

Image formation for future radio telescopes

Naghizadeh, Shahrzad

DOI

[10.4233/uuid:653d520b-07c0-4fd8-bd40-4f4ee74d8bf5](https://doi.org/10.4233/uuid:653d520b-07c0-4fd8-bd40-4f4ee74d8bf5)

Publication date

2018

Document Version

Final published version

Citation (APA)

Naghizadeh, S. (2018). *Image formation for future radio telescopes*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:653d520b-07c0-4fd8-bd40-4f4ee74d8bf5>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

IMAGE FORMATION FOR FUTURE RADIO TELESCOPES

IMAGE FORMATION FOR FUTURE RADIO TELESCOPES

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T. H. J. J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 7 december 2018 om 12:30 uur

door

Shahrzad NAGHIBZADEH

Electrotechnisch Ingenieur, Technische Universiteit Delft, Nederland

geboren te Mashhad, Iran.

I leave Sisyphus at the foot of the mountain. One always finds one's burden again. But Sisyphus teaches the higher fidelity that negates the gods and raises rocks. He too concludes that all is well. This universe henceforth without a master seems to him neither sterile nor futile. Each atom of that stone, each mineral flake of that night-filled mountain, in itself, forms a world. The struggle itself toward the heights is enough to fill a man's heart. One must imagine Sisyphus happy.

Albert Camus, *The myth of Sisyphus* (1942)

CONTENTS

Summary	xi
Samenvatting	xiii
I Part I: Preliminaries	1
1 Introduction	3
1.1 Radio astronomy	4
1.2 Radio interferometry	5
1.3 Radio interferometric imaging	7
1.4 Next generation radio telescopes	9
1.5 SKA overview	9
1.5.1 Science cases	9
1.5.2 The instrument	11
1.5.3 Imaging and data reduction pipeline.	12
1.5.4 Imaging challenges	12
1.6 Scope of this dissertation	13
1.6.1 DRIFT	13
1.6.2 Problem statement and research goals.	14
1.6.3 Outline and contributions	14
1.6.4 List of publications.	16
2 Data Model and Problem Formulation	19
2.1 Symbols, notations, operators and matrix relations	20
2.1.1 Symbols	20
2.1.2 Operators	20
2.1.3 Matrix product relations	20
2.2 Measurement model for image formation.	21
2.2.1 Measurement model in conventional radio astronomy	21
2.2.2 Array processing formulation	27
2.2.3 Some further remarks and generalizations.	29
2.2.4 SKA processing pipeline	32
2.3 Interpretations of the measurement equation.	33
2.3.1 Discrete linear inverse problem	33
2.3.2 Second-order statistics estimation	34
2.4 Radio interferometric imaging problem formulation	38
2.4.1 Beamforming-based estimation	38
2.4.2 Least squares estimation.	40
2.4.3 Maximum likelihood estimation	40

2.5	Conclusions.	41
3	Ill-posedness and Regularization	43
3.1	Analysis of the measurement system	43
3.1.1	Ill-posedness of the integral equation	43
3.1.2	Conditioning analysis of the discrete measurement equation	44
3.1.3	Principal solution and invisible distributions	47
3.1.4	Normal equations and the "dirty image"	48
3.2	Regularization	51
3.2.1	Regularizing assumptions in radio interferometric imaging	52
3.2.2	Bayesian problem formulation and regularization	55
3.3	Preconditioning as a tool for regularization	56
3.3.1	Preconditioning in numerical mathematics	57
3.3.2	Preconditioning of ill-posed problems	58
3.3.3	Statistical preconditioning: priorconditioning	59
3.3.4	Preconditioners for sparse representation	59
3.3.5	Antenna/baseline selection	60
3.3.6	Preconditioning in radio astronomy	61
3.4	Conclusions.	62
II	Part II : Data-driven regularization	63
4	Data-Driven Regularization for Direct Deconvolution	65
4.1	Introduction	65
4.2	Radio interferometric deconvolution problem	65
4.3	Proposed methods	66
4.3.1	Weighted truncated eigenvalue decomposition	67
4.3.2	Beamforming-based conditioning	67
4.4	Simulation results.	68
4.5	Experimental results	69
4.6	Conclusions.	72
5	Fast Imaging with General Regularization	73
5.1	Introduction	73
5.1.1	State of the art imaging algorithms.	74
5.1.2	Results	75
5.2	Proposed solution method	76
5.2.1	Problem reformulation.	76
5.2.2	Selection of the prior covariance	77
5.2.3	Discussion and generalizations	78
5.3	The PRIFIRA algorithm	82
5.3.1	Implementation using Krylov subspace methods	82
5.3.2	Stopping criteria	83
5.3.3	Combined regularizing effect of prior-conditioning and early stop- ping	85
5.3.4	Implementation details	86
5.3.5	Computational complexity of PRIFIRA.	86

5.4	One-dimensional Simulation results	87
5.4.1	Terminology	87
5.4.2	One-dimensional tests	87
5.5	Effect of missing autocorrelations.	90
5.6	Afterword	92
6	PRIFIRA Experimental Results	95
6.1	Introduction	95
6.2	Tests on model images	95
6.3	Single LOFAR station data.	101
6.4	LOFAR core stations.	103
6.5	Conclusions.	107
III	Part III : Sparsity-based regularization	113
7	Dual-Basis Sparse Recovery	115
7.1	Introduction	115
7.2	Gradient descent-based imaging methods	117
7.2.1	The CLEAN-based imaging pipeline	118
7.2.2	The active set-based imaging pipeline	120
7.3	Dual-basis NNLS	122
7.3.1	Numerical Results and Performance Analysis	124
7.4	Multi-basis generalization of IR0-PRIFIRA	128
7.5	Synthesis-based sparse imaging with convex optimization	128
7.6	Analysis versus synthesis sparse recovery	129
7.7	Simulation results.	130
7.7.1	One-dimensional simulations	130
7.7.2	Two-dimensional simulations	133
7.8	Conclusions.	135
8	Facet-Based Regularization for Scalable Imaging	137
8.1	Introduction	137
8.2	Radio-interferometric Imaging	139
8.2.1	Problem description	139
8.2.2	Compressive sensing and data splitting	139
8.3	Proposed approach	140
8.3.1	Facet-based Imaging.	140
8.3.2	Proposed algorithm	141
8.4	Simulation Results	144
8.5	Conclusion and Future Works.	147
IV	Part IV : Epilogue	149
9	Conclusions and suggestions for future work	151
9.1	Conclusions and contributions	151
9.2	Suggestions for future work	153

Acknowledgements	157
Curriculum Vitae	159
A List of abbreviations	161
Bibliography	165

SUMMARY

Fundamental scientific questions such as how the first stars were formed or how the universe came into existence and evolved to its present state drive us to observe weak radio signals impinging on the earth from the early days of the universe. During the last century, radio astronomy has been vastly advancing. Important discoveries on the formation of various celestial objects such as pulsars, neutron stars, black holes, radio galaxies and quasars are the result of radio astronomical observations.

To study celestial objects and the astrophysical processes that are responsible for their radio emissions, images must be formed. This is done with the help of large radio telescope arrays. Next generation radio telescopes such as the Low Frequency Array Radio Telescope (LOFAR) [1] and the Square Kilometer Array (SKA) [2], bring about increasingly more observational evidence for the study of the radio sky by generating very high resolution and high fidelity images.

In this dissertation, we study radio astronomical imaging as the problem of estimating the sky spatial intensity distribution over the field of view of the radio telescope array from the incomplete and noisy array data. The increased sensitivity, resolution and sky coverage of the new instruments pose additional challenges to the current radio astronomical imaging pipeline. Namely, the large amount of data captured by the radio telescopes cannot be stored and needs to be processed quasi-realtime. Many pixel-based imaging algorithms, such as the widely-used CLEAN [3] algorithm, are not scalable to the size of the required images and perform very slow in high resolution scenarios. Therefore, there is an urgent need for new efficient imaging algorithms. Moreover, regardless of the amount of collected data, there is an inherent loss of information in the measurement process due to physical limitations. Therefore, to recover physically meaningful images additional information in the form of constraints and regularizing assumptions are necessary. The central objective of the current dissertation is to introduce advanced algebraic techniques together with custom-made regularization schemes to speed up the image formation pipeline of the next generation radio telescopes.

Signal processing provides powerful tools to address these issues. In the current work, following a signal processing model of the radio astronomical observation process, we first analyze the imaging system based on tools from numerical linear algebra, sampling, interpolation and filtering theory to investigate the inherent loss of information in the measurement process. Based on these results, we show that the imaging problem in radio astronomy is highly ill-posed and regularization is necessary to find a stable and physically meaningful image. We continue by deriving an adequate model for the imaging problem in radio interferometry in the context of statistical estimation theory. Moreover, we introduce a framework to incorporate regularization assumptions into the measurement model by borrowing the concept of preconditioning from numerical linear algebra.

Radio emissions observed by radio telescopes appear either as distributed radiation

from diffuse media or as compact emission from isolated point-like sources. Based on this observation, different source models need to be applied in the imaging problem formulation to obtain the best reconstruction performance. Due to the ill-posedness of the imaging problem in radio astronomy, to guarantee a reliable image reconstruction, proper modeling of the source emissions and regularizing assumptions are of utmost importance. We integrate these assumptions implementing a multi-basis dictionary based on the proposed preconditioning formalism. In traditional radio astronomical imaging methods, the constraints and prior models, such as positivity and sparsity, are employed for the complete image. However, large radio sky images usually manifest individual source occupancy regions in a large empty background. Based on this observation, we propose to split the field of view into multiple regions of source occupancy. Leveraging a stochastic primal dual algorithm we apply adequate regularization on each facet. We demonstrate the merits of applying facet-based regularization in terms of memory savings and computation time by realistic simulations.

The formulation of the radio astronomical imaging problem has a direct consequence on the radio sky estimation performance. We define the astronomical imaging problem in a Bayesian-inspired regularized maximum likelihood formulation. Based on this formalism, we develop a general algorithmic framework that can handle diffuse as well as compact source models. Leveraging the linearity of the radio astronomical imaging problem, we propose to directly embody the regularization operator into the system by right preconditioning. We employ an iterative method based on projections onto Krylov subspaces to solve the subsequent system. The proposed algorithm is named PRIor-conditioned Fast Iterative Radio Astronomy (PRIFIRA). We motivate the use of a beam-formed image as an efficient regularizing prior-conditioner for diffuse emission recovery. Different sparsity-based regularization priors are incorporated in the algorithmic framework by generalizing the core algorithm with iterative re-weighting schemes. We evaluate the performance of PRIFIRA based on simulated measurements as well as astronomical data and compare to the state-of-the-art imaging methods. We conclude that the proposed method maintains competitive reconstruction quality with the current techniques while remaining flexible in terms of different regularization schemes. Moreover, we show that the imaging efficiency can be greatly improved by exploiting Krylov subspace methods together with an appropriate noise-based stopping criterion.

Based on the results from this thesis we can conclude that with the help of advanced techniques from signal processing and numerical linear algebra, customized algorithms can be designed to tackle some of the challenges in the next generation radio telescope imaging. We note that since radio interferometric imaging can be considered as an instance of the broad area of inverse imaging problems, the numerical techniques as well as regularization methods developed in this dissertation have a direct impact on many different imaging application areas, such as biomedical and geophysics/seismic imaging.

SAMENVATTING

Fundamentele wetenschappelijke vragen, zoals hoe de eerste sterren gevormd zijn, of hoe het universum ontstaan is en is geëvolueerd tot zijn huidige toestand, motiveren ons om zwakke radiosignalen waar te nemen die ons bereiken vanuit de begindagen van het universum. In de vorige eeuw heeft de radio-astronomie een grote ontwikkeling doorgemaakt. Belangrijke ontdekkingen over de vorming van diverse hemelobjecten, zoals pulsars, zwarte gaten, radiostelsels en quasars zijn het resultaat van radio-astronomische waarnemingen.

Om hemelobjecten en de bijbehorende astrofysische processen te bestuderen moeten afbeeldingen gemaakt worden. Dit wordt gedaan met behulp van grote arrays van radiotelescopen. De volgende generatie van radiotelescopen, zoals de Low Frequency Array Radio Telescope (LOFAR) [1] en de Square Kilometer Array (SKA) [2], leveren toenemend bewijs omtrent de studie van de hemel door hun hoge resolutie en nauwkeurigheid van de gemaakte afbeeldingen.

In dit proefschrift bestuderen we radio-astronomische beeldvorming als het probleem van het schatten van de spatiële verdeling van de intensiteit van de hemel over het beeldvlak van het array van radiotelescopen, uit de onvolledige en verstoorde waarnemingen. De toegenomen gevoeligheid, resolutie en hemelafdekking van de nieuwe instrumenten resulteren in nieuwe uitdagingen aan de huidige manier van het maken van afbeeldingen. Namelijk, de grote hoeveelheid gegevens die de radiotelescoop produceert kan niet worden opgeslagen en moetequasi-realttime worden verwerkt. Veel pixelgebaseerde beeldvormings-algoritmes, zoals het veelgebruikte CLEAN [3] algoritme, schalen niet op naar de beeldafmetingen die nu nodig zijn, en worden erg langzaam als hoge resolutie gevraagd wordt. Het is daarom nodig om nieuwe efficiënte beeldvormings-algoritmes te ontwerpen. Afgezien van de hoeveelheid gemeten data, is er bovendien een inherent verlies aan informatie in het waarnemingsproces vanwege fysische beperkingen. Om fysisch betekenisvolle beelden te maken is extra informatie nodig, in de vorm van randvoorwaarden en regularisatie-aannames. Het centrale doel van dit proefschrift is het introduceren van geavanceerde algebraïsche technieken en geschikte vormen van regularisatie om de beeldvorming voor de nieuwe generatie van radiotelescopen te versnellen.

Signaalbewerking geeft krachtige gereedschappen om deze problemen aan te pakken. Uitgaande van een signaalbewerkingsmodel van het radio-astronomische waarnemingsproces, analyseren we in dit proefschrift eerst het beeldvormingssysteem, met behulp van gereedschappen uit de numerieke lineaire algebra en bemonstering-, interpolatie- en filtertheorie, om hiermee het inherente verlies aan informatie in het meetproces te beschrijven. Dit toont aan dat het probleem van beeldvorming in radio-astronomie een wiskundig slecht bepaald probleem is, zodat regularisering nodig is om een stabiel en betekenisvolle afbeelding te krijgen. Vervolgens leiden we een geschikt model voor het beeldvormingsprobleem af in de context van statistische schattingsthe-

orie. Bovendien introduceren we een kader waarmee regularisatie-aannames in het meetmodel ingebracht kunnen worden, door het overnemen van het concept van preconditionering uit de numerieke lineaire algebra.

Radiostraling opgevangen door radio-telescopen is afkomstig ofwel van straling die uitgesmeerd werd door diffuse media, ofwel door straling van geïsoleerde compacte puntbronnen. Om hier rekening mee te houden moeten verscheidene bronmodellen toegepast worden in de formulering van het beeldvormingsprobleem. Omdat dit probleem slecht bepaald is, is het van het grootste belang dat deze bronstraling en regularisatie-aannames goed gemodelleerd worden. We integreren deze aannames middels een multi-basis woordenboek gebaseerd op het voorgestelde preconditioneringsformalisme.

De randvoorwaarden en voorkennis in traditionele radio-astronomische beeldvormingstechnieken zijn gebaseerd op positiviteit en “schaarsheid” (puntbronnen). Echter, grote hemelkaarten hebben meestal gebieden met bronnen temidden van grote lege gebieden. Hiervan maken we gebruik door het gebied dat we willen afbeelden te splitsen in kleinere deelgebieden waar bronnen aanwezig zijn. Gebruik makend van een stochastisch primal-dual algoritme passen we een geschikte regularisatie toe op ieder deelgebied (“facet”). Met realistische simulaties demonstreren we de voordelen van deze facet-geïoriënteerde regularisatie in termen van verminderde hoeveelheid geheugen en rekentijd.

De formulering van een radio-astronomisch beeldvormingsprobleem heeft directe gevolgen voor de kwaliteit van het schattingsprobleem. We definiëren het astronomische beeldvormingsprobleem als een Bayes-achtig geregulariseerd maximum likelihood probleem. Gebaseerd hierop ontwikkelen we algemene algoritmische kaders waarbinnen zowel diffuse als compacte bronnen behandeld kunnen worden. Gebruikmakend van de lineariteit van het radio-astronomische probleem stellen we voor om de regularisatie-operator direct in de systeemvergelijking mee te nemen, door middel van een preconditioner aan de rechterkant. We gebruiken een iteratieve methode gebaseerd op projecties op Krylov deelruimtes om het onderliggende systeem van vergelijkingen op te lossen. Het voorgestelde algoritme noemen we PRIor-conditioned Fast Iterative Radio Astronomy (PRIFIRA). We motiveren het gebruik van een afbeelding verkregen door bundelvorming als een efficiënte regulariserende prior-conditioner geschikt voor diffuse bronnen. Diverse regulariserende priors gebaseerd op “schaarsheid” worden in het algoritmische kader ingebouwd, door het centrale algoritme te herhalen met iteratief bepaalde wegingen. We evalueren de resultaten van PRIFIRA door middel van gesimuleerde metingen en astronomische data, en vergelijken dit met de huidige standaard gebruikte beeldvormingstechnieken. We concluderen dat de voorgestelde methode een vergelijkbare reconstructie-kwaliteit heeft, terwijl het flexibel is in termen van de gebruikte regularisatie. We laten ook zien dat de efficiëntie van beeldvorming drastisch verbeterd kan worden door gebruik te maken van methodes gebaseerd op Krylov deelruimtes, samen met een geschikte drempel om het algoritme te stoppen.

Gebaseerd op de resultaten van dit proefschrift concluderen we dat door gebruik te maken van geavanceerde technieken van signaalbewerking en numerieke lineaire algebra, het mogelijk is algoritmes te ontwerpen die toegespitst zijn op de uitdagingen van beeldvorming met de volgende generatie van radio telescopen. Omdat beeldvorming

voor radio-astronomie een voorbeeld is voor meer algemene problemen van beeldvorming, zijn de numerieke technieken en regularisatie-methoden die in dit proefschrift ontwikkeld werden ook van toepassing op veel andere gebieden, zoals biomedische en geofysische/seismische beeldvorming.

I

PART I : PRELIMINARIES

1

INTRODUCTION

We are just an advanced breed of monkeys on a minor planet of a very average star. But we can understand the Universe. That makes us something very special.

Stephen Hawking

It all started with the Big Bang! After the explosion in about 13.8 billion years ago, in a fraction of a second, the universe started to expand and gradually cool down to form the first particles. After about 300,000 years, the first atoms, mainly hydrogen and helium, were formed. In the subsequent 6 to 7 hundred million years galaxies and quasars began to form in an otherwise dark universe. About a billion years ago the first stars began to shine and the cosmic dawn started. Fast forward to today, astronomers and engineers work hand-in-hand to build powerful telescopes to unravel the history of the universe with increasingly more observational evidence.

Figure 1.1 shows an artist's impression of the stages in Big Bang cosmology. The cosmic phenomena and timelines are derived by observing waves from the early universe. Radio waves are in particular informative due to radio emissions from the neutral hydrogen, the first and most abundant element in the universe. These emissions are caused by a change in spin direction of the electron in a neutral hydrogen (HI) atom and result in an emission of the 21 cm spectral line (1420 MHz). This discovery had a huge impact on the importance of radio astronomy. Probing HI resulted in mapping of the Cosmic Microwave Background (CMB), the massive radiation that followed the Big Bang. CMB maps provide valuable information about the early universe and is itself the best evidence of the Big Bang theory. Furthermore, due to the expansion of the universe, HI emissions are observed at different red-shifts (a.k.a. Dopplers). This phenomena helps astronomers to put a date on cosmic events.

With the success of radio astronomy in resolving mysteries about formation of the universe, increasingly more accurate and powerful radio telescopes are being developed. The Square Kilometre Array (SKA) [4] is the most sensitive and accurate radio telescope

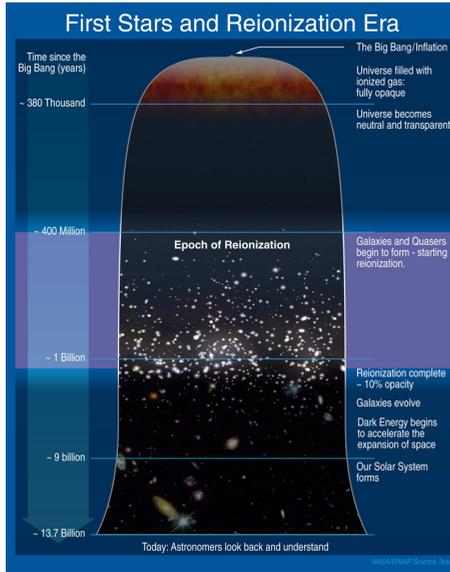


Figure 1.1: Creation of the universe. (Image courtesy of NASA/WMAP science team)

that has ever been designed. The SKA is a future radio telescope, currently under development, that is expected to generate an unprecedented amount of data. Storage and processing of the SKA data for scientific research purposes are extremely challenging tasks that need to be addressed before the instrument becomes fully functional.

In this chapter, we give a brief overview of the field of radio astronomy and in particular radio interferometric imaging. We investigate the challenges of future radio telescopes and introduce the motivation behind the current work and the research questions we have pursued with the current research.

1.1. RADIO ASTRONOMY

In observational astronomy, electromagnetic (EM) waves originating from cosmic sources, that impinge on earth are measured with the help of telescopes. Astronomers study the cosmic objects and phenomena based on these measurements. The received waves occupy a large portion of the EM spectrum with each frequency range providing a different view on the cosmos. As an example, different views of the Milky Way galaxy observed at different frequencies are shown in Figure 1.2.

In the early 1930s, with the discovery of the cosmic radio waves by Karl Jansky, the field of radio astronomy was born. Radio astronomy (RA) is a branch of astronomy that is concerned with observations of the sky at radio frequencies using radio telescopes. The observable radio waves constitute the EM spectrum from around 30 MHz up to 40 GHz (wavelength 10 m to 7 mm). Waves with higher wavelength than 10 m are blocked by the earth's ionosphere and the ones with lower wavelength than about 1 cm are absorbed by atmospheric gases [6]. Radio waves are minimally affected by the atmosphere (in contrast to optical waves) as shown in Figure 1.3. Therefore even under cloudy skies and

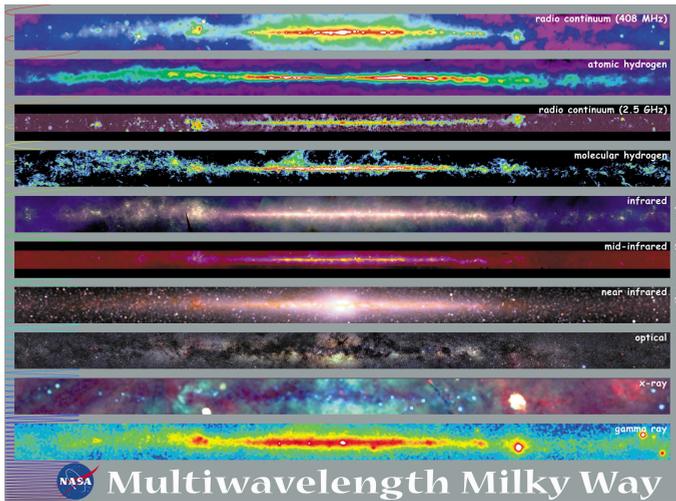


Figure 1.2: Multi-wavelength view of the Milky Way galaxy. (Image courtesy of [5])

during daytime observations at radio frequencies can be performed.

Apart from blockage by the atmosphere, optical telescopes only observe the light emitted from the stars. On the other hand, highly sensitivity radio telescopes can look back in time by observing weak radio emissions originating from the hydrogen line and ionized gases even before the first stars were created. These observations provide us with evidence to study the creation of the universe and formations of the stars. In Big Bang cosmology, the era when the first stars started forming and the neutral gases started becoming ionized again is referred to as the "Epoch of Reionization" (EoR). EoR provides a fundamental case of study for radio astronomical research. In particular, studies of HI emissions are fundamental to radio astronomy. These studies have already resulted in the discovery of the CMB. As can be seen from Figure 1.2, the interstellar medium appears completely dark (empty) in optical observations. However, at radio frequencies, this medium is almost entirely filled with a faint background radiation (CMB). CMB is the oldest detectable radiation in the universe. Furthermore, radio frequency observations reveal more accurately the scale, structure and dynamics of the galaxies. In particular, radio galaxies, i.e. active galaxies that emit radio frequencies due to the synchrotron process, are best (or only) observable at radio frequencies. As an example, the view of the radio galaxy Centaurus A at optical as well as radio frequencies is shown in Figure 1.4. This figure clearly shows that the radio extent of this galaxy (shown as a blue-green radiation) is observable far beyond its optical extent.

1.2. RADIO INTERFEROMETRY

The first purpose-built radio telescope was a 9-meter parabolic dish constructed by Grote Reber in his backyard in Wheaton, Illinois in 1937. Using this telescope, he conducted a sky survey which is often considered as the beginning of the field of radio astronomy.

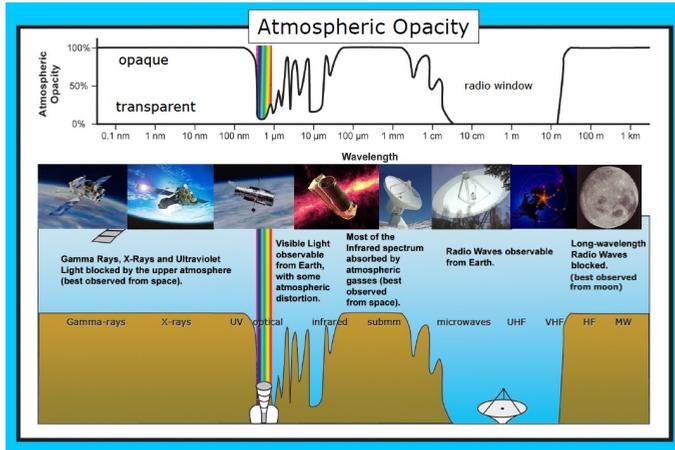


Figure 1.3: Atmospheric absorption (Image courtesy of [7])

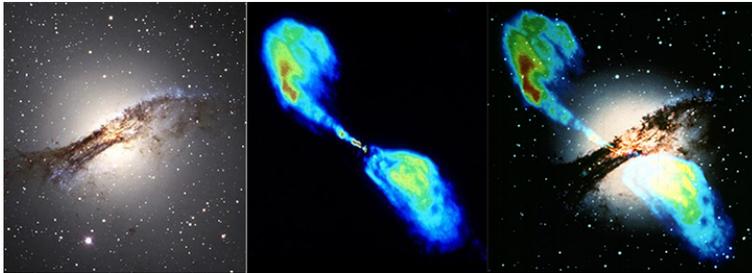


Figure 1.4: From left to right: Centaurus A galaxy at optical frequencies, at radio frequencies and superposition of the two views. (Image credit: NASA and ESA)

The resolution of a telescope is proportional to the wavelength of observation divided by its diameter. Radio waves are about 1 million times longer than optical waves. Therefore, to attain a similar resolution as optical telescopes, very large radio telescopes are required. Furthermore, radio telescopes must be able to detect weak signals from very distant objects. As a consequence, sensitivity is a defining factor for radio telescopes. Sensitivity of a radio telescope is proportional to its total collecting area. Hence, to make high resolution and high dynamic range images of the radio sky, telescopes with large apertures and large collecting areas are required.

Building extremely large dishes to achieve high resolution images is impractical. Instead, in a technique called radio interferometry (RI), or aperture synthesis, a large telescope is synthesized by combining the received signals from an array of radio telescopes with smaller apertures [8] as shown in Figure 1.5. In this technique, the aperture of the resulting telescope is equal to the maximum distance between the individual telescopes and its total collecting area is the sum of the collecting areas of the individual elements.

Radio interferometry was first introduced by Martin Ryle in 1946 [9] and has become

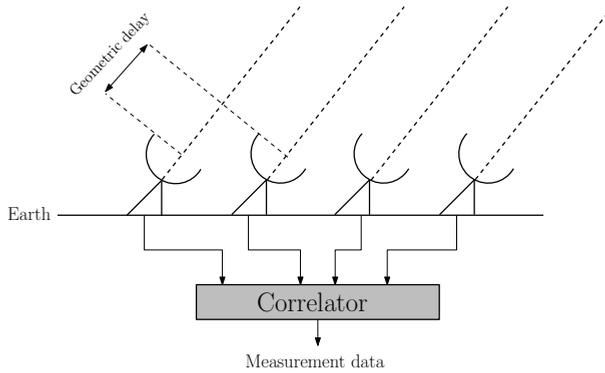


Figure 1.5: Interferometry

the principal instrument for the study of distant radio sources. In this technique, arrays of tens of elements, extending from hundreds of meters to nearly the diameter of the earth with dimensions up to hundreds of millions of wavelengths are interconnected and used in the study of the radio sky.

As shown in Figure 1.5, radio interferometry is based on estimating the propagation time (geometric delay) of celestial signals impinging on the telescope array, as this relates to the angle of arrival. In this technique, the direction and intensity of cosmic signals are obtained by measuring the cross-correlation of signals received by an array of telescopes placed at different locations. Doing so, an image of the radio sky over the Field of View (FoV) of the telescope array is formed by combining the cross-correlations of the telescope pairs that constitute the interferometer. Since the celestial sources are statistically stationary, not all interferometer elements need to be present simultaneously. In this regard, in a technique called “earth rotation synthesis” [8] the earth’s rotation is used to collect more measurement points. Moreover, radio telescopes can collect observations over a wide range of frequency channels.

1.3. RADIO INTERFEROMETRIC IMAGING

Based on observations at radio frequencies, celestial sources and phenomena that originate radio emissions can be studied. To study the celestial objects and the astrophysical processes that are responsible for radio emissions, images of the radio sky must be formed. Proper astrophysical interpretations require the image to have a high resolution and dynamic range. The objective of radio interferometric imaging is to obtain an estimate of the intensity and Direction of Arrival (DoA) of the celestial sources over the FoV of the telescope array. This translates to a linear inverse problem: given the radio telescope measurements the aim is to invert the measurement process to attain an estimate of the intensity of the celestial sources that originated the observed radio waves. The measurement process in radio astronomy can be described as a linear measurement equation of the form

$$\tilde{\mathbf{r}} = \mathbf{M}\boldsymbol{\sigma} + \mathbf{e} \quad (1.1)$$

where σ is the image of the radio sky to be estimated based on correlation data $\tilde{\mathbf{r}}$. \mathbf{M} describes the measurement process and \mathbf{e} is the noise affecting the measurements. This equation will be described in more detail in Chapter 2. A schematic of the inverse problem in radio astronomy is shown in Figure 1.6.

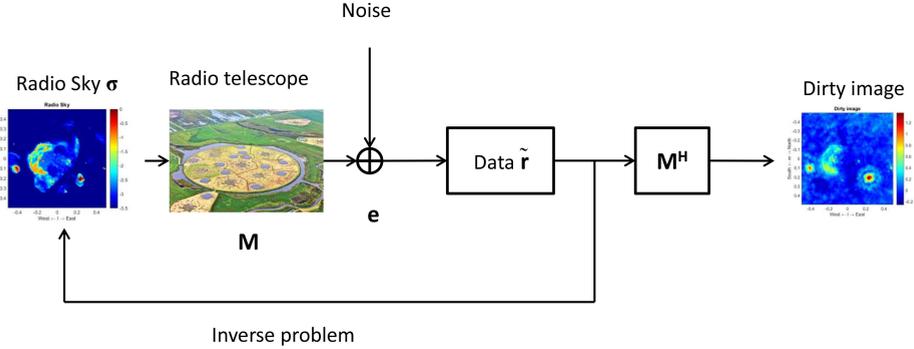


Figure 1.6: Inverse problem in radio astronomy

In this figure, dirty image refers to the image obtained by back-projecting the received signals by multiplication with \mathbf{M}^H . As can be seen from the figure, the dirty image contains many artifacts and cannot be accepted as a good estimate of the true radio sky. One major issue in radio interferometric imaging is the sparse sampling of the mutual coherence by the available baselines. Due to the physical limitations of the radio telescopes, the measurement set is inevitably incomplete and due to the inherent loss of information in the measurement process the inverse problem is ill-posed. Therefore, trying to directly invert the measurement process to obtain an estimate of the sky map results in noise amplification. Therefore, to achieve astronomically plausible images, side information and constraints available about the sky image need to be incorporated in the imaging process. This ensures to contain the noise amplification effect on the intensity estimate to within the bound as dictated by the constraints. Therefore, correct regularizing assumptions are required to ensure a unique and physically meaningful sky map estimate.

The measurement accuracy is further degraded by propagation and instrumental defects. These issues are mainly addressed in the calibration process that is done before imaging and/or in combination with the imaging process. In spite of complicated calibration procedures, radio telescope measurements are unavoidably affected with errors [10]. As a result of these issues, advanced and expensive processing, for both image reconstruction and calibration, is necessary to reconstruct meaningful images of radio sources from the incomplete, inexact and noisy data. [10]

Finally, inverse problems are not restricted to radio interferometric imaging. Similar problems arise in different imaging fields where an array of sensors is used to make indirect observations of an object by observing waves originating (or scattered) from that object. These fields include, but are not restricted to, biomedical imaging (tomography,

ultrasound, etc.) and geophysics/seismic imaging. The mathematical techniques developed for one field can be easily carried over to another field with minor changes. Therefore, advances in imaging techniques in radio astronomy will also have a direct impact on the related imaging areas.

1.4. NEXT GENERATION RADIO TELESCOPES

Since the invention of radio interferometry in 1946, increasingly larger and more sensitive radio interferometers are being built with the aim to achieve high resolution and high dynamic range images of the radio sky. Driven by ambitious science goals, currently the Square Kilometre Array (SKA) is being designed and is expected to be the largest and most sensitive radio telescope that has been built to date [4]. The SKA contains several instruments that together cover most of the radio spectrum. The telescope sites are planned to be in the southern hemisphere, in radio-quiet zones in Southern Africa and Australia. Some of the path-finders and precursors of the SKA are currently in function providing high resolution, high dynamic range and wide field of view observations of the radio sky. These telescopes include, but are not limited to, Australian SKA Pathfinder (ASKAP) [11] and the Murchison Wide-field Array (MWA) [12] in Western Australia, MeerKAT [13] and Hydrogen Epoch of Reionization Array (HERA) [14] in Southern Africa and the Low Frequency Array (LOFAR) [1] in Europe. Figure 1.7 shows an overview of these telescopes.

1.5. SKA OVERVIEW

The broad goal of the SKA is to study the history of the universe in more detail. In fact, the main question that derived the design and construction of the SKA was “What size radio telescope would it take to permit us to read the history of the Universe as written in the language of its most abundant constituent, Hydrogen?” [18]. The answer is reflected in the name of the instrument, one square kilometer.

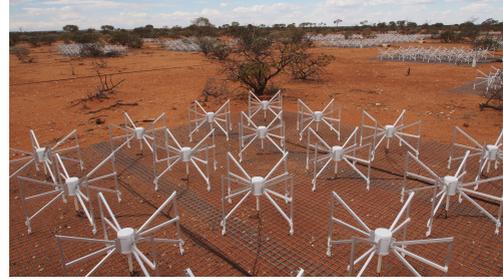
1.5.1. SCIENCE CASES

While a broad group of engineers are designing and implementing the SKA, astronomers are identifying in more detail the science goals to be achieved from this large and multidisciplinary project. The astrophysical goals expected to be achieved with the SKA are detailed in [18]. Some of the key scientific drivers of the SKA are as briefly stated below.

- Investigating how the very first stars and galaxies formed just after the big bang by probing the EoR based on observations of the HI in much greater detail.
- Studying the galaxy evolutions by probing neutral Hydrogen using both wide and deep surveys of the galaxy.
- Challenging Einstein's general theory of relativity to the limits by probing into extreme environments such as black holes and orbiting neutron stars by the help of high precision pulsar timing to observe the gravitational waves.
- Helping scientists understand the nature of a mysterious force known as the dark energy with more observational evidence to model this phenomenon.



(a) ASKAP



(b) MWA



(c) MeerKAT



(d) HERA



(e) LOFAR

Figure 1.7: New generation radio telescope arrays (images courtesy of CSIRO [15], the SKA Telescope [16] and LOFAR [17])

- Understanding the vast magnetic fields which permeate the cosmos (cosmic magnetism) and have a vast influence on the cosmic processes by providing a detailed magnetic map of the Milky Way.
- Observing the bursting sky to map the universe's plasma content by the study of fast radio bursts.
- Understanding why the rate of star formation has changed over the course of the universe by detecting signals and making resolved images of super star clusters when the universe was 0.5 to 1 billion years old.
- Studying the formation of the planets from coalescing particles by observing at exact radio frequencies matched to the size of the particles.

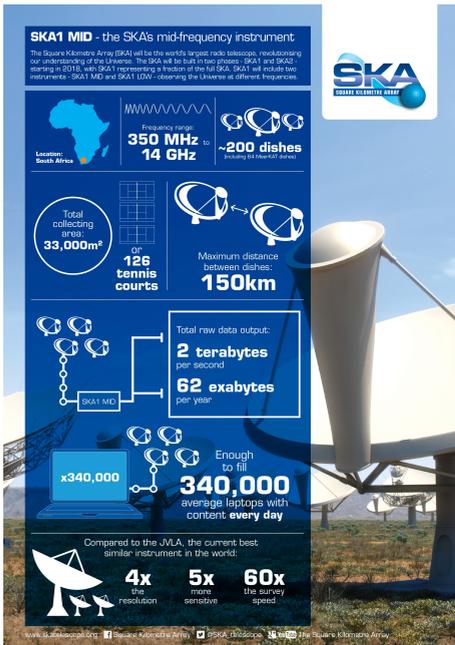
Apart from these main science goals, the increased observational power of the SKA may lead to discoveries of new astrophysical phenomena [19, 16].

1.5.2. THE INSTRUMENT

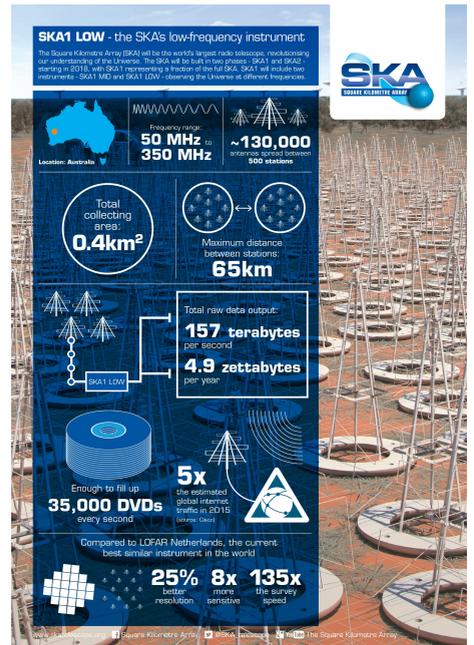
The science cases of the SKA can be divided into imaging and non-imaging cases [20]. The imaging cases are concerned with providing high accuracy sky images from which quantities such as source count or background noise level can be inferred. The non-imaging cases deal with the transient sky and short-time phenomena such as pulsars and bursts. In this dissertation, we focus on the imaging cases.

For imaging purposes, the SKA is expected to produce large images with more than 10^8 pixels and a dynamic range of over 60 dB [21]. It will be one to two orders of magnitude more sensitive than the current radio telescopes. To achieve the high resolution requirements, the SKA will be composed of a large aperture consisting of millions of coherently connected antennas extended over an area of about 3000 kilometers. Furthermore, to provide the required sensitivity, the SKA has a large total collecting area of about one square kilometer [4]. The SKA is composed of several instruments that together cover a large frequency band between 50 MHz to 20 GHz. [16]. The antennas used to build the aperture array for the SKA are composed of sparse aperture arrays and dense aperture arrays for low frequencies, parabolic reflectors with phased array feeds and reflectors with single pixel feeds [4].

The design and construction of the SKA is divided into two phases. Phase-one SKA (SKA1) baseline design is detailed in [19] followed by a more feasible re-baselined design detailed in [22]. The exact design for the second phase of the SKA is still under consideration and is expected to follow the current phase-one design with more collecting area and larger maximal baselines. The most accurate design of SKA currently available is the re-baselined SKA1 design. SKA1 will be located in South Africa and Australia containing two different instruments. The Australian and South African core arrays are respectively called "SKA1-LOW" and "SKA1-MID" denoting they operate in low-frequency and mid-frequency range of the SKA, respectively. Both of these core arrays contain remote stations. Some of the specifications of these instruments, together with an artist's impression of the telescope types are shown in Figure 1.8. At the time of writing, one SKA1-LOW station has been commissioned, as well as 64 dishes for the SKA1-MID [23].



SKA1 MID (image courtesy of [24])



SKA1 LOW (image courtesy of [25])

Figure 1.8: SKA1 instruments

1.5.3. IMAGING AND DATA REDUCTION PIPELINE

The imaging science cases heavily depend on efficient signal and image processing of the SKA. Focusing on the imaging science cases, a potential imaging pipeline for SKA instruments, similar to the current processing pipeline for LOFAR has been described in [16, 26]. The data acquisition and signal processing steps are broken down into three main stages: (i) Station Processing, which involves digitalization of the station data and division into frequency channels, (ii) Central Signal Processing (CSP), which involves the formation of the raw correlation data from the station signals and (iii) Science Data Processing (SDP), where the calibration and imaging operations are performed to form radio astronomical images. In the next chapter we will discuss the data model and the processing pipeline in more detail.

1.5.4. IMAGING CHALLENGES

The imaging science requirements of the SKA are very ambitious and result in an exceptional need for accurate measurements. On the other hand, the increased sensitivity, resolution and sky coverage of the new instruments pose additional challenges to the current radio astronomical imaging processing pipeline.

A wide field of view of the individual antennas of the phased array telescope elements, long non-coplanar baselines, very large amounts of data and the required computing power with low energy consumption are among the technical challenges of the

SKA. More precisely, to satisfy high fidelity imaging requirements of radio interferometric images, next generation telescopes are expected to produce streams of extremely large data volumes. The expected data volumes for SKA1-MID instrument is 2 terabytes per second and for the low-frequency instrument is 157 terabytes per second [16]. These data volumes cannot be stored for long astronomical observations and have to be processed and reduced by post-processing in near-realtime [20, 26]. The acquired data deluge pushes the computing requirements of the SKA to Exascale limits [21]. Transport of the large amount of data and the processing power required to form astronomically sound images from the captured data introduce significant challenges. These issues make the SKA a data and compute-intensive application.

Image formation is the main computational bottleneck of the SKA. On the other hand, image formation provides a means of data reduction by performing transformation from the size of the data to image size. As a result, the burden in dealing with large volumes of data is shifted to imaging algorithms. These algorithms are required to be extremely efficient in dealing with data and produce reliable high fidelity images in soft-realtime. However, the current imaging algorithms are not scalable to the image and data size requirements of the SKA. Furthermore, the imaging computational requirements imposed by the SKA are not affordable by the current high performance computers and pushes the computational requirements to Exascale limit. Therefore, imaging algorithms have to be adjusted to the scale of SKA to reduce the burden from the hardware without sacrificing the quality of the reconstructed image. Many pixel-based imaging algorithms, such as the widely-used CLEAN [3] algorithm, are not scalable to the size of the required images and perform very slow in high resolution scenarios. Therefore, there is an urgent need for new efficient imaging algorithms.

Moreover, regardless of the amount of collected data, there is an inherent loss of information in the measurement process due to the physical limitations. Therefore, to recover physically meaningful images additional information in the form of constraints and regularizing assumptions are necessary. Regularization is applied by imposing prior knowledge and assumptions about the sky map on the solution of the imaging problem. These prior assumptions depend on the science case under study. Choosing a correct regularization that can be applied in an efficient manner is very important for the next generation of radio telescopes.

1.6. SCOPE OF THIS DISSERTATION

In this section, we present the scope, goals and research questions pursued in this dissertation. In addition, we give a brief summary of the main results from each chapter.

1.6.1. DRIFT

The work presented in this dissertation was supported by the NWO DRIFT (Data Reduction and Image formation for Future radio Telescopes) project under the context of the ASTRON-IBM DOME project [27] collaboration. Netherlands Institute for Radio Astronomy (ASTRON) and IBM are co-developing a computing system that is expected to become the IT backbone of the SKA. They face many hurdles how to gather, transmit, store and analyze an incredible volume of radio signals generated by the SKA. This project

is considered as the ultimate big-data challenge and is introduced under the name of Big Bang Big Data (BBBD). The purpose of DRIFT was to consider novel processing and imaging algorithms that are feasible at the scales required by the SKA.

1.6.2. PROBLEM STATEMENT AND RESEARCH GOALS

Image formation is the most computationally expensive step in the SKA processing pipeline [26]. Imaging methods are implemented in the SDP pipeline as iterative algorithms consisting of many iterations of transforms between the domain of correlation data and the image. Massive data sizes of the SKA cause these repeated steps to be very demanding in terms of both amount of computations and storage. As analyzed by Jongerius et al. [26], the image-data domain transform operations are expected to be responsible for about 50% of the compute load of the SKA [20]. Furthermore, due to the unprecedented data volumes, current storage facilities are not able to store the data and there is a need for fast imaging algorithms that are able to process the data in near-realtime. Near-realtime means that the observations must not be halted due to processing of the data. Furthermore, reliable image reconstruction for the science cases of the SKA (high sensitivity and dynamic range) requires novel regularization schemes. Therefore, efficient and accurate regularization methods need to be integrated in the imaging algorithm.

Based on these arguments, the main requirements of the image formation algorithms for the next generation radio telescopes can be summarized in three categories: (i) computationally efficient, (ii) near-realtime performance and (iii) reliable and accurate regularization. In this context, the main question of the current dissertation is *how to design novel imaging algorithms for the SKA science cases that are fast, efficient and scalable to the extent of the SKA and produce reliable results?* In consequence, the main aim of the current work is to *introduce advanced algebraic techniques together with custom-made regularization schemes to speed up the image formation pipeline of the next generation radio telescopes.*

The main research question can be further broken down into a number of research objectives that are pursued in this dissertation.

- A framework for introducing prior knowledge on the image (beyond simple point sources) as required to regularize the ill-posed inversion problem;
- New efficient numerical techniques to solve the resulting optimization problems in quasi-realtime;
- Advanced regularization methods for scalable imaging.

The solutions are in turn required to be accurate (free of artifacts), robust, computationally efficient, scalable, fully automated yet with some tuning flexibility.

1.6.3. OUTLINE AND CONTRIBUTIONS

In this dissertation, we focus on introducing new imaging methods for the imaging pipeline of the next generation radio telescopes. Therefore, we consider the radio interferometric imaging problem in much detail. We present the topics discussed in each chapter of this dissertation in the following outline.

Chapter 2: In this chapter, we introduce the radio interferometric measurement model, in familiar terms for astronomers as well as in array signal processing formulations and show both perspectives result in the same discrete measurement model. We further align the measurement model with the proposed processing pipeline of the SKA. We interpret the measurement equation from the point of view of inverse problems and show the analogy with compressive second-order statistic estimation. We finalize the chapter with introducing two main statistical estimation frameworks for the imaging problem formulation.

Chapter 3: In this chapter, we analyze the imaging system based on the measurement model developed in Chapter 2 to investigate the identifiability of the sky intensity distribution based on the measured telescope data. We conclude that the radio interferometric imaging problem is highly ill-posed and thus finding a numerically stable and physically meaningful solution requires regularization. We continue by introducing regularization methods and reviewing some common regularization methods applied in RA. Furthermore, we borrow the concept of preconditioning from numerical linear algebra and introduce a framework to incorporate regularization assumptions into the measurement model.

Chapter 4: In this chapter, we start our quest towards efficient regularization for radio interferometric imaging by proposing two regularization methods for the direct solution method. The first method, benefits from a weighted truncation of the eigenvalue decomposition of the image deconvolution matrix and the other is based on the prior knowledge of the “dirty image” using the available data. The latter has the design of the data-driven regularization method that we apply on the iterative algorithm proposed later on in Chapter 5. The methods are evaluated by simulations as well as by actual data from the LOFAR station.

Chapter 5: In this chapter, we present a general algorithmic framework based on a Bayesian-inspired regularized maximum likelihood formulation of the radio astronomical imaging problem with a focus on diffuse emission recovery from limited noisy correlation data. The algorithm is dubbed PRIor-conditioned Fast Iterative Radio Astronomy (PRIFIRA) and is based on a direct embodiment of the regularization operator into the system by right preconditioning. The resulting system is then solved using an iterative method based on projections onto Krylov subspaces. We motivate the use of a beamformed image (which includes the classical “dirty image”) as an efficient preconditioner. Iterative reweighting schemes generalize the algorithmic framework and can account for different regularization operators that encourage sparsity of the solution. We show the generality of the proposed method in terms of regularization schemes while maintaining a competitive reconstruction quality with the current reconstruction techniques. Furthermore, we show that exploiting Krylov subspace methods together with proper noise-based stopping criteria results in a great improvement in imaging efficiency.

Chapter 6: In this chapter, we present some further realistic simulation results to compare PRIFIRA with the state-of-the-art radio interferometric imaging techniques. Furthermore, we test the performance of PRIFIRA on data sets from (i) a single station of LOFAR and (ii) core LOFAR stations (Superterp).

Chapter 7: In this chapter, we introduce a remedy on the design of a right preconditioner for differentiating between point sources and extended emissions that can co-exist in a radio interferometric image. We show that by designing a multi-dictionary basis we can guarantee super-resolution recovery for point sources as well as reliable recovery of extended emissions. In a preliminary effort, we generalize the greedy method of active-set to incorporate multi-dictionary extensions. Furthermore, we apply the multi-dictionary extension on the PRIFIRA-based sparse recovery methods presented in Chapter 5.

Chapter 8: In this chapter, we focus on introducing advanced and scalable regularization assumptions into the radio interferometric imaging problem. More precisely, we propose a scalable regularization scheme by splitting the image under scrutiny into source occupancy regions which we call facets. Each facet incorporates its own complex regularization assumptions. Leveraging a stochastic primal dual algorithm, we show the potential merits of applying facet-based regularization on the radio-interferometric images which results in both computation time and memory requirement savings.

Chapter 9: In this chapter, we present a summary of the main results and conclude the dissertation. Furthermore, we suggest some possible future directions for radio interferometric imaging with the future radio telescopes.

1.6.4. LIST OF PUBLICATIONS

Some of the works presented in this dissertation have already resulted in publications in a journal as well as conferences/workshops. A list of the publications are as follows.

JOURNALS

- S. Naghibzadeh and A.-J. van der Veen, “PRIFIRA: General regularization using prior-conditioning for fast radio interferometric imaging”. In *Monthly Notices of the Royal Astronomical Society* (2018).
- “Semi-realtime joint covariance sketching and imaging for future radio telescopes”. In preparation.

CONFERENCES AND WORKSHOPS

- S. Naghibzadeh and A.-J. van der Veen, “Bayesian-inspired regularization using prior-conditioning for fast radio interferometric imaging”. In *proceedings of the international BASP frontiers workshop*, February 2019, invited poster.
- S. Naghibzadeh, A. Repetti, A.-J. van der Veen, Y. Wiaux, “Facet-Based Regularization for scalable radio-interferometric imaging”. In *proceedings of the 26th European Signal Processing Conference (EUSIPCO)*, pp. 2696 – 2700, September 2018.

- S. Naghibzadeh and A.-J. van der Veen, “Fast and accurate radio interferometric imaging using krylov subspaces”. In *proceedings of the 7th IEEE international workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 1 – 5, December 2017.
- S. Naghibzadeh and A.-J. van der Veen, “Radioastronomical Least Squares image reconstruction with iteration regularized Krylov subspaces and beamforming-based prior conditioning”. In *Proceedings the 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* , pp. 3385 – 3389, March 2017.
- S. Naghibzadeh, A. Mouri Sardarabadi and A.-J. van der Veen, “Point and beam-sparse radio astronomical source recovery using non-negative least squares”. In *Proceedings of the 9th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 1 – 5, July 2016.
- S. Naghibzadeh, A. Mouri Sardarabadi and A.-J. van der Veen, “Radioastronomical image reconstruction with regularized least squares”. In *Proceedings the 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* , pp. 3316 – 3320, March 2016.

2

DATA MODEL AND PROBLEM FORMULATION

Every man takes the limits of his own field of vision for the limits of the world.

Arthur Schopenhauer

In this chapter, we introduce the signal processing data model for radio interferometric imaging. The imaging techniques developed in the rest of the chapters are based on the data model introduced in this chapter. We start by introducing some of the most important symbols, notations, operators and matrix relations used in this chapter and the rest of the thesis. We continue with describing the radio interferometric imaging measurement model that is widely used in the astronomical society. Furthermore, we consider radio interferometric imaging as an array processing problem and describe the array processing data model for radio interferometric imaging.

Moreover, based on the developed data model for radio interferometric imaging, we discuss the properties of the radio astronomical imaging problem from the point of view of inverse problems and show the analogy with the power spectrum estimation problem. Furthermore, we formulate the radio interferometric imaging problem as an intensity estimation problem.

2.1. SYMBOLS, NOTATIONS, OPERATORS AND MATRIX RELATIONS

2.1.1. SYMBOLS

a, A	plain lowercase and uppercase letters denote scalars
\mathcal{A}	calligraphic letters denote continuous functions
\mathcal{A}	calligraphic boldface letters denote operators
\mathbf{a}	boldface letters denote column vectors
\mathbf{A}	boldface capital letters denote matrices
\mathbf{a}_i	for a matrix \mathbf{A} denotes the i th column of \mathbf{A}
$a_{i,j}$	for a matrix \mathbf{A} denotes the i, j th entry
$\mathbf{1}$	vector with all the elements equal to 1
\mathbf{I}	identity matrix of appropriate size
\mathbf{I}_p	$p \times p$ identity matrix
\mathbf{e}_i	vector with all zero elements with the i th element equal to one
$\mathbf{0}$	vector with all the elements equal to 0
j	square root of -1

2.1.2. OPERATORS

$E\{\cdot\}$	expectation operator
$(\cdot)^T$	transpose operator
$(\cdot)^*$	complex conjugate operator
$(\cdot)^H$	Hermitian transpose
$(\cdot)^\dagger$	Moore-Penrose psuedo-inverse operator
$\ \mathbf{a}\ _p$	p -norm of a vector \mathbf{a} defined as $\ \mathbf{a}\ _p^p = \sum a_i ^p$
$\ \mathbf{A}\ _F$	Frobenius norm (matrix norm)
$\text{trace}(\mathbf{A})$	computes the sum of the diagonal elements of a matrix
$\det(\mathbf{A})$	the determinant of the matrix \mathbf{A}
$\text{vect}(\mathbf{A})$	stacks the columns of the argument matrix \mathbf{A} to form a vector
$\text{vectdiag}(\mathbf{A})$	stacks the diagonal elements of the argument matrix \mathbf{A} to form a vector
$\text{diag}(\mathbf{a})$	is a diagonal matrix with its diagonal entries from the argument vector \mathbf{a}
$\text{diag}(\mathbf{A})$	$= \text{diag}(\text{vectdiag}(\mathbf{A}))$ if the argument is a matrix (\mathbf{A})
\otimes	Kronecker product
\circ	Khatri-Rao product (column-wise Kronecker product)
\odot	Hadamard (element-wise) product
\mathbb{R}	the set of real numbers
$\mathbb{R}^{m \times n}$	the set of real-valued m by n arrays
\mathbb{C}	the set of complex numbers
\in	belongs to

2.1.3. MATRIX PRODUCT RELATIONS

Kronecker, Khatri-Rao and Hadamard products are used in this thesis. Based on model matrices \mathbf{A} and \mathbf{B} , these products are defined here and some useful relations are presented. Assuming a_{ij} is the element of \mathbf{A} corresponding to the i th row and j th column,

the Kronecker product is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \quad (2.1)$$

The Khatri-Rao product is a column-wise Kronecker product defined as

$$\mathbf{A} \circ \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \mathbf{a}_2 \otimes \mathbf{b}_2, \dots] \quad (2.2)$$

where \mathbf{a}_i and \mathbf{b}_j denote the i th and j th columns of \mathbf{A} and \mathbf{B} , respectively. Furthermore, the Hadamard product of \mathbf{A} and \mathbf{B} is defined as

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots \\ a_{21}b_{21} & a_{22}b_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \quad (2.3)$$

The following properties are used throughout the thesis (for matrices and vectors with compatible dimensions):

$$\begin{aligned} (\mathbf{B}^T \otimes \mathbf{A}) \text{vect}(\mathbf{X}) &= \text{vect}(\mathbf{A}\mathbf{X}\mathbf{B}) \\ (\mathbf{B} \otimes \mathbf{A})^H &= (\mathbf{B}^H \otimes \mathbf{A}^H) \\ (\mathbf{B} \otimes \mathbf{A})^{-1} &= (\mathbf{B}^{-1} \otimes \mathbf{A}^{-1}) \\ (\mathbf{B}^T \circ \mathbf{A})\mathbf{x} &= \text{vect}(\mathbf{A}\text{diag}(\mathbf{x})\mathbf{B}) \\ (\mathbf{B}\mathbf{C} \otimes \mathbf{A}\mathbf{D}) &= (\mathbf{B} \otimes \mathbf{A})(\mathbf{C} \otimes \mathbf{D}) \\ (\mathbf{B}\mathbf{C} \circ \mathbf{A}\mathbf{D}) &= (\mathbf{B} \otimes \mathbf{A})(\mathbf{C} \circ \mathbf{D}) \\ (\mathbf{B}^H \mathbf{C} \circ \mathbf{A}^H \mathbf{D}) &= (\mathbf{B} \circ \mathbf{A})^H (\mathbf{C} \circ \mathbf{D}) \\ \text{vectdiag}(\mathbf{A}^H \mathbf{X} \mathbf{A}) &= (\mathbf{A}^* \circ \mathbf{A})^H \text{vect}(\mathbf{X}). \end{aligned}$$

2.2. MEASUREMENT MODEL FOR IMAGE FORMATION

In this section, we discuss the data model (i) as developed in the conventional radio astronomy and (ii) from an array processing point of view. Furthermore, we compare the two approaches and show that they result in a similar discrete measurement model. The measurement equation is the cornerstone for developing imaging and data reduction algorithms.

2.2.1. MEASUREMENT MODEL IN CONVENTIONAL RADIO ASTRONOMY

An interferometer is a device that measures the correlation of the electric field received by a pair of telescopes (antennas). The electric field is generated from the celestial sources. Due to the large distance of celestial sources from the earth, in the far field context, the sources are assumed to be placed on an imaginary sphere of unit radius around the earth, called the “celestial sphere”.

