

Fish-Eye Observing with Phased Array Radio Telescopes

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 2 maart 2010 om 12:30 uur
door

Stefan Jeroen WIJNHOLDS

natuurkundig ingenieur
doctorandus in de sterrenkunde
geboren te Groningen.

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. A.J. van der Veen

Samenstelling promotiecommissie:

Rector Magnificus

Prof. dr. ir. A.J. van der Veen

Prof. dr. ir. J. Biemond

Prof. dr. ir. M. Verhaegen

Prof. dr. A.G. de Bruyn

Prof. dr. W.N. Brouw

Prof. dr. Ir. M. Moonen

Prof. dr. M. Viberg

voorzitter

Technische Universiteit Delft, promotor

Technische Universiteit Delft

Technische Universiteit Delft

Rijksuniversiteit Groningen

Rijksuniversiteit Groningen

Katholieke Universiteit Leuven, België

Chalmers University of Technology,

Göteborg, Zweden

Copyright © 2010 by S.J. Wijnholds

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the prior permission of the author. An exception is made for retrieval from the World Wide Web for personal use only.

Printed in The Netherlands by Wöhrmann Print Service

Typeset by the author with the L^AT_EX Documentation System.

Author email: stefan@wjnholds.nl

Published and distributed by S.J. Wijnholds

ISBN 978-90-9025180-6

Fish-Eye Observing with Phased Array Radio Telescopes

Stefan J. Wijnholds

Contents

Preface	ix
1 Introduction	1
1.1 Phased array technology	3
1.2 LOFAR	5
1.3 Hierarchical system architecture	8
1.4 Scope of this thesis	10
1.5 Outline and summary of main results	11
1.6 Publications and other contributions	14
2 Calibration and imaging with LOFAR	17
2.1 Coordinate systems	18
2.1.1 Direction of arrival	18
2.1.2 Polarization	19
2.1.3 Antenna locations	20
2.2 Highlights from early prototypes	20
2.2.1 The Ten Heterogeneous Element Test Array	20
2.2.2 The LOFAR Prototype Station	22
2.2.3 The Initial Test Station	24
2.2.4 Core Station 10	27
2.3 Station calibration	27
2.3.1 Challenges	29
2.3.2 Station calibration pipeline	33
2.3.3 Demonstrations	34
2.4 Fish-eye imaging	38
2.4.1 DFT versus least squares imaging	39
2.4.2 Statistical performance	41
2.4.3 All-sky least squares synthesis imaging	43

3 Data model	47
3.1 Physical model	48
3.2 Signal processing models	53
3.2.1 Short term covariance matrix	53
3.2.2 Polarimetry	54
3.2.3 Commonly used data models	57
3.3 Noise subspace and signal subspace	58
3.4 Narrowband assumption	61
3.5 Spatial filtering of source signals	66
3.5.1 Scalar beam forming	66
3.5.2 Array response	67
3.6 From snapshots to synthesis observations	70
4 Cramer-Rao Bound	73
4.1 Definition of the CRB	74
4.2 Analytic expressions for the CRB	76
4.2.1 Partitioning the FIM	76
4.2.2 Jacobians	78
4.2.3 Constituents of the FIM	82
4.3 Impact of gain phase constraints	84
4.3.1 Constraints on the gain phase solution	84
4.3.2 The constrained CRB	85
4.3.3 Single calibration source	86
4.3.4 Multiple calibration sources	88
4.3.5 Application example	89
4.3.6 Concluding remarks	91
4.4 Dealing with missing data	92
4.4.1 Problem statement	92
4.4.2 Derivation of the CRB	93
4.4.3 Discussion	96
5 Calibration	99
5.1 Introduction	99
5.2 Problem statement	103
5.3 Algorithm development	104
5.3.1 Generalized least squares formulation	104
5.3.2 Estimation of direction independent gains	105
5.3.3 Source power estimation	110
5.3.4 Estimating receiver noise powers	110
5.3.5 DOA estimation	112

5.3.6 (Weighted) alternating least squares	113
5.3.7 Computational complexity	115
5.4 Algorithm validation	116
5.5 Non-diagonal noise covariance matrix	124
5.5.1 Motivation	124
5.5.2 Modification of the WALS method	125
5.6 Summary of the main results	129
6 Imaging	131
6.1 Introduction	131
6.2 Fundamental imaging limits	133
6.2.1 Imaging and deconvolution	134
6.2.2 Effective noise	143
6.2.3 Implications	149
6.2.4 Conclusions	151
6.3 Least squares synthesis imaging	153
6.3.1 Generic formulation	153
6.3.2 Scalar imaging with known or negligible noise	157
6.3.3 Scalar imaging with unknown noise	158
6.3.4 Full polarized imaging with known or negligible noise	158
6.3.5 Full polarized imaging with unknown noise	160
6.3.6 Treatment of more complicated noise models	161
6.A Derivation of closed form least squares imaging solutions	161
6.A.1 Scalar imaging with known or negligible noise	161
6.A.2 Scalar imaging with unknown noise	163
6.A.3 Full polarized imaging with known or negligible noise	166
6.A.4 Full polarized imaging with unknown noise	168
7 Conclusions and suggestions for further research	171
7.1 Summary of the main results	171
7.2 Suggestions for further research	174
7.2.1 Calibration	174
7.2.2 Imaging	176
7.3 Towards SKA	178
A Abbreviations and acronyms	181
B Notation	183
B.1 Symbols	183
B.2 Operators	184

C Matrix product relations	185
References	187
Samenvatting	203
About the author	207

Preface

In some sense, this thesis is living proof that any plan is valid until it is confronted with reality. After following a crash course in radio telescope integration and testing by Jaap Bregman during my internship at the Netherlands Institute for Radio Astronomy (ASTRON), I decided to apply for a job as research assistant with ASTRON instead of pursuing a Ph.D. I thus started working on system testing of the early LOFAR (Low Frequency Array) prototypes THETA (Ten Heterogeneous Element Test Array) and ITS (Initial Test Station). During this period, I spent even more hours in Jaap's office to learn the tricks of the trade. The main goal of these prototypes was a performance evaluation of an antenna array in the field. This naturally led to an in depth study of in situ calibration of antenna arrays, on which I gave a status update during a calibration workshop on May 31, 2005. Alle-Jan van der Veen was one of the attendees to that meeting and invited me for further discussion in Delft. This led to a new series of fruitful discussions introducing me to the world of signal processing and ultimately to this thesis. I would therefore like to express my deepest gratitude to both Jaap and Alle-Jan for all their guidance, support and advice.

Working on a big project like LOFAR implies that this research was not done in splendid isolation. I would therefore like to thank my colleagues at ASTRON for all inspiring conversations, their support on the hardware and software needed to do the measurements and for creating a pleasant environment to work in. Taking the risk of regretful omissions, I would like to thank a few persons more specifically. This research was, to a considerable extent, driven by the "features" found in the data obtained from the prototype stations. These measurements would not have been so successful without the hardware support by Yde Koopman, Gijs Schoonderbeek, Andre Gunst and Menno Norden and the software support by Klaas-Jan Wierenga and Chris Broekema. ITS required a focused one-week debug session to align hardware and on-line and off-line software together with Michiel Brentjens, who was always available to reflect on new ideas and challenges as well. It was more than just convenient to have Albert-Jan Boonstra, who graduated under supervision of Alle-Jan in 2005, sitting in the same building to discuss signal processing concepts and issues.

This research also benefitted from the insights of Ger de Bruyn, Johan Hamaker, Wim Brouw and Jan Noordam, who shared their lifelong experience with the calibration and imaging challenges of the Westerbork Synthesis Radio Telescope (WSRT) and from discussions with Ronald Nijboer. The latter, together with Albert-Jan, Jan and Jaap, also provided the personal support to start this thesis project in the first place.

Although there has been considerable synergy between the research presented in this thesis and the demands from the LOFAR project, writing papers and especially writing a thesis forms a major distraction from project work. I would therefore like to thank Arnold van Ardenne, former director of the R&D department, Marco de Vos, former LOFAR project manager and successor of Arnold, and Dion Kant, group leader of the system design and integration group, for supporting my thesis work by allowing me to spend time on this. Especially in the last year, this has been a great help in getting things finished.

Finally, I would like to thank the committee members, Jan Biemond, Michel Verhaegen, Ger de Bruyn, Wim Brouw, Marc Moonen and Mats Viberg, for the time invested in thoroughly reviewing my thesis. This has been a great help in improving the quality and readability of this thesis.

Stefan Wijnholds
Dwingeloo, The Netherlands, January 31, 2010

Chapter 1

Introduction

Astronomers study the physical phenomena outside the Earth’s atmosphere by observing cosmic particles and electromagnetic waves incident on the Earth. Each type of observation provides another perspective on the universe thereby unraveling some mysteries while raising new questions. Over the years, astronomy has become a truly multi-wavelength science. A nice demonstration is provided in Fig. 1.1. In this image, the neutral hydrogen gas observed with the Westerbork Synthesis Radio Telescope (WSRT) exhibits a warped structure that is completely invisible in the optical image from the Sloan digital sky survey [106]. The radio observations therefore provide a radically different view on the dynamics of this galaxy.

Images like Fig. 1.1 are only possible if the instruments used to observe different parts of the electromagnetic spectrum provide a similar resolution. This poses quite a challenge since the resolution of any telescope is determined by the ratio of wavelength and telescope diameter. Consequently, the aperture of radio telescopes has to be 5 to 6 orders of magnitude larger than that of an optical telescope to provide comparable resolution, i.e. radio telescopes should have an aperture of several hundred kilometers. Although it is not feasible to make a dish of this size, it is possible to synthesize an aperture by building an interferometer, i.e., an array of receiving elements.

An interferometer measures the correlation between signals received by two antennas spaced at a certain distance from each other. Initially used to study a single source passing over the sky, the principle was used in optical astronomy in the Michelson stellar interferometer (1890, 1920); the first radio observations using two dipoles were done by Ryle and Vonberg in 1946 [96,97]. Examples of subsequent instruments were the Cambridge One Mile Telescope, the 3 km WSRT in Westerbork, The Netherlands (1970, 14 dishes with 25 m diameter), the 36 km Very Large Array

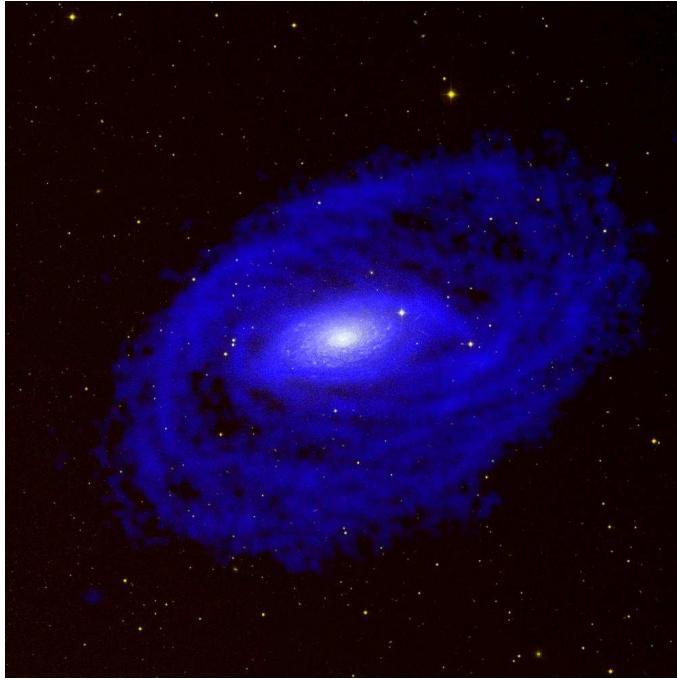


Figure 1.1: Image of the spiral galaxy NGC 5055, showing the structure of the neutral hydrogen gas observed with the Westerbork Synthesis Radio Telescope (blue) superimposed on an optical image of the same galaxy from the Sloan digital sky survey (white) [40].

(VLA) in Socorro, New Mexico, USA (1980, 27 movable 25-m dishes) and the 25 km Giant Meter-wave Radio Telescope (GMRT) in Pune, India (1988, 30 dishes with 45 m diameter). These telescopes use Earth rotation to obtain a sequence of correlations for varying antenna baseline geometries w.r.t. the observed sources, resulting in high-resolution images via synthesis mapping. Even larger baselines (up to a few thousand kilometers) were obtained by combining these instruments into a single instrument using a technique called very long baseline interferometry (VLBI), where the telescope outputs are time-stamped and post-processed by correlation at a central location. An extensive historical overview is presented in [116].

The radio astronomical community is currently building or developing several new radio telescopes for observations below 300 MHz, such as the Low Frequency Array (LOFAR) [15, 127], the Murchison Widefield Array (MWA) [72, 73] and the

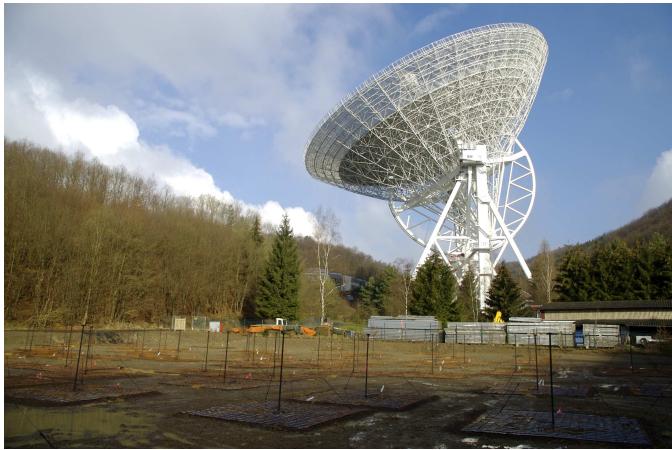


Figure 1.2: Photograph of the LOFAR antennas near the 100-m telescope in Effelsberg.

Long Wavelength Array (LWA) [33, 57]. LOFAR, designed to operate from 10 to 90 MHz and from 110 to 240 MHz, is under construction in the Netherlands with extensions throughout Europe. The MWA is being deployed in the Western Australian Desert and covers the frequency range from 80 to 300 MHz. Finally, the LWA is being developed in the United States to observe between 20 and 88 MHz, emphasizing the world wide interest in low frequency radio astronomy. Earth based observations of extraterrestrial radio waves at the lowest frequencies of the electromagnetic spectrum are hampered by the ionosphere, which becomes effectively opaque at frequencies below 10 MHz. Several studies have therefore investigated the feasibility of a radio astronomical facility on the Moon [35, 62, 154] or in space [9, 54]. The sensitivity, resolution and spectral flexibility of these new instruments will provide an unprecedented view at the low frequency universe.

1.1 Phased array technology

These instruments look vastly different from current radio telescopes as nicely demonstrated by a picture of the LOFAR field near the 100-m telescope in Effelsberg shown in Fig. 1.2. These instruments use phased array technology, which is new to the radio astronomical community. A phased array radio telescope consists of a large number of small antennas instead of a small number of large dishes. Each of these antennas

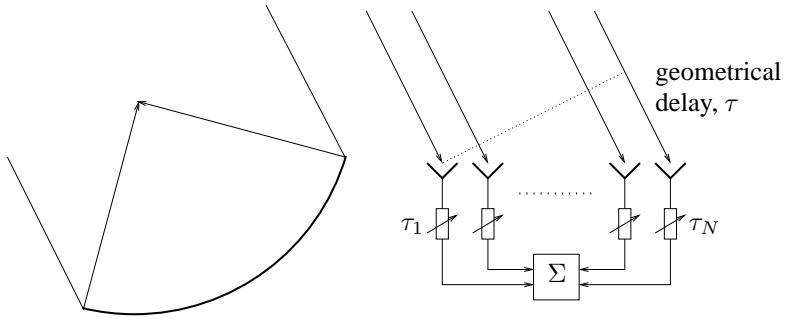


Figure 1.3: Schematic comparison between focusing of waves from a specific direction by a classical radio telescope (left) and a phased array telescope (right).

has a very large field-of-view (FOV), which means that they receive signals from all over the sky simultaneously. The array is focused on the signal of interest by combining the signals from many antennas after correction for the geometrical delay over the array. This is schematically compared to focusing of an incoming wave by a dish in Fig. 1.3.

Beam forming is the process of spatial filtering of the received signals to select the signal of interest. Beam forming may be implemented using either analog or digital electronics and aims to maximize the sensitivity of the array in the direction of interest while minimizing its sensitivity in other directions. Different optimization schemes have been developed for distinct constraints on the desired sensitivity pattern [23, 69, 122, 126] or to suppress radio frequency interference (RFI) [14, 19]. The sensitivity pattern of the array is called the array beam pattern or simply array beam. In a similar way, the sensitivity pattern of an antenna is referred to as antenna beam pattern or antenna beam.

We can map out the celestial sphere by scanning over the sky using a beam former. Another option is to correlate the signals from the antennas with each other. The correlation between the signals from a pair of antennas provides information on the spatial coherency of the incoming electromagnetic field on the spatial frequency corresponding to the baseline between the two antennas. By measuring these correlations, or visibilities, for all possible antenna pairs within the array, we thus sample the spatial coherency in an aperture described by all baseline vectors. These visibilities are the Fourier transform of the brightness distribution on the sky weighted by the antenna beam pattern [59, 116] and can therefore be used to construct an image of the sky. This process is called imaging. While a beam former can only apply one delay and one gain to each antenna at a time, an imaging routine, or imager, can assign one

delay and one gain to each visibility, i.e. to each antenna pair, and is therefore much more flexible. An imager can also optimize for all directions simultaneously which allows an imager to disentangle the source structure from the instrumental response, an operation called deconvolution.

The combination of antennas with a wide FOV and an electronic beam former or a correlator provides phased array telescopes with unprecedented capabilities. Depending on the system design, available bandwidth and available computing power, it is possible to form multiple beams simultaneously. This can be exploited to increase the instantaneous FOV of the telescope. If the signals from all antennas are correlated, an imager can produce snapshot images of their entire FOV. Ultimately, the telescope can be used as an all-sky monitor. Practical limitations will generally reduce either the bandwidth or the FOV or both. This will limit the all-sky monitoring capabilities which are attractive for the study of transient phenomena. Fortunately, if there is an appropriate transient detection scheme, a beam of a phased array telescope can be steered to a detected transient within seconds or even shorter time scales. This response time is much faster than the few minutes required by a classical dish telescope to slew to another part of the sky. Phased arrays are therefore potentially powerful instruments for the study of transient phenomena and generally provide a much larger instantaneous FOV than current radio telescopes.

This flexibility comes, of course, at a price. The sensitivity of a radio telescope is directly proportional to its collecting area. The Square Kilometre Array (SKA) [31, 44], the telescope of the future, is envisaged to be two orders of magnitude more sensitive than current radio telescopes. In optimal phased array designs, the collecting area per antenna roughly scales with wavelength squared. This implies that the number of antenna elements, which is closely related to the number of signal paths, increases with frequency squared. This makes phased arrays prohibitively expensive at frequencies higher than about 1500 MHz. Although the exact turnover frequency may increase as hardware costs decrease due to further advances in integrated circuit manufacturing, this sets the limit on the range of applicability of phased arrays for SKA. This is reflected in the specifications of the Electronic Multi-beam Radio Astronomy Concept (EMBRACE) [1, 2, 84] prototype which operates between 400 and 1550 MHz.

1.2 LOFAR

The author has been deeply involved in the development and verification of LOFAR, which is a nice example of a distributed phased array telescope. LOFAR is designed to operate between 10 and 240 MHz [127]. It was originally conceived as a survey telescope for frequencies below 100 MHz. A survey telescope is optimized to map out

large fractions of the sky efficiently. Its key figure of merit is its survey speed, which is proportional to the product of instantaneous FOV and sensitivity squared. At these frequencies, phased array technology provides the most cost effective way to create the large effective area and the large FOV needed to get at least an order of magnitude improvement in survey speed over existing radio telescopes [15]. The driving force behind the sensitivity requirement was the need to calibrate the distortions of the incoming radio waves by the ionosphere. Calibratability of the telescope has therefore been the key argument in the design of the LOFAR lay-out [17, 148, 155, 159]. The emphasis put on the calibration of the telescope is also reflected in the fact that the calibration studies have resulted in three Ph.D. theses. This thesis concentrates on station level signal processing, i.e. processing of the signals from individual receivers within a station. The other two, written by Sebastiaan van der Tol [121] and Huib Intema [51], concentrate on (ionospheric) calibration at array level, i.e. between stations (groups of antennas).

During early discussions on the LOFAR concept, people realized that an extension of this telescope to the 110 – 240 MHz frequency range would create an instrument that should be able to detect the very weak signals from the Epoch of Reionization (EoR). During this era, the first objects formed in the universe after the birth of the universe itself. These objects reionized the neutral hydrogen in their surroundings. With LOFAR, we should be able to detect this process by looking for a transition in the intensity of the 21 cm neutral hydrogen line at redshifts of about 10. Detection of these signals will put strict constraints on cosmological models describing how the structures we observe in the universe today were formed. This led to the introduction of an additional type of antenna to cover this frequency band.

An attractive feature of phased array telescopes is that their receiving elements are, in principle, omni-directional, which makes it possible, in principle, to build an all-sky monitor. This makes them very suitable instruments to search for transient phenomena and to study variable sources. A particular form of transient, which also led to LOFAR’s first scientific result by the LOFAR Prototype Station (LOPES) in Karlsruhe [34], are the radio flashes produced by ultra-high energy cosmic rays (UHECR) when colliding with particles in the Earth’s atmosphere or the surface of the Moon. These UHECRs consist of particles or γ -photons at energies over 10^{20} eV. Such particles cannot be produced by particle accelerators on Earth and may thus provide a new view on the most energetic processes in our universe.

The four science cases alluded to above (survey, EoR, transient and cosmic rays) led to a set of specifications that defined a very versatile instrument. LOFAR exploits two types of antennas, low band antennas (LBA) operating between 10 and 90 MHz and high band antennas (HBA) covering the 110 – 240 MHz frequency range. These antennas are grouped in stations. The antennas within each station are connected by coaxial cables to an on-site station backend containing receiver units (RCU) and

mode	input connector	frequency band (MHz)	ADC clock (MHz)
1	LBA-1	10 – 90	200
2	LBA-1	30 – 90	200
3	LBA-2	10 – 90	200
4	LBA-2	30 – 90	200
5	HBA	110 – 190	200
6	HBA	170 – 230	160
7	HBA	210 – 240	200

Table 1.1: Observing modes supported by the RCUs.

digital processing boards. The RCUs are equipped with three input connectors, two for a single polarization from an LBA and one for a single polarization from a HBA tile, several analog band pass filters and a 12-bit ADC that can either run at 160 MHz or at 200 MHz. A complete overview of supported observing modes is provided in Table 1.1. The digitized antenna signals are processed further in dedicated FPGA-based remote station processing (RSP) boards. In the standard processing pipeline the signals are split into 512 subbands, which are either 156 or 195 kHz wide depending on the ADC clock frequency, using a polyphase filter bank. After subband selection, 48 MHz total bandwidth is beam formed towards the direction of interest before the beam formed signal is sent over a 2 Gbit/s Ethernet link to the central processing facility (CEP) located in Groningen. At CEP each subband signal is split into 256 channels and correlated with the corresponding signals from other stations. This produces the raw visibility data for the LOFAR array which has 1 s time resolution and ~ 1 kHz frequency resolution.

The Dutch LOFAR stations have 96 LBAs, 48 HBA tiles and 96 RCUs. A tile is a compound element in which the signals from a 4×4 array of 16 HBAs are combined using an analog beam former. During standard observations, 48 LBAs or 48 HBA tiles will be selected by choosing the appropriate input channel of each RCU. The HBA tiles on the stations in the inner 1.5 km, the core area, are arranged in two subarrays of 24 HBA tiles instead of a single array of 48 HBA tiles. This allows each core station to emulate two 24-tile stations instead of a single 48-tile station. This increases the number of unique baseline vectors, and thereby the amount of instantaneously available spatial frequency information, considerably. Figure 1.4 shows an aerial photograph of a core station. The Dutch LOFAR stations support baselines up to 150 km. Collaboration with foreign institutes has led to extensions of



Figure 1.4: Aerial photograph of a LOFAR core station.

LOFAR throughout Europe with stations in Germany, the United Kingdom, France and Sweden. This is nicely illustrated by Fig. 1.2 showing the LBA station in front of the 100-m telescope in Effelsberg. These stations will have 96 LBAs and 96 HBA tiles. They will also have a backend that supports 192 signal paths, which allows all 96 LBAs or all 96 HBA tiles to be used at the same time.

The RSP boards can also produce a number of statistics of the incoming data. The real-time station correlator can correlate all input channels for a single subband. This data is used by the on-line station calibration process that will be discussed in Sec. 2.3. The correlated data can also be stored on the local control unit (LCU) for further off-line processing. This option has been used extensively to obtain data for validation of the calibration and imaging approaches discussed in this thesis. The RSP boards can also provide real-time autocorrelation spectra of all input channels and the beam former output, which is very helpful for system health monitoring.

1.3 Hierarchical system architecture

The Dutch LOFAR system is envisaged to consist of at least 36 stations. This implies that the full system consists of ~ 3500 LBAs and ~ 28000 HBAs. It is too costly to sample and process the signal from every individual antenna. LOFAR therefore

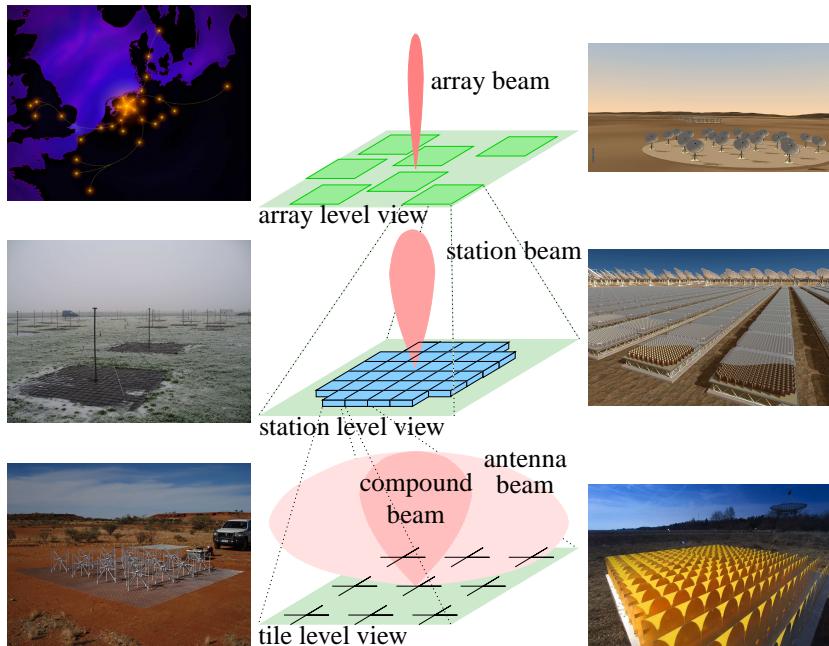


Figure 1.5: (*Center column*) The beam forming hierarchy with the array beam produced by an array of stations at the top and the antenna beam at the bottom. Subsequent levels in the hierarchy have beams that are narrower and more sensitive. (*Left column*) the corresponding concept layout of LOFAR [70], a LOFAR low band antenna station (photograph by Menno Norden) and an MWA [77] tile. (*Right column*) a concept for SKA consisting of an array of stations, each with small dishes [105], a concept for the SKA core station [105], and a SKA demonstrator tile consisting of Vivaldi antennas [4].

has a hierarchical system architecture that reduces the number of signal paths at each level to order 100: 96 single polarization antenna or tile signals are combined to produce the dual polarized station output and the resulting 36 dual polarization station signals are processed further at CEP. The signals from 16 HBAs are combined by an analog beam former in a tile, because the cost per unit collecting area would become unacceptably high if each HBA would be directly connected to an RCU. This would require 16 times as many RCUs in the LOFAR system with a corresponding increase in other components like cables, power supplies and digital processing boards.

This system design leads to a hierarchy of beams with increasingly smaller FOV

but higher sensitivity. This hierarchy starts with the antenna beam pattern. The analog tile beam former then forms a tile beam, the station beam former combines the signals from all tiles within a station to form a station beam, whose output is combined with the beam formed signals from other stations to form the array beam of the entire LOFAR array. This is an important conceptual difference with current telescopes using dishes. Those instruments normally have only two beams: the primary beam, i.e. the beam of a single dish, and the array beam. The station beam can be regarded as the single dish beam of classical radio telescopes, but is far more flexible because it is not the base of the beam forming hierarchy. This beam forming hierarchy is illustrated in Fig. 1.5 with examples from LOFAR, MWA and SKA.

1.4 Scope of this thesis

The main question of this thesis is “*What are the possibilities and limitations of imaging with a phased array station if one has access to the individual elements and how can this be optimized?*”. There are two ways to look at the first part of this question. From a fundamental perspective, one may consider a perfectly calibrated phased array and ask how well such an instrument would be able to discern sources spatially. Ideally, the station should achieve high-fidelity, noise limited imaging at diffraction limited resolution. Although the theoretical imaging performance may not be achieved in practice, imagining such an instrument may provide insight in the way a phased array works and how one could optimize it by design. From a practical perspective, one may ask how the station should be calibrated and how the calibration errors affect the station beam pattern and thereby its ability to discern sources spatially.

Calibration and imaging are the key aspects. Both of them pose new challenges due to the large FOV of the individual elements. The half power beam width (HPBW) of such an element is about 60 degrees and it can easily detect sources close to the horizon. Observing with a station can therefore be aptly called fish-eye observing. This has a number of important consequences from a calibration perspective. First of all, even initial calibration should be done using an all-sky source model that may contain very complicated source structures like the extended diffuse emission from our own Galaxy, since the antennas cannot be steered to a single calibrator. Secondly, the propagation conditions may vary over the FOV causing direction dependent perturbations of the source model. This naturally leads to the following research questions regarding *calibration*:

- How can we describe the calibration problem mathematically and what is an appropriate parameterization?
- How can we solve this problem in a computationally efficient way?

- What is the numerical complexity of this algorithm?
- What is the theoretical accuracy of the calibration parameter estimates?
- Can we achieve this accuracy in practice?
- What is the impact of calibration errors on the beam forming or imaging capabilities of the station?

In many imaging routines it is assumed that the FOV is sufficiently small to approximate the part of the celestial sphere within the FOV by a plane. This assumption is made to reduce the numerical complexity of the imaging algorithm, but it obviously does not hold for fish-eye observations. The fact that the station beam pattern varies over the FOV of the individual elements, over time and over frequency and that the propagation conditions may vary as well, implies that the imaging routine has to reconstruct the source structure while the direction dependent gain towards each source is varying over time, frequency and direction. We therefore need to answer the following research questions regarding *imaging*:

- How can we reconstruct the source structure exploiting knowledge of the propagation conditions and the instrument over time, frequency and space from, e.g., the calibration?
- What is the maximum achievable dynamic range in such an image?
- What is the numerical complexity of the algorithm?
- How close can we get to the theoretical limits in practice?

The phrase “if one has access to the individual receiving elements” was added to the main research question to emphasize that the results reported in this thesis were obtained under this assumption. Although most of this work is directly applicable to stations of compound elements, such as LOFAR’s HBA tiles, this analysis lacks a proper calibration strategy for the individual antennas within the compound element. Knowledge of this strategy is required to compute the calibration errors on the individual elements which ultimately affect the station beam as well. However, once these errors are determined, the framework developed in this thesis can be applied to stations of compound elements as well. This implies that the analysis in this thesis can also be applied to higher levels in the beam forming hierarchy.

1.5 Outline and summary of main results

The research presented in this thesis has been done in close collaboration with the LOFAR team. This interaction provided the background for my research questions and the unique opportunity to apply and validate the results on actual LOFAR data. Chapter 2 starts with a presentation of a few important historical milestones. I describe how the calibration algorithm described in Ch. 5 was integrated in the on-line

LOFAR station calibration process and demonstrate that the algorithm still finds reasonable solutions for the instrumental parameters under conditions of strong ionospheric scintillation. Chapter 2 also provides a number of imaging examples including time and frequency synthesis imaging to demonstrate the full range of situations to which the results from Ch. 6 may be applied. One of the imaging experiments also demonstrates that the proposed algorithm is able to reach the Cramer-Rao bound on actual data. The Cramer-Rao bound is a lower bound on the total variance of the estimated parameters when using an unbiased estimator.

These results can only be obtained using a proper model of the antenna signals. The output signals of the receiving elements are the result of the actual electromagnetic field incident on the array, the instrumental response of the individual antennas and noise. Chapter 3 starts with an electromagnetic description of an antenna array to derive a general data model or measurement equation for an aperture array radio telescope. This model is compared to models commonly exploited in the array signal processing literature to demonstrate that many results in the literature are applicable to aperture array radio telescopes as well. This chapter will also demonstrate how the data model can be extended to synthesis observations in which multiple snapshot observations are combined to improve the calibration and imaging results.

The Cramer-Rao bound puts a fundamental limit on the parameter estimation accuracy. It is therefore used to assess the statistical efficiency of the algorithms proposed in this thesis. Chapter 4 derives all the results required for this assessment in other chapters. The results confirm two properties of the estimation process that can be intuitively understood but had not been explicitly derived earlier in the literature. The first result is the impact of the boundary condition imposed to provide a phase reference on the variance of the phases of the direction independent receiver path gains. The variance of the phase reference is added to the variance of phase estimates for the individual receiver paths. Therefore a phase reference based on the average phase of all receiver paths provides a lower total variance than one based on a single element [151]. The second result is that the variance on the direction independent gain estimates is the same regardless whether a particular entry of the array covariance matrix is affected by an additive nuisance parameter that has to be included in the estimation process or it is simply ignored [119]. This result can be understood intuitively by regarding the matrix equation describing the data model as a set of linear equations. This provides a convenient tool to mark specific entries of the array covariance matrix as bad data in a Cramer-Rao bound analysis or in an estimation problem.

Source modeling for calibration of aperture array radio telescopes is complicated by the complex source structure on the sky, which consists of a very large number of bright discrete sources and extended diffuse emission from our own Galaxy. The single source array calibration method proposed in [12] is therefore extended to han-

dle multiple sources in Ch. 5. It is demonstrated that this method is asymptotically statistically efficient by comparing the results from Monte Carlo simulations with the Cramer-Rao bound, while the numerical complexity is similar to that for less statistically efficient algorithms in the literature [149]. Using a weighted alternating least squares approach, this calibration method is extended to handle direction dependent effects such as apparent variations in the calibration source powers and small shifts in their positions [165] as well as correlated noise [164]. With these extensions, the algorithm can handle modest ionospheric scintillation and extended emission from large scale source structures like the Galactic plane, which proves to be a necessity in the analysis of data from a LOFAR LBA station as demonstrated in Ch. 2.

In Ch. 6 an analytic expression describing the relation between the true brightness distribution on the sky and the correlations between the received signals is used to derive an expression for the covariance of the image values due to propagated calibration errors. Together with the source confusion and the noise in the measurements, this contributes to the effective noise in the image [161]. Although these results are derived for snapshot observations, the framework is suitable for an analysis of synthesis observations as well, since the data model for a synthesis observation has the same mathematical form as demonstrated in Ch. 3. One of the fundamental challenges of synthesis imaging with an array of aperture array stations is that the station beams will change considerably more than the primary beams of current radio telescopes during an observation. The analytic least squares solution to the imaging problem handles this naturally, as will be mathematically demonstrated in Ch. 6.

In Ch. 7 the results are summarized and linked to the research questions presented in Sec. 1.4. The main contribution of this study, however, may be that it consistently uses methods from the array signal processing literature independent from the algorithms in todays radio astronomical data reduction packages. This is a fundamentally new way of looking at the data that gradually attracts more attention from people working in the field and should provide the basis for the next generation of self-calibration and imaging routines. However, the reduction of data from synthesis observations still requires answers to many fundamental and practical questions. This provides room for further research, which will be discussed in Ch. 7 as well.

Three appendices have been added to this thesis for convenience of the reader. Appendix A provides a list of used acronyms while App. B explains the notational conventions used in equations throughout this thesis. The derivations often exploit properties of special matrix products, such as the element-wise or Hadamard product, the Kronecker product and the Khatri-Rao or column-wise Kronecker product, that may not be familiar to most readers. These properties are listed in App. C.

1.6 Publications and other contributions

The work done for this thesis has led to the following publications, conference contributions and internal reports. Interested readers should feel free to contact the author about work that is not publicly available.

International journal papers

- Stefan J. Wijnholds, Jaap D. Bregman and Alle-Jan van der Veen. LOFAR Station Calibration. *Experimental Astronomy*, 2010, in preparation (invited paper).
- J.D. Bregman et al. LOFAR Synthesis Array and Station Configuration. *Experimental Astronomy*, 2010, in preparation (invited paper).
- Stefan J. Wijnholds, Sebastiaan van der Tol, Ronald Nijboer and Alle-Jan van der Veen. Calibration Challenges for Future Radio Telescopes. *IEEE Signal Processing Magazine*, 27(1):30-42, January 2010.
- Stefan J. Wijnholds and Alle-Jan van der Veen. Multisource Self-calibration for Sensor Arrays. *IEEE Transactions on Signal Processing*, 57(9):3512-3522, September 2009.
- Stefan J. Wijnholds and Alle-Jan van der Veen. Fundamental Imaging Limits of Radio Telescope Arrays. *IEEE Journal of Selected Topics in Signal Processing*, 5(2):613-623, October 2008.
- Stefan J. Wijnholds and Alle-Jan van der Veen. Effects of Parametric Constraints on the CRLB in Gain and Phase Estimation Problems. *IEEE Signal Processing Letters*, 13(10):620-623, October 2006.
- H. Falcke et al. (LOPES collaboration). Detection and imaging of atmospheric radio flashes from cosmic ray air showers. *Nature*, 435:313-316, 19 May 2005.
- Stefan J. Wijnholds, Jaap D. Bregman and Albert-Jan Boonstra. Sky Noise Limited Snapshot Imaging in the Presence of RFI with LOFAR's Initial Test Station. *Experimental Astronomy*, 17(1-3):35:42, 2004.

Refereed international conference papers

- Stefan J. Wijnholds and Alle-Jan van der Veen. Self-Calibration of Radio Astronomical Arrays with Non-Diagonal Noise Covariance Matrix. In *Proceedings of the 17th European Signal Processing Conference (EuSiPCo)*, Glasgow, United Kingdom, 24-28 August 2009.
- Stefan J. Wijnholds. Fish-Eye Imaging With A LOFAR Station. In *Proceedings of the International Union of Radio Science (URSI) Benelux Forum 2009*, Delft, The Netherlands, 8 June 2009.

- Stefan J. Wijnholds. Autonomous Online LOFAR Station Calibration. In *Proceedings of the XXIXth General Assembly of the International Union of Radio Science (URSI GA)*, Chicago (Ill.), USA, 7-16 August 2008.
- Stefan J. Wijnholds. Calibration of Phased Array Radio Telescopes. In *Proceedings of the International Union of Radio Science (URSI) Benelux Forum*, Brussels, Belgium, 30 May 2008, invited paper.
- Stefan J. Wijnholds and Albert-Jan Boonstra. A Multisource Calibration Method for Phased Array Radio Telescopes. In *Proceedings of the fourth IEEE workshop on Sensor Array and Multichannel Processing (SAM)*, Waltham (Mass.), USA, 12-14 July 2006.
- Sebastiaan van der Tol and Stefan J. Wijnholds. CRB Analysis of the Impact of Unknown Receiver Noise in Phased Array Calibration. In *Proceedings of the Fourth IEEE workshop on Sensor Array and Multichannel Processing (SAM)*, Waltham (Mass.), USA, 12-14 July 2006.
- Stefan J. Wijnholds. Confusion Limited All Sky Imaging With LOFAR's Initial Test Station Applying Wide Field Calibration Techniques. In *Proceedings of the XXVIIIth General Assembly of the International Union of Radio Science (URSI GA)*, New Delhi, India, 23-29 October 2005.
- A.J. Boonstra, S.J. Wijnholds, S. van der Tol and B. Jeffs. Calibration, Sensitivity and RFI Mitigation Requirements for LOFAR. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia (Penn.), USA, 18-23 March 2005.

Other conference contributions

- Stefan J. Wijnholds. Least Squares All-Sky Imaging With A LOFAR Station. In *SKA Calibration and Imaging Workshop (CalIm)*, Socorro (New Mexico), USA, 30 March - 3 April 2009.
- Stefan J. Wijnholds. Autonomous Online LOFAR Station Calibration - Methods and Results. In *SKA Calibration and Imaging Workshop (CalIm)*, Perth, Australia, 7-9 April 2008.
- Stefan J. Wijnholds. Effective Noise. In *SKA Calibration and Imaging Workshop (CalIm)*, Perth, Australia, 7-9 April 2008.
- Stefan J. Wijnholds. Reducing the Impact of Station Level Spatial Filtering Limitations. In *SKA Calibration and Imaging Workshop (CalIm)*, Cape Town, South Africa, 4-6 December 2006.
- Stefan J. Wijnholds. Self-noise in Full Sky LOFAR Images. In *Dutch Astronomers Conference (NAC)*, Hollum, The Netherlands, 10-12 May 2006.

Internal reports and memos

- Stefan J. Wijnholds. Mutual Coupling, Inter-tile Spacing and Inter-station Rotation in the HBA array. Technical Report LOFAR-ASTRON-MEM-245, ASTRON, Dwingeloo, The Netherlands, 6 August 2008.
- Stefan J. Wijnholds. LBA station configuration. Technical Report LOFAR-ASTRON-MEM-241, ASTRON, Dwingeloo, The Netherlands, 4 December 2007.
- Stefan J. Wijnholds. CS1 system validation. Technical Report LOFAR-ASTRON-RPT-124, ASTRON, Dwingeloo, The Netherlands, 10 April 2007.
- Stefan J. Wijnholds. Station Calibration - Pitfalls & Possibilities 2006. Technical Report LOFAR-ASTRON-MEM-217, ASTRON, Dwingeloo, The Netherlands, 13 October 2006.
- Stefan J. Wijnholds. Study on LOFAR Core Calibratability Based on Source Statistics. Technical Report LOFAR-ASTRON-MEM-205, ASTRON, Dwingeloo, The Netherlands, 20 June 2006.
- Stefan J. Wijnholds. Statistical Performance Analysis of Station Calibration Algorithms. Technical Report LOFAR-ASTRON-MEM-188, ASTRON, Dwingeloo, The Netherlands, 8 November 2005.
- Stefan J. Wijnholds. ITS Results 2005 - Filling the gaps. Technical Report LOFAR-ASTRON-RPT-056, ASTRON, Dwingeloo, The Netherlands, 23 September 2005.
- Stefan J. Wijnholds. Theory of Polarized Imaging With Phased Array Antennas. Technical Report LOFAR-ASTRON-MEM-183, ASTRON, Dwingeloo, The Netherlands, 26 July 2005.
- Stefan J. Wijnholds. One year in the bushed - Results from ITS 2004. Technical Report LOFAR-ASTRON-RPT-052, ASTRON, Dwingeloo, The Netherlands, 22 February 2005.
- Stefan J. Wijnholds. Theoretical Analysis of the Main Lobe Effects of Beam Squint. Technical Report LOFAR-ASTRON-MEM-151, ASTRON, Dwingeloo, The Netherlands, 14 June 2004.
- Stefan J. Wijnholds. THETA experimental results. Technical Report LOFAR-ASTRON-MEM-120, ASTRON, Dwingeloo, The Netherlands, 10 March 2004.

Chapter 2

Calibration and imaging with LOFAR

This thesis is the result of strong interaction with the LOFAR team. The work on system integration, validation and design has led to insight in the calibration and imaging issues that LOFAR had to face and that needed to be resolved to get this new radio telescope operational. On the other hand the mathematical framework provided by the signal processing community helped to make an in-depth analysis of these issues and to find solutions to a number of challenging problems. This interplay led to a number of interesting results. This chapter provides a historical overview of some major project milestones achieved with early prototypes, in which I have been involved, such as LOFAR’s first all-sky survey [145] and the successful detection of ultra-high energy cosmic rays [34]. This chapter also demonstrates the applicability of the ideas and methods developed in later chapters to actual LOFAR data. These demonstrations include successful calibration of a LOFAR station during a period of strong ionospheric turbulence and statistically efficient imaging performance on actual data [162].

After an introduction of the coordinate systems used to present the results, a few successes from early LOFAR prototypes are presented. The calibration approach described in Ch. 5 is the most noticeable contribution of this thesis project to LOFAR, since it has been implemented on the LCU to perform the unsupervised station calibration. This pipeline is discussed in detail in Sec. 2.3. The least squares imaging approach discussed in Ch. 6 is well suited for an all-sky monitor with either a LOFAR station or the central core.

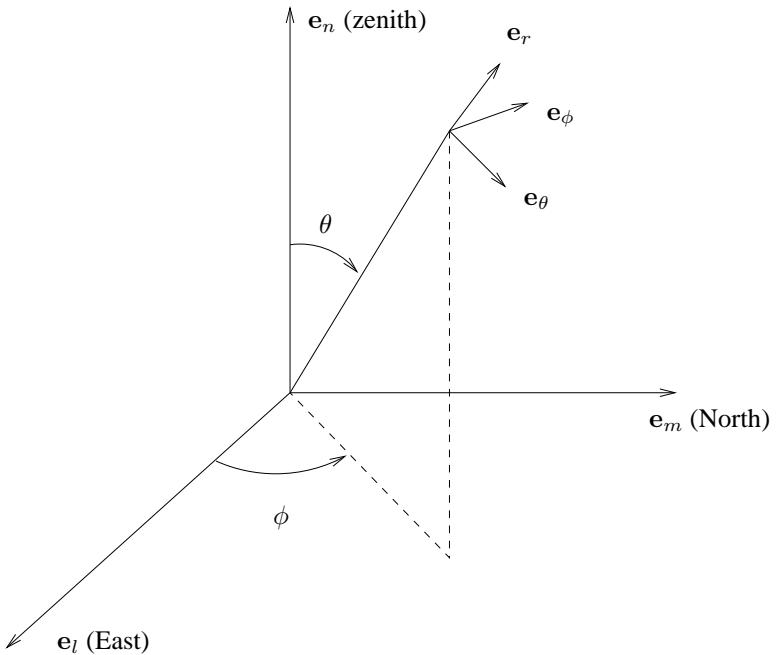


Figure 2.1: Coordinate systems to describe the direction of arrival along the unit vectors \mathbf{e}_l , \mathbf{e}_m and \mathbf{e}_n or using the angles θ and ϕ and polarization of the electromagnetic waves incident on the antenna array in components along the unit vectors \mathbf{e}_r , \mathbf{e}_θ and \mathbf{e}_ϕ .

2.1 Coordinate systems

2.1.1 Direction of arrival

Sky catalogs, like the third Cambridge catalog of radio sources [8], or 3C catalog for short, generally specify the positions of astronomical sources either in right ascension α and declination δ , a system of longitude and latitude using the ecliptic as equatorial plane, or in Galactic longitude and latitude, a system which uses the Galactic plane as equatorial plane. These coordinates need to be converted to azimuth az and elevation el at the location of the telescope at the time of observation. Azimuth is the azimuthal angle running from North (0°) through East (90°) along the horizon while the elevation is measured from the horizon (0°) positive towards the zenith (90°).

The literature on antenna design and engineering uses a spherical coordinate sys-

tem fixed to the antenna. The direction from which radiation is received or in which radiation is transmitted is specified in terms of an azimuthal angle ϕ and the angle from bore sight, the point straight in front of the antenna, θ . For a phased array placed on a horizontal surface, bore sight corresponds to the zenith. It is then only a minor step to fix the azimuthal angle to the quarters of the compass, as depicted in Fig. 2.1. It follows directly from these definitions that the θ and ϕ are related to azimuth and elevation by

$$\theta = 90^\circ - el \quad (2.1)$$

$$\phi = 90^\circ - az \quad (2.2)$$

The field-of-view (FOV) of most radio telescopes is sufficiently small to project the observed part of the celestial sphere on a plane touching the celestial sphere in the center of the FOV without introducing too much distortion in the imaging process. The position vectors projected on this plane are generally denoted by l pointing towards East and m pointing to the North pole of the chosen spherical coordinate system. Such a coordinate system consisting of directional cosines turns out to be highly convenient for phased array telescopes, so we define

$$l = \sin \theta \cos \phi = \cos(el) \sin(az) \quad (2.3)$$

$$m = \sin \theta \sin \phi = \cos(el) \cos(az) \quad (2.4)$$

$$n = \cos \theta = \sin(el) \quad (2.5)$$

Thus, referring to Fig. 2.1, $\mathbf{e}_r = l\mathbf{e}_l + m\mathbf{e}_m + n\mathbf{e}_n$. Since $n = \sqrt{1 - l^2 - m^2}$, l and m suffice to specify a point $\mathbf{l} = [l, m, n]^T$ on the unit sphere. These definitions are chosen such that l is pointing East and m is pointing North thereby following the common practice in radio astronomy. Although this coordinate system was defined to conveniently express positions within small FOVs, we will see that with appropriate modifications in the underlying data reduction algorithms it can be applied to extremely wide FOVs up to all sky observations.

2.1.2 Polarization

Figure 2.1 also shows the mutually orthogonal unit vectors \mathbf{e}_r , \mathbf{e}_θ and \mathbf{e}_ϕ . These can be used to specify the polarization of the incoming radiation. \mathbf{e}_r is oriented parallel to the direction of arrival, while \mathbf{e}_θ and \mathbf{e}_ϕ lie in a plane perpendicular to the direction of arrival pointing in the direction of increasing θ and increasing ϕ respectively. For a plane wave, the electromagnetic field is always perpendicular to the direction of propagation. Therefore \mathbf{e}_θ and \mathbf{e}_ϕ suffice to characterize the polarization of an incoming plane wave.

2.1.3 Antenna locations

The position of the receiving elements within a phased array station can be specified using Cartesian coordinates $\xi = [x, y, z]^T$, where the positive x -axis is pointing East, the positive y -axis is pointing North and the positive z -axis is pointing towards the zenith, i.e. $\mathbf{e}_x \parallel \mathbf{e}_l$, $\mathbf{e}_y \parallel \mathbf{e}_m$ and $\mathbf{e}_z \parallel \mathbf{e}_n$. As explained in Sec. 1.1, the directions of arrival of the incoming signals are estimated by measuring the phase difference between pairs of receiving elements at a specific observing frequency. These phase differences are measured by correlating the output signals of the receivers. In radio astronomy, these correlations are called visibilities. Since visibilities are related to pairs of receiving elements, they are associated with a so called baseline vector pointing from the i th element to the j th element of the array $\mathbf{u}_{ij} = [u, v, w]^T = \xi_j - \xi_i$. Each baseline vector describes a point in the visibility space with axis denoted by u , v and w such that $\mathbf{e}_u \parallel \mathbf{e}_x$, $\mathbf{e}_v \parallel \mathbf{e}_y$ and $\mathbf{e}_w \parallel \mathbf{e}_z$.

2.2 Highlights from early prototypes

2.2.1 The Ten Heterogeneous Element Test Array¹

System setup

The Ten Heterogeneous Element Test Array (THETA) was built during the summer of 2003 for verification of the system concept. It consisted of 10 inverted V-shaped dipoles arranged in a Y-configuration, each leg being a 4-element non-redundant linear array. Each dipole was connected to an RCU with a 40 – 80 MHz band pass filter with 60 dB stopband attenuation and 6 MHz wide transition bands from 37 to 43 MHz and from 77 to 83 MHz. The filtered signals were sampled by a 12-bit 80-MHz ADC and transmitted over a high speed optical link to a twin input module (TIM) board in one of the backend PCs.

A TIM board [86] is a PCI card equipped with two high speed link receivers and 2 Gbyte of RAM. Since the 12-bit ADC samples were stored as 16-bit numbers for easy handling by the data acquisition PCs, each TIM board was capable of buffering 6.4 s of data from both its high speed link receivers. These data could be processed on a small cluster consisting of four data acquisition machines and a central processing and storage machine interconnected by 1 Gbit/s Ethernet interfaces.

¹The material in this section was taken from [138]

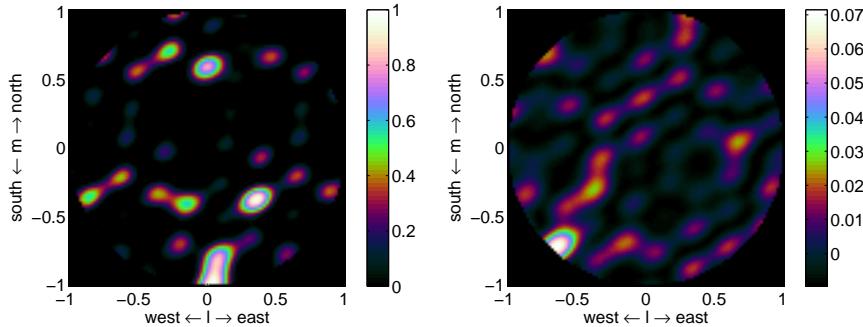


Figure 2.2: Full sky images taken on 28 October 2003 at 12:04 (left) and 15:04 (right) local time showing a clearly detected coronal mass ejection. Three hours after the initial detection at 12:04 local time, the intensity of the sun has decreased to about 7% of the original power of the CME.

Detection of a coronal mass ejection

During a 24-hour measurement campaign started on October 28, 2003 at 9:04 local time, a coronal mass ejection (CME) was detected by THETA over its full operational frequency range. The system was set up to simultaneously capture 50 ms of data for all antennas once every hour. The data was split into 4096 frequency channels using a non-overlapping 8192-point FFT. 100 consecutive frequency channels were selected in the RFI free region around 46.3 MHz. For each channel, the instrumental phases were determined based on the first row of the correlation matrix assuming that the sun was the single dominant source in the sky and that the antennas were identical. The more advanced calibration algorithms presented in Ch. 5 were not available at that time. Based on this rudimentary calibration, the correlation matrix for each channel was corrected for the instrumental gain phases. After forming an image of the sky for each frequency channel using simple Fourier transform imaging (see Ch. 6), the 100 images were added. Fig. 2.2 shows the result for the snapshots taken at 12:04 and 15:04 local time.

The image of the sun is very suitable for calibration based on a single point source since the sun's diameter of about 0.5° is small compared to the resolution of the THETA system, which is about 10° at 46.3 MHz. A lot of side lobe structure is still visible in the image as expected from the array geometry. This side lobe structure is frequency dependent. It can therefore be suppressed further by averaging over a larger frequency range. Side lobe structure can also be removed from the image by

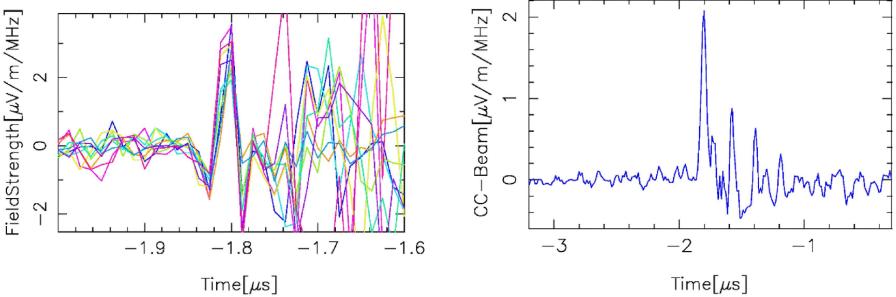


Figure 2.3: Detection of a cosmic ray event. The left panel shows the electric field detected by the individual dipoles after correcting for the instrumental and geometrical delays towards the air shower. The right panel shows the resulting signal after beam forming towards the air shower. These results were originally published in [34].

disentangling the source structure and the instrumental response in a process called deconvolution. This is done implicitly by the least squares imaging method discussed in Ch. 6 as demonstrated by the full sky imaging results presented in Sec. 2.4

2.2.2 The LOFAR Prototype Station

Using hardware identical to the THETA hardware, the 10-element LOFAR Prototype Station (LOPES) was deployed co-located with the Karlsruhe Shower Core and Array Detector (KASCADE). In this experiment, the TIM boards were used as cyclic buffer for temporary data storage until an interrupt was generated by the KASCADE instrument. The combination of the two instruments provided a facility that cannot only detect the air shower of secondary particles produced by ultra-high energy cosmic rays (UHECR) when they collide with particles in the Earth's atmosphere, but also the radio waves produced when the charged particles in the air shower are deflected by the Earth's magnetic field. Even after RFI removal, the usable bandwidth was still 33 MHz allowing a time resolution of 30 ns, which was far better than the $\sim 1 \mu\text{s}$ in historic experiments. The spatial resolution was about 2° , which was an order of magnitude better than previous experiments.

The first results were presented in 2005 [34] and indicate that low frequency radio telescopes like LOFAR have sufficient spatial and temporal resolution to provide complementary data for experiments with particle detector arrays. A typical event is shown in Fig. 2.3. The instrumental and geometrical delays were corrected by shifting the time series data of the dipole followed by a frequency dependent phase

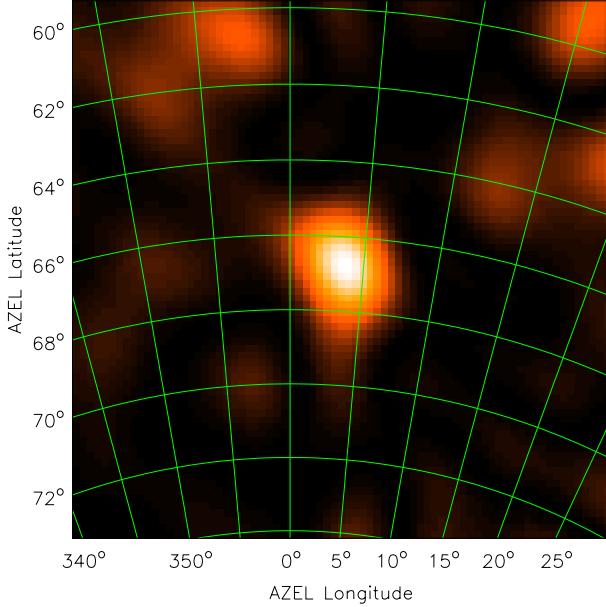


Figure 2.4: Image of an air shower made with the LOPES array [34].

rotation. The result is shown in the left panel. The signal is almost fully coherent at $-1.8 \mu\text{s}$, the arrival time of the shower. Thereafter, the signal is dominated by radio emission produced by the photomultipliers of the particle detectors. This interference is incoherent as demonstrated in the right panel of Fig. 2.3 showing the beam formed signal from the direction of the air shower. As a result, the interference from the photomultipliers is strongly reduced and the detected cosmic ray event clearly stands out.

Although a cosmic ray event lasts for only a few nanoseconds, the SNR of events like the one shown in Fig. 2.3 is sufficiently high to image the air shower. This is demonstrated in Fig. 2.4. This image in azimuth and elevation (North is to the top, East to the right) was made using a near field imager focused at an altitude of 2000 m, i.e. with a fixed radius of curvature for each pixel. The cosmic ray event is seen as the bright blob in the center of the image. The “noise” in the map is mainly caused by the side lobes of the LOPES array sensitivity pattern. No image deconvolution has been applied to remove the instrumental response.

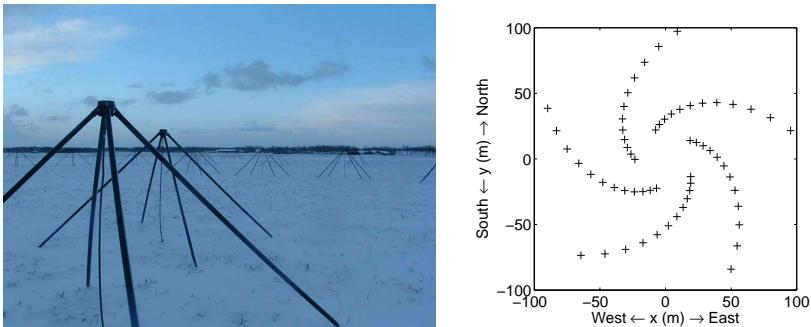


Figure 2.5: Photo showing the ITS antennas (left) and the ITS lay-out (right).

2.2.3 The Initial Test Station

The initial test station (ITS), which was built in December 2003, was the first full scale prototype of a LOFAR station. It consisted of 60 sky noise limited [140] inverted V-shaped dipoles with East-West orientation arranged in a five-armed spiral configuration. Figure 2.5 shows a photograph of the ITS dipoles and the ITS lay-out. ITS offered an instantaneous synthesized aperture of almost 200 m diameter.

The system architecture of ITS was similar to that of THETA and LOPES. Each dipole was connected to an RCU by a coaxial cable. In the RCUs, the signals were filtered by a 10 – 40 MHz band pass filter and digitized by a 12-bit 80-MHz ADC. The digitized signals were transported over optical high speed links to the 32 TIM boards in the 16 data acquisition PCs (2.4 GHz Intel Xeon, 1 Gbyte RAM). The data acquisition PCs could also be used for first stage processing of the data, such as the application of a Fourier transform or polyphase filter bank to obtain spectral information. The raw or pre-processed data could be sent over a 1 Gbit/s Ethernet data network to a central processing and storage machine (dual 2.0 GHz Intel Xeon, 512 Mbyte RAM).

Being the first full scale prototype, much practical experience on roll-out of a station and system testing was gained [142]. ITS also has been instrumental in shaping our ideas on station processing, such as RFI detection and mitigation, station calibration and treatment of polarized signals [142, 144]. Here, I will focus on the single result that appeared prominently in many LOFAR presentations for almost three years: the first LOFAR all-sky map observed at 30 MHz. This map, shown in Fig. 2.6, was officially presented in [145]. The discussion below is based on that paper.

Between September 24 and October 9, 2004 a total of 86 hemispheric 6.7 s snapshot observations were done with ITS. A non-overlapping 8192-point FFT with Han-

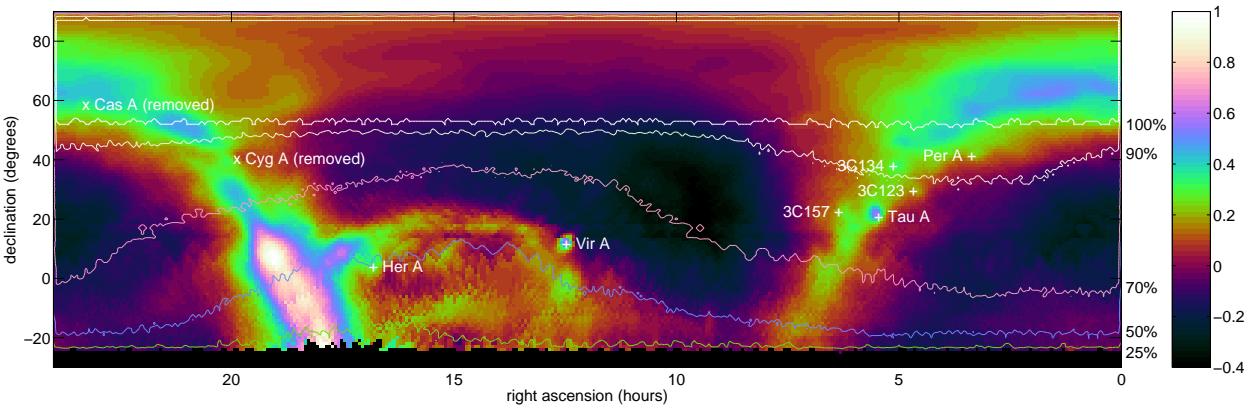


Figure 2.6: This image shows LOFAR’s first all-sky map, which is based on 56 6.7 s hemispheric snapshot observations using all RFI free 9.77 kHz frequency channels between 29.5 and 30.5 MHz. Cas A and Cyg A have been removed to avoid contamination of the weak sources by their side lobes. The contours indicate the distribution of the square root of the product of total bandwidth and integration time over all snapshots as percentage of the maximum value of $5.7 \cdot 10^3$.

ning window was applied to the data to produce spectra with 9.77 kHz frequency bins. Only the channels within the pass band of the analog filter (channels 1050 through 3900) were correlated to reduce the computational load. From this data, all RFI free channels between 29.5 and 30.5 MHz were selected for further processing using a median filter of 51 channels around the channel tested for RFI. Snapshots with less than 50 RFI free channels in this frequency range were rejected for further processing leaving 56 snapshots.

The data were calibrated using the method described in [13] with a sky model consisting of four point sources: Cassiopeia A (Cas A), Cygnus A (Cyg A), Virgo A (Vir A) and Taurus A (Tau A). The locations and source powers were assumed known. The method used is based on an element-wise division of the measured covariance matrix by the covariance matrix predicted based on the source model. It is therefore not as statistically optimal as the method presented in Ch. 5, which was used to calibrate later prototype stations and will be used to calibrate the stations in the LOFAR production system. There is always some extended emission in snapshot images of the sky above a station due to the galactic plane and galactic loops like the north polar spur. This emission is hard to model with a discrete point source model. Therefore extended emission was filtered out by only using baselines longer than four wavelengths during calibration.

