

Imperial College London
Department of Electrical and Electronic Engineering

Gridless processing for multi-band MIMO sensing

Nuofan Mao
Supervised by Dr. Wei Dai

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Electrical and Electronic Engineering of the Imperial College
London and the Diploma of Imperial College London, September 2024

Statement of Originality

I declare that this thesis is original and is the result of my own work. In the thesis, material, including information and ideas, derived from the published and unpublished work of others has been fully acknowledged. The content of the thesis has not been and will not be submitted to any other institution or university for fulfilling the requirements of another degree.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

In the field of multiple-input multiple-output (MIMO) sensing, traditional studies assume the probing signals are narrowband. With modern 5G/6G and sensing systems operating on multi-bands, models and methods tailored for multi-band active sensing are necessary. This thesis develops the model and method tailored for MIMO sensing with multi-band transmitted signals. The multi-band array can jointly sense gridless direction-of-arrivals (DoAs) and spectral responses of targets. In this thesis, we use the greatest common divisor (GCD) of multi-band for joint signal processing. The spectral responses can be different at different frequencies. We show that the gridless DoA and spectral response can be estimated jointly in numerical simulations. Due to the multi-band, the number of targets can be more than the antennas even for uniform linear arrays (ULA), which has exceeded the traditional limit of maximum resolvable targets. Subsequently, this thesis discusses the DoAs and ranges sensing in frequency diversity arrays (FDA). We formulate the problem as low rank matrix recovery problem, and solve it. We show that DoAs and ranges can be jointly sensed with a limited number of snapshots and low computational complexity. Finally, this thesis further discuss nonconvex nonsmooth optimisation problems. We analyse the algorithms: proximal gradient method, PDOM and ADMM.

Acknowledgements

First of all, I would like express my gratitude to my supervisor, Dr. Wei Dai, for his continuous supervision and support during my PhD research. He encourages me to become an independent researcher. "Learning by doing" is the most important idea I learned from him. I learned a lot from both group and individual meetings. The group lunches he organised made my PhD life happy. With his help, I can successfully attend two academic conferences: CAMSAP in 2023 and SPAWC in 2024. I deeply appreciate his dedication and guidance.

I would like to express my gratitude to all of my colleagues in the Communications and Signal Processing group, for both personal and academical help. I would also want to thank all of members in the research group of my supervisor. I learned a lot from them.

Next, I want to express deep gratitude to my family, especially my parents, for their countless support and love in my life. I cherish every reunion with them. When I have issues in my life, they can always provide their advice from their personal experience. My parents always remind me of the importance of the physical and mental health, which benefits me a lot.

Finally, I want to express my special thanks to my maternal grandparents : Zhibin Deng (1938 - 2023) and Chunxian Pu (1942 - 2010). They always accompanied and cared me during my childhood. They provided me with endless love. Although they have passed away, I deeply love them forever.

"He that falls in love with himself will have no rivals."

- Benjamin Franklin

Contents

Statement of Originality	i
Copyright Declaration	ii
Abstract	iii
Acknowledgements	iv
1 Introduction	1
1.1 Overview	1
1.1.1 MIMO array sensing	2
1.1.2 Gridless processing	2
1.1.3 Bandwidth	3
1.1.4 Frequency diversity array	3
1.1.5 Sparse signal recovery problem	4
1.2 Research objectives	5
1.3 Contributions and outline	6
1.4 Notations	7
1.5 Publications	7
2 Background	9
2.1 Direction-of-Arrival estimation	9
2.1.1 Steering vector and 1-D array geometry	9
2.1.2 Signal model in ULA	11
2.1.3 Identifiability and resolution in ULA	12
2.1.4 Sparse linear array	12

2.1.5	MUSIC	13
2.2	Structured matrix and decomposition	14
2.2.1	Hankel matrix	14
2.2.2	SVD	15
2.3	Nonconvex and Nonsmooth Optimisation	16
2.3.1	ADMM	16
2.3.2	Majorization Minimization	17
2.3.3	Line Search Method	18
2.3.4	Proximal Gradient Method	19
2.3.5	PDOM	21
3	Multi-band Active Array	23
3.1	Introduction	23
3.2	System Model and Assumptions	26
3.2.1	The Standard Model	26
3.2.2	Model for Multi-band ULA	27
3.2.3	Model for Sparse Linear Arrays	27
3.2.4	Multiple Snapshots	28
3.3	Some Useful Concepts and Assumptions	29
3.3.1	Spectral Response Model for Targets	29
3.3.2	Hankel Matrix	30
3.3.3	Greatest Common Divisor (GCD) of Rational Numbers	31
3.3.4	Assumptions	32
3.4	DoAs and Spectral Responses Sensing	33
3.4.1	Linear Operator Notations	33
3.4.2	One Vector Fits All	33
3.4.3	System Model Reformulation	34
3.4.4	Low Rank Square Hankel Matrix Formulation	35
3.4.5	Optimisation Formulation and Solving	36
3.4.6	Target Sensing	37
3.5	Discussion	38
3.5.1	What Defines a Wideband for an Array	38

3.5.2	Why can we sense more targets than the number of physical antennas in ULA?	40
3.6	Numerical Tests	43
3.6.1	Comparison with NNM	43
3.6.2	Type of Target Localisation and Identification	45
3.6.3	More Targets Sensed by the Less Number of Antennas	46
3.6.4	Wideband Versus Narrowband in terms of the number of antennas in ULA	47
3.7	Summary	49
4	Frequency Diversity Array	51
4.1	Introduction	51
4.2	System Model	52
4.3	Low Rank Matrix Formulation	54
4.3.1	DoA and Range Variables Decoupling	54
4.3.2	Hankel Matrix Formulation	55
4.3.3	Problem Solving	57
4.3.4	MUSIC and identifiability	60
4.4	Numerical Tests	61
4.4.1	Joint Sensing Performance	61
4.4.2	RMSE versus SNR	61
4.5	Summary	64
5	Nonconvex Algorithms Analysis	65
5.1	Introduction	65
5.2	Proximal gradient method and PDOM	66
5.2.1	Introduction	66
5.2.2	Proximal Gradient Method details	66
5.2.3	PDOM details	68
5.2.4	Cramér–Rao Bound Derivation	70
5.2.5	Numerical tests	74
5.3	ADMM in nonconvex optimisation	76
5.3.1	Introduction	76
5.3.2	Short-and-sparse blind deconvolution formulation	78