We consider a setting in which an interferometer is observing a portion of the celestial sphere at a narrow frequency band centered at f . Figure 2.1 shows a schematic of a

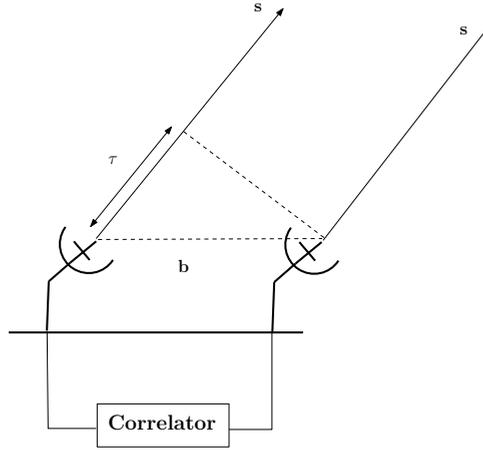


Figure 2.1: A two-element interferometer

two-element interferometer. The unit norm vector \mathbf{s} denotes the direction of arrival of a celestial source. We denote the intensity (brightness) of the source as a function of \mathbf{s} by $I(\mathbf{s})$. In this setting, denoting the difference between the antenna pair locations by \mathbf{b} , for a narrow-band observation, this difference in location determines a delay computed as $\tau = \frac{\mathbf{s}^T \mathbf{b}}{c}$, with c indicating the speed of light. Therefore, the corresponding phase delay is computed as $2\pi f\tau = 2\pi \mathbf{s}^T \mathbf{b} / \lambda$ where $\lambda = \frac{c}{f}$ is wavelength of the impinging electric field. Normalizing the difference between the antenna pair locations by the wavelength, we call $\mathbf{b}_\lambda = \mathbf{b} / \lambda$ the baseline vector.

Correlations, a.k.a. “visibilities”, are complex quantities that are related to the sky brightness distribution via the relation [28]

$$\mathcal{V}(\mathbf{b}_\lambda) \approx \int_{4\pi} I(\mathbf{s}) e^{-2\pi j \mathbf{s}^T \mathbf{b}_\lambda} d\Omega. \quad (2.4)$$

In this relation, $\mathcal{V}(\mathbf{b}_\lambda)$ is the visibility as a function of the frequency-dependent baseline vector \mathbf{b}_λ and $d\Omega$ is the surface area over the sphere. For this integration it has been assumed that the celestial sources are spatially incoherent. Therefore, the elements $d\Omega$ over the source intensity distribution are uncorrelated and the integration can be applied. The integration in Equation (2.4) is calculated over the entire celestial sphere (4π). However, in practice the FoV is limited by factors such as the primary beam of the interferometer. Therefore, the integration region can be reduced to the FoV of the telescope array. The FoV is mainly defined based on the angular scale of the primary beam of the radio interferometer [8, 29].

Observing Equation (2.4), we see a relationship in the form of a Fourier transform between the intensity distribution function $I(\mathbf{s})$ and the visibility function $\mathcal{V}(\mathbf{b}_\lambda)$ which is known as the Van-Cittert-Zernike theorem for radio interferometry [8]. As a result, by measuring the complex visibility function of an astronomical object, we can reconstruct a map of two-dimensional brightness distribution of the object.

“Synthesis mapping” or imaging is to determine the intensity distribution $I(\mathbf{s})$ for

the area of interest on the sky using the earth-bound visibility measurements $\mathcal{V}(\mathbf{b}_\lambda)$. In synthesis imaging, a point at the center of the field of view of the interferometer is chosen as the “phase reference position” or “phase tracking center”. The position of this point in the FoV is denoted by the unit norm vector \mathbf{s}_0 . The position of any other point \mathbf{s} in the FoV is calculated with respect to the phase reference point, i.e., $\mathbf{s} = \mathbf{s}_0 + \boldsymbol{\rho}$ as shown in Figure 2.2. Denoting the antenna collecting area in direction \mathbf{s}_0 as A_0 and introducing the

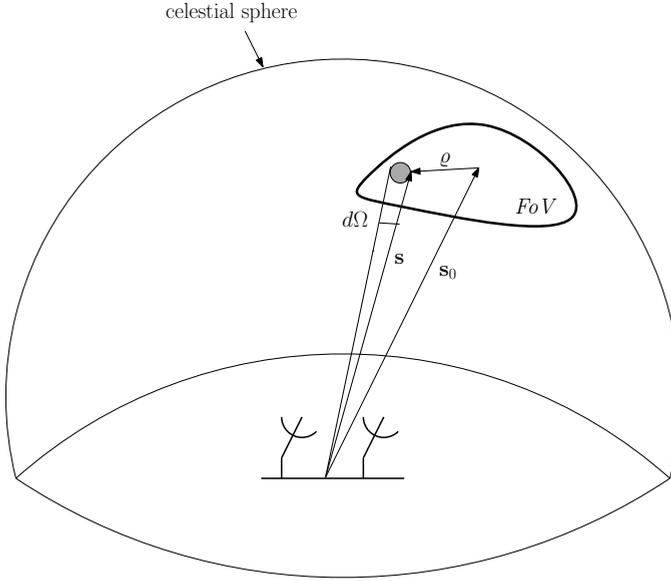


Figure 2.2: Source and interferometer geometry

normalized reception pattern as $A_N(\boldsymbol{\rho}) = A(\boldsymbol{\rho})/A_0$, the complex visibility, with amplitude $|\mathcal{V}|$ and phase ϕ_v , is defined as

$$\mathcal{V}(\mathbf{b}_\lambda) = |\mathcal{V}| e^{j\phi_v} = \int_{FoV} A_N(\boldsymbol{\rho}) I(\boldsymbol{\rho}) e^{-2\pi j \boldsymbol{\rho}^T \mathbf{b}_\lambda} d\Omega, \quad (2.5)$$

[8, 28, 30]. We can infer from this equation that $\mathcal{V}(\mathbf{b}_\lambda) = \mathcal{V}^*(-\mathbf{b}_\lambda)$.

To identify the measurement equation for radio interferometric imaging a Cartesian coordinate system is used. This coordinate system is shown schematically in Figure 2.3 for a two-element interferometer. We assume the antennas are placed in a field on the surface of the earth denoted by \mathcal{M} . A point in the approximate center of the measurement field is chosen as the origin of the coordinate system, denoted by O . Assuming a field of view \mathcal{F} on the celestial sphere, a point at the approximate center of the field of view is chosen as the phase reference center, denoted by C [31], with the vector \mathbf{s}_0 pointing towards it. The antennas in the interferometer track the phase reference center over the course of observation. In the conventional coordinate system in radio astronomy, a right-handed Cartesian coordinate system is used to measure the baseline vector. The magnitude of the baseline vector is represented in units of wavelength at the center frequency of the observing band. The components of the baseline coordinate system are

(u, v, w) . In this system, the w axis points towards the phase reference center and the w component is measured in direction \mathbf{s}_0 . The u and v axes lie in a plane perpendicular to the direction of the phase reference center and the w axis. This plane is called the $u-v$ -plane or baseline plane.

2

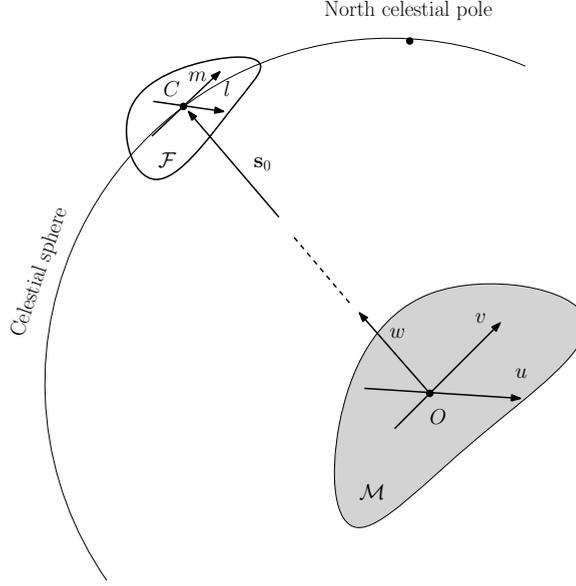


Figure 2.3: Coordinate system of the baseline and the source under observation

The intensity distribution is measured by projecting the celestial sphere on a plane parallel to the $u-v$ plane centered at C . This plane has axis l and m parallel to u and v in the baseline plane. Positions on the sky are defined in the l and m components which are the direction cosines measured with respect to the u and v axis, respectively. In this notation, l is the dimension along the equatorial plane pointing towards the east and m is pointing towards the north. It is also customary in radio interferometry to introduce a third auxiliary dimension along the reference direction \mathbf{s}_0 , called n where $n = \sqrt{1 - l^2 - m^2}$. This dimension accounts for the spherical nature of the image plane by posing limitations on the values of l and m . The (l, m, n) components describe the direction cosines of a source.

Starting from Equation (2.5) and considering the aforementioned Cartesian coordinate system, the parameters in Equation (2.5) in terms of the coordinate system are presented as

$$\mathbf{b}_\lambda^T \mathbf{s}_0 = w \quad (2.6)$$

$$\mathbf{b}_\lambda^T \mathbf{s} = ul + vm + w\sqrt{1 - l^2 - m^2} \quad (2.7)$$

$$\mathbf{b}_\lambda^T \boldsymbol{\varrho} = \mathbf{b}_\lambda^T \mathbf{s} - \mathbf{b}_\lambda^T \mathbf{s}_0 \quad (2.8)$$

$$d\Omega = \frac{dl dm}{\sqrt{1 - l^2 - m^2}} \quad (2.9)$$

where $n = \sqrt{1 - l^2 - m^2}$ is the third direction cosine with respect to the w axis and dl and dm are the increments on the image plane and define a square cross section. Inserting in Equation (2.5) and indicating $\mathbf{b}_\lambda = [u, v, w]^T$, the visibility measurements $\mathcal{V}(u, v, w)$ and the intensity distribution $I(l, m)$ are related as

$$\mathcal{V}(u, v, w) = \int_l \int_m A_N(l, m) I(l, m) e^{-j2\pi[ul+vm+w(n-1)]} \frac{dl dm}{n}. \quad (2.10)$$

This equation shows how the complex visibility function is related to the “modified” intensity $A_N(l, m)I(l, m)$. This equation is valid for a narrow frequency band around the center frequency f .

There exists a more general treatment of the radio interferometric measurement equation which considers wave polarization, Direction-Dependent Effects (DDEs) and Direction Independent Effects (DIEs) given by Hamaker et al. [32] and more recently by Smirnov [33]. These effects are mainly considered in the calibration process. In this thesis, we focus on the imaging aspect of the measurement equation and therefore abstract the DDEs and DIEs from the measurement equation. Furthermore, the primary beam $A_N(l, m)$ and the effect of the w component of the telescope array can also be considered as DDEs and abstracted from the imaging problem.

DISCRETIZATION OF THE MEASUREMENT EQUATION

In practice, the synthesis telescope obtains discrete samples of the visibility function in the baseline plane. An intensity map (image) of radio sources can be reconstructed based on these samples [10]. The visibility measurements are time dependent due to the rotation of the earth. To capture more visibility measurements, the earth’s rotation can be used to gain more visibility measurements by tracking the phase reference center over the course of observation which can last up to 6 to 12 hours. This process is called “earth rotation synthesis” [8, 28].

Assuming a single frequency band and one snapshot of visibility data, we also discretize the FoV over a grid of pixels to obtain the discrete version of the measurement equation based on the measured visibility samples. An example of a pixel grid on (l, m) of half of the celestial sphere for $l^2 + m^2 \leq 1$ is shown in Figure 2.4.

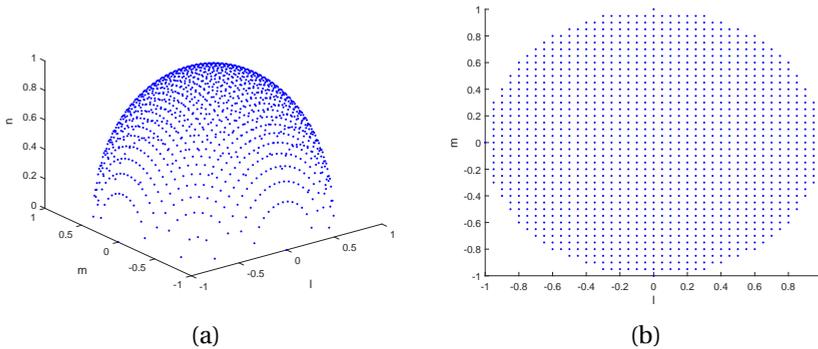


Figure 2.4: (a) Grid of the celestial sphere based on l and m direction cosines, (b) top view

Assuming M visibility measurements, and a pixel size of Δl and Δm , defined based on the resolution of the telescope array, and the total number of pixels N_l and N_m over the l and m dimensions of the image, neglecting the DDEs, the discrete version of Equation (2.10) can be stated as

$$\mathcal{V}(u_k, v_k, w_k) = \sum_{p=1}^{N_l} \sum_{q=1}^{N_m} I(l_p, m_q) e^{-2\pi j(u_k l_p + v_k m_q + w_k(\sqrt{1-l_p^2-m_q^2}-1))} \frac{\Delta l \Delta m}{\sqrt{1-l_p^2-m_q^2}}, \quad (2.11)$$

$$k = 1, 2, \dots, M.$$

In another way, expressing the total number of pixels over the complete grid in l and m dimension as $Q = N_l \times N_m$ and neglecting the constant factor $\Delta l \Delta m$, the discretized measurement equation becomes

$$\mathcal{V}(u_k, v_k, w_k) = \sum_{q=1}^Q \frac{I(l_q, m_q)}{\sqrt{1-l_q^2-m_q^2}} e^{-2\pi j(u_k l_q + v_k m_q + w_k(\sqrt{1-l_q^2-m_q^2}-1))}, \quad (2.12)$$

$$k = 1, 2, \dots, M.$$

Vectorizing the visibility samples in vector \mathbf{r} as

$$\mathbf{r} = \begin{bmatrix} \mathcal{V}(u_1, v_1, w_1) \\ \mathcal{V}(u_2, v_2, w_2) \\ \vdots \\ \mathcal{V}(u_M, v_M, w_M) \end{bmatrix}, \quad (2.13)$$

Equation (2.12) can be stated in matrix form as

$$\mathbf{r} = \mathbf{M}\boldsymbol{\sigma}, \quad (2.14)$$

where

$$\boldsymbol{\sigma} = \begin{bmatrix} \frac{I(l_1, m_1)}{\sqrt{1-l_1^2-m_1^2}} \\ \frac{I(l_2, m_2)}{\sqrt{1-l_2^2-m_2^2}} \\ \vdots \\ \frac{I(l_Q, m_Q)}{\sqrt{1-l_Q^2-m_Q^2}} \end{bmatrix}, \quad (2.15)$$

and

$$m_{k,q} = e^{-2\pi j(u_k l_q + v_k m_q + w_k(\sqrt{1-l_q^2-m_q^2}-1))} \quad (2.16)$$

is the (k, q) -th component of the kernel matrix that relates the discretized intensity points over the image grid to the visibility samples. The kernel matrix \mathbf{M} is known for a certain configuration of the source and array. If the telescopes are not tracking the reference source, we can express the (k, q) -th component of the kernel matrix as

$$m_{k,q} = e^{-2\pi j(u_k l_q + v_k m_q + w_k \sqrt{1-l_q^2-m_q^2})}. \quad (2.17)$$

Expressing the k th baseline components in a vector as $\mathbf{b}_k = [u_k, v_k, w_k]^T$ and source components for the q th pixel as

$$\mathbf{z}_q = \begin{bmatrix} l_q \\ m_q \\ \sqrt{1 - l_q^2 - m_q^2} \end{bmatrix}, \quad (2.18)$$

the components of the kernel matrix \mathbf{M} can be stated in terms of the inner product between the baseline and source components as

$$m_{k,q} = e^{-2\pi j \mathbf{b}_k^T \mathbf{z}_q}. \quad (2.19)$$

In practice, the available visibility data is noisy and non-uniformly sampled. The imaging problem is to estimate the source $\boldsymbol{\sigma}$ from the noisy discrete visibility measurements $\hat{\mathbf{r}}$. Therefore, the discrete measurement equation boils down to

$$\hat{\mathbf{r}} = \mathbf{M}\boldsymbol{\sigma} + \mathbf{e}, \quad (2.20)$$

where \mathbf{e} expresses the measurement noise and the quantization error. As a further comment, measurements from multiple time snapshots and/or frequency bands can be stacked in the measurement vector $\hat{\mathbf{r}}$.

The aim of radio interferometric imaging is to estimate intensity from the visibility measurements. This discrete measurement model is widely used in developing imaging algorithms especially in the context of compressed sensing e.g. [34, 35, 36]. Additive noise \mathbf{e} on the observed visibilities is a result of sky, receiver and ground pick up, etc. noise. As suggested in Chapter 16 of [28], this noise, to a very good approximation, can be considered Gaussian having equal variance in the imaginary and real parts of the complex visibility.

2.2.2. ARRAY PROCESSING FORMULATION

In this section, we develop the array processing data model for radio interferometric imaging. We build our data model for RA imaging in the array signal processing framework based on prior efforts presented in [37, 38, 39, 40]. The data model is developed based on a perfectly calibrated array without DDEs and DIES. We briefly present these effects in the data model in the next section.

We assume a telescope array of P distinct receiving elements (antennas or beam-formed stations) with an arbitrary arrangement. The celestial signals are received by the array elements with different delays depending on the placement of the elements. The raw analog output signals of the array elements are digitized and divided into narrow frequency sub-bands. We assume that the narrow-band condition holds, so that propagation delays across the array can be replaced by complex phase shifts. Although the sources are considered stationary, because of the earth's rotation the apparent position of the celestial sources will change with time. For this reason the data is split into short blocks or "snapshots" of N samples, where the exact value of N depends on the resolution of the instrument. The sampled output station signals of the telescope array are stacked into $P \times 1$ vectors $\mathbf{x}_k[n]$, where $n = 1, \dots, N$ is the sample index, and $k = 1, \dots, K$

denotes the snapshot times the frequency channel index (assuming ordered stacking of frequency channels and time snapshots). We can stack all the short-time samples and define a data matrix \mathbf{X}_k of size $P \times N$. Similarly, assuming Q mutually independent source signals $s_q[n]$ impinging on the array, we stack them into $Q \times 1$ vectors $\mathbf{s}_k[n]$. We model the receiver noise as mutually independent zero mean Gaussian signals stacked in a $P \times 1$ vector $\mathbf{n}_k[n]$.

The output of the telescope array is a linear combination of the source signals and receiver noise:

$$\mathbf{x}_k[n] = \mathbf{A}_k \mathbf{s}_k[n] + \mathbf{n}_k[n] \quad (2.21)$$

where $\mathbf{A}_k = [\mathbf{a}_1, \dots, \mathbf{a}_Q]$ of size $P \times Q$ is called the array response matrix, and \mathbf{a}_q is its q th column. Ideally, entry (p, q) of \mathbf{A}_k follows from the geometric delay of source q arriving at antenna p :

$$a_{p,q} = \frac{1}{\sqrt{P}} e^{-j \frac{2\pi}{\lambda} \mathbf{y}_p^T \mathbf{z}_q}, \quad (2.22)$$

where the scaling by \sqrt{P} is such that $\|\mathbf{a}_q\| = 1$. For an incoming monochromatic wave at frequency f , $\lambda = \frac{c}{f}$ denotes the wavelength of the received signal, \mathbf{y}_p is a 3×1 vector of the Cartesian location of the p th array element (at time-frequency-index k) with respect to a chosen origin in the field of array, and \mathbf{z}_q contains the direction cosines of the q th pixel in the image plane (defined in Equation (2.18)).

In practice, the array also suffers from antenna-dependent gains and direction-dependent gains that need to be estimated and multiplied with \mathbf{A}_k . This estimation is done in an outer loop (the selfcal loop) and therefore, for the purpose of the research in this thesis, we can assume that \mathbf{A}_k is known (although not necessarily of exactly the form (2.22)) based on the known geometry of the array elements and the image grid. Nonetheless, before selfcal has converged, the data will suffer from a model mismatch.

Without loss of generality, we will from now on consider only a single snapshot and frequency channel k , and will drop the index k . Assuming the signals and the receiver noise are uncorrelated and the noise on different antennas are mutually uncorrelated, the data covariance matrix of the received signals is modeled as

$$\mathbf{R} := E\{\mathbf{x}[n]\mathbf{x}^H[n]\} = \mathbf{A}\mathbf{\Sigma}_s\mathbf{A}^H + \mathbf{\Sigma}_n, \quad (2.23)$$

where $\mathbf{\Sigma}_s = \text{diag}\{\boldsymbol{\sigma}_s\}$ and $\mathbf{\Sigma}_n = \text{diag}\{\boldsymbol{\sigma}_n\}$ represent the covariance matrices associated with the source signals and the received noise, $\boldsymbol{\sigma}_s = [\sigma_{s,1}^2, \sigma_{s,2}^2, \dots, \sigma_{s,Q}^2]^T$ and $\boldsymbol{\sigma}_n = [\sigma_{n,1}^2, \sigma_{n,2}^2, \dots, \sigma_{n,P}^2]^T$. We assume that the receiver noise powers $\mathbf{\Sigma}_n$ are known from the calibration process.

An estimate of the data covariance matrix is obtained using the available received data samples. The sample covariance matrix for a single snapshot is calculated as

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}[n]\mathbf{x}^H[n] = \frac{1}{N} \mathbf{X}\mathbf{X}^H, \quad (2.24)$$

and is used as an estimate of the true covariance matrix \mathbf{R} .

The radio astronomical imaging process amounts to estimating the image pixel intensities $\boldsymbol{\sigma}$ based on the covariance data measured by a telescope array $\hat{\mathbf{R}}$ over the FoV

of the array. To obtain a linear measurement model, we vectorize the covariance data model (2.23) as

$$\mathbf{r} = (\mathbf{A}^* \circ \mathbf{A})\boldsymbol{\sigma} + \mathbf{r}_n = \mathbf{M}\boldsymbol{\sigma} + \mathbf{r}_n, \quad (2.25)$$

where $\mathbf{r} = \text{vect}(\mathbf{R})$, $\mathbf{r}_n = \text{vect}(\mathbf{R}_n) = (\mathbf{I} \circ \mathbf{I})\boldsymbol{\sigma}_n$, and $\mathbf{M} = \mathbf{A}^* \circ \mathbf{A}$ is the system matrix of the linear measurement model of size $P^2 \times Q$. Based on (2.22), one element of \mathbf{M} corresponding to the baseline between the i th and j th antenna and the q th pixel is computed as:

$$m_{ij,q} = a_{i,q}^* a_{j,q} = \frac{1}{P} e^{j \frac{2\pi}{\lambda} (\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{z}_q}. \quad (2.26)$$

This expression is modified in the presence of calibration parameters (antenna-dependent gains and direction-dependent gains) which we assume to be known at this point.

Similarly, we vectorize the covariance measurement matrix as

$$\hat{\mathbf{r}} = \text{vect}(\hat{\mathbf{R}}), \quad (2.27)$$

and compensate $\hat{\mathbf{r}}$ for the (known) receiver noise powers,

$$\tilde{\mathbf{r}} = \hat{\mathbf{r}} - \mathbf{r}_n, \quad \tilde{\mathbf{R}} = \hat{\mathbf{R}} - \mathbf{R}_n. \quad (2.28)$$

This results in a linear measurement equation for estimating $\boldsymbol{\sigma}$ based on the measured $\tilde{\mathbf{r}}$:

$$\tilde{\mathbf{r}} = \mathbf{M}\boldsymbol{\sigma} + \mathbf{e}, \quad (2.29)$$

where \mathbf{e} represents the error due to the finite sample modeling of the covariance data. \mathbf{e} is zero-mean with covariance [41]

$$\mathbf{C}_e = E\{(\hat{\mathbf{r}} - \mathbf{r})(\hat{\mathbf{r}} - \mathbf{r})^H\} = \frac{1}{N}(\mathbf{R}^T \otimes \mathbf{R}). \quad (2.30)$$

For a large number of samples N we can assume that \mathbf{e} is distributed according to a zero-mean complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \mathbf{C}_e)$ where $\mathbf{C}_e = \frac{1}{N}(\mathbf{R}^T \otimes \mathbf{R})$ [41], which will be estimated from $\hat{\mathbf{R}}$.

Comparing Equations (2.20) and (2.29) we see that the two methods result in a similar measurement model. However, applying array processing models for the radio interferometric imaging problem, the structure of the measurement matrix \mathbf{M} can be expressed in terms of the array response vector (as stated in Equation (2.25)). Furthermore, statistics of the noise on the correlation measurements can be inferred from the statistics of the thermal noise on the array elements.

2.2.3. SOME FURTHER REMARKS AND GENERALIZATIONS

We present some remarks on the measurement model (2.29) and briefly discuss the cases where the actual radio interferometric measurements may deviate from (2.29) and introduce generalizations to the model.

1. In actual instruments, the autocorrelations of the data are often not formed (or at least not used for imaging) because they are considered too much contaminated. In that case, $\tilde{\mathbf{r}}$ is not computed from (2.28), but rather by omitting the autocorrelation terms from $\hat{\mathbf{r}}$ (they correspond to the nonzero entries of \mathbf{r}_n). Equation (2.29)

holds but some rows of \mathbf{M} have been dropped. Unfortunately, with the missing autocorrelation data we lose the estimate for \mathbf{C}_e .

However, in the signal processing community, measuring the autocorrelations is considered essential since without the autocorrelations, $\hat{\mathbf{R}}$ would not constitute sufficient statistics for the collected data $\{x_k[n]\}$, i.e., information is lost. If these autocorrelations are deemed to be “more noisy” then that should be represented in the data model. Estimates of the variance of the measurements are equivalent information and usually available in RA via the natural weights [42] or System Equivalent Flux Density (SEFD) [1]. We believe that in any case the autocorrelations will have been used in the calibration and subsequent whitening of the system noise, so that we arrive at the implicit assumptions where $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$ and astronomical signals much weaker than the noise. Furthermore, it is shown in [43] that the Cramer-Rao lower bound for the estimation of $\boldsymbol{\sigma}$ increases if the information from the autocorrelations is not used. This means that we are only able to estimate the source powers worse than when the autocorrelations are present.

2. Many recent papers on radio astronomy image formation start from (2.20) and model the covariance of \mathbf{e} as spatially white, $\mathbf{C}_e \propto \mathbf{I}$. However, this is correct only under two assumptions, (i) the additive noise \mathbf{n} is much stronger than the astronomical signals \mathbf{s} , and (ii) the additive noise is spatially white, $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$. This requires a whitening operation after calibration of the antenna-dependent gain parameters. These assumptions are usually considered valid in radio astronomy practice.
3. In the general case where we have measurements for multiple time snapshots and/or frequency channels, assuming the measurements from different snapshots and channels are uncorrelated, we can stack them into one vector as

$$\hat{\mathbf{r}} = \begin{bmatrix} \hat{\mathbf{r}}_1 \\ \hat{\mathbf{r}}_2 \\ \vdots \\ \hat{\mathbf{r}}_K \end{bmatrix}. \quad (2.31)$$

Accordingly, the covariance data model becomes

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_K \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_K \end{bmatrix} \boldsymbol{\sigma} + \begin{bmatrix} \mathbf{r}_{n,1} \\ \mathbf{r}_{n,2} \\ \vdots \\ \mathbf{r}_{n,K} \end{bmatrix} = \mathbf{M}\boldsymbol{\sigma} + \mathbf{r}_n, \quad (2.32)$$

and therefore the measurement equation can be stated as

$$\begin{aligned} \hat{\mathbf{r}} &= \mathbf{r} + \mathbf{e} \\ &= \mathbf{M}\boldsymbol{\sigma} + \mathbf{r}_n + \mathbf{e} \end{aligned} \quad (2.33)$$

where $\mathbf{e} = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_K]^T$ and similar as before by defining $\tilde{\mathbf{r}} = \hat{\mathbf{r}} - \hat{\mathbf{r}}_n$ we arrive at Equation (2.29). However, in this case we have a larger amount of data (better $u-v$ coverage) from multiple frequencies and/or snapshots.

4. In practice, the radio signals impinging on a telescope array are corrupted by instrumental and propagation effects. The measurement model (2.29) is derived based on ideal and identical array elements by neglecting the propagation effects. Neglecting these effects in imaging will result in erroneous intensity estimates.

The propagation and instrumental effects can be included in the measurement model with the aim to compensate for the errors introduced in the imaging. Calibration is the process of compensating (correcting) for these effects. The calibration procedure is applied on the data before (external calibration) and during (self calibration) the imaging procedure in the SDP.

We briefly introduce parameters for DIE and DDE calibration in measurement model (2.29) based on [44, 45, 40, 46]. With a similar setting as in Section 2.2.2, the array elements are assumed to have the same direction-dependent gains denoted in matrix form as $\mathbf{B} = \text{diag}([b_1, b_2, \dots, b_Q])$ where b_q is the gain factor towards the q th source as received by the telescope array. The direction independent gains and phases are stated as $\mathbf{G} = \text{diag}([g_1, g_2, \dots, g_P])$ where $g_p = \gamma_p e^{j\Phi_p}$ with γ_p and $e^{j\Phi_p}$ are the gain and phase factors on the p th array element, respectively. Therefore, we can write the output of the telescope array as

$$\mathbf{x}[n] = \mathbf{GABs}[n] + \mathbf{n}[n], \quad n = 1, \dots, N. \quad (2.34)$$

where \mathbf{G} is the $P \times P$ matrix of direction dependent gains and phases and \mathbf{B} is the $Q \times Q$ matrix of direction independent gains. In this case the autocorrelation function can be stated as

$$\mathbf{R} = \mathbf{GAB}\Sigma_s\mathbf{B}^H\mathbf{A}^H\mathbf{G}^H + \mathbf{R}_n, \quad (2.35)$$

Assuming known directional response of antennas, \mathbf{B} can be absorbed in \mathbf{A} . If the directional gains are not known, the directional response and pixel powers can be estimated simultaneously by defining

$$\begin{aligned} \Sigma &= \mathbf{B}\Sigma_s\mathbf{B}^H \\ &= \text{diag}([|b_1|^2\sigma_{s,1}^2, |b_2|^2\sigma_{s,2}^2, \dots, |b_Q|^2\sigma_{s,Q}^2]) = \text{diag}(\boldsymbol{\sigma}). \end{aligned} \quad (2.36)$$

with real and positive values $\boldsymbol{\sigma} = [\sigma_1^2, \sigma_2^2, \dots, \sigma_Q^2]$. Therefore, the data model in (2.35) can be restated as

$$\mathbf{R} = \mathbf{GA}\Sigma\mathbf{A}^H\mathbf{G}^H + \mathbf{R}_n \quad (2.37)$$

Similarly as before, we can vectorize the covariance data model (2.37) as

$$\mathbf{r} = (\mathbf{G}^* \mathbf{A}^* \circ \mathbf{GA})^H \boldsymbol{\sigma} + \mathbf{r}_n, \quad (2.38)$$

where $\mathbf{r} = \text{vec}(\mathbf{R})$ and $\mathbf{r}_n = \text{vec}(\mathbf{R}_n) = (\mathbf{I} \circ \mathbf{I})\boldsymbol{\sigma}_n$. Subsequently, by vectorizing the sample covariance matrix, we arrive at the measurement model

$$\hat{\mathbf{r}} = \mathbf{r} + \mathbf{e}. \quad (2.39)$$

The main difference of the data model obtained in this section with the pure imaging data model presented in Section 2.2 is in the presence of the direction-

independent and the complex direction-dependent gain matrices, \mathbf{G} and \mathbf{B} , applied to \mathbf{A} to account for the instrumental and propagation effects. DDEs are applied on the right of \mathbf{A} and appear as a multiplication in the image domain as they are dependent on the direction of the source. DDEs contain effects such as the antenna beams, ionospheric propagation and the non-coplanarity of the array. DDEs are related to the complex gain (phase and gain) of the individual antennas and appear on the left of \mathbf{A} . The main aim of calibration is to estimate the parameters in \mathbf{B} and \mathbf{G} . In fact, if we assume identical array elements with equal gains i.e. $\mathbf{G} = \mathbf{I}$ we arrive at a similar measurement model as Equation (2.29).

2.2.4. SKA PROCESSING PIPELINE

We have briefly mentioned the proposed data acquisition and processing pipeline for the SKA in Chapter 1, Section 1.5.3. We show that this pipeline nicely aligns with the array processing data model presented in Section 2.2.2. This pipeline is summarized in Figure 2.5 and its relation to the array processing data model is explained below.



Figure 2.5: End-to-end processing pipeline of the SKA

Station processing Raw analogue measured signals of radio telescopes are digitized, split into regular narrow frequency channels (sub-bands) of width Δf . SKA1-LOW contains phased array telescope systems. In these systems, some antennas are grouped into a station and mimic a single large telescope by beamforming [22]. Beamforming is an operation that is performed electronically on the output of the individual antennas in a phased array that results in pointing the array towards a specific portion of the sky. This is done by introducing phase delays on the array signals depending on the beam direction and integrating the resulting signals. After station beamforming, the signals are time-sampled regularly at the Nyquist rate of $2\Delta f$ Hz. The station data is split into short blocks of N time samples (depending on the resolution of the instrument). We can associate the output of this stage to the data matrix \mathbf{X}_k .

Central signal processing The measured data from beamformed stations and dishes are transported to the CSP for correlation. In this step, correlation of the data stream from all pairs of stations and/or dishes, a.k.a. baselines, are performed and integrated (time averaged) over a short time span T . The output of this step is the snapshot- and frequency-dependent sample covariance matrix which we denote by matrix $\hat{\mathbf{R}}_k$ of size $P \times P$.

Science data processing The measured correlations are calibrated to correct for instrumental and propagation errors. Furthermore, an image of the radio sky over the FoV is formed. The output of this step is the radio astronomical image that we indicate

by vector $\hat{\sigma}$. We will focus on this step and design imaging algorithms in the following chapters.

2.3. INTERPRETATIONS OF THE MEASUREMENT EQUATION

We focus our attention on the linear measurement equation for radio interferometric imaging presented in Equation (2.29). In radio interferometric imaging, the aim is to estimate the sky brightness map σ from the measured correlations (or visibilities) $\tilde{\mathbf{r}}$. Although the measurement system is linear, due to the existence of noise, $\tilde{\mathbf{r}}$ is not in the range of \mathbf{M} . Therefore, to estimate σ we should minimize the distance of $\tilde{\mathbf{r}}$ from the model $\mathbf{M}\sigma$. Having shown the connections of radio astronomy with array processing, in this section we show that we can apply different interpretations to radio interferometric imaging and therefore benefit from the tools developed in relevant areas for estimating the intensity distribution from the correlation (visibility) measurements. In particular, we consider “calibrated” visibility measurements and investigate approaches to estimate an image based on the visibility measurements.

2.3.1. DISCRETE LINEAR INVERSE PROBLEM

Measurement equation (2.29) describes a linear “forward problem”. This means that, if the true sky intensity distribution is known, correlation (visibility) measurements can be calculated based on the known measurement process described by the measurement matrix \mathbf{M} . In the imaging process, the objective is to invert the measurement process to obtain an estimate of the sky intensity distribution from the visibility measurements. This is an instance from the broad area of “inverse problems” [47]. In this problem, since we are unable to measure the source intensities directly, we obtain the noisy covariance measurements by indirect observations from the far field of the sources by the telescope array pointing at the FoV of interest. Due to the physical limitations of the radio telescopes, the measurement set is inevitably incomplete and corrupted by noise. Due to the inherent loss of information in the measurement process the inverse problem is “ill-posed”. A problem is said to be ill-posed if (i) the solution to the problem does not exist, (ii) the solution is not unique, or (iii) the solution is not stable, i.e. it does not depend continuously on the data.

The inverse problem in radio interferometric imaging can be formulated as to estimate a sky map that agrees with the measured visibility data within a bound imposed by the level of the noise on the data. This is a linear regression analysis problem. Least squares is a powerful tool for linear regression analysis problems. The simplest problem formulation for radio astronomical imaging is based on the Least Squares (LS) formulation. That is

$$\{\hat{\sigma}\} = \underset{\sigma}{\operatorname{argmin}} \|\tilde{\mathbf{r}} - \mathbf{M}\sigma\|_2^2, \quad (2.40)$$

where $\{\hat{\sigma}\}$ indicates the set of all the possible solutions when the imaging problem is underdetermined.

If the problem is underdetermined, the solution to problem (2.40) is not unique. Otherwise, the solution may not exist or if it exists it may be completely deprived of physical meaning. The latter is due to the propagation and amplification of the noise from the data to the estimate of the solution [48] and sensitivity of the solution to modeling errors.

To achieve a physically meaningful solution, it is required to regularize the problem by imposing constraints on the image based on some prior information.

2.3.2. SECOND-ORDER STATISTICS ESTIMATION

We present this section as a side-track to illustrate the relation of radio interferometric imaging with compressive covariance sampling. The imaging system in RI is an incoherent imaging system in which the sources emit radiations which are received by the telescope array. Since the emission is incoherent, the narrow-band radiation from different parts of the source can be considered uncorrelated. In incoherent imaging systems, the intensity distribution of the sources and autocorrelation function in the array plane are related via a Fourier transform relationship [49], i.e. Equation (2.4). In the spatial domain, estimating the Power Spectral Density (PSD) corresponds to incoherent imaging [49] and the PSD can be regarded as the intensity distribution of the sources.

FOURIER TRANSFORM RELATIONSHIP

Radio interferometers perform the transform between two spaces; (i) the real space of the sky image and (ii) the Fourier space of the visibility measurements. Observations by radio interferometers are in the Fourier space and the sampling of the Fourier space depends on the geometry of the telescope array. The largest spatial scale that is measured by the telescope array (FoV) is determined based on the smallest measured spatial frequency. Furthermore, the natural sensitivity of the telescope array to different spatial scales is determined based on the sample density within the measured range [50]. Based on the power spectrum estimation interpretation, the domain of the intensity image is called the “real” domain and its transform domain is called the “spatial frequency” or Fourier domain. Baselines in spatial domain can be regarded as the lags in time domain.

If the visibility samples are placed regularly on a rectangular grid, which is the transform of the image pixel grid, the Fourier transform relationship can easily be implemented as a Fast Fourier Transform (FFT). The Fourier relation is shown schematically in Figure 2.6.

We assume, as described in Section 2.2.1, that the image field is divided into equally-distant grid points in the l and m dimension with N_l and N_m total pixels, respectively in l and m . The Nyquist-Shannon sampling criterion can be expressed in terms of the maximum baseline length in the direction of the u -axis and v -axis, u_{max} and v_{max} , as [51]

$$\Delta l \leq \frac{1}{2u_{max}}, \quad \Delta m \leq \frac{1}{2v_{max}}, \quad (2.41)$$

Furthermore, the angular extent of the image in the direction of l - and m -axis is defined by [51],

$$N_l \Delta l \geq \frac{1}{u_{min}}, \quad N_m \Delta m \geq \frac{1}{v_{min}}, \quad (2.42)$$

where u_{min} and v_{min} are the minimum baseline length (discarding the autocorrelations) measured in u and v axis, respectively. If we consider $\Delta u = u_{min}$ and $\Delta v = v_{min}$, the relation of the grid spacing in the u - and v -axis to the angular region of the sky covered in the FoV becomes

$$\Delta u = (N_l \Delta l)^{-1}, \quad \Delta v = (N_m \Delta m)^{-1}. \quad (2.43)$$

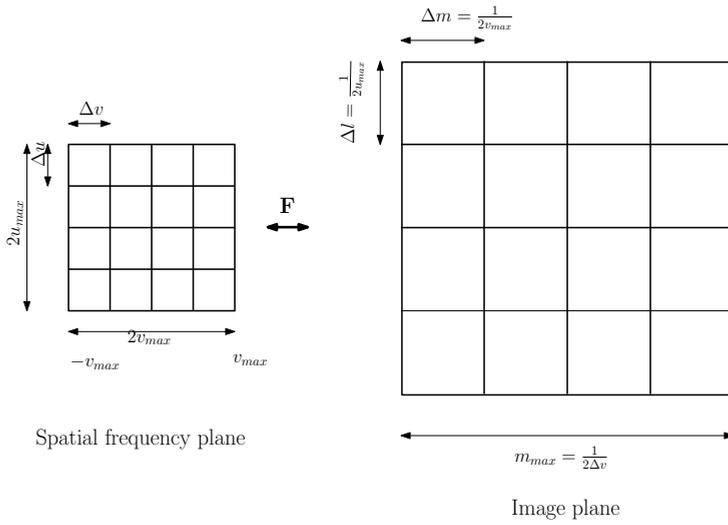


Figure 2.6: Image plane and the spatial frequency plane relation

To compare with the conventional PSD estimation, we can replace the temporal indexes with the image (angular) domain indexes (l, m, n) and the frequency indexes with the indexes in the spatial frequency domain (u, v, w) . Therefore, we can use interpretations from the time-frequency domain in the angular-spatial frequency domain.

The extent of the $u-v$ plane should ideally be identified based on the spatial band of the sources in the sky map. In RA, due to the existence of point-like sources, the source signals are not spatially band limited. In fact, point sources require an infinite aperture for perfect reconstruction. However, due to the physical constraints of the aperture plane, it is not possible to make an infinitely large aperture. In practice, the size of the image is determined by the FoV of the telescope array. Extrapolation schemes can be used to address some of these issues for point sources if some prior information is available. Furthermore, the data samples (visibility measurements) are incomplete and are not placed regularly on the uniform grid points. Therefore, there is a need for an interpolation scheme to make FFT possible. These schemes are called gridding in RA. The gridding process is a demanding part of the imaging computations. In practice, to grid the measured $u-v$ points, they are first convolved with a smooth “gridding kernel”, or “gridding convolution function” and then re-sampled at the center of the grid cells (interpolation operation)[28].

COMPRESSIVE COVARIANCE ESTIMATION

We show the Fourier transform relationship and the effect of irregular sampling of the telescope array. Assuming that all the spatial samples are available in \mathbf{y} , the sub-sampled vector \mathbf{x} can be indicated as

$$\mathbf{x} = \Phi \mathbf{y} \quad (2.44)$$

where Φ indicates the (sparse) sampling matrix. The aim is to reconstruct the angular (spatial) PSD based on the measured samples \mathbf{x} . The covariance measurements have a

2-dimensional Fourier transform relation with the PSD

$$\mathbf{R}_y = E\{\mathbf{y}\mathbf{y}^H\} = \mathbf{F}\mathbf{\Sigma}_s\mathbf{F}^H \quad (2.45)$$

where $\mathbf{\Sigma}_s = \text{diag}(\boldsymbol{\sigma})$ and \mathbf{F} is the Discrete Fourier Transform (DFT) matrix. Vectorizing \mathbf{R}_y we obtain

$$\mathbf{r}_y = \text{vect}(\mathbf{R}_y) = (\mathbf{F}^* \otimes \mathbf{F})\boldsymbol{\sigma}. \quad (2.46)$$

We know based on the derivations of Section 2.2.2 that the covariance data model for the measured samples \mathbf{x} can be stated as

$$\mathbf{R}_x := E\{\mathbf{x}[n]\mathbf{x}^H[n]\} = \mathbf{A}\mathbf{\Sigma}_s\mathbf{A}^H \quad (2.47)$$

Based on Equation (2.44) we can write

$$\mathbf{R}_x = \mathbf{\Phi}\mathbf{R}_y\mathbf{\Phi}^H = \mathbf{\Phi}\mathbf{F}\mathbf{\Sigma}_s\mathbf{F}^H\mathbf{\Phi}^H \quad (2.48)$$

Therefore, we conclude that \mathbf{A} can be written as

$$\mathbf{A} = \mathbf{\Phi}\mathbf{F} \quad (2.49)$$

and the linear relation between the vectorized measured correlations $\mathbf{r}_x = \text{vect}(\mathbf{R}_x)$ with the PSD (or source intensities) is

$$\mathbf{r}_x = (\mathbf{\Phi}^* \otimes \mathbf{\Phi})(\mathbf{F}^* \circ \mathbf{F})\boldsymbol{\sigma} = (\mathbf{\Phi}^*\mathbf{F}^* \circ \mathbf{\Phi}\mathbf{F})\boldsymbol{\sigma}. \quad (2.50)$$

This corresponds to a sub-sampled Fourier relation between the measured covariance data and the source intensity distribution. $(\mathbf{\Phi}^* \otimes \mathbf{\Phi})$ denotes the sampling of baselines. We note that the sampling matrix $\mathbf{\Phi}$ need not be defined on a regular grid and therefore non-regular sampling patterns are also possible. This is the case in radio interferometry where the placement of the baselines is irregular. Comparing to Equation (2.29) we notice that the measurement matrix can be written as

$$\mathbf{M} = \mathbf{\Phi}^*\mathbf{F}^* \circ \mathbf{\Phi}\mathbf{F}. \quad (2.51)$$

Matrix \mathbf{M} performs the transform between the image domain and the visibility domain and contains the sampling by the telescope array and the Fourier transform relationship.

In radio interferometric imaging we are only interested in estimating the second-order statistics of the cosmic signals (the power spectrum) and not the signals themselves. It is shown that under this assumption, we do not need the Nyquist-sampled complete covariance matrix. In a field called the compressive covariance sampling, the second-order statistics are estimated based on a subset of the covariance samples. It is shown that under certain conditions, this allows for perfect reconstruction of the signal powers [52]. The field of compressive covariance sampling is concerned with designing a compression scheme for the correlation measurements, i.e. designing $\mathbf{\Phi}$ in Equation (2.44), such that perfect reconstruction of the second-order statistics \mathbf{R}_y or $\boldsymbol{\sigma}$ is possible based on the sub-sampled correlation \mathbf{R}_x . In fact, radio interferometric imaging is an application of second-order statistics estimation based on inherently compressive

covariance samples. This is because the baseline plane is, by design, compressively sampled.

In the compressive covariance estimation, the concept of co-array is used [49]. The co-array is a virtual array by considering all the possible differences of the antennas (baselines) within the telescope array. It is shown that in a considerably sub-sampled array, if the sub-sampling is such that all the lags of the antennas are preserved (with minimum repetition), then the second-order statistics can be estimated perfectly. This is the main concept in designing minimum redundancy arrays [53]. However, if these conditions are not satisfied, i.e. not all the spatial lags are present, it is not possible to retrieve the exact solution and we must resort to approximate solutions.

The correlation function \mathbf{R}_x is estimated based on temporal samples of $\mathbf{x}[n]$ by temporal averaging. The sample covariance matrix is called $\hat{\mathbf{R}}$. We note that in incoherent imaging, not only the number of spatial samples are limited but also the number of time samples to construct the sample covariance matrix $\hat{\mathbf{R}}$ is limited by the coherence time between the source and the telescope array. The former is similar to applying a spatial windowing on the spatial frequency measurements and the latter imposes a temporal window on the length of the available temporal samples to construct the correlation matrix. The larger the aperture of the telescope array, the smaller the coherence time and the smaller the temporal window.

The concept of a virtual array can be explained based on the construction of the measurement matrix \mathbf{M} . For simplicity we assume an array of two elements with positions denoted by vectors \mathbf{y}_1 and \mathbf{y}_2 and a celestial source on the celestial sphere with direction cosine \mathbf{z}_q as shown in Figure 2.7. The q th column of the measurement matrix \mathbf{M} ,

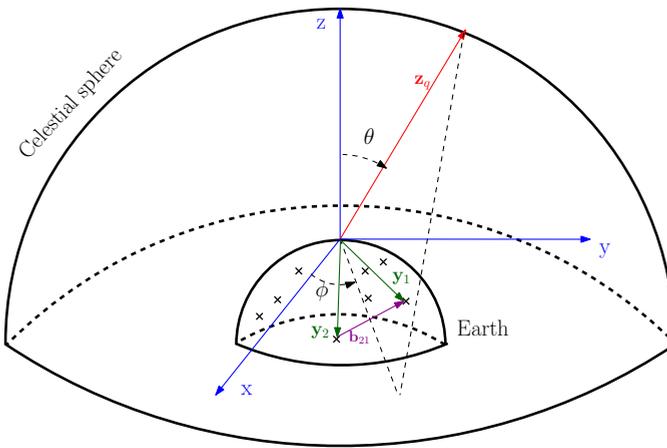


Figure 2.7: Measurement geometry

denoted by \mathbf{m}_q , is computed based on the corresponding column of the array response vector, \mathbf{a}_q , as

$$\mathbf{m}_q = \mathbf{a}_q^* \otimes \mathbf{a}_q \quad (2.52)$$

As an example, the second row of \mathbf{m}_q is computed as

$$a_{1,q}^* a_{2,q} = \frac{1}{P} e^{j \frac{2\pi}{\lambda} \mathbf{y}_1^T \mathbf{z}_q} e^{-j \frac{2\pi}{\lambda} \mathbf{y}_2^T \mathbf{z}_q} \quad (2.53)$$

$$= e^{-j \frac{2\pi}{\lambda} (\mathbf{y}_2 - \mathbf{y}_1)^T \mathbf{z}_q} \quad (2.54)$$

$$= e^{-j \frac{2\pi}{\lambda} (\mathbf{b}_{21})^T \mathbf{z}_q} \quad (2.55)$$

where \mathbf{b}_{21} denotes the baseline pointing from element 1 to element 2. For a p -element telescope array, \mathbf{m}_q contains the response of all the pairwise antenna interactions (baselines) to the q th source. In fact, the rows of the measurement matrix \mathbf{M} simulate a larger virtual array with elements located at $\mathbf{y}_i - \mathbf{y}_j$ with $1 \leq i, j \leq P$. In other words, \mathbf{m}_q shows the array response vector of the “difference coarray” towards the q th pixel. Therefore, the measurement equation shows a mapping between the sky intensity domain as described by $\boldsymbol{\sigma}$ to the measured covariances $\tilde{\mathbf{r}}$ with a phase delay compensation per baseline by the elements of the measurement matrix \mathbf{M} .

2.4. RADIO INTERFEROMETRIC IMAGING PROBLEM FORMULATION

In this section, we formulate the imaging problem in radio astronomy as an intensity estimation problem with the aim of finding the best approximation to the unknown sky map based on the correlation measurements. The problem can be presented in terms of a covariance matching estimation problem [41] that differs based on the statistical assumptions on the noise and sky map.

2.4.1. BEAMFORMING-BASED ESTIMATION

In incoherent imaging, spatial processing and beamforming techniques (parametric and non-parametric) can be regarded as spectrum estimation techniques in the spatial domain and the analysis can be carried over. More precisely, an initial estimate of the image can be obtained via beamforming. In this case, the i th pixel of the image is estimated as

$$\boldsymbol{\sigma}_{B,i} = \mathbf{w}_i^H (\mathbf{R} - \mathbf{R}_n) \mathbf{w}_i, \quad i = 1, \dots, Q \quad (2.56)$$

where \mathbf{w}_i is a spatially dependent beamformer (a spatial filter). In practice, since we do not have access to the true covariance matrix, we use the sample covariance matrix $\hat{\mathbf{R}}$ instead to obtain the beamformed image, i.e.,

$$\hat{\boldsymbol{\sigma}}_{B,i} = \mathbf{w}_i^H \hat{\mathbf{R}} \mathbf{w}_i = \mathbf{w}_i^H (\hat{\mathbf{R}} - \mathbf{R}_n) \mathbf{w}_i, \quad i = 1, \dots, Q. \quad (2.57)$$

For a linear beamformer, $\boldsymbol{\sigma}_{B,i}$ is the expected value of $\hat{\boldsymbol{\sigma}}_{B,i}$.

We consider two common beamforming approaches: Matched Filtering (MF) and Minimum Variance Distortionless Response (MVDR) beamforming [54]. The image estimate obtained by the MF beamformer is obtained by setting $\mathbf{w}_i = \mathbf{a}_i$, so that

$$\hat{\boldsymbol{\sigma}}_{MF,i} = \mathbf{a}_i^H (\hat{\mathbf{R}} - \mathbf{R}_n) \mathbf{a}_i \quad \Leftrightarrow \quad \hat{\boldsymbol{\sigma}}_{MF} = \mathbf{M}^H \tilde{\mathbf{r}}. \quad (2.58)$$

This estimate is known as the “dirty image” in the RA community. The expected value of this image is

$$\boldsymbol{\sigma}_{\text{MF},i} = \mathbf{a}_i^H (\mathbf{R} - \mathbf{R}_n) \mathbf{a}_i, \quad i = 1, \dots, Q. \quad (2.59)$$

Dirty image estimation from the correlation measurements, in terms of the power spectrum estimation, corresponds to the periodogram [55], more precisely the Welch periodogram, estimate of the power spectrum [49].

Similarly, the MVDR beamformer is defined as [39]¹

$$\mathbf{w}_i = \frac{\mathbf{R}^{-1} \mathbf{a}_i}{\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i}, \quad i = 1, \dots, Q \quad (2.60)$$

leading to the MVDR dirty image

$$\boldsymbol{\sigma}_{\text{MVDR},i} = \frac{\mathbf{a}_i^H \mathbf{R}^{-1} (\mathbf{R} - \mathbf{R}_n) \mathbf{R}^{-1} \mathbf{a}_i}{(\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i)^2} = \frac{1}{\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i} - \frac{\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{R}_n \mathbf{R}^{-1} \mathbf{a}_i}{(\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i)^2}. \quad (2.61)$$

In this expression, the first term is the “classical” MVDR solution, while the second term is a correction for the unwanted contribution of the noise covariance to this image. MVDR beamforming in incoherent imaging corresponds to the minimum variance spectrum estimation [55]. Similarly, many other techniques from spectrum estimation can be carried over to incoherent imaging.

We note that under the typical radio astronomical assumptions, i.e. white noise and weak signals, the factors \mathbf{R}^{-1} in the nominator and denominator of (2.60) cancel each other and the MVDR reduces to the matched filter beamformer in (2.59). Furthermore, if the autocorrelations are not available, as may be the case with visibility measurements in radio interferometry, \mathbf{R}^{-1} cannot be formed and we need to resort to the MF estimate.

It is shown in [56] that, if the corrections by \mathbf{R}_n are ignored in (2.59), (2.61), then

$$\mathbf{0} \leq \boldsymbol{\sigma}_{\text{true}} \leq \boldsymbol{\sigma}_{\text{MVDR}} \leq \boldsymbol{\sigma}_{\text{MF}} \quad (2.62)$$

Without ignoring \mathbf{R}_n , we can prove that the same result holds at least if $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$ (see below). This indicates that the MVDR dirty image is always closer to the true image than the MF beamformer.

PROOF OF (2.62)

It is known that [56]

$$\mathbf{a}_i^H \mathbf{R} \mathbf{a}_i \geq \frac{1}{\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i}.$$

Therefore, it is sufficient to prove that

$$\mathbf{a}_i^H \mathbf{R}_n \mathbf{a}_i \leq \frac{\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{R}_n \mathbf{R}^{-1} \mathbf{a}_i}{(\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i)^2}.$$

¹Actually, a correct derivation based on minimization of (2.56) subject to $\mathbf{w}_i^H \mathbf{a}_i = 1$ would give a result where \mathbf{R}^{-1} is replaced by $(\mathbf{R} - \mathbf{R}_n)^{-1}$ in (2.60), but this inverse is not numerically stable if \mathbf{R} is replaced by its estimate $\hat{\mathbf{R}}$.

This is the same as

$$\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i \cdot \mathbf{a}_i^H \mathbf{R}_n \mathbf{a}_i \cdot \mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i \leq \mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{R}_n \mathbf{R}^{-1} \mathbf{a}_i.$$

For this to hold, it is sufficient if

$$\mathbf{a}_i \mathbf{a}_i^H \mathbf{R}_n \mathbf{a}_i \mathbf{a}_i^H \leq \mathbf{R}_n.$$

While this is not true in general, the relation holds if \mathbf{a}_i is an eigenvector of \mathbf{R}_n , which includes special cases such as $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$, this relation holds. \square

2.4.2. LEAST SQUARES ESTIMATION

The most straightforward formulation of the source intensity estimation problem is via LS. In this problem formulation, no statistical assumptions are made about the sources, only the available measurements are fitted to the model in a LS sense. Due to the absence of probabilistic assumptions on $\boldsymbol{\sigma}$, claims about statistical optimality of the solution and its statistical performance cannot be made [57].

Equation (2.29) shows a parametric model of the sky. LS estimation can be regarded as “model fitting”. Based on the parametric model, the expected measurements obtained from the imaging equation are fitted to the measurements. Statistically, the appropriate way for model fitting is via the maximum likelihood method if the error statistics of the observations are known. A sub-optimal way is via the LS method. In this method, the implicit assumption is that the measurement errors are actually white Gaussian.

The LS RA imaging problem can be stated as

$$\hat{\boldsymbol{\sigma}} = \arg \min_{\boldsymbol{\sigma}} \|\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}\|_2^2. \quad (2.63)$$

The solution to (2.63) satisfies the normal equations

$$\mathbf{M}^H \mathbf{M} \hat{\boldsymbol{\sigma}} = \mathbf{M}^H \tilde{\mathbf{r}}, \quad (2.64)$$

where the left hand side shows the convolution of the image pixels with the beam pattern of the array via $\mathbf{M}^H \mathbf{M}$, and the right hand side $\hat{\boldsymbol{\sigma}}_{\text{MF}} = \mathbf{M}^H \tilde{\mathbf{r}}$ is recognized as the MF dirty image which is the same as the image obtained by matched filtering the data.

In RA imaging, the columns of \mathbf{M} corresponding to neighboring pixels are nearly parallel, making $\mathbf{M}^H \mathbf{M}$ poorly conditioned and the problem ill-posed. For large Q , $\mathbf{M}^H \mathbf{M}$ is not even invertible and a unique solution cannot be obtained without regularizing assumptions. Usually the LS cost function is regarded as the “data fidelity” term since it represents the fidelity of the expected measurements from the model to the measured data. However, the data fidelity term is in general insufficient to obtain a unique and physically meaningful estimate of the sky map.

2.4.3. MAXIMUM LIKELIHOOD ESTIMATION

Equation (2.29) shows the linear measurement model for RA imaging. For such a model, Maximum Likelihood Estimation (MLE) results in an efficient estimator that is also a Minimum Variance Unbiased (MVU) estimator [57]. If $\boldsymbol{\sigma}$ is considered deterministic

(i.e., a parameter vector without associated stochastic model), the likelihood function for (2.29) with complex Gaussian noise \mathbf{e} is

$$p(\tilde{\mathbf{r}}|\boldsymbol{\sigma}) = \frac{1}{\pi^{P^2} \det(\mathbf{C}_e)} \exp[-(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})^H \mathbf{C}_e^{-1} (\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})], \quad (2.65)$$

where, as mentioned before, $\mathbf{C}_e = \frac{1}{N}(\mathbf{R}^T \otimes \mathbf{R})$. MLE suggests finding the $\boldsymbol{\sigma}$ that results in the most probable set of measurements $\tilde{\mathbf{r}}$. Maximizing the likelihood function is equivalent to minimizing the cost function

$$J(\boldsymbol{\sigma}) = (\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})^H \mathbf{C}_e^{-1} (\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}) \quad (2.66)$$

which results in the weighted least squares (WLS) formulation of the imaging problem as shown in [40, 39]

$$\hat{\boldsymbol{\sigma}} = \arg \min_{\boldsymbol{\sigma}} \|\Gamma(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2, \quad (2.67)$$

where $\mathbf{C}_e^{-1} = \Gamma^H \Gamma$. The corresponding normal equations are

$$\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M} \hat{\boldsymbol{\sigma}} = \mathbf{M}^H \mathbf{C}_e^{-1} \tilde{\mathbf{r}}, \quad (2.68)$$

and the WLS (or MLE) solution is given by

$$\hat{\boldsymbol{\sigma}} = (\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M})^{-1} \mathbf{M}^H \mathbf{C}_e^{-1} \tilde{\mathbf{r}}. \quad (2.69)$$

As before, the inversion problem is ill-posed. The weighting by \mathbf{C}_e is omitted if the autocorrelations are not known, or if we may assume that the noise is white and much stronger than the sources.