After successful calibration of the individual frequency channels, the result produced by Fourier transform imaging depends on the distribution of sources within the field-of-view and the array beam pattern. The latter may cause contamination of weak sources in the field by the side lobes of strong sources. If the position and power of a strong source are known, it can be subtracted from the measured visibilities. This was used to remove the signals from Cas A and Cyg A from the individual frequency channels before integrating over all selected channels of the respective snapshots. These integrated snapshot images were projected on an (α, δ) -grid with 1° resolution and averaged weighted by the square root of bandwidth and integration time of the snapshot observation. The result is shown in Fig. 2.6.

The contours indicate the distribution of the square root of the product of total bandwidth and integration time over all snapshots as percentage of the maximum value of $5.7 \cdot 10^3$. This is directly related to the sensitivity of the measurement. At the time of observation, the sources at hour angles between 10 h and 20 h were observed during day time. Since more RFI is detected during day time than during night time, the usable number of frequency channels during day time is smaller, which reduces the SNR achieved in those measurements. The gradient with declination is easily explained by the fact that sources at lower declination are observable over a shorter period of time. This restricts the achieved integration time for sources at low declinations.

The extended emission from the Galactic plane and the north polar spur is clearly

visible in the image thanks to the considerable number of short baselines in the ITS configuration. A number of well-known sources is indicated. The vague blob below Virgo A is caused by the sun which appeared at a slightly different position on the sky w.r.t. the stars on each consecutive day during the observing run.

The resolution of a phased array telescope varies with elevation. Since the elevation at transit is declination dependent, this results in a varying resolution over an all-sky map. Most sources, however, are observed with an effective resolution of about 4° . With such a resolution about 260 independent pointings per steradian can be defined. To detect individual sources separately, i.e. uncluttered, there should be at least 10 resolution elements per source in the image to avoid source confusion. This implies that ITS is able to detect individual sources as long as their fluxes are above its detection limit and as long as the source density is less than 26 sources per steradian. A close examination of a $20^\circ \times 20^\circ$ region around Taurus A revealed that the map above the 70% of peak sensitivity contour is confusion limited [142].

2.2.4 Core Station 10

After completing the ITS, the project team dedicated itself to the development of the next prototype station that would sample the incoming signals at 200 MHz and process these data in real-time. This prototype was built on the field on which core station number 10 was planned and was therefore named CS10. CS10 initially consisted of 48 LBAs in an exoshell configuration and an on-site real-time processing platform with practically the same architecture as the final design described in Ch. 1. In a later stage, 30 HBAs and 4 HBA tiles were added in varying configurations. This station became operational in September 2007. The map shown in Fig. 2.6 has therefore been the key demonstrator of LOFAR's future abilities for almost three years. However, the ability of CS10 to process the incoming signals in real-time paved the path for further experiments, some of which will be presented in the next sections.

2.3 Station calibration

A LOFAR station tracks a source by assigning complex valued weights to the antenna signals in such a way that signals originating from the direction of the source are added coherently while unwanted signals, like RFI, are suppressed. These beam former weights should thus correct the geometrical delays over the array, modify the beam pattern as desired (e.g. lowering side lobes by tapering and nulling of RFI) and correct for instrumental gain and phase variations between signal paths. As can be calculated using the gain error propagation formulas derived in Ch. 6, it can be shown that for an array of 100 elements, 1% errors in the beam former weights cause a rel-

ative error in the main beam of the station sensitivity pattern of -27 dB and relative errors of -17 and -5.5 dB on -20 and -40 dB side lobes respectively [148]. These numbers clearly demonstrate the importance of choosing the appropriate beam former weights to form a well defined beam to facilitate further processing at central level.

Calculation of these beam former weights requires knowledge of [150]:

1. the position of each receiving element;
2. the orientation of each receiving element;
3. the complex electronic gains as function of frequency;
4. the system noise as function of frequency;
5. the element beam pattern as function of frequency.

This should either be provided by the station calibration or by exploiting prior knowledge. In principle, the positions of the individual receivers can be derived from visibility data if a good source model is available. Using the Cramer-Rao bound, which is the lowest possible total variance of an unbiased estimator, it could be shown that even if the calibration observations are done when the source geometry is most optimal, the SNR of the standard station calibration measurements will limit the position estimation accuracy to about 30 cm, which is much higher than the 2 cm RMS accuracy of antenna placement during roll-out. It was therefore decided to measure the antenna positions physically [150].

The same conclusion was drawn for the orientations. Since the element beam pattern is very smooth, a small change in the orientation of an antenna will only cause a minor change in the direction dependent gains towards the calibration sources. An attempt to estimate the orientation based on a known or assumed element beam pattern will thus result in large estimation errors, while $\sim 1^\circ$ RMS placement accuracy can easily be achieved in the field. This is sufficiently accurate for station processing.

In [150] an overview is given of experiments to validate element voltage beam predictions from full EM simulations on ITS against actual data. The final conclusion was that the prediction from the full EM simulations agreed quantitatively with the observations within the measurement errors. It was thus concluded that direction dependent gains could confidently be derived from the element beam predicted by the EM simulations and would not have to be estimated by the station calibration.

The station calibration algorithm should therefore focus on estimation of the complex valued electronic gains of the individual signal paths and their system noise powers. This is a common problem in the array signal processing literature [12, 39, 79, 90], but practical experience with the LOFAR prototypes indicates that we have to deal with a number of additional challenges that are discussed in the next section. The resulting calibration methods are described in Ch. 5.

2.3.1 Challenges

Autonomous process

Although the LOFAR antennas are sufficiently stable to apply corrections known from tables after initial station calibration, the calibration routine will run continuously to support system health monitoring. Furthermore, this is not a single centralized process, but a process running at each station. Human interaction is thus impractical and undesirable. The station calibration routine should therefore run autonomously. This implies that it should recognize unsuitable data and verify that subtle effects, like very low power RFI, did not affect the calibration results.

Limited computing resources

The station correlator can real-time correlate a single subband. The shortest possible integration time is one second. The station correlator can thus scan over all 512 subbands in the current observing mode in 512 seconds to produce a dataset containing the cross- and autocorrelations of all dipoles for each subband. The station calibration process should keep up with this rate of data generation if the time resolution provided by the station correlator is to be maximally exploited. Since the station calibration software runs on the LCU, which is a single standard computer, this begs for an efficient algorithm that can process this data at the desired rate using only one or two concurrent CPUs.

Short baseline effects

A LOFAR station contains many short baselines with a length of only a few wavelengths. These baselines are very sensitive to large scale structures on the sky like the emission from our own Galaxy as is nicely demonstrated by the all sky map presented in Fig. 2.6. This is a nuisance for calibration because the instrumental response to a mixture of large scale extended emission and point sources is hard to model.

The sensitivity pattern of a phased array provides a variable resolution because the cross section of the array varies with elevation. As long as the resolution is lower than the structure size of the extended emission, the total power received by a beam pointed at that structure will increase if the resolution decreases, i.e. the main beam gets larger. A point source has, by definition, a size smaller than the resolution of the instrument. If the resolution decreases, the point source still fits within the main beam and therefore the power received from the source remains the same. Extended emission and point sources thus require different modeling, which requires two separate sky models and is computationally demanding.

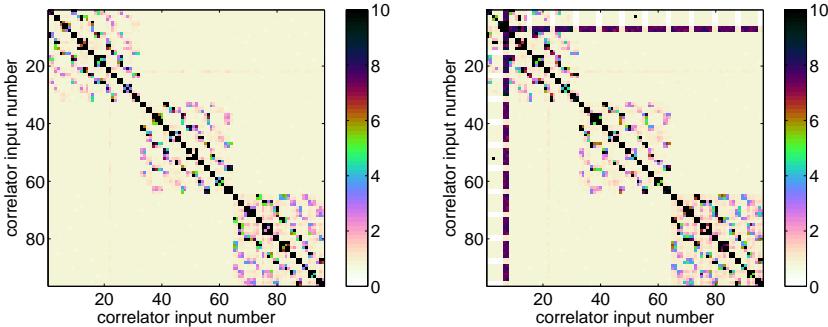


Figure 2.7: These plots show the median amplitude of the array covariance matrix elements over frequency expressed in standard deviations while the ADC clock is running at 160 MHz (left) and 200 MHz (right). The structures in the correlations of the first eight RCUs at 200 MHz are not due to crosstalk but due to bad synchronization between the RSP boards, a problem that was solved shortly after this measurement.

This computationally intensive modeling of the calibration data only works if we know the instrumental response. Unfortunately, the electromagnetic coupling between closely packed antennas poses another challenge. Full EM simulations can provide decent approximations to the beam patterns of antenna elements in a large regular array or those of isolated antennas. Even with the availability of a supercomputer the intermediate situation of a randomly distributed group of antennas with separate small ground planes is prohibitively expensive to simulate. This is the case for the LBA station. The experiments with the 327 MHz array demonstrate how difficult this modeling is [93]. The conclusion here is that the correlations on short baselines should be regarded unsuitable for calibration. It would thus be convenient if we could ignore the data on the short baselines.

Crosstalk

Flagging of specific visibilities is also desirable in view of the crosstalk found in the station backend at CS10. This is demonstrated by the results shown in Fig. 2.7 obtained from measurements on 29 through 31 January 2007. In these experiments, $75\ \Omega$ loads were mounted on the RCU inputs instead of antenna cables. The thermal noise of these loads should be uncorrelated. The station correlator was swept over all 512 subbands integrating over 64 s for each subband. This sweep was made once with the ADC clock running at 160 and once at 200 MHz. Figure 2.7 shows the median

amplitude of each visibility over frequency. Since the input signals are uncorrelated, we should obtain crosscorrelations that only contain noise. To facilitate verification, the color scales have therefore been expressed in standard deviations. This shows that there is some structure in the crosscorrelations that should not be there.

Additional measurements [153] indicate that the analog receiver chain with filters and amplifiers picks up signals radiated by the backplane connectors of neighboring RCUs in the same subrack. Since the RCUs within a Dutch LOFAR station are distributed over three subracks, we see three distinct groups in both plots in Fig. 2.7. The structure within each subrack is caused by the numbering of the RCUs and the fact that the crosstalk has a preferential direction towards the neighboring RCUs faced by the analog receiver chain.

Celestial sources as calibrators

Calibration using celestial sources poses several challenges. Since the LOFAR antennas are mechanically fixed and have a hemispheric FOV, they cannot be steered towards a single calibrator. Calibration should therefore exploit the sources that are present at the time of observation. These sources will generally be Cas A, Cyg A, Tau A, Vir A and the Sun. These sources may be simultaneously present. The calibration algorithm thus has to handle multiple sources.

Although the aforementioned sources are the strongest celestial sources in the sky, their power is still only $\sim 1\%$ of the system noise. This implies that the SNR of these sources in a 195 kHz 1 s snapshot is less than 5 per baseline. The calibration method should therefore be able to work at low SNR and should preferably be statistically efficient to obtain as accurate gain estimates as possible.

In view of the low SNR of the celestial sources, it has been proposed to use RFI for calibration, especially those sources that are the single dominant source in their subband. This would allow for a much simpler calibration routine, which exploits a single strong calibrator. However, this method also has a few disadvantages. Although a few demonstrations on ITS data [117, 140] indicate that it is possible to image astronomical sources in the presence of moderately strong RFI, up to 60 dB above the power of Cas A, RFI free subbands are generally more suitable for astronomical observations. A small glance at a typical LBA spectrum as shown in the left panel of Fig. 2.10 in Sec. 2.3.3 reveals that the RFI is not nicely distributed over the band and is virtually absent in the frequency range in which the antennas have their highest gains. Calibration on RFI would therefore not provide calibration data for the subbands that are most interesting from an astronomical point of view. Another disadvantage of RFI sources is that most transmitters are located on the ground. Most RFI is therefore located at the horizon, while most observations will be done at elevations above 30° . At low elevations, effects like reflections in the station en-

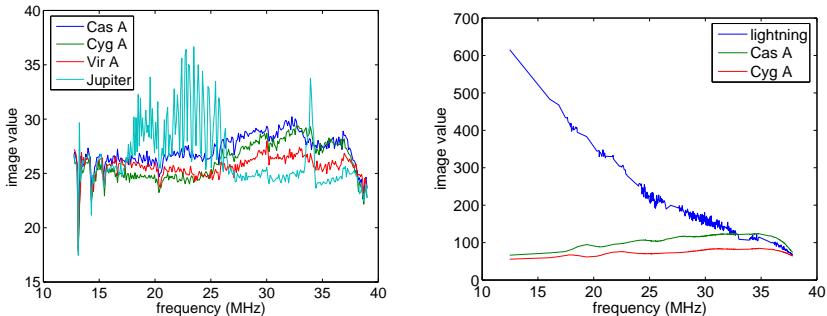


Figure 2.8: Autocorrelation spectra of a few sources derived from hemispheric images without absolute flux calibration from ITS during a Jupiter burst (left) [142] and a lightning bolt (right) [144].

vironment and shadowing of antennas will perturb the incoming signals. The single source calibration model that would be assumed for the RFI may therefore turn out to be an oversimplification. For these reasons it was decided to use celestial sources as calibrators [150].

Transients

Calibration on celestial sources using a fixed sky model is occasionally hampered by transient phenomena. We have already seen a very noticeable example: the coronal mass ejection detected by THETA. Figure 2.8 shows two other examples detected in ITS observations. The left panel shows autocorrelation spectra derived from hemispheric images over the full ITS frequency range at the location of the indicated sources. Below 26 MHz there is strong sharply peaked emission from a Jupiter outburst caused by the magnetic interaction between Jupiter and its moon Io [74, 169]. The right panel shows the presence of a continuum source with an even steeper spectrum than Cas A and Cyg A. This is a typical spectrum produced by lightning. With the tools developed for transients like the UHECRs observed with LOPES it is even possible to make a movie showing how the lightning bolt ionizes its way through the atmosphere!

In many cases, the characteristics of such transient phenomena are more similar to those of regular celestial sources than to those of RFI. They may therefore escape detection by an RFI flagger while they are sufficiently strong to have a severe impact on the calibration results. This implies that the station calibration pipeline should verify that its results are reasonable. This can be achieved by assuming that the

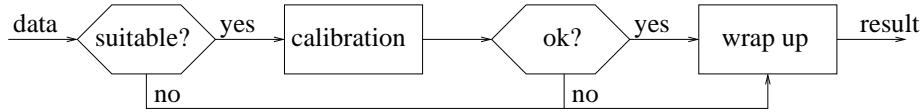


Figure 2.9: Overview of the station calibration pipeline.

complex valued electronic gains vary smoothly over time and frequency.

2.3.2 Station calibration pipeline

An overview of the station calibration pipeline [150, 156, 160, 167] is shown in Fig. 2.9. The calibration data are produced by a single frequency sweep of the station correlator over all subbands with one second integration per subband. Before calibration, the data are checked for suitability. The calibration routine assumes a source model consisting of celestial sources. The calibration results will thus be distorted if the subband contains RFI and will be inaccurate for subbands outside the passband of the analog filters. These subbands are flagged before calibration to save computing resources.

The calibration itself is done using the modified weighted alternating least squares method described in Sec. 5.5. This method exploits nuisance parameters to “flag” the visibilities that are prone to modeling errors, for example due to the short baseline effects. The validity of the calibration results is checked by assuming that the complex electronic gains of the receiver paths vary smoothly over time and frequency. After this detector, we will only have valid correction factors for a limited number of subbands. The other subbands either contained RFI or did not produce a correct result, for example due to lightning as illustrated in the previous section. The corrections required for those subbands can be obtained by fitting a band pass model to the available calibration results.

The RFI detection method used at the LOFAR stations is based on the Frobenius norm of the array covariance matrix [156, 160]. This allows the algorithm to exploit differences in spatial as well as spectral characteristics between the interfering sources and the calibrators. It is therefore sensitive to even very weak RFI sources that have an apparent power similar to sources like Cas A and Cyg A. Since the main purpose of this detector is to remove subbands that will likely produce erroneous calibration results, this detector may be very tolerant, i.e. it may have a high false acceptance rate to ensure a low false rejection rate. A detailed study based on actual data from CS10 indicates that the test statistic has a Gaussian distribution [81], which greatly simplifies quantitative analysis of the false rejection rate.

The data captured by the station correlator for consecutive subbands is acquired

only 1 s apart. The state of the instrument and the source configuration hardly changes over such a small time interval. The response of the system is also expected to vary only slowly with frequency. The validity of the calibration results can thus be checked by putting a threshold on the difference between the antenna based complex gain solutions for consecutive subbands [160]. Since the station calibration works autonomously, the calibration results will hardly be checked by a human operator. False acceptance of an erroneous result may thus have serious effects. This detector should therefore be very strict. For LOFAR, the threshold was determined using data from monitoring campaigns.

2.3.3 Demonstrations

In this section I demonstrate the operation of the station calibration pipeline using data from two monitoring campaigns with the LBAs at CS10. The first run was a three day measurement covering the 10 – 90 MHz frequency range started on 18 January 2008 at 8:23:58 UTC in which the station correlator was set up to sweep over all 512 195 kHz subbands continuously with an integration time of 1 s per subband. This provides good frequency coverage but each individual frequency is sampled at a time interval that is a little larger than 8.5 minutes. A second observation was done with fine sampling in time but coarse sampling in frequency and the same hardware settings. In this monitoring campaign, which was started on 14 February 2008 at 15:31:38 UTC, each of the subbands 160, 192, 224, ..., 416 was monitored for 300 s with 1 s integration per snapshot observation before switching to the next subband. This cycle was repeated 19 times.

Detection of RFI and devious results

The left panel in Fig. 2.10 shows an autocorrelation spectrum from the first frequency sweep during the January campaign showing the 10 – 90 MHz band pass filter and RFI in the low (<25 MHz) and high (>88 MHz) frequency ends of the observing band. The autocorrelation spectrum after blanking is shown in the right panel of Fig. 2.10. The data in the stop band will be ignored during calibration.

A tolerant RFI detection threshold increases the probability that low power RFI escapes detection. This is nicely demonstrated in the left panel of Fig. 2.11 showing the phases of the complex gain solutions. One element has been used as phase reference. The phase behavior shown in these plots is therefore the difference in phase behavior between two signal paths showing, amongst others, the effect of the antenna resonance. The detection threshold per element was set to a maximum relative error of 20%. If at least 40 out of 48 antennas had a smaller error, the data were regarded as

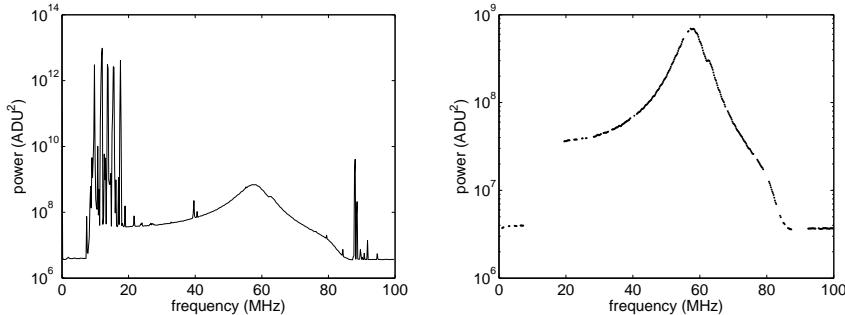


Figure 2.10: A typical LBA spectrum before (left) and after (right) blanking of RFI occupied subbands [160].

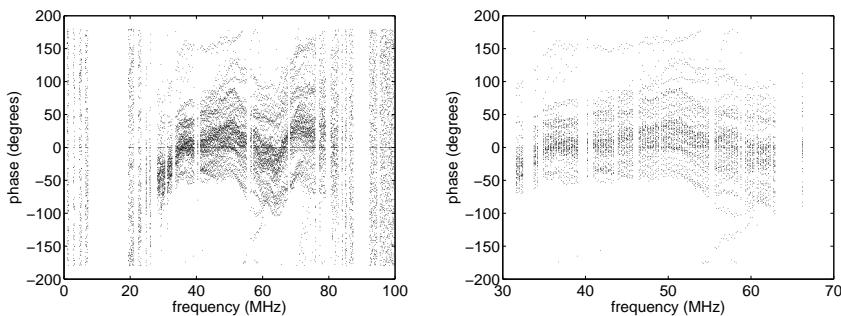


Figure 2.11: The phase of the complex gain solutions before (left) and after (right) blanking of erroneous results [160].

good. These data are shown in the right panel of Fig. 2.11. This shows that the outermost frequency points are no longer available stressing the importance of a good band pass description to ensure correct extrapolation to the lower and upper frequencies in the observing band.

Dealing with short baseline effects

The station calibration algorithm that is currently implemented on the LCU at each station solves for a complex valued direction independent gain per receiver path, a noise model that spans all baselines shorter than four wavelengths and the direction dependent gains towards the calibrators. For CS10, this implies estimation of 48

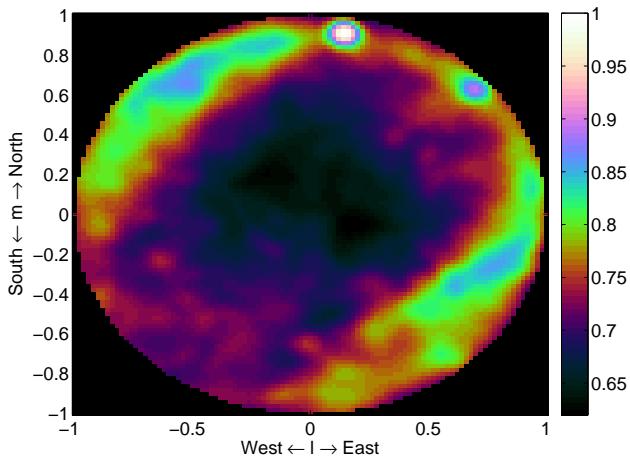


Figure 2.12: Calibrated hemispheric image for the array of x -dipoles on CS10 in a single 195 kHz subband centered at 50 MHz observed on 15 February 2008 at 1:42:07 UTC.

antenna based gains, 47 antenna based phases, the source power ratio of the two brightest sources and 764 real valued nuisance parameters describing the additive terms on the baselines shorter than four wavelengths for a total of 860 real valued parameters per polarization. Using data from a 1 s snapshot observation in the 195 kHz subband at 50 MHz captured on 15 February 2008 at 1:42:07 UTC, I show that use of such additive nuisance parameters can reduce the complex source structure on the sky to a simple model with only two point sources.

Figure 2.12 shows a calibrated fish-eye image for the array of x -dipoles. There are two bright point sources, Cas A and Cyg A, near the northeastern horizon. The image also shows a lot of extended emission from the Galactic plane (near the northwestern horizon) and the north polar spur (near the eastern horizon). This extended emission is hard to model accurately, but only affects the short baselines since short distances in the aperture plane of a phased array correspond to the low spatial frequencies, that describe the structure on large spatial scales.

This observation was calibrated using the method described in Sec. 5.5. The estimated noise model can be interpreted as an estimate of the combined effect of extended source structure, noise coupling and receiver noise powers. This is nicely demonstrated by the image of the estimated noise model after calibration shown in the left panel of Fig. 2.13. The right panel shows the difference between the all-sky

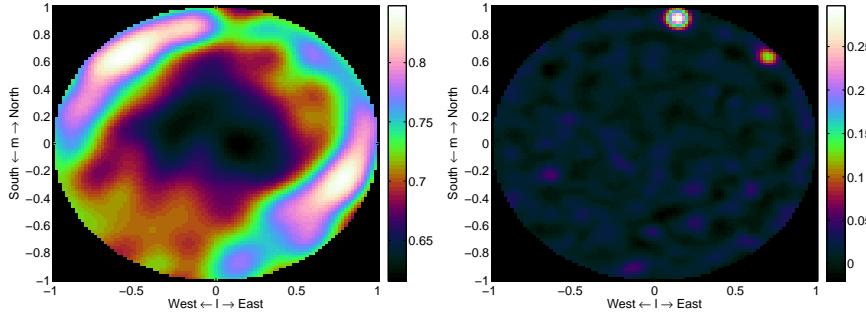


Figure 2.13: Calibrated hemispheric image of the noise model for baselines shorter than four wavelengths (left) and difference between the calibrated fish-eye map shown in Fig. 2.12 and the noise model indicating that the remainder can be accurately modeled by only two point sources.

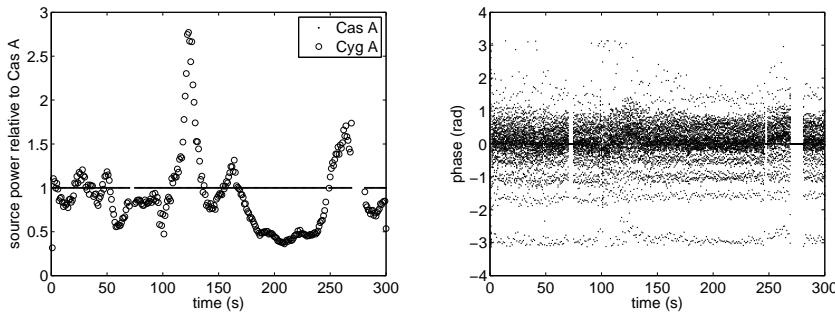


Figure 2.14: Apparent source power of Cas A and Cyg A (left) and phase of the complex valued direction independent gains (right) during scintillating ionospheric conditions [158].

map shown in Fig. 2.12 and the image of the noise model. This indicates that after subtraction of the noise model from the data, the data can be described by a model consisting of only two point sources. This reduces the original calibration problem to one that has been discussed extensively in the array signal processing literature.

Ionospheric scintillation

It may seem surprising that we have to estimate the apparent source powers of well-known astronomical calibrators like Cas A and Cyg A, because it should be possible to model them based on the instrumental response and their intrinsic power. Such an effort is hampered by the presence of the ionosphere which introduces a time, frequency and direction dependent gain effect. This can be illustrated by the calibration solutions found during the 300 s measurement started at 2:18:47 UTC on 15 February 2008 in a 195 kHz subband at 50 MHz. This measurement, originally reported in [158], was affected by strong ionospheric scintillation. The left panel of Fig. 2.14 shows the apparent source powers of Cas A and Cyg A over these 300 s indicating that their power ratio varies by a factor 7 with a steepest gradient of a factor 4 in just 15 s. Note that Cas A has been used as power reference, causing scintillation of Cas A to be absorbed in the gain solutions for the receiver paths. The right panel of Fig. 2.14 shows the receiver path phase solution for this observation, which remains almost flat, as expected for a stable instrument. The stability of the instrument can thus only be demonstrated if the data model allows for variability in the source model. The assumption that the sources are inherently stable would have resulted in a direction independent gain solution per receiver path suggesting a highly unstable instrument.

Impact of station calibration on telescope sensitivity

The impact of station calibration on the system performance can most clearly be illustrated by determining the gain of the station, usually expressed as the average effective area divided by the system temperature per dipole A/T , before and after calibration. For this experiment data were used from a single frequency sweep started at 8:49:58 UTC on 18 January 2008 during the first monitoring campaign. The raw visibility data for baselines longer than four wavelengths of the x -dipoles were imaged in a small region around Cas A to determine the apparent power of Cas A per baseline. The system noise power was determined by taking the mean of the autocorrelations. The A/T ratio is directly related to the power ratio of Cas A and the system noise [153]. The same ratio was determined based on the source and noise power estimates obtained during calibration. Both results are presented in Fig. 2.15 suggesting almost a factor two improvement in the sensitivity of CS10 due to calibration.

2.4 Fish-eye imaging

The fish-eye images and the survey map shown so far were made by computing the direct Fourier transform (DFT) of the visibilities. Imaging can also be considered an estimation problem in which the intensity of every pixel in the map needs to be

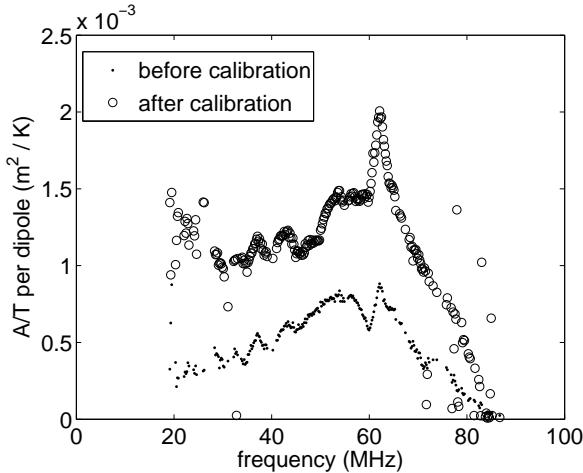


Figure 2.15: Average A/T per dipole measured towards Cas A before and after calibration obtained from a measurement started at 8:49:58 UTC on 18 January 2008.

estimated. This approach is described in Ch. 6 leading to a closed form expression for the least squares solution to this problem. I will refer to this approach as least squares imaging, mainly to signify that other optimization schemes, like l_1 -optimization [22, 94], are also possible. In the literature, least squares imaging with appropriate weighting is referred to as optimal map making [114, 115]. In this section, I show that least squares imaging implicitly takes care of the deconvolution of the true image and the array response [163] and that it appears to produce statistically efficient results on actual data [162]. This chapter ends with a survey map based on CS10 data that was made by combining the calibration method described in Sec. 5.5 with least squares synthesis imaging. Although the resulting map indicates that there is room for improvement and further studies, it clearly shows the potential of least squares imaging for LOFAR and SKA core observations, in which complex source structures like our Galaxy will be observed with a large FOV.

2.4.1 DFT versus least squares imaging

The attractive point of the least squares imaging technique is the deconvolution of complex source structures. It can also handle arbitrarily large FOVs. It should thus be well suited for fish-eye observations with a LOFAR station. On 8 November 2008

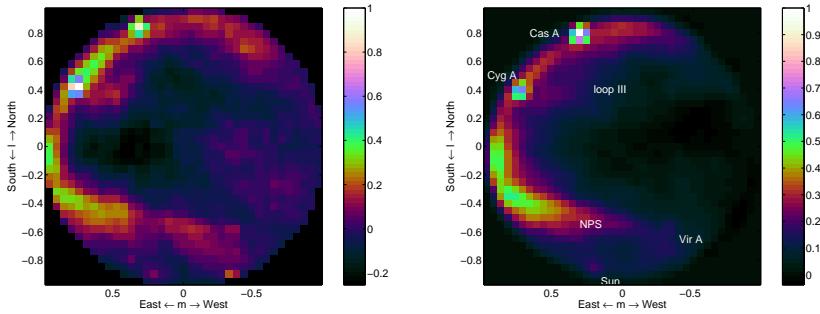


Figure 2.16: These total intensity maps were made by combining data from 27 156 kHz subbands evenly distributed between 45.3 and 67.3 MHz and 10 seconds of integration per channel. The left panel shows the result after applying a direct Fourier transform to the raw visibility data excluding the autocorrelations followed by summation over frequency while the right panel shows the result obtained using the least squares imaging technique proposed in Sec. 6.3 after calibration of the individual frequency channels using the method described in Sec. 5.5 [163].

between 10:21:59 and 10:26:45 UTC data were captured from the 48 LBAs of CS10. The station correlator was used to correlate 27 156 kHz wide RFI free subbands evenly distributed between 45.3 and 67.3 MHz. The integration time per subband was 10 s. The arrays of x -dipoles and y -dipoles were calibrated independently using the weighted alternating least squares approach described in Sec. 5.5 using Cas A and Cyg A as calibrators. Least squares imaging was applied to the calibrated xx -visibilities and yy -visibilities separately before the resulting images were added to form a pseudo-intensity map. The final result is shown in Fig. 2.16. For comparison, a DFT was applied to the raw visibility data.

In the least squares image a number of well known sources have been indicated. The most subtle feature visible is the Galactic loop emerging from Cyg A. This bit of extended emission can be identified as loop III in the Haslam survey [49]. This shows the improvement of the least squares image over the DFT image. In the DFT image, there is some extended emission emerging from Cyg A, but there is also a large region of extended emission near the western horizon that cannot be associated with known sources. In the least squares image, this emission disappears while the emission near Cyg A remains. We can therefore conclude that the extended emission near the western horizon in the DFT image is actually caused by the array response to the Galactic plane and the north polar spur. This demonstrates that least squares imaging ade-

quately removes the instrumental response thereby showing only the real emission in the final image. This simple observation already involves an interesting deconvolution challenge since the image contains both strong point sources as well as extended emission that are convolved with a frequency dependent station beam and element beam. With conventional imaging techniques, this would require a multi-resolution CLEAN [29]. The least squares imaging technique performs this deconvolution task implicitly.

2.4.2 Statistical performance

The statistical performance of an algorithm can be assessed by comparing the variance of the parameter estimates in multiple realizations with the Cramer-Rao bound (CRB). These realizations are normally obtained by Monte Carlo simulations, an approach that will also be exploited in Ch. 5. If the instrument, the environment and the test signal are stable, multiple realizations may be obtained by repeating a single measurement over time. Browsing through the data from the monitoring campaign in February 2008, I found a 300 s observation on the 195 kHz subband at 50 MHz consisting of 298 snapshots with 1 s integration each that did not seem to be affected by ionospheric disturbances or other obvious irregularities. These data were captured by the station correlator at CS10 starting at 15:46:44 UTC on 14 February 2008.

The data for each 1 s snapshot observation were calibrated using the method described in Sec. 5.5. The calibration solutions were constant over the 300 s interval confirming that the individual snapshots could be regarded as realizations of the same measurement. The calibration results were averaged over the complete interval to improve their accuracy and avoid additional noise in the image due to varying calibration error propagation (see Ch. 6). The calibration results were applied to the data during the imaging process. The top left panel of Fig. 2.17 shows the average image over all 298 snapshots.

The CRB was calculated using the result for source power estimation obtained in Ch. 4 by regarding each pixel in the image as a discrete source. The result is shown in the top right panel of Fig. 2.17. Note that the CRB is not constant over the image. This is an important result, because it shows that the interplay between the array beam pattern and the source structure in the imaging and deconvolution process causes a non-uniform noise distribution over the map that neither follows the source structure nor is homogeneous over the sky. Initially this was explained by the fact that LOFAR is a sky noise dominated instrument, which implies that the source noise is a significant fraction of the total system noise [60, 147]. However, the more detailed analysis presented in Ch. 6 indicates that that explanation holds for DFT imagers, but does not necessarily hold after deconvolution.

The variance of the estimated image values was determined by computing the

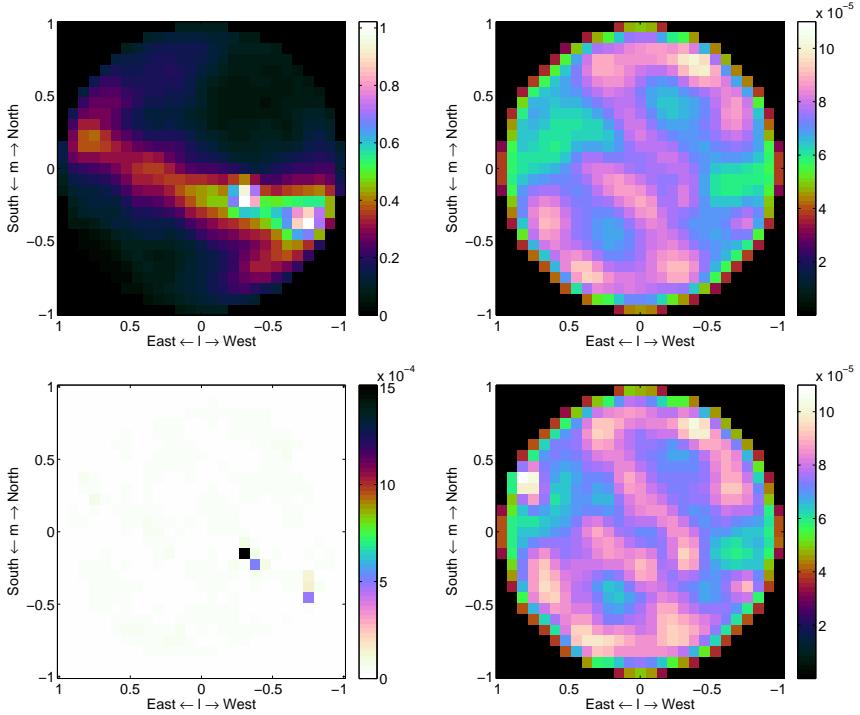


Figure 2.17: These plots demonstrate the excellent noise performance of least squares imaging using 298 consecutive 1 s observations in the same subband. The top left panel shows the all sky map, the top right panel shows the expected variance of the image pixels based on the Cramer-Rao bound, the bottom left panel shows the actual variance of the image values if Earth rotation is ignored and the bottom right panel shows the variance of the image values if Earth rotation is handled properly [162].

variance for every pixel in the map over the 298 snapshots. The result is shown in the bottom left panel of Fig. 2.17. The variance on Cas A and Cyg A appears to be an order of magnitude higher than predicted by the CRB while the values and overall structure in the rest of the image seem to agree quite well. During a 5 minute interval, the Earth rotates by 1.25° , which is a significant fraction of a 5° beam. We are therefore not measuring the variance of the imaging process but the rate at which Cas A and Cyg A are moving in and out of a station beam directed towards a particular point in the sky. The estimation procedure for the variance was improved

by calculating the average variance per 10 consecutive snapshots. This led to the result shown in the bottom right panel of Fig. 2.17 indicating quantitative agreement with the the CRB. The least squares imaging method described in Sec. 6.3 therefore appears to operate statistically efficient on actual data.

2.4.3 All-sky least squares synthesis imaging

The hemispheric image shown in Fig. 2.16 demonstrates that the least squares imaging method described in Sec. 6.3 can produce a frequency synthesis image. Mathematically, however, there is no difference between snapshots over frequency and snapshots over time. We should thus be able to combine multiple observations taken at different times into a synthesis map. After 24 hour, a LOFAR station has seen the entire sky visible from its geographical location and this synthesis map should be an all-sky image, covering up to 4π steradian for a station placed at the equator. This can be demonstrated using data captured during the first monitoring campaign with the LBAs at CS10 started on 18 January 2008 at 8:28:39 UTC and lasting until 9:37:04 UTC on 21 January 2008. During this period the station correlator swept continuously over all 512 195 kHz subbands with 1 s integration time per subband.

The data for the array of x -dipoles were calibrated using the station calibration pipeline discussed in Sec. 2.3.2. The standard calibration routine was extended with position estimation of the calibrators to reduce the impact of ionospheric effects on the antenna based gain estimates. After calibration and subtraction of Cas A and Cyg A, least squares imaging was used to produce the all-sky map shown in Fig. 2.18 using all RFI free data between 40 and 70 MHz assuming an isotropic element beam pattern. The actual element beam pattern is still an object of study and can easily be included as soon as an appropriate model becomes available.

Comparison with the ITS survey map shown in Fig. 2.6 clearly shows the impact of the implicit deconvolution performed by least squares imaging; the extended structures like the galactic plane are less smeared out, the contrast of the bright sources Vir A and Tau A is much improved and the “empty” parts of the sky are now much darker. In the CS10 map the sun is better localized near about 20 hours right ascension and -20° declination, since the measurements are spread over only three days instead of two weeks.

The image also shows that there is room for further improvement. The flux of Cas A and Cyg A was estimated by the calibration routine based on baselines longer than four wavelengths. This baseline restriction should ensure that the flux of the underlying galactic plane is not attributed to these point sources. In Fig. 2.13 one can already see that this does not work perfectly if the galactic plane is close to the horizon. In that case, the projection of a four wavelength baseline in the direction of the galactic plane is much shorter making the baseline restriction less effective. This

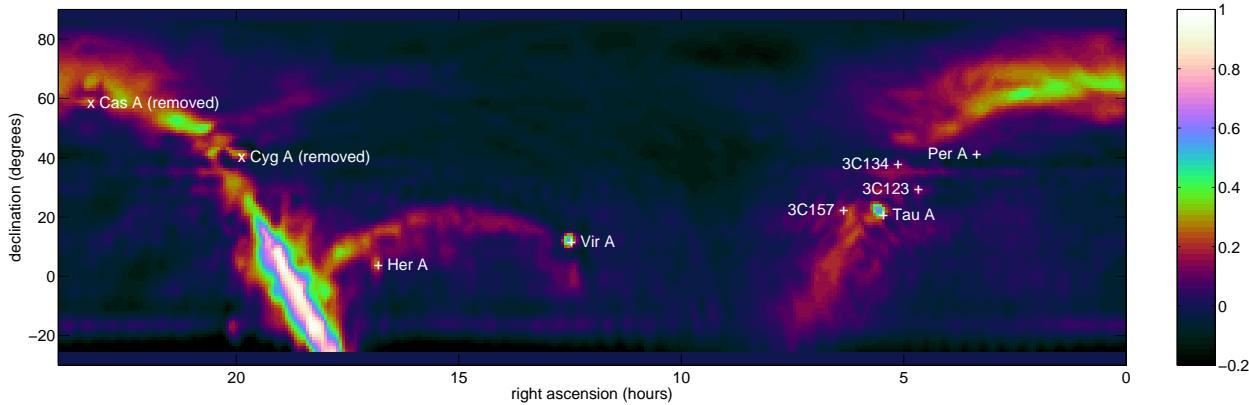


Figure 2.18: CS10 survey map based on 508 frequency scans with the station correlator using all RFI free 195 kHz subbands between 40 and 70 MHz. Cas A and Cyg A have been removed after calibration of the individual snapshots but before least squares imaging.

causes small artefacts at the locations of Cas A and Cyg A. Such artefacts were not visible in the ITS survey map shown in Fig. 2.6, which demonstrates the least squares imaging is far more sensitive to errors in the preparatory data reduction than DFT imaging. This is an important asset, since it implies that data reduction errors are not covered up by the imaging routine.

The grid of pixels was chosen such that the maximum achievable resolution towards a source at a given declination is properly taken into account. This gives rise to a declination dependent grid spacing. The smallest spacing is defined by a source that transits in the zenith and is $0.75\lambda_0/D$ where D is the diameter of the station and λ_0 is the minimum wavelength, in this case 4.28 m at 70 MHz. This grid may cause artifacts like the horizontal feature at the declination of the sun and the horizontal feature just below 3C134. Better results can probably be obtained if the grid is chosen based on the data instead of using a predefined grid. This would also include a more objective criterion for the optimal imaging resolution, which is now determined by hand based on several trials using the condition number of the deconvolution matrix as figure of merit.

Although there is enough room for further improvement, least squares imaging seems to be a promising technique for imaging with relatively small arrays like the LOFAR and SKA cores with a large FOV observing complex source structures. As demonstrated with the examples above and discussed in Sec. 6.3, its attractive features are:

1. it can handle arbitrary direction dependent effects;
2. it can handle arbitrary visibility corrections;
3. it can handle non-sparse fields, i.e. complex source structures;
4. it appears to be statistically efficient on actual data.

The results presented in this chapter demonstrate that the calibration and imaging methods described in Secs. 5.5 and 6.3 work well on actual data. The foundations for this success are laid in the next chapters. Chapter 3 describes the data model based on an accurate physical description of a phased array antenna system. Expressions for the Cramer-Rao bound are derived in Ch. 4. This allows a thorough assessment of statistical performance of the proposed calibration and imaging algorithms, which led to several refinements over the years. This is nicely illustrated by the calibration approaches applied to THETA, ITS and CS10 data mentioned in this chapter. However, a statistically efficient method based on an erroneous data model will not produce proper results on actual data, so the next chapter is devoted to the data model.

Chapter 3

Data model

The data model, often called measurement equation in the radio astronomical community, is the foundation on which all calibration and imaging or, in general, all data reduction algorithms are built. It provides a mathematical model of the measurement system and shows the assumptions underlying the data reduction methods. Due to its fundamental importance all well-written literature on (array) signal processing starts with a description of the data model. I will therefore mention only a few references suitable for introductory reading. Thompson, Moran and Swenson [116] and Perley, Schwab and Bridle [88] provide a thorough introduction to the field of radio astronomy. A key contribution to the measurement equation came from Hamaker, Bregman and Sault [45], who recognized that correct treatment of measurements on polarized sources required the use of Jones matrices, i.e. the use of a matrix equation per antenna instead of a scalar equation per feed. Leshem and Van der Veen [64], Van der Veen, Leshem and Boonstra [123] and Boonstra [14] have reformulated the measurement equation in the formalism used in the array signal processing literature in which the data model for the entire array is formulated as a single matrix equation. This paves the way for the use of results from linear algebra for both mathematical analysis and extraction of parameters from the measured data. With the advent of phased arrays in radio astronomy, such as ThEA [56, 141], LOFAR [15, 127], EMBRACE [1, 2, 84] and MWA [72, 73], the study of the electromagnetic interaction between the receiving elements in a tile or station array becomes increasingly important. Lanne [61] describes a practical data model for antenna arrays including mutual coupling.

The description in this chapter starts from antenna theory. This leads to a rather complex data model that imposes only minor restrictions on its applicability. Working with this data model is rather inconvenient when focusing on a specific aspect of the

system or data reduction. The analysis in the rest of this thesis will therefore be done on simplified variants of the generic model presented here. This presentation should provide a clear view on the assumptions made in the discussions.

3.1 Physical model

The electromagnetic field probed by the antennas in a phased array can be described as a superposition of electromagnetic waves coming from all directions at different frequencies. Most signals relevant to radio astronomers originate from such vast distances that the curvature of these waves over the array may be neglected. This allows us to describe the electromagnetic wave with wave vector \mathbf{k} impinging on the receiving element located at position $\boldsymbol{\xi}$ at time t as the plane wave

$$\mathbf{E}(\boldsymbol{\xi}, \mathbf{k}, t) = \mathbf{E}_0(\mathbf{k}, t) e^{-j(\mathbf{k}^T \boldsymbol{\xi} - |\mathbf{k}|ct)}. \quad (3.1)$$

The wave vector $\mathbf{k} = -2\pi f \mathbf{l}/c$ completely describes the plane wave coming from direction \mathbf{l} at frequency f and j is defined as the negative square root of -1 . The plane wave spectrum $\mathbf{E}_0(\mathbf{k}, t) = [E_{0\theta}, E_{0\phi}, E_{0r}]^T$ describes the complex amplitude distribution along the three polarization directions \mathbf{e}_θ , \mathbf{e}_ϕ and \mathbf{e}_r . Since a plane wave does not have a field component along the direction of propagation, the component along \mathbf{e}_r is zero and therefore it is sufficient to describe the field amplitude by just two components, i.e. $\mathbf{E}_0(\mathbf{k}, t) = [E_{0\theta}, E_{0\phi}]^T$.

The amount of power absorbed by the p th receiving element subject to an electromagnetic wave with wavenumber \mathbf{k} and polarization $[E_{0\theta}, E_{0\phi}]^T$ depends on the sensitivity of the receiving element to radiation from that direction at that frequency with that polarization. Using reciprocity, this sensitivity pattern can be determined by measuring the radiation pattern of the array while applying an input current I_0 to the p th element and leaving the terminals of all other elements open. This is called the embedded open circuit loaded radiation pattern \mathbf{E}_p [128, 129] and will be referred to as the element beam pattern. The isolated loaded radiation pattern provides a reasonable approximation if the array is sufficiently sparse [48, 130]. The open circuit voltage at the antenna terminals is thus given by [129]

$$v_{oc,p}(\mathbf{k}, t) = \frac{4\pi j \xi_0 e^{j|\mathbf{k}|\xi_0}}{|\mathbf{k}| \eta I_0} \mathbf{E}(\boldsymbol{\xi}_p, \mathbf{k}, t) \cdot \mathbf{E}_p(\mathbf{k}) \quad (3.2)$$

where ξ_0 is the distance at which the embedded open circuit loaded radiation pattern is measured and $\eta = \sqrt{\mu/\epsilon}$, with μ and ϵ as the permeability and dielectric constant of the propagation medium, is the intrinsic impedance, which is 377Ω for free space.

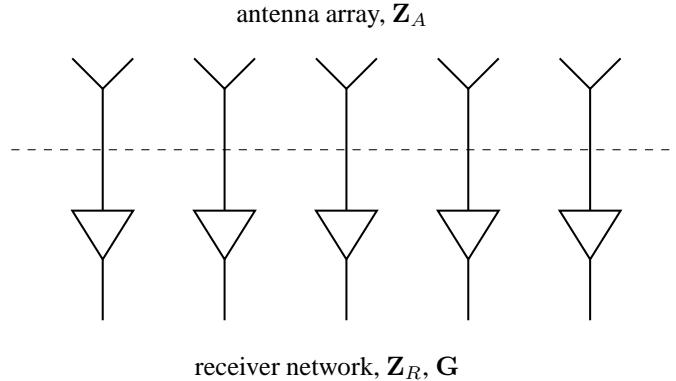


Figure 3.1: Schematic representation of an antenna array characterized by its mutual impedance matrix \mathbf{Z}_A and its receiving network characterized by its mutual input impedance matrix \mathbf{Z}_R and its gain matrix \mathbf{G} .

The open circuit voltages of all P elements of the antenna array can be stacked in a $P \times 1$ column vector $\mathbf{v}_{oc} = [v_{oc,1}, v_{oc,2}, \dots, v_{oc,P}]^T$. The impedance of the output terminals of the array can be described by a mutual impedance matrix \mathbf{Z}_A . The array is loaded by a receiver chain network, generally consisting of a low noise amplifier behind every receiving element as depicted in Fig. 3.1, characterized by its mutual input impedance matrix \mathbf{Z}_R and voltage gains g_p for each receiver chain which can be stacked in a $P \times 1$ column vector $\mathbf{g} = [g_1, g_2, \dots, g_P]^T$. The voltages at the output of the receiver chain network can thus be described by [128]

$$\begin{aligned}\mathbf{v}(\mathbf{k}, t) &= \mathbf{G}(f) \mathbf{Z}_R(f) (\mathbf{Z}_R(f) + \mathbf{Z}_A(f))^{-1} \mathbf{v}_{oc}(\mathbf{k}, t) \\ &= \mathbf{G}(f) \mathbf{Q}(f) \mathbf{v}_{oc}(\mathbf{k}, t)\end{aligned}\quad (3.3)$$

where $\mathbf{G} = \text{diag}(\mathbf{g})$. The matrix $\mathbf{Q} = \mathbf{Z}_R(\mathbf{Z}_R + \mathbf{Z}_A)^{-1}$ describes the effect of mutual coupling between the receiving elements of the array. The output of each element is a mixture of the response of the element itself and the scaled and delayed responses of the other elements in the array. The antenna array and its receiving network can also be described using the scattering matrix \mathbf{S} . This approach is followed in [130] and leads to a similar result, i.e. $\mathbf{v} = \mathbf{GQv}_{oc}$ is still an accurate description of the mutual coupling in the antenna array. The third representation is based on the mutual admittance matrix \mathbf{Y} . These representations are equivalent and if either \mathbf{Z}_A (and \mathbf{Z}_R), \mathbf{S} or \mathbf{Y} is known, the other two (and therefore \mathbf{Q}) can be calculated [61].

The output voltages of the receiver network $\mathbf{v}(t)$ are the result of the superposition of voltages induced by electromagnetic waves coming from all directions at all

frequencies. If these signals originate from celestial sources, they are stochastic in nature. Although each celestial source has its own power spectral density, their signal in individual narrowband [170] frequency channels may be modeled as independent identically distributed (i.i.d.) complex Gaussian noise. The fundamental reason for this is provided by quantum mechanics. From quantum mechanics, we know that electromagnetic radiation may not only be described as waves but also as particles, called photons. Each photon carries an energy $E = hf$ where $h = 6.626 \cdot 10^{-34}$ Js is the Planck constant. This implies that photons at radio frequencies, the lowest frequencies of the electromagnetic spectrum, carry very small amounts of energy. For example, a single photon at 178 MHz, the frequency at which the 3C catalog was observed [8], carries $1.18 \cdot 10^{-25}$ J. According to the 3C catalog, Cyg A has a flux of $8100 \text{ Jy} = 8100 \cdot 10^{-26} \text{ Wm}^{-2}\text{Hz}^{-1}$. If this source is observed by a LOFAR high band antenna tile with an effective area of 25 m^2 in a Nyquist sampled frequency channel, each time sample already represents $2.03 \cdot 10^{-21}$ J which corresponds to over 17000 photons. This number is sufficiently large to ensure that the Poisson distribution of a photon counting experiment closely resembles a Gaussian distribution. In reality, observers integrate over even longer time scales because the signal-to-noise ratio in a single sample is insufficient to detect even a strong astronomical source like Cyg A.

Due to the stochastic nature of the signals of interest, radio astronomical measurements are based on measuring the spatial coherency between antenna signals, the visibilities. This can either be done by first computing the crosscorrelation function in time before calculating the cross power spectrum or by applying a Fourier transform to the antenna signals first and then compute the cross power spectrum by channel-wise multiplication. A correlator architecture based on the first approach is an XF-type correlator while an architecture based on the second approach is an FX-type correlator. It follows from the Wiener-Kinchin theorem that these approaches are equivalent. A derivation of this theorem [83] shows that this theorem holds if the signals are jointly stationary and have a finite coherence length in time. Both requirements are satisfied by celestial signals, but not by, e.g., chirped radar signals.

Since correlation can be described by a multiplication in the frequency domain, we will assume that the narrowband condition [170] holds. This can be described mathematically by assuming an infinitely narrow frequency response of the system centered around f_0 , i.e.

$$\begin{aligned} \mathbf{v}(\mathbf{l}, f_0, t) &= \int_{-\infty}^{\infty} \mathbf{G}(f) \mathbf{Q}(f) \mathbf{v}_{oc}(f, \mathbf{l}, t) \delta(f - f_0) df \\ &= \mathbf{G}(f_0) \mathbf{Q}(f_0) \mathbf{v}_{oc}(f_0, \mathbf{l}, t), \end{aligned} \quad (3.4)$$

where $\delta(f - f_0)$ denotes the Dirac delta function. For notational convenience we

will stop mentioning the frequency dependence explicitly from this point. The output voltages $\mathbf{v}(t)$ can now be obtained by integrating (3.4) over direction, i.e.

$$\mathbf{v}(t) = \int_{-\pi}^{\pi} \int_0^{\pi/2} \mathbf{GQv}_{oc}(\mathbf{l}, t) |\sin \theta| d\theta d\phi = \int_{\Omega} \mathbf{GQv}_{oc}(\mathbf{l}, t) d\Omega. \quad (3.5)$$

The details of $\mathbf{v}_{oc}(\mathbf{l}, t)$ follow from (3.1) and (3.2). By splitting $\mathbf{E}_0(\mathbf{l}, t) = \mathbf{e}_0(\mathbf{l}, t) s(\mathbf{l}, t)$ in a unit polarization vector $\mathbf{e}_0(\mathbf{l}, t)$ and a source signal $s(\mathbf{l}, t)$, we can introduce a direction dependent gain for the p th element of the array

$$g_{0p}(\mathbf{l}, t) = \frac{4\pi j \xi_0 e^{j|\mathbf{k}|\xi_0}}{|\mathbf{k}| \eta I_0} \mathbf{e}_0(\mathbf{l}, t) \cdot \mathbf{E}_p(\mathbf{l}). \quad (3.6)$$

Although in this presentation, we only use direction dependent gains to describe the element beam patterns, additional complex valued multiplicative factors may be added to describe distortions of the incoming wave fronts due to, e.g., varying propagation conditions or near-field effects.

These gains can be stored in a $P \times 1$ column vector $\mathbf{g}_0(\mathbf{l}, t) = [g_{01}(\mathbf{l}, t), g_{02}(\mathbf{l}, t), \dots, g_{0P}(\mathbf{l}, t)]^T$. The phase of an incoming plane wave at the p th receiving element can be denoted by a phasor $a_p(\mathbf{l}, t) = \exp(-j(\mathbf{k}^T \boldsymbol{\xi}_p - |\mathbf{k}| ct))$. These phasors can be stacked in a $P \times 1$ column vector $\mathbf{a}(\mathbf{l}, t) = [a_1(\mathbf{l}, t), a_2(\mathbf{l}, t), \dots, a_P(\mathbf{l}, t)]^T$. If the receiver locations are stacked in a matrix $\boldsymbol{\Xi} = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_P]^T$, the array response vector $\mathbf{a}(\mathbf{l}, t)$ can be written in the convenient form

$$\mathbf{a}(\mathbf{l}, t) = \frac{1}{\sqrt{P}} e^{-j(\boldsymbol{\Xi k} - |\mathbf{k}| ct)}, \quad (3.7)$$

which is normalized such that $\|\mathbf{a}(\mathbf{l}, t)\|^2 = \mathbf{a}^H(\mathbf{l}, t) \mathbf{a}(\mathbf{l}, t) = 1$. Equation (3.5) can now be written as

$$\mathbf{v}(t) = \int_{\Omega} \mathbf{GQ}(\mathbf{a}(\mathbf{l}, t) \odot \mathbf{g}_0(\mathbf{l}, t)) s(\mathbf{l}, t) d\Omega. \quad (3.8)$$

The array covariance matrix $\mathbf{R} = \mathcal{E}\{\mathbf{v}(t) \mathbf{v}^H(t)\}$ is a matrix containing the expected value of the visibilities. Since the output voltages of the receiver network are a superposition of source signals and noise which are mutually uncorrelated, the array covariance matrix can be split in a noise term and a signal term. The noise term consists of a contribution from the homogenous sky background noise picked up by the receiving elements and the noise power of the receiver chains. The noise covariance matrix can therefore be modeled as [48]

$$\mathbf{R}_n = k_b T_{rec} B \mathbf{I} + k_b T_{sky} B \Psi \quad (3.9)$$

where T_{rec} and T_{sky} denote the receiver and sky noise temperature respectively, $k_b = 1.38 \cdot 10^{-23}$ J/K is the Boltzmann constant, B is the observing bandwidth and Ψ is the element pattern overlap matrix whose elements are given by [48, 130]

$$\Psi_{p_1 p_2} = \frac{1}{2\eta P_{el}} \int_{\Omega} \bar{\mathbf{E}}_{p_1}(\mathbf{l}) \cdot \mathbf{E}_{p_2}(\mathbf{l}) d\Omega. \quad (3.10)$$

This matrix describes the overlap or correlated power received by the array elements integrated over the homogenous sky background. The element pattern overlap matrix is normalized such that its diagonal elements are unity.

The contribution of the source signals to the array covariance matrix can be derived directly from (3.5):

$$\mathbf{R}_s = \mathcal{E} \left\{ \left(\int_{\Omega} \mathbf{G} \mathbf{Q} \mathbf{v}_{oc}(\mathbf{l}, t) d\Omega \right) \left(\int_{\Omega} \mathbf{G} \mathbf{Q} \mathbf{v}_{oc}(\mathbf{l}, t) d\Omega \right)^H \right\}. \quad (3.11)$$

If the source signals are spatially uncorrelated, i.e. the signal coming from \mathbf{l} is not correlated with the signal coming from $\mathbf{l}' \neq \mathbf{l}$, (3.11) reduces to

$$\mathbf{R}_s = \mathbf{G} \mathbf{Q} \left(\int_{\Omega} \mathcal{E} \{ \mathbf{v}_{oc}(\mathbf{l}, t) \mathbf{v}_{oc}^H(\mathbf{l}, t) \} d\Omega \right) \mathbf{Q}^H \mathbf{G}^H. \quad (3.12)$$

Using the factorization of $\mathbf{v}_{oc}(\mathbf{l}, t)$ introduced in (3.8) and taking expected values, this can be written as

$$\mathbf{R}_s = \mathbf{G} \mathbf{Q} \left(\int_{\Omega} (\mathbf{a}(\mathbf{l}) \odot \mathbf{g}_0(\mathbf{l})) \boldsymbol{\sigma}(\mathbf{l}) (\mathbf{a}(\mathbf{l}) \odot \mathbf{g}_0(\mathbf{l}))^H d\Omega \right) \mathbf{Q}^H \mathbf{G}^H, \quad (3.13)$$

where $\boldsymbol{\sigma}(\mathbf{l}) = \mathcal{E} \{ s(\mathbf{l}, t) \bar{s}(\mathbf{l}, t) \}$.

If the signals incident on the array originate from a number of discrete sources, the surface integral can be replaced by a summation over the Q source signals. In practice, we always probe the sky with a finite resolution, so the number of resolution elements is limited. A discrete source model can therefore handle all practical situations, even in the presence of extended emission. We can thus write

$$\mathbf{R}_s = \mathbf{G} \mathbf{Q} \left(\sum_{q=1}^Q (\mathbf{a}_q \odot \mathbf{g}_{0q}) \boldsymbol{\sigma}_q (\mathbf{a}_q \odot \mathbf{g}_{0q})^H \right) \mathbf{Q}^H \mathbf{G}^H. \quad (3.14)$$

Introducing the $P \times Q$ matrices $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_Q]$ and $\mathbf{G}_0 = [\mathbf{g}_{01}, \mathbf{g}_{02}, \dots, \mathbf{g}_{0Q}]$, the $Q \times 1$ column vector $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \dots, \sigma_Q]^T$ and the $Q \times Q$ matrix $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma})$, this can be simplified further to

$$\mathbf{R}_s = \mathbf{G} \mathbf{Q} (\mathbf{A} \odot \mathbf{G}_0) \boldsymbol{\Sigma} (\mathbf{A} \odot \mathbf{G}_0)^H \mathbf{Q}^H \mathbf{G}^H. \quad (3.15)$$

Since the signals at the receiver network output terminals are a superposition of the source and noise signals received by the antennas and the noise introduced by the receiver path electronics, the array covariance matrix at these output terminals is described by the addition of \mathbf{R}_n given by (3.9) and \mathbf{R}_s given by (3.15), i.e.

$$\mathbf{R} = \mathbf{GQ} (\mathbf{A} \odot \mathbf{G}_0) \Sigma (\mathbf{A} \odot \mathbf{G}_0)^H \mathbf{Q}^H \mathbf{G}^H + k_b T_{sky} B \Psi + k_b T_{rec} B \mathbf{I}. \quad (3.16)$$

3.2 Signal processing models

Although (3.16) provides a complete description of a radio astronomical phased array based on the physics of the system, it is more complex than required in most practical applications, in which simplifying assumptions can often be made. We will thus make a transition from the physical description in the previous section to the representation of array systems commonly used in the array signal processing literature by discretizing the signal in time to model the sampling by an ADC and introducing Jones matrices to describe the response of a dual polarized antenna to the incoming signals. These steps will facilitate a comparison with other array signal processing models.

3.2.1 Short term covariance matrix

If the signal is sampled with period T , the n th time sample of the array signal vector $\mathbf{x}[n]$ follows from (3.5) by

$$\mathbf{x}[n] = \int_{-\infty}^{\infty} \mathbf{v}(t) \delta(t - nT) dt = \mathbf{v}(nT). \quad (3.17)$$

As argued in the previous section, this signal can be modeled by a superposition of mutually independent source and noise signals. Equation (3.8) provides an accurate model for the source signals while the sampled noise signals may be represented by $\mathbf{n}(t)$ thereby leaving out the details of their origin for the moment. The array signal vector can thus be described by

$$\mathbf{x}[n] = \int_{\Omega} \mathbf{GQ} (\mathbf{a}(\mathbf{l}, nT) \odot \mathbf{g}_0(\mathbf{l}, nT)) s(\mathbf{l}, nT) d\Omega + \mathbf{n}(nT) \quad (3.18)$$

or, after discretizing the source model as demonstrated in the previous section, by

$$\mathbf{x}[n] = \mathbf{GQ} \left(\sum_{q=1}^Q (\mathbf{a}_q(nT) \odot \mathbf{g}_{0q}(nT)) s_q(nT) \right) + \mathbf{n}(nT). \quad (3.19)$$

The Q source signals can be stacked in a $Q \times 1$ source vector $\mathbf{s}(nT) = [s_1(nT), s_2(nT), \dots, s_Q(nT)]^T$. Using the matrices \mathbf{A} and \mathbf{G}_0 as defined in the previous section, we can now rewrite (3.19) as

$$\mathbf{x}[n] = \mathbf{G}\mathbf{Q}(\mathbf{A} \odot \mathbf{G}_0)\mathbf{s}(nT) + \mathbf{n}(nT). \quad (3.20)$$

N samples can be stacked into a $P \times N$ matrix $\mathbf{X} = [\mathbf{x}[1], \mathbf{x}[2], \dots, \mathbf{x}[N]]$, which is a data set used for short term integration. Over this short term integration interval, often referred to as a snapshot, the array covariance matrix $\mathbf{R} = \mathcal{E}\{\mathbf{x}[n]\mathbf{x}^H[n]\}$ is estimated by the short term covariance matrix estimate

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{X}\mathbf{X}^H, \quad (3.21)$$

which is computed by a correlator and stored for later use. Its expected value based on (3.20) is

$$\mathbf{R} = \mathbf{G}\mathbf{Q}(\mathbf{A} \odot \mathbf{G}_0)\Sigma(\mathbf{A} \odot \mathbf{G}_0)^H\mathbf{Q}^H\mathbf{G}^H + \Sigma_n \quad (3.22)$$

where $\Sigma = \mathcal{E}\{\mathbf{s}(nT)\mathbf{s}^H(nT)\}$ is the $Q \times Q$ source covariance matrix and $\Sigma_n = \mathcal{E}\{\mathbf{n}(nT)\mathbf{n}^H(nT)\}$ is the $P \times P$ noise covariance matrix. The reader should note that (3.22) has the same form as (3.16) with $\Sigma_n = k_b T_{sky} B\Psi + k_b T_{sys} B\mathbf{I}$. One should also realize that, although it is often assumed that Σ and Σ_n are diagonal matrices, this is not a restriction imposed by this model. However, parametric constraints on these matrices are often required in practical estimation problems.

