors of the direction vector. The theoretical attenuation limit fits in reasonably well with the the observed attenuation numbers, both for the projection filter and the subtraction filter. Note that the attenuation for the autocorrelations is lower than for the cross correlations.

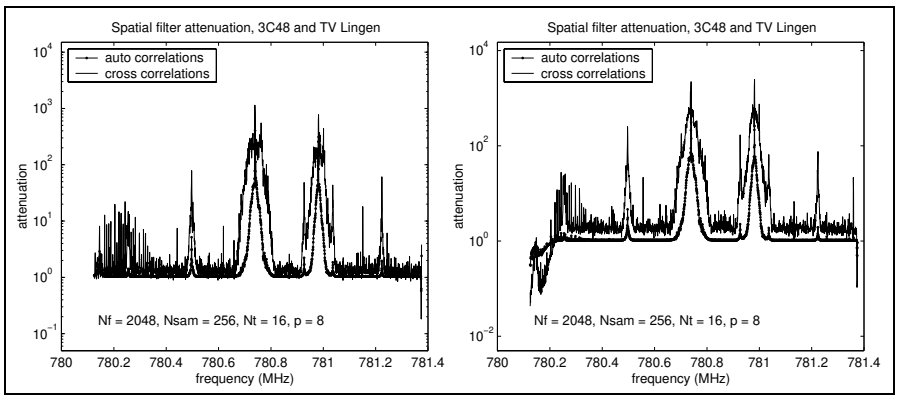


Figure 6.8. Spatial filter attenuation of TV Lingen transmissions of the projection filter (left), and subtraction filter (right).

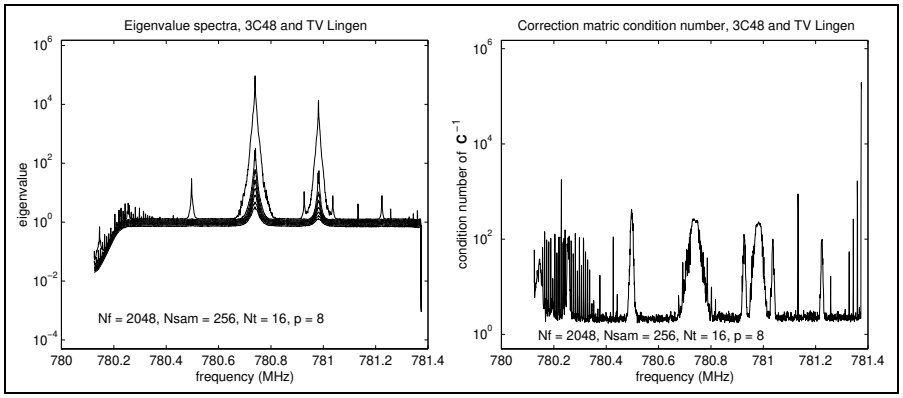


Figure 6.9. Eigenvalue decomposition of observed data with 3C48 and TV Lingen transmissions (left), and projection matrix correction matrix condition numbers (right).

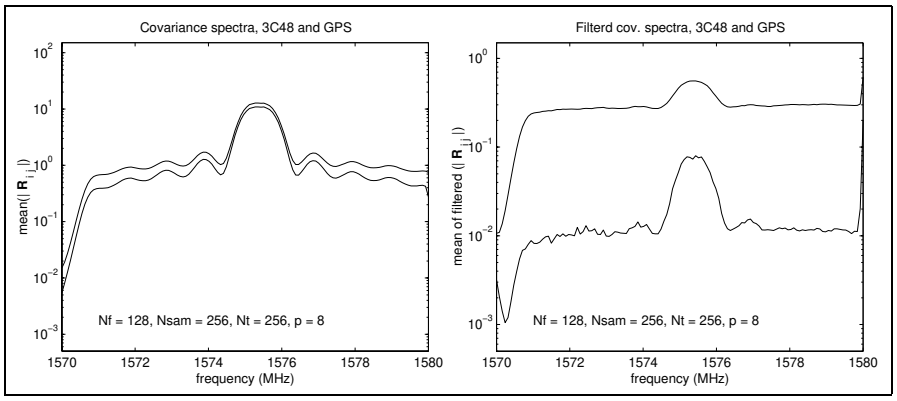


Figure 6.10. Observation of source 3C48 and GPS satellite transmissions. The left-hand figure shows autocorrelation (above) and cross-correlation spectra (below). The right-hand figure shows filtered spectra using a projection filter: autocorrelation curve (above) and cross-correlation curve (below).

This is caused by the fact that the autocorrelation data have an "offset" due to the system noise.

Figure 6.9, left-hand figure, shows the eigenvalue decomposition of the whitened covariance matrix. In the interference-free frequency bins, the expected ratio between the largest and smallest eigenvalues is two (cf. equation 6.30), which fits in with the observed values. The ratio between the largest and second-largest eigenvalue is about 250. This means that by removing the (scaled) subspace corresponding to the largest eigenvalue, the interfering power is reduced by the same factor of 250. This corresponds well to the numbers mentioned above.

Processing the data sets for various frequency bin widths shows that the eigenvalue ratios are not affected by the frequency resolution. The conclusion is that the narrow-band assumption is not violated up to a level of an eigenvalue ratio $\frac{\lambda_1}{\lambda_2}$ of approximately 250. This is in line with the expected value, see for instance equation 6.30.

Figure 6.9, right-hand figure, shows the observed condition number of the correction matrix for the projection filter. In frequency regions without interference, the condition number lies in the range 3 to 4, but when there is interference, the condition number increases to roughly 200. This is expected as the duration of the TV Lingen observation was only 6.7 seconds, and the fringe of the WSRT does not rotate fast enough in these time scales, see for instance section 6.6. Condition numbers of observations of different types of transmitters, with integration times up to 400 s, are shown in section 6.7.3.

GPS transmissions in an observation of 3C48

Another example of an astronomical observation affected by man-made trans-

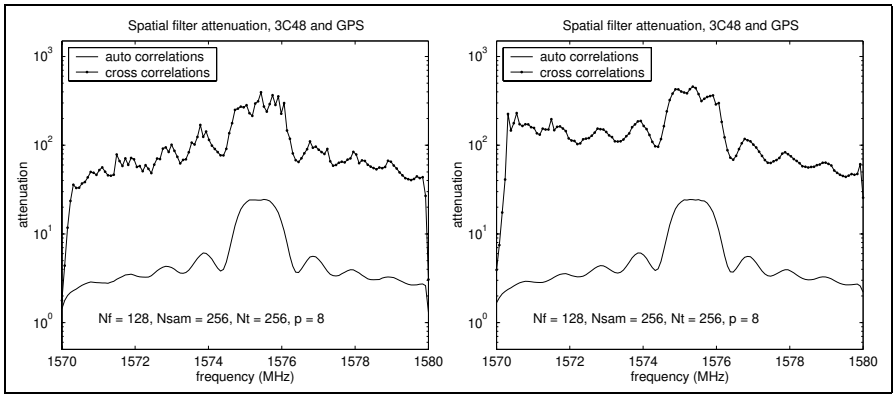


Figure 6.11. Observation of source 3C48 and GPS satellite transmissions. Attenuation spectra are shown for the projection filter (left) and subtraction filter (right).

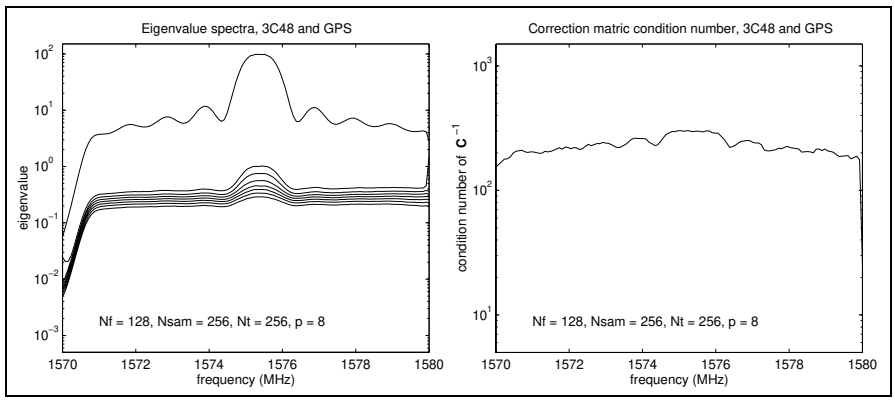


Figure 6.12. Observation of source 3C48 and GPS satellite transmissions. The left-hand figure shows an eigenvalue decomposition of the observed data, the right-hand figure shows the observed condition number spectrum.

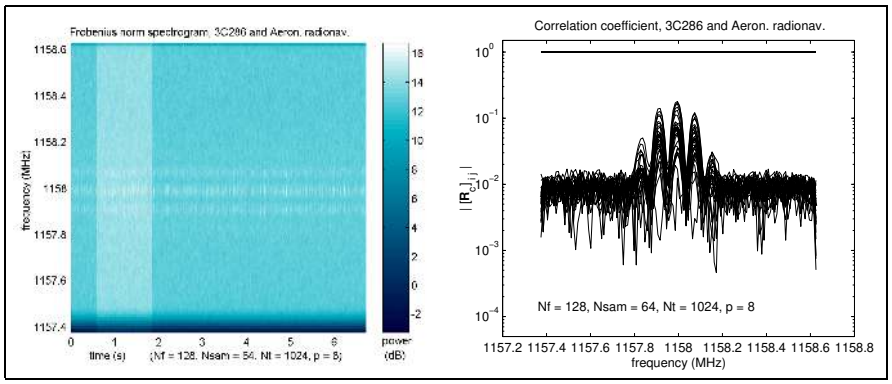


Figure 6.13. Frobenius norm spectrogram of 3C48 and aeronautical radio navigation, Eelde airport (left-hand figure), and auto and cross correlation coefficients (right-hand figure).

missions is shown in figure 6.10. A spectrum is visible of an observation of the source 3C48 together with transmissions at 1575.42 MHz from the GPS satellite system. The left-hand figure shows auto correlation spectra (upper curve) and cross correlation spectra (lower curve). The right-hand figure shows the spectra after projection filtering. The data set was whitened using the diagonal of $\widehat{\mathbf{R}}$ at the edges of the band, at 1571 and 1579 MHz.

Attenuation graphs of projection and subtraction filtering are shown in figure 6.11. Both filters have similar attenuation performance and the maximum observed attenuation at the centre of the band is close to the theoretical value, that is 256, see for instance equation 6.30. This number matches the observed difference between the largest and second-largest eigenvalue ratio. Also here, the observed condition number is about 200. Although the GPS satellites orbit the earth in 12 hours, the resulting fringe rate is too low to generate low condition numbers for the duration of the observation, which was 0.838 s.

6.7.2 Application example of projection filter for intermittent interference: influence of short time scale integration

Radiolocation signals, Eelde airport radar, in an observation of 3C286

Figure 6.13 shows a Frobenius norm spectrogram (left) of a radio location transmission, and the normalized covariance spectra (right). The transmissions are from the Eelde airport radar and consist of sub-ms emitted bursts, visible in the spectrum as a comb with five peaks. Note that in this case the interference is weak: $\mathrm{INR}_{\mathrm{before}}$ as defined in equation (6.27) is less than one. The transmitter is clearly visible in the cross-correlation spectra, but is barely visible in the autocorrelation spectra. The 1.25 s duration interval of increased power in the spectrogram plot, starting at 0.6 s, is caused by switching on the calibration

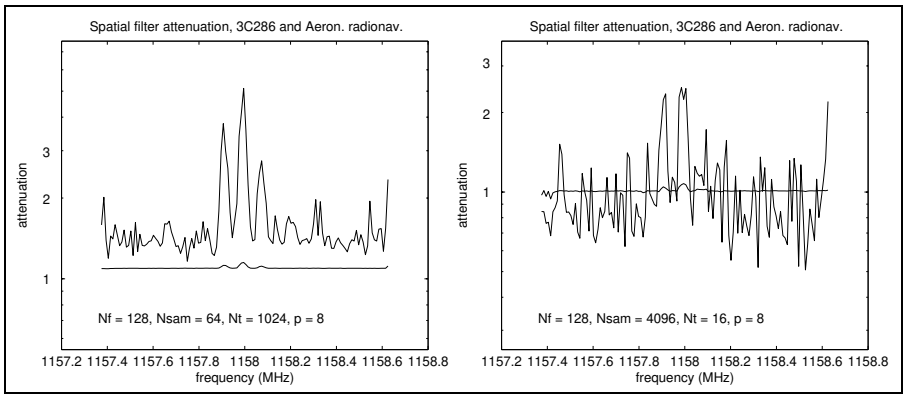


Figure 6.14. Projection filter attenuation for aeronautical radio navigation transmissions, for short time scale integrations $N_{sam} = 64$ (left), and $N_{sam} = 4096$ (right). The smooth curves are autocorrelation curves, the more erratic curves are cross-correlation curves.

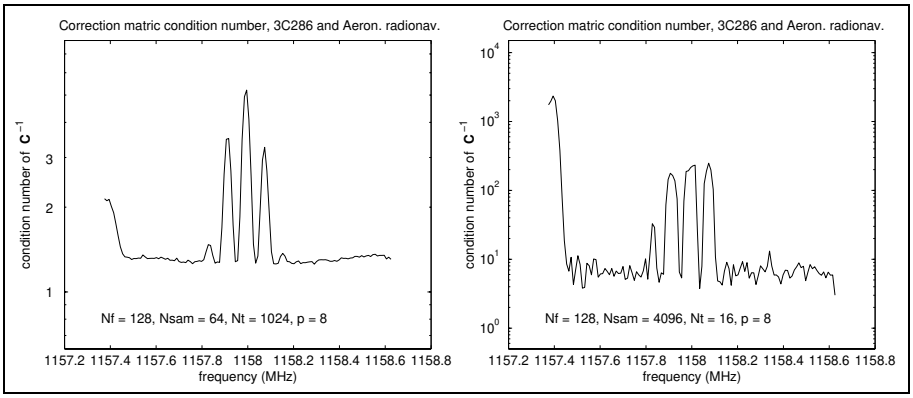


Figure 6.15. Condition number of the projection filter for aeronautical radio navigation transmissions, for short time scale integrations $N_{sam} = 64$ (left), and $N_{sam} = 4096$ (right).

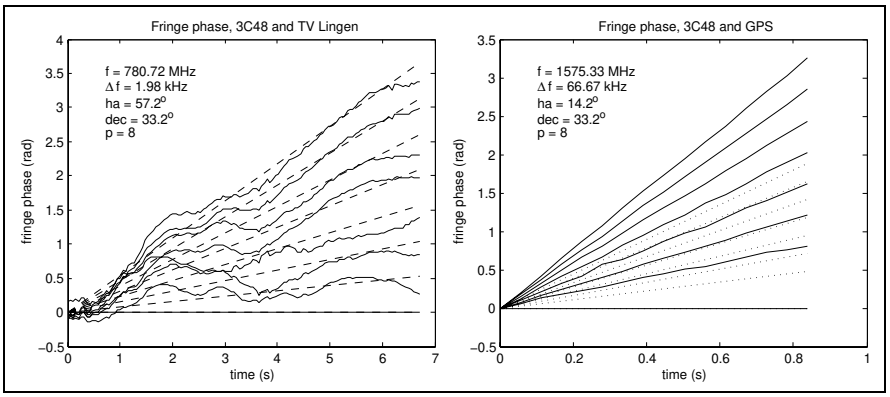


Figure 6.16. The left-hand figure shows telescope fringe phases for TV Lingen observations, measured (solid curves), and predicted (dashed curves). The right-hand figure shows the predicted curves for a transmitter which is fixed to the horizon (dotted curves), and observed curves for the GPS satellite signals.

noise source at the WSRT. The calibration sources generate i.i.d. noise, and can easily be accounted for in the calibration and whitening processes.

Figure 6.14 shows the interference attenuation curves for a projection filter with distortion correction, for sample integrations $N_{sam} = 64$ (left) and for $N_{sam} = 2048$ (right). For low N_{sam} values, the measurement sample interval is close to the transmitted bursts. For large N_{sam} intervals, the transmitted energy is smeared out, which leads to a poorer estimate of the source direction and to a poorer interferer suppression number.

Figure 6.15 shows observed condition numbers for the two filters with the different N_{sam} numbers. The condition number at the frequency where the transmitter has the largest power, is 5 for $N_{sam} = 64$ and 200 for $N_{sam} = 2048$. The number 200 matches the observed number for the TV Lingen transmitter. The condition number 5 for $N_{sam} = 2048$ is much lower than is expected, and is caused by the fact that the sample time is close to the transmitter burst time. The consequence of this match is that there are many observed samples without interference, leading to projection matrices with arbitrary subspace structure.

6.7.3 Observed fringe rates and long-term condition numbers

So far, most of the observed projection matrix condition numbers, in observations with durations of 0.839 to 6.7 seconds, appeared to be in the range 200 to 300. The aim of this section is to investigate the condition number behaviour for longer observation durations, and to relate it to the theoretically expected values derived in section 6.6.

As a first check, predicted and observed fringe phases at the WSRT are

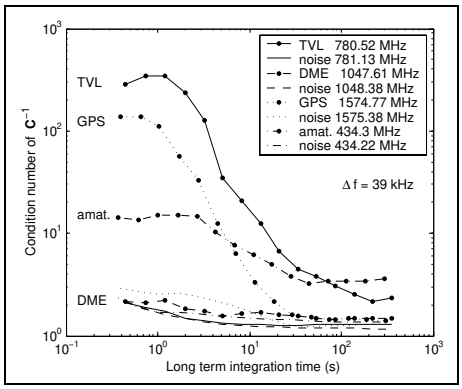


Figure 6.17. The condition number of the correction matrix $\mathbf{C}^{-1}$ for the transmitters TV Lingen, GPS satellites, amateur radio, and DME, are shown as curves with large dots indicating the measured data points. The condition numbers for "clean" parts of the spectrum are shown as curves without large dots.

compared for different transmitters. Figure 6.16 shows them for TV Lingen and GPS. The left-hand figure shows the theoretical fringe phase for the source 3C48 as a function of time for the eight telescope baselines (dashes) and the observed fringe phases (solid). The phase trends are correct, but the phases fluctuate around the predicted curve, most likely because of multipath effects. The right-hand figure also shows the fringe phase for source 3C48 as a function of time for the eight telescope baselines (dotted) and the observed fringe phases (solid). There is a mismatch of a factor 1.8, caused by the fact that GPS is not a transmitter fixed to the horizon, but orbiting the earth (in 12 hours).

Now that it is verified that the observed fringe phases behave as expected, the condition numbers are obtained for a range of long-term integration times. These integration times range from 0.4 to 400 s. The results for four types of interference are shown in figure 6.17 as curves with large dots. An analysis of the expected fringe rates shows that for these transmitters, the theoretical $5.5 \times 2\pi$ criterion (cf. section 6.6) is reached after 20 to 200 seconds. The exact number depends on telescope pointing and frequency. These numbers match with the observed condition number behaviour. The DME curve in the figure lies close to the curves for the RFI-free cases. The reason for this was explained in the previous section. The GPS curve lies below the TV Lingen curve because the fringe rate is large, due to the satellite motion, and because it is observed at a higher frequency. The amateur radio curve lies between the GPS curve and the DME curve. The reason for this is not entirely clear; it could be caused by multipath effects.

6.7.4 Further examples of applied spatial projection filters

Radio location transmissions in an observation of 3C48.

As an additional example, the spatial projection algorithm was applied to an observation of 3C48 containing radiolocation and amateur radio transmissions. Figure 6.18 shows a spectrogram of the 20-minute duration observation (left). In the channel at 433.9 MHz (the one with the very strong intermittent signals) up to about six eigenvalues were affected. After applying the projections, we recovered the original signal by the inversion process described earlier in this paper. Recall that to be able to reconstruct the astronomical signal after projection by applying the inverted $\mathbf{C}$ matrix operation, it is required that the projection matrix $\mathbf{P}_k$ varies sufficiently over the interval which is to be processed. We found that in most cases the interfering signal varies enough to ensure that the correction matrix was invertible. We observed condition numbers in the range of 3 to 20. Only for the continuous interference at 434.3 MHz we found an extremely high value. This is probably related to the fact that the interfering signal was only visible in one of the telescopes, whereas the other signals were visible in all telescopes. The result of the projection is shown in figure 6.18, right. The upper figure shows the mean of all cross-correlations, before and after applying the spatial filter. Both the time-continuous and intermittent interference are suppressed significantly. The exception is the signal at 434.3 MHz; the autocorrelation plot in the lower figure shows that only one of the telescopes receives a strong interfering signal. The conclusion is that, apparently, the interference at the telescope shows enough fluctuations to fulfill the stationarity requirements of the projection matrix. Further study has to demonstrate which fraction of the non-stationarity is caused by fringe stopping (an instrumental phase rotation) and which fraction is caused by other effects (multi-path fading).

Glonass satellite transmissions and observed OH emissions.

Figure 6.19 shows an observation of astronomical OH maser lines at 1612.5 MHz, together with Glonass satellite transmissions. The left-hand figure shows an eigenvalue decomposition which clearly shows that the Glonass transmission at 1614 MHz occupies only one dimension in subspace. Therefore, in principle it could be filtered effectively. But as the $\mathrm{INR}_{\mathrm{before}}$ is low, the filter does not work effectively at these time scales.

6.8 Conclusions and further research

In this chapter the effectiveness of spatial filters was derived and demonstrated for observed data. Both projection and subtraction filters can be applied. Each has its own specific advantages and disadvantages. The advantage of projection filters is that only the direction of the interferer is required, not estimates of its power. The main disadvantage is that projection filters need a distortion correction which puts constraints on the stationarity of the observed transmitter signals and which requires $\mathcal{O}(p^3)$ operations. An advantage of subtraction filters

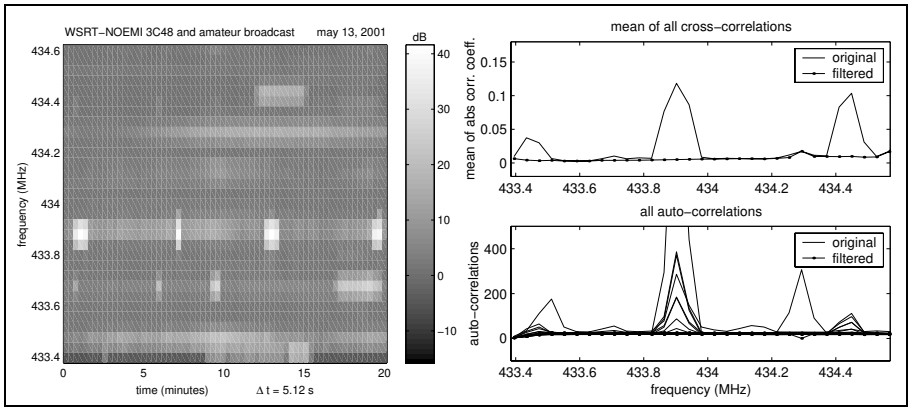


Figure 6.18. Observation of the astronomical source 3C48, in which amateur radio and/or radio location transmissions are present. Both time-continuous and intermittent transmissions are present. Shown to the left is a spectrogram of the largest eigenvalue obtained via eigenvalue decomposition of the telescope covariance matrices. The right-hand figure shows the mean cross correlation spectrum and all auto- correlation spectra before applying spatial projections (above), and after applying the spatial projections (below).

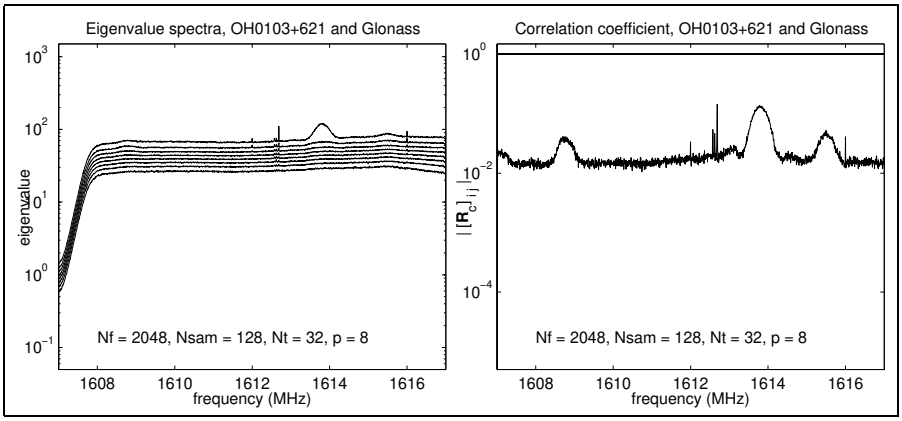


Figure 6.19. Astronomical OH maser source at 1612.5 MHz and Glonass satellite signals: eigenvalue decomposition (left), and mean of all absolute cross correlations (right).

is that they are relatively simple to apply, but they require more knowledge of the system noise and they require a good estimate of the interferer power.

Further research, for instance a careful analysis of the biases involved in the subtraction procedures, would be useful. Also a study of combining information in adjacent frequency channels, for the purpose of spatial filtering, would be useful.

Chapter 7

Spatial filtering using a reference

Assuming an array of telescopes, we have previously considered spatial filtering techniques based on projecting out the interferer array signature vector. In this chapter[1], we will consider extending the astronomical array with a reference antenna (or array), and develop spatial filtering algorithms for this situation. It is shown that the information from the reference antenna improves the quality of the interferer signature vector estimation, hence more of the interference can be projected out. The conditioning of the problem improves as well. The algorithms are tested both on simulated and experimental data.

7.1 Introduction

With an interferometer array of p telescope dishes, spatial filtering techniques are applicable. The desired instrument outputs in this case are $p \times p$ correlation matrices, integrated to 10 s. Based on short-term correlation matrices (integration to e.g., 10 ms) and narrow subband processing, the array signature vector of an interferer can be estimated and subsequently projected out [119]. This technique was described in chapter 6.

To improve the performance for weak or stationary interferers, in this chapter we will consider extending the telescope array with one or more reference antennas. In general, a higher gain (interference-to-noise ratio) than that obtained from an omnidirectional antenna is needed to expect any benefits. Most flexibility is obtained by using a phased array which can adaptively be pointed towards the strongest interferers. We have actually built a demonstrator set-up along these lines, utilising a wideband phased array of 64 elements (see section 4.4). The reference signal is correlated along with the telescope signals as if it

[1]This chapter was published in [152]

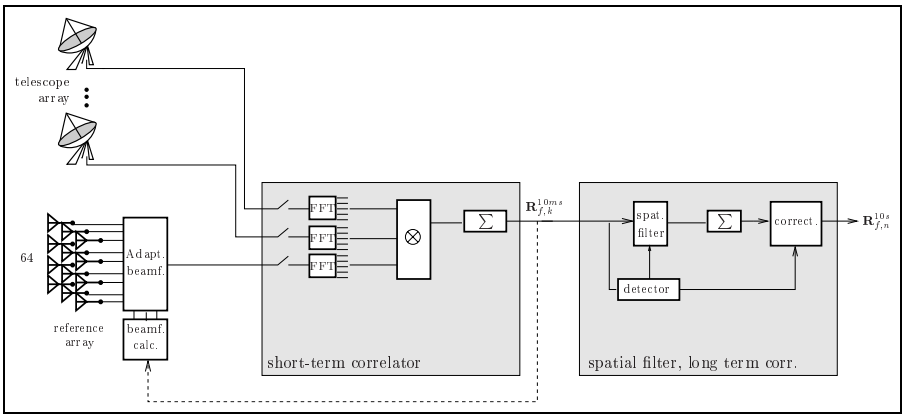


Figure 7.1. Telescope array augmented with a reference phased array

were an additional telescope, and spatial filtering algorithms can be applied to the resulting short-term integrated covariance matrices. This set-up is shown in figure 7.1. Spatial filtering on extended arrays was first considered by Briggs et al. [29] for a single dual-polarised telescope (two channels) and two reference antennas. With their technique a single interferer can be cancelled. Jeffs et al. [77, 78] propose spatial filtering algorithms along the lines of [119]; we will summarise their approach in section 7.3.2 and subsequently make extensions which may improve performance.

7.2 Problem definition

7.2.1 Data model

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

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

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

where $\mathbf{x}_0(t) = [x_1(t), \ldots, x_{p_0}(t)]^t$ is the $p_0 \times 1$ vector of telescope signals at time t, $\mathbf{v}_0(t)$ is the received sky signal, assumed on the time scale of 10 s to be a stationary Gaussian vector with covariance matrix $\mathbf{R}_{v,0}$ (the astronomical "visibilities"), and $\mathbf{n}_0(t)$ is the $p_0 \times 1$ noise vector with independent identically

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

distributed Gaussian entries and covariance matrix $\sigma_0^2\mathbf{I}$. The astronomer is interested in $\mathbf{R}_{v,0}$.

If an interferer is present and the processing bandwidth is sufficiently narrow, then the primary array output is modelled as

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

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

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

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

It is assumed here that the contribution of the astronomical sources to the reference signals is negligible. The noise on the reference antennas is assumed to be i.i.d. Gaussian with covariance matrix $\sigma_1^2\mathbf{I}$. Stacking all antenna signals in a single vector $\mathbf{x}^T = [\mathbf{x}_0^T \;\; \mathbf{x}_1^T]^T$, we obtain

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

We make the following additional assumptions with regard to this model:

(A1) The noise variances σ_0^2 and σ_1^2 are known from calibration.

(A2) $\mathbf{R}_{v,0} \ll \sigma_0^2\mathbf{I}$. This is reasonable as even the strongest sky sources are about 15 dB under the noise floor.

(A3) The interferer signature $\mathbf{a}(t)$ is stationary over short processing times (say 10 ms). It may or may not vary over longer periods.

This was the model considered in [119]. The model is easily extended to multiple interferers, in which case we obtain

$$\mathbf{x}(t) = \mathbf{v}(t) + \mathbf{A}(t)\mathbf{s}(t) + \mathbf{n}(t) \Leftrightarrow \left\{ \begin{array}{rcll} \mathbf{x}_0(t) & = & \mathbf{v}_0(t) + & \mathbf{A}_0(t)\mathbf{s}(t) + \mathbf{n}_0(t) \\ \mathbf{x}_1(t) & = & & \mathbf{A}_1(t)\mathbf{s}(t) + \mathbf{n}_1(t) \end{array} \right.$$

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

7.2.2 Covariance model

Let observations $\mathbf{x}[n] = \mathbf{x}(nT_s)$ be given, 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 $\widehat{\mathbf{R}}_k$,

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

where M is the number of samples per short-term average. All interference filtering algorithms in this chapter are based on applying operations to each $\widehat{\mathbf{R}}_k$ 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 $\widehat{\mathbf{R}}_k$ is denoted by $\mathbf{R}_k$, which can be written in block-partitioned form as

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

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

$$\begin{aligned} \mathbf{R}_k &= \mathbf{\Psi} + \mathbf{A}_k\mathbf{A}_k^H = \mathbf{R}_v + \mathbf{\Sigma} + \mathbf{A}_k\mathbf{A}_k^H \\ &= \left[\begin{array}{c|c} \mathbf{R}_{v,0} + \mathbf{A}_{0,k}\mathbf{A}_{0,k}^H + \sigma_0^2\mathbf{I} & \mathbf{A}_{0,k}\mathbf{A}_{1,k}^H \\ \hline \mathbf{A}_{1,k}\mathbf{A}_{0,k}^H & \mathbf{A}_{1,k}\mathbf{A}_{1,k}^H + \sigma_1^2\mathbf{I} \end{array}\right] \end{aligned} \tag{7.1}$$

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

7.3 Algorithms

7.3.1 Traditional subtraction technique

In array signal processing, a classical technique for interference removal using a reference antenna is based on taking the covariance of the primary antennas, $\mathbf{R}_{00,k}$, and subtracting the estimated contribution of the interferers, $\mathbf{A}_{0,k}\mathbf{A}_{0,k}^H$. In effect, the rank deficiency of the interference term

$$\mathbf{A}\mathbf{A}^H = \begin{bmatrix} \mathbf{A}_0\mathbf{A}_0^H & \mathbf{A}_0\mathbf{A}_1^H \\ \mathbf{A}_1\mathbf{A}_0^H & \mathbf{A}_1\mathbf{A}_1^H \end{bmatrix}$$

is exploited: if $q \le p_1$ and moreover $\mathbf{A}_1 : p_1 \times q$ has full column rank q, then the first p_0 columns must be linear combinations of the remaining p_1. Under these conditions,

$$\mathbf{A}_0\mathbf{A}_0^H = \mathbf{A}_0\mathbf{A}_1^H(\mathbf{A}_1\mathbf{A}_1^H)^\dagger\mathbf{A}_1\mathbf{A}_0^H$$

where $\dagger$ indicates the pseudo-inverse, and hence a 'clean' instantaneous covariance estimate is

$$\mathbf{\Psi}_{00,k} = \widehat{\mathbf{R}}_{00,k} - \widehat{\mathbf{R}}_{01,k}\widehat{\mathbf{R}}_{11,k}^\dagger\widehat{\mathbf{R}}_{10,k}$$

(ignoring the effect of the noise term $\sigma^2\mathbf{I}$). The final 'clean' covariance estimate is obtained by averaging over N such matrices to obtain a long-term estimate

$$\mathbf{\Psi}_{00} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{\Psi}_{00,k}\,.$$

Briggs et al. [29] essentially derive this algorithm and several variants of it, for the special case of $q = 1$ and $p_1 = 2$. Jeffs et al. [77] describe the same technique as a generalisation of the classical Multiple Sidelobe Canceller.

The conditions mentioned on $\mathbf{A}_1$ entail that this technique can be used for at most p_1 interferers, and only if the reference antennas are sufficiently independent so that they receive independent linear combinations of the interferers. Unlike some of the techniques to be discussed in later sections, the technique does not rely on the variation of $\mathbf{A}_k$: in principle, $\mathbf{A}_k$ can be stationary. Also, no detection of the number of interferers is done, nor of any noise powers. This simplifies the algorithm but might also limit its performance.

7.3.2 Spatial filtering using projections

In chapter 6, a spatial filtering algorithm based on projections was introduced. Although this algorithm did not assume the presence of reference antennas, it can also be used in our current situation. Briefly, the method is as follows.

Suppose that an orthogonal basis $\mathbf{U}_k$ of the subspace spanned by interferer spatial signatures $\mathrm{span}(\mathbf{A}_k)$ is known. We can then form a spatial projection matrix $\mathbf{P}_k := \mathbf{I} - \mathbf{U}_k\mathbf{U}_k^H$ which is such that $\mathbf{P}_k\mathbf{A}_k = 0$. When this spatial filter is applied to the data covariance matrix, all the energy due to the interferer will be nulled: let

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

then

$$\mathcal{E}\{\widehat{\mathbf{Q}}_k\} = \mathbf{P}_k\mathbf{\Psi}\mathbf{P}_k$$

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

$$\widehat{\mathbf{Q}} := \frac{1}{N}\sum_{k=1}^{N}\widehat{\mathbf{Q}}_k = \frac{1}{N}\sum_{k=1}^{N}\mathbf{P}_k\widehat{\mathbf{R}}_k\mathbf{P}_k\,. \tag{7.2}$$

$\widehat{\mathbf{Q}}$ is an estimate of $\mathbf{\Psi}$, but it is biased due to the projection. To correct for this we first write the two-sided multiplication as a single-sided multiplication employing the matrix identity $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^t \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$, This gives

$$\mathrm{vec}(\widehat{\mathbf{Q}}) = \frac{1}{N}\sum_{k=1}^{N}\mathbf{C}_k\mathrm{vec}(\widehat{\mathbf{R}}_k) \quad \text{where } \mathbf{C}_k := \mathbf{P}_k^t \otimes \mathbf{P}_k\,. \tag{7.3}$$

If the interference was completely removed, then

$$\mathcal{E}\mathrm{vec}(\widehat{\mathbf{Q}}) = \frac{1}{N}\sum_{k=1}^{N}\mathbf{C}_k\mathrm{vec}(\mathbf{\Psi}) = \mathbf{C}\mathrm{vec}(\mathbf{\Psi})\,; \quad \mathbf{C} := \frac{1}{N}\sum_{k=1}^{N}\mathbf{C}_k\,. \tag{7.4}$$

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

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

If the interference was completely projected out, then $\mathbf{\Psi}$ is an unbiased estimate of the covariance matrix without interference. This was the algorithm introduced in chapter 6.