There exists a third statistical formulation of the imaging problem by making statistical assumptions about the sky map $\boldsymbol{\sigma}$. We will consider this problem formulation in more detail in Chapter 3.

2.5. CONCLUSIONS

In this chapter, we derived the radio interferometric imaging measurement equation. From this measurement equation we introduced the radio interferometric imaging problem and made interpretations to gain insight into the problem. We have briefly mentioned the ill-posedness of the radio interferometric imaging problem. In the following chapter we will analyze the ill-posedness in more detail and introduce methods to compensate for the loss of information in the measurement process.

3

ILL-POSEDNESS AND REGULARIZATION

A lack of information cannot be remedied by any mathematical trickery.

Cornelius Lanczos

In this chapter, we analyze the imaging system to investigate the identifiability of the sky intensity distribution based on the measured telescope correlation data. We show that due to the ill-posedness of the radio interferometric imaging problem, finding a numerically stable and physically meaningful solution requires regularization. We further discuss some regularization methods that can be applied to the radio interferometric imaging problem. We conclude with a discussion on preconditioning and how it can be used to integrate the regularization assumptions and priors into the imaging system.

3.1. ANALYSIS OF THE MEASUREMENT SYSTEM

In this section, we analyze the measurement system in RI based on tools from numerical linear algebra, sampling, interpolation and filtering theory to investigate the inherent loss of information in the measurement process. Based on these results, we show that the imaging problem in radio astronomy is highly ill-posed and regularization is necessary to find a stable and physically meaningful image.

3.1.1. ILL-POSEDNESS OF THE INTEGRAL EQUATION

An interpretation of the information loss in the measurement process in RI can be drawn based on the continuous forward measurement model presented in Equation (2.10) in Chapter 2. Neglecting the DDEs, Equation (2.10) is restated here as

$$\mathcal{V}(u, v, w) = \int_{-1}^1 \int_{-1}^1 I(l, m) e^{-j2\pi[ul+vm+w(\sqrt{1-l^2-m^2}-1)]} \frac{dl dm}{\sqrt{1-l^2-m^2}}. \quad (3.1)$$

This equation describes the measurement process in the continuous case in the form of an integral equation. This integral is in the form of a first-kind Fredholm integral equation [58] as the unknown appears only under the integral sign. This equation shows the underlying linear relationship between the visibility function $\mathcal{V}(u, v, w)$ and the sky map $I(l, m)$ where the model is described by the function

$$K(u, v, w, l, m, n) = e^{-j2\pi\{ul+vm+w(\sqrt{1-l^2-m^2}-1)\}} \quad (3.2)$$

This function is called the kernel function of the integral. The joint effect of the integral and multiplication with K is a compact operator. This compact operator has a smoothing effect on the intensity function $I(l, m)$. This smoothing effect can be described as the higher frequencies (in terms of more rapidly changing components) of I are damped in comparison with the lower frequencies [59]. This effect results in an inherent loss of information. The frequencies lost in the forward measurement process cannot be retrieved in the inverse imaging process. The Riemann–Lebesgue lemma [60] mathematically describes this effect for any arbitrary kernel. Actually, to invert the effect of the integral equation when going from \mathcal{V} to I (the inverse problem), high frequency components of \mathcal{V} must be amplified. If there is a perturbation in \mathcal{V} on the high frequencies, the perturbation will also be amplified. This is an inherent issue in inverse problems that causes the solution to become unstable. Therefore, the inverse problem based on this integral equation is ill-posed.

3.1.2. CONDITIONING ANALYSIS OF THE DISCRETE MEASUREMENT EQUATION

We further explain the loss of information in the radio interferometric measurement process based on the discrete measurement system (2.29). As discussed in Chapter 2, this matrix equation can be derived based on the discretization of the integral equation (3.1). A direct conclusion can be drawn: since the integral problem is ill-posed, the condition number of the discrete equation can be extremely large [48]. A finer discretization for the ill-posed problems, implies a larger condition number of the corresponding discrete problem [48]. Estimating σ from $\tilde{\mathbf{r}}$ based on (2.29) depends on the properties of \mathbf{M} . As mentioned, \mathbf{M} is ill-conditioned and in some cases where the requested resolution (number of pixels) Q is very large it may become wide, so that no unique solution exists. Therefore, the RA imaging problem is ill-posed and in some cases under-determined.

If the sampling of the interferometer is regular, \mathbf{M} can be diagonalized by the Fourier transform. Bracewell and Roberts [61] analyzed the effect of the convolution kernel \mathbf{M} by Fourier transform and frequency domain interpretations. However, in general \mathbf{M} is not exactly a Fourier transform matrix. To be more general, we investigate the effect of smoothing by the antenna pattern by analyzing the singular value spectrum of the measurement matrix \mathbf{M} . We show that we can achieve similar results as presented in [61] with this more general analysis. Only if the telescope array is arranged as a Uniform Linear Array (ULA), the Singular Value Decomposition (SVD) is equivalent to the Discrete Fourier Transform (DFT). However, for an arbitrary geometry of the telescope array, we use the SVD as an analysis tool from numerical linear algebra.

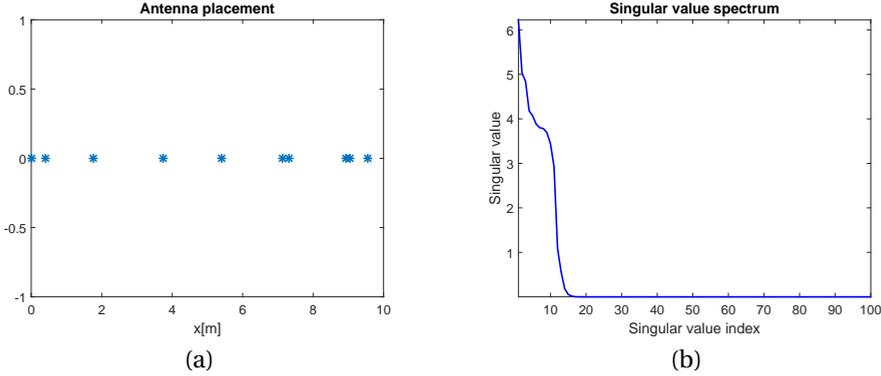


Figure 3.1: (a) one-dimensional irregular array (b) singular value spectrum of \mathbf{M}

The SVD of $\mathbf{M} \in \mathbb{C}^{P^2 \times Q}$ can be written as

$$\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H = \sum_{i=1}^{\rho} \mathbf{u}_i \lambda_i \mathbf{v}_i^H, \quad (3.3)$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{P^2}]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_Q]$ respectively contain the left and right singular vectors, \mathbf{u}_i and \mathbf{v}_i , and $\mathbf{\Lambda}$ is a diagonal matrix containing the singular values of \mathbf{M} , i.e. $\lambda_1, \lambda_2, \dots, \lambda_{\rho}$, on its diagonal in decreasing order. We assume that ρ indicates the rank of matrix \mathbf{M} . In case the rank of \mathbf{M} cannot be defined, ρ corresponds to a truncation length on the singular values.

The singular value spectrum of \mathbf{M} for a simple one-dimensional case with $P = 10$ antennas, positioned as shown in Figure 3.1(a), is shown in Figure 3.1(b). The image is composed of $Q = 201$ pixels. The effect introduced by the Fredholm integral equation in the discrete case translates into the measurement matrix \mathbf{M} having a very large condition number as most of the high frequency singular values of \mathbf{M} concentrate very close to zero. Ill-posed problems are characterized by a large number of small singular values close to zero and by the smooth transition of the singular values towards zero that results in an ill-defined rank. In fact, the main difference between the ill-posed problems and numerically rank-deficient problems is that in the former, in contrary to the latter, there does not exist a distinct gap between the small and large singular values. This effect can be seen in Figure 3.1(b).

The SVD provides a way to compute the inverse of the matrix \mathbf{M} . In case \mathbf{M} is singular, an approximate inverse, a.k.a pseudo-inverse is computed as

$$\mathbf{M}^{\dagger} = \sum_{i=1}^{\rho} \mathbf{v}_i \frac{1}{\lambda_i} \mathbf{u}_i^H, \quad (3.4)$$

where \mathbf{M}^{\dagger} indicates the pseudo-inverse of \mathbf{M} . Moreover, the minimum-norm solution of the LS problem indicated in (2.40) can be computed based on SVD as

$$\hat{\boldsymbol{\sigma}} = \mathbf{M}^{\dagger} \bar{\mathbf{r}} = \sum_{i=1}^{\rho} \frac{\mathbf{u}_i^H \bar{\mathbf{r}}}{\lambda_i} \mathbf{v}_i. \quad (3.5)$$

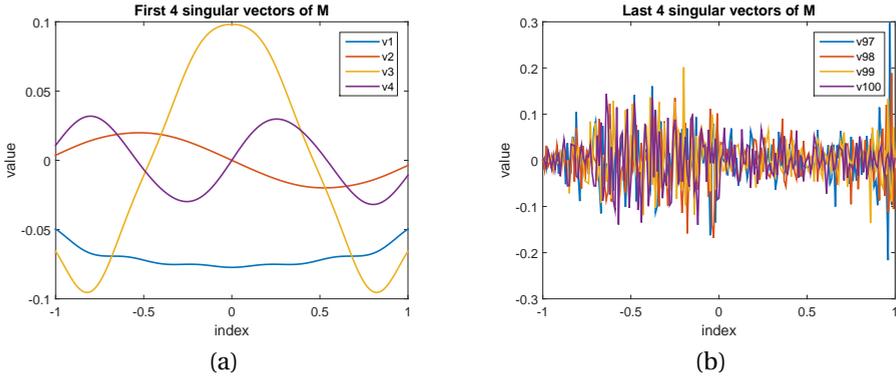


Figure 3.2: (a) First 4 singular vectors of \mathbf{M} (b) last 4 singular vectors of \mathbf{M}

Following [62], we call this solution obtained by direct inversion of the measurement matrix \mathbf{M} the “naive” solution.

Based on Equation (3.5), the naive solution $\hat{\boldsymbol{\sigma}}$ can be represented as an expansion based on the right singular vectors \mathbf{v}_i with coefficients $\frac{\mathbf{u}_i^H \hat{\mathbf{r}}}{\lambda_i}$. We demonstrate the basis vectors \mathbf{v}_i with an example. The first 4 and last 4 singular vectors corresponding to the array shown in Figure 3.1 are shown in Figure 3.2(a) and (b), respectively. As can be seen from the figure, as the index i increases, \mathbf{v}_i exhibits more oscillations.

We can assume that \mathbf{v}_i s are representing the frequencies in the sky map where increasing the index i results in larger oscillations and higher frequency components. Therefore, the small singular values carry mainly high-frequency information of the image. These high frequency components are not well-presented by the measurements $\hat{\mathbf{r}}$ as the high frequencies are largely suppressed in the response of the interferometer. By inverting \mathbf{M} , since the high frequency components are divided by the corresponding small singular values, the recovered value is significantly amplified. Therefore, the recovered values of these frequencies are subject to substantial errors and this causes the solution to become unstable. This explains the numerical instability of the inverse solution.

By computing the pseudo-inverse via truncated SVD (TSVD), by choosing the right truncation length, the contribution of the high frequency components to the image is disregarded. This is in effect similar to applying a low-pass rectangular window on the frequency spectrum of the sky map. However, applying the TSVD on an ill-posed problem results in spurious and implausible oscillations. Furthermore, the main issue in computing the TSVD in ill-posed problems is choosing the right truncation length. This is due to the fact that the system matrix in ill-posed problems exhibits no distinct gap between the large and small singular values and the notion of rank is ill-define. We show this effect with a single unit point source in the middle of the FoV for the array presented in Figure 3.1(a). Theoretically, we can compute the rank of the imaging system based on the number of unique baselines. This amounts to $2P - 1$ for a ULA and to $\frac{P(P-1)}{2}$ for a more general array with non-overlapping antennas. Choosing the truncation length equal to $2P - 1 = 19$ for this example, the TSVD result is shown in Figure 3.3(a). The result is reminiscent of the impulse response of an ideal low-pass filter. On the other

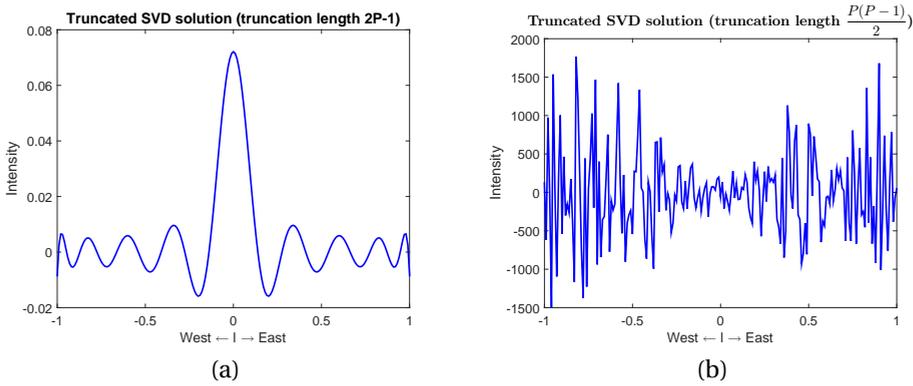


Figure 3.3: TSVD result for a unit point source in the middle of the FoV (a) truncation length $2P - 1$ (b) truncation length $\frac{P(P-1)}{2}$

hand, setting the truncation length equal to $\frac{P(P-1)}{2} = 45$, which is the theoretical rank for this example, results in allowing many small singular values to the computation of the pseudo-inverse which causes instability, dominance of high frequency oscillations, of the solution as shown in Figure 3.3(b). In this case, the solution is devoid of physical meaning.

We explain the sidelobes in Figure 3.3(a), in which the effect of the small singular values are negligible, via Gibbs oscillations. It is well-known that if the Fourier series or Fourier integrals describing a discontinuous function are truncated, i.e. not infinite, oscillations occur in the function representation especially around discontinuities. This is known as the “Gibbs phenomenon” or “Gibbs oscillations”. The same effect happens in radio interferometric image reconstruction based on TSVD. Considering the singular values are roughly presenting the spatial frequencies, TSVD implies a band-limiting effect by truncating the high spatial frequencies. We note that these frequencies are actually not measured by the system due to the finite extent of the array in the spatial frequency plane. In restoring an image containing point sources, spurious oscillations occur around the discontinuities. This is called the ringing effect [48].

3.1.1.3. PRINCIPAL SOLUTION AND INVISIBLE DISTRIBUTIONS

There are two major deficiencies in the sampling of the spatial frequencies of the sky intensity distribution by the radio interferometer: (i) sparse and incomplete sampling that results in the existence of holes in the $u - v$ coverage and (ii) limited spatial bandwidth since the high spatial frequencies are not measured by the interferometer. The former is a cause of aliasing and sidelobes and the latter performs a low-pass filtering. Therefore, the fine details of the observed distribution are irreversibly lost in the measurement process, while the less fine details may be substantially modified [61]. Considering Equation (2.29), if we directly invert the measurement matrix \mathbf{M} to achieve the solution, this means that all the unsampled visibility points are set to zero. This corresponds to the naive solution derived in Section 3.1.2 and is known in RA as the “principal solution” [47, 10]. The intensity distributions containing only the unsampled spatial fre-

quencies produce no output when observed by the telescope array. These distributions are called the “invisible distributions” or “ghosts” for the telescope array [61].

We can interpret the invisible distributions in terms of the subspaces of the measurement matrix \mathbf{M} . In case there is no observation noise, i.e. $\mathbf{e} = \mathbf{0}$, the measured correlations $\tilde{\mathbf{r}}$ lie in the range of \mathbf{M} . However, since \mathbf{M} is ill-conditioned, we can assume a null-space for \mathbf{M} . Therefore, the invisible distributions correspond to the null space of \mathbf{M} , i.e.,

$$\mathbf{M}\boldsymbol{\sigma}_{\text{null}} = \mathbf{0}, \quad (3.6)$$

where $\boldsymbol{\sigma}_{\text{null}}$ is a member of the null-space of \mathbf{M} , i.e. $\boldsymbol{\sigma}_{\text{null}} \in \mathcal{N}(\mathbf{M})$. This shows the inherent loss of information in the measurement process by the telescope array. The null-space of \mathbf{M} does not produce any visible data. In fact, resorting only to Equation (2.29), if $\hat{\boldsymbol{\sigma}}$ is a solution, then $\hat{\boldsymbol{\sigma}} + \alpha\boldsymbol{\sigma}_{\text{null}}$ is also a solution where α denotes any arbitrary scaling. Therefore, the invisible distributions are responsible for the non-uniqueness of the solution. In fact, it would be possible to make a solution containing an element of the null-space arbitrary large and hence unphysical. Furthermore, due to the noise on the measurements, $\tilde{\mathbf{r}}$ is not in the range of \mathbf{M} , i.e. $\tilde{\mathbf{r}} \notin \mathcal{R}(\mathbf{M})$. Instead, we minimize the difference between the data and the model by means of a LS. If \mathbf{M} has a clear null-space the corresponding singular values are exactly zero and the TSVD results in the minimum-norm solution among the set of all possible solutions. However, this is not the case in ill-posed problems. The main problem of radio interferometric imaging is to choose adequate invisible distributions to add to the principal solution to obtain a plausible distribution [28].

These issues, known to radio astronomers [61], signify the ill-posedness claim about the inverse problem in RA: (i) due to the existence of the null-space, the solution of the imaging problem is not unique, (ii) for a noisy distribution that is not in the range of \mathbf{M} , the solution of the imaging problem does not exist and (iii) if we have two close intensity distributions, the reconstructed distributions may be very distant. Therefore, the solution of the imaging problem does not depend continuously on the initial distributions [48]. Therefore, we can see that all the criteria for well-posedness of the problem are violated.

3.1.4. NORMAL EQUATIONS AND THE “DIRTY IMAGE”

Assuming there is no noise in the measurement process, i.e., $\mathbf{e} = \mathbf{0}$ in Equation (2.29), in this section we obtain interpretations based on the normal equations of the measurement system. The normal equations of the measurement system can be written as

$$\mathbf{M}^H \mathbf{M} \boldsymbol{\sigma} = \mathbf{M}^H \tilde{\mathbf{r}}. \quad (3.7)$$

We recognize the right hand side as the dirty image, i.e.,

$$\boldsymbol{\sigma}_{\text{MF}} = \mathbf{M}^H (\mathbf{M} \boldsymbol{\sigma}). \quad (3.8)$$

These equations show the convolution process of the true sky intensity $\boldsymbol{\sigma}$ with the array pattern where multiplication of the image with $\mathbf{M}^H \mathbf{M}$ represents the convolution operation. The dirty image is obtained by back-projecting, represented by multiplication by \mathbf{M}^H , the received signals $\tilde{\mathbf{r}}$ to the source.

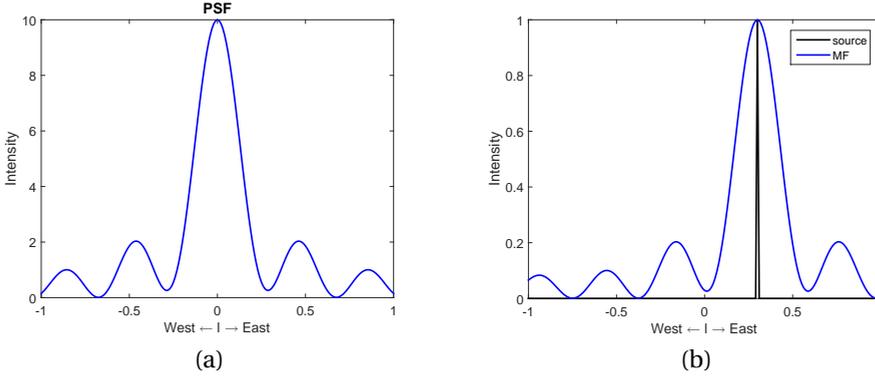


Figure 3.4: (a) PSF (b) dirty image (overlapped with the source)

To obtain the impulse response of the imaging system, we consider the image only contains a unit norm point source in the middle of the FoV. Thus we can rewrite (3.8) as

$$\mathbf{b} = (\mathbf{M}^H \mathbf{M}) \mathbf{e}_{\text{mid}}, \quad (3.9)$$

where \mathbf{e}_{mid} is the unit vector with the element in the middle of the FoV equal to 1. In this case the dirty image is called the dirty beam, indicated by \mathbf{b} . We note that the Point Spread Function (PSF), the impulse response of the telescope array or the beam pattern of the array all refer to the dirty beam. If we insert an arbitrary image σ in the FoV of the array, the resulting σ_{MF} would be the convolution of the dirty beam with the array. Another way to obtain \mathbf{b} is via

$$\mathbf{b} = \mathbf{M}^H \mathbf{1}. \quad (3.10)$$

This equation shows that the array pattern is the matched filtered response of a telescope array when all the antennas are receiving unit power.

For the one-dimensional irregular array presented in Figure 3.1(a), we present the PSF and the dirty image of a unit point source in Figure 3.4(a) and (b), respectively. In this case the normalization for computing the PSF is such that the height of the peak corresponds to the total number of array elements. We can see from the figure that the dirty image in this case is equivalent to the scaled PSF that is shifted to the source position. For the one-dimensional array, the beam pattern of the antenna array is calculated as

$$b(l_q) = \sum_{m=1}^P \sum_{p=1}^P e^{-j \frac{2\pi}{\lambda} (x_m - x_p) l_q}. \quad (3.11)$$

Based on this equation we infer that the half power beam width of the array is approximately $\text{HPBW} \approx \frac{\lambda}{\Delta x_{\text{max}}}$ radians [63]. We see that the size of the main beam, indicating the resolution of the array, is determined by the maximum baseline.

We present some further remarks about the PSF and the dirty image below:

- The linear imaging system of Equation (2.29) is completely characterized by its PSF. In fact, two imaging systems with the same PSFs are equivalent [49]. The point

spread function is completely defined based on the system matrix \mathbf{M} by knowing the geometry of the array and the weightings that can be applied on the measured correlations [49, 28]. In fact, the PSF is the transform of the sampling function of the baseline plane to the image domain [28].

- The PSF is the response of the system to a point source. Due to the sampling deficiencies of the radio telescope; namely (i) spatial band limit and (ii) existence of holes in the $u - v$ coverage, the PSF contains sidelobes. The former is responsible for the existence of sidelobes in the PSF, similar to the effect of applying a rectangular window on the Fourier domain measurements of a signal. On the other hand, the latter may cause grating lobes in the PSF that can cause aliasing in the dirty image.
- The existence of sidelobes in the PSF results in an unavoidable confusion in determining the true intensity distribution. This confusion can only be resolved by introducing some side information about the intensity distribution. For example, consider the dirty image resulting from a unit norm point source as presented in Figure 3.4. In this case, by only looking at the dirty image, we cannot distinguish if the source is a point source or shaped like the dirty beam [28]. Furthermore, the sidelobes resulting from strong sources may obscure fainter emissions in the dirty image.
- We can regard the radio interferometer as a spatial filter. This interpretation can be described by writing $\mathbf{M}^H \mathbf{M} \boldsymbol{\sigma}$ as

$$\hat{\boldsymbol{\sigma}}_{\text{MF},j} = \boldsymbol{\sigma}_j + \sum_{k=1, k \neq j}^Q \boldsymbol{\sigma}_k \mathbf{m}_k^H \mathbf{m}_j \quad (3.12)$$

where $\hat{\boldsymbol{\sigma}}_{\text{MF},j}$ indicates the j th pixel of the MF dirty image and \mathbf{m}_i corresponds to the i th column of \mathbf{M} . We can write $\mathbf{m}_k^H \mathbf{m}_j$ based on the expression for \mathbf{M} in (2.26) as

$$\mathbf{m}_k^H \mathbf{m}_j = \frac{1}{P^2} \sum_{i=1}^{P^2} e^{j \frac{2\pi}{\lambda} (\mathbf{u}_i)^T (\mathbf{z}_j - \mathbf{z}_k)}, \quad (3.13)$$

where \mathbf{u}_i corresponds to the i th baseline. Comparing with Equations (3.11), we see that the function $\mathbf{m}_k^H \mathbf{m}_j$ for $k = 1, 2, 3, \dots, Q$ indicates the PSF centered at pixel j . Furthermore, Equation (3.13) corresponds to a sum of complex exponential functions where the angle depends on the difference $\mathbf{z}_j - \mathbf{z}_k$ between pixel j and k of the image. Since the baselines appear in conjugate pairs, the exponential function can be reduced to a cosine. Therefore, $\mathbf{m}_k^H \mathbf{m}_j$ indicates a sum of cosine functions and the first null location is related to the inverse of the maximum baseline. This shows for a nearby pixel k the damping of the term $\mathbf{m}_k^H \mathbf{m}_j$ in computing $\hat{\boldsymbol{\sigma}}_{j,\text{MF}}$ is not very large and $\boldsymbol{\sigma}_k$ is added to $\boldsymbol{\sigma}_j$ almost in phase. A similar effect happens if pixel k is in a sidelobe of the function $\mathbf{m}_k^H \mathbf{m}_j$ for $k = 1, 2, 3, \dots, Q$. As a result, if extended structures and resolved sources are present in the image, the dirty image will highly over-estimate the value of each pixel in the extended structure due to the additive contribution from the nearby pixels and sidelobes. An example of this

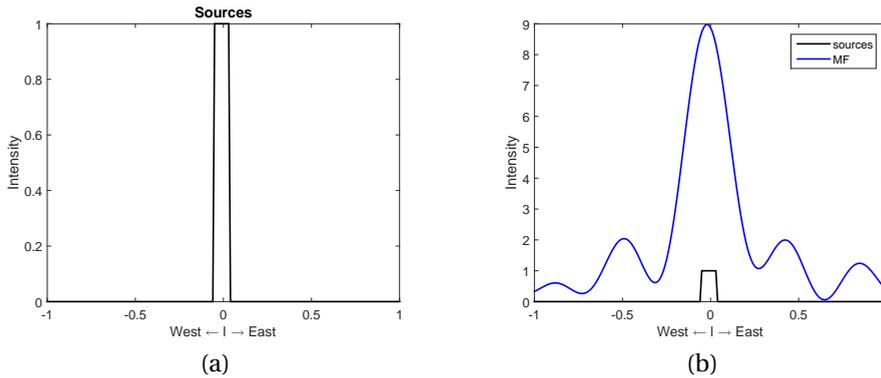


Figure 3.5: (a) Extended source (b) Dirty image (overlapped with the source)

phenomenon for the one-dimensional array with the PSF shown in Figure 3.4(a) is shown in Figure 3.5. In this example, an extended structure is simulated by 9 adjacent pixels with intensity equal to one. The rest of the pixels are set to zero. We can see from Figure 3.5(b) that the dirty image highly over-estimates the source structure such that the pixel in the middle of the image has an intensity approximately equal to the sum of all the sources. This is while for the single point source shown in Figure 3.4(b) the intensity of the corresponding pixel is correctly estimated by the dirty image.

Furthermore, we can see that Equation (3.12) indicates a moving average-type filtering on the image. Therefore, application of this filter on the image results in smoothing of the small-scale variations (e.g. point sources) and enhances the large-scale trends (e.g. extended emissions). As a result, the details of the image that are smaller than the width of the main beam of the PSF are lost in the measurements and cannot be retrieved in the restored image without providing additional information.

3.2. REGULARIZATION

As we have discussed, the radio interferometric imaging problem is inherently ill-posed. Due to the intrinsic loss of information in the measurement process, we cannot opt for an exact solution and we need to resort to an approximation. The set of plausible approximate solutions is very broad [48]. However, not all of the possible solutions are physically meaningful. In fact, as shown before, due to the noise amplification some of the solutions are very unstable, wildly oscillating and deprived of physical meaning. The idea of regularization is to constrain the approximate solution based on a priori information such that the physically implausible solutions are disregarded. In fact, the prior information that is imposed on the sky map must be carefully chosen since it biases the solution towards the preferred solutions by the constraint. The additional information can be imposed in terms of additional constraints on the model fitting problem. The constraints help to discriminate between meaningful and spurious solutions and

replace the ill-posed problem with a nearby well-posed problem.

Assuming a maximum likelihood-type problem formulation, we can write the general regularized WLS radio interferometric imaging problem as

$$\hat{\boldsymbol{\sigma}} = \arg \min_{\boldsymbol{\sigma}} \|\boldsymbol{\Gamma}(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau \mathcal{R}(\boldsymbol{\sigma}) \quad (3.14)$$

where $\boldsymbol{\Gamma}$ is a LS weight, τ is a regularization parameter, $\mathcal{R}(\cdot)$ denotes the regularization operator and $\mathcal{R}(\boldsymbol{\sigma})$ is referred to as the penalty term. The first term in the cost function assures the fidelity of the solution to the data and the second term acts as a penalty to control some (undesired) features of the unknown presented by the function $\mathcal{R}(\boldsymbol{\sigma})$. The regularization parameter controls how much emphasis is put on each of the two terms. The penalty term controls overfitting of the model to the data and avoids numerical problems (instabilities) that make the MLE problem unfeasible for noisy data. Alternatively, we can rewrite the problem formulation (3.14) as

$$\min_{\boldsymbol{\sigma}} \mathcal{R}(\boldsymbol{\sigma}) \quad \text{subject to} \quad \|\boldsymbol{\Gamma}(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 \leq \epsilon. \quad (3.15)$$

There is a data-dependent one-to-one correspondence between τ in (3.14) and ϵ in (3.15) for which the two optimization problems have the same solution and can be used interchangeably. In fact, the former can be achieved from the latter by incorporating the constraint via a Lagrange multiplier τ [59]. Both ϵ and τ are related to the level of noise in the data.

Regularization is needed, in the form of prior knowledge, structure, or other constraints on the solution $\boldsymbol{\sigma}$. Many choices for the regularization operator are possible, a typical regularization operator is the general ℓ_p norm which in general can be stated by $\mathcal{R}(\boldsymbol{\sigma}) = \|\mathcal{F}(\boldsymbol{\sigma})\|_p$. Some examples include $\|\boldsymbol{\sigma}\|_2^2$ (an ℓ_2 constraint or Tikhonov regularization) or $\|\boldsymbol{\sigma}\|_1$ or $\|\boldsymbol{\sigma}\|_0$ (an ℓ_1 or ℓ_0 constraint) a total variation constraint or a maximum entropy constraint. The quality of the solution depends on the properties of the penalty term. These options induce smoothness or sparsity of the solution. Another structural constraint is to require the image pixels to be positive. It is clear that, in this generic form, the problem has been widely studied in many areas of mathematics, engineering, signal processing, and computer science (e.g., machine learning) [64]. In recent years, some of these techniques are now gradually being introduced in the radio astronomy context.

3.2.1. REGULARIZING ASSUMPTIONS IN RADIO INTERFEROMETRIC IMAGING

In this section, we briefly mention some of the common regularizing assumptions and constraints that are extensively used in radio interferometric imaging. In fact, different imaging algorithms can be categorized based on the regularization operator, $\mathcal{R}(\boldsymbol{\sigma})$, used. Usually, a non-linear and iterative method is performed to remove the sidelobes of the PSF from the image and attain a “clean” image. In these procedures, the value of the visibility function at the unmeasured spatial frequencies are estimated. This is done by interpolating the unsampled $u-v$ points (holes) and extrapolating the unmeasured spatial band. To do so, an optimization problem is defined by introducing constraints and assumptions on the sky intensity distribution [55]. For example, extrapolation can be

done by introducing discontinuity assumptions on the sky map such as applying sparsity constraints. Some theoretical studies of the latter has been presented under the name of "super resolution" by Candes and Fernandez-Granda [65].

First of all, there are some basic assumptions that result from the physical features of the sky intensity distribution. Mainly, the distribution is always real-valued and non-negative. That is, $\sigma \geq 0$ and $\sigma \in \mathbb{R}^Q$. Nonnegativity is the main regularizing assumption imposed in the Non-Negative Least Squares (NNLS) algorithm [66] which has been introduced in radio interferometric imaging by Briggs [42] and optimized in this context by Sardarabadi et al. [56]. The objective function for the NNLS can be stated as

$$\min_{\sigma} \|\tilde{\mathbf{r}} - \mathbf{M}\sigma\|_2^2 \quad \text{subject to} \quad \sigma \geq \mathbf{0}. \quad (3.16)$$

Furthermore, Sardarabadi et al. showed that additional upper bounds on the image values can also be imposed using a beamforming-based estimate of the image. They defined the imaging problem via a maximum likelihood estimation formalism with the objective

$$\min_{\sigma} \|\mathbf{\Gamma}(\tilde{\mathbf{r}} - \mathbf{M}\sigma)\|_2^2 \quad \text{subject to} \quad \mathbf{0} \leq \sigma \leq \sigma_{\text{prior}}. \quad (3.17)$$

where in this case σ_{prior} refers to the initial image obtained from beamforming.

The CLEAN algorithm [3] is undoubtedly the most popular imaging algorithm in radio interferometry. Many various modifications have been proposed to the core algorithm such as Clark [67], Cotton-Schwab [68], Multi-Scale CLEAN (MS-CLEAN) [69], multi-resolution CLEAN [70] and a recent optimized version by Offringa and Smirnov [71]. These modified versions have been developed with the purpose to make CLEAN more computationally efficient and increase the accuracy of the estimate. The basic assumption of CLEAN is that the image is composed of a distinct set of point sources and employs a non-linear and greedy algorithm to obtain the support and the source intensity estimates. The regularization provided by CLEAN is based on this assumption that the radio sources can be represented by a number of point sources and the rest of the FoV is empty. It is shown by Marsh and Richardson [72] that the regularizing constraints in CLEAN can be represented as

$$\min_{\sigma} \sum_j \sigma_j \quad \text{subject to} \quad \|\mathbf{\Gamma}(\tilde{\mathbf{r}} - \mathbf{M}\sigma)\|_2^2 \leq \epsilon \quad \text{and} \quad \sigma \geq \mathbf{0}. \quad (3.18)$$

which is actually an ℓ_1 regularized problem due to the non-negativity of the image pixels. Furthermore, it is shown in [69, 34] that CLEAN is a Matching Pursuit (MP) algorithm [73] which is in the family of greedy sparse recovery methods.

We note that the NNLS algorithm as proposed by Lawson and Hanson [66] is based on the active-set method. The active set method is a greedy method similar to CLEAN and in effect also assumes a sparse point source model.

Instead of pixel-based discretization of the sources (point source assumption), they can be discretized over arbitrary basis functions, or dictionaries, based on the physical properties of the astronomical sources [36], that is,

$$\sigma(\ell) = \sum_{j=1}^J \alpha_j \Psi_j(\ell),$$

where α_j s are the basis coefficients, Ψ_j s are the basis functions, ℓ shows the continuous coordinates of the source and J is the total number of basis functions. This is the assumption for the Multi-Scale version of the CLEAN algorithm (MS-CLEAN) [69] where the image is expanded over a set of basis functions (modeled by truncated parabola) of different scales. Moreover, wavelet basis expansion was introduced in astronomy by Starck, Pantin and Murtagh [74] and further developed together with the convex optimization formalism by Carrillo et al. [36] and many efficiently implemented variants followed from the initial average sparsity assumption over the multi-wavelet representation of the intensity distribution [75, 35]. The initial formalism is based on an analysis-based weighted ℓ_1 stated as

$$\min_{\sigma \in \mathbb{R}^Q} \|\mathbf{W}\Psi^H \sigma\|_1 \quad \text{subject to} \quad \|\tilde{\mathbf{r}} - \mathbf{M}\sigma\|_2^2 \leq \epsilon \quad \text{and} \quad \sigma \geq \mathbf{0}. \quad (3.19)$$

which is known as the "Sparsity Averaging Re-weighted Analysis" approach (SARA). By leveraging the versatility of optimization theory in incorporating complex signal models into the image recovery, the SARA formalism has been extensively shown to provide significant improvements in imaging quality in comparison to CLEAN [34, 36, 75, 76].

Another well-known radio interferometric imaging algorithm is the Maximum Entropy Method (MEM) [77, 78]. This algorithm is based on a constrained LS formalism of the imaging problem where the constraint is applied into the problem via the entropy-based penalty function. The most famous version applies the relative image entropy $\mathcal{R}(\sigma) = \sigma \ln(\frac{\sigma}{\sigma_{\text{prior}}})$. In this formulation, the image estimate is biased towards the priori image σ_{prior} [50]. σ_{prior} can be chosen as a low resolution initial estimate of the image such as the MF or MVDR dirty images. Usually a flat image is chosen as the prior image and thus the solution is biased towards being smooth. Therefore, in this case MEM is more suitable for the recovery of extended emissions. MEM is also widely used in power spectrum estimation. In this interpretation, it extrapolates the autocorrelation function outside of the aperture plane defined by the sampling of the antennas [55]. This mitigates the effect of the windowing due to the constraints of the aperture plane. The extrapolation is based on a randomness assumption and is obtained by assuming that the image is as flat as possible [55]. Moreover, $\ln(\sigma)$ imposes a positivity constraint on the image pixels. Entropy in the context of imaging is a broad term. One of the interpretations is "lowest common denominator" which states that entropy is a measure that when minimized produces a positive image with a compressed range in pixel values [47]. Furthermore, Donoho et al. [79] have shown that if the underlying sky map is nearly black, the MEM method results in signal to noise enhancements as well as super-resolution. One possible objective for MEM can be stated as

$$\min_{\sigma} \sigma \ln\left(\frac{\sigma}{\sigma_{\text{prior}}}\right) \quad \text{subject to} \quad \|\tilde{\mathbf{r}} - \mathbf{M}\sigma\|_2^2 \leq \epsilon. \quad (3.20)$$

Since sampling in the spatial frequency domain is bounded and incomplete, linear regularization methods such as Wiener or Tikhonov result in band-limited (smooth) image estimates. This is not optimal for images where there exist point sources [74] since point sources have a flat frequency response that is not band-limited. However, linear regularizations can be applied for the recovery of band-limited extended emissions.

Furthermore, regularization based on truncated iterative methods where the signal subspaces are retrieved in the initial iterations has a similar smoothing effect and provides a linear regularization.

Finally, if some prior knowledge about the existence of isolated point sources in the image is present, this can be imposed via the sparsity-promoting norms ℓ_0 or ℓ_1 into the problem. This results in non-linear methods that provide super-resolution by extrapolating the measurements in the spatial frequency domain to attain sparse images [65].

3.2.2. BAYESIAN PROBLEM FORMULATION AND REGULARIZATION

An alternative approach to regularization is via the Bayesian framework. In this framework, both the noise term and the image are assumed random variables, and a prior distribution $p(\boldsymbol{\sigma})$ is posed. The Maximum A Posteriori (MAP) estimator is defined as [57]

$$\begin{aligned}\hat{\boldsymbol{\sigma}} &= \arg \max_{\boldsymbol{\sigma}} p(\boldsymbol{\sigma}|\tilde{\mathbf{r}}) = \arg \max_{\boldsymbol{\sigma}} \frac{p(\tilde{\mathbf{r}}|\boldsymbol{\sigma})p(\boldsymbol{\sigma})}{\int p(\tilde{\mathbf{r}}|\boldsymbol{\sigma})p(\boldsymbol{\sigma})d\boldsymbol{\sigma}} \\ &= \arg \max_{\boldsymbol{\sigma}} p(\tilde{\mathbf{r}}|\boldsymbol{\sigma})p(\boldsymbol{\sigma}).\end{aligned}\quad (3.21)$$

Here, $p(\boldsymbol{\sigma}|\tilde{\mathbf{r}})$ denotes the posterior probability density function of the image given the observation, and Bayes' rule is used to replace it by $p(\tilde{\mathbf{r}}|\boldsymbol{\sigma})p(\boldsymbol{\sigma})$, which is a product of the likelihood of the observation given an image with the prior probability of that image. The likelihood is given in (2.65). Assuming for simplicity that the prior for the image is also distributed according to a Gaussian distribution, with mean $\boldsymbol{\mu}_{\boldsymbol{\sigma}}$ and covariance $\mathbf{C}_{\boldsymbol{\sigma}}$, then $\boldsymbol{\sigma} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\sigma}}, \mathbf{C}_{\boldsymbol{\sigma}})$, or

$$p(\boldsymbol{\sigma}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\sigma} - \boldsymbol{\mu}_{\boldsymbol{\sigma}})^T \mathbf{C}_{\boldsymbol{\sigma}}^{-1}(\boldsymbol{\sigma} - \boldsymbol{\mu}_{\boldsymbol{\sigma}})\right]. \quad (3.22)$$

The log of the posterior likelihood is then

$$\begin{aligned}\log p(\boldsymbol{\sigma}|\tilde{\mathbf{r}}) &\propto \\ & -(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})^H \mathbf{C}_e^{-1}(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}) - \frac{1}{2}(\boldsymbol{\sigma} - \boldsymbol{\mu}_{\boldsymbol{\sigma}})^T \mathbf{C}_{\boldsymbol{\sigma}}^{-1}(\boldsymbol{\sigma} - \boldsymbol{\mu}_{\boldsymbol{\sigma}}).\end{aligned}\quad (3.23)$$

If we define the Cholesky factorization of the inverse image covariance matrix as

$$\mathbf{C}_{\boldsymbol{\sigma}}^{-1} = \mathbf{L}^T \mathbf{L}, \quad (3.24)$$

we can equivalently write this as

$$\log p(\boldsymbol{\sigma}|\tilde{\mathbf{r}}) \propto -\|\boldsymbol{\Gamma}(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 - \frac{1}{2}\|\mathbf{L}(\boldsymbol{\sigma} - \boldsymbol{\mu}_{\boldsymbol{\sigma}})\|_2^2, \quad (3.25)$$

where $\mathbf{C}_e^{-1} = \boldsymbol{\Gamma}^H \boldsymbol{\Gamma}$. Therefore, maximizing the posterior likelihood is equivalent to solving the minimization problem

$$\hat{\boldsymbol{\sigma}} = \arg \min_{\boldsymbol{\sigma}} \|\boldsymbol{\Gamma}(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau \|\mathbf{L}(\boldsymbol{\sigma} - \boldsymbol{\mu}_{\boldsymbol{\sigma}})\|_2^2, \quad (3.26)$$

where $\tau = \frac{1}{2}$. This is also known as ridge regression and a specific case of (3.14), with the advantage that there is some insight in the role of \mathbf{L} . E.g., if we have accurate prior

knowledge, then \mathbf{C}_σ is small and \mathbf{L} is large, and the solution $\hat{\boldsymbol{\sigma}}$ will be close to $\boldsymbol{\mu}_\sigma$. If instead of a Gaussian prior we assume a Laplace distribution for $\boldsymbol{\sigma}$,

$$p(\sigma_i; \mu_i, b) = \frac{1}{2b} \exp\left(-\frac{|\sigma_i - \mu_i|}{b}\right),$$

we obtain an ℓ_1 constrained MLE formalism of the problem, i.e.,

$$\hat{\boldsymbol{\sigma}} = \operatorname{argmin}_{\boldsymbol{\sigma}} \|\boldsymbol{\Gamma}(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau \|\boldsymbol{\sigma} - \boldsymbol{\mu}\|_1, \quad (3.27)$$

The Laplace distribution is more concentrated around zero and has long tails, which models images that are mostly zero with occasional outliers, explaining why ℓ_1 constraints lead to sparse solutions. Similarly, a lognormal density prior will lead to constraints that generate a maximum-entropy solution [64], and such a prior was used in RESOLVE (Radio Extended SOURces Lognormal deconVolution Estimator) [80]. Thus, the Bayesian framework is a general method to derive constrained optimization problems.

Returning to the Gaussian prior, we can rewrite (3.26) as

$$\hat{\boldsymbol{\sigma}} = \operatorname{argmin}_{\boldsymbol{\sigma}} \left\| \begin{bmatrix} \boldsymbol{\Gamma}\mathbf{M} \\ \sqrt{\tau}\mathbf{L} \end{bmatrix} \boldsymbol{\sigma} - \begin{bmatrix} \boldsymbol{\Gamma}\tilde{\mathbf{r}} \\ \sqrt{\tau}\mathbf{L}\boldsymbol{\mu}_\sigma \end{bmatrix} \right\|_2^2.$$

The corresponding normal equations are

$$(\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M} + \tau \mathbf{C}_\sigma^{-1}) \boldsymbol{\sigma} = \mathbf{M}^H \mathbf{C}_e^{-1} \tilde{\mathbf{r}} + \tau \mathbf{C}_\sigma^{-1} \boldsymbol{\mu}_\sigma, \quad (3.28)$$

and the solution is

$$\hat{\boldsymbol{\sigma}} = (\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M} + \tau \mathbf{C}_\sigma^{-1})^{-1} (\mathbf{M}^H \mathbf{C}_e^{-1} \tilde{\mathbf{r}} + \tau \mathbf{C}_\sigma^{-1} \boldsymbol{\mu}_\sigma).$$

For the specific case where $\boldsymbol{\mu}_\sigma = \mathbf{0}$, and assuming white processes $\mathbf{C}_e = v^2 \mathbf{I}$ and $\mathbf{C}_\sigma = \eta^2 \mathbf{I}$, Equation (3.28) can be written as

$$(\mathbf{M}^H \mathbf{M} + \tau \mathbf{I}) \boldsymbol{\sigma} = \mathbf{M}^H \tilde{\mathbf{r}}, \quad \tau = \frac{1}{2} \left(\frac{v}{\eta}\right)^2 \quad (3.29)$$

which is recognized as a Tikhonov regularized LS problem [48]. Thus, these standard regularization methods are all included in the Bayesian framework.

More generally, many of the regularization techniques can also be formulated in a Bayesian framework, where $\boldsymbol{\sigma}$ is modeled as a random variable, and prior knowledge on $\boldsymbol{\sigma}$ is given in the form of a prior statistical distribution $p(\boldsymbol{\sigma})$, often containing unknown parameters (e.g., scale) which can be modeled statistically as well using hyperpriors [81, 82].

3.3. PRECONDITIONING AS A TOOL FOR REGULARIZATION

In this section, we discuss data weighting and regularizing assumptions on the image under the unifying framework of preconditioning. This section serves as a reference that relates the rest of the chapters and methods developed in this thesis. We do not follow

the formal definition of preconditioning where the main purpose is to speed up the convergence of iterative methods. In fact, in discrete ill-posed problems convergence with the conventional meaning is not the main interest due to the huge condition number of system matrices. The main interest is however in finding a “good” estimate of the image and stopping criteria are chosen with this aim. Therefore, in this section we use preconditioning to incorporate the regularization operator directly into the system matrix such that linear iterative methods can be easily applied. This framework has been introduced in tackling Bayesian estimation problems from a statistical point of view by Calvetti et al [83] and is termed prior-conditioning. We broaden this definition and borrow the name preconditioner superficially as a left or right multiplicand matrix, linear operator, on the system matrix. Borrowing the name preconditioner, we show that we can present statistical regularization as well as selection, support information, baseline weighting and image model basis functions under this unifying framework.

3.3.1. PRECONDITIONING IN NUMERICAL MATHEMATICS

The concept of preconditioning is extensively used in numerical linear algebra to speed up the convergence of iterative solution methods. Iterative solution methods are the method of choice for large scale image reconstruction such as the radio interferometric imaging problem. It has been shown that many iterative methods for image reconstruction can be considered variations of one basic method where the difference is in the type of preconditioner used [84].

The speed of convergence of iterative algorithms is mainly determined by the condition number of the system matrix. A large condition number usually results in the slow convergence of algorithms. The aim of preconditioning in its prime meaning is to cluster the singular values of the system matrix around one to improve the conditioning of the system. This results in faster convergence of the iterative solution methods. The operator that applies the preconditioning into the system matrix is called a preconditioner. There are two variants of preconditioners distinguished based on the placement of the operator, i.e. right and left preconditioners. If the preconditioner is applied at the left of the system matrix, it is called a left preconditioner. A similar definition holds for a right preconditioner that is applied from the right to the system matrix [85].

Based on these definitions, we can transform the radio interferometric imaging system by applying right and left preconditioners as

$$\bar{\mathbf{M}} = \mathbf{\Pi}\mathbf{M}\mathbf{\Theta}, \quad (3.30)$$

where $\bar{\mathbf{M}}$ denotes the transformed system and $\mathbf{\Pi}$ and $\mathbf{\Theta}$ respectively denote left and right preconditioners. Therefore, the measurement system (2.29) is transformed into a preconditioned system as

$$\mathbf{\Pi}\bar{\mathbf{r}} = \mathbf{\Pi}\mathbf{M}\mathbf{\Theta}\boldsymbol{\alpha} + \mathbf{\Pi}\mathbf{e}. \quad (3.31)$$

Applying the right preconditioner changes the image model as $\boldsymbol{\sigma} = \mathbf{\Theta}\boldsymbol{\alpha}$ where $\boldsymbol{\alpha}$ are the coefficients of the new model. To cluster the singular values of \mathbf{M} around one, in a well-posed problem, right and left preconditioners approximate an inverse of \mathbf{M} . This is not feasible in ill-posed problems since a good inverse of \mathbf{M} would be highly unstable [86].

3.3.2. PRECONDITIONING OF ILL-POSED PROBLEMS

Preconditioning is largely used to accelerate the convergence of well-posed problems by clustering the singular values around one. However in ill-posed problems, there exist a large number of small singular values that cluster around zero and if the preconditioning is not done appropriately, the algorithm will converge fast to a poorly approximated solution [84]. In ill-posed problems the singular value distribution of the system matrix provides important information about the signal and noise subspaces as mentioned in Chapter 2. Mainly, the large singular values correspond to subspaces that contain well-presented information about the image (signal subspaces) and the small singular values correspond to noisy subspaces. By conventional preconditioning, these singular values are clustered around one and the information about signal and noise subspaces is lost and cannot be retrieved by the iterative method. This in turn results in fast convergence of the iterative method to a noise corrupted solution. In ill-posed problems, preconditioning must be applied in a way such that the singular values corresponding to the signal subspace (corresponding to large singular values) are modified and the small singular values are not affected [87]. This is in order not to destroy the information on the signal and noise subspaces which provide valuable means for regularization. In fact, if preconditioners are applied correctly in ill-posed problems, they can be used as a means for regularizing the problem.

Some iterative methods for solving a linear system similar to (2.29) exhibit a “preconvergence” behavior. Preconvergence refers to the property of some iterative methods to obtain a better estimate of the underlying variable at the intermediate iterations than at convergence. Preconvergence is due to the fact that the components of the solution corresponding to the signal subspace are mainly obtained during the initial iterations. As the iteration continues, some components of the noise subspace are also recovered resulting in instability in the solution. In these methods, the iterations may be truncated before convergence to avoid recovery of the solution components presented by the noisy subspaces. In these methods, the truncation length is used as a type of regularization where the regularization parameter is defined based on a function of the iteration count. Iterative regularization [59] is based on truncating the iterations prematurely such that the components from the noise subspace do not obscure the solution.

Dealing with ill-posed problems, preconditioning provides a tool for applying regularizing assumptions, constraints, prior models of the sky intensity distribution and statistical information directly into the measurement system. We call these preconditioners “regularizing preconditioners”. These preconditioners may not increase the rate of convergence of the iterative methods or they may even slow down the convergence but the solution is guaranteed to be in accordance with the prior and thus credible.

We have seen in Equation (3.31) that applying regularization changes the system. A right preconditioner is a linear operator that affects σ and can be used to represent additional requirements and prior models for the solution. On the other hand, the left preconditioner is applied on the measurement vector $\tilde{\mathbf{r}}$ and can be used to influence the measurements in some way. As an example we can see that in the MLE problem formulation from Chapter 2, statistical assumptions about the measurement error \mathbf{e} are applied by the matrix $\mathbf{\Gamma}$ into the system. $\mathbf{\Gamma}$ can be viewed as a left preconditioner.

In the rest of this section, we show some information that can be incorporated into

the radio interferometric imaging system matrix by a preconditioner to regularize the solution. We benefit from these regularizations in later chapters to develop efficient custom-made regularized iterative solution methods for the radio interferometric imaging problem.

3.3.3. STATISTICAL PRECONDITIONING: PRIORCONDITIONING

We reconsider the Bayesian problem with Gaussian priors stated in Section 3.2.2, Equation (3.26). As stated earlier, this formulation corresponds to the Generalized Tikhonov regularized solution and the second term penalizes the growth of the solution affected by the operator \mathbf{L} . However, the operators \mathbf{L} and $\mathbf{\Gamma}$ are computed based on the prior statistical assumption about the signal and noise. In fact these operators are computed as the Cholesky factor of the inverse of the prior signal and noise covariance matrices, i.e. $\mathbf{C}_\sigma^{-1} = \mathbf{L}^H \mathbf{L}$ and $\mathbf{C}_e^{-1} = \mathbf{\Gamma}^H \mathbf{\Gamma}$.

We restate problem (3.26) with a single LS to observe the resulting system of equations and apply the transform $\boldsymbol{\alpha} = \mathbf{L}\boldsymbol{\sigma}$. The resulting system can be stated as

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \left\| \begin{bmatrix} \mathbf{\Gamma} \mathbf{M} \mathbf{L}^{-1} \\ \sqrt{\tau} \mathbf{I} \end{bmatrix} \boldsymbol{\alpha} - \begin{bmatrix} \mathbf{\Gamma} \bar{\mathbf{r}} \\ \sqrt{\tau} \mathbf{L} \boldsymbol{\mu}_\sigma \end{bmatrix} \right\|_2^2. \quad (3.32)$$

The system matrix in this case is $\bar{\mathbf{M}} = \mathbf{\Gamma} \mathbf{M} \mathbf{L}^{-1}$. We see that operators $\mathbf{\Gamma}$ and \mathbf{L}^{-1} appear as right and left preconditioners. These preconditioners that are computed based on the prior statistical assumptions about the signal and the noise are called "priorconditioners" [83]. In fact, they apply whitening of the prior and the noise, i.e. $\text{Cov}(\boldsymbol{\alpha}) = \mathbf{I}_Q$ and $\text{Cov}(\mathbf{\Gamma} \mathbf{e}) = \mathbf{I}_{p^2}$.

This type of preconditioner can be used in combination with the iterative methods that exhibit preconvergence. The truncation length applies a regularization based on the signal and noise subspaces and applying priorconditioners is shown to enrich the signal subspace [88]. In fact they have been shown to accelerate the rate of convergence of the portion of the solution corresponding to the signal subspace while leaving the noise subspace unaffected [88].

3.3.4. PRECONDITIONERS FOR SPARSE REPRESENTATION

As mentioned earlier, right preconditioners provide a means to incorporate assumptions about the model of the sky map directly into the measurement system. In the context of compressive sensing, where the underlying signal exhibits some sparsity, right preconditioners can be used to incorporate available information about the support of the signal into the system. This is especially interesting in radio astronomy as most of the sky is empty and source emissions appear either as clusters of diffuse emission or isolated point sources. Therefore, the image exhibits a bounded support. We recall the preconditioned system matrix stated in Equation (3.30) as $\bar{\mathbf{M}} = \mathbf{\Pi} \mathbf{M} \boldsymbol{\Theta}$. If there is some prior knowledge about the support of the image, e.g. via a low resolution initial image, this can be inserted into $\boldsymbol{\Theta}$. Therefore, in this case $\boldsymbol{\Theta}$ is a diagonal masking matrix defined as

$$[\boldsymbol{\Theta}]_{i,i} = \begin{cases} 1, & \text{if } [\boldsymbol{\sigma}_{\text{true}}]_i > 0 \\ 0, & \text{if } [\boldsymbol{\sigma}_{\text{true}}]_i = 0. \end{cases} \quad (3.33)$$

Applying the right preconditioner as defined corresponds to a column selection operation on the system matrix \mathbf{M} . Applying the support information this way, reduces the size of the system as many pixels are discarded (set to zero) for image formation.

Furthermore, if the sky map can be presented sparsely in some (overdetermined) basis, the right preconditioner can be adjusted to incorporate the sparsity basis. In this case, the modeling of the image changes as $\boldsymbol{\sigma} = \boldsymbol{\Theta}\boldsymbol{\alpha}$ where $\boldsymbol{\Theta}$ represents the new basis for image representation and $\boldsymbol{\alpha}$ is a sparse vector containing the coefficients of the image in the new basis.

3

3.3.5. ANTENNA/BASELINE SELECTION

The concept of sensor selection is used in incoherent imaging and compressive sampling to design antenna arrays. The goal is mainly to find the best (based on some optimization objective) arrangement of the antennas in the spatial frequency field such that the minimum number of antennas are applied. The motivation is to have an inexpensive array especially when the sensors, in our case radio telescopes, are expensive.

Selection can be applied on the antennas or directly on the baselines. We call the former baseline selection and the latter sensor selection. Baseline selection is particularly interesting for data reduction in radio interferometry. The motivation in this case is to reduce the amount of data that is used in image formation without sacrificing the image reconstruction performance such that it results in savings in data transport and imaging computations. Since doing so, the size of the data is reduced and the redundant data are neglected.

We can use a left preconditioner to model sensor/baseline selection. Sensor selection can be applied based on Equation (2.44) from Chapter 2. We define the sensor selection matrix as $\boldsymbol{\Phi} = \text{diag}(\mathbf{w})$ where \mathbf{w} is the selection vector containing only zeros and ones. This translates into a left preconditioning matrix on the measurement system defined as $\boldsymbol{\Xi} = (\boldsymbol{\Phi}^* \otimes \boldsymbol{\Phi})$, corresponding to the cross-correlation of all the weights applied on the sensors. For baseline selection we can directly define the left preconditioning matrix as $\boldsymbol{\Xi} = \text{diag}(\mathbf{w})$ where in this case \mathbf{w} is the baseline selection vector of size P^2 with zero and one elements. Therefore, in this case the size of the correlation data will reduce as $\boldsymbol{\Xi}\tilde{\mathbf{r}}$. Applying baseline selection instead of sensor selection increases the degrees of freedom [49]. Baseline or sensor selection can be viewed as a row selection operation on matrix \mathbf{M} to discard non-informative rows.

The point spread function provides a characterization of the sensor placement geometry and the weighting applied on the baselines. Based on the formulation of Chapter 2, we can integrate the left preconditioning in the definition of PSF, i.e.,

$$\mathbf{b} = \mathbf{M}^H \boldsymbol{\Pi}^H \mathbf{1}. \quad (3.34)$$

Discarding the redundant baselines, the shape of the PSF does not change. This concept is used in minimum redundancy [53] array or sparse ruler [89] array designs to minimize the number of redundant baselines. However by discarding elements, the sensitivity of the array, corresponding to the peak of the PSF, is altered. This can be remedied by increasing the integration time for correlation formation.

3.3.6. PRECONDITIONING IN RADIO ASTRONOMY

The concept of preconditioning has been used in radio astronomy to refer to baseline weighting [28, 90]. We can regard baseline weighting as a left preconditioner similar to the one applied in baseline/sensor selection where the weights can be arbitrary numbers (rather than zeros and ones). The purpose of baseline weighting is to adjust the shape of the PSF such that the quality of image reconstruction is improved. There are mainly two criteria for selection of weights; (i) maximizing sensitivity of the instrument, or (ii) minimizing PSF sidelobe level [91]. In fact, by applying weights to the baselines, we can artificially construct arrays for which the PSF is a subset of the PSF of the original array [49]. Selection of weights differs based on the science case.