5.3.3	Problem solving	79
5.3.4	Numerical Test: image deblurring	80
5.4	Summary	81
6	Conclusion	83
6.1	Summary of Thesis Achievements	83
6.2	Applications	84
6.3	Future Work	85
	Bibliography	85
	Appendices	94

List of Tables

3.1	MSE of both DoAs and spectral response, $(M_T, M_R, S) = (1, 6, 3), SNR = 10dB, L = 10, S = 3$, Frequency set is $\mathcal{F}_1$. DoA set $\boldsymbol{\theta} = \{-60^\circ, 10^\circ, 45^\circ\}$	47
5.1	PSNR and average running time for image deblurring	81

List of Figures

1.1	NASA Applied Remote Sensing Training Program [1].	3
2.1	The setup of DoA problem of one target in ULA. As the targets is far-field, the DoA for each antenna is the same.	10
2.2	An example of SLA from 6-antenna ULA. Black points are the physical antennas. White points are holes, which means no antenna in that position.	13
3.1	Active sensing	25
3.2	An example of virtual array for $(M_T, M_R) = (1, 6)$. White points are holes, which means no antenna in that position. The antenna spacing $d = \lambda_{max}/2$. (a) Single frequency case $\mathcal{F}_s = 1$ GHz. (b) Multiple frequency case $\mathcal{F}_m = \{1, 2\}$ GHz.	42
3.3	Resolution comparison in terms of formulation, with one snapshot, $M_T = M_R = 5$, $K = 1$, SNR = 50 dB: (a) (b) $\mathcal{F} = \{1\}$ kHz, the DoA $\boldsymbol{\theta} = \{0^\circ, 15^\circ\}$. (c) (d) $\mathcal{F} = \{1, 3\}$ kHz, the DoA $\boldsymbol{\theta} = \{0^\circ, 8^\circ, 27^\circ, 46^\circ\}$	44
3.4	Target localisation plot for $(M_T, M_R, S) = (1, 6, 3)$ with only $L = 10$ snapshots and frequency set $\mathcal{F}_1$. DoA set $\boldsymbol{\theta} = \{-60^\circ, 10^\circ, 45^\circ\}$. SNR= 10 dB. (a) $k = 1$. $MSE_\theta = 0.04, MSE_\Gamma = 2.57 \times 10^{-4}$. (b) $k = 2$. $MSE_\theta = 0.39, MSE_\Gamma = 5.97 \times 10^{-3}$	46
3.5	MUSIC spectrum for more targets ($S = 7$) sensed by the less number of antennas $(M_T, M_R) = (1, 6)$ in ULA with only 30 snapshots. (Blue: $\mathcal{F}_1, (MSE_\theta, MSE_\Gamma) = (1.73, 7.21 \times 10^{-3})$. Orange: $\mathcal{F}_2, (MSE_\theta, MSE_\Gamma) = (146.44, 1.91)$). DoA set $\boldsymbol{\theta} = \{-49^\circ, -28^\circ, -12^\circ, 2^\circ, 15^\circ, 32^\circ, 55^\circ\}$. The spectral response is independent with frequencies ($k = 1$). SNR = 10 dB.	48

- 3.6 MUSIC spectrum for more targets ($S = 6$) sensed by the less number of antennas (M_T, M_R) = (1, 6) in SLA with only 40 snapshots. (Blue: $\mathcal{F}_1, (MSE_\theta, MSE_r) = (3.21, 0.02)$. Orange: $\mathcal{F}_2, (MSE_\theta, MSE_r) = (32.58, 0.02)$). DoA set $\boldsymbol{\theta} = \{-49^\circ, -12^\circ, 2^\circ, 15^\circ, 32^\circ, 55^\circ\}$. The spectral response is independent with frequencies ($k = 1$). SNR = 10 dB. 48
- 3.7 The probability of success cases ($RMSE_\theta < 1$) in terms of the number of antennas in ULA with $M_T = M_R = 6$. $L = 10$. $\mathcal{F}_3 = \{1.0, 1.5\}$ GHz. $K = 1$. SNR = 30 dB. From (3.40). From our wideband criteria (3.32), $\mathcal{F}_3$ is wideband when $M \geq 4$. The wideband case can detect two closely targets with higher probability. 49
- 4.1 Co-located transmitters and receivers with M -element in FDA, and their corresponding transmitted carrier frequencies. 52
- 4.2 MUSIC spectrum for the targets sensed by 10-antenna FDA with only $L = 10$ snapshots ($M_T = M_R = 10$). The carrier frequency $f_1 = 1$ GHz. The frequency increment $\Delta f = 75$ MHz. The groundtruth DoAs $\boldsymbol{\theta} = \{-40^\circ, 20^\circ\}$ and range $\mathbf{r} = \{0.2, 1.5\}$ km. SNR = 30 dB. ($RMSE_\theta, RMSE_r$) = (0.99, 5.59×10^{-2}). . . 62
- 4.3 MUSIC spectrum for the targets sensed by 10-antenna FDA with only $L = 10$ snapshots ($M_T = M_R = 10$). The carrier frequency $f_1 = 1$ GHz. The frequency increment $\Delta f = 75$ MHz. The groundtruth DoAs $\boldsymbol{\theta} = \{-40^\circ, 5^\circ, 20^\circ\}$ and range $\mathbf{r} = \{0.4, 1.1, 1.8\}$ km. SNR = 30 dB. ($RMSE_\theta, RMSE_r$) = (1.84, 5.59×10^{-2}). . 63
- 4.4 RMSE versus SNR in a limited snapshots $L = 10$ and $L = 5$. The carrier frequency $f_1 = 1$ GHz. The frequency increment $\Delta f = 75$ MHz. The groundtruth DoAs $\boldsymbol{\theta} = \{-40^\circ, 20^\circ\}$ and range $\mathbf{r} = \{0.2, 1.5\}$ km. 64
- 5.1 The value of objective function versus the number of iterations for various penalty parameter α in (3.28). (M_T, M_R) = (1, 20), $L = 5$, $N = 1$, $K = 1$, $\boldsymbol{\theta} = \{-20^\circ, 45^\circ, 60^\circ\}$. SNR = 30 dB. 75
- 5.2 The value of scaled f_1 (5.35): αf_1 versus the number of iterations for various penalty parameter α . (M_T, M_R) = (1, 20), $L = 5$, $N = 1$, $K = 1$, $\boldsymbol{\theta} = \{-20^\circ, 45^\circ, 60^\circ\}$. SNR = 30 dB. 75
- 5.3 The value of f_2 (5.36) versus the number of iterations for various penalty parameter α . (M_T, M_R) = (1, 20), $L = 5$, $N = 1$, $K = 1$, $\boldsymbol{\theta} = \{-20^\circ, 45^\circ, 60^\circ\}$. SNR = 30 dB. 76