The reconstructed covariance matrix is size $p \times p$. In the present case, we are only interested in the submatrix corresponding to the primary antennas. Hence, the estimate produced by the algorithm is the $p_0 \times p_0$ submatrix in the top-left-hand corner, $\mathbf{\Psi}_{00}$. This is one of the algorithms introduced in [77].

The spatial signature of the interferer is generally unknown, but it can be estimated from an eigen-analysis of the sample covariance matrices $\widehat{\mathbf{R}}_k$. To do this, recall that the noise powers on the two antenna arrays are not necessarily the same, and they will first have to be made equal. This noise whitening is done by working with $\mathbf{\Sigma}^{-1/2}\widehat{\mathbf{R}}_k\mathbf{\Sigma}^{-1/2}$. Without interference and assuming that $\mathbf{R}_v$ is negligible compared to $\mathbf{\Sigma}$, all eigenvalues of this matrix are expected to be close to 1. With q interferers, q eigenvalues become larger, and the eigenvectors corresponding to these eigenvalues are an estimate of $\mathrm{span}(\mathbf{A}_k)$.

7.3.3 Improved spatial filter with projections

We will now derive an improved algorithm. Computing the projections and long-term average of the projected estimates $\widehat{\mathbf{Q}}$ as before in (7.2). Then (7.4) applies:

$$\mathcal{E}\mathrm{vec}(\widehat{\mathbf{Q}}) = \mathbf{C}\mathrm{vec}(\mathbf{\Psi})\,.$$

Based on this, we previously set $\mathrm{vec}(\mathbf{\Psi}) = \mathbf{C}^{-1}\mathrm{vec}(\widehat{\mathbf{Q}})$, which is the solution in Least Squares sense to the covariance model error minimisation problem, $\|\mathrm{vec}(\widehat{\mathbf{Q}}) - \mathbf{C}\mathrm{vec}(\mathbf{\Psi})\|^2$. Now, instead of this, partition $\mathbf{\Psi}$ into 4 submatrices as in (7.1)s. Since we are only interested in recovering $\mathbf{\Psi}_{00}$, the other submatrices in $\mathbf{\Psi}$ are replaced by their expected values, $\mathbf{\Psi}_{01} = \mathbf{0}$, $\mathbf{\Psi}_{10} = \mathbf{0}$, $\mathbf{\Psi}_{11} = \sigma_1^2\mathbf{I}$ respectively. This corresponds to solving the reduced-size covariance model error minimisation problem,

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

The solution to this problem can be reduced to a standard LS problem by separating the knowns from the unknowns. Thus, the entries of $\mathrm{vec}(\mathbf{\Psi})$ are rearranged into

$$\left[\begin{array}{c} \mathrm{vec}(\mathbf{\Psi}_{00}) \\ \hline \sigma_1^2\mathbf{1} \\ \mathbf{0} \end{array}\right]$$

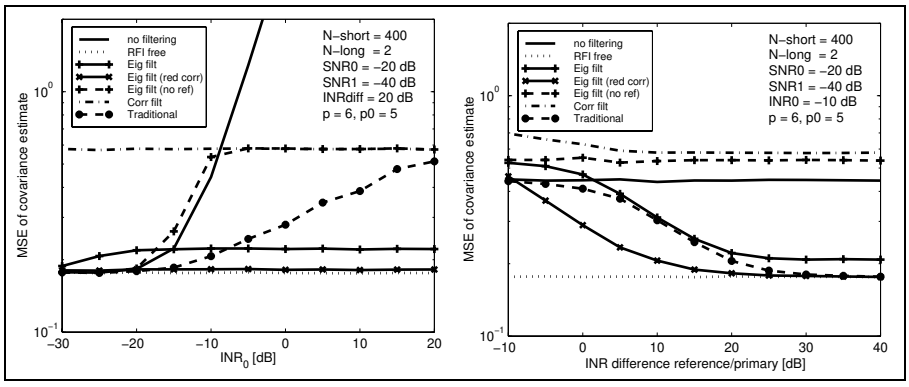


Figure 7.2. MSE, left as a function of interferer power at the reference antenna, right as a function of the interferer power difference between the reference antenna and the primary array elements.

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

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

The advantage compared to the preceding algorithm is that $\mathbf{C}_1$ is a tall matrix, and better conditioned than $\mathbf{C}$. This improves the performance of the algorithm in cases where $\mathbf{C}$ is ill-conditioned, for example, for stationary interferers, or an interferer entering at only a single telescope. Asymptotically for large INR of the reference array, the algorithm is found to behave similar to the traditional subtraction technique.

7.4 Simulations

We will first test the performance of the algorithms in a simulation set-up. We will use $p = 6$ antennas, with $p_0 = 5$ primary antennas (telescopes) and $p_1 = 1$ reference antenna. For simplicity, the array is a uniform linear array with half-wavelength spacing and the same noise power on all antennas. The astronomical source is simulated by a source with a constant direction-of-arrival of 10° with respect to array broadside. The source has $\mathrm{SNR}_0 = -20$ dB with respect to each primary array element, and $\mathrm{SNR}_1 = -40$ dB for the reference antenna. The interferer is simulated by a source with a randomly generated

and varying complex $\mathbf{a}_k$ (i.i.d. entries, each complex normal distributed), and varying INRs as explained in the simulations. This corresponds to a Rayleigh fading interferer.

The following algorithms are compared:

- the subtraction method in section 7.3.1 denoted 'traditional',
- the spatial filtering algorithm using projections and eigenvalue computations, section 7.3.2, denoted 'eig filt',
- the spatial filtering algorithm with reduced-size covariance reconstruction, section 7.3.3, denoted 'eig filt (red corr)',
- for comparison, the spatial filtering technique without reference antenna, denoted 'eig filt (no ref)', the covariance estimate without RFI ('RFI free'), and the estimate obtained without any filtering ('no filtering').

Figure 7.2 shows the mean-squared-error (MSE) of the primary filtered covariance estimate compared to the theoretical value $\mathbf{R}_{v,0} + \sigma_0^2\mathbf{I}$, for varying interferer power INR_0 and interferer array gain $INR_1 - INR_0$ respectively. Here, we took $M = 400$ short-term samples and $N = 2$ long-term averages, which is unrealistically small but serves to illustrate the effect of limited variability of $\mathbf{a}_k$ (only two different vectors).

It is seen that the new algorithm has a great advantage over the spatial filtering algorithm without reference antenna in case the $\mathbf{a}_k$-vector does not sufficiently vary. The MSE performance is flat for varying INR and INR difference, which is very desirable. Moreover, it is very close to the RFI-free case. The new algorithm is also often better than the subtraction technique.

Additional simulations, not shown here, indicate that if the interferer enters only at one telescope and at the reference antenna, then the algorithm without a reference antenna performs poorly: it cannot reconstruct the contaminated dimension. The algorithm with reference antennas performs fine.

7.5 Experiment

To test the algorithm on actual data, we have made a short observation of the strong astronomical source 3C48 contaminated by Afristar satellite signals. The set-up follows figure 7.1. The primary array consists of $p_0 = 6$ of the 14 telescope dishes of the Westerbork Synthesis Radio Telescope (WSRT). As a reference we use the THEA tile discussed in section 4.4. One beam was pointed approximately at the satellite, the other was used for scanning. We recorded 65 kSamples at 20 MS/s, and processed these offline. After short-term windowed Fourier transforms, the data was split into 64 frequency bins, correlated, and averaged over 32 samples to obtain 16 short-term covariance matrices.

The resulting auto- and cross-correlation spectra after filtering are shown in figure 7.3. The autocorrelation spectra are almost flat, and close to 1 (the

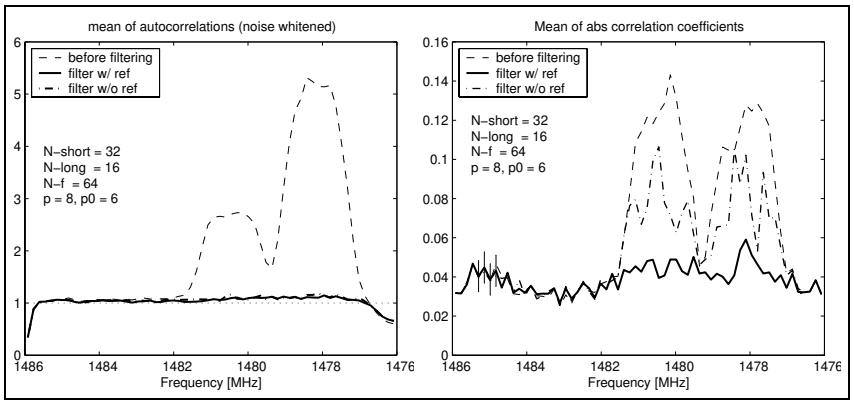


Figure 7.3. Averaged autocorrelation spectrum before and after filtering (left), averaged cross-correlation spectrum before and after filtering (right)

whitened noise power). The cross-correlation spectra show that spatial filtering with reference antenna much better removed the interference than filtering without reference antenna. The residual correlation of about 4% is known to be the SNR of the astronomical source. The lines are noisy due to the finite sample effect; the predicted standard deviation (based on number of samples averaged) is indicated for a few frequencies.

Chapter 8

Gain calibration

In radio telescope arrays, the complex receiver gains and sensor noise powers are initially unknown and have to be calibrated. Gain calibration can enhance the quality of astronomical sky images and, moreover, improve the effectiveness of array signal processing techniques for interference mitigation and spatial filtering. It is challenging that the signal-to-noise ratio is usually much below 0 dB even for the brightest sky sources. The calibration method considered here consists of observing a single point source and extracting the gain and noise parameters from the estimated covariance matrix. We will present several closed form and iterative identification algorithms for this[1]. Weighted versions of the algorithms will be proved asymptotically efficient. The algorithms will be validated by simulations and by application to experimental data observed at the Westerbork Synthesis Radio Telescope.

8.1 Introduction

In contrast to most communications systems, in radio astronomy the sources of interest are usually much weaker than the instantaneous system noise levels, with signal to noise (SNR) levels much below 0 dB even for the strongest sources. Integration times of several hours to more than ten hours are needed to obtain sky images with acceptable sensitivity. Astronomical sources can be broadband or narrowband, and accurate estimates of the telescope gains are necessary to compute the astronomical source power distribution, in radio astronomy also known as the surface brightness.

For an array of telescopes (figure 8.1), not only the gain of the main beam of each telescope needs to be estimated, but also the phase differences between the telescopes, and the power of the noise present at each receiver. Two techniques are widely used. In imaging applications, a "blind" iterative self-calibration technique [113,146] starts from an initial estimate of the parameters and adapts

[1]This chapter was published in [21]

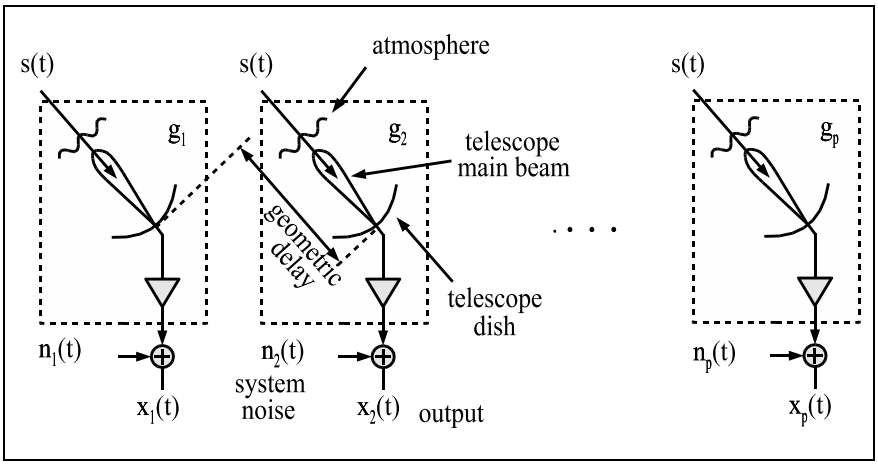


Figure 8.1. Radio telescope array

them until the resulting image matches a prior parametric model of the field of interest (usually a point source model). A second standard technique is to obtain dedicated calibration observations of a part of the sky that contains a single known, relatively strong point source, and this is the technique that we are considering in this chapter.

Several algorithms to estimate the parameters from these observations have existed for a long time [138, 37, 169], and until recently their performance was considered satisfactory. For example, a typical 12-hour observation at the Westerbork Synthesis Radio Telescope (WSRT, a 3-km linear array of 14 telescope dishes of 25 meter diameter located in The Netherlands) is usually calibrated with two dedicated short calibration observations prior to and after the 12-hour observation run. At 1420 MHz and under optimal atmospheric and ionospheric conditions, the resulting gain phase accuracy is about 5 degrees, and the gain magnitude accuracy about a few percent. The gain estimates are assumed to be valid for any direction in the sky. If a better accuracy is required or if the atmospheric and ionospheric conditions are unstable, intermediate calibration measurements need to be carried out, and calibration sources closer to the astronomical source of interest can be selected. Since these are usually weaker, longer integration periods are needed to reach sufficient accuracy. Observation time is expensive, and there is a growing need for data-efficient estimation algorithms.

The advent of a new generation of radio telescopes such as the Square Kilometer Array radio telescope or SKA [134] has sparked new interest in the issue. SKA will be a phased array of about 10^6 elements with receivers that are not as well matched and much noisier than the classical arrays. Hence array calibration will be both harder and more important. A proposed low-frequency array (LOFAR) will suffer from atmospheric disturbances that can vary within minutes, and calibration will have to be online and work with short data sets,

so that efficient algorithms are needed. A second reason for renewed interest in gain calibration issues is the recent attention for radio frequency interference (RFI) suppression techniques. Advanced array signal processing techniques such as spatial filtering are very adequate in this context [98], but most algorithms rely on spatially white noise models which are valid only after accurate calibration. In this chapter we will study calibration techniques valid for the 14-dish Westerbork array, but the results are more generally applicable.

As indicated before, we will consider the standard procedure for estimating the complex gain and noise power of each telescope, which starts by pointing the telescopes at a relatively strong astronomical source (e.g., SNR= -20 dB). Each telescope output signal is the sum of the telescope system noise (uncorrelated among the telescopes) and the astronomical source flux, which is correlated, multiplied by the telescope gain. The source flux is the same for each of the telescopes, but telescope gain and noise power usually are not. The gain consists of the combined effect of atmospheric disturbances, telescope geometry, receiver characteristics, and electronic (amplifier) gain. The noise consists mostly of thermal system noise, and differs between the receivers by several dBs.

The output of the back-end processing stage is a sequence of covariance matrices formed by cross-correlation of all telescope outputs. In this chapter we aim to estimate the complex gain factors and the system noise power from a recorded covariance matrix, assuming that we observe a single astronomical source whose flux is known from tables. We will present four algorithms to extract these parameters.

The telescope gains are frequency dependent, and therefore the received frequency band is divided into sub-bands and the parameter estimation is repeated for each frequency. We assume that the sub-bands are sufficiently narrow, so that the geometric delay compensation (i.e. the delay of the astronomical signal across the array) can be carried out via phase rotations. Under the narrowband assumption, it is sufficient to discuss the parameter estimation at a single frequency.[2]

As we will show in section 8.2, the calibration problem is essentially reduced to estimating the parameters of a $p \times p$ covariance matrix of the form

$$\mathbf{R} = \mathbf{g}\mathbf{g}^H + \mathbf{D} \tag{8.1}$$

where $\mathbf{g}$ is a gain vector and $\mathbf{D}$ a diagonal matrix representing the noise covariance. This problem is not entirely unknown in signal processing. There is some recent literature on direction-of-arrival estimation in colored noise, with models of the form $\mathbf{R} = \mathbf{a}(\theta)\mathbf{a}(\theta)^H + \mathbf{D}$, where $\mathbf{a}(\theta)$ is an array response vector corresponding to the direction of arrival of a source. This problem and generalisations to multiple sources have been considered for example in [114, 140]. There is also a significant body of earlier research that considers more general

[2] If a model is assumed for the frequency behaviour of the parameters, then the information from different frequencies should obviously be combined. This is outside the scope of this chapter.

structured noise models, see [55, 111] and references therein. Our problem differs in that we consider an *unstructured* complex vector $\mathbf{g}$ and a *diagonal* noise covariance matrix $\mathbf{D}$. This permits some interesting closed-form estimators.

The model (8.1) also appears in the statistical literature as a (rank-1) *factor analysis* problem [105,91]. Factor analysis is a mature field which has seen much activity in the 1960-1980s. Although quite relevant to array signal processing, the results are apparently little known in this field. Many results also need to be translated to the complex domain.

Our main contribution in this chapter is that we will give several new algorithms for estimating the gain and noise parameters, including simple closed-form algorithms. We will start by posing the data model, deriving the Cramer-Rao bound, and formulating a Maximum Likelihood (ML) estimation problem (section 8.2). This does not directly lead to useful algorithms, since the number of parameters to be estimated can be large ($(3p-1)$ where the number of sensors p is in the order of 10–60 for classical arrays, and potentially much more in future arrays). We will then present an asymptotically efficient least squares (LS) cost function (section 8.3.1), and subsequently derive several iterative and closed-form algorithms (sections 8.3.2 to 8.3.5). The algorithms will first be verified using simulations (section 8.4), and then applied to experimental data collected at the WSRT (section 8.5).

8.2 Data model and preliminary results

8.2.1 Data model description

Consider a telescope array as in figure 8.1, and assume that during the calibration observation the telescopes are pointed at a single radio source in the sky, placed in the center of the field of view. Let $x_i(t)$ be the complex baseband signal at a certain frequency f at the output of the receiver of element i at time t. We assume that the frequency bin is sufficiently narrow for the maximal propagation delay of a signal across the array to be much smaller than the inverse bandwidth, so that it may be represented by a phase shift. Then $x_i(t)$ can be modelled as

$$x_i(t) = \gamma_i e^{\imath\phi_i} a_i s(t) + n_i(t)$$

where γ_i is the overall amplitude gain of the receiver system and the atmospheric disturbances, ϕ_i is the corresponding phase shift, $s(t)$ is the flux of the impinging source signal, $a_i = e^{-\imath 2\pi f \tau_i}$ is the phase shift due to the propagation delay τ_i of $s(t)$ across the array, as compared to its arrival at the first element, and n_i is the system noise.

Assuming that we have p elements, we can stack the $x_i(t)$ into a vector $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_p(t)]^t$. Similarly, we define

$$\mathbf{g} = \left[\ \gamma_1 e^{\imath\phi_1} a_1, \cdots, \gamma_p e^{\imath\phi_p} a_p\ \right]^t$$

and $\mathbf{n}(t) = [n_1(t), n_2(t), \cdots, n_p(t)]^t$. We thus arrive at the data model

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

We make the following model assumptions:

1. The source signal is zero mean Gaussian, temporally independent identically distributed (i.i.d.), with signal power $\sigma_s^2 = \mathcal{E}\{|s(t)|^2\}$ known from tables.

2. The noise signal is zero mean Gaussian, temporally i.i.d., spatially uncorrelated among the sensors, with unknown power $\mathcal{E}\{|n_i|^2\} = \sigma_i^2$, and independent of the source signal.

3. The gains γ_i and phases ϕ_i are unknown constants.

4. The geometry and looking directions of the telescopes are known, so that the phase shifts a_i are known.

Assume that we have collected N independent samples $\mathbf{x}_n = \mathbf{x}(nT)$, $n = 0, \cdots, N-1$, where T is the sample period. Since the signals are Gaussian, all information on the parameters is located in the second order covariance. Let the true covariance matrix $\mathbf{R}$ and its sample estimate $\widehat{\mathbf{R}}$ be

$$\mathbf{R} = \mathcal{E}\{\mathbf{x}_n \mathbf{x}_n^H\}, \quad \widehat{\mathbf{R}} = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{x}_n \mathbf{x}_n^H \tag{8.2}$$

Since the noise is uncorrelated to the signal, we obtain

$$\mathbf{R} = \sigma_s^2 \mathbf{g}\mathbf{g}^H + \mathbf{D} \tag{8.3}$$

where $\mathbf{D} = \mathrm{E}\{\mathbf{n}\mathbf{n}^H\} = \mathrm{diag}[\sigma_1^2, \cdots, \sigma_p^2]$ is the noise covariance matrix, a diagonal matrix containing the noise powers.

Given $\widehat{\mathbf{R}}$, our objective in this chapter will be to estimate $\mathbf{g}$ and $\mathbf{D}$. Since σ_s^2 is known from tables, we can make it equal to 1 without loss of generality. Also, the a_i are known from the geometry and looking direction of the telescopes; hence without loss of generality we make $a_i = 1$ as well.

We will use the following real-valued parametrisation of the model. Define a factorisation of $\mathbf{g}$ into a magnitude vector $\boldsymbol{\gamma}$ and a phase vector $e^{\imath\boldsymbol{\phi}}$ as

$$\mathbf{g} = \boldsymbol{\gamma} \odot e^{\imath\boldsymbol{\phi}}$$

It is clear that the phases in $\mathbf{g}$ are underdetermined, so that we define the phase of the first entry as zero, $\phi_1 = 0$. The parameter vector to be estimated is thus

$$\boldsymbol{\theta} = [\boldsymbol{\gamma}^t,\ \boldsymbol{\phi}^t,\ \mathbf{d}^t]^t = [\gamma_1, \cdots, \gamma_p,\ \phi_2, \cdots, \phi_p,\ d_1, \cdots, d_p]^t$$

where $d_i = \sigma_i^2$.

8.2.2 Cramer-Rao lower bound

The Cramer-Rao Bound (CRB) gives a lower bound to the variance of any unbiased estimator $\hat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$. In our situation, we assume that the source signal and the channel noise are complex independent Gaussian distributed with zero mean, and satisfy the model in equation (8.3) with $\sigma_s = 1$. Following standard techniques [65, 81], the CRB is known to be given by the diagonal entries of

$$\mathbf{C}_{CRB} = \frac{1}{N}\mathbf{M}^{-1}$$

where $\mathbf{M}$ is the (scaled) Fisher information matrix (FIM), which can be written as (e.g., [142])

$$\mathbf{M} = \mathbf{F}_0^H(\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1})\mathbf{F}_0 \tag{8.4}$$

Here, $\mathbf{F}_0 := \mathbf{F}(\boldsymbol{\theta}_0)$ is the Jacobian evaluated at the true value of the parameters $\boldsymbol{\theta}_0$,

$$\mathbf{F}(\boldsymbol{\theta}) = [\mathbf{F}_\gamma,\ \mathbf{F}_\phi,\ \mathbf{F}_d](\boldsymbol{\theta}) \tag{8.5}$$

$$\mathbf{F}_\gamma(\boldsymbol{\theta}) \quad = \quad \frac{\partial \mathrm{vec}(\mathbf{R})}{\partial \boldsymbol{\gamma}^t}(\boldsymbol{\theta}) = \left[\frac{\partial \mathrm{vec}(\mathbf{R})}{\partial \gamma_1}, \cdots, \frac{\partial \mathrm{vec}(\mathbf{R})}{\partial \gamma_p}\right](\boldsymbol{\theta})$$

$\mathbf{F}_\phi(\boldsymbol{\theta})$ and $\mathbf{F}_d(\boldsymbol{\theta})$ are expressed similarly.

Further expansion of $\mathbf{M}$, presented in appendix D, shows that the gain phase parameters are decoupled from the other parameters, which suggests that they can be estimated separately (indeed, we will derive such an algorithm). Moreover, $\mathbf{M}$ is independent of $\boldsymbol{\phi}$: the estimation accuracy bounds are independent of the particular values of the phases.

8.2.3 Maximum likelihood formulation

In principle, asymptotically efficient estimates of the model parameters $\boldsymbol{\theta}$, or $\mathbf{g}$ and $\mathbf{D}$, can be obtained via a maximum likelihood (ML) formulation. Since all signal waveforms are i.i.d. Gaussian sequences, the derivation is standard, and ML parameter estimates for N independent samples are obtained by minimising the negative log likelihood function

$$\{\hat{\mathbf{g}}, \widehat{\mathbf{D}}\} = \arg\min_{\mathbf{g},\mathbf{D}} \ln|\mathbf{R}_\theta| + \mathrm{tr}(\mathbf{R}_\theta^{-1}\widehat{\mathbf{R}})$$

where $\mathbf{R}_\theta$ is a function of $\boldsymbol{\theta}$, and $\widehat{\mathbf{R}}$ is the sample covariance matrix defined in (8.2)). This would lead to ML-optimal, asymptotically efficient, estimates.

Even if the model is somewhat simpler than in [111], it does not seem possible to solve this minimisation problem in closed form, and we have to resort to numerical optimisation methods, for example, Newton-Raphson or scoring methods [81, 57, 91]. Stable implementations of such methods are complicated; global convergence is not guaranteed and a good initial point is needed. The

computational complexity is dominated by the repeated evaluation of second-order derivatives.

In the following section we will propose algorithms that are based on the least squares (LS) optimisation of the model errors. Although there is no guarantee that the solution converges to the global optimum here either, the advantage of the LS algorithms is reduced computational complexity. We will also derive simple closed-form approximate solutions to the LS cost function.

8.3 Gain decomposition algorithms

We will start by describing the Least Squares cost function, after which we will derive four Least Squares estimation algorithms, two iterative and two in closed form.

8.3.1 Generalised Least Squares Formulation

As discussed in [111], and following techniques described in detail in [102, chap. 9], a Weighted Least Squares covariance matching approach is known to lead to estimates that for a large number of samples are equivalent to ML estimates, hence are asymptotically efficient and reach the CRB.

Therefore, for a given $\widehat{\mathbf{R}}$, the Least Squares covariance model fitting problem is, for estimation of $\mathbf{g}$ and $\mathbf{D}$, defined as

$$\{\hat{\mathbf{g}}, \widehat{\mathbf{D}}\} = \arg\min_{\mathbf{g},\mathbf{D}} \|\widehat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H - \mathbf{D}\|_F^2 \tag{8.6}$$

By writing $\mathbf{g}$ and $\mathbf{D}$ in terms of their parametrization $\boldsymbol{\theta}$, we correspondingly consider the minimisation of the cost function $\kappa(\boldsymbol{\theta})$:

$$\kappa(\boldsymbol{\theta}) = \|\widehat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H - \mathbf{D}\|_F^2 = \frac{1}{2}\mathbf{f}(\boldsymbol{\theta})^H\mathbf{f}(\boldsymbol{\theta}) \tag{8.7}$$

where $\mathbf{f}(\boldsymbol{\theta}) = \mathrm{vec}(\widehat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H - \mathbf{D})$. A more general Weighted Least Squares problem is obtained, for a Hermitian weighting matrix $\mathbf{W}$, as the optimisation of

$$\kappa_{\mathbf{W}}(\boldsymbol{\theta}) = \frac{1}{2}\mathbf{f}(\boldsymbol{\theta})^H\mathbf{W}\mathbf{f}(\boldsymbol{\theta}) = \frac{1}{2}\mathbf{f}_{\mathbf{W}}(\boldsymbol{\theta})^H\mathbf{f}_{\mathbf{W}}(\boldsymbol{\theta}) \tag{8.8}$$

where $\mathbf{f}_{\mathbf{W}}(\boldsymbol{\theta}) := \mathbf{W}^{\frac{1}{2}}\mathbf{f}(\boldsymbol{\theta})$. The weighting can be used to obtain estimators with a reduced variance. The optimal weight is known to be the inverse of the asymptotic covariance of the residuals (cf. [111]), $\mathrm{E}\{\mathbf{f}(\boldsymbol{\theta}_0)\mathbf{f}(\boldsymbol{\theta}_0)^H\}$, where θ_0 is the "true" value of the parameters, or $\mathbf{f}(\boldsymbol{\theta}_0) = \mathrm{vec}(\widehat{\mathbf{R}} - \mathbf{R})$ (corresponding to the true covariance matrix $\mathbf{R}$). Since all sources are Gaussian, we find

$$\mathbf{W}_{opt} = (\bar{\mathbf{R}} \otimes \mathbf{R})^{-1} = \bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}\,. \tag{8.9}$$

Let $\hat{\boldsymbol{\theta}} = \arg\min \kappa(\boldsymbol{\theta})$. Before discussing algorithms to compute $\hat{\boldsymbol{\theta}}$, we will summarise the statistical performance of this LS estimator, as it follows from

[102] and from similar results, for example in [81,156,111]. It is well known that the (W)LS estimator is asymptotically unbiased and consistent. The asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is Gaussian, and the large sample covariance matrix $\mathbf{C} = \mathrm{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^H\}$ of the parameter estimate $\hat{\boldsymbol{\theta}}$ is given by

$$\mathbf{C} \;=\; \frac{1}{N}\mathbf{F}_0^{\dagger}\,(\bar{\mathbf{R}} \otimes \mathbf{R})\,\mathbf{F}_0^{\dagger H}$$

where $\mathbf{F}_0 = \mathbf{F}(\boldsymbol{\theta}_0)$ is the Jacobian in (8.5) evaluated at $\boldsymbol{\theta}_0$. For weighted estimates $\hat{\boldsymbol{\theta}}_{\mathbf{W}}$, the large sample covariance matrix is derived as

$$\mathbf{C}_{\mathbf{W}} = \frac{1}{N}({\mathbf{F}_0}^H\mathbf{W}\mathbf{F}_0)^{-1}{\mathbf{F}_0}^H\mathbf{W}(\bar{\mathbf{R}} \otimes \mathbf{R})\mathbf{W}\mathbf{F}_0({\mathbf{F}_0}^H\mathbf{W}\mathbf{F}_0)^{-1}$$

At the optimal weight (8.9),

$$\mathbf{C}_{\mathbf{W}_{opt}} = \frac{1}{N}\left(\mathbf{F}_0^H\;\left(\bar{\mathbf{R}} \otimes \mathbf{R}\right)^{-1}\;\mathbf{F}_0\right)^{-1} = \mathbf{C}_{CRB}\,.$$

Hence, at the optimal weight (or a consistent estimate of it), the WLS estimator is asymptotically efficient.

8.3.2 Gauss-Newton iterations (GNLS)

Assuming that we have a good initial point for $\boldsymbol{\theta}$, the minimisation of the (weighted) LS cost function (8.6) can be carried out using the Gauss-Newton method [57].

Let $\mathbf{F}(\boldsymbol{\theta})$ denote the Jacobian (8.5). For the unweighted cost function, the gradient at $\boldsymbol{\theta}$ is

$$\mathbf{p}(\boldsymbol{\theta}) = \mathrm{Re}(\mathbf{F}(\boldsymbol{\theta})^H\mathbf{f}(\boldsymbol{\theta})) = \mathbf{F}(\boldsymbol{\theta})^H\mathbf{f}(\boldsymbol{\theta}) \tag{8.10}$$

Note that due to the Hermitian symmetries, the product $\mathbf{F}^H\mathbf{f}$ is already real. The Hessian of the cost function at $\boldsymbol{\theta}$ is given by

$$\mathbf{H}(\boldsymbol{\theta}) = \mathrm{Re}(\mathbf{F}^H(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta})) = \mathbf{F}(\boldsymbol{\theta})^H\mathbf{F}(\boldsymbol{\theta}) \tag{8.11}$$

The Gauss-Newton update step is then

$$\boldsymbol{\theta}_{k+1} \;=\; \boldsymbol{\theta}_k \;-\; \mu_k\mathbf{H}(\boldsymbol{\theta}_k)^{-1}\mathbf{p}(\boldsymbol{\theta}_k) \;=\; \boldsymbol{\theta}_k \;-\; \mu_k\mathbf{F}_k^{\dagger}\mathbf{f}_k \tag{8.12}$$

where $\mathbf{F}_k = \mathbf{F}(\boldsymbol{\theta}_k)$, and $\mathbf{f}_k = \mathbf{f}(\boldsymbol{\theta}_k)$. $\mu_k \in (0, 1]$ is a step size; with a good initial point we can take $\mu_k = 1$, but in practice a line search would be necessary to ensure proper convergence. For numerical stability, the pseudo-inverse of $\mathbf{F}_k$ can be regularised by incorporating a certain threshold on its singular values. A similar iteration can be derived for the weighted cost function.

An initial point is needed for the GNLS recursion, and can often be obtained from an SVD of $\widehat{\mathbf{R}} - \mathrm{diag}(\widehat{\mathbf{R}})$, because the astronomical source power is usually much smaller than the noise powers (or $\mathrm{diag}(\mathbf{g}\mathbf{g}^H) \ll \mathbf{D}$). Another possibility is applying one of the closed-form algorithms described later in sections 8.3.4

and 8.3.5. The initial point can also be used to generate a consistent estimate of the optimal weight (since the latter depends on the true $\mathbf{R}$ and is unknown).

In the CRB derivations it was shown that the gain magnitudes and the gain phases are decoupled. By analogy with the derivation in Appendix D, we can show that $\mathbf{F}_\gamma^H \mathbf{F}_\phi = 0$, and $\mathbf{F}_d^H \mathbf{F}_\phi = 0$, which somewhat simplifies the Hessian. As in [111], it is also possible to concentrate the cost function, eliminating $\mathbf{D}$ and the scaling of $\mathbf{g}$, but the remaining parameters are more strongly coupled and the derivatives more complex to evaluate. In our experience, the complexity of the resulting Gauss-Newton scheme is higher and the convergence not better.

8.3.3 Minimisation using alternating least squares (ALS)

Unweighted ALS algorithm

A straightforward technique to optimise a cost function over many parameters is to alternatingly minimise over a subset, keeping the remaining parameters fixed. In our case, assume that we have an estimate $\widehat{\mathbf{D}}_k$ at the k-th iteration. The next step is to minimise the LS cost function (8.6) with respect to the gain vector only:

$$\hat{\mathbf{g}}_k = \arg\min_{\mathbf{g}} \|\widehat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H - \widehat{\mathbf{D}}_k\|_F^2 \tag{8.13}$$

The minimum is found from the eigenvalue decomposition $\widehat{\mathbf{R}} - \widehat{\mathbf{D}}_k = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where the matrix $\mathbf{U} = [\mathbf{u}_1\,,\ \cdots\,,\ \mathbf{u}_p]$ contains the eigenvectors $\mathbf{u}_i$, and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues λ_i. The gain estimate minimising (9.3) is given by

$$\hat{\mathbf{g}}_k = \mathbf{u}_1 \sqrt{\lambda_1} \tag{8.14}$$

where λ_1 is the largest eigenvalue and $\mathbf{u}_1$ is the corresponding eigenvector. The second step is minimising (8.6) with respect to the system noise matrix $\mathbf{D}$, keeping the gain vector fixed:

$$\widehat{\mathbf{D}}_{k+1} = \arg\min_{\mathbf{D}} \|\widehat{\mathbf{R}} - \hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H - \mathbf{D}\|^2 \tag{8.15}$$

where $\mathbf{D}$ is constrained to be diagonal with non-negative entries. The minimum is obtained by subtracting $\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H$ from $\widehat{\mathbf{R}}$ and discarding all off-diagonal elements:

$$\widehat{\mathbf{D}}_{k+1} = \mathrm{diag}(\widehat{\mathbf{R}} - \mathbf{g}_k \mathbf{g}_k^H)$$

The condition that the diagonal elements of $\widehat{\mathbf{D}}_{k+1}$ should be positive can be implemented by subsequently setting the negative entries at zero. The two minimisations steps (9.3) and (9.4) are repeated until the model error (8.7) converges. Since each of the minimising steps in the iteration loop reduces the model error, we obtain monotonic convergence to a local minimum. Although the iteration is very simple to implement, simulations indicate that convergence is usually very slow, especially in the absence of a reasonable initial point.