Translating baseline weighting to left preconditioning, $\mathbf{\Pi}$ is a diagonal matrix with arbitrary values. In radio astronomy $\mathbf{\Pi}$ can be decomposed as

$$\mathbf{\Pi} = \mathbf{KTD} \quad (3.35)$$

where \mathbf{K} , \mathbf{T} and \mathbf{D} are all diagonal matrices. \mathbf{K} indicates the reliability of the visibility sample depending on the integration time, system temperature and bandwidth at the corresponding baseline. This parameter cannot be controlled or changed during image formation. However, \mathbf{T} and \mathbf{D} can be adjusted to control the beam shape. \mathbf{T} is the baseline tapering function to control, "fine tune", the shape of the PSF. \mathbf{T} is implemented as a smooth function, usually a Gaussian, used to weigh down the data at the outer edge of the $u-v$ coverage. This results in a suppression of the small-scale sidelobes at the expense of broadening the main beam (reducing the resolution). This is similar to sidelobe reduction windows, Bartlett, Hamming, Hann etc., in spectral analysis [55]. Furthermore, \mathbf{D} is the weighting applied to control the effect of sampling density in the $u-v$ plane in image reconstruction [28]. \mathbf{D} is used to simulate a more uniform sampling in the $u-v$ plane by reducing the effect of high sampling in the center of the plane and to lessen the sidelobes due to the existence of gaps in the $u-v$ coverage. This is done by applying weights to each measurement with a reciprocal of the local measurement density [28].

Different weighting schemes exist in radio interferometric imaging; Natural weighting applies equal weights to all the samples and therefore, preserves the natural resolution and the peak sensitivity of the instrument. Therefore, it gives the best SNR for detecting weak sources. However, this scheme does not provide enough resolution for images containing large- and small-scale sources. If the measurement noise is white, natural weighting can be applied as $\mathbf{K} = \mathbf{I}$ or in case of non-white noise as $\mathbf{K} = \mathbf{\Gamma}$. Sardarabadi et al. [56] have shown the correspondence between natural weighting and the statistical weighting applied in the ML formulation. In fact, left preconditioning affects the statistics of the noise on the system. Therefore, left preconditioning must be applied with care as it may affect the statistical assumptions (such as maximum likelihood) made in the problem formulation [84].

By applying uniform density weights to all the samples the main beam of the PSF can be narrowed resulting in an increase in the resolution of the instrument. This concept is used in the uniform weighting scheme. In uniform weighting, \mathbf{D} is such that the value of each visibility sample is divided by the sampling density in a symmetric region around it. Size and shape of the region can be controlled. Uniform weighting is used to obtain the

maximum resolution and to emphasize small-scale structure of bright sources whereas natural weighting is better suited to emphasize extended structures of faint sources.

There exist some intermediate weighting schemes such as the robust weighting [42] and adaptive weighting [91]. Baseline weighting has recently been used as a left preconditioning method to accelerate the convergence of a primal-dual distributed imaging algorithm [92]. Onose et al. [92] have shown that convergence acceleration can be achieved using the sampling density information, especially for highly nonuniform sampling patterns. They have shown that by incorporating the sampling density information, the algorithm step size can be adjusted such that it makes a large step toward the solution in each iteration. This in turn accelerates the convergence. In fact, by applying left preconditioning, both sampling density and the noise statistics of visibility measurements can be incorporated in the algorithmic structure resulting in an increase in convergence speed and increase in image reconstruction quality.

3.4. CONCLUSIONS

In this chapter, we analyzed the radio interferometric imaging system and concluded that the radio interferometric imaging problem is highly ill-posed. Furthermore, we reviewed the available regularization methods. Last but not least, we proposed a framework based on preconditioning to incorporate the regularization operator directly into the measurement system. In the following chapters, we use this formulation to design regularized radio interferometric imaging algorithms and methods.

II

PART II : DATA-DRIVEN REGULARIZATION

4

DATA-DRIVEN REGULARIZATION FOR DIRECT DECONVOLUTION

4.1. INTRODUCTION

In this chapter, based on a MLE formulation of the radio interferometric imaging problem, we propose two regularization methods for the deconvolution problem in RA. The first method is based on weighted truncation of the eigenvalue decomposition of the image deconvolution matrix. In the second method, we introduce a conditioning of the deconvolution matrix based on the prior knowledge of the dirty image using the available data. The effectiveness of the proposed regularization methods is demonstrated by simulations and on actual data from a LOFAR [1] station.

4.2. RADIO INTERFEROMETRIC DECONVOLUTION PROBLEM

We recall the MLE problem formulation for RA imaging stated in Chapter 2, Equation (2.69),

$$\hat{\boldsymbol{\sigma}} = \underset{\boldsymbol{\sigma}}{\operatorname{argmin}} \|\boldsymbol{\Gamma}(\tilde{\mathbf{r}} - \mathbf{M})\boldsymbol{\sigma}\|_2^2,$$

and the corresponding normal equations:

$$\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M} \hat{\boldsymbol{\sigma}} = \mathbf{M}^H \mathbf{C}_e^{-1} \tilde{\mathbf{r}},$$

where $\mathbf{C}_e^{-1} = \boldsymbol{\Gamma}^H \boldsymbol{\Gamma}$ and $\mathbf{C}_e = \frac{1}{N} (\mathbf{R}^T \otimes \mathbf{R})$. The closed form solution of this problem can be stated as

$$\hat{\boldsymbol{\sigma}} = (\boldsymbol{\Gamma} \mathbf{M})^\dagger \boldsymbol{\Gamma} \tilde{\mathbf{r}} = (\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M})^{-1} \mathbf{M}^H \mathbf{C}_e^{-1} \tilde{\mathbf{r}}. \quad (4.1)$$

Part of this chapter is published as: S.Naghizadeh, A. Mouri Sardarabadi and A.J. van der Veen. Radioastronomical image reconstruction with regularized least squares. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3316-3320, 2016.

This formulation assumes the receiver noise power \mathbf{r}_n is known and subtracted from the measurements to obtain $\tilde{\mathbf{r}}$. However, if \mathbf{r}_n is not known, as proposed by Wijnholds [44], the weighting Γ can be changed to a modified weighting $\tilde{\Gamma}$ obtained as

$$\tilde{\Gamma} = \Gamma - \Gamma((\mathbf{I} \circ \mathbf{I})^H \mathbf{C}_e^{-1} (\mathbf{I} \circ \mathbf{I}))^{-1} (\mathbf{I} \circ \mathbf{I})^H \Gamma^H \Gamma. \quad (4.2)$$

Equation (4.1) describes the deconvolution problem in RA where $(\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M})^{-1}$ represents the deconvolution operation. With this operation, the effect of the convolution by the PSF is removed from the image. In this equation, $\mathbf{H} = \mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M}$ is called the deconvolution matrix and $\hat{\boldsymbol{\sigma}}_{\text{MF}} = \mathbf{M}^H \mathbf{C}_e^{-1} \tilde{\mathbf{r}}$ denotes the MF dirty image. To enable the astronomers to draw proper interpretations of the cosmic objects and phenomena, the image is required to have a high dynamic range. In the dirty image, due to the convolution of the true distribution with the PSF there exist high sidelobes that limit the dynamic range of the image significantly. In fact, some weak sources may be completely drawn under the sidelobes of the bright sources. Therefore, to achieve a high dynamic range, deconvolution is performed to remove the effect of the sidelobes of the PSF from the dirty image to obtain a reasonable estimate of the sky image.

If multiple frequency channels and/or time snapshots are measured, e.g. K frequency channels and/or snapshots, and the sky map does not change over the K frequency channels and snapshots, we can write the aggregate normal equations as

$$\sum_{k=1}^K \mathbf{H}_k \hat{\boldsymbol{\sigma}} = \sum_{k=1}^K \hat{\boldsymbol{\sigma}}_{\text{MF},k} \quad k = 1, 2, 3, \dots, K. \quad (4.3)$$

We call $\mathbf{H} = \sum_{k=1}^K \mathbf{H}_k$ and $\hat{\boldsymbol{\sigma}}_{\text{MF}} = \sum_{k=1}^K \hat{\boldsymbol{\sigma}}_{\text{MF},k}$. Therefore, the deconvolution problem for the aggregate snapshot and frequency channels becomes

$$\hat{\boldsymbol{\sigma}} = \mathbf{H}^{-1} \hat{\boldsymbol{\sigma}}_{\text{MF}}. \quad (4.4)$$

Since in RA the deconvolution matrix \mathbf{H} is highly ill-conditioned, instead of a direct inversion, a pseudo-inverse, indicated by \mathbf{H}^\dagger is applied. A linear deconvolution is obtained by the direct inversion of the deconvolution matrix by means of a (truncated) SVD, or since the deconvolution matrix is Hermitian via a (truncated) Eigen Value Decomposition (EVD). This is first considered in the context of RA by Briggs [42]. However, this method cannot be applied to large scale problems due to the storage and computational limitations. Furthermore, since the deconvolution problem is ill-posed this results in many complications due to noise amplifications.

4.3. PROPOSED METHODS

The authors in [93] have proposed a direct data-driven model-based least squares method based on Karhunen-Loeve transform (KLT) to obtain the signal power estimates $\hat{\boldsymbol{\sigma}}$ from the noisy covariance data. As the results in [93] suggest, the KLT-based deconvolution process introduces a ripple effect, similar to the Gibbs oscillations introduced in Chapter 2, Section 3.1.2, in the reconstructed image due to the truncation of the eigenvalues with a rectangular window. In this section, we propose two regularization methods to improve the reconstructed image quality based on a linear deconvolution performed by the EVD.

4.3.1. WEIGHTED TRUNCATED EIGENVALUE DECOMPOSITION

The first method is based on introducing a weighting matrix Φ in the eigenvalue decomposition of the deconvolution matrix \mathbf{H} to perform smoother spectral windowing [94] and to reduce the ripple effect caused by the KLT-based method proposed in [93].

The eigenvalue decomposition of the deconvolution matrix is stated as

$$\mathbf{H} = \mathbf{V}\mathbf{\Omega}\mathbf{V}^H, \quad (4.5)$$

where $\mathbf{\Omega}$ is a diagonal matrix containing the eigenvalues of \mathbf{H} sorted in decreasing order and the columns of the matrix \mathbf{V} contain the corresponding eigenvectors. According to [93], by truncating the eigenvalues of \mathbf{H} , the deconvolution matrix can be represented by the set of dominant eigenvalues and eigenvectors as

$$\mathbf{H} \approx \hat{\mathbf{V}}\hat{\mathbf{\Omega}}\hat{\mathbf{V}}^H, \quad (4.6)$$

where $\hat{\mathbf{\Omega}}$ is a diagonal matrix composed of the significant eigenvalues of \mathbf{H} and the columns of $\hat{\mathbf{V}}$ contain the corresponding significant eigenvectors. This method is known as numerical filtering [95].

To reduce the effect of spectral truncation with a sharp rectangular window, we propose to introduce a diagonal weighting matrix Φ to correct for the different dominance of the remaining spatial frequencies. Defining \mathbf{H}^\dagger as the weighted inverse of the truncated eigenvalue decomposition of \mathbf{H} , the proposed method computes the image estimate as

$$\hat{\boldsymbol{\sigma}} = \mathbf{H}^\dagger \hat{\boldsymbol{\sigma}}_{\text{MF}} = \hat{\mathbf{V}}\Phi\hat{\mathbf{\Omega}}^{-1}\hat{\mathbf{V}}^H \hat{\boldsymbol{\sigma}}_{\text{MF}}. \quad (4.7)$$

Truncating the eigenvalues acts in a similar way as low-pass spatial filtering. Based on the analysis in Chapter 3, Section 3.1.2, we know that the smaller eigenvalues correspond to higher frequency information of the image which is mostly captured by the longer baselines. Since usually the longer baselines are sampled more sparsely, they contain less information. Therefore, the corresponding eigenvalues have smaller magnitude. This affects the condition number of the deconvolution matrix \mathbf{H} adversely.

TSVD implies a band-limiting by truncating the high spatial frequencies. We note that these frequencies are actually not measured by the system due to the finite extent of the array in the spatial frequency plane. In restoring an image containing point sources, spurious oscillations occur around the discontinuities. This is called the ringing effect [48].

Furthermore, the high frequency information of the image is not measured by the interferometer since the spatial frequency coverage by the interferometer is inevitably band-limited. Truncated EVD imposes a hard threshold to discard the eigenvalues corresponding to these spatial frequencies and by doing so, causes spurious Gibbs oscillations around the discontinuities. Using smoother filter shapes such as a triangular window instead of the rectangular window, that is normally used in numerical filtering, reduces the sidelobes at the cost of a decreased resolution [55].

4.3.2. BEAMFORMING-BASED CONDITIONING

The second proposed regularization method uses the prior knowledge that the expected value of the dirty image is an upper bound on the desired source power and that the

source powers are positive [56], that is,

$$\mathbf{0} \leq \boldsymbol{\sigma} \leq \boldsymbol{\sigma}_{\text{MF}}, \quad (4.8)$$

where $\boldsymbol{\sigma}_{\text{MF}} = \mathbf{M}^H \mathbf{r}$ represents the expected value of the dirty image. This relation has been further explained in Chapter 2, Section 2.4.1. By defining a conditioning vector \mathbf{d} based on the true dirty image, the inequality condition (4.8) can be translated into conditioning weights d_i satisfying $0 \leq d_i \leq 1$, that is,

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_{\text{MF}} \odot \mathbf{d}. \quad (4.9)$$

The conditioning weights are then computed as

$$\boldsymbol{\sigma}_{\text{MF}} = \mathbf{H}\boldsymbol{\sigma} = \mathbf{H}\text{diag}\{\boldsymbol{\sigma}_{\text{MF}}\}\mathbf{d} \quad (4.10)$$

$$\mathbf{d} = (\mathbf{H}\text{diag}\{\boldsymbol{\sigma}_{\text{MF}}\})^\dagger \boldsymbol{\sigma}_{\text{MF}}. \quad (4.11)$$

In practice, we use the dirty image obtained from the noisy data, i.e. $\hat{\boldsymbol{\sigma}}_{\text{MF}}$, to obtain an estimate of \mathbf{d} denoted as $\hat{\mathbf{d}}$. Furthermore, we compute the pseudo-inverse mentioned in (4.11) by $(\cdot)^\dagger$ by a truncated singular value decomposition of the deconvolution matrix \mathbf{H} conditioned by weights obtained from the dirty image. That is,

$$(\mathbf{H}\text{diag}\{\hat{\boldsymbol{\sigma}}_{\text{MF}}\})^\dagger = \hat{\mathbf{Y}}\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{U}}^H, \quad (4.12)$$

where $\boldsymbol{\Psi}$ is a diagonal matrix containing the significant singular values of $(\mathbf{H}\text{diag}\{\hat{\boldsymbol{\sigma}}_d\})$ and the columns of $\hat{\mathbf{Y}}$ and $\hat{\mathbf{U}}$ contain the corresponding right and left singular vectors. Therefore, the source power estimates are then obtained by

$$\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}_{\text{MF}} \odot \hat{\mathbf{d}}.$$

The idea behind conditioning the deconvolution matrix with the dirty image is that the discrete source model assumes a point source per image pixel. However, if the resolution is chosen high, source distributions are spread over multiple pixels which causes a linear dependence in the columns of the deconvolution matrix. Applying the conditioning weights to the deconvolution matrix \mathbf{H} based on the dirty image promotes the columns with more available information, i.e. source power while demoting the columns related to the empty parts of the sky, that contribute to less power in the dirty image, before applying the inversion.

4.4. SIMULATION RESULTS

In this section, the performances of the proposed methods are evaluated using a set of simulations. A uniform linear array (ULA) configuration with $P = 15$ antennas, placed at half-wavelength spacing, with scanning angles in the range $[-90^\circ, 90^\circ]$ is considered. The resolution of the image is chosen to be 0.1 of the main beam which results in $Q = 235$ image pixels. 5 point sources with intensities $[5, 5, 0.8, 5, 5]$ are placed at angles $[-23^\circ, -17^\circ, 7.5^\circ, 23^\circ, 32.5^\circ]$. $N = 10^5$ time samples are used to construct the sample covariance matrix and the eigenvalues are truncated to the length $T = 2P - 1$ due to the

fact that in the special case of a ULA matrix \mathbf{V} represents a spatial Fourier transform with the unique weights defined by the number of unique baselines, i.e. $2P - 1$. The obtained dirty image is shown in figure 4.1(a). Following section 4.1, figure 4.1(b) shows the source power estimates obtained by a truncated eigenvalue decomposition. As can be seen, the truncation causes sinc-like ripple effects, the two nearby sources are not distinguishable and the smallest source is completely buried in the sidelobes of the other sources. Figure 4.1(c) illustrates the effect of applying the triangular weighting function on the truncated eigenvalues. The result shows significantly reduced sidelobes at the expense of reduced resolution, the close sources are still not distinguishable but the small source is recovered. The triangular weighting function with weights ϕ_i is computed as [94]

$$\phi_i = \begin{cases} 1 - \frac{i-1}{T} & i \leq T \\ 0 & i > T \end{cases}. \quad (4.13)$$

and is scaled by window power per length, i.e. $\frac{1}{T} \sum_{i=1}^T |\phi_i|^2$.

Furthermore, the effect of the proposed conditioning method on the point source estimates, as discussed in section 4.2, is demonstrated in Figure 4.1(d). As can be seen, by using the prior knowledge from the dirty image in the deconvolution process, the dynamic range is significantly increased, the sidelobes are reduced and a sharper estimate of the source positions with less bias is obtained. Moreover, the nearby sources are clearly distinguishable. Focusing on Figure 4.1(d), we see that the reconstructed image exhibits negative sidelobes. We think that this effect is caused since multiplying \mathbf{H} on the right by the dirty image as discussed in Equation (4.11), hampers the positive definiteness of the deconvolution matrix \mathbf{H} which exhibits itself in the reconstructed image.

In the special case of a regularly-sampled linear array, the unitary spectral decomposition matrix \mathbf{U} equals a Fourier transform matrix \mathbf{F} and a rectangular window on the frequency content of the signal causes a sinc function in the spatial content. By applying a triangular window, the sidelobes of the sinc function are reduced at the expense of a wider main beam.

4.5. EXPERIMENTAL RESULTS

Actual data from the LOFAR telescope [17] was used to investigate the effect of the proposed methods on the reconstructed image quality. For comparison reasons, the same data set as introduced in [93] was used. The data is captured by 48 semi-randomly spaced planar antennas and consists of 25 time snapshots of 10 seconds each in 25 distinct frequency channels of bandwidth 156 kHz between 45.156 and 67.188 MHz. Since we have multiple frequency channels and time snapshots, we use Equation (4.3) to construct the deconvolution problem. The image is sampled regularly on the projected plane and the same number of image pixels, $Q = 8937$ and eigenvalue truncation threshold of $\lambda_{max}/200$ as in [93] was chosen, where λ_{max} denotes the maximum eigenvalue of the deconvolution matrix \mathbf{H} . The normalized dirty image obtained from matched filter beamforming is shown in Figure 4.2(a).

Using a weighted LS formulation, the resulting normalized KLT-based image as proposed in [93] is shown in Figure 4.2(b). Next, the triangular spectral weighting method was used to reduce the ripple effect in the resulting image. The resulting normalized

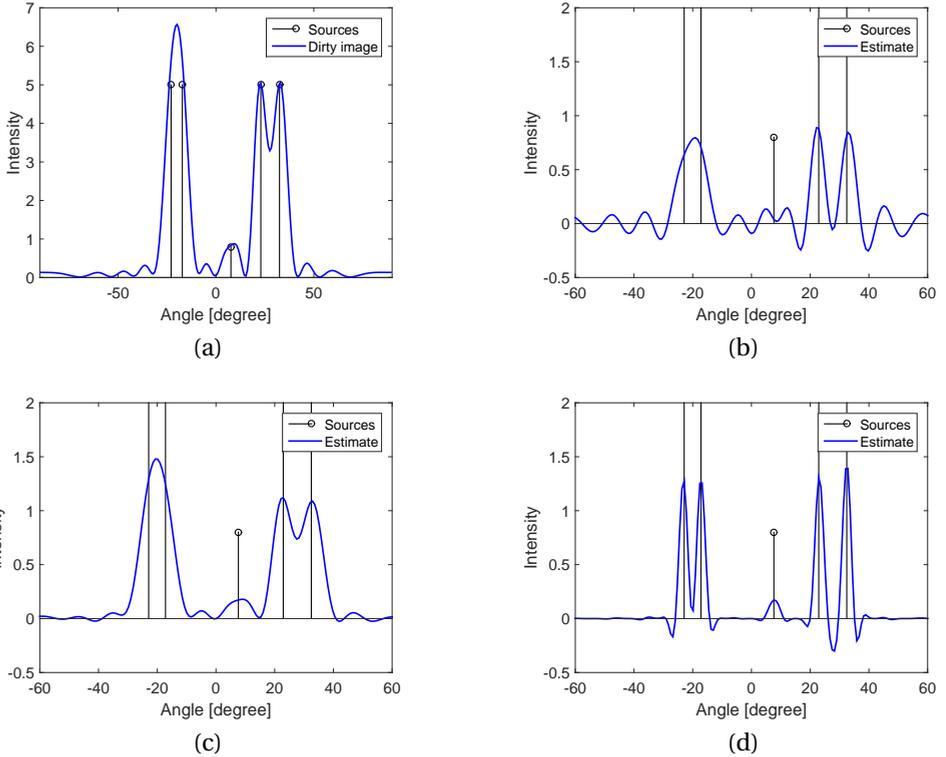


Figure 4.1: Simulation results for (a) dirty image, and point source estimates based on (b) truncated eigenvalue decomposition, (c) triangular-weighted eigenvalue decomposition, (d) dirty image based conditioning.

reconstructed image is shown in Figure 4.2(c). Moreover, the normalized reconstructed image using the dirty image based conditioning method is shown in Figure 4.2(d). As the figure suggests, conditioning based on the computed dirty image results in a sharper image with less sidelobes and higher dynamic range as compared with the KLT-based method.

We focus on the image of Cassiopeia A supernova remnant and Cygnus A radio galaxy seen as two point-like emissions on the right edge of the image. These two objects can be considered as discontinuities in the otherwise rather smooth image. These ringing effects are due to the Gibbs phenomena. Comparing Figure 4.2(a) and (b) we can see that the definition of the position of these discontinuities is more precise in 4.2(b) but this comes at a cost of distortion by “ringing” around the discontinuities. As mentioned in Section 3.1.2, we expect the discontinuities to result in ringing effects in restoration by TSVD as can be seen in Figure 4.2(b) due to the Gibbs phenomena. By applying smooth weights in Figure 4.2(c) we trade resolution for ringing. On the other hand, in Figure 4.2(d), the dynamic range and the sharpness of the image is increased however the ringing effects are still observable. We can see that the combined regularizing effect

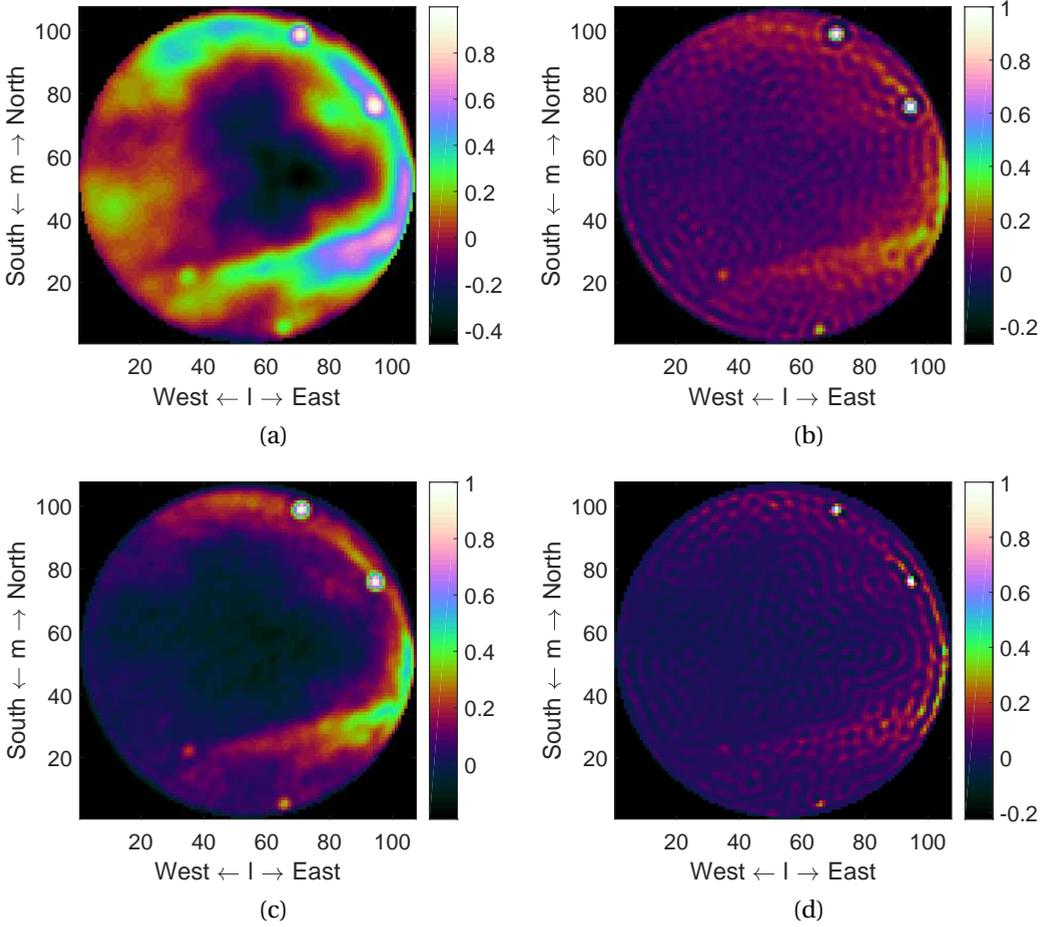


Figure 4.2: Image estimates based on (a) MF dirty image (b) truncated eigenvalue decomposition, (c) triangular-weighted eigenvalue decomposition, (d) dirty image based conditioning.

of TSVD and weighting by the dirty image helps in increasing the sharpness of point-like sources but the recovery for extended emissions are not satisfactory.

4.6. CONCLUSIONS

Image deconvolution for radio astronomy is a highly ill-posed problem. In this chapter, two regularization methods are proposed to control stability of the solution. The first method uses a weighting function to control the dominance of the available spatial frequencies in the data and the second uses the prior information on the dirty image to condition the spatial distribution of the celestial sources. The proposed regularization methods are validated by actual data from the LOFAR radio telescope and through simulations. We can see that by applying windowing, we trade sharpness of the image and ringing effects with resolution. Furthermore, we see that applying the dirty image conditioning results in a higher dynamic range and sharpness of the image. We would like to note that, due to the considerable memory and computational demands of the SVD, solving large-scale radio interferometric problems using the direct method is not feasible.

In the next chapter, we derive a similar data-driven regularization method starting from a Bayesian RA imaging formulation. Estimating the covariance of the image from an MVDR dirty image, we show that we can obtain a regularized MLE formulation such that the system matrix \mathbf{M} is right-preconditioned with a dirty image. This provides a means for regularization. Applying the weights directly on \mathbf{M} rather than \mathbf{H} eliminates the adverse effect of one-sided preconditioning of \mathbf{H} . We further generalize the framework to account for sparsity-promoting regularizations and propose efficient iterative methods to obtain the solution of the regularized problem.

5

FAST IMAGING WITH GENERAL REGULARIZATION

5.1. INTRODUCTION

The advent and development of increasingly large radio interferometers such as LOFAR [1] and the SKA [4] has sparked renewed interest in the image formation task. More precisely, image formation is expected to be the main computational bottleneck in the processing pipeline of next generation radio telescopes [26]. The main issue in image reconstruction in the context of next generation radio telescopes is the huge scale of the problem both in terms of the size of the measurements as well as the size of the image due to the scientific requirements. Owing to the considerable memory and computational demands, solving large-scale interferometric problems using the direct method, as discussed in Chapter 4, is not feasible. More precisely, the SVD requires storage of the measurement matrix \mathbf{M} and computations of order cube of the problem size which is infeasible on the scales of SKA. As an alternative, custom-made iterative methods should be applied.

We have introduced the radio interferometric imaging measurement equation in Chapter 2 and highlighted its ill-posedness and the need for regularization together with some of the most commonly used regularization methods in RA in Chapter 3. Different iterative methods not only differ on the optimization problem that they tackle but also on the algorithm used for finding the solution of the optimization problem. Some iterative solution methods, such as the conjugate gradient method, exhibit a property such that they can provide for another form of regularization by prematurely stop the iterations, restricting the solution to a particular data-dependent subspace.

Part of this chapter is published as S. Naghibzadeh and A.J. van der Veen. PRIFIRA: General regularization using prior-conditioning for fast radio interferometric imaging. Monthly Notices of Royal Astronomical Society. June 2018

Considerations in algorithm selection are (1) the accuracy (fidelity) of the resulting image, related to the definition of the optimization problem, (2) computational complexity, related to the scalable solution of the optimization problem, and (3) the automation and flexibility of the process regarding the selection of unknown parameters or settings such as iteration counts. For future radio telescopes, not all measurement data can be stored and image formation has to be done in an automated and quasi-real time manner. One of the main advantages of the iterative methods is that there is no need to store and invert \mathbf{M} . Instead, the action of the linear operator on a vector in the form of matrix-vector multiplications with \mathbf{M} and \mathbf{M}^H must be applied. These multiplications can be implemented as operators which eliminate the need for construction and storage of the matrix \mathbf{M} and allows for the implementation of fast methods based on the structure of the measurement matrix. Furthermore, the cost of the iterative methods is determined based on the amount of computations per iteration times the total number of iterations needed for a good restoration of the image [96, 84].

This chapter is a quest to introduce efficient algebraic techniques for radio interferometric imaging with the aim to tackle some of the deficiencies in the imaging pipeline of the next generation radio telescopes. A wealth of research has been dedicated to the recovery of discrete emissions from distant galaxies, with sparsity as the main regularizing assumption. In this chapter we focus on the recovery of extended emissions. We present a general algorithmic framework based on a Bayesian-inspired regularized maximum likelihood formulation of the radio astronomical imaging problem with a focus on diffuse emission recovery from limited noisy correlation data. The algorithm is dubbed PRIor-conditioned Fast Iterative Radio Astronomy (PRIFIRA) and is based on a direct embodiment of the regularization operator into the system by right preconditioning. The resulting system is then solved using an iterative method based on projections onto Krylov subspaces. We motivate the use of a beamformed image (which includes the classical “dirty image”) as an efficient prior-conditioner. Iterative reweighting schemes generalize the algorithmic framework and can account for different regularization operators that encourage sparsity of the solution. The performance of the proposed method is evaluated based on simulated one- and two-dimensional array arrangements as well as actual data from the core stations of the Low Frequency Array radio telescope antenna configuration, and compared to state-of-the-art imaging techniques. The simulation and experimental results of PRIFIRA are included in Chapter 6. We show the generality of the proposed method in terms of regularization schemes while maintaining a competitive reconstruction quality with the current reconstruction techniques. Furthermore, we show that exploiting Krylov subspace methods together with the proper noise-based stopping criteria results in a great improvement in imaging efficiency.

5.1.1. STATE OF THE ART IMAGING ALGORITHMS

Classical radio astronomical imaging algorithms are based on the CLEAN algorithm [3, 68] and its multiresolution and multiscale variants [69, 70, 97, 71]. The considered cost function is the LS objective, implicitly regularized by an ℓ_0 constraint [72] which favors maximal sparsity of the solution. The CLEAN algorithm was recently interpreted as a gradient descent method combined with a “greedy” procedure to find the support of the image [75]. We will explain this in more detail in Chapter 7.

Alternatively, the problem can be regularized by posing a non-negativity constraint on the solution [42]. The resulting Nonnegative Least Squares (NNLS) optimization can be implemented using the active set method [56] and similarly consists of two levels of iterations: (i) an outer loop to iteratively find the sparse support of the image and (ii) an inner loop in which a dimension reduced version of the LS problem is solved. The increased resolution, bandwidth, sensitivity and sky coverage of these instruments result in many more sources, including unresolved sources and extended structures, rendering the traditional greedy imaging algorithms based on point source detection, such as CLEAN and NNLS, iterations less effective.

Another classical RA imaging algorithm is the maximum entropy method (MEM) [77, 78]. The regularization term is the entropy function $\sigma^T \log(\sigma)$, and the problem is solved using computationally expensive nonlinear optimization methods such as Newton-Raphson.

Finding the nonzero support of an image using an ℓ_0 constraint is an NP-complete problem. Instead, this constraint may be weakened to an ℓ_1 constraint, which still promotes sparsity of the solution, but admits a solution based on the theory of compressed sensing and convex optimization, for which efficient techniques exist. Recently, many algorithms in this direction have been proposed [34, 35, 36, 75, 98, 99, 100]. These methods are based on a gradient descent approach. Instead of a constraint on $\|\sigma\|_1$ (sparse image), also a more general constraint $\|\Psi^T \sigma\|_1$ or $\|\alpha\|_1$ where $\sigma = \Psi \alpha$ can be used, in which Ψ is an overcomplete dictionary of orthonormal bases. E.g., Sparsity Averaging Reweighted Analysis (SARA) [36] employs a concatenation of wavelet dictionaries. The advantage of the methods based on convex optimization is the simplicity of imposing additional constraints on the solution, the existence of many well-developed methods with guaranteed convergence and the ability to split the work into simpler, parallelizable sub-problems [101, 75]. The disadvantage of these algorithms is that the gradient descent steps make the algorithm convergence rather slow. Also, as remarked in [71], many of these algorithms have not yet been tested on real data.

Taking another direction, the RESOLVE algorithm introduced techniques from Bayesian statistics to propose priors that regularize the solution [80], aimed specifically at extended sources, and modeled these a priori using log-normal distributions. Unfortunately, the resulting method appears to be extremely slow [80].

5.1.2. RESULTS

In this chapter, our interest is in developing a new method for science cases where a considerable amount of complex diffuse emissions is present such as in the studies of galactic magnetism, the epoch of reionization, and polarized imaging.

We start from a Bayesian statistical approach for regularization, but formulate a shortcut that immediately connects to a numerical method called prior-conditioning, i.e., a data-dependent Jacobi-like right preconditioner that scales the columns of \mathbf{M} . In this general framework, the prior conditioner can take the form of a beamformed image, such as the classical dirty image, or the MVDR image, or any other low resolution prior image that is strictly positive on the true support of the image. This could also be determined iteratively, which gives a connection to reweighted least squares solutions, often used to approximate ℓ_0 or ℓ_1 norm optimization by least squares optimization, in par-

ticular the FOCUSS algorithm [102] and the algorithm presented in [103].

Next, we propose to solve the obtained regularized LS problem by a fast and efficient iterative algorithm based on the Krylov subspace-based method of LSQR [104]. Krylov methods often exhibit a faster convergence than methods based on gradient descent [105]. Therefore, they appear to be good candidates as alternative iterative solution methods for the RA imaging problem. The stopping criterion of the LSQR algorithm is based on the norm of the residual, which provides another form of regularization, called iterative regularization or semiconvergence [84, 59]. The resulting algorithm is straightforward to implement and computationally very efficient.

We compare the proposed method to classical RA imaging methods as well as methods based on convex optimization both in terms of speed and quality of the estimate. It will be seen that the proposed method is accurate and converges extremely fast (around 10 iterations).

5.2. PROPOSED SOLUTION METHOD

We consider the Bayesian problem formulation framework from Chapter 3, Section 3.2.2. The main question in the Bayesian framework is the selection of a suitable prior. E.g., we can select a Gaussian prior where $\boldsymbol{\mu}_\sigma$ is the currently best known estimate for the image (the current sky map), with \mathbf{C}_σ related to the accuracy of that knowledge. As it is hard to quantify this, \mathbf{C}_σ could be modeled as a diagonal matrix, with the unknown variances on the diagonal modeled in turn as statistical parameters, for which a distribution (with unknown parameters called hyperparameters) has to be proposed. The estimation of these hyperpriors from the data is known as Sparse Bayesian Learning (SBL) [81] and in the context of our problem has been worked out by [82]. The RESOLVE method [80] follows a similar approach. Unfortunately, the computational complexity is reported to be rather high.

Since the prior in this framework is data-dependent, the question at this point is whether it would be possible to use a (perhaps less optimal) data-dependent prior that is easier to estimate.

5.2.1. PROBLEM REFORMULATION

We focus on the Tikhonov regularized WLS problem formulation and will use $\boldsymbol{\mu}_\sigma = \mathbf{0}$ and restrict \mathbf{L} to be diagonal. Our aim is thus to propose a suitable \mathbf{L} . Since $\mathbf{C}_\sigma^{-1} = \mathbf{L}^H \mathbf{L}$, the diagonal entries of \mathbf{L} model the *precision* of our prior knowledge, and a large entry of \mathbf{L} will result in a dark pixel (since $\boldsymbol{\mu}_\sigma = \mathbf{0}$), whereas a small entry of \mathbf{L} will make that pixel be determined by the data.

With change of variables $\boldsymbol{\alpha} = \mathbf{L}\boldsymbol{\sigma}$, we can rewrite the objective function (3.14) in terms of $\boldsymbol{\alpha}$ as

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\Gamma(\bar{\mathbf{r}} - \mathbf{M}\mathbf{L}^{-1}\boldsymbol{\alpha})\|_2^2 + \tau \|\boldsymbol{\alpha}\|_2^2, \quad (5.1)$$

The image can be recovered from $\hat{\boldsymbol{\alpha}}$ by the linear transform $\hat{\boldsymbol{\sigma}} = \mathbf{L}^{-1}\hat{\boldsymbol{\alpha}}$.

Equation (5.1) is equivalent to the solution of

$$(\mathbf{L}^{-H}\mathbf{M}^H\mathbf{C}_e^{-1}\mathbf{M}\mathbf{L}^{-1} + \tau\mathbf{I})\boldsymbol{\alpha} = \mathbf{L}^{-H}\mathbf{M}^H\mathbf{C}_e^{-1}\bar{\mathbf{r}}. \quad (5.2)$$

With the choice of $\mathbf{C}_\sigma^{-1} = \mathbf{L}^H \mathbf{L}$ and $\mathbf{C}_e^{-1} = \mathbf{\Gamma}^H \mathbf{\Gamma}$ and change of variables $\tilde{\mathbf{M}} = \mathbf{\Gamma} \mathbf{M} \mathbf{L}^{-1}$ and $\tilde{\mathbf{r}} = \mathbf{\Gamma} \tilde{\mathbf{r}}$ we can rewrite this as

$$(\tilde{\mathbf{M}}^H \tilde{\mathbf{M}} + \tau \mathbf{I}) \boldsymbol{\alpha} = \tilde{\mathbf{M}}^H \tilde{\mathbf{r}}. \quad (5.3)$$

Such a scaling of the columns of \mathbf{M} by a matrix \mathbf{L}^{-1} related to the prior distribution is known as prior-conditioning [83], as it is similar in shape to the preconditioning that is sometimes done in iterative solvers to improve convergence. The difference is that preconditioning only involves \mathbf{M} whereas prior-conditioning is not just based on \mathbf{M} but on the interaction of \mathbf{M} with the data $\tilde{\mathbf{r}}$.

5.2.2. SELECTION OF THE PRIOR COVARIANCE

To obtain a prior, data-dependent, estimate of \mathbf{C}_σ , the idea is to compute from the data an unbiased estimate for the image, using minimal assumptions, i.e., we consider $\boldsymbol{\sigma}$ deterministic. The variance of this estimate can then be used as estimate for \mathbf{C}_σ .

The best possible estimate under this assumption is the MLE estimate, in this case equal to the WLS estimate

$$\hat{\boldsymbol{\sigma}}_{\text{MLE}} = (\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M})^{-1} \mathbf{M}^H \mathbf{C}_e^{-1} \tilde{\mathbf{r}}. \quad (5.4)$$

It is known that this estimator is an efficient MVU estimator [57] with covariance

$$\mathbf{C}_{\hat{\boldsymbol{\sigma}}} = (\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M})^{-1}, \quad (5.5)$$

where $\mathbf{C}_e = \frac{1}{N} (\mathbf{R}^T \otimes \mathbf{R})$. Therefore

$$\begin{aligned} \mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M} &= N (\mathbf{A}^* \circ \mathbf{A})^H (\mathbf{R}^{-T} \otimes \mathbf{R}^{-1}) (\mathbf{A}^* \circ \mathbf{A}) \\ &= N (\mathbf{A}^T \mathbf{R}^{-T} \mathbf{A}^*) \circ (\mathbf{A}^H \mathbf{R}^{-1} \mathbf{A}). \end{aligned} \quad (5.6)$$

If we denote the variance of $\hat{\boldsymbol{\sigma}}$ as $\text{Var}(\hat{\boldsymbol{\sigma}})$, it consists of the diagonal elements of $\mathbf{C}_{\hat{\boldsymbol{\sigma}}}$, i.e.

$$\text{Var}(\hat{\boldsymbol{\sigma}}) = \text{vectdiag}(\mathbf{C}_{\hat{\boldsymbol{\sigma}}}). \quad (5.7)$$

Based on Equation (5.6), the i th diagonal element of $\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M}$ can be computed as $N (\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i)^2$. Although Equation (5.6) shows that the estimated pixel intensities are correlated, we ignore that and set $\mathbf{C}_{\hat{\boldsymbol{\sigma}}} \approx \text{diag}(\text{Var}(\hat{\boldsymbol{\sigma}}))$ where

$$\text{Var}(\hat{\sigma}_i) = \frac{1}{N (\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i)^2}, \quad i = 1, 2, \dots, Q, \quad (5.8)$$

with $\text{Var}(\hat{\sigma}_i)$ denoting the variance of the i th pixel estimate. Comparing (2.61) and (5.8) we conclude that (if \mathbf{R}_n is ignored in (2.61))

$$\mathbf{C}_{\hat{\boldsymbol{\sigma}}} \approx \text{diag}(\text{Var}(\hat{\boldsymbol{\sigma}})) = \frac{1}{N} \text{diag}(\boldsymbol{\sigma}_{\text{MVDR}})^2. \quad (5.9)$$

Since the true data covariance matrix is not available, we will use the sample covariance matrix $\hat{\mathbf{R}}$, and obtain the estimated MVDR image $\hat{\boldsymbol{\sigma}}_{\text{MVDR}}$. This leads to the choice to set

$$\mathbf{L}^{-1} = \text{diag}(\hat{\boldsymbol{\sigma}}_{\text{MVDR}}) \quad (5.10)$$

as regularizing operator. (A factor \sqrt{N} is absorbed in τ .) Subsequently, we set $\bar{\mathbf{M}} = \mathbf{\Gamma ML}^{-1}$ and solve Equation (5.3).

While this choice is obviously a shortcut from a truly Bayesian approach (e.g., the mean value of the initial image is ignored and only the variance is taken into account), we will show in the simulations that this simple idea is very effective in obtaining regularized solutions. Moreover, it is computationally not very involved as it amounts to constructing a beamformed image (similar to computing the classical dirty image), followed by solving Equation (5.3). We propose to use Krylov subspace iterations to do this efficiently (Section 5.3.1).

5.2.3. DISCUSSION AND GENERALIZATIONS

Before we develop an efficient algorithm for finding the solution of the problem stated in Section 5.2.1, we discuss some of the properties of the problem and address potential generalizations of the framework.

1) RA images contain substantial black background of radio quiet zones. The working principle of greedy algorithms such as CLEAN and NNLS is to first obtain the support of the image, also called the active set, and to solve only for the elements of the image in the active set. Therefore, as shown by [72], these methods solve the regularized LS or MLE problem (3.14) with $\mathcal{R}(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}\|_0$,

$$\hat{\boldsymbol{\sigma}} = \arg \min_{\boldsymbol{\sigma}} \|\mathbf{\Gamma}(\bar{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau \|\boldsymbol{\sigma}\|_0, \quad (5.11)$$

with the addition of a nonnegativity constraint for NNLS. Minimizing the ℓ_0 norm produces satisfactory results both in terms of the support of the image and the intensity estimates if the underlying image is sufficiently sparse and only consists of scattered point sources.

In line with our problem formulation, the ℓ_0 constraint can be translated into a right preconditioner. If we assume for the moment the knowledge of the true $\boldsymbol{\sigma}$, denoted as $\boldsymbol{\sigma}_{\text{true}}$, we can define a diagonal matrix \mathbf{D} as

$$[\mathbf{D}]_{i,i} = \begin{cases} 1, & \text{if } [\boldsymbol{\sigma}_{\text{true}}]_i > 0 \\ 0, & \text{if } [\boldsymbol{\sigma}_{\text{true}}]_i = 0. \end{cases} \quad (5.12)$$

Therefore, in terms of the LS formulation, we need to solve the problem

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \|\mathbf{\Gamma}(\bar{\mathbf{r}} - \mathbf{MD}\boldsymbol{\alpha})\|_2^2 + \tau \|\boldsymbol{\alpha}\|_2^2, \quad (5.13)$$

where the image estimate is found by the transform $\hat{\boldsymbol{\sigma}} = \mathbf{D}\hat{\boldsymbol{\alpha}}$. Thus, $\hat{\boldsymbol{\sigma}}$ will be zero where $\mathbf{D}_{i,i}$ is zero, and \mathbf{D} would be the optimal prior conditioner. In reality we do not know $\boldsymbol{\sigma}_{\text{true}}$. Finding the active set in greedy algorithms is done iteratively through outer iterations. This increases the cost of the algorithms substantially. Clearly, Problem (5.13) is connected to Problem (5.1) considered in Section 5.2.1 via $\mathbf{D} = \mathbf{L}^{-1}$. In this context, our use of a beamformed image $\mathbf{D} = \text{diag}(\hat{\boldsymbol{\sigma}}_{\text{MVDR}})$ or $\mathbf{D} = \text{diag}(\hat{\boldsymbol{\sigma}}_{\text{MF}})$ can be interpreted as a surrogate for this.

2) Low resolution initial estimates of the image can be obtained via MF or MVDR beamforming. Previously, we suggested in [106] and [107] to use the MF dirty image

$\text{diag}(\hat{\boldsymbol{\sigma}}_{\text{MF}})$ for regularization purposes. Moreover, we showed the relation to Bayesian estimation when applying MVDR-based right prior-conditioning weights in [108]. If the noise is lower or comparable to the signal, we have the Relation (2.62),

$$\mathbf{0} \leq \boldsymbol{\sigma}_{\text{true}} \leq \boldsymbol{\sigma}_{\text{MVDR}} \leq \boldsymbol{\sigma}_{\text{MF}}. \quad (5.14)$$

Therefore, the prior variance $\mathbf{C}_{\boldsymbol{\sigma}}$ based on the MF dirty image is higher than when the MVDR dirty image is used, and the latter provides a better start. The correction by $\mathbf{R}_{\mathbf{n}}$ introduced in Equation (2.61) moves the MVDR image even further towards the true image. It is also important to note from this relation that the true image is black wherever the initial image is black. This way the initial estimate provides a rough estimate for the true support of the image.

If we do not know the autocorrelations in the measurement data, we can use the MF estimate without the diagonals or the MVDR image obtained by diagonal loading. However, we must take care that all brightness estimates are strictly positive by adding a constant value to the image since negative weights will be completely wrong while zero weights will result in pixels that will stay black throughout the iterations and in the final solution. We note that without autocorrelation information the results will be sub-optimal. Section 5.5 gives a brief analysis and provides additional remarks related to the proposed imaging techniques.

3) We show that applying MF or MVDR dirty images as prior conditioners favors smooth reconstructions and is therefore more interesting for the recovery of diffuse structures and smooth features of the sky map rather than point sources. We motivate our claim for prior-conditioning with the MF Analysis for MVDR-based prior-conditioning would be similar.

Assuming for the moment that there is no noise and error on the covariance measurements, i.e. $\mathbf{e} = \mathbf{0}$ and $\mathbf{r}_{\mathbf{n}} = \mathbf{0}$, based on Equation (2.58) we can write the true dirty image as

$$\boldsymbol{\sigma}_{\text{MF}} = \mathbf{M}^H(\mathbf{M}\boldsymbol{\sigma}). \quad (5.15)$$

If we consider an image that only contains a unit norm point source in the middle of the FoV, we can rewrite Equation (5.15) as

$$\mathbf{b} = (\mathbf{M}^H\mathbf{M})\mathbf{e}_{\text{mid}}, \quad (5.16)$$

where \mathbf{e}_{mid} is the unit vector with the element in the middle of the FoV equal to 1. In this case the dirty image is called the dirty beam, indicated by \mathbf{b} . The dirty beam is also known as Point Spread Function (PSF) (impulse response of the imaging system). If we insert an arbitrary image $\boldsymbol{\sigma}$ in the FoV of the array, resulting $\boldsymbol{\sigma}_{\text{MF}}$ would be the convolution of dirty beam with the image. Dirty beam by construction acts as a low-pass filter with the main beam corresponding to the resolution of the array. Therefore, the resulting dirty image will be a low-pass filtered version of the sky map and is smooth.

When we use the dirty image as a prior, the smoothness will be preserved in the resulting image from Equation (5.1). Since the extended emissions exhibit a smooth structure, in the reconstruction they will be preserved. By the same token, isolated point sources will not be imaged sharper than the resolution of the instrument and will be

spread out. This spreading is similar to the post-processing applied to the CLEAN solution to restore the natural resolution of the telescope array. In our method, since the prior obeys the resolution of the array, spreading is done automatically.

By relating Equation (5.1) to the effect of right preconditioning explained in Chapter 3, when $\mathbf{L}^{-1} = \text{diag}(\hat{\boldsymbol{\sigma}}_{\text{MF}})$ we find another interpretation. We have seen that right preconditioning implies modeling of the image. In this case, the image $\boldsymbol{\sigma}$ is confined to the model $\boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}}_{\text{MF}} \circ \boldsymbol{\alpha}$. This reduces the degrees of freedom for $\boldsymbol{\sigma}$ to a smooth model assigned by $\hat{\boldsymbol{\sigma}}_{\text{MF}}$. Therefore, the solution by definition favors smooth reconstructions. We note that confining $\boldsymbol{\sigma}$ by $\hat{\boldsymbol{\sigma}}_{\text{MF}} \circ \boldsymbol{\alpha}$ is not sufficient to obtain a unique solution. The additional regularization term $\|\boldsymbol{\alpha}\|_2^2$ also constrains the norm of free parameter $\boldsymbol{\alpha}$ such that a unique estimate is obtained.

4) Next, we show that applying the regularization operator as a prior-conditioner opens the door to various regularizations. This can be applied in cases where the underlying sky map contains isolated point sources.

It is well-known that ℓ_1 constraints result in sparse solutions. The associated regularized MLE problem is

$$\hat{\boldsymbol{\sigma}} = \arg \min_{\boldsymbol{\sigma}} \|\Gamma(\bar{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau \|\boldsymbol{\sigma}\|_1. \quad (5.17)$$

One way to solve Equation (5.17) is via the Iteratively Reweighted Least Squares method [103]. The ℓ_1 constraint is transformed to an ℓ_2 constraint by

$$\begin{aligned} \|\boldsymbol{\sigma}\|_1 &= \sum_{i=1}^Q |\sigma_i| = \sum_{i=1}^Q \frac{|\sigma_i|^2}{|\sigma_i|} \\ &= \|\mathbf{W}\boldsymbol{\sigma}\|_2^2, \quad \text{where } \mathbf{W} = \text{diag}(\boldsymbol{\sigma}^{-1/2}). \end{aligned} \quad (5.18)$$

Equation (5.18) suggests that $\|\boldsymbol{\sigma}\|_1$ can be computed from a properly weighted ℓ_2 -norm. Although this optimal weight is unknown, we can enter an iteration wherein, at each step, the weight is based on the solution obtained at the previous step. It is thus sufficient to solve only weighted LS problems. Specifically, we define the weight matrix at iteration k as $\mathbf{W}_k = \text{diag}(\hat{\boldsymbol{\sigma}}_{k-1}^{-1/2})$ where $\hat{\boldsymbol{\sigma}}_{k-1}$ is the solution obtained at the previous iteration $k-1$ ¹. Therefore, Equation (5.17) is replaced by

$$\hat{\boldsymbol{\sigma}}_k = \arg \min_{\boldsymbol{\sigma}} \|\Gamma(\bar{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau \|\mathbf{W}_k \boldsymbol{\sigma}\|_2^2 \quad (5.19)$$

which can be transformed into a right preconditioned system using the transform $\boldsymbol{\alpha} = \mathbf{W}_k \boldsymbol{\sigma}$,

$$\hat{\boldsymbol{\alpha}}_k = \arg \min_{\boldsymbol{\alpha}} \|\Gamma(\bar{\mathbf{r}} - \mathbf{M}\mathbf{W}_k^{-1} \boldsymbol{\alpha})\|_2^2 + \tau \|\boldsymbol{\alpha}\|_2^2. \quad (5.20)$$

After the estimate $\hat{\boldsymbol{\sigma}}_k$ is obtained, Problem (5.20) is solved again with the new weights. Therefore, this method requires solving Problem (5.20) multiple times where the outer iterations are indicated by k . Comparing to Problem (5.1), we see that Problem (5.20) is a prior-conditioned problem where the prior is iterated upon as more accurate images are being computed.

¹A modification may be required by adding a small ε to the weights, i.e. $\mathbf{W}_k^{-1} = \text{diag}(\hat{\boldsymbol{\sigma}}_{k-1}^{1/2} + \varepsilon)$ for stability reasons. Similar modifications are proposed in [109] and [103].

If we start the iteration with $\mathbf{W}_0 = \mathbf{I}$, then the first estimate $\hat{\boldsymbol{\sigma}}_1$ will be the MLE estimate (presented in Equation (2.69)). The next iteration will solve a right preconditioned system where the square-root of this image is the prior.

In contrast, the prior we proposed in Equation (5.10) uses the MVDR image and omits the square-root. Nonetheless, it is interesting to consider what happens if we iterate this estimate in the same way as (5.20), but with $\mathbf{W}_k = \text{diag}(|\hat{\boldsymbol{\sigma}}_{k-1}|^{-1})$. If this converges to a fixed point, the corresponding constraint is

$$\|\mathbf{W}\boldsymbol{\sigma}\|_2^2 = \sum_{i=1, \sigma_i \neq 0}^Q \frac{|\sigma_i|^2}{|\sigma_i|^2} = \|\boldsymbol{\sigma}\|_0. \quad (5.21)$$

This shows that iteratively minimizing (5.20) in this way is a surrogate for using the ℓ_0 norm as regularizer, and will result in a very sparse image (even if the true image is not sparse). We show this effect with simulations in a one dimensional setting.

Iterations of this form have been proposed in the context of our problem by [102], and are known as the FOCUSS algorithm. As mentioned in that paper, sparsity by itself does not form a sufficient constraint to obtain a unique estimate for an underdetermined problem, and the use of a low-resolution initial estimate provides the necessary additional constraint. The paper proves the quadratic convergence to a local fixed point in the neighborhood of the initialization, and also mentions a technique to impose a positivity constraint on the solution. Unfortunately, the proposed solution method is based on the truncated Singular Value Decomposition (SVD) and is not applicable for large-scale problems.

Overall, the regularization penalty $\|\boldsymbol{\sigma}\|_0$ assumes the image is composed of a set of separate point sources. On the other hand, the $\|\boldsymbol{\sigma}\|_2$ penalty favors solutions with similar intensity levels over different pixels to minimize the overall power and is suitable for the recovery of diffuse emissions. The $\|\boldsymbol{\sigma}\|_1$ penalty is intermediate between smoothness and sparsity penalties but is not specifically designed for extended or point sources. It is known that if the ℓ_0 -constrained problem (5.11) contains a sufficiently sparse solution, the ℓ_1 -constrained problem (5.17) as a surrogate for (5.11) will recover it [110]. However, in cases where both resolved and unresolved sources co-exist in the sky map, (5.17) recovers both types of sources but it is not optimal for any of them. In Section 7.4 we discuss some attempts to generalize the formulation such that both resolved and unresolved sources can be recovered. This is done by means of introducing overcomplete dictionaries. Section 7.4 shows the capability of the proposed framework for generalization.

We see that by applying outer iterations, based on the prior-conditioning formulation, we are able to also impose sparsity-promoting norms into the framework of Section 5.2.1. Doing so, we are able to benefit from the efficient algebraic algorithms that exist for solving the ℓ_2 regularized problem to also recover images with sparsity priors. In the next section, we present an efficient algorithmic framework to solve the prior-conditioned problem.

5.3. THE PRIFIRA ALGORITHM

The proposed solution method from Section 5.2 is now further worked out into an algorithm which we call PRIFIRA.

5.3.1. IMPLEMENTATION USING KRYLOV SUBSPACE METHODS

We solve Problem (5.1) by an iterative method based on projections onto Krylov subspaces. Let $\bar{\mathbf{M}} = \mathbf{\Gamma ML}^{-1}$ and $\bar{\mathbf{r}} = \mathbf{\Gamma r}$, then (5.1) is written as

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \|\bar{\mathbf{r}} - \bar{\mathbf{M}}\boldsymbol{\alpha}\|_2^2 + \tau \|\boldsymbol{\alpha}\|_2^2. \quad (5.22)$$

Define the t -dimensional Krylov subspace \mathcal{K}_t , for $t = 1, 2, \dots$, as

$$\begin{aligned} \mathcal{K}_t(\bar{\mathbf{M}}^H \bar{\mathbf{M}}, \bar{\mathbf{M}}^H \bar{\mathbf{r}}) \\ = \text{span}\{\bar{\mathbf{M}}^H \bar{\mathbf{r}}, (\bar{\mathbf{M}}^H \bar{\mathbf{M}})\bar{\mathbf{M}}^H \bar{\mathbf{r}}, \dots, (\bar{\mathbf{M}}^H \bar{\mathbf{M}})^{t-1} \bar{\mathbf{M}}^H \bar{\mathbf{r}}\}. \end{aligned} \quad (5.23)$$

Krylov subspace methods instead solve the problem

$$\begin{aligned} \hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \|\bar{\mathbf{r}} - \bar{\mathbf{M}}\boldsymbol{\alpha}\|_2^2 \\ \text{subject to } \boldsymbol{\alpha} \in \mathcal{K}_t(\bar{\mathbf{M}}^H \bar{\mathbf{M}}, \bar{\mathbf{M}}^H \bar{\mathbf{r}}) \end{aligned} \quad (5.24)$$

for $t = 1, 2, \dots$. As the iteration count t increases, the Krylov subspace gradually increases in dimension as well, so that the residual $\|\bar{\mathbf{r}} - \bar{\mathbf{M}}\hat{\boldsymbol{\alpha}}_t\|_2^2$ decreases while $\|\boldsymbol{\alpha}_t\|_2^2$ usually increases.

Due to the ill-posedness of the problem, $\|\boldsymbol{\alpha}_t\|_2^2$ will grow out of bound as the iteration progresses [111]. One way to stop the iterations while the solution is still numerically stable is via the discrepancy principle [111], [112]. In this case, the iteration is stopped at iteration T once $\|\bar{\mathbf{r}} - \bar{\mathbf{M}}\boldsymbol{\alpha}_T\|_2^2 \leq \epsilon$, which then gives an approximate solution to (5.22).

The restriction to the Krylov subspace before it spans the complete space provides a regularization, called semiconvergence [59]. If the iteration is allowed to continue, then the residual converges to zero and the solution converges to the pseudo-inverse minimum norm solution (2.69), so that we obtain the unregularized solution. In contrast to Tikhonov regularization or truncated SVD where the regularization only depends on \mathbf{M} and not on the measured data, the regularization provided by Krylov subspace methods adapts to the data via the initial vector $\bar{\mathbf{r}}$. While problem (5.22) is not exactly equivalent to (5.24), their solutions are considered very similar [113].