3.2.2 Polarimetry

Up to this point, the array system has been described by an array signal vector containing the signals from every individual receiving element in the array. In this description each receiving element is treated the same way. If the array has been designed to fully reconstruct the polarization of the incident electromagnetic field, the array generally consists of antenna elements, each consisting of two (quasi-)orthogonal feeds. Hamaker, Bregman and Sault [45, 99] recognized that one should consider the behavior of these receiving element pairs to fully understand measurements on polarized sources. We will therefore return to (3.6) to see how pairing of the receiving elements affects the array signal vector and the array covariance matrix.

In the Sec. 3.1 we introduced the direction dependent gain per element $g_{0p}(\mathbf{l}, t)$ by taking the inner product of the polarized element voltage beam pattern $\mathbf{E}_p(\mathbf{l})$ and the polarization of the incoming plane wave $\mathbf{e}_0(\mathbf{l}, t)$ (see (3.6)). If we (arbitrarily) label one receiving element of the p th antenna as x -element and the other as y -element, we can stack the direction dependent gains of both elements in a 2×1 column vector

$$\mathbf{g}_{0p}(t) = \begin{bmatrix} g_{0px} \\ g_{0py} \end{bmatrix} = \frac{4\pi j\xi_0 e^{j|\mathbf{k}|\xi_0}}{|\mathbf{k}| \eta I_0} \begin{bmatrix} \mathbf{E}_{px}(\mathbf{l}) \cdot \mathbf{e}_0(\mathbf{l}, t) \\ \mathbf{E}_{py}(\mathbf{l}) \cdot \mathbf{e}_0(\mathbf{l}, t) \end{bmatrix}. \quad (3.23)$$

This can be conveniently written as

$$\mathbf{g}_{0p}(t) = \frac{4\pi j\xi_0 e^{j|\mathbf{k}|\xi_0}}{|\mathbf{k}| \eta I_0} \begin{bmatrix} \mathbf{E}_{px}^T(\mathbf{l}) \\ \mathbf{E}_{py}^T(\mathbf{l}) \end{bmatrix} \mathbf{e}_0(\mathbf{l}, t) = \mathbf{J}_p(\mathbf{l}) \mathbf{e}_0(\mathbf{l}, t) \quad (3.24)$$

where we have introduced a 2×2 Jones matrix for the p th antenna $\mathbf{J}_p(\mathbf{l})$.

It is now straightforward to construct a model analogous to (3.19) for the array signal vector of an array of dual polarized antennas. Define $\mathbf{x}[n] = [x_{1x}[n], x_{1y}[n], x_{2x}[n], x_{2y}[n], \dots, x_{Px}[n], x_{Py}[n]]^T$, i.e. stack the signals from the respective x - and y -elements alternatingly. Stack the direction independent gains in the same way, i.e. let $\mathbf{g} = [g_{1x}, g_{1y}, g_{2x}, g_{2y}, \dots, g_{Px}, g_{Py}]^T$ so that $\mathbf{G} = \text{diag}(\mathbf{g})$. The matrix \mathbf{Q} remains essentially unaltered except that it has now size $2P \times 2P$. The array response vector $\mathbf{a}_q(nT)$ should be replaced by a $2P \times 2$ matrix $\mathbf{A}_q(nT)$ with two identical columns since the two gain factors of each receiving element for the respective components of the electromagnetic field should be multiplied by the same phasor to compensate for the geometrical delay. The ordering of the elements in the columns should be the same as in $\mathbf{x}[n]$ and \mathbf{g} . If the x -feeds and y -feeds of all antennas are co-located, it is sufficient to know the array response vector of the either the array of x -elements $\mathbf{a}_{qx}(nT)$ or the array of y -elements $\mathbf{a}_{qy}(nT)$, since the $2P \times 2$ array response matrix for the full array can now be constructed by $\mathbf{A}_q(nT) = \mathbf{a}_{qx}(nT) \otimes \mathbf{1}_{2 \times 2}$, where $\mathbf{1}_{2 \times 2}$ denotes a 2×2 matrix filled with ones. The direction dependent gain vector \mathbf{g}_{0q} is obtained by stacking $\mathbf{g}_{0p}(nT)$, i.e. $\mathbf{g}_q(nT) = [\mathbf{J}_{1q}^T, \mathbf{J}_{2q}^T, \dots, \mathbf{J}_{Pq}^T]^T \mathbf{e}_{0q}(nT) = \mathbf{J}_q \mathbf{e}_{0q}(nT)$. Note that, with some abuse of notation, we have defined the $2P \times 2$ matrix \mathbf{J}_q in which we have stacked the Jones matrices of the individual antennas, \mathbf{J}_{pq} . Finally, $\mathbf{n}(nT)$ also becomes a $2P \times 1$ vector with its element arranged in the same order as $\mathbf{x}[n]$. Putting all these definitions together, we may reformulate (3.19) as

$$\mathbf{x}[n] = \mathbf{G}\mathbf{Q} \left(\sum_{q=1}^Q (\mathbf{A}_q(nT) \odot \mathbf{J}_q) \mathbf{e}_{0q}(nT) s_q(nT) \right) + \mathbf{n}(nT). \quad (3.25)$$

The signal from the q th source is measured as a series of time samples characterized by their amplitude $s_q(nT)$ and their polarization $\mathbf{e}_{0q}(nT) = e_\theta \mathbf{e}_\theta(nT) + e_\phi \mathbf{e}_\phi(nT)$. Since these samples originate from a stochastic process, we need to look for their stochastic properties. Hamaker, Bregman and Sault [45, 46, 99] introduced the concept of source coherency, which is the 2×2 covariance matrix of the dual polarized source signal, i.e.

$$\mathbf{E}_q = \mathcal{E} \{ \mathbf{e}_{0q}(nT) s_q(nT) \bar{s}_q(nT) \mathbf{e}_{0q}^H(nT) \} = \begin{bmatrix} E_q^{\theta\theta} & E_q^{\theta\phi} \\ E_q^{\phi\theta} & E_q^{\phi\phi} \end{bmatrix}. \quad (3.26)$$

With this definition, we can formulate the model for the short term array covariance matrix as

$$\mathbf{R} = \mathbf{GQ} \left(\sum_{q=1}^Q (\mathbf{A}_q \odot \mathbf{J}_q) \mathbf{E}_q (\mathbf{J}_q \odot \mathbf{A}_q)^H \right) \mathbf{Q}^H \mathbf{G}^H + \boldsymbol{\Sigma}_n \quad (3.27)$$

assuming that the sources are mutually independent. By defining the source covariance matrix as the $2Q \times 2Q$ block diagonal matrix $\boldsymbol{\Sigma} = \text{bdiag}(\mathbf{E}_q)$ with the source coherencies \mathbf{E}_q on the main diagonal and stacking the array response matrices in a $2P \times 2Q$ matrix $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_q]$ and the corresponding Jones matrices in a $2P \times 2Q$ matrix $\mathbf{G}_0 = [\mathbf{J}_1, \mathbf{J}_2, \dots, \mathbf{J}_Q]$, this can be rewritten as

$$\mathbf{R} = \mathbf{GQ} (\mathbf{A} \odot \mathbf{G}_0) \boldsymbol{\Sigma} (\mathbf{A} \odot \mathbf{G}_0)^H \mathbf{Q}^H \mathbf{G}^H + \boldsymbol{\Sigma}_n \quad (3.28)$$

which has exactly the same form as (3.22). Note that although the assumption that the source signals are independent was introduced in an intermediate step to improve the readability of this derivation, the final result can easily handle correlated sources.

Each source is characterized by its coherency matrix $\mathbf{E}_q^{\theta\phi}$, where the superscript has been added to emphasize that the polarization has been characterized by the field components along \mathbf{e}_θ and \mathbf{e}_ϕ . For celestial objects, astronomers generally use right ascension α and declination δ as the principal axis of the coordinate system. In that case, the source coherencies could be denoted as $\mathbf{E}_q^{\alpha\delta}$. Another important coherency matrix is $\mathbf{E}_{p_1 p_2}^{xy}$, the coherency on the baseline from antenna p_1 to antenna p_2 . Ignoring the receiver gain, mutual coupling and geometrical delay (or assuming they are appropriately compensated) $\mathbf{E}_{p_1 p_2}^{xy}$ is related to $\mathbf{E}_q^{\theta\phi}$ by the Jones matrices of the two antennas involved

$$\mathbf{E}_{p_1 p_2}^{xy} = \mathbf{J}_{p_1 q} \mathbf{E}_q^{\theta\phi} \mathbf{J}_{p_2 q}^H. \quad (3.29)$$

These Jones matrices describe the two voltage beam patterns of the two feeds of the antenna.

Section 6.3 discusses least squares synthesis imaging. If this is applied to polarized data, we have to estimate four real valued parameters per pointing direction in the image to describe the source coherency in that direction $\mathbf{E}_q^{\theta\phi}$. Since $\mathbf{E}_q^{\theta\phi}$ is Hermitian, an intuitive choice for a real valued coherency vector is

$$\mathbf{e}_q = \left[\mathbf{E}_{q11}^{\theta\phi}, \text{Re}(\mathbf{E}_{q12}^{\theta\phi}), \text{Im}(\mathbf{E}_{q12}^{\theta\phi}), \mathbf{E}_{q22}^{\theta\phi} \right]^T. \quad (3.30)$$

The measured visibility based coherencies $\mathbf{e}_{p_1 p_2} = \text{vec}(\mathbf{E}_{p_1 p_2}^{xy})$ produced by the q th source are related to the source coherency \mathbf{e}_q through the instrumental response towards the q th source that can be described by a 4×4 beam polarization matrix $\mathbf{B}_{p_1 p_2 q}$, such that

$$\mathbf{e}_{p_1 p_2} = \mathbf{B}_{p_1 p_2 q} \mathbf{e}_q, \quad (3.31)$$

which is merely the vectorized form of (3.29).

3.2.3 Commonly used data models

For the development of the data model presented here the work of Warnick et al. [48, 128–130] and Lanne [61] was used as starting point. This model was augmented with the theory of polarimetry developed by Hamaker, Bregman and Sault [45,46,99]. This route was chosen to obtain a generalized data model that covers all viable array signal processing models proposed in the literature. Such a generalized data model is quite cumbersome to work with if one’s interests lie with a particular aspect of the system. Most studies therefore assume a simplified version of the data model, for example by ignoring mutual coupling or assuming a calibrated array. This section presents a brief overview of the assumptions commonly made in the literature.

Although mutual coupling may have significant impact on the array response in dense phased arrays, its impact on array signal processing is often ignored, except by a few authors. Svantesson and Viberg [112] and Svantesson [111] have proposed modified versions of commonly used algorithms like multiple signal classification (MUSIC) [101] and estimation of signal parameters via rotational invariance techniques (ESPRIT) [95] to handle arrays affected by mutual coupling under the assumption that the array is calibrated, i.e. $\mathbf{G} = \mathbf{I}$. However, high resolution DOA estimation techniques are known to be very sensitive to calibration errors. A number of recent studies therefore focuses on phased array calibration and DOA estimation in the presence of mutual coupling. Examples are Lanne [61] and Warnick et al. [48, 128–130].

Ignoring mutual coupling implies $\mathbf{Q} = \mathbf{I}$ which simplifies (3.22) to

$$\mathbf{R} = \mathbf{G} (\mathbf{A} \odot \mathbf{G}_0) \Sigma (\mathbf{A} \odot \mathbf{G}_0)^H \mathbf{G}^H + \boldsymbol{\Sigma}_n, \quad (3.32)$$

where $\boldsymbol{\Sigma}$ is commonly assumed to be diagonal and \mathbf{G}_0 can be a full matrix. In this case \mathbf{G} can be absorbed in \mathbf{G}_0 , which then contains complex gain factors per receiving element and per source. This model is adopted by Van der Tol, Jeffs and Van der Veen [120] to describe the direction dependent calibration of LOFAR. They recognize that a calibration routine allowing for one complex gain factor per station per source has too many degrees of freedom to result in a unique solution. They solved this ambiguity by exploiting the spatial, spectral and temporal continuity of these gain factors. Wijnholds and Van der Veen [165] studied this problem under the constraint that all receiving elements have the same directional response, an assumption also made by Weiss and Friedlander [134]. Calibration assuming this scenario will be discussed extensively in Ch. 5.

The fact that a single complex gain factor per visibility does not lead to a meaningful solution because such a model will fit to any data, is recognized by Cornwell

and Fomalont [26], who describe the self-calibration assumption, which states that all gain effects in a radio telescope system are telescope based. This reduces the data model to

$$\mathbf{R} = \mathbf{G}\mathbf{A}\Sigma\mathbf{A}^H\mathbf{G}^H + \Sigma_n, \quad (3.33)$$

where Σ is often assumed to be diagonal. This model is not as restrictive as it may seem, since in self-calibration the source structure described by Σ and the direction independent gains \mathbf{G} are estimated simultaneously. Since the source structure is assumed unknown, Σ effectively absorbs the direction dependent gains \mathbf{G}_0 . The data model described by (3.33) is commonly used in papers on array calibration [12, 37, 39, 90, 149].

Studies on receiving element position calibration, DOA estimation or imaging parameterize the array response matrix \mathbf{A} and either absorb the direction independent gains \mathbf{G} in this matrix or assume that the array is calibrated. This leads to

$$\mathbf{R} = \mathbf{A}(\boldsymbol{\theta})\Sigma\mathbf{A}^H(\boldsymbol{\theta}) + \Sigma_n. \quad (3.34)$$

This model is very suitable for analyzing methods based on subspace techniques [124], which include high resolution DOA estimation techniques such as MUSIC [101], ESPRIT [95] and weighted subspace fitting (WSF) [125]. This model also lends itself well for the development and analysis of detectors for source signals [65, 91, 92, 131].

3.3 Noise subspace and signal subspace

Eigenvalue decomposition has proven to be a very powerful technique for the analysis of array covariance matrices as illustrated by subspace based techniques for parameter fitting [124], high resolution DOA estimation [101, 125] and source detection [14, 91, 92]. Some of these techniques may be used to determine the number of sources that has to be included in the source model used for calibration or to estimate the source positions, as suggested in Ch. 5. This section therefore provides some background on the eigenvalues and eigenvectors on the array covariance matrix.

The eigenvalue decomposition of the $P \times P$ array covariance matrix

$$\mathbf{R} = \mathbf{V}\Lambda\mathbf{V}^H \quad (3.35)$$

decomposes the array covariance matrix in a set of orthonormal eigenvectors stacked in a full rank $P \times P$ matrix \mathbf{V} and their corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_P$ which form the main diagonal of the $P \times P$ diagonal matrix Λ . Based on the distribution of eigenvalues this decomposition is often conceptually split in eigenvectors

associated with the source signals and eigenvectors associated with the noise in the measurement, i.e.

$$\mathbf{R} = \mathbf{V}_s \boldsymbol{\Lambda}_s \mathbf{V}_s^H + \mathbf{V}_n \boldsymbol{\Lambda}_n \mathbf{V}_n^H. \quad (3.36)$$

The space spanned by the Q signal eigenvectors stacked in the $P \times Q$ matrix \mathbf{V}_s is called the signal subspace while its orthogonal complement associated with the noise eigenvectors stacked in the $P \times (P - Q)$ matrix \mathbf{V}_n is referred to as noise subspace.

In the remainder of this section, I present a number of simple examples to illustrate the physical significance of the eigenvalues and eigenvectors of the array covariance matrix. First, consider the array covariance matrix for a single source with power σ_1 and array response vector \mathbf{a}_1 observed by a noise free perfectly calibrated system

$$\mathbf{R} = \sigma_1 \mathbf{a}_1 \mathbf{a}_1^H. \quad (3.37)$$

This model has rank 1. Since $\|\mathbf{a}_1\|^2 = \mathbf{a}_1^H \mathbf{a}_1 = 1$, the eigenvector associated with this source will be $\mathbf{v}_1 = \mathbf{a}_1$ and the corresponding eigenvalue will be $\lambda_1 = \sigma_1$. All other eigenvalues will be 0, since we assumed a noise free scenario. The eigenvalue decomposition will thus yield

$$\mathbf{R} = \mathbf{v}_1 \sigma_1 \mathbf{v}_1^H + \mathbf{V}_n \text{diag}(\mathbf{0}) \mathbf{V}_n^H. \quad (3.38)$$

This simple scenario becomes a little more realistic if we add receiver gains and mutual coupling to our model, i.e. if we expand (3.37) to

$$\mathbf{R} = \sigma_1 \mathbf{G} \mathbf{Q} \mathbf{a}_1 \mathbf{a}_1^H \mathbf{Q}^H \mathbf{G}^H. \quad (3.39)$$

This model may actually be used to model the response of a very strong point source. After introducing $\mathbf{a}'_1 = \mathbf{G} \mathbf{Q} \mathbf{a}_1$, we can rewrite (3.39) as

$$\mathbf{R} = \sigma_1 \mathbf{a}'_1 \mathbf{a}'_1^H, \quad (3.40)$$

which shows that the receiver gains and mutual coupling do not increase the rank of the source model, although they do alter the associated eigenvalues and eigenvectors, since the norm of \mathbf{a}'_1 may not be equal to 1.

From these examples, one may get the impression that the eigenvectors in the source subspace are directly related to the array response vectors of the individual sources \mathbf{a}_q , which may include the receiver gains and mutual coupling, and that the corresponding eigenvalues provide information on the source powers. Unfortunately, this relation breaks down as soon as there are two or more sources involved. The simple argument is that the array response vectors of the sources are not always orthogonal while the eigenvectors are orthonormal by definition. There can therefore be no one-to-one relation between the array response vectors and the eigenvector spanning the signal subspace.

The fact that there is no one-to-one correspondence between the eigenvalue decomposition of the signal subspace and the sources, is also nicely illustrated by the eigenvalues for a rank 2 model for which an analytic solution exists. Consider the rank 2 model

$$\mathbf{R} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^H = [\mathbf{a}_1, \mathbf{a}_2] \begin{bmatrix} \sigma_1 & \sigma_{12} \\ \sigma_{21} & \sigma_2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1^H \\ \mathbf{a}_2^H \end{bmatrix}. \quad (3.41)$$

The solution for the two non-zero eigenvalues is given by [170]

$$\begin{aligned} \lambda_{1,2} &= \left(\frac{1}{2} (\sigma_1 + \sigma_2) + \operatorname{Re} \{ \sigma_{12} \mathbf{a}_2^H \mathbf{a}_1 \} \right) \\ &\times \left(1 \pm \sqrt{1 - 4 \frac{(1 - |\mathbf{a}_1^H \mathbf{a}_2|^2)(\sigma_1 \sigma_2 - |\sigma_{12}|^2)}{(\sigma_1 + \sigma_2 + 2\operatorname{Re} \{ \sigma_{12} \mathbf{a}_2^H \mathbf{a}_1 \})^2}} \right). \end{aligned} \quad (3.42)$$

This result shows that even if the signals are uncorrelated, i.e. $\sigma_{12} = \sigma_{21} = 0$, the eigenvalues are not equal to the source powers unless $\mathbf{a}_1^H \mathbf{a}_2 = 0$, i.e. unless the two array response vectors are orthogonal. Only in that case, the eigenvector corresponding to the largest eigenvalue will be aligned with the array response vector of the strongest source and the other eigenvector with the array response vector of the other source.

Finally, we consider the impact of noise on the eigenvalue structure. Suppose we have an array with equal but uncorrelated system noise powers on all receivers and that the eigenvalue decomposition of the array covariance matrix before addition of the noise is $\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$, i.e.

$$\mathbf{R} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H + \sigma_n \mathbf{I}. \quad (3.43)$$

Then the eigenvalue decomposition will yield

$$\mathbf{R} = \mathbf{V}(\boldsymbol{\Lambda} + \sigma_n \mathbf{I})\mathbf{V}^H. \quad (3.44)$$

This shows that the eigenvectors remain the same. The corresponding eigenvalues, however, are augmented by the addition of σ_n . This property of the eigenvalue decomposition follows directly from the orthonormality of the eigenvectors; since $\mathbf{V}\mathbf{V}^H = \mathbf{I}$, it is easily seen that $\mathbf{V}\sigma_n \mathbf{I}\mathbf{V}^H = \sigma_n \mathbf{I}$. This also shows that if the system noise varies per receiver or is correlated between receivers, the eigenvectors are affected, since

$$\boldsymbol{\Sigma}_n^{-1/2} (\mathbf{R}_s + \boldsymbol{\Sigma}_n) \boldsymbol{\Sigma}_n^{-1/2} = \boldsymbol{\Sigma}_n^{-1/2} \mathbf{R}_s \boldsymbol{\Sigma}_n^{-1/2} + \mathbf{I} \neq \mathbf{V}_s \boldsymbol{\Lambda}_s \mathbf{V}_s^H + \mathbf{I} \quad (3.45)$$

Equation (3.44) suggests that it is fairly simple to discern the signal subspace from the noise subspace if the system noise powers of the individual receivers are the same and uncorrelated. Unfortunately, the analysis above only applies to the expected value of the array covariance matrix. In practice, the noise on the data introduces a structure in the estimated eigenvalues that makes it more difficult to discern the eigenvalues associated with the signal subspace from those associated with the noise subspace. Edelman [32] has made a thorough analysis of the eigenvalues and condition numbers of real and complex valued random matrices, including matrices of the form $\mathbf{R} = \sigma_n \mathbf{I}$ for which he found that the largest and smallest eigenvalues, λ_1 and λ_P , of short term realizations of N samples are approximately given by

$$\lambda_{1,P} = \sigma_n \left(1 \pm \sqrt{\frac{P}{N}} \right)^2. \quad (3.46)$$

This shows that the largest eigenvalue tends to be somewhat larger than the expected σ_n while the smallest eigenvalue is somewhat lower. The symmetry of the largest and smallest eigenvalue around the expected value suggests a linear relation between the P eigenvalues. Although this works reasonably well for large N , the actual relation is best described by an exponential function. Quinlan et al. [91] empirically established that the p th eigenvalue can be derived from the first by

$$\lambda_p = \lambda_1 (e^{-2a})^{p-1}, \quad (3.47)$$

where

$$\lambda_1 = \sigma_n \left(\frac{1 - e^{-2a}}{1 - e^{-2aP}} \right) P \quad (3.48)$$

$$a = \sqrt{\frac{15}{2(P^2 + 2)} \left(1 - \sqrt{1 - \frac{4P(P^2 + 2)}{5N(P^2 - 1)}} \right)}. \quad (3.49)$$

3.4 Narrowband assumption

In theoretical derivations it is convenient to assume an infinitely small bandwidth, but practical systems will always produce spectrally separated data with finite channel width. Fortunately, we may still apply a single frequency model if the system satisfies the narrowband condition. In view of its importance, this assumption is discussed in detail. This discussion provides insight in the physical relevance of this assumption and leads to a quantitative relation between the decorrelation (gain) loss, the array

size, the channel width and the zenith angle of observation that is useful in the design of radio telescope systems.

The main reason for invoking the narrowband condition is that the differences in time of arrival of the incoming signals at different receiving elements, the geometrical delays, can be modeled by a phase rotation of the complex signals, i.e. by a complex multiplication. The theoretical basis is provided by the shift theorem for the Fourier transform [83]. The constraint that a signal with bandwidth Δf should not infringe the applicability of the shift theorem, leads to the commonly used narrowband criterion [116, 135]

$$\Delta f \ll (2\pi\tau)^{-1} \quad (3.50)$$

where τ is the time delay over which the signal should be coherently shiftable, i.e. the coherence time. The associated physical distance $c\tau$ is called the coherence length of the signal. This suggests that the bandwidth requirement imposed on the antenna array depends on the size of the array.

This point is clearly demonstrated by an analysis of the impact of bandwidth on the structure of the array covariance matrix as performed independently by Zatman [170] and Boonstra [14]. Zatman defines narrowband as the regime in which the second eigenvalue of a single source model covariance matrix is lower than the combined noise level of the source covariance matrix and the noise covariance matrix given the SNR of the measurement. In his analysis he models the array covariance matrix with two sources, each representing half the source power, at the same location but separated in frequency. Mathematically this can be formulated as

$$\mathbf{R} = \frac{1}{2}\sigma_s (\mathbf{a}_1\mathbf{a}_1^H + \mathbf{a}_2\mathbf{a}_2^H) + \sigma_n \mathbf{I} \quad (3.51)$$

where

$$\mathbf{a}_1 = \frac{1}{\sqrt{P}} e^{-j\frac{2\pi f_0}{c} \mathbf{\Xi} \mathbf{l}_s} \quad (3.52)$$

$$\mathbf{a}_2 = \frac{1}{\sqrt{P}} e^{-j\frac{2\pi(f_0 + \Delta f)}{c} \mathbf{\Xi} \mathbf{l}_s} = \mathbf{a}_1 \odot e^{-j\frac{2\pi\Delta f}{c} \mathbf{\Xi} \mathbf{l}_s}. \quad (3.53)$$

He then derives a relation to determine when statistically efficient high resolution DOA estimators will start to interpret this single broad band source as two individual sources at two different locations. Intuitively, the broadband source modeled by these two sources will be interpreted as two sources if the phase difference between the two array signal vectors represents a vector larger than the system noise. Equation (3.53) shows that this will not only depend on the SNR of the measurement, but also on the size of the array, since the inner products of receiver locations with the source

location $\Xi \mathbf{l}_s$ will produce larger values for larger arrays and therefore smaller values of Δf are sufficient to produce the same phase change.

This problem can also be studied by writing the data model for a single source as

$$\mathbf{R} = \frac{\sigma_s}{2\Delta f} \int_{f_0-\Delta f}^{f_0+\Delta f} \mathbf{a}(f) \mathbf{a}^H(f) df + \sigma_n \mathbf{I}, \quad (3.54)$$

thus by explicitly including the integration over the observing bandwidth. Boonstra [14] followed this approach using a second order Taylor expansion for $\mathbf{a}(f)$ which is given by

$$\mathbf{a}(f) \mathbf{a}_0^{(0)} + (f - f_0) \mathbf{a}_0^{(1)} + \frac{1}{2} (f - f_0)^2 \mathbf{a}_0^{(2)} \quad (3.55)$$

with Taylor coefficients

$$\mathbf{a}_0^{(0)} = \frac{1}{\sqrt{P}} e^{-j \frac{2\pi f_0}{c} \Xi \mathbf{l}_s} \quad (3.56)$$

$$\mathbf{a}_0^{(1)} = -j \frac{2\pi}{c} \Xi \mathbf{l}_s \odot \mathbf{a}_0^{(0)} \quad (3.57)$$

$$\mathbf{a}_0^{(2)} = -\frac{4\pi^2}{c^2} \Xi \mathbf{l}_s \odot \Xi \mathbf{l}_s \odot \mathbf{a}_0^{(0)}. \quad (3.58)$$

Equation (3.54) can now be approximated by the two source single frequency model

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

where

$$\sqrt{\sigma_1} \mathbf{a}_1 = \sqrt{\sigma_s} \left(\mathbf{a}_0^{(0)} + \frac{1}{6} \Delta f^2 \mathbf{a}_0^{(2)} \right) \quad (3.60)$$

$$\sqrt{\sigma_2} \mathbf{a}_2 = \frac{\sqrt{\sigma_s}}{\sqrt{3}} \Delta f \mathbf{a}_0^{(1)} \quad (3.61)$$

such that $\| \mathbf{a}_1 \|^2 = \| \mathbf{a}_2 \|^2 = 1$. Since (3.59) assumes a two source model, the signal subspace can be described by just two eigenvectors. The values of the corresponding eigenvalues can therefore be found using (3.42). Since \mathbf{a}_1 and \mathbf{a}_2 can be expressed in terms of Taylor coefficients whose phase and gain depend on the inner product of the antenna and source locations, $\Xi \mathbf{l}_s$, this alternative derivation leads to the same conclusion: the maximum allowable bandwidth to satisfy the narrowband condition not only depends on the SNR but also on the size of the array.

This can be quantified by considering the case in which a single signal at frequency f_0 from direction \mathbf{l}_0 impinges on the array. To compensate for the geometrical

delays between time of arrival at different antennas in the array, we have to weight the incoming signals with the conjugate of $\mathbf{w} = \mathbf{a}(l_0, f_0)$. If we apply these weights to a signal coming from the same direction at frequency $f_0 + \Delta f$, we find that the weighted sum over all array elements is

$$\begin{aligned}\mathbf{w}^H \mathbf{a} &= \frac{1}{P} \left(e^{-j\frac{2\pi f_0}{c} \Xi l_0} \right)^H \left(e^{-j\frac{2\pi f}{c} \Xi l_0} \right) \\ &= \frac{\mathbf{1}^T}{P} \left(e^{j\frac{2\pi f_0}{c} \Xi l_0} \odot e^{-j\frac{2\pi(f_0 + \Delta f)}{c} \Xi l_0} \right) \\ &= \frac{\mathbf{1}^T}{P} e^{-j\frac{2\pi \Delta f}{c} \Xi l_0}\end{aligned}\quad (3.62)$$

instead of 1 due to the phase mismatch caused by the frequency mismatch. Assuming that the narrowband condition holds in practical situations thus implies a gain loss at all in-band frequencies other than the reference frequency. The gain loss at frequency $f \neq f_0$ is given by

$$D(f) = 1 - |\mathbf{w}^H(f_0) \mathbf{a}(f)|^2. \quad (3.63)$$

This gain loss is often referred to as decorrelation loss.

In view of the discussion above, one would expect the decorrelation loss to depend on the array geometry. It is therefore instructive to derive an approximate formula for the decorrelation loss over a 1-D uniformly sampled array aperture, in which we can indicate positions by the scalar x . In general, if the aperture is spatially weighted by a window $w(x)$, the decorrelation loss can be calculated as

$$D(f) = 1 - \left| \frac{\int_{-\infty}^{\infty} e^{-j\frac{2\pi(f-f_0)}{c} xl_0} w(x) dx}{\int_{-\infty}^{\infty} w(x) dx} \right|^2 \quad (3.64)$$

where we have substituted (3.62). In the case of a uniformly tapered finite aperture, $w(x)$ becomes a rectangular window function. Therefore

$$D(f) = 1 - \left| \frac{\int_{-x_{max}}^{x_{max}} e^{-j\frac{2\pi(f-f_0)}{c} xl_0} dx}{\int_{-x_{max}}^{x_{max}} dx} \right|^2. \quad (3.65)$$

Integrating over x and reintroducing $\Delta f = f - f_0$ gives

$$D(\Delta f) = 1 - \left| \frac{-\frac{c}{2\pi j \Delta f l_0} \left(e^{-j\frac{2\pi \Delta f l_0}{c} x_{max}} - e^{j\frac{2\pi \Delta f l_0}{c} x_{max}} \right)}{2x_{max}} \right|^2. \quad (3.66)$$

The difference between the two exponential functions in the numerator can be replaced by the sine function. The sine function has Taylor series expansion $\sin(x) = x - x^3/3! + \dots$ for small values of x , so the decorrelation loss given in (3.66) may be approximated by

$$\begin{aligned}
D(\Delta f) &\approx 1 - \left| \frac{-\frac{c}{2\pi j \Delta f l_0} 2j \left(\frac{2\pi \Delta f l_0}{c} x_{max} - \frac{1}{3!} \left(\frac{2\pi \Delta f l_0}{c} \right)^3 x_{max}^3 \right)}{2x_{max}} \right|^2 \\
&= 1 - \left(1 - \frac{1}{6} \left(\frac{2\pi \Delta f l_0}{c} \right)^2 x_{max}^2 \right)^2 \\
&= 1 - \left(1 - \frac{1}{3} \frac{4\pi^2 \Delta f^2 l_0^2}{c^2} x_{max}^2 + \frac{1}{36} \frac{4\pi^4 \Delta f^4 l_0^4}{c^4} x_{max}^4 \right) \\
&\approx \frac{4\pi^2 \Delta f^2 l_0^2}{3c^2} x_{max}^2.
\end{aligned} \tag{3.67}$$

This result shows that the decorrelation loss is related to the width of the frequency band, the size of the array and the DOA of the signal of interest. If the source of interest is in the zenith, the decorrelation due to channel width even vanishes for a horizontal planar array since the incoming plane wave arrives at all elements simultaneously. The fact that decorrelation occurs for all other positions on the sky can physically be attributed to a mispointing of the array. This becomes apparent if we interpret the phase mismatch at the p th receiving element due to the frequency mismatch as mispointing, which can be formulated mathematically as

$$\frac{2\pi \Delta f \xi_p^T \mathbf{l}_0}{c} = \frac{2\pi f_0 \xi_p^T \Delta \mathbf{l}}{c}, \tag{3.68}$$

which implies that

$$\frac{\Delta f}{f_0} \xi_p^T \mathbf{l}_0 = \xi_p^T \Delta \mathbf{l}. \tag{3.69}$$

This shows that if $\xi_p^T \mathbf{l}_0 = 0$, i.e. $\xi_p \perp \mathbf{l}_0$, no shift in \mathbf{l}_0 occurs. For a horizontal planar array this corresponds to the situation where \mathbf{l}_0 is pointing to the zenith. Since there is a two dimensional subspace in which no mismatch occurs, we can reduce the problem to a one dimensional problem by defining a plane including \mathbf{l}_0 . For a planar array in the (x, y) -plane, it is convenient to define this plane such that it includes both \mathbf{l}_0 and the z -axis, since this allows the use of simple projections. Denoting the projection of ξ_p by ξ_p and the projection of \mathbf{l}_0 by \mathbf{l}_0 , we can restate (3.69) as

$$\Delta \mathbf{l}_0 = \frac{\Delta f}{f_0} \mathbf{l}_0 \tag{3.70}$$

where ξ_p has dropped out, thus showing that the mispointing is independent of projected array size. This mispointing is usually called beam squint. Equation (3.70) also shows that for a horizontal planar array beam squint depends linearly on the projected zenith distance $l_0 = \sin(\theta)$ and on the fractional bandwidth, a conclusion which was also found in an earlier analysis [139] and in simulations [19].

Regardless of the way the decorrelation loss is physically explained, (3.64) and (3.67) are key relations for the design of a phased array radio telescope. Since the size of the array is determined by the desired spatial resolution and l_0 is fixed by specifying the desired observable part of the sky, the only way to keep the decorrelation loss within acceptable limits is to make an appropriate choice for the spectral resolution of the instrument. Even the signals from broadband sources need to be treated at this spectral resolution to avoid decorrelation losses due to premature integration.

3.5 Spatial filtering of source signals

3.5.1 Scalar beam forming

In section 3.4 it was already noted that the phase difference between the antennas due to geometrical delays for an incoming plane wave from direction \mathbf{l}_0 at frequency f_0 can be compensated by weighting the antenna signals with the conjugate of $\mathbf{w} = \mathbf{a}(\mathbf{l}_0, f_0)$. The operation of forming a weighted sum of the antenna signals is called beam forming. The goal of beam forming is to add the signal from a specific direction \mathbf{l}_0 at a specific frequency f_0 coherently while suppressing all other signals. The decorrelation discussed in the previous section describes how well these weights suppress signals coming from the same direction at slightly different frequencies. Below we will assess the suppression of signals at the same frequency coming from different directions.

Consider a single source with unit power $\sigma_s = 1$ and array response vector $\mathbf{a}_s(\mathbf{l}_s)$ received by a noise free array of antennas with isotropic element patterns ($\mathbf{G}_0 = \mathbf{I}$), no coupling between the antennas ($\mathbf{Q} = \mathbf{I}$) and unit receiver gain ($\mathbf{G} = \mathbf{I}$). These simplifications allow us to study beam forming without complicating factors. In this case, the voltage output of the beam former when pointed to \mathbf{l}_0 will be

$$y(t) = \mathbf{w}^H(\mathbf{l}_0) \mathbf{a}_s(\mathbf{l}_s) s_s(t). \quad (3.71)$$

Due to the stochastic nature of the source signal, the voltage output of the beam former is stochastic as well, so we have to consider the power output of the beam

former $\mathcal{E} \left\{ |y(t)|^2 \right\}$, which is given by

$$\begin{aligned} y_{BF} &= \mathcal{E} \left\{ \mathbf{w}^H(\mathbf{l}_0) \mathbf{a}_s(\mathbf{l}_s) s_s(t) \bar{s}_s(t) \mathbf{a}_s^H(\mathbf{l}_s) \mathbf{w}(\mathbf{l}_0) \right\} \\ &= \mathbf{w}^H(\mathbf{l}_0) \mathbf{R}_s(\mathbf{l}_s) \mathbf{w}(\mathbf{l}_0) \end{aligned} \quad (3.72)$$

where $\mathbf{R}_s(\mathbf{l}_s) = \mathbf{a}_s(\mathbf{l}_s) \sigma_s \mathbf{a}_s^H(\mathbf{l}_s) = \mathbf{a}_s(\mathbf{l}_s) \mathbf{a}^H(\mathbf{l}_s)$. This shows that the output power of the beam former obtained by adding the antenna signals in the voltage domain followed by averaging in the power domain gives the same result as coherent addition of the visibilities, which is generally referred to as imaging. The equivalence of an image made by consecutive pointings using a beam former and an image produced by applying the same antenna based weights to the visibilities was discussed in [140]. The key difference between imaging and beam forming is that a beam former can only apply antenna based weights while an imager can assign weights to individual baselines, i.e. individual pairs of antennas, which provides far more flexibility.

The advantage of the beam former described by (3.72) is that it effectively describes a direct Fourier transform, which is especially attractive for dense regular arrays where the fast Fourier transform can be exploited. Its disadvantage is that it therefore suffers from the same sampling effects as the Fourier transform which causes imperfect spatial separation between sources, as will be demonstrated in Sec. 3.5.2. Therefore more sophisticated schemes have been proposed imposing additional constraints to reduce the power received from other directions than the direction of interest. The disadvantage of these more advanced methods is that information on the array covariance matrix is required. Examples of such schemes are the minimum variance distortionless response (MVDR) beam former [23, 122] and the robust Capon beam former (RCB) [69, 126]. The use of these more advanced beam formers has been proposed [64] and demonstrated on actual data [117, 118] in the context of beam forming and imaging in the presence of strong interfering sources.

3.5.2 Array response

The direction dependent gain of the array or array beam pattern for a single polarization can be studied using (3.71) to obtain an array voltage beam or (3.72) to obtain an array power beam by pointing the array in a specific direction while scanning the sky with a probe. Consider a two-dimensional uniform rectangular array (URA), i.e. a planar array with its equidistantly spaced elements placed on a rectangular grid, consisting of 16×16 elements with an isotropic element beam pattern. Figure 3.2 shows its power beam pattern when the array is pointed towards $(l, m) = (0, -0.6)$ for the frequencies at which the array is 0.5λ spaced and 0.8λ spaced

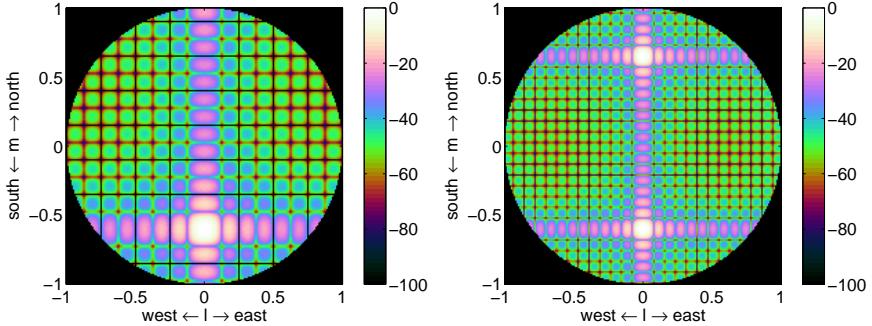


Figure 3.2: Array power beam patterns for a 16×16 URA pointed towards $(l, m) = (0, -0.6)$ if the array is 0.5λ spaced (left) and 0.8λ spaced (right).

In section 3.5.1 it was noted that the basic beam former describes a direct Fourier transform relating the far field array voltage response to the voltage distribution of the electromagnetic field sampled by the receiving elements. A URA samples a rectangular area on a regular grid, so based on Fourier theory we may expect the array voltage beam of such an array to have a shape described by the sinc-function when the array is homogeneously excited. Hence the array power beam will follow a sinc^2 -function. This is nicely demonstrated by the left panel of Fig. 3.2 which shows the array power beam for a 0.5λ spaced or Nyquist sampled array. The right panel shows the corresponding beam for a 0.8λ spaced or undersampled array. This causes aliasing which leads to a grating lobe with the same sensitivity as the main lobe appearing in the northern part of the sky. The presence of grating lobes implies that the array is not able to separate signals received by the grating lobes and the signals received in the main lobe because all these narrowband signals induce exactly the same phase differences in the array elements.

Antenna arrays with all their receiving elements arranged in a single plane will experience a distortion in their array beam pattern due to projection of their aperture plane on the plane perpendicular to the direction of arrival. This projection tends to zero at the horizon leaving the array with no spatial selectivity at zero elevation. As a result, the size of the main beam will grow towards lower elevations. This distortion is not seen in Fig. 3.2 due to the use of (l, m) -coordinates. These coordinates project the celestial sphere on the plane of the local horizon, which exactly compensates the projection effect suffered by the array. This is nicely demonstrated in Fig. 3.3 which shows the same beam in an (l, m) -projection and with linearly increasing zenith angle on the radii.

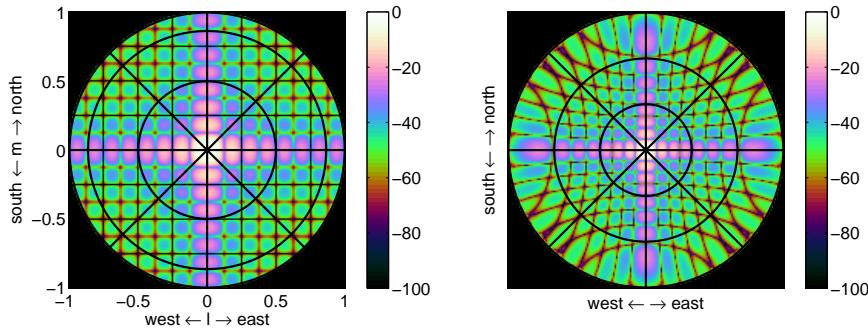


Figure 3.3: Array power beam pattern for a 16×16 half wavelength spaced URA pointed towards the zenith using the (l, m) -projection and polar coordinates with linearly increasing zenith angle on the radii. The circles indicate zenith angles of 30, 60 and 90 degrees, the radii are equally spaced at azimuth intervals of 45 degrees.

Another advantage of (l, m) -coordinates is that a change in pointing of the array causes only a shift of the array beam pattern, i.e. the pattern itself does not change. This is readily demonstrated by the left panels of Figs. 3.2 and 3.3. This can be explained by the fact that the array beam pattern is determined by the array configuration and the amplitudes of the weights applied to its elements. As mentioned earlier, the array beam pattern is the Fourier transform of the spatial sampling within the array aperture. The array thus defines a spatial window function. An incoming plane wave corresponds to a single spatial frequency, which can be described with a delta function, that gets convolved with the Fourier transform of this window function. If this plane wave comes from a different direction, its delta peak in the spatial spectrum shifts and the pattern it is convolved with, will shift along.

This convolution also implies that the ability of the array to disentangle signals coming from two different directions only depends on the separation of their DOAs in (l, m) -coordinates. This is easily seen if the inner product of $\mathbf{w}(l)$ and $\mathbf{a}(l_0)$ describing the array beam pattern is written down explicitly:

$$\mathbf{w}^H(l) \mathbf{a}(l_0) = \mathbf{a}^H(l) \mathbf{a}(l_0) = \frac{1}{P} \left(e^{-j \frac{2\pi f}{c} \Xi l} \right)^H \left(e^{-j \frac{2\pi f}{c} \Xi l_0} \right) = \frac{\mathbf{1}^H}{P} e^{j \frac{2\pi f}{c} \Xi(l-l_0)}. \quad (3.73)$$

Since the inner product is also a measure of the angle between two vectors, in this case in the P -dimensional vector space of array response vectors, perfect separation between signals implies that the corresponding array response vectors either are orthogonal or are made orthogonal by application of appropriate weights to the array

signal vector.

The convolution of the array beam pattern with the source signal has an important consequence for polarimetric measurements with arrays of identical antennas with co-located x -elements and y -elements: the polarization measured in the side lobe response is determined by the instrumental polarization in the main lobe and the intrinsic polarization of the source. Stated differently, the polarization measured in the side lobes is not determined by the instrumental polarization at the position of the side lobe but related to the instrumental polarization in the main lobe [143].

3.6 From snapshots to synthesis observations

The data model described by (3.22) is only valid over a limited time interval and frequency range. The time interval is limited by the stability of the instrument, which determines the update rate for the instrumental parameters, and by the angular velocity of the Earth, which enforces an update rate for the beam former weights depending on the size of the array. The frequency range is limited by the narrowband assumption as discussed in Sec. 3.4. These constraints on the integration time and bandwidth limit the number of samples that can be combined to obtain the short term covariance matrix estimate. This limits the SNR of the, generally weak, astronomical sources. We therefore have to combine several snapshot observations to improve the result, typically one or more images or source parameter estimates. This process is called time or frequency synthesis, depending along which axis the data are combined. This section describes how several snapshot observations can be combined into a synthesis observation.

Equation (3.22) gives a general data model for a single snapshot. To emphasize this, we introduce an additional subscript m indicating the snapshot index and rewrite (3.22) to

$$\mathbf{R}_m = \mathbf{G}_m \mathbf{Q}_m (\mathbf{A}_m \odot \mathbf{G}_{0m}) \boldsymbol{\Sigma} (\mathbf{A}_m \odot \mathbf{G}_{0m})^H \mathbf{Q}_m^H \mathbf{G}_m^H + \boldsymbol{\Sigma}_{nm}. \quad (3.74)$$

These snapshots may either be observations at a different time or at a different frequency. If the snapshot observations are based on independent time-frequency samples, the covariance matrix for the synthesis observation is given by

$$\mathbf{R}_{syn} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{R}_M \end{bmatrix} = \text{bdiag}(\mathbf{R}_m) \quad (3.75)$$

where we have introduced $\text{bdiag}(\cdot)$ as a short hand notation for a block diagonal matrix.

Theoretically, \mathbf{R}_{syn} has the same role in synthesis observations as \mathbf{R}_m has for the m th snapshot. For example, if $\mathbf{w}_{syn} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_M^T]^T$, then $\mathbf{w}_{syn}^H \mathbf{R}_{syn} \mathbf{w}_{syn}$ is the output of the scalar beam former applied to the synthesis observation just like $\mathbf{w}_m^H \mathbf{R}_m \mathbf{w}_m$ describes its output for the m th snapshot. In practice, the independence of the individual snapshots can often be exploited to reduce processing cost and memory usage. For example $\mathbf{w}_{syn}^H \mathbf{R}_{syn} \mathbf{w}_{syn}$ requires $O(M^2 P^2)$ multiplications and storage of an $MP \times MP$ matrix, while $\sum_{m=1}^M \mathbf{w}_m^H \mathbf{R}_m \mathbf{w}_m$ requires $O(MP^2)$ multiplications and storage of $M P \times P$ matrices. It is therefore convenient to vectorize (3.75) such that only the non-zero elements are retained. This can be done by defining

$$\mathbf{r}_{syn} = \begin{bmatrix} \text{vec}(\mathbf{R}_1) \\ \text{vec}(\mathbf{R}_2) \\ \vdots \\ \text{vec}(\mathbf{R}_M) \end{bmatrix}. \quad (3.76)$$

Using $\text{vec}(\mathbf{w}_m^H \mathbf{R}_m \mathbf{w}_m) = (\overline{\mathbf{w}}_m \otimes \mathbf{w}_m)^H \text{vec}(\mathbf{R}_m)$ and defining

$$\mathbf{w}_{syn} = \left[(\overline{\mathbf{w}}_1 \otimes \mathbf{w}_1)^T, (\overline{\mathbf{w}}_2 \otimes \mathbf{w}_2)^T, \dots, (\overline{\mathbf{w}}_M \otimes \mathbf{w}_M)^T \right]^T, \quad (3.77)$$

the output of the scalar beam former is simply $\mathbf{w}_{syn}^H \mathbf{r}_{syn}$ which only requires the necessary MP^2 multiplications.

In this thesis, calibration is done per snapshot to track the variations in instrumental and, as demonstrated in Ch. 2, environmental parameters. The calibrated snapshots are then combined to form an image using the procedure described in Sec. 6.3. For imaging purposes, we will exploit the fact that $\Sigma = \text{diag}(\boldsymbol{\sigma})$, so we can write (3.74) as

$$\text{vec}(\mathbf{R}_m) = (\overline{\mathbf{A}}_{sm} \circ \mathbf{A}_{sm}) \boldsymbol{\sigma} + \text{vec}(\Sigma_{nm}), \quad (3.78)$$

where we have introduced $\mathbf{A}_{sm} = \mathbf{G}_m \mathbf{Q}_m (\mathbf{A}_m \odot \mathbf{G}_{0m}) \mathbf{I}_{sm}$. The selection matrix \mathbf{I}_{sm} , which determines whether a source is detected in the m th snapshot, describes sources moving in and out of the FOV during the synthesis observation. Focusing only on the overall structure, we write this as

$$\text{vec}(\mathbf{R}_m) = \mathbf{M}_m \boldsymbol{\sigma} + \mathbf{n}_m. \quad (3.79)$$

For the polarized case, $\text{vec}(\mathbf{R}_m)$ can be expressed in the same form. We will demonstrate this for an array of identical antennas where standard matrix products

can be exploited. First note that, for an array of identical elements, (3.27) becomes

$$\begin{aligned} \mathbf{R}_m &= \\ &= \mathbf{G}_m \mathbf{Q}_m \left(\sum_{q=1}^Q (\mathbf{A}_{qm} \odot (\mathbf{1} \otimes \mathbf{I})) \mathbf{J}_{qm} \mathbf{E}_q \mathbf{J}_{qm}^H (\mathbf{A}_{qm} \odot (\mathbf{1} \otimes \mathbf{I}))^H \right) \mathbf{Q}_m^H \mathbf{G}_m^H + \\ &\quad + \boldsymbol{\Sigma}_{mn}. \end{aligned} \quad (3.80)$$

This can also be written as

$$\text{vec}(\mathbf{R}_m) = \sum_{q=1}^Q \left(\overline{\mathbf{A}}'_{qm} \otimes \mathbf{A}'_{qm} \right) \mathbf{B}_{qm} \mathbf{e}_q + \text{vec}(\boldsymbol{\Sigma}_{mn}). \quad (3.81)$$

where $\mathbf{A}'_{qm} = \mathbf{G}_m \mathbf{Q}_m (\mathbf{A}_{qm} \odot (\mathbf{1} \otimes \mathbf{I}))$ and used \mathbf{B}_{qm} as defined in (3.31). Taking

$$\mathbf{M}_m = \left[\overline{\mathbf{A}}'_{1m} \otimes \mathbf{A}'_{1m}, \dots, \overline{\mathbf{A}}'_{Qm} \otimes \mathbf{A}'_{Qm} \right] \mathbf{I}_{sm} \text{bdiaig}(\mathbf{B}_{qm}) \quad (3.82)$$

and defining $\boldsymbol{\sigma} = [\mathbf{e}_1^T, \mathbf{e}_2^T, \dots, \mathbf{e}_Q^T]^T$, the full polarized data model given by (3.27) can be expressed as (3.79).

Using the definition for \mathbf{r}_{syn} given by (3.76) and assuming that the source powers, $\boldsymbol{\sigma}$, are constant over the synthesis observations, we find that

$$\mathbf{r}_{syn} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_M \end{bmatrix} \boldsymbol{\sigma} + \begin{bmatrix} \text{vec}(\boldsymbol{\Sigma}_{n1}) \\ \text{vec}(\boldsymbol{\Sigma}_{n2}) \\ \vdots \\ \text{vec}(\boldsymbol{\Sigma}_{nM}) \end{bmatrix} = \mathbf{M}\boldsymbol{\sigma} + \mathbf{n}. \quad (3.83)$$

This result is used in Sec. 6.3 to derive a generic weighted least squares solution to the imaging problem, while (3.78) and (3.81) are used to derive specific expressions for single and full polarization data, that can be applied to actual LOFAR data as demonstrated in Ch. 2. This shows that, although much more data is involved in a synthesis observation than in a snapshot observation, the mathematical formulation is identical. This implies that many results derived for snapshot observations can be applied to synthesis observations as well in a straightforward manner. The analysis of snapshot observations presented in this thesis is therefore a major step towards the analysis of synthesis observations.

Chapter 4

Cramer-Rao Bound

The Cramer-Rao bound (CRB) is a lower bound on the variance of an unbiased estimator. An estimator is efficient if it is unbiased and the covariance of the estimation error achieves the CRB. Such an estimator is a maximum-likelihood (ML) estimator. Although it is generally not feasible to derive such an estimator for practical problems, practical estimators may be asymptotically efficient, which means that if the number of samples N tends to infinity, i.e. is sufficiently large, the estimator tends to behave like an ML estimator. An ML estimator has a number of attractive properties: it is a consistent estimator, its estimates are asymptotically normally distributed and it is an asymptotically efficient estimator. For a more detailed treatment on estimation theory, the reader is referred to the excellent text books written by Kay [58] and Moon and Stirling [75].

In this thesis, the CRB is used in two ways. In the first place, it is used to evaluate the performance of the proposed parameter estimation methods. This is done by comparing the variance of the parameter estimates in Monte Carlo simulations with the CRB. The results from the Monte Carlo simulations also indicate whether the estimates are unbiased or not. These simulations therefore provide an indication of the statistical efficiency of the proposed estimators. Secondly, the CRB is used to compute the minimum covariance of the calibration parameters as a starting point for the analysis of calibration error propagation in the imaging process in Ch. 6.

This chapter therefore provides the ground work for later chapters. We start with the basic definition of the CRB and specialize to the case in which the data samples have a Gaussian distribution. The resulting relation for the CRB clearly shows that the covariance matrix describing the CRB can be partitioned in blocks associated with specific groups of parameters. This will be exploited in Sec. 4.2 in which we provide the results used in later chapters. Section 4.3, which is based on [151], discusses the

impact of the boundary condition imposed to fix the gain phase solution. It is shown that the commonly used condition that the gain phase of one of the elements is set to zero results in a variance on the phase estimates which is a factor $2P/(P - 1)$ larger than under the constraint that the average phase is zero. Another result, originally described in [119], is that the CRB on the direction independent gain estimates is the same regardless whether an entry of the array covariance matrix is affected by an additive nuisance parameter that has to be included in the parameter estimation process or it is simply ignored. The details are discussed in Sec. 4.4.

4.1 Definition of the CRB

Suppose we want to estimate a parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_L]^T$ from a series of sample vectors

$$\mathbf{x}(t) = \mathbf{s}(t; \boldsymbol{\theta}) + \mathbf{n}(t) \quad (4.1)$$

which are composed of a parameterized signal vector $\mathbf{s}(t; \boldsymbol{\theta})$ and a noise vector $\mathbf{n}(t)$. Assume that the cumulative probability density function (pdf) $p(\mathbf{X}; \boldsymbol{\theta})$ of the sample vectors stacked in $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ is known. The covariance matrix $\mathbf{C}_{\boldsymbol{\theta}}$ of the estimated parameter vector then satisfies [58, 75]

$$\mathbf{C}_{\boldsymbol{\theta}} \geq \mathbf{C}_{CRB} = \mathbf{F}^{-1} \quad (4.2)$$

where the elements of the Fisher information matrix (FIM) \mathbf{F} are given by

$$F_{ij} = -\mathcal{E} \left\{ \frac{\partial^2 \ln p(\mathbf{X}; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\}. \quad (4.3)$$

This is the Cramer-Rao theorem and \mathbf{C}_{CRB} is the Cramer-Rao bound.

Although the CRB provides a powerful tool for statistical analysis of estimation problems and proposed estimators, its definition leads to rather cumbersome calculations when applied to practical problems. Fortunately, a lot of ground work has already been done if the data samples can be described by complex Gaussian noise with covariance $\mathbf{R}(\boldsymbol{\theta}) = \mathcal{E}\{\mathbf{x}(t)\mathbf{x}^H(t)\}$ and mean $\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathcal{E}\{\mathbf{x}(t)\}$, i.e. $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{R}(\boldsymbol{\theta}))$. In this case, the pdf is given by [58]

$$p(\mathbf{X}; \boldsymbol{\theta}) = \prod_{n=1}^N \frac{1}{\pi^P |\mathbf{R}(\boldsymbol{\theta})|} \exp \left(-\frac{1}{2} (\mathbf{x}_n - \boldsymbol{\mu}(\boldsymbol{\theta}))^H \mathbf{R}^{-1}(\boldsymbol{\theta}) (\mathbf{x}_n - \boldsymbol{\mu}(\boldsymbol{\theta})) \right). \quad (4.4)$$

It is straightforward, although tedious, to derive an expression for the FIM, and therefore the CRB, by substitution of this pdf in (4.3). Since the complete derivation can

be found in [58], we only state the result,

$$\begin{aligned} F_{ij} &= N \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i} \right]^H \mathbf{R}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j} \right] \\ &\quad + N \text{tr} \left(\mathbf{R}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{R}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_j} \right) \end{aligned} \quad (4.5)$$

where $\text{tr}(\cdot)$ denotes the trace of its argument,

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i} = \begin{bmatrix} \frac{\partial \mu_1(\boldsymbol{\theta})}{\partial \theta_i} \\ \frac{\partial \mu_2(\boldsymbol{\theta})}{\partial \theta_i} \\ \vdots \\ \frac{\partial \mu_P(\boldsymbol{\theta})}{\partial \theta_i} \end{bmatrix} \quad (4.6)$$

and

$$\frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_i} = \begin{bmatrix} \frac{\partial R_{11}(\boldsymbol{\theta})}{\partial \theta_i} & \frac{\partial R_{12}(\boldsymbol{\theta})}{\partial \theta_i} & \dots & \frac{\partial R_{1P}(\boldsymbol{\theta})}{\partial \theta_i} \\ \frac{\partial R_{21}(\boldsymbol{\theta})}{\partial \theta_i} & \frac{\partial R_{22}(\boldsymbol{\theta})}{\partial \theta_i} & \dots & \frac{\partial R_{2P}(\boldsymbol{\theta})}{\partial \theta_i} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial R_{P1}(\boldsymbol{\theta})}{\partial \theta_i} & \frac{\partial R_{P2}(\boldsymbol{\theta})}{\partial \theta_i} & \dots & \frac{\partial R_{PP}(\boldsymbol{\theta})}{\partial \theta_i} \end{bmatrix}, \quad (4.7)$$

where the derivatives are evaluated at the true value of $\boldsymbol{\theta}$.

If the signals can be described as i.i.d. complex Gaussian noise, $\boldsymbol{\mu} = \mathbf{0}$ such that $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}))$. In that case, (4.5) simplifies to Bangs' Formula [5, 109]

$$\begin{aligned} F_{ij} &= N \text{tr} \left(\mathbf{R}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{R}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_j} \right) \\ &= N \left[\frac{\partial \text{vec}(\mathbf{R}(\boldsymbol{\theta}))}{\partial \theta_i} \right]^H \left(\overline{\mathbf{R}}^{-1}(\boldsymbol{\theta}) \otimes \mathbf{R}^{-1}(\boldsymbol{\theta}) \right) \left[\frac{\partial \text{vec}(\mathbf{R}(\boldsymbol{\theta}))}{\partial \theta_j} \right]. \end{aligned} \quad (4.8)$$

The FIM for i.i.d. complex Gaussian signals can thus be written in compact form as

$$\mathbf{F} = N \mathbf{J}^H \left(\overline{\mathbf{R}}^{-1}(\boldsymbol{\theta}) \otimes \mathbf{R}^{-1}(\boldsymbol{\theta}) \right) \mathbf{J} \quad (4.9)$$

where we have introduced the Jacobian matrix

$$\mathbf{J} = \frac{\partial \text{vec}(\mathbf{R}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^T} = \left[\frac{\partial \text{vec}(\mathbf{R}(\boldsymbol{\theta}))}{\partial \theta_1}, \frac{\partial \text{vec}(\mathbf{R}(\boldsymbol{\theta}))}{\partial \theta_2}, \dots, \frac{\partial \text{vec}(\mathbf{R}(\boldsymbol{\theta}))}{\partial \theta_L} \right]. \quad (4.10)$$

Once the covariance of the parameter vector $\mathbf{C}_{\boldsymbol{\theta}}$ has been determined, the covariance of a derived parameter vector $\tilde{\boldsymbol{\theta}} = \Theta(\boldsymbol{\theta})$ can be obtained by standard error propagation formulas, i.e. [10, 58]

$$\mathbf{C}_{\tilde{\boldsymbol{\theta}}} = \left[\frac{\partial \Theta(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right]^H \mathbf{C}_{\boldsymbol{\theta}} \left[\frac{\partial \Theta(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right] \quad (4.11)$$

Since the CRB is a covariance matrix, this error propagation formula can also be used to derive the CRB for $\tilde{\boldsymbol{\theta}}$ from the CRB for $\boldsymbol{\theta}$. This property can be used to demonstrate that the CRB calculations actually lead to very intuitive results. Suppose we have a function $\Theta(\text{vec}(\mathbf{R}))$ to derive $\boldsymbol{\theta}$ from $\text{vec}(\mathbf{R})$. The CRB for estimating the elements of \mathbf{R} is given by

$$\begin{aligned} \mathbf{C}_{\text{vec}(\mathbf{R})} &= \left(\left[\frac{\partial \text{vec}(\mathbf{R})}{\partial \text{vec}(\mathbf{R})^T} \right]^H N (\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) \left[\frac{\partial \text{vec}(\mathbf{R})}{\partial \text{vec}(\mathbf{R})^T} \right] \right)^{-1} \\ &= \left(\mathbf{I}^H \left(\frac{1}{N} (\bar{\mathbf{R}} \otimes \mathbf{R}) \right)^{-1} \mathbf{I} \right)^{-1} \end{aligned} \quad (4.12)$$

Since $\text{cov}(\mathbf{R}) = (\bar{\mathbf{R}} \otimes \mathbf{R}) / N$ for a complex Gaussian signal $\mathbf{x}(t)$ as derived in [14, 82], we immediately see that $\mathbf{C}_{\text{vec}(\mathbf{R})} = \text{cov}(\mathbf{R})$, i.e. that the CRB for estimating the elements of $\text{vec}(\mathbf{R})$ is simply the covariance of these elements, an expected result. The CRB for $\boldsymbol{\theta} = \Theta(\text{vec}(\mathbf{R}))$ is therefore simply propagated measurement noise if $\boldsymbol{\theta}$ is estimated based from the measured covariance matrix and not from the measured samples. In practice, $\Theta(\mathbf{R})$ is often only known implicitly, which makes calculation of the CRB for $\boldsymbol{\theta}$ from the inversion of the FIM given by (4.9) the most obvious approach.

4.2 Analytic expressions for the CRB

4.2.1 Partitioning the FIM

Another consequence of (4.9) is that if $\boldsymbol{\theta}$ can be partitioned in subgroups $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots$, then \mathbf{J} can be partitioned in subgroups $\partial \text{vec}(\mathbf{R}(\boldsymbol{\theta})) / \partial \boldsymbol{\theta}_1^T, \partial \text{vec}(\mathbf{R}(\boldsymbol{\theta})) / \partial \boldsymbol{\theta}_2^T, \dots$ and therefore the FIM can be partitioned in blocks,

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{\boldsymbol{\theta}_1 \boldsymbol{\theta}_1} & \mathbf{F}_{\boldsymbol{\theta}_1 \boldsymbol{\theta}_2} & \cdots \\ \mathbf{F}_{\boldsymbol{\theta}_2 \boldsymbol{\theta}_1} & \mathbf{F}_{\boldsymbol{\theta}_2 \boldsymbol{\theta}_2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \quad (4.13)$$

As we will see, it is often feasible to derive an analytic solution for the submatrices $\mathbf{F}_{\theta_i \theta_j}$ in which the Kronecker product in (4.9) is reduced to a Khatri-Rao product or even a Hadamard product. This reduces the sizes of the matrices involved considerably thus reducing the memory requirements for their storage. This allows us to apply the results not only to toy examples with $P \sim 10$ often used in the literature, but also to a full European LOFAR station with $P = 192$. The analytic solutions also provide insight in the sensitivity of the calibration measurements to the model parameters.