Weighted ALS

The optimal weight for the LS cost function is $\mathbf{W} = \bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}$. Due to this Kronecker structure, the WLS cost function (8.8) can also be written as

$$\{\hat{\mathbf{g}}, \widehat{\mathbf{D}}\} = \arg\min_{\mathbf{g},\mathbf{D}} \|\mathbf{W}_c(\widehat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H - \mathbf{D})\mathbf{W}_c\|_F^2 \tag{8.16}$$

where $\mathbf{W}_c = \mathbf{R}^{-1/2}$. As before, if we have an estimate $\widehat{\mathbf{D}}_k$ of $\mathbf{D}$ at the k-th iteration, an estimate of $\mathbf{g}$ follows from

$$\hat{\mathbf{g}}_k = \arg\min_{\mathbf{g}} \|\mathbf{W}_c(\widehat{\mathbf{R}} - \widehat{\mathbf{D}}_k)\mathbf{W}_c - (\mathbf{W}_c\mathbf{g})(\mathbf{W}_c\mathbf{g})^H\|_F^2 \tag{8.17}$$

After computing the eigenvalue decomposition $\mathbf{W}_c(\widehat{\mathbf{R}} - \widehat{\mathbf{D}}_k)\mathbf{W}_c = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, the estimate $\hat{\mathbf{g}}_k$ follows as

$$\hat{\mathbf{g}}_k = \mathbf{W}_c^{-1}\mathbf{u}_1\sqrt{\lambda_1} \tag{8.18}$$

The second step is minimising the cost function with respect to $\mathbf{D}$ while keeping $\mathbf{g}$ fixed:

$$\widehat{\mathbf{D}}_{k+1} = \arg\min_{\mathbf{D}} \|\mathbf{W}_c(\widehat{\mathbf{R}} - \hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H)\mathbf{W}_c - \mathbf{W}_c\mathbf{D}\mathbf{W}_c\|^2 \tag{8.19}$$

Let $\mathbf{d} = \text{vecdiag}(\mathbf{D})$, then (using several Kronecker relations [59])

$$\begin{aligned}\hat{\mathbf{d}}_{k+1} &= \arg\min_{\mathbf{d}} \|(\bar{\mathbf{W}}_c \otimes \mathbf{W}_c)\text{vec}(\widehat{\mathbf{R}} - \hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H) - (\bar{\mathbf{W}}_c \circ \mathbf{W}_c)\mathbf{d}\|^2 \\ &= (\bar{\mathbf{W}}_c^2 \odot \mathbf{W}_c^2)^{-1}\text{vecdiag}(\mathbf{W}_c^2(\widehat{\mathbf{R}} - \hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H)\mathbf{W}_c^2)\end{aligned}$$

This is a closed-form solution for $\widehat{\mathbf{D}}_{k+1}$. Note that unless $\mathbf{W}_c$ is diagonal, the result in general is not equal to $\text{diag}(\widehat{\mathbf{R}} - \hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^H)$.

The optimal weight $\mathbf{W}_c = \mathbf{R}^{-1/2}$ depends on the true covariance matrix $\mathbf{R}$, which is unknown. Asymptotically the same results are obtained by replacing $\mathbf{R}$ by a consistent estimate, for example $\widehat{\mathbf{R}}$ or the result of one of the closed-form estimates in the following sections. At the optimum, the statistical properties of the ALS and W-ALS estimators are as described in section 8.3.1.

8.3.4 Closed form using logarithmic least squares (LOGLS)

Unweighted LOGLS algorithm

An alternative closed form estimate (as in use at the Westerbork Synthesis Radio Telescope, WSRT [138] since 1980) is obtained by minimising the mean squared error of the *logarithms* of the model. As we will show, taking the logarithm has several effects. The equations become linear in the parameters as the products of gains become sums:

$$\ln(r_{ij}) = \ln(\gamma_i) + \ln(\gamma_j) + \imath(\phi_i - \phi_j) \bmod 2\pi\imath \qquad (i \neq j)\,.$$

It is seen that a least squares model fitting can be applied to the gain magnitude and gain phase separately. Unfortunately, a modulo 2π phase ambiguity is

introduced because of the complex properties of the logarithm. This makes phase unwrapping necessary in the decomposition algorithm.

In a matrix formulation, we minimise an LS cost function after taking the element-wise logarithm:

$$\{\hat{\mathbf{g}}, \widehat{\mathbf{D}}\} = \arg\min_{\mathbf{g},\mathbf{D},\mathbf{k}} \| \ln(\widehat{\mathbf{R}}) - \ln(\mathbf{g}\mathbf{g}^H + \mathbf{D}) + \mathbf{K}2\pi\imath \|_F^2 \tag{8.20}$$

where $\mathbf{K}$ is a nuisance parameter matrix of integers. For any estimate $\hat{\mathbf{g}}$, the optimal estimate of $\widehat{\mathbf{D}}$ is still given by $\widehat{\mathbf{D}} = \mathrm{diag}(\widehat{\mathbf{R}} - \hat{\mathbf{g}}\hat{\mathbf{g}}^H)$. Substituting this back into the cost function shows that the main diagonal of the argument to the Frobenius norm is equal to zero. Thus, the cost function is compressed as

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g},\mathbf{k}} \| \mathbf{J}[\mathrm{vec}\ln(\widehat{\mathbf{R}}) - \ln(\bar{\mathbf{g}} \otimes \mathbf{g}) + \mathbf{k}2\pi\imath] \|^2 \tag{8.21}$$

where $\mathbf{k} = \mathrm{vec}(\mathbf{K})$. Note that

$$\begin{aligned} \ln(\bar{\mathbf{g}} \otimes \mathbf{g}) &= \mathbf{1} \otimes \ln(\mathbf{g}) + \ln(\bar{\mathbf{g}}) \otimes \mathbf{1} \\ &= \mathbf{1} \otimes \ln(\boldsymbol{\gamma}) + \ln(\boldsymbol{\gamma}) \otimes \mathbf{1} \ + \ \imath(\mathbf{1} \otimes \boldsymbol{\phi} - \boldsymbol{\phi} \otimes \mathbf{1}) \\ &= [\mathbf{1} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{1}] \ln(\boldsymbol{\gamma}) \ + \ \imath[\mathbf{1} \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{1}]\boldsymbol{\phi} \end{aligned} \tag{8.22}$$

Defining $\mathbf{v}_R = \mathrm{vec}(\mathrm{Re}\ \ln(\widehat{\mathbf{R}}))$ and $\mathbf{v}_I = \mathrm{vec}(\mathrm{Im}\ \ln(\widehat{\mathbf{R}}))$ it is seen that the optimisation problem separates into independent optimisations over $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ (corresponding to the real and imaginary components), namely

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= \arg\min_{\boldsymbol{\gamma}} \| \mathbf{J}\mathbf{v}_R \ - \ \mathbf{J}[\mathbf{1} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{1}] \ln(\boldsymbol{\gamma}) \|^2 \\ \hat{\boldsymbol{\phi}} &= \arg\min_{\boldsymbol{\phi},\mathbf{k};\phi_1=0} \| \mathbf{J}(\mathbf{v}_I + \mathbf{k}2\pi) \ - \ \mathbf{J}[\mathbf{1} \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{1}]\boldsymbol{\phi} \|^2 \end{aligned}$$

Each cost function is linear and easily optimised in closed form. For the first cost function, using $\mathbf{J}^H\mathbf{J} = \mathbf{I} \otimes \mathbf{I} - (\mathbf{I} \circ \mathbf{I})(\mathbf{I} \circ \mathbf{I})^H$ and several Kronecker relations [59], we obtain

$$\ln(\hat{\boldsymbol{\gamma}}) = (\mathbf{J}[\mathbf{1} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{1}])^\dagger \mathbf{J}\mathbf{v}_R = \mathbf{B}^{-1}\mathbf{q} \tag{8.23}$$

where $\mathbf{B} = 2((p-2)\mathbf{I} + \mathbf{1}\mathbf{1}^t)$, $\mathbf{q} = 2(\mathrm{Re}\ \ln(\widehat{\mathbf{R}}) \odot \mathbf{I}_c)\mathbf{1}$, and $\mathbf{I}_c = \mathbf{1}\mathbf{1}^t - \mathbf{I}$. An explicit expression of $\mathbf{B}^{-1}$ is easily obtained using Woodbury's identity. Note also that $\mathbf{B}^{-1}\mathbf{q}$ can be computed very efficiently. Similarly, solving the second cost function is reduced to solving

$$\mathbf{C}_e\hat{\boldsymbol{\phi}} = \mathbf{v}_e + 2\pi\mathbf{k}_e$$

where

$$\mathbf{C}_e = \begin{bmatrix} \mathbf{C} \\ \mathbf{e}_1^t \end{bmatrix}, \quad \mathbf{v}_e = \begin{bmatrix} \mathbf{v}_I \\ \phi_1 \end{bmatrix}, \quad \mathbf{k}_e = \begin{bmatrix} \mathbf{k} \\ 0 \end{bmatrix}$$

and $\mathbf{C} = \mathbf{1} \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{1}$. The last row in $\mathbf{C}_e$ implements the chosen phase uniqueness constraint, $\phi_1 = 0$. The system can be solved as

$$\hat{\boldsymbol{\phi}} = (\mathbf{C}_e^H\mathbf{C}_e)^{-1}\mathbf{C}_e^H(\mathbf{v}_e + 2\pi\mathbf{k}_e) \tag{8.24}$$

once the integers $\mathbf{k}_e$ are known. Suppose we have an initial estimate $\boldsymbol{\phi}^\circ$ of $\boldsymbol{\phi}$, for example from the first column $\hat{\mathbf{r}}_1$ of $\widehat{\mathbf{R}}$ as $\boldsymbol{\phi}^\circ = \mathrm{Im}\ln(\hat{\mathbf{r}}_1)$. Then $\mathbf{k}$ (and thus $\mathbf{k}_e$) is obtained by rounding the entries of $\mathbf{C}\boldsymbol{\phi}^\circ - \mathbf{v}_I$ to the nearest multiple of 2π: $\mathbf{k} = \mathrm{round}\left((2\pi)^{-1}(\mathbf{C}\boldsymbol{\phi}^\circ - \mathbf{v}_I)\right)$. Using Woodbury's identity and several Kronecker relations, equation (8.24) can be worked out further, producing

$$(\mathbf{C}_e^H\mathbf{C}_e)^{-1} = \frac{1}{2}\left(\mathbf{E} + \frac{(\mathbf{E}\mathbf{1})(\mathbf{E}\mathbf{1})^H}{1-\mathrm{tr}(\mathbf{E})}\right)$$

$$\mathbf{C}_e^H(\mathbf{v}_e + 2\pi\mathbf{k}_e) = 2[(\mathrm{Im}\ln(\widehat{\mathbf{R}}) + 2\pi\mathbf{K}) \odot \mathbf{I}_c]\mathbf{1}$$

where $\mathbf{E} = \frac{1}{2}(2p\mathbf{I} + \mathbf{e}_1\mathbf{e}_1^t)$ and $\mathbf{K} = \mathrm{unvec}(\mathbf{k})$.

Weighted LOGLS

The LOGLS method has zero bias and is consistent, as the logarithm function is a smooth monotonous transformation, but the logarithm operator prevents the LOGLS method from being statistically efficient. In general, there is no weighting matrix which makes the LOGLS method asymptotically efficient, but special cases for which asymptotical efficiency can be reached do exist. This will be shown by comparing the LOGLS cost function with the weighted LS cost function, the latter leading to asymptotically efficient estimates. We start from the compressed LOGLS cost function (8.21). After replacing the selection matrix $\mathbf{J}$ by $\mathbf{J}^H\mathbf{J}$ (which does not change the norm and is more convenient since $\mathbf{J}^H\mathbf{J}$ is diagonal), we will introduce a weighting matrix $\mathbf{W}_{log}$ to obtain

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g},\mathbf{k}} \|\mathbf{W}_{log}^{\frac{1}{2}}\mathbf{J}^H\mathbf{J}\mathrm{vec}[\ln(\widehat{\mathbf{R}}) - \ln(\mathbf{g}\mathbf{g}^H) + \mathbf{k}2\pi\imath]\|^2 \tag{8.25}$$

Assuming that $\hat{\mathbf{g}}$ is close to $\mathbf{g}$ and that none of the gain amplitudes γ are zero, we can use the Taylor approximation $\ln(x) \approx x - 1$. Let $\ominus$ denote an element-wise matrix division, then

$$\begin{aligned}
&\mathbf{J}^H\mathbf{J}\mathrm{vec}[\ln(\widehat{\mathbf{R}}) - \ln(\mathbf{g}\mathbf{g}^H)] \\
&= \mathbf{J}^H\mathbf{J}\mathrm{vec}[\ln(\widehat{\mathbf{R}} \ominus (\mathbf{g}\mathbf{g}^H))] \\
&\approx \mathbf{J}^H\mathbf{J}\mathrm{vec}[(\widehat{\mathbf{R}} \ominus (\mathbf{g}\mathbf{g}^H) - \mathbf{I})] \\
&= \mathbf{J}^H\mathbf{J}\mathrm{vec}[(\mathrm{diag}(\mathbf{g}))^{-1}(\widehat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H)(\mathrm{diag}(\overline{\mathbf{g}}))^{-1}] \\
&= \mathbf{J}^H\mathbf{J}\,(\mathrm{diag}(\overline{\mathbf{g}} \otimes \mathbf{g}))^{-1}\mathrm{vec}(\widehat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H)
\end{aligned}$$

Note that for the cost function the phase of the diagonal matrix $\mathrm{diag}(\overline{\mathbf{g}}\otimes\mathbf{g})^{-1}$ is irrelevant, so it can be replaced by $(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})^{-1}$, where $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$. The weighted LOGLS cost function close to the optimum can thus be written as

$$\min_{\mathbf{g}} \| \mathbf{W}_{log}^{\frac{1}{2}}\mathbf{J}^H\mathbf{J}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})^{-1}\mathrm{vec}(\widehat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H) \|^2$$

If we compare this to the cost function for the weighted LS, $\min_{\mathbf{g}} \|\mathbf{W}_{ls}^{\frac{1}{2}}\mathbf{J}^H\mathbf{J}\mathrm{vec}(\widehat{\mathbf{R}} - \mathbf{g}\mathbf{g}^H) \|^2$ we see that these two cost functions give the same solution if

$$\mathbf{W}_{log}^{\frac{1}{2}}\mathbf{J}^H\mathbf{J}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})^{-1} = \mathbf{W}_{ls}^{\frac{1}{2}}\mathbf{J}^H\mathbf{J} \tag{8.26}$$

Because $\mathbf{J}^H\mathbf{J}$ is singular, a solution is possible only in special cases, for example diagonal weighting matrices. For low SNRs, SNR $= \mathbf{g}^H\mathbf{g}/\mathrm{tr}(\mathbf{D}) \ll 1$, which is true for most radio astronomical observations, the $\mathbf{W}_{ls}$ weighting matrix is close to diagonal: $\mathbf{W}_{ls} = \mathbf{D}^{-1} \otimes \mathbf{D}^{-1}$, and in this case the choice $\mathbf{W}_{log}^{\frac{1}{2}} = (\mathbf{D}^{-\frac{1}{2}} \otimes \mathbf{D}^{-\frac{1}{2}})(\mathbf{\Gamma} \otimes \mathbf{\Gamma}) = \mathbf{D}^{-\frac{1}{2}}\mathbf{\Gamma} \otimes \mathbf{D}^{-\frac{1}{2}}\mathbf{\Gamma}$ satisfies (8.26). Here we used the diagonal and commutative properties of $\mathbf{J}^H\mathbf{J}$. Defining $\mathbf{W} = \mathbf{D}^{-\frac{1}{2}}\mathbf{\Gamma}$, the weighted LOGLS cost function can be expressed as

$$\min_{\mathbf{g},\mathbf{k}} \parallel \mathbf{J}(\mathbf{W} \otimes \mathbf{W})[\mathrm{vec}\, \ln(\widehat{\mathbf{R}}) - \ln(\overline{\mathbf{g}} \otimes \mathbf{g}) + 2\pi\mathbf{k}\imath] \parallel^2$$

Further defining $\mathbf{v}'_R = \mathrm{vec}(\mathrm{Re}\ \mathbf{W}\ln(\widehat{\mathbf{R}})\mathbf{W})$, $\mathbf{v}'_I = \mathrm{vec}(\mathrm{Im}\ \mathbf{W}\ln(\widehat{\mathbf{R}})\mathbf{W})$, and $\mathbf{k}' = (\mathbf{W} \otimes \mathbf{W})\mathbf{k}$, and noting that (viz. (8.22))

$$(\mathbf{W} \otimes \mathbf{W})\ln(\overline{\mathbf{g}} \otimes \mathbf{g}) = [\mathbf{w} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{w}]\mathbf{W}\ln\gamma + \imath[\mathbf{w} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{w}]\mathbf{W}\boldsymbol{\phi}$$

where $\mathbf{w} = \mathbf{W1}$, it follows that we need to solve

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} &= \arg\min_{\boldsymbol{\gamma}} \parallel \mathbf{J}\mathbf{v}'_R - \mathbf{J}[\mathbf{w} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{w}]\mathbf{W}\ln(\boldsymbol{\gamma}) \parallel^2 \\
\hat{\boldsymbol{\phi}} &= \arg\min_{\boldsymbol{\phi},\mathbf{k};\phi_1=0} \parallel \mathbf{J}(\mathbf{v}'_I + \mathbf{k}'2\pi) - \mathbf{J}[\mathbf{w} \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{w}]\mathbf{W}\boldsymbol{\phi} \parallel^2
\end{aligned}$$

It is clear that this can be done in closed form as before; we omit the details.

The conclusion is that for low SNRs, we can weight by $\mathbf{W}_{log}$ and make the weighted LOGLS method asymptotically efficient to a good approximation.

8.3.5 Closed form using column ratios (COLR)

Unweighted COLR algorithm

Finally, we will set out to find a closed-form estimate of $\mathbf{g}$, which recovers $\mathbf{g}$ exactly when applied to $\mathbf{R}$ (hence asymptotically for $\widehat{\mathbf{R}}$). The crux of this method is the observation that the off-diagonal entries of $\mathbf{g}\mathbf{g}^H$ are equal to those of $\mathbf{R}$, and are known, so that we only need to reconstruct the diagonal entries of $\mathbf{g}\mathbf{g}^H$. We further note that $\mathbf{g}\mathbf{g}^H$ is rank 1, so any submatrix of $\mathbf{R}$ that does not contain elements from the main diagonal is also rank 1. This property can be used to estimate the ratio between any pair of columns of $\mathbf{R}$ away from the diagonal, and subsequently to estimate how the main diagonal of $\mathbf{R}$ has to be changed so that the resulting $\mathbf{R}'$ is rank 1, or $\mathbf{R}' = \mathbf{g}\mathbf{g}^H$. The gain vector $\mathbf{g}$ can then be extracted by an eigenvalue decomposition.

For any two elements g_i and g_j of the complex gain vector $\mathbf{g}$, define the ratio $\alpha_{ij} = g_i/g_j$. This ratio can be estimated from the data $\mathbf{R}$ by solving

$$\mathbf{c}_i = \alpha_{ij}\mathbf{c}_j$$

where $\mathbf{c}_i$ and $\mathbf{c}_j$ are the i-th and j-th column of the matrix $\mathbf{R}$, not including the entries r_{ii}, r_{ij}, r_{ji} and r_{jj} because r_{ii} and r_{jj} also depend on the unknown

system noise d_i. Solving for α_{ij} in the least squares sense gives

$$\hat{\alpha}_{ij} = (\mathbf{c}_j^H \mathbf{c}_j)^{-1} \mathbf{c}_j^H \mathbf{c}_i = \frac{\sum_{k \neq i,j} r_{kj}^* r_{ki}}{\sum_{k \neq i,j} r_{kj}^* r_{kj}}$$

We can subsequently estimate $|g_i|^2$ as $|g_i|^2 = \hat{\alpha}_{ij} r_{ij}$, for any choice of $j \neq i$. This estimate can be improved if all $(p-1)$ available column ratios are used, and the fact that $|g_i|^2$ is real:

$$|g_i|^2 = \frac{1}{p-1} \mathrm{Re}\Big(\sum_{\substack{j=1 \\ j \neq i}}^{p} \hat{\alpha}_{ij} r_{ij} \Big)$$

The next step is to form $\mathbf{R}'$ equal to $\mathbf{R}$ but with the diagonal entries replaced by the estimates of $|g_i|^2$ obtained above. The resulting matrix $\mathbf{R}'$ is an estimate of $\mathbf{g}\mathbf{g}^H$, and $\hat{\mathbf{g}}$ is found from the eigenvalue decomposition $\mathbf{R}' = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, similarly as in (8.14). $\mathbf{D}$ is obtained as $\mathbf{D} = \mathrm{diag}(\mathbf{R} - \mathbf{g}\mathbf{g}^H)$.

With measured data, we follow the same procedure but replace $\mathbf{R}$ by the sample estimate $\widehat{\mathbf{R}}$. Although there is no claim that this procedure minimises the LS cost function, its estimates are rather close and asymptotically (large N) the true $\mathbf{g}$ and $\mathbf{D}$ are obtained. In the simulations and in the experimental results, the column ratio method will be denoted by the acronym "COLR".

Weighted COLR

Because of the ad-hoc derivation of the algorithm, it is difficult to establish the statistical efficiency, and the weighting matrix which might achieve this. The analysis of the other three gain estimation methods suggests that a weighting matrix of the form $\mathbf{W}_{col}^{\frac{1}{2}} = (\mathbf{D} \otimes \mathbf{D})^{-\frac{1}{2}}$ might improve the statistical efficiency in some cases (namely, very non-uniform noise power). In the section on simulations we will show that this is indeed the case. Since $\mathbf{D}$ is initially unknown, we can construct the weighting matrix only after a first (unweighted) estimate of the parameters has been made. The COLR algorithm is subsequently applied to $\mathbf{D}^{-\frac{1}{2}} \widehat{\mathbf{R}} \mathbf{D}^{-\frac{1}{2}}$. The resulting gain estimate $\hat{\mathbf{g}}_w$ is converted to the unweighted gain by $\hat{\mathbf{g}} = \mathbf{D}^{\frac{1}{2}} \hat{\mathbf{g}}_w$. The diagonal estimate $\widehat{\mathbf{D}}$ is obtained as before as $\widehat{\mathbf{D}} = \mathrm{diag}(\widehat{\mathbf{R}} - \hat{\mathbf{g}}\hat{\mathbf{g}}^H)$.

8.3.6 Computational complexity

When a fast update rate for the parameter estimation is needed or when a very large number of telescopes/antenna sensors is used, the computational complexity of the algorithms becomes important. This is especially the case for future generations of large radio telescopes, such as SKA and LOFAR , where the number of telescopes will be very large ($p \geq 100$). Table 8.1 lists the order of the number of multiplicative operations required for each of the algorithms, and the additional number of multiplications for computing weighted estimates.

Table 8.1. Number of multiplications. (p: number of sensors, N_{it}: number of iterations.)

Method	unweighted	weighted (add'l)
GNLS	$34\,N_{it}p^3$	$40\,N_{it}p^3$
ALS	$8\,N_{it}p^3$	$16\,N_{it}p^3$
COLR	$20\,p^3$	$2\,p^2$
LOGLS	$2\,p^2$	$16\,p^2$

The computations for determining the initial points for the GNLS and the ALS methods have not been taken into account.

It is seen that the GNLS method is the most complex, whereas the LOGLS method is computationally the fastest. For the GNLS method the most demanding operation is the calculation of the inverse of the Hessian; for the ALS and the COLR methods it is the repeated evaluation of the SVD. (In the table, we did not take into account the fact that faster estimators for the dominant singular vector exist.) The LOGLS algorithm only requires a few matrix-vector multiplications as the LOGLS estimators have simple closed form expressions, and it is therefore the fastest method.

8.4 Simulations

First we will verify the performance of the estimation algorithms by computer simulations, and then show results based on experimental data collected at the WSRT.

In all simulations we use $p = 8$ telescope channels. Unless denoted otherwise, the gain magnitude γ_i $(i = 1, \cdots, p)$ is kept fixed for each simulation. The γ_i are chosen randomly in the interval $\gamma_0 - \Delta\gamma$ and $\gamma_0 + \Delta\gamma$, where $\gamma_0 = 1$ is the nominal gain magnitude and $\Delta\gamma$ is a spreading parameter. The gain phase ϕ_i lies uniformly distributed between 0 and $\Delta\phi$, where $\Delta\phi \leq 2\pi$. The gain phase is randomly chosen and is also kept fixed during most simulations. The astronomical source is a zero mean complex Gaussian signal with unit power. The signal-to-noise ratio is defined as the total received power due to the source, divided by the total noise power, or $\mathrm{SNR} = \mathbf{g}^H\mathbf{g}/\mathrm{tr}(\mathbf{D})$. Finally, the system noise magnitude d_i lies uniformly distributed between $\mathrm{SNR}^{-1}(1 - \Delta d)$ and $\mathrm{SNR}^{-1}(1 + \Delta d)$, where Δd is the noise spreading parameter, $\Delta d \leq 1$.

For a typical online gain calibration measurement at a radio observatory, astronomical sources are used with SNRs in the range -20 dB to -10 dB, the integration time of the correlation data is usually several seconds to a few minutes for a typical frequency bin resolution of 10 to 100 kHz. Therefore, for most of the simulations, the following parameter settings are chosen: $\Delta\gamma = 0.1$, $\Delta d = 0.1$, $\Delta\phi = 2\pi$, and $N = 2^{15}$.

In the simulations, N is the number of complex samples on which the covariance matrix is based and to which the gain decomposition algorithms are applied. Finally, N_{sim} is the number of simulation runs from which the estimation standard deviation is derived. Choosing $N_{sim} = 256$ gives a reasonable standard deviation accuracy.

8.4.1 Convergence of GNLS and ALS

Figure 8.2 (left) shows the convergence speed (Frobenius norm of the covariance matrix) of the weighted algorithms, when applied to a true covariance matrix $\mathbf{R}$ (infinite sample case). The ALS method shows a linear decrease of the error; the Gauss-Newton curve shows a quadratic decrease in the estimation error, as is expected and well known from literature. The closed-form estimators produce the correct values in one step, up to computer accuracy. Figure 8.2 (right) shows the convergence speed for covariance matrices with noise. In this case, the GNLS method does not show quadratic convergence because after a few iteration steps the estimation accuracy is dominated by the noise (the CRB is reached within a few iteration steps).

For the chosen SNRs and spread in gain and noise power, applying weights improves the convergence speed of the ALS method, which is reduced by a factor two. For the GNLS method, weighting does not have a significant effect on the convergence speed for the parameter settings used.

For the ALS and the GNLS methods, an initial point is needed which is relatively close to the true values. The results of the non-iterative methods could be used as initial point. In the simulations, however, the true gain and noise values with small perturbations were used as initial points.

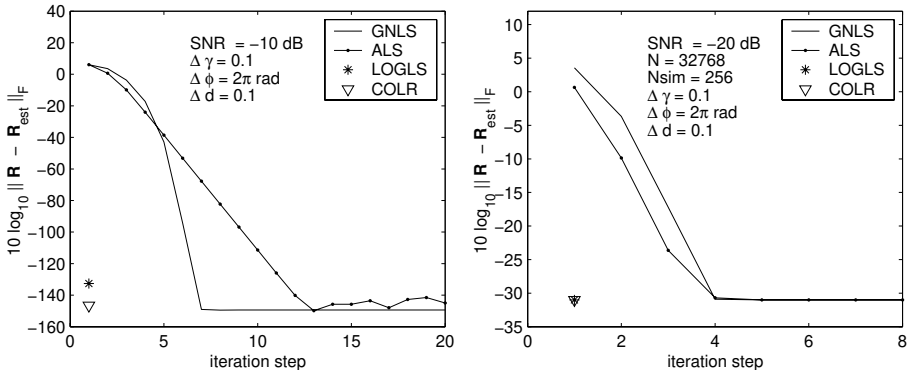


Figure 8.2. Convergence speed of the cost function, weighted gain estimation algorithms applied to the true $\mathbf{R}$ (left), and applied to finite sample covariance matrices (right).

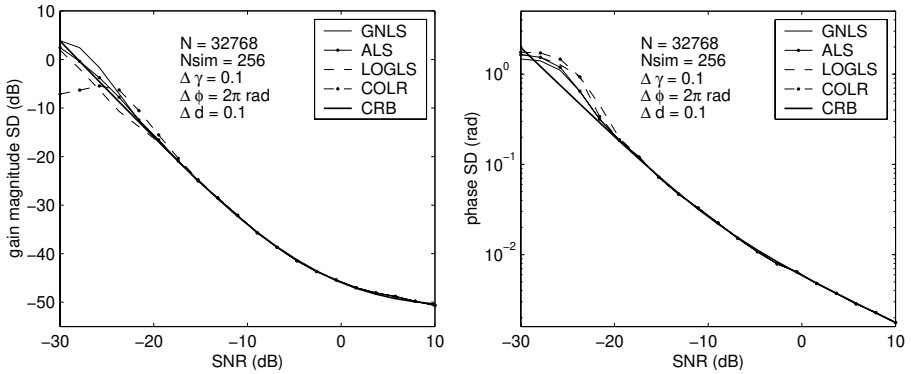


Figure 8.3. Weighted gain estimate standard deviation versus SNR, gain magnitude SD (left), and gain phase SD (right).

8.4.2 Influence of SNR and number of samples

Figure 8.3 shows the results of a gain estimation simulation in which the gain estimation standard deviation is plotted versus SNR for $N = 32768$ samples. The four estimator results are plotted together with the statistical efficiency bound (CRB). For the parameter range under consideration, all methods perform close to the bound, except for SNRs < -15 dB in which case the gain estimators are biased, as is shown in figure 8.4. The figures show that the LOGLS and COLR methods break down at slightly higher SNRs than the ALS and GNLS methods.

For high SNR values, the variance of the gain estimates saturates towards a fixed value, determined by the number of samples. Indeed, at very large SNR, the direction of $\mathbf{g}$ can be estimated accurately even with a single sample, but its scaling $\|\mathbf{g}\|$, or the source power estimate, suffers from the finite sample effect.

In figure 8.5 the weighted gain estimates are plotted as a function of the number of observed time samples for an SNR of -10 dB . The performance of the four weighted methods does not differ much for the parameter setting used.

8.4.3 Influence of parameter spread

We will next investigate the influence of deviations of the gains and system noise values from their nominal values. The number of samples was fixed at $N = 2^{15}$ and the SNR at nominal -10 dB. The gain magnitude and noise parameter ranges are defined by $\boldsymbol{\gamma} = 10^{0.1\Delta\gamma\ \beta_\gamma}$ and $\mathbf{d} = \mathrm{SNR}^{-1} 10^{0.1\Delta d\ \beta_d}$, where $\Delta\gamma$ and Δd are spreading parameters with values ranging from 0 to 10 dB, and β_γ and β_d are random vectors with elements uniformly distributed in the interval [-1,1].

Figures 8.6 and 8.7 show the gain estimation standard deviation as a function of gain spreading for the non-weighted and weighted simulations respectively. For $\Delta\gamma = 0$ dB, all gains are unity; for $\Delta\gamma = 10$ dB, the gains vary between

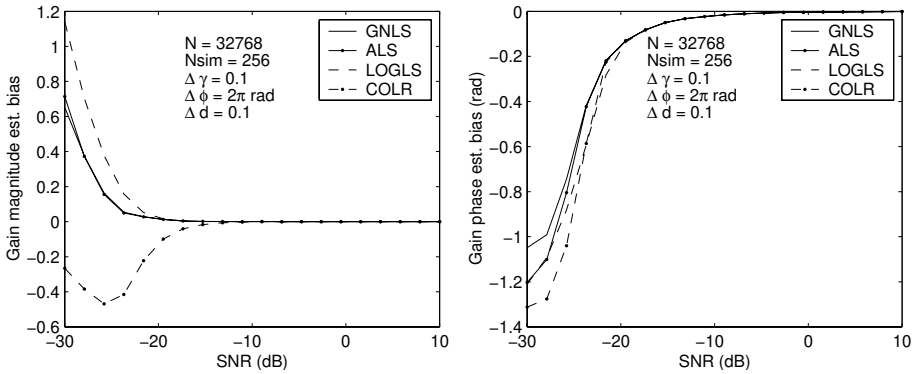


Figure 8.4. Weighted gain estimate bias versus SNR, gain magnitude bias (left), and gain phase bias (right).

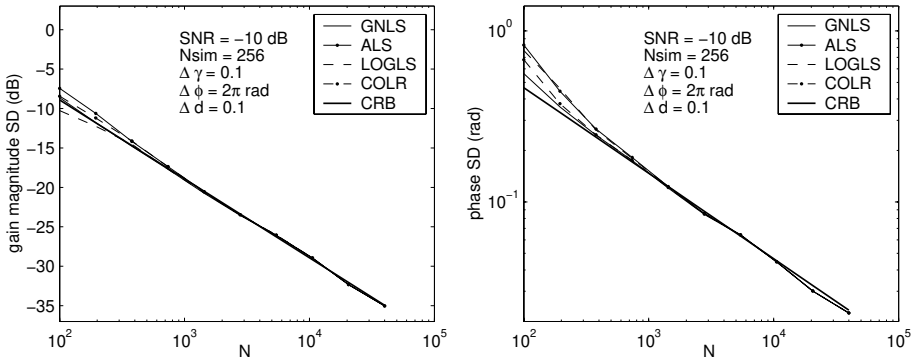


Figure 8.5. Weighted gain estimate standard deviation versus number of samples, gain magnitude SD (left), and gain phase SD (right).

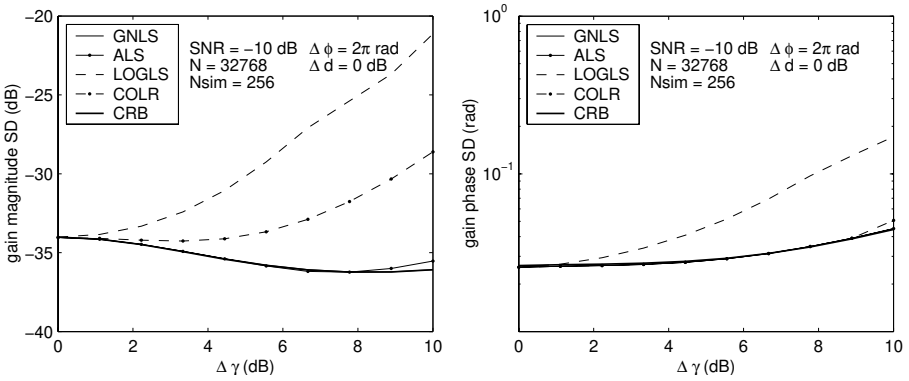


Figure 8.6. Non-weighted gain estimation standard deviation versus dispersion of the gain magnitudes, gain magnitude SD (left), and gain phase SD (right).

−10 dB and 10 dB (with respect to the nominal value of 1). The effect of the weighting on the statistical efficiency of the LOGLS method is dramatic for a large gain spreading parameter. The weighting does not influence the gain magnitude estimation of the COLR method. This is expected as the COLR weighting is a real diagonal matrix. The ALS and GNLS methods coincide with the CRB curve, except for a large gain spread parameter value in the non-weighted case.

Figures 8.8 and 8.9 show the gain estimation standard deviation as a function of noise spreading for the non-weighted and weighted simulations respectively. All non-weighted method standard deviations lie well above the CRB curves for $\Delta d > 2$ dB and coincide with the CRB curves in the weighted case, except the gain magnitude estimation for the COLR method.

We conclude that weighting does not have much influence if the gain and noise values of the different telescopes are approximately equal, as in this case the methods are already asymptotically efficient. The improvement, however, is large if gain and noise spread become significant. In that case weighting makes the methods (except COLR) asymptotically efficient.

8.5 Experimental results

8.5.1 Measurement setup

So far, the accuracy of gain parameter estimations has been verified by means of simulations. In this section, we present an experimental verification based on collected radio telescope data. In the experiment, we observe a strong point source in the sky with a radio telescope interferometer array, in our case the Westerbork Synthesis Radio Telescope. The point source requirement is that

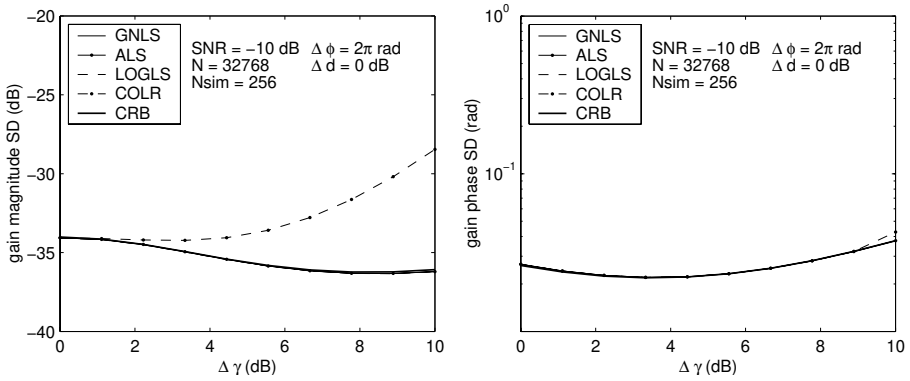


Figure 8.7. Weighted gain estimation standard deviation versus dispersion of the gain magnitudes, gain magnitude SD (left), and gain phase SD (right).