Krylov subspace methods are attractive because we do not need to store $\bar{\mathbf{M}}$, rather we need to provide functions that return matrix-vector products of the form $\bar{\mathbf{M}}\mathbf{u}$ and $\bar{\mathbf{M}}^H \mathbf{v}$ [105]. With the functional form of \mathbf{M} as given in (2.26), we can implement such a subroutine. This greatly reduces storage requirements. Related details are in Section 5.3.4.

To solve (5.24) iteratively for a non-square $\bar{\mathbf{M}}$ with arbitrary rank, the LSQR method [104] is appropriate [114]. This is analytically equivalent to the Conjugate Gradient method applied to the normal equations, but is numerically preferred [111]. The LSQR method is based on the Golub-Kahan (GK) bidiagonalization algorithm [115], also referred to as

Lanczos iterations. First define

$$\begin{aligned}\beta_1 &= \|\bar{\mathbf{r}}\|, & \mathbf{u}_1 &= \bar{\mathbf{r}}/\beta_1, \\ \alpha_1 &= \|\bar{\mathbf{M}}^H \mathbf{u}_1\|, & \mathbf{v}_1 &= \bar{\mathbf{M}}^H \mathbf{u}_1/\alpha_1.\end{aligned}$$

Using these as initialization, in the GK process, at iteration t , two orthonormal vectors \mathbf{v}_t and \mathbf{u}_t are computed as

$$\begin{aligned}\beta_t \mathbf{u}_t &:= \bar{\mathbf{M}} \mathbf{v}_{t-1} - \alpha_{t-1} \mathbf{u}_{t-1}, \\ \alpha_t \mathbf{v}_t &:= \bar{\mathbf{M}}^H \mathbf{u}_t - \beta_t \mathbf{v}_{t-1}\end{aligned}\tag{5.25}$$

where β_t and α_t are chosen such that \mathbf{u}_t and \mathbf{v}_t are normalized. Let

$$\mathbf{B}_t = \begin{bmatrix} \alpha_1 & & & & & \\ \beta_2 & \alpha_2 & & & & \\ & \beta_3 & \ddots & & & \\ & & \ddots & \alpha_t & & \\ & & & & \beta_{t+1} & \end{bmatrix}.\tag{5.26}$$

It can then be shown that solving (5.24) reduces to solving the bidiagonalized LS problem

$$\min_{\mathbf{y}_t} \|\mathbf{B}_t \mathbf{y}_t - \beta_1 \mathbf{e}_1\|_2^2,\tag{5.27}$$

where \mathbf{e}_1 is the unit vector with its first element equal to one. LSQR uses QR updates to obtain \mathbf{y}_t at each iteration t [104].

The complete algorithm to solve (5.24) and compute the estimate of the image $\hat{\boldsymbol{\sigma}}$ is summarized in algorithm 1.

If we are going to apply the generalized reweighted prior-conditioning as discussed in Section 5.2.3, we can do so by defining an outer iteration loop around Algorithm 1 where the weights are obtained using the values of $\hat{\boldsymbol{\sigma}}$ at the previous iteration. The reweighted algorithm is summarized as Algorithm 2, where $f(\boldsymbol{\sigma})$ refers to an arbitrary function applied to $\boldsymbol{\sigma}$ which depends on the constraint as discussed in Section 5.2.3, and PRIFIRA(\cdot) denotes Algorithm 1 with the mentioned input. The outer loop also allows for applying more constraints such as projecting the solution into the real and positive orthant but comes at a greater computation expense due to the repeated application of the LSQR algorithm. As initialization, we can choose the MVDR dirty image, the MF dirty image, or simply set $\boldsymbol{\sigma}_0 = \mathbf{1}$.

5.3.2. STOPPING CRITERIA

Algorithm 1 requires an appropriate stopping rule. As mentioned, this goes back to Equation (5.24), where we increase the iteration count t until $\|\bar{\mathbf{r}} - \bar{\mathbf{M}} \boldsymbol{\alpha}_t\|_2^2 \leq \epsilon$. This is known as the discrepancy principle [116]. The threshold ϵ on the residual norm can be set using the expected error on the data at the “true” solution $\boldsymbol{\alpha}$,

$$E\|\bar{\mathbf{r}} - \bar{\mathbf{M}} \boldsymbol{\alpha}\|_2^2 = E\|\boldsymbol{\Gamma} \mathbf{e}\|_2^2 = \text{trace}(\text{Cov}(\boldsymbol{\Gamma} \mathbf{e})),\tag{5.28}$$

where $\boldsymbol{\Gamma} \mathbf{e}$ is the whitened error on the data, and $\text{Cov}(\cdot)$ denotes the covariance. Note that, by definition of $\boldsymbol{\Gamma}$,

$$\text{Cov}(\boldsymbol{\Gamma} \mathbf{e}) = E\{\boldsymbol{\Gamma} \mathbf{e} \mathbf{e}^H \boldsymbol{\Gamma}^H\} = \mathbf{I},\tag{5.29}$$

Algorithm 1: PRIFIRA (based on LSQR)

input : $\bar{\mathbf{r}}, \bar{\mathbf{M}}$ (or operator function), $\mathbf{L}^{-1}, \epsilon$
output: image $\hat{\sigma}$

- 1 Initialize: $\beta_1 \mathbf{u}_1 := \bar{\mathbf{r}}, \alpha_1 \mathbf{v}_1 := \bar{\mathbf{M}}^H \mathbf{u}_1, \omega_1 := \mathbf{v}_1, \alpha_1 := \mathbf{0}, \bar{\Phi}_1 := \beta_1, \bar{\rho}_1 := \alpha_1, t := 1;$
- 2 **while** *stopping criteria not satisfied* **do**
- 3 $\beta_{t+1} \mathbf{u}_{t+1} := \bar{\mathbf{M}} \mathbf{v}_t - \alpha_t \mathbf{u}_t,$
- 4 $\alpha_{t+1} \mathbf{v}_{t+1} := \bar{\mathbf{M}}^H \mathbf{u}_{t+1} - \beta_t \mathbf{v}_t;$
- 5 Construct and apply orthogonal transform:
- 6 $\rho_t = (\bar{\rho}_t^2 + \beta_{t+1}^2)^{1/2}$
- 7 $c_t = \bar{\rho}_t / \rho_t, s_t = \beta_{t+1} / \rho_t$
- 8 $\theta_{t+1} = s_t \alpha_{t+1}, \bar{\rho}_{t+1} = -c_t \alpha_{t+1}$
- 9 $\bar{\Phi}_t = c_t \bar{\Phi}_t, \bar{\Phi}_{t+1} = s_t \bar{\Phi}_t;$
- 10 Update:
- 11 $\alpha_{t+1} = \alpha_t + (\bar{\Phi}_t / \rho_t) \omega_t$
- 12 $\omega_{t+1} = \mathbf{v}_{t+1} - (\theta_{t+1} / \rho_t) \omega_t$
- 13 $t := t + 1;$
- 14 **end**
- 15 $\hat{\alpha} = \alpha_t;$
- 16 Transform to the image: $\hat{\sigma} = \mathbf{L}^{-1} \hat{\alpha}$

5

Algorithm 2: Reweighted PRIFIRA

- 1 Initialize:
- 2 $\sigma_0 = \mathbf{1}, \mathbf{W}_1^{-1} = \text{diag}(f(\sigma_0)), \bar{\mathbf{M}}_1 = \Gamma \mathbf{M} \mathbf{W}_1^{-1};$
- 3 **for** $k = 1, 2, \dots, K$ **do**
- 4 $\sigma_k = \text{PRIFIRA}(\bar{\mathbf{r}}, \bar{\mathbf{M}}_k, \mathbf{W}_k^{-1}, \epsilon);$
- 5 $\mathbf{W}_{k+1}^{-1} = \text{diag}(f(\sigma_k));$
- 6 $\bar{\mathbf{M}}_{k+1} = \Gamma \mathbf{M} \mathbf{W}_{k+1}^{-1},$
- 7 **end**
- 8 $\hat{\sigma} = \sigma_K$

where \mathbf{I} is a $P^2 \times P^2$ identity matrix. Therefore, we set $\epsilon = P^2$, or a slightly larger value to account for finite sample noise.

If the autocorrelations of the measurements are not available, we need to resort to the unweighted LS estimator ($\mathbf{\Gamma} = \mathbf{I}$). The stopping criteria are based on an estimate of the noise on the visibilities.

5.3.3. COMBINED REGULARIZING EFFECT OF PRIOR-CONDITIONING AND EARLY STOPPING

In this Section we discuss the combined regularizing effect of the prior-conditioning and early stopping on the solution of Problem (5.24). The SVD is a powerful tool for the analysis of least squares problems. It is well-known that the minimum norm solution of a linear least squares problem is obtained via the pseudo-inverse [66]. For example we consider Problem (2.69). If the SVD of $\mathbf{\Gamma M}$ can be stated as $\mathbf{\Gamma M} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$ where \mathbf{U} and \mathbf{V} contain the left and right singular vectors respectively and $\mathbf{\Lambda}$ is a diagonal matrix containing the singular values of $\mathbf{\Gamma M}$, the minimum norm solution can be expressed as

$$\hat{\boldsymbol{\sigma}} = (\mathbf{\Gamma M})^\dagger \bar{\mathbf{r}} = \mathbf{V}\mathbf{\Lambda}^\dagger \mathbf{U}^H \bar{\mathbf{r}}. \quad (5.30)$$

Unfortunately, for ill-posed problems the pseudo-inverse solution is unstable due to the noise amplification by the inversion of small singular values [59].

Applying regularization stabilizes the solution. The solution of many regularized least squares problems can be stated in the form of a filtered SVD [113]. If regularization is applied on (2.69), the solution in terms of filtered SVD can be stated as

$$\hat{\boldsymbol{\sigma}} = \mathbf{V}\mathbf{\Phi}\mathbf{\Lambda}^\dagger \mathbf{U}^H \bar{\mathbf{r}}, \quad (5.31)$$

where $\mathbf{\Phi}$ is a diagonal matrix containing the regularization filter factors and is dependent on the type of regularization applied. The purpose of the regularizing filter factors is to filter out the effect of small singular values in $\mathbf{\Lambda}$ that cause noise amplification and instability of the estimated solution when inversion is performed.

We present the solution of (5.24) in terms of filtered SVD to show the regularizing effect of the iteration count and the prior-conditioner. Assuming in this case the SVD of $\bar{\mathbf{M}} = \mathbf{\Gamma M L}^{-1}$ is given as $\bar{\mathbf{U}}\bar{\mathbf{\Lambda}}\bar{\mathbf{V}}^H$, starting from (5.23) and following the approach from [117, 113] we obtain:

$$\begin{aligned} \mathcal{K}_t(\bar{\mathbf{M}}^H \bar{\mathbf{M}}, \bar{\mathbf{M}}^H \bar{\mathbf{r}}) \\ = \text{span}\{\bar{\mathbf{V}}\bar{\mathbf{\Lambda}}\bar{\mathbf{U}}^H \bar{\mathbf{r}}, \bar{\mathbf{V}}\bar{\mathbf{\Lambda}}^3 \bar{\mathbf{U}}^H \bar{\mathbf{r}}, \dots, \bar{\mathbf{V}}\bar{\mathbf{\Lambda}}^{2t-1} \bar{\mathbf{U}}^H \bar{\mathbf{r}}\}. \end{aligned} \quad (5.32)$$

Since the solution is a linear combination of vectors, the filtered SVD solution of $\hat{\boldsymbol{\sigma}}$ can be stated as

$$\hat{\boldsymbol{\sigma}} = \mathbf{L}^{-1} \bar{\mathbf{V}} \bar{\mathbf{\Phi}}_t \mathbf{\Lambda}^\dagger \bar{\mathbf{U}}^H \bar{\mathbf{r}}, \quad (5.33)$$

where $\bar{\mathbf{\Phi}}_t$ is a diagonal matrix of the form $\bar{\mathbf{\Phi}}_t = \mathcal{P}_t(\bar{\mathbf{\Lambda}}^2) \bar{\mathbf{\Lambda}}^2$ where \mathcal{P}_t indicates a polynomial of degree smaller than $t - 1$. This polynomial is shown to be dominated by large singular values in the initial iterations. As the iteration continues, more singular values are recovered and the effect of small singular values becomes prominent. Therefore, choosing the right stopping iteration, T , limits the influence of the small singular values and therefore stabilizes the solution [59].

We can conclude from Equation (5.33) that the sky map obtained using PRIFIRA is in the form of a regularized least squares solution. In this solution, both the prior-conditioning weights and the iteration count contribute to the regularization filter factors for filtering out the small singular values and thus stabilizing the inversion.

5.3.4. IMPLEMENTATION DETAILS

As mentioned earlier, the system matrix $\tilde{\mathbf{M}}$ is a full matrix. However it exhibits a “data sparse” structure. Since $\mathbf{M} = \mathbf{A}^* \circ \mathbf{A}$, we can represent the system matrix \mathbf{M} of dimension $P^2 \times Q$ with a lower dimensional matrix \mathbf{A} of dimension $P \times Q$. In the case of $\tilde{\mathbf{M}}$ we need to apply the proper right and left preconditioners to \mathbf{A} . Considering the Cholesky factorization

$$\hat{\mathbf{R}}^{-1} = \mathbf{B}^H \mathbf{B}, \quad (5.34)$$

we find $\mathbf{\Gamma} = N^{1/2} (\mathbf{B}^* \otimes \mathbf{B})$ and therefore $\tilde{\mathbf{M}} = \tilde{\mathbf{A}}^* \circ \tilde{\mathbf{A}}$ with $\tilde{\mathbf{A}} := N^{1/4} \mathbf{B} \mathbf{A} L^{-1/2}$. If the dimensions of the imaging problem are such that we can store matrix $\tilde{\mathbf{A}}$ in memory, we implement the matrix vector operations $\tilde{\mathbf{M}}\mathbf{v}$ and $\tilde{\mathbf{M}}^H \mathbf{u}$ as

$$\begin{aligned} \tilde{\mathbf{M}}\mathbf{v} &= \text{vect}(\tilde{\mathbf{A}} \text{diag}(\mathbf{v}) \tilde{\mathbf{A}}^H), \\ \tilde{\mathbf{M}}^H \mathbf{u} &= \text{vect} \text{diag}(\tilde{\mathbf{A}}^H \mathbf{U} \tilde{\mathbf{A}}) = [\tilde{\mathbf{a}}_i^H \mathbf{U} \tilde{\mathbf{a}}_i]_{i=1}^Q, \end{aligned} \quad (5.35)$$

where $\text{diag}(\mathbf{v})$ is a diagonal matrix with the vector \mathbf{v} on its main diagonal, $\text{vect} \text{diag}(\cdot)$ selects the diagonal of a matrix and stores it in a vector, and \mathbf{U} is a $P \times P$ matrix such that $\mathbf{u} = \text{vect}(\mathbf{U})$. The diagonal matrices are stored in a sparse manner for memory considerations.

If the dimensions of \mathbf{A} are also higher than the available physical memory, the matrix-vector multiplications can be directly implemented through the function representation of matrix \mathbf{M} as denoted in Equation (2.26), or more efficiently through the W -projection algorithm or its various implementations [118].

5.3.5. COMPUTATIONAL COMPLEXITY OF PRIFIRA

As can be seen from the description of Algorithm 1, the computational complexity of PRIFIRA is dominated by the two matrix-vector multiplications $\tilde{\mathbf{M}}\mathbf{v}_t$ and $\tilde{\mathbf{M}}^H \mathbf{u}_t$. Therefore, the implementation of these matrix-vector multiplications determines the computational complexity of the algorithm. The first operation is in fact equivalent to computing correlation data from an image (sky model) using the measurement equation, while the second corresponds to the computation of an MF dirty image from correlation data. These are standard operations in any radio astronomy imaging toolbox, and many fast algorithms (based on gridding and FFTs) have been proposed and implemented.

Assuming that no fast transform is used to obtain the matrix-vector multiplications, the complexity of computing $\tilde{\mathbf{M}}\mathbf{v}_t$ is $\mathcal{O}(P^2 Q)$, and computing $\tilde{\mathbf{M}}^H \mathbf{u}_t$ has the same complexity. The complexity of PRIFIRA is thus $\mathcal{O}(T P^2 Q)$ where T denotes the required number of iterations until the stopping criteria are satisfied. Simulations indicate that T is usually quite small (around 5 to 10). In case of the reweighted PRIFIRA, the complexity increases to $\mathcal{O}(K T P^2 Q)$ where K is the total number of reweighted outer iterations.

To compare this complexity to the existing imaging algorithms, we first note that all of them require basic operations of the form $\mathbf{M}\mathbf{v}$ (a forward step, computing correla-

tion data from a sky model) and $\mathbf{M}^H \mathbf{u}$ (a backward step, computing a dirty image from correlation data), and often these are implemented efficiently using gridding, FFTs, and W-projections.

Existing algorithms can be classified into (1) greedy algorithms such as CLEAN and NNLS, which require an exhaustive search over all the pixels of the image to find potential sources; and (2) compressed-sensing based algorithms, implemented using convex optimization, such as SARA implemented using ADMM. The first category requires a large number of iterations where the basic work is comparable to PRIFIRA, although significant reductions are possible by utilizing multi-resolution and image partitioning techniques. NNLS also requires a number of sub-iterations for each iteration to solve a system of equations and is therefore more costly. These algorithms are sensitive to grid mismatch due to misalignment of the sources with the grid points.

Like PRIFIRA, compressed sensing algorithms consider the complete image at once and therefore are less sensitive to grid mismatch. CS algorithms are based on gradient descent steps which in the end boil down to matrix-vector multiplications with \mathbf{M} and \mathbf{M}^H . Less costly sub-iterations and (nonlinear) outer loops are used to satisfy positivity and sparsity constraints through proximity operators. While the amount of work per iteration is therefore comparable, the simulations presented in Section 5.4 show that the prior-conditioning used by PRIFIRA provides an order of magnitude faster convergence. A good estimate of the image is already obtained after a few iterations, resulting in major savings on the overall cost of the imaging algorithm.

5.4. ONE-DIMENSIONAL SIMULATION RESULTS

In this section, we show the performance of PRIFIRA based on a simple one-dimensional case. Further realistic simulation and experimental results are deferred to Chapter 6.

5.4.1. TERMINOLOGY

We proposed several variants of the PRIFIRA algorithm, based on the initial prior-conditioners and the optional use of reweighting iterations. We therefore indicate the right prior-conditioner as a prefix to the name of the algorithm; i.e., X -PRIFIRA where X indicates the prior-conditioner. We consider X as MF, MVDR, IR0 and IR1 where MF and MVDR respectively denote the matched filtered and MVDR dirty images, and IR0 and IR1 indicate the iteratively reweighted PRIFIRA resulting in ℓ_0 and ℓ_1 image norm minimizations respectively, as discussed in Section 5.2.3. For comparison and to show the effect of right preconditioners on the reconstruction quality, we also consider the LSQR algorithm which is equivalent to PRIFIRA when there is no right preconditioner applied.

5.4.2. ONE-DIMENSIONAL TESTS

We first demonstrate the effects of prior-conditioning using a one-dimensional test example. For this simulation, we use a nonuniform linear array with $P = 10$ elements as shown in Figure 5.1(a). The conditioning of matrix \mathbf{M} is shown via its singular value spectrum in Figure 5.1(b). Two Gaussian sources with the same height 2 and different width positioned at direction cosines $l = -0.5$ and $l = 0.5$ are used to model resolved

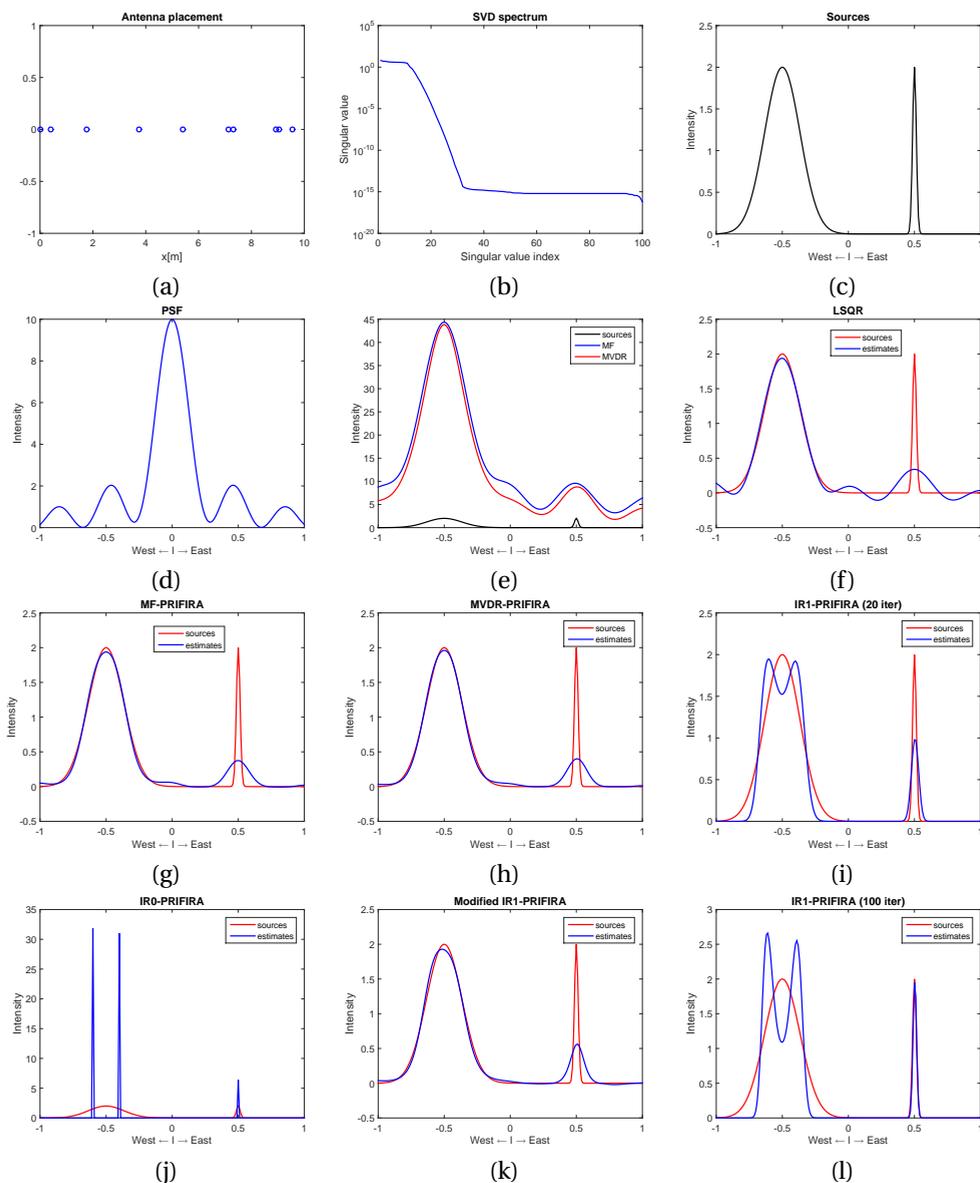


Figure 5.1: (a) Antenna positions, (b) singular value distribution of \mathbf{M} , (c) source distribution, (d) PSE, (e) MF and MVDR dirty images, (f) LSQR reconstruction, (g) MF-PRIFIRA reconstruction, (h) MVDR-PRIFIRA reconstruction, (i) IR1-PRIFIRA reconstruction, (j) IR0-PRIFIRA reconstruction, (k) modified IR1-PRIFIRA reconstruction, (l) IR1-PRIFIRA reconstruction after 100 outer iterations

and unresolved sources, respectively. The sources are shown in Figure 5.1(c). We discretize the line “image” into $Q = 201$ pixels. The operating frequency is set to 80 MHz and the covariance data contains correlated noise with power 100. The correlation data $\hat{\mathbf{R}}$ is constructed from $N = 10^5$ samples.

The PSF of the antenna array is shown in Figure 5.1(d) and as can be seen contains large sidelobes. We can see based on the PSF that the left Gaussian source, with a width larger than the main beam, can be considered as a resolved source and the right Gaussian, that is significantly narrower than the main beam, can be considered as an unresolved source. The MF and the MVDR dirty images are plotted in Figure 5.1(e). As can be seen, due to the large noise power, the MF and MVDR images are relatively close. Figure 5.1(f), (g), (h), (i) and (j) respectively show the reconstruction results for LSQR, MF-PRIFIRA, MVDR-PRIFIRA, IR1-PRIFIRA and IR0-PRIFIRA. The total number of iterations until the stopping criteria are achieved for LSQR, MF-PRIFIRA, MVDR-PRIFIRA, IR1-PRIFIRA and IR0-PRIFIRA are 4, 3, 3, 60 and 60, respectively. 20 outer iterations are used for IR1- and IR0-PRIFIRA. For IR0-PRIFIRA, the non-zero coefficients in $\hat{\boldsymbol{\alpha}}_k$ converge to 1 after 20 iterations. Therefore, the solution $\hat{\boldsymbol{\theta}}$ becomes invariant with the increase of outer iterations. We have added two more reconstructions based on IR1-PRIFIRA. Figure 5.1(k) is a modified version of IR1-PRIFIRA such that the prior-conditioning weights are computed as $\mathbf{W}_k^{-1} = \text{diag}(\hat{\boldsymbol{\theta}}_{k-1}^{1/2} + \varepsilon)$ with $\varepsilon = 0.2$. Similar modifications are proposed in [109] and [103] for stability reasons. The number of outer iterations is kept at 20 for this result. Furthermore, Figure 5.1(l) shows an extreme case of IR1-PRIFIRA with 100 outer iterations.

The figure shows that the LSQR reconstructed image has many sidelobes, some of which are negative. MF-PRIFIRA and MVDR-PRIFIRA stabilize the solution such that the sidelobes disappear to a large extent with MVDR-PRIFIRA being more successful in this regard. Both MF-PRIFIRA and MVDR-PRIFIRA recover the resolved source reliably and smear out the unresolved source as was expected from the third argument in Section 5.2.3. IR1-PRIFIRA attempts to narrow the smearing while IR0-PRIFIRA aims for an optimally sparse and spiky solution which is not the preferred solution in cases where retrieving extended emissions are of interest. The modified version of IR1-PRIFIRA is more faithful to the recovery of the extended emission while smearing the unresolved source. In the extreme case, IR1-PRIFIRA recovers the unresolved source almost perfectly while narrowing the extended emission into two peaks similar to the effect observed with the recovery of IR0-PRIFIRA. Both IR1- and IR0-PRIFIRA do not observe the natural resolution of the instrument while MF and MVDR-PRIFIRA maintain this resolution.

In Section 5.5, we discuss the effect of removing the autocorrelations analytically and based on a simulation for a one-dimensional scenario.

We now look more closely into the Krylov basis vectors produced by the various algorithms. As mentioned in Section 5.3, Krylov subspace-based methods restrict the solution space to the first t Krylov vectors. When applying the LSQR algorithm, the Krylov vectors are reorthogonalized as Lanczos vectors indicated by \mathbf{v}_t at iteration t . Therefore, the solution space is spanned by $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t]$. It is informative to look at the Lanczos vectors with and without the application of prior-conditioners. We show these effects for the simple one-dimensional test case. Figure 5.2 shows the first four initial Lanczos vectors. It is seen that the LSQR basis has a non-zero support where the true image is

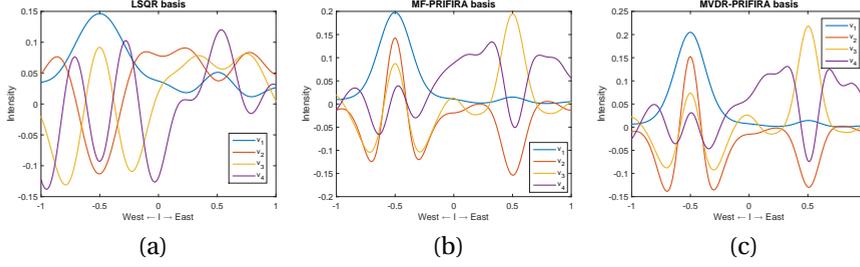


Figure 5.2: (a) LSQR basis vectors, (b) MF-PRIFIRA basis vectors, (c) MVDR-PRIFIRA basis vectors

zero while MF- and MVDR-PRIFIRA bases capture the support of the image already in the initial Lanczos vectors. This indicates that the latter bases provide a good space to represent the image.

5.5. EFFECT OF MISSING AUTOCORRELATIONS

Traditionally in radio astronomy, the autocorrelations are not measured or are discarded for image formation, as they are considered inaccurate due to the addition of large noise power terms. We briefly discuss the effect of missing autocorrelations on the proposed method.

If the autocorrelations are not available, we need to change the data model accordingly. We redefine $\tilde{\mathbf{r}}$ as

$$\tilde{\mathbf{R}} = \hat{\mathbf{R}} - \text{diag}(\hat{\mathbf{R}}), \quad \tilde{\mathbf{r}} = \text{vect}(\tilde{\mathbf{R}}), \quad (5.36)$$

which includes “zero” entries in place of the missing autocorrelations. It is straightforward to derive that $\tilde{\mathbf{r}}$ is related to $\hat{\mathbf{r}}$ as

$$\tilde{\mathbf{r}} = \mathbf{\Pi} \hat{\mathbf{r}}, \quad (5.37)$$

where

$$\mathbf{\Pi} = \mathbf{I}_{p^2} - (\mathbf{I}_p \circ \mathbf{I}_p)(\mathbf{I}_p \circ \mathbf{I}_p)^H \quad (5.38)$$

is an orthogonal projection matrix that projects out the diagonal entries from $\hat{\mathbf{r}}$. The resulting data model is

$$\tilde{\mathbf{r}} = \tilde{\mathbf{M}} \boldsymbol{\sigma} + \tilde{\mathbf{e}},$$

where $\tilde{\mathbf{M}} = \mathbf{\Pi} \mathbf{M}$, and $\tilde{\mathbf{e}} = \mathbf{\Pi} \mathbf{e}$ is the finite sample noise, modeled as complex Gaussian with zero mean and variance

$$\tilde{\mathbf{C}}_e = \mathbf{\Pi} \mathbf{C}_e \mathbf{\Pi} = \frac{1}{N} \mathbf{\Pi} (\mathbf{R}^T \otimes \mathbf{R}) \mathbf{\Pi}.$$

This has a number of consequences:

1. $\tilde{\mathbf{r}}$ does not correspond to a positive (correlation) matrix;
2. A straightforward estimate of $\tilde{\mathbf{C}}_e$ is unknown as $\hat{\mathbf{R}}$ is unavailable. Moreover, $\tilde{\mathbf{C}}_e$ is not invertible. Thus, the weight matrix $\mathbf{\Gamma}$ in the regularized WLS Problem (5.1) is not available, and we need to resort to the unweighted LS formulation

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\text{argmin}} \|\mathbf{\Pi}(\tilde{\mathbf{r}} - \mathbf{M} \mathbf{L}^{-1} \boldsymbol{\alpha})\|_2^2 + \tau \|\boldsymbol{\alpha}\|_2^2; \quad (5.39)$$

3. The MVDR beamformer weights cannot be formed, for the same reason. We used this to form an initial image for the regularization operator \mathbf{L} . Instead, we should resort to the MF (classical) dirty image (omitting autocorrelation terms), $\tilde{\boldsymbol{\sigma}}_{\text{MF}} = \tilde{\mathbf{M}}^H \tilde{\mathbf{r}}$ (cf. (2.58)) and set $\mathbf{L}^{-1} = \text{diag}(\tilde{\boldsymbol{\sigma}}_{\text{MF}})$ as a surrogate.

Under the usual assumptions in radio astronomy (noise much stronger than the sources, noise powers have been whitened), it can be argued that the difference between WLS and LS is small, and also the difference between MF and MVDR is small. Alternatively, we can apply diagonal loading and replace $\tilde{\mathbf{R}}$ by $\tilde{\mathbf{R}} + \eta \mathbf{I}$, where η is a noise variance estimate.

More important is the fact that $\tilde{\mathbf{r}}$ does not correspond to a positive matrix. The resulting MF dirty image $\tilde{\boldsymbol{\sigma}}_{\text{MF}}$ does not have to be positive and sources can have negative sidelobes. Similarly, the PSF, or dirty beam, is defined as $\mathbf{b} = \mathbf{M}^H \mathbf{1}$, and becomes $\tilde{\mathbf{b}} = \tilde{\mathbf{M}}^H \mathbf{1}$. Since $\mathbf{M} = \mathbf{A}^* \circ \mathbf{A}$, and assuming normalized array response vectors $\|\mathbf{a}_j\| = 1$, it can be shown that $\tilde{\mathbf{b}} = \mathbf{b} - \mathbf{1}$. Thus, also the modified PSF can have negative sidelobes, although it is straightforward to correct this.

The negative sidelobes in $\tilde{\boldsymbol{\sigma}}_{\text{MF}}$ makes this unsuitable to be used as weight in (5.39). Some entries in this vector may be close to zero, causing the resulting solution to have a black pixel at that location. Negative values should be avoided by shifting up all the pixels by (at least) the smallest negative value of the sidelobes. If we assume the entries of \mathbf{A} to contain only phases, as in (2.22), then all entries have equal magnitude, and it is straightforward to show that the difference between the original MF image and the MF image without autocorrelations is a constant, equal to the total neglected power. (This is essentially because the MF dirty beam is spatially invariant.) Thus, to correct the MF dirty image we only have to estimate a single shift common to all the pixels. For MVDR, the PSF is spatially variant and we cannot use a single common shift to obtain the MVDR image where autocorrelations are available.

Discarding autocorrelations results in the PSF, MF and MVDR image to have negative sidelobes. Since MF and MVDR are applied as weights to the columns of the \mathbf{M} , any zero value in the weights will enforce zero values in the estimated coefficients and eventually in the solution. The upper bound property of MF and MVDR ensures that none of the non-zero image pixels will not be set to zero. However, when autocorrelations are missing we should avoid zero values by shifting up all the pixels in the MF and MVDR image by the smallest negative value of the sidelobes.

In Figure 5.3, we illustrate with a one-dimensional simulation the effect of dropping the autocorrelations on the PSF, MF, and MVDR image, and on the reconstructed MVDR-PRIFIRA image (MF-PRIFIRA would give similar results). We use a similar setting as for the one-dimensional example presented in Section 5.4 but with a different antenna placing to better show the effect of the negative sidelobes. We choose for this experiment two Gaussian sources of heights 5 and 1 centered at direction cosines $l = -0.5$ and $l = 0.5$ respectively from left to right. Figure 5.3(a),(b),(c) and (d) respectively show the antenna placement, PSF, MF dirty image and MVDR dirty image. Lines related to the setting where we have access to the complete correlation matrix are shown in blue, while the red lines represent the case where the autocorrelations are not available. Figure 5.3(e) and (f) represent MVDR-PRIFIRA, and MVDR-PRIFIRA when we do not have the autocorrelations. Figure 5.3(g) is when we make the MVDR image strictly positive by adding the smallest negative sidelobe to the image and apply it as the right precon-

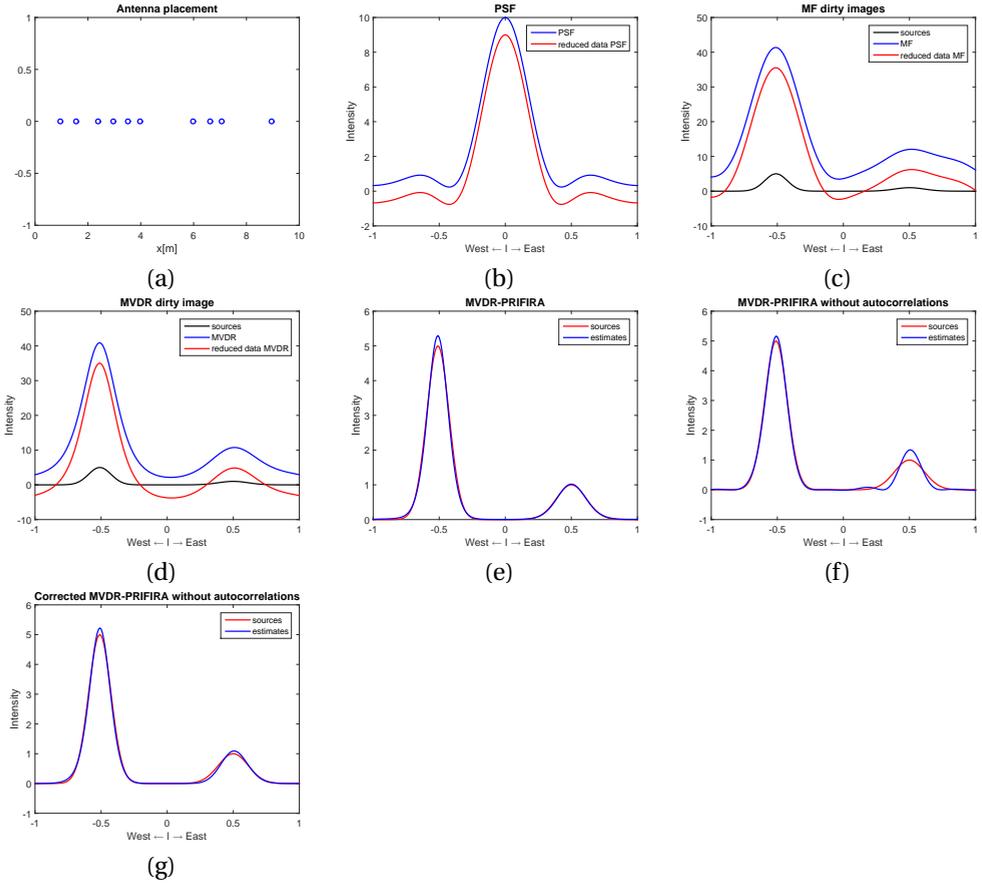


Figure 5.3: Effect of missing autocorrelations: (a) Antenna positions, (b) PSF (c) MF dirty image, (d) MVDR dirty image, (e) MVDR-PRIFIRA, (f) MVDR-PRIFIRA without autocorrelations, (g) MVDR-PRIFIRA without autocorrelations with correction to make sidelobes positive.

ditioner in MVDR-PRIFIRA. As can be seen, this correction improves the estimation performance, although the performance of the algorithm is still suboptimal compared to the case where we have full information.

5.6. AFTERWORD

In this chapter, we have introduced an algorithmic framework to efficiently solve radio astronomical imaging problems with a focus on the recovery of extended emissions. An initial image based on beamforming techniques is used to regularize the Maximum Likelihood image estimation problem by means of prior (or right) preconditioning. We further generalize the proposed framework to also handle images with sparsity priors. To achieve an efficient implementation, we have proposed the use of Krylov subspace methods. We call the algorithmic framework PRIFIRA which consists of different vari-

ants referring to the type of regularization applied. In the next chapter, we evaluate the methods proposed in this chapter based on realistic simulations and real radio astronomical data.

6

PRIFIRA EXPERIMENTAL RESULTS

6.1. INTRODUCTION

In this chapter, we present some realistic simulation and experimental results to test the performance of the PRIFIRA algorithmic framework, presented in Chapter 5, on simulated as well as calibrated real astronomical data. Via simulations, we compare the performance of PRIFIRA with the state-of-the-art radio interferometric imaging methods. The experiments are performed in MATLAB for (i) a data set from a single station of LOFAR, (ii) data set from core stations of LOFAR (Superterp).

We recall the main problems that are tackled in Chapter 5 here. First, we have proposed to state the radio interferometric imaging problem based on a left and right preconditioning formalism, with the change of variables $\boldsymbol{\alpha} = \mathbf{L}\boldsymbol{\sigma}$, as

$$\hat{\boldsymbol{\alpha}} = \operatorname{argmin}_{\boldsymbol{\alpha}} \|\Gamma(\tilde{\mathbf{r}} - \mathbf{M}\mathbf{L}^{-1}\boldsymbol{\alpha})\|_2^2 + \tau\|\boldsymbol{\alpha}\|_2^2, \quad (6.1)$$

by choosing $\mathbf{L}^{-1} = \operatorname{diag}(\hat{\boldsymbol{\sigma}}_{\text{MVDR}})$ or $\mathbf{L}^{-1} = \operatorname{diag}(\hat{\boldsymbol{\sigma}}_{\text{MF}})$. The resulting problem is solved based on the Krylov subspace-based method of LSQR benefiting from the semiconvergence property. The resulting framework is referred to as the MF- or MVDR-PRIFIRA depending on the choice of the right preconditioner.

We have further generalized this framework to incorporate sparsity-based regularization by introducing outer iterations. Defining a right preconditioned system using the transform $\boldsymbol{\alpha} = \mathbf{W}_k\boldsymbol{\sigma}$, in each outer iteration k , the problem

$$\hat{\boldsymbol{\alpha}}_k = \operatorname{argmin}_{\boldsymbol{\alpha}} \|\Gamma(\tilde{\mathbf{r}} - \mathbf{M}\mathbf{W}_k^{-1}\boldsymbol{\alpha})\|_2^2 + \tau\|\boldsymbol{\alpha}\|_2^2. \quad (6.2)$$

is solved using LSQR. In this context, \mathbf{W}_k is chosen as either $\mathbf{W}_k = \operatorname{diag}(\hat{\boldsymbol{\sigma}}_{k-1}^{-1/2})$ or $\mathbf{W}_k = \operatorname{diag}(|\hat{\boldsymbol{\sigma}}_{k-1}|^{-1})$ where the former promotes ℓ_1 sparsity and the latter promotes ℓ_0 sparsity in the solution. The resulting frameworks are called IR1- and IR0-PRIFIRA, respectively.

6.2. TESTS ON MODEL IMAGES

We compare variants of PRIFIRA with the MATLAB implementations of some of the state-of-the-art algorithms. Among the greedy sparse reconstruction methods we use

the NNLS optimization implemented using the active set algorithm as discussed in [56]. The CLEAN algorithm [3] is implemented in MATLAB with both minor cycles and occasional major cycles. MEM is implemented based on Newton-Raphson iterations. Among the compressed sensing techniques based on convex optimization we focus on ℓ_1 norm minimization and the SARA formalism [36] implemented based on the Alternating Direction Method of Multipliers (ADMM) [119]. Furthermore, we compare the results with the conventional deconvolution method of Richardson-Lucy (RL) [120].

We mention that there is no shortage of RA imaging algorithms and it is not possible to compare the proposed method with all the implementations of the present methods. Therefore, we have categorized the imaging methods and compare our algorithm with the basic implementation of the main methods. It is also worth noting that many of the imaging methods have been optimized both in software and hardware to perform faster. There are many possibilities to also optimize PRIFIRA in the future but for the current chapter we focus on the most basic implementation and for a fair comparison compare it with basic implementations of the state-of-the-art algorithms.

We test the proposed methods on the noisy simulated data using the configuration of antennas from the core stations of the LOFAR telescope. As test image, we consider a normalized image of the W28 supernova remnant, shown in Figure 6.1(a), obtained from <https://casaguides.nrao.edu/index.php>. The core stations contain $P = 273$ antennas with a maximum baseline length of about 326 m as shown in Figure 6.1(b). The operating frequency is chosen as 58.975 MHz and a single time snapshot is considered. The u-v coverage of the antenna array is shown in Figure 6.1(c). Figure 6.1(d) illustrates the PSF of the array showing the limited resolution of the array and the existence of sidelobes. To construct the sampled covariance matrix, \mathbf{R} is generated from the test image, $\mathbf{R}^{1/2}$ is used to shape white Gaussian noise into data vectors $\mathbf{x}[n]$ that has the required covariance structure, and white Gaussian receiver noise with variance $\sigma_n^2 = 4$ is added. $N = 10^5$ data samples $\mathbf{x}[n]$ are used to construct $\hat{\mathbf{R}}$. The image is discretized into $Q = 84681$ pixels. The dirty image obtained from the matched filtered beamformer is shown in Figure 6.1(e), and the MVDR dirty image is shown in Figure 6.1(f). The simulations were performed in MATLAB R2014b on a computer with Intel i5-4670 CPU 3.40 GHz under 64-bit Windows 7 with a 8 GB RAM. The images are shown in logarithmic scale and for demonstration and for comparison reasons are limited to scales in the range 1 to $10^{-3.5}$.

Figure 6.2 compares the reconstructed images for the various imaging algorithms. Figure (a) and (b) respectively show the CLEAN and NNLS [56] reconstructions after applying post-processing with a Gaussian main beam which was fitted to the PSE. 10 major cycles of 500 minor cycles are chosen for running the CLEAN algorithm. Figure 6.2(c) shows the ADMM reconstruction with an ℓ_1 sparsity constraint. Figure 6.2(d) is the reconstructed image based on the SARA formalism [36], implemented with ADMM. Figure 6.2(e) is the reconstruction based on the Richardson-Lucy algorithm [120]. Figures 6.2(f) is the maximum entropy reconstruction [77] based on the implementation [121]. This method is very sensitive to the choice of the regularization parameter and the starting vector, and we chose the scaled MF dirty image as the starting vector. Figures 6.2(g), (h), (i), (j) and (k) show the results for LSQR, MF-PRIFIRA, MVDR-PRIFIRA, IR1-PRIFIRA and IR0-PRIFIRA, respectively. 5 outer iterations are chosen for IR0-PRIFIRA and IR1-PRIFIRA.

Qualitatively, Figure 6.2 shows that CLEAN and NNLS have less resolution than the

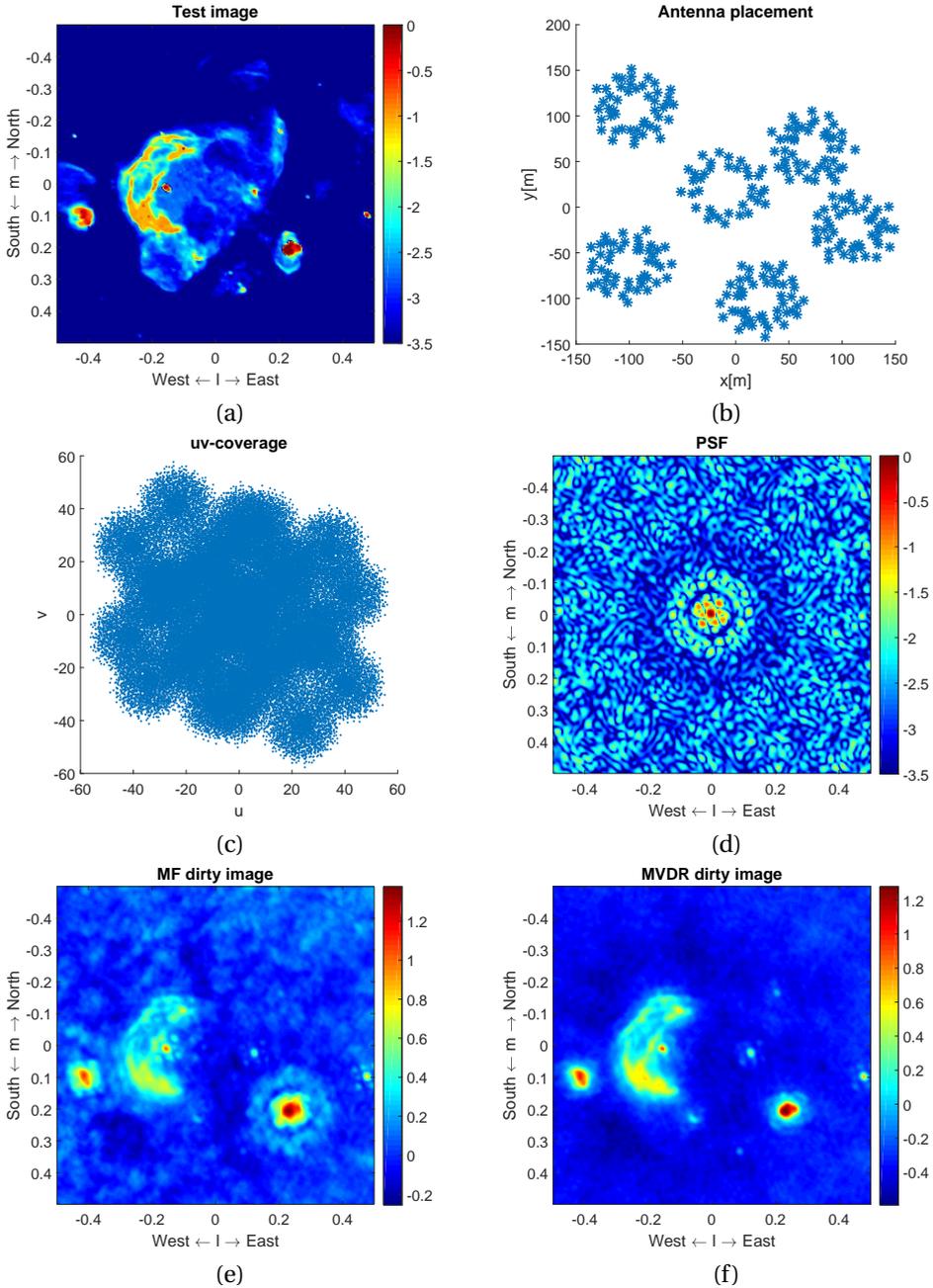
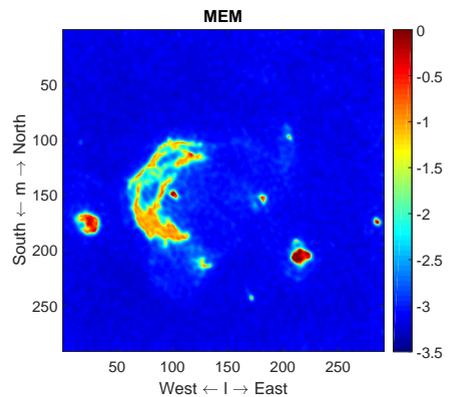
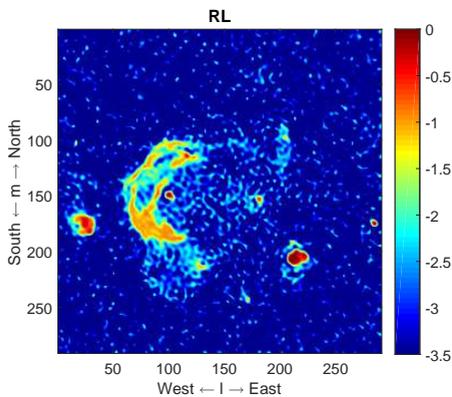
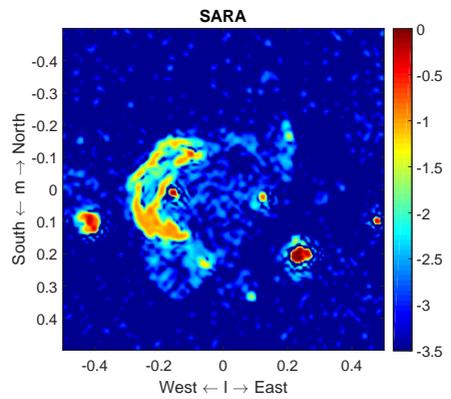
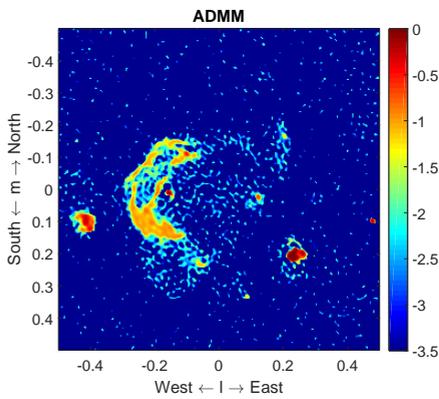
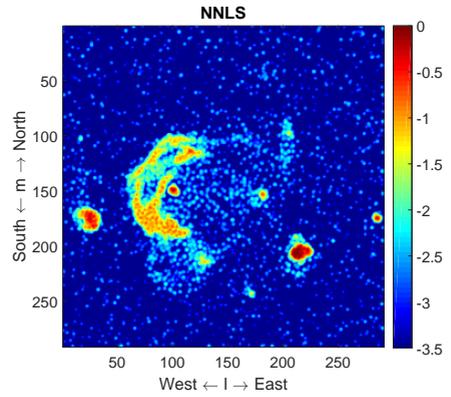
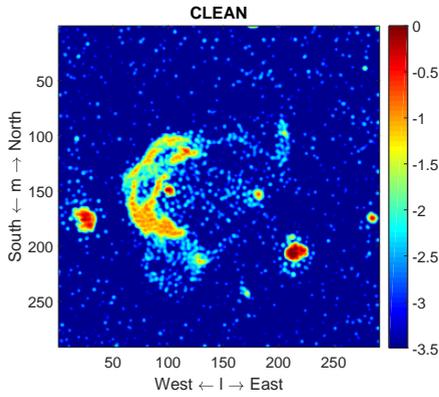


Figure 6.1: (a) Test image in logarithmic scale, (b) antenna placement, (c) u-v coverage, (d) normalized PSF in logarithmic scale, (e) MF dirty image in logarithmic scale, (f) MVDR dirty image in logarithmic scale



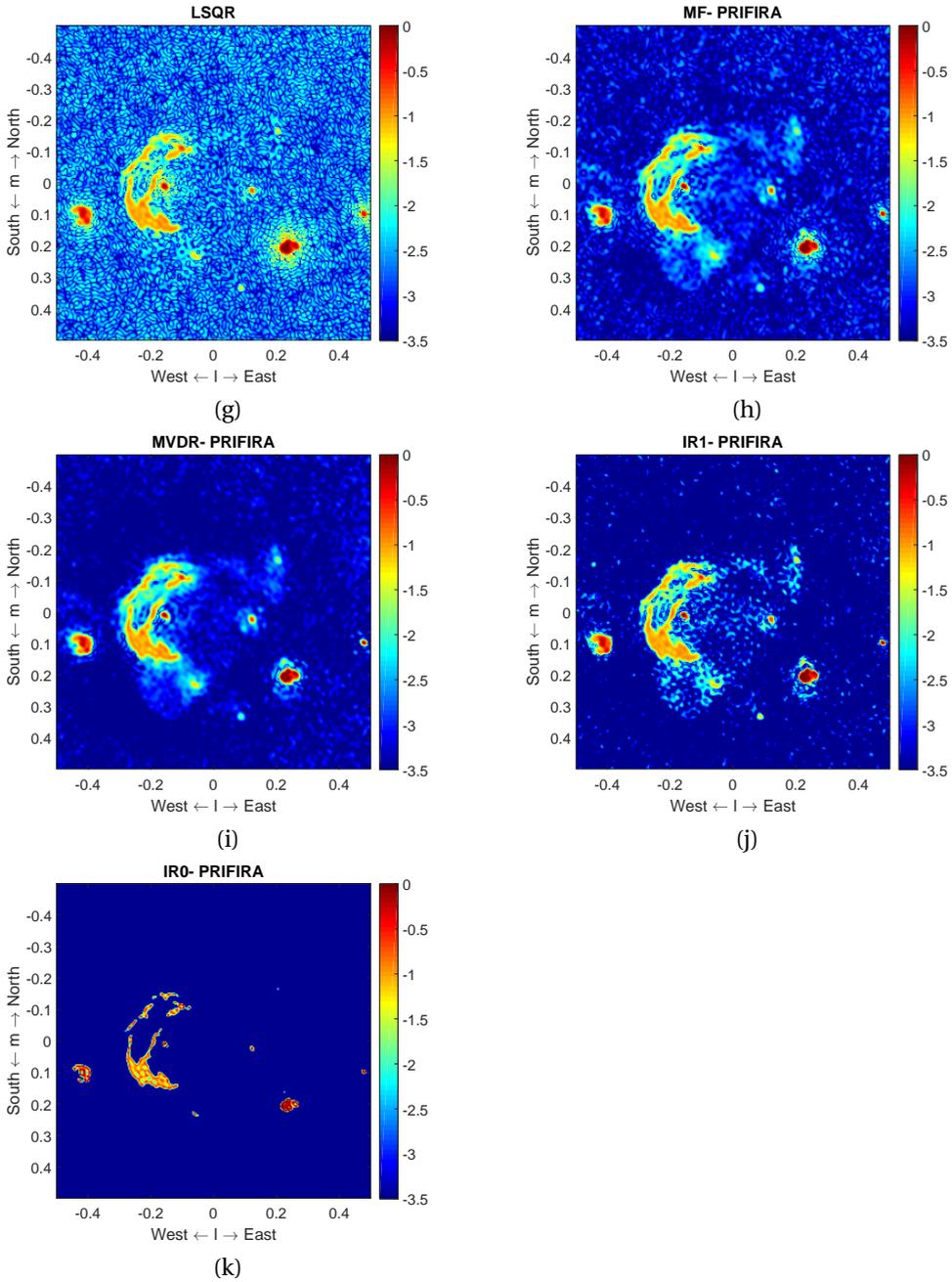


Figure 6.2: (a) CLEAN, (b) NNLS, (c) ADMM, (d) SARA, (e) RL, (f) MEM, (g) LSQR, (h) MF-PRIFIRA, (i) MVDR-PRIFIRA, (j) IR1-PRIFIRA, (k) IR0-PRIFIRA

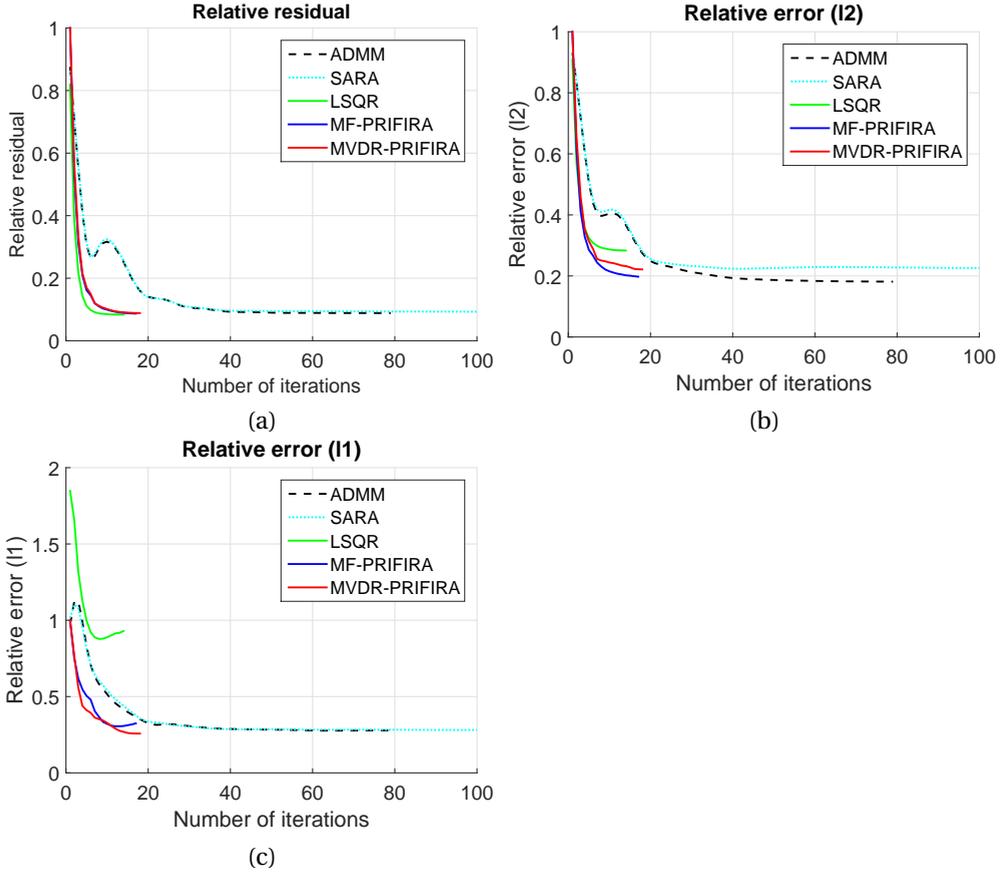


Figure 6.3: (a) Relative residual per iteration, (b) ℓ_2 norm error, (c) ℓ_1 norm error

other methods due to the correction with the main beam. MEM results in a rather “flat” image outside the area of significant emission. LSQR has significant remaining side lobes (ringing effects indicating insufficient regularization), and MF-PRIFIRA also shows this to a lesser extent. MVDR-PRIFIRA and IR1-PRIFIRA are comparable, although the latter results in a “sharper” image due to the imposed sparsity. IR0-PRIFIRA has converged to a very sparse solution, indicating it is not suitable to capture extended structures. These observations are consistent with the 1-D case.