5.4	RMSE (in degrees) of estimating algorithms versus SNR for the single frequency. $\mathcal{F} = 1GHz, (M_T, M_R, S) = (1, 10, 3), K = 1, L = 3$. DoA set $\boldsymbol{\theta} = \{-30^\circ, 5^\circ, 45^\circ\}$	77
5.5	RMSE (in degrees) of estimating algorithms versus SNR for the two frequencies. $\mathcal{F} = \{1, 2\}GHz, (M_T, M_R, S) = (1, 5, 2), K = 1, L = 10$. DoA set $\boldsymbol{\theta} = \{-20^\circ, 25^\circ\}$	77

Chapter 1

Introduction

1.1 Overview

With the development of 5G and 6G communication [2], the problem of target sensing has been a great research interest. The applications include wireless communication [3], autonomous driving [4], and medical imaging. To sense the targets accurately, the following parameters are important: Direction-of-Arrival (DoA), spectral response and range.

The DoA sensing is the process that retrieving the direction information of receiving signals from the output of antenna arrays [5]. Since the early 1900s, there has been a great progress in this field. Some classical methods include multiple signal classification (MUSIC) [6] and estimation of signal parameters via rotational invariant techniques (ESPRIT) [7]. Apart from the DoA of targets, the range information can be another important parameters for targets sensing. There has been many classic probing signals to sense the range, such as frequency-modulated continuous wave (FMCW) and Orthogonal frequency-division multiplexing (OFDM). If we want to distinguish the types of targets, the spectral response of targets is also required. The spectral response sensing has been applied in the field of geoscience and remote sensing [8] [9].

In this thesis, we focus on the target sensing by using 1-D antenna arrays. 1-D antenna arrays is convenient to manufacture. Compared with 2-D and 3-D antenna arrays, 1-D antenna arrays only need to process parameters sensing in one direction. It means that 1-D antenna arrays

have lower computational burden. The programming language we use is Julia [10], which is an open source programming language. The editor is chosen as Visual Studio Code. We will review the MIMO array and bandwidth in the following subsections.

1.1.1 MIMO array sensing

Multiple-input multiple-output (MIMO) array is an array that has multiple transmitters and receivers. MIMO array has played a key role in the field of wireless communication. The first generation of MIMO array is only one antenna, which serves as transmitter and receiver. It is called as single-input single-output (SISO). Compared with SISO, the benefit of MIMO is its better spatial resolution. The number of resolvable targets can be improved as well [11]. The transmitted vector-signals have more diversity to sense targets. The FMCW and OFDM are two signals, which has been widely applied in industries. In this thesis, we focus on the array with colocated antennas, which means the transmitters and receivers are in the same array. We will show discuss both SIMO(single-input multiple-output) and MIMO array.

Passive and active sensing: In MIMO array, there are two types of sensing technique: passive and active sensing. Passive sensing technique is that antennas can only receive the signals from targets. In contrast, active sensing technique is that antennas can illuminate targets to sense them. Active sensing has been widely applied in the remote sensing, such as multispectral and hyperspectral imaging. For example, National Aeronautics and Space Administration (NASA) use active sensing in satellite, which is shown in Fig.1.1. In this thesis, we focus on the active sensing.

1.1.2 Gridless processing

The sparse methods for MIMO sensing can be classified into three types: on-grid, off-grid and gridless [5]. The on-grid sparse method assume parameters to be estimated should be on a set of fixed grid points. It means the parameters are on discrete domain. Off-grid can allow the parameters not to be on the grid, but the grid is still required. For gridless processing, there is no grid for parameters. The parameters are totally on continuous domain.

Active Sensors

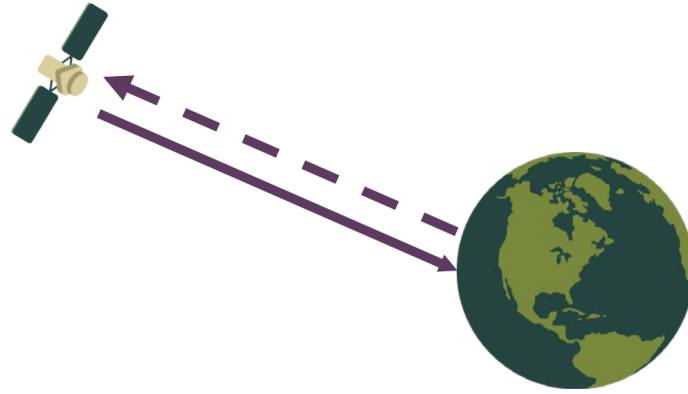


Figure 1.1: NASA Applied Remote Sensing Training Program [1].