The calibration and imaging problems [146, 149, 151, 161, 164, 165] presented in this thesis can all be described using the data model

$$\mathbf{R} = \mathbf{G}\mathbf{A}(\mathcal{L})\mathbf{G}_0\boldsymbol{\Sigma}\mathbf{G}_0^H\mathbf{A}^H(\mathcal{L})\mathbf{G}^H + \boldsymbol{\Sigma}_n \quad (4.14)$$

where $\mathcal{L} = [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_Q]$. This is a specialization of (3.22) with $\mathbf{Q} = \mathbf{I}$ and diagonal \mathbf{G}_0 . The first constraint is merely introduced for notational convenience since mutual coupling can be absorbed in \mathbf{A} if the mutual coupling matrix \mathbf{Q} is known. The latter condition implies that we assume that all receiving elements in the array have the same sensitivity pattern. If the diagonal elements of \mathbf{G}_0 are unknown, the apparent source powers of the calibration sources are not known either. This implies that we cannot exploit our knowledge, if available, of the intrinsic power of these calibrators to model the array covariance matrix for, e.g., complex receiver path gain estimation. We can therefore introduce $\boldsymbol{\Sigma}_s = \mathbf{G}_0\boldsymbol{\Sigma}\mathbf{G}_0^H$ and simplify (4.14) to

$$\mathbf{R} = \mathbf{G}\mathbf{A}(\mathcal{L})\boldsymbol{\Sigma}_s\mathbf{A}^H(\mathcal{L})\mathbf{G}^H + \boldsymbol{\Sigma}_n. \quad (4.15)$$

The direction independent receiver path gains \mathbf{G} can be written as $\mathbf{G} = \boldsymbol{\Gamma}\boldsymbol{\Phi}$ where $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$ describes the gain amplitudes and $\boldsymbol{\Phi} = \text{diag}(\exp(j\phi))$ describes the gain phases. We can now identify the following groups of parameters from which different subsets will be estimated using the model described by (4.15):

- the amplitudes of the direction independent gains, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_P]$;
- the phases of the direction independent gains, $\boldsymbol{\phi} = [\phi_1, \phi_2, \dots, \phi_P]$. Since phases can only be measured with respect to a phase reference, an appropriate constraint is required to find an unambiguous solution;
- the source positions, $\mathcal{L} = [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_Q]$. If the phases of the direction independent gains are estimated simultaneously, a common rotation or shift of the source positions can be compensated by changing the direction independent gain phase solutions $\boldsymbol{\phi}$. An appropriate constraint is therefore required to resolve this ambiguity;
- the apparent source powers $\boldsymbol{\Sigma}_s = \text{diag}(\boldsymbol{\sigma})$ where $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \dots, \sigma_Q]$. If the amplitudes of the direction dependent gains are estimated simultaneously, a common scalar factor applied to all elements of $\boldsymbol{\sigma}_s$ can be compensated by

scaling the elements of γ by the square root of its inverse. An appropriate boundary condition is required to resolve this;

- the receiver noise powers $\Sigma_n = \text{diag}(\sigma_n)$ where $\sigma_n = [\sigma_{n1}, \sigma_{n2}, \dots, \sigma_{nP}]$. Although we have extended the calibration routines to deal with non-diagonal noise covariance matrices [164], this case has not yet been assessed using a CRB analysis.

Let $\tilde{\mathbf{l}} = [l_1, l_2, \dots, l_Q]$ be a vector containing the l -coordinates of the sources and let $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{n}}$ be corresponding vectors for the m - and n -coordinates. If all parameters would be estimated simultaneously, the Jacobian-matrix could be partitioned as [165]

$$\mathbf{J} = \left[\mathbf{J}_\gamma, \mathbf{J}_\phi, \mathbf{J}_\sigma, \mathbf{J}_{\tilde{\mathbf{l}}}, \mathbf{J}_{\tilde{\mathbf{m}}}, \mathbf{J}_{\tilde{\mathbf{n}}}, \mathbf{J}_{\sigma_n} \right]. \quad (4.16)$$

The corresponding partitioning of the FIM is given by

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{\gamma\gamma} & \mathbf{F}_{\gamma\phi} & \mathbf{F}_{\gamma\sigma} & \mathbf{F}_{\gamma\tilde{\mathbf{l}}} & \mathbf{F}_{\gamma\tilde{\mathbf{m}}} & \mathbf{F}_{\gamma\tilde{\mathbf{n}}} & \mathbf{F}_{\gamma\sigma_n} \\ \mathbf{F}_{\gamma\phi}^H & \mathbf{F}_{\phi\phi} & \mathbf{F}_{\phi\sigma} & \mathbf{F}_{\phi\tilde{\mathbf{l}}} & \mathbf{F}_{\phi\tilde{\mathbf{m}}} & \mathbf{F}_{\phi\tilde{\mathbf{n}}} & \mathbf{F}_{\phi\sigma_n} \\ \mathbf{F}_{\gamma\sigma}^H & \mathbf{F}_{\phi\sigma}^H & \mathbf{F}_{\sigma\sigma} & \mathbf{F}_{\sigma\tilde{\mathbf{l}}} & \mathbf{F}_{\sigma\tilde{\mathbf{m}}} & \mathbf{F}_{\sigma\tilde{\mathbf{n}}} & \mathbf{F}_{\sigma\sigma_n} \\ \mathbf{F}_{\tilde{\mathbf{l}}\tilde{\mathbf{l}}}^H & \mathbf{F}_{\tilde{\mathbf{l}}\phi}^H & \mathbf{F}_{\tilde{\mathbf{l}}\sigma}^H & \mathbf{F}_{\tilde{\mathbf{l}}\tilde{\mathbf{l}}} & \mathbf{F}_{\tilde{\mathbf{l}}\tilde{\mathbf{m}}} & \mathbf{F}_{\tilde{\mathbf{l}}\tilde{\mathbf{n}}} & \mathbf{F}_{\tilde{\mathbf{l}}\sigma_n} \\ \mathbf{F}_{\tilde{\mathbf{l}}\tilde{\mathbf{m}}}^H & \mathbf{F}_{\tilde{\mathbf{l}}\tilde{\mathbf{m}}}^H & \mathbf{F}_{\tilde{\mathbf{m}}\tilde{\mathbf{m}}}^H & \mathbf{F}_{\tilde{\mathbf{l}}\tilde{\mathbf{m}}} & \mathbf{F}_{\tilde{\mathbf{m}}\tilde{\mathbf{m}}} & \mathbf{F}_{\tilde{\mathbf{m}}\tilde{\mathbf{n}}} & \mathbf{F}_{\tilde{\mathbf{m}}\sigma_n} \\ \mathbf{F}_{\tilde{\mathbf{l}}\tilde{\mathbf{n}}}^H & \mathbf{F}_{\tilde{\mathbf{l}}\tilde{\mathbf{n}}}^H & \mathbf{F}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}}^H & \mathbf{F}_{\tilde{\mathbf{l}}\tilde{\mathbf{n}}} & \mathbf{F}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}} & \mathbf{F}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}} & \mathbf{F}_{\tilde{\mathbf{n}}\sigma_n} \\ \mathbf{F}_{\gamma\sigma_n}^H & \mathbf{F}_{\phi\sigma_n}^H & \mathbf{F}_{\sigma\sigma_n}^H & \mathbf{F}_{\tilde{\mathbf{l}}\sigma_n}^H & \mathbf{F}_{\tilde{\mathbf{m}}\sigma_n}^H & \mathbf{F}_{\tilde{\mathbf{n}}\sigma_n}^H & \mathbf{F}_{\sigma_n\sigma_n} \end{bmatrix}. \quad (4.17)$$

Note that we only have to know the constituents in the upper triangle and on the main diagonal since the other parts of the FIM follow from its Hermiticity. Below we will first derive expressions for the components of the Jacobian and then demonstrate how the results may be used to reduce the Kronecker product in (4.9). The similarity in structure of the submatrices of the FIM will be sufficiently clear by then to limit the presentation to those terms that provide interesting insights and leave the others as an exercise for the reader.

4.2.2 Jacobians

The Jacobian for the direction independent gains is, by definition,

$$\mathbf{J}_\gamma = \left[\frac{\partial \text{vec}(\mathbf{R})}{\partial \gamma_1}, \frac{\partial \text{vec}(\mathbf{R})}{\partial \gamma_2}, \dots, \frac{\partial \text{vec}(\mathbf{R})}{\partial \gamma_P} \right] \quad (4.18)$$

where we have omitted the explicit statement of the parameterization of \mathbf{R} . For notational convenience, we introduce $\mathbf{R}_0 = \mathbf{A}(\mathcal{L}) \Sigma_s \mathbf{A}^H(\mathcal{L})$. Focusing on the deriva-

tive with respect to the scalar γ_i we find that

$$\frac{\partial \text{vec}(\mathbf{R})}{\partial \gamma_i} = \text{vec}\left(\frac{\partial \mathbf{R}}{\partial \gamma_i}\right) = \text{vec}\left(\mathbf{E}_{ii}\Phi\mathbf{R}_0\mathbf{G}^H + \mathbf{G}\mathbf{R}_0\Phi^H\mathbf{E}_{ii}\right), \quad (4.19)$$

where \mathbf{E}_{ij} denotes the elementary matrix in which only the (i, j) th element is non-zero and equal to 1. An elementary matrix can be written as the product of two unit vectors or elementary vectors, i.e. $\mathbf{E}_{ij} = \mathbf{e}_i\mathbf{e}_j^H$. If $i = j$, then the elementary matrix can be treated as a diagonal matrix with the corresponding elementary vector on the main diagonal, i.e. $\mathbf{E}_{ii} = \text{diag}(\mathbf{e}_i)$. Using (C.6), this can be exploited to rewrite (4.19) to

$$\frac{\partial \text{vec}(\mathbf{R})}{\partial \gamma_i} = \left(\left(\Phi\mathbf{R}_0\mathbf{G}^H\right)^T \circ \mathbf{I}\right)\mathbf{e}_i + \left(\mathbf{I}^T \circ \mathbf{G}\mathbf{R}_0\Phi^H\right)\mathbf{e}_i. \quad (4.20)$$

This particular form was chosen because it allows us to obtain a compact expression for \mathbf{J}_γ . Before stacking the derivatives to the gains of the individual receiver paths to form the Jacobian matrix, we note that \mathbf{G} and Φ are diagonal matrices with complex valued elements on their main diagonal and that their Hermitian transpose can thus be replaced by conjugation. We also note that $\mathbf{R}_0^T = \overline{\mathbf{R}_0}$ since \mathbf{R}_0 is a Hermitian matrix. The Jacobian for the direction independent gains is therefore given by

$$\mathbf{J}_\gamma = \overline{\mathbf{G}\mathbf{R}_0\Phi} \circ \mathbf{I} + \mathbf{I} \circ \mathbf{G}\mathbf{R}_0\overline{\Phi}. \quad (4.21)$$

Gain phases can only be determined with respect to some phase reference. This reference can be formulated as a constraint on the data. Here, we will use the simple constraint that $\phi_1 = 0$. The impact of this choice will be discussed in more detail in Sec. 4.3. The advantage of this simple constraint is that the value of the phase of the first receiver path is set, so it does not need to be estimated. We can thus determine the Jacobian for all phases and then apply a selection matrix \mathbf{I}_s to remove the first column in the computation of the CRB. The i th column of the \mathbf{J}_ϕ is given by

$$\frac{\partial \mathbf{R}}{\partial \phi_i} = \text{jvec}\left(\Gamma\mathbf{E}_{ii}\Phi\mathbf{R}_0\mathbf{G}^H - \mathbf{G}\mathbf{R}_0\Phi^H\mathbf{E}_{ii}\Gamma^H\right). \quad (4.22)$$

Using (C.6), this can be written as

$$\frac{\partial \text{vec}(\mathbf{R})}{\partial \phi_i} = \text{j}\left(\left(\overline{\mathbf{G}\mathbf{R}_0\mathbf{G}} \circ \mathbf{I}\right)\mathbf{e}_i - \left(\mathbf{I} \circ \mathbf{G}\mathbf{R}_0\overline{\mathbf{G}}\right)\mathbf{e}_i\right). \quad (4.23)$$

The Jacobian matrix for the gain phases including the boundary condition $\phi_1 = 0$ can therefore be expressed as

$$\mathbf{J}_\phi = \text{j}\left(\overline{\mathbf{G}\mathbf{R}_0\mathbf{G}} \circ \mathbf{I} - \mathbf{I} \circ \mathbf{G}\mathbf{R}_0\overline{\mathbf{G}}\right)\mathbf{I}_s. \quad (4.24)$$

The Jacobian for the apparent source powers $\boldsymbol{\sigma}$ is easily found after recognizing that

$$\text{vec}(\mathbf{R}) = (\overline{\mathbf{GA}} \circ \mathbf{GA}) \boldsymbol{\sigma} + \text{vec}(\boldsymbol{\Sigma}_n), \quad (4.25)$$

where we have omitted the explicit statement of the dependence of \mathbf{A} on \mathcal{L} , which is irrelevant at this point. It then follows that

$$\frac{\partial \text{vec}(\mathbf{R})}{\partial \sigma_i} = (\overline{\mathbf{GA}} \circ \mathbf{GA}) \mathbf{e}_i. \quad (4.26)$$

Therefore

$$\mathbf{J}_{\boldsymbol{\sigma}} = \overline{\mathbf{GA}} \circ \mathbf{GA}. \quad (4.27)$$

In the previous section, it was noted that the direction independent gains and the apparent source powers share a common scaling factor. This can, for example, be resolved by imposing the constraint $\sigma_1 = 1$, which can be reflected in the computation of the Jacobian by applying an appropriate selection matrix.

The product rule allows us to write the derivative of $\text{vec}(\mathbf{R})$ with respect to l_i as

$$\frac{\partial \text{vec}(\mathbf{R})}{\partial l_i} = \text{vec} \left(\mathbf{G} \frac{\partial \mathbf{A}(\mathcal{L})}{\partial l_i} \boldsymbol{\Sigma}_s \mathbf{A}^H(\mathcal{L}) \mathbf{G}^H + \mathbf{GA}(\mathcal{L}) \boldsymbol{\Sigma}_s \left[\frac{\partial \mathbf{A}(\mathcal{L})}{\partial l_i} \right]^H \mathbf{G}^H \right). \quad (4.28)$$

All elements of $\mathbf{A}(\mathcal{L})$ that do not depend on l_i will be zero after differentiation. Therefore

$$\frac{\partial \mathbf{A}(\mathcal{L})}{\partial l_i} = \frac{\partial}{\partial l_i} \mathbf{a}_i(\mathbf{l}_i) \mathbf{e}_i^H, \quad (4.29)$$

where \mathbf{a}_i is the i th column of $\mathbf{A}(\mathcal{L})$. An expression for $\mathbf{a}(\mathbf{l}, t)$ is given by (3.7). Here, we only need to take into account the geometrical delay. Differentiating with respect to the l -coordinate then gives

$$\frac{\partial \mathbf{a}_i(\mathbf{l}_i)}{\partial l_i} = \frac{\partial}{\partial l_i} \frac{1}{\sqrt{P}} e^{-j\frac{2\pi}{\lambda} \boldsymbol{\Xi} \mathbf{l}_i} = -j \frac{1}{\sqrt{P}} \frac{2\pi}{\lambda} \text{diag}([x_1, x_2, \dots, x_P]) e^{-j\frac{2\pi}{\lambda} \boldsymbol{\Xi} \mathbf{l}_i}. \quad (4.30)$$

Introducing $\mathbf{G}_x = \text{diag}([x_1, x_2, \dots, x_P]) \mathbf{G}$, we can write (4.28) as

$$\frac{\partial \text{vec}(\mathbf{R})}{\partial l_i} = -j \frac{1}{\sqrt{P}} \frac{2\pi}{\lambda} \text{vec}(\mathbf{G}_x \mathbf{A} \mathbf{E}_{ii} \boldsymbol{\Sigma}_s \mathbf{A}^H \mathbf{G}^H - \mathbf{GA} \boldsymbol{\Sigma}_s \mathbf{E}_{ii} \mathbf{A}^H \mathbf{G}_x^H). \quad (4.31)$$

Vectorization using (C.6) gives

$$\frac{\partial \text{vec}(\mathbf{R})}{\partial l_i} = -j \frac{1}{\sqrt{P}} \frac{2\pi}{\lambda} (\overline{\mathbf{GA}} \circ \mathbf{G}_x \mathbf{A} - \overline{\mathbf{G}_x \mathbf{A}} \circ \mathbf{GA}) \boldsymbol{\Sigma}_s \mathbf{e}_i, \quad (4.32)$$

which implies that

$$\mathbf{J}_{\tilde{\mathbf{l}}} = j \frac{1}{\sqrt{P}} \frac{2\pi}{\lambda} (\overline{\mathbf{G}_x \mathbf{A}} \circ \mathbf{G} \mathbf{A} - \overline{\mathbf{G} \mathbf{A}} \circ \mathbf{G}_x \mathbf{A}) \Sigma_s. \quad (4.33)$$

In a similar way, we find

$$\mathbf{J}_{\tilde{\mathbf{m}}} = j \frac{1}{\sqrt{P}} \frac{2\pi}{\lambda} (\overline{\mathbf{G}_y \mathbf{A}} \circ \mathbf{G} \mathbf{A} - \overline{\mathbf{G} \mathbf{A}} \circ \mathbf{G}_y \mathbf{A}) \Sigma_s \quad (4.34)$$

and

$$\mathbf{J}_{\tilde{\mathbf{n}}} = j \frac{1}{\sqrt{P}} \frac{2\pi}{\lambda} (\overline{\mathbf{G}_z \mathbf{A}} \circ \mathbf{G} \mathbf{A} - \overline{\mathbf{G} \mathbf{A}} \circ \mathbf{G}_z \mathbf{A}) \Sigma_s \quad (4.35)$$

where $\mathbf{G}_y = \text{diag}([y_1, y_2, \dots, y_P]) \mathbf{G}$ and $\mathbf{G}_z = \text{diag}([z_1, z_2, \dots, z_P])$ respectively.

Equations (4.33), (4.34) and (4.35) show that the entries of the Jacobian related to derivatives with respect to the l -, m - and n -coordinates of the sources are proportional to the x -, y - and z -coordinates of the receiving elements respectively. The physical interpretation of this relation is that a plane wave propagating along the coordinate axis of the coordinate to be estimated provides a better test signal to estimate the source location than a signal propagating perpendicular to this axis.

As noted in the previous section, a common rotation or shift of the source positions can be compensated by adjusting the gain phases. If source locations and gain phases are estimated simultaneously, an appropriate boundary condition should be imposed on the solution. If the position of the first source is fixed, the impact of this constraint on the CRB can be described by selection matrices.

For the last component of the Jacobian, \mathbf{J}_{σ_n} , we will consider two scenarios. In the first scenario, all receiving elements have possibly different noise powers such that $\Sigma_n = \text{diag}(\sigma_n)$. In the second scenario, all system noise powers are assumed equal such that $\Sigma_n = \sigma_n \mathbf{I}$. Beginning with the first scenario, we note that

$$\frac{\partial \text{vec}(\mathbf{R})}{\partial \sigma_{ni}} = \text{vec}(\mathbf{E}_{ii}) = \mathbf{e}_i \circ \mathbf{e}_i. \quad (4.36)$$

Collecting the derivatives for all values of i and stacking them in a single matrix, we obtain

$$\mathbf{J}_{\sigma_n} = \mathbf{I} \circ \mathbf{I}. \quad (4.37)$$

If all receiving elements have equal noise power, the second scenario, the Jacobian is given by

$$\mathbf{J}_{\sigma_n} = \frac{\partial \text{vec}(\mathbf{R})}{\partial \sigma_n} = \frac{\partial \text{vec}(\sigma_n \mathbf{I})}{\partial \sigma_n} = \text{vec}(\mathbf{I}). \quad (4.38)$$

4.2.3 Constituents of the FIM

Using the expressions for the Jacobians for the appropriate parameter sets derived in the previous section, it is straightforward to compute the FIM from (4.9) and invert the result to obtain the CRB for the problem at hand. This works conveniently for toy examples with $P = 10$, but becomes impractical for real life situations. For example, if we want to compute the CRB for an array with $P = 100$ for estimation of direction independent gains and phases and receiver noise powers ($L = 299$), we need to multiply a 299×10^4 matrix with a $10^4 \times 10^4$ matrix and multiply the result with another $10^4 \times 299$ matrix. This requires a considerable amount of memory, that is not available on most commonly used desktop computers. The sizes of the matrices involved are driven by the Kronecker and Khatri-Rao products. Fortunately, there are a number of convenient relations that may be exploited to reduce Kronecker products to Khatri-Rao products and Khatri-Rao products to Hadamard products, each reducing one matrix dimension from P^2 to P . These relations are summarized in App. C. We will illustrate this reduction by computing $\mathbf{F}_{\gamma\gamma}$, $\mathbf{F}_{\gamma\phi}$ and $\mathbf{F}_{\phi\phi}$ and leave the other constituents of the FIM as an exercise for the reader.

The first constituent, $\mathbf{F}_{\gamma\gamma}$, follows from (4.9) as

$$\mathbf{F}_{\gamma\gamma} = N \mathbf{J}_{\gamma}^H \left(\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1} \right) \mathbf{J}_{\gamma}. \quad (4.39)$$

Substitution of (4.21) gives

$$\begin{aligned} \mathbf{F}_{\gamma\gamma} &= N \left(\overline{\mathbf{G}} \overline{\mathbf{R}}_0 \Phi \circ \mathbf{I} + \mathbf{I} \circ \mathbf{G} \mathbf{R}_0 \overline{\Phi} \right)^H \left(\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1} \right) \\ &\quad \times \left(\overline{\mathbf{G}} \overline{\mathbf{R}}_0 \Phi \circ \mathbf{I} + \mathbf{I} \circ \mathbf{G} \mathbf{R}_0 \overline{\Phi} \right). \end{aligned} \quad (4.40)$$

Using (C.10), the Kronecker product can be replaced by a Khatri-Rao product. The result is

$$\begin{aligned} \mathbf{F}_{\gamma\gamma} &= N \left(\overline{\mathbf{G}} \overline{\mathbf{R}}_0 \Phi \circ \mathbf{I} + \mathbf{I} \circ \mathbf{G} \mathbf{R}_0 \overline{\Phi} \right)^H \\ &\quad \times \left(\overline{\mathbf{R}}^{-1} \overline{\mathbf{G}} \overline{\mathbf{R}}_0 \Phi \circ \mathbf{R}^{-1} \mathbf{I} + \overline{\mathbf{R}}^{-1} \mathbf{I} \circ \mathbf{R}^{-1} \mathbf{G} \mathbf{R}_0 \overline{\Phi} \right). \end{aligned} \quad (4.41)$$

The Khatri-Rao products may be reduced to Hadamard products by applying (C.11). This gives

$$\begin{aligned} \mathbf{F}_{\gamma\gamma} &= N \left(\overline{\Phi} \overline{\mathbf{R}}_0 \mathbf{G} \overline{\mathbf{R}}^{-1} \overline{\mathbf{G}} \overline{\mathbf{R}}_0 \Phi \odot \mathbf{R}^{-1} + \overline{\Phi} \overline{\mathbf{R}}_0 \mathbf{G} \overline{\mathbf{R}}^{-1} \odot \mathbf{R}^{-1} \mathbf{G} \mathbf{R}_0 \overline{\Phi} + \right. \\ &\quad \left. + \overline{\mathbf{R}}^{-1} \overline{\mathbf{G}} \overline{\mathbf{R}}_0 \Phi \odot \mathbf{F} \mathbf{R}_0 \overline{\mathbf{G}} \overline{\mathbf{R}}^{-1} + \overline{\mathbf{R}}^{-1} \odot \mathbf{F} \mathbf{R}_0 \overline{\mathbf{G}} \overline{\mathbf{R}}^{-1} \mathbf{G} \mathbf{R}_0 \overline{\Phi} \right). \end{aligned} \quad (4.42)$$

Note that the first and last terms are each other's complex conjugates as are the second and third terms. We can therefore simplify this result further to

$$\mathbf{F}_{\gamma\gamma} = 2N \operatorname{Re} \{ \mathbf{M}_1 + \mathbf{M}_2 \} \quad (4.43)$$

where we have introduced

$$\mathbf{M}_1 = \overline{\mathbf{R}}^{-1} \odot \Phi \mathbf{R}_0 \overline{\mathbf{G}} \mathbf{R}^{-1} \mathbf{G} \mathbf{R}_0 \overline{\Phi} \quad (4.44)$$

$$\mathbf{M}_2 = \overline{\Phi} \mathbf{R}_0 \mathbf{G} \overline{\mathbf{R}}^{-1} \odot \mathbf{R}^{-1} \mathbf{G} \mathbf{R}_0 \overline{\Phi}. \quad (4.45)$$

Equation (4.24) shows that \mathbf{J}_ϕ has a more complicated structure than \mathbf{J}_γ since the factors with a Khatri-Rao product are pre-multiplied by a scalar and post-multiplied by a selection matrix. The submatrices $\mathbf{J}_{\tilde{l}}$, $\mathbf{J}_{\tilde{m}}$ and $\mathbf{J}_{\tilde{n}}$ have a similar form. The derivation of the corresponding blocks of the FIM is therefore similar to the derivation for $\mathbf{F}_{\gamma\phi}$ and $\mathbf{F}_{\phi\phi}$. Starting with the first, we have

$$\begin{aligned} \mathbf{F}_{\gamma\phi} &= jN (\overline{\mathbf{G}} \mathbf{R}_0 \Phi \circ \mathbf{I} + \mathbf{I} \circ \mathbf{G} \mathbf{R}_0 \overline{\Phi})^H (\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) \\ &\quad \times (\overline{\mathbf{G}} \mathbf{R}_0 \mathbf{G} \circ \mathbf{I} - \mathbf{I} \circ \mathbf{G} \mathbf{R}_0 \overline{\mathbf{G}}) \mathbf{I}_s. \end{aligned} \quad (4.46)$$

Reduction of the Kronecker product using (C.10) gives

$$\begin{aligned} \mathbf{F}_{\gamma\phi} &= jN (\overline{\mathbf{G}} \mathbf{R}_0 \Phi \circ \mathbf{I} + \mathbf{I} \circ \mathbf{G} \mathbf{R}_0 \overline{\Phi})^H \\ &\quad \times (\overline{\mathbf{R}}^{-1} \overline{\mathbf{G}} \mathbf{R}_0 \mathbf{G} \circ \mathbf{R}^{-1} \mathbf{I} - \overline{\mathbf{R}}^{-1} \mathbf{I} \circ \mathbf{R}^{-1} \mathbf{G} \mathbf{R}_0 \overline{\mathbf{G}}) \mathbf{I}_s. \end{aligned} \quad (4.47)$$

This result can be reduced further using (C.11), which results in

$$\begin{aligned} \mathbf{F}_{\gamma\phi} &= jN (\overline{\Phi} \mathbf{R}_0 \mathbf{G} \overline{\mathbf{R}}^{-1} \overline{\mathbf{G}} \mathbf{R}_0 \mathbf{G} \odot \mathbf{R}^{-1} - \overline{\Phi} \mathbf{R}_0 \mathbf{G} \overline{\mathbf{R}}^{-1} \odot \mathbf{R}^{-1} \mathbf{G} \mathbf{R}_0 \overline{\mathbf{G}} + \\ &\quad + \overline{\mathbf{R}}^{-1} \overline{\mathbf{G}} \mathbf{R}_0 \mathbf{G} \odot \Phi \mathbf{R}_0 \overline{\mathbf{G}} \mathbf{R}^{-1} - \overline{\mathbf{R}}^{-1} \odot \Phi \mathbf{R}_0 \overline{\mathbf{G}} \mathbf{R}^{-1} \mathbf{G} \mathbf{R}_0 \overline{\mathbf{G}}) \mathbf{I}_s. \end{aligned} \quad (4.48)$$

Using the Hadamard product relations given by (C.8) and (C.9), we may rewrite this as

$$\mathbf{F}_{\gamma\phi} = jN (\overline{\mathbf{M}}_1 \Gamma - \mathbf{M}_2 \Gamma + \overline{\mathbf{M}}_2 \Gamma - \mathbf{M}_1 \Gamma) \mathbf{I}_s. \quad (4.49)$$

Note that each pair of complex conjugates are now subtracted leaving only their imaginary parts. Therefore

$$\mathbf{F}_{\gamma\phi} = 2N \operatorname{Im} \{ \mathbf{M}_1 + \mathbf{M}_2 \} \Gamma \mathbf{I}_s. \quad (4.50)$$

If the factor $\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}$ in (4.9) is pre- and post-multiplied by \mathbf{J}_ϕ , the Hermitian transpose ensures that the selection matrix does not end up between the factors with Khatri-Rao and Kronecker products, i.e.

$$\begin{aligned}\mathbf{F}_{\phi\phi} &= N\mathbf{I}_s^H (\bar{\mathbf{G}}\bar{\mathbf{R}}_0\mathbf{G} \circ \mathbf{I} - \mathbf{I} \circ \mathbf{G}\mathbf{R}_0\bar{\mathbf{G}})^H (\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) \\ &\quad \times (\bar{\mathbf{G}}\bar{\mathbf{R}}_0\mathbf{G} \circ \mathbf{I} - \mathbf{I} \circ \mathbf{G}\mathbf{R}_0\bar{\mathbf{G}}) \mathbf{I}_s.\end{aligned}\quad (4.51)$$

The Kronecker and Khatri-Rao products can therefore be reduced as before giving

$$\begin{aligned}\mathbf{F}_{\phi\phi} &= N\mathbf{I}_s^H (\bar{\mathbf{G}}\bar{\mathbf{R}}_0\mathbf{G} \circ \mathbf{I} - \mathbf{I} \circ \mathbf{G}\mathbf{R}_0\bar{\mathbf{G}})^H \\ &\quad \times (\bar{\mathbf{R}}^{-1} \mathbf{G}\bar{\mathbf{R}}_0\mathbf{G} \circ \mathbf{R}^{-1} - \bar{\mathbf{R}}^{-1} \circ \mathbf{R}^{-1} \mathbf{G}\mathbf{R}_0\bar{\mathbf{G}}) \mathbf{I}_s \\ &= N\mathbf{I}_s^H (\Gamma\bar{\mathbf{M}}_1\Gamma - \Gamma\mathbf{M}_2\Gamma - \Gamma\bar{\mathbf{M}}_2\Gamma + \Gamma\mathbf{M}_1\Gamma) \mathbf{I}_s\end{aligned}\quad (4.52)$$

The constituents of each pair of complex conjugates has the same sign, so the result is two times their real parts. Since Γ is a real valued diagonal matrix, it can be moved outside the Re-operator. Therefore, $\mathbf{F}_{\phi\phi}$ is given by

$$\mathbf{F}_{\phi\phi} = 2N\mathbf{I}_s^H \Gamma \text{Re}\{\mathbf{M}_1 - \mathbf{M}_2\} \Gamma \mathbf{I}_s.\quad (4.53)$$

The submatrices \mathbf{J}_σ and \mathbf{J}_{σ_n} have the same structure as \mathbf{J}_γ and the structure of $\mathbf{J}_{\tilde{l}}$, $\mathbf{J}_{\tilde{m}}$ and $\mathbf{J}_{\tilde{n}}$ is identical to that of \mathbf{J}_ϕ . The derivations of the other FIM submatrices are therefore similar to the ones presented here. The reader should thus find it straightforward to obtain expressions for those components by following the steps above.

4.3 Impact of gain phase constraints

4.3.1 Constraints on the gain phase solution

The problem of estimating the direction independent gain and phase characteristics of the receiving elements requires a boundary condition to obtain a unique solution. The physical reason for this is that the phase of an element within the array can only be determined with respect to some phase reference. The unconstrained solution to this problem will therefore produce a phase solution that correctly predicts the phase differences between the elements but has an arbitrary phase offset. This arbitrary phase offset becomes problematic if multiple frequency channels or multiple subarrays are calibrated independently, but the results are combined at a later stage. Therefore an

appropriate constraint needs to be imposed to fix this offset. It has become common practice in the signal processing literature to impose the constraint that the first element of the array has zero phase, as illustrated in [3, 12, 13, 25, 39, 107]. In mathematical derivations, this choice can be made without loss of generality, and in some practical cases, this choice is required due to the necessity of a well-defined phase reference.

In this section, which was published earlier in [151], analytic expressions are derived for the CRB for the gain and phase estimation problem when calibrating an array of identical elements on a single point source under the constraint that the first element has zero phase and under the constraint that the average phase of all elements is zero. We show that under the latter constraint, the CRB for the phase estimate decreases by a factor $2P/(P - 1)$ as compared to the first constraint. It is then shown that the proposed constraint does not only give a lower CRB but is actually the constraint that minimizes the CRB. We subsequently demonstrate that this constraint is also optimal in the case of an arbitrary source model and arrays of non-identical elements. Finally, the analysis is confirmed by a simulation focusing on the calibration problem of a phased array radio telescope.

4.3.2 The constrained CRB

We consider the estimation of a real valued $L \times 1$ parameter vector θ . Stoica and Ng [108] have shown that if the estimate $\hat{\theta}$ is subject to K continuously differentiable constraints $\mathbf{h}(\theta) = \mathbf{0}$, the CRB has the form

$$\mathbf{C}_\theta = \mathcal{E} \left\{ (\hat{\theta} - \theta) (\hat{\theta} - \theta)^T \right\} \geq \mathbf{U} (\mathbf{U}^T \mathbf{F} \mathbf{U})^{-1} \mathbf{U}^T. \quad (4.54)$$

In (4.54), \mathbf{F} denotes the FIM for the unconstrained parameter estimation problem and \mathbf{U} is an $L \times (L - K)$ matrix whose columns form an orthonormal basis for the null space of the $K \times L$ gradient matrix of $\mathbf{h}(\theta)$,

$$\mathbf{H}(\theta) = \frac{\partial \mathbf{h}(\theta)}{\partial \theta^T}. \quad (4.55)$$

This implies that $\mathbf{H}(\theta) \mathbf{U} = \mathbf{0}$ while $\mathbf{U}^T \mathbf{U} = \mathbf{I}$. Earlier, Gorman and Hero [43] have found a similar expression for the CRB under parametric constraints using a derivation based on the Barankin bound.

For our application, the complex gain estimation problem for sensor arrays can be parameterized in terms of gain amplitudes and gain phases, i.e.

$$\theta = [\gamma^T, \phi^T]^T = [\gamma_1, \gamma_2, \dots, \gamma_P \mid \phi_1, \phi_2, \dots, \phi_P]^T. \quad (4.56)$$

The aim of this analysis is to compare the CRB under the following two parametric constraints:

$$h_1(\boldsymbol{\theta}) = \phi_1 = 0 \quad (4.57)$$

$$h_2(\boldsymbol{\theta}) = \sum_{p=1}^P \phi_p = 0. \quad (4.58)$$

The gradient matrices for these constraints are

$$\mathbf{H}_1(\boldsymbol{\theta}) = [\mathbf{0}_P^T \mid \mathbf{1}, \mathbf{0}_{P-1}^T] \quad (4.59)$$

$$\mathbf{H}_2(\boldsymbol{\theta}) = [\mathbf{0}_P^T \mid \mathbf{1}_P^T] \quad (4.60)$$

where $\mathbf{0}_p$ and $\mathbf{1}_p$ denote column vectors of size $p \times 1$ filled with zeros and ones respectively. Orthonormal bases for the null spaces of these gradient matrices are

$$\mathbf{U}_1 = \left[\begin{array}{c|c} \mathbf{I}_P & \mathbf{0}_{P \times (P-1)} \\ \hline \mathbf{0}_{P \times P} & \mathbf{0}_{P-1}^T \\ & \mathbf{I}_{P-1} \end{array} \right] \quad (4.61)$$

$$\mathbf{U}_2 = \left[\begin{array}{c|c} \mathbf{0}_{P-1}^T & \mathbf{1}_P^T / \sqrt{P} \\ \hline \mathbf{I}_{P-1} & \mathbf{0}_{(P-1) \times P} \\ \hline \mathbf{0}_{P \times (P-1)} & \mathbf{I}_P - \mathbf{1}_P \mathbf{1}_P^T / P \end{array} \right] \quad (4.62)$$

where \mathbf{I}_p denotes the $p \times p$ identity matrix and $\mathbf{0}_{p \times q}$ and $\mathbf{1}_{p \times q}$ denote $p \times q$ matrices filled with zeros and ones respectively. The unconstrained FIM for this problem can be found by taking $\mathbf{I}_s = \mathbf{I}$ in (4.24) and is given by

$$\mathbf{F} = 2N \begin{bmatrix} \text{Re}\{\mathbf{M}_1 + \mathbf{M}_2\} & \text{Im}\{\mathbf{M}_1 + \mathbf{M}_2\} \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} \text{Im}\{\mathbf{M}_1^T + \mathbf{M}_2^T\} & \boldsymbol{\Gamma} \text{Re}\{\mathbf{M}_2 - \mathbf{M}_1\} \boldsymbol{\Gamma} \end{bmatrix} \quad (4.63)$$

where \mathbf{M}_1 and \mathbf{M}_2 are given by (4.44) and (4.45) respectively.

4.3.3 Single calibration source

In this section, we specialize to calibration based on a single point source; the case $Q > 1$ is considered in Sec. 4.3.4. Without loss of generality, we can assume that the single source is positioned in the phase center of the array and has unit power. This simplifies the source covariance model to $\mathbf{R}_0 = \mathbf{1}_P \mathbf{1}_P^T$. Assume also that the array consists of identical elements, so we may take $\boldsymbol{\Gamma} = \mathbf{I}$, $\boldsymbol{\Phi} = \mathbf{I}$ and $\boldsymbol{\Sigma}_n = \sigma_n \mathbf{I}$, where σ_n is the noise power of a single element. The covariance matrix model then

simplifies to $\mathbf{R} = \mathbf{1}_P \mathbf{1}_P^T + \sigma_n \mathbf{I}$, which can be inverted using the Sherman-Morrison formula [6],

$$(\mathbf{B} + \mathbf{b}\mathbf{b}^H)^{-1} = \mathbf{B}^{-1} - \frac{\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^H\mathbf{B}^{-1}}{1 + \mathbf{b}^H\mathbf{B}^{-1}\mathbf{b}}. \quad (4.64)$$

This yields

$$\mathbf{R}^{-1} = \frac{1}{\sigma_n} \left(\mathbf{I} - \frac{1}{\sigma_n + P} \mathbf{1}_P \mathbf{1}_P^T \right). \quad (4.65)$$

Using these simplifications, (4.44) and (4.45) reduce to

$$\mathbf{M}_1 = \frac{P}{\sigma_n (\sigma_n + P)} \left(\mathbf{I} - \frac{1}{\sigma_n + P} \mathbf{1}_P \mathbf{1}_P^T \right) \quad (4.66)$$

$$\mathbf{M}_2 = \frac{1}{(\sigma_n + P)^2} \mathbf{1}_P \mathbf{1}_P^T. \quad (4.67)$$

The FIM is then described by

$$\mathbf{F} = \frac{2}{\sigma_n (\sigma_n + P)} \begin{bmatrix} P\mathbf{I} + \frac{\sigma_n - P}{\sigma_n + P} \mathbf{1}_P \mathbf{1}_P^T & \mathbf{0}_{P \times P} \\ \mathbf{0}_{P \times P} & P\mathbf{I} - \mathbf{1}_P \mathbf{1}_P^T \end{bmatrix}. \quad (4.68)$$

Note (as expected) that \mathbf{F} is singular with a one-dimensional null space spanned by the vector

$$\mathbf{v}_0 = \frac{1}{\sqrt{P}} \begin{bmatrix} \mathbf{0}_P \\ \mathbf{1}_P \end{bmatrix}. \quad (4.69)$$

Thus, a constraint to avoid this singularity is necessary.

Substitution of (4.61) and (4.68) in (4.54) gives the CRB under the constraint $\phi_1 = 0$ as

$$\mathbf{C}_{\boldsymbol{\theta}} \geq \frac{\sigma_n (\sigma_n + P)}{2P} \begin{bmatrix} \mathbf{I}_P - \frac{\sigma_n - P}{2\sigma_n P} \mathbf{1}_P \mathbf{1}_P^T & \mathbf{0}_{P \times P} \\ \mathbf{0}_{P \times P} & \begin{array}{c|c} 0 & \mathbf{0}_{P-1}^T \\ \hline \mathbf{0}_{P-1} & \mathbf{I}_{P-1} + \mathbf{1}_{P-1} \mathbf{1}_{P-1}^T \end{array} \end{bmatrix}. \quad (4.70)$$

This implies that

$$\text{var}(\gamma_p) = \frac{\sigma_n (\sigma_n + P)}{2P} \left(1 - \frac{\sigma_n - P}{2\sigma_n P} \right) \quad (4.71)$$

for all elements and that

$$\text{var}(\phi_p) = \frac{\sigma_n (\sigma_n + P)}{P} \quad (4.72)$$

for all elements except the first. The variance of the phase of the first element is zero, since $\phi_1 = 0$ by definition.

The CRB under the constraint $\sum_{p=1}^P \phi_p = 0$ can be found in a similar way using (4.54), (4.62) and (4.68) as

$$\mathbf{C}_{\boldsymbol{\theta}} \geq \frac{\sigma_n (\sigma_n + P)}{2P} \begin{bmatrix} \mathbf{I}_P - \frac{\sigma_n - P}{2\sigma_n P} \mathbf{1}_P \mathbf{1}_P^T & \mathbf{0}_{P \times P} \\ \mathbf{0}_{P \times P} & \mathbf{I}_P - \mathbf{1}_P \mathbf{1}_P^T / P \end{bmatrix}. \quad (4.73)$$

This means that

$$\text{var}(\gamma_p) = \frac{\sigma_n (\sigma_n + P)}{2P} \left(1 - \frac{\sigma_n - P}{2\sigma_n P} \right) \quad (4.74)$$

for all elements. This result is the same as the result obtained under the constraint $\phi_1 = 0$. This outcome was expected since neither constraint affects the gain estimates. For the variance of the phase estimates, the result is

$$\text{var}(\phi_p) = \frac{\sigma_n (\sigma_n + P)}{2P} \left(1 - \frac{1}{P} \right) \quad (4.75)$$

which differs from the previous result by a factor $2P/(P-1)$. If P is reasonably large, this factor is approximately equal to 2.

In [24], De Carvalho et al. pose that among all sets of a minimal number of independent constraints, *the pseudo-inverse of the unconstrained FIM \mathbf{F}^\dagger yields the lowest value for the total variance on all estimated parameters*. An eigenvalue decomposition on \mathbf{F} will yield $r = \text{rank}(\mathbf{F})$ eigenvalues $\lambda_i \neq 0$ with corresponding eigenvectors \mathbf{v}_i . If $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$ and $\Lambda = \text{diag}([\lambda_1, \lambda_2, \dots, \lambda_r])$, the pseudo-inverse inverse can be computed by

$$\mathbf{F}^\dagger = \mathbf{V} \Lambda^{-1} \mathbf{V}^T = \mathbf{V} (\mathbf{V}^T \mathbf{F} \mathbf{V})^{-1} \mathbf{V}^T \quad (4.76)$$

which has the same form as (4.54). Therefore, to achieve the lowest total variance on all estimated parameters, we need to set $\mathbf{U} = \mathbf{V}$ or (more in general) select \mathbf{U} such that the column span of \mathbf{U} equals the column span of \mathbf{V} . In other words, the boundary conditions should be selected such that their gradient matrix spans the orthogonal complement of \mathbf{V} , the range of the FIM. This complement space is described by the eigenvectors corresponding to the zero-valued eigenvalues of the FIM.

In the case studied here, one eigenvalue of \mathbf{F} in (4.63) is equal to zero with corresponding eigenvector $\mathbf{v}_0 = [\mathbf{0}_P^T, \mathbf{1}_P^T]^T / \sqrt{P}$. This is precisely the space spanned by the gradient matrix $\mathbf{H}_2(\boldsymbol{\theta})$ in (4.60). This proves that $\sum_{p=1}^P \phi_p = 0$ is a constraint that minimizes the total variance on the estimated parameters.

4.3.4 Multiple calibration sources

When studying more complicated source models following the steps outlined in the previous section, one does not obtain a pedagogically interesting expression. We will

show below, however, that \mathbf{v}_0 still lies in the null space of \mathbf{F} in the general case with an arbitrary source covariance model \mathbf{R}_0 and for arbitrary direction independent gains and phases, i.e. for $\boldsymbol{\Gamma} \neq \mathbf{I}$ and $\boldsymbol{\Phi} \neq \mathbf{I}$. This suggests that the constraint $\sum_{p=1}^P \phi_p = 0$ also minimizes the total variance on all parameters in the more general case.

We need to show that $\mathbf{F}\mathbf{v}_0 = \mathbf{0}$. Since \mathbf{F} is a positive semi-definite matrix, it suffices to prove that $\mathbf{v}_0^T \mathbf{F} \mathbf{v}_0 = 0$. Using the structure of \mathbf{v}_0 , this reduces to proving that, for the $(2, 2)$ -block of \mathbf{F} ,

$$\mathbf{1}_P^T \operatorname{Re} \{ \boldsymbol{\Gamma} (\mathbf{M}_2 - \mathbf{M}_1) \boldsymbol{\Gamma} \} \mathbf{1}_P = 0. \quad (4.77)$$

As a starting point, we note that

$$\boldsymbol{\Gamma} \mathbf{M}_1 \boldsymbol{\Gamma} = \boldsymbol{\Gamma} \overline{\mathbf{R}}^{-1} \boldsymbol{\Gamma} \odot (\boldsymbol{\Phi} \mathbf{R}_0 \overline{\boldsymbol{\Phi}}) (\boldsymbol{\Gamma} \mathbf{R}^{-1} \boldsymbol{\Gamma}) (\boldsymbol{\Phi} \mathbf{R}_0 \overline{\boldsymbol{\Phi}}) = \overline{\mathbf{Y}} \odot \mathbf{Z} \mathbf{Y} \mathbf{Z} \quad (4.78)$$

where we have introduced the Hermitian matrices $\mathbf{Y} = \boldsymbol{\Gamma} \mathbf{R}^{-1} \boldsymbol{\Gamma}$ and $\mathbf{Z} = \boldsymbol{\Phi} \mathbf{R}_0 \boldsymbol{\Phi}^H$. In a similar way, we find that

$$\boldsymbol{\Gamma} \mathbf{M}_2 \boldsymbol{\Gamma} = \overline{\mathbf{Z} \mathbf{Y}} \odot \mathbf{Y} \mathbf{Z}. \quad (4.79)$$

By inserting (4.78) and (4.79) in (4.77), we obtain

$$\begin{aligned} \mathbf{1}_P^T \operatorname{Re} \{ \boldsymbol{\Gamma} (\mathbf{M}_2 - \mathbf{M}_1) \boldsymbol{\Gamma} \} \mathbf{1}_P &= \\ &= \mathbf{1}_P^T \operatorname{Re} \{ \overline{\mathbf{Z} \mathbf{Y}} \odot \mathbf{Y} \mathbf{Z} - \mathbf{Z} \mathbf{Y} \mathbf{Z} \odot \overline{\mathbf{Y}} \} \mathbf{1}_P \\ &= \operatorname{Re} \left\{ \operatorname{diag} \left((\overline{\mathbf{Z} \mathbf{Y}})^T \mathbf{Y} \mathbf{Z} - (\mathbf{Z} \mathbf{Y} \mathbf{Z})^T \overline{\mathbf{Y}} \right)^T \right\} \mathbf{1}_P \\ &= \operatorname{Re} \left\{ \operatorname{tr} \left((\mathbf{Z} \mathbf{Y} \mathbf{Z} \mathbf{Y})^H - \overline{\mathbf{Z} \mathbf{Y} \mathbf{Z} \mathbf{Y}} \right) \right\} \\ &= 0, \end{aligned} \quad (4.80)$$

which completes the proof.

4.3.5 Application example

We will illustrate the multi-source case by an example. For this example, we will use the antenna configuration of LOFAR's initial test station (ITS) [140], which is shown in the right panel of Fig. 2.5. ITS consisted of $P = 60$ inverted V-shaped dipoles arranged in a five-armed spiral configuration. We will assume a sky model at 30 MHz consisting of the strongest $Q = 10$ astronomical sources that were visible in the sky above ITS on January 26, 2005 at midnight. The source locations and power ratios were taken from the third Cambridge (3C) catalog of radio sources [8]. The total power of these sources was assumed to be 1% of the system noise power

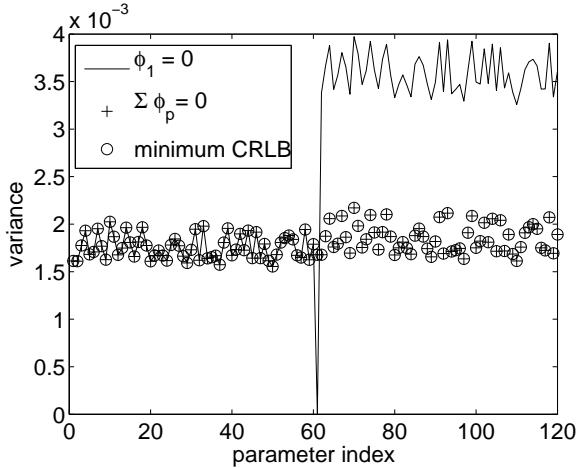


Figure 4.1: CRBs for multi-source complex gain estimation with $\boldsymbol{\theta} = [\gamma_1, \gamma_2, \dots, \gamma_P, \phi_1, \phi_2, \dots, \phi_P]^T$ under the constraints $\phi_1 = 0$ and $\sum_{p=1}^P \phi_p = 0$ as well as the CRB based on the pseudo-inverse of the FIM that should give the lowest total variance on the estimated parameters. Note that the latter coincides with the CRB under the constraint $\sum_{p=1}^P \phi_p = 0$, indicating that this boundary condition leads to the lowest possible CRB.

of the individual antennas. The computations were done assuming integration over $N = 156250$ samples.

We compare the CRBs for both boundary conditions ($\phi_1 = 0$ and $\sum_{p=1}^P \phi_p = 0$) with the lower bound given by \mathbf{F}^\dagger . We take identical elements as parameter values, i.e. $\boldsymbol{\Gamma} = \mathbf{I}$ and $\boldsymbol{\Phi} = \mathbf{I}$. The results are shown in the right panel of Fig. 4.1. The horizontal axis is the index of the individual parameters in the parameter vector $\boldsymbol{\theta}$ in (4.56). The vertical axis is the computed bound on the variance.

The plot shows that, as in the single source case, the constraints have no impact on the variance of the gain estimates and that the variance on the phase estimates decreases by approximately a factor 2. The plot also shows that the variance varies from one element to another. This can be explained by signal-to-noise ratio differences between baselines. As a result, some elements have a less favorable position in the array than others and therefore have less accurate gain and phase estimates.

In the previous subsection, it was demonstrated that \mathbf{v}_0 lies in the null space of the FIM, suggesting that the boundary condition $\sum_{p=1}^P \phi_p = 0$ gives the same CRB as the CRB obtained by taking the pseudo-inverse of the FIM, which gives

the lowest total variance on the estimated parameters. In Fig. 4.1, we also plotted the latter CRB for the multiple source case considered here. The CRB under the constraint $\sum_{p=1}^P \phi_p = 0$ coincides within the numerical accuracy of the simulation with the CRB based on the pseudo-inverse of the FIM as expected based on our earlier analysis.

4.3.6 Concluding remarks

The CRB for estimation of the direction independent gains and phases was studied under the constraints $\phi_1 = 0$ and $\sum_{p=1}^P \phi_p = 0$. It was found that for calibration on a single point source of an array of identical elements, the variance on the phase estimates is larger by a factor $2P/(P-1)$ under the first boundary condition as compared to the latter, which is in fact the constraint that minimizes the total variance on the estimated parameters. It was also shown that similar conclusions hold for more general cases involving arbitrary source covariance models and arrays of non-identical elements.

We offer the following intuition for this loss by a factor of almost 2. Suppose that the phase parameter vector is estimated without a proper phase constraint. In that case, all phases are shifted by an arbitrary offset that cannot be identified. An implementation of the first constraint would set $\phi_1 = 0$ by subtracting ϕ_1 from all other phase estimates. This reduces the variance of the estimate of ϕ_1 to zero but adds its original variance to the variance on the phases of all elements, thus doubling the variance on their phase estimates. The second boundary condition would compute the average phase and subtract it from all elements, which only leads to a small increase in the variance on the phase estimates. This intuition can be followed in many similar cases as well, e.g. in the context of blind source separation and equalization, where the phase is often not uniquely identifiable.

It can be shown that if the estimation problem is reformulated in terms of $P-1$ phase differences, the variance on these $P-1$ parameters are the same under both constraints. This indicates that if the system only uses knowledge of the phase differences between the elements, the result is not affected by the choice for either of the two boundary conditions discussed in this section. An example of such a situation is the power output of the beam former of an array of antennas. The difference in the variance on the phase estimates does matter, if phase stability of the whole system is required. An example of such a system is an array of subarrays in which the beam formed output of the subarrays is processed further at system level. If the subarrays are constantly recalibrated independently, their phase references should be as stable as possible. In such cases, the average phase may provide a more suitable reference than the phase of a single element. Such a system would require careful tracking of the phases, since the constraint $\sum_{p=1}^P \phi_p = 0$ is insensitive to a phase offset that is

an integral multiple of $2\pi/P$. If the phase variations between consecutive calibration measurements are too large, the phase reference may exhibit a phase jump of an integral multiple of $2\pi/P$.

4.4 Dealing with missing data

4.4.1 Problem statement

The CRBs for radio astronomical calibration problems commonly assume the use of the full array covariance matrix $\hat{\mathbf{R}}$ [12, 120, 149, 165], while the existing algorithms for estimation of instrumental and source parameters effectively discard the diagonal entries of $\hat{\mathbf{R}}$ [12, 149, 165]. This makes sense because the system noise powers of the receiving elements are assumed unknown and mutually uncorrelated, which results in a diagonal but unknown noise covariance matrix Σ_n . A general formulation of the data model applicable to all the commonly used models [12, 13, 52, 140, 165] is therefore

$$\mathbf{R} = \mathbf{R}_s(\boldsymbol{\theta}) + \Sigma_n, \quad (4.81)$$

where $\mathbf{R}_s(\boldsymbol{\theta})$ is the parameterized model of the noise-free array covariance matrix. Existing algorithms often only estimate $\boldsymbol{\theta}$ and simply discard the diagonal of $\hat{\mathbf{R}}$. The lower bound for the case that the diagonal of $\hat{\mathbf{R}}$ is used is of course also a lower bound for the case in which it is discarded, because discarding data can only increase the variance. However, the question remains whether the CRB is higher if we take into account that the data has been discarded. If so, the diagonal contains some useful information to estimate $\boldsymbol{\theta}$ and there might exist an algorithm that exploits that information. In this section, whose material was published earlier in [119], we investigate this issue using a CRB analysis.

Intuitively, we expect that it does not matter whether the diagonal of $\hat{\mathbf{R}}$ is discarded or not if Σ_n is unknown. This intuition is based on the form of the commonly used least squares cost function. If Σ_n is known, the least squares estimation problem is described by

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\| \hat{\mathbf{R}} - (\mathbf{R}_s(\boldsymbol{\theta}) + \Sigma_n) \right\|_F^2 \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i=1}^P \sum_{j=1}^P \left| \hat{R}_{ij} - (R_{s,ij}(\boldsymbol{\theta}) + \Sigma_{n,ij}) \right|^2. \end{aligned} \quad (4.82)$$

For unknown $\Sigma_n = \text{diag}(\boldsymbol{\sigma}_n)$ where $\boldsymbol{\sigma}_n = [\sigma_{n1}, \sigma_{n2}, \dots, \sigma_{nP}]^T$, this becomes

$$\left\{ \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\sigma}}_n \right\} = \underset{\boldsymbol{\theta}, \boldsymbol{\sigma}_n}{\text{argmin}} \sum_{i=1}^P \sum_{j=1}^P \left| \widehat{R}_{ij} - (R_{s,ij}(\boldsymbol{\theta}) + \sigma_{ni} \delta_{ij}) \right|^2. \quad (4.83)$$

If $\boldsymbol{\sigma}_n$ is unknown, its constituents σ_{ni} only appear in one term of the sum, so we are free to choose σ_{ni} such that

$$\left| \widehat{R}_{ii} - (R_{s,ii}(\boldsymbol{\theta}) + \sigma_{ni}) \right|^2 = 0. \quad (4.84)$$

Therefore, this term is always zero and the solution for $\boldsymbol{\theta}$ only depends on the remaining terms, i.e.

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{argmin}} \sum_{\substack{1 \leq i, j \leq P \\ i \neq j}} \left| \widehat{R}_{ij} - R_{s,ij}(\boldsymbol{\theta}) \right|^2. \quad (4.85)$$

The availability of \widehat{R}_{ii} does not matter for the estimation of $\boldsymbol{\theta}$ since it does not appear in (4.85). If \widehat{R}_{ii} is missing, the term in (4.84) cannot be evaluated, so it needs to be removed from the least squares problem. The remaining problem is given by (4.85).

We thus conclude that in solving for $\boldsymbol{\theta}$ using the least squares method it does not matter whether either σ_{ni} is unknown or \widehat{R}_{ii} is unavailable or both. This suggests that the theoretically best attainable statistical performance is the same for all cases. To see whether this is true for any estimator, we will evaluate the CRB for four cases whether we consider σ_{ni} as known or unknown and whether \widehat{R}_{ii} is available or not.

4.4.2 Derivation of the CRB

If the $P \times 1$ array signal vectors $\mathbf{x}[n]$ are i.i.d. with a complex normal distribution with zero mean and variance \mathbf{R} , i.e. $\mathbf{x}[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$, the array covariance matrix estimate

$$\widehat{\mathbf{R}} = \sum_{n=1}^N \mathbf{x}[n] \mathbf{x}^H[n] \quad (4.86)$$

will have a complex Wishart distribution with N degrees of freedom, i.e. $\widehat{\mathbf{R}} \sim \mathcal{CW}(\mathbf{R}, N)$. The probability density function of the complex Wishart distribution is given by [18]

$$p\left(\widehat{\mathbf{R}}; \mathbf{R}\right) = \frac{1}{c_{P,N} |\mathbf{R}|^N} \left| \widehat{\mathbf{R}} \right|^{N-P} e^{-\text{tr}(\mathbf{R}^{-1} \widehat{\mathbf{R}})} \quad (4.87)$$

where $c_{P,N}$ is a normalization constant. This distribution has the following properties:

$$\mathcal{E} \left\{ \widehat{\mathbf{R}} \right\} = N\mathbf{R} \quad (4.88)$$

$$\text{cov} \left(\widehat{\mathbf{R}} \right) = N\overline{\mathbf{R}} \otimes \mathbf{R} \quad (4.89)$$

Note that these are the expected properties of $\widehat{\mathbf{R}}$ if the samples $\mathbf{x}[n]$ are simply added instead of averaged as was done in Ch. 3.

If a parameter is estimated from samples x_1, x_2, \dots, x_N , but the estimator only uses the samples x_1, x_2, \dots, x_{N-1} , i.e. the N th sample is discarded or not available, the calculation of the CRB requires the pdf of x_1, x_2, \dots, x_{N-1} . Suppose that the pdf of the samples x_1, x_2, \dots, x_N is known and given by $p(x_1, x_2, \dots, x_N)$. The pdf of the samples x_1, x_2, \dots, x_{N-1} can then be found by integrating $p(x_1, x_2, \dots, x_N)$ over x_N from $-\infty$ to ∞ [63], i.e.

$$p(x_1, x_2, \dots, x_{N-1}) = \int_{-\infty}^{\infty} p(x_1, x_2, \dots, x_N) dx_N. \quad (4.90)$$

From the Wishart distribution and the definition of the FIM given in (4.3), one can derive the FIM for Gaussian sources. The result is Bangs' formula which can be written in the compact form given in (4.9). Now we consider the elements that do not use the first element of the main diagonal of the array covariance matrix, i.e. we consider the estimators that either discard or do not have knowledge of \widehat{R}_{11} . The FIM for these estimators is based on the pdf of the remaining data. Based on (4.90), this pdf can be found by integrating the pdf of the Wishart distribution given in (4.87) over \widehat{R}_{11} . For convenience of notation we define $\widehat{\mathbf{R}}_0$ to be equal to $\widehat{\mathbf{R}}$ except for the (1, 1)th element, which is equal to zero. We also define $\widehat{\mathbf{R}}_{1,1}$ as the submatrix obtained from $\widehat{\mathbf{R}}$ by omitting its first row and first column. We can then express the pdf for our problem as

$$\begin{aligned} p \left(\widehat{\mathbf{R}}_0; \mathbf{R} \right) &= \frac{1}{c_{P,N} |\mathbf{R}|^N} \left| \widehat{\mathbf{R}}_{1,1} \right|^{N-P} \left([\mathbf{R}^{-1}]_{11} \right)^{N-P+1} \\ &\times \exp \left(-\text{tr} \left(\mathbf{R}^{-1} \widehat{\mathbf{R}}_0 \right) + \frac{[\mathbf{R}^{-1}]_{11}}{\left[\widehat{\mathbf{R}}_0^{-1} \right]_{11}} \right). \end{aligned} \quad (4.91)$$

From this pdf, the FIM for the case in which \widehat{R}_{11} is unavailable but σ_{n1} is known can be derived. The derivation is somewhat tedious and does not provide any additional insight. We will therefore only state the result, which can be expressed in a

simple form; the FIM is Bangs' formula minus a penalty for discarding the data:

$$\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \mathbf{J}^H \left(N \left(\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1} \right) - (N - P + 1) \mathbf{P} \right) \mathbf{J}, \quad (4.92)$$

where \mathbf{P} is a “penalty” matrix defined as

$$\mathbf{P} = \frac{\text{vec}(\mathbf{R}^{-1} \mathbf{E}_{11} \mathbf{R}^{-1}) \text{vec}(\mathbf{R}^{-1} \mathbf{E}_{11} \mathbf{R}^{-1})^H}{[\mathbf{R}^{-1}]_{11}^2}. \quad (4.93)$$

If we modify the parameter vector to $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}^T, \sigma_{n1}]^T$ to include the unknown parameter σ_{n1} , the FIM is extended by one row and one column, i.e.

$$\mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}} = \begin{bmatrix} \mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \mathbf{F}_{\boldsymbol{\theta}\sigma_{n1}} \\ \mathbf{F}_{\sigma_{n1}\boldsymbol{\theta}} & \mathbf{F}_{\sigma_{n1}\sigma_{n1}} \end{bmatrix}. \quad (4.94)$$

The CRB for $\tilde{\boldsymbol{\theta}}$ is the inverse of $\mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}$ which can be partitioned as

$$\mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1} = \begin{bmatrix} [\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}]_{\boldsymbol{\theta}\boldsymbol{\theta}} & [\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}]_{\boldsymbol{\theta}\sigma_{n1}} \\ [\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}]_{\sigma_{n1}\boldsymbol{\theta}} & [\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}]_{\sigma_{n1}\sigma_{n1}} \end{bmatrix}. \quad (4.95)$$

Since our interests lie with the impact on the parameter vector $\boldsymbol{\theta}$ we will focus on the upper left block. Using the Schur complement [75], this block can be written as

$$[\mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1}]_{\boldsymbol{\theta}\boldsymbol{\theta}} = \left(\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \mathbf{F}_{\boldsymbol{\theta}\sigma_{n1}} \mathbf{F}_{\sigma_{n1}\sigma_{n1}}^{-1} \mathbf{F}_{\sigma_{n1}\boldsymbol{\theta}} \right)^{-1}. \quad (4.96)$$

We thus define the modified FIM as

$$\tilde{\mathbf{F}}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \mathbf{F}_{\boldsymbol{\theta}\sigma_{n1}} \mathbf{F}_{\sigma_{n1}\sigma_{n1}}^{-1} \mathbf{F}_{\sigma_{n1}\boldsymbol{\theta}}. \quad (4.97)$$

The CRB for the original parameter vector $\boldsymbol{\theta}$ in the presence of the additional parameter σ_{n1} is given by the inverse of the modified FIM. If we apply (4.97) to either the case in which \hat{R}_{11} is discarded using the pdf given in (4.91) or the case in which the full array covariance matrix is available using Bangs' formula given in (4.9), we obtain

$$\tilde{\mathbf{F}}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \mathbf{J}^H \left(N \left(\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1} \right) - N \mathbf{P} \right) \mathbf{J}. \quad (4.98)$$

This shows that if σ_{n1} is unknown, the corresponding entry of the array covariance matrix \hat{R}_{11} can be discarded without changing the bound. In other words, if σ_{n1} is not known, an algorithm that does not use R_{11} can be statistically efficient. For convenience of the reader, all results are summarized in Table 4.1.