The convergence in terms of relative residual, ℓ_1 -norm error and ℓ_2 -norm error per iteration are compared in Figure 6.3. The relative ℓ_i -norm error for $i = 1, 2$ at iteration t , $e_{i,t}$, is defined as

$$e_{i,t} = \frac{\|\hat{\boldsymbol{\sigma}}_t - \boldsymbol{\sigma}\|_i}{\|\boldsymbol{\sigma}\|_i}, \quad (6.3)$$

where $\boldsymbol{\sigma}$ is the model true image and $\hat{\boldsymbol{\sigma}}_t$ is the reconstructed image at iteration t . We use $e_{1,t}$ as an indicator of how accurately the algorithm is capable of retrieving the source positions as well as the intensities whereas $e_{2,t}$ is mostly concerned with retrieving the cor-

	# iterations	Reconstruction time (s)	$\ \hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\ _2$	$\ \hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\ _1$
CLEAN	5000	69.53	2.86	116.7
NNLS	2656	11.88 (hrs)	3.08	124.8
ADMM	79	134.61	1.45	3.01
SARA	200	353.66	1.76	2.85
RL	200	342.9	1.7	2.61
MEM	30	43.9(min)	1.8	94.4
LSQR	12	22.57	2.27	9.8
MF-PRIFIRA	15	35.08	1.57	3.5
MVDR-PRIFIRA	15	32.57	1.76	2.79
IR1-PRIFIRA	66	128	1.34	2.53
IR0-PRIFIRA	84	165.45	6.42	8.34

Table 6.1: Performance summary

rect overall intensity in the image. ADMM, LSQR, MF-PRIFIRA and MVDR-PRIFIRA are shown in blue, black and red graphs, respectively. For comparison reasons, LSQR, MF-PRIFIRA and MVDR-PRIFIRA are run beyond the stopping threshold. The figure shows that methods based on LSQR exhibit a substantially faster convergence than steepest descent-based ADMM while maintaining comparable reconstruction quality.

The performance of the imaging algorithms is summarized in Table 6.1, which shows the number of iterations, reconstruction time, and error norm for the considered algorithms. The table shows that methods based on LSQR and PRIFIRA with one iteration level (i.e. no outer iterations), namely MF- and MVDR-PRIFIRA, exhibit greatly reduced number of iterations and reconstruction time. We can see that SARA, ADMM and RL exhibit good reconstruction qualities but are considerably slower than the PRIFIRA-based methods. Among the PRIFIRA-based methods IR1-PRIFIRA, MF-PRIFIRA and MVDR-PRIFIRA exhibit the best reconstruction quality.

6.3. SINGLE LOFAR STATION DATA

Next, we test the proposed imaging algorithm on measured correlation data from a single LOFAR station. The data set as introduced in [93, 44] is used. The station consists of an array of 48 antennas as shown in Figure 6.4(a). An observation from a single 10 second snapshot at frequency 50.3125 MHz is considered to construct an image with $Q = 8937$ pixels. The $u - v$ coverage and PSF of the array is shown in Figure 6.4 (b) and (c), respectively. The normalized MF and MVDR images are shown in Figure 6.5(a) and (b), respectively. The power of the additive noise on the antennas is unknown, and we compute an estimate of it as

$$\hat{\boldsymbol{\sigma}}_{\mathbf{n}} = |\hat{\mathbf{R}}^{-1}|^{-\odot 2} \text{vectdiag}(\hat{\mathbf{R}}^{-1}), \quad (6.4)$$

as discussed in [44], where the notation $|\cdot|^{-\odot 2}$ denotes entrywise taking the absolute value, inverting and squaring. MF and MVDR images are computed based on the noise corrected covariance data $\hat{\mathbf{R}} - \hat{\mathbf{R}}_{\mathbf{n}}$.

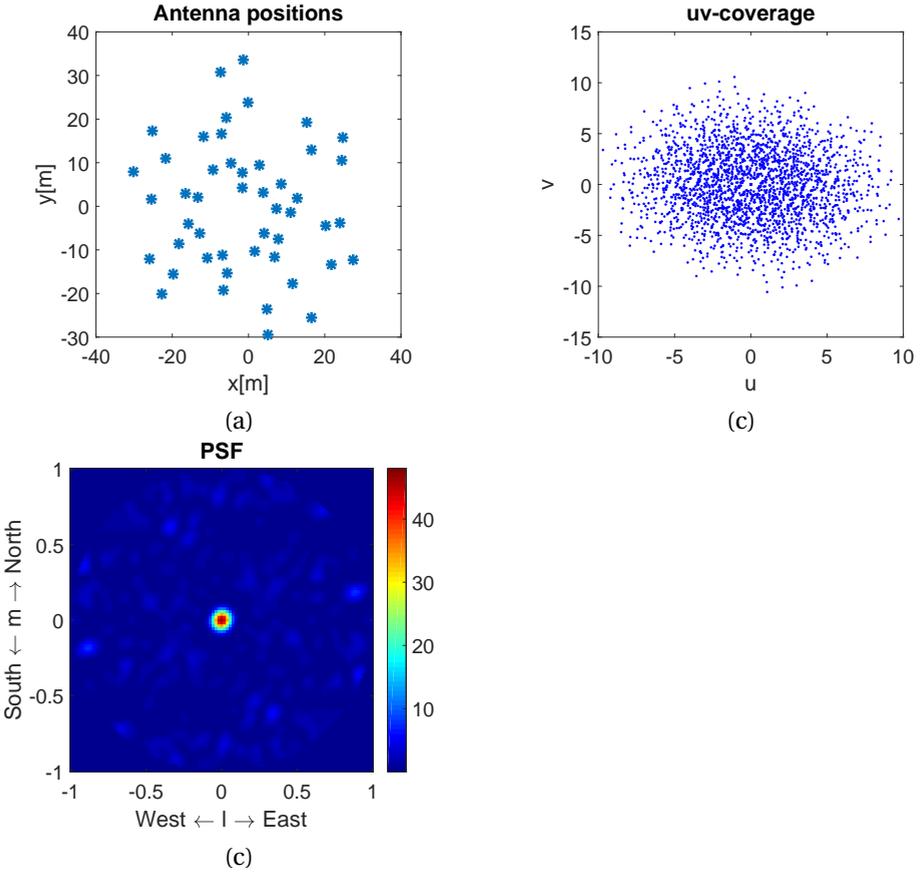


Figure 6.4: (a) LOFAR single station antenna position, (b) $u - v$ coverage, (c) PSF

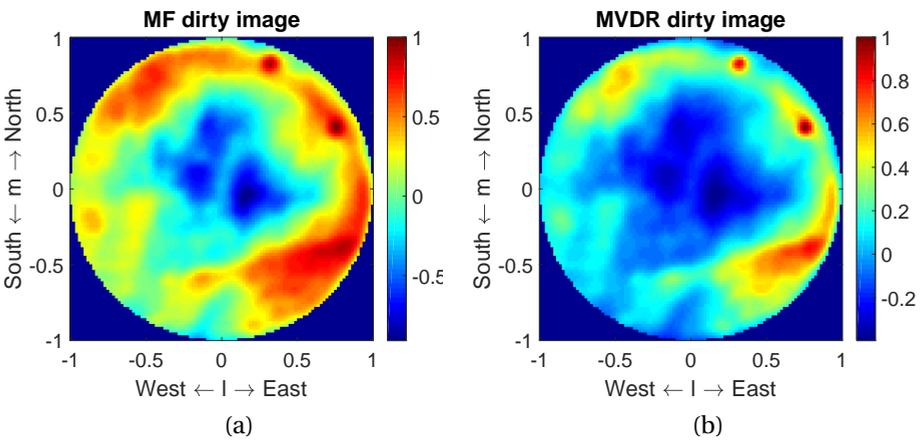


Figure 6.5: (a) MF dirty image, (b) MVDR dirty image

The reconstruction results with the X-PRIFIRA family are shown in Figure 6.6. Since the ground truth of the sky map is unknown with the real data, we show the residual image after reconstruction as is customary in the radio astronomical community. The residual image is computed as

$$\boldsymbol{\delta} = \mathbf{M}^H (\tilde{\mathbf{r}} - \mathbf{M}\hat{\boldsymbol{\sigma}}), \quad (6.5)$$

where $\boldsymbol{\delta}$ indicates the residual image and $\hat{\boldsymbol{\sigma}}$ is the estimated image.

Figures 6.6(a),(b) and (c) respectively show the reconstruction results for LSQR, MF-PRIFIRA, MVDR-PRIFIRA after 7 iterations and Figures 6.7(a),(b) and (c) show the corresponding residual images. Figures 6.6(d) and (e) respectively show the reconstructed images using IR1-PRIFIRA and IR0-PRIFIRA after 3 inner iterations and 4 outer iterations and Figures 6.7(d) and (e) are the corresponding residual images. The image scales on the residual images are cropped at $[-0.2, 0.2]$ for ease of comparison.

As a reference to judge the reconstruction quality of the methods, we use the least squares image generated from combining 25 frequency channels and 10 seconds integration per channel as discussed in [44]. The bright sources are identified as Cyg A and Cas A and the presence of a Galactic loop emerging from Cyg A is identified as loop III in the Haslam survey [44]. Most of the middle and west part of the image do not contain recognizable emissions. This example is interesting as the data contains the contribution from both point sources as well as extended emissions. It is worth noting that we only use data from one frequency and snapshot to test our algorithm. We can see from the LSQR reconstruction in Figure 6.6(a) that there is considerable excess power in the middle and west part of the image due to the instability of the solution. MF-PRIFIRA and MVDR-PRIFIRA correct to a large extent for the faulty reconstruction in the middle and west part of the image while reducing the residual power with MVDR-PRIFIRA showing the smallest and smoothest residual image. IR1-PRIFIRA seems to capture most of the relevant emissions with a similar residual level. However, IR0-PRIFIRA only captures the point sources discarding the extended emissions as predicted and is more appropriate for images with only point sources.

6.4. LOFAR CORE STATIONS

Furthermore, we present the results of X-PRIFIRA applied on the data from 6 core stations of LOFAR (called Superterp). Each station is composed of 48 low-band dipole antennas resulting in a total of 288 antennas [1]. However, 15 antennas have been flagged during the observations and the number of active antennas is reduced to $P = 273$. The station and antenna layout is similar to the layout used for simulations in Section 6.2 shown in Figure 6.1(b) and a picture of the LOFAR Superterp is shown in Figure 6.8.

The data is collected in the context of the Amsterdam-ASTRON Radio Transient Facility And Analysis Center (AARTFAAC) project [122]. The goal of the AARTFAAC project is to implement a near real-time, 24×7 All-Sky Monitor (ASM) for the LOFAR. Since ASM is aimed at providing full-sky images, the data before station beamforming is used for imaging by individual treatment of the antennas contained in a station. To avoid time and bandwidth smearing, the spectral and temporal resolution of the measurements are set to 24 kHz and 1 second, respectively. The ASM pipeline requires low-latency snapshot images to provide near real-time full-sky images for transient detection.

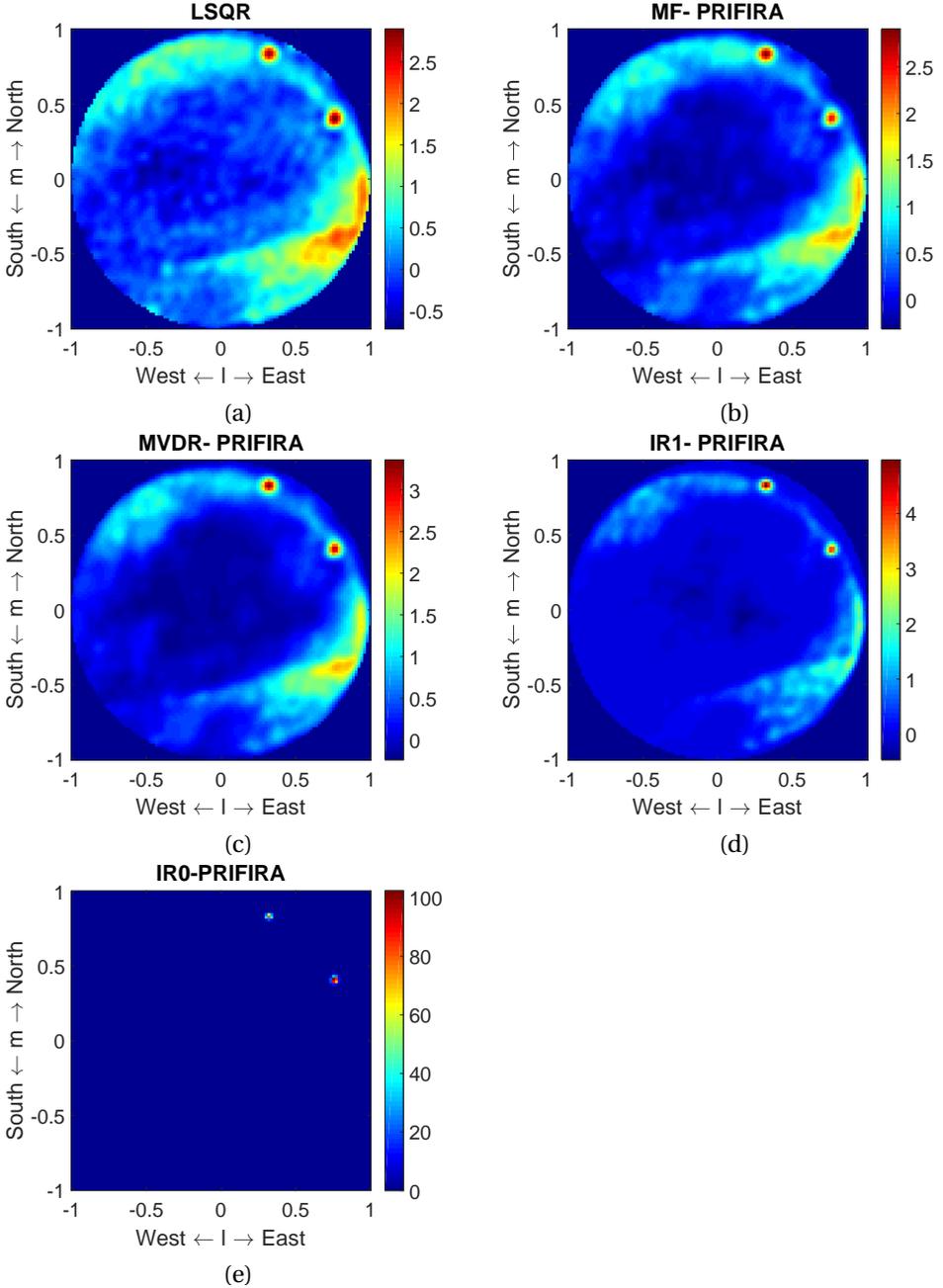


Figure 6.6: Reconstruction results of (a) LSQR, (b) MF-PRIFIRA, (c) MVDR-PRIFIRA, (d) IR1-PRIFIRA, (e)IR0-PRIFIRA

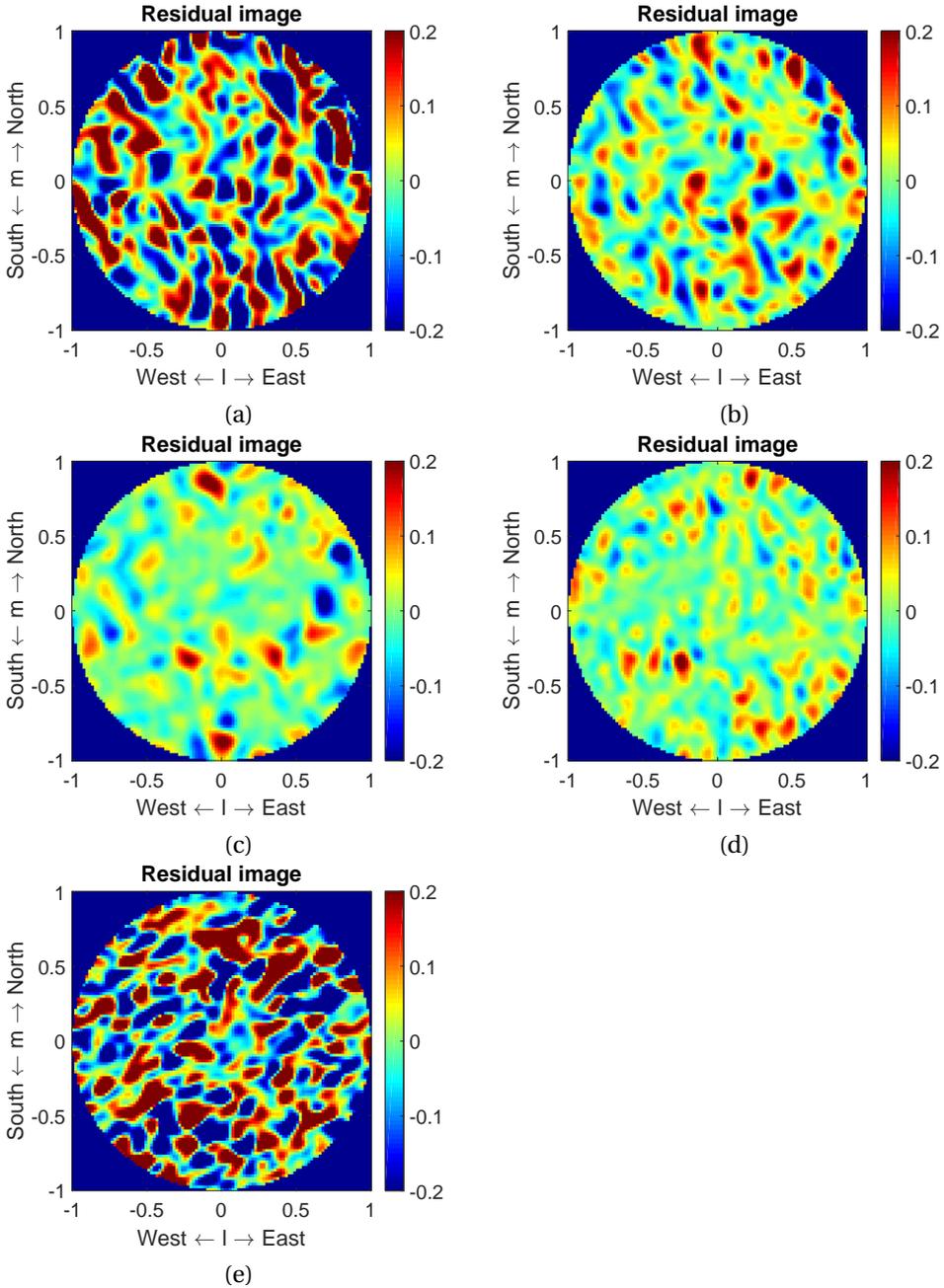


Figure 6.7: Residual images of reconstruction with (a) LSQR , (b) MF-PRIFIRA, (c) MVDR-PRIFIRA, (d) IR1-PRIFIRA, (e) IR0-PRIFIRA



Figure 6.8: LOFAR Superterp. (Image courtesy of [1])

The requirements of ASM fit very well with the characteristics of PRIFIRA. In fact, PRIFIRA can provide a fast imaging framework for near-realtime snapshot imaging. The snapshot data under consideration is at the observing frequency of 58.975 MHz and the observation time (UTC) is 2012 – 07 – 11 – 22 : 22 : 20. The $u - v$ coverage for this experiment is shown in Figure 6.10(a). The data set is calibrated in accordance with the procedure described in [122] prior to performing the imaging. The FoV is discretized into $Q = 206685$ pixels. Since an estimate of the receiver noise is not known, we use the correlation data $\hat{\mathbf{R}}$ for imaging. The normalized full-sky PSF and the MF and MVDR dirty images are shown in Figure 6.10(b),(c) and (d), respectively. We use the logarithmic scale to better represent the dynamic range of the images. The color scale range for the PSF is limited from 1 to 10^{-3} and the normalized dirty images are presented at their actual logarithmic range which is between 1 and about 0.2.

Figure 6.11 (a), (b) and (c) present the normalized reconstruction results for LSQR, MF- and MVDR-PRIFIRA, respectively. The results are obtained after 10 iterations. The normalized reconstructed images are also shown in logarithmic scale and for comparison reasons the scales are limited from 1 to 10^{-3} . Furthermore, the results for IR1- and IR0-PRIFIRA obtained by 10 inner iterations and 5 outer iterations are shown in Figure 6.11 (d) and (e) with the same logarithmic range as before. More outer iterations are required for IR1-PRIFIRA to approximate the point sources. Furthermore, the absolute value of the normalized residual image obtained from LSQR, MF-,MVDR-,IR1- and IR0-PRIFIRA are shown in Figure 6.12(a), (b), (c), (d) and (e), respectively.

Comparing with the dirty images in Figure 6.10, the reconstructed images in Figure 6.11 contain significantly higher dynamic range and the co-existence of emission

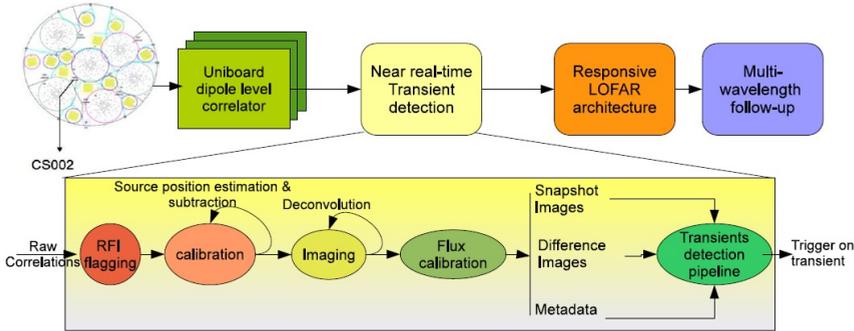
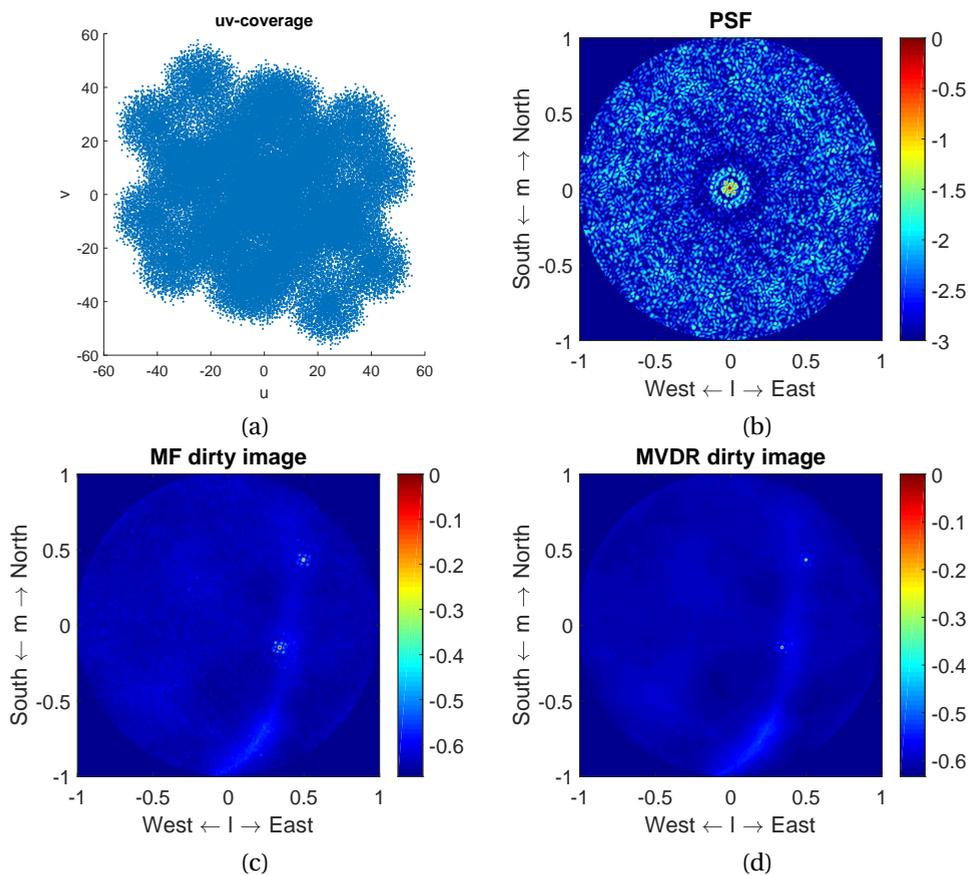


Figure 6.9: AARTFAAC ASM main components. (Image courtesy of [122])

clouds as well as point sources are more prominent. The LSQR reconstruction exhibits ripple effects as expected and this effect is largely reduced by regularizing by the MF or MVDR dirty images. However, the ripple effect is less prominent than the results of Section 6.3 that were obtained using the coverage of a single LOFAR station due to the better $u - v$ coverage of the core stations that results in a better conditioning of the system. Considering MF- and MVDR-PRIFIRA reconstructions, we recognize multiple fainter point sources besides the two bright point sources. These fainter sources are not visible in the dirty images. As before, due to the large level of receiver noise, MF- and MVDR-PRIFIRA provide similar results. We see that IR1- and IR0-PRIFIRA are not the method of choice for this image due to the existence of large emission clouds with similar intensity as the fainter point sources. In fact, ℓ_1 and ℓ_0 sparsity prior on the image do not provide a good prior for this image as it is dominated by extended emissions. From the residual images in Figure 6.12 we arrive at a similar conclusion. While the residual for MF- and MVDR-PRIFIRA are mainly noise-like throughout the FoV, we see the remainder of the dominant point sources in IR1- and IR0-PRIFIRA residual images. This suggests that these methods require many more iterations to completely recover the point sources and therefore are not favorable for quasi-realtime snapshot imaging.

6.5. CONCLUSIONS

We have compared the performance of PRIFIRA with several state-of-the-art imaging algorithms and have shown the computational savings and improvements in accuracy of the estimations. In particular, prior-conditioning using a MF or MVDR dirty image is seen to provide very fast convergence of the Krylov iterations to a solution that has comparable reconstruction quality as the state-of-the-art methods at a significantly reduced computational cost. Furthermore, we have shown the performance of PRIFIRA on realistic data sets from LOFAR to demonstrate the applicability of the method to real astronomical data. These results suggest that regularizing assumptions can be applied on numerical methods based on projections onto Krylov subspaces via prior-conditioning to provide comparable reconstruction results with the existing reconstruction techniques

Figure 6.10: (a) $u-v$ coverage (b) PSF, (c) MF dirty image, (d) MVDR dirty image

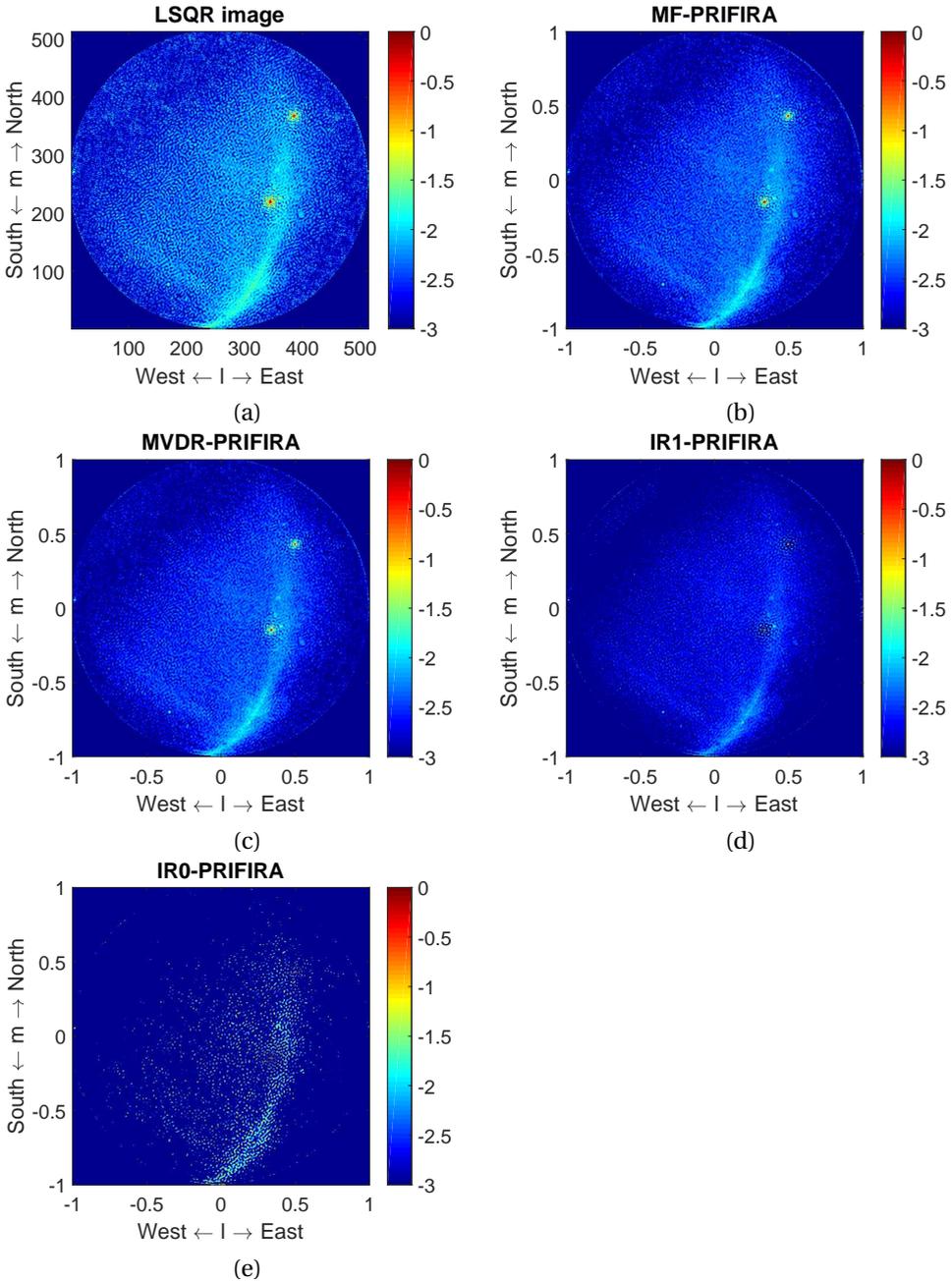


Figure 6.11: Reconstruction results of (a) LSQR, (b) MF-PRIFIRA, (c) MVDR-PRIFIRA, (d) IR1-PRIFIRA, (e)IR0-PRIFIRA

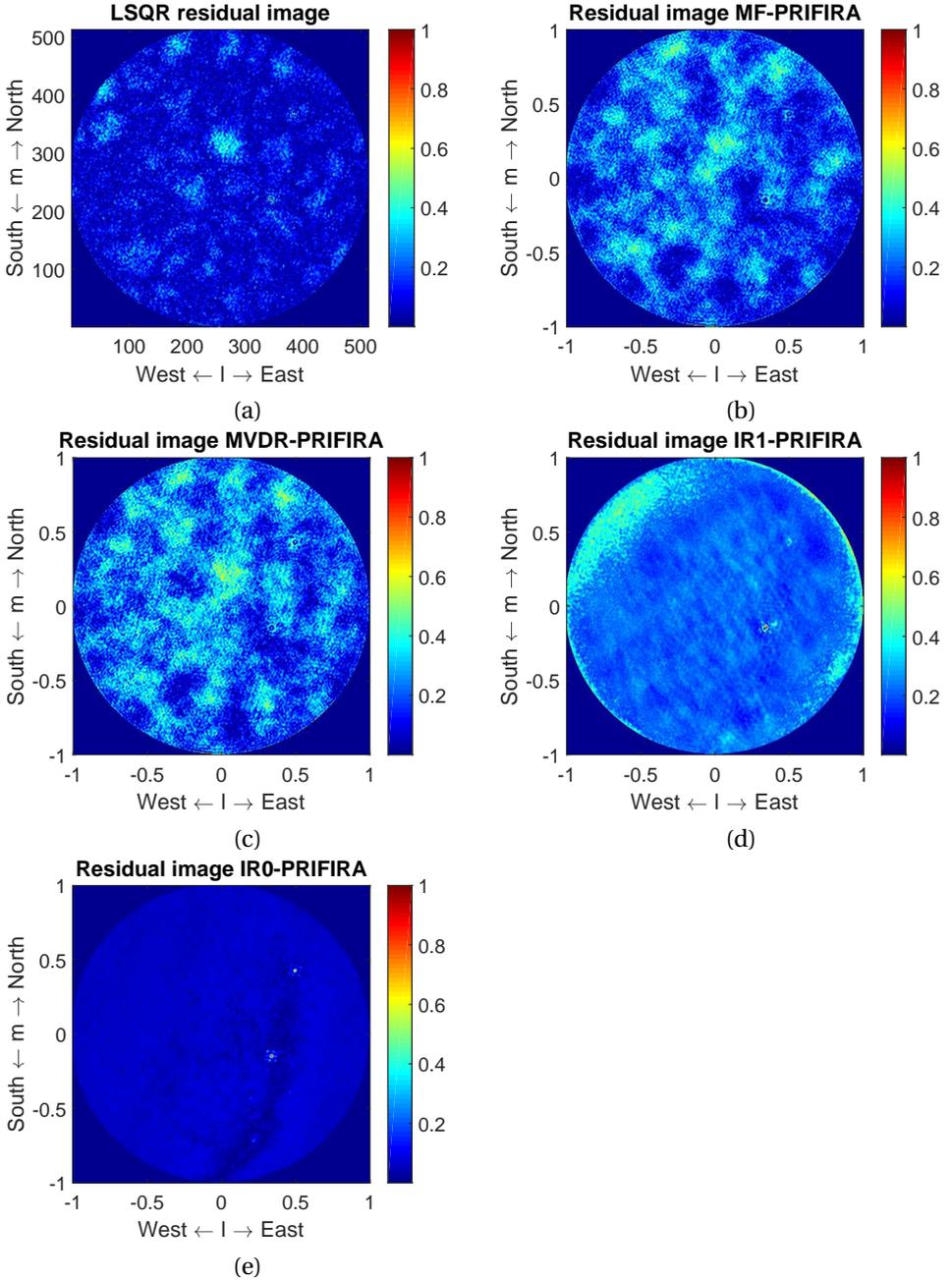


Figure 6.12: Residual images of reconstruction with (a) LSQR , (b) MF-PRIFIRA, (c) MVDR-PRIFIRA, (d) IR1-PRIFIRA, (e) IR0-PRIFIRA

while resulting in computational savings. However, applying sparsity regularization on images which contain sparse as well as extended structures produce suboptimal results. In the next chapter, we investigate ways to solve this issue with introducing overcomplete dictionaries.

III

PART III : SPARSITY-BASED REGULARIZATION

7

DUAL-BASIS SPARSE RECOVERY

They were never meant to be together. Some signals just cannot be represented efficiently in an orthonormal basis.

Joel A. Tropp

7.1. INTRODUCTION

Radio emissions observed by the radio telescopes can be categorized as (i) isolated discrete point-like emissions due to the distant galaxies and quasars and (ii) diffuse extended emissions which are defining features of the radio emissions. In fact, radio telescopes have made it possible to detect such emissions far beyond the optical extent of the galaxies.

Discrete point-like sources are called “unresolved” sources and extended emissions or closely-spaced sources that cannot be distinguished based on the finite resolution of the telescope array are called “resolved” sources. Resolved and unresolved sources cannot be distinguished based on the dirty image without proper side information. Regularizing assumptions and deconvolution procedures for resolved and unresolved sources must ideally be different. However, this requires a remedy to distinguish resolved and unresolved emissions.

In this chapter we introduce such a remedy based on the design of a right preconditioner. Recalling the right preconditioner from Chapter 3, the right-preconditioned measurement system is represented as

$$\tilde{\mathbf{r}} = \mathbf{M}\Theta\boldsymbol{\alpha} + \mathbf{e}. \tag{7.1}$$

Part of this chapter is published as: S.Naghizadeh, A. Mouri Sardarabadi and A.J. van der Veen. Point and beam-sparse radio astronomical source recovery using non-negative least squares. 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), 1-5, 2016.

We see that applying the right preconditioner changes the image model as $\boldsymbol{\sigma} = \boldsymbol{\Theta}\boldsymbol{\alpha}$ where $\boldsymbol{\alpha}$ contains the coefficients of the new model.

It is known that redundant systems, called dictionaries, can be deployed for representing complicated functions [123, 124]. In the presence of extended emissions, the sky map $\boldsymbol{\sigma}$ is not sparse. However, if presented in a correct (redundant) dictionary the coefficients of the sky map in the corresponding dictionary will be sparse. Therefore, right preconditioners can be adjusted to incorporate the (overdetermined) sparsity basis (dictionary). In this case, $\boldsymbol{\Theta}$ represents the new basis for image representation and $\boldsymbol{\alpha}$ is a sparse vector containing the coefficients of image in the dictionary.

Based on these definitions, the sparse coefficient recovery problem can be stated as

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\Gamma(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\Theta}\boldsymbol{\alpha})\|_2^2 + \tau\|\boldsymbol{\alpha}\|_0. \quad (7.2)$$

We see that $\mathcal{R}(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_0$ performs the regularization task. The problem in this case is to recover the sparse coefficients (called atoms) from a redundant dictionary based on the model and the available data. Problem (7.2) corresponds to a synthesis-based sparse recovery problem [123, 125].

If the sky map only contains a sparse set of discrete point sources, the pixel basis is sufficient, i.e. we can set $\boldsymbol{\Theta} = \mathbf{I}$. The working principle of greedy algorithms such as CLEAN and NNLS is to first obtain the support of the image, also called the active set, and to solve only for the elements of the image in the active set. Therefore, as shown by [72], these methods solve the regularized LS or MLE Problem (3.14) with $\mathcal{R}(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}\|_0$,

$$\hat{\boldsymbol{\sigma}} = \underset{\boldsymbol{\sigma}}{\operatorname{argmin}} \|\Gamma(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2 + \tau\|\boldsymbol{\sigma}\|_0, \quad (7.3)$$

with the addition of a nonnegativity constraint for NNLS. Minimizing the ℓ_0 norm produces satisfactory results both in terms of the support of the image and the intensity estimates if the underlying image is sufficiently sparse and only consists of scattered point sources.

CLEAN and NNLS are members of a broader family of algorithms which we call Matching Pursuit-type (MP-type) [126, 127] algorithms. MP-type algorithms are iterative greedy sparse recovery algorithms that at each step select the component in the dictionary $\boldsymbol{\Theta}$ that is best correlated with the residual part of the image. Then they produce a new approximation by projecting the signal onto these already-selected elements [124]. CLEAN and NNLS only consider a pixel basis and encounter limitations in deconvolving extended emissions since these emissions cannot be presented sparsely in the pixel-basis. More specifically, the convergence of these algorithms is very slow and the intensity estimates are not reliable in existence of extended emissions. Cornwell [69] has presented a multi-scale version of CLEAN that allows for representing the image by components with different size scales.

We categorize the radio interferometric imaging algorithms into three groups: (i) greedy (sparse) reconstruction methods such as CLEAN and NNLS, (ii) algorithms based on convex optimization such as MEM and compressive sensing formulations and (iii) projection-based methods. The first two groups have been extensively studied in the context of RA and the third method has been proposed for radio interferometric imaging applications in Chapter 5.

The first part of this chapter is an attempt on extending the greedy sparse method presented in [56] to recover sky maps where resolved and unresolved sources co-exist. We employ a simple beamforming-based basis function to show the effectiveness of the method. We show based on simulations that employing a simple overcomplete dictionary greatly improves the reconstruction quality.

Furthermore, we show a similar sparsity basis can be introduced for the projection-based method of PRIFIRA introduced in Chapter 5. We show the effect of the sparsity basis based on the iteratively-reweighted scheme, namely IR0-PRIFIRA, presented in Chapter 5.

To integrate more advanced regularization schemes we require advanced and highly flexible algorithms that allow for inclusion of sophisticated dictionaries and priors. In the next chapter we show that benefiting from proximal operators and advanced stochastic block coordinate descent algorithms we are able to incorporate advanced regularization assumptions and allow for scalable implementations by block-parallelism on the data as well as in the image. These algorithms fit in the category of convex optimization-based algorithms.

7.2. GRADIENT DESCENT-BASED IMAGING METHODS

We have shown in Chapter 3 that the imaging problem in RA boils down to constrained optimization problems where the objective is to minimize the model fitting error in a LS or MLE sense subject to some constraints based on a prior model on the sky intensity distribution. Iterative algorithms are employed to solve the large-scale imaging problem in RA.

Iterative methods in general start from an initial estimate (guess) of $\boldsymbol{\sigma}$ and find refinements to the estimate iteratively. In general, in the iterative solution method, the solution of the optimization problem

$$\hat{\boldsymbol{\sigma}} = \underset{\boldsymbol{\sigma}}{\operatorname{argmin}} J(\boldsymbol{\sigma}, \tilde{\mathbf{r}}) \quad (7.4)$$

where $J(\cdot)$ is the objective function to be optimized, is found by iterations of the form

$$\boldsymbol{\sigma}^{(t+1)} = \boldsymbol{\sigma}^{(t)} + \delta_t \mathbf{P}^{-1} \Delta \boldsymbol{\sigma}^{(t)}, \quad t = 1, 2, 3, \dots \quad (7.5)$$

where t is the iteration count, $\boldsymbol{\sigma}^{(t)}$ indicates the intensity estimate at iteration t , $\Delta \boldsymbol{\sigma}^{(t)}$ is the refinement to the solution at iteration t , δ_t is an iteration-dependent scaling constant (step size) and \mathbf{P}^{-1} is an arbitrary preconditioner. The term $\delta_t \mathbf{P}^{-1} \Delta \boldsymbol{\sigma}^{(t)}$ indicates a refinement on the previous solution [85].

More precisely, considering the MLE imaging problem formulation, the objective function can be stated as

$$J(\boldsymbol{\sigma}, \tilde{\mathbf{r}}) = \|\boldsymbol{\Gamma}(\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})\|_2^2. \quad (7.6)$$

The cost function is quadratic, i.e.,

$$J(\boldsymbol{\sigma}, \tilde{\mathbf{r}}) = (\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma})^H \mathbf{C}_e^{-1} (\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}). \quad (7.7)$$

The cost function is optimized when $\nabla J(\boldsymbol{\sigma}, \tilde{\mathbf{r}}) = \mathbf{0}$, where $\nabla J(\cdot)$ indicates the gradient of the objective function. The gradient of (7.6) is equal to

$$\nabla J(\boldsymbol{\sigma}, \tilde{\mathbf{r}}) = \mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M}\boldsymbol{\sigma} - \mathbf{M}^H \mathbf{C}_e^{-1} \tilde{\mathbf{r}} = \mathbf{M}^H \mathbf{C}_e^{-1} (\mathbf{M}\boldsymbol{\sigma} - \tilde{\mathbf{r}}). \quad (7.8)$$

At the minimum, the gradient vanishes ($\nabla J(\boldsymbol{\sigma}) = \mathbf{0}$) and we have

$$\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M} \boldsymbol{\sigma} = \mathbf{M}^H \mathbf{C}_e^{-1} \tilde{\mathbf{r}} \quad (7.9)$$

which results in the normal equations for the MLE problem.

The iterative solution method used in many convex optimization problems is based on the gradient descent algorithm. CLEAN can be regarded as a constrained optimization problem and as stated by Carillo et al. [35] and Onose et al. [75], the CLEAN procedure can be interpreted as a regularized gradient descent method. In the gradient descent method the correction term $\Delta \boldsymbol{\sigma}^{(t)}$ in (7.5) is the negative of the gradient function which for the MLE problem (7.6) is computed as

$$\Delta \boldsymbol{\sigma}^{(t)} = -\nabla J(\boldsymbol{\sigma}^{(t)}) = \mathbf{M}^H \mathbf{C}_e^{-1} (\tilde{\mathbf{r}} - \mathbf{M} \boldsymbol{\sigma}^{(t)}) \quad (7.10)$$

$\Delta \boldsymbol{\sigma}^{(t)}$ in the context of radio interferometric imaging is called the residual image and the term $\Delta \mathbf{r} = \tilde{\mathbf{r}} - \mathbf{M} \boldsymbol{\sigma}$ is called the residual. The square norm of the residual is dubbed “discrepancy” [48].

To incorporate regularizing assumptions and constraints into the gradient descent procedure, proximity operators [101] can be exploited. Therefore, most of the traditional RA imaging problems can be identified under the broad category of the gradient descent methods. Incorporating the proximity operator, the gradient descent-based methods in RA can be described in a general form as

$$\boldsymbol{\sigma}^{(t+1)} = \mathcal{P}(\boldsymbol{\sigma}^{(t)} + \mathcal{T}(\mathbf{M}^H \mathbf{C}_e^{-1} (\tilde{\mathbf{r}} - \mathbf{M} \boldsymbol{\sigma}^{(t)}))), \quad t = 1, 2, 3, \dots \quad (7.11)$$

where \mathcal{P} is a projection or proximity operator to take some priors such as positivity of the image into account and \mathcal{T} can be considered as a nonlinear deconvolution operator such as the one used in CLEAN, a step size and a preconditioner $\alpha_k \mathbf{P}^{-1}$ similar to the formalism of Equation (7.5) or merely a step size scalar.

In CLEAN, \mathcal{T} refers to a nonlinear deconvolution operator that finds the occupied pixels and removes a scaled version of the array beam pattern from the identified location to remove the effect of that source from the image [75].

7.2.1. THE CLEAN-BASED IMAGING PIPELINE

In this section, we discuss the working principle of the CLEAN algorithm in more detail. CLEAN is the primary implemented imaging method in the pipeline of many radio telescope arrays, including LOFAR. Therefore, it serves as a reference for the iterative imaging algorithms for RA. We explain the radio interferometric imaging pipeline based on the CLEAN iterations.

CLEAN was initially introduced as a procedural approach. Schwarz [128] analyzed CLEAN mathematically and showed that it is equivalent to solving a system of linear equations by iterative methods based on a least squares fitting of the real and the imaginary part of the visibility function to the measurements. As mentioned earlier, CLEAN has been identified as a greedy sparse MP-type algorithm as well as a method based on gradient descent [35, 75]. CLEAN is a sequential source removal technique in which at each step of the algorithm, the strongest source position and power are estimated and the effect is removed from the residual image until the residual is noise-like. The procedure in the CLEAN algorithm can be described as:

1. Initialization: $\boldsymbol{\sigma}^{(0)} = \mathbf{0}$ or a priori known model image
2. Major cycle (iteration count t): Compute the residual image from the model image

$$\Delta\boldsymbol{\sigma}^{(t)} = \mathbf{M}^H \mathbf{C}_e^{-1} (\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}^{(t)})$$

The major cycle is performed in two steps

- Forward step: computing the model visibility (must be done with high accuracy)

$$\mathbf{r}^{(t)} = \mathbf{M}\boldsymbol{\sigma}^{(t)}$$

- Backward step: Computing the residual image

$$\Delta\boldsymbol{\sigma}^{(t)} = \mathbf{M}^H \mathbf{C}_e^{-1} \Delta\mathbf{r}^{(t)}$$

where $\Delta\mathbf{r}^{(t)} = \tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}^{(t)}$ is the residual visibility at iteration t

3. Minor cycle (iteration count k): finding update to the model image $\boldsymbol{\sigma}^{(t+1)}$ based on $\Delta\boldsymbol{\sigma}^{(t)}$ and $\boldsymbol{\sigma}^{(t)}$

- while $\Delta\boldsymbol{\sigma}^{(t)} > \boldsymbol{\epsilon}$ ($\boldsymbol{\epsilon}$ is the noise level threshold)
 - $j = \operatorname{argmax}(\Delta\boldsymbol{\sigma}^{(t)})$
 - Update the corresponding pixel j in the model image

$$\sigma_j^{(t),(k+1)} = \sigma_j^{(t),(k)} + \alpha_k \Delta\sigma_j^{(k)}$$

- Update the residual image

$$\Delta\boldsymbol{\sigma}^{(t),(k+1)} = \Delta\boldsymbol{\sigma}^{(t),(k)} - [\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M}]_j \alpha_k \Delta\sigma_j^{(k)}$$

where $[\mathbf{M}^H \mathbf{C}_e^{-1} \mathbf{M}]_j$ is the j -th column of the Hessian which is actually the PSF for pixel j .

4. After a number of minor cycles, n , update the model image and return to the major cycle

$$\boldsymbol{\sigma}^{(t+1)} = \boldsymbol{\sigma}^{(t),(n)}.$$

We can identify two iteration levels for CLEAN: (i) an outer loop to iteratively find the sparse support of the image and (ii) an inner loop in which the dimension-reduced version of the LS problem is solved. The inner loop works purely in the image domain and performs the deconvolution of the PSF, explained in Chapter 3, Section 3.1.4, from the residual image and the outer loop validates the estimated pixel values by fitting the obtained values in the model to the visibility measurements.

The major cycle corresponds to going back and forth between the visibility domain and the image domain (Forward and backward steps). This step is the most costly step of the CLEAN algorithm that requires two matrix vector multiplications with \mathbf{M} and \mathbf{M}^H . These multiplications are implemented based on convolutional gridding and FFT. Assuming $\mathbf{G} \in \mathbb{C}^{Q \times P^2}$ represents the precomputed interpolation and resampling matrix, the

- Update all the pixels in the set \mathcal{F} by solving

$$\boldsymbol{\sigma}_{\mathcal{F}}^{(t+1)} = \underset{\boldsymbol{\sigma}_{\mathcal{F}}}{\operatorname{argmin}} \|\Gamma(\tilde{\mathbf{r}} - \mathbf{M}_{\mathcal{F}} \boldsymbol{\sigma}_{\mathcal{F}}^{(t)})\|_2^2 \quad (7.12)$$

for example with Krylov methods.

4. If some of the elements in $\boldsymbol{\sigma}_{\mathcal{F}}^{(t+1)}$ do not satisfy the constraints additional steps are required.
5. Otherwise go to step 2 to recompute the residual image.

In short, the active-set algorithm for NNLS [66] iteratively reduces the residual image, which is the negative of the gradient vector, until it converges to the noise threshold vector of the image. This algorithm is based on the definition of free set \mathcal{F} and active set \mathcal{A} . At the start, all the pixels of the image are in the active set (the black background). In each iteration, the residual image is computed based on the current estimate of the image, $\boldsymbol{\sigma}^{(t)}$, and the pixel corresponding to the maximum of the residual image is removed from the active set and placed in the free set. Then a reduced MLE problem, Equation (7.12), consisting only of the pixels of the image that are in the free set is solved to find an estimate of the model image with the pixels in the free set ($\mathbf{M}_{\mathcal{F}}$ only contains the columns of \mathbf{M} corresponding to the pixels in the free set). If some of the pixels in the model image do not satisfy the positivity constraints, projection steps are performed to ensure the positivity of the pixels. The NNLS processing pipeline is shown in Figure 7.2.

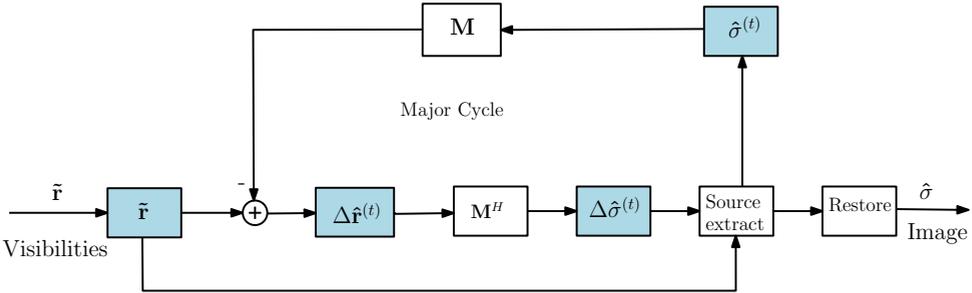


Figure 7.2: The NNLS pipeline

Based on this figure, we can see the procedural similarity of CLEAN and the active-set algorithm where the main difference is in the source extraction procedure. In the active-set method, the minor cycle is replaced by a solving an MLE problem of reduced order. Sardarabadi et al. [56] have proposed to solve the reduced MLE problem at each step by a Krylov-subspace solution method.

By iteratively selecting the columns of \mathbf{M} , with the underlying assumption of the sparsity of the sources in a black background, active-set performs a sparsity-promoting regularization. This is since, for a sufficiently sparse image, most of the columns of \mathbf{M} correspond to zero pixel values and remain in the active set. Furthermore, the gradient calculation and maximum selection is similar to performing a matched-filtering

on the residual vector $\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\sigma}^{(t)}$ at the t th iteration. The authors in [129] have investigated the similarity of the active-set algorithm with the orthogonal matching pursuit algorithm [130] in terms of sparse recovery. It was shown that the NNLS algorithm obtains good position and intensity estimates in scenarios where the underlying image is composed of a sparse set of point sources [42, 56]. However, the convergence rate and the intensity estimates are adversely affected in scenarios where extended emissions are present.

7.3. DUAL-BASIS NNLS

In order to tackle the problems of the active-set method where point sources and extended emissions co-exist in an image, we propose to represent the image with a dual-basis dictionary as

$$\boldsymbol{\sigma} = \boldsymbol{\Theta}\boldsymbol{\alpha}, \text{ where } \boldsymbol{\Theta} = [\mathbf{I}, \mathbf{B}]. \quad (7.13)$$

In this formulation, $\boldsymbol{\Theta}$ is a $Q \times 2Q$ matrix that is composed of two $Q \times Q$ matrices \mathbf{I} and \mathbf{B} . \mathbf{I} is the identity matrix that represents pixel basis for the recovery of point sources. We choose this basis to retain the super-resolution point source recovery feature of the NNLS algorithm. We propose to capture the extended emissions with the actual resolution of the antenna array. For this purpose, we choose a Gaussian function with the same Full Width at Half Maximum (FWHM) as the Half Power Beam Width (HPBW) of the antenna array. Each column of \mathbf{B} is a shifted version of the aforementioned Gaussian function with the peak shifted to the location of the corresponding pixel and normalized such that the total underlying energy is 1.

With the choice of a positive basis matrix $\boldsymbol{\Theta}$, we restrict our attention to the positive coefficient set $\boldsymbol{\alpha}$. Therefore, we can reformulate the image reconstruction problem as

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \|\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\Theta}\boldsymbol{\alpha}\|_2^2 + \tau \|\boldsymbol{\alpha}\|_0 \quad \text{subject to } \boldsymbol{\alpha} \geq \mathbf{0}, \quad (7.14)$$

and correspondingly the image estimate is obtained as $\hat{\boldsymbol{\sigma}} = \boldsymbol{\Theta}\hat{\boldsymbol{\alpha}}$. We see that in this case the basis matrix $\boldsymbol{\Theta}$ is the right preconditioner. The prior information about the image that is incorporated in this problem is: the radio-astronomical image is positive and is composed of both point sources and extended emissions with a substantial black background. We apply the modified active-set algorithm to the aforementioned problem to recover the basis coefficients $\boldsymbol{\alpha}$. The proposed algorithm works on the implicit assumption of the sparsity of the signal over the predefined basis to obtain a unique solution. We call this new method the dual-basis non-negative least squares as explained in Algorithm 3.

As illustrated in Algorithm 3, we begin with an empty sky (zero basis coefficients). Therefore, all of the indices are in the active set \mathcal{A} and the free set is empty. The initial residual basis coefficient vector is computed in step 2 as $\Delta\boldsymbol{\alpha} = \boldsymbol{\Theta}^H \mathbf{M}^H (\tilde{\mathbf{r}} - \mathbf{M}\boldsymbol{\Theta}\boldsymbol{\alpha})$.

In fact $\Delta\boldsymbol{\alpha}$ represents the matched filter output power projected in the basis coefficient space. The stopping criteria for the algorithm as explained in step 3 are when there are no more indices in the active set and when all of the indices in the residual coefficient vector $\Delta\boldsymbol{\alpha}$ have reached the corresponding detection threshold. In order to choose the pixel-based detection threshold ϵ , based on Equation (2.30), when all the sources are

Algorithm 3: Dual-basis non-negative least squares

input : $\bar{\mathbf{r}}, \mathbf{M}, \Theta, \epsilon$
output: Basis coefficient vector α

- 1 Initialize: $\mathcal{F} = \emptyset$, $\mathcal{A} = \{1, 2, 3, \dots, 2Q\}$ and $\alpha = \mathbf{0}$;
- 2 Compute the initial residual basis coefficient vector $\Delta\alpha = \Theta^H \mathbf{M}^H (\bar{\mathbf{r}} - \mathbf{M}\Theta\alpha)$;
- 3 **while** $\mathcal{A} \neq \emptyset$ and $\Delta\alpha_t > \epsilon_t$ for any $t \in \mathcal{A}$ **do**
- 4 $m = \underset{t}{\operatorname{argmax}} \{\Delta\alpha_t \mid t \in \mathcal{A}\}$;
- 5 add m to \mathcal{F} and remove from \mathcal{A} ;
- 6 Solve for the coefficients in the set \mathcal{F} $\beta_{\mathcal{F}} = \underset{\alpha_{\mathcal{F}}}{\operatorname{argmin}} \|\bar{\mathbf{r}} - \mathbf{M}\Theta_{\mathcal{F}}\alpha_{\mathcal{F}}\|_2^2$;
- 7 Define $\beta_n := 0$ for $n \in \mathcal{A}$;
- 8 **if** $\beta_u \leq \epsilon_u$ for any $u \in \mathcal{F}$ **then**
- 9 **while** $\beta_u \leq \epsilon_u$ for any $u \in \mathcal{F}$ **do**
- 10 $h = \operatorname{argmin} \frac{\alpha_u}{\alpha_u - \beta_u}$ for $u \in \mathcal{F}$, $\beta_u \leq \epsilon_u$;
- 11 Set $\delta = \frac{\alpha_h}{\alpha_h - \beta_h}$;
- 12 Set $\alpha = \alpha + \delta(\beta - \alpha)$;
- 13 Move from \mathcal{F} to \mathcal{A} all indexes s for which $|\alpha_s| < \epsilon_s$;
- 14 Repeat steps 6 and 7;
- 15 **end**
- 16 **else**
- 17 Set $\alpha = \beta$;
- 18 Compute $\Delta\alpha = \Theta^H \mathbf{M}^H (\bar{\mathbf{r}} - \mathbf{M}\Theta\alpha)$
- 19 **end**
- 20 **end**

estimated the covariance of the residual basis coefficient vector is

$$\text{Cov}(\Delta\alpha) = \frac{1}{N} \Theta^H \mathbf{M}^H (\mathbf{R}^* \otimes \mathbf{R}) \mathbf{M} \Theta. \quad (7.15)$$

To set the detection threshold, we compute an estimate of the standard deviation of the residual basis coefficient vector based on the available data $\hat{\mathbf{R}}$ and allow for a false alarm rate of 0.01 %. Therefore, for the $2Q \times 1$ vector of index-based detection threshold we have

$$\epsilon = 6 \times \frac{1}{\sqrt{N}} (\text{vectdiag}(\Theta^H \mathbf{M}^H (\hat{\mathbf{R}}^* \otimes \hat{\mathbf{R}}) \mathbf{M} \Theta))^{\circ \frac{1}{2}}. \quad (7.16)$$

At step 4 of the algorithm, the coefficient corresponding to the maximum output power of the matched filter projected in the coefficient space is computed. This is the step that explains the key idea of the proposed method. The maximum not only shows the location (Direction of Arrival, DoA) of the potential source but it also provides the most likely shape of the source by choosing the coefficient in the basis that maximizes the output power. This way, we can distinguish between point sources and extended sources and estimate the source power in the next steps based on the chosen basis as done in step 6. The regularized minimization problem in step 6 is solved by an iterative algorithm such as LSQR [104]. Steps 8 to 14 check if all of the computed coefficients in the free set satisfy the corresponding bound. If the bound is violated for some of the coefficients, the maximum step size, δ , towards the bound is found and the coefficients and the sets are updated accordingly in step 11 and 12. Afterwards, the minimization in step 6 is recomputed until all of the coefficients in the free set satisfy the bound, after which the residual basis coefficients are recomputed in step 12 and the iteration is continued until convergence is reached. The imaging pipeline of dual-basis NNLS is shown in Figure 7.3.