1.1.3 Bandwidth

The bandwidth W is defined as the frequency range from lowest frequency f_L to highest frequency f_H [12]. Traditionally, there is no unified criteria of narrowband and wideband. For ultra-wideband, FCC (Federal Communications Commission) [13] defines it is ultra-wideband when the bandwidth is greater than 0.2 of the central frequency $(f_H - f_L)/2$ or 500 MHz. Wideband signals can carry more diversity of information than narrowband signals. In this thesis, we will use wideband signals in the chapter 3, and narrowband in the chapter 4.

1.1.4 Frequency diversity array

Although MIMO array received many research interest, the range-dependent targets cannot be sense accurately. It is because the DoA and range cannot be unambiguously sensed from the beamforming peaks [14]. To solve this issue, one of practical methods is using frequency diversity array (FDA) in MIMO [15]. FDA is an array that use a small frequency increment Δf from antenna-to-antenna when transmitting signals.

It is noted that there is one difference between Frequency Diversity Array (FDA) and OFDM-MIMO systems. Each antenna of FDA transmits signals with a single frequency. In contrast, each antenna of OFDM-MIMO transmits signals with all frequencies. The phase difference

between adjacent antennas is:

$$\Delta\phi = \frac{2\pi f d \sin \theta}{c} - \frac{2\pi r \Delta f}{c} + \frac{2\pi \Delta f d \sin \theta}{c}, \quad (1.1)$$

where f is the carrier frequency at the antenna with low frequency. d is the spacing between adjacent antennas. c is the propagation speed of transmitted signals in the medium. θ and r are the DoA and range of one of targets. In (1.1), the first term $\frac{2\pi f d \sin \theta}{c}$ is the phase difference in traditional MIMO array. Its derivation is shown in section 2.1. This term depends on the DoA. The second term $-\frac{2\pi r \Delta f}{c}$ shows that the phase difference also depends on the range and frequency increment. The first two terms are also distinct with each other. The third term $\frac{2\pi \Delta f d \sin \theta}{c}$ is always assumed to be negligible as $\Delta f \ll f$. This assumption imply our signal is narrowband. [15] shows that both the DoA and range is more sensitive to detect in beam pattern, compared with traditional MIMO arrays.

1.1.5 Sparse signal recovery problem

Signal recovery problem is one of important problems in signal processing. Given adequate measurements, signals are not hard to be recovered. In many real applications, such as medial imaging and radar imaging, only a limited number of measurements can be collected. In this thesis, we model such problem as sparse signal recovery problem. Usually, the sparse signals can be modelled as the superposition of sinusoids [16]. The observed signal $\mathbf{x} \in \mathbb{C}^M$ at the index t is

$$\mathbf{x}[t] = \sum_{s=1}^S c_s e^{-j2\pi f_s t}, \quad t = 0, \dots, M-1 \quad (1.2)$$

where c_s is the amplitude of s -th sinusoids. To recover the S sinusoids from $\mathbf{x}$, the frequency $f_s \in [0, 1)$ is the key parameter to be estimated. This is because f_s is highly nonlinear with the sinusoids. Now the sparse signal recovery problem is converted to the frequency estimation problem. If $S < M$, there has been many successful methods to recover the frequency f_s from $\mathbf{x}$, such as Prony's method and MUSIC [17]. In chapter 3, we model our DoA sensing problem as sparse signal recovery problem. In chapter 4, we model our DoA and range sensing problem

as sparse signal recovery problem as well, which is the joint frequency estimation.

1.2 Research objectives

As the thesis title "Gridless processing for multi-band MIMO sensing" mentions, the thesis aims to discuss the MIMO sensing with multi-band transmitted signals. Although there is a great progress in the field of MIMO sensing, there are still some problems. The problems and our corresponding research objectives are the following.

- **Multi-band:** Although there are many studies discussing the sensing with multi-band signals [18], some studies focus on specific signals, such as FMCW. Previous work always used the linear approximation or beamforming to cope with multi-band. However, these methods require a lot of snapshots. The multi-band can provide more information diversity compared with single-band [19]. Without any transmitted signals assumption, how can we use every band efficiently? In this thesis, we aim to use all frequency bands efficiently to sense targets. The diversity information provided by multi-band is also used to improve estimation ability. As a result, the angle, range and spectral response of targets can be estimated.
- **MIMO array sensing:** The main studies about spectral response sensing is in the field of geoscience and remote sensing. There are few studies about spectral response sensing in MIMO arrays. The other problem is that how multi-band signals can improve the performance of MIMO array sensing. We aim to use multi-band signals to sense spectral response and other parameters in MIMO.
- **Fast sensing:** We aim to use a limited number of data to sense targets with high accuracy. In this thesis, we call it as fast sensing. Traditional MIMO sensing requires sufficient measurements. For example, the Fourier-based spectral analysis need sufficient measurements to do Fourier transform. Inspired from compressive sensing, we can highly reduce the number of measurements by using sparsity-based methods.

1.3 Contributions and outline

The contributions of this thesis can be summarised as follows:

- **Multi-band signals processing:** In the chapter 3, we use Greatest Common Divisor (GCD) to process the multiple bands in MIMO arrays. The GCD can fit all bands into one steering vector. This method can convert the problem to the more virtual antennas with single frequency. Due to the diversity provided by multi-band, the more parameters can be sensed compared with single-band signals.
- **Multiple parameters sensing:** In the chapter 3, our multi-band array can estimate both DoAs and the target spectral response. In the chapter 4, we can sense DoAs and ranges with lower computational complexity in FDA.
- **Fast sensing:** In the chapter 3 and 4, we can use a limited number of snapshots to sense targets in both multi-band arrays and FDA. Inspired by the concept of compressive sensing, we construct low rank Hankel matrix to include the unknown parameters. We formulate the sensing problem as the low rank matrix recovery problem. As the problem is nonconvex and nonsmooth, we apply proximal gradient method to solve it.