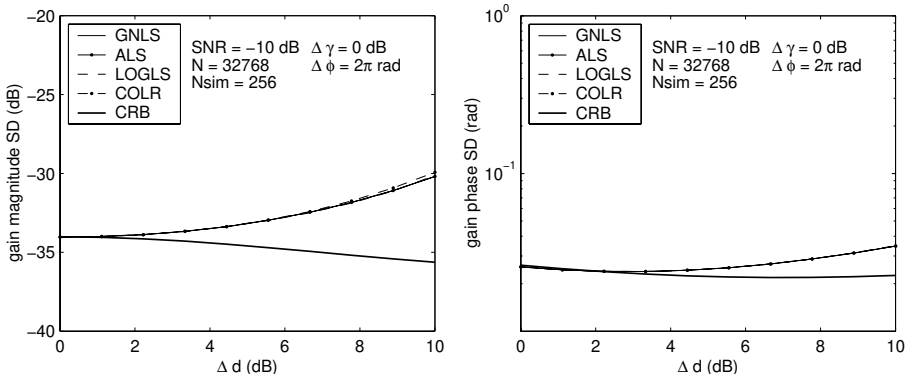


Figure 8.8. Non-weighted gain estimation standard deviation versus dispersion of the system noise, gain magnitude SD (left), and gain phase SD (right).

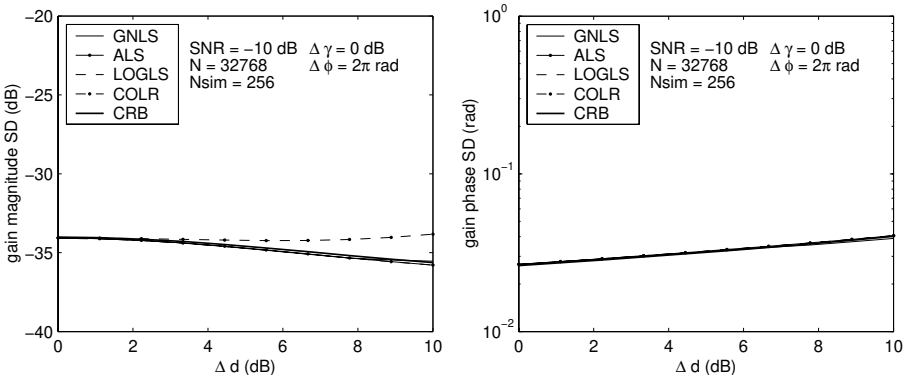


Figure 8.9. Weighted gain estimation standard deviation versus dispersion of the system noise, gain magnitude SD (left), and gain phase SD (right).

the source angular size is much smaller than the telescope main beam power.

The time series at the baseband digital output of the telescopes are segmented into short intervals, Fourier transformed, and subsequently spatially cross-correlated resulting in complex covariance matrices for each of the frequency bins, to which the gain decomposition algorithms are applied.

In our experiments, we used $p = 8$ of the 14 WSRT telescopes, in single linear polarisation mode, with a maximum distance (baseline) of 1 km. The telescopes tracked the strong astronomical point source "3C48" at a sky frequency of 1420.4 MHz with a receiver bandwidth of 1.25 MHz. An eight-channel data recorder, equipped with eight ADCs and 8×32 MBytes of memory, was connected to the telescope baseband IF system outputs. The time sample data was recorded on CDROM and processed offline. In our experiment the earth-rotation related phase drift was compensated for, which means that during the experiment the telescope–interferometer phase was constant. We split the data into 32 frequency bins, each with a bandwidth of 39 kHz, which fits the narrowband assumption reasonably well (propagation delays across the array).

The data model in the experiment is $\mathbf{R} = \mathbf{g}\sigma_s^2\mathbf{g}^H + \mathbf{D}$, where σ_s is the source flux (known from tables). The outcome of the algorithms forms estimates for $\mathbf{g}$ and $\mathbf{D}$. To compare this to known telescope parameters, we first note that the gain vector $\mathbf{g}$ and the noise vector $\mathbf{n}$ contain a common unknown electronic amplification factor $\mathbf{\Gamma}_{el}$ (a diagonal matrix) which is frequency dependent and cannot be obtained separately. Thus define

$$\mathbf{g} = \mathbf{\Gamma}_{el}\underline{\mathbf{g}}\,, \qquad \mathbf{D} = \underline{\mathbf{D}}\mathbf{\Gamma}_{el}^2$$

where $\underline{\mathbf{g}}$ and $\underline{\mathbf{D}}$ are the telescope gain vector and telescope noise covariance matrix at a location in the system prior to amplification. Note that the ratio $\underline{\mathbf{D}}^{-\frac{1}{2}}\underline{\mathbf{g}}$ is independent from the electronic gain. Hence $\sigma_s\mathbf{D}^{-\frac{1}{2}}\mathbf{g} = \sigma_s\underline{\mathbf{D}}^{-\frac{1}{2}}\underline{\mathbf{g}}$. The right hand side is the signal-to-noise ratio at the input of the low-noise

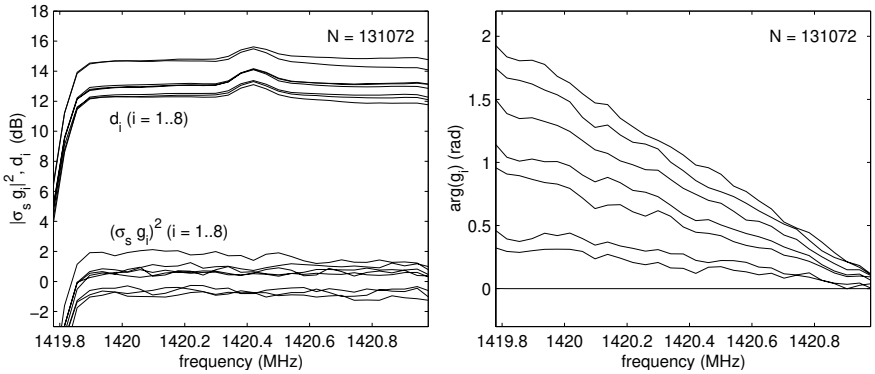


Figure 8.10. Gain magnitude and noise power estimates (left), and gain phase estimates (right), as function of frequency. Estimates are obtained using the weighted LOGLS method, based on an observation the astronomical source 3C48

amplifiers. Its nominal value depends on the construction of the telescopes (collecting area), the telescope system noise and the source flux, and is known from calibration tables. Thus, we can compare the estimate for $\sigma_s \mathbf{D}^{-\frac{1}{2}}\mathbf{g}$ to the literature. In our experiment, the 3C48 source emits spectrally continuous radiowaves at 1420.4 MHz with a source power, according to calibration tables, which lies 13 dB below the WSRT system noise, or $\sigma_s^2|g_i|^2/d_i \approx 0.05$.

8.5.2 Experimental results

The weighted LOGLS method was applied to a WSRT telescope data set consisting of $p = 8$ telescopes and $N = 131072$ samples in 32 frequency bins. The estimates of $\sigma_s^2|g_i|^2$, d_i and ϕ_i for $i = 1, \cdots, 8$ as function of frequency are shown in figure 8.10.

Figure 8.10 (left) confirms that the received SNR is about -13 dB for each antenna. Note that there is a bump in the noise power curves at 1420.4 MHz. This corresponds to the spectral line of neutral hydrogen, and is caused by the galactic emission of our Milky Way. As the Milky Way is a spatially wide source of radiowaves, it is not resolved by the WSRT interferometers, and is therefore visible only in the noise estimates. Figure 8.10 (right) shows the estimated phases of the telescopes, which are frequency dependent because of the geometric delay of the incident source across the array. The horizontal line corresponds to the first telescope, the steepest line to the farthest. The phase slope over a frequency band Δf is given by $\Delta\phi_i = 2\pi(\Delta f)L_i/c$, where L_i is the geometric delay of telescope i. For the longest telescope distance, the calculated phase slope over the passband is 1.9 rad, which matches the observed value.

Next, all four weighted gain estimation methods are applied to the same dataset, but for one frequency bin only. The ratio of the gains and noise vector

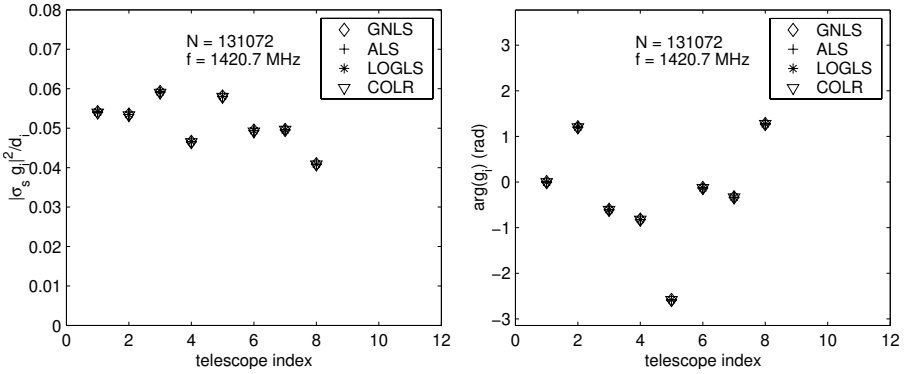


Figure 8.11. Observed signal to noise ratio (left), and estimated gain phase (right), for astronomical source 3C48 at $f = 1420.7$ MHz and channel bandwidth $\Delta f = 39$ kHz.

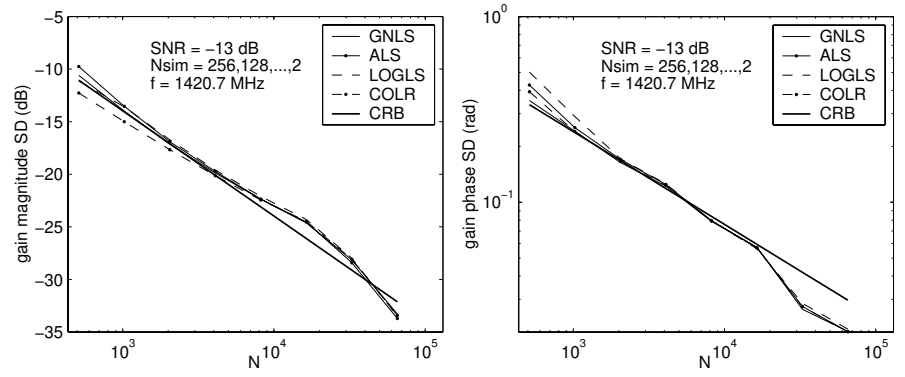


Figure 8.12. Gain estimation standard deviation as function of the integration length ($f = 1420.7$ MHz, $\Delta f = 39$ kHz), gain magnitude SD (left), and gain phase SD (right).

components, $\sigma_s^2|g_i|^2/d_i$, and the observed phases are plotted in figure 8.11. All four estimators yield the same ratios of about 0.05, the expected value. Note that, while the variation among the telescopes of gain magnitudes and noise power in figure 8.10 is about 3 dB, the fluctuation in the gain-noise *ratios* (figure 8.11) is only 25% around the average. This is because the electronic gain variation is quite large, and this is factored out by the division.

Finally, we investigate the effect of the data sample size N on the estimates. The data set was split into subsets increasing in size from $N = 512$ to 65536 in steps of a factor two. The number of subsets decreased correspondingly from $N_{sim} = 256$ down to 2, as the total number of available samples is constant (131072). Both the estimation standard deviation and the CRB are shown in figure 8.12, where the CRB was derived from the 'true' values estimated from the complete data set. There is a fair match between the observed standard deviation and the theoretical bound, except at the edges where either N or N_{sim} are too small to obtain reliable statistics.

8.6 Conclusions

In this chapter we have derived several algorithms for estimating gain and noise parameters of a phased array from an observed covariance matrix. We have shown asymptotically efficient weighted least square estimators, and have derived several closed form estimators which, under certain conditions, are also asymptotically efficient. For low SNRs, in the regime where all algorithms under consideration are biased, the LS estimators perform slightly better than the closed form ad-hoc algorithms. Nonetheless, the closed-form algorithms provide the essential starting points for the iterative methods (alternating LS and Gauss-Newton LS optimization).

For uniform gain and noise vectors, the performance of the unweighted algorithms is close to the CRB and weighting does not improve much. However, for parameters with a large spread (> 3 dB), the weighting dramatically improves the performance, making all algorithms (except COLR) asymptotically efficient, and improving the speed of convergence for the ALS algorithm by a factor two.

An advantage of the LOGLS method is its low computational expenditure, proportional to p^2 rather than p^3 as for the other algorithms. Unfortunately, this method is not easily generalised to the estimation of multiple gain vectors, which is needed for example in the calibration of dual-polarised telescope arrays.

Chapter 9

Polarisation gain calibration

9.1 Introduction

For unpolarised telescope arrays, a standard calibration procedure is to point the telescopes at a strong astronomical source, and to estimate a covariance matrix $\widehat{\mathbf{R}}$, containing all correlation products between the telescope output signals. Asymptotically, $\widehat{\mathbf{R}}$ converges to its expected value $\mathbf{R}$ which has the model $\mathbf{R} = \mathbf{g}\sigma_s^2\mathbf{g}^H + \mathbf{D}$. Here, σ_s^2 is the known source flux, $\mathbf{g}$ is a vector containing the complex gains to be estimated, and $\mathbf{D}$ is a diagonal matrix containing the unknown noise powers per antenna element (it is assumed that the noise power is uncorrelated from one antenna to another). This is essentially the model considered by [138] [37]. Improved estimation algorithms using iterative and closed form least squares techniques have recently been derived [18]: by incorporating proper weighting, these methods are proved to be asymptotically statistically efficient [21].

For dual polarised telescope arrays, much less is known. In 1995, Hamaker et al. [63] [62] developed a matrix formalism in which the polarisation properties of the astronomical signals and their propagation through the ionosphere and the astronomical receiving instrument were efficiently incorporated. An iterative procedure similar to SelfCal [146] was used to estimate the polarisation gain coefficients, but it was not known to what solution it would converge.

In [19], the scalar gain calibration methods of [21] [18] were extended to polarised arrays. In this chapter[1] we will follow the notation and derivations of [19] and [62], and extend them with more efficient algorithms to obtain least squares solutions for the gain factors. We will also verify that at least three sky sources with different polarisation states are needed to find the gain factors.

[1] This chapter was published in [20]

9.2 Data model

9.2.1 Coherency

In aperture synthesis radio astronomy, the output of the interferometers is the correlation of the field strengths at the different telescopes, also known as coherencies [146]. The electric field at the location of an antenna element can be described by two linear polarisation components, stacked in a 2×1 vector: $\mathbf{e}_i = [e_{ix},\ e_{iy}]^t$. The correlation between two different telescopes i and j is a 2×2 interferometer coherency matrix $\boldsymbol{E}_{ij} = \mathcal{E}\{\mathbf{e}_i \mathbf{e}_j^H\}$. If there are p telescopes, each with two polarisations, then the $2p$ observed electric fields can similarly be stacked in one vector: $\mathbf{e} = \left(\mathbf{e}_1^t, \cdots, \mathbf{e}_p^t\right)^t$. The $2p \times 2p$ Hermitian coherency matrix $\mathbf{E}$ is defined by $\mathbf{E} = \mathcal{E}\{\mathbf{e}\mathbf{e}^H\}$ which can be written in terms of interferometer coherency matrices $\boldsymbol{E}_{ij}$ as

$$\boldsymbol{E}_{i,j} = \begin{bmatrix} \mathcal{E}\{e_{ix}\overline{e}_{jx}\} & \mathcal{E}\{e_{ix}\overline{e}_{jy}\} \\ \mathcal{E}\{e_{iy}\overline{e}_{jx}\} & \mathcal{E}\{e_{iy}\overline{e}_{jy}\} \end{bmatrix}, \ \mathbf{E} = \begin{bmatrix} \boldsymbol{E}_{11} & \cdots & \boldsymbol{E}_{1p} \\ \vdots & \ddots & \vdots \\ \boldsymbol{E}_{p1} & \cdots & \boldsymbol{E}_{pp} \end{bmatrix}$$

$\mathbf{E}$ is dependent on frequency and time, but for our analysis we will assume that we will work in a narrow subband and estimate the coherencies at sufficiently short time scales.

9.2.2 Observed covariance matrix

Instead of the field strengths, each telescope measures a voltage vector $\mathbf{v}_i$. Their relation is given by $\mathbf{v}_i = \boldsymbol{J}_i \mathbf{e}_i$, where $\boldsymbol{J}_i$ is a 2×2 matrix called the Jones matrix. It also incorporates the various ionospheric and atmospheric distortions, gain phase rotations and antenna feed polarisation leakage. Hence, the $\boldsymbol{J}_i$ are unknown and have to be estimated.

The observed voltages of the dual polarisation output signals of the telescopes i and j are cross-correlated into covariance matrices $\boldsymbol{R}_{ij}$, for which $\boldsymbol{R}_{ij} := \mathcal{E}\{\mathbf{v}_i \mathbf{v}_j^H\} = \boldsymbol{J}_i \boldsymbol{E}_{ij} \boldsymbol{J}_j^H$. Stacking the telescope output voltages $\mathbf{v}_i$ into a $2p$-dimensional vector $\mathbf{v} = [\mathbf{v}_1^t, \cdots, \mathbf{v}_p^t]^t$, and defining

$$\mathbf{J} = \begin{bmatrix} \boldsymbol{J}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{J}_p \end{bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} \boldsymbol{R}_{11} & \cdots & \boldsymbol{R}_{1p} \\ \vdots & \ddots & \vdots \\ \boldsymbol{R}_{p1} & \cdots & \boldsymbol{R}_{pp} \end{bmatrix}$$

it follows that the $2p \times 2p$ covariance matrix $\mathbf{R}$ is given by $\mathbf{R} = \mathbf{J}\mathbf{E}\mathbf{J}^H$.

In practice the observations are distorted by noise. The system noise signals of each of the two polarisation channels, $\mathbf{n}_i = [n_{ix}, n_{iy}]^t$ are stacked into a vector: $\mathbf{n} = \left(\mathbf{n}_1^t, \cdots, \mathbf{n}_p^t\right)^t$. The noise signals are uncorrelated between the telescopes, and up to a certain level also uncorrelated between the two polarisations of a telescope. In our analysis we assume that this is the case. Then the noise matrix

$\mathbf{D} = \mathcal{E}\{\mathbf{n}\mathbf{n}^H\}$ is diagonal: $\mathbf{D} = \mathrm{diag}(\sigma_{1x}^2, \sigma_{1y}^2, \cdots, \sigma_{px}^2, \sigma_{py}^2)$. The system noise can be considered additive, so that the covariance matrix of the received data can be written as $\mathbf{R} = \mathbf{J}\,\mathbf{E}\,\mathbf{J}^H + \mathbf{D}$.

9.2.3 Point source model

Under certain conditions, the electric field can be modelled as the contributions of a finite number of point sources:

$$\mathbf{R} \;=\; \sum_{\ell} \mathbf{J}_\ell\, \mathbf{E}_\ell\, \mathbf{J}_\ell^H \;+\; \mathbf{D}$$

where ℓ is the source direction and $\mathbf{E}_\ell$ is the coherency due to a single source from direction ℓ. Suppose that the source has sky brightness $\mathbf{B}_\ell$ (a 2×2 matrix determined by the source flux polarisation components or Stokes parameters). The relation of $\mathbf{B}_\ell$ to $\mathbf{E}_\ell$ can be written as $\boldsymbol{E}_{ij,\ell} = w_{ij,\ell}\, \mathbf{B}_\ell$ where $w_{ij,\ell}$ is the phase shift due to the geometric delay in an interferometer pair i-j [146]. Note that $w_{ij,\ell} = w_{i,\ell}\overline{w_{j,\ell}}$, where $w_{i,\ell}$ is the phase shift at a single telescope. Note that it is the same for the x and the y polarisation of this telescope. Thus define

$$\boldsymbol{W}_{i,\ell} = \begin{bmatrix} w_{i,\ell} & 0 \\ 0 & w_{i,\ell} \end{bmatrix}, \qquad \mathbf{W}_\ell = [\boldsymbol{W}_{1,\ell}^t, \cdots, \boldsymbol{W}_{p,\ell}^t]^t,$$

then $\boldsymbol{E}_{ij,\ell} = \boldsymbol{W}_{i,\ell}\, \mathbf{B}_\ell\, \boldsymbol{W}_{j,\ell}^H$, and $\mathbf{E}_\ell = \mathbf{W}_\ell\, \mathbf{B}_\ell\, \mathbf{W}_\ell^H$. The overall observed point source model thus becomes

$$\mathbf{R} \;=\; \sum_{\ell} \mathbf{J}_\ell \mathbf{W}_\ell \mathbf{B}_\ell \mathbf{W}_\ell^H \mathbf{J}_\ell^H \;+\; \mathbf{D}\,. \tag{9.1}$$

9.3 Gain calibration observations

During a calibration observation, the telescopes are pointed at a single dominant point source in the sky, with known sky brightness. The sum in equation (9.1) is reduced to a single term. Because the geometry of the telescope array is known, the delay matrix $\mathbf{W}_\ell$ is known as well. We thus obtain the observation model

$$\mathbf{R} = \mathbf{G}\mathbf{B}\mathbf{G}^H + \mathbf{D} \tag{9.2}$$

where we define the $2p \times 2$ gain matrix $\mathbf{G}$ by $\mathbf{G} = \mathbf{J}\mathbf{W}$. Our objective is to estimate $\mathbf{G}$ and $\mathbf{D}$, assuming that estimates of $\mathbf{R}$ and $\mathbf{B}$ are available. Since $\mathbf{W}_\ell$ is known, $\mathbf{J}$ is easily determined from $\mathbf{G}$. Alternatively, $\mathbf{R}$ can be corrected in advance for $\mathbf{W}_\ell$, after which we can assume without loss of generality that $\mathbf{W}_\ell = \mathbf{I}$ and that $\mathbf{G} = \mathbf{J}$ is direction-independent.

$\mathbf{R}$ is estimated by an observation covariance matrix $\widehat{\mathbf{R}}$, obtained by cross-correlation of N samples $\mathbf{x}_n$ of the telescope output signal vector, $\widehat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^H$.

9.4 Maximum likelihood and least squares

As will be shown later in this chapter, observations are required of three astronomical sources which must have different polarisation states ($\mathbf{R}_m = \mathbf{G}\mathbf{B}_m\mathbf{G}^H + \mathbf{D},\ m = 1, \cdots, 3$). As these observations are independent, the joint pdf can be written as a product of the pfd's belonging to the three individual observations. This leads to the following negative logarithmic maximum likelihood formula [168].

$$l(\mathbf{G}, \mathbf{D}) = \sum_{m=1}^{3} \left(\log |\mathbf{R}_m| + \mathrm{tr}(\mathbf{R}_m^{-1}\widehat{\mathbf{R}}_m) \right)$$

Finding the maximum of this formula involves complicated complex derivatives, and a solution can probably only be found in an iterative form. We therefore formulate the following least squares problem for finding $\mathbf{G}$ and $\mathbf{D}$.

$$\{\widehat{\mathbf{G}}, \widehat{\mathbf{D}}\} = \arg\min_{\mathbf{G},\mathbf{D}} \sum_{m=1}^{3} || \widehat{\mathbf{R}}_m - (\mathbf{G}\mathbf{B}_m\mathbf{G}^H + \mathbf{D}) ||_F^2$$

We assume that the noise power $\mathbf{D}$ is identical for all three observations.

9.5 Factor analysis algorithms

In this section, we will describe factor analysis algorithms which are needed for the polarisation gain estimations. The gain estimation algorithms themselves will be described in the next section. We will consider the factor model [91] $\mathbf{R} = \mathbf{A}\mathbf{A}^H + \mathbf{D}$, where the factor $\mathbf{A}$ is rank-two, and we will present two computationally efficient techniques.

9.5.1 Alternating Least Squares

A straightforward technique for trying to optimise a cost function over many parameters is to alternately minimise over a subset, keeping the remaining parameters fixed. This was explained in section 8.3.3 and is summarised here. In our case, let us assume at the k-th iteration that we have an estimate $\widehat{\mathbf{D}}[k]$. The next step is to minimise the LS cost function with respect to the gain vector only:

$$\widehat{\mathbf{A}}[k] \ = \ \arg\min_{\mathbf{A}} \ \| \widehat{\mathbf{R}} - \mathbf{A}\mathbf{A}^H - \widehat{\mathbf{D}}[k] \|_F^2 \tag{9.3}$$

The minimum is found from the eigenvalue decomposition $\widehat{\mathbf{R}} - \widehat{\mathbf{D}}[k] = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where the matrix $\mathbf{U} = [\mathbf{u}_1\ ,\ \cdots\ ,\ \mathbf{u}_{2p}]$ contains the eigenvectors $\mathbf{u}_i$, and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues λ_i, sorting in descending order. The factor minimising (9.3) is given by $\widehat{\mathbf{A}}[k] = [\mathbf{u}_1\lambda_1^{1/2} \quad \mathbf{u}_2\lambda_2^{1/2}]$. The second step is minimising with respect to the system noise matrix $\mathbf{D}$, keeping the gain vector fixed:

$$\widehat{\mathbf{D}}[k+1] \ = \ \arg\min_{\mathbf{D}} \ \| \widehat{\mathbf{R}} - \widehat{\mathbf{A}}[k]\widehat{\mathbf{A}}[k]^H - \mathbf{D} \|_F^2 \tag{9.4}$$

where $\mathbf{D}$ is constrained to be diagonal with nonnegative entries. The minimum is obtained by subtracting $\widehat{\mathbf{A}}[k]\widehat{\mathbf{A}}[k]^H$ from $\widehat{\mathbf{R}}$ and discarding all off-diagonal elements. Since each of the minimising steps in the iteration loop reduces the model error, we obtain monotonic convergence to a local minimum. Simulations indicate that in the absence of a reasonable initial point, convergence can be very slow.

9.5.2 Closed-form algorithm

The crux of this method is the observation that the off-diagonal entries of $\mathbf{AA}^H$ are equal to those of $\mathbf{R}$, and are known, so that we only need to reconstruct the diagonal entries of $\mathbf{AA}^H$. We further note that $\mathbf{AA}^H$ is rank two, so any submatrix of $\mathbf{R}$ that does not contain elements from the main diagonal is also rank two. This property can be used to estimate the ratio between any triplet of columns of $\mathbf{R}$ away from the diagonal, and subsequently to estimate how the main diagonal of $\mathbf{R}$ has to be changed so that the resulting $\mathbf{R}'$ is rank two, or $\mathbf{R}' = \mathbf{AA}^H$. The gain factor $\mathbf{A}$ can then be extracted by an eigenvalue decomposition.

To illustrate the idea, let (i, j, k) be a triplet of column indices, and let $\mathbf{M}$ be a submatrix of $\mathbf{R}$ consisting of columns (i, j, k), and all rows with indices unequal to i, j, k. Then $\mathbf{M}$ has three columns, and rank two, so that there exists a vector $\mathbf{v} = [v_1, v_2, v_3]^t$ such that $\mathbf{Mv} = 0$. The vector can be found from an SVD of $\mathbf{M}$. It follows that $[r'_{ii}, r_{ij}, r_{ik}]\mathbf{v} = 0$, so that $\hat{r}'_{ii} = -(r_{ij}v_2 + r_{ik}v_3)/v_1$. This estimate can be improved by considering all possible triplets containing i, and combining the ratios. After filling in all diagonal entries of $\mathbf{R}'$ in this way, a rank-two factorisation of $\mathbf{R}' = \mathbf{AA}^H$ provides an estimate for the factor $\mathbf{A}$. An estimate for $\mathbf{D}$ is subsequently found from $\mathbf{R} - \mathbf{AA}^H$.

9.6 Polarisation gain estimation algorithms

9.6.1 Closed form algorithm

One reference source
Consider a single source, $\mathbf{R} = \mathbf{GBG}^H + \mathbf{D}$, where $\mathbf{R}$ has been estimated and $\mathbf{B}$ is known from sky tables. Using factor analysis, we can find $\mathbf{D}$ and a factor $\mathbf{A}$ such that $\mathbf{R} = \mathbf{AA}^H + \mathbf{D}$. However, $\mathbf{A}$ is not unique: for any 2×2 unitary matrix $\mathbf{Q}$, we have $\mathbf{AA}^H = (\mathbf{AQ})(\mathbf{Q}^H\mathbf{A}^H)$. Hence, we can estimate $\mathbf{A}$ only up to a unitary factor. It follows that $\mathbf{G} = \mathbf{AQB}^{-1/2}$, where $\mathbf{Q}$ is unknown. It is not possible to estimate $\mathbf{G}$ in more detail using only a single reference source.

Two reference sources
With two reference sources, we have

$$\begin{aligned} \mathbf{R}_1 &= \mathbf{GB}_1\mathbf{G}^H + \mathbf{D} \\ \mathbf{R}_2 &= \mathbf{GB}_2\mathbf{G}^H + \mathbf{D} \end{aligned}$$

$\mathbf{R}_1$ and $\mathbf{R}_2$ are observed, $\mathbf{B}_1$ and $\mathbf{B}_2$ are the known polarisation matrices from the reference sources, and $\mathbf{D}$ is known from a factor analysis. Again, they are unique only up to unknown 2×2 unitary factors $\mathbf{Q}_1$, $\mathbf{Q}_2$.

A generalised eigenvalue decomposition of the pair $(\mathbf{B}_1, \mathbf{B}_2)$ provides the factorisations

$$\mathbf{B}_1 \;=\; \mathbf{M}\mathbf{\Lambda}_1\mathbf{M}^H, \quad \mathbf{B}_2 = \mathbf{M}\mathbf{\Lambda}_2\mathbf{M}^H\,, \tag{9.5}$$

where $\mathbf{M}$ is a square invertible matrix and $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2$ are positive diagonal matrices. It is assumed that the generalised eigenvalues are distinct.

The same decomposition on $(\mathbf{R}_1 - \mathbf{D}, \mathbf{R}_2 - \mathbf{D})$ gives

$$\begin{array}{rcl} \mathbf{R}_1 & = & \mathbf{A}\mathbf{\Lambda}_1\mathbf{A}^H + \mathbf{D} \\ \mathbf{R}_2 & = & \mathbf{A}\mathbf{\Lambda}_2\mathbf{A}^H + \mathbf{D} \end{array} \tag{9.6}$$

Note that the $\mathbf{\Lambda}_1$ in both cases is the same, and also $\mathbf{\Lambda}_2$. This is because

$$(\mathbf{R}_1 - \mathbf{D}) - \lambda(\mathbf{R}_2 - \mathbf{D}) = \mathbf{G}(\mathbf{B}_1 - \lambda\mathbf{B}_2)\mathbf{G}^H$$

so that $(\mathbf{B}_1, \mathbf{B}_2)$ and $(\mathbf{R}_1, \mathbf{R}_2)$ have the same rank reducing numbers λ.

Combining the two equations, we immediately obtain

$$\mathbf{A} = \mathbf{G}\mathbf{M}\bar{\mathbf{\Phi}} \quad \Rightarrow \quad \mathbf{G} = \mathbf{A}\mathbf{\Phi}\mathbf{M}^{-1} \tag{9.7}$$

where $\mathbf{\Phi} = \mathrm{diag}(\boldsymbol{\phi})$ is an unknown diagonal matrix with unimodular diagonal entries, representing phase ambiguities. Without further information, these cannot be further resolved.

An alternative computation that leads to (9.7) would go via factor analysis of $\mathbf{R}_1$ and $\mathbf{R}_2$ separately:

$$\begin{array}{rcl} \mathbf{R}_1 & = & \mathbf{A}_1\mathbf{A}_1^H + \mathbf{D}_1 = (\mathbf{A}_1\mathbf{Q}_1)(\mathbf{Q}_1^H\mathbf{A}_1^H) + \mathbf{D}_1 \\ \mathbf{R}_2 & = & \mathbf{A}_2\mathbf{A}_2^H + \mathbf{D}_2 = (\mathbf{A}_2\mathbf{Q}_2)(\mathbf{Q}_2^H\mathbf{A}_2^H) + \mathbf{D}_2 \end{array}$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are unknown unitary matrices. Setting $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{D}$, comparing the two equations with the model (9.6), and inserting (9.5), we obtain

$$\mathbf{G} \;=\; \mathbf{A}_1\mathbf{Q}_1\mathbf{\Lambda}_1^{-1/2}\mathbf{M}^{-1} \;=\; \mathbf{A}_2\mathbf{Q}_2\mathbf{\Lambda}_2^{-1/2}\mathbf{M}^{-1}$$

The latter equality relates $\mathbf{Q}_1$ to $\mathbf{Q}_2$ as

$$\mathbf{A}_1^\dagger\mathbf{A}_2 = \mathbf{Q}_1(\mathbf{\Lambda}_1^{-1/2}\mathbf{\Lambda}_2^{1/2})\mathbf{Q}_2^H\,.$$

This has the form of an SVD, and $\mathbf{Q}_1$ and $\mathbf{Q}_2$ can be computed as the left and right singular vectors of $\mathbf{A}_1^\dagger\mathbf{A}_2$. However, these are unique only up to an unknown diagonal phase matrix $\mathbf{\Phi}$. Hence we obtain

$$\mathbf{G} \;=\; (\mathbf{A}_1\mathbf{Q}_1)\mathbf{\Phi}(\mathbf{\Lambda}_1^{-1/2}\mathbf{M}^{-1})$$

where only $\mathbf{\Phi}$ is unknown. This is of the same form as (9.7).

Thus, we can estimate $\mathbf{G}$ only up to two unknown phases. With some effort, this can be converted to a more convenient normalisation: if we define a normalised $\mathbf{G}$ to have positive real entries on its first row, then the normalised $\mathbf{G}$ is unique. This is the best that can be expected using two reference sources.

Three reference sources
If a third observation of a point source is available, $\mathbf{R}_3 = \mathbf{G}\mathbf{B}_3\mathbf{G}^H + \mathbf{D}$, then the ambiguity can be further reduced to a single common phase. Indeed, after a similar generalised eigenvalue decomposition on the pair $(\mathbf{R}_1, \mathbf{R}_3)$, we have two sets of equations available,

$$\mathbf{G} = \mathbf{A}_1\mathbf{\Phi}_1\mathbf{M}_1^{-1} = \mathbf{A}_2\mathbf{\Phi}_2\mathbf{M}_2^{-1}$$

or, denoting by $\circ$ the Khatri-Rao (column-wise Kronecker) product,

$$\text{vec}(\mathbf{G}) = (\mathbf{M}_1^{-T} \circ \mathbf{A}_1)\boldsymbol{\phi}_1 = (\mathbf{M}_2^{-T} \circ \mathbf{A}_2)\boldsymbol{\phi}_2 \tag{9.8}$$

or

$$[(\mathbf{M}_1^{-T} \circ \mathbf{A}_1) \quad -(\mathbf{M}_2^{-T} \circ \mathbf{A}_2)] \begin{bmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \end{bmatrix} = 0\,.$$

Thus, $[\boldsymbol{\phi}_1^T,\ \boldsymbol{\phi}_2^T]^T$ is the unique solution in the null space, and determined up to a scaling ϕ. Inserting in (9.8), we obtain $\mathbf{G}$ up to an unknown scaling by ϕ.