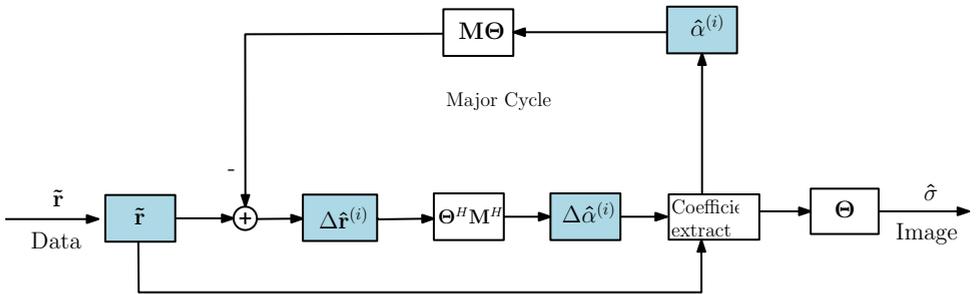


Figure 7.3: Dual-basis NNLS imaging pipeline

7.3.1. NUMERICAL RESULTS AND PERFORMANCE ANALYSIS

We consider a one-dimensional test example to evaluate the performance of the proposed algorithm. A random non-uniform linear array with $P = 30$ antenna elements is considered. The FoV of the array in terms of the θ angle is in the range $(-90^\circ, 90^\circ)$. The wavelength of the radio frequency is chosen to be $\lambda = 2$ m. Assuming the placement

of the antenna elements on the x axis and indicating the position of the p th antenna element as x_p , the beam pattern of the antenna array is calculated as

$$f(l_q) = \frac{1}{P} \sum_{m=1}^P \sum_{p=1}^P e^{-j \frac{2\pi}{\lambda} (x_m - x_p) l_q}. \quad (7.17)$$

The maximum of the beam pattern is P and the half power beam width of the array is approximately $\text{HPBW} \approx \frac{\lambda}{\Delta x_{max}}$ radians [63] where Δx_{max} is the maximum distance between all of the antenna pairs. This indicates the resolution of the system. The image resolution is chosen as $\Delta\theta = 0.1 \times \text{HPBW} \approx 0.4^\circ$; Therefore, the number of image pixels is $Q = 429$. The direction cosine on the sky for the q th pixel is computed as $l_q = \sin\theta_q$ [8]. We fit a Gaussian function to the main beam of the array. The Gaussian function with the same maximum amplitude as the beam pattern function centered at zero is

$$g(l_q) = P \exp\left(-\frac{l_q^2}{2\rho^2}\right), \quad q = 1, 2, 3, \dots, Q. \quad (7.18)$$

The FWHM for the Gaussian function happens at $2\rho\sqrt{2\ln 2}$. Setting $\lambda/\Delta x_{max} = 2\rho\sqrt{2\ln 2}$, the width of the Gaussian to fit the main beam of the array is $\rho = \frac{\lambda}{2\Delta x_{max}\sqrt{2\ln 2}}$. An instance of the non-uniform linear array and the corresponding array beam pattern and the Gaussian function fit to its main beam is shown in Figure 7.4. We then normalize this Gaussian function such that the contained energy over all the discrete points is summed to 1. The array beam pattern and the Gaussian fit are shown in Figure 7.4(a).

To investigate the performance of the proposed algorithm, we apply the algorithm on the aforementioned non-uniform linear array. The underlying source intensities and the matched-filtered dirty image are shown in Figure 7.4(b) and (c). The sources are composed of two point sources with the intensities 4 and 3 placed at the angles 8.8° and 35.6° respectively and two Gaussian sources with the peak intensities 5 and 1 with different widths placed respectively at the angles -54.1° -30.6° . To construct the sampled covariance matrix, Gaussian receiver noise with variance $\sigma_n = 0.5$ is added to the covariance matrix \mathbf{R} and $N = 10^5$ data samples are used to construct the sample covariance matrix $\hat{\mathbf{R}}$. The image obtained by applying the NNLS algorithm is shown in Figure 7.4(d). As can be seen, the extended sources are approximated by a large number of point sources and the intensity estimates are much larger than the actual intensities. Figure 7.4(e) shows the result of convolving the result of NNLS with the normalized Gaussian beam times the pixel width. This post processing approximately retains the shape of the extended sources at the expense of reducing the resolution of point sources; Furthermore, the peak intensity values are not correctly restored, i.e. the responses of the point sources have been broadened resulting in a lower peak and the responses of the extended sources have been narrowed resulting in a higher peak. However, the integrated power over the source response in this case remains a correct representation of the source powers. Figure 7.4(f) shows the result of applying the proposed dual-basis NNLS algorithm. As can be seen, both the extended emissions and the point sources are estimated correctly with super-resolution estimate of the point sources and the peak and the shape of the intensity estimates are very close to the true intensities.

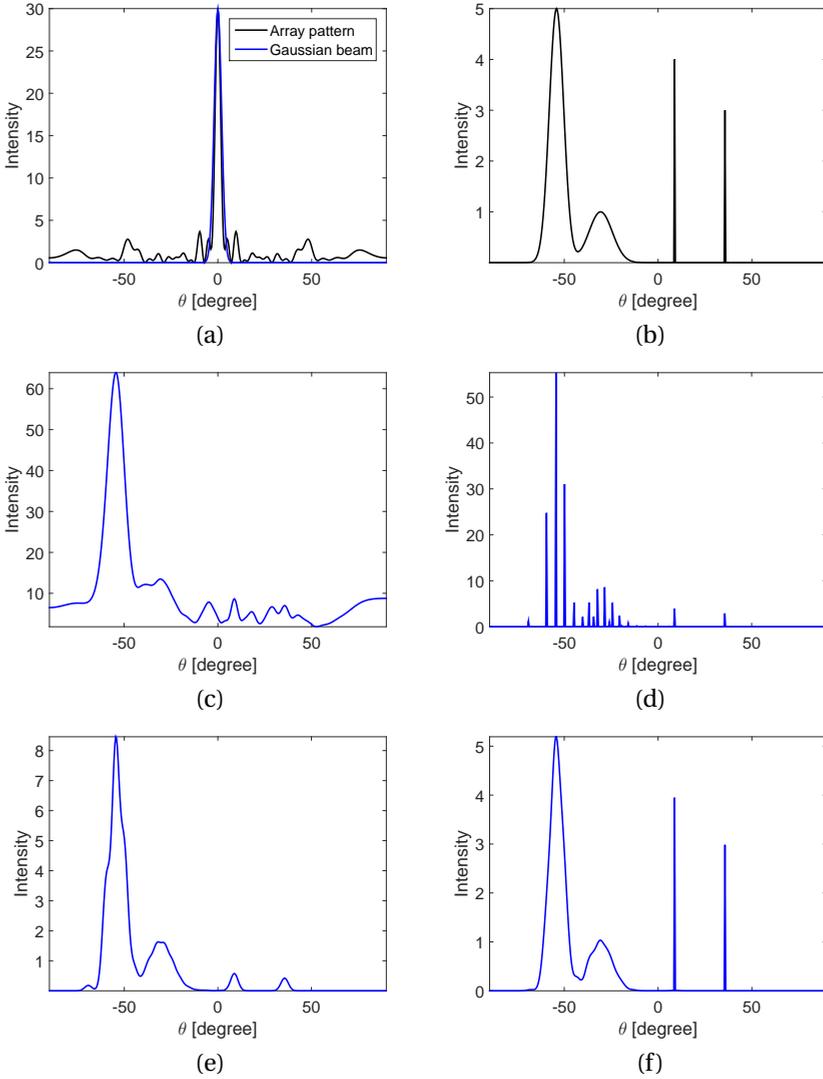


Figure 7.4: (a) Array pattern and the Gaussian fit, (b) The underlying source intensity distribution (c) Matched-filter dirty image, (d) Recovered image based on the NNLS algorithm, (e) Gaussian beam applied on the result of NNLS and (f) Recovered image based on the dual-basis NNLS algorithm.

Number of antennas	20	30	40	50
Number of pixels	303	429	611	757
Average computation time (seconds)				
NNLS	0.53	0.67	1.47	4
Dual-basis NNLS	0.34	0.69	1.46	3.25
Average number of iterations				
NNLS	20.73	21.68	27.5	39.55
Dual-basis NNLS	15.49	20.26	24.05	31
Average number of nonzero basis coefficients				
NNLS	19.73	20.52	26.5	37.57
Dual-basis NNLS	10.69	18.39	22.05	28.5
Average error norm				
NNLS	54.36	66.5	79.92	72.83
Post-processed NNLS	21.74	14.36	7.4	5.1
Dual-basis NNLS	1.08	0.75	1.49	1.42

Table 7.1: Performance analysis

The performance of the dual-basis NNLS algorithm was compared to the performance of the NNLS algorithm and the post-processed version of NNLS by running the algorithms for 100 noise instances for random linear arrays with different number of elements. The pixel resolution was kept at 0.1 of the main beam of the array. Table 7.1 displays the number of antenna elements, the associated number of pixels, the algorithm computation time in seconds, the number of iterations until convergence to the threshold, the number of non-zero elements needed to represent the sources and the average error norm $\|\sigma - \hat{\sigma}\|_2 / 100$. Based on the results of the experiments, we conclude that by applying the dual-basis NNLS algorithm we are able to capture the actual intensity of both point sources and extended emissions with a small number of basis elements with high accuracy. Furthermore, the number of iterations and computing time are decreased with respect to NNLS.

Generally speaking, synthesis-based greedy sparse recovery methods face prohibitive challenges when employed for large and high resolution image recovery problems. For large images and overcomplete dictionaries, the coefficient vector α becomes very large. The greedy methods search through the complete coefficient vector to find its sparse support. Therefore, their complexity increases with the increase in the size of the coefficient vector. Furthermore, the solution of the problem is highly dictionary-dependent and if the chosen dictionary is not well-representative of the underlying image, these methods provide erroneous results. This problem is alleviated by choosing a highly redundant dictionary, which in turn increases the computational complexity. Furthermore, since in each iteration the complete dictionary is used to compute the residual basis coefficient vector $\Delta\alpha$, dictionaries for which a fast transform exists must be employed to keep the computational complexity manageable.

7.4. MULTI-BASIS GENERALIZATION OF IRO-PRIFIRA

As we have shown in Chapter 5, iteratively reweighted schemes, namely IRO-PRIFIRA, can also be used to estimate a minimum ℓ_0 norm solution, i.e. solve problem (7.2). However, IRO-PRIFIRA is not reliable in estimating images in which extended emissions and point-sources coexist as it aggregates the diffused power in a few pixels trying to achieve a sparse representation. This is due to the fact that the ℓ_0 norm by definition only regularizes the support of the image and is not reliable for recovering intensity estimates. An interesting feature of IRO-PRIFIRA in comparison with the greedy sparse methods is that it is not a pixel-based approach. Therefore it is resilient to pixel mismatch and linear dependence of the number of iterations to the number of pixels. We show here that imposing the proposed simple dictionary in Section 7.3 significantly increases the fidelity of IRO-PRIFIRA for one-dimensional problems where point sources and extended emissions co-exist.

Similarly as before, we assume that we can model the sky map via an overcomplete dictionary Θ as $\sigma = \Theta\alpha$ such that the coefficient α is sparse. We propose to solve the synthesis-based sparse recovery problem (7.2), i.e.

$$\hat{\alpha} = \arg \min_{\alpha} \|\Gamma(\tilde{\mathbf{r}} - \mathbf{M}\Theta\alpha)\|_2^2 + \tau \|\alpha\|_p, \quad (7.19)$$

where $p = 0$, using the iteratively-reweighted schemes proposed in Chapter 5, Section 5.2.3. More specifically, we apply IRO-PRIFIRA to Problem (7.19) to obtain an estimate $\hat{\alpha}$. The image estimate $\hat{\sigma}$ can be obtained as $\hat{\sigma} = \Theta\hat{\alpha}$.

As discussed in Chapter 5, Section 5.2.3, the iteratively reweighted schemes solve Problem (7.19) by replacing the ℓ_p -norm regularization by an ℓ_2 -norm regularization. This is done by introducing outer iterations, indexed by k , using the transform $\beta_k = \mathbf{W}_k\alpha_k$ as

$$\hat{\beta}_k = \arg \min_{\beta} \|\Gamma(\tilde{\mathbf{r}} - \mathbf{M}\Theta\mathbf{W}_k^{-1}\beta)\|_2^2 + \tau \|\beta\|_2^2, \quad (7.20)$$

where the weight matrix \mathbf{W}_k is defined as $\mathbf{W}_k = \text{diag}(\alpha_{k-1}^{(p-2)/2})$. This formalism presents a generalization of the concept of right preconditioning to provide an alternative to greedy methods for synthesis-based sparse image recovery.

7.5. SYNTHESIS-BASED SPARSE IMAGING WITH CONVEX OPTIMIZATION

So far we have discussed solution methods for the synthesis-based sparse recovery problem based on (i) greedy sparse recovery methods, namely CLEAN and NNLS and (ii) iteratively reweighted schemes that generalize the numerical recovery methods by an iterated application of an reweighted right preconditioner. Additionally, if sparsity is imposed by an ℓ_1 norm, i.e. $p = 1$ in Problem (7.19), convex optimization can also be applied to find a solution to the synthesis-based sparse recovery problem (7.19). More specifically, Problem (7.19) is composed of a convex and differentiable function (the data fidelity term) and a convex and non-differentiable function (the ℓ_1 regularization term). It is well known that proximal gradient methods are well-suited to these problems. They provide operator splitting for composite functions by incorporating the non-

differentiable function in the gradient descent by its proximal operator similar to Equation (7.11). The proximity operator of the ℓ_1 -norm regularization function is the soft-thresholding (shrinkage) operator. The soft thresholding operator for the vector \mathbf{u} is defined as

$$\mathcal{P}_\tau(\mathbf{u}) = \begin{cases} -\mathbf{u} + \tau & \text{if } \mathbf{u} < -\tau \\ 0 & \text{if } -\tau < \mathbf{u} < \tau \\ \mathbf{u} - \tau & \text{otherwise.} \end{cases} \quad (7.21)$$

The Fast Iterative Shrinkage- Thresholding Algorithm (FISTA) [131] is a well-known algorithm to address ℓ_1 -regularized sparse recovery problems. FISTA has been generalized to address synthesis-based sparse recovery problems similar to Problem (7.19). The synthesis-based formalism of FISTA using a starlet dictionary has been considered for imaging with LOFAR in [132]. FISTA solves the sparse recovery problem by performing accelerated proximal gradient descent steps as described in Algorithm 4. In this algo-

Algorithm 4: Synthesis-based sparse recovery FISTA

input : $\tilde{\mathbf{r}}, \mathbf{M}, \Theta, \delta, K$

output: Image estimate $\hat{\sigma}$

1 Initialize:

2 $\alpha_0 = \mathbf{0}, \mathbf{z}_1 = \alpha_0, t_1 = 1;$

3 **for** $k = 1, 2, \dots, K$ **do**

4 $\alpha_k = \mathcal{P}_\tau(\mathbf{z}_k - \delta \Theta^H \mathbf{M}^H (\mathbf{M} \Theta \mathbf{z}_k - \tilde{\mathbf{r}}));$

5 $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2};$

6 $\mathbf{z}_{k+1} = \alpha_k + (\frac{t_k - 1}{t_{k+1}})(\alpha_k - \alpha_{k-1}),$

7 **end**

8 $\hat{\sigma} = \Theta \alpha_K$

rithm, \mathbf{z}_k is an auxiliary variable of the same size as the coefficient vector α and δ is the gradient descent step size which is computed based on the Lipschitz constant of the composite operator $\mathbf{M}\Theta$ such that $0 < \delta < \frac{1}{\|\Theta^H \mathbf{M}^H \mathbf{M} \Theta\|}$ to ensure convergence. As can be inferred from Algorithm 4, the obtained solution with FISTA depends on the choice of τ .

7.6. ANALYSIS VERSUS SYNTHESIS SPARSE RECOVERY

Another method to impose sparsity in the transform domain for images that are not naturally sparse is via the analysis-based formulation of the imaging problem, i.e.

$$\hat{\sigma} = \arg \min_{\sigma} \|\Gamma(\tilde{\mathbf{r}} - \mathbf{M}\sigma)\|_2^2 + \tau \|\Omega\sigma\|_p, \quad (7.22)$$

assuming $p = 0$ or 1 . The matrix Ω in this case is called the analysis operator. In this formulation, the image is considered sparse in the transform domain $\Omega\sigma$ and the problem is directly solved for σ (rather than the coefficient vector α). Therefore, in this formulation, in contrast to the synthesis-based formulation, the images σ is not restricted to be a linear combination of the atoms in the dictionary matrix. Note that for redundant

dictionaries there is no equivalence between Problems (7.19) and (7.22). Furthermore, for overcomplete, and thus non-invertible, dictionaries there is no stable way to transform the analysis-based formulation (7.22) to an iteratively reweighted scheme similar to (7.20) without the need to invert $\mathbf{\Omega}$.

If the analysis operator is tight frame, i.e. $\mathbf{\Omega}^H \mathbf{\Omega} = \mathbf{I}$, FISTA can be applied on the analysis-based sparse recovery problem. In this case, the soft thresholding operator is replaced by the projected soft thresholding operator, i.e. $\mathbf{\Omega}^H \mathcal{P}_r(\mathbf{\Omega} \mathbf{u})$ [133] and the solution is directly found for the image $\boldsymbol{\sigma}$ (rather than the basis coefficients $\boldsymbol{\alpha}$). However, if the analysis operator is not tight frame, FISTA cannot be easily implemented to find the solution of the analysis sparse recovery problem.

Elad et al. [125] have compared the analysis and synthesis-based ℓ_1 -sparse recovery problems theoretically. They have pointed out that despite the similar structure of the two problems, there are multiple differences; More specifically (i) in the synthesis-based approach, the image $\boldsymbol{\sigma}$ is represented as a sparse linear combination, given by the coefficient vector $\boldsymbol{\alpha}$, of the elements of the dictionary matrix $\mathbf{\Theta}$, i.e. $\boldsymbol{\sigma} = \mathbf{\Theta} \boldsymbol{\alpha}$. This is while the analysis-based approach provides a sparsification of the forward projection of the image on the basis elements of matrix $\mathbf{\Omega}$. Therefore, increasing the redundancy in the dictionary matrix $\mathbf{\Theta}$ in general results in a better representation of the image. However, since the image has to agree with all the rows of $\mathbf{\Omega}$ simultaneously, increased redundancy reduces the usefulness of the analysis operator. (ii) Moreover, if a small dictionary matrix $\mathbf{\Theta}$ with insufficient basis functions is used, significance of each individual basis function increases. In this case, a wrong choice of basis by the algorithm leads to a sequential erroneous selection of bases to compensate for the error. On the other hand, in the analysis-based method, all the bases contribute evenly to describing the image. Therefore, the individual dependence is reduced which in turn leads to a more stable solution. (iii) Furthermore, the analysis-based optimization problems, for overcomplete dictionaries, are easier to solve due to the smaller dimension of the unknowns. (iv) On the other hand, the synthesis-based formulations are more intuitive and provide a versatile structure, e.g. can be easily incorporated in the iteratively reweighted schemes.

7.7. SIMULATION RESULTS

7.7.1. ONE-DIMENSIONAL SIMULATIONS

For the one-dimensional simulations, we choose the simple dictionary $\mathbf{\Theta} = [\mathbf{I}, \mathbf{B}]$ similar to Section 7.3, i.e., \mathbf{I} is the identity matrix, pixel basis, to model the point sources and \mathbf{B} is the normalized Gaussian clean beam basis matrix. We show the effect of applying the overcomplete dictionary on IR0-PRIFIRA based on a one-dimensional test with a point source with amplitude 4 and a Gaussian source with amplitude 2, modeling an extended emission as shown in Figure 7.5(a). We take as before $P = 10$ antennas with a random linear placement as shown in Figure 7.5(b). The number of pixels in the image is $Q = 201$ as with the previous simulations. Gaussian receiver noise with variance $\sigma_n^2 = 100$ is added to the measurements. The beam pattern and the clean beam used in defining the dictionary are superimposed in Figure 7.5(c). The MF and MVDR dirty images are shown in Figure 7.5(d). The estimated basis coefficients $\hat{\boldsymbol{\alpha}}$ and the reconstructed image based on the multi-basis version of IR0-PRIFIRA superimposed with the image are

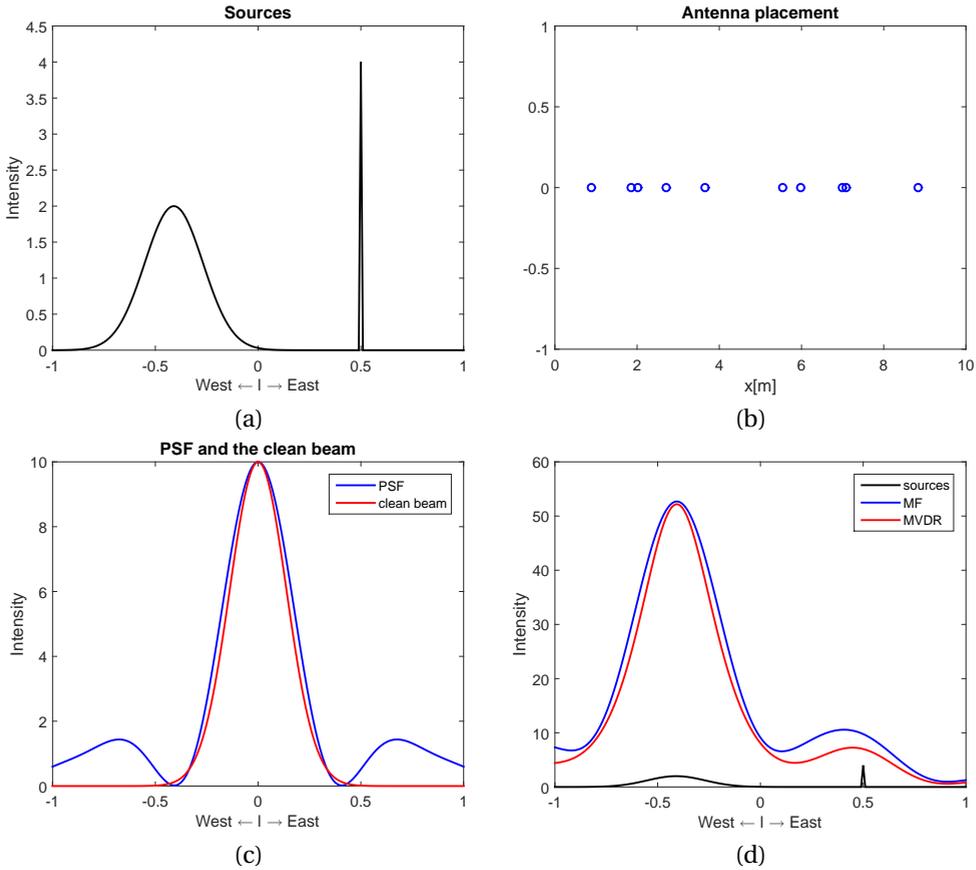


Figure 7.5: (a) Sources, (b) antenna positions, (c) PSF and the clean beam, (d) MF and MVDR dirty images

shown in Figure 7.6(a) and (b), respectively. For this estimate 20 outer iterations of IRO-PRIFIRA are performed with a total number of iterations (outer and inner) amounting to 100 iterations. For comparison, the estimates obtained by FISTA (without applying a dictionary) and dual-basis synthesis FISTA with the aforementioned dual-basis dictionary are provided in Figures 7.6 (c) and (d), respectively. The total number of iterations for FISTA-based recoveries are kept at 100 iterations and a $\tau = 10^{-3}$ is chosen. These simulation results show that dual-basis IRO-PRIFIRA, with a choice of dictionary that is well-representative (models the underlying image very well) of the true image, provides an accurate estimate of the original image. Furthermore, we see that simultaneous recovery of the point and extended source becomes hazardous with FISTA. More specifically, while the extended source is recovered well the point source is mainly discarded. This issue has been addressed to some extent by introducing a dual-basis with synthesis-based FISTA which yields a more balanced recovery. However, full recovery of the point source requires a large number of iterations due to the approximation of the ℓ_0 norm with the ℓ_1 norm which in turn hampers the recovery of the extended source.

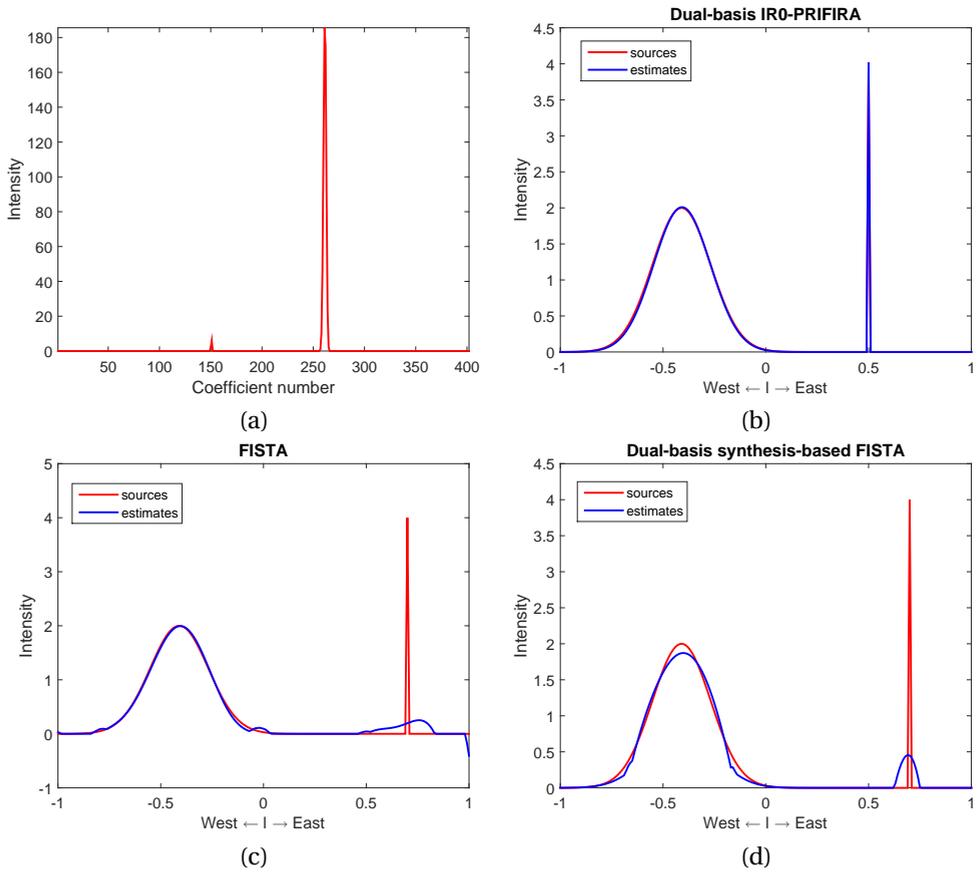


Figure 7.6: (a) Estimated basis coefficients, (b) dual-basis IR0-PRIFIRA estimates, (c) FISTA estimates (d) dual-basis synthesis-based FISTA estimates

7.7.2. TWO-DIMENSIONAL SIMULATIONS

The proposed dictionary in the one-dimensional simulations provides a very good representation of the underlying image and in that sense is rather simplistic. To show the application of the proposed methods to radio astronomy, we show the performance of a multi-basis extension of IRX-PRIFIRA ($X = 0 \text{ or } 1$) based on the realistic two-dimensional simulation setting described in Chapter 6, Section 6.2.

In general, it is hard to find a dictionary that is well-representative of the underlying image. Especially if the underlying image is not known a priori. The general observation in radio astronomical images is that they contain point sources and extended emissions of different scales. These scales, correspond to different spatial frequency contents in the image. In order to effectively restore high and low frequency content of the image at the same time, the wavelet transform of multiple scales can be used. The wavelet transform is a mathematical tool to decompose a signal into different frequency bands and it has been used in radio astronomy for image reconstruction [74, 134, 36]. Following [36], we use 8 Daubechies wavelet bases, with 2 levels each, and a Dirac basis (for point sources) as the multi-basis dictionary. Furthermore, there exist fast implementations for applications of the wavelet-basis on the signals. Since in this case the chosen dictionary satisfies the tight frame condition, an extension of FISTA based on projected soft thresholding can be implemented to solve the analysis-based sparse recovery problem (Equation (7.22)). We compare the results of the multi-basis synthesis-based sparse recovery using the multi-wavelet dictionary implemented by IR0-PRIFIRA, IR1-PRIFIRA and FISTA in Figures 7.7(a), (b) and (c), respectively. Furthermore, the result of applying an analysis-based formalism with the multi-wavelet analysis operator implemented by FISTA is shown in Figure 7.7. For reference, the FISTA recovery (without bases) and analysis-based FISTA with multi-wavelet dictionaries of 4 levels each are shown in Figures 7.7(e) and (f). The reference reconstructions using basis-free IR0- and IR1-PRIFIRA can be obtained from Chapter 6, Section 6.2. For IR0- and IR1-PRIFIRA, 2 outer iterations are applied (amounting to a total of 34 iterations) and the FISTA implementations are stopped after 100 iterations. For comparison reasons, the scales of all the reconstructed image are restricted to the range 1 and $10^{-3.5}$.

As can be inferred from the Figure, in the synthesis-based formulation, IR0-PRIFIRA and FISTA yield similar results where the weaker extended emissions are not recovered reliably. This is while IR1-PRIFIRA recovers the weaker extended emissions to some extent. The analysis-based FISTA also mostly ignores the extended emissions when wavelets with two levels are used but recovers the edges of sharper features more reliably. This effect is reduced when using 4-level wavelet bases. However, the sharper features of the image are smoothed in the latter case. We note that using 4 levels for the wavelet-bases in the synthesis formalism results in more pronounced hazardous effects close to the edges of the sharper features. This effect is due to the fact that in the synthesis-based formalism, the image is modeled as a linear combination of the dictionary elements and in this case the wavelet dictionary has difficulty modeling the edges of the sharper features such that the corresponding coefficients are sufficiently sparse. In fact, for reliable reconstruction, the synthesis-based formulation requires a highly redundant dictionary that can model all the features of the image accurately.

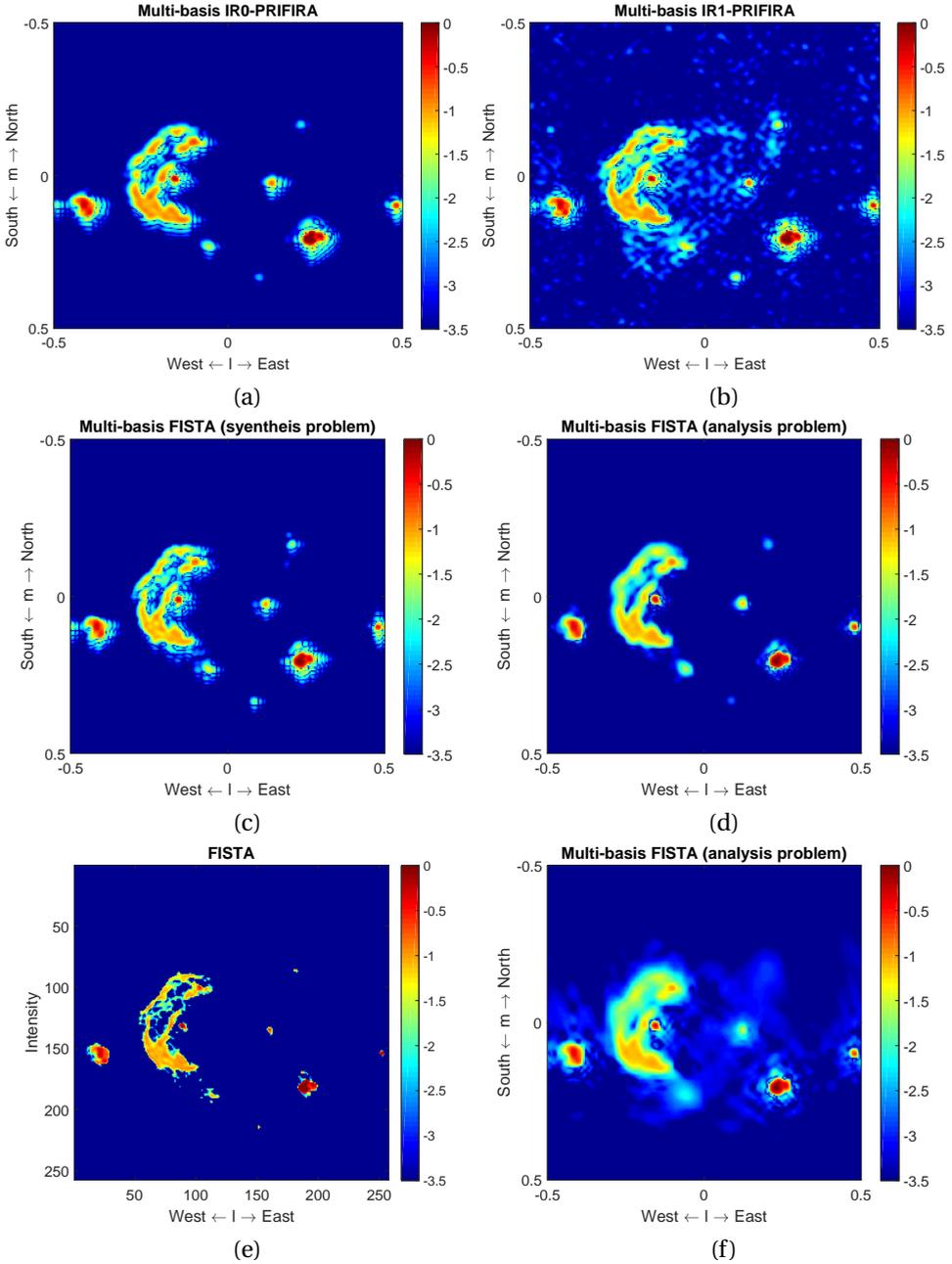


Figure 7.7: Reconstruction results (a) multi-basis IR0-PRIFIRA, (b) multi-basis IR1-PRIFIRA, (c) multi-basis FISTA (synthesis problem), (d) multi-basis FISTA (analysis problem), (e) FISTA (without bases) and (f) multi-basis FISTA (analysis problem) with wavelets with 4 levels

7.8. CONCLUSIONS

In this chapter, we have proposed a problem formulation to incorporate sparse recovery of resolved and unresolved sources based on the preconditioning concept described in Chapter 3. We have investigated greedy sparse recovery methods for the solution of the corresponding problem. Furthermore, we have proposed a generalization for the IR0-PRIFIRA scheme, presented in Chapter 5, to incorporate source representation dictionaries. We have tested both algorithms on one-dimensional examples in which sparse and extended emissions co-exist. For the one-dimensional tests, we have used a simple dictionary composed of Dirac delta functions to represent point sources and a Gaussian function approximated from the main beam of the antenna array to capture the extended emissions. We see that both methods have resulted in a super-resolution (finer resolution than prescribed by the main beam of the antenna array pattern) estimate for the point sources and reliable recovery for the extended emissions.

We have presented a distinction between analysis-based and synthesis-based sparse recovery problems and identified different classes of algorithms that can be used to implement each formalism. To show the application of multi-basis generalization of the sparse recovery problem to radio astronomy, we have tested the performance of the IR0- and IR1-PRIFIRA on a two-dimensional example by exploiting 8 Daubechies wavelets with different scales in combination with a point source basis. We have compared the results with multi-basis extensions of the FISTA algorithm (for synthesis as well as analysis problem). We concluded that the synthesis-based IR0-PRIFIRA and FISTA attain similar results which are not satisfactory since the chosen dictionary cannot sparsely model the underlying image (especially at the edges of the sharp image features). These effects are reduced in the synthesis-based IR1-PRIFIRA implementation. Furthermore, the results of the analysis-based FISTA implementations suggest that non-exact bases representations can be handled more constructively with an analysis formalism. Apart from the importance of the right choice of the dictionary, the synthesis-based formalism is prohibitive in large-scale problems due to the large dimension of the of overcomplete basis coefficients.

In this chapter we have introduced the multi-basis generalizations of the imaging methods (greedy, iteratively reweighted and convex optimization) in incorporating sparse regularization assumptions on non-sparse images. In the next chapter, we focus on the analysis sparse recovery problem and show the versatility of the advanced convex optimization methods in efficiently incorporating complex regularization assumptions into the radio astronomical imaging problem.

8

FACET-BASED REGULARIZATION FOR SCALABLE IMAGING

Nothing is particularly hard if you divide it into small jobs.

Henry Ford

8.1. INTRODUCTION

The imaging problem in radio interferometry is highly ill-posed and the choice of the prior model of the sky is of utmost importance to guarantee a reliable reconstruction. Any attempt to image reconstruction requires the regularization of the problem by postulating an appropriate signal model. Point sources and extended emissions can coexist in a radio image and require different models to obtain the best reconstruction performance as studied in [135]. Traditionally, one or more regularization terms (e.g. sparsity and positivity) are applied for the complete image. However, radio sky images can often contain individual source regions in a large empty background. Based on the observation that most of the radio sky is empty and sources appear as clusters of complex emissions, we develop a regularization formalism. We propose to divide radio images into source occupancy regions (facets) and apply relevant regularizing assumptions for each facet. This regularization incorporates two modeling assumptions, (i) bounded support

Part of this chapter is published as: S.Naghibzadeh, A. Repetti, A.J. van der Veen and Y. Wiaux. Facet-Based Regularization for Scalable Radio-Interferometric Imaging. 2018 IEEE European Signal Processing Conference (EUSIPCO), 2018.

The work presented in this chapter has been conducted in collaboration with the Biomedical and Astronomical Signal Processing research group, Heriot-Watt University, Edinburgh, UK.

of the source clusters (called facets) and (ii) multi-scale wavelet modeling of the sources within clusters.

Furthermore, the design of the imaging algorithms for the next generation radio telescopes faces extreme challenges due to the unprecedented data volumes and the demand for recovery of sky images at a new range of resolutions and sensitivities. In fact, the inverse problem in radio astronomy is high dimensional both in the data domain and the image domain. In the previous chapters we have discussed that numerical methods can be adjusted to incorporate regularizing assumptions into the imaging problem and provide efficient imaging methods for image formation with the next generation radio telescopes. However, these methods are not flexible enough to incorporate complex regularizing and source modeling assumptions. In this context, advanced convex optimization methods can be exploited. Furthermore, modern convex optimization distributed algorithmic structures enable to split both data and image into an arbitrary number of blocks which can be handled in parallel. This increases the scalability of the imaging algorithms. Among these algorithms, the primal dual algorithm [136, 137, 138, 139] enables efficient full splitting of operators involved in the composite convex optimization problem. An additional randomization functionality of the primal-dual algorithm enables to visit a subset of the blocks at each iteration. A recent work has studied the benefits of the primal-dual algorithm for radio-interferometric imaging by splitting the data into blocks, and randomizing over the data blocks [75].

The present work is a preliminary attempt to investigate the computational benefit of splitting the image under scrutiny into blocks as well, here called facets, using optimization theory. Faceting is a common technique in RI. Traditionally, they have been introduced to handle the so-called direction-dependent effects (DDEs) [140]. Recently, Tasse et al. [141] proposed a faceting approach to parallelize the facet computation, and accelerate the global image reconstruction. This method is a CLEAN-based technique, assuming sparsity in the image domain (e.g. ℓ_1 regularization). The method we develop in this article is similar to [141] in the parallelization of facet computation, but leverages advanced stochastic optimization techniques rather than greedy approaches. Moreover, optimization theory allows to use versatile regularization terms and benefits from convergence guaranties [138]. In fact, the proposed method can be seen as a multi-facet generalization of the data-block SARA method developed in [75].

In particular, beyond the obvious benefit of parallelization of the facet computation at each iteration, we focus on the following two aspects. Firstly, our work stems from the realization that a large class of radio images of interest are mostly empty, with few structured sources appearing in distinct facets. In this context, it should be computationally more efficient to apply a very simple prior model (i.e. ℓ_1) on the background, and confine sophisticated regularization priors (i.e. sparsity averaging) to specific source regions (facets). We assume prior knowledge of the facet decomposition, which can for example be obtained from low-resolution reconstruction. Secondly, we investigate how the computation can be further lightened by relying on a stochastic functionality and propose a procedure to process only a fraction of the facets. Exploiting stochastic block coordinate primal-dual proximal algorithms, we develop a scalable algorithm for facet-based regularization of radio interferometric imaging problems. Facet-based modeling can be represented as a right preconditioner which performs two tasks, (i) dictating the support

of image by discarding the empty columns of \mathbf{M} (column-selection) and (ii) complex facet-based modeling by multi-Wavelet sparsity dictionary.

The remainder of the chapter is organized as follows. In Section 8.2 we restate the RI imaging problem and recall the data splitting approach developed in [75]. We introduce our facet-based approach and our algorithm in Section 8.3. We illustrate the performance of our approach on simulated data in Section 8.4. Finally, we draw conclusions and suggest future works in Section 8.5.

8.2. RADIO-INTERFEROMETRIC IMAGING

8.2.1. PROBLEM DESCRIPTION

Following the data model presented in Chapter 2, the telescope array contains P distinct antennas that capture radio signals that are split into narrow frequency bands. The FoV of the radio telescope is decomposed into Q pixels that can be independently treated as the source signals impinging on the array. The received signals are contaminated by the receiver noise modeled as mutually independent zero mean Gaussian signals. The objective is to find an estimate $\hat{\boldsymbol{\sigma}}$ of the original unknown sky brightness distribution $\boldsymbol{\sigma} \in \mathbb{R}^Q$ from the degraded measurement $\tilde{\mathbf{r}} \in \mathbb{C}^D$ (also called visibilities), corresponding to the pairwise correlation of the signals output from the telescope array ($D = P^2$). We restate the measurement model as

$$\tilde{\mathbf{r}} = \mathbf{M}\boldsymbol{\sigma} + \mathbf{e} \quad (8.1)$$

where $\mathbf{M} \in \mathbb{C}^{D \times Q}$ is the linear measurement operator and $\mathbf{e} \in \mathbb{C}^D$ is a realization of an additive random noise with bounded energy, i.e. there exists $\epsilon > 0$ such that $\|\mathbf{e}\|_2 \leq \epsilon$.

We assume that the measurement operator \mathbf{M} is perfectly known, such that no calibration is needed [142]. In this chapter, \mathbf{M} is modeled and implemented as the product between a matrix \mathbf{G} and an oversampled Fourier operator \mathbf{F} (implemented using the Fast Fourier Transform - FFT). The matrix \mathbf{G} contains the compact support kernels enabling the computation of the continuous Fourier samples from the discrete Fourier coefficients provided by the FFT [143, 144]. The combined operator is known as the Non-Uniform FFT (NUFFT) [144]. Note that the w -term can be also incorporated in this matrix in the case when the non-coplanarity of the baselines is taken into account [118, 145].

8.2.2. COMPRESSIVE SENSING AND DATA SPLITTING

In [36], the authors have proposed to define the estimate of $\boldsymbol{\sigma}$ as a solution to a convex minimization problem, leveraging compressive sensing theory [146, 147]. In this context, the sky image is assumed to have a sparse representation into a given basis $\boldsymbol{\Psi} \in \mathbb{C}^{Q \times L}$, with $L \geq Q$. In other words, it is assumed that $\boldsymbol{\Psi}^H \boldsymbol{\sigma}$ has only few non-zero coefficients, where $\boldsymbol{\Psi}^H$ denotes the Hermitian transpose matrix of $\boldsymbol{\Psi}$. This sparsity-aware approach has been adopted in multiple works such as [148, 36, 75, 142] with different choices of sparsity basis $\boldsymbol{\Psi}$. In [36], the authors proposed to promote average sparsity by choosing $\boldsymbol{\Psi}$ to be the concatenation of the Dirac basis with the first 8 Daubechies wavelet transforms [149]. More recently, to achieve a highly parallelizable algorithmic structure, it has been proposed in [75] to split the visibilities into $n_d > 0$ blocks. Formally,

$\tilde{\mathbf{r}} = (\tilde{\mathbf{r}}_j)_{1 \leq j \leq n_d}$ where $\tilde{\mathbf{r}}_j \in \mathbb{C}^{M_j}$ is the j th data block of dimension D_j . In this context, the linear system (8.1) can be rewritten as follows:

$$(\forall j \in \{1, \dots, n_d\}) \quad \tilde{\mathbf{r}}_j = \mathbf{M}_j \boldsymbol{\sigma} + \mathbf{e}_j \quad (8.2)$$

where $\mathbf{M}_j = \mathbf{G}_j \mathbf{F} \in \mathbb{C}^{D_j \times Q}$ is the associated block of the measurement matrix and $\mathbf{e}_j \in \mathbb{C}^{M_j}$ is the j th block of the additive noise \mathbf{e} . Since the additive noise is assumed to have a bounded energy, for every $j \in \{1, \dots, n_d\}$, there exists $\epsilon_j > 0$ such that $\|\mathbf{e}_j\|_2 \leq \epsilon_j$. Using this structure, the authors have proposed to define the estimate of the sky image as a solution to the constrained analysis-based sparse recovery problem

$$(\forall j \in \{1, \dots, n_d\}) \quad \underset{\boldsymbol{\sigma} \in \mathbb{R}_+^Q}{\text{minimize}} \quad \|\Psi^H \boldsymbol{\sigma}\|_1 \quad \text{subject to} \quad \|\mathbf{M}_j \boldsymbol{\sigma} - \tilde{\mathbf{r}}_j\|_2 \leq \epsilon_j, \quad (8.3)$$

or in terms of the indicator function as

$$(\forall j \in \{1, \dots, n_d\}) \quad \underset{\boldsymbol{\sigma} \in \mathbb{R}_+^Q}{\text{minimize}} \quad \|\Psi^H \boldsymbol{\sigma}\|_1 + \sum_{j=1}^{n_d} \iota_{\mathcal{B}_2(\tilde{\mathbf{r}}_j, \epsilon_j)}(\mathbf{M}_j \boldsymbol{\sigma}), \quad (8.4)$$

where $\iota_{\mathcal{B}_2(\tilde{\mathbf{r}}_j, \epsilon_j)}$ denotes the indicator function of the ℓ_2 ball $\mathcal{B}_2(\tilde{\mathbf{r}}_j, \epsilon_j)$ centered in $\tilde{\mathbf{r}}_j$ with radius ϵ_j , i.e. $\mathcal{B}_2(\tilde{\mathbf{r}}_j, \epsilon_j) = \{\mathbf{z} \in \mathbb{C}^{D_j} \mid \|\mathbf{z} - \tilde{\mathbf{r}}_j\|_2 \leq \epsilon_j\}$. The indicator function of the ℓ_2 ball at a point \mathbf{z} is equal to 0 if $\mathbf{z} \in \mathcal{B}_2(\tilde{\mathbf{r}}_j, \epsilon_j)$, and $+\infty$ otherwise¹.

8.3. PROPOSED APPROACH

8.3.1. FACET-BASED IMAGING

Giga-pixel images of the radio sky $\boldsymbol{\sigma}$ usually consist of a large black background which is mostly empty and separate extended structures which can be clustered together. An example of this type of radio image is shown in Fig. 8.1(top) which shows the radio emission from the Cygnus A radio galaxy image with separate emissions and a large emission-free background. We propose to take advantage of this particular sky image structure and split these images into non-overlapping n_c facets. Formally, we define $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_k)_{1 \leq k \leq n_c}$ where, for every $k \in \{1, \dots, n_c\}$, $\boldsymbol{\sigma}_k \in \mathbb{R}^{Q_k}$ consists of a subpart of the image, and $Q = Q_1 + \dots + Q_{n_c}$. In Fig. 8.1(bottom), we show an example of the image splitting of the Cygnus A image. We can see that 3 main facets are identified, corresponding to the 3 main structures of the image. In addition, the background (in black in Fig. 8.1(bottom)) is considered to be a 4-th facet. In this work, we assume that a pre-processing clustering step has been performed (e.g. using a low resolution estimate), and that we have access to the support of each facet.

In this context, problem (8.1) can be rewritten for the j th data block as

$$\tilde{\mathbf{r}}_j = \sum_{k=1}^{n_c} [\mathbf{M}_j \boldsymbol{\sigma}]_k + \mathbf{e}_j, \quad (8.5)$$

where $[\cdot]_k$ denotes the k th block of its argument and for the NUFFT implementation of the measurement operator amounts to $[\mathbf{M}_j \boldsymbol{\sigma}]_k = [\mathbf{G}_j \mathbf{F} \boldsymbol{\sigma}]_k$. We note that the FFT operator

¹In general, the indicator function of a set \mathcal{C} at point \mathbf{z} , denoted as $\iota_{\mathcal{C}}(\mathbf{z})$, is by definition equal to 0 when $\mathbf{z} \in \mathcal{C}$ and $+\infty$ otherwise.

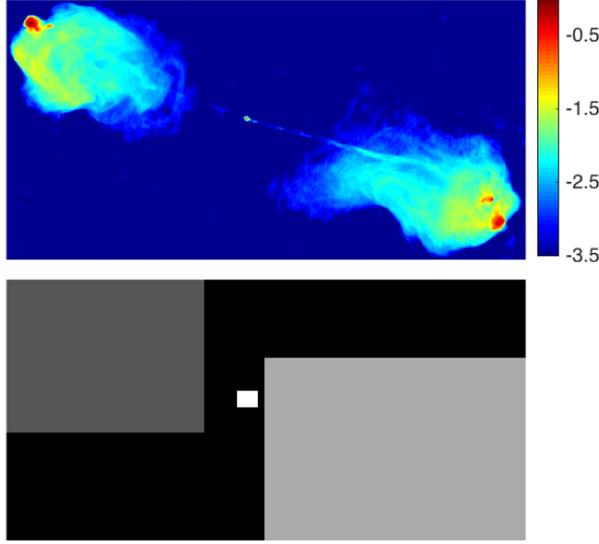


Figure 8.1: Top: Image of Cygnus A ($Q = 512 \times 1024$) in log scale. Bottom: Image showing the 4 considered facets (3 facets with structures, and 1 facet for the background).

is not easily separable on the image facets and requires the complete image σ . However for a direct implementation of the measurement matrix (instead of approximation by NUFFT), we have $[\mathbf{M}_j \sigma]_k = \mathbf{M}_{j,k} \sigma_k$.

Since the considered facets can contain either complex structures, or almost only zero coefficients for the background, we propose to choose different sparsity regularization terms for the different facets. In particular, the facet associated with the background, chosen to be the last facet $k = n_c$, is mostly empty in the image domain. To avoid losing any weak emission that might be present in the background, we propose to choose a simple regularization, the ℓ_1 norm, to regularize this particular facet. For all the other facets containing sophisticated structures, we use the average sparsity regularization introduced in [36]. The associated minimization problem can be formulated as

$$\underset{\sigma = (\sigma_1, \dots, \sigma_{n_c}) \in \mathbb{R}_+^Q}{\text{minimize}} \sum_{k=1}^{n_c} \gamma_k \|\Psi_k^H \sigma_k\|_1 + \sum_{j=1}^{n_d} \iota_{\mathcal{B}_2(\bar{\mathbf{r}}_j, \epsilon_j)} \left(\sum_{k=1}^{n_c} [\mathbf{M}_j \sigma]_k \right), \quad (8.6)$$

where, for every $k \in \{1, \dots, n_c\}$, $\gamma_k > 0$ is a regularization parameter, associated with the k th facet. In this formulation, for every $k \in \{1, \dots, n_c - 1\}$, $\Psi_k \in \mathbb{R}^{L_k \times Q_k}$ corresponds to the concatenation of the Dirac basis and the first 8 Daubechies wavelet transforms associated to the k th facet σ_k . The last facet σ_{n_c} being dedicated to the background, Ψ_k is chosen to be the Dirac basis (i.e. identity matrix with $L_k = Q_k$). This formulation allows to separate the regularization terms for the disjoint facets.

8.3.2. PROPOSED ALGORITHM

Problem (8.6) consists in minimizing a sum of composite non-differentiable convex functions containing linear operators. Primal-dual proximal algorithms are particularly effi-

cient to solve such problems [139] and provide efficient splitting of all the operators involved. In these methods, the solution to the original problem is computed iteratively by solving a sequence of smaller subproblems, where each subproblem involves only one of the terms appearing in the objective function, handling each of the involved operators separately [139]. This class of iterative optimization algorithms have been already used during the last decade in the context of RI, when sparsity-based regularization terms are considered [34, 36, 75]. In particular, in [75], the authors proposed to leverage the stochastic primal-dual proximal algorithm developed in [138], to solve problem (8.4). Basically, in this work, the authors proposed to utilize the stochastic properties to reduce the computational complexity of the algorithm by activating, randomly, only a subset of the data blocks per iteration.

The primal dual algorithm is based on iterative alternation between solving a primal problem and the associated dual optimization problem [139]. Writing all the constraints in the objective function 8.6 via their indicator functions, we can distinguish the primal problem and the associated dual problem. More specifically, the primal problem can be reformulated as

$$\underset{\boldsymbol{\sigma}=(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_{n_c})}{\text{minimize}} \sum_{k=1}^{n_c} (f_k(\boldsymbol{\sigma}_k) + g_k(\boldsymbol{\Psi}_k^H \boldsymbol{\sigma}_k)) + \sum_{j=1}^{n_d} h_j \left(\sum_{k=1}^{n_c} [\mathbf{M}_j \boldsymbol{\sigma}]_k \right), \quad (8.7)$$

where, for every $k \in \{1, \dots, n_c\}$ and $j \in \{1, \dots, n_d\}$, $f_k = \iota_{\mathbb{R}_+^{Q_k}}$, $g_k(\mathbf{z}) = \gamma_k \|\mathbf{z}\|_1$ and $h_j = \iota_{\mathcal{B}_2(\bar{\mathbf{r}}_j, \epsilon_j)}$. The dual problem associated with Problem 8.7 can be stated as

$$\underset{\substack{\mathbf{v}_k=(\mathbf{v}_1,\dots,\mathbf{v}_{n_c}), \\ \mathbf{z}_j=(\mathbf{z}_1,\dots,\mathbf{z}_{n_d})}}{\text{minimize}} \sum_{k=1}^{n_c} f_k^*(-\boldsymbol{\Psi}_k \mathbf{v}_k - \sum_{j=1}^{n_d} [\mathbf{M}_j^H \mathbf{z}_j]_k) + g_k^*(\mathbf{v}_k) + \sum_{j=1}^{n_d} h_j^*(\mathbf{z}_j), \quad (8.8)$$

where $\mathbf{v}_k \in \mathbb{R}^{\tilde{Q}_k}$ and $\mathbf{z}_j \in \mathbb{C}^{D_j}$ are the dual variables for the k th source cluster and the j th data block and f_k^* , g_k^* and h_j^* denote the conjugates of the functions f_k , g_k and h_j .²

Furthermore, in the proposed primal-dual facet-based imaging approach, not only the data are divided into blocks, but also the the image is divided into facets. We propose to use the block-coordinate structure of the stochastic primal-dual proximal algorithms developed in [138] to handle efficiently the block data terms and the facet-based regularization terms. However, the FFT operator is not easily separable and has to be applied on the complete image. The block-coordinate structure relies on sweeps of data and image blocks at each iteration [138]. The resulting *faceting primal dual algorithm*, to solve problem (8.6), is described in Algorithm 5.

In Algorithm 5, we can distinguish three main parallel loops. The first loop (step 6) is used to update the facets. The second loop (step 10) is used to handle the ℓ_1 facet-based regularization terms. Finally, the third loop (step 21) is used to handle the ℓ_2 -ball constraints related to the data blocks.

To handle the positivity and the ℓ_2 -ball constraints, in steps 7 and 24 respectively, projection steps are performed. The projection onto a convex, closed, non-empty subset

²The conjugate of a function J is defined as $J^*(\mathbf{z}) := \sup_{\mathbf{y}} (\mathbf{z}^H \mathbf{y} - J(\mathbf{y}))$, where \sup denotes the supremum.