	σ_{n1} known	σ_{n1} unknown
\widehat{R}_{11} included	$\mathbf{S} = \mathbf{Q}$	$\mathbf{S} = \mathbf{Q} - N\mathbf{P}$
\widehat{R}_{11} excluded	$\mathbf{S} = \mathbf{Q} - (N - P + 1)\mathbf{P}$	$\mathbf{S} = \mathbf{Q} - N\mathbf{P}$

Table 4.1: The FIM for the four different cases is given by $\mathbf{J}^H \mathbf{S} \mathbf{J}$. The matrix \mathbf{Q} is defined as $\mathbf{Q} = N \left(\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1} \right)$.

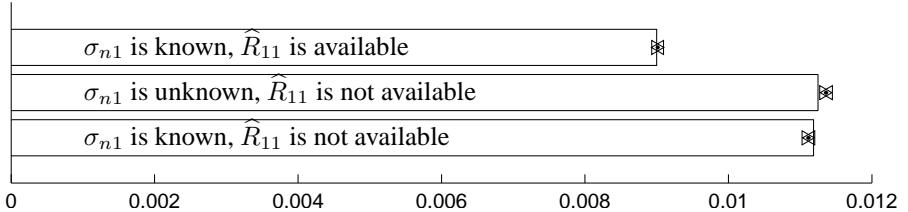


Figure 4.2: CRBs (bars) and mean square errors of the absolute gain of the first element obtained by a MLE in Monte Carlo simulations. The markers indicate the 99% confidence intervals.

4.4.3 Discussion

This analysis also shows that if \widehat{R}_{11} has not been observed, it still matters whether σ_{n1} is known or not. This may seem surprising at first glance but a second thought quickly suggests that this may be related to the fact that the statistical performance of least squares algorithms often improves if appropriate weighting is applied to the data and that the required weights often depend on Σ_n . This argument indicates that the additional information may indeed be exploited. We have verified this conjecture by implementing the maximum likelihood estimator (MLE) for three cases:

1. σ_{n1} is known and \widehat{R}_{11} is available;
2. σ_{n1} is unknown and \widehat{R}_{11} is not available;
3. σ_{n1} is known and \widehat{R}_{11} is not available.

We have performed a Monte Carlo simulation with 10^6 runs for each scenario with $P = 3$ antennas and $N = 100$ samples. The parameter vector consisted of the absolute gains and the phases, i.e. $\boldsymbol{\theta} = [\gamma_1, \gamma_2, \gamma_3, \phi_2, \phi_3]$.

In Fig. 4.2 the CRB and the mean square error of the absolute gain of the first element have been plotted. The figure shows that the MLE performs better when σ_{n1} is known. The mean square error even lies slightly below the CRB. This is possible

because the MLE for this problem is slightly biased at small N . For large N the performance equals the bound, but also the differences between the CRB for the case in which σ_{n1} is known and the CRB for the scenario in which it is not known, goes to zero. In any practical scenario the number of samples N is much larger than the number of sensors P and then the difference becomes negligible. Apart from this consideration, it is questionable whether there exists any practical situation in which one knows σ_{n1} but does not know \hat{R}_{11} .

The most important conclusion therefore remains that if the noise powers are unknown, an algorithm that discards or ignores the diagonal entries while estimating θ can still be statistically efficient since discarding the diagonal of $\hat{\mathbf{R}}$ has no influence on the CRB. It also indicates that the CRB in estimation problems in which some entries of the array covariance matrix are discarded may be computed by extending the parameter vector with additive nuisance parameters on the appropriate entries. The intuitive rationale for this is provided by considering the contribution of the individual entries to the value of a least squares cost function.

Chapter 5

Calibration

5.1 Introduction

The calibration of aperture array radio telescopes will require novel approaches compared to the methods currently used to operate the classical radio telescopes based on a single receiver in the focal plane of a large parabolic dish like the WSRT and the VLA. This statement is motivated by three reasons. First of all, aperture arrays have to sample their entire collecting area while classical radio telescopes use their parabolic dishes to focus all received power at the receiver. Aperture arrays therefore require a large number of receiving elements to provide sufficient collecting area to meet the desired sensitivity. Instruments like LOFAR [15, 127], MWA [72, 73] and EMBRACE [1, 2, 84] will have $\sim 10,000$ elements, which is three orders of magnitudes more than current radio telescopes like the WSRT (14 dishes) and the VLA (27 dishes). The SKA [31, 44] is envisaged to have $\sim 10^6$ elements. These numbers enforce clever distributed signal processing schemes to keep the data streams manageable and clever algorithms which can run on suitable high performance computing hardware.

In practice, this challenge is dealt with by adopting a hierarchical system design as described in Sec. 1.3 and depicted in Fig. 1.5. In such a design, the array is subdivided in subarrays commonly referred to as stations. Each station consists of a number of antennas, which may either be individual receiving elements or compound elements. These compound elements or tiles consist of multiple receiving elements whose output signals are combined in an analog beam former to produce the tile output. A nice example of such a system that uses all the aforementioned aggregation levels, is the LOFAR HBA system [127]. Sixteen dual polarized HBAs form a 4×4

tile and 48 or 96 HBA tiles are connected on-site to digital processing hardware to operate as a station. Each station sends its beam formed output to a central processing facility where all station data streams are correlated to provide the basic data for (self-)calibration of and imaging with the full LOFAR array. Other systems like EMBRACE and MWA follow a similar approach. The hierarchical system design limits the number of signal paths at each level to ~ 100 thus keeping the data streams and processing requirements manageable.

The second reason for considering novel approaches to calibration is the wide field-of-view. The receiving elements of an aperture array station have a very broad beam. This allows us to steer the telescope electronically by means of changing the beam former delays or weights instead of mechanically. This reduces the amount of time spent on slewing to another source to just a few seconds, which opens many new observing possibilities such as interrupting regular observing schedule to study transient phenomena. The intrinsically wide FOV of aperture arrays also makes them attractive for the SKA at the lowest frequencies (≤ 1.5 GHz), since survey speed, which is regarded a key figure of merit for the SKA, is directly proportional to the FOV [16, 20, 55]. However, due to their large FOV, the individual receiving elements will detect signals from a multitude of sources all over the sky and cannot be focused on a single source for calibration. Calibration techniques based on a single point source (see, e.g., [12]) may therefore no longer be applicable, especially not for aperture array stations. Another complicating factor in large FOVs is that the propagation conditions may vary over the FOV due to ionospheric or tropospheric disturbances. If the array has a large physical extent or is insufficiently sparse to ignore mutual coupling, the gain variations over the FOV may even differ from one receiving element to another.

Finally, these new instruments will have a much higher sensitivity than current instruments and therefore have the ability to detect very weak signals. All signal processing algorithms applied to the new radio telescopes should thus be able to detect these very weak sources amidst all the known strong signals. This requires a dynamic range of a few million to one in observations close to the strongest sources in the sky. In practice, this requirement can be reformulated by stating that the signal processing algorithms should not unnecessarily raise the noise in the final data products. This can only be judged if one knows the best possible performance, which is the main motivation for the analysis in Chapters 4 and 6 of this thesis.

The calibration problems of radio telescopes can qualitatively be divided into the following four regimes depending on the size of the FOV of the individual receivers and the baselines between them [71]:

1. All antennas and all lines of sight sample the same propagation path. The array manifold is therefore not distorted. This is the regime with small FOVs and

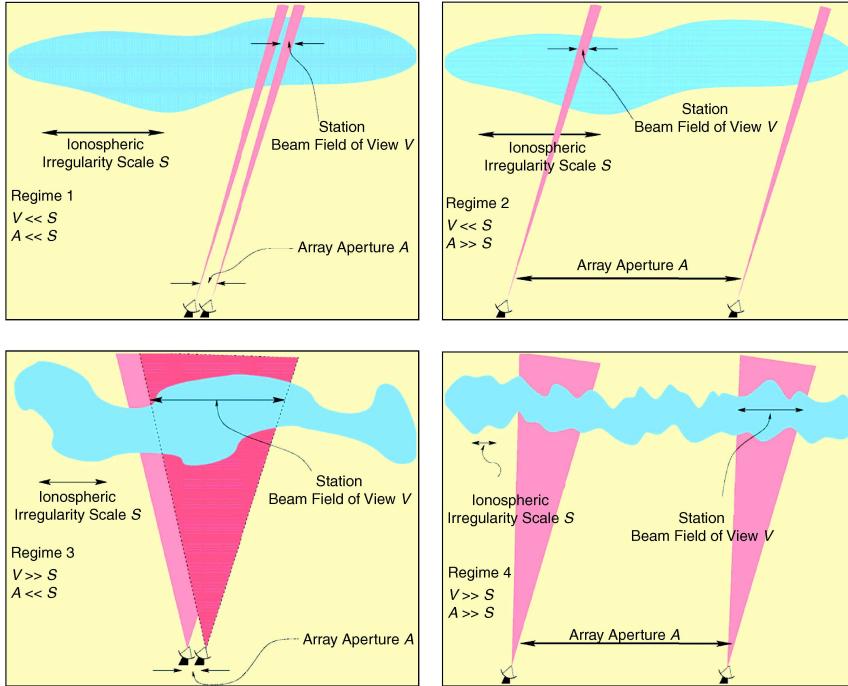


Figure 5.1: From top left to bottom right: graphical representation (after [71]) of regime 1 (small FOV, short baselines), regime 2 (small FOV, long baselines), regime 3 (large FOV, short baselines) and regime 4 (large FOV, long baselines).

short baselines as shown in the top left panel of Fig. 5.1.

2. If the FOV is small but the baselines between the receiving elements are long, the lines of sight towards a given source may experience different propagation conditions, but all lines of sight within the FOV of a single receiving element experience the same propagation path. In this regime, pictorially represented in the top right panel of Fig. 5.1, the different propagation conditions only cause antenna based gain effects.
3. If the FOV is large while the baselines are short, all lines of sight towards a single source go through the same propagation path, but there may be consid-

erable differences in the propagation paths towards distinct sources within the FOV. The environment thus imposes a direction dependent gain that is the same for all elements. This scenario is depicted in the bottom left panel of Fig. 5.1.

4. In large arrays consisting of elements with a large FOV, the lines of sight from the receiving elements towards each source within the FOV may encounter different propagation conditions. The required gain corrections therefore differ per source and per element as shown in the bottom right panel of Fig. 5.1.

The first two regimes can be handled under the self-calibration assumption used in radio astronomy, which states that errors are telescope based thereby reducing the estimation problem to a single direction independent gain factor per telescope [26]. The problem of estimating these direction independent gains based on a single calibrator source is treated in [12], while the multiple source case has been discussed in [149]. The problem of finding a complex gain per antenna per source, the fourth regime, is not tractable without further assumptions that allow to parameterize the behavior of the gains over space, time and/or frequency [120]. In this chapter, based on [164, 165], we will focus on the third regime, thus filling the gap in the available literature. This should allow us to calibrate closely packed groups of antennas such as a LOFAR station or a subarray of the MWA, which forms a valuable step towards calibration of the whole array. In Sec. 5.5 we will discuss a few adaptations to the basic algorithm presented in Sec. 5.3 that were required for the LOFAR station application.

Although our primary interests lie with aperture array radio telescopes, the problem is stated in general terms in Sec. 5.2, since the problem plays a key role in a range of applications ranging from underwater acoustics to antenna arrays. This explains the persistent interest in on-line calibration, autocalibration or self-calibration of sensor arrays [3, 36, 37, 89, 104, 120]. In most applications, the main driver for studies on array calibration is to improve the DOA estimation accuracy. Many studies in this field therefore try to solve for the DOAs and a number of array parameters. In [36, 37, 103, 133] self-calibration schemes are presented which solve for the direction independent gains and the sensor positions, i.e. the directional response of the sensors is assumed to be known. Other studies assume a controlled environment to calibrate the array by measuring the calibrator sources one at a time [79, 90] or exploit the array geometry, e.g. a uniform linear array (ULA) having a Toeplitz matrix as array covariance matrix [3, 68]. Weiss and Friedlander [134] have presented a technique for almost fully blind signal estimation. Their work, however, focuses on estimation and separation of source signals, not on characterizing the array itself. The problem at hand thus also forms an interesting addition to the literature available on sensor array calibration in general.

5.2 Problem statement

The algorithms developed in Sec. 5.3 and validated using Monte Carlo simulations in Sec. 5.4 are based on the data model given by (3.33), i.e. for our analysis we will initially adopt the array covariance model

$$\mathbf{R} = \mathbf{G}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^H\mathbf{G}^H + \boldsymbol{\Sigma}_n. \quad (5.1)$$

Calibration then means solving for \mathbf{G} , $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_n$ under the assumption that \mathbf{A} is known or parametrically known.

This data model is commonly used in papers on sensor array calibration (see e.g. [12, 39, 90, 149]). Flanagan and Bell [36, 37], Weiss and Friedlander [133] and See [103] effectively use the same model, but focus on position calibration of the array elements and are therefore more explicit on the form of \mathbf{A} . If the source positions and locations of the sensors within the array are known, an explicit formula for \mathbf{a}_q like (3.7) can be used to compute the spatial signature vectors. Based on (3.7) and introducing $\mathcal{L} = [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_Q]$, the spatial signature matrix \mathbf{A} is described by

$$\mathbf{A} = \frac{1}{\sqrt{P}} \exp \left(-j \frac{2\pi}{\lambda} \boldsymbol{\Xi} \mathcal{L} \right) \quad (5.2)$$

where the exponential function is applied element-wise to its argument. The time dependence in (3.7) has dropped out due to correlation. In the remainder of this chapter we will specialize to a planar array having $z_p = 0$ for convenience of presentation but without loss of generality.

From (5.1) we observe that \mathbf{G} and $\boldsymbol{\Sigma}$ share a common scalar factor. We may therefore impose the boundary condition $\sigma_1 = 1$. Furthermore, we will take the first element as phase reference, i.e. we will impose $\phi_1 = 0$, to resolve the ambiguity in the phase solution of \mathbf{G} . In Sec. 4.3 it was shown that $\sum_{p=1}^P \phi_p = 0$ is the optimal constraint for this problem. However, it turns out that the choice of constraint only matters when signals from different stations are combined. The demonstrations and examples in Ch. 2 were all done at station level, so this subtlety did not have impact on the results. The choice for the constraint $\phi_1 = 0$ also simplifies our analysis in combination with the constraints required to uniquely identify the apparent source powers and source locations. When solving for source locations, a single rotation of all DOA vectors can be compensated by the direction independent gain phase solution. We will therefore fix the position of the first source to solve this ambiguity.

In Sec. 5.3 we will address four related sensor array calibration problems based on the data model described in (5.1). These scenarios are summarized below where the parameter vectors adhering to the aforementioned boundary conditions are stated explicitly.

1. The sensor noise powers are the same for all elements, i.e. $\sigma_{n1} = \sigma_{n2} = \dots = \sigma_{nP}$, such that $\Sigma_n = \sigma_n \mathbf{I}$ where \mathbf{I} is the identity matrix. In this scenario, the parameter vector to be estimated is $\boldsymbol{\theta} = [\gamma^T, \phi_2, \dots, \phi_P, \sigma_2, \dots, \sigma_Q, \sigma_n]^T$.
2. The sensor noise powers are allowed to differ from one element to another, i.e. $\Sigma_n = \text{diag}(\boldsymbol{\sigma}_n)$. In this case, the parameter vector is $\boldsymbol{\theta} = [\gamma^T, \phi_2, \dots, \phi_P, \sigma_2, \dots, \sigma_Q, \sigma_{n1}, \dots, \sigma_{nP}]^T$.
3. $\Sigma_n = \sigma_n \mathbf{I}$ and $\mathbf{A} = \mathbf{A}(\mathcal{L})$, i.e. similar to the first scenario but with unknown source locations. In this case, $\boldsymbol{\theta} = [\gamma^T, \phi_2, \dots, \phi_P, \sigma_2, \dots, \sigma_Q, \sigma_n, \mathbf{l}_2^T, \dots, \mathbf{l}_Q^T]^T$.
4. $\Sigma_n = \text{diag}(\boldsymbol{\sigma}_n)$ and $\mathbf{A} = \mathbf{A}(\mathcal{L})$, giving $\boldsymbol{\theta} = [\gamma^T, \phi_2, \dots, \phi_P, \sigma_2, \dots, \sigma_Q, \sigma_{n1}, \dots, \sigma_{nP}, \mathbf{l}_2^T, \dots, \mathbf{l}_Q^T]^T$.

Note that all scenarios assume a diagonal noise matrix. From the underlying physical model described in Ch. 3 it is clear that this may not provide a realistic description of actual data. In Ch. 2 we have demonstrated that a non-diagonal noise covariance matrix may also be used to model extended emission, allowing us to do our calibration on a point source model. Therefore an extension with a non-diagonal noise covariance matrix will be discussed in Sec. 5.5.

5.3 Algorithm development

5.3.1 Generalized least squares formulation

An asymptotically efficient estimate of the model parameters in $\boldsymbol{\theta}$ can be obtained via the ML formulation. Since all signals are assumed to be i.i.d. Gaussian signals, the derivation is standard and the ML parameter estimates for N independent samples are obtained by minimizing the negative log-likelihood function [82]

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left(\ln |\mathbf{R}(\boldsymbol{\theta})| + \text{tr} \left(\mathbf{R}^{-1}(\boldsymbol{\theta}) \hat{\mathbf{R}} \right) \right) \quad (5.3)$$

where $\mathbf{R}(\boldsymbol{\theta})$ is the model covariance matrix as a function of $\boldsymbol{\theta}$ and $\hat{\mathbf{R}}$ is the sample covariance matrix $N^{-1} \mathbf{X} \mathbf{X}^H$.

It does not seem possible to solve this minimization problem in closed form. As discussed in [82] a weighted least squares covariance matching approach is known to lead to estimates that are, for a large number of samples, equivalent to ML estimates and are therefore asymptotically efficient and reach the CRB.

The least squares covariance model fitting problem can be defined as

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} \left\| \widehat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta}) \right\|_F^2. \quad (5.4)$$

Equivalently, we consider minimization of the cost function

$$\kappa(\boldsymbol{\theta}) = \left\| \widehat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta}) \right\|_F^2 = \mathbf{f}^H(\boldsymbol{\theta}) \mathbf{f}(\boldsymbol{\theta}). \quad (5.5)$$

The more general weighted least squares problem is obtained by introducing a weighting matrix \mathbf{W} and minimizing

$$\kappa_W(\boldsymbol{\theta}) = \mathbf{f}^H(\boldsymbol{\theta}) \mathbf{W} \mathbf{f}(\boldsymbol{\theta}). \quad (5.6)$$

The optimal weight is known to be the inverse of the asymptotic covariance of the residuals, $\mathcal{E}\{\mathbf{f}(\boldsymbol{\theta}_0) \mathbf{f}^H(\boldsymbol{\theta}_0)\}$, where $\boldsymbol{\theta}_0$ is the true value of the parameters [82]. The optimal weight for Gaussian sources is thus

$$\mathbf{W}_{opt} = (\overline{\mathbf{R}} \otimes \mathbf{R})^{-1} = \overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}. \quad (5.7)$$

The Kronecker structure of \mathbf{W}_{opt} allows us to introduce $\mathbf{W}_c = \mathbf{R}^{-1/2}$ and write the weighted least squares (WLS) cost function as

$$\kappa_W(\boldsymbol{\theta}) = \left\| \mathbf{W}_c (\widehat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta})) \mathbf{W}_c \right\|_F^2. \quad (5.8)$$

As mentioned earlier, this estimator is asymptotically unbiased and asymptotically efficient [82].

We propose to solve the least squares problems for the four scenarios defined in Sec. 5.2 by alternating between least squares solutions to subsets of parameters. The subsets are chosen such that the solution is either available in the literature or can be derived analytically. We will have four subsets of parameters: the complex direction independent gains of the receiving elements \mathbf{g} , the apparent source powers $\boldsymbol{\sigma}$, the receiver noise powers $\boldsymbol{\sigma}_n$ and the source locations \mathcal{L} . In the following subsections we will develop unweighted and weighted least squares solutions for these subsets before putting them together to form the alternating least squares (ALS) or weighted alternating least squares (WALS) method respectively.

5.3.2 Estimation of direction independent gains

The least squares problem to find the omnidirectional gains $\widehat{\mathbf{g}}$ based on the weighted cost function $\kappa_W(\boldsymbol{\theta})$ can be formulated as

$$\widehat{\mathbf{g}} = \operatorname{argmin}_{\mathbf{g}} \left\| (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \left(\operatorname{vec}(\widehat{\mathbf{R}} - \boldsymbol{\Sigma}_n) - \operatorname{diag}(\operatorname{vec}(\mathbf{R}_0)) (\overline{\mathbf{g}} \otimes \mathbf{g}) \right) \right\|_F^2 \quad (5.9)$$

where we have introduced $\mathbf{R}_0 = \mathbf{A}\Sigma\mathbf{A}^H$; in this subsection \mathbf{R}_0 is assumed to be known. This problem can be solved using standard techniques by regarding \mathbf{g} and $\overline{\mathbf{g}}$ as independent vector parameters and alternately solving for them until convergence. In this approach, the solution for \mathbf{g} is

$$\begin{aligned}\widehat{\mathbf{g}} &= \underset{\mathbf{g}}{\operatorname{argmin}} \left\| (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \left(\operatorname{vec}(\widehat{\mathbf{R}} - \Sigma_n) - ((\mathbf{R}_0 \mathbf{G}^H) \circ \mathbf{I}) \mathbf{g} \right) \right\|_F^2 \\ &= \left(\overline{\mathbf{R}}_0^H \overline{\mathbf{G}}^H \overline{\mathbf{R}}^{-1} \overline{\mathbf{G}} \mathbf{R}_0 \odot \mathbf{R}^{-1} \right)^{-1} \left(\overline{\mathbf{R}}^{-1} \overline{\mathbf{G}} \mathbf{R}_0 \odot \mathbf{R}^{-1} \right)^H \operatorname{vec}(\widehat{\mathbf{R}} - \Sigma_n).\end{aligned}\quad (5.10)$$

Since the result for $\overline{\mathbf{g}}$ is simply the complex conjugate of this relation, it is sufficient to apply (5.10) repeatedly until convergence. Although (W)ALS ensures that the value of the cost function decreases in each iteration, it does not guarantee convergence to the global minimum, especially if the initial estimate is poor. The number of iterations required for convergence also depends strongly on the vicinity of the initial estimate to the true value. We are thus interested in finding a good initial estimate for \mathbf{g} .

Fuhrmann [39] has proposed to use a suboptimal closed form solution to initialize the Newton iterations used to solve the ML cost function under the assumption that Σ_n is known (or actually no noise is present). A similar problem (with unknown Σ_n and with \mathbf{R}_0 a rank 1 matrix, i.e. a single calibrator source) was studied in [12], where a “column ratio method” was proposed. Below we generalize and improve that technique following and extending the discussion in [149]. All these techniques are suboptimal in general in the case of multiple calibrator sources, but they can be used to provide the starting point for iterative refinement.

The closed form solution suggested in [39] can be derived as follows. If the structure in $\mathbf{u} = \overline{\mathbf{g}} \otimes \mathbf{g}$ is neglected and Σ_n is known or negligible, we can solve for \mathbf{u} in the least squares sense by

$$\widehat{\mathbf{u}} = \operatorname{diag}(\operatorname{vec}(\mathbf{R}_0))^{-1} \operatorname{vec}(\widehat{\mathbf{R}} - \Sigma_n) \quad (5.11)$$

or equivalently

$$\widehat{\mathbf{g}} \widehat{\mathbf{g}}^H = (\widehat{\mathbf{R}} - \Sigma_n) \oslash \mathbf{R}_0. \quad (5.12)$$

Note that the weights cancel because we solve for one parameter for each entry of $\widehat{\mathbf{R}}$. Subsequently, use the structure of \mathbf{u} and assume that the estimate of $\mathbf{g}\mathbf{g}^H$ obtained in the first step has rank 1. An eigenvalue decomposition is used to extract $\widehat{\mathbf{g}}$ as the dominant eigenvector from $\widehat{\mathbf{u}}$. It is clear, however, that the noise on specific elements of $\widehat{\mathbf{u}}$ may be increased considerably if some entries of \mathbf{R}_0 have a small value. Also, the method is not applicable if Σ_n is unknown.

In [149] another approach is therefore suggested based on the observation that $g_i \bar{g}_k = R_{ik}/R_{0,ik}$ holds for all off-diagonal elements of \mathbf{R} , i.e. for $i \neq k$. This implies that

$$\frac{g_i}{g_j} = \frac{g_i \bar{g}_k}{g_j \bar{g}_k} = \frac{R_{ik}/R_{0,ik}}{R_{jk}/R_{0,jk}} = \frac{R_{ik}R_{0,jk}}{R_{0,ik}R_{jk}} \quad (k \neq i, j) \quad (5.13)$$

This relation is similar to the closure amplitude relation in astronomy [26, 85], which states that in an observation on a *single* point source or, more generally, in the case of a rank 1 model, the amplitude ratios are related as indicated by the second equality sign in (5.13).

Since the index k can be chosen freely as long as $k \neq i, j$, we can introduce $\mathbf{c}_{1,ij}$ being the column vector containing the values $R_{ik}R_{0,jk}$ and $\mathbf{c}_{2,ij}$ being the column vector containing the values $R_{0,ik}R_{jk}$ for all possible values of $k \neq i, j$. We can now write (5.13) in the more general form

$$\frac{g_i}{g_j} \mathbf{c}_{2,ij} = \mathbf{c}_{1,ij} \quad (5.14)$$

which has the well-known solution

$$\frac{g_i}{g_j} = \mathbf{c}_{2,ij}^\dagger \mathbf{c}_{1,ij}. \quad (5.15)$$

All possible gain ratios can be collected in a matrix \mathbf{M} with entries $M_{ij} = \mathbf{c}_{2,ij}^\dagger \mathbf{c}_{1,ij}$. Since the model for \mathbf{M} is $M_{ij} = g_i/g_j$, the matrix \mathbf{M} is expected to be of rank 1 and \mathbf{g} can be extracted from this matrix using an eigenvalue decomposition: let \mathbf{v}_1 be the eigenvector corresponding to the largest eigenvalue of \mathbf{M} , then $\mathbf{g} = \alpha \mathbf{v}_1$, where the scaling factor α needs to be determined separately since the quotient g_i/g_j is insensitive to modification by a constant scaling factor applied to all gains. This factor can be found by minimizing

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \left\| \alpha \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \bar{\alpha} - \hat{\mathbf{R}} \right\|_F^2 \quad (5.16)$$

for all off-diagonal elements. By introducing the vectors $\mathbf{r}_0 = \operatorname{vec}_-(\mathbf{G} \mathbf{R}_0 \mathbf{G}^H)$ and $\mathbf{r} = \operatorname{vec}_-(\hat{\mathbf{R}})$, where $\operatorname{vec}_-(\cdot)$ operates like the $\operatorname{vec}(\cdot)$ operator but leaves out the elements on the main diagonal of its argument, this cost function can be rewritten as

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \left\| |\alpha|^2 \mathbf{r}_0 - \mathbf{r} \right\|_F^2 \quad (5.17)$$

which has least squares solution $\hat{\alpha} = (\mathbf{r}_0^\dagger \mathbf{r})^{1/2}$.

Equation (5.15) extends the column ratio method in [12] to all elements of the matrix. The column ratio method was originally introduced to reconstruct the main diagonal of a rank 1 source model, which allows to estimate this rank 1 model using an eigenvalue decomposition without distortion of the results due to unknown receiver noise powers. In our case, this allows us to neglect the unknown receiver noise powers. The CRB analysis in Sec. 4.4 shows that it does not matter whether one simultaneously estimates the receiver noise powers and the omnidirectional complex gains exploiting all data, i.e. including the autocorrelations, or ignores the autocorrelations and only solves for the direction independent complex gains. Another advantage of this method is, that it is easily adapted to scenarios with a non-diagonal noise matrix, as is demonstrated in Sec. 5.5.

The Monte Carlo simulations presented in [149] suggest that the method outlined above provides a statistically efficient estimate of $\hat{\mathbf{g}}$; the pseudo-inverse ensures that the noise on the entries of \mathbf{M} does not dramatically increase due to small values in either $\hat{\mathbf{R}}$ or \mathbf{R}_0 . It also provides a modification of the initial estimate proposed by Fuhrmann [39] which only finds a near optimal solution to the least squares gain estimation problem without further optimization using e.g. the Newton algorithm. The proposed method will therefore generally save computational effort. We will use (5.15) in the simulations presented in Sec. 5.4 to demonstrate that this method gives statistically efficient results without requiring additional iterations using (5.10). Moreover, if three or more iterations are required to ensure convergence using (5.10), it also requires less computational effort as discussed in Sec. 5.3.7.

Although the pseudo-inverse in (5.15) ensures robustness against small entries in either $\hat{\mathbf{R}}$ or \mathbf{R}_0 , the fact that the method relies on gain ratios may lead to poor performance if there are small entries in \mathbf{g} , e.g. due to failing array elements. This risk can be mitigated by rewriting (5.14) to

$$\mathbf{c}_{2,ij}g_i = \mathbf{c}_{1,ij}g_j. \quad (5.18)$$

By defining

$$\begin{aligned} \mathbf{y}_i &= [\mathbf{c}_{2,i1}^T, \mathbf{c}_{2,i2}^T, \dots, \mathbf{c}_{2,iP}^T]^T \\ \mathbf{Y}_i &= \begin{bmatrix} \mathbf{c}_{1,i1} & & & \\ & \mathbf{c}_{1,i2} & & \\ & & \ddots & \\ & & & \mathbf{c}_{1,iP} \end{bmatrix} \\ \mathbf{C}_1 &= [\mathbf{Y}_1^T, \mathbf{Y}_2^T, \dots, \mathbf{Y}_P^T]^T \end{aligned}$$

$$\mathbf{C}_2 = \begin{bmatrix} \mathbf{y}_1 \\ & \mathbf{y}_2 \\ & & \ddots \\ & & & \mathbf{y}_P \end{bmatrix}$$

we can enumerate i and j and collect all relations in a single matrix equation

$$\mathbf{C}_2\mathbf{g} = \mathbf{C}_1\mathbf{g}. \quad (5.19)$$

This suggests that \mathbf{g} can be found by searching for the null space of $\mathbf{C}_2 - \mathbf{C}_1$, which has size $P(P-1)^2 \times P$. By substituting the details of $\mathbf{c}_{1,ij}$ and $\mathbf{c}_{2,ij}$ in $\mathbf{C}_2 - \mathbf{C}_1$, it is easily seen that \mathbf{g} indeed lies in the null space and that this null space is one dimensional. However, this is only true for noise free data. In the practical case of noisy data, there will not be a null space and \mathbf{g} will be in the noise subspace instead. Furthermore, finding \mathbf{g} will involve a singular value decomposition on a very large matrix which may be computationally prohibitive.

An alternative approach follows from recognizing that \mathbf{g} is obtained from the eigenvector corresponding to the eigenvalue 1 of

$$\begin{aligned} \mathbf{C}_2^\dagger \mathbf{C}_1 &= \left[\frac{\sum_k \bar{R}_{0,ik} \bar{R}_{jk} R_{0,jk} R_{ik}}{\sum_j \sum_k \bar{R}_{0,ik} \bar{R}_{jk} R_{0,ik} R_{jk}} \right] \\ &= \left[\frac{g_i \bar{g}_j \sum_k |g_k|^2 |R_{0,ik}|^2 |R_{0,jk}|^2}{\sum_j |g_j|^2 \sum_k |g_k|^2 |R_{0,ik}|^2 |R_{0,jk}|^2} \right], \end{aligned} \quad (5.20)$$

which is easily found by substitution of the definition of $\mathbf{c}_{1,ij}$ and $\mathbf{c}_{2,ij}$ and use of the Moore-Penrose left inverse. Equation (5.20) shows that \mathbf{g} is the eigenvector of $\mathbf{C}_2^\dagger \mathbf{C}_1$ corresponding to eigenvalue 1 in the noise free case and that all other eigenvalues are lower than 1, i.e. we can find the eigenvector associated with \mathbf{g} by finding the eigenvector corresponding to the largest eigenvalue. Simulations with completely randomized \mathbf{R}_0 and \mathbf{g} suggest that the other eigenvalues are not just lower than 1 but are considerably lower than 1. Since it is easily demonstrated that the noise on the actual data will lower the main eigenvalue, this is an important result; the contrast between the largest and the second largest eigenvalue determines the susceptibility of this method to noise on the data.

Note that these methods are still insensitive to a constant scaling factor applied to all gains. This ambiguity is solved by (5.17). Also note that all methods presented above enumerate over $i, j \neq k$ to avoid the diagonal entries of the array covariance matrix which are affected by the system noise. By imposing more stringent restrictions on i and j , these methods can also handle cases in which the noise covariance

matrix Σ_n is non-diagonal, e.g. due to correlated noise between specific pairs of receiving elements, but still has sufficient zero entries to allow the methods above to extract the minimal amount of information for every combination of g_i and g_j .

5.3.3 Source power estimation

Based on the unweighted cost function $\kappa(\theta)$, $\hat{\sigma}$ is found by

$$\begin{aligned}\hat{\sigma} &= \underset{\sigma}{\operatorname{argmin}} \left\| \widehat{\mathbf{R}} - \mathbf{G}\mathbf{A}\Sigma\mathbf{A}^H\mathbf{G}^H - \Sigma_n \right\|_F^2 \\ &= \underset{\sigma}{\operatorname{argmin}} \left\| \operatorname{vec} \left((\widehat{\mathbf{R}} - \Sigma_n) - (\mathbf{G}\mathbf{A}) \Sigma (\mathbf{G}\mathbf{A})^H \right) \right\|_F^2 \\ &= \underset{\sigma}{\operatorname{argmin}} \left\| \operatorname{vec} \left(\widehat{\mathbf{R}} - \Sigma_n \right) - (\overline{\mathbf{GA}}) \circ (\mathbf{GA}) \sigma \right\|_F^2 \\ &= (\overline{\mathbf{GA}} \circ \mathbf{GA})^\dagger \operatorname{vec} \left(\widehat{\mathbf{R}} - \Sigma_n \right)\end{aligned}\quad (5.21)$$

If weighting is applied, $\hat{\sigma}$ follows from

$$\begin{aligned}\hat{\sigma} &= \underset{\sigma}{\operatorname{argmin}} \left\| \mathbf{W}_c \left(\widehat{\mathbf{R}} - \Sigma_n \right) \mathbf{W}_c - \mathbf{W}_c \mathbf{G} \mathbf{A} \Sigma \mathbf{A}^H \mathbf{G}^H \mathbf{W}_c \right\|_F^2 \\ &= \underset{\sigma}{\operatorname{argmin}} \left\| (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \operatorname{vec} \left(\widehat{\mathbf{R}} - \Sigma_n \right) - (\overline{\mathbf{W}}_c \overline{\mathbf{GA}}) \circ (\mathbf{W}_c \mathbf{GA}) \sigma_n \right\|_F^2 \\ &= ((\overline{\mathbf{W}}_c \overline{\mathbf{GA}}) \circ (\mathbf{W}_c \mathbf{GA}))^\dagger (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \operatorname{vec} \left(\widehat{\mathbf{R}} - \Sigma_n \right).\end{aligned}\quad (5.22)$$

Using (C.10) and (C.11) to reduce the Kronecker and Khatri-Rao products and substituting $\mathbf{W}_c = \mathbf{R}^{-1/2}$ we obtain

$$\begin{aligned}\hat{\sigma} &= \left((\overline{\mathbf{A}}^H \overline{\mathbf{G}}^H \mathbf{R}^{-1} \overline{\mathbf{GA}}) \odot (\mathbf{A}^H \mathbf{G}^H \mathbf{R}^{-1} \mathbf{GA}) \right)^{-1} \times \\ &\quad \operatorname{vecdiag} \left(\mathbf{A}^H \mathbf{G}^H \mathbf{R}^{-1} (\widehat{\mathbf{R}} - \Sigma) \mathbf{R}^{-1} \mathbf{GA} \right)\end{aligned}\quad (5.23)$$

This result confirms the observation by Ottersten, Stoica and Roy [82] that although the derivation involves the square root of the array covariance matrix, the final result only depends on the array covariance matrix itself and its inverse.

5.3.4 Estimating receiver noise powers

If $\Sigma_n = \operatorname{diag}(\sigma_n)$ and no weighting is applied to the cost function, then $\hat{\sigma}_n$ is found by solving

$$\hat{\sigma}_n = \underset{\sigma_n}{\operatorname{argmin}} \left\| \widehat{\mathbf{R}} - \mathbf{G}\mathbf{A}\Sigma\mathbf{A}^H\mathbf{G}^H - \Sigma_n \right\|_F^2.\quad (5.24)$$

This estimation problem is the same as the sensor noise estimation problem treated in [12], so we just state the result, which is found to be

$$\hat{\boldsymbol{\sigma}}_n = \text{vecdiag} \left(\widehat{\mathbf{R}} - \mathbf{G} \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \mathbf{G}^H \right). \quad (5.25)$$

If $\boldsymbol{\Sigma}_n = \sigma_n \mathbf{I}$, a similar derivation gives

$$\hat{\boldsymbol{\sigma}}_n = \frac{1}{P} \text{tr} \left(\widehat{\mathbf{R}} - \mathbf{G} \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \mathbf{G}^H \right). \quad (5.26)$$

This result is just the average of the sensor noise estimates obtained when they are estimated individually.

If the weighted cost function $\kappa_W(\boldsymbol{\theta})$ is used, we can follow a derivation similar to the one that led us to the estimate for $\hat{\boldsymbol{\sigma}}$ in the previous section. We first note that

$$\begin{aligned} \hat{\boldsymbol{\sigma}}_n &= \underset{\boldsymbol{\sigma}_n}{\text{argmin}} \left\| \mathbf{W}_c \left(\widehat{\mathbf{R}} - \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \right) \mathbf{W}_c - \mathbf{W}_c \boldsymbol{\Sigma}_n \mathbf{W}_c \right\|_F^2 \\ &= \underset{\boldsymbol{\sigma}_n}{\text{argmin}} \left\| (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \text{vec} \left(\widehat{\mathbf{R}} - \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \right) - (\overline{\mathbf{W}}_c \circ \mathbf{W}_c) \boldsymbol{\sigma}_n \right\|_F \\ &= (\overline{\mathbf{W}}_c \circ \mathbf{W}_c)^\dagger (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \text{vec} \left(\widehat{\mathbf{R}} - \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \right). \end{aligned} \quad (5.27)$$

Applying Eqs. (C.10) and (C.11) once more and inserting $\mathbf{W}_c = \mathbf{R}^{-1/2}$, we get

$$\hat{\boldsymbol{\sigma}}_n = \left(\overline{\mathbf{R}}^{-1} \odot \mathbf{R}^{-1} \right)^{-1} \text{vecdiag} \left(\mathbf{R}^{-1} \left(\widehat{\mathbf{R}} - \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \right) \mathbf{R}^{-1} \right). \quad (5.28)$$

A similar derivation for $\boldsymbol{\Sigma}_n = \sigma_n \mathbf{I}$ gives

$$\hat{\boldsymbol{\sigma}}_n = \|\mathbf{R}^{-1}\|_F^{-2} \text{tr} \left(\mathbf{R}^{-1} \left(\widehat{\mathbf{R}} - \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \right) \mathbf{R}^{-1} \right). \quad (5.29)$$

The true value of the array covariance matrix \mathbf{R} is not known in practical situations. The measured array covariance matrix $\widehat{\mathbf{R}}$ is therefore generally used as estimate of \mathbf{R} . It can be shown that this conventional approach results in a noticeable bias in the estimate $\hat{\boldsymbol{\sigma}}_n$ for a finite number of samples N . For simplicity of the argument, we will prove this statement for the case where $\mathbf{R} = \sigma_n \mathbf{I}$, such that $\mathbf{G} \mathbf{R}_0 \mathbf{G}^H = \mathbf{0}$. Inserting this in (5.29) gives

$$\begin{aligned} \hat{\boldsymbol{\sigma}}_n &= \left\| \widehat{\mathbf{R}}^{-1} \right\|_F^{-2} \text{tr} \left(\widehat{\mathbf{R}}^{-1} \widehat{\mathbf{R}} \widehat{\mathbf{R}}^{-1} \right) \\ &= \left\| \widehat{\mathbf{R}}^{-1} \right\|_F^{-2} \text{tr} \left(\widehat{\mathbf{R}}^{-1} \right). \end{aligned} \quad (5.30)$$

Since $\mathbf{R}^{-1} = \sigma_n^{-1}\mathbf{I}$, we can write $\widehat{\mathbf{R}}^{-1} = \sigma_n^{-1}(\mathbf{I} + \mathbf{E})$ where $\sigma_n^{-1}\mathbf{E} = \widehat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}$ represents a specific realization of the noise on the elements of $\widehat{\mathbf{R}}^{-1}$. After substitution in (5.30) we obtain

$$\begin{aligned}\widehat{\sigma}_n &= \|\sigma_n^{-1}(\mathbf{I} + \mathbf{E})\|_F^{-2} \operatorname{tr}(\sigma_n^{-1}(\mathbf{I} + \mathbf{E})) \\ &= \sigma_n \operatorname{tr}((\mathbf{I} + \mathbf{E})(\mathbf{I} + \mathbf{E})^H)^{-1} \operatorname{tr}(\mathbf{I} + \mathbf{E}) \\ &= \sigma_n (\operatorname{tr}(\mathbf{I}) + \operatorname{tr}(\mathbf{E}) + \operatorname{tr}(\mathbf{E}^H) + \operatorname{tr}(\mathbf{E}\mathbf{E}^H))^{-1} (\operatorname{tr}(\mathbf{I}) + \operatorname{tr}(\mathbf{E})).\end{aligned}\quad (5.31)$$

Since \mathbf{E} represents the noise on the data, the expected value of $\operatorname{tr}(\mathbf{E})$ and $\operatorname{tr}(\mathbf{E}^H)$ is zero. This implies that

$$\begin{aligned}\widehat{\sigma}_n &\approx \sigma_n (\operatorname{tr}(\mathbf{I}) + \operatorname{tr}(\mathbf{E}\mathbf{E}^H))^{-1} \operatorname{tr}(\mathbf{I}) \\ &= \sigma_n \frac{\|\mathbf{I}\|_F^2}{\|\mathbf{I}\|_F^2 + \|\mathbf{E}\|_F^2}.\end{aligned}\quad (5.32)$$

This shows that the presence of noise systematically reduces the value of the estimate $\widehat{\sigma}_n$ by a bias $\|\mathbf{E}\|_F^2 / (\|\mathbf{I}\|_F^2 + \|\mathbf{E}\|_F^2)$. Since $\widehat{\mathbf{R}}^{-1}$ converges asymptotically in N to the true value \mathbf{R}^{-1} , \mathbf{E} converges to zero. This result therefore also shows that the estimate of σ_n asymptotically converges to the true value if the number of samples N approaches infinity.

However, this bias can be avoided by using the best available knowledge of the estimated parameters, i.e. by using

$$\mathbf{R} = \widehat{\boldsymbol{\Gamma}} \mathbf{A} \left(\widehat{\mathcal{L}} \right) \Sigma \mathbf{A}^H \left(\widehat{\mathcal{L}} \right) \widehat{\boldsymbol{\Gamma}}^H + \widehat{\boldsymbol{\Sigma}}_n \quad (5.33)$$

in (5.29). In the first iteration of an alternating least squares algorithm, initial estimates of the parameters are used. These values are replaced by increasingly accurate estimated values in consecutive iterations.

5.3.5 DOA estimation

The problem of estimating \mathcal{L} is that of estimating the direction of arrival of signals incident on a sensor array and has been studied extensively. MUSIC [101] and weighted subspace fitting (WSF) [124, 125] are well-known statistically efficient DOA estimation methods applicable to arbitrary sensor arrays. In either case, the eigenvalue decomposition of \mathbf{R} is interpreted in terms of a noise subspace and a signal subspace

(see Sec. 3.3), i.e.

$$\mathbf{R} = \sum_{p=1}^P \lambda_p \mathbf{v}_p \mathbf{v}_p^H = \mathbf{V}_s \boldsymbol{\Lambda}_s \mathbf{V}_s^H + \mathbf{V}_n \boldsymbol{\Lambda}_n \mathbf{V}_n^H \quad (5.34)$$

where $\lambda_1 > \dots > \lambda_Q > \lambda_{Q+1} \geq \dots \geq \lambda_P$. The noise eigenvalues of the whitened true array covariance matrix are all equal and the number of sources can be derived from the distribution of the eigenvalues of \mathbf{R} ; we will assume that the number of sources is known. In practical situations, in which only an estimate of the true array covariance matrix is available, methods like the exponential fitting test [91, 92] or information theoretic criteria [131] must be used to determine the number of signals.

It is quite straightforward to adapt the WSF method to our models. The data model assumed in [125] can be described as

$$\mathbf{R} = \tilde{\mathbf{A}}(\mathcal{L}) \boldsymbol{\Sigma} \tilde{\mathbf{A}}^H(\mathcal{L}) + \sigma_n^2 \mathbf{I} \quad (5.35)$$

where we have used $\tilde{\mathbf{A}}$ to avoid confusion with \mathbf{A} as commonly used in this thesis. To map our data model on the data model described by (5.35) we should whiten our array covariance matrix using

$$\mathbf{R}_w = \boldsymbol{\Sigma}_n^{-1/2} \mathbf{R} \boldsymbol{\Sigma}_n^{-1/2} = \boldsymbol{\Sigma}_n^{-1/2} \mathbf{G} \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \mathbf{G}^H \boldsymbol{\Sigma}_n^{-1/2} + \mathbf{I} \quad (5.36)$$

and then set $\tilde{\mathbf{A}} = \boldsymbol{\Sigma}_n^{-1/2} \mathbf{G} \mathbf{A}$. With these substitutions it is straightforward to implement the procedures described in [125].

5.3.6 (Weighted) alternating least squares

The ingredients of the preceding four subsections can be combined to formulate a (Weighted) Alternating Least Squares ((W)ALS) solution to the stated optimization problems. We start by introducing an algorithm handling the first two scenarios identified in Sec. 5.2. DOA estimation is then added in a straightforward way.

To estimate $\hat{\mathbf{g}}$, $\hat{\boldsymbol{\sigma}}$ and $\hat{\sigma}_n$ we propose the following (W)ALS algorithm:

1. *Initialization* Set the iteration counter $i = 1$ and initialize $\hat{\boldsymbol{\sigma}}^{[0]}$ based on knowledge of $\boldsymbol{\sigma}$ and the directional response of the sensors. For WALS, define the weight $\mathbf{W}_c = \hat{\mathbf{R}}^{-1/2}$. Initialize $\hat{\mathcal{L}}^{[0]}$ based on knowledge of the nominal positions of the calibrator sources.
2. *Estimate* $\hat{\mathbf{g}}^{[i]}$ by an eigenvalue decomposition of the matrix \mathbf{M} with its entries obtained by averaging (5.13) over all $k \neq i, j$ or by using (5.15) or by an eigenvalue decomposition of the matrix $\mathbf{C}_2^\dagger \mathbf{C}_1$ as given by (5.20) using \mathbf{A} and

$\hat{\sigma}^{[i-1]}$ as prior knowledge. In Sec. 5.4, I compare gain estimation using (5.13) and (5.15) demonstrating that the latter gives statistically efficient results. Note that neither gain calibration approach does require knowledge of the sensor noise powers σ_n and that the latter approach is advisable if some elements may have a very low gain.

3. Estimate $\hat{\sigma}_n^{[i]}$ using either (5.25) or (5.26) (resp. (5.28) or (5.29)) applying available knowledge of $\hat{g}^{[i]}$, $\hat{\sigma}_n^{[i-1]}$ and \mathbf{A} .
4. Estimate $\hat{\sigma}^{[i]}$ using (5.21) (resp. (5.23)) and knowledge of $\hat{g}^{[i]}$, $\hat{\sigma}_n^{[i]}$ and \mathbf{A} .
5. If the DOAs are inaccurately known (scenarios 3 and 4), estimate $\hat{\mathcal{L}}^{[i]}$ using WSF as described in Sec. 5.3.5 using knowledge of $\hat{g}^{[i]}$ and $\hat{\sigma}_n^{[i]}$ and initial estimate $\hat{\mathcal{L}}^{[i-1]}$.
6. *Check for convergence or stop criterion* If $|\theta^{[i-1]\dagger}\theta^{[i]} - 1| < \epsilon$ or $i > i_{max}$, stop, otherwise increase i by 1, update the weighting matrix $\mathbf{W}_c = \mathbf{R}^{-1/2}$ as proposed by (5.33) and continue with step 2.

The proposed criterion for convergence is based on a measure of the average relative error in all parameters and will work even if the parameter values differ by orders of magnitude. The extension with step 5 is referred to as the Extended ALS (xALS) or Extended WALS (xWALS) algorithm respectively.

An algorithm that alternatingly optimizes for distinct groups of parameters can be proven to converge if the value of the cost function decreases in each iteration. In [124, 125] it is demonstrated that WSF minimizes the least squares cost function w.r.t. the parameterization of \mathbf{A} , thus providing a partial solution to the least squares problem considered here. In step 2, \hat{g} is estimated using a method that only provides a near optimal solution to the least squares cost function. Its solution could, however, not be discerned from the true solution in Monte Carlo simulations, as demonstrated in Sec. 5.4. Optionally, step 2 could be augmented with one or two iterations of (5.10) to assure minimization of the least squares cost function. Alternatively, one could use the proposed estimate in the first iteration and use (5.10) in consecutive iterations in which a proper initial estimate is available from the previous iteration. The other parameters are estimated using well-known standard solutions for least squares estimation problems. The value of the cost function is therefore reduced in each step, thereby ensuring convergence.

ALS	$\hat{\mathbf{g}}$	$4\frac{1}{2}P^3 + PQ^2 + o(\dots)$
	$\hat{\sigma}$	$2P^2Q + 12PQ^2 + 9Q^3 + o(\dots)$
	$\hat{\sigma}_n$	0
	$\hat{\sigma}_n$	0
	total	$4\frac{1}{2}P^3 + 2P^2Q + 13PQ^2 + 9Q^3 + o(\dots)$
WALS	$\hat{\mathbf{g}}$	$5P^3 + PQ^2 + o(\dots)$
	$\hat{\mathbf{g}}$ (Eq. (5.10))	$N_{iter} \left(2\frac{2}{3}P^3 + 4P^2Q + 2Q^2P + o(\dots) \right)$
	$\hat{\sigma}$	$\frac{2}{3}P^3 + 2P^2Q + PQ^2 + \frac{2}{3}Q^3 + o(\dots)$
	$\hat{\sigma}_n$	$\frac{5}{3}P^3 + o(\dots)$
	$\hat{\sigma}_n$	$P^3 + o(\dots)$
	total	$7\frac{1}{3}P^3 + 2P^2Q + 2PQ^2 + \frac{2}{3}Q^3 + o(\dots) (\hat{\sigma}_n)$
		$6\frac{2}{3}P^3 + 2P^2Q + 2PQ^2 + \frac{2}{3}Q^3 + o(\dots) (\hat{\sigma}_n)$

Table 5.1: Numerical complexity of the ALS and WALS iterations expressed in terms of the number of receiving elements in the array P , the number of source signals Q and the number of iterations required for convergence when using (5.10) for direction independent gain estimation N_{iter} .

5.3.7 Computational complexity

Table 5.1 summarizes the numerical complexity of different stages of the ALS and WALS algorithms per iteration expressed in the number of complex multiplications. N_{iter} is the number of iterations required for convergence when using (5.10). In this table, the notation $o(\dots)$ is used to denote all the neglected lower order terms. The complexity of the omnidirectional complex gain estimate is dominated by the eigenvalue decomposition on \mathbf{M} , which requires approximately $4P^3$ multiplications assuming use of the divide and conquer method [168]. This step dominates the overall complexity as well, so it may be worthwhile to estimate the number of iterations of the power method [75] required to obtain sufficient accuracy, especially if P is large.

The pseudo-inverse in (5.21) can be implemented by a singular value decomposition, which is computationally more efficient than direct computation of the Moore-Penrose inverse [75]. Some intermediate results, such as $\mathbf{G}\mathbf{A}\Sigma\mathbf{A}^H\mathbf{G}^H$, are required more than once. In our calculation, we assume that these terms are computed only once and are stored for future use. Under this assumption, the number of complex multiplications required to compute the noise power reduces to zero. The number of array elements P will generally be considerably larger than the number of source

signals Q . The Q^3 -, PQ^2 - and P^2Q -terms may therefore be considered negligible compared to the P^3 -term in most cases. These results suggest that the better statistical performance of the WALS algorithm compared to the ALS algorithm, as demonstrated by the simulation results presented in the next section, comes with only a minor increase in complexity.

Both algorithms can be augmented with source position estimates using WSF. In [125] it is demonstrated that WSF can be implemented with $O(PQ^2)$ complex operations per Gauss-Newton type iteration of the proposed modified variable projection algorithm once the result from the eigenvalue decomposition is available. The cost of WSF is therefore dominated by the eigenvalue decomposition on $\widehat{\mathbf{R}}$ requiring about $4P^3$ complex multiplications. This makes the computational cost of source location estimation comparable to the cost of estimating the direction independent complex gains.

5.4 Algorithm validation

The methods proposed in the previous section were tested using Monte Carlo simulations and compared with the CRB. The results presented in this section were presented earlier in [149] and [165]. In the first simulation, I compare the variance of the direction independent gain estimates obtained from (5.13) and (5.15) with the CRB assuming that $\Sigma_n = \text{diag}(\sigma_n)$ is unknown. This is a simplified version of the second scenario defined in Sec. 5.2, which zooms in on the gain estimation problem. I will use this case to demonstrate that (5.15) seems to provide an asymptotically statistically efficient solution.

For this Monte Carlo simulation we will use the ITS antenna configuration introduced earlier in Sec. 4.3.5. We assume a sky model at 30 MHz consisting of the $Q = 10$ strongest astronomical sources which were visible in the sky above ITS on January 26, 2005 at midnight. The source locations and power ratios (see Table 5.2) were taken from the third Cambridge catalog of radio sources [8], which provides source fluxes at 178 MHz as known in the early 1960s. Although these values are not physically accurate at 30 MHz in the current epoch, they do provide typical power ratios that can be used to make a realistic source model. The total power of these sources was assumed to be 1% of the system noise power of the individual antennas. The data were averaged over $N = 512, 1024, \dots, 262144$ samples.

The results for the unweighted variant of the algorithm based on (5.13) are shown in Fig. 5.2. The plots show that the CRB not only varies with N , but also from element to element within the array. This is a geometrical effect. Interference between the source signals causes higher signal levels in some visibilities compared to others, while the noise power in all correlations is approximately the same. As a result, the

name	flux (Jy ^a)	az ^b (deg)	el (deg)
Cassiopeia A	11000	-21.3	26.0
Cygnus A	8100	9.0	4.1
Taurus A	1420	-111.1	42.3
Virgo A	970	110.7	30.1
Cygnus X (part)	410	4.9	3.2
Cygnus X (part)	230	4.9	8.3
3C157	210	-122.1	48.6
3C123	175	-91.5	39.5
3C416.2	160	2.1	4.6
Tycho's SNR ^c	134	-25.5	34.3

^a $1 \text{ Jy} = 10^{-26} \text{ Wm}^{-2}\text{Hz}^{-1}$

^b Starting from North through East

^c Super nova remnant

Table 5.2: Details of the discrete source model used in the first simulation. These values were taken from the third Cambridge catalog of radio sources [8], which provides source fluxes at 178 MHz.

net SNR is better for some visibilities than for others. This also explains why two elements show much larger variance on their gain and phase estimates.

The scaling factor α limits the gain amplitudes if the SNR of the calibrator sources is too low to obtain a meaningful dominant eigenvector from the eigenvalue decomposition on \mathbf{M} . This causes the variance of the gain estimate to be lower than the CRB for small N . The variance on the phase estimates is limited by the range of possible values (the interval 0 through 2π), as demonstrated by the asymptotic value reached for small N . Both gain and phase estimates do not attain the CRB. For large N , the variance on the estimates remains approximately five times as large as the minimum variance predicted by the CRB.

If the entries of \mathbf{M} are computed using (5.15), the algorithm does attain the CRB as shown in Fig. 5.3. The variance on the gain and phase estimates of the individual elements in the array is now far less sensitive to the variations of the SNR over different baselines, which causes the large variance values shown in Fig. 5.2. The plots indicate that the algorithm attains the CRB for $N \geq 32768$ samples when starting with an instantaneous SNR of 0.01. This translates to an effective SNR of 2 per

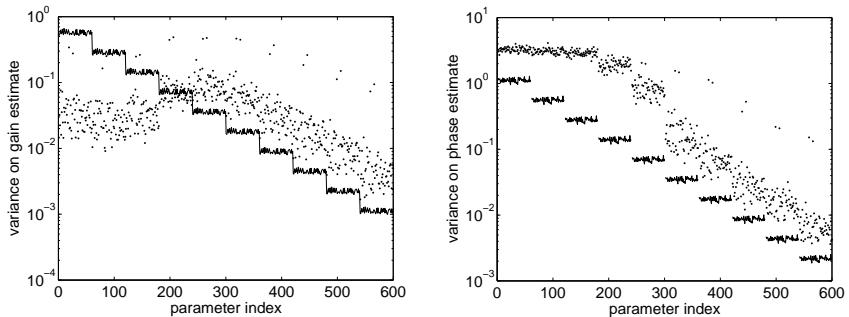


Figure 5.2: These plots show the variance on the gain (left) and phase (right) estimates in the Monte Carlo simulations (dots) and the CRB (solid line) for the unweighted variant of the direction independent complex gain estimation algorithm. In these simulations, the amplitude was estimated as a dimensionless factor and the phases were expressed in radians. The parameter vector for these plots was created by concatenating the parameter vectors of the individual Monte Carlo simulations, i.e. indices 1 through 60 represent the results for the first simulation ($N = 512$), 61 through 120 the results for the second simulation ($N = 1024$), etc. The variance on the phase of the first element is not plotted since this element is used as phase reference causing its variance to be zero by definition.

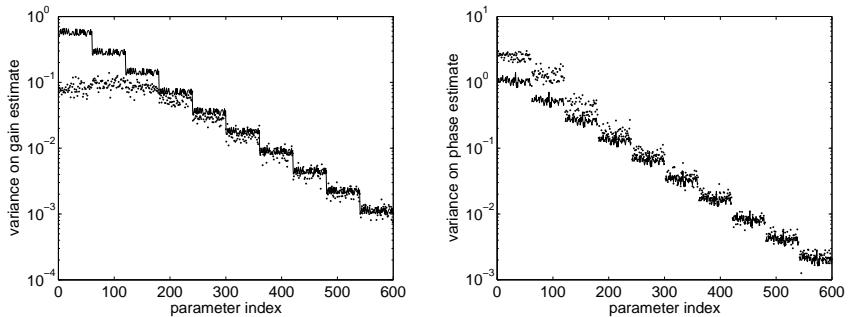


Figure 5.3: These plots show the variance on the gain (left) and phase (right) estimates in the Monte Carlo simulations (dots) and the CRB (solid line) for the weighted variant of the direction independent gain estimation method. The conventions used in these plots are the same as in Fig. 5.2.

#	l	m	σ_q
1	0.24651	-0.71637	1.00000
2	-0.34346	0.76883	0.88051
3	-0.13125	-0.31463	0.79079
4	-0.29941	-0.52339	0.74654
5	0.39290	0.58902	0.69781

Table 5.3: Source powers and source locations used in the second and third Monte Carlo simulation.

visibility at the moment the CRB is reached.

The results in Figs. 5.2 and 5.3 suggest that an eigenvalue decomposition on \mathbf{M} based on (5.13) does not provide a statistically efficient solution to the direction independent complex gain estimation problem while the result based on (5.15) does. Intuitively, this result was expected based on the choice of the weights applied while collecting the terms over $i, j \neq k$.

For the next two simulations, a five armed array was defined, each arm being an eight-element one wavelength spaced ULA. The first element of each arm formed an equally spaced circular array with half wavelength spacing between the elements. The source model is presented in Table 5.3. This source model was created using random number generators to demonstrate that the proposed approach indeed works for arbitrary source models.

In the second simulation, the direction independent gains, the source powers and the receiver noise powers were estimated assuming an array of identical elements, i.e. the parameter vector was defined as $\boldsymbol{\theta} = [\gamma^T, \phi_2, \dots, \phi_{40}, \sigma_2, \dots, \sigma_5, \sigma_n]^T$ corresponding to the first scenario formulated in Sec. 5.2. Data were generated assuming $\sigma_n = 10$ and $N = 10^5$. This value for σ_n implies an instantaneous SNR on the strongest source of only 0.1 per array element. Such a scenario, in which the instantaneous SNR per baseline is considerably lower than one is quite common in radio astronomy. Figure 5.4 shows the variance of the estimated parameters based on 1000 runs and compares these values with the CRB. As expected, the WALS method appears to be asymptotically statistically efficient while the ALS approach does not attain the CRB and shows clear outliers.

Figure 5.5 shows the variance found for a number of representative parameters as function of the number of samples and compares this with the corresponding CRBs. This plot confirms the conjecture stated in the previous paragraph that the WALS method is asymptotically statistically efficient. This is further corroborated by Fig.

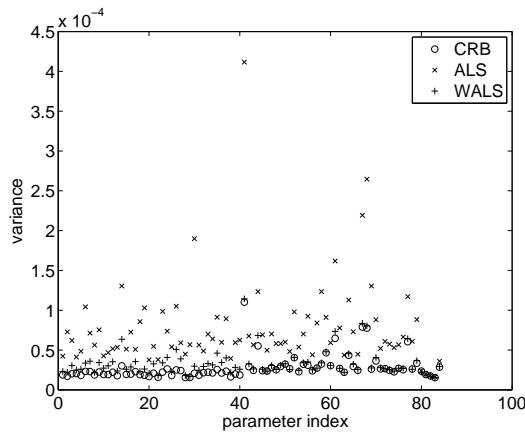


Figure 5.4: Variance of the parameter vector $\theta = [\gamma^T, \phi_2, \dots, \phi_{40}, \sigma_2, \dots, \sigma_5, \sigma_n]^T$ estimates obtained in Monte Carlo simulations for the weighted and unweighted versions of the alternating least squares algorithm and compared these results with the CRB.

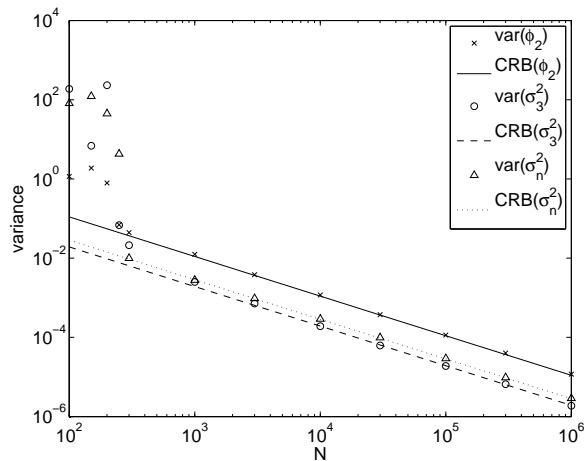


Figure 5.5: Variances on ϕ_2 , σ_3 and σ_n as function of the number of samples N , compared to the corresponding CRBs.

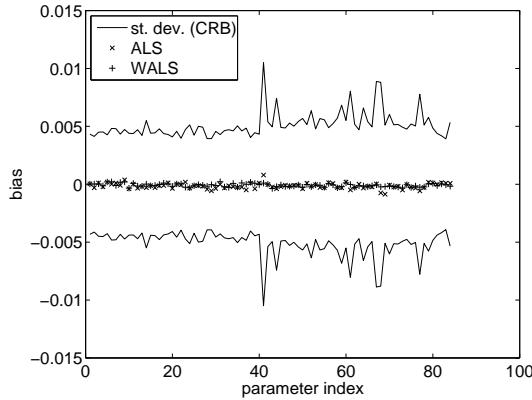


Figure 5.6: Bias on the estimated parameters for the weighted and unweighted version of the alternating least squares algorithm in the Monte Carlo simulations ($\boldsymbol{\theta} = [\gamma^T, \phi_2, \dots, \phi_{40}, \sigma_2, \dots, \sigma_q, \sigma_n]^T$). The bias is compared with the standard deviation on the estimates derived from the CRB.

5.6, which shows the bias found in the Monte Carlo simulations and compares this with the statistical error of 1σ for a single realization based on the CRB. This result indicates that both methods are unbiased for all parameters.

Figure 5.7 shows the difference of the parameter value at the end of each iteration and its final value obtained from one run of the Monte Carlo simulation for a representative case, i.e. similar results were found for other parameters and other runs. The standard deviation based on the CRB is plotted as reference to facilitate the interpretation of the vertical scale. The results indicate that only one or two iterations are needed to reach the CRB in this scenario. With regard to other scenarios, this result suggests exponential convergence, i.e. each iteration adds about one significant digit to the parameter estimate and it indicates that the WALS method converges more rapidly than the ALS method. In these simulations the stop criterion was $|\boldsymbol{\theta}^{[i-1]\dagger} \boldsymbol{\theta}^{[i]} - 1| < 10^{-10}$.