9.6.2 Parallel factor analysis

In [150] [101] [175] parallel factor analysis solutions are given for the model $||\widehat{\mathbf{R}} - \mathbf{ABC}||_F^2$. Our model, which is based on a combination of three sources $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_3$, is slightly different. Let $\widetilde{\mathbf{R}} = \left[\widehat{\mathbf{R}}_1^t\ \widehat{\mathbf{R}}_2^t\ \widehat{\mathbf{R}}_3^t\right]^t$, and let $\widetilde{\mathbf{D}} = [\mathbf{D}^t\ \mathbf{D}^t\ \mathbf{D}^t]^t$. Define also $\widetilde{\mathbf{G}}_\mathbf{B} = [(\mathbf{G}\mathbf{B}_1)^t\ (\mathbf{G}\mathbf{B}_2)^t(\mathbf{G}\mathbf{B}_3)^t]^t$, then, assuming that $\mathbf{D}$ is found by using one of the factor analysis algorithms described earlier, the Least Squares cost function to minimise is given by:

$$\{\widehat{\mathbf{G}}\} = \arg\min_{\mathbf{G}}\ ||\,\widetilde{\mathbf{R}} - \widetilde{\mathbf{D}} - (\widetilde{\mathbf{G}}_\mathbf{B}\mathbf{G}^H)\,||_F^2 \tag{9.9}$$

Our model differs from [175] [150] in the sense that the $\mathbf{B}_k$ matrices in general are not diagonal, also our model has more structure than in [101] as the two $\mathbf{G}$ matrices are identical. A solution for the cost function above can be found by iteratively estimating the gain matrix $\widehat{\mathbf{G}}$ by $\widehat{\mathbf{G}} = (\widetilde{\mathbf{G}}_\mathbf{B})^\dagger(\widetilde{\mathbf{R}} - \widetilde{\mathbf{D}}))^H$ and inserting it in $\widetilde{\mathbf{G}}_\mathbf{B}$. For this algorithm, a good initial point is needed; one of the two proposed closed form algorithms can be used for this purpose.

9.7 Simulations

The performance of the closed-form and parafac dual polarisation gain estimation methods is studied by applying them to three generated covariance matrices

$\widehat{\mathbf{R}}_k$, based on the "true" values of $\mathbf{G}$, $\mathbf{D}$, and $\mathbf{B}_k$. The performance is quantified by calculating the model error $||\widetilde{\mathbf{R}} - \widetilde{\mathbf{D}} - \widetilde{\mathbf{G}}_{\mathbf{B}}\widehat{\mathbf{G}}||_F$. Here, the noise matrices $\widehat{\mathbf{D}}$ or $\widetilde{\mathbf{D}}$ are estimated using the closed-form factor analysis algorithms. In the simulations, the $\mathbf{B}_k$ matrices used are:

$$\mathbf{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{B}_2 = \begin{bmatrix} 1.2 & 0.1 \\ 0.1 & 0.8 \end{bmatrix}, \mathbf{B}_3 = \begin{bmatrix} 1 & -0.1\imath \\ 0.1\imath & 1 \end{bmatrix}$$

The gain magnitudes are nominally one, with gain magnitude variations (over the 2p telescope channels) up to 10%, and phase variations up to 2π radians.

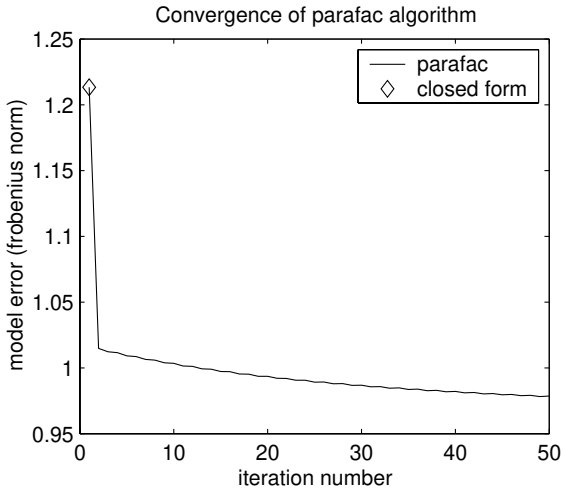


Figure 9.1. Convergence of the parafac algorithm, using the closed form algorithm as a starting point ($Nsam = 1000$, $p = 5$, $\mathbf{D} = \mathbf{I}$, $N_{it} = 256$).

Figure 9.1 shows the convergence of the parafac method, using the closed form method as initial point. The improvement of the parafac method with regard to the closed form method is not dramatic. Figure 9.2 shows the model error of the closed form method compared to the model error of the true gains. The closed form model error is apparently very close to the model error of the true gains.

9.8 Conclusions

We have shown that observations of at least three astronomical sources are required to fully solve the dual polarisation gain estimation problem of radio telescope arrays. We have presented several closed-form and iterative solutions to the dual polarisation gain estimation problem. These solutions are not optimal as they do not fully exploit the structure which is present in the model.

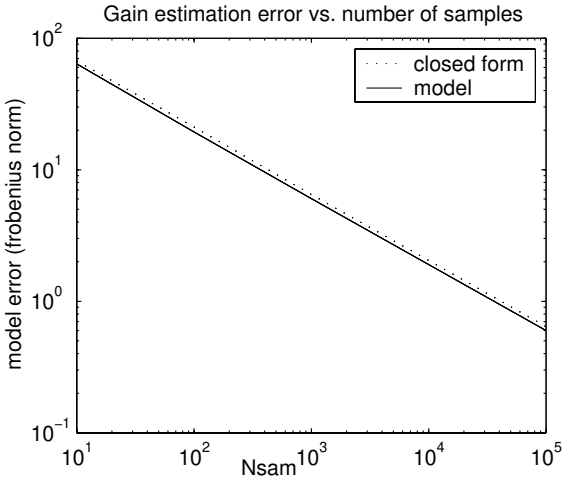


Figure 9.2. Performance of the closed form algorithm, compared to the model error using the true gains. ($p = 5$, $\mathbf{D} = 10\,\mathbf{I}$, $N_{it} = 128$).

However, initial simulations show that, for these solutions, the model error is close to the model error using the true gain values.

Chapter 10

Implications for future telescopes

10.1 Introduction

In previous chapters, multichannel detection, excision, and spatial filtering techniques were presented. Both theoretical analyses and experimental results were shown. The influence of several system and signal parameters on interference mitigation effectiveness was analysed as well. In this chapter, implications of this knowledge for the next generation of radio telescopes will be discussed. The focus will be on the LOFAR telescope, a large distributed aperture synthesis array, consisting of many antenna elements. However, many of the implications for LOFAR will also be applicable to other large-scale aperture synthesis telescopes, such as SKA . This chapter will describe the LOFAR interference mitigation strategy, and interference mitigation options in the LOFAR design. Finally, some initial interference filtering results from the LOFAR Initial Test Station (ITS) will be presented.

10.2 LOFAR interference mitigation strategy

10.2.1 The LOFAR telescope

The Low Frequency Array (LOFAR) [27, 85, 32] is a next-generation radio telescope which is currently being designed and constructed in the Netherlands, with possible extensions to neighbouring countries. The observational frequency window of LOFAR will lie in the 10-250 MHz range. It is an aperture array synthesis telescope [146, 117] and will consist of order 100 telescopes, also called LOFAR stations. These stations will be positioned in a spiral configuration over an area with a radius of a few hundred kilometers, as well as in a more densely populated central region, also known as the central core. Figure 10.1 shows an

artist impression of a possible LOFAR configuration.

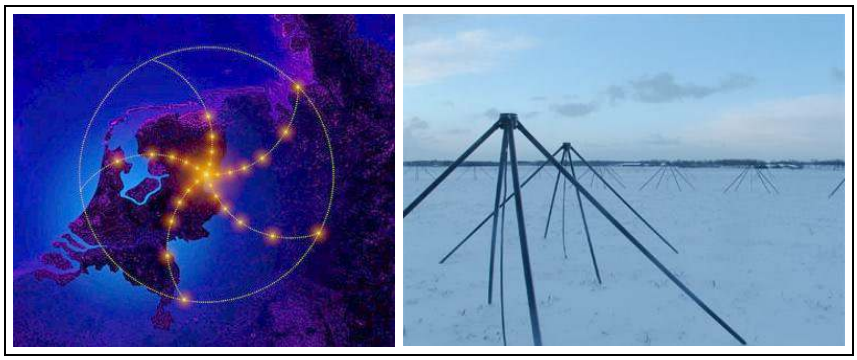


Figure 10.1. Artist impression of the distribution of LOFAR telescope stations in the Netherlands and Germany (left), photograph of the LOFAR initial test station (ITS) in Borger-Odoorn (right).

Each of the stations will consist of about 100 sky noise limited dual polarisation antennas[1]. These antennas are combined in such a way that station beams can be formed for each of the desired look directions or pointings. The preliminary LOFAR design defines multiple beam capabilities, (non-contiguous) 4 MHz wide bands which are composed of order 200 kHz wide subbands. The station signals are transported to a central location where they are time-delayed in order to compensate for their geometric delay differences. After delay correction, the signals are further split into 1 kHz wide frequency channels, after which they are correlated, integrated in time, and stored. These correlation products are subsequently processed to form astronomical sky maps. The station and central processing signal chains are schematically shown in figure 10.2. About one quarter of the LOFAR stations will be located in a central area with a diameter of a few kilometers, also known as the central core. Not all the design and implementation details of LOFAR were fixed at the time of writing of this thesis; the final LOFAR design may therefore deviate from the one sketched here.

In addition to synthesis imaging capabilities, LOFAR also supports other modes, such as:

- transient detection using wideband waveform signal processing or space-time array processing (STAP) [67].
- tied array beamforming, that is, summing the station output signals coherently instead of correlating them.

[1] Actually, LOFAR will have different antenna concepts for different frequency bands, the Low Band Antenna for frequencies below 90 MHz, and the High Band Antenna for frequencies above 110 MHz

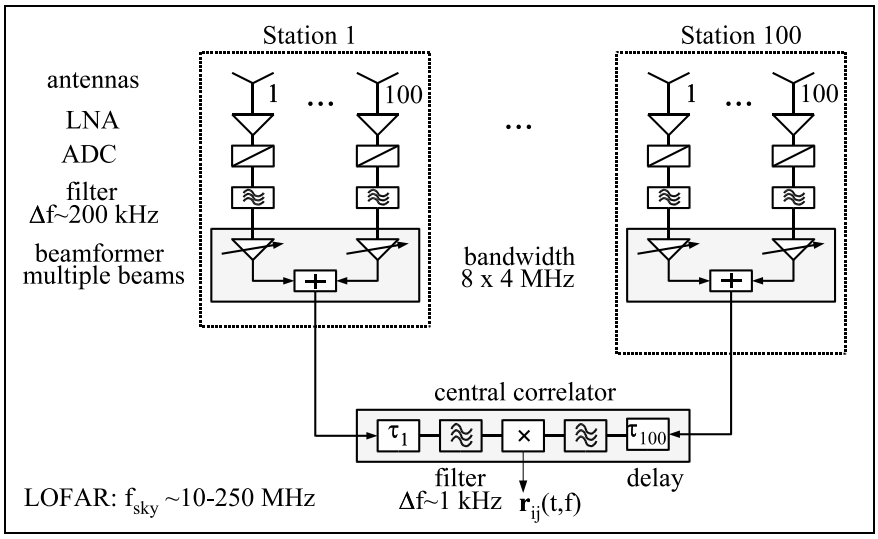


Figure 10.2. Schematic overview of LOFAR . The antenna signals in a station are amplified using low noise amplifiers (LNAs), filtered and digitised, and added using appropriate complex weights (beamforming). After beamforming, the signals are compensated for their relative geometric delay differences, filtered further, and finally are correlated between the stations. These correlation products $\mathbf{r}_{ij}(t, f)$ are used to generate astronomical sky images.

- antenna based buffering of 1 second duration at full-digitised bandwidth for limited detection/triggering capabilities.

However, the focus in this chapter will be on aperture synthesis, as this is the main observing mode.

The LOFAR initial operations phase is scheduled to start in 2006; the target date to have LOFAR fully operational is 2008. Figure 10.1 shows a photograph of the LOFAR Initial Test Station (ITS), which became operational in December 2003.

10.2.2 Interference mitigation strategy

LOFAR will operate in bands allocated to radio astronomy as well as in bands where other spectrum users are active. It is estimated that 4 MHz wide frequency bands will be needed for which 80% bandwidth can be recovered in order to achieve the sensitivity and the required dynamic range specified in [27, 85]. Here it is assumed that 20% bandwidth loss due to interference, outside the bands allocated to radio astronomy, does not lead to a dramatic decrease in sensitivity of the instrument as a whole. For the 80 % cleanest frequency bins within a 4 MHz band, it is also assumed that the interference at station level

is either at a 1-3 (station) sky noise sigma level, or can be reduced to this level by RFI mitigation techniques. If the interference and the telescope meet certain requirements, the interference can be reduced to power levels which lie below the sky map noise. The rationale behind this is explained in the following sections. Both the LOFAR interference mitigation approach and some of the factors influencing mitigation effectiveness will be described. The focus is on mitigation in the digital domain; it is assumed that the analog sections of the receiver remain linear. In [13] it is shown that a 12 bit ADC has sufficient dynamic range to allow observations close to strong transmitters.

Interference mitigation approach

Which fraction of the spectrum can be used for radio astronomy depends on spectrum occupancy statistics and on the interference power levels relative to the noise in the sky maps. It also depends on telescope design, and on interference mitigation effectiveness. LOFAR will be one of the first radio telescopes in which RFI mitigation techniques form an integral part of the system design. Initially, LOFAR will be equipped with relatively simple mitigation techniques. An important aim of the design is that it should be possible to add new interference mitigation functionality at a later stage, without major hardware and image processing redesigns. As LOFAR is, to a large extent, a software telescope[2] this flexibility will be built-in.

The calibration and imaging capabilities of LOFAR [109] include removal of relatively strong cosmic sky sources such as Cas.A from the observed data sets and images. This process not only removes the source from the map, but also its spatial sidelobes, which might otherwise overlap with the sky region of interest. This suggests that mitigation techniques need to suppress interference at least down to the Cas.A flux level. If the interferer meets certain requirements, then it spatially behaves just like a cosmic source. This would mean that it can be removed with standard signal processing approaches. The interferer requirements are: the interfering source must be spatial point source, and it must be quasi stationary with a fixed or slowly varying spatial signature vector. It is also assumed that the number of interferers that need to be suppressed is limited, as computing resources are limited.

Clearly, much research is needed to verify these assumptions. However, spectrum monitoring observations [12] and initial test observations at both the ITS in the 10-40 MHz band [170, 17] and the WSRT in the 110-180 MHz band indicate that successful observations can be carried out. The LOFAR interference mitigation strategy is based on the assumptions mentioned above and distinguishes three steps:

- Selecting the cleanest frequency subbands within the LOFAR bands.
- Reducing RFI levels by RFI mitigation down to Cas.A power flux levels.

[2]Software telescope is defined analogously to the term software radio, a concept in which radio waves are sampled and signal processing (decoding, demodulation) is done in (embedded) software.

- Further reducing RFI levels to levels below the noise in the sky maps.

Step one: selection of clean subbands

The first interference defence is to select subband frequencies in such a way that the number of received interfering signals is minimised. This obviously only works for spectral continuum observations, not for spectral line work. Monitoring observations [12, 41, 35] have been carried out to investigate spectral occupancies. Figure 10.3 shows a monitoring observation carried out in 2002 at the LOFAR central core site, for which the median and 90 percentile spectra are plotted. In this particular example, the system noise is dominated by the receiver noise, and less than 20% of the frequency bands (outside the 87.5-108 MHz FM radio band) were occupied with signal powers exceeding 0 dBμVm^{-1}.

Many transmitters in the LOFAR bands have bandwidths of approximately 1-3 kHz. This roughly matches the LOFAR frequency resolution of 1 kHz. This implies that affected frequency bins can be excised with only a moderate decrease in system sensitivity, assuming that less than 20% of the selected 4 MHz (non-contiguous) spectrum bands are affected. The remaining 80 % needs to be free from interference, before or after interference mitigation.

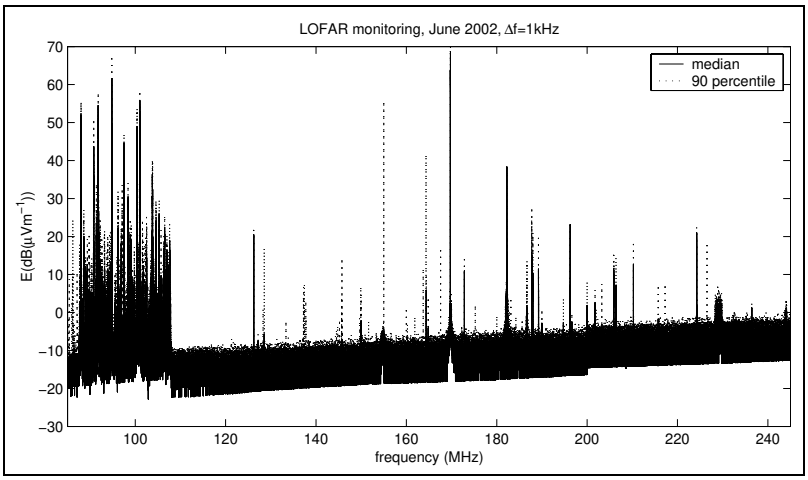


Figure 10.3. LOFAR monitoring spectrum, obtained in 2002 in Drenthe, The Netherlands. Visible are the FM radio band (87.5-108.0 MHz), T-DAB digital radio (229 MHz), mobile use such as paging and marine telephone (170,160 MHz), and TV carrier waves (182.25,187.75 MHz). The band 150.05-153.00 MHz has a (shared) radio astronomy allocation.

Step two: reduction of interference down to the level of Cas.A

The second step in the strategy is to suppress interference in the subbands down to the Cas.A power flux level, using interference mitigation techniques. These techniques will be applied at time scales up to order 10 seconds. In order

to estimate the required interference attenuation levels, the observed spectrum power fluxes need to be related to the LOFAR sensitivity. Figure 10.4 illustrates the levels involved. The vertical axis represents radio wave flux levels and sensi-

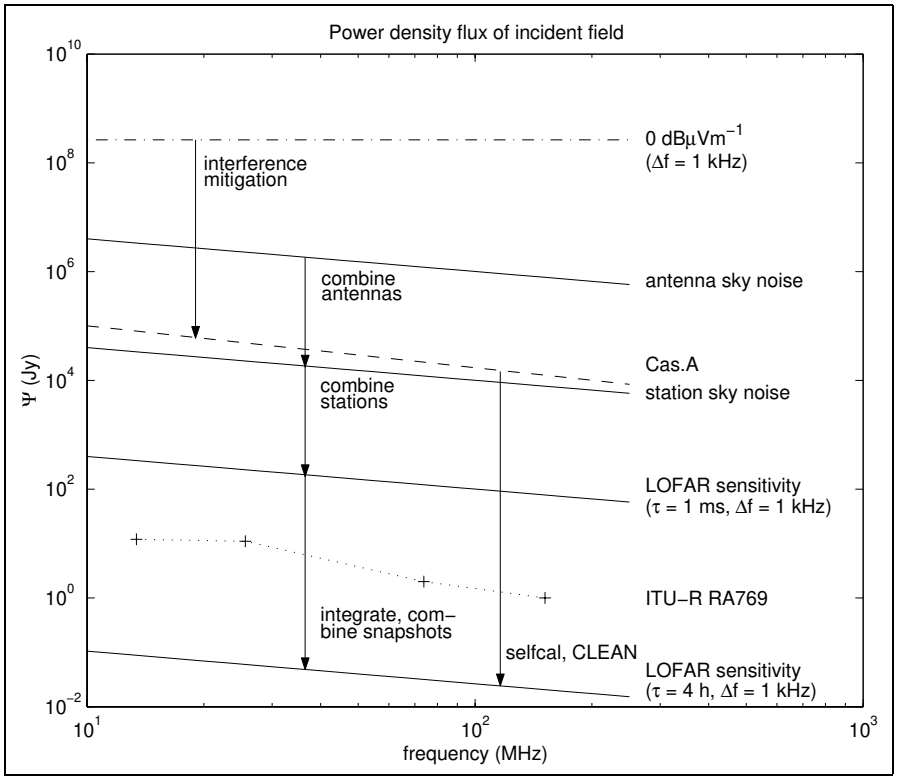


Figure 10.4. LOFAR sensitivity levels.

tivities in Jy, where 1 Jy = 10^{-26} W m^2 Hz^{-1}, the horizontal axis representing the frequency. The upper curve corresponds to a transmitter flux level[3] of 0 dBμVm^{-1}, where it is assumed that the transmitter power is concentrated in a 1 kHz band. The curve "antenna sky noise" shows the sky noise flux Ψ as a function of frequency, as would be observed with a single polarisation LOFAR antenna dipole. It is based on the sky temperature, given by the approximate formula [27, 87]

$$T_{sky} \approx T_{so}\lambda^{2.55}\ (K) \tag{10.1}$$

and on a formula [87, 124] relating sky temperature T_{sky} to flux density Ψ.

$$\Psi = \frac{2k}{\lambda^2} T_{sky}\Omega\ (Jy) \tag{10.2}$$

[3]The electric field power unit dBμVm^{-1} for an electric field E in Vm^{-1}is defined by 20 ^{10}log(10^6E)

Here Ω is the antenna solid angle (assumed to be 4 Sr), k the Boltzmann constant, T_{so} is 60 ± 20 K for angular distances to the galactic plane larger that 10°, and λ is the wavelength in m. According to the ITU radio noise recommendation [70], the environmental RF noise floor in rural, residential and business areas typically lies 5 to 20 dB above sky background noise, and for quiet rural areas, it lies ≈ 5 to 10 dB below background noise. In practical situations, these levels are strongly time and frequency dependent and also dependent on location and height. Note that for the LOFAR central core location in the Netherlands the sky background noise has already been detected for several bands [170], which means that in those cases the "rural curve" mentioned in [70] is too pessimistic. In order to keep the number of curves in the figure limited, these radio noise curves have not been drawn.

Curve "Cas.A" represents the flux level of the astronomical source Cas.A [5], one of the strongest sky sources in the radio frequency regime. A few dBs below Cas.A lies the station beam noise level, which will be calculated next.

Effect of station beamforming on SNR

In a station, order 100 antennas are combined in a phased array beamformer . One or multiple output beams are constructed by summing the antenna signals, after applying suitable beamformer weights $\mathbf{w}$. Assume that there is a single dominant source at a direction $\mathbf{s}_\ell$ with spatial signature vector $\mathbf{a}_\ell$. Further assume that the source has a signal waveform $s_\ell(t)$ and power $\sigma_\ell^2 = \mathcal{E}\{s_\ell(t)^2\}$. Let p be the number of antennas in the two-dimensional array, and stack the array element position vectors in an antenna location matrix $\mathcal{R}$, as in equation (2.19). For $\mathbf{a}_\ell$ this yields:

$$\mathbf{a}_\ell = \boldsymbol{\gamma} \odot e^{-\jmath 2\pi \lambda^{-1} \mathcal{R} \mathbf{s}_\ell} \tag{10.3}$$

where $\boldsymbol{\gamma}$ is a vector containing the antenna element gains, which may differ among the antennas. Note that $\mathcal{R}$ is a $(p \times 3)$ matrix and $\mathbf{s}_\ell$ a vector of length 3, also note that $\mathbf{a}_\ell^H \mathbf{a}_\ell = \boldsymbol{\gamma}^H \boldsymbol{\gamma}$. Define the array steering vector $\mathbf{w}_o$ for the direction $\mathbf{s}_o$ by

$$\mathbf{w}_o = \frac{1}{\sqrt{p}}\, e^{-\jmath 2\pi \lambda^{-1} \mathcal{R} \mathbf{s}_o} \tag{10.4}$$

For this classical beamformer the steering vector is normalised, $\mathbf{w}_o^H \mathbf{w}_o = 1$. Dropping time indices, let the array output vector $\mathbf{x}$ be defined by

$$\mathbf{x} = \mathbf{a}_l s_l + \mathbf{n} \tag{10.5}$$

where $\mathbf{n}$ is the i.i.d. array noise vector with $\mathcal{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2 \mathbf{I}$. For this system, the array beamformer output $y = \mathbf{w}_o^H \mathbf{x}$ and its power, the "dirty beam" [95, 151] $I_D(\mathbf{s}_\ell, \mathbf{s}_0) = \mathcal{E}\{|y|^2\}$ can be described by

$$\begin{aligned} y &= \mathbf{w}_o^H \mathbf{a}_\ell \; s_\ell + \mathbf{w}_o^H \mathbf{n} && (10.6) \\ I_D(\mathbf{s}_\ell, \mathbf{s}_0) &= \mathbf{w}_o^H \mathbf{a}_\ell \mathbf{a}_\ell^H \mathbf{w}_o \; \sigma_\ell^2 + \sigma_n^2 \\ &= \frac{1}{p} \sigma_l^2 |\boldsymbol{\gamma}^t e^{-\jmath 2\pi \lambda^{-1} \mathcal{R}(\mathbf{s}_\ell - \mathbf{s}_o)}|^2 + \sigma_n^2 && (10.7) \end{aligned}$$

The latter expression is derived using relation (B.14). When the source direction and look direction coincide, $\mathbf{s}_\ell = \mathbf{s}_0$, and when antenna gains are unity: $\boldsymbol{\gamma} = \mathbf{1}$, the beamformer output can be written as:

$$I_D(\mathbf{s}_\ell, \mathbf{s}_\ell) = p\, \boldsymbol{\sigma}_\ell^2 + \sigma_n^2 \tag{10.8}$$

The SNR scales linearly with the number of antennas in a phased array, a well-known fact in array processing. For 100 antennas in a phased array beamformer this means an increase in sensitivity of 20 dB. This explains that the station sensitivity curve lies 20 dB below the single-antenna sky noise curve in figure 10.4.

In the lower part of figure 10.4, two "LOFAR sensitivity" curves are drawn, both for 1 kHz bandwidth and for a synthesis array of 100 stations. The two curves differ in integration time: 1ms for the upper curve (corresponding to one sample moment or "snapshot"), and 4h integration time for the lower curve.

As a reference, the ITU-R RA769 radio astronomy protection criterion [69] is given. It roughly states the RFI level at which the error in determining signal power exceeds 10%, for an integration time of 2000 s. The curve is based on bandwidth values mentioned in [69], which are larger than the 1 kHz LOFAR frequency resolution. For example, the depicted value of 11 dBJy at 13.385 MHz corresponds to 50 kHz bandwidth and 2000 s integration. The value at 13.385 MHz for four hours of integration and 1 kHz bandwidth would be $10\,^{10}\log(\sqrt{50 \times 2000/1 \times 4 \times 3600}) = 4$ dB higher.

Step three: reduction of interference down to levels below the sky noise in the map

The third step in the strategy is reducing interference from 'Cas.A' level down to sky image noise levels. This step is closely related to the image making process, and it involves long integration times. Assuming that the interferers are spatial point sources, removing them should in principle not be too different from removing astronomical sources such as Cas.A. In addition to selfcal and CLEAN, other methods such as spatial filtering in the image formation process [94], and new calibration approaches [110] could be used as well. Long-term and short-term stationarity issues as well as estimation biases [98,119,147] are relevant here and need careful consideration. However, this is not pursued any further in this thesis.

Spectral dilution

Two additional effects help reduce the observed interference. The first is spectral dilution. Suppose that in a LOFAR band a single narrowband RFI spatial point source with spatial signature vector $\mathbf{a}_r$ and power σ_r^2 is present. The corresponding covariance model for a single frequency channel, as defined before, is $\mathbf{R} = \sigma_r^2 \mathbf{a}_r \mathbf{a}_r^H + \sigma_n^2 \mathbf{I}$, where σ_n^2 is the i.i.d. telescope noise power. The power of the RFI source in the covariance product r_{ij} is diluted by a factor N_f^{-1} when averaging all N_f frequency channels in the band. Its standard deviation

$\mathrm{std}(r_{ij})$ can be expressed as

$$\mathrm{std}(r_{ij}) = \sqrt{\frac{1}{N_f} r_{ii} r_{jj}} \tag{10.9}$$

(cf. equation C.15), assuming that the signals in different frequency channels are independent, and assuming that RFI can be modelled as Gaussian noise. The standard deviation decreases with $\sqrt{N_f}$, whereas the RFI power in the covariance product r_{ij} decreases with N_f. This leads to a spectral dilution factor which scales with $\sqrt{N_f}$. Further investigation is needed to determine the power level above which excision is better than spectral dilution for practical cost-effective systems.

Spatial dilution

The second effect which reduces interference is spatial dilution. In the aperture synthesis mode, snapshot images are made and are further integrated to form a sky map. Sky sources move with respect to the baseline vectors during an observation. The data processing corrects for this motion, so that the snapshots can be added coherently. Interferers, on the other hand, do not move with respect to the baseline vectors or move differently. This means that in the snapshot averaging process RFI sources are diluted. This is not (necessarily) true for the dilution of the RFI spatial sidelobes, which may overlap with a sky region of interest.

The dilution of an RFI point source is comparable to the dilution of a narrowband signal due to frequency averaging: the system noise standard deviation decreases with the square root of the number of snapshots N whereas the RFI is reduced (as it moves around the map) by the number of snapshots. This means that if a point source is reduced to the station noise level, it will be reduced to below the integrated noise level by further integration (snapshot averaging). This statement will be now be explained.

Consider a simplified model in which there is one interfering source with a snapshot spatial signature vector $\mathbf{a}_k$, and interferer source direction $\mathbf{s}_k$, so that

$$\mathbf{a}_k = \boldsymbol{\gamma} \odot e^{-\jmath 2\pi \lambda^{-1} \mathcal{R} \mathbf{s}_k} \tag{10.10}$$

Note that, as before, $\mathbf{a}_k^H \mathbf{a}_k = \boldsymbol{\gamma}^H \boldsymbol{\gamma}$, and that $\mathbf{a}_k^H \mathbf{a}_k = p$ for $\boldsymbol{\gamma} = \mathbf{1}$. The signature vector $\mathbf{a}_k$ for each of the snapshots k may differ, the interferer power σ_r^2 is assumed constant. Further assume i.i.d. noise with power σ_n^2, so that the array covariance for the k^{th} snapshot is given by:

$$\mathbf{R}_k = \sigma_r^2 \mathbf{a}_k \mathbf{a}_k^H + \sigma_n^2 \mathbf{I} \tag{10.11}$$

The snapshot beam output power $I_D^s(\mathbf{s}_k^o)$ is given by $I_D^s(\mathbf{s}_k^o) = \mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k$, where the steering vector $\mathbf{w}_k$ corresponding to the direction $\mathbf{s}_k^o$ is given by

$$\mathbf{w}_k = \frac{1}{\sqrt{p}} \, e^{-\jmath 2\pi \lambda^{-1} \mathcal{R} \mathbf{s}_k^o} \tag{10.12}$$

It has unit norm, $\mathbf{w}_k^H \mathbf{w}_k = 1$. The beam power I_D of all combined N snapshots yields:

$$I_D = \frac{1}{N} \sum_{k=1}^{N} I_D^s(\mathbf{s}_k^o) = \sigma_n^2 + \sigma_r^2 \frac{1}{N} \sum_{k=1}^{N} |\mathbf{w}_k^H \mathbf{a}_k|^2 \tag{10.13}$$

Now assume that for the m^{th} time index, $k = m$, the steered beam is pointed at the interfering source: $\mathbf{w}_m = \mathbf{a}_m / \sqrt{\mathbf{a}_m^H \mathbf{a}_m}$, and that the interference enters the beam via its sidelobes for $k \neq m$. Under these assumptions the beam takes the form

$$I_D = \sigma_n^2 + \frac{1}{N} \sigma_r^2 \boldsymbol{\gamma}^H \boldsymbol{\gamma} + \frac{1}{N} \sigma_r^2 \sum_{\substack{k=1 \\ k \neq m}}^{N} |\mathbf{w}_k^H \mathbf{a}_k|^2 \tag{10.14}$$

The second term of the expression above is the interferer power term, spatially diluted with the factor N. Assuming $\boldsymbol{\gamma} = \mathbf{1}$ and ignoring the finite sample effect ($\hat{I}_D = \frac{1}{N} \sum_k \mathbf{w}_k^H \widehat{\mathbf{R}}_k \mathbf{w}_k$), this term reduces to $p\,\sigma_r^2/N$. The third term is the "sidelobe term". If the assumption is made that the steering vector and the interferer source spatial signature vector are independent from each other, then the third term could be modelled as a random walk problem. This can be seen as follows. Insert the expressions for the vectors $\mathbf{w}_k$ and $\mathbf{a}_k$ in $|\mathbf{w}_k^H \mathbf{a}_k|^2$, where the directions $\mathbf{s}_k^o$ and $\mathbf{s}_k$ are assumed to be random and independent:

$$\begin{aligned} |\mathbf{w}_k^H \mathbf{a}_k|^2 &= \frac{1}{p} \left| \left[e^{-\jmath 2\pi \lambda^{-1} \mathcal{R} \mathbf{s}_k^o} \right]^H \left[e^{-\jmath 2\pi \lambda^{-1} \mathcal{R} \mathbf{s}_k} \right] \right|^2 \\ &= \frac{1}{p} \left| \mathbf{1}^t e^{-\jmath 2\pi \lambda^{-1} \mathcal{R} (\mathbf{s}_k - \mathbf{s}_k^o)} \right|^2 \end{aligned} \tag{10.15}$$

Under the assumptions made, this expression represents a two dimensional random walk problem [166]: $z = \sum_{i=1}^{p} e^{\jmath \Theta_i}$, with Θ a uniformly distributed random variable in the interval $[0, 2\pi)$, and with $\mathcal{E}\{|z|^2\} = p$. This results in the following approximate equation

$$I_D \approx \sigma_n^2 + \frac{1}{N} p\,\sigma_r^2 + \sigma_r^2 \tag{10.16}$$

The second term of this equation is the spatially diluted interferer signal. The third term is the interferer sidelobe contribution. It shows that the interferer signals entering via the sidelobes are not spatially diluted.

The beamformer output variance is calculated next, in order to relate the sky map noise to the diluted interference power as described above.

Sky map variance

In the sky map noise analysis it is assumed that there is only system noise, and no strong cosmic source or interference. The beamformer output noise power can be described by using the array beamforming model as described before

$$I_D = \mathcal{E}\{|y|^2\} = \mathcal{E}\{\mathbf{w}^H \hat{\mathbf{x}} \hat{\mathbf{x}}^H \mathbf{w}\} = \mathbf{w}^H \mathbf{R} \mathbf{w} = \sigma_n^2 \tag{10.17}$$

The beam output power variance $\sigma_b^2 = \mathrm{var}(\hat{I}_D)$ is given by:

$$\begin{aligned}
\sigma_b^2 &= \mathrm{var}(\mathbf{w}^H \widehat{\mathbf{R}} \mathbf{w}) && (10.18)\\
&= (\mathbf{w} \odot \overline{\mathbf{w}})^H \mathrm{var}(\widehat{\mathbf{R}})(\mathbf{w} \odot \overline{\mathbf{w}}) && (10.19)\\
&= (\mathbf{w} \odot \overline{\mathbf{w}})^H \mathrm{vecdiag}(\overline{\mathbf{R}})(\mathrm{vecdiag}(\mathbf{R}))^t (\mathbf{w} \odot \overline{\mathbf{w}}) && (10.20)\\
&= \sigma_n^4 (\mathbf{w} \odot \overline{\mathbf{w}})^H \mathbf{1}\mathbf{1}^t (\mathbf{w} \odot \overline{\mathbf{w}}) && (10.21)
\end{aligned}$$

The variance of $\hat{I}_D$ can be approximated, assuming that p is large and that the locations of the LOFAR stations are not regular so that they yield arbitrary phases ϕ_i. In this case, expression (10.12) takes the form $\mathbf{w} = (1/\sqrt{p})$ $(e^{j\phi_1}, \cdots, e^{j\phi_p})^t$, where ϕ_i is considered a random variable. Now $(\mathbf{w} \odot \overline{\mathbf{w}})^H$ $\mathbf{1}\mathbf{1}^t$ $(\mathbf{w} \odot \overline{\mathbf{w}})$ can be written as follows

$$(\mathbf{w} \odot \overline{\mathbf{w}})^H \mathbf{1}\mathbf{1}^t (\mathbf{w} \odot \overline{\mathbf{w}}) = \sum_{i=1}^{p} \overline{w_i^2}\, w_i^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{p} \overline{w_i^2}\, w_j^2 \qquad (10.22)$$

This expression also represents a two dimensional random walk problem [166]. The first term is obviously equal to p^{-1}, the expected value of the second term is zero. The power variance due to system noise only, can therefore be approximated by

$$\sigma_b^2 = \frac{\sigma_n^4}{p} \qquad (10.23)$$

So far the analysis was based on one snapshot, that is, one sample moment in time. For N snapshots, the (u, v) aperture is spatially sampled with pN sample points, which means that for N snapshots, the p values in equation (10.23) must be replaced by pN. This leads to the following conclusion: if an interferer, moving relative to the sky, crosses the field of view under consideration, and if this interference is reduced to the station noise power levels, then further snapshot averaging will reduce it to below the noise in the sky map, the reason being that the noise rms in the map decreases proportionally to $(p\,N)^{-\frac{1}{2}}$, and the spatially diluted interference level decreases at least as fast as $(p\,N)^{-1}$. This is not true in case only the interferer spatial sidelobes enter the field of view under consideration.