Algorithm 5: Faceting primal-dual algorithm

```

1 Initialization:  $\sigma_0 \in \mathbb{R}^Q$ ,  $\zeta_0 \in \mathbb{R}^Q$ ,  $\mathbf{v}_0 \in \mathbb{R}^Q$ ,  $\nu \in ]0, +\infty[$ ,  $\tau \in ]0, +\infty[$ ,
   ( $\forall k \in \{1, \dots, n_c\}$ )  $\mathbf{v}_{k,0} \in \mathbb{R}^{L_k}$ , and  $\gamma_k \in ]0, +\infty[$ , and ( $\forall j \in \{1, \dots, n_d\}$ )  $\mathbf{z}_{j,0} \in \mathbb{R}^{M_j}$ 
2 Iterations:
3 for  $n = 0, 1, \dots$  do
4   Choose randomly  $\mathbb{S}_n \subset \{1, \dots, n_c\}$ 
5   Choose randomly  $\mathbb{D}_n \subset \{1, \dots, n_d\}$ 
6   for  $k \in \{1, \dots, n_c\}$  (facet parallel update) do
7      $\sigma_{k,n+1} = \Pi_{\mathbb{R}^Q}(\sigma_{k,n} - \tau(\zeta_{k,n} + \mathbf{v}_{k,n}))$ 
8      $\tilde{\sigma}_{k,n} = 2\sigma_{k,n+1} - \sigma_{k,n}$ 
9   end
10  for  $k \in \{1, \dots, n_c\}$  (regularization parallel step) do
11    if  $k \in \mathbb{S}_n$  then
12       $\mathbf{u}_{k,n} = \mathbf{v}_{k,n} + \gamma_k \Psi_k^H \tilde{\sigma}_{k,n}$ 
13       $\mathbf{v}_{k,n+1} = \mathbf{u}_{k,n} - \gamma_k \mathcal{P}_{\gamma_k^{-1}}(\gamma_k^{-1} \mathbf{u}_{k,n})$ 
14       $\mathbf{v}_{k,n} = \Psi_k \mathbf{v}_{k,n}$ 
15    else
16       $\mathbf{v}_{k,n+1} = \mathbf{v}_{k,n}$ 
17       $\mathbf{v}_{k,n+1} = \mathbf{v}_{k,n}$ 
18    end
19  end
20   $\mathbf{b}_n = \mathbf{F} \tilde{\sigma}_n$ 
21  for  $j \in \{1, \dots, n_d\}$  (data parallel step) do
22    if  $j \in \mathbb{D}_n$  then
23       $\mathbf{s}_{j,n} = \mathbf{z}_{j,n} + \nu \mathbf{G}_j \mathbf{b}_{j,n}$ 
24       $\mathbf{z}_{j,n+1} = \mathbf{s}_{j,n} - \nu \Pi_{\mathcal{B}_2(\bar{\mathbf{r}}_j, \epsilon_j)}(\nu^{-1} \mathbf{s}_{j,n})$ 
25       $\mathbf{g}_{j,n+1} = \mathbf{G}_j^H \mathbf{z}_{j,n+1}$ 
26    else
27       $\mathbf{z}_{j,n+1} = \mathbf{z}_{j,n}$ 
28       $\mathbf{g}_{j,n+1} = \mathbf{g}_{j,n}$ 
29    end
30  end
31   $\zeta_{n+1} = \mathbf{F}^H \mathbf{g}_{n+1}$ 
32 end

```

\mathcal{C} of \mathbb{R}^Q of $\mathbf{c} \in \mathbb{R}^P$, is defined as $\Pi_{\mathcal{C}}(\mathbf{c}) = \underset{\boldsymbol{\theta} \in \mathbb{R}^P}{\operatorname{argmin}} \|\boldsymbol{\theta} - \mathbf{c}\|_2$. On the one hand, in the case when $\mathcal{C} = \mathbb{R}_+^Q$, this projection reduces to

$$(\forall \boldsymbol{\sigma} \in \mathbb{R}^Q) \quad \Pi_{\mathbb{R}_+^Q}(\boldsymbol{\sigma}) = \max\{\mathbf{0}, \boldsymbol{\sigma}\}, \quad (8.9)$$

where the max operation is performed component-wise. On the other hand, in the case when $\mathcal{C} = \mathcal{B}_2(\tilde{\mathbf{r}}_j, \epsilon_j)$, for every $j \in \{1, \dots, n_d\}$, we have, for every $\mathbf{s}_j \in \mathbb{C}^j$,

$$\Pi_{\mathcal{B}_2(\tilde{\mathbf{r}}_j, \epsilon_j)}(\mathbf{s}_j) = \begin{cases} \mathbf{s}_j & \text{if } \mathbf{s}_j \in \mathcal{B}_2(\tilde{\mathbf{r}}_j, \epsilon_j), \\ \tilde{\mathbf{r}}_j + \epsilon_j \frac{\mathbf{s}_j - \tilde{\mathbf{r}}_j}{\|\mathbf{s}_j - \tilde{\mathbf{r}}_j\|_2} & \text{otherwise.} \end{cases} \quad (8.10)$$

Similarly, to handle the ℓ_1 regularization terms, soft-thresholding operations are performed in step 13. For every facet $k \in \{1, \dots, n_c\}$, for every $\gamma_k > 0$ and $\mathbf{u}_k \in \mathbb{R}^{L_k}$, this operator is defined as follows [101]:

$$\mathcal{P}_{\gamma_k^{-1}}(\mathbf{u}_k) = \begin{cases} -\mathbf{u}_k^{(i)} + \gamma_k^{-1} & \text{if } \mathbf{u}_k < -\gamma_k^{-1} \\ 0 & \text{if } -\gamma_k^{-1} < \mathbf{u}_k < \gamma_k^{-1} \\ \mathbf{u}_k - \gamma_k^{-1} & \text{otherwise.} \end{cases} \quad (8.11)$$

Algorithm 5 is a stochastic algorithm in the sense that, at each iteration $n \in \mathbb{N}$, a subset \mathbb{S}_n of the n_c facet-based regularization terms and a subset \mathbb{D}_n of the n_d data terms are selected randomly (steps 4 and 5, respectively), and only the related variables are updated. It is in particular interesting to notice that the sparsifying operator Ψ_k is only applied if the facet $k \in \{1, \dots, n_c\}$ is selected (see parallel loop in step 10). Similarly, as proposed in [75], the gridding matrix \mathbf{G}_j is only applied if the data block $j \in \{1, \dots, n_d\}$ is selected (see parallel step 21). However, the oversampled FFT operator \mathbf{F} and its adjoint must be performed at each iteration (steps 20 and 31, respectively). Note that when $n_c = 1$, then the algorithm proposed in [75] is recovered. In addition, if, at each iteration $n \in \mathbb{N}$, we choose $\mathbb{S}_n = \{1, \dots, n_c\}$ and $\mathbb{D}_n = \{1, \dots, n_d\}$, then the algorithm reduces to a deterministic primal dual algorithm [136, 137].

Finally, it is important to emphasize that the proposed algorithm benefits from the convergence properties of the general stochastic primal dual algorithms developed in [138]. Let $\nu > 0$, $\tau > 0$ and, for every $k \in \{1, \dots, n_c\}$, let $\gamma_k > 0$ satisfying $\tau^{-1} > \nu \|\mathbf{M}\|_S^2 + \sum_{k=1}^{n_c} \gamma_k \|\Psi_k\|_S^2$, where $\|\cdot\|_S$ denotes the spectral norm. Then, the sequence $(\boldsymbol{\sigma}_n)_{n \in \mathbb{N}}$ generated by Algorithm 5 converges almost surely to a random variable $\hat{\boldsymbol{\sigma}}$ solution to Problem (8.6).

8.4. SIMULATION RESULTS

We evaluate the performance of Algorithm 5 on simulated radio-interferometric data. We choose as the test image the Cygnus A radio galaxy image of size $Q = 512 \times 1024$ shown in in Fig. 8.1(top), in log scale. The image is split into the $n_c = 4$ facets shown in Fig. 8.1(bottom). In this image, the black facet represents the background, indexed by $k = n_c$. The visibilities are generated according to model (8.1), using a random Gaussian undersampling $u-v$ coverage, with $D = Q$. The $u-v$ coverage, split into $n_d = 16$ blocks, is

H

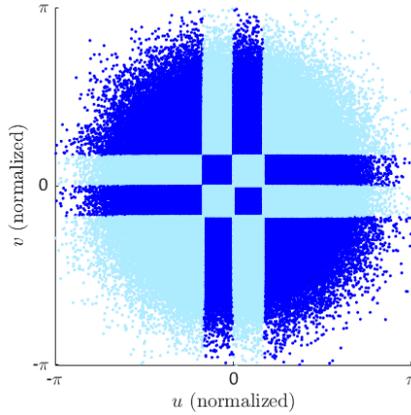


Figure 8.2: Normalized random Gaussian $u-v$ coverage. The light blue and dark blue colors emphasize the different n_d data blocks

shown in Fig. 8.2. In splitting the data, similar to [75] we consider equal sizing of the data blocks such that the load is balanced over different nodes among which the data term computations are distributed. Furthermore, we consider a compact support of each of the data blocks over the $u-v$ plane such that the computation of the FFT can be localized in the $u-v$ plane. In model (8.1), the additive noise is generated following to a zero-mean complex Gaussian distribution, considering an input signal-to-noise ratio (iSNR) of 20 dB, where $\text{iSNR} = 20 \log_{10} \left(\frac{\|F\|_2}{\sqrt{M\sigma}} \right)$, σ^2 being the variance of the noise. For the details of data generation and the estimation of the bounds $(\epsilon_j)_{1 \leq j \leq n_d}$ the reader is referred to [75].

We compare the performances of the proposed faceting primal-dual algorithm, with and without randomization over the facet-based regularization terms, with the primal-dual algorithm developed in [75] to judge the benefit of faceting. Since the data blocks are handled exactly in the same manner in Algorithm 5 and [75], in our simulations, we focus on the interest of using a facet-based approach, with possible randomization. Consequently, in our simulations, we choose, for every $n \in \mathbb{N}$, $\mathbb{D}_n = \{1, \dots, n_d\}$ (i.e. no randomization over the data terms). The reader is referred to [75] for a complete investigation of randomizing the primal dual algorithm over the data blocks. Concerning \mathbb{S}_n , we investigate two cases. In the first case, at each iteration $n \in \mathbb{N}$, we choose $\mathbb{S}_n = \{1, \dots, n_c\}$. It corresponds to a deterministic version of Algorithm 5 where all the facet-based regularization terms are selected at each iteration. This first configuration is used to emphasize the advantage of the proposed facet-based approach, considering the simple ℓ_1 regularization term on the background. In the second case, we use the randomization property of Algorithm 5. In this case, we choose to activate, at each iteration $n \in \mathbb{N}$ the complete background (i.e. facet $k = n_c = 4$) and select randomly one of the three other facets. In this context, we choose $\mathbb{S}_n = \{k_n, 4\}$, where $k_n \in \{1, 2, 3\}$ is chosen following a uniform distribution.

We present in Fig. 8.3 the results of our simulations. Fig. 8.3(a) shows the signal-to-noise ratio (SNR) as a function of the iterations $n \in \mathbb{N}$. The SNR (in dB), for the current

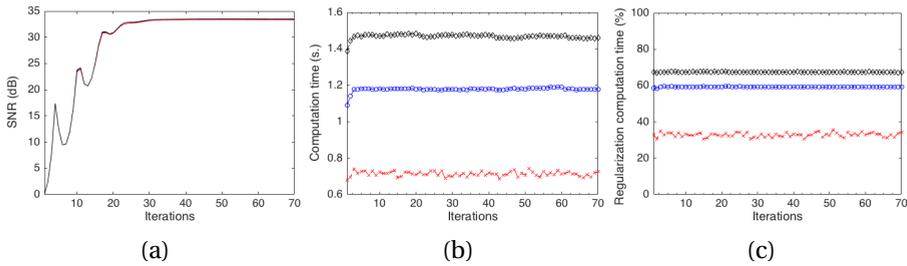


Figure 8.3: Results obtained using the primal dual algorithm from [75] (black - diamond marks), Algorithm 5 without randomization (blue - circle marks), and Algorithm 5 with randomization on the facet-based regularization terms (red - cross marks). (a) SNR (dB) as a function of iterations. The SNR is an averaged SNR for 10 realizations of noise and under-sample distributions. (b) Computation time per global iteration. (c) Computation time to handle the regularization term (i.e. step 10 in Algorithm 5), as a % of the time necessary to perform a complete iteration. For (b)-(c), the computation time is an averaged computation time for 10 realizations of noise and under-sample distributions, with 10 runs on each setting.

iterate σ_n is defined as

$$\text{SNR} = 20 \log_{10} \left(\frac{\|\sigma\|_2}{\|\sigma - \sigma_n\|_2} \right). \quad (8.12)$$

These curves emphasize that the proposed facet-based regularization term is not affecting the reconstruction quality. In Figs. 8.3(b) and (c), for the three considered algorithms, we give the time necessary to compute each global iteration, and the % of time spent to perform the regularization steps (i.e. step 10 in Algorithm 5) per global iteration, respectively. By comparing the black curve ([75]) and the blue curve (Algorithm 5 without randomization), we can conclude that the proposed facet-based regularization term reduces the total computational time. This is due to the fact that the wavelet transforms are only performed on the facets with complex structures. In particular, the theoretical complexity to perform the wavelet decomposition with Algorithm 5 (without randomization) is $\sum_{k=1}^{n_c-1} \mathcal{O}(L_k)$, which is smaller than the complexity of performing a wavelet decomposition on the global image (which is equal to $\mathcal{O}(L)$). The red curves are associated with the randomized version of Algorithm 5. It can be observed that leveraging the stochastic properties of this algorithm allows to divide by 2 the computation time per iteration (w.r.t. [75]). In addition, with the randomized version, performing the regularization steps only requires $\sim 30\%$ of the iteration computation time, while it requires $\sim 60\%$ without randomization and $\sim 70\%$ for [75]. We note that in this comparison, the facets being identified manually, the computational contribution from the facet-selection procedure is not included in Fig. 8.3.

Lastly, we presents some remarks. (i) In this work we assume that the facets are known based on an initial low-resolution image assuming the cost of forming the facets is a small percentage of the total computational burden. (ii) The presented algorithmic framework allows for general facetting schemes which also include overlapping facets. (iii) In a very recent work, Tasse et. al. [141] have presented facet-based wide-field deconvolution algorithms based on the greedy matching pursuit-type schemes mainly to tackle DDEs. Leveraging recent work on self-calibration for DDEs by block-coordinate algorithms [142], DDEs can be incorporated in the presented imaging algorithm by block-

processing with convergence guarantees. All in all, stochastic block-coordinate algorithms together with proximal operators provide a robust and versatile scheme to incorporate (randomized and parallelized) block operations (both in the image and the $u - v$ plane) and regularizations with convergence guarantees. Our scheme benefits from two regularizing assumptions, (i) bounded support of the image and (ii) facet-based regularizing assumption.

8.5. CONCLUSION AND FUTURE WORKS

We have reported our attempts to achieve a facet-based regularization for scalable interferometric imaging. Firstly, we have extended the primal-dual algorithm developed in [75] to incorporate image facets and enable randomization of the facet computation at each iteration. Secondly, we have shown that based on the nature of radio images where individual source facets exist in a sea of mostly-empty background, we can apply the regularization more wisely in order to save computations without sacrificing the reconstruction quality. Thirdly, we have shown that randomization over facets further saves computations. We can conclude that, due to the flexibility of the primal-dual proximal methods, we can incorporate advanced regularization schemes while benefiting from the scalability and computational savings that are provided by randomization. As we have mentioned earlier, we can categorize the radio interferometric imaging problems into three main categories, namely (i) greedy methods, (ii) methods based on convex optimization and (iii) projection methods. Each of these methods have their pros and cons. Based on the results from this chapter, comparing with Chapter 7 and Chapter 5, we conclude that the methods based on convex optimization provide a more versatile scheme for incorporating advanced regularizations.

For future works, we plan to implement an automated faceting schemes based on a low resolution initial image and integrate general faceting schemes which also include overlapping facets in the algorithm by introducing consensus steps as presented in [138]. In addition, similar to the faceting approach proposed in [141], we will perform splitting of the measurement operator over the facets. Finally, we plan to develop a facet-based DDE calibration version of the proposed method, by leveraging a block-coordinate approach [142, 150].

IV

PART IV : EPILOGUE

9

CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

The reward of our work is not what we get, but what we become.

Paulo Coelho

9.1. CONCLUSIONS AND CONTRIBUTIONS

In radio interferometry, the radio emissions from the sky are observed via a telescope array with the objective to construct radio sky images over the field of view of the telescope array. Radio telescope arrays probe the sky through incomplete and noisy measurements of the sky image of interest in the observation plane, leading to an ill-posed inverse problem for image recovery. This is a common problem in the broad area of inverse imaging problems which is encountered in many fields such as biomedical imaging, geophysics/seismic imaging, optical spectrometry and remote sensing.

The imaging pipeline of the next generation radio telescopes, and in particular the SKA, can be considered as a challenging application for the field of array signal processing. Array processing is a well-established field in signal processing where antenna arrays are studied and processing of the data acquired by these arrays are addressed. Array processing and signal processing in general have much to contribute to the next generation radio telescopes both in modeling the data acquisition and processing pipeline and subsequently in designing and analyzing radio interferometric imaging algorithms.

The research on radio astronomy and in particular radio interferometric imaging, has been around for about 80 years. However, the introduction of the next generation radio telescopes, and in particular the SKA, have brought about many new challenges. As a result, imaging techniques developed for the previous generation of radio telescopes cannot be easily applied to the next generation of telescopes. Some of the challenges for the next generation radio telescopes are due to

- the unprecedented data volumes that are beyond the current storage facilities. This highlights a need for fast imaging algorithms that are able to process the data in near-realtime. Considering the well-developed imaging method of CLEAN as the benchmark, new imaging methods are required to be at most of similar computational complexity as CLEAN.
- SKA is going to be able to see the sky in a new range of sensitivity and dynamic range and the ambitious science cases of the SKA demand advanced regularization schemes to produce reliable results. As a result, point source models (the basis for CLEAN) are probably not suitable to reach the desired sensitivity limits.

Therefore, the imaging techniques for the next generation radio telescopes are required to be accurate (free of artifacts), robust, computationally efficient, scalable and fully automated yet with some tuning flexibility. In this thesis, we categorized the current radio interferometric imaging techniques into two categories, i.e. (i) greedy (sparse) reconstruction methods such as CLEAN and NNLS and (ii) convex optimization-based methods.

We summarize our contributions in this thesis as follows

- We have proposed a regularization scheme based on priorconditioning, both for the imaging and the deconvolution problem (Chapter 3), that paves the way for efficient regularization of the projection-based solution methods (in particular the PRIFIRA framework). We further proposed a data-driven regularization which is derived from beamforming of the telescope array data (Chapter 4).
- We have proposed a third category for the radio interferometric imaging algorithms based on projections onto Krylov subspaces and have shown that the proposed priorconditioning scheme can be incorporated in these algorithms to provide efficient regularization. Furthermore, we generalized our algorithmic framework, called PRIFIRA, to increase its flexibility in incorporating different regularization schemes such as sparsity (Chapter 5). We have shown that PRIFIRA is computationally much less demanding than CLEAN and can be a good candidate for the snapshot imaging pipeline of the next generation radio telescopes (Chapter 6).
- We have investigated the applicability of the greedy sparse methods, in particular the NNLS algorithm, to science cases of the SKA and concluded that for the scale and science cases of the SKA, these methods exhibit deficiencies both in incorporating regularization and in computations. Furthermore, we have investigated the analysis vs. synthesis formalism of the radio interferometric imaging problem for images in which extended emissions and point sources co-exist and have inspected the applicability of the three categories of algorithms to such problems (Chapter 7).
- Finally, we have proposed a method to incorporate complex regularization assumptions into the radio interferometric imaging problem via faceting and have reported our experience with the convex optimization methods. We have concluded that these methods provide a versatile scheme for incorporating advanced

and complex regularization assumptions which will be required for the science cases of the SKA (Chapter 8).

We can conclude that the proposed projection-based methods together with the prior-conditioning regularization scheme, i.e. the PRIFIRA framework, are the most computationally efficient and would satisfy the near-realtime requirements of the SKA. Therefore, they are good candidates to be integrated in the imaging pipeline as methods for fast snapshot imaging. However, these methods are not very flexible in incorporating complex regularization terms when the underlying image contains a mixture of point sources and complex extended structures. On the other hand, if computations and convergence times can be tolerated, the advanced convex optimization methods introduced in Chapter 8 provide a versatile scheme for incorporating complex regularization terms (such as facet-based regularization) and due to their proven potential for distributed implementation they can provide for the required scalability. In conclusion, we would like to highlight that there is no “best” method to solve all the challenges of the imaging pipeline of the SKA and each of the existing and proposed imaging categories have their pros and cons. Depending on the computation and science requirements we need to choose the most adequate method for imaging.

Last but not least, the efficient numerical techniques as well as regularization methods developed in this thesis are not restricted to radio astronomy and find direct application in and have a direct impact on many different imaging application areas.

9.2. SUGGESTIONS FOR FUTURE WORK

The present work addresses some of the many challenges of the processing pipelines of the next generation radio telescopes. We have reported our attempts in designing efficient algorithmic frameworks as well as adequate regularization frameworks for the processing pipeline of the next generation of radio telescopes. There are still many hurdles remaining in the design of the SKA processing pipelines. Furthermore, various questions arise from the research results we presented in this dissertation. We propose the following directions for future research on the topic.

Joint calibration and imaging with PRIFIRA We have assumed that we have access to purely calibrated data for imaging. However, if the visibility measurements are poorly calibrated, the calibration parameters and the intensity estimates are obtained via an iterative loop that acts on the correlation data and estimates calibration parameters and the intensity estimates in turns. This loop is called the self calibration (selfcal) loop [8, 151].

We show the inter-play of the CLEAN-based imaging and self-calibration pipeline in the following diagram (Figure 9.1). Other imaging algorithms should be translated in this framework and combined with the self-calibration pipeline. Figure 9.1 shows the Science Data Processing pipeline of the SKA where imaging and calibration is performed. This schematic is based on the implementation of the CLEAN algorithm and shows the joint imaging and calibration pipeline. The pipeline consists of three nested loops: self-calibration, major cycle and minor cycle. First, an initial estimate of the sky map is obtained based on the measured visibilities. Afterwards, the sky model is updated after

several major and minor cycle iterations. After the sky model has been updated, new calibration parameters are determined based on the current sky model and new iterations are initiated to modify the sky model [26].

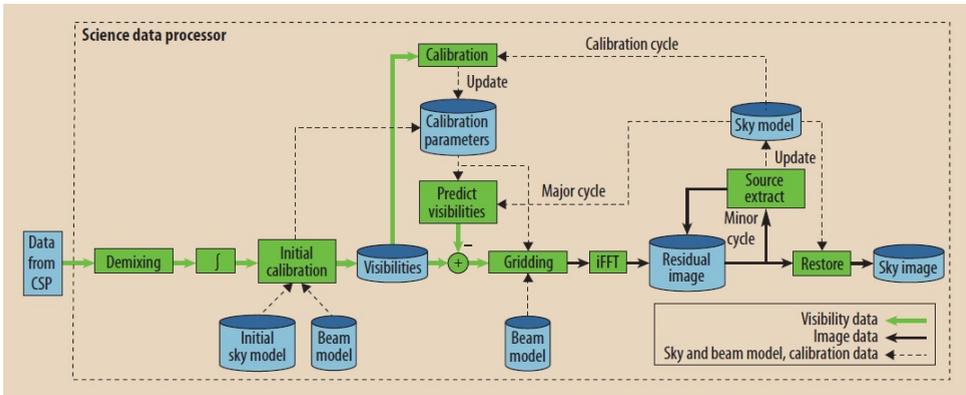


Figure 9.1: Science data processing pipeline (image courtesy of [26])

To be able to integrate PRIFIRA in the processing pipeline of the SKA, we must combine PRIFIRA with the existing selfcal loops in the pipeline that are primarily designed to be compatible with the CLEAN pipeline.

Joint data reduction and imaging Exploiting the PRIFIRA framework, we can implement adaptive methods for data reduction by integrating online algorithms that attain a (data-driven) data reduction framework. With this regard, we can benefit from the recently proposed baseline-dependent averaging [152, 153, 29] frameworks to reduce the amount of data required for image formation by updating the data from short baselines less frequently than the longer baselines over snapshots time.

Hierarchical Bayesian models for PRIFIRA We started the formulation of PRIFIRA from a Bayesian problem formulation and based on approximately estimating the signal covariance matrix by beamforming we made a shortcut to the MLE formalism. However, we can continue with the Bayesian formalism and devise a systematic way to estimate the hyper-parameters needed for the statistical modeling of the image. We expect that this will result in increasing the computational complexity of the framework. However, it will improve the estimation performance. This has similarities with the Bayesian learning framework from the machine learning community.

Automated faceting and splitting the measurement operator for facet-based imaging In Chapter 8, we presented a facet-based framework for scalable radio interferometric imaging. We mainly benefited from the faceting to obtain additional regularization and to distribute the regularization term among the facets. However, we have not proposed a systematic way to obtain the facets. This can become a future direction of research

together with obtaining a splitting of the measurement operator that can result in significant computational savings.

Purely image-domain PRIFIRA The main difference between solving the imaging and deconvolution problem, if implemented by iterative methods, is that in the former we need to alternate between the image domain and the visibility domain during the iterations while the latter is completely performed in the image domain and only requires computing the dirty image from the visibility data once. Therefore, by tackling the deconvolution problem, we can eliminate the need for frequent travels to the visibility domain. We are able to adjust PRIFIRA by replacing LSQR with a different Krylov-based iterative method, namely MINRES-QLP [154], to obtain a purely image-domain algorithm by tackling the deconvolution problem instead of the imaging problem. We can further exploit the structure present in the deconvolution matrix to speed-up the iterations.

PRIFIRA for calibration MF- and MVDR-PRIFIRA are mainly aimed at the recovery of smooth and extended source structures. Similar circumstances are observed in multi-frequency calibration where the variation of the parameters are considered smooth. It would be fruitful to investigate the applicability of PRIFIRA in estimating the calibration parameters.

Exploiting the applicability of the proposed methods in other applications The impact of the methods developed in this thesis is beyond only radio interferometry. An interesting research direction would be to figure out the applicability of these methods in different areas of research such as biomedical imaging.

Parallel implementation of PRIFIRA PRIFIRA is based on the Krylov subspace-based method of LSQR. For highly scalable implementations of PRIFIRA, it is important to investigate parallel implementations of LSQR (or similar Krylov-based methods). Parallelism can be applied both over the image pixels as well as the visibility measurements.

Implementation of PRIFIRA in the imaging pipeline PRIFIRA as well as many new compressive sensing-based algorithms are implemented in MATLAB. However, in MATLAB we do not have access to advanced implementations of W-projection and W-stacking for forward and backward operations. To benefit from the cutting-edge fast implementations of the forward and backward operations available in the WSClean imager software package¹ [155], we have made a Python implementation of PRIFIRA. We have interfaced this implementation based on the PYWSCLEAN wrapper (developed by Andre Offringa) with WSClean for performing the forward and backward operations. A schematic of the platforms for performing the imaging by PRIFIRA is shown in Figure 9.2.

This scheme provides a means for comparing the performance of PRIFIRA with the state-of-the-art implementations of CLEAN by directly working on the real radio interferometric data available in the Measurement Set (MS) format. Due to some compatibility

¹<https://sourceforge.net/p/wsclean/wiki/Home/>

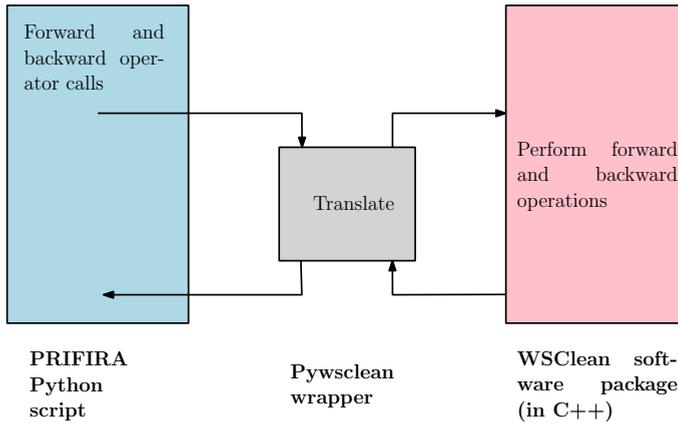


Figure 9.2: PRIFIRA-WSclean platform relations

issues, we have not yet reached a consistent implementation. Therefore, we present this as the content of future work.

ACKNOWLEDGEMENTS

The brain appears to possess a special area which we might call poetic memory and which records everything that charms or touches us, that makes our lives beautiful...

Milan Kundera, *The Unbearable Lightness of Being*

Now that the bitter-sweet journey of my PhD, and certainly an influential part of my life, is about to end, I want to express my deepest gratitude towards all the people who joined me, supported me and influenced me along the way. It goes without saying that this letter will be perfectly imperfect and incomplete in addressing all the influential people who shaped me personally and scientifically during the past four years. However, I hope I can briefly acknowledge at least some of them.

First and foremost, I want to express my deepest appreciations towards my supervisor/promoter Prof. Alle-Jan van der Veen. I have learned, more than I can express appreciations for, from Alle-Jan. His infinite patience, pragmatic leadership, authenticity, integrity and great wisdom have always been sources of inspiration and admiration for me. During this past four years, I have always benefited from a great freedom in research together with his motivating comments and, when required, his critical attitude and high standards of perfection to become a responsible researcher in academia. He deeply cares about his PhD students as human beings, rather than temporary students, and provides them with many opportunities to grow scientifically and personally.

I want to express my gratitude towards Dr. Stefan Wijnholds who has been a very knowledgeable and always accessible contact for me and has patiently and comprehensively answered all my questions, provided me with the simulations/data initials to get me started on my research and given me the opportunity to expand my research cycle by inviting me to conferences. I have always benefited from his constructive comments and especially his comprehensive comments on my PhD thesis draft. I appreciate his time and effort and cooperative character. I want to thank Prof. Patrick Dewilde for occasionally spending time in Delft and giving me the opportunity for in-depth discussions, reading and commenting my work and providing me with highly thought-provoking questions. I want to thank Prof. Geert Leus for our scientific discussions and his constructive feedbacks and critical comments on my thesis.

I am grateful to Dr. Millad Mouri Sardarabadi for our long discussions, his patience and willingness to share his knowledge and experience on radio astronomical signal processing with me, all our brainstorming sessions and all his motivating comments during the initial stages of my PhD. I want to thank Dr. Sundeep Chopuri for our scientific discussions and his career and scientific advices.

I am grateful to all the other committee members, Prof. Per Christian Hansen, Prof. Dimitri Van de Ville and Prof. Nico de Jong who have accepted to read my thesis and take

the hassle to attend my PhD defense to judge my work.

I would like to acknowledge TU Delft, Astron and NWO for funding my research. Special thanks to the Astron team, in particular Dr. Albert-Jan Boonstra, Dr. Ronald Nijboer, Dr. Andre Offringa, Dr. Bas van der Tol, Dr. Tammo-Jan Dijkema and Bram Veenboer for all the fruitful meetings, technical discussions and supports and for providing me with data.

My appreciation goes to the BASP group at the Heriot-Watt University, and in particular Prof. Yves Wiaux for accepting me in his team and letting me benefit from their knowledge. Thanks to Arwa, Alex, Jasleen, Marica for welcoming me and patiently answering my questions and special thanks to Dr. Audrey Repetti for her cooperation, willingness to share her knowledge and experience and for all the things she taught me.

It has been a great pleasure for me to be a part of the Circuits and Systems (CAS) research group. Through CAS I have met and worked among invaluable colleagues in a stimulating and friendly environment. CAS has always been my home away from home and I am grateful to my colleagues for all the lunches that we enjoyed together, all the memories we made in group outings and Christmas dinners, all the travels we went together and all the conversations we shared while enjoying our coffees.

My boundless love and appreciation goes to my parents, Mahmoud and Nayyereh, for who I am now. They have never failed to support me throughout my life and although my studies have pulled us apart, their selfless love has remained a constant source of endorsement and reassurance for me. Great thanks to my brother Bahram for always supporting me and being a motivating example for me. It is sad that visa processing times and long distances have prevented you from attending my defense ceremony and from sharing this memorable event with me.

Heartfelt thanks to my always cheerful and high-spirited friend Fahime for all the happy and memorable times we spent together in Delft. Special thanks to Ali for his patience and kind support during the last year of my PhD, when I was perhaps the most intolerable. Many thanks to my friends at TU Delft and beyond, Behnam, Samira, Robin, Naghmeah and Fabricio who made the PhD life more enjoyable and fun. Thanks to my dearest friends in Germany, Faezeh and Fatemeh who heart-fully supported me and shared some delightful time off from PhD with me. Thanks to Nasim, Mitra, Arash, Mohammad who kept me close to their hearts in spite of being so far.

Shahrzad

Delft, November 2018.

CURRICULUM VITÆ

Shahrzad Naghibzadeh was born in Mashhad, Iran, on June 6, 1990. She earned her B.Sc. degree (with honors) from the Ferdowsi University of Mashhad (FUM), Mashhad, Iran, in July 2012 and her M.Sc. degree (cum laude) from Delft University of Technology (TU Delft), Delft, The Netherlands, in July 2014. During her BSc studies, she did an internship on network architecture and wireless sensor networks in the IP-PBX Research Laboratory, FUM. For her BSc thesis project, she worked on biomedical signal processing for EEG, in the Biomedical Research Laboratory, FUM. For her MSc thesis project, she joined the Lighting group of Philips Research, Eindhoven, The Netherlands, where she conducted research on source localization, tracking and communication systems



with ultrasonic sensor arrays under the joint supervision of Dr. Ashish Pandharipande, Philips Research Eindhoven, and Prof. Geert Leus, TU Delft, The Netherlands. In July 2014, she joined the Circuits and Systems (CAS) group at the Microelectronics department, TU Delft, The Netherlands, where she has been working towards obtaining her PhD degree under the supervision of Prof. Alle-Jan van der Veen. Her PhD research project is funded by NWO and is conducted in a collaboration with ASTRON, The Netherlands. In 2016 and 2017 she visited the Biomedical and Astronomical Signal Processing (BASP) group, Edinburgh, United Kingdom, led by Prof. Yves Wiaux. For the latter visit she obtained an Erasmus mobility grant. During her PhD, she has obtained the best poster award at the URSI Benelux Forum 2018. Furthermore, she has been the runner-up for the best three minutes thesis presentation (3MT) contest at EUSIPCO, 2018, in which she competed among PhD candidates from different countries on a three minute presentation of her PhD thesis work. She has also obtained multiple conference travel grants. Her general research interests lie in the areas of statistical signal processing, array signal processing, inverse problems and optimization methods.

A

LIST OF ABBREVIATIONS

A

RI	radio interferometry
RA	radio astronomy
LS	least squares
WLS	weighted least squares
MLE	maximum likelihood estimation
MVU	minimum variance unbiased
MF	matched filter
MVDR	minimum variance distortionless response
SKA	square kilometer array
LOFAR	low frequency aperture array
ASTRON	Netherlands institute for radio astronomy
AARTFAAC	Amsterdam-ASTRON radio transient facility and analysis center
DRIFT	data reduction and image formation for future radio telescopes
MWA	Murchison wide-field array
EOR	epoch of reionization
CMB	cosmic microwave background
FoV	field of view
DDE	direction-dependent effect
DIE	direction-independent effect
ME	measurement Equation
EM	electromagnetic
HI	neutral Hydrogen
CSP	central signal processing
SDP	science data processing
PSD	power spectral density
ULA	uniform linear array
DoA	direction of arrival
SVD	singular value decomposition
FFT	fast Fourier transform
FWHM	full width at half maximum
HPBW	half power beam width
DFT	discrete Fourier transform
MP	matching pursuit
OMP	orthogonal matching pursuit
SNR	signal to noise ratio
SVD	singular value decomposition
TSVD	truncated singular value decomposition
CSP	central signal processing
SDP	science data processing
PSF	point spread function
EVD	eigen value decomposition
ASM	all-sky monitor
SARA	sparsity averaging re-weighted analysis
FISTA	Fast Iterative Shrinkage- Thresholding Algorithm
MEM	maximum entropy method
NNLS	non-negative least squares

BIBLIOGRAPHY

- [1] M. P. Van Haarlem *et al.*, “LOFAR: The low-frequency array,” *Astronomy & Astrophysics*, vol. 556, p. A2, 2013.
- [2] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. W. Lazio, “The Square Kilometre Array,” *Proc. IEEE*, vol. 97, no. 8, pp. 1482–1496, 2009.
- [3] J. A. Högbom, “Aperture synthesis with a non-regular distribution of interferometer baselines,” vol. 15, p. 417, 1974.
- [4] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. Lazio, “The Square Kilometre Array,” *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1482–1496, 2009.
- [5] “The multiwave milky way.” https://asd.gsfc.nasa.gov/archive/mmw/mmw_product.html. Accessed: 2018-05-22.
- [6] C. H. Papas, *Theory of electromagnetic wave propagation*. Courier Corporation, 2014.
- [7] “Atmospheric absorption and transmission.” http://gsp.humboldt.edu/olm_2015/Courses/GSP_216_Online/lesson2-1/atmosphere.html. Accessed: 2018-05-22.
- [8] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy*. Wiley-VCH, 2007.
- [9] M. Ryle, “Radio telescopes of large resolving power,” *Reviews of Modern Physics*, vol. 47, no. 3, p. 557, 1975.
- [10] S. Haykin, J. H. Justice, N. L. Owsley, J. Yen, and A. C. Kak, “Array signal processing,” 1985.
- [11] D. R. DeBoer, R. G. Gough, J. D. Bunton, T. J. Cornwell, R. J. Beresford, S. Johnston, I. J. Feain, A. E. Schinckel, C. A. Jackson, M. J. Kesteven, *et al.*, “Australian SKA pathfinder: A high-dynamic range wide-field of view survey telescope,” *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1507–1521, 2009.
- [12] C. J. Lonsdale *et al.*, “The Murchison Widefield Array: Design overview,” *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1497–1506, 2009.
- [13] J. L. Jonas, “MeerKAT—the south african array with composite dishes and wide-band single pixel feeds,” *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1522–1530, 2009.
- [14] D. R. DeBoer *et al.*, “Hydrogen epoch of reionization array (HERA),” *Publications of the Astronomical Society of the Pacific*, vol. 129, no. 974, p. 045001, 2017.

- [15] “Commonwealth scientific and industrial research organisation.” <https://www.csiro.au/>. Accessed: 2018-05-22.
- [16] “SKA telescope.” <https://www.skatelescope.org/>. Accessed: 2018-05-22.
- [17] M. De Vos, A. W. Gunst, and R. Nijboer, “The LOFAR telescope: System architecture and signal processing,” *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1431–1437, 2009.
- [18] R. Braun, E. Keane, T. Bourke, J. A. Green, and J. Wagg, “Advancing astrophysics with the Square Kilometre Array,” *PoS*, p. 174, 2015.
- [19] P. Dewdney, W. Turner, R. Millenaar, R. McCool, J. Lazio, and T. Cornwell, “SKA1 system baseline design,” *Document number SKA-TEL-SKO-DD-001 Revision*, vol. 1, no. 1, 2013.
- [20] R. Jongerius, “Exascale computer system design: the Square Kilometre Array,” *Department of Electrical Engineering, Ph. D. dissertation, Eindhoven Univ. Technol., Eindhoven, the Netherlands*, 2016.
- [21] S. J. Wijnholds, A.-J. van der Veen, F. De Stefani, E. La Rosa, and A. Farina, “Signal processing challenges for radio astronomical arrays,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 5382–5386, IEEE, 2014.
- [22] P. Dewdney, W. Turner, R. Braun, J. Santander-Vela, M. Waterson, and G. Tan, “SKA1 system baseline v2 description,” *SKA Organisation (November 2015). Google Scholar*, 2015.
- [23] “SKA precursor telescope MeerKAT inaugurated in south africa.” <https://www.skatelescope.org/news/ska-precursor-telescope-meerkat-inaugurated-south-africa/>. Accessed: 2018-07-24.
- [24] “SKA1-MID.” <https://www.skatelescope.org/multimedia/image/ska-infographic-ska1-mid/>. Accessed: 2018-5-14.
- [25] “SKA1-LOW.” <https://www.skatelescope.org/multimedia/image/ska-infographic-ska1-low/>. Accessed: 2018-5-14.
- [26] R. Jongerius, S. Wijnholds, R. Nijboer, and H. Corporaal, “An end-to-end computing model for the Square Kilometre Array,” *Computer*, no. 9, pp. 48–54, 2014.
- [27] P. C. Broekema *et al.*, “Dome: towards the ASTRON & IBM center for exascale technology,” in *Proceedings of the 2012 Workshop on High-Performance Computing for Astronomy Data*, pp. 1–4, ACM, 2012.
- [28] R. A. Perley, F. R. Schwab, and A. H. Bridle, “Synthesis imaging in radio astronomy,” 1989.

- [29] M. Atemkeng, O. Smirnov, C. Tasse, G. Foster, and J. Jonas, "Using baseline-dependent window functions for data compression and field-of-interest shaping in radio interferometry," *Monthly Notices of the Royal Astronomical Society*, vol. 462, no. 3, pp. 2542–2558, 2016.
- [30] M. A. Garrett and H. J. van Langevelde, "Radio astronomy course lecture notes," 2015.
- [31] T. Carozzi and G. Woan, "A generalized measurement equation and van cittert-zernike theorem for wide-field radio astronomical interferometry," *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 3, pp. 1558–1568, 2009.
- [32] J. Hamaker, J. Bregman, and R. Sault, "Understanding radio polarimetry. i. mathematical foundations," *Astronomy and Astrophysics Supplement Series*, vol. 117, no. 1, pp. 137–147, 1996.
- [33] O. M. Smirnov, "Revisiting the radio interferometer measurement equation-i. a full-sky jones formalism," *Astronomy & Astrophysics*, vol. 527, p. A106, 2011.
- [34] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, and P. Vandergheynst, "Compressed sensing imaging techniques for radio interferometry," *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 3, pp. 1733–1742, 2009.
- [35] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "PURIFY: a new approach to radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 439, no. 4, pp. 3591–3604, 2014.
- [36] R. Carrillo, J. McEwen, and Y. Wiaux, "Sparsity averaging reweighted analysis (SARA): a novel algorithm for radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 2, pp. 1223–1234, 2012.
- [37] A.-J. Boonstra, *Radio frequency interference mitigation in radio astronomy*. PhD thesis, Delft University of Technology, Delft, The Netherlands, 2005.
- [38] A.-J. van der Veen, A. Leshem, and A.-J. Boonstra, "Array signal processing for radio astronomy," in *The Square Kilometre Array: An Engineering Perspective*, pp. 231–249, Springer, 2005.
- [39] A.-J. van der Veen and S. J. Wijnholds, "Signal processing tools for radio astronomy," in *Handbook of Signal Processing Systems*, pp. 421–463, Springer, 2013.
- [40] S. J. Wijnholds and A.-J. van der Veen, "Fundamental imaging limits of radio telescope arrays," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 5, pp. 613–623, 2008.
- [41] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.

- [42] D. S. Briggs, *High fidelity deconvolution of moderately resolved sources*. PhD thesis, The new Mexico Institute of Mining and Technology, Socorro, New Mexico, 1995.
- [43] S. Van Der Tol and S. J. Wijnholds, "CRB analysis of the impact of unknown receiver noise on phased array calibration," in *Sensor Array and Multichannel Processing, 2006. Fourth IEEE Workshop on*, pp. 185–189, IEEE, 2006.
- [44] S. J. Wijnholds, *Fish-eye observing with phased array radio telescopes*. PhD thesis, Delft University of Technology, Delft, The Netherlands, 2010.
- [45] S. Chiarucci and S. J. Wijnholds, "Blind calibration of radio interferometric arrays using sparsity constraints and its implications for self-calibration," *Monthly Notices of the Royal Astronomical Society*, vol. 474, no. 1, pp. 1028–1040, 2017.
- [46] S. J. Wijnholds and A.-J. Van Der Veen, "Multisource self-calibration for sensor arrays," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3512–3522, 2009.
- [47] G. B. Taylor, C. L. Carilli, and R. A. Perley, "Synthesis imaging in radio astronomy ii," in *Synthesis Imaging in Radio Astronomy II*, vol. 180, 1999.
- [48] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*. CRC press, 1998.
- [49] R. T. Hoctor and S. A. Kassam, "The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging," *Proceedings of the IEEE*, vol. 78, no. 4, pp. 735–752, 1990.
- [50] U. R. Venkata, *Parameterized Deconvolution for Wide-Band Radio Synthesis Imaging*. PhD thesis, The new Mexico Institute of Mining and Technology, Socorro, New Mexico, 2010.
- [51] M. Yashar and A. Kembal, "TDP calibration & processing group CPG memo# 3 computational costs of radio imaging algorithms dealing with the non-coplanar baselines effect: I," 2009.
- [52] D. Romero, D. D. Ariananda, Z. Tian, and G. Leus, "Compressive covariance sensing: Structure-based compressive sensing beyond sparsity," *IEEE signal processing magazine*, vol. 33, no. 1, pp. 78–93, 2016.
- [53] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Transactions on antennas and propagation*, vol. 16, no. 2, pp. 172–175, 1968.
- [54] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE signal processing magazine*, vol. 13, no. 4, pp. 67–94, 1996.
- [55] M. H. Hayes, *Statistical digital signal processing and modeling*. John Wiley & Sons, 2009.
- [56] A. M. Sardarabadi, A. Leshem, and A.-J. van der Veen, "Radio astronomical image formation using constrained least squares and krylov subspaces," *Astronomy & Astrophysics*, vol. 588, p. A95, 2016.

- [57] S. M. Kay, *Fundamentals of statistical signal processing, volume I: estimation theory*. Prentice Hall, 1993.
- [58] R. Kress, V. Maz'ya, and V. Kozlov, *Linear integral equations*, vol. 17. Springer, 1989.
- [59] P. C. Hansen, *Discrete inverse problems: insight and algorithms*, vol. 7. SIAM, 2010.
- [60] E. C. Titchmarsh *et al.*, *Introduction to the theory of Fourier integrals*, vol. 2. Clarendon Press Oxford, 1948.
- [61] R. N. Bracewell and J. A. Roberts, "Aerial smoothing in radio astronomy," *Australian Journal of Physics*, vol. 7, no. 4, pp. 615–640, 1954.
- [62] P. C. Hansen, J. G. Nagy, and D. P. O'leary, *Deblurring images: matrices, spectra, and filtering*, vol. 3. Siam, 2006.
- [63] C. A. Balanis, *Antenna theory: analysis and design*. John Wiley & Sons, 2016.
- [64] J. P. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, vol. 160 of *Applied Mathematical Sciences*. Springer, 2004. ISBN 0-387-22073-9.
- [65] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.
- [66] C. L. Lawson and R. J. Hanson, *Solving least squares problems*, vol. 161. SIAM, 1974.
- [67] B. Clark, "An efficient implementation of the algorithm 'CLEAN'," *Astronomy and Astrophysics*, vol. 89, p. 377, 1980.
- [68] F. R. Schwab, "Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry," *The Astronomical Journal*, vol. 89, pp. 1076–1081, 1984.
- [69] T. J. Cornwell, "Multiscale CLEAN deconvolution of radio synthesis images," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 5, pp. 793–801, 2008.
- [70] B. Wakker and U. Schwarz, "The multi-resolution CLEAN and its application to the short-spacing problem in interferometry," *Astronomy and Astrophysics*, vol. 200, pp. 312–322, 1988.
- [71] A. R. Offringa and O. Smirnov, "An optimized algorithm for multi-scale wideband deconvolution of radio astronomical images," *Monthly Notices of the Royal Astronomical Society*, vol. 471, pp. 301–316, Oct. 2017.
- [72] K. Marsh and J. Richardson, "The objective function implicit in the CLEAN algorithm," *Astronomy and Astrophysics*, vol. 182, pp. 174–178, 1987.
- [73] S. Mallat and Z. Zhang, "Matching pursuits with time frequency dictionaries," vol. 41, pp. 3397–3415, Dec. 1993.

- [74] J. Starck, E. Pantin, and F. Murtagh, "Deconvolution in astronomy: A review," *Publications of the Astronomical Society of the Pacific*, vol. 114, no. 800, p. 1051, 2002.
- [75] A. Onose, R. E. Carrillo, A. Repetti, J. D. McEwen, J.-P. Thiran, J.-C. Pesquet, and Y. Wiaux, "Scalable splitting algorithms for big-data interferometric imaging in the SKA era," *Monthly Notices of the Royal Astronomical Society*, vol. 462, no. 4, pp. 4314–4335, 2016.
- [76] A. Dabbech, A. Onose, A. Abdulaziz, R. A. Perley, O. M. Smirnov, and Y. Wiaux, "Cygnus A super-resolved via convex optimisation from vla data," *arXiv preprint arXiv:1710.08810*, 2017.
- [77] T. J. Cornwell and K. Evans, "A simple maximum entropy deconvolution algorithm," *Astronomy and Astrophysics*, vol. 143, pp. 77–83, 1985.
- [78] R. Narayan and R. Nityananda, "Maximum entropy image restoration in astronomy," *Annual review of astronomy and astrophysics*, vol. 24, no. 1, pp. 127–170, 1986.
- [79] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, "Maximum entropy and the nearly black object," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 41–81, 1992.
- [80] H. Junklewitz, M. Bell, M. Selig, and T. Enßlin, "RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy," *Astronomy & Astrophysics*, vol. 586, p. A76, 2016.
- [81] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Machine Learning Res.*, vol. 1, pp. 211–244, 2001.
- [82] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Proc.*, vol. 52, pp. 2153–2164, Aug. 2004.
- [83] D. Calvetti and E. Somersalo, "Priorconditioners for linear systems," *Inverse problems*, vol. 21, no. 4, p. 1397, 2005.
- [84] S. Berisha and J. G. Nagy, "Iterative methods for image restoration," in *Image, Video Processing and Analysis, Hardware, Audio, Acoustic and Speech Processing* (R. Chellappa and S. Theodoridis, eds.), vol. 4 of *Academic Press Library in Signal Processing*, ch. 7, pp. 193–247, Academic Press, 2013.
- [85] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical mathematics*, vol. 37. Springer Science & Business Media, 2010.
- [86] J. M. Bardsley and J. G. Nagy, "Covariance-preconditioned iterative methods for nonnegatively constrained astronomical imaging," *SIAM journal on matrix analysis and applications*, vol. 27, no. 4, pp. 1184–1197, 2006.
- [87] M. Hanke, *Conjugate gradient type methods for ill-posed problems*. Routledge, 2017.

- [88] D. Calvetti, F. Pitolli, E. Somersalo, and B. Vantaggi, "Bayes meets Krylov: preconditioning cgls for underdetermined systems," *arXiv preprint arXiv:1503.06844*, 2015.
- [89] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Transactions on signal processing*, vol. 60, no. 9, pp. 4775–4789, 2012.
- [90] U. Rau, S. Bhatnagar, M. Voronkov, and T. Cornwell, "Advances in calibration and imaging techniques in radio interferometry," *Proc. IEEE*, vol. 97, pp. 1472–1481, Aug 2009.
- [91] S. Yatawatta, "Adaptive weighting in radio interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 444, no. 1, pp. 790–796, 2014.
- [92] A. Onose, A. Dabbech, and Y. Wiaux, "An accelerated splitting algorithm for radio-interferometric imaging: when natural and uniform weighting meet," *Monthly Notices of the Royal Astronomical Society*, vol. 469, no. 1, pp. 938–949, 2017.
- [93] S. J. Wijnholds and A.-J. Van der Veen, "Data driven model based least squares image reconstruction for radio astronomy," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 2704–2707, IEEE, 2011.
- [94] M. Bertero, C. De Mol, and E. R. Pike, "Linear inverse problems with discrete data: Ii. stability and regularisation," *Inverse problems*, vol. 4, no. 3, p. 573, 1988.
- [95] S. Twomey, "The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements," *Journal of the Franklin Institute*, vol. 279, no. 2, pp. 95–109, 1965.
- [96] D. P. Bertsekas, *Nonlinear programming*. Athena scientific Belmont, 1999.
- [97] U. Rau and T. J. Cotton, "A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry," *Astronomy & Astrophysics*, vol. 532, p. A71, 2011.
- [98] J. McEwen and Y. Wiaux, "Compressed sensing for wide-field radio interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 413, no. 2, pp. 1318–1332, 2011.
- [99] A. Dabbech, C. Ferrari, D. Mary, E. Slezak, O. Smirnov, and J. Kenyon, "MORE-SANE, model reconstruction by synthesis-analysis estimators," *Astronomy and Astrophysics*, vol. 576, 2014.
- [100] J. Girard, H. Garsden, J. Girard, J.-L. Starck, S. Corbel, A. Woiselle, C. Tasse, J. McKean, and J. Bobin, "Sparse representations and convex optimization as tools for LOFAR radio interferometric imaging," in *2nd Int. Summer School on Intelligent Signal Processing for Frontier Research and Industry*, (Univ. Paris-Diderot Campus, Paris, France), IOP Publishing, Aug. 2015.
- [101] P. L. Combettes and J.-C. Pesquet, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, ch. Proximal splitting methods in signal processing, pp. 185–212. Springer, New York, 2011.

- [102] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Proc.*, vol. 45, pp. 600–616, Mar. 1997.
- [103] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.
- [104] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM transactions on mathematical software*, vol. 8, no. 1, pp. 43–71, 1982.
- [105] Y. Saad, "Krylov subspace methods for solving large unsymmetric linear systems," *Mathematics of computation*, vol. 37, no. 155, pp. 105–126, 1981.
- [106] S. Naghibzadeh, A. M. Sardarabadi, and A.-J. van der Veen, "Radioastronomical image reconstruction with regularized least squares," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3316–3320, IEEE, 2016.
- [107] S. Naghibzadeh and A.-J. van der Veen, "Radioastronomical least squares image reconstruction with iteration regularized Krylov subspaces and beamforming-based prior conditioning," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3385–3389, IEEE, Mar. 2017.
- [108] S. Naghibzadeh and A.-J. van der Veen, "Fast and accurate radio interferometric imaging using Krylov subspaces," in *2017 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 173–177, IEEE, Dec. 2017.
- [109] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Acoustics, speech and signal processing, 2008. ICASSP 2008. IEEE international conference on*, pp. 3869–3872, IEEE, 2008.
- [110] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.
- [111] M. Hanke, *Conjugate gradient type methods for ill-posed problems*, vol. 327. CRC Press, 1995.
- [112] A. S. Nemirovskii, "The regularizing properties of the adjoint gradient method in ill-posed problems," *USSR Computational Mathematics and Mathematical Physics*, vol. 26, no. 2, pp. 7–16, 1986.
- [113] P. C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, vol. 4. Siam, 2005.
- [114] S.-C. T. Choi, *Iterative methods for singular linear equations and least-squares problems*. PhD thesis, Stanford University, 2006.

- [115] G. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix," *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, vol. 2, no. 2, pp. 205–224, 1965.
- [116] V. A. Morozov, *Methods for solving incorrectly posed problems*. Springer Science & Business Media, 2012.
- [117] T. K. Jensen and P. C. Hansen, "Iterative regularization with minimum-residual methods," *BIT Numerical Mathematics*, vol. 47, no. 1, pp. 103–120, 2007.
- [118] T. J. Cornwell, K. Golap, and S. Bhatnagar, "The noncoplanar baselines effect in radio interferometry: The w-projection algorithm," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 5, pp. 647–657, 2008.
- [119] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [120] W. H. Richardson, "Bayesian-based iterative method of image restoration," *JOSA*, vol. 62, no. 1, pp. 55–59, 1972.
- [121] M. Hanke, J. G. Nagy, and C. Vogel, "Quasi-newton approach to nonnegative image restorations," *Linear Algebra and its applications*, vol. 316, no. 1, pp. 223–236, 2000.
- [122] P. Prasad and S. J. Wijnholds, "AARTFAAC: Towards a 24x7, all-sky monitor for LO-FAR," *arXiv preprint arXiv:1205.3056*, 2012.
- [123] S. Chen, D. L. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
- [124] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [125] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, p. 947, 2007.
- [126] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on signal processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [127] L. Schwardt, "Compressed sensing and the CLEAN algorithm," April 2008.
- [128] U. Schwarz, "Mathematical-statistical description of the iterative beam removing technique (method CLEAN)," *Astronomy and Astrophysics*, vol. 65, p. 345, 1978.
- [129] S. Foucart and D. Koslicki, "Sparse recovery by means of nonnegative least squares," *IEEE Signal processing Letters*, vol. 21, no. 4, pp. 498–502, 2014.
- [130] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

- [131] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [132] H. Garsden, J. Girard, J.-L. Starck, S. Corbel, C. Tasse, A. Woiselle, J. McKean, A. S. Van Amesfoort, J. Anderson, I. Avruch, *et al.*, "LOFAR sparse image reconstruction," *Astronomy & Astrophysics*, vol. 575, p. A90, 2015.
- [133] Y. Liu, Z. Zhan, J.-F. Cai, D. Guo, Z. Chen, and X. Qu, "Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging," *IEEE transactions on medical imaging*, vol. 35, no. 9, pp. 2130–2140, 2016.
- [134] E. Pantin and J.-L. Starck, "Deconvolution of astronomical images using the multi-scale maximum entropy method," *Astronomy and Astrophysics Supplement Series*, vol. 118, no. 3, pp. 575–585, 1996.
- [135] S. Yatawatta, "Fundamental limitations of pixel based image deconvolution in radio astronomy," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2010 IEEE*, pp. 69–72, IEEE, 2010.
- [136] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, 2013.
- [137] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comp. Math.*, vol. 38, no. 3, pp. 667–681, 2013.
- [138] J.-C. Pesquet and A. Repetti, "A class of randomized primal-dual algorithms for distributed optimization," *Journal of Nonlinear and Convex Analysis*, vol. 16, no. 12, pp. 2353–2490, 2015.
- [139] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal dual approaches for solving large-scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, 2015.
- [140] L. Kogan and E. W. Greisen, "Faceted imaging in aips," *AIPS memo*, 2009.
- [141] C. Tasse, B. Hugo, M. Mirmont, O. Smirnov, M. Atemkeng, L. Bester, M. Hardcastle, R. Lakhoo, S. Perkins, and T. Shimwell, "Faceting for direction-dependent spectral deconvolution," *Astronomy & Astrophysics*, vol. 611, p. A87, 2018.
- [142] A. Repetti, J. Birdi, A. Dabbech, and Y. Wiaux, "Non-convex optimization for self-calibration of direction-dependent effects in radio interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 470, no. 4, pp. 3981–4006, 2017.
- [143] L. Greengard and J.-Y. Lee, "Accelerating the nonuniform fast fourier transform," *SIAM review*, vol. 46, no. 3, pp. 443–454, 2004.

- [144] J. Fessler and B. Sutton, "Nonuniform fast fourier transforms using min-max interpolation," *IEEE Tran. Sig. Proc.*, vol. 51, pp. 560–574, Feb 2003.
- [145] A. Dabbech, L. Wolz, L. Pratley, J. D. McEwen, and Y. Wiaux, "The w-effect in interferometric imaging: from a fast sparse measurement operator to superresolution," *Monthly Notices of the Royal Astronomical Society*, vol. 471, no. 4, pp. 4300–4313, 2017.
- [146] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [147] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [148] S. J. Hardy, "Direct deconvolution of radio synthesis images using L1 minimisation," *A&A*, vol. 557, p. A134, 2013.
- [149] I. Daubechies, *Ten lectures on wavelets*, vol. 61. Siam, 1992.
- [150] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "A block coordinate variable metric forward-backward algorithm," *Journal of Global Optimization*, vol. 66, no. 3, pp. 457–485, 2016.
- [151] S. J. Wijnholds, S. Van Der Tol, R. Nijboer, and A.-J. van der Veen, "Calibration challenges for future radio telescopes," *IEEE Signal Processing Magazine*, vol. 27, no. 1, 2010.
- [152] S. Salvini and S. J. Wijnholds, "High-compression baseline dependent averaging," in *General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS), 2017 XXXIInd*, pp. 1–4, IEEE, 2017.
- [153] M. Atemkeng, O. Smirnov, C. Tasse, G. Foster, A. Keimpema, Z. Paragi, and J. Jonas, "Baseline-dependent sampling and windowing for radio interferometry: data compression, field-of-interest shaping, and outer field suppression," *Monthly Notices of the Royal Astronomical Society*, vol. 477, no. 4, pp. 4511–4523, 2018.
- [154] S.-C. T. Choi, C. C. Paige, and M. A. Saunders, "MINRES-QLP: A Krylov subspace method for indefinite or singular symmetric systems," *SIAM Journal on Scientific Computing*, vol. 33, no. 4, pp. 1810–1836, 2011.
- [155] A. R. Offringa *et al.*, "WSCLEAN: an implementation of a fast, generic wide-field imager for radio astronomy," *Monthly Notices of the Royal Astronomical Society*, vol. 444, no. 1, pp. 606–619, 2014.