In the third simulation, the direction independent gains, the apparent source powers, the source locations and the receiver noise powers were estimated assuming an array of identical elements. This corresponds to the third scenario described in Sec. 5.2. The parameter vector to be estimated is thus $\boldsymbol{\theta} = [\gamma^T, \phi_2, \dots, \phi_{40}, \sigma_2, \dots, \sigma_5, \sigma_n, l_2, \dots, l_5, m_2, \dots, m_5]^T$. Data were generated assuming $\sigma_n = 10$ and $N = 1000$. This again implies an instantaneous SNR of only 0.1 per array element, but in this

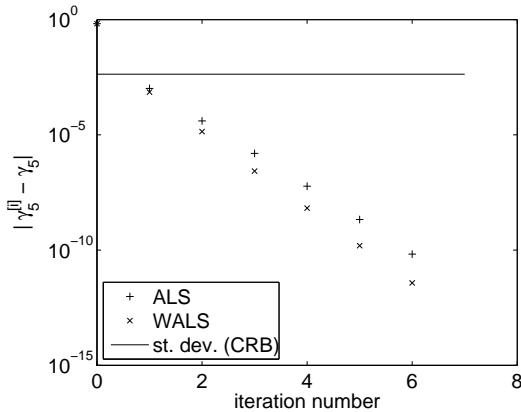


Figure 5.7: The difference between the value of $\hat{\gamma}_5$ at the end of each iteration and its final value is plotted versus the iteration number for ALS and WALS obtained in one of the Monte Carlo simulation runs. The behavior shown in this plot is representative for the results obtained for other parameters and other runs. The standard deviation based on the CRB is also shown to demonstrate that one but preferably two iterations are sufficient to obtain a sufficiently accurate results.

case the SNR per receiver is only 3.2 for the strongest source after integration. Figure 5.8 shows the variance on the estimated parameters based on 1000 runs and compares these results with the CRB. In these simulations, we only used the xWALS algorithm, since the previous simulation already demonstrated that the ALS method gives inferior results compared to the WALS approach. These results show that the variance on the estimated parameters is close to the CRB.

Figure 5.9 shows the bias on the estimated parameters found in these Monte Carlo simulations. These results indicate that the estimates are unbiased.

The convergence of a representative parameter is shown in Fig. 5.10. Again, one but preferably two iterations are sufficient to get accurate results. However, in this scenario the stop criterion $|\boldsymbol{\theta}^{[i-1]\dagger} \boldsymbol{\theta}^{[i]} - 1| < 10^{-10}$ is not reached and the algorithm stops because the maximum number of iterations (15) is reached. This indicates that the algorithm tries to interpret the noise as real signal. In the first iteration, the omnidirectional gains, apparent source powers and noise power are estimated first. These values are then used in the WSF algorithm to find the source locations. The source locations are then used to update the sky model, leading to an update of $\hat{\mathbf{g}}$, $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{\sigma}}_n$ and the cycle is complete. In the second simulation, the SNR after integration

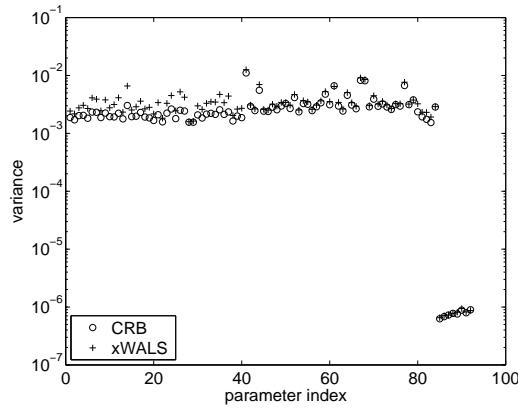


Figure 5.8: Variance on the parameter vector $\theta = [\gamma^T, \phi_2, \dots, \phi_{40}, \sigma_2, \dots, \sigma_5, \sigma_n, l_2, \dots, l_5, m_2 \dots, m_5]^T$ estimates obtained in Monte Carlo simulations for the xWALS method, compared with the CRB.

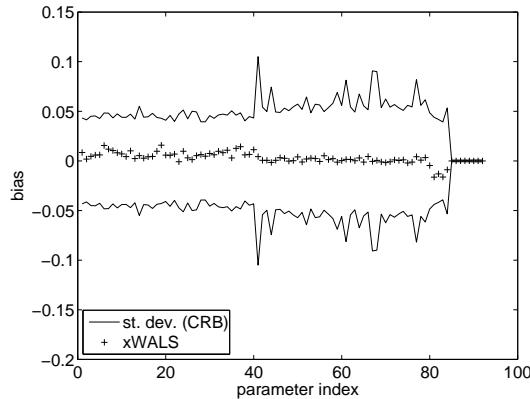


Figure 5.9: Bias on the estimated parameters ($\theta = [\gamma^T, \phi_2, \dots, \phi_{40}, \sigma_2, \dots, \sigma_5, \sigma_n, l_2, \dots, l_5, m_2 \dots, m_5]^T$) for the xWALS algorithm obtained in Monte Carlo simulations. The results are compared with the standard deviation on the estimated parameters derived from the CRB.

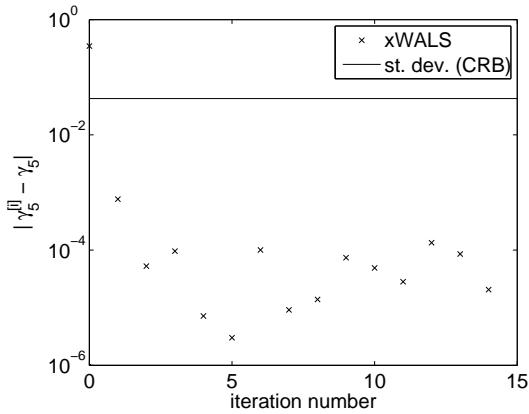


Figure 5.10: The difference between the value of $\hat{\gamma}_5$ at the end of each iteration and its final value is plotted versus the iteration number for one of the Monte Carlo simulation runs. The behavior shown in this plot is representative for the results obtained for other parameters and from other runs. The standard deviation based on the CRB is also shown to demonstrate that one but preferably two iterations are needed to obtain a sufficiently accurate result.

defined as $\sqrt{N}\sigma_p/\sigma_n$ was a factor 10 better than in this simulation in which the SNR per array element per source after integration is only 2.2 to 3.2.

5.5 Non-diagonal noise covariance matrix

5.5.1 Motivation

The prime goal of array calibration is to find the direction independent gains of the antennas and the direction dependent gain of the array towards each calibration source. The Monte Carlo simulations in the previous section suggest that the weighted alternating least squares approach proposed in Sec. 5.3 provides an asymptotically statistically efficient solution and the analysis in Sec. 5.3.7 shows that it is reasonably computationally efficient as well. However, the calibration of actual arrays is generally even more challenging than the scenarios defined in Sec. 5.2. For example, the calibration may be complicated by the fact that the covariances of signals from closely separated receiving elements are sensitive to noise coupling, as described by the physical model in Ch. 3. In the case of a radio astronomical phased array station

with short baselines, the point source model may not be an accurate description of the received signals due to bright emission from our Galaxy extending over the entire sky. Fortunately, this emission is spatially smooth, which implies that it is dominant on the short baselines. Since both effects affect only the short baselines, we propose to model these effects phenomenologically by a non-diagonal noise covariance matrix with unknown non-zero entries on all short baselines. The short baseline effects are thus absorbed in additive nuisance parameters. This approach was originally proposed in [164].

Direction finding problems for calibrated arrays in the presence of unknown correlated noise have been extensively studied in the 1990s. It was proven that the general problem is not tractable without imposing some appropriate constraints on the noise covariance matrix or exploiting differences in temporal characteristics between source and noise signals [110]. Radio astronomical signals generally behave like noise, thus temporal techniques (instrumental variables) are not applicable. Instead, we should rely on an appropriately constrained parameterized model of the noise covariance matrix. Starting with [11], a series of papers were published; see [42] for an overview. ML estimators for the source and instrument parameters under a generalized noise covariance parameterization are provided in [38, 42], whereas non-linear least squares estimators were studied in [38, 82, 132]. In either case, an analytic source and instrument parameter dependent solution is derived for the noise model parameters, which is substituted back into the cost function. This cost function then has to be minimized using a generalized solving technique, such as Newton iterations. This approach works well if the number of instrument and source parameters is small. For larger problems (in Ch. 2 we have seen an example with 96 antenna and source parameters and 764 noise covariance parameters), it is convenient to exploit suboptimal but closed-form analytic solutions, at least for initialization. We therefore propose to modify the WALS method outlined in Sec. 5.3 by replacing the estimate of the diagonal noise covariance matrix Σ_n by a parameterized estimate of a non-diagonal noise covariance matrix.

5.5.2 Modification of the WALS method

Least squares estimation of non-diagonal noise covariance matrix

As in [11] and subsequent papers, we will model the unknown noise covariance matrix as a linear sum of known matrices. The simplest set of matrices is provided by elementary matrices \mathbf{E}_{ij} , which are zero everywhere except for a '1' in entry (i, j) ,

$$\Sigma_n = \sum_{(i,j) \in \mathcal{S}} \sigma_{ij} \mathbf{E}_{ij}. \quad (5.37)$$

The set \mathcal{S} contains the index pairs of short baselines, including the autocorrelation entries (i, i) . In the Monte Carlo simulation presented later in this section and in the application to LOFAR data in Ch. 2, all baselines shorter than four wavelengths were regarded as short baselines. The unknown coefficients of σ_{ij} are the nuisance parameters. The noise covariance matrix is an array covariance matrix and should therefore be Hermitian. This knowledge can be exploited to obtain a real valued parameterization by splitting the parameter vector in the autocorrelations, the real parts of the crosscorrelations and the imaginary parts of the crosscorrelations, i.e.

$$\Sigma_n = \sum_{i=1}^P \sigma_i^{ac} \mathbf{E}_{ii} + \sum_{(i,j) \in \mathcal{S}} \sigma_{ij}^{Re} (\mathbf{E}_{ij} + \mathbf{E}_{ji}) + \sum_{(i,j) \in \mathcal{S}} \sigma_{ij}^{Im} \mathbf{j} (\mathbf{E}_{ij} - \mathbf{E}_{ji}), \quad (5.38)$$

where the set \mathcal{S} contains the index pairs of the non-zero entries above the main diagonal of Σ_n . All unique scalar real valued parameters required to describe Σ_n , i.e. σ_i^{ac} , σ_{ij}^{Re} and σ_{ij}^{Im} , can be stacked in a vector $\boldsymbol{\sigma}_n$. With a representation such as given by (5.37) or (5.38), the parameter vector $\boldsymbol{\sigma}_n$ can be related to Σ_n using a selection matrix \mathbf{I}_s such that $\text{vec}(\Sigma_n) = \mathbf{I}_s \boldsymbol{\sigma}_n$. By choosing the selection matrix such that it represents (5.38), we can ensure that the estimated Σ_n is Hermitian.

The weighted least squares estimate $\hat{\boldsymbol{\sigma}}_n$ can now be obtained by solving

$$\begin{aligned} \hat{\boldsymbol{\sigma}}_n &= \underset{\boldsymbol{\sigma}_n}{\text{argmin}} \left\| \mathbf{W}_c \left(\hat{\mathbf{R}} - \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \right) \mathbf{W}_c - \mathbf{W}_c \Sigma_n \mathbf{W}_c \right\|_F^2 \\ &= \underset{\boldsymbol{\sigma}_n}{\text{argmin}} \left\| (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \text{vec} \left(\hat{\mathbf{R}} - \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \right) - (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \mathbf{I}_s \boldsymbol{\sigma}_n \right\|_F^2, \end{aligned} \quad (5.39)$$

where $\mathbf{W}_c = \mathbf{R}^{-1/2}$ provides optimal weighting as discussed in Sec. 5.3.1. The solution to this problem is

$$\begin{aligned} \hat{\boldsymbol{\sigma}}_n &= ((\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \mathbf{I}_s)^\dagger (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \text{vec} \left(\hat{\mathbf{R}} - \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \right) \\ &= \left(\mathbf{I}_s^H (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c)^H (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \mathbf{I}_s \right)^{-1} \mathbf{I}_s^H (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c)^H \\ &\quad \times (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \text{vec} \left(\hat{\mathbf{R}} - \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \right) \\ &= \left(\mathbf{I}_s^H (\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) \mathbf{I}_s \right)^{-1} \mathbf{I}_s^H (\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) \text{vec} \left(\hat{\mathbf{R}} - \mathbf{G} \mathbf{R}_0 \mathbf{G}^H \right). \end{aligned} \quad (5.40)$$

Estimation of a diagonal noise covariance matrix as described by (5.28) is a special case of (5.40). In that case, $\mathbf{I}_s = \mathbf{I} \circ \mathbf{I}$ where \mathbf{I} is the $P \times P$ identity matrix.

This form of the selection matrix simplifies (5.40) to (5.28) after reduction of the Kronecker and Khatri-Rao products using (C.10) and (C.11).

The modified WALS algorithm

The WALS algorithm formulated in Sec. 5.3.6 can now be modified to

1. *Initialization* Set the iteration counter $i = 1$ and initialize $\hat{\sigma}^{[0]}$ based on knowledge of σ and the directional response of the sensors. Define the weight $\mathbf{W}_c = \mathbf{R}^{-1/2}$. Initialize $\hat{\mathcal{L}}^{[0]}$ based on knowledge of the nominal position of the calibrator sources.
2. *Estimate* $\hat{\mathbf{g}}^{[i]}$ by an eigenvalue decomposition of the matrix \mathbf{M} with its entries obtained from (5.15) or by an eigenvalue decomposition of the matrix $\mathbf{C}_2^\dagger \mathbf{C}_1$ given by (5.20) using \mathbf{A} and $\hat{\sigma}^{[i-1]}$ as prior knowledge. In the simulation presented later in this section and the LOFAR examples in Ch. 2, I used (5.15), which appears to be sufficiently robust and has the lowest computational cost. Note that neither approach requires knowledge of Σ_n .
3. *Estimate* $\hat{\sigma}_n$ using (5.40) applying available knowledge of $\hat{\mathbf{g}}^{[i]}$, $\hat{\sigma}^{[i-1]}$ and \mathbf{A} .
4. *Estimate* $\hat{\sigma}^{[i]}$ using (5.23) and knowledge of $\hat{\mathbf{g}}^{[i]}$, $\hat{\sigma}_n^{[i]}$ and \mathbf{A} .
5. If the DOAs are inaccurately known, *Estimate* $\hat{\mathcal{L}}^{[i]}$ using WSF as described in Sec. 5.3.5 using knowledge of $\hat{\mathbf{g}}^{[i]}$, $\hat{\sigma}_n^{[i]}$ and initial estimate $\hat{\mathcal{L}}^{[i-1]}$.
6. *Check for convergence or stop criterion* If $\left| \hat{\theta}^{[i-1]\dagger} \hat{\theta}^{[i]} - 1 \right| > \epsilon$ or $i > i_{max}$, stop, otherwise increase i by 1, update $\mathbf{W}_c = \mathbf{R}^{-1/2}$ as proposed by (5.33) and continue with step 2.

In some cases, for example in estimating the direction independent gains, it is possible to simply ignore the entries of the array covariance matrix corresponding to the non-zero entries of Σ_n . In Sec. 4.4 it was even shown that ignoring these entries instead of including them using nuisance parameters, does not change the CRB of the parameters of interest. In practice, however, it may be hard to develop an algorithm for other scenarios that ignores the corrupted entries in a statistically efficient way.

Improving the computational efficiency

The calculation of $\hat{\sigma}_n$ using (5.40) forms the most expensive part of the modified WALS algorithm in terms of CPU and memory usage due to the Kronecker products.

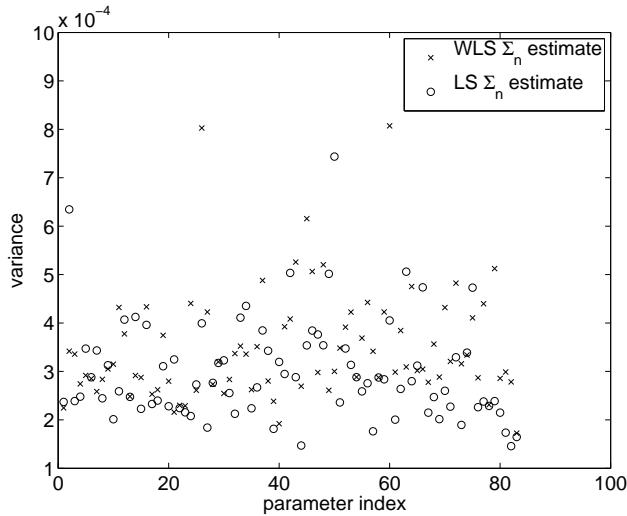


Figure 5.11: Comparison of the variance on the direction independent complex gain and source power estimates obtained in Monte Carlo simulations with weighted least squares estimate of $\hat{\sigma}_n$ and unweighted least squares estimate of $\hat{\sigma}_n$.

These Kronecker products can only be reduced to simpler Khatri-Rao or Hadamard products in a number of special cases, such as the case of a diagonal noise covariance matrix mentioned above. However, the parameterization of Σ_n chosen here implies that each entry of the array covariance matrix with a contribution from Σ_n is affected by a unique additive parameter. Intuitively, one would therefore expect that the weighting in (5.40) would not make much difference. Omitting this would reduce the CPU and memory requirements considerably.

This idea was therefore tested in Monte Carlo simulations. For these simulations, the five armed array described earlier and the source model presented in Table 5.3 were used. Figure 5.11 compares the variance on the estimates for the omnidirectional complex gains of the receiving elements and the source powers obtained after 100 runs of a Monte Carlo simulation with weighted least squares estimate of $\hat{\sigma}_n$ using (5.40) and the computationally more efficient unweighted least squares estimate of $\hat{\sigma}_n$. The results indicate that the variance on these estimates is the same in both cases within the accuracy provided by the simulations. We therefore conclude that it is viable to discard the weighting in (5.40). With this modification, all 860 free parameters in the experiment described in Ch. 2 could be extracted from the actual data using Matlab on a standard dual core 2.4 GHz CPU in only 0.4 seconds.

5.6 Summary of the main results

Calibration and imaging of fish-eye observations with arrays like LOFAR is very challenging due to the complex source structure, which includes multiple bright point sources and extended emission from, e.g. the Galactic plane. In this chapter, a gain calibration method was presented that can handle a multi-source sky model while ignoring short baselines. This method shows statistically efficient performance in Monte Carlo simulations, even with an SNR per receiving element per calibrator between 2.2 and 3.2, while having the same $O(P^3)$ numerical complexity as other asymptotically efficient algorithms in the literature that solve comparable calibration problems.

Using a weighted alternating least squares approach, this method can be extended to include estimation of apparent source powers, apparent source positions and a non-diagonal noise covariance matrix. These extensions do not increase the $O(P^3)$ numerical complexity, but more computations are required per step. It was also demonstrated that the computational burden can be reduced with almost no loss in direction independent gain estimation accuracy by using the unweighted least squares solution to find the noise covariance matrix. Even large problems, like the 860-parameter LOFAR station calibration problem described in Ch. 2, can be handled by a single 2.4 GHz CPU in only 0.4 seconds.

Chapter 6

Imaging

6.1 Introduction

The resolution of a telescope is determined by the ratio of observing wavelength and diameter of the telescope aperture. Since radio waves are the longest waves in the electromagnetic spectrum, the resolution achievable by radio telescopes consisting of a single parabolic reflector is fairly limited. At 21-cm wavelength, for example, even the largest single dish radio telescopes, the 300-m dish in Arecibo and the 100-m telescopes in Green Bank and Effelsberg, provide a resolution that is still an order of magnitude worse than the human eye. The successful construction of the first radio interferometer by Ryle and Vonberg [96, 97] in 1946 was therefore a major breakthrough, since it allowed radio astronomers to sample the spatial coherence function of the incoming signals in the aperture plane. The measured spatial coherencies can be stored in an array covariance matrix. In Sec. 3.5.1 it was demonstrated how the brightness distribution on the sky may be probed by a beam former and that the beam former output power can be described by $\mathbf{w}^H \mathbf{R} \mathbf{w}$. With $\mathbf{w} = \mathbf{a}$ using \mathbf{a} as defined by (3.7), the brightness distribution on the sky is simply the Fourier transform of these visibilities. This relation is known as the Van Cittert-Zernike theorem [88, 116].

Initially, the synthesized apertures were created by movable antennas until the principle of Earth rotation synthesis was introduced in 1962 [98]. The basic principle here is to exploit the rotation of the Earth to move the array w.r.t. to the sky to obtain denser sampling of the aperture with a fixed array. This led to the development and construction of the WSRT and the VLA during the 1960s and 1970s.

The 25-m dishes of the WSRT are placed on an East-West line. This ensures that during Earth rotation synthesis observations all visibility are measured in a single

plane. This allows us to produce the dirty image in a computationally efficient way by applying a fast Fourier transform (as opposed to the direct Fourier transform) after gridding the data onto a regular grid. If the visibilities do not lie on a single plane, as is the case with the samples produced by the VLA, they should be projected on a single plane taking into account the resulting delay to allow the use of the computationally efficient fast Fourier transform. This operation is called *w*-projection [28, 29].

Another attractive feature of the WSRT compared to the VLA is its equatorial mount. This ensures that the sensitivity pattern of the dishes does not rotate w.r.t. the sky during the observation. This implies that all points within the FOV are measured with the same direction dependent gain during the entire observation and therefore that the apparent source powers can be corrected to actual source powers by applying a single direction dependent gain correction at the end. When reducing VLA data, one has to apply these direction dependent gain corrections to every individual snapshot. These simplifications proved to be a vital advantage given the available computing power during the 1970s and 1980s. With the advent of todays computationally demanding imaging routines, both telescopes have been able to operate closer to their respective theoretical dynamic range limits.

The radio astronomical community is currently building a number of new instruments for the lowest frequencies of the radio spectrum, such as LOFAR and MWA. Imaging with these new telescopes will be even more challenging than imaging with the classical radio telescopes for several reasons:

- *Beam stability* These new instruments are large phased arrays in which the antennas are clustered in subarrays. Each subarray, or station, points its beam to the desired location, but these beams will vary much more than those of the classical mechanical dishes. This will cause more and less predictable variations in the beam pattern and hence require more complex direction dependent corrections.
- *Propagation conditions* The low frequency radio waves observed by these new instruments are strongly affected by ionospheric scintillation and refraction. These effects need to be compensated by complex direction dependent corrections.
- *Wide field-of-view* The next generation of radio telescopes is designed to have a wide FOV. The celestial sphere in the region of interest can therefore no longer be approximated by a plane without introducing unacceptable distortions. The current strategy is to split the image in facets that are sufficiently small to treat each of them as a plane, and produce a simultaneous solution for all facets. This approach is called facet imaging [87] and may become computationally

expensive for very large fields, such as the all-sky FOV provided by the LOFAR stations as demonstrated in Ch. 2.

- *Large data sets* With the advent of these new instruments, the number of elements in the array will increase by two orders of magnitude compared with the WSRT and the VLA. These instruments will therefore produce a few terabyte of data per day. The data should be analyzed in an amount of time comparable with the time used to acquire the data to ensure effective use of telescope time. The data should therefore be handled very efficiently due to restrictions in available computing power.

This combination of complex imaging problems and processing restrictions motivates the continuous efforts to improve and expand existing algorithms and to develop new methods. Much research is focused on finding clever short-cuts to reduce the amount of processing required, such as the aforementioned *w*-projection [28, 29] and facet imaging [87] algorithms and various variants of CLEAN [27, 30]. The validity and quality of these methods is generally assessed by practical experience. Attempts to do a rigorous analysis are done for some aspects and cases [60, 102, 113, 147], but rules of thumb are used most of the time. Section 6.2 presents the first comprehensive mathematical framework capable of describing the fundamental limits of radio astronomical imaging problems. This analysis was presented earlier in [157, 161]. Although this analysis is done for snapshot observations, i.e. variations in time and frequency are not considered, it is straightforward to extend the analysis to synthesis observations using a multi-measurement data model as described in Sec. 3.6 and in [64, 120, 123]. This point is illustrated in Sec. 6.3 by deriving closed form expressions for synthesis images based on minimization of the least squares cost function.

6.2 Fundamental imaging limits

The resolution of the final image, or map, is normally determined by the size and configuration of the array and the spatial taper function. Under the assumption that the sky is mainly empty, i.e., that the image contains only a few sources, maps with higher resolution than predicted by the array configuration (superresolution) can be made using CLEAN. The maximum entropy method (MEM) [78] imposes a similar constraint by aiming for a solution that is as featureless as possible. In the array processing literature, superresolution is achieved by high-resolution DOA estimation techniques such as MUSIC [101] and WSF [124, 125] or by high-resolution beam formers such as the MVDR beam former [23, 122] or the RCB [69, 126]. In all these approaches, the goal is to disentangle the spatial response of the array and the source structure, a process called deconvolution. In Sec. 6.2.1, imaging is formulated as a

least squares estimation problem, a form of model based imaging, and a closed form expression is obtained for its solution. This allows us to formulate the deconvolution problem as a matrix inversion problem. This provides a powerful tool to assess the tractability of the deconvolution problem and demonstrate the impact of the array configuration on the deconvolution limits and the redistribution of noise in the imaging and deconvolution process.

The next generation of radio telescopes should not only provide a high resolution, but also a large dynamic range. The dynamic range of an image is generally defined as the power ratio between the strongest and weakest meaningful features in the map. In practice, the limitations of an instrument are more conveniently described by the achievable noise floor in an imaging observation since the dynamic range strongly depends on the strength of the strongest source within the field-of-view (FOV) and because the noise may vary over the map. This noise floor is a combination of calibration errors, thermal noise and confusion noise. The term “effective noise” will be used to refer to the net result of these constituents. In Sec. 6.2.2 closed form expressions are derived that describe the components of effective noise in terms of the covariance of the image values, a concept that will be referred to as image covariance. The consequences of these expressions are illustrated with a few examples in Sec. 6.2.3. These examples suggest that the contribution of propagated calibration errors to the image covariance is considerably smaller than the contribution of thermal noise, even if the calibration is done on data with similar SNR. They also indicate that self-calibration causes a higher covariance between source power estimates than pure imaging does.

6.2.1 Imaging and deconvolution

Beam forming versus model based imaging

In Sec. 3.5.1 the scalar beam former was introduced. If this scalar beam former is used to probe the power coming from different directions, the source signal will be convolved with the array beam pattern as explained in Sec. 3.5.2. Imaging is another approach to map out the spatial distribution of source signals. This section briefly explains the difference between imaging and beam forming following the presentation in [161].

The imaging process transforms the covariances of the received signals, which are referred to as visibilities in radio astronomy, to an image of the source structure within the FOV of the receivers, while the beam former scans the FOV by forming a weighted superposition of the received signals for each distinct direction. In array processing terms this can be described as follows [64]. The operation $y(t) = \mathbf{w}^H \mathbf{x}(t)$ (see (3.71)) is called beam forming and can be regarded as a spatially matched filter. The weight vector $\mathbf{w} = \mathbf{a}$ described in Sec. 3.5.1 provides the

most basic beam former that assumes the presence of only a single source and only corrects the signal delays due to the array geometry. These weights can be adapted to correct the complex gain differences between the receiving elements \mathbf{g} derived from calibration measurements (see Ch. 5), nulling of interfering sources [123] and spatial tapering of the array [152].

The image value at \mathbf{l} is equal to the expected output power of the beam former when pointed into that direction and can be computed directly from the array covariance matrix $\widehat{\mathbf{R}}$ as

$$\widehat{i}(\mathbf{l}) = \mathcal{E} \left\{ |y(t)|^2 \right\} = \mathbf{w}^H \widehat{\mathbf{R}} \mathbf{w}. \quad (6.1)$$

For $\mathbf{w} = \mathbf{a}$, this is known as direct Fourier transform imaging. To create an image, \mathbf{w} is scanned over all relevant directions \mathbf{l} . The required weights can be stacked into a single matrix \mathbf{W} . Since $\mathbf{w}^H \widehat{\mathbf{R}} \mathbf{w} = (\overline{\mathbf{w}} \otimes \mathbf{w})^H \text{vec}(\widehat{\mathbf{R}})$, we can stack all image values in a single vector $\widehat{\mathbf{i}}$ and write

$$\widehat{\mathbf{i}}_{BF} = (\overline{\mathbf{W}} \circ \mathbf{W})^H \text{vec}(\widehat{\mathbf{R}}). \quad (6.2)$$

If we only want to image at the source locations, we have $\mathbf{W} = \mathbf{A}$. A typical model assumption is that there is a source present at every pixel location, in which case

$$\widehat{\mathbf{i}}_{BF} = (\overline{\mathbf{A}} \circ \mathbf{A})^H \text{vec}(\widehat{\mathbf{R}}). \quad (6.3)$$

This is the classical dirty image. Let us assume momentarily that $\mathbf{G} = \mathbf{I}$ and $\Sigma_n = \mathbf{0}$, i.e. that the gains of all receiver paths are equal and the data is noise free. Inserting the data model given in (3.33), or $\text{vec}(\mathbf{R}) = (\overline{\mathbf{A}} \circ \mathbf{A}) \boldsymbol{\sigma}$, into (6.3) gives

$$\begin{aligned} \mathbf{i}_{BF} &= \mathcal{E} \left\{ \widehat{\mathbf{i}}_{BF} \right\} = (\overline{\mathbf{A}} \circ \mathbf{A})^H (\overline{\mathbf{A}} \circ \mathbf{A}) \boldsymbol{\sigma} \\ &= (\overline{\mathbf{A}}^H \overline{\mathbf{A}} \odot \mathbf{A}^H \mathbf{A}) \boldsymbol{\sigma}. \end{aligned} \quad (6.4)$$

This shows that the dirty image is not equal to the true source structure $\boldsymbol{\sigma}$. To understand the physical meaning of the factor $\overline{\mathbf{A}}^H \overline{\mathbf{A}} \odot \mathbf{A}^H \mathbf{A}$, consider the product $\mathbf{a}_i^H \mathbf{a}_j$, where the indices i and j refer to the respective columns of \mathbf{A} . Using (3.7), this can be written explicitly as

$$\begin{aligned} \mathbf{a}_i^H \mathbf{a}_j &= \frac{1}{P} \exp \left(-j \frac{2\pi}{\lambda} \boldsymbol{\Xi} \mathbf{l}_i \right)^H \exp \left(-j \frac{2\pi}{\lambda} \boldsymbol{\Xi} \mathbf{l}_j \right) \\ &= \frac{1}{P} \sum_{p=1}^P \exp \left(j \frac{2\pi}{\lambda} \boldsymbol{\xi}_p^T (\mathbf{l}_i - \mathbf{l}_j) \right). \end{aligned} \quad (6.5)$$

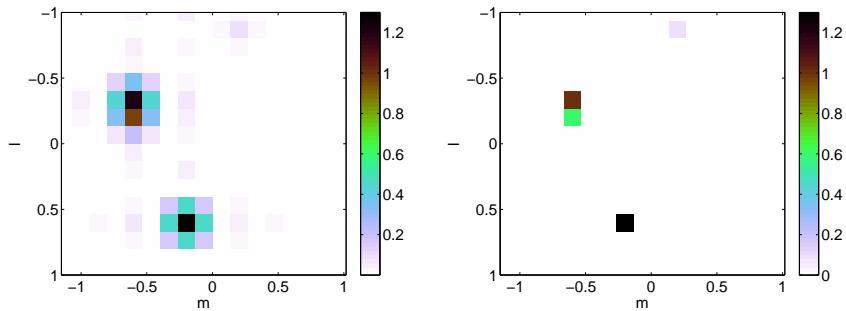


Figure 6.1: Image obtained by direct Fourier transform imaging without deconvolution as in (6.1) showing the sources and their side lobe patterns (left) and image obtained by model based imaging as in (6.12), which estimates the power emanating from each direction on the grid simultaneously, resulting in a deconvolved image showing only the sources without the array response.

This result matches (3.73). The physical interpretation of the inner product between the two spatial signature vectors is that it measures the sensitivity of the array to signals coming from direction \mathbf{l}_j while the array is steered towards \mathbf{l}_i and vice versa. The product $\mathbf{A}^H \mathbf{a}_j$ thus describes the array sensitivity for all directions of interest stacked in \mathcal{L} towards a source located at \mathbf{l}_j . It therefore provides the array voltage response or array voltage beam pattern centered around \mathbf{l}_j ,

$$\mathbf{b}_V(\mathbf{l}_j) = \mathbf{A}^H \mathbf{a}_j. \quad (6.6)$$

With \mathbf{a} defined as in (3.7), this shows that the voltage beam pattern is just the Fourier transform of the spatial sampling function provided by the array configuration and the weighting of the array elements. The corresponding power beam pattern can be calculated as

$$\mathbf{b}_P(\mathbf{l}_j) = \overline{\mathbf{b}}_V(\mathbf{l}_j) \odot \mathbf{b}_V(\mathbf{l}_j) = \overline{\mathbf{A}}^H \overline{\mathbf{a}}_j \odot \mathbf{A}^H \mathbf{a}_j. \quad (6.7)$$

The factor $\overline{\mathbf{A}}^H \overline{\mathbf{A}} \odot \mathbf{A}^H \mathbf{A}$ in (6.4) can thus be interpreted as a convolution by the Fourier transform of the spatial distribution of baseline vectors, which is known as the array beam pattern or dirty beam [116]. A more extensive discussion on the array beam pattern including some examples was presented in Sec. 3.5.2.

The convolution is illustrated in the left panel of Fig. 6.1. This image is the result of a simulated observation with an 8×8 half wavelength spaced, i.e. spatially Nyquist sampled, 2-D uniform rectangular array (URA). The grid of image values on the sky

is taken such that the first Nyquist zone is appropriately sampled. The source model contains four sources at (l, m) -grid points $(-0.33, -0.6)$, $(-0.2, -0.6)$, $(0.6, -0.2)$ and $(0.87, 0.2)$ respectively and $\sigma = [1, 0.6, 1.3, 0.1]^T$. This source model and array configuration will also be used in Sec. 6.2.3 unless stated otherwise. This map clearly shows the four sources (or three, if one regards the two sources on neighboring grid points as a single extended source) being convolved with the array beam pattern.

Following a model based approach, the deconvolution problem can be formulated as a maximum likelihood (ML) estimation problem, that should provide a statistically efficient estimate of the image parameters. Since all signals are assumed to be i.i.d. Gaussian signals, the derivation is standard and the ML estimates are obtained by minimizing the negative log-likelihood function [82], i.e.

$$\hat{\sigma} = \underset{\sigma}{\operatorname{argmin}} \left(\ln |\mathbf{R}(\sigma)| + \operatorname{tr} \left(\mathbf{R}^{-1}(\sigma) \hat{\mathbf{R}} \right) \right). \quad (6.8)$$

It does not seem possible to solve this minimization problem in closed form, but a weighted least squares covariance matching approach is known to lead to estimates, that are, for a large number of samples, equivalent to ML estimates and therefore asymptotically statistically efficient [82]. Inserting the data model given by (3.33), the problem can thus be reformulated as

$$\hat{\sigma} = \underset{\sigma}{\operatorname{argmin}} \left\| \mathbf{W}_c \left(\hat{\mathbf{R}} - \Sigma_n \right) \mathbf{W}_c - \mathbf{W}_c \mathbf{G} \mathbf{A} \Sigma \mathbf{A}^H \mathbf{G}^H \mathbf{W}_c \right\|_F^2. \quad (6.9)$$

This problem is identical to the source power estimation discussed in Sec. 5.3.3 and the solution is given by (5.22), i.e.

$$\hat{\sigma} = ((\overline{\mathbf{W}}_c \overline{\mathbf{G}} \mathbf{A}) \circ (\mathbf{W}_c \mathbf{G} \mathbf{A}))^\dagger (\overline{\mathbf{W}}_c \otimes \mathbf{W}_c) \operatorname{vec} \left(\hat{\mathbf{R}} - \Sigma_n \right). \quad (6.10)$$

Optimal weighting is provided by $\mathbf{W}_c = \mathbf{R}^{-1/2}$. Since radio astronomical sources are generally very weak with the strongest source in the field having an instantaneous SNR per receiving element in the order of 0.01, we may introduce the approximation $\mathbf{R} \approx \sigma_n \mathbf{I}$ for an array of identical elements for convenience of notation. This reduces (6.10) to

$$\hat{\sigma} = (\overline{\mathbf{G}} \mathbf{A} \circ \mathbf{G} \mathbf{A})^\dagger \operatorname{vec} \left(\hat{\mathbf{R}} - \Sigma_n \right). \quad (6.11)$$

One may argue that this requires knowledge of the source locations before doing the imaging. This is generally solved by simultaneously estimating the source locations and source powers. Although the CLEAN algorithm has not yet been fully analyzed, it can be regarded as an iterative procedure to do this [64]. It is instructive,

however, to use (6.11) for imaging by estimating the power on every image point (pixel), i.e., by assuming a data model with a source present at every pixel. We can simplify (6.11) by replacing the pseudo-inverse by the Moore-Penrose left inverse to obtain the image vector

$$\begin{aligned}\hat{\mathbf{i}} &= \left((\overline{\mathbf{GA}} \circ \mathbf{GA})^H (\overline{\mathbf{GA}} \circ \mathbf{GA}) \right)^{-1} (\overline{\mathbf{GA}} \circ \mathbf{GA})^H \text{vec}(\widehat{\mathbf{R}} - \Sigma_n) \\ &= \left(\overline{\mathbf{A}}^H \Gamma^2 \overline{\mathbf{A}} \odot \mathbf{A}^H \Gamma^2 \mathbf{A} \right)^{-1} (\overline{\mathbf{GA}} \circ \mathbf{GA}) \text{vec}(\widehat{\mathbf{R}} - \Sigma_n).\end{aligned}\quad (6.12)$$

The first factor in this equation represents the deconvolution operation. It is therefore convenient to introduce the deconvolution matrix $\mathbf{M} = \overline{\mathbf{A}}^H \Gamma^2 \overline{\mathbf{A}} \odot \mathbf{A}^H \Gamma^2 \mathbf{A} = |\mathbf{A}^H \Gamma^2 \mathbf{A}|^{\odot 2}$. This provides a powerful check on the sampling of the image plane. If the image plane is oversampled, i.e. if too many image points are defined, this matrix will be singular. This property demonstrates that high resolution imaging is only possible if a limited number of sources is present, i.e. if the number of sources is much smaller than the number of resolution elements within the FOV. The condition number of the deconvolution matrix, which provides a measure of the magnification of measurement noise, will be discussed in more detail later. This mostly empty field-of-view is commonly assumed in astronomical imaging and this assumption is one of the reasons why CLEAN and MEM work in practice. The right panel of Fig. 6.1 shows the image obtained by applying (6.12) to the 8×8 URA. Comparison with the image obtained using (6.1) clearly shows the effectiveness of the model based imaging approach in suppressing the array beam pattern.

Noise redistribution

If imaging is done without deconvolution by using (6.3), the thermal noise adds a constant value to all image values. This can be illustrated by assuming that $\mathcal{E}\{\widehat{\mathbf{R}}\} = \Sigma_n$, i.e. by assuming that the image is completely dominated by thermal noise. The expected value of the image then becomes

$$\begin{aligned}\mathbf{i}_{BF} &= (\overline{\mathbf{A}} \circ \mathbf{A})^H \text{vec}(\Sigma_n) \\ &= (\overline{\mathbf{A}} \odot \mathbf{A})^H \boldsymbol{\sigma}_n \\ &= \frac{\mathbf{1}^T \boldsymbol{\sigma}_n \mathbf{1}}{P^2} \mathbf{1}\end{aligned}\quad (6.13)$$

where we used the fact that all elements of \mathbf{A} have amplitude $1/P$. This equation describes an image where all values are equal to the average thermal noise per baseline.

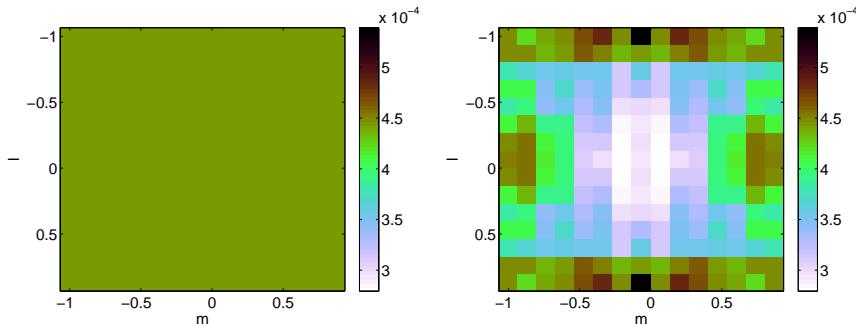


Figure 6.2: (a) Imaging with deconvolution using an 8×8 half wavelength spaced array for a Nyquist sampled image assuming $\mathbf{R} = 0.1\mathbf{I}$ (an empty field with only thermal noise). (b) Imaging result for a five armed array, each arm being an eight element half wavelength spaced ULA.

If the imaging process involves deconvolution, the result is described by (6.12). For simplicity, we will assume that we have an array of identical elements, so we can set $\mathbf{G} = \mathbf{I}$. Further, to illustrate the effect, we momentarily omit the correction by Σ_n in (6.12). In this case, the expected value of the image is

$$\begin{aligned}\mathbf{i} &= \left(\overline{\mathbf{A}}^H \overline{\mathbf{A}} \odot \mathbf{A}^H \mathbf{A} \right)^{-1} (\overline{\mathbf{A}} \odot \mathbf{A})^H \text{vec}(\Sigma_n) \\ &= \left(|\mathbf{A}^H \mathbf{A}|^{\odot 2} \right)^{-1} (\overline{\mathbf{A}} \odot \mathbf{A})^H \boldsymbol{\sigma}_n \\ &= \left(|\mathbf{A}^H \mathbf{A}|^{\odot 2} \right)^{-1} \frac{\mathbf{1}^T \boldsymbol{\sigma}_n}{P^2} \mathbf{1}.\end{aligned}\quad (6.14)$$

In this case, the homogeneity of the thermal noise distribution over the map depends on the row sums of $\left(|\mathbf{A}^H \mathbf{A}|^{\odot 2} \right)^{-1}$ being constant. If this is true, the noise distribution in the model based image obtained from (6.12) is analogous to that in the beam formed image obtained using (6.3). The situation in which the columns of \mathbf{A} are orthonormal forms a special case in which this holds.

Otherwise the structure is more complex. This is illustrated in Fig. 6.2 which compares the noise distribution over the image of the 8×8 -element URA by assuming $\mathbf{R} = 0.1\mathbf{I}$ with the corresponding image for a five armed array, each arm being an eight element half wavelength spaced uniform linear array (ULA). The impact of the redistribution of noise can be reduced by estimating the receiver noise powers and subtracting these estimates from the array covariance matrix as described by

(6.12). In most astronomical imaging algorithms, the autocorrelations are generally ignored completely thus effectively introducing a small negative system noise since the autocorrelations represent the power sum of the source signals and the noise.

Deconvolution matrix condition number

The deconvolution matrix \mathbf{M} not only causes a redistribution of noise over the map, but also determines whether the deconvolution is a well conditioned problem. If the deconvolution matrix is not invertible, the problem is ill-posed and additional constraints are required to obtain a unique solution. Different choices for these constraints or even the rigor with which they are applied, lead to different imaging results for CLEAN and MEM based on the same data. This may even lead to different interpretations of the final maps in some cases [78]. These problems arise due to over-interpretation of the data by allowing for more image points (parameters) than can be justified by the data. In such situations, the condition number of the deconvolution matrix will be infinitely large. Even if the deconvolution matrix is invertible, its condition number may be unacceptably high in view of the SNR of the data: the condition number is a measure for the magnification of measurement noise [41]. The condition number thus provides a powerful diagnostic tool to assess the feasibility of the deconvolution problem at hand.

It is instructive to analyze a half wavelength spaced 1-D ULA with identical elements, i.e. with $\mathbf{G} = \mathbf{I}$, sampling the sky on a regular grid. In this case, \mathbf{A} represents a Fourier transform mapping the spatial frequencies of the source structure on the spatial samples describing the electromagnetic field over the array aperture. As demonstrated in the previous section, these spatial frequencies will be convolved with the Fourier transform of the array aperture taper or voltage beam pattern during the imaging process. The voltage beam pattern for $\mathbf{l} = \mathbf{0}$ is easily calculated:

$$\mathbf{b}_V(\mathbf{0}) = \mathbf{A}^H \mathbf{a}(\mathbf{0}) = \mathcal{FT} \left(\begin{bmatrix} \mathbf{1}_P \\ \mathbf{0}_{Q-P} \end{bmatrix} \right). \quad (6.15)$$

Here, \mathcal{FT} denotes the Fourier transform and Q is the total number of image points. The corresponding power beam pattern is

$$\begin{aligned} \mathbf{b}_P(\mathbf{0}) &= \overline{\mathbf{b}}_V(\mathbf{0}) \odot \mathbf{b}_V(\mathbf{0}) \\ &= \mathcal{FT} \left(\begin{bmatrix} \mathbf{1}_P \\ \mathbf{0}_{Q-P} \end{bmatrix} \circledast \begin{bmatrix} \mathbf{1}_P \\ \mathbf{0}_{Q-P} \end{bmatrix} \right), \end{aligned} \quad (6.16)$$

where \circledast denotes the circular convolution of two vectors, i.e. for vectors of length N , $[\mathbf{a} \circledast \mathbf{b}]_m = \sum_{n=0}^{N-1} \mathbf{a}_n \mathbf{b}_{m-n \bmod N}$. If the columns of \mathbf{A} are ordered such that they

describe the array response vector for regularly spaced DOAs starting with $\mathbf{l}_1 = \mathbf{0}$, it is easily seen that

$$\mathbf{M} = \overline{\mathbf{A}}^H \overline{\mathbf{A}} \odot \mathbf{A}^H \mathbf{A} = \text{circulant}(\mathbf{b}_P(\mathbf{0})), \quad (6.17)$$

i.e. that the deconvolution matrix for a 1-D ULA equidistantly sampling the image plane is a circulant matrix. Since \mathbf{M} is a circulant matrix, its eigenvalues $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_Q]^T$ are given by the Fourier transform of $\mathbf{b}_P(\mathbf{0})$ [75], or

$$\begin{aligned} \boldsymbol{\lambda} &= \mathcal{FT}(\mathbf{b}_P(\mathbf{0})) \\ &= \mathcal{FT}\left(\mathcal{FT}\left(\begin{bmatrix} \mathbf{1}_P \\ \mathbf{0}_{Q-P} \end{bmatrix} \circledast \begin{bmatrix} \mathbf{1}_P \\ \mathbf{0}_{Q-P} \end{bmatrix}\right)\right) \\ &= \begin{bmatrix} \mathbf{1}_P \\ \mathbf{0}_{Q-P} \end{bmatrix} \circledast \begin{bmatrix} \mathbf{1}_P \\ \mathbf{0}_{Q-P} \end{bmatrix} \end{aligned} \quad (6.18)$$

since $\mathcal{FT}(\cdot) = \mathcal{FT}^{-1}(\cdot)$ for real symmetric functions.

For Hermitian matrices, the condition number κ is given by the ratio of the largest and smallest eigenvalue, i.e. $\kappa = \lambda_{\max}/\lambda_{\min}$ [75]. If the image plane is Nyquist sampled, $Q = 2P - 1$ and

$$\boldsymbol{\lambda} = [1, 2, \dots, P-1, P, P-1, \dots, 2, 1]^T. \quad (6.19)$$

In this case, the condition number of \mathbf{M} is

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{P}{1} = P, \quad (6.20)$$

thus \mathbf{M} is invertible. The deconvolution problem is therefore well-posed and has a unique solution.

If the image plane is undersampled symmetrically around $\mathbf{l} = \mathbf{0}$ with $Q < 2P - 1$ samples, then

$$\boldsymbol{\lambda} = \left[P - \frac{Q-1}{2}, \dots, P-1, P, P-1, \dots, P - \frac{Q-1}{2} \right]^T \quad (6.21)$$

and $\kappa = 2P/(2P - (Q - 1))$. The deconvolution problem in itself is thus well-posed and has a unique solution. However, we know from Fourier theory that aliasing effects may occur due to undersampling.

If the image plane is oversampled symmetrically around $\mathbf{l} = \mathbf{0}$ with $Q > 2P - 1$ samples, then

$$\boldsymbol{\lambda} = [\dots, 0, 1, 2, \dots, P-1, P, P-1, \dots, 2, 1, 0, \dots]^T \quad (6.22)$$

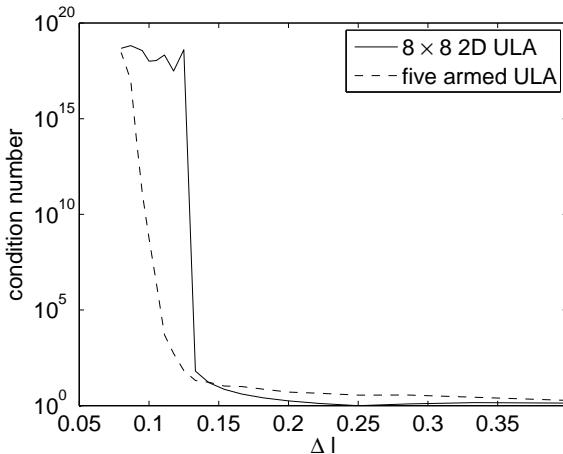


Figure 6.3: This plot shows the condition number of the deconvolution matrix as function of the image resolution for the 8×8 half wavelength spaced URA and the five armed array with each arm being an eight element half wavelength spaced ULA.

and $\kappa = \infty$. In this case, the deconvolution problem is ill-posed and thus not solvable without introducing additional boundary conditions to constrain the problem.

This analysis shows that, for a 1-D ULA, the condition number slowly increases up to Nyquist sampling of the image plane and then jumps to infinity. Since a URA is just the 2-D analog of a 1-D ULA, this behavior is also expected for the 8×8 URA introduced earlier. This conjecture is confirmed in Fig. 6.3 which shows the condition number of the deconvolution matrix as function of image resolution. This figure also shows the corresponding result for the five armed array introduced earlier to demonstrate the impact of less regular and sparser sampling of the array aperture. Although the array diameter is nearly twice as large, it does not provide twice the resolution due to sparser sampling of the aperture plane. This plot also demonstrates that a less regular array also has a less strict cut-off: there is a gradual transition of the condition number from small values to infinity. For array processing problems, this means that the user should decide which value of the condition number (or noise enhancement) is still acceptable.

Regularization is commonly used to avoid uninvertibility of matrices. In radio astronomical imaging where most sources have a low SNR, this would lead to imperfect deconvolution causing the weakest sources in the field to be drowned in the imperfectly removed array response pattern of the strongest sources. However, sev-

eral forms of implicit regularization have been studied to handle special cases, like strong interference, with some impressive results [117].

6.2.2 Effective noise

Equation (6.12) shows that calibration and imaging problems are strongly coupled. Knowledge of the instrumental parameters is required to obtain the proper image. People have approached this problem in two ways. In the first approach, calibration and imaging are treated as separate steps, i.e. the instrument parameters are estimated first from a calibration measurement and consecutively applied to the actual measurement data. The second approach is self-calibration, which regards the estimation of instrumental and image parameters as a single parameter estimation problem [26, 36, 37, 85, 120].

In either case, the achievable dynamic range is limited by the combination of estimation errors, thermal noise and confusion noise. Together, they determine the effective noise in the image which need not be homogeneous over the field of interest. In this section, a number of analytic expressions are derived that describe these contributions in terms of the data model presented in Ch. 3. The implications will be discussed in Sec. 6.2.3.

Noise in self-calibrated images

In self-calibration, the instrumental and image parameters are estimated simultaneously. Self-calibration based on the data model described by (3.33) can thus be described as simultaneous estimation of the direction independent complex gains, the apparent source powers, the source locations and the receiver noise powers, i.e. estimation of the parameter vector

$$\boldsymbol{\theta} = [\gamma_1, \dots, \gamma_P, \phi_2, \dots, \phi_P, \sigma_2, \dots, \sigma_Q, \sigma_{n1}, \dots, \sigma_{nP}, l_2, \dots, l_Q, m_2, \dots, m_Q]^T. \quad (6.23)$$

This corresponds to the fourth scenario described in Sec. 5.2. As before, ϕ_1 , σ_1 , l_1 and m_1 are omitted because they are set to constants to ensure identifiability of the problem. The calibration problem can be formulated as a least squares estimation problem and solved using the WALS method introduced in Sec. 5.3.

The minimum variance of an unbiased estimator is given by the CRB. Closed form expressions for the CRB are presented in Sec. 4.2 including a discussion on their implications. These expressions allow us to compute the CRB for the self-calibration problem. The variance of the estimated image values, i.e. the noise on the image values due to estimation inaccuracy, is given by the diagonal of the block $\mathbf{C}_{\sigma\sigma}$ of

this matrix, following the partitioning of $\boldsymbol{\theta}$. In general, $\mathbf{C}_{\boldsymbol{\sigma}\boldsymbol{\sigma}}$ is not a diagonal matrix. The non-diagonal entries in this submatrix describe the covariance of the noise on the pixels, i.e. the correlation between the noise on distinct pixels. This is associated with false structures.

Propagation of calibration errors

If the instrumental parameters are extracted from separate calibration data, the minimum variance on these estimated values is given by the CRB on the instrumental parameters in the calibration experiment, $\mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ where $\boldsymbol{\theta} = [\gamma_1, \dots, \gamma_P, \phi_2, \dots, \phi_P, \sigma_{n1}, \dots, \sigma_{nP}]^T$. This is a simplified version of the second scenario described in Sec. 5.2 and is easily handled by the WALS algorithm described in Sec. 5.3 skipping the apparent source power estimation. The CRB follows from the results presented in Sec. 4.2, assuming that the calibration measurement adheres to the same data model. The propagation of the calibration errors to the image is described by

$$\text{cov}(\mathbf{i}) = \left(\frac{\partial \mathbf{i}}{\partial \boldsymbol{\theta}^T} \right) \mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}} \left(\frac{\partial \mathbf{i}}{\partial \boldsymbol{\theta}^T} \right)^T. \quad (6.24)$$

We thus need to find $\partial \mathbf{i} / \partial \boldsymbol{\gamma}^T$, $\partial \mathbf{i} / \partial \boldsymbol{\phi}^T$ and $\partial \mathbf{i} / \partial \boldsymbol{\sigma}_n^T$. The derivative of the image values with respect to γ_k is given by

$$\frac{\partial \mathbf{i}}{\partial \gamma_k} = \frac{\partial}{\partial \gamma_k} \mathbf{M}^{-1} (\overline{\mathbf{GA}} \circ \mathbf{GA})^H \text{vec}(\mathbf{R} - \boldsymbol{\Sigma}_n) \quad (6.25)$$

where \mathbf{M} and \mathbf{G} depend on $\boldsymbol{\gamma}$ and thus on γ_k . Applying the formula for the derivative of an inverted matrix with respect to one of its elements [75], this can be rewritten as

$$\begin{aligned} \frac{\partial \mathbf{i}}{\partial \gamma_k} &= \left(-\mathbf{M}^{-1} \left(\frac{\partial}{\partial \gamma_k} \mathbf{M} \right) \mathbf{M}^{-1} (\overline{\mathbf{GA}} \circ \mathbf{GA})^H \right. \\ &\quad \left. + \mathbf{M}^{-1} \gamma_k (e^{-j\phi_k} \mathbf{E}_{kk} \overline{\mathbf{A}} \circ \mathbf{GA} + \overline{\mathbf{GA}} \circ e^{j\phi_k} \mathbf{E}_{kk} \mathbf{A})^H \right) \text{vec}(\mathbf{R} - \boldsymbol{\Sigma}_n) \end{aligned} \quad (6.26)$$

where \mathbf{E}_{kk} is the elementary matrix with all its entries set to zero except element E_{kk} which is set to 1. Inserting the vectorized version of (3.33) in (6.26) and removing the Khatri-Rao products using (C.11) we obtain

$$\frac{\partial \mathbf{i}}{\partial \gamma_k} = -2\gamma_k (2 - \gamma_k) \mathbf{M}^{-1} \text{Re} \left\{ \overline{\mathbf{A}}_{k:}^H \overline{\mathbf{A}}_{k:} \odot \mathbf{A}^H \boldsymbol{\Gamma}^2 \mathbf{A} \right\} \boldsymbol{\sigma}. \quad (6.27)$$

We have introduced the notation $\mathbf{A}_{k:} = \mathbf{E}_{kk}\mathbf{A}$, i.e. $\mathbf{A}_{k:}$ has only zero valued entries except on the k th row where the elements are equal to the corresponding elements of \mathbf{A} . The goal of this derivation is to obtain an expression for $\partial \mathbf{i} / \partial \gamma^T$. We will thus have to stack the expression for $\partial \mathbf{i} / \partial \gamma_k$ for all values of k in a single matrix. This is facilitated by introducing \mathbf{a}_k as the k th row of \mathbf{A} and rewriting (6.27) as

$$\frac{\partial \mathbf{i}}{\partial \gamma_k} = -2\gamma_k (2 - \gamma_k) \mathbf{M}^{-1} \operatorname{Re} \left\{ \mathbf{a}_k^T \odot \mathbf{A}^H \boldsymbol{\Gamma}^2 \mathbf{A} \boldsymbol{\Sigma} \mathbf{a}_k^H \right\}. \quad (6.28)$$

By stacking all vectors $\partial \mathbf{i} / \partial \gamma_k$ in a single matrix, we therefore get

$$\frac{\partial \mathbf{i}}{\partial \gamma^T} = -2\mathbf{M}^{-1} \operatorname{Re} \left\{ \mathbf{A}^T \odot \mathbf{A}^H \boldsymbol{\Gamma}^2 \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \right\} (2\mathbf{I} - \boldsymbol{\Gamma}) \boldsymbol{\Gamma}. \quad (6.29)$$

The corresponding result for $\partial \mathbf{i} / \partial \phi_k$ can be derived in a similar way, so we only present the main steps.

$$\begin{aligned} \frac{\partial \mathbf{i}}{\partial \phi_k} &= \frac{\partial}{\partial \phi_k} (\overline{\mathbf{G}\mathbf{A}} \circ \mathbf{G}\mathbf{A})^\dagger \operatorname{vec}(\mathbf{R} - \boldsymbol{\Sigma}_n) \\ &= \mathbf{M}^{-1} \left(\frac{\partial}{\partial \phi_k} \overline{\mathbf{G}\mathbf{A}} \circ \mathbf{G}\mathbf{A} \right)^H ((\overline{\mathbf{G}\mathbf{A}} \circ \mathbf{G}\mathbf{A}) \boldsymbol{\sigma}) \\ &= \mathbf{M}^{-1} (-je^{-j\phi_k} \boldsymbol{\Gamma} \overline{\mathbf{A}}_{k:} \circ \mathbf{G}\mathbf{A} + je^{j\phi_k} \overline{\mathbf{G}\mathbf{A}} \circ \boldsymbol{\Gamma} \mathbf{A}_{k:})^H ((\overline{\mathbf{G}\mathbf{A}} \circ \mathbf{G}\mathbf{A}) \boldsymbol{\sigma}). \end{aligned} \quad (6.30)$$

Removal of the Khatri-Rao products by reducing them to Hadamard products using (C.11) gives

$$\frac{\partial \mathbf{i}}{\partial \phi_k} = -2\gamma_k^2 \mathbf{M}^{-1} \operatorname{Im} \left\{ (\overline{\mathbf{A}}_{k:}^H \overline{\mathbf{A}}_{k:} \odot \mathbf{A}^H \boldsymbol{\Gamma}^2 \mathbf{A}) \right\} \boldsymbol{\sigma}. \quad (6.31)$$

Note that this has the same form as (6.27), so it can be rewritten in a similar way. This gives

$$\frac{\partial \mathbf{i}}{\partial \phi_k} = -2\gamma_k^2 \mathbf{M}^{-1} \operatorname{Im} \left\{ \mathbf{a}_k^T \odot \mathbf{A}^H \boldsymbol{\Gamma}^2 \mathbf{A} \boldsymbol{\Sigma} \mathbf{a}_k^H \right\} \boldsymbol{\sigma}. \quad (6.32)$$

and therefore

$$\frac{\partial \mathbf{i}}{\partial \phi^T} = -2\mathbf{M}^{-1} \operatorname{Im} \left\{ \mathbf{A}^T \odot \mathbf{A}^H \boldsymbol{\Gamma}^2 \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \right\} \boldsymbol{\Gamma}^2. \quad (6.33)$$

Finally, the partial derivative of the image values with respect to σ_{nk} is given by

$$\begin{aligned}\frac{\partial \mathbf{i}}{\partial \sigma_{nk}} &= \frac{\partial}{\partial \sigma_{nk}} \mathbf{M}^{-1} (\overline{\mathbf{GA}} \circ \mathbf{GA})^H \text{vec}(\mathbf{R} - \Sigma_n) \\ &= \mathbf{M}^{-1} (\overline{\mathbf{GA}} \circ \mathbf{GA})^H \text{vec}(-\mathbf{E}_{kk}) \\ &= -\mathbf{M}^{-1} \left(|\mathbf{GA}|^{\odot 2} \right)^H \text{vecdiag}(\mathbf{E}_{kk}).\end{aligned}\quad (6.34)$$

Therefore

$$\frac{\partial \mathbf{i}}{\partial \sigma_n^T} = -\mathbf{M}^{-1} \left(|\mathbf{GA}|^{\odot 2} \right)^H. \quad (6.35)$$

If $\Sigma_n = \sigma_n \mathbf{I}$ this reduces further to

$$\frac{\partial \mathbf{i}}{\partial \sigma_n} = -\mathbf{M}^{-1} \left(|\mathbf{GA}|^{\odot 2} \right)^H \mathbf{1}. \quad (6.36)$$

The partial derivatives as well as the CRB (see Sec. 4.2) contain terms involving $\mathbf{A}^H \mathbf{A}$, often weighted by the gains of the receiving elements. Given the physical interpretation of this factor discussed in Sec. 6.2.1, this suggests that the error pattern introduced by calibration errors follows the structures in the dirty image. This is confirmed by the example in Sec. 6.2.3. Since the CRB is inversely proportional to N , which is equal to the product of bandwidth and integration time, the image covariance due to calibration errors decreases proportional to bandwidth and integration time.

Thermal noise

In this section we derive an expression for the covariance of the image values due to the thermal noise in the data. We will therefore assume that perfect knowledge of the thermal noise power Σ_n is available to avoid confusion between the thermal noise contribution and the contribution of propagated calibration errors. The covariance of the image values is by definition given by

$$\begin{aligned}\text{cov}(\hat{\mathbf{i}}) &= \mathcal{E} \left\{ (\hat{\mathbf{i}} - \mathbf{i})(\hat{\mathbf{i}} - \mathbf{i})^H \right\} \\ &= \mathcal{E} \left\{ (\overline{\mathbf{GA}} \circ \mathbf{GA})^\dagger \left(\text{vec}(\hat{\mathbf{R}} - \Sigma_n) - \text{vec}(\mathbf{R} - \Sigma_n) \right) \right. \\ &\quad \times \left. \left(\text{vec}(\hat{\mathbf{R}} - \Sigma_n) - \text{vec}(\mathbf{R} - \Sigma_n) \right)^H (\overline{\mathbf{GA}} \circ \mathbf{GA})^{\dagger H} \right\}.\end{aligned}\quad (6.37)$$

This shows that under the assumption of perfectly known noise covariance matrix Σ_n , this noise covariance matrix drops out. Furthermore, $(\overline{\mathbf{GA}} \circ \mathbf{GA})^\dagger$ can be moved

outside the expectation operator, since it contains no estimated values. Therefore

$$\text{cov}(\hat{\mathbf{i}}) = \mathbf{M}^{-1} (\overline{\mathbf{GA}} \circ \mathbf{GA})^H \text{cov}(\hat{\mathbf{R}}) (\overline{\mathbf{GA}} \circ \mathbf{GA}) \mathbf{M}^{-1}. \quad (6.38)$$

For Gaussian models

$$\text{cov}(\hat{\mathbf{R}}) = \frac{1}{N} (\overline{\mathbf{R}} \otimes \mathbf{R}) \quad (6.39)$$

and we find that

$$\text{cov}(\hat{\mathbf{i}}) = \frac{1}{N} \mathbf{M}^{-1} (\overline{\mathbf{GA}} \circ \mathbf{GA})^H (\overline{\mathbf{R}} \otimes \mathbf{R}) (\overline{\mathbf{GA}} \circ \mathbf{GA}) \mathbf{M}^{-1}. \quad (6.40)$$

This can be rewritten using standard Kronecker and Khatri-Rao product relations (see App. C) as

$$\begin{aligned} \text{cov}(\hat{\mathbf{i}}) &= \frac{1}{N} \mathbf{M}^{-1} (\overline{\mathbf{GA}} \circ \mathbf{GA})^H (\overline{\mathbf{RGA}} \circ \mathbf{RGA}) \mathbf{M}^{-1} \\ &= \frac{1}{N} \mathbf{M}^{-1} |\mathbf{A}^H \mathbf{G}^H \mathbf{RGA}|^{\odot 2} \mathbf{M}^{-1}. \end{aligned} \quad (6.41)$$

Finally, substituting the data model presented in (3.33), we get

$$\text{cov}(\hat{\mathbf{i}}) = \frac{1}{N} \mathbf{M}^{-1} |\mathbf{A}^H \boldsymbol{\Gamma}^2 \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \boldsymbol{\Gamma}^2 \mathbf{A} + \mathbf{A}^H \boldsymbol{\Gamma}^2 \boldsymbol{\Sigma}_n \mathbf{A}|^{\odot 2} \mathbf{M}^{-1}. \quad (6.42)$$

It is interesting to note that for an array of identical elements, i.e. for $\mathbf{G} = \mathbf{I}$, diagonalization of $\mathbf{A}^H \mathbf{A}$ does not only ensure a homogeneous noise distribution over the map after the deconvolution operation as demonstrated in Sec. 6.2.1, but also diagonalizes the image covariance due to thermal noise, thus ensuring that the noise on the pixels is uncorrelated. The Gram matrix $\mathbf{A}^H \mathbf{A}$ describes the amount of linear independence (or orthogonality) of the direction of arrival vectors within the field-of-view of the array, which can be visualized as the array beam pattern. This observation therefore suggests that an array with a low side lobe pattern does not only provide good spatial separation between source signals, but also ensures a relatively low covariance between image values after deconvolution.