The outline of this thesis is the following:

- **Chapter 2: Background:** In this chapter, we review related concepts and techniques for MIMO sensing, including DoA, structured matrix and nonconvex and nonsmooth optimisation. For structured matrix, we focus on Hankel matrix. For nonconvex and nonsmooth optimisation, we review the ADMM method. Followed by majorization minimization and line search method. Based on these two methods, we review proximal gradient method and PDOM.
- **Chapter 3: Multi-band Active Array:** In this chapter, we use MIMO arrays to sense both DoAs and spectral response. We firstly model the multi-band array with active sensing. We also model the spectral response for targets. Then we use GCD of all bands to fit all bands in one steering vector. We reformulate the system model, and formulate the problem as low rank matrix recovery problem. We use both proximal gradient method

to solve our problem, and compare the results. The benefits of our multi-band active array is the following. (a) The number of targets can be more than the antennas in ULA. (b) The array can sense DoA and spectral response of targets. (c) Only limited number of snapshots is required.

- **Chapter 4: Frequency Diversity Array:** In this chapter, we use frequency diversity array (FDA) to sense DoAs and range. Similarly to the chapter 3, we formulate the problem as low rank matrix recovery problem. The benefits of our FDA can sense DoAs and range with lower computational complexity and limited snapshots.
- **Chapter 5: Nonconvex Algorithms Analysis:** In this chapter, we further discuss the nonconvex and nonsmooth optimisation. We discuss proximal gradient method, PDOM and ADMM. We compare proximal gradient method with PDOM. For ADMM, We use a blind deconvolution to show it can also be applied in nonconvex problems.
- **Chapter 6: Conclusion:** In this chapter, we summarise this thesis achievements, applications, and point out the corresponding future work.

1.4 Notations

We list the following notations that will be used in this thesis. We denotes uppercase letter as matrix, and lowercase letter as vectors. $\mathbb{R}$ denotes real number set. $\mathcal{A}(\cdot)$ denotes linear operator. For norms, $\|\cdot\|_2$ denotes Euclidean norm and $\|\cdot\|_F$ denotes Frobenius norm. $\|\cdot\|_0$ denotes pseudo-norm for vectors. $\|\cdot\|_{r,2,0}$ denotes Euclidean norm for each row in a matrix, and pseudo-norm for the vectors with results of each row. $\mathbf{1}$ denotes the matrix whose all elements are 1. $\mathbf{A}^T$ is the transpose of matrix $\mathbf{A}$. $\mathbf{A}^H$ is the conjugate transpose of $\mathbf{A}$. $\bar{\mathbf{A}}$ is the real part of $\mathbf{A}$. $\tilde{\mathbf{A}}$ is the imaginary part of $\mathbf{A}$.

1.5 Publications

Until September 15, 2024, the following papers have been published or accepted:

- N. Mao and W. Dai, "A Multi-Band Active Array for Direction-of-Arrival Estimation",

2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Herradura, Costa Rica, 2023, pp. 341-345, doi: 10.1109/CAMSAP58249.2023.10403467.

- N. Mao and W. Dai, ” Multi-band Active Arrays for Gridless DoA and Spectral Response Sensing,” 2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2024), Lucca, Italy, 2024. (Accepted)

Chapter 2

Background

2.1 Direction-of-Arrival estimation

The far-field direction-of-arrival (DoA) estimation by arrays has been a great research interest for decades. It has been widely applied in a range of domains, such as wireless communication [3] and autonomous driving [4]. To solve the DoA, the most classical method is multiple signal classification (MUSIC) [6]. The other classical method is estimation of signal parameters via rotational invariant techniques (ESPRIT) [7].

2.1.1 Steering vector and 1-D array geometry

We start from uniform linear array (ULA), which means the all antennas are equally spaced in 1-D linear array. Before we derive the signal model of received signal, it is necessary to show the steering vector $\mathbf{a}(\theta)$. Steering vector is an vector that can represent the DoA of a target in an antenna array. $\theta \in [-90^\circ, 90^\circ)$ represents the angle of the the target. Considering an ULA with M antennas. The spacing between adjacent antennas is d .

We assume that the targets are far-field, so the DoA θ are all the same for each antenna in the ULA. This setup is shown in Fig. 2.1. In Fig.2.1, we set the first antenna as our reference antenna. At the time index l , the signal from the targets is $x(l)$. The noise-free received signal