It may be necessary to suppress the interference to levels below those described in the paragraphs above. One reason for this may be that the interference does not fulfill the requirements given earlier. A second reason may be that the calibration process is hampered by the interference. Small interference residuals may introduce calibration distortions which become apparent only after long integration times. However, it seems feasible that even in densely populated regions such as the Netherlands, the station sky noise level can be reached for wide frequency ranges. It is also expected that the integrated sky noise levels can be reached for several frequency bands.

10.2.3 Interference mitigation options

This section will briefly describe interference mitigation options for LOFAR, followed by some of the choices that have been made. Figures 10.5 and 10.6 schematically show the signal chain for a LOFAR station and for the central data processing. The difference with figure 10.2 is that it also shows interference mitigation options at various locations in the signal processing chain.

The interference mitigation measures mentioned in the second step of the strategy can be applied at several locations in the signal chain. It strongly depends on the system design which is the optimal location. This is determined by various factors, such as number of data bits of the beams, data transport load, number of correlator input bits, linearity of the RFI mitigation methods, computational load, etcetera.

For LOFAR it was decided to apply only fixed or very slowly time-varying spatial nulls at station level. Fast-changing interferer nulling directions would lead to fast changes in the (sidelobe) gains and this would hamper the calibration.

Flagging or excising can best be done on a central level because interference is often localised in space. This means that interference may be present in one or a few stations, but not visible in all interferometer outputs. Flagging or excising at station level would, therefore, often remove too many data. More mitigation options will be described in the next section.

Interference mitigation options at the stations

The main purpose of a LOFAR station is to digitise the dual polarisation antenna outputs, to form subbands and to sum the antenna signals in order to create (multiple) beams. In a station, wide-band burst-like interfering signals can be detected and excised directly after the ADC. Examples of such signals are lightning and radio emissions from sparking high power voltage lines. The excised samples have to be replaced with zeros or with some value determined from the signal statistics. Another requirement is that the fraction of blanked samples is recorded, in order to account for the calibration of the sensitivity loss due to excising.

Another method, besides blanking, is to apply spatial filtering in the beamforming stage. In a separate calibration process at the station, the direction or spatial signature of the interferer is measured. Spatial nulls toward the interferer are subsequently applied by adjusting the beamformer weights. Beamforming in the LOFAR stations is done with order 200 kHz beams, a choice which was made for reasons of cost and computational load. This bandwidth is relatively large, which means that spatial nulls will be optimal for one frequency only. Multiple nulls and spectrally and spatially wide nulls are possible, but these reduce the degree of freedom. Another issue is whether the spatial filter should be fixed or adaptive. Optimising the spatial filtering capabilities is an important topic which requires further study.

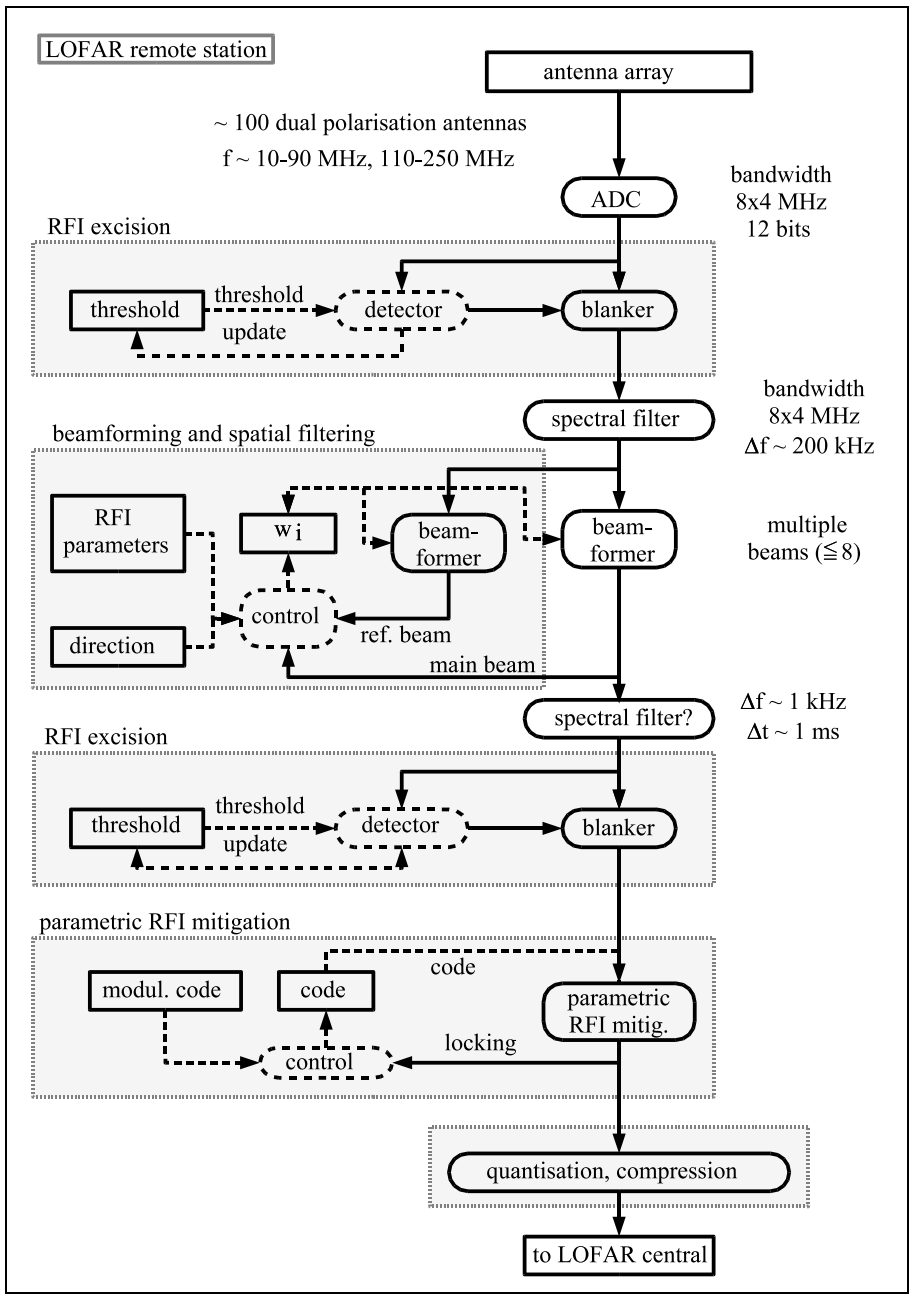


Figure 10.5. Interference mitigation options in LOFAR at station level.

After the second spectral filter, interference excision can be done for each of the 1 kHz channels. An advantage of doing this excision at the station level, and not centrally, is that the data rate will be reduced. The reason is that strong interfering signals need many bits to represent both the interference and the weak cosmic signal. A White Gaussian Noise (WGN) cosmic signal can be described accurately by one to three bits [146]. For several reasons, however, it was decided to shift the 1 kHz spectral filter from the station to the central processing area.

Parametric methods can be applied as well. In the figure they are located after the beamformer stage, but they could be located before the beamformer as well.

Finally, the output data stream of a station can be compressed. However, most astronomical signals are white Gaussian noise signals whose waveform must be preserved up to the correlation process. Data compression therefore only works if interference can be removed or if only part of the data bit range is used.

Interference mitigation options in LOFAR central processing

The signals from the stations are correlated and integrated at the central processing site, as is shown in figure 10.6. Before correlation, the signals can be redistributed. This opens up the possibility to add a station reference beam, directed at an interferer, to the correlation process. This reference beam will have a high INR ratio, and can be used for improved post-correlation spatial filtering purposes. The aim is to select bands free from transmitters or interference, and to reduce occurring interference to levels below the Cas.A flux level. Many post-correlation spatial filtering and interference subtraction techniques can be carried out. Further studies will have to determine the most suitable ones. Spatial filtering in LOFAR central can also be done before the correlation process, but it is computationally and financially more efficient to do it after correlation.

Another type of mitigation is excision, which can be done conveniently after correlation on 1 to 100 ms data. This is an important option, as many transmitters and interferers in the LOFAR bands are intermittent.

Finally, interference mitigation techniques and mitigation distortion corrections can be applied parallel to the map-making process. The goal here is to suppress the interference to levels below the noise in the map.

Design choices

Initially only relatively simple interference mitigation measures will be applied to LOFAR. In a later phase, more complicated scenarios will be added. The LOFAR design is such that new methods can, in principle, be added without major hardware or software redesigns. The main reasons for a limited complexity in the initial phases of LOFAR will now be described.

A first obvious constraint on complexity is that the computing power required for interference mitigation should be an order of magnitude less than

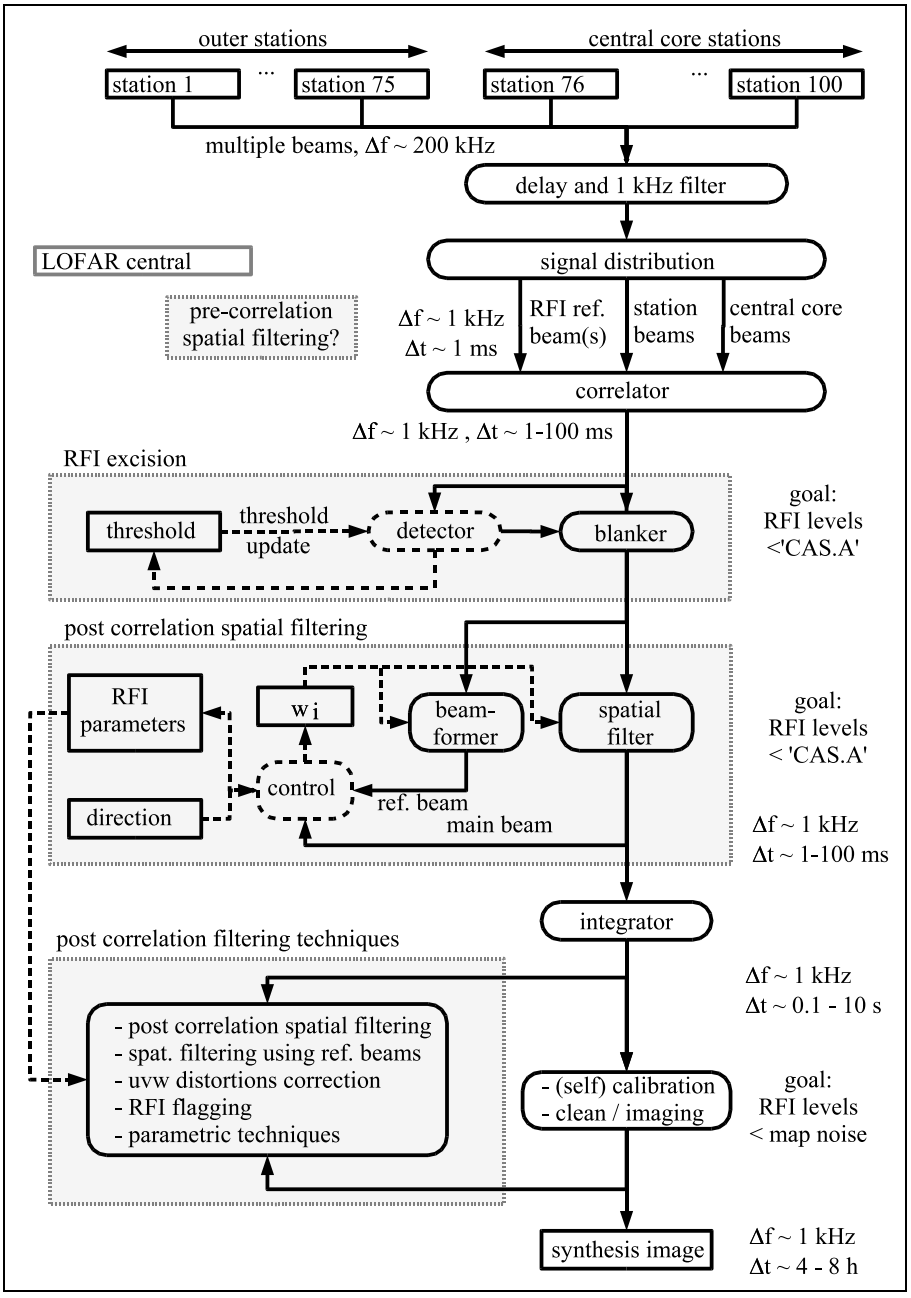


Figure 10.6. Interference mitigation options in LOFAR central processing.

that required for the astronomical signal processing. Only in special cases is spending a major portion of the computing resources on RFI mitigation acceptable. A second reason for relatively simple techniques is that the calibration of LOFAR [109,110] requires stable station beams. Only slowly varying (sidelobe) gains are allowed, otherwise the calibration process will not converge. For this reason, only spatial filters with fixed or slowly varying nulls are considered at the station level, as fast interference tracking would change the station beams too rapidly. A third reason is that for radio astronomy, interference mitigation is a relatively new field, and that the effects of interference mitigation related distortions cannot (yet) be fully removed from the data sets. Finally, even without interference mitigation it is a challenge to build LOFAR, to calibrate it, and to make sky maps with very high sensitivities.

10.3 LOFAR interference mitigation: initial results

The purpose of this section is to demonstrate the spatial filtering capabilities of a LOFAR station. The aim is also to verify some of the claims in the LOFAR interference mitigation strategy. It is specifically shown that spatial filtering at the station level recovers the strongest astronomical sky sources. The spatial filter tests were carried out with ITS data.

10.3.1 Beamforming and imaging with ITS

ITS consists of 60 sky noise limited single polarisation dipoles, and is configured in a five-armed spiral. This configuration is shown in figure 10.7, together with its snapshot (u, v) configuration. The antenna outputs are connected to low noise amplifiers, filtered by a 10-35 MHz filter, and digitised with a sampling frequency of 80 MHz. ITS operates in the frequency band 10-40 MHz, and the observed signals are directly digitised without the use of mixers. The data can be stored either as time series or as covariance matrices. The effect of spatial filters was demonstrated by applying them to antenna covariance matrices obtained with ITS. For the experiments, a 8192 sample length, Hanning tapered, FFT was used. This yields a frequency resolution of 9.77 kHz. The spectra were correlated and integrated to 6.7 seconds, yielding a 60×60 covariance matrix for each of the 4096 frequency bins. This station covariance output mode is not a standard LOFAR mode, but is used for testing and calibration purposes. With ITS, all-sky images were made. The same beamforming data processing was used as described in section 10.2.2. The procedure starts with stacking the antenna outputs in a vector $\mathbf{x}$. Its covariance matrix $\mathbf{R} = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}$ is composed of an astronomical sources matrix $\mathbf{R}_v$, and a diagonal noise matrix $\mathbf{D}$. After a whitening step applied to $\mathbf{R}$, a sky image was made by means of beamforming. Each pixel of the sky map was obtained by applying a direction weight vector $\mathbf{w}$ which represents the geometric phase rotations. Define the array output signal

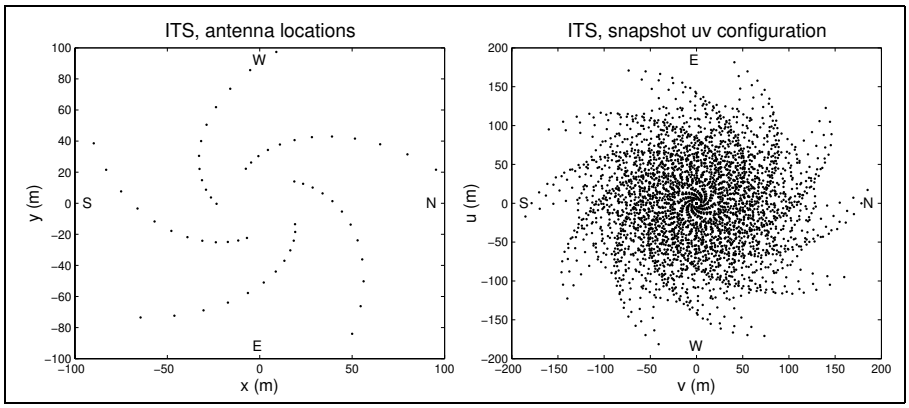


Figure 10.7. LOFAR ITS station antenna configuration (left) and station snapshot uv coverage (right).

y by $y = \mathbf{w}^H\mathbf{x}$, and its power $P(\theta, \phi)$ by $P(\theta, \phi) = \mathcal{E}\{y^2\}$, then

$$P(\theta, \phi) = \mathbf{w}^H\mathbf{R}\mathbf{w} = \mathbf{w}^H\mathbf{R}_v\mathbf{w} + \sigma_n^2 \tag{10.24}$$

where (θ, ϕ) are the zenith and azimuth spherical coordinate angles. Calculating $P(\theta, \phi)$ for each of the beam directions over the entire sky yields an all-sky map. The results are plotted in the direction cosine coordinates (l, m). This imaging process closely resembles the two-dimensional Fourier transform traditionally used in synthesis imaging. One of the differences is that a regular telescope positioning grid is not required in the beamforming process.

10.3.2 Spatial filtering at station level

On February 26^{th} 2004, a 6.7 s duration night time observation was carried out with ITS in the frequency range $10 - 40$ MHz. The spectra showed that at that time a major portion of the band did not contain transmitters and interference at levels above the sky noise. Figure 10.8 shows the eigenvalue distribution of a part of the observed band with one of the strongest transmitters.

The eigenvalues were calculated after whitening the covariance matrix $\mathbf{R}$. The dominant eigenvalue of the interferer data is 20 dB above the noise. The second eigenvalue lies $\approx 1\%$ above the mean noise level. This can be fully explained by the finite sample effect, assuming that the interferer is a narrowband point source. The estimation of the interferer direction is distorted by noise, which leads to an underestimation of its power. The remaining energy then leaks to the other eigenvalues, especially to the second as they are sorted according to magnitude. The increase of the second eigenvalue is of the order of the rms differences between the eigenvalues of the no-interference case. This assumes that the interferer is a strong source with $\lambda_1 \gg \lambda_2$.

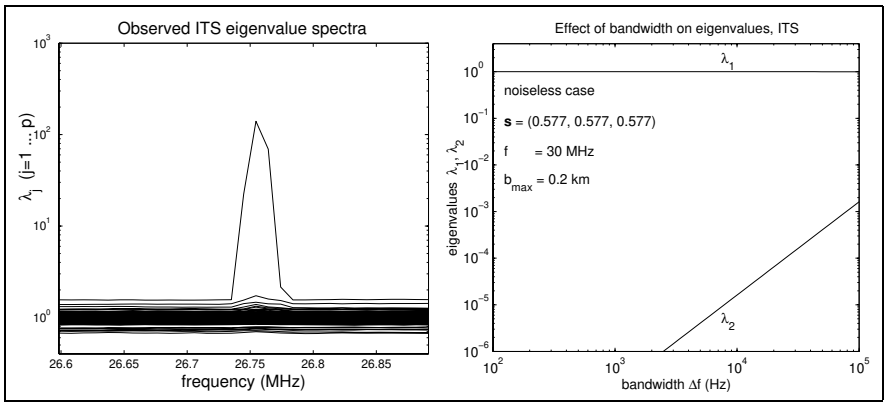


Figure 10.8. Eigenvalue distribution, for N_{sam} = 131000, Δ t = 6.7 s, Δf = 9.77 kHz (left), and a simulation showing the effect of bandwidth on the largest eigenvalues for the noiseless case (right).

The spatial filter which will be applied is tuned at the frequency of the interference, at 26.75 MHz. In actual LOFAR stations, however, beams will be made with order 200 kHz bandwidth. Figure 10.8 therefore shows a theoretical calculation of the two largest eigenvalues for a transmitter with varying bandwidth, see for instance section 3.3.1. This figure indicates that strong interference can be suppressed up to at least 30 dB using a rank one model. Figure 10.9 shows

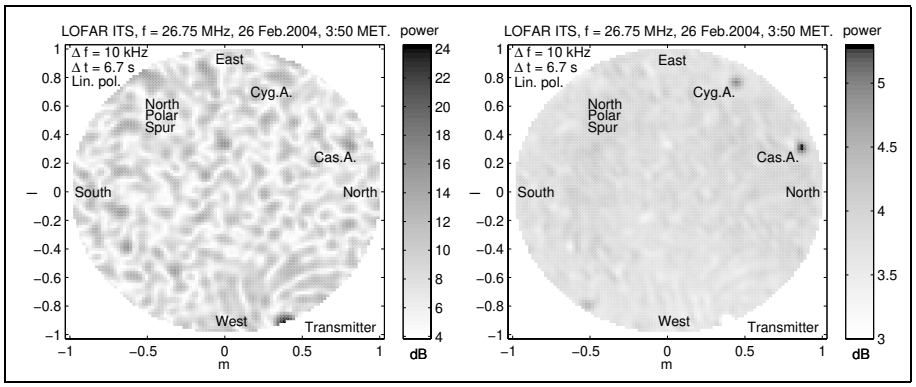


Figure 10.9. LOFAR all-sky image with strong interference close to the east horizon(left), and the same image after spatial projection filtering (right).

an ITS all-sky image, which is obscured by a transmitter on the horizon in the east-south-east direction. The sources Cas.A, Vir.A, and the North Polar Spur are not visible due to the transmitter spatial sidelobes. The data set was

filtered by the standard procedure described in chapter 6. The covariance matrix $\mathbf{R}$ is whitened by data at nearby frequencies, which are not affected by interference. The next step was to apply an eigenvalue decomposition on the whitened data $\mathbf{R}_w$. An interferer projection matrix $\mathbf{P}_\perp$ was formed using the eigenvector belonging to the largest eigenvector. Filtered sky images were made by beamforming, using the filtered covariance matrices:

$$P(\theta, \phi) = \mathbf{w}^H \mathbf{P}_\perp \mathbf{R}_w \mathbf{P}_\perp \mathbf{w} \tag{10.25}$$

No distortion correction was applied at this stage. The resulting sky image is shown in figure 10.9. The sources Cas.A, Vir.A, and the north galactic spur are visible again, and the interferer on the horizon is suppressed below the sky noise. Its spatial sidelobes are suppressed to levels clearly below Cas.A, and Vir.A, and slightly above the station sky noise. Further analysis [170] shows that this interference can be suppressed to the sky noise background by additional calibration procedures. What this shows is that moderately intense interferers can be suppressed to levels below Cas.A and even down to station noise levels. This result supports the validity of the second step in the interference mitigation strategy. However, this result is based on one example only, and more study is needed.

10.4 Conclusion

In this chapter a LOFAR interference mitigation strategy was presented. The strategy relates interferer powers to the system noise levels of the antenna, the stations and the sky maps. Interference mitigation options were discussed and some design considerations in relation to mitigation were presented. It was shown that moderately intense interferers can be suppressed to levels close to the station noise. On the basis of a theoretical analysis it is expected that this interference can be further reduced to levels below the noise in the sky maps. The analysis presented assumes that several requirements concerning the interferer and the system are met. Many of the tools needed to verify these requirements are described and used in this thesis, and are easily adapted for use in LOFAR or SKA.

Further study is needed, for example to investigate the influence of multipaths, the bandwidth, and the extendedness of the interferers on the effectiveness of mitigation. Signal distortions due to mitigation, and interference residuals after mitigation also need further study. Another issue deserving further study is the suppression of intermodulation products. As LOFAR is going to operate in an environment with very strong transmitters, some intermodulation products will probably be visible in parts of the band. A simplified version of the intermodulation product model described in section 2.5.2 was used to predict its spatial behaviour [17]. This study shows that the model describes the phenomenon accurately, and that even intermodulation products can be spatially suppressed.

Chapter 11

Conclusion

11.1 Main results

In this thesis, modern array processing techniques have been applied to narrowband multichannel interference detection and excision, and for narrowband spatial interference mitigation filters in radio astronomy. By exploiting the structure in the array output covariance matrices, new results were found, such as upper limits on interference residuals after excision and spatial filtering. This result was derived by means of a theoretical analysis, and effects which limit the detection and estimation accuracies have been investigated. Using a subspace approach, the effect of bandwidth, extendedness of the impinging sources, and multipath effects on the detection and estimation effectiveness were found. In addition, the effect of the finite sample effect[1] has been quantified. For the WSRT these effects were presented in term of eigenvalue ratios. Interference attenuation numbers of 20 to 40 dB have been found at the WSRT and at other radio telescopes. In the literature, reports of numbers exceeding 40 dB are very rare. In view of the effects described above, this is to be expected.

The advantage of a multichannel approach in detection, excision and spatial filtering was demonstrated by using WSRT data; the attenuation numbers observed match with the theoretical ones. For the spatial projection filter, a distortion correction was applied. The telescope-transmitter stationarity requirements for this correction filter were expressed in terms of matrix condition numbers. For four different types of transmitters and modulation types, these condition numbers were determined experimentally. It was found that for the WSRT , approximately one minute integration time is needed to ensure acceptable signal distortion levels.

By applying reference antennas to a telescope array, spatial filtering effectiveness can be improved. Adding a reference antenna has the additional advantage of not reducing the degrees of freedom. This probably leads to less signal dis-

[1] Finite sample effect: estimation error due to a limited number of samples.

tortions, but it is required that the reference antenna noise level is lower than the telescope noise levels; otherwise the system noise increases unacceptably. Phased-array tiles are highly directional, such as ASTRON's THEA tile. Using this tile as a reference at the WSRT it was shown that the effectiveness of the spatial filter improves indeed. It was also shown that it is advantageous to use as much a-priori information of the covariance matrix structure as possible; it leads, for example, to a more stable filter.

For a single polarisation array with unknown complex gains which observes one dominant source, it was shown that weighted least square gain estimators are asymptotically efficient. Closed-form estimators were derived as well, and under certain conditions these are also asymptotically efficient. The Cramer-Rao Lower Bound on the parameter accuracies, a technique previously unknown in radio astronomy, was established to estimate the gain phase and gain magnitude estimation errors. It was shown that the phase and magnitude estimates are independent.

A factor analysis approach was applied to a dual polarised array model, in order to estimate the model components. A closed-form solution was found for the least squares minimisation of the model error; it turned out that at least three sky sources with different polarisation states are needed to find the telescope complex gain factors. This three-source requirement follows directly from the model solutions, a result which is difficult to achieve otherwise.

Finally, an interference mitigation strategy was given for the LOFAR telescope. The interference power levels observed in spectrum monitoring data were linked to the LOFAR sensitivity levels. It was shown that under certain conditions, interference mitigation techniques and spatial dilution effects will reduce the interference to levels below the integrated noise levels. In two-dimensional aperture arrays such as LOFAR, spatial dilution is the two-dimensional equivalent of fringe rotation dilution in a one-dimensional array and in interferometer-based analysis.

11.2 Suggestions for further research

So far, most research in interference mitigation was done on pre-correlation or post-correlation data-sets of seconds to minutes. Only very limited research was dedicated to the effects of interference and interference mitigation residues on the astronomical end products [96, 67]. Obviously, as the astronomical end product is what matters in the end, this is an essential field; devoting new research efforts to it would be well-justified.

In radio telescopes, the approaches often used for calibration and image making are selfcal [66] and CLEAN [130]. An essential step in map making is estimating the astronomical (point) source direction and then subtracting the source from the map. This may also be done in the Fourier transformation domain, the uvw space. Estimation errors lead to residues in the map. The question is whether the theoretical noise levels in the map can be reached, given

a certain power flux versus the source numbers distribution of astronomical sources. A related question is whether these low noise levels can be reached in case there is interference which may or may not be attenuated by mitigation methods. These questions are currently particularly relevant for telescopes such as LOFAR. A first step in finding answers could be defining a representative model and deriving Cramer-Rao bounds for the parameter estimators.

There are several ways to improve the proposed multichannel detectors and spatial filters. As the focus in this thesis is on processing single narrowbands, an extension of the detectors and filters would be combining adjacent frequency channels. In case an interferer is present in all of those frequency channels, this approach will improve the effectiveness of mitigation.

Beam-former stages in large-scale phased array telescopes such as LOFAR may be implemented as wide-band beam-formers. The bands are narrow enough to ensure acceptable beam squint[2], but are too wide to fulfil the narrow-band requirement. Filters which are designed under the narrow-band assumption will therefore be less effective. Multiple narrow-band filters can be implemented as wide-band filters and vice versa. These are issues which are closely connected to implementation in telescope hardware. Studies on wide-band interference mitigation are therefore highly relevant and require further attention.

The data models derived in this thesis can be expanded and effectively used for studies of the effects of instrumental errors on the astronomical end products. For example, intermodulation modelling [17] yields practical information of the effects of intermodulation products on the sky maps.

The filter residues of the spatial subtraction filter also require further study. For the spatial subtraction filter more knowledge of the involved signals is needed than for the projection filter, but it is not yet clear which of the two filters leads to the lowest residues. This could also be studied in the context of image making, in which Clean is the subtraction filter equivalent.

Given the spectrum trends towards a denser use of digital wide-band communications [139], it is worthwhile to direct more resources to interference mitigation using a-priori knowledge [43, 162].

This thesis has shown that the mitigation effectiveness is determined by parameter estimation and detection accuracies. These accuracies are influenced by telescope properties such as spatial telescope distribution, system noise level, integration time, and channel bandwidth. Transmitter and interference properties are important too, such as emitted power and distances to the different telescopes of an interferometer array. It has also been shown in this thesis that applying fixed spatial filters before and after correlation is mathematically equivalent. This assumes that the signal parameters mentioned above are the same before and after correlation. In a more general sense the question now arises where in the signal chain a specific mitigation method should be applied in order to be most effective. To answer this question a comparative study is needed of the effects of specific telescope and transmitter conditions on filtering

[2]Beam squint: different sections in the frequency bands have different phased array beam shapes and beam directions.

effectiveness. For example, spatial filtering at a phased-array telescope station is to be preferred to post-correlation filtering if the interferer is present at only one or a few stations. In case the channel bandwidth of a phased-array telescope station is much larger than the post-correlation channel bandwidth and the interferer bandwidth, then post-correlation spatial filtering is advantageous.

A closely related research topic is stacking mitigation methods, that is, simultaneously using different mitigation techniques. It would be interesting to find out under which conditions it is useful to stack these techniques, and which combined attenuation levels can be reached.

11.3 Conclusion

This thesis presents a mathematical framework for interference mitigation in the digital signal processing domain. Such a framework yields an efficient way to represent the filtering problem, to solve it, and to estimate effectiveness. The framework was originally introduced in Radio Astronomy by Leshem and van der Veen [93], and by Hamaker et al [63]. In this thesis, the mathematical models are extended, applied to experimental data, and their effectiveness is analysed. The rigour of this mathematical approach makes it a promising tool for further interference mitigation studies.

Appendix A

Abbreviations

ASTRON	Netherlands Foundation for Research in Astronomy, Dwingeloo, The Netherlands
ATA	Allen Telescope Array
CDMA	Code Division Multiple Access
CRB	Cramer-Rao lower bound
DAB	Digital Audio Broadcast
DOA	Direction Of Arrival
DSP	Digital Signal Processor
DVB	Digital Video Broadcast
EMC	Electro Magnetic Compatibility
FIM	Fisher Information Matrix
FM	Frequency modulation
FOV	Field of view
GLONASS	Global Positioning System, Russian Federation
GPS	US Global Positioning System
GSM	Global System for Mobile communications
HTSC	High Temperature Super Conducting
i.i.d.	Independent and identically distributed
INR	Interference to noise ratio
ITS	LOFAR Initial Test Station
IRIDIUM	Low earth orbit based global communications system
LNA	Low Noise Amplifier
LOFAR	LOw Frequency ARray
LS	Least squares
ML	Maximum Likelihood
NOEMI	Nulling Obstructing Electro Magnetic Interferers
Pdf	Probability density function
RAST	Radio astronomy service
RF	Radio Frequency
RFI	Radio Frequency Interference

SINR	Signal plus interference to noise ratio
SKA	Square Kilometer Array
SNR	Signal to noise ratio
SNIR	Signal to noise plus interference ratio
SVD	Singular Value Decomposition
TDMA	Time Division Multiple Access
THEA	THousand Element Array
TUDelft	Delft University of Technology, Delft, The Netherlands
UWB	Ultra Wide Band
VLA	Very Large Array
VLBI	Very Long Baseline Interferometry
WGN	White Gaussian Noise
WSRT	The Westerbork Synthesis Radio Telescope

Appendix B

Mathematical relations

B.1 Vectors and matrices: products and operators

This section lists properties of matrix and vector products which are used throughout this thesis. For complex matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{E}$ and vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ of compatible dimensions, and for diagonal matrices $\mathbf{D} = \mathrm{diag}(\mathbf{d})$, $\mathbf{D}_1 = \mathrm{diag}(\mathbf{d}_1)$, $\mathbf{D}_2 = \mathrm{diag}(\mathbf{d}_2)$, the following relations hold.