Confusion noise

The contributions to the effective noise from calibration errors and thermal noise scale inversely with the number of samples N , which is equal to the product of bandwidth and integration time. This implies that, theoretically, these sources of image noise can be reduced to arbitrarily low levels. In practice, a radio astronomical array will detect

more sources with every reduction of the noise in the map. At some point, the number of detected sources will become larger than the number of resolution elements in the image, which will turn the map into one blur of sources. The maximum density of discernable sources is the classical confusion limit and relates to the resolution of the image.

In terms of self-calibration, having more detectable sources within the field-of-view requires more source parameters to describe the source model. At some point, the self-calibration problem will become ill-posed. We will refer to this as the self-calibration confusion limit. Although the exact limit depends on the minutiae of the array and source configuration, we can easily compute an upper limit on the tractable number of sources based on the argument that the number of unknowns should be smaller than the number of equations. The data model provides a relation between the parameters and the data. For a P -element array, the covariance matrix contains P^2 independent real values, so the data model can be regarded as P^2 independent equations. Solving for the direction independent complex gains requires $2P - 1$ real valued parameters, estimation of the receiver element noise powers requires another P parameters and the Q sources are described by $3Q - 3$ parameters, namely the apparent source powers relative to the first source and two coordinates per source with an appropriate constraint to avoid a common position shift of all sources. The self-calibration problem is therefore constrained by

$$P^2 \geq 2P - 1 + P + 3Q - 3 = 3P + 3Q - 4 \quad (6.43)$$

implying that

$$Q \leq \frac{P^2 - 3P + 4}{3}. \quad (6.44)$$

The spatial Nyquist sampling with the 8×8 -element URA allows an image grid of $15 \times 15 = 225$ image values. This resolution was confirmed by the condition number analysis presented in Fig. 6.3. However, the upper limit based on the analysis above for a 64 element array is 1302. The mismatch between this upper limit and the actual number of uniquely solvable image values can be attributed to the redundancy in the array. Due to this redundancy, the crosscorrelations of many antenna pairs provide the same spatial information instead of providing additional information on the spatial distribution of signals over the sky. In terms of the argument leading to (6.44), there is linear dependence between the equations and therefore the number of equations that provide additional information on the parameters is reduced. The five armed array performs much better in this regard. For $P = 40$, the upper limit on the number of sources given by (6.44) is 494. Since $\sqrt{494} \approx 22$, we can form an image grid of 22×22 points, thus providing a resolution of $\Delta l = 0.09$. Fig. 6.3 shows that the condition number for this array goes numerically to infinity at $\Delta l = 0.08$,

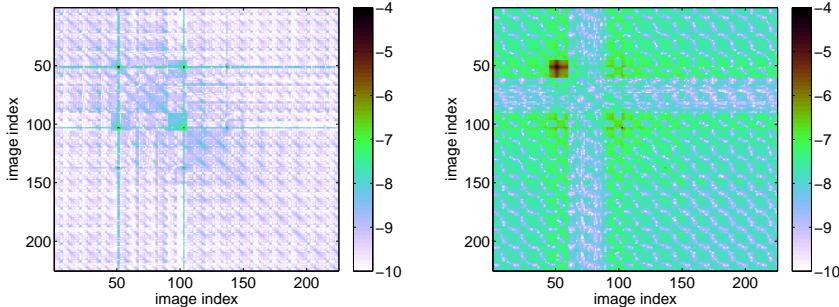


Figure 6.4: The logarithm (base 10) of the image covariance matrix due to calibration errors (left) and due to the measurement noise (right) on the same color scale. The former is almost everywhere two orders of magnitude lower. The four sources are located at grid points with indices 51, 52, 103 and 137 with power 1, 0.6, 1.3 and 0.1 respectively.

showing that the five armed array approaches its theoretical self-calibration confusion limit. This example illustrates that if processing power is cheap compared to antenna hardware, a non-redundant array should be preferred over a redundant array if the confusion limit should be reached without introducing an ill-posed deconvolution problem.

In this section, we addressed the classical confusion limit for pure imaging problems and the self-calibration confusion limit for self-calibration problems. This type of confusion is often called source confusion as opposed to side lobe confusion, which refers to blurring of the image by side lobe leftovers introduced in the CLEAN process. In terms of the analysis presented here, side lobe confusion is part of the deconvolution problem and is therefore implicitly included in the analysis of calibration error and thermal noise propagation and does not need to be addressed separately. Source confusion does require a separate treatment, because it involves the source density distribution as a function of source brightness.

6.2.3 Implications

Thermal noise vs. propagated calibration errors

I compare the image covariance due to calibration errors to the image covariance due to the noise in the data in a simulation. For this simulation, I used the 4-source sky model introduced in Sec. 6.2.1. I will assume that the calibration parameters are

calculated from a separate data set with the same data model and integration time. I computed the CRB for the 8×8 -element URA using the relations presented in Sec. 4.2 assuming simultaneous estimation of the omni-directional complex gains, \mathbf{g} , and the system noise power, σ_n , which was assumed to be the same for all array elements. We assumed a short term integration over $N = 16384$ samples. This CRB was used in (6.24) to compute the image covariance due to calibration errors. The magnitudes of the entries of the image covariance matrix $\text{cov}(\mathbf{i})$ are shown in the left panel of Fig. 6.4 on a logarithmic scale.

The image covariance matrix due to the noise in the measurements was determined using (6.42). The result is shown in the right panel of Fig. 6.4. This shows that the covariance of the image values due to calibration errors is more concentrated at the source locations than the covariance due to the system noise and is generally more than two orders of magnitude lower. These results indicate that the calibration errors only represent a minor contribution to the total effective noise, even when the calibration measurements are done on the same (short) time scales using sources of similar strength as observed in the actual measurements, i.e. when the calibration measurement is similar to the actual measurement.

Calibration observations vs. self-calibration

In the previous section, we discussed the situation in which the array is calibrated in a separate measurement. This scheme requires an extremely stable instrument. In most practical applications, the calibration is therefore done on the same data that is also used to provide the final image (self-calibration). It is interesting to see how these scenarios compare. To this end, we used the relationships presented in Sec. 4.2 to compute the CRB for simultaneous estimation of the omni-directional complex gains, the apparent source powers, the source locations and the system noise power for the 8×8 -element URA. This corresponds to the third scenario described in Sec. 5.2.

Figure 6.5 compares the CRBs for these two cases. The expected covariance of the gains and phases in the self-calibration experiment is higher since more parameters have to be estimated simultaneously. The behavior of the CRB on the phases in the self-calibrated observation (sloped upwards for increasing parameter index) can be explained by the interaction between the source parameters and the gain phases in combination with the choice of the phase reference element in the corner of the array.

Table 6.1 shows, for each of the two cases, the covariance matrices of the apparent source powers, i.e. the covariance of the image values at the locations of the sources. The scaling factor ambiguity between \mathbf{G} and Σ in the self-calibration case is resolved by putting $\sigma_1 = 1$ to constrain the problem. Therefore, only the covariance values of the other three sources are tabulated. For the case of a separate calibration

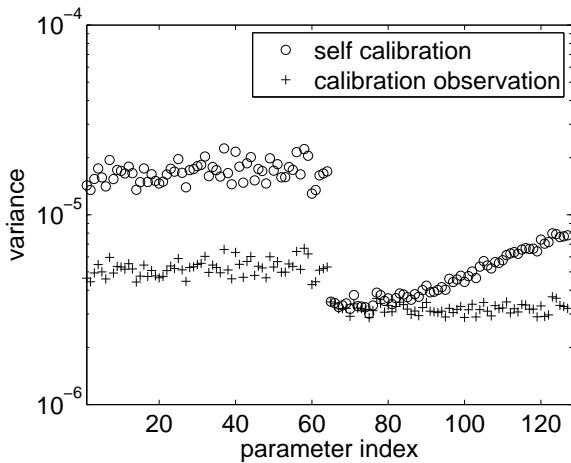


Figure 6.5: The CRB for the omni-directional complex gain amplitudes (parameters 1 through 64) and phases (parameters 65 through 127) for a separate calibration observation and the self-calibration approach.

observation, the covariance matrix was determined by the sum of the image covariances due to calibration errors and system noise. The results in the table indicate that the variance in the source power estimates in both cases are comparable, although the source power estimates are slightly better when gain calibration data is available from a separate measurement. The covariance values found for a separate calibration stage are much lower than the corresponding values for self-calibration. This suggests that pure imaging is more capable of separating source signals from different directions than self-calibrated imaging.

6.2.4 Conclusions

In this discussion on fundamental imaging limits, a closed form solution for snapshot imaging including deconvolution was presented based on a data model (measurement equation) for the antenna signal covariance matrix or visibilities. The presented comprehensive framework is sufficiently flexible to enable extension of this analysis to synthesis observations, since the data model for a synthesis observation has the same form as the data model for a snapshot observation presented in (3.33). This was demonstrated in Sec. 3.6. This similarity will become even more apparent when we derive expressions for least squares synthesis imaging in Sec. 6.3. This framework

self-calibration				
index	2	3	4	
2	0.266×10^{-4}	0.287×10^{-4}	0.022×10^{-4}	
3	0.287×10^{-4}	1.237×10^{-4}	0.048×10^{-4}	
4	0.022×10^{-4}	0.048×10^{-4}	0.007×10^{-4}	

separate calibration				
index	2	3	4	
2	0.280×10^{-4}	0.072×10^{-4}	0.006×10^{-4}	
3	0.072×10^{-4}	0.772×10^{-4}	0.012×10^{-4}	
4	0.006×10^{-4}	0.012×10^{-4}	0.005×10^{-4}	

Table 6.1: Covariance of source power estimates in a self-calibrated image and in an image produced after a separate calibration step on data with the same source model.

allowed us to make the first rigorous assessment of the effective noise floor, which is the combined effect of propagated calibration errors, thermal noise and source confusion, in the image in terms of the covariance of the image values. The simulations for a URA presented in Sec. 6.2.3 indicate that the effects of propagated calibration errors are strongly concentrated at the source locations but are considerably smaller than the thermal noise at other image points. The results also suggest that if the instrument is sufficiently stable, a separate calibration step is to be preferred over a self-calibrated image since it allows better source separation in the imaging process.

The effects of deconvolution can be described by a deconvolution matrix that describes the amount of linear independence (orthogonality) of the spatial signature vectors weighted by the actual gains of the receiving elements. A diagonal deconvolution matrix not only ensures the best possible spatial separation between the sources, but also ensures a homogeneous noise distribution over the map. This begs the question whether this matrix can be diagonalized by array design or by applying appropriate weights to the array elements. Since the deconvolution matrix is related to the array beam pattern, the latter is equivalent to finding weights that suppress the side lobe pattern at least in the direction of other sources, which suggests that techniques like robust Capon beam forming could provide the requested weighting [117]. The con-

dition number of the deconvolution matrix can be used to assess the quality of the solution to the deconvolution problem.

Compared to a redundant array (ULA, URA), an array without redundant element spacings provides much better possibilities to approach the maximum number of solvable image points for a fixed number of antenna elements, thereby allowing the system to reach the theoretical self-calibration confusion limit.

6.3 Least squares synthesis imaging

This section presents closed form solutions for weighted least squares synthesis imaging. In the context of this thesis, this demonstrates two points. First, it shows that it is straightforward to apply the effective noise analysis presented in Sec. 6.2.2 to synthesis observations, although the mathematical minutiae may become cumbersome. Second, it outlines how the WALS method presented in Sec. 5.3 can be extended to handle polarized sources, which implies that the WALS approach may be turned into a full polarized self-calibration algorithm. This excursion therefore emphasizes the fact that the applicability of the results presented in this thesis is not limited to single polarization snapshot observations. This point is also demonstrated by the LOFAR images presented in Sec. 2.4.

Faced with the challenges posed by the advent of the next generation of radio telescopes, people in academia are currently reconsidering the entire imaging process [7, 21, 64, 67, 76]. Although vastly different methods are applied by distinct research groups, the common denominator in all these methods is the model based approach and the application of signal processing techniques. The least squares image vector estimators derived here fall in the same category and are formulated as explicit expressions for radio astronomical imaging. They can therefore be regarded as specialized versions of the optimal map making approach used in cosmic microwave background imaging, originally discussed by Texmark [114, 115]. It is known that l_1 -minimization provides better results when the data is sparse [22, 94]. Sparsity implies that the image contains a number of discrete sources and is mostly empty. This does not hold for station data as is clear from the images presented in Ch. 2, making least squares imaging a proper deconvolution method for imaging with a LOFAR station, but also for imaging of the complex source structures in studies on, e.g., Galactic polarization and the EoR that will be conducted by LOFAR and SKA.

6.3.1 Generic formulation

As discussed in [82] the weighted least squares covariance matching approach leads to estimates, that are, for a large number of samples, equivalent to maximum likeli-

hood estimators and are therefore asymptotically efficient. The weighted least squares cost function is

$$\kappa_W(\boldsymbol{\theta}) = (\hat{\mathbf{r}}_{syn} - \mathbf{r}_{syn}(\boldsymbol{\theta}))^H \mathbf{W}^2 (\hat{\mathbf{r}}_{syn} - \mathbf{r}_{syn}(\boldsymbol{\theta})) = \|\mathbf{W}(\hat{\mathbf{r}}_{syn} - \mathbf{r}_{syn}(\boldsymbol{\theta}))\|_F^2, \quad (6.45)$$

where \mathbf{r}_{syn} is the visibility vector for the entire synthesis observation as defined in (3.76), which is parameterized by the model parameters stacked in $\boldsymbol{\theta}$. The optimal weight is known to be the inverse of the asymptotic covariance of the residuals [82]. The optimal weight for Gaussian sources is therefore $\mathbf{W} = \text{cov}(\mathbf{r}_{syn})^{-1/2}$. Since the covariance of a single snapshot observation, $\text{cov}(\text{vec}(\mathbf{R}_m))$, is known to be $(\bar{\mathbf{R}}_m \otimes \mathbf{R}_m) / N$ and the snapshots are assumed to be uncorrelated, the optimal weights are given by

$$\mathbf{W} = \text{bdiag}\left(\bar{\mathbf{R}}_m^{-1/2} \otimes \mathbf{R}_m^{-1/2}\right). \quad (6.46)$$

In this section we will formulate the WLS estimation problem and its analytic solution for two scenarios:

1. $\boldsymbol{\theta} = [\boldsymbol{\sigma}^T, \boldsymbol{\sigma}_n^T]$, where $\boldsymbol{\sigma}_n$ is the parameterization of the noise term, i.e. the image values and the noise parameters are both unknown;
2. $\boldsymbol{\theta} = \boldsymbol{\sigma}$, i.e. the image values are unknown, but the noise in the measurements is either known or negligible.

We shall first present the solution in a very generic form to emphasize the overall structure, which turns out to be the same for both scenarios and for scalar as well as for full polarized imaging. This presentation without all the distracting mathematical details also helps to highlight a few important characteristics of this method.

The WLS estimation for the first scenario can be formulated as

$$\{\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}_n\} = \underset{\boldsymbol{\sigma}, \boldsymbol{\sigma}_n}{\text{argmin}} \|\mathbf{W}(\hat{\mathbf{r}}_{syn} - \mathbf{M}\boldsymbol{\sigma} - \mathbf{n})\|_F^2, \quad (6.47)$$

where the noise vector \mathbf{n} is parameterized by $\boldsymbol{\sigma}_n$ as discussed in Sec. 3.6. This least squares problem may be solved in two ways. One may solve for $\boldsymbol{\sigma}_n$ as a function of $\boldsymbol{\sigma}$, substitute the result back in the least squares cost function and consecutively solve for $\boldsymbol{\sigma}$. The second approach is to solve for $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}_n$ using an alternating least squares approach. Here we will follow the first approach assuming that $\mathbf{n} = \mathbf{M}_n \boldsymbol{\sigma}_n$, where $\boldsymbol{\sigma}_n = [\boldsymbol{\sigma}_{n1}^T, \boldsymbol{\sigma}_{n2}^T, \dots, \boldsymbol{\sigma}_{nM}^T]^T$ is the parameterization of \mathbf{n} . The solution for

σ_n is

$$\begin{aligned}\sigma_n &= \underset{\sigma_n}{\operatorname{argmin}} \| \mathbf{W} (\hat{\mathbf{r}}_{syn} - \mathbf{M}\sigma) - \mathbf{WM}_n\sigma_n \|_F^2 \\ &= (\mathbf{WM}_n)^\dagger \mathbf{W} (\hat{\mathbf{r}}_{syn} - \mathbf{M}\sigma) \\ &= \widetilde{\mathbf{M}}_n (\hat{\mathbf{r}}_{syn} - \mathbf{M}\sigma)\end{aligned}\quad (6.48)$$

Substitution of this result in (6.47) shows that σ can be found by

$$\begin{aligned}\hat{\sigma} &= \underset{\sigma}{\operatorname{argmin}} \| (\mathbf{W} - \mathbf{WM}_n \widetilde{\mathbf{M}}_n) (\hat{\mathbf{r}}_{syn} - \mathbf{M}\sigma) \|_F^2 \\ &= \underset{\sigma}{\operatorname{argmin}} \| \widetilde{\mathbf{W}} \hat{\mathbf{r}}_{syn} - \widetilde{\mathbf{W}} \mathbf{M}\sigma \|_F^2,\end{aligned}\quad (6.49)$$

where $\widetilde{\mathbf{W}} = \mathbf{W} - \mathbf{WM}_n \widetilde{\mathbf{M}}_n$. The solution to this problem is given by

$$\begin{aligned}\hat{\sigma} &= (\widetilde{\mathbf{W}} \mathbf{M})^\dagger \widetilde{\mathbf{W}} \hat{\mathbf{r}}_{syn} \\ &= (\mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} \mathbf{M})^{-1} \mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} \hat{\mathbf{r}}_{syn} \\ &= \mathbf{M}_d^{-1} \sigma_d,\end{aligned}\quad (6.50)$$

where $\mathbf{M}_d = \mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} \mathbf{M}$ is the deconvolution matrix and $\sigma_d = \mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} \hat{\mathbf{r}}_{syn}$ is the dirty image.

The matrices introduced in this presentation all have a very distinct physical meaning. The linear mapping of the noise vector \mathbf{n} on its parameterization σ_n is given by the noise model \mathbf{M}_n . Depending on the noise model and the weighting \mathbf{W} applied to the measured visibilities $\hat{\mathbf{r}}_{syn}$, a new weighting matrix $\widetilde{\mathbf{W}}$ can be constructed which filters the noise from the measurements. Based on these weights and the instrument model \mathbf{M} , a dirty image σ_d can be constructed. The deconvolution matrix \mathbf{M}_d describes how the flux in each pixel in the true image is distributed over the pixels in the dirty image during the imaging process. The inversion of this matrix therefore handles the deconvolution and the primary beam correction implicitly.

Although this description is very general, a number of important conclusions can already be drawn:

- The deconvolution matrix describes how the values in the dirty image are related to the values in the true image. More specifically, each dirty image value is a weighted sum over all true image values. The weights depend on the overall sensitivity pattern of the array and its behavior during the entire observation.

- Since the instrument model describes how the intensity distribution over the sky maps on the visibilities for each individual snapshot, the instrument model can describe arbitrary time and frequency dependent beam patterns, both for the individual receivers as well as for the whole array, as long as the time and frequency span of the individual snapshots is sufficiently short to allow a constant model for each snapshot.
- The formulation of the instrument model presented in Ch. 3 uses a direct Fourier transform to describe the geometrical delays over the array for distinct DOAs. It can therefore handle arbitrary array geometries and extremely wide FOVs, up to full sky imaging.
- The sizes of the deconvolution matrix and the dirty image are solely determined by the number of image values. If the image consists of Q points, the deconvolution matrix has size $Q \times Q$ and the dirty image is a $Q \times 1$ column vector regardless of the amount of data stacked in the $MP^2 \times 1$ visibility vector.
- The inverse of the deconvolution matrix does not have to be explicitly known to solve the imaging problem. This fact should be exploited to save computational effort and ensure numerical stability. The problem can be solved using, e.g., Gaussian elimination.
- The deconvolution matrix is Hermitian. Since σ_d and σ are both real valued (the physical reason for this is that their elements represent powers), the deconvolution matrix must be real valued as well and is therefore symmetric.

If the noise is either known or negligible, as in the second scenario, the image values can be found by solving

$$\hat{\sigma} = \underset{\sigma}{\operatorname{argmin}} \| \mathbf{W} (\hat{\mathbf{r}}_{syn} - \mathbf{n}) - \mathbf{W} \mathbf{M} \sigma \|_F^2. \quad (6.51)$$

This has a solution similar to (6.50):

$$\hat{\sigma} = (\mathbf{W} \mathbf{M})^\dagger \mathbf{W} (\hat{\mathbf{r}}_{syn} - \mathbf{n}) = \mathbf{M}_d^{-1} \sigma_d \quad (6.52)$$

with deconvolution matrix $\mathbf{M}_d = \mathbf{M}^H \mathbf{W}^H \mathbf{W} \mathbf{M}$ and $\sigma_d = \mathbf{M}^H \mathbf{W}^H \mathbf{W} (\hat{\mathbf{r}}_{syn} - \mathbf{n})$. Since the structure of this result is the same as the structure of the result for the first scenario, the inferences made above also hold here. In the next subsections, we will describe the deconvolution matrix and the dirty image in more detail for both scenarios and for scalar as well as full polarized imaging.

6.3.2 Scalar imaging with known or negligible noise

The results for this scenario are derived in Sec. 6.A.1 in the appendix to this chapter. Here, I will only state the results. The deconvolution matrix is given by

$$\mathbf{M}_d = \sum_{m=1}^M |\mathbf{A}_{sm}^H \mathbf{R}^{-1} \mathbf{A}_{sm}|^{\odot 2}, \quad (6.53)$$

where $\mathbf{A}_{sm} = \mathbf{G}_m \mathbf{Q}_m (\mathbf{A}_m \odot \mathbf{G}_{0m}) \mathbf{I}_{sm}$ as introduced in (3.78). The dirty image is given by

$$\boldsymbol{\sigma}_d = \sum_{m=1}^M \left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \circ \mathbf{R}_m^{-1} \mathbf{A}_{sm} \right)^H (\text{vec}(\mathbf{R}_m) - \text{vec}(\boldsymbol{\Sigma}_{nm})). \quad (6.54)$$

If we assume a diagonal noise covariance matrix per snapshot, this can be simplified further to

$$\boldsymbol{\sigma}_d = \sum_{m=1}^M \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \right) \mathbf{1} - \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \right) \boldsymbol{\sigma}_{nm}. \quad (6.55)$$

The following inferences can be made based on these results in addition to the conclusions drawn earlier:

- The deconvolution matrix generally has no special structure (block matrix, band matrix, circulant matrix, etc.) that can be exploited to reduce the numerical complexity of the algorithm apart from the fact that it is a real symmetric matrix.
- A careful look at the individual terms contributing to the dirty image vector reveals that each entry of this vector is the result of a summation containing pairs of values and their complex conjugate. Therefore, each term only contains real valued elements. Thus, the dirty image vector is a real valued vector, as expected based on its physical significance.
- Both the deconvolution matrix and the dirty image vector are the result of a summation over the individual snapshots. This implies that the full data set can be subdivided in snapshots over time and frequency which can be processed independently. This is the reason why the sizes of the deconvolution matrix and the dirty image do not depend on the amount of data. It also shows that the computation of the deconvolution matrix and the dirty image vector can be implemented as an embarrassingly parallel algorithm.

- If the noise covariance matrix is diagonal, all Kronecker and Khatri-Rao products can be removed from the equations. This reduces the sizes of the matrices involved considerably thereby saving processing power and memory usage.

6.3.3 Scalar imaging with unknown noise

As demonstrated in Sec. 6.A.2, the deconvolution matrix in the more general case in which the noise is unknown and has to be estimated as well, is given by

$$\begin{aligned} \mathbf{M}_d = & \sum_{m=1}^M |\mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \mathbf{A}_{sm}|^{\odot 2} + \\ & - \sum_{m=1}^M |\mathbf{A}_{sn}^H \mathbf{R}_m^{-1}|^{\odot 2} |\mathbf{R}_m^{-1}|^{-\odot 2} |\mathbf{R}_m^{-1} \mathbf{A}_{sm}|^{\odot 2}. \end{aligned} \quad (6.56)$$

Comparison with the deconvolution matrix in the noise free case presented in (6.53) shows that the first term is just the noise free deconvolution matrix. This implies that the second term represents a correction for the noise filtering operation.

The dirty image vector

$$\boldsymbol{\sigma}_d = \sum_{m=1}^M \left((\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H) \mathbf{1} - |\mathbf{A}_{sm}^H \mathbf{R}_m^{-1}|^{\odot 2} |\mathbf{R}_m^{-1}|^{-\odot 2} \text{vecdiag}(\mathbf{R}_m^{-1}) \right) \quad (6.57)$$

closely resembles the dirty image vector in the noise free case given by (6.55). The only difference is that $\boldsymbol{\sigma}_{nm}$ is replaced by $|\mathbf{R}_m^{-1}|^{-\odot 2} \text{vecdiag}(\mathbf{R}_m^{-1})$, which can be interpreted as an estimate of the noise parameter vector $\boldsymbol{\sigma}_{nm}$. The $\text{vecdiag}(\cdot)$ operator produces a vector containing the elements on the main diagonal of its argument.

Note that all the remarks on the physical interpretation and on the attractive properties of the deconvolution matrix and the dirty image vector made for the noise free case still hold if we have to filter the noise from the data. The summation over snapshots is retained due to the fact that each snapshot has its own noise estimates, which ensures that the snapshots are treated independently during the estimation process. Equations (6.56) and (6.57) have been used to produce the least squares imaging results presented in Sec. 2.4.

6.3.4 Full polarized imaging with known or negligible noise

Extension of the imaging problem to the full polarized case implies that we have to estimate four values per image point. This aspect appears nicely if we compare the

deconvolution matrix for polarized imaging with known noise derived in Sec. 6.A.3,

$$\mathbf{M}_d = \sum_{m=1}^M \mathbf{I}_{sm}^H \begin{bmatrix} \ddots & & \vdots \\ \cdots & \mathbf{M}_{3q_1 q_2 m} & \cdots \\ & \vdots & \ddots \end{bmatrix} \mathbf{I}_{sm} \quad (6.58)$$

where

$$\mathbf{M}_{3q_1 q_2 m} = \mathbf{B}_{q_1 m}^H \left(\overline{\mathbf{A}}_{q_1 m}^H \overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{q_2 m}' \otimes \mathbf{A}_{q_1 m}^H \mathbf{R}_m^{-1} \mathbf{A}_{q_2 m}' \right) \mathbf{B}_{q_2 m}, \quad (6.59)$$

with the deconvolution matrix for the corresponding scalar case given by (6.53). The basic structure of the solution is the same, but the Hadamard product has been replaced by a Kronecker product to produce a 4×4 submatrix per source pair and a non-scalar gain factor has appeared in the form of the 4×4 matrix \mathbf{B}_{qm} which represents the transformation between real valued image values to complex valued components of the electromagnetic field.

A similar observation holds for the dirty image vector, $\epsilon_d = [\mathbf{e}_1^T, \mathbf{e}_2^T, \dots, \mathbf{e}_Q^T]^T$, in which the source coherency \mathbf{e}_q for each image point is stacked (see Sec. 3.2.2 for a more detailed discussion). The reader can verify this by comparing

$$\epsilon_d = \sum_{m=1}^M \mathbf{I}_{sm}^H \begin{bmatrix} & & \vdots & \\ \mathbf{B}_{qm}^H \left((\overline{\mathbf{A}}_{qm}^H \overline{\mathbf{R}}_m^{-1} \circ \mathbf{A}_{qm}^H) \mathbf{1} - (\overline{\mathbf{A}}_{qm}^H \overline{\mathbf{R}}_m^{-1} \circ \mathbf{A}_{qm}^H \mathbf{R}_m^{-1}) \boldsymbol{\sigma}_{nm} \right) & & & \\ & & \vdots & \end{bmatrix} \quad (6.60)$$

with the result for the scalar case given by (6.55). A derivation of this result can be found in Sec. 6.A.3.

The matrix \mathbf{B}_{qm} plays a key role in ensuring a real valued deconvolution matrix and a real valued dirty image vector, because the replacement of the Hadamard product by a Kronecker product (deconvolution matrix) or a Khatri-Rao product (dirty image vector) voids the warrant of real valued results. The matrix \mathbf{B}_{qm} can be made snapshot independent by a careful definition of the Jones matrices included in \mathbf{A}_{qm} . These Jones matrices should describe the sensitivity of the antennas to radiation from a snapshot dependent direction with its electromagnetic field components defined in a snapshot independent coordinate system. The latter ensures that \mathbf{B}_{qm} is snapshot independent and therefore does not have to align the polarization definitions of the individual snapshots. Although the mathematics involved are slightly more laborious to work with or implement, the convenient properties of the deconvolution matrix and dirty image vector described earlier still hold.

6.3.5 Full polarized imaging with unknown noise

The final and most general case is full polarization imaging with unknown noise. Expressions for the deconvolution matrix and dirty image vector are derived in Sec. 6.A.4. The deconvolution matrix is given by

$$\mathbf{M}_d = \sum_{m=1}^M \mathbf{I}_{sm}^H \begin{bmatrix} \ddots & & \vdots \\ \cdots & \mathbf{M}_{3q_1 q_2 m} - \mathbf{M}_{4q_1 q_2 m} & \cdots \\ & \vdots & \ddots \end{bmatrix} \mathbf{I}_{sm}, \quad (6.61)$$

where $\mathbf{M}_{3q_1 q_2 m}$ is given by (6.59) and $\mathbf{M}_{4q_1 q_2 m}$ by

$$\begin{aligned} \mathbf{M}_{4q_1 q_2 m} &= \mathbf{B}_{q_1 m}^H \left(\overline{\mathbf{A}}_{q_1 m}' \overline{\mathbf{R}}_m^{-1} \otimes \mathbf{A}_{q_1 m}' \mathbf{R}_m^{-1} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \\ &\times \left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{q_2 m}' \otimes \mathbf{R}_m^{-1} \mathbf{A}_{q_2 m}' \right) \mathbf{B}_{q_2 m}. \end{aligned} \quad (6.62)$$

The corresponding dirty image vector is described by

$$\begin{aligned} \boldsymbol{\epsilon}_d &= \sum_{m=1}^M \mathbf{I}_{sm}^H \begin{bmatrix} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}_{qm}' \overline{\mathbf{R}}_m^{-1} \circ \mathbf{A}_{qm}' \right) \mathbf{1} \\ \vdots \\ \vdots \end{bmatrix} + \\ &- \sum_{m=1}^M \mathbf{I}_{sm}^H \mathbf{M}_{2m} \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \text{vecdiag}(\mathbf{R}_m^{-1}), \end{aligned} \quad (6.63)$$

where

$$\mathbf{M}_{2m} = \begin{bmatrix} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}_{qm}' \overline{\mathbf{R}}_m^{-1} \circ \mathbf{A}_{qm}' \mathbf{R}_m^{-1} \right) \\ \vdots \end{bmatrix}. \quad (6.64)$$

Comparison of these results with the corresponding expression for the noise free case shows that, as in the scalar case, the deconvolution matrix and the dirty image vector can be split in two terms. The first term describes the noise free imaging process while the second term can be considered as a correction to compensate for the changes in the weighting matrix to filter the noise from the data.

6.3.6 Treatment of more complicated noise models

The results in the previous sections were derived assuming that the noise covariance matrices of the individual snapshots Σ_{nm} are diagonal. This results in a very simple noise model with $\mathbf{M}_{nm} = (\mathbf{I} \circ \mathbf{I}) \boldsymbol{\sigma}_{nm}$ and $\mathbf{M}_n = (\mathbf{I} \otimes (\mathbf{I} \circ \mathbf{I})) \boldsymbol{\sigma}_n$. These simple expressions allowed to reduce several Kronecker and Khatri-Rao products to Hadamard products thereby reducing the sizes of the matrices involved considerably. In many practical cases, a more complicated noise model may be required to describe effects like noise coupling or correlator offsets. In these cases, \mathbf{M}_n may no longer have a simple description that allows to simplify the expressions after substitution of the solution for $\boldsymbol{\sigma}_n$ or may even require a description based on a weighted sum of predefined matrices as in [42, 132]. If such a noise model is required, it may not be profitable to follow the approach presented above, which involves back substitution of the solution for the noise parameters. Instead, it may be computationally more efficient to solve for noise and image parameters alternatingly until convergence, i.e. to follow a similar approach as introduced in Sec. 5.5.2 to handle non-diagonal noise covariance matrices in self-calibration problems. The Monte Carlo simulations presented in that section suggest that such an approach may even converge within only three or four iterations.

6.A Derivation of closed form least squares imaging solutions

6.A.1 Scalar imaging with known or negligible noise

In this appendix, expressions for the deconvolution matrix and dirty image vector are derived for scalar imaging with known or negligible noise. From the generalized expression given by (6.52) it follows that $\mathbf{M}_d = \mathbf{M}^H \mathbf{W}^H \mathbf{W} \mathbf{M}$. The optimal weights are given by (6.46), thus

$$\begin{aligned} \mathbf{W}^H \mathbf{W} &= \text{bdiag}\left(\overline{\mathbf{R}}_m^{-1/2} \otimes \mathbf{R}_m^{-1/2}\right)^H \text{bdiag}\left(\overline{\mathbf{R}}_m^{-1/2} \otimes \mathbf{R}_m^{-1/2}\right) \\ &= \text{bdiag}\left(\overline{\mathbf{R}}_m^{-1} \otimes \mathbf{R}_m^{-1}\right). \end{aligned} \quad (6.65)$$

The instrument model implied by (3.78) is

$$\mathbf{M} = \begin{bmatrix} \overline{\mathbf{A}}_{s1} \circ \mathbf{A}_{s1} \\ \overline{\mathbf{A}}_{s2} \circ \mathbf{A}_{s2} \\ \vdots \\ \overline{\mathbf{A}}_{sM} \circ \mathbf{A}_{sM} \end{bmatrix}. \quad (6.66)$$

Pre-multiplication of $\mathbf{W}^H \mathbf{W}$ with \mathbf{M}^H gives

$$\mathbf{M}^H \mathbf{W}^H \mathbf{W} = \left(\text{bdiag} \left(\overline{\mathbf{R}}_m^{-1} \otimes \mathbf{R}_m^{-1} \right)^H \begin{bmatrix} \vdots \\ \overline{\mathbf{A}}_{sm} \circ \mathbf{A}_{sm} \\ \vdots \end{bmatrix} \right)^H, \quad (6.67)$$

where the second factor indicates a matrix consisting of M blocks $\overline{\mathbf{A}}_{sm} \circ \mathbf{A}_{sm}$ stacked column-wise and indexed by m . The Hermitian transpose is used to invert the order of multiplication, so we may use (C.10). This gives

$$\mathbf{M}^H \mathbf{W}^H \mathbf{W} = \left[\cdots, \left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \circ \mathbf{R}_m^{-1} \mathbf{A}_{sm} \right)^H, \cdots \right]. \quad (6.68)$$

Post-multiplying with \mathbf{M} now gives the deconvolution matrix

$$\begin{aligned} \mathbf{M}_d &= \sum_{m=1}^M \left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \circ \mathbf{R}_m^{-1} \mathbf{A}_{sm} \right)^H (\overline{\mathbf{A}}_{sm} \circ \mathbf{A}_{sm}) \\ &= \sum_{m=1}^M \overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \odot \mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \mathbf{A}_{sm} \\ &= \sum_{m=1}^M |\mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \mathbf{A}_{sm}|^{\odot 2}, \end{aligned} \quad (6.69)$$

where we used the Khatri-Rao product property given by (C.11). This result is stated in (6.53).

The dirty image vector implied by (6.52), $\boldsymbol{\sigma}_d = \mathbf{M}^H \mathbf{W}^H \mathbf{W} (\hat{\mathbf{r}}_{syn} - \mathbf{n})$, can also be written as a sum over the snapshots using the definition of \mathbf{r}_{syn} given by (3.76). This directly leads to (6.54). Note that we have omitted the hat on \mathbf{R}_m while replacing $\hat{\mathbf{r}}_{syn}$. This is done for notational convenience. This is allowable since in practical situations, $\widehat{\mathbf{R}}_m$ is used to compute the weighting matrix \mathbf{W} . This implies that \mathbf{R}_m in the first factor, originating from \mathbf{W} may be treated as being equal to \mathbf{R}_m in the second factor, originating from $\hat{\mathbf{r}}_{syn}$. If the snapshot noise covariance matrices are diagonal, we can exploit the fact that

$$\text{vec}(\boldsymbol{\Sigma}_{nm}) = \text{vec}(\text{diag}(\boldsymbol{\sigma}_{nm})) = (\mathbf{I} \circ \mathbf{I}) \boldsymbol{\sigma}_{nm}. \quad (6.70)$$

Noting that $\text{vec}(\mathbf{R}_m) = (\mathbf{I} \circ \mathbf{R}_m) \mathbf{1}$, we can reduce the Khatri-Rao products to

Hadamard products. This leads to (6.55) in a straightforward way:

$$\begin{aligned}\boldsymbol{\sigma}_d &= \sum_{m=1}^N \left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \circ \mathbf{R}_m^{-1} \mathbf{A}_{sm} \right)^H ((\mathbf{I} \circ \mathbf{R}_m) \mathbf{1} - (\mathbf{I} \circ \mathbf{I}) \boldsymbol{\sigma}_{nm}) \\ &= \sum_{m=1}^M \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \right) \mathbf{1} - \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \right) \boldsymbol{\sigma}_{nm}. \quad (6.71)\end{aligned}$$

6.A.2 Scalar imaging with unknown noise

The generic solution for this scenario is given by (6.50). This implicitly defines the deconvolution matrix as

$$\mathbf{M}_d = \mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} \mathbf{M} \quad (6.72)$$

and the dirty image vector as

$$\boldsymbol{\sigma}_d = \mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} \hat{\mathbf{r}}_{syn}, \quad (6.73)$$

where

$$\widetilde{\mathbf{W}} = \mathbf{W} - \mathbf{W} \mathbf{M}_n \left(\mathbf{M}_n^H \mathbf{W}^H \mathbf{W} \mathbf{M}_n \right)^{-1} \mathbf{M}_n^H \mathbf{W}^H \mathbf{W}. \quad (6.74)$$

In our derivation we will assume that the noise covariance matrix of the individual snapshots is diagonal with unknown elements on its main diagonal, i.e. $\boldsymbol{\Sigma}_{nm} = \text{diag}(\boldsymbol{\sigma}_{nm})$. This implies that $\mathbf{M}_{nm} = \mathbf{I}_P \circ \mathbf{I}_P$ and that $\mathbf{M}_n = \mathbf{I}_M \otimes (\mathbf{I}_P \circ \mathbf{I}_P)$ where \mathbf{I}_n denotes the $n \times n$ identity matrix.

We will first concentrate on finding a suitable expression for $\widetilde{\mathbf{W}}$. Since

$$\begin{aligned}\mathbf{W} \mathbf{M}_n &= \text{bdiag} \left(\overline{\mathbf{R}}_m^{-1/2} \otimes \mathbf{R}_m^{-1/2} \right) \text{bdiag} (\mathbf{I}_P \circ \mathbf{I}_P) \\ &= \text{bdiag} \left(\left(\overline{\mathbf{R}}_m^{-1/2} \otimes \mathbf{R}_m^{-1/2} \right) (\mathbf{I}_P \circ \mathbf{I}_P) \right) \\ &= \text{bdiag} \left(\overline{\mathbf{R}}_m^{-1/2} \circ \mathbf{R}_m^{-1/2} \right), \quad (6.75)\end{aligned}$$

where we have used (C.10) to reduce the Kronecker product. It follows that

$$\begin{aligned}\mathbf{M}_n^H \mathbf{W}^H \mathbf{W} \mathbf{M}_n &= \text{bdiag} \left(\overline{\mathbf{R}}_m^{-1/2} \circ \mathbf{R}_m^{-1/2} \right)^H \text{bdiag} \left(\overline{\mathbf{R}}_m^{-1/2} \circ \mathbf{R}_m^{-1/2} \right) \\ &= \text{bdiag} \left(\overline{\mathbf{R}}^{-1} \odot \mathbf{R}_m^{-1} \right), \quad (6.76)\end{aligned}$$

where we have used (C.11) to reduce the Khatri-Rao products to a Hadamard product.

Since

$$\begin{aligned}\mathbf{M}_n^H \mathbf{W}^H \mathbf{W} &= \text{bdiag} \left(\overline{\mathbf{R}}_m^{-1/2} \circ \mathbf{R}_m^{-1/2} \right)^H \text{bdiag} \left(\overline{\mathbf{R}}_m^{-1/2} \otimes \mathbf{R}_m^{-1/2} \right) \\ &= \text{bdiag} \left(\left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \right),\end{aligned}\quad (6.77)$$

$\widetilde{\mathbf{W}}$ is described by

$$\widetilde{\mathbf{W}} = \mathbf{W} - \text{bdiag} \left(\left(\overline{\mathbf{R}}_m^{-1/2} \circ \mathbf{R}_m^{-1/2} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \right). \quad (6.78)$$

We can now find an expression for $\widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}}$, a factor that appears in the expression for the deconvolution matrix as well as the expression for the dirty image vector.

$$\begin{aligned}\widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} &= \mathbf{W}^H \mathbf{W} + \\ &\quad - \mathbf{W}^H \text{bdiag} \left(\left(\overline{\mathbf{R}}_m^{-1/2} \circ \mathbf{R}_m^{-1/2} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \right) + \\ &\quad - \text{bdiag} \left(\left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-H} \left(\overline{\mathbf{R}}_m^{-1/2} \circ \mathbf{R}_m^{-1/2} \right)^H \right) \mathbf{W} + \\ &\quad + \text{bdiag} \left(\left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-H} \left(\overline{\mathbf{R}}_m^{-1/2} \circ \mathbf{R}_m^{-1/2} \right)^H \right) \times \\ &\quad \times \text{bdiag} \left(\left(\overline{\mathbf{R}}_m^{-1/2} \circ \mathbf{R}_m^{-1/2} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \right) \\ &= \text{bdiag} \left(\overline{\mathbf{R}}_m^{-1} \otimes \mathbf{R}_m^{-1} - \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \right).\end{aligned}\quad (6.79)$$

The next step to obtain expressions for the deconvolution matrix and the dirty image vector is to pre-multiply this factor with \mathbf{M}^H . This gives

$$\mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} = \left[\dots, \begin{array}{c} \left(\overline{\mathbf{A}}_{sm} \circ \mathbf{A}_{sm} \right)^H \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right) - \\ \left(\overline{\mathbf{A}}_{sm} \circ \mathbf{A}_{sm} \right)^H \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right) \times \\ \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \end{array}, \dots \right], \quad (6.80)$$

which reduces to

$$\begin{aligned} \mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} &= \\ &= \left[\cdots, \begin{array}{c} \left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \circ \mathbf{R}_m^{-1} \mathbf{A}_{sm} \right)^H - \\ \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H, \cdots \end{array} \right]. \end{aligned} \quad (6.81)$$

Post-multiplication with \mathbf{M} gives the deconvolution matrix

$$\begin{aligned} \mathbf{M}_d &= \sum_{m=1}^M \left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \circ \mathbf{R}_m^{-1} \mathbf{A}_{sm} \right)^H (\overline{\mathbf{A}}_{sm} \circ \mathbf{A}_{sm}) + \\ &\quad - \sum_{m=1}^M \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \times \\ &\quad \times \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H (\overline{\mathbf{A}}_{sm} \circ \mathbf{A}_{sm}) \\ &= \sum_{m=1}^M \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \odot \mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \mathbf{A}_{sm} \right) + \\ &\quad - \sum_{m=1}^M \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \times \\ &\quad \times \left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \odot \mathbf{R}_m^{-1} \mathbf{A}_{sm} \right). \end{aligned} \quad (6.82)$$

This can be simplified to (6.56).

In a similar way, we can find an expression for the dirty image vector by post-multiplying (6.81) with $\widehat{\mathbf{r}}_{syn}$, i.e.

$$\begin{aligned} \boldsymbol{\sigma}_d &= \left[\cdots, \begin{array}{c} \left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \circ \mathbf{R}_m^{-1} \mathbf{A}_{sm} \right)^H - \\ \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \right) \times \\ \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \end{array}, \cdots \right] \text{bdiag}(\mathbf{I} \circ \mathbf{R}_m) \mathbf{1}, \end{aligned} \quad (6.83)$$

which may also be formulated as

$$\begin{aligned}
\boldsymbol{\sigma}_d &= \sum_{m=1}^M \left(\left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{sm} \circ \mathbf{R}_m^{-1} \mathbf{A}_{sm} \right)^H - \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \right) \times \right. \\
&\quad \left. \times \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \right) (\mathbf{I} \circ \mathbf{R}_m) \mathbf{1} \\
&= \sum_{m=1}^M \left(\left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \right) \mathbf{1} + \right. \\
&\quad \left. - \left(\overline{\mathbf{A}}_{sm}^H \overline{\mathbf{R}}_m^{-1} \odot \mathbf{A}_{sm}^H \mathbf{R}_m^{-1} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \text{vecdiag}(\mathbf{R}_m^{-1}) \right), \tag{6.84}
\end{aligned}$$

which simplifies to (6.57).

6.A.3 Full polarized imaging with known or negligible noise

Section 3.6 shows that the full polarization data model can be cast in the same form as the scalar data model. The only difference is, that the visibility vector \mathbf{r}_{syn} , the image vector $\boldsymbol{\sigma} = [\mathbf{e}_1^T, \mathbf{e}_2^T, \dots, \mathbf{e}_Q^T]$ and the noise vector \mathbf{n} have four times as many elements, since each visibility or image point is now represented by four values. The solution to the imaging problem in the case of known or negligible noise can therefore still be expressed as (6.52), which implicitly defines a deconvolution matrix $\mathbf{M}_d = \mathbf{M}^H \mathbf{W}^H \mathbf{W} \mathbf{M}$ and a dirty image vector $\boldsymbol{\epsilon}_d = \mathbf{M}^H \mathbf{W}^H \mathbf{W} (\hat{\mathbf{r}}_{syn} - \mathbf{n})$, where we used $\boldsymbol{\epsilon}_d$ instead of $\boldsymbol{\sigma}_d$ to indicate that we are estimating a full polarized image vector.

With $\mathbf{M} = [\mathbf{M}_1^T, \mathbf{M}_2^T, \dots, \mathbf{M}_M^T]^T$, where the submatrices \mathbf{M}_m are given by (3.82), it follows that

$$\begin{aligned}
\mathbf{M}^H \mathbf{W}^H \mathbf{W} &= \left[\cdots, \text{bdiag}(\mathbf{B}_{qm})^H \mathbf{I}_{sm}^H \left[\begin{array}{c} \vdots \\ \left(\overline{\mathbf{A}}'_{qm} \otimes \mathbf{A}'_{qm} \right)^H \\ \vdots \end{array} \right], \cdots \right] \times \\
&\quad \times \text{bdiag} \left(\overline{\mathbf{R}}_m^{-1} \otimes \mathbf{R}_m^{-1} \right) \\
&= [\cdots, \mathbf{I}_{sm}^H \mathbf{M}_{1m}, \cdots], \tag{6.85}
\end{aligned}$$

where we have exploited the fact that \mathbf{I}_{sm} and $\text{bdiag}(\mathbf{B}_{qm})$ commute and we have

introduced

$$\begin{aligned} \mathbf{M}_{1m} &= \begin{bmatrix} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}_{qm}' \otimes \mathbf{A}_{qm}' \right)^H \\ \vdots \end{bmatrix} \text{bdiag} \left(\overline{\mathbf{R}}_m^{-1} \otimes \mathbf{R}_m^{-1} \right) \\ &= \begin{bmatrix} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}_{qm}'^H \overline{\mathbf{R}}_m^{-1} \otimes \mathbf{A}_{qm}'^H \mathbf{R}_m^{-1} \right) \\ \vdots \end{bmatrix}. \end{aligned} \quad (6.86)$$

Post-multiplication with \mathbf{M} gives the deconvolution matrix

$$\begin{aligned} \mathbf{M}_d &= [\cdots, \mathbf{I}_{sm}^H \mathbf{M}_{1m}, \cdots] \begin{bmatrix} \vdots \\ \mathbf{M}_m \\ \vdots \end{bmatrix} \\ &= \sum_{m=1}^M \mathbf{I}_{sm}^H \mathbf{M}_{1m} \mathbf{M}_m, \end{aligned} \quad (6.87)$$

which can be written as

$$\mathbf{M}_d = \sum_{m=1}^M \mathbf{I}_{sm}^H \begin{bmatrix} \ddots & & \vdots \\ \cdots & \mathbf{M}_{3q_1 q_2 m} & \cdots \\ & \vdots & \ddots \end{bmatrix} \mathbf{I}_{sm}, \quad (6.88)$$

where

$$\mathbf{M}_{3q_1 q_2 m} = \mathbf{B}_{q_1 m}^H \left(\overline{\mathbf{A}}_{q_1 m}'^H \overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}_{q_2 m}' \otimes \mathbf{A}_{q_1 m}'^H \mathbf{R}_m^{-1} \mathbf{A}_{q_2 m} \right) \mathbf{B}_{q_2 m}, \quad (6.89)$$

which was obtained using (C.2) and (C.4).

To obtain an expression for the dirty image vector ϵ_d , we have to multiply (6.85) with $\hat{\mathbf{r}}_{syn} - \mathbf{n}$. Noting that $\text{vec}(\mathbf{R}_m) = (\mathbf{I} \circ \mathbf{R}_m) \mathbf{1}$ and assuming a diagonal noise covariance matrix such that $\text{vec}(\Sigma_{nm}) = \text{vec}(\text{diag}(\sigma_{nm})) = (\mathbf{I} \circ \mathbf{I}) \sigma_{nm}$, this

gives

$$\begin{aligned}
\epsilon_d &= [\cdots, \mathbf{I}_{sm}^H \mathbf{M}_{1m}, \cdots] (\text{bdiag}(\mathbf{I} \circ \mathbf{R}_m) \mathbf{1} - \text{bdiag}(\mathbf{I}_P \circ \mathbf{I}_P) \boldsymbol{\sigma}_n) \\
&= \sum_{m=1}^M \mathbf{I}_{sm}^H \left[\begin{array}{c} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}_{qm}^H \overline{\mathbf{R}}_m^{-1} \circ \mathbf{A}_{qm}^H \right) \\ \vdots \end{array} \right] \mathbf{1} + \\
&\quad - \sum_{m=1}^M \mathbf{I}_{sm}^H \left[\begin{array}{c} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}_{qm}^H \overline{\mathbf{R}}_m^{-1} \circ \mathbf{A}_{qm}^H \mathbf{R}_m^{-1} \right) \\ \vdots \end{array} \right] \boldsymbol{\sigma}_{nm} \tag{6.90}
\end{aligned}$$

where we have used (C.10) to combine the Kronecker and Khatri-Rao products.

6.A.4 Full polarized imaging with unknown noise

The solution for imaging with unknown noise is given by (6.50), which implicitly defines the deconvolution matrix as $\mathbf{M}_d = \mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} \mathbf{M}$ and the dirty image vector as $\epsilon_d = \mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} \hat{\mathbf{r}}_{syn}$, where $\widetilde{\mathbf{W}}$ is a weighting matrix that filters the noise from the data while providing optimal weighting to ensure statistical efficiency. If we assume that $\Sigma_{nm} = \text{diag}(\boldsymbol{\sigma}_{nm})$, $\widetilde{\mathbf{W}}$ is given by (6.79).

With $\mathbf{M} = [\mathbf{M}_1^T, \mathbf{M}_2^T, \cdots, \mathbf{M}_M^T]^T$, where the submatrices \mathbf{M}_m are given by (3.82), it follows that

$$\begin{aligned}
\mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} &= \\
&= \left[\cdots, \mathbf{I}_{sm} \left[\begin{array}{c} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}_{qm}^H \otimes \mathbf{A}_{qm}^H \right)^H \\ \vdots \end{array} \right], \cdots \right] \times \\
&\quad \times \left(\text{bdiag} \left(\overline{\mathbf{R}}_m^{-1} \otimes \mathbf{R}_m^{-1} \right) + \right. \\
&\quad \left. - \text{bdiag} \left(\left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \right) \right). \tag{6.91}
\end{aligned}$$

For convenience of notation we introduce

$$\begin{aligned} \mathbf{M}_{2m} &= \begin{bmatrix} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}'_{qm} \otimes \mathbf{A}'_{qm} \right)^H \\ \vdots \end{bmatrix} \text{bdiag} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right) \\ &= \begin{bmatrix} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}'^H_{qm} \overline{\mathbf{R}}_m^{-1} \circ \mathbf{A}'^H_{qm} \mathbf{R}_m^{-1} \right) \\ \vdots \end{bmatrix}. \end{aligned} \quad (6.92)$$

Together with \mathbf{M}_{1m} introduced in the previous section, we can now write (6.91) as

$$\begin{aligned} \mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} &= \\ &= \left[\cdots, \mathbf{I}_{sm}^H \left(\mathbf{M}_{1m} - \mathbf{M}_{2m} \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \right), \cdots \right]. \end{aligned} \quad (6.93)$$

Post-multiplication with \mathbf{M} gives the deconvolution matrix

$$\begin{aligned} \mathbf{M}_d &= \\ &= \sum_{m=1}^M \mathbf{I}_{sm}^H \mathbf{M}_{1m} \mathbf{M} - \sum_{m=1}^M \mathbf{I}_{sm}^H \mathbf{M}_{2m} \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H \mathbf{M} \\ &= \sum_{m=1}^M \left(\mathbf{I}_{sm}^H \mathbf{M}_{1m} \mathbf{M} - \mathbf{I}_{sm}^H \mathbf{M}_{2m} \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \mathbf{M}_{2m}^H \mathbf{I}_{sm} \right) \end{aligned} \quad (6.94)$$

The first term is easily recognized as the deconvolution matrix in the noise free case given by (6.88). Introducing

$$\begin{aligned} \mathbf{M}_{4q_1 q_2 m} &= \mathbf{B}_{q_1 m}^H \left(\overline{\mathbf{A}}'^H_{q_1 m} \overline{\mathbf{R}}_m^{-1} \otimes \mathbf{A}'^H_{q_1 m} \mathbf{R}_m^{-1} \right) \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \times \\ &\quad \times \left(\overline{\mathbf{R}}_m^{-1} \overline{\mathbf{A}}'_{q_2 m} \otimes \mathbf{R}_m^{-1} \mathbf{A}'_{q_2 m} \right) \mathbf{B}_{q_2 m} \end{aligned} \quad (6.95)$$

and using $\mathbf{M}_{3q_1 q_2 m}$ defined in (6.89), we can formulate the expression for the deconvolution matrix as

$$\mathbf{M}_d = \sum_{m=1}^M \mathbf{I}_{sm}^H \begin{bmatrix} \ddots & & \vdots \\ \cdots & \mathbf{M}_{3q_1 q_2 m} - \mathbf{M}_{4q_1 q_2 m} & \cdots \\ & \vdots & \ddots \end{bmatrix} \mathbf{I}_{sm}. \quad (6.96)$$

Using $\mathbf{r}_{syn} = \text{bdiag}(\mathbf{I} \circ \mathbf{R}_m) \mathbf{1}$, an expression for the dirty image vector can be derived as follows:

$$\begin{aligned}
\epsilon_d &= \mathbf{M}^H \widetilde{\mathbf{W}}^H \widetilde{\mathbf{W}} \text{bdiag}(\mathbf{I} \circ \mathbf{R}_m) \mathbf{1} \\
&= \sum_{m=1}^M \mathbf{I}_{sm}^H \left[\begin{array}{c} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}_{qm}^H \overline{\mathbf{R}}_m^{-1} \otimes \mathbf{A}_{qm}^H \mathbf{R}_m^{-1} \right) \\ \vdots \end{array} \right] (\mathbf{I} \circ \mathbf{R}_m) \mathbf{1} + \\
&\quad - \sum_{m=1}^M \mathbf{I}_{sm}^H \mathbf{M}_{2m} \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \left(\overline{\mathbf{R}}_m^{-1} \circ \mathbf{R}_m^{-1} \right)^H (\mathbf{I} \circ \mathbf{R}_m) \mathbf{1} \\
&= \sum_{m=1}^M \mathbf{I}_{sm}^H \left[\begin{array}{c} \vdots \\ \mathbf{B}_{qm}^H \left(\overline{\mathbf{A}}_{qm}^H \overline{\mathbf{R}}_m^{-1} \circ \mathbf{A}_{qm}^H \right) \mathbf{1} \\ \vdots \end{array} \right] + \\
&\quad - \sum_{m=1}^M \mathbf{I}_{sm}^H \mathbf{M}_{2m} \left(\overline{\mathbf{R}}_m^{-1} \odot \mathbf{R}_m^{-1} \right)^{-1} \text{vecdiag}(\mathbf{R}_m^{-1}), \tag{6.97}
\end{aligned}$$

where the $\text{vecdiag}(\cdot)$ operator produces a vector containing the elements of the main diagonal of its argument.

Chapter 7

Conclusions and suggestions for further research

7.1 Summary of the main results

In this thesis, I have discussed the possibilities and limitations of imaging with an aperture array station assuming one has access to the signals from the individual receiving elements. I have shown how this may be applied in practice using data from LOFAR (prototype) stations in several examples. A nice example in which the theoretical limits were attained, was presented in Sec. 2.4 in which the variance of the image values in a fish-eye observation with a LOFAR station were compared with the CRB and found to show good quantitative agreement. Such results can only be obtained by choosing an appropriate data model, often referred to as measurement equation. A very generic model was derived in Ch. 3 based on an electromagnetic description of a phased array. This provides a clear view on the underlying assumptions. Commonly used data models are specializations of this generic model tailored to the problem at hand as discussed in 3.2.3. Chapter 3 also provides explicit extensions for modeling full polarized data and data from synthesis observations to stress that the assumed model does not restrict the methods discussed in this thesis to single polarization snapshot observations. This is also demonstrated by the all-sky map based on a synthesis observation with CS10 presented in Sec. 2.4 and the derivation of expressions for full polarization synthesis imaging in Secs. 6.A.3 and 6.A.4.

The CRB was used to determine the theoretical limits of the calibration and imaging problems. The required expressions were derived in Ch. 4. These expressions were used to confirm two intuitive properties of the complex valued receiver path

gain estimation problem. It was shown that for calibration on a single point source, a phase reference based on the average phase solution reduces the variance on the phase estimate by a factor $2P/(P - 1)$ compared to a phase reference based on a single element. The reference based on the average phase is also the constraint that minimizes the total variance on the estimated parameters. A similar conclusion holds for more general cases involving arbitrary source covariance models and arrays of non-identical elements. An intuition for this result is offered in Sec. 4.3.6.

The second result is that the variance on the direction independent gain estimates is the same regardless whether a particular entry of the array covariance matrix is affected by an additive nuisance parameter that has to be included in the estimation process or is simply ignored. This result was exploited in Ch. 5 which presents a direction independent gain estimation method that can ignore particular entries of the array covariance matrix and that seems to be asymptotically statistically efficient. Interestingly, it was also found that, even if the entry is ignored, knowledge of the value of that entry is still valuable. This was attributed to the fact that parameter estimates can often be improved by applying appropriate weights to the data. However, this effect is so small that it will hardly affect practical situations.

Aperture arrays provide an extremely wide FOV. In the case of a LOFAR LBA station, the FOV even spans the full hemisphere and thus enables observations that can aptly be called fish-eye observations. Calibration and imaging of such observations is very challenging due to the complex source structure on the sky, which includes a number of bright point sources and diffuse emission from our Galaxy. In chapter 5, a gain calibration method is presented that can handle a multi-source sky model while ignoring short baselines. This method shows asymptotically statistically efficient performance in Monte Carlo simulations, even with an SNR per receiving element per calibrator between 2.2 to 3.2, while having the same $O(P^3)$ numerical complexity as other asymptotically efficient algorithms in the literature that solve comparable calibration problems. Using a weighted alternating least squares approach, this method was extended to include estimation of apparent source powers, apparent source positions and a non-diagonal noise covariance matrix. These extensions do not increase the $O(P^3)$ numerical complexity, but more computations are required per step. It was demonstrated that the computational burden can be reduced with almost no loss in direction independent gain estimation accuracy by using the unweighted least squares solution to find the noise covariance matrix. With this modification, the 860-parameter calibration problem with actual LOFAR data presented in Sec. 2.3.3 could be solved in only 0.4 seconds using Matlab on a single core of a standard dual core 2.4 GHz CPU. Since LOFAR's station correlator produces at most one array covariance matrix per second, the algorithm has sufficient computational performance to keep up with the data and is therefore well suited for application in the station calibration pipeline described in Sec. 2.3.2.

Chapter 6 presented a closed form solution for least squares optimized snapshot imaging including deconvolution based on the data model presented in Ch. 3. This result allowed the first rigorous assessment of the effective noise floor, which is the combined effect of propagated calibration errors, thermal noise and source confusion in terms of the covariance of the image values. Effective noise is a more suitable quality measure for imaging performance than dynamic range, since the latter strongly depends on the flux of the strongest source in the field and is therefore highly observation dependent. The simulations for a URA presented in Sec. 6.2.3 indicate that the effects of propagated calibration errors are strongly concentrated at source locations, but are considerably smaller than the thermal noise at other image points. The concentration at source locations becomes evident if we assess the impact of direction dependent gain errors on the station beam pattern. The distortions in the power beam pattern are roughly proportional to the square root of this power beam pattern. For a typical 1% error in the direction independent gains, this causes relative errors of -27 dB in the main beam, -17 dB on a -20 dB side lobe and -5.5 dB on a -40 dB side lobe corresponding to absolute error levels of -27 dB, -37 dB and -45.5 dB relative to the main beam. The results also indicate that, if the stability of the instrument is such that the variations in the instrumental parameters between calibration measurements are comparable to or smaller than the accuracy with which they can be estimated during self-calibration, a separate calibration step is to be preferred over a self-calibrated image, since it allows better source separation in the imaging process.

The presented mathematical framework is sufficiently flexible to enable extension to synthesis observations, since the data model for a synthesis observation has the same mathematical form as the data model for a snapshot observation as demonstrated in Sec. 3.6. This becomes even more clear from the derivation of the least squares imaging expressions presented in Sec. 6.3.

The effects of deconvolution can be described by a deconvolution matrix that describes the amount of linear independence (orthogonality) of the spatial signature vectors weighted by the actual gains of the receiving elements. A diagonal deconvolution matrix not only ensures the best possible spatial separation between the sources, but also ensures a homogeneous noise distribution over the map. The deconvolution matrix condition number can be used to assess whether the deconvolution problem is well-posed.