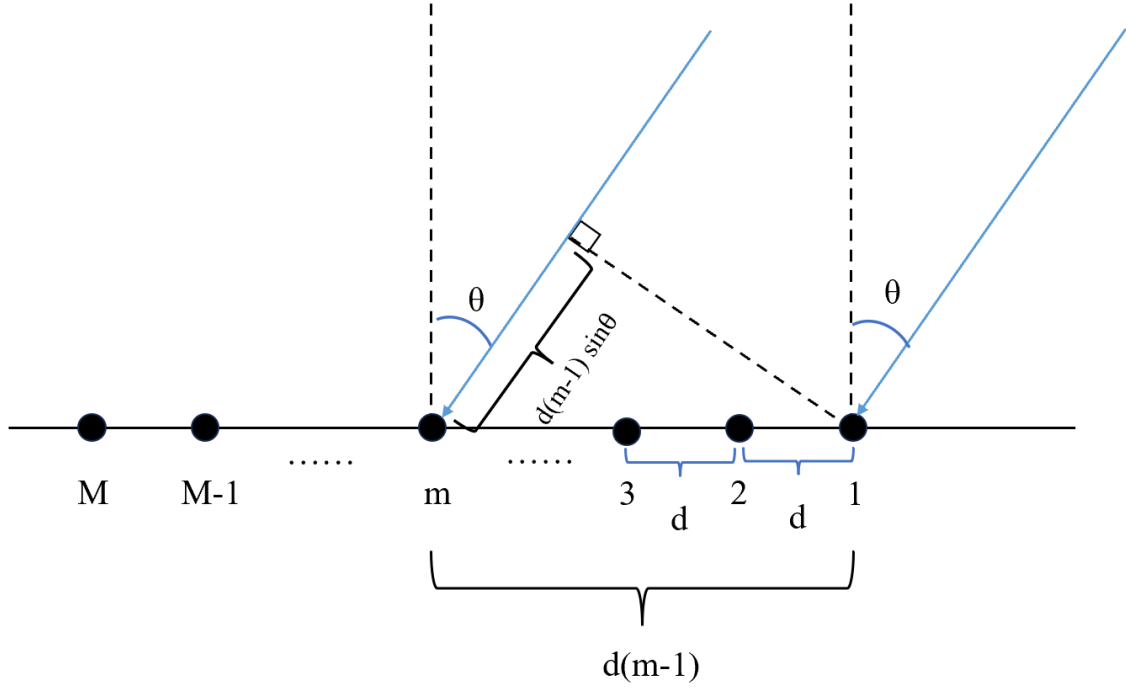


Figure 2.1: The setup of DoA problem of one target in ULA. As the targets is far-field, the DoA for each antenna is the same.

at the first antenna is:

$$y_1(l) = x(l). \quad (2.1)$$

It is noted that there should be an additional reference phase shift caused by the propagation delay from the target to the first antenna. As this delay is negligible, we do not take this phase shift into account. For the m -th antenna, there is a time delay from the first antenna. Now the noise-free received signal is:

$$y_m(l) = e^{-j(m-1)2\pi f \frac{d}{v} \sin \theta} x(l), \quad (2.2)$$

where v is the signal speed in the medium. f is the carrier frequency of the signal. To avoid the spatial aliasing, it is noted that the spacing d is chosen as the half wavelength of the signal $\lambda/2 = v/2f$. Then, if we collect all exponential part in one vector. The steering vector $\mathbf{a}_M(f, \theta)$

is:

$$\mathbf{a}_M(f, \theta) = \left[1, e^{-j2\pi f \frac{d}{v} \sin \theta}, \dots, e^{-j(M-1)2\pi f \frac{d}{v} \sin \theta} \right]^T. \quad (2.3)$$

It is noted that steering vector has the general Vandermonde structure. Due to its structure, Vandermonde structure is important in the area of signal processing. The structure is defined as:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_s & x_s^2 & \cdots & x_s^{n-1} \end{bmatrix}. \quad (2.4)$$

In the chapter 3 and 4, we will use Vandermonde decomposition of Hankel matrix. Vandermonde decomposition is important in frequency and DoA estimation [5]. After Vandermonde decomposition, the Hankel matrix can be decomposed as Vandermonde matrix, diagonal matrix and the conjugate transpose of that Vandermonde matrix.

2.1.2 Signal model in ULA

Considering S far-field targets from directions θ_s , where $s = 1, \dots, S$. The number of antennas in ULA is M . Then the signal model [5] is:

$$\begin{aligned} \mathbf{y}(l) &= \sum_{s=1}^S \mathbf{a}_M(f, \theta_s) x_s(l) + \mathbf{e}(l) \\ &= \mathbf{A}(f) \mathbf{x}(l) + \mathbf{e}(l), \quad l = 1, \dots, L. \end{aligned} \quad (2.5)$$

There are L snapshots. At l -th snapshot, $\mathbf{y}(l) \in \mathbb{C}^M$ is the received signals. $\mathbf{a}(\theta_s) \in \mathbb{C}^M$ is the steering vector of θ_s . $x_s(l)$ is the signal from the s -th target. $\mathbf{e}(l) \in \mathbb{C}^M$ is the measurement noise. The steering matrix $\mathbf{A}(f) = [\mathbf{a}_M(f, \theta_1), \dots, \mathbf{a}_M(f, \theta_S)]$. The signal vector $\mathbf{x}(l) = [x_1(l), \dots, x_S(l)]^T$. If we collect all snapshots in one matrix, we can write (2.5) in

matrix form:

$$\mathbf{Y} = \mathbf{A}(f)\mathbf{X} + \mathbf{E}, \quad (2.6)$$

where the matrix of received signal $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(L)]$. The signal matrix $\mathbf{X} = [\mathbf{x}(1), \dots, \mathbf{x}(L)]^T$. The noise matrix $\mathbf{E} = [\mathbf{e}(1), \dots, \mathbf{e}(L)]^T$.

2.1.3 Identifiability and resolution in ULA

The identifiability is the maximum number of resolvable targets in an array, which has been discussed since 1980s. The ULA case was discussed in [20]. For M -antenna ULA in (2.6), the array can resolve up to:

$$S < \frac{M + \text{rank}(\mathbf{X})}{2}. \quad (2.7)$$

The detailed proof for (2.7) can be referred in [20]. For the single snapshot case ($L = 1$), $\text{rank}(\mathbf{X}) = 1$. (2.7) shows that the M -antenna ULA can measure up to $\frac{M+1}{2} - 1$ targets. When the number of snapshots is greater than M , $\text{rank}(\mathbf{X}) = M$. The M -antenna ULA can measure up to $M - 1$ targets. For the resolution, the Rayleigh angle resolution limit for the M -antenna ULA is $2/M$ radians [21]. The detailed proof can be referred in [22].

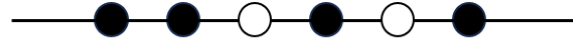
2.1.4 Sparse linear array

Due to lower cost and complexity compared with ULA, sparse linear arrays (SLA) have received much attention, such as nested array [23] and co-prime array [24]. The SLA can be also seen as a part of antennas from ULA. In an SLA, the inter-element spacing can be integer multiple of the fundamental element spacing d . That is, the locations of the antennas are given by the set

$$\mathcal{M}d := \{m_1d, m_2d, \dots, m_Pd\}, \quad (2.8)$$



(a) 6-antenna ULA



(b) Corresponding SLA

Figure 2.2: An example of SLA from 6-antenna ULA. Black points are the physical antennas. White points are holes, which means no antenna in that position.

where $\mathcal{M} = \{m_1, m_2, \dots, m_P\}$, $0 = m_1 < m_2 < \dots < m_P$ are all integers, and P denotes the number of physical antennas in SLA. An example of SLA from 6-antenna ULA is shown in Fig. 2.2.