Kronecker products

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{E}) = \mathbf{AC} \otimes \mathbf{BE} \tag{B.1}$$

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \tag{B.2}$$

$$(\mathbf{A} \otimes \mathbf{B})^{H} = \mathbf{A}^{H} \otimes \mathbf{B}^{H} \tag{B.3}$$

$$\ln(\mathbf{a} \otimes \mathbf{b}^{H}) = \ln(\mathbf{a}) \otimes \mathbf{1} + \mathbf{1} \otimes \ln(\mathbf{b}) \tag{B.4}$$

$$\frac{\partial(\mathbf{a} \otimes \mathbf{b})}{\partial \mathbf{c}^{t}} = \frac{\partial \mathbf{a}}{\partial \mathbf{c}^{t}} \otimes \mathbf{b} + \mathbf{a} \otimes \frac{\partial \mathbf{b}}{\partial \mathbf{c}^{t}} \tag{B.5}$$

$$\mathrm{vecdiag}(\overline{\mathbf{A}} \otimes \mathbf{A}) = \mathrm{vec}(\mathrm{vecdiag}(\mathbf{A})(\mathrm{vecdiag}(\mathbf{A})^{H})) \tag{B.6}$$

Hadamard product relations

$$\mathbf{a}\mathbf{b}^{t} \odot \mathbf{c}\mathbf{d}^{t} = \mathbf{a}\mathbf{d}^{t} \odot \mathbf{c}\mathbf{b}^{t} \tag{B.7}$$

$$(\mathbf{AD}) \odot \mathbf{C} = \mathbf{A} \odot (\mathbf{CD}) \tag{B.8}$$

$$\mathbf{A} \odot (\mathbf{CD}) = (\mathbf{A} \odot \mathbf{C})\mathbf{D} \tag{B.9}$$

$$(\mathbf{A} \odot \mathbf{B})^{H} = \mathbf{A}^{H} \odot \mathbf{B}^{H} \tag{B.10}$$

$$(\mathbf{A} \ominus \mathbf{B})^{H} = \mathbf{A}^{H} \ominus \mathbf{B}^{H} \tag{B.11}$$

$$\mathbf{a}^{H}(\mathbf{b} \odot \mathbf{c}) = \mathbf{b}^{t}(\overline{\mathbf{a}} \odot \mathbf{c}) \tag{B.12}$$

$$(\mathbf{a} \odot e^{\jmath\mathbf{b}})^{H}(\mathbf{a} \odot e^{\jmath\mathbf{b}}) = \mathbf{a}^{H}\mathbf{a} \tag{B.13}$$

$$(\mathbf{a} \odot e^{\jmath\mathbf{b}})^{H}(\mathbf{c} \odot e^{\jmath\mathbf{d}}) = (\overline{\mathbf{a}} \odot \mathbf{c})^{t} e^{\jmath(\mathbf{d}-\mathbf{b})} \tag{B.14}$$

Khatri-Rao product relations

$$
\begin{aligned}
(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\circ\mathbf{E}) &= \mathbf{AC}\circ\mathbf{BE} && \text{(B.15)}\\
(\mathbf{A}\circ\mathbf{B})^H(\mathbf{C}\circ\mathbf{E}) &= \mathbf{A}^H\mathbf{C}\odot\mathbf{B}^H\mathbf{E} && \text{(B.16)}\\
(\mathbf{D}_1\mathbf{D}_2)\circ\mathbf{A} &= (\mathbf{D}_2\circ\mathbf{A})\mathbf{D}_1 && \text{(B.17)}\\
(\mathbf{A}\circ\mathbf{B})^H &= \mathbf{A}^H\circ\mathbf{B}^H && \text{(B.18)}\\
(\mathbf{I}_p\circ\mathbf{I}_p)^H(\mathbf{I}_p\circ\mathbf{I}_p) &= \mathbf{I}_p && \text{(B.19)}\\
& && \text{(B.20)}
\end{aligned}
$$

Vectorisation

$$
\begin{aligned}
\mathrm{vec}(\mathbf{ABC}) &= (\mathbf{C}^t\otimes\mathbf{A})\mathrm{vec}(\mathbf{B}) && \text{(B.21)}\\
\mathrm{vec}(\mathbf{ADC}) &= (\mathbf{C}^t\circ\mathbf{A})\mathbf{d} && \text{(B.22)}
\end{aligned}
$$

Trace operations

$$
\begin{aligned}
\mathrm{tr}\,(\mathbf{AB}) &= \mathrm{tr}\,(\mathbf{BA}) && \text{(B.23)}\\
\mathrm{tr}(\mathbf{AB}) &= \mathrm{vec}(\mathbf{A}^t)^t\mathrm{vec}(\mathbf{B}) && \text{(B.24)}\\
\mathbf{a}^H\mathbf{Bc} &= \mathrm{tr}\,(\mathbf{a}^H\mathbf{Bc}) = \mathrm{tr}\,(\mathbf{B}\,\mathbf{ca}^H) && \text{(B.25)}\\
\mathrm{tr}\,(\mathbf{b}^H\mathbf{a}) &= \mathbf{b}^H\mathbf{a} = \mathrm{tr}\,(\mathbf{ab}^H) && \text{(B.26)}\\
\|\mathbf{A}\|_F^2 &= (\mathrm{vec}(\mathbf{A})^H\mathrm{vec}(\mathbf{A}))\\
&= \mathrm{tr}\,(\mathrm{vec}(\mathbf{A})\mathrm{vec}(\mathbf{A})^H) && \text{(B.27)}\\
\mathrm{tr}\,(\mathbf{A}+\mathbf{B}) &= \mathrm{tr}\,(\mathbf{A}) + \mathrm{tr}\,(\mathbf{B}) && \text{(B.28)}\\
\mathrm{tr}\,(\mathbf{x}^H\mathbf{Ax}) &= \mathrm{tr}\,(\mathbf{Axx}^H) && \text{(B.29)}
\end{aligned}
$$

Selection matrices

Define $\mathbf{J}:(p^2-p)\times p^2$ such that for any $p\times p$ matrix $\mathbf{A}$, $\mathrm{vec}'(\mathbf{A})=\mathbf{J}\mathrm{vec}(\mathbf{A})$. Note that $\mathbf{J}$ is easily constructed from a $p^2\times p^2$ identity matrix by removing its rows $1,(p+1)+1,2(p+1)+1,\cdots,p^2$. Then

$$
\begin{aligned}
\mathbf{J} &\equiv [\mathbf{I}_{p-1}\otimes[\mathbf{0}_p,\mathbf{I}_p],\mathbf{0}_{p^2-p}] && \text{(B.30)}\\
\mathbf{J}^H\mathbf{J} &= \mathbf{I}\otimes\mathbf{I}-(\mathbf{I}\circ\mathbf{I})(\mathbf{I}\circ\mathbf{I})^H && \text{(B.31)}\\
\mathbf{J}\mathbf{J}^H &= \mathbf{I}_p && \text{(B.32)}\\
(\mathbf{J}^H\mathbf{J})^2 &= \mathbf{J}^H\mathbf{J} \quad (\mathbf{J}^H\mathbf{J}\text{ is idempotent}) && \text{(B.33)}
\end{aligned}
$$

Miscellaneous product relations

$$\begin{aligned}
\mathbf{D}_1\mathbf{A}\mathbf{D}_2^H &= \mathbf{A}\odot(\mathbf{d}_1\mathbf{d}_2^H) && \text{(B.34)}\\
\mathbf{D}_1^{-1}\mathbf{A}\mathbf{D}_2^{-H} &= \mathbf{A}\ominus(\mathbf{d}_1\mathbf{d}_2^H) && \text{(B.35)}\\
(\mathbf{A}_1\mathbf{B}\cdots\mathbf{A}_p\mathbf{B}) &= (\mathbf{A}_1\cdots\mathbf{A}_p)(\mathbf{I}_p\otimes\mathbf{B}) && \text{(B.36)}\\
\mathbf{a}\mathbf{b}^H &= \overline{\mathbf{b}}\mathbf{a}^t && \text{(B.37)}\\
\sum_{k=1}^{K}\mathbf{a}_k b_k &= [\mathbf{a}_1\cdots\mathbf{a}_K]\begin{bmatrix} b_1\\ \vdots\\ b_K\end{bmatrix}=\mathbf{A}\mathbf{b} && \text{(B.38)}\\
(e^{\jmath\mathbf{w}})^H e^{\jmath\mathbf{a}} &= \mathbf{1}^t e^{\jmath(\mathbf{a}-\mathbf{w})} && \text{(B.39)}\\
\mathrm{diag}(\mathbf{D}_1\mathbf{A}\mathbf{D}_2^H) &= \mathbf{D}_1\mathbf{D}_2^H\mathrm{diag}(\mathbf{A}) && \text{(B.40)}
\end{aligned}$$

B.2 Multivariate complex PDF

The multivariate joint probability density function for a normally distributed real or complex valued vector $\mathbf{x}$ can be found by using characteristic functions [104,167]. Let $\mathbf{X}$ be defined as a sequence of N time sampled instances of $\mathbf{x}$: $\mathbf{X} = (\mathbf{x}_1\cdots\mathbf{x}_N)$, and let $\boldsymbol{\Theta}$ be the model parameters. The real valued $\mathcal{N}(\mu,\mathbf{R}_{\boldsymbol{\Theta}})$ and complex valued $\mathcal{CN}(\mu,\mathbf{R}_{\boldsymbol{\Theta}})$ multivariate distributed joint probability density function $p_{\mathcal{N}}(\mathbf{X}|\boldsymbol{\Theta})$ and $p_{\mathcal{CN}}(\mathbf{X}|\boldsymbol{\Theta})$ are given by [81]:

$$\begin{aligned}
p_{\mathcal{N}}(\mathbf{X}|\boldsymbol{\Theta}) &= \prod_{n=1}^{N}\frac{1}{(2\pi)^{\frac{1}{2}p}|\mathbf{R}_\Theta|^{\frac{1}{2}}}e^{-\frac{1}{2}(\mathbf{x}_n-\boldsymbol{\mu})^H\mathbf{R}_\Theta^{-1}(\mathbf{x}_n-\boldsymbol{\mu})} && \text{(B.41)}\\
&= \frac{1}{(2\pi)^{\frac{1}{2}Np}|\mathbf{R}_\Theta|^{\frac{1}{2}N}}e^{-\frac{1}{2}N\ \mathrm{tr}(\mathbf{R}_\Theta^{-1}\widehat{\mathbf{R}})} && \text{(B.42)}\\
p_{\mathcal{CN}}(\mathbf{X}|\boldsymbol{\Theta}) &= \prod_{n=1}^{N}\frac{1}{\pi^p|\mathbf{R}_\Theta|}e^{-(\mathbf{x}_n-\boldsymbol{\mu})^H\mathbf{R}_\Theta^{-1}(\mathbf{x}_n-\boldsymbol{\mu})} && \text{(B.43)}\\
&= \frac{1}{\pi^{Np}|\mathbf{R}_\Theta|^N}e^{-N\ \mathrm{tr}(\mathbf{R}_\Theta^{-1}\widehat{\mathbf{R}})} && \text{(B.44)}
\end{aligned}$$

These normally distributed probability density functions are completely determined by the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\mathbf{R}_{\boldsymbol{\Theta}}$, and all higher order moments can be expressed in term of $\boldsymbol{\mu}$ and $\mathbf{R}_{\boldsymbol{\Theta}}$. The fourth order moments for the $\mathcal{N}(\boldsymbol{\mu},\mathbf{R}_{\boldsymbol{\Theta}})$ distribution, for example, can be expressed [104, 167] as

$$\begin{aligned}
\mathcal{E}\{x_1x_2x_3x_4\} &= \mathcal{E}\{x_1x_2\}\mathcal{E}\{x_3x_4\}+\mathcal{E}\{x_1x_3\}\mathcal{E}\{x_2x_4\}+\\
&\quad\ \mathcal{E}\{x_1x_4\}\mathcal{E}\{x_2x_3\} && \text{(B.45)}
\end{aligned}$$

and the fourth order moments for the $\mathcal{CN}(\boldsymbol{\mu},\mathbf{R}_{\boldsymbol{\Theta}})$ distribution resembles the real case. For zero-mean, circular symmetry complex random variables however,

the third term $\mathcal{E}\{\bar{x}_1\bar{x}_3\}\mathcal{E}\{x_2x_4\}$ is zero [167, 81, 104]. The resulting expression is:

$$\mathcal{E}\{\bar{x}_1x_2\bar{x}_3x_4\} = \mathcal{E}\{\bar{x}_1x_2\}\mathcal{E}\{\bar{x}_3x_4\} + \mathcal{E}\{\bar{x}_1x_4\}\mathcal{E}\{x_2\bar{x}_3\} \tag{B.46}$$

B.3 Matrix inversion

This section gives several expressions for matrix inverses. In these expressions it is assumed that the matrices have compatible dimensions and that they are invertible. The first expression is the inverse of the matrix product **AB** which can be written as

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{B.47}$$

Another useful formula, the matrix inversion lemma, can be expressed in several forms:

$$\begin{aligned}(\mathbf{A}+\mathbf{BCD})^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1}\right)^{-1}\mathbf{DA}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{BC}\left(\mathbf{I}+\mathbf{DA}^{-1}\mathbf{BC}\right)^{-1}\mathbf{DA}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{I}+\mathbf{CDA}^{-1}\mathbf{B}\right)^{-1}\mathbf{CDA}^{-1}\end{aligned} \tag{B.48}$$

A special case of the matrix inversion lemma is Woodbury's identity:

$$(\mathbf{A}+\mathbf{bb}^H)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{bb}^H\mathbf{A}^{-1}}{1+\mathbf{b}^H\mathbf{A}^{-1}\mathbf{b}} \tag{B.49}$$

Appendix C

Covariance of matrix estimates

For the analysis of the data models, the simulations and the interpretation of the processed observed data, an estimate of the variance of the covariance matrix estimate $\widehat{\mathbf{R}}$ is essential. In this appendix, the formula for the covariance of the covariance matrix estimate is given, together with a proof. Both the non-weighted and weighted covariances are given.

C.1 Matrix covariance

For calculation of the covariance $\text{cov}(\widehat{\mathbf{R}})$ of the covariance matrix estimate $\widehat{\mathbf{R}}$, the starting point is the definition of $\widehat{\mathbf{R}}_n$ in terms of the array output complex voltage vector $\mathbf{x}_n$ at time instance n: $\widehat{\mathbf{R}}_n = \mathbf{x}_n \mathbf{x}_n^H$. The time N-sample average of the covariance estimate is given by:

$$\widehat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n{}^H \tag{C.1}$$

and it is assumed that the array output vector $\mathbf{x}$ is i.i.d. WGN with $\mathcal{E}\{\widehat{\mathbf{R}}\} = \mathcal{E}\{\widehat{\mathbf{R}}_n\} = \mathbf{R}$. The $p^2 \times p^2$ covariance $\text{cov}(\mathbf{R})$ of $\widehat{\mathbf{R}}$ is defined as

$$\text{cov}(\widehat{\mathbf{R}}) \equiv \mathcal{E}\{ [\text{vec}(\widehat{\mathbf{R}}) - \text{vec}(\mathbf{R})] [\text{vec}(\widehat{\mathbf{R}}) - \text{vec}(\mathbf{R})]^H \} \tag{C.2}$$

where

$$\text{vec}(\widehat{\mathbf{R}}) = \frac{1}{N} \sum_{n=1}^{N} \text{vec}(\mathbf{x}_n \mathbf{x}_n^H) = \frac{1}{N} \sum_{n=1}^{N} \bar{\mathbf{x}}_n \otimes \mathbf{x}_n \tag{C.3}$$

Expanding the expression for $\text{cov}(\widehat{\mathbf{R}})$ and inserting $\mathcal{E}\{\text{vec}(\widehat{\mathbf{R}})\} = \text{vec}(\mathbf{R})$ yields:

$$\text{cov}(\widehat{\mathbf{R}}) = \mathcal{E}\{ \text{vec}(\widehat{\mathbf{R}}) \text{vec}(\widehat{\mathbf{R}})^H \} - \text{vec}(\mathbf{R}) \text{vec}(\mathbf{R})^H \tag{C.4}$$

Inserting the expression for $\widehat{\mathbf{R}}$ in the previous equation yields

$$\mathrm{cov}(\widehat{\mathbf{R}}) = \mathcal{E}\{\ \frac{1}{N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}(\bar{\mathbf{x}}_n \otimes \mathbf{x}_n)\ (\bar{\mathbf{x}}_{n'} \otimes \mathbf{x}_{n'})^H\ \} - \mathrm{vec}(\mathbf{R})\mathrm{vec}(\mathbf{R})^H \quad \text{(C.5)}$$

If $n \neq n'$, assuming that the array output sample vectors $\mathbf{x}_n$ at different times are independent, then the $(n,n')^{th}$ element of the first term of the previous equation can be written as

$$\begin{aligned}\mathcal{E}\{\ (\bar{\mathbf{x}}_n \otimes \mathbf{x}_n)\ (\bar{\mathbf{x}}_{n'} \otimes \mathbf{x}_{n'})^H\ \} &= \mathcal{E}\{\ (\bar{\mathbf{x}}_n \otimes \mathbf{x}_n)\}\ \mathcal{E}\{\ (\bar{\mathbf{x}}_{n'} \otimes \mathbf{x}_{n'})^H\ \} \\ &= \mathcal{E}\{\mathrm{vec}(\widehat{\mathbf{R}}_n)\}\ \mathcal{E}\{\mathrm{vec}(\widehat{\mathbf{R}}_{n'})^H\} \quad \text{(C.6)} \\ &= \mathrm{vec}(\mathbf{R})\mathrm{vec}(\mathbf{R})^H \quad \text{(C.7)}\end{aligned}$$

If $n = n'$, then the $(n,n)^{th}$ element of the first term of the previous equation can be written as

$$\begin{aligned}\mathcal{E}\{\ (\bar{\mathbf{x}}_n \otimes \mathbf{x}_n)\ (\bar{\mathbf{x}}_{n'} \otimes \mathbf{x}_{n'})^H\ \} &= \mathcal{E}\{\ \bar{\mathbf{x}}_n\bar{\mathbf{x}}_n^H \otimes \mathbf{x}_n\mathbf{x}_n^H\ \} \quad \text{(C.8)} \\ &= \mathcal{E}\{\ \widehat{\bar{\mathbf{R}}}_n \otimes \widehat{\mathbf{R}}_n\ \} \quad \text{(C.9)}\end{aligned}$$

Inserting these results in the equation for $\mathrm{cov}(\widehat{\mathbf{R}})$ yields

$$\mathrm{cov}(\widehat{\mathbf{R}}) = \frac{1}{N}\mathcal{E}\{\ \widehat{\bar{\mathbf{R}}}_n \otimes \widehat{\mathbf{R}}_n\ \} - \frac{1}{N}\mathrm{vec}(\mathbf{R})\mathrm{vec}(\mathbf{R})^H \quad \text{(C.10)}$$

Note that $\mathcal{E}\{\ \widehat{\bar{\mathbf{R}}}_n \otimes \widehat{\mathbf{R}}_n\ \} = \bar{\mathbf{R}} \otimes \mathbf{R} + \mathrm{vec}(\mathbf{R})\mathrm{vec}(\mathbf{R})^H$, which can be derived in a straightforward manner, by using the following mixed moment relationship between the $\mathcal{CN}(\mathbf{0},\mathbf{R})$ distributed variables x_{1n},x_{2n},x_{3n}, and x_{4n} (cf. section B.2)

$$\begin{aligned}\mathcal{E}\{\bar{x}_{1n}x_{2n}\bar{x}_{3n}x_{4n}\} &= \mathcal{E}\{\bar{x}_{1n}x_{2n}\}\mathcal{E}\{\bar{x}_{3n}x_{4n}\} \\ &\quad + \mathcal{E}\{\bar{x}_{1n}x_{4n}\}\mathcal{E}\{x_{2n}\bar{x}_{3n}\} \quad \text{(C.11)}\end{aligned}$$

where x_{in} is the i^{th} array sensor output at time instance n. Recall that $\mathcal{E}\{x_{in}\overline{x_{jn}}\}$ $= \mathbf{R}_{ijn}$. This leads to the following expression for $\mathrm{cov}(\widehat{\mathbf{R}})$

$$\boxed{\mathrm{cov}(\widehat{\mathbf{R}}) = \tfrac{1}{N}\bar{\mathbf{R}} \otimes \mathbf{R}} \quad \text{(C.12)}$$

In [111], a slightly different approach is chosen to calculate $\mathrm{cov}(\widehat{\mathbf{R}})$, but with the same result.

Define the variance of the covariance matrix estimate as follows:

$$\mathrm{var}(\widehat{\mathbf{R}}) \equiv \mathcal{E}\{(\widehat{\mathbf{R}} - \mathbf{R}) \odot \overline{(\widehat{\mathbf{R}} - \mathbf{R})}\} \quad \text{(C.13)}$$

Using the formula for the covariance of the covariance matrix estimate, this can be written as

$$\mathrm{var}(\widehat{\mathbf{R}}) = \frac{1}{N}\mathrm{unvec}(\mathrm{vecdiag}(\bar{\mathbf{R}} \otimes \mathbf{R})) \quad \text{(C.14)}$$

or as

$$\boxed{\mathrm{var}(\widehat{\mathbf{R}}) = \tfrac{1}{N}\ \mathrm{vecdiag}(\mathbf{R})\ \mathrm{vecdiag}(\mathbf{R})^H} \quad \text{(C.15)}$$

C.2 Weighted matrix covariance

Suppose a covariance matrix $\mathbf{R}$ is weighted with matrix $\mathbf{W}$, giving the weighted matrix $\mathbf{R}_w$ and its estimate $\widehat{\mathbf{R}}_w$:

$$\mathbf{R}_w \equiv \mathbf{W}\mathbf{R}\mathbf{W}^H \tag{C.16}$$

$$\widehat{\mathbf{R}}_w = \mathbf{W}\widehat{\mathbf{R}}\mathbf{W}^H \tag{C.17}$$

The covariance of $\widehat{\mathbf{R}}_w$, $\mathrm{cov}(\widehat{\mathbf{R}}_w) \equiv \mathcal{E}\{[\mathrm{vec}(\widehat{\mathbf{R}}_w) - \mathrm{vec}(\mathbf{R}_w)][\mathrm{vec}(\widehat{\mathbf{R}}_w) - \mathrm{vec}(\mathbf{R}_w)]^H\}$ is found by inserting the expressions for $\mathbf{R}_w$ and $\widehat{\mathbf{R}}_w$ in equation (C.12), and using relation (B.1). The result is:

$$\boxed{\mathrm{cov}(\widehat{\mathbf{R}}_w) = (\overline{\mathbf{W}} \otimes \mathbf{W})\ \mathrm{cov}(\widehat{\mathbf{R}})\ (\overline{\mathbf{W}} \otimes \mathbf{W})^H} \tag{C.18}$$

where N is the number of samples. Following equation (C.15), the variance is given by

$$\boxed{\mathrm{var}(\widehat{\mathbf{R}}_w) = \tfrac{1}{N}\ \mathrm{vecdiag}(\mathbf{W}\mathbf{R}\mathbf{W}^H)\ \mathrm{vecdiag}(\mathbf{W}\mathbf{R}\mathbf{W}^H)^H} \tag{C.19}$$

For example, suppose that the data model is given by $\mathbf{R} = \mathbf{R}_o + \mathbf{D}$, where $\mathbf{R}_o$ is a matrix representing the astronomical and interference signal covariance, and where $\mathbf{D}$ is the diagonal noise matrix. A diagonal weighting matrix $\mathbf{W}_D$ leads to a simple form for the variance. Let $\mathbf{W}_D \equiv \mathbf{D}^{-\frac{1}{2}}$, where $\mathbf{D}$ is assumed to be known from calibration measurements, then the normalised covariance matrix $\mathbf{R}_{W_D}$ is given by:

$$\mathbf{R}_{W_D} = \mathbf{W}_D\mathbf{R}\mathbf{W}_D^H = \mathbf{D}^{-\frac{1}{2}}\mathbf{R}_o\mathbf{D}^{-\frac{1}{2}} + \mathbf{I} \tag{C.20}$$

and, using (B.40), its variance by

$$\begin{aligned} \mathrm{var}(\widehat{\mathbf{R}}_{W_D}) &= \mathbf{D}^{-1}\ \mathrm{var}(\widehat{\mathbf{R}})\ \mathbf{D}^{-1} \\ &= \frac{1}{N}\mathbf{D}^{-1}\ \mathrm{vecdiag}(\mathbf{R})\mathrm{vecdiag}(\mathbf{R})\ \mathbf{D}^{-1} \end{aligned} \tag{C.21}$$

As a second example (cf. sections 2.3.1, 2.6, 10.2.2, and 10.3.1) consider the beamformer output or "dirty beam" I_D: $I_D = \mathbf{w}^H\mathbf{R}\mathbf{w}$, where $\mathbf{w}$ is the beamsteering or weighting vector, and $\mathbf{R}$ is the covariance of the array output signals. As $\mathbf{w}^H\mathbf{R}\mathbf{w}$ is a scalar, $\mathrm{vecdiag}(\mathbf{w}^H\mathbf{R}\mathbf{w}) = \mathbf{w}^H\mathbf{R}\mathbf{w}$, and:

$$\mathrm{var}(\hat{I}_D) = \mathrm{var}(\mathbf{w}^H\widehat{\mathbf{R}}\mathbf{w}) = \frac{1}{N}\ |\mathbf{w}^H\mathbf{R}\mathbf{w}|^2 = \frac{1}{N}I_D^2 \tag{C.22}$$

which means that the variance of the beamformer output power scales with I_D^2, which is a well known fact.

Appendix D

Derivation of FIM components

In this appendix, we elaborate on the form of the Fisher information matrix in (8.4)–(8.5). $\mathbf{F}_\gamma$ can be written as $\mathbf{F}_\gamma = \frac{\partial(\bar{\mathbf{g}} \otimes \mathbf{g})}{\partial \boldsymbol{\gamma}^t} = \frac{\partial \bar{\mathbf{g}}}{\partial \boldsymbol{\gamma}^t} \otimes \mathbf{g} + \bar{\mathbf{g}} \otimes \frac{\partial \mathbf{g}}{\partial \boldsymbol{\gamma}^t}$ where $\frac{\partial \bar{\mathbf{g}}}{\partial \boldsymbol{\gamma}^t} \otimes \mathbf{g} = (\mathbf{I} \circ \mathbf{G})\bar{\boldsymbol{\Phi}}$ and where $\bar{\mathbf{g}} \otimes \frac{\partial \mathbf{g}}{\partial \boldsymbol{\gamma}^t} = (\bar{\mathbf{G}} \circ \mathbf{I})\boldsymbol{\Phi}$. It follows readily that

$$\begin{aligned} \mathbf{F}_\gamma &= (\bar{\mathbf{G}} \circ \mathbf{I})\boldsymbol{\Phi} + (\mathbf{I} \circ \mathbf{G})\bar{\boldsymbol{\Phi}} && \text{(D.1)} \\ \mathbf{F}_\phi &= \imath \left((\bar{\mathbf{G}} \circ \mathbf{I})\boldsymbol{\Gamma}\boldsymbol{\Phi} - (\mathbf{I} \circ \mathbf{G})\boldsymbol{\Gamma}\bar{\boldsymbol{\Phi}} \right) \mathbf{I}_s && \text{(D.2)} \\ \mathbf{F}_d &= \mathbf{I} \circ \mathbf{I} \end{aligned}$$

where $\mathbf{I}_s$ is a selection matrix, equal to the identity matrix with its first column removed, so that the derivative to ϕ_1 is omitted.

The FIM can be partitioned as

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{\gamma\gamma} & \mathbf{M}_{\gamma\phi} & \mathbf{M}_{\gamma d} \\ \mathbf{M}_{\phi\gamma} & \mathbf{M}_{\phi\phi} & \mathbf{M}_{\phi d} \\ \mathbf{M}_{d\gamma} & \mathbf{M}_{d\phi} & \mathbf{M}_{dd} \end{bmatrix} \tag{D.3}$$

We now show that

$$\begin{aligned} \mathbf{M}_{\gamma\phi} &= \mathbf{F}_\gamma^H (\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) \mathbf{F}_\phi = 0 && \text{(D.4)} \\ \mathbf{M}_{d\phi} &= \mathbf{F}_d^H (\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) \mathbf{F}_\phi = 0 && \text{(D.5)} \end{aligned}$$

so that[1]

$$\mathbf{M}^{-1} = \begin{bmatrix} (\mathbf{M}_{\gamma\gamma} - \mathbf{M}_{\gamma d}\mathbf{M}_{dd}^{-1}\mathbf{M}_{d\gamma})^{-1} & 0 & * \\ 0 & \mathbf{M}_{\phi\phi}^{-1} & 0 \\ * & 0 & (\mathbf{M}_{dd} - \mathbf{M}_{d\gamma}\mathbf{M}_{\gamma\gamma}^{-1}\mathbf{M}_{\gamma d})^{-1} \end{bmatrix}$$

[1] $*$ denotes certain sections of the matrix that are not of interest.

which indicates that the phase parameters are decoupled from the gain magnitude and noise parameters.

To show (D.4), we start with equation (D.4): $\mathbf{M}_{\gamma\phi} = \mathbf{F}_{\gamma}^{H}(\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1})\mathbf{F}_{\phi}$, where inserting the Jacobians (D.1), (D.2) produces

$$\begin{aligned}\mathbf{M}_{\gamma\phi} &= \left((\bar{\mathbf{G}} \circ \mathbf{I})\boldsymbol{\Phi} + (\mathbf{I} \circ \mathbf{G})\bar{\boldsymbol{\Phi}}\right)^{H} (\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) \cdot \\ &\quad \cdot \imath \left((\bar{\mathbf{G}} \circ \mathbf{I})\boldsymbol{\Gamma}\boldsymbol{\Phi} - (\mathbf{I} \circ \mathbf{G})\boldsymbol{\Gamma}\bar{\boldsymbol{\Phi}}\right)\mathbf{I}_s\end{aligned}$$

Factoring out this equation produces four terms,

$$\begin{aligned}\mathbf{M}_{\gamma\phi} &= \imath\bar{\boldsymbol{\Phi}}((\bar{\mathbf{G}}^{H}\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}) \odot \mathbf{R}^{-1})\boldsymbol{\Gamma}\boldsymbol{\Phi}\mathbf{I}_s \\ &\quad -\imath\bar{\boldsymbol{\Phi}}((\bar{\mathbf{G}}^{H}\bar{\mathbf{R}}^{-1}) \odot (\mathbf{R}^{-1}\mathbf{G}))\boldsymbol{\Gamma}\bar{\boldsymbol{\Phi}}\mathbf{I}_s \\ &\quad +\imath\boldsymbol{\Phi}((\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}) \odot (\mathbf{G}^{H}\mathbf{R}^{-1}))\boldsymbol{\Gamma}\boldsymbol{\Phi}\mathbf{I}_s \\ &\quad -\imath\boldsymbol{\Phi}(\bar{\mathbf{R}}^{-1} \odot (\mathbf{G}^{H}\mathbf{R}^{-1}\mathbf{G}))\boldsymbol{\Gamma}\bar{\boldsymbol{\Phi}}\mathbf{I}_s\end{aligned}$$

This can be expanded using $\mathbf{G} = \mathbf{g}\mathbf{1}^{t} = \boldsymbol{\Phi}\boldsymbol{\gamma}\mathbf{1}^{t}$, and by defining $\mathbf{R} = \boldsymbol{\Phi}\mathbf{R}_a\bar{\boldsymbol{\Phi}}$ where $\mathbf{R}_a$ is a matrix containing the absolute values of the elements of the matrix $\mathbf{R}$. The result is:

$$\begin{aligned}\mathbf{M}_{\gamma\phi} &= \imath\bar{\boldsymbol{\Phi}}((\mathbf{1}\boldsymbol{\gamma}^{t}\boldsymbol{\Phi}\bar{\boldsymbol{\Phi}}\mathbf{R}_a^{-1}\boldsymbol{\Phi}\bar{\boldsymbol{\Phi}}\boldsymbol{\gamma}\mathbf{1}^{t}) \odot (\boldsymbol{\Phi}\mathbf{R}_a^{-1}\bar{\boldsymbol{\Phi}}))\boldsymbol{\Gamma}\boldsymbol{\Phi}\mathbf{I}_s \\ &\quad -\imath\bar{\boldsymbol{\Phi}}((\mathbf{1}\boldsymbol{\gamma}^{t}\boldsymbol{\Phi}\bar{\boldsymbol{\Phi}}\mathbf{R}_a^{-1}\boldsymbol{\Phi}) \odot (\boldsymbol{\Phi}\mathbf{R}_a^{-1}\bar{\boldsymbol{\Phi}}\boldsymbol{\Phi}\boldsymbol{\gamma}\mathbf{1}^{t}))\boldsymbol{\Gamma}\bar{\boldsymbol{\Phi}}\mathbf{I}_s \\ &\quad +\imath\boldsymbol{\Phi}((\bar{\boldsymbol{\Phi}}\mathbf{R}_a^{-1}\boldsymbol{\Phi}\bar{\boldsymbol{\Phi}}\boldsymbol{\gamma}\mathbf{1}^{t}) \odot (\mathbf{1}\boldsymbol{\gamma}^{t}\bar{\boldsymbol{\Phi}}\boldsymbol{\Phi}\mathbf{R}_a^{-1}\bar{\boldsymbol{\Phi}}))\boldsymbol{\Gamma}\boldsymbol{\Phi}\mathbf{I}_s \\ &\quad -\imath\boldsymbol{\Phi}((\bar{\boldsymbol{\Phi}}\mathbf{R}_a^{-1}\boldsymbol{\Phi}) \odot (\mathbf{1}\boldsymbol{\gamma}^{t}\bar{\boldsymbol{\Phi}}\boldsymbol{\Phi}\mathbf{R}_a^{-1}\bar{\boldsymbol{\Phi}}\boldsymbol{\Phi}\boldsymbol{\gamma}\mathbf{1}^{t}))\boldsymbol{\Gamma}\bar{\boldsymbol{\Phi}}\mathbf{I}_s\end{aligned}$$

By noting that $\boldsymbol{\Phi}\bar{\boldsymbol{\Phi}} = \mathbf{I}$, it follows that all factors $\boldsymbol{\Phi}$ and $\bar{\boldsymbol{\Phi}}$ cancel each other. Since the second and the third term of the equation cancel each other, and so do the first and fourth term, $\mathbf{M}_{\gamma\phi} = 0$. In the same way it can be shown that $\mathbf{M}_{\phi d} = 0$. The nonzero FIM components can be derived following the same procedure. The results are, using the definition $\alpha = \boldsymbol{\gamma}^{t}\mathbf{R}_a^{-1}\boldsymbol{\gamma}$, and using the fact that $\mathbf{M}$ is Hermitian:

$$\begin{aligned}\mathbf{M}_{\gamma\gamma} &= 2(\alpha\mathbf{R}_a^{-1} + \mathbf{R}_a^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^{t}\mathbf{R}_a^{-1}) \\ \mathbf{M}_{\gamma\phi} &= \mathbf{M}_{\phi\gamma} = 0 \\ \mathbf{M}_{\gamma d} &= \mathbf{M}_{d\gamma} = 2(\mathbf{1}\boldsymbol{\gamma}^{t}\mathbf{R}_a^{-1} \odot \mathbf{R}_a^{-1}) \\ \mathbf{M}_{\phi\phi} &= 2\mathbf{I}_s^{t}\boldsymbol{\Gamma}(\alpha\mathbf{R}_a^{-1} - \mathbf{R}_a^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^{t}\mathbf{R}_a^{-1})\boldsymbol{\Gamma}\mathbf{I}_s \\ \mathbf{M}_{\phi d} &= \mathbf{M}_{d\phi} = 0 \\ \mathbf{M}_{dd} &= \bar{\mathbf{R}}^{-1} \odot \mathbf{R}^{-1} = \mathbf{R}_a^{-1} \odot \mathbf{R}_a^{-1}\end{aligned}$$

The FIM components are real and do not depend on $\boldsymbol{\phi}$. This proves that the estimation accuracy bounds are independent of the particular values of the phases.

Bibliography

[1] S. Alliot, M. Soudani, and J. Bregman. Comparison of filters with a polyphase structure applied to large embedded systems for telescopes. In *Proc. 3^{rd} IEEE Benelux Signal Processing Symposium (SPS-2002)*, Leuven, Belgium, March 2002.

[2] A. v. Ardenne. Concepts of the square kilometer array; towards the new generation radio telescopes. *IAU Symp. 199, Puna , The Universe at Low Frequencies*, December 1999.

[3] AT. Jaarverslag 2003. Technical report, Agentschap Telecom (AT), 2003.

[4] W.A. Baan, P.A. Fridman, and R.P. Millenaar. RFI mitigation at WSRT: Algorithms testobservations, system implementation. *Proc. International Union of Radio Science (URSI), 27^{th} General Assembly, Maastricht, The Netherlands*, August 2002.

[5] J.W.M. Baars, T. Genzel, I.I.K. Pauliny-Toth, and A. Witzel. The absolute spectrum of Cas.A; an accurate flux densityscale and a set of secondary calibrators. *Astron. Astrophys.*, 61:99–106, 1977.

[6] C. Barnbaum and R.F. Bradley. A new approach to interference excision in radio astronomy: Real-time adaptive cancellation. *The Astronomical Journal*, (115):2598–2614, November 1998.

[7] C.M. Beaudet, G. Watts, J. Acree, and S.J. E. Radford. RFI Survey at the ALMA Site at Chajnantor, ALMA MEMO no. 470 . Technical report, National Radio Astronomy Observatory, July 2003.

[8] J. Bell, et al. Software radio telescope: interference atlas and mitigation strategies. In *Perspectives on Radio Astronomy: Technologies for Large Antenna Arrays*. ASTRON, April 1999.

[9] A.J. Boonstra. Interference mitigation strategies for radio astronomy: RFI research areas for SKA. *SKA Workshop, Technology Pathways to the Square Kilometre Array*, August 2000.

[10] A.J. Boonstra. LOFAR RFI mitigation strategy. Technical report, ASTRON, LOFAR, Dwingeloo, The Netherlands, October 2002.

[11] A.J. Boonstra. RFI mitigation strategies. *SKA Workshop*, August 2002.

[12] A.J. Boonstra. LOFAR RFI studies for Drenthe. Technical report, ASTRON / LOFAR, Dwingeloo, The Netherlands, March 2003.