Since a diagonal deconvolution matrix seems to be profitable, it is interesting to ask whether this matrix can be diagonalized by array design or by applying appropriate weights to the array elements. Since the deconvolution matrix is related to the array beam pattern, the latter is equivalent to finding weights that suppress the side lobe patterns at least in the direction of other sources, which implies that techniques like robust Capon beam forming could provide the desired weighting [117]. The physical interpretation of a diagonal deconvolution matrix is that the array has negli-

gibly low side lobes. An array with low side lobe levels is thus to be preferred from an imaging perspective. A low side lobe level is most easily achieved with a dense regular array, especially if an appropriate taper is applied. However, a regular array is also highly redundant and the analysis in this thesis suggests that an array without redundant spacings is more likely to approach the maximum number of solvable image points for a fixed number of antenna elements thereby allowing the system to reach the theoretical self-calibration confusion limit. If processing power is cheap compared to antenna hardware, a non-redundant array may be more attractive than a redundant array if the observations are confusion limited. A trade-off between these two aspects should thus be made based on the goals of the system.

The imaging expressions derived in Ch. 6 were applied to actual LOFAR data. The examples in Sec. 2.4 nicely demonstrate that the deconvolution is handled implicitly, that this imaging approach appears to be statistically efficient on actual data and that it can straightforwardly be applied to both time and frequency synthesis observations. These results and the derivations presented in Sec. 6.3 demonstrate a few attractive features of least squares imaging: the computational and memory requirements for the deconvolution step are independent of the amount of data, the method can handle arbitrarily large fields (up to 4π sr) and it can handle complex source structures like the Galactic plane and the North polar spur. This makes this method very suitable for reduction of observations with the cores of LOFAR and SKA aimed at full polarization imaging of the Galactic plane and detection of the omni-present signals from the epoch of reionization.

The latter conclusion corroborates the feeling that the main contribution of this thesis may be that it consistently uses methods from the array signal processing literature independent from the algorithms available in todays radio astronomical data reduction packages. This fundamentally new approach to radio astronomical data handling gradually attracts more attention from people working in this field and should form the basis for the next generation of self-calibration and imaging routines. However, the reduction of data from synthesis observations still requires answers to many fundamental and practical questions, which provides room for further research.

7.2 Suggestions for further research

7.2.1 Calibration

The calibration method described in Ch. 5 is based on least squares minimization of the difference between the measured and the modeled visibilities. The increase in station sensitivity towards Cas A due to calibration demonstrated by Fig. 2.15 indicates that calibration improves the performance of the station considerably. The

calibration error propagation equations presented in Ch. 6 suggest that this is achieved by increased sensitivity in the pointing direction and improved pointing accuracy. The fact that calibration improves the image quality begs the following questions:

- Is there a cost function based on some measure of image quality, e.g. contrast or total power optimization, which can be used to calibrate the telescope?
- Does such a technique allow us to calibrate the telescope without providing a source model, possibly with the exception of a systematic pointing offset?
- Does such a technique work in general or are there limitations in terms of source structure (e.g. the presence of point sources), solvable instrument parameters (e.g. knowledge of the antenna positions may be required), array configuration (e.g. availability of redundant baselines) or other constraints?

Calibration techniques that do not require knowledge of the available calibrators may be very valuable for future space based instruments aimed at the frequency range below 10 MHz. At these frequencies, our knowledge on the universe in general and standard calibration sources in particular is fairly limited. Such instruments are currently in the phase of concept and feasibility studies [9, 35, 54, 62, 154].

One such technique already exists: redundancy calibration [136]. This method exploits redundant baselines in the array which measure the same visibility, i.e. the same spatial frequency. This method proved to be very valuable for the highly redundant WSRT [80]. The LOFAR HBA station arrays and the SKA phased array stations envisaged for frequencies below 1.5 GHz are highly redundant as well. The following questions are thus timely:

- Since redundancy calibration only exploits redundant baselines, it does not use all available information. How much calibration accuracy is lost by ignoring this information?
- Phased array receiving elements are subject to far more electromagnetic coupling than the feeds in the WSRT dishes. Can redundancy still be exploited for the calibration of phased arrays?

The ability of phased arrays to detect Galactic neutral hydrogen was demonstrated using the Thousand Element Array (ThEA), an SKA pathfinder built by ASTRON, in 2002 – 2003 [56, 141]. These phased array tiles were calibrated using a holographic measurement method in which the analog beam former in each tile is used to scan over the sky forming different superpositions of antenna signals while measuring the complex valued voltage output of the tile using a well defined reference signal [47]. Originally, these measurements were done in an anechoic chamber, but this method was demonstrated to work in the field as well if a strong calibrator is available [137]. The applicability of this method is severely limited by the required availability of a single dominant point source. This begs the question

- How can this tile calibration method be modified to allow a multi-source sky model? This is highly desirable in view of the LOFAR HBA stations and the envisaged SKA phased array stations.

The multi-source calibration algorithm described in Ch. 5 is a self-calibration algorithm that assumes that the number of calibrators is *a priori* known. If the uncalibrated array has sufficient performance to form an interpretable, but probably very messy, image using a DFT imager, the number of sources and initial estimates for their locations and powers may be extracted from the data. Based on this intuition, I pose the following questions:

- If the number of available calibration sources is not *a priori* known, how can the self-calibration methods presented in this thesis be modified to include order estimation of the source model?
- If such an extension is possible, what are its practical limitations in terms of the SNR of the sources, the number of sources or other restrictions?

Good candidates for estimation of the number of sources are the exponential fitting test (EFT) [91, 92], the minimum description length (MDL) and the Akaike information criterion (AIC) [131]. The minimization of either the AIC or MDL criterion proposed in [131] can, in principle, be extended to a combined detection and source parameter estimation solution. This solution, however, is highly non-linear and computationally expensive. It may therefore be more attractive to use the EFT or the AIC to obtain a, probably too high, estimate of the number of sources and then use the consistency of the source parameter estimates in consecutive iterations of the WALS approach to find the true number of usable calibrators.

7.2.2 Imaging

In the least squares imaging method described in Sec. 6.3, the condition number of the deconvolution matrix signifies whether the imaging problem is well-posed. The physical interpretation of “well-posed” is the independency of the measured visibility points. This is nicely illustrated by the aforementioned conclusion that an irregular array is able to get closer to the theoretical self-calibration confusion limit than a regular, and therefore redundant, array. There are still a number of open issues regarding the physical significance of the deconvolution matrix:

- What does independency between visibilities mean? For example, even if visibilities are strongly mutually dependent, they may still contribute distinct information to the image if they have sufficient SNR.
- Does the deconvolution matrix also provide information on the presence of gaps in the visibility domain, i.e. on unmeasured visibilities?

- If so, how does it deal with it?

Based on this physical insight regarding the relation between the deconvolution matrix condition number, dependency between visibilities and SNR, it is expected that an optimal imaging result can only be obtained by a data dependent image grid. Such a grid will probably provide high resolution in regions with bright extended emission or near strong point sources and low resolution in the “empty” parts of the image. The method as described in this thesis requires manual optimization of the grid by imaging the data several times while modifying the grid between iterations. This is a rather cumbersome process, so this begs the following question:

- Can the least squares imaging method be adapted in such a way that it determines a suitable grid automatically based on the data itself?

This may be considered a search for appropriate base functions to describe the image, a problem for which several methods, like the Karhunen-Loève expansion [75], are readily available.

It is interesting to note that the least squares synthesis all-sky map based on CS10 data shown in Fig. 2.18 was made at a resolution of $0.75\lambda_0/D$ where λ_0 is the minimum wavelength. This resolution is higher than expected based on the diffraction limit of a uniformly weighted aperture. As mentioned above, this superresolution is possible due to the high SNR of the data. This effect is well-known from high resolution DOA estimation techniques, but in this case we image the entire sky, i.e. not a few distinct sources. During the experiments it was also noted that the resolution limit can be pushed even further by applying regularization to the deconvolution matrix at the expense of a bias in source power estimation. These observations lead to the following questions:

- What are the theoretical and practical resolution limits of least squares image estimation?
- What are these limits if regularization is exploited?
- Is it possible to predict and correct the bias introduced by regularization?

From a practical perspective, we note that least squares imaging up to 512×512 -pixel images is tractable, but that larger images require excessive memory and processing resources. A full polarization 2048×2048 -pixel image, for example, would involve a $4096^2 \times 4096^2$ deconvolution matrix requiring 2.05 Pbyte of memory and about 4.72 Zflop for inversion. Fortunately, least squares imaging is particularly suited for imaging of the complex diffuse emission that will be examined in studies on Galactic magnetism and the epoch of reionization. This diffuse emission will typically be observed with the LOFAR and SKA cores that have a low resolution limit. The images produced by the full LOFAR and SKA arrays will typically consist

of many discrete sources, i.e. will typically be sparse. Since least squares optimization is not well-suited for sparse data, other methods are probably more suitable for these images, which will have typical sizes of 4096×4096 pixels or larger. These imaging challenges still require further study:

- Can the properties and physical significance of the deconvolution matrix be exploited to improve the computational efficiency of least squares imaging?
- Can these insights be used to expand the applicability of current techniques like CLEAN to temporally, spatially and spectrally varying beam patterns?
- What can we learn from alternative imaging techniques like Least Squares Minimum Variance Imaging [7, 67], l_1 -imaging [66] and compressed sensing [100]?

7.3 Towards SKA

Although this research was primarily conducted for the LOFAR project, its results are relevant steps towards data processing for the SKA as well. SKA is envisaged as an array of stations that may consist of aperture arrays, dishes with single pixel feeds and dishes with phased array feeds, often referred to as focal plane arrays (FPA). This implies that the full beam forming hierarchy depicted in Fig. 1.5 will be present with corresponding levels of calibration. The calibration studies done for LOFAR have resulted in a thorough treatment of calibration scenarios 3 and 4 depicted in Fig. 5.1. Scenario 3 is discussed in Ch. 5 of this thesis while scenario 4 is discussed in [120] and the Ph.D. thesis of Sebastiaan van der Tol [121]. The data model described in Ch. 3 is easily adapted to a tile-based architecture as demonstrated in [166] and can therefore be applied to FPAs as well. A nice introduction to FPA signal processing can be found in [53]. Although there are still a number of open issues, as mentioned above, we may thus conclude that considerable progress on all aspects of SKA calibration has been made over the last years, particularly in the framework of the LOFAR project.

The framework describing the fundamental imaging limits in Ch. 6 is applicable to any radio telescope, thus also to SKA. SKA is still in its design phase. The framework laid down in Ch. 6 can be used to analyze the calibration challenges and expected image quality of different design choices thus providing a nice addition to the set of tools used to make and support the SKA design decisions. This analysis also led to a study on least squares imaging, a technique that may prove useful for (all-sky) imaging with the SKA stations and the SKA core.

Apart from these direct contributions to the solution of SKA calibration and imaging challenges, signal processing techniques also provide a proper framework for quantitative assessment of the quality of the results. We may thus safely conclude

that signal processing techniques have the potential to revolutionize the way astronomical data is handled on the path towards SKA.

Appendix A

Abbreviations and acronyms

3C	third Cambridge (catalog)
ADC	analog-to-digital converter
AIC	Akaike information criterion
ALS	alternating least squares
ASTRON	Netherlands Institute for Radio Astronomy
CEP	central processing facility
CLEAN	iterative beam removing technique [50, 102]
CPU	central processing unit
CRB	Cramer-Rao bound
CS10	Core Station 10
DFT	direct Fourier transform
DOA	direction of arrival
EFT	exponential fitting test
EM	electromagnetic
EMBRACE	Electronic Multi-beam Radio Astronomy Concept
ESPRIT	estimation of signals via rotational invariance techniques
EoR	epoch of reionization
FFT	fast Fourier transform
FIM	Fisher information matrix
FOV	field-of-view
FPA	focal plane array
GMRT	Giant Meter-wave Radio Telescope
HBA	high band antenna
HPBW	half power beam width
i.i.d.	independent identically distributed

ITS	Initial Test Station
KASCADE	Karlsruhe Shower Core and Array Detector
LBA	low band antenna
LCU	local control unit
LOFAR	Low Frequency Array
LOPES	LOFAR Prototype Station
LWA	Long Wavelength Array
MDL	minimum description length
MEM	maximum entropy method
ML	maximum likelihood
MLE	maximum likelihood estimator
MUSIC	multiple signal classification
MVDR	minimum variance distortionless response
MWA	Murchison Widefield Array
PC	personal computer
PCI	peripheral component interconnect
pdf	probability density function
psf	point spread function
RAM	random access memory
RCB	robust Capon beam former
RCU	receiver unit
RFI	radio frequency interference
RMS	root mean square
RSP	remote station processing
SKA	Square Kilometre Array
SNR	signal-to-noise ratio
ThEA	Thousand Element Array
THETA	Ten Heterogeneous Element Test Array
TIM	twin input module
UHECR	ultra-high energy cosmic ray
ULA	uniform linear array
URA	uniform rectangular array
VLA	Very Large Array
VLBI	Very Long Baseline Interferometry
WALS	weighted alternating least squares
WLS	weighted least squares
WSF	weighted subspace fitting
WSRT	Westerbork Synthesis Radio Telescope
xALS	extended alternating least squares
xWALS	extended weighted alternating least squares

Appendix B

Notation

B.1 Symbols

a, A	italic lower or upper case characters denote scalars
\mathbf{a}	bold lower case characters denote column vectors
\mathbf{A}	bold upper case characters denote matrices
a_i	element i from vector \mathbf{a}
A_{ij}	element (i, j) from matrix \mathbf{A}
\mathbf{I}	identity matrix of appropriate size
\mathbf{I}_n	$n \times n$ identity matrix
\mathbf{I}_s	selection matrix
\mathbf{E}_{ij}	elementary matrix with entry (i, j) equal to 1 and all others set to zero
\mathbf{e}_i	unit vector with entry i equal to 1 and all others set to zero
$\mathbf{1}$	vector or matrix of appropriate size filled with ones
$\mathbf{1}_P$	$P \times 1$ column vector filled with ones
$\mathbf{1}_{P \times Q}$	$P \times Q$ matrix filled with ones
$\mathbf{0}$	vector or matrix of appropriate size filled with zeros
$\mathbf{0}_P$	$P \times 1$ column vector filled with zeros
$\mathbf{0}_{P \times Q}$	$P \times Q$ matrix filled with zeros
$\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$	unit vectors in the (x, y, z) -coordinate system
j	negative square root of -1
e	Euler's number, 2.71828...

B.2 Operators

$(\cdot)^T$	transpose
$(\cdot)^H$	Hermitian or conjugate transpose
$(\cdot)^\dagger$	pseudo-inverse
$(\cdot)^{1/2}$	symmetric matrix square root
$(\cdot)^{\odot n}$	element-wise matrix exponent
$e^{(\cdot)}, \exp(\cdot)$	element-wise application of the exponential function
$\bar{\cdot}$	conjugation
$\text{vec}(\cdot)$	vector formed by stacking the columns of its matrix argument
$\text{vec}_-(\cdot)$	as $\text{vec}(\cdot)$ but leaving out the elements on the main diagonal
$\text{vecdiag}(\cdot)$	vector formed from the main diagonal of its argument
$\text{diag}(\cdot)$	diagonal matrix with the vector argument on the main diagonal
$\text{bdiag}(\cdot)$	block diagonal matrix with the matrices described by its argument on the main diagonal
\cdot	inner product of two vectors
\circledast	circular convolution of two vectors
\odot	Hadamard or element-wise product of two matrices or vectors
\oslash	element-wise division of two matrices or vectors
\otimes	Kronecker product of two matrices or vectors
\circ	Khatri-Rao or column-wise Kronecker product of two matrices
$[\cdot]_{ij}$	(i, j) th element from matrix or vector argument
$\text{Re}\{\cdot\}$	real part
$\text{Im}\{\cdot\}$	imaginary part
$\ \cdot\ _F$	Frobenius norm
$\ \cdot\ $	Euclidian norm
$ \cdot $	absolute value of a scalar
	length of a vector
	determinant of a matrix
$\text{rank}(\cdot)$	rank of a matrix
$\text{tr}(\cdot)$	trace of a matrix
$\mathcal{FT}(\cdot)$	Fourier transform
$\text{circulant}(\cdot)$	circulant matrix having the vector argument as first row
$\mathcal{E}\{\cdot\}$	expectation operator
$\hat{\cdot}$	expected value
$\text{cov}(\cdot)$	covariance of the elements of its vector argument
$\arg\min_{\boldsymbol{\theta}} \kappa(\boldsymbol{\theta})$	minimize the cost function $\kappa(\boldsymbol{\theta})$ w.r.t. parameter vector $\boldsymbol{\theta}$
$\mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$	i.i.d. complex Gaussian noise with mean $\boldsymbol{\mu}$ and covariance \mathbf{R}
$\mathcal{CW}(\mathbf{R}, N)$	complex Wishart distribution with N degrees of freedom and covariance \mathbf{R}

Appendix C

Matrix product relations

The properties of the Kronecker, Khatri-Rao and Hadamard products are frequently exploited. This appendix presents an overview of the definition of these products and lists a number of important relations.

The Kronecker product \otimes of an $N \times M$ matrix \mathbf{A} and a $P \times Q$ matrix \mathbf{B} produces the $NP \times MQ$ matrix

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \cdots & A_{1M}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \cdots & A_{2M}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N1}\mathbf{B} & A_{N2}\mathbf{B} & \cdots & A_{NM}\mathbf{B} \end{bmatrix}. \quad (\text{C.1})$$

For matrices \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{E} of appropriate size, the Kronecker product has the following useful properties:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{E}) = \mathbf{AC} \otimes \mathbf{BE} \quad (\text{C.2})$$

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \quad (\text{C.3})$$

$$(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H \quad (\text{C.4})$$

The Khatri-Rao product \circ is the column-wise Kronecker product. The Khatri-Rao product of an $N \times M$ matrix \mathbf{A} and a $P \times M$ matrix \mathbf{B} produces the $NP \times M$ matrix

$$\mathbf{A} \circ \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \mathbf{a}_2 \otimes \mathbf{b}_2, \dots, \mathbf{a}_M \otimes \mathbf{b}_M], \quad (\text{C.5})$$

where \mathbf{a}_m and \mathbf{b}_m denote the m th columns of \mathbf{A} and \mathbf{B} respectively. The Khatri-Rao product can be very useful to vectorize matrix products, for example

$$\text{vec}(\mathbf{ADB}) = (\mathbf{B}^T \circ \mathbf{A}) \mathbf{d}, \quad (\text{C.6})$$

where $\mathbf{D} = \text{diag}(\mathbf{d})$.

The Hadamard product \odot is the entry-wise product of two matrices. The Hadamard product of two $N \times M$ matrices \mathbf{A} and \mathbf{B} is thus given by

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} A_{11}B_{11} & A_{12}B_{12} & \cdots & A_{1M}B_{1M} \\ A_{21}B_{21} & A_{22}B_{22} & \cdots & A_{2M}B_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N1}B_{N1} & A_{N2}B_{N2} & \cdots & A_{NM}B_{NM} \end{bmatrix} \quad (\text{C.7})$$

For matrices \mathbf{A} and \mathbf{B} and diagonal matrices \mathbf{D} with appropriate size, the Hadamard product has the following properties:

$$(\mathbf{AD}) \odot \mathbf{B} = \mathbf{A} \odot (\mathbf{BD}) \quad (\text{C.8})$$

$$\mathbf{A} \odot (\mathbf{BD}) = (\mathbf{A} \odot \mathbf{B}) \mathbf{D} \quad (\text{C.9})$$

The Kronecker and Khatri-Rao products tend to produce very large matrices. The following relations, applicable to matrices of appropriate size, may help to reduce a Kronecker product to a Khatri-Rao product and a Khatri-Rao product to a Hadamard product:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \circ \mathbf{E}) = \mathbf{AC} \circ \mathbf{BE} \quad (\text{C.10})$$

$$(\mathbf{A} \circ \mathbf{B})^H (\mathbf{C} \circ \mathbf{E}) = \mathbf{A}^H \mathbf{C} \odot \mathbf{B}^H \mathbf{E} \quad (\text{C.11})$$

References

Several references are hard to find of one does not have access to specific resources. Interested readers should feel free to contact the author regarding such references.

- [1] A. van Ardenne, P.N. Wilkinson, P.D. Patel and J.G. bij de Vaate. Electronic Multi-Beam Radio Astronomy Concept: Embrace a Demonstrator for the European SKA Program. *Experimental Astronomy*, 17(1-3):65–77, 2004.
- [2] A. van Ardenne, J.D. Bregman, W.A. van Cappellen, G.W. Kant and J.G. bij de Vaate. Extending the Field of View With Phased Array Techniques: Results of European SKA Research. *Proceedings of the IEEE*, 97(8):1531–1542, August 2009.
- [3] D. Astély, A.L. Swindlehurst and B. Ottersten. Spatial Signature Estimation for Uniform Linear Arrays with Unknown Receiver Gains and Phases. *IEEE Transactions on Signal Processing*, 47(8):2128–2138, August 1999.
- [4] ASTRON - Netherlands Institute for Radio Astronomy. online: <http://www.astron.nl>.
- [5] W.J. Bangs. *Array Processing with Generalized Beamformers*. Ph.D. thesis, Yale University, New Haven, USA, 1971.
- [6] M.S. Bartlett. An Inverse Matrix Adjustment Arising in Discriminant Analysis. *Annals of Mathematical Statistics*, 22(1):107–111, March 1951.
- [7] Chen Ben-David and Amir Leshem. Parametric High Resolution Techniques for Radio Astronomical Imaging. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):670–684, October 2008.
- [8] A.S. Bennett. The Revised 3C Catalog of Radio Sources. *Memoirs of the Royal Astronomical Society*, 68:163–172, 1962.

- [9] M.J. Bentum, C.J.M. Verhoeven, A.J. Boonstra, A.J. van der Veen and E.K.A. Gill. A Novel Astronomical Application for Formation Flying Small Satellites. In *60th International Astronautical Congress*, Daejeon, Korea, 12-16 October 2009.
- [10] Philip R. Bevington and D. Keith Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw Hill, second edition, 1992.
- [11] J.F. Böhme and D. Kraus. On Least Squares Methods for Direction of Arrival Estimation in the Presence of Unknown Noise Fields. In *Proceedings of the IEEE 1988 International Conference on Acoustics, Speech and Signal Processing (ICASSP-88)*, volume 5, pages 2833–2836, New York (NY), USA, 11-14 April 1988.
- [12] Albert-Jan Boonstra and Alle-Jan van der Veen. Gain Calibration Methods for Radio Telescope Arrays. *IEEE Transactions on Signal Processing*, 51(1):25–38, January 2003.
- [13] A.J. Boonstra, S.J. Wijnholds, S. van der Tol and B. Jeffs. Calibration, Sensitivity and RFI Mitigation Requirements for LOFAR. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia (Penn.), USA, 18-23 March 2005.
- [14] Albert-Jan Boonstra. *Radio Frequency Interference Mitigation in Radio Astronomy*. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 2005.
- [15] J.D. Bregman. Design Concepts for a Sky Noise Limited Low Frequency Array. In A.B. Smolders and M.P. van Haarlem, editors, *Perspectives on Radio Astronomy: Technologies for Large Antenna Arrays*, pages 23–32, Dwingeloo, The Netherlands, 12-14 April 1999.
- [16] Jaap D. Bregman. System Optimization of Multi-beam Aperture Synthesis Arrays for Survey Performance. *Experimental Astronomy*, 17(1-3):365–380, 2004.
- [17] J.D. Bregman et al. LOFAR Synthesis Array and Station Configuration. *Experimental Astronomy*, 2010. in preparation (invited paper).
- [18] David R. Brillinger. *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, Inc., 1975.

- [19] P.C. Broekema. Numerical analysis of the LOFAR remote station beamformer. Master's thesis, University of Groningen, Groningen, The Netherlands, 16 May 2003. Also available as LOFAR-ASTRON-MEM-136.
- [20] John D. Bunton and Stuart G. Hay. SKA Cost Model for Wide Field-of-View Options. *Experimental Astronomy*, 17(1-3):5–16, 2004.
- [21] E.C. Button. Bayesian Interferometric Imaging. In *Proceedings of the XXIXth General Assembly of the International Union of Radio Science (URSI GA)*, Chicago (Ill.), USA, 7-16 August 2008.
- [22] Emmanuel J. Candès and Michael B. Wakin. An Introduction to Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008.
- [23] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, August 1969.
- [24] E. de Carvalho, J. Cioffi and D. Slock. Cramér-Rao Bounds for Blind Multichannel Estimation. In *IEEE Global Telecommunications Conference 2000 (GLOBECOM '00)*, volume 2, pages 1036–1040, 27 November - 1 December 2000.
- [25] Q. Cheng, Y. Hua and P. Stoica. Asymptotic Performance of Optimal Gain-and-Phase Estimators of Sensor Arrays. *IEEE Transactions on Signal Processing*, 48(12):3587–3590, December 2000.
- [26] Tim Cornwell and Edward B. Fomalont. Self-Calibration. In R.A. Perley, F.R. Schwab and A.H. Bridle, editors, *Synthesis Imaging in Radio Astronomy*, volume 6 of *Astronomical Society of the Pacific Conference Series*. BookCrafters Inc., 1994.
- [27] T. Cornwell, R. Braun and D.S. Briggs. Deconvolution. In G.B. Taylor, C.L. Carilli and R.A. Perley, editors, *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, pages 151–170. Astronomical Society of the Pacific, San Francisco (Calif.), USA, 1999.
- [28] T.J. Cornwell, K. Golap and Sanjay Bhatnagar. w -projection: A New Algorithm for Wide Field Imaging with Radio Synthesis Arrays. In P.L. Shopbell, M.C. Britton and R. Ebert, editors, *Astronomical Data Analysis Software & Systems (ADASS) XIV*, volume 347 of *Astronomical Society of the Pacific Conference Series*, pages 86–90. Astronomical Society of the Pacific, San Francisco (Calif.), USA, 2005.

- [29] Tim J. Cornwell, Kumar Golap and Sanjay Bhatnagar. The Noncoplanar Baselines Effect in Radio Interferometry: The W-Projection Algorithm. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):647–657, October 2008.
- [30] Tim J. Cornwell. Multiscale CLEAN Deconvolution of Radio Synthesis Images. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):793–801, October 2008.
- [31] P.E. Dewdney, P.J. Hall, R.T. Schilizzi and T.J.L.W. Lazio. The Square Kilometre Array. *Proceedings of the IEEE*, 97(8):1482–1496, August 2009.
- [32] Alan Edelman. *Eigenvalues and Condition Numbers of Random Matrices*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge (Mass.), USA, May 1989.
- [33] S.W. Ellington et al. The Long Wavelength Array. *Proceedings of the IEEE*, 97(8):1421–1430, August 2009.
- [34] H. Falcke et al. (LOPES collaboration). Detection and imaging of atmospheric radio flashes from cosmic ray air showers. *Nature*, 435:313–316, 19 May 2005.
- [35] H. Falcke and S. Jester. Moon-Based Ultra-Long-Wavelength Radio Astronomy: Science Case and Technology Outlook. In *Proceedings of the XXIXth General Assembly of the International Union of Radio Science (URSI GA)*, Chicago (Ill.), USA, 7–16 August 2008.
- [36] B.P. Flanagan and K.L. Bell. Array Self-Calibration with Large Sensor Position Errors. In *Conference Record of the 33rd IEEE Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 258–262, 1999.
- [37] Brian P. Flanagan and Kristine L. Bell. Array Self Calibration with Large Sensor Position Errors. *Signal Processing*, 81(10):2201–2214, October 2001.
- [38] Benjamin Friedlander and Anthony J. Weiss. Direction Finding Using Noise Covariance Modeling. *IEEE Transactions on Signal Processing*, 43(7):1557–1567, July 1995.
- [39] Daniel R. Fuhrmann. Estimation of Sensor Gain and Phase. *IEEE Transactions on Signal Processing*, 42(1):77–87, January 1994.
- [40] G. Battaglia, F. Fraternali, T. Oosterloo and R. Sancisi. HI study of the warped spiral galaxy NGC 5055: a disk/dark matter halo offset? *Astronomy & Astrophysics*, 447:49–62, February 2006.

- [41] G.H. Golub and C.F. van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore (Md.), USA, 1984.
- [42] Bo Göransson and Björn Ottersten. Direction Estimation in Partially Unknown Noise Fields. *IEEE Transactions on Signal Processing*, 47(9):2375–2385, September 1999.
- [43] J.D. Gorman and A.O. Hero. Lower Bounds for Parametric Estimation with Constraints. *IEEE Transactions on Information Theory*, 26(6):1285–1301, November 1990.
- [44] P.J. Hall. The Square Kilometre Array: an International Engineering Perspective. *Experimental Astronomy*, 17(1-3):5–16, 2004.
- [45] 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.
- [46] 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.
- [47] Grant A Hampson and A. Bart Smolders. A Fast and Accurate Scheme for Calibration of Active Phased Array Antennas. In *IEEE AP-S International Symposium and USNC/URSI National Radio Science Meeting*, July 1999.
- [48] Chad K. Hansen, Karl F. Warnick, Brian D. Jeffs, J. Richard Fisher and Richard Bradley. Interference Mitigation Using a Focal Plane Array. *Radio Science*, 40(5), June 2005.
- [49] C.G.T. Haslam, C.J. Salter, H. Stoffel and W.E. Wilson. A 408 MHz all-sky continuum survey II - The atlas of contour maps. *Astronomy and Astrophysics Supplement Series*, January 1982.
- [50] J.A. Högblom. Aperture Synthesis with a Non-regular Distribution of Interferometer Baselines. *Astronomy & Astrophysics Supplement*, 15:417–426, 1974.
- [51] H. Intema. *A sharp view on the low frequency radio sky*. Ph.D. thesis, Leiden University, Leiden, The Netherlands, August 2009.
- [52] Brian D. Jeffs, Sebastiaan van der Tol and Alle-Jan van der Veen. Direction Dependent Self Calibration of Large Distributed Sensor Arrays. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal*

- Processing (ICASSP)*, volume 4, pages 1069–1072, Toulouse, France, 14-19 May 2006.
- [53] Brian D. Jeffs, Karl F. Warnick, Jonathan Landon, Jacob Waldron, David Jones, J. Richard Fisher and Roger D. Norrod. Signal Processing for Phased Array Feeds in Radio Astronomical Telescopes. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):635–646, October 2008.
 - [54] D. Jones et al. The Astronomical Low Frequency Array: A Proposed Explorer Mission for Radio Astronomy. In Robert G. Stone, Kurt W. Weiler, Melvyn L. Goldstein and Jean-Louis Bougeret, editors, *Radio Astronomy at Long Wavelengths*, volume 119 of *Geophysical Monograph Series*. American Geophysical Union, Washington (D.C.), USA, 2000.
 - [55] D. Jones. SKA Science Requirements, Version 2. Technical Report SKA memo 45, SKA Project, 26 February 2004. online: <http://www.skatelescope.org>.
 - [56] G.W. Kant, A.W. Gunst, A.B.J. Kokkeler and A.B. Smolders. Receiver Architecture of the Thousand Element Array (THEA). In *Proceedings of the IEEE SPIE Conference*, volume 4015, Munich, Germany, 2000.
 - [57] N.E. Kassim. Towards the Long Wavelength Array. In *Proceedings of the XXIXth General Assembly of the International Union of Radio Science (URSI GA)*, Chicago (Ill.), USA, 7-16 August 2008.
 - [58] Steven M. Kay. *Fundamentals of Statistical Signal Processing - Volume I: Estimation Theory*. Prentice Hall Signal Processing Series. Prentice Hall, 1993.
 - [59] John D. Kraus. *Radio Astronomy*. Cygnus-Quasar Books, 2nd edition, 1986.
 - [60] S.R. Kulkarni. Self-Noise in Interferometers: Radio and Infrared. *Astronomical Journal*, 98(3):1112–1130, September 1989.
 - [61] Maria Lanne. *Phased array calibration and beamforming using signal processing*. Ph.D. thesis, Chalmers University of Technology, Göteborg, Sweden, 2007.
 - [62] J. Lazio. The Dark Ages Lunar Interferometer. In *Proceedings of the XXIXth General Assembly of the International Union of Radio Science (URSI GA)*, 7-16 August 2008.
 - [63] Alberto Leon-Garcia. *Probability and Random Processes for Electrical Engineering*. Addison-Wesley, Reading (MA), USA, 1994.

- [64] Amir Leshem and Alle-Jan van der Veen. Radio-Astronomical Imaging in the Presence of Strong Radio Interference. *IEEE Transactions on Information Theory*, 46(5):1730–1747, August 2000.
- [65] Amir Leshem and Alle-Jan van der Veen. Multichannel Detection of Gaussian Signals with Uncalibrated Receivers. *IEEE Signals Processing Letters*, 8(4):120–122, April 2001.
- [66] R. Levanda and A. Leshem. Radio astronomical image formation using sparse reconstruction techniques. In *Proceedings of the IEEE 25th Convention of Electrical and Electronics Engineers in Israel*, pages 716–720, December 2008.
- [67] Ronny Levanda and Amir Leshem. Synthetic Aperture Radio Telescopes. *IEEE Signal Processing Magazine*, 27(1):14–29, January 2010.
- [68] H. Li, P. Stoica and J. Li. Computationally Efficient Maximum Likelihood Estimation of Structured Covariance Matrices. *IEEE Transactions on Signal Processing*, 47(5):1314–1323, May 1999.
- [69] Jian Li, P. Stoica and Zhisong Wang. On Robust Capon Beamforming and Diagonal Loading. *IEEE Transactions on Signal Processing*, 51(7):1702–1715, July 2003.
- [70] LOFAR project. online: <http://www.lofar.org>.
- [71] C. Lonsdale. Calibration approaches. Technical Report LFD memo 015, MIT Haystack, December 8, 2004.
- [72] C. Lonsdale. The Murchison Widefield Array. In *Proceedings of the XXIXth General Assembly of the International Union of Radio Science (URSI GA)*, Chicago (Ill.), USA, 7-16 August 2008.
- [73] C.J. Lonsdale et al. The Murchison Widefield Array: Design Overview. *Proceedings of the IEEE*, 97(8):1497–1506, August 2009.
- [74] Marykutty Michael. Jupiter and Io: The unique celestial couple. *Europhysics News*, 34(3):85–88, May / June 2003.
- [75] Todd K. Moon and Wynn C. Stirling. *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, Upper Saddle River (New Jersey), 2000.

- [76] M.F. Morales. Optimal Map Making and the Analysis of the Next Generation Interferometric Data Sets. In *Proceedings of the XXIXth General Assembly of the International Union of Radio Science (URSI GA)*, Chicago (Ill.), USA, 7-16 August 2008.
- [77] MWA project. online: <http://www.mwatelescope.org>.
- [78] Ramesh Narayan and Rajaram Nityananda. Maximum Entropy Image Restoration in Astronomy. *Annual Review of Astronomy & Astrophysics*, 24:127–170, 1986.
- [79] B.C. Ng and C.M.S. See. Sensor Array Calibration Using a Maximum Likelihood Approach. *IEEE Transactions on Antennas and Propagation*, 44(6):827–835, June 1996.
- [80] J.E. Noordam and A.G. de Bruyn. High dynamic range mapping of strong radio sources, with application to 3C84. *Nature*, 299(5884):597–600, 14 October 1982.
- [81] Parisa Noorishad. Appropriate RFI Detection Threshold for Error Prevention in LOFAR Station Calibration. Master's thesis, Chalmers University of Technology, Göteborg, Sweden, June 2008.
- [82] 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.
- [83] Athanasios Papoulis. *Signal Analysis*. McGraw-Hill Book Company, 1984.
- [84] Parbhoo D. Patel, Dion W. Kant, Erik v.d. Wal and Arnold van Ardenne. Towards Completion of Aperture Array Developments - EMBRACE Update. In *Proceedings of the XXIXth General Assembly of the International Union of Radio Science (URSI GA)*, Chicago (Ill.), USA, 7-16 August 2008.
- [85] T.J. Pearson and A.C.S. Readhead. Image Formation by Self-Calibration in Radio Astronomy. *Annual Review of Astronomy and Astrophysics*, 22:97–130, 1984.
- [86] H.J. Pepping. Board description of the TIM module. Technical Report LOFAR-ASTRON-EDD-001, ASTRON, Dwingeloo, The Netherlands, December 2003. rev. nr. 1.0.

- [87] Richard A. Perley. Wide Field Imaging II: Imaging with Non-coplanar Arrays. In Richard A. Perley, Frederic R. Schwab and Alan H. Bridle, editors, *Synthesis Imaging in Radio Astronomy*, volume 6 of *Astronomical Society of the Pacific Conference Series*, pages 139–165. BookCrafters Inc., 1994.
- [88] Richard A. Perley, Frederic R. Schwab and Alan H. Bridle. *Synthesis Imaging in Radio Astronomy*, volume 6 of *Astronomical Society of the Pacific Conference Series*. BookCrafters Inc., 1994.
- [89] M. Pesavento, A.B. Gershman and K.M. Wong. Direction Finding in Partly Calibrated Sensor Arrays Composed of Multiple Subarrays. *IEEE Transactions on Signal Processing*, 50(9):2103–2115, September 2002.
- [90] J. Pierre and M. Kaveh. Experimental Performance of Calibration and Direction Finding Algorithms. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) '91*, volume 2, pages 1365–1368, Toronto, Canada, May 1991.
- [91] Angela Quinlan, Jean-Pierre Barbot and Pascal Larzabal. Automatic detection of the number of targets present when using the time reversal operator. *Journal of the Acoustic Society of America*, 119(4):2220–2225, April 2006.
- [92] Angela Quinlan, Jean-Pierre Barbot, Pascal Larzabal and Martin Haardt. Model Order Selection for Short Data: An Exponential Fitting Test. *EURASIP Journal on Advances in Signal Processing*, 2007.
- [93] Alan E.E. Rogers, Preethi Pratab and Eric Kratzenberg. Calibration of active antenna arrays using a sky brightness model. *Radio Science*, 39(RS2023), 2004.
- [94] Justin Romberg. Imaging via Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2):14–20, March 2008.
- [95] Richard Roy and Thomas Kailath. ESPRIT - Estimation of Signal Parameters Via Rotational Invariance Techniques. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(7):984–995, July 1989.
- [96] M. Ryle and D.D. Vonberg. Solar Radiation on 175 Mc./s. *Nature*, 158:339–340, 7 September 1946.
- [97] M. Ryle. A New Radio Interferometer and Its Application to the Observation of Weak Stars. *Proceedings of the Royal Society of London A: Mathematical and Physical Sciences*, 211(1106):351–375, March 1952.

- [98] M. Ryle. The New Cambridge Radio Telescope. *Nature*, 194:517–518, May 1962.
- [99] R.J. Sault, J.P. Hamaker and J.D. Bregman. Understanding radio polarimetry - II. Instrumental calibration of an interferometer array. *Astronomy & Astrophysics Supplement Series*, 117:149–159, May 1996.
- [100] Anna Scaife, Yves Wiaux, Laurent Jacques, Gilles Puy and Pierre Vandergheynst. Image reconstruction using Compressed Sensing. In *SKA Calibration & Imaging Workshop (CalIm)*, Socorro (New Mexico), USA, 30 March -3 April 2009.
- [101] Ralph O. Schmidt. Multiple Emitter Location and Parameter Estimation. *IEEE Transactions on Antennas and Propagation*, AP-34(3):276–280, March 1986.
- [102] U.J. Schwarz. Mathematical-statistical Description of the Iterative Beam Removing Technique (Method CLEAN). *Astronomy & Astrophysics*, 65:345–356, 1978.
- [103] C.M.S. See. Method for array calibration in high-resolution sensor array processing. *IEE Proc. Radar, Sonar and Navig.*, 142(3), June 1995.
- [104] C.M.S. See and A.B. Gershman. Direction of Arrival Estimation in Partly Calibrated Subarray-Based Sensor Arrays. *IEEE Transactions on Signal Processing*, 52(2):329–338, February 2004.
- [105] SKA project. online: <http://www.skatelescope.org>.
- [106] Sloan digital sky survey. online: <http://www.sdss.org>.
- [107] V.C. Soon, L. Tong, Y.F. Huang and R. Liu. A Subspace Method for Estimating Sensor Gains and Phases. *IEEE Transactions on Signal Processing*, 42(4):973–976, April 1994.
- [108] P. Stoica and B.C. Ng. On the Cramér-Rao Bound Under Parametric Constraints. *IEEE Signal Processing Letters*, 5(7):177–179, July 1998.
- [109] Petre Stoica, Björn Ottersten, Mats Viberg and Randolph L. Moses. Maximum Likelihood Array Processing for Stochastic Coherent Sources. *IEEE Transactions on Signal Processing*, 44(1):96–105, January 1996.
- [110] Petre Stoica and Torsten Söderström. On Array Signal Processing in Spatially Correlated Noise Fields. *IEEE Transactions on Circuits and Systems - II. Analog and Digital Signal Processing*, 39(12):879–882, December 1992.

- [111] Thomas Svantesson. Modeling and Estimation of Mutual Coupling in a ULA of Dipoles. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) '99*, pages 2961–2964, Phoenix (Arizona), USA, March 1999.
- [112] Thomas Svantesson and Mats Viberg. Mutual Coupling in Antenna Arrays: Effects and Cures. In *Proceedings PCC Workshop*, pages 99–103, Stockholm, Sweden, November 1998.
- [113] S.M. Tan. An Analysis of the Properties of CLEAN and Smoothness Stabilized CLEAN—Some Warnings. *Monthly Notices of the Royal Astronomical Society*, 220:971–1001, 1986.
- [114] Max Texmark. How to Make Maps from Cosmic Microwave Background Data Without Losing Information. *Astrophysical Journal Letters*, 480:L87–L90, May 1997.
- [115] Max Texmark. How to Measure CMB Power Spectra Without Losing Information. *Physical Review D*, 55(10):5895–5907, May 1997.
- [116] A. Richard Thompson, James M. Moran and George W. Swenson. *Interferometry and Synthesis in Radio Astronomy*. Krieger Publishing Company, Malabar (Florida), USA, 1998. Reissue with minor corrections, copyright 1986 by John Wiley & Sons, Inc.
- [117] Sebastiaan van der Tol and Alle-Jan van der Veen. Application of Robust Capon Beamforming to Radio Astronomical Imaging. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2005*, volume 4, pages 1089–1092, Philadelphia (Pennsylvania), USA, 18-23 March 2005.
- [118] 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.
- [119] Sebastiaan van der Tol and Stefan J. Wijnholds. CRB Analysis of the Impact of Unknown Receiver Noise in Phased Array Calibration. In *Proceedings of the fourth IEEE Workshop on Sensor Array and Multi-channel Processing (SAM)*, Waltham (Mass.), USA, 12-14 July 2006.
- [120] Sebastiaan van der Tol, Brian D. Jeffs and Alle-Jan van der Veen. Self Calibration for the LOFAR Radio Astronomical Array. *IEEE Transactions on Signal Processing*, 55(9):4497–4510, September 2007.

- [121] S. van der Tol. *Bayesian Estimation for Ionospheric Calibration in Radio Astronomy*. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, November 2009.
- [122] Harry L. van Trees. *Optimum Array Processing - Part IV of Detection, Estimation and Modulation Theory*. John Wiley & Sons Inc., New York, USA, 2002.
- [123] Alle-Jan van der Veen, Amir Leshem and Albert-Jan Boonstra. Array Signal Processing for Radio Astronomy. *Experimental Astronomy*, 17(1-3):231–249, 2004.
- [124] Mats Viberg and Björn Ottersten. Sensor Array Processing Based on Subspace Fitting. *IEEE Transactions on Signal Processing*, 39(5):1110–1121, May 1991.
- [125] Mats Viberg, Björn Ottersten and Thomas Kailath. Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting. *IEEE Transactions on Signal Processing*, 39(11):2436–2449, November 1991.
- [126] S.A. Vorobyov, A.B. Gershman and Zhi-Quan Luo. Robust Adaptive Beamforming Using Worst-case Performance Optimization: A Solution to the Signal Mismatch Problem. *IEEE Transactions on Signal Processing*, 51(2):313–324, February 2003.
- [127] Marco de Vos, Andre W. Gunst and Ronald Nijboer. The LOFAR Telescope: System Architecture and Signal Processing. *Proceedings of the IEEE*, 97(8):1431–1437, August 2009.
- [128] Karl F. Warnick, Leonid Belostotski and Peter Russer. Minimizing the Noise Penalty Due to Mutual Coupling for a Receiving Array. In *Electrical and Computer Engineering Faculty Publications*. Brigham Young University, 28 March 2008. online: <https://dspace.bye.edu/handle/1877/621>.
- [129] Karl F. Warnick and Brian D. Jeffs. Gain and Aperture Efficiency for a Reflector Antenna With an Array Feed. *IEEE Antennas and Wireless Propagation Letters*, 5(1):499–502, December 2006.
- [130] Karl F. Warnick and Michael Jensen. Effects of Mutual Coupling on Interference Mitigation With a Focal Plane Array. *IEEE Transactions on Antennas and Propagation*, 53(8):2490–2498, August 2005.

- [131] Mati Wax and Thomas Kailath. Detection of Signals by Information Theoretic Criteria. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(2):387–392, April 1985.
- [132] Mati Wax, Jacob Sheinvald and Anthony J. Weiss. Detection and Localization in Colored Noise via Generalized Least Squares. *IEEE Transactions on Signal Processing*, 44(7):1734–1743, July 1996.
- [133] A.J. Weiss and B. Friedlander. Array Shape Calibration Using Sources in Unknown Locations - A Maximum Likelihood Approach. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12), December 1989.
- [134] A.J. Weiss and B. Friedlander. 'Almost Blind' signal estimation using second order moments. *IEE Proc. Radar, Sonar and Navig.*, 142(5):213–217, October 1995.
- [135] Anthony D. Whalen. *Detection of Signals in Noise*. Academic Press Inc., San Diego (California), 1971.
- [136] M.H. Wieringa. An investigation of the telescope based calibration methods 'redundancy' and 'self-cal'. *Experimental Astronomy*, 2(4):203–225, 1992.
- [137] Stefan J. Wijnholds. Evaluation and Demonstration of THEP as a Radio Astronomical Observing Facility. Master's thesis, University of Groningen, Groningen, The Netherlands, 2003.
- [138] Stefan J. Wijnholds. THETA experimental results. Technical Report LOFAR-ASTRON-MEM-120, ASTRON, Dwingeloo (The Netherlands), 10 March 2004.
- [139] Stefan J. Wijnholds. Theoretical analysis of the main lobe effects of beamsquint. Technical Report LOFAR-ASTRON-MEM-151, ASTRON, Dwingeloo, The Netherlands, 14 June 2004.
- [140] Stefan J. Wijnholds, Jaap D. Bregman and Albert-Jan Boonstra. Sky Noise Limited Snapshot Imaging in the Presence of RFI With LOFAR's Initial Test Station. *Experimental Astronomy*, 17(1-3):35–42, 2004.
- [141] Stefan J. Wijnholds, A. Ger de Bruyn, Jaap D. Bregman and Jan-Geralt bij de Vaate. Hemispheric Imaging of Galactic Neutral Hydrogen with a Phased Array Antenna System. *Experimental Astronomy*, 17(1-3):59–64, 2004.

- [142] Stefan J. Wijnholds. One year in the bushes - Results from ITS 2004. Technical Report LOFAR-ASTRON-RPT-052, ASTRON, Dwingeloo, The Netherlands, 22 February 2005.
- [143] Stefan J. Wijnholds. Theory of Polarized Imaging With Phased Array Antennas. Technical Report LOFAR-ASTRON-MEM-183, ASTRON, Dwingeloo, The Netherlands, 26 July 2005.
- [144] Stefan J. Wijnholds. ITS Results 2005 - Filling the gaps. Technical Report LOFAR-ASTRON-RPT-056, ASTRON, Dwingeloo, The Netherlands, 23 September 2005.
- [145] Stefan J. Wijnholds. Confusion Limited All Sky Imaging with LOFAR's Initial Test Station Applying Wide Field Calibration Techniques. In *Proceedings of the XXVIIIth General Assembly of the International Union of Radio Science (URSI GA)*, New Delhi, India, 23-29 October 2005.
- [146] Stefan J. Wijnholds. Statistical Performance Analysis of Station Calibration Algorithms. Technical Report LOFAR-ASTRON-MEM-188, ASTRON, Dwingeloo, The Netherlands, 8 November 2005.
- [147] Stefan J. Wijnholds. Self-noise in full sky LOFAR images. In *Dutch Astronomers Conference (NAC)*, Hollum, The Netherlands, 10-12 May 2006.
- [148] Stefan J. Wijnholds. Study on LOFAR Core Calibratability Based on Source Statistics. Technical Report LOFAR-ASTRON-MEM-205, ASTRON, Dwingeloo, The Netherlands, 20 June 2006.
- [149] Stefan J. Wijnholds and Albert-Jan Boonstra. A Multisource Calibration Method for Phased Array Telescopes. In *Proceedings of the fourth IEEE Workshop on Sensor Array and Multi-channel Processing (SAM)*, pages 200–204, Waltham (Mass.), USA, 12-14 July 2006.
- [150] Stefan J. Wijnholds. Station Calibration - Pitfalls & Possibilities 2006. Technical Report LOFAR-ASTRON-MEM-217, ASTRON, 13 October 2006.
- [151] Stefan J. Wijnholds and Alle-Jan van der Veen. Effects of Parametric Constraints on the CRLB in Gain and Phase Estimation Problems. *IEEE Signal Processing Letters*, 13(10):620–623, October 2006.
- [152] Stefan J. Wijnholds. Reducing the Impact of Station Level Spatial Filtering Limitations. In *SKA Calibration & Imaging Workshop (CalIm)*, Cape Town, South Africa, 4-6 December 2006.

- [153] Stefan J. Wijnholds. CS1 system validation. Technical Report LOFAR-ASTRON-RPT-124, ASTRON, Dwingeloo, The Netherlands, 10 April 2007.
- [154] Stefan J. Wijnholds. LOFAR on the Moon: Science & System Concept. In *Dutch Astronomers Conference (NAC)*, Veldhoven, The Netherlands, 14-16 May 2007.
- [155] Stefan J. Wijnholds. LBA station configuration. Technical Report LOFAR-ASTRON-MEM-241, ASTRON, Dwingeloo, The Netherlands, 4 December 2007.
- [156] Stefan J. Wijnholds. Autonomous Online LOFAR Station Calibration - Methods and Results. In *SKA Calibration and Imaging Workshop (CalIm)*, Perth, Australia, 7-9 April 2008.
- [157] Stefan J. Wijnholds. Effective Noise. In *SKA Calibration and Imaging Workshop (CalIm)*, Perth, Australia, 7-9 April 2008.
- [158] Stefan J. Wijnholds. Calibration of Phased Array Radio Telescopes. In *Proceedings of the International Union of Radio Science (URSI) Benelux Forum*, Brussels, Belgium, 30 May 2008. invited paper.
- [159] Stefan J. Wijnholds. Mutual Coupling, Inter-tile Spacing and Inter-station Rotation in the HBA Array. Technical Report LOFAR-ASTRON-MEM-245, ASTRON, Dwingeloo, The Netherlands, 6 August 2008.
- [160] Stefan J. Wijnholds. Autonomous Online LOFAR Station Calibration. In *Proceedings of the XXIXth General Assembly of the International Union of Radio Science (URSI GA)*, Chicago (Ill.), USA, 7-16 August 2008.
- [161] Stefan J. Wijnholds and Alle-Jan van der Veen. Fundamental Imaging Limits of Radio Telescope Arrays. *IEEE Journal of Selected Topics in Signal Processing*, 5(2):613–623, October 2008.
- [162] Stefan J. Wijnholds. Least Squares All-Sky Imaging With A LOFAR Station. In *SKA Calibration & Imaging Workshop (CalIm)*, Socorro (New Mexico), USA, 30 March - 3 April 2009.
- [163] Stefan J. Wijnholds. Fish-Eye Imaging With A LOFAR Station. In *Proceedings of the International Union of Radio Science (URSI) Benelux Forum*, Delft, The Netherlands, 8 June 2009.

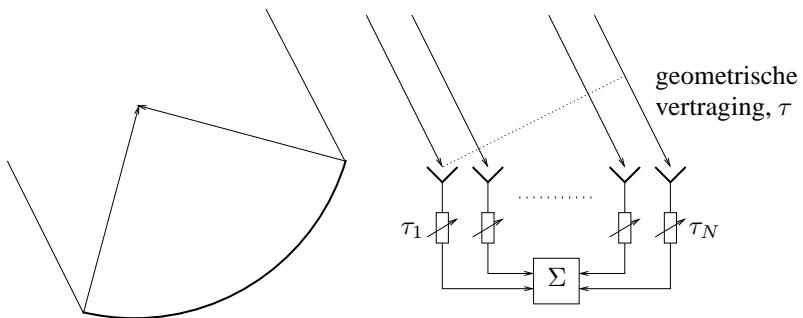
- [164] Stefan J. Wijnholds and Alle-Jan van der Veen. Self-Calibration of Radio Astronomical Arrays with Non-Diagonal Noise Covariance Matrix. In *Proceedings of the 17th European Signal Processing Conference (EuSiPCo)*, Glasgow, United Kingdom, 24-28 August 2009.
- [165] Stefan J. Wijnholds and Alle-Jan van der Veen. Multisource Self-calibration for Sensor Arrays. *IEEE Transactions on Signal Processing*, 57(9):3512–3522, September 2009.
- [166] Stefan J. Wijnholds, Sebastiaan van der Tol, Ronald Nijboer and Alle-Jan van der Veen. Calibration Challenges for Future Radio Telescopes. *IEEE Signal Processing Magazine*, 27(1):30–42, January 2010.
- [167] Stefan J. Wijnholds, Jaap D. Bregman and Alle-Jan van der Veen. LOFAR Station Calibration. *Experimental Astronomy*, 2010. in preparation (invited paper).
- [168] Wikipedia. Divide-and-conquer eigenvalue algorithm. http://en.wikipedia.org/wiki/Divide-and-conquer_eigenvalue_algorithm.
- [169] P. Zarka. Fast radio imaging of Jupiter’s magnetosphere at low-frequencies with LOFAR. *Planetary and Space Science*, 52(15):1455–1467, December 2004.
- [170] M. Zatman. How narrow is narrowband. *IEE Proc. Radar, Sonar and Navig.*, 145(2):85–91, April 1998.

Samenvatting

Het doen van fish-eye waarnemingen met radiotelescopen gebaseerd op phased array technologie

Sterrenkunde richt zich op het bestuderen van natuurkundige verschijnselen buiten de dampkring van de Aarde. Een van de manieren om dat te doen is door het meten van radiogolven uit het heelal. Deze metingen worden gedaan met radiotelescopen zoals de Westerbork Synthese Radiotelescoop in Nederland en de Very Large Array in Nieuw Mexico. Deze telescopen bestaan uit respectievelijk 14 en 27 schotelantennes met een diameter van 25 m. Beide telescopen hebben dus meerdere ontvangende elementen en zijn derhalve zogenaamde *arrays*. Door de signalen ontvangen door de afzonderlijke elementen binnen het array op de juiste wijze te combineren, kan een virtuele telescoop gesynthetiseerd worden met een diameter zo groot als de afstand tussen de verstuut uit elkaar staande elementen van het array. Aangezien de resolutie evenredig is met de diameter van de telescoop, kan met een array van een aantal kleine schotels een veel hogere resolutie gehaald worden dan met een enkele grote schotel.

Op dit moment wordt er binnen de radiosterrenkunde gewerkt aan de ontwikkeling en de bouw van een aantal nieuwe radiotelescopen gericht op het doen van waarnemingen beneden de 300 MHz. Voorbeelden hiervan zijn de Low Frequency Array (LOFAR, Nederland), de Murchison Widefield Array (MWA, Australië) en de Long Wavelength Array (LWA, VS). Daarnaast zijn er al verschillende haalbaarheidsstudies gedaan naar telescopen op de Maan en in de ruimte voor waarnemingen beneden 10 MHz, die niet vanaf de Aarde gedaan kunnen worden vanwege de ionosfeer. Al deze telescopen zijn zogenaamde *apertuur arrays* en werken fundamenteel anders dan de “klassieke” radiotelescopen met schotelantennes. Dit verschil is geschatst in Fig. 1. Bij een schotelantenne worden de inkomende radiogolven geconcentreerd in het brandpunt van de schotel waar zich de ontvanger bevindt. Bij een apertuur array zijn alle ontvangers rechtstreeks op de hemel gericht. Deze ontvangers hebben een zeer groot zichtsveld en ontvangen dus tegelijk signalen uit vele verschillende richtin-



Figuur 1: Vergelijking tussen de werking van een schotelantenne (links) en een apertuur array (rechts).

gen. Een signaal uit één bepaalde richting komt bij de ene ontvanger eerder aan dan bij de andere ontvanger. Door dit tijdsverschil weg te regelen voordat de signalen van de verschillende ontvangers opgeteld worden, wordt het gewenste signaal versterkt en worden alle andere signalen onderdrukt.

Dit proefschrift richt zich op de vraag wat de mogelijkheden en beperkingen zijn voor het maken van beelden (*imaging*) van de hemel op radiofrequenties met dergelijke apertuur arrays. Hierbij wordt er vanuit gegaan dat de signalen van de afzonderlijke ontvangers beschikbaar zijn. Het maken van beelden gaat in twee stappen. In de eerste stap, de calibratie, wordt de karakteristiek van de telescoop tijdens de waarneming bepaald. Deze informatie wordt gebruikt in de tweede stap, de eigenlijke imaging. In de praktijk wordt vaak gekozen voor een aanpak waarbij deze stappen een aantal keer herhaald worden om zo een betere oplossing, een scherper beeld, te krijgen.

Waar schotelontvangers gericht kunnen worden op één sterke bron voor (initiële) calibratie, zien de ontvangers van een apertuur array permanent de hele hemel. Daardoor ontvangen ze altijd de signalen van meerdere discrete bronnen en de diffuse emissie van de Melkweg. Daarnaast kunnen de radiogolven op deze lage frequenties behoorlijk verstoord worden door de ionosfeer. Het gevolg daarvan is dat de richtingsafhankelijke gevoeligheid van de telescoop met de tijd kan variëren. In dit proefschrift wordt een wiskundig model opgesteld, waarmee het signaal van de afzonderlijke ontvangers beschreven kan worden. Met dit model kan het calibratieprobleem beschreven worden als een schattingsprobleem. Dit probleem wordt vervolgens opgesplitst in een aantal deelproblemen, waarvoor analytische oplossingen gepresenteerd worden of beschikbaar zijn in de literatuur. Op basis hiervan wordt een rekenkundig efficiënt calibratie-algoritme voorgesteld. Dit algoritme is toegepast op data van

een LOFAR prototype station. Uit deze tests bleek o.a. dat het algorithme in staat was om 860 vrije parameters op te lossen op een enkele standaard 2.4 GHz core in slechts 0.4 s en dat het algorithme nog steeds betrouwbare oplossingen vindt bij een sterk verstorende ionosfeer.

Dankzij de wiskundige formulering van het probleem, is het mogelijk om de Cramer-Rao limiet (CRL) te bepalen. De CRL is een theoretische limiet voor de nauwkeurigheid waarmee een schatter zonder systematische afwijkingen de vrije parameters kan schatten. De nauwkeurigheid van de voorgestelde calibratiemethode is middels simulaties vergeleken met de CRL. Uit deze simulaties blijkt dat de voorgestelde methode geen systematische afwijkingen heeft en qua nauwkeurigheid de CRL haalt zodra het vermogen van de bronnen een factor 3 hoger is dan de ruis. Hiermee is gedemonstreerd dat er geen algorithmes kunnen bestaan die voor hetzelfde probleem op basis van dezelfde data een nog hogere nauwkeurigheid kunnen halen.

Gezien het feit dat de nauwkeurigheid van de calibratie fundamenteel beperkt is, is de logische volgende vraag in hoeverre dit de kwaliteit van de gemaakte beelden beperkt. Vaak wordt hiervoor gekeken naar het dynamische bereik van een kaart, gedefinieerd als het verschil in vermogen tussen de sterkste en de zwakste bron in de kaart. Dit blijkt in het algemeen geen goede maat te zijn voor de kwaliteit van een telescoop of dataverwerkingsmethode, omdat dit sterk afhangt van het vermogen van de sterkste bron in het beeldveld, hetgeen per beeldveld behoorlijk kan verschillen. In dit proefschrift wordt daarom de effectieve ruis als maat genomen. De effectieve ruis is een maat voor de totale ruis op ieder punt in een kaart, die opgebouwd is uit de ruis in de meting, de fouten in de kaart t.g.v. calibratiefouten en confusieruis. De laatste bijdrage is het gevolg van het feit dat een telescoop bij hogere gevoeligheid steeds meer bronnen zal detecteren tot het punt dat de afzonderlijke bronnen niet meer uit elkaar te houden zijn omdat er op ieder punt in de kaart een bron staat. Waarnemingen doen met nog hogere gevoeligheid heeft dan pas weer zin als de resolutie van de telescoop verder verbeterd wordt. Met behulp van het opgebouwde wiskundige formalisme, is het mogelijk om de impact van zowel de ruis in de metingen als de calibratiefouten op de uiteindelijke hemelkaart te analyseren. Hieruit blijkt dat calibratiefouten vooral doorwerken op de plaats van de sterke bronnen en dat ze minder invloed hebben op de “lege” delen van het beeld dan de ruis in de metingen. De analyse laat ook zien, dat als de telescoop zo stabiel is, dat de instrumentele variaties tussen calibratiemetingen vergelijkbaar of kleiner zijn dan de nauwkeurigheid waarmee het instrument tijdens de waarneming zelf gecalibreerd kan worden, aparte calibratiemetingen de voorkeur verdienen boven calibratie tijdens de waarneming zelf.

Ook imaging kan beschouwd worden als een schattingsprobleem, waarbij we het vermogen willen weten dat in ieder beelpunt in de kaart zit. Het mooie van een dergelijke, op een wiskundig model gebaseerde, aanpak is, dat grote uitdagingen bij

imaging met aperture arrays, zoals het feit dat de gevoeligheid in de richting van elke bron in het beeldveld per frequentie en per tijdstip anders kan zijn en het extreem grote beeldveld, impliciet opgelost worden. In dit proefschrift worden analytische oplossingen afgeleid voor verschillende imaging scenario's. Hieruit blijkt dat de rekenkundige complexiteit van de feitelijke imaging stap slechts afhangt van het aantal beeldpunten in de kaart en niet van de hoeveelheid data. Een mooie illustratie hiervan is de kaart van de hele hemel zichtbaar boven Exloo geproduceerd op basis van data van een driedaagse waarnemingscampagne met een LOFAR prototype station. Met data van datzelfde station wordt tevens gedemonstreerd dat deze imaging-methodiek de Cramer-Rao limiet haalt op echte data. Dat laat tevens zien dat het gepresenteerde wiskundige model voor de data goed aansluit bij de werkelijkheid.

Hoewel de gepresenteerde benadering voor imaging nog verder verbeterd kan worden, laten de resultaten behaald met LOFAR-data zien, dat deze aanpak weinig problemen heeft met complexe bronstructuren zoals de Melkweg. Dit maakt deze methode geschikt voor waarnemingen met LOFAR, en in de toekomst de Square Kilometre Array (SKA), die gericht zijn op imaging van de Melkweg en detectie van de geboorte van de eerste sterren. Deze waarnemingen worden gekenmerkt door een relatief lage resolutie en een zo hoog mogelijke gevoeligheid op complexe bronstructuren. Deze blik in de toekomst versterkt het gevoel dat de belangrijkste bijdrage van dit proefschrift wel eens zou kunnen liggen in het feit dat er consistent gebruik wordt gemaakt van methoden uit de signaalverwerkingsliteratuur i.p.v. voort te bordenen op de methoden die beschikbaar zijn in de huidige dataverwerkingspakketten binnen de radiosterrenkunde. De op wiskundige modellen gebaseerde aanpak uit de statistische signaalverwerking krijgt langzamerhand meer aandacht binnen de radiosterrenkunde en zal de basis moeten vormen van de volgende generatie calibratie- en imaging-methoden.

Stefan J. Wijnholds
25 januari 2010

About the author

Stefan Jeroen Wijnholds was born in Groningen, The Netherlands, on October 21, 1978. He finished his secondary education (Gymnasium) at the Willem Lodewijk Gymnasium in Groningen in 1997. In the same year, he began his studies at the University of Groningen. In 2003, he received the M.Eng. degree in Applied Physics (cum laude) for a study on a measurement technique to evaluate the optics of the Herschel Space Observatory at sub-millimeter wavelengths with the Space Research Organization of the Netherlands (SRON) in Groningen. In the same year, he also received the M.Sc. degree in Astronomy (cum laude) for a study on the integration and the evaluation of a phased array antenna system for radio astronomy with the Netherlands Institute for Radio Astronomy (ASTRON) in Dwingeloo, The Netherlands.

After his graduation he joined the R&D department of ASTRON, where he works with the system design and integration group on the development of the next generation of radio telescopes. He has been involved in integration and testing of all LOFAR (prototype) stations deployed since 2003 with a focus on station calibration and system characterization. Since 2006 he has also been affiliated with the Delft University of Technology in Delft, The Netherlands, for pursuing his Ph.D. degree. His research interests lie in the area of array signal processing, specifically calibration and imaging.