In an SLA, there may be the ambiguity issue when estimating DoAs. Typically we expect that there is a unique solution when estimating DoAs. The ambiguity issue arises when there are multiple sets of solutions [25]. In this thesis, we use a sparse matrix to represent the received signals before we estimate DoAs. All zeros are for the signals in the non-active antenna location.

2.1.5 MUSIC

MUSIC (Multiple Signal Classification) is a classical algorithm that is used to solve frequency estimation problem, such as DoA problem [6]. Based on our DoA signal model (2.6), the MUSIC algorithm is summarised as:

In the chapter 3 and 4, we will use MUSIC to estimate DoA and range. Although the input matrix is a Hankel matrix rather than covariance matrix, the MUSIC algorithm is still the same from the step 2 in algorithm 1.

Algorithm 1: MUSIC algorithm

Input: Received signal $\mathbf{Y} \in \mathbb{C}^{M \times L}$, the number of targets S , the steering vector $\mathbf{a}(\theta) = [1, e^{-j2\pi f \frac{d}{v} \sin \theta}, \dots, e^{-j(M-1)2\pi f \frac{d}{v} \sin \theta}]$.

1. Compute the covariance matrix $\mathbf{R}$ of $\mathbf{Y}$: $\mathbf{R} = \frac{1}{L} \mathbf{Y} \mathbf{Y}^H$. ;
2. SVD $\mathbf{R} = [\mathbf{U}_s, \mathbf{U}_n] \begin{bmatrix} \mathbf{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_n \end{bmatrix} [\mathbf{V}_s, \mathbf{V}_n]^H$, where $\mathbf{U}_s, \mathbf{V}_s \in \mathbb{C}^{M \times S}$ and $\mathbf{\Sigma}_s \in \mathbb{C}^{S \times S}$;
3. Compute $\phi(\sin \theta) = \|\mathbf{a}(\theta)\|_2 / \|\mathbf{U}_n^H \mathbf{a}(\theta)\|_2$, where $\sin \theta \in [-1, 1]$. ;
4. Collect $\sin \theta$ set by finding the S largest value from $\phi(\sin \theta)$. ;
5. Compute the DoA set from $\sin \theta$ set by using the function: $\arcsin(\cdot)$. ;

Output: the DoA set;

2.2 Structured matrix and decomposition

2.2.1 Hankel matrix

A Hankel matrix is a symmetric matrix of which all the entries in any given skew diagonal are the same. In the chapter 3 and 4, we use Hankel matrix for matrix decomposition and system model reformulation. The general form of Hankel matrix $\mathbf{H} \in \mathbb{C}^{p \times q}$ can be generated from a vector $\mathbf{v} = [v_1, v_2, \dots, v_{p+q-1}] \in \mathbb{C}^{p+q-1}$ by using the Hankel linear operator $\mathcal{H}_{p,q} : \mathbb{C}^{p+q-1} \rightarrow \mathbb{C}^{p \times q}$.

$$\mathbf{v} \mapsto \mathcal{H}_{p,q}(\mathbf{v}) = \begin{bmatrix} v_1 & v_2 & \cdots & v_q \\ v_2 & v_3 & \cdots & v_{q+1} \\ \vdots & \vdots & \ddots & \vdots \\ v_p & v_{p+1} & \cdots & v_{p+q-1} \end{bmatrix}. \quad (2.9)$$

The value of p and q should be known when using $\mathcal{H}_{p,q}$. Note that the above Hankel matrix has a rank- S if $S < \min(M_T, M_R)$. Otherwise the rank is $\min(M_T, M_R)$. The weakness of non-square Hankel matrix is its rank is limited by $\min(M_T, M_R)$, such as $M_T = 1$. To solve the limitation of the rank, typically, we let the Hankel matrix close to square matrix [26]. Now we use Hankel linear operator $\mathcal{H}_{h_1, h_2} : \mathbb{C}^{p+q-1} \rightarrow \mathbb{C}^{h_1 \times h_2}$ to construct the square (or close to

square) Hankel matrix as:

$$\mathbf{v} \mapsto \mathcal{H}_{h_1, h_2}(\mathbf{v}) = \begin{bmatrix} v_1 & v_2 & \cdots & v_{h_2} \\ v_2 & v_3 & \cdots & v_{h_2+1} \\ \vdots & \vdots & \ddots & \vdots \\ v_{h_1} & v_{h_1+1} & \cdots & v_{p+q-1} \end{bmatrix}. \quad (2.10)$$

where $h_1 \approx h_2 \approx (p + q - 1)/2$. In details, $h_1 = h_2 = (p + q)/2$ when $p + q$ is even number. $h_1 = (p + q + 1)/2$ and $h_2 = (p + q - 1)/2$ when $p + q$ is odd number. Specifically, the output of $\mathcal{H}_{p,q}$ and $\mathcal{H}_{h_1, h_2}$ are the same when $p = q$. Square Hankel matrix have been adopted for spectral estimation as they admit Vandermonde decomposition [5].