[13] A.J. Boonstra, J.D. Bregman, and A.A. Mohamoud. LOFAR spectrum monitoring: Dynamic range and spectral occupancy issues. *SKA Workshop, Technology Pathways to the Square Kilometre Array*, August 2000.

[14] A.J. Boonstra and A.J.Van der Veen. Calibration method, device, and computer program. Patent WO2004017090, February 26 2002.

[15] A.J. Boonstra, A. Leshem, A.J. van der Veen, A. Kokkeler, and G. Schoonderbeek. The effect of blanking of TDMA interference on radio-astronomical observations: Experimental results. *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3546–3549, June 2000.

[16] A.J. Boonstra, and S. van der Tol S. Wijnholds, and B. Jeffs. Calibration, sensitivity and RFI mitigation requirements for LOFAR. In *2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, PA, USA*. IEEE, March 2005.

[17] A.J. Boonstra and S. van der Tol. Spatial filtering of interfering signals at the intial LOFAR phased array test station. In *Radio Science, accepted for publication*, 2005.

[18] A.J. Boonstra and A.J. van der Veen. Gain decomposition methods in sensor array systems. *11th IEEE Workshop on Statistical Signal Processing, Singapore*, August 2001.

[19] A.J. Boonstra and A.J. van der Veen. Gain estimation for polarized radio telescope arrays. *Proc. International Union of Radio Science (URSI), 27^{th} General Assembly*, August 2002.

[20] A.J. Boonstra and A.J. van der Veen. Dual polarization gain estimation for radio telescope arrays. *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, April 2003.

[21] A.J. Boonstra and A.J. van der Veen. Gain calibration methods for radio telescope arrays. *IEEE Transactions on Signal Processing*, 51(1):25–38, January 2003.

[22] A.J. Boonstra, A.J. van der Veen, and J. Raza. Spatial filtering of continuous interference in radio astronomy. *Proc. IEEE ICASSP, Orlando (FL)*, 3:2933–2936, May 2002.

[23] M Born and E. Wolf. *Principles of Optics*. Pergamon Press, 1964.

[24] A. Bos. *On instrumental effects in spectral line synthesis observations.* PhD thesis, University of Leiden, June 1985.

[25] G.E.P. Box. A general distribution theory for a class of likelihood criteria. *Biometrika*, 36(317-346), 1949.

[26] R. Braun. *The Westerbork Observatory, continuing adventure in radio astronomy*, volume 207 of *Astrophysics and Space Sciences Library*, pages 167–183. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[27] J. D. Bregman. Concept design for a low-frequency array. In *Proc. SPIE Vol. 4015, p. 19-32, Radio Telescopes, Harvey R. Butcher; Ed.*, pages 19–32, July 2000.

[28] J.D. Bregman. Design concepts for a sky noise limited low frequency array. In *Perspectives on Radio Astronomy - Technologies for Large Antenna Arrays*, pages 23–32. ASTRON, 1999.

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

[30] A.G. de Bruyn. The Westerbork Synthesis Radio Telescope, a second lease on life. In E. Raimond and R. Genee, editors, *The Westerbork Observatory, Continuing Adventure in Radio Astronomy*, pages 109–125. ASTRON, Kluwer Academic Publishers, 1996.

[31] J.F. Burger. *Cryogenic Microcooling, A micromachined cold stage operating with a sorption compressor in a vapor compression cycle.* PhD thesis, University of Twente, The Netherlands, 2001.

[32] H. R. Butcher. LOFAR: first of a new generation of radio telescopes. In *Proceedings of the SPIE, Volume 5489, pp. 537-544 (2004)*, pages 537–544, October 2004.

[33] J. Capon and N. R. Goodman. Probability Distributions for Estimators of the Frequency-Wavenumber Spectrum. In *Proc. IEEE, Volume 58, p. 1785-1786*, pages 1785–1786, 1970.

[34] W.A. van Cappellen and G.W. Kant. Design of the Low Frequency Array (LOFAR) active antenna amplifier. In *European Microwave Conference, Milan*, September 2002.

[35] A. Chippendale, M. Storey, and P. Hall (ed). Low Frequency Array RFI Site Report, Mileura Station Western Australia. Technical report, CSIRO Australia Telescope National Facility, Sydney, Australia, March 2003.

[36] V. Clerc, R. Weber, L. Denis, and C. Rosolen. High performance receiver for RFI mitigation in radio astronomy : Application at decameter wavelengths. *EUSIPCO'02, Toulouse, France*, September 2002.

[37] T.J. Cornwell and P.N. Wilkinson. A New Method for Making Maps with Unstable Radio Interferometers. *Mon. Not. R. Astron. Soc.*, 196:1067–1086, 1981.

[38] CRAF. *CRAF handbook for radio astronomy.* ESF, 1997.

[39] CRAF. Power line communications and LOFAR, May 2003.

[40] X. Dong. *Wideband Adaptive Beamforming in Phased Array Based Radio Telescope.* Post masters thesis report, Eindhoven University of Technology (SAI/SPS/TUE) / ASTRON, November 2002.

[41] N. Duric, et al. RFI Report for the US South-West. Technical report, University of New Mexico, Los Alamos National Laboratory, University of Texas, March 2003.

[42] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4), 1988.

[43] G.W. Ellingson and G.A. Hampson. A subspace-tracking approach to interference nulling for phased array-based radio telescopes. *IEEE Transactions on Antennas and Propagation*, 50(1):25–30, January 2002.

[44] S. Ellingson. RFI measurement protocol for candidate SKA sites. Technical report, SKA, S. Ellingson (ed.), May 2003.

[45] S. Ellingson and W. Cazemier. Efficient multibeam synthesis with interference nulling for large arrays. *IEEE Transactions on Antennas and Propagation*, 51(3):503–511, March 2003.

[46] S.W. Ellingson. RFI Suppression for Radio Astronomy: Frequency-, Time-, Space-, and Multidomain Approaches. *URSI General Assembly, Toronto*, August 1999.

[47] S.W. Ellingson. Beamforming and interference cancelling with very large wideband arrays. *IEEE Transactions on Antennas and Propagation*, 51(6):1338–1346, June 2003.

[48] S.W. Ellingson, J. Bunton, and J.F. Bell. Removal of the GLONASS C/A signal from OH spectral line observations using a parametric modeling. *Astrophysical Journal Supplement*, 135(1):88–93, July 2001.

[49] J.R. Fisher. Multi-path measurements using a 1.3 GHz radar signal received by the GBT. *IUCAF RFI mitigation workshop*, March 2001.

[50] S.R. Fleurke, H. Dehling, A.J. Boonstra, A.D. Binkerink, and A.D. den Besten. Monte carlo methods for measuring spectrum occupancy. *IUCAF RFI Mitigation Workshop*, March 2001.

[51] P.A. Fridman. A change point detection method for elimination of industrial interference in radio astronomy receivers. *8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing, Corfu, Greece*, pages 264–266, June 1996.

[52] P.A. Fridman. Robust power spectrum estimation in the presence of interferences, in signal and data processing of small targets. *Proc. Spie, San-Diego*, (4473):1–12, July-August 2001.

[53] P.A. Fridman and W.A. Baan. RFI mitigation methods in radio astronomy. *Astronomy and Astrophysics*, 378:327–344, October 2001.

[54] S.H. Friedberg, A.J. Insel, and L.E. Spence. *Linear Algebra*. Prentice-Hall, 1997.

[55] B. Friedlander and A. Weiss. Direction finding using noise covariance modeling. *IEEE Trans. Signal Processing*, 43(1557-1567), July 1995.

[56] B. Friedlander and A.J. Weiss. Direction finding in the presence of mutual coupling. *IEEE Transactions on Antennas and Propagation*, pages 273–284, March 1991.

[57] P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization*. Academic Press, London, UK, 1995.

[58] M. Goris. Categories of RF interference. Technical report, ASTRON report, February 1998.

[59] A. Graham. *Kronecker Products and Matrix Calculation*. Ellis Horwood Limited, Chichester, England, 1981.

[60] W.H. Greene. *Econometric analysis*. Prentice Hal International, Inc., 2000.

[61] A. W. Gunst, A. B. J. Kokkeler, and G. W. Kant. A/D Converter Research for SKA. In A.B. Smolders and M.P. van Haarlem, editors, *Perspectives on Radio Astronomy: Technologies for Large Antenna Arrays*, pages 261–264. ASTRON, April 12-14 1999.

[62] J.P Hamaker. Understanding radio polarimetry. IV. the full-coherency analogue of scalar self-calibration: Self-alignment, dynamic range and polarimetric fidelity. *Astronomy & Astrophysics Supplement Series*, 143:515–534, May 2000.

[63] J.P. Hamaker, J.D. Bregman, and R.J. Sault. Understanding radio polarimetry. I. mathematical foundations. *Astronomy & Astrophysics Supplement Series*, 117:137–147, May 1996.

[64] G. Hampson, M. Goris, and A. Joseph. The adaptive antenna demonstrator. *IEEE Digital Signal Processing Workshop, Bryce Canyon, Utah, USA*, April 28 1998.

[65] S. Haykin, J. Litva, and T.L. Shepherd, editors. *Radar Array Processing*, volume 25 of *Springer Series in Infromation Sciences*. Springer Verlag, Berlin, 1993.

[66] J.A. Hogbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astron. Astrophys. Suppl.*, 15:417–426, 1974.

[67] A. Horneffer, et al. Lopes: detecting radio emission from cosmic ray air showers. *Proceedings of SPIE*, 5500, June 23-24 2004.

[68] J.E. Hudson. *Adaptive Array Principles*. IEE Electromagnetic Wave Series 11. Peter Peregrinus Ltd., London, U.K., 1981.

[69] ITU. Protection Criteria Used for Radio Astronomical Measurements. Recommendation ITU-R RA.769-1.

[70] ITU. Radio Noise, Recommendation ITU-R. PI.372-6, 1994.

[71] ITU. *Handbook on radio astronomy*. Radiocommunication Bureau, ITU, 1995.

[72] ITU. Protection of the radio astonomy service in frequency bands shared with other services, 1995.

[73] ITU. VHF and UHF propagation curves for the frequency range from 30 MHz to 1000 MHz, 1995.

[74] ITU. Propagation by diffraction, 1999.

[75] ITU. *Radio Regulations*, volume 1, art.1. International Telecommunication Union (ITU), Geneva, 2004.

[76] K.G. Jansky. Electrical disturbances apparently of extraterrestrial origin. *Proc. IRE*, 21:1387, 1933.

[77] B.D. Jeffs, K.Warnick, and L. Li. Improved interference cancellation in synthesis array radio astronomy using auxiliary antennas. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, 5:77–80, April 2003.

[78] Li L. Jeffs, B.D. and K Warnick. Auxiliary antenna-assisted interference mitigation for radio astronomy array. *IEEE Transactions on Signal Processing*, 53(2):439–451, February 2005.

[79] F.A. Jenkins and H.E.White. *Fundamentals of Optics*. Mc Graw-Hill, 4^{th} edition, 1976.

[80] B.L. Kasper, F.S. Chute, and D. Routledge. Excising terrestial interference in low frequency radio astronomy. *Mon. Not. R. Astr. Soc.*, 199:345–354, 1982.

[81] S.M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*, volume 1. Prentice Hall, Inc., New Jersey, 1993.

[82] S.M. Kay. *Fundamentals of Statistical Signal Processing: Detection Theory*, volume 2. Prentice Hall, Inc., New Jersey, 1998.

[83] A.B.J. Kokkeler, P. Fridman, and A. van Ardenne. Degradation due to quantisation noise in radio astronomy phased arrays. *Experimental Astronomy*, 11:33–56, 2001.

[84] O.M. Kolkman, editor. *The Westerbork Synthesis Radio Telescope User Documentation*. Dwingeloo, The Netherlands, 1.0.0 edition, July 16 1993.

[85] H. Kollen. LOFAR System Requirements Specification SDE. *Report, LOFAR-ASTRON-SRS-001, rev.3.0.*, May 2004.

[86] M. Kouwenhoven. *Pulsar observations with the Westerbork Synthesis Radio Telescope*. PhD thesis, University of Utrecht, The Netherlands, October 2000.

[87] J.D. Kraus. *Radio Astronomy*. Cygnus Quasar Books, second edition, 1986.

[88] H. Krim and M. Viberg. Two decades of array signal processing research. *IEEE Signal Processing Magazine*, July 1996.

[89] W. Krzanowski. *Principles of multivariate analysis*. Oxford University Press, Oxford, 1988.

[90] H. de Lassus and A. Lecacheux. Automatic recognition of low frequency radio planetary signals. In *4th International Workshop on Radio Emissions from Planetary Magnetospheres, Graz,*, September 1996.

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

[92] A. Leshem, A.J. Boonstra, A.J. van der Veen, A. Kokkler, and G. Schoonderbeek. Blanking of TDMA interference and its effect on radio-astronomical correlation measurements: Experimatal results. *STW/IEEE/ProRISC99 Workshop on Circuits, Systems and Signal Processing*, pages 275–280, November 1999.

[93] A. Leshem and A.J. van der Veen. Introduction to interference mitigation techniques in radio astronomy. *Perspectives on Radio Astronomy: Technologies for Large Atenna Arrays*, pages 201–223, April 1999.

[94] A. Leshem and A.J. van der Veen. The effect of adaptive interference suppression on radio astronomical image formation. *Radio Telescopes, Proceedings of SPIE*, 4015:341–352, 2000.

[95] A. Leshem and A.J. van der Veen. Adaptive interference suppression and its effect on radio astronomical image formation. *Proceedings of IEEE SPS-2000. Hilvarenbeek, The Netherlands*, March 2000.

[96] A. Leshem and A.J. van der Veen. Adaptive suppression of RFI and its effect on radio-astronomical image formation. *Proceedings of the International Conference on Image Processing*, 3:616–619, October 2001.

[97] A. Leshem and A.J. van der Veen. Multichannel detection of gaussian signals with uncalibrated receivers. *IEEE Signal Processing Letters*, 8:120–122, April 2001.

[98] A. Leshem, A.J. van der Veen, and A.J. Boonstra. Multichannel interference mitigation techniques in radio astronomy. *The Astrophysical Journal Supplement Series*, 131(1):355–373, November 2000.

[99] A. Leshem, A.J. van der Veen, and E.F. Deprettere. Detection and blanking of GSM interference in radio-astronomical observations. *Proc. IEEE workshop on Signal Processing Advances in Wireless Communication, Annapolis (MD)*, pages 374–377, May 1999.

[100] L. Li and B.D. Jeffs. Analysis of adaptive array algorithm performance for satellite interference cancellation in radio astronomy. *Proc. International Union of Radio Science (URSI), 27^{th} General Assembly*, 2002.

[101] X. Liu and N.D. Siridopoulis. Parafac methods for blind beamfoming: Multilinear ALS performance and CRB.

[102] L. Ljung. *System Identification, Theory for the User*. Prentice Hall, 1987.

[103] A. Manikas and N. Fistas. Modelling and estimation of mutual coupling between array elements. *Proc. ICASSP*, pages 553–556, April 1994.

[104] D.G. Manolakis, V.K. Ingle, and S.M. Kogon. *Statistical and Adaptive Signal Processing*. Mac Graw Hill, 2000.

[105] K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Probability and Mathematical Statistics. Academic Press, London, 1979.

[106] S. Maslakovic, I.R. Linscott, M. Oslick, and J.D. Twicken. Excising radio frequency interference using the discrete wavelet transform. *In Proceedings of the IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis*, pages 349–352, 1996.

[107] J. Morawietz, J.F.N. Roosjen, G.W. Kant, and J.G. Bij de Vaate. A flexible, cost effective, broadband-receiver unit for the next generation radio telescopes. *IEEE conference proceedings, 12th International Conference for Microwave & Telecommunication Technology, Sevastopol*, September 2002.

[108] D.M. Morgan. *A Handbook for EMC Testing and Measurement*. Peter Peregrinus Ltd, UK, 1994.

[109] J.E. Noordam. Generalized self-calibration for LOFAR. *Proc. International Union of Radio Science (URSI), 27^{th} General Assembly*, August 2002.

[110] J.E.N. Noordam. LOFAR calibration challenges. *SPIE conference on astronomical telescopes and instrumentation, Glasgow UK*, June 2004.

[111] B. Ottersten, P. Stoica, and R. Roy. Covariance Matching Estimation Techniques for Array Signal Processing Applications. *Digital Signal Processing, A Review Journal*, 8:185–210, July 1998.

[112] A. Papoulis. *Signal Analysis*. McGraw-Hill, New York, USA, 1984.

[113] R.A. Perley, F.R. Schwab, and A.H. Bridle. *Synthesis Imaging in Radio Astronomy*, volume 6. Astronomical Society of the Pacific Conference Series, 1994.

[114] M. Pesavento and A.B. Gershman. Array processing in the presence of unknown nonuniform sensor noise: a maximum likelihood direction finding algorithm and Cramer-Rao bounds. *IEEE Wokshop on Statistical Signal Processing*, 2000.

[115] P. Picard, D. Aubry, E. Gerard, and I. Thomas. A short pulse blanker for wide band radio spectroscopy. *In Proc. XXVII th General Assembly of the International Union of Radio Science*, August 2002.

[116] J.G. Proakis and D.G. Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications*. Macmillan Publishing Company, New York, USA, 2^{nd} edition, 1992.

[117] E. Raimond and R. Genee, editors. *The Westerbork Observatory, Continuing Adventure in Radio Astronomy*, volume 207 of *Astrophysics and Space Library*. Kluwer Academic Publishers, 1996.

[118] P. Ravier and R. Weber. Robustness in the RFI detection for time blanking. *In Proc. XXVII th General Assembly of the International Union of Radio Science,*, August 2002.

[119] J. Raza, A.J. Boonstra, and A.J. van der Veen. Spatial filtering of RF interference in radio astronomy. *IEEE Signal Processing Letters*, 9(2):64–67, February 2002.

[120] G. Reber. Cosmic static. *Astrophysical Journal*, 91:621–624, June 1940.

[121] L.v.d. Ree. A lumped element highpass filter for the UHF-high band. Technical report, ASTRON, November 1999.

[122] J.H. Reed. *Software Radio: A Modern Approach to Radio Engineering*. Prentice Hall, May 2002.

[123] D.W. Robinson and G.T. Reid, editors. *Interferogram Analysis: digital fringe pattern measurement techniques*. IOP Publishing Ltd, London, 1993.

[124] K. Rohlfs. *Tools of Radio Astronomy*. Springer-Verlag, Heidelberg Germany, 1990.

[125] C. Rosolen, V. Clerc, and A. Lecacheux. High dynamic range, interferences tolerant, digital receivers for radio astronomy. *The Radio Science Bulletin, U.R.S.I.*, (291):6–12, December 1999.

[126] K. Rothammel. *Antennenbuch*. Telekosmos-Verlag, Franckische Verlagshandlung, Stuttgart, Germany, 1984.

[127] S. Roy et al. Ultrawideband radio design: the promise of high-speed, short-range wireless connectivity. *Proceedings of the IEEE*, 92(2), February 2004.

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

[129] F.H. Sanders, B. J. Ramsey, and V. S. Lawrence. Broadband spectrum survey at Los Angeles, California. Technical report, NTIA Report 97-336., May 1997.

[130] U. J. Schwarz. Mathematical-statistical Description of the Iterative Beam Removing Technique (Method CLEAN). *Astron. Astrophys.*, 65:345–356, April 1978.

[131] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, July and October 1948.

[132] K. Sharman, T. Durani, M. Wax, and T. Kailath. Asymptotic performance of eigenstructure spectral analysis methods. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84*, 9:440–443, March 1984.

[133] A.B. Smolders, J.G. Bij de Vaate, G.W. Kant, A. van Ardenne, D. Schaubert, and T.H. Chio. Dual-beam wide-band beamformer with integrated anenna array. *IEEE Millenium Conference on Antennas & Propagation*, April 2000.

[134] A.B. Smolders and M.P. van Haarlem, editors. *Perspectives in Radio Astronomy: Technologies for Large Antenna Arrays*. Conference proceedings, ASTRON. ASTRON, May 2000.

[135] A.B. Smolders, et al. Receiver architectures of the thousand element array (THEA). *Proc. SPIE Conf. 4015 Radio Telescopes, München, Germany*, March 2000.

[136] B. Smolders and G. Hampson. Deterministic RF nulling in phased arrays for the next generation of radio telescopes. *IEEE Antennas and Propagation Magazine*, 44(4):13–22, August 2002.

[137] W. Soares-Filho, J.M. Seixas, and L.P. Caloba. Enlarging neural class detection capacity in passive sonar systems. *ISCAS, Phoenix, USA*, May 2002.

[138] H. van Someren-Greve. Logarithmic least square gain decomposition algorithm for the WSRT, 1980. IWOS software documentation, ASTRON internal document.

[139] G. Staple and K. Werbach. The end of spectrum scarcity. *IEEE Signal Processing Magazine*, 41(3):48–52, March 2004.

[140] P. Stoica and M. Cederval. Detection tests for array processing in unknown correlated noise fields. *IEEE Trans. Signal Processing*, 45:2351–2362, Sept 1997.

[141] P. Stoica and A. Nehorai. MUSIC, maximum likelihood, and Cramer-Rao bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(5):720–741, May 1989.

[142] P. Stoica, B. Ottersten, M. Viberg, and R.L. Moses. Maximum likelihood array processing for stochastic coherent sources. *IEEE Transactions on Signal Processing*, 44(1):96–105, January 1996.

[143] T. Svantesson. *Direction Finding in the Presence of Mutual Coupling*. PhD thesis, Chalmers University of Sweden, 1999.

[144] J.C. Tarter. SETI is still alive: Results from one year of high resolution microwave survey observations and a progress report on Project PHOENIX. *American Astronomical Society, 184th AAS Meeting; Bulletin of the American Astronomical Society*, 26:948, May 1994.

[145] J. H. Taylor, R. N. Manchester, and A. G. Lyne. Catalog of 558 pulsars. *Astrophysical Journal Supplement Series*, 88(2):529–568, October 1993.

[146] A.R. Thompson, J.R. Moran, and G.W. Swenson. *Interferometry and Synthesis in Radio Astronomy*. John Wiley & Sons, New York, first edition, 1986.

[147] S. van der Tol. and A.J. van der Veen. Performance analysis of spatial filtering of RF interference in radio astronomy. *IEEE Transactions on Signal Processing*, 53(3):896–910, March 2005.

[148] S. van der Tol, A.J. van der Veen, and A.J. Boonstra. Mitigation of continuous interference in radio astronomy using spatial filters. *Proc. International Union of Radio Science (URSI), 27^{th} General Assembly*, August 2002.

[149] H.L. van Trees. *Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory*. Wiley Interscience, John Wiley & Sons, 2002.

[150] A.J. van der Veen. Joint diagonalisation via subspace fitting techniques. *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2001.

[151] A.J. van der Veen, A. Leshem, and A.J. Boonstra. Signal processing for radio astronomical arrays. *IEEE Sensor Array and Multichannel Signal Procesing workshop (SAM), Sitges, Barcelona, Spain.*, July 2004.

[152] A.J. van der Veen and A.J. Boonstra. Spatial filtering of RF interference in radio astronomy using a reference antenna. *IEEE, ICASSP 2004, Montreal, Canada*, May 2004.

[153] A.J. van der Veen and E. Deprettere. *Signal processing for communications, an Algebraic approach*. PATO Course on Signal Procesing, 1996.

[154] B.D. van Veen and K.M. Buckley. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Magazine*, April 1988.

[155] T.L. Venkatasubramani et al. Conceptual design of the analog receiver for SKA. ASTRON, April 12-14 1999.

[156] M. Viberg and B. Ottersten. Sensor array processing based on subspace fitting. *IEEE Trans. Acoust. Speech Signal Processing*, 39:1110–1121, May 1991.

[157] L. Voute. *The many shapes of Giant Pulses*. PhD thesis, University of Amsterdam, The Netherlands, October 2001.

[158] S.O. Wallage. *Superconducting transmission lines in microwave filters*. Delft University Press, Delft, 1997.

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

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

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

[162] R. Weber, V. Clerc, L. Denis, and C. Rosolen. Robust receiver for RFI mitigation in radio astronomy. *In Proc. XXVII th General Assembly of the International Union of Radio Science*, August 2002.

[163] R. Weber and C. Faye. Real time detector for cyclostationary RFI in radio astronomy. *EUSIPCO'98, Greece*, September 1998.

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

[165] S. Weinreb. Noise temperature estimates for a next generation very large microwave array. *IEEE MTT-S International Microwave Symposium*, 2:673–676, 1998.

[166] E.W. Weisstein. *CRC Concise Encyclopedia of mathematics*. Chapmann & Hall, 2003.

[167] A.D. Whalen. *Detection of Signals in Noise*. Academic Press Inc., San Diego, 1971.

[168] B. Widrow and S.D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, 1985.

[169] M.H. Wieringa. An Investigation of the Telescope Based Calibration Methods 'Redudancy' and 'Self-Cal'. *Experimental Astronomy*, 2:203–225, 1992.

[170] S.J. Wijnholds, J.D. Bregman, and A.J. Boonstra. Sky noise limited snapshot imaging in the presence of RFI with the LOFAR Initial Test Station. *Experimental Astronomy, special issue on SKA. accepted for publication, 2005.*

[171] E.E.M. Woestenburg. Final report on interference suppression in radio astronomy receivers, subproject 2 of the TMR-LSF RTD project, enhancing the European network of radio telescopes. Technical report, ASTRON, 2001.

[172] E.E.M. Woestenburg, J-G. bij de Vaate, and R.H. Witvers. Wide band low noise active antenna for radio astronomical arrays. *AP2000 Millenium Conf. Antennas and Propagation, Davos, Switzerland*, SP-444:414, April 2000.

[173] E.E.M. Woestenburg and K.F Dijkstra. Noise characterization of a phased array tile. *Proceedings of the 33rd European Microwave Conference, Munich, Germany*, October 2003.

[174] R. Wohlleben, H. Mattes, and T. Kirchbaum. *Interferometry in Radio Astronomy and Radar Techniques*. Kluwer Academic Publishers, Dordrecht, The Neterlands, 1991.

[175] A. Yeredor. Non-orthogonal joint diagonalisation in the least squares sense wit application in blind source separation. *IEEE Transactions on Signal Processing*, 50:1545–1553, July 2002.

[176] P.H. Young. *Electronic Communications Techniques*. Merrill's International Series in Engineering Technology. Macmillan Publishing Company, New York, 3 edition, 1994.

[177] M. Zatman. How narrowband is narrowband. *IEE Proc.-Radar, Sonar Navig.*, 145(2):85–91, April 1998.

[178] J. Zhou, M.J Lancaster, and F. Huang. HTS narrow band filters at UHF band for radio astronomy applications. *Applied Superconductivity Conference, ASOC, Jacksonville, Florida*, October 2004.

Samenvatting

Radiofrequentie storingsonderdrukking in de radioastronomie

Met radiotelescopen worden waarnemingen gedaan ten behoeve van onderzoek naar de structuur en dynamiek van het heelal. De waar te nemen objecten staan op grote afstand en zijn doorgaans erg zwak. Om toch zinvolle waarnemingen te kunnen doen is een extreem grote gevoeligheid nodig. Deze gevoeligheid wordt in de radioastronomie bereikt door gebruik te maken van grote telescoopspiegel- en antenneoppervlakten, door in brede waarneemfrequentiebanden te meten, door versterkers met lage ruiskarakteristieken te gebruiken en door gedurende lange tijd een object waar te nemen (integreren in tijd). De Westerbork Synthese Telescoop bijvoorbeeld heeft als standaard meettijd 12 uur. Deze aanpak levert een gevoeligheid op die tien orden van grootte beter is dan die in de meeste communicatiesystemen.

Door de technologische vooruitgang in de microelectronica in de afgelopen decade kunnen telescopen in principe nog veel gevoeliger gemaakt worden. Voorbeelden hiervan zijn de Low Frequency Array (LOFAR), een Nederlandse "phased array" radio telescoop die op dit moment in de constructiefase is, en de Square Kilometer Array (SKA), een internationale telescoop die zich nog in de concept studiefase bevindt. Echter, de technologische vooruitgang heeft er ook voor gezorgd dat nieuwe radiocommunicatiesystemen ontwikkeld zijn. Er is een toename zowel in het aantal systemen als in verscheidenheid. Nieuwe signaalmodulatie en coderingssystemen zijn ontwikkeld, zoals bijvoorbeeld Ultra Wideband technieken. Door deze ontwikkeling is het gebruik van het frequentiespectrum intensiever geworden en komt het vaker voor dat communicatiesystemen de radioastronomische waarnemingen verstoren. Deze verstoring wordt radiofrequentie interferentie (RFI) of electromagnetische interferentie (EMI) genoemd.

Om de invloed van storende radiosignalen op de radio astronomische waarnemingen te reduceren, kunnen filtertechnieken toegepast worden. Om te beginnen dienen de telescoopontvangers voldoende lineair gedimensioneerd te zijn. Dit gebeurt onder andere door het toepassen van geschikte componenten, mengschema's en analoge filters. Na digitalisatie van de ontvangen signalen kunnen de

stoorsignalen gereduceerd worden met behulp van digitale filtertechnieken. Dit proefschrift richt zich op dit soort digitale technieken, toegepast op apertuursynthese waarnemingen. Apertuursynthese is een waarneemtechniek waarbij een grote telescoop gesynthetiseerd wordt door de signalen van vele kleintjes, een zogenaamd "array", te combineren. Deze waarneemmode is gekozen om diverse redenen: ten eerste omdat de nieuwe generatie radiotelescopen van het synthese-type zullen zijn vanwege praktische grenzen aan de grootte van telescoopspiegels; ten tweede omdat array-filtertechnieken, ten tijde van het starten van het onderzoek, niet of nauwelijks onderzocht waren in de context van de radioastronomie.; en ten slotte omdat de onderzochte technieken "blind" zijn in die zin dat er weinig of geen a-priori infomatie nodig is om de filters goed te laten werken. Het voordeel hiervan is dat de methoden relatief eenvoudig zijn en toepasbaar op een uitgebreide klasse van stoorsignalen.

Het onderzoek in dit proefschrift bouwt voort op de datamodellen, detectietechnieken en interferentiefilters van A.J. van der Veen en A. Leshem [93] en op de polarisatiemodellering van onder andere J. Hamaker [63]. De door hen afgeleide matrixmodellen voor de signaalverwerking worden in deze thesis op een aantal punten verder uitgewerkt en uitgebreid met onder andere een multipadmodel voor stoorsignalen en een oplossingsmethode voor polarisatie calibratie. Hoewel de modellen naast een matrixrepresentatie evengoed als sommaties van scalaire grootheden uitgedrukt kunnen worden, biedt de matrixrepresentatie een aantal voordelen. Ten eerste is er het voordeel van compacte notatie en ten tweede nodigt een overzichtelijke schrijfwijze uit tot het vinden van onderliggende structuren in het probleem en tot nieuwe opossingsbenaderingen. Een bijkomend voordeel is dat er in de statistische signaalverwerkingsliteratuur oplossingen gevonden zijn voor vraagstukken die dezelfde mathematische matrixmodellen blijken te hebben als sommige modellen in de radioastronomie.

In deze thesis zijn storingsdetectietechnieken en ruimtelijke filtertechnieken onderzocht, gesimuleerd en uitgetest op radioastronomische waarnemingen. Deze methoden zijn gebaseerd op lineaire algebra "subspace" technieken. De effectiviteit van deze filters is onderzocht en de beperkingen ervan zijn in relatie gebracht tot systeemparameters zoals de gebruikte kanaalbandbreedte, integratie tijd, signaal-ruis verhouding, maar ook tot spatiële bronafmetingen en multipadeffecten. Verder is de waargenomen effectiviteit vergeleken met de theoretische verwachting.

Zowel voor detectie als voor het ruimtelijk filteren worden in deze thesis eigenwaarde ontbindingen gebruikt. Deze techniek kan alleen goed toegepast worden als de ruis van de telescopen zodanig geschaald wordt dat de waargenomen ruisvermogens voor alle telescopen gelijk zijn. Om dit te bereiken zijn calibratiemethoden onderzocht. In deze thesis zijn methoden uit de factoranalyse gebruikt, maar ook zijn nieuwe methoden ontwikkeld. Twee van deze ontwikkelde algoritmen zijn gepatenteerd. Verder is bij deze analyse de Cramer-Rao bound gebruikt om de theoretische calibratieparameter schattingsnauwkeurigheid vast te stellen. Deze techniek is een standaardtechniek in bijvoorbeeld de statistische signaalverwerking, maar binnen de radio astronomie is deze

techniek nieuw.

Tenslotte is onderzocht hoe interferentie filtertechnieken het beste kunnen worden toegepast in de LOFAR telescoop. Naast een analyse zijn er spatiële filtertesten uitgevoerd met het LOFAR testsation ITS.

Albert-Jan Boonstra, 14 juni 2005.

Acknowledgements

In the first place I would like to express my gratitude to my supervisor, prof. A.J. van der Veen, for his gentle coaching and for his lucid support. I would also like to thank prof. P.M. Dewilde for supervising and encouraging me. Thanks to prof. J.T. Fokkema (Rector Magnificus), prof. A.J. van der Veen, prof. P.M. Dewilde , prof. M.H.G. Verhaegen, prof. A.G. de Bruyn, prof. B.D. Jeffs, Ir. A. van Ardenne, and Dr. A. Leshem for participating in the jury.

I also want to thank TU Delft, ASTRON, and STW for funding my research. Ir. A. van Ardenne, then Head R&D of ASTRON, gave me the opportunity to do this research, and my then boss, Dr. W.Baan, who encouraged me to turn the results into a PhD thesis. Arnold and Willem, thank you both for your support and encouragement.

I am grateful to the NOEMI team for creating a stimulating environment in which my PhD research could take shape. A.J. van der Veen and A. Leshem set me an example by being energetic and creative researchers. Alle-Jan and Amir, you have been to me what "Newtons giants" were to him. I also want to thank G. Schoonderbeek, Jérôme Dromer, and H. Lincklaen Arriëns for their hardware and software support, N. Petrochilos for helping me put the finishing touches on the Cramer-Rao bound solution, and Bas van der Tol for helping me with the intermodulation analysis.

Then I want to thank Dr W. Baan, Director of the Westerbork Radio Observatory, and the Westerbork crew for the telescope time and support at the WSRT to do our measurements. I would specifically like to mention M. Bentum, R.Vermeulen, the telescope operators J. Sluman, G. Kuper and B. Harms, and the electronic engineers H. Weggemans and H.J. Stiepel. Thanks to the LOFAR team, especially Stefan Wijnholds, for supplying me with the data for the last chapter.

I am grateful to my parents, my close relatives, and all my friends for their interest and their moral support. I wish my father could have seen the completion of this book. Special thanks also to my friends Wander Lowie and Jacqueline van der Poel for sympathising with me and helping me keep my spirits up at

difficult moments during the writing process - they know what it is like to live with a PhD thesis. I also want to thank my friend Leendert Hooites for all the chats about life, the universe, and everything.

Thanks to Pien Rotterdam, Marijke van der Horst, and Robert Braun for correcting the English and style of (parts of) my thesis.

Thanks to Pien for her support, without which the thesis enterprise would have been impossible for me, and to Marieke for fetching me when supper was ready and making all further work impossible.

About the author

Albert-Jan Boonstra was born in Nijmegen, the Netherlands, in 1961. He received B.Sc. and M.Sc. degrees in Applied Physics from Groningen University, the Netherlands, in 1984 and 1987 respectively.

He worked at the Laboratory for Space Research Groningen from 1987 to 1991, where he was involved in developing the Short Wavelength Spectrometer (SWS) for the Infrared Space Observatory (ISO).

In 1992 he joined ASTRON, the Netherlands Foundation for Research in Astronomy, initially at the Radio Observatory Westerbork, The Netherlands. He is currently with the ASTRON R&D Department, Dwingeloo, where he heads the Digital and Embedded Signal Processing group.

From 1999 to 2003 he also worked at Delft University of Technology, pursuing a Ph.D. degree. His research interests lie in the area of signal processing, specifically in radio frequency interference (RFI) mitigation techniques for radio astronomy. In 2002 he was granted the KIvI Telecommunication Section Best Thesis Award. Also in 2002, an array calibration patent was issued, based on work done by A.J.Boonstra and A.J. van der Veen.