2.2.2 SVD

SVD (Singular Value Decomposition) is a factorisation of a matrix into rotation, rescaling and followed by another rotation. Given a matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$, its SVD is:

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H. \quad (2.11)$$

The $\mathbf{U} \in \mathbb{C}^{m \times m}$ and $\mathbf{V} \in \mathbb{C}^{n \times n}$ are unitary matrix, whose columns are orthonormal. The $\mathbf{\Sigma} \in \mathbb{C}^{m \times n}$ is a diagonal matrix with non-negative diagonal entries. $\Sigma_{i,i} = \sigma_i$ is i -th the singular value of the matrix $\mathbf{A}$. It is noted that SVD is not unique. If we sort σ_i in descending order in $\mathbf{\Sigma}$. The SVD can be written as:

$$\begin{aligned} \mathbf{A} &= \sum_{i=1}^r \sigma_i \mathbf{u}_{:,i} \mathbf{v}_{:,i}^H \\ &= \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^H, \end{aligned} \quad (2.12)$$

where $r \leq \min\{m, n\}$ is the rank of $\mathbf{A}$. This SVD is called compact SVD. $\mathbf{U}_r \in \mathbb{C}^{m \times r}$ and $\mathbf{V} \in \mathbb{C}^{n \times r}$ are the first r columns of $\mathbf{U}$ and $\mathbf{V}$ respectively. $\mathbf{\Sigma}_r \in \mathbb{C}^{r \times r}$ is a square diagonal matrix.

2.3 Nonconvex and Nonsmooth Optimisation

In this section, we will review the optimisation methods for nonconvex and nonsmooth objective function. In this thesis, we review ADMM, followed by the optimisation methods in the scope of majorization minimization, including proximal gradient method and PDOM.

2.3.1 ADMM

The Alternating Direction Method of Multipliers (ADMM) is the algorithm which usually solves the problems in the form:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}} \quad & f(\mathbf{x}) + g(\mathbf{z}) \\ \text{s.t.} \quad & \mathbf{Ax} + \mathbf{Bz} = \mathbf{c}. \end{aligned} \tag{2.13}$$

The Lagrangian for this problem is:

$$L(\mathbf{x}, \mathbf{z}, \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{y}^T(\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}), \tag{2.14}$$

where $\mathbf{y}$ is the Lagrange multiplier. The augmented Lagrangian is:

$$L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{y}^T(\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}) + \frac{\rho}{2} \|\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}\|_2^2, \tag{2.15}$$

where $\rho > 0$ is the penalty parameter. The role of the augmented term is enhancing stability and allowing fixed learning rate. More details about ADMM could be referred in paper [27]. ADMM can solve convex problem with convergence successfully. Sometimes it is convergent on nonconvex problem in practice. However, there still needs more research on convergence analysis for nonconvex problem.

Paper [28] discusses one of nonconvex problem: nonnegative matrix factorization. The paper proves the convergence of ADMM in three perspectives: the size of consecutive difference of dual variables, the augmented Lagrangian, and combine these two to show global convergence for objective function, dual variable and residual.

We extend the problem to nonconvex and nonsmooth optimisation. the paper [29] shows the ADMM can converge with weaker conditions for this optimisation with convergence in the form:

$$\begin{aligned} & \min_{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}} f(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}) \\ \text{s.t. } & \mathbf{A}_0 \mathbf{x}_0 + \mathbf{A}_1 \mathbf{x}_1 + \dots + \mathbf{A}_p \mathbf{x}_p + \mathbf{B} \mathbf{y} = \mathbf{b}. \end{aligned} \quad (2.16)$$

To solve this problem by using ADMM, the paper [29] define a new variable $\mathbf{x} = [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_p]^T$ and then update variable block $\mathbf{x}_{<i}$ and $\mathbf{x}_{>i}$ by using ADMM. The induction technique is used for convergence analysis.

2.3.2 Majorization Minimization

To reduce the computational cost for nonconvex problems, the majorization minimization (MM) is proposed in 1970s. Now it has been successfully applied to a number of nonconvex and nonsmooth problems, such as sparse regression, matrix factorization and covariance estimation. MM is an algorithm which aims to find minima in an optimisation problem [30]. We want to minimise the problem $f(x)$ with no constraint. To minimise this problem, there are two stages in MM:

Majorization: we construct the surrogate function $\phi(\cdot|x^i)$ at the point $x^i, (i \in \mathbb{N})$:

$$\phi(\cdot|x^i) \geq f(x) + c^i. \quad (2.17)$$

This surrogate function is to approximate the objective function $f(x)$, where $c^i = g(x^i|x^i) - f(x^i)$. To construct the surrogate function, the following methods can be used: Taylor expansion with the first order or second order, convexity inequality, and other special inequality.

Minimization: we minimise the surrogate function ϕ , and find the next point x^{i+1} :

$$x^{i+1} = \arg \min_x \phi(x|x^i). \quad (2.18)$$

After majorization, $\phi(\cdot)$ function is easy to minimise. As a result, the sequence of $f(x^i)$ for all

i is decreasing, which can guarantee the convergence of MM:

$$f(x^{i+1}) \leq f(x^i) \quad (2.19)$$

The advantage of MM is that it can guarantee strict convergence of the objective function. However, MM for optimisation problem with constraints is not well understood. The paper [31] explores the optimisation problem with orthonormality constraint, but its algorithm cannot be applied to the nonconvex and nonsmooth optimisation directly. Another disadvantage is constructing surrogate function is hard in practice, because the local minima rather than global minima may be found in the second stage. MM could also get stuck on stationary points, which is not any minima.

2.3.3 Line Search Method

For unconstrained minimisation problems, the line search method is an iterative algorithm which generates the series of point x_i from the initial point x_0 [32]. To calculate the global minima, we expect the function value $f(x_i)$ at each iteration:

$$f(x_{i+1}) < f(x_i) \quad (2.20)$$

To find the x_{i+1} at each iteration, we choose the search direction p_i and the step length α_i . The formula is given by:

$$x_{i+1} = x_i + \lambda_i p_i \quad (2.21)$$

For choosing step length λ_i , one of popular methods is the Wolfe conditions. The Wolfe conditions are the combination of the following two conditions [33]:

Armijo condition: Ensuring the step length is not too large.

$$f(x_i + \lambda_i p_i) \leq f(x_i) + c_1 \lambda_i \nabla f(x_i)^T p_i \quad (2.22)$$

Curvature condition: Ensuring the step length is not close to 0.

$$\nabla f(x_i + \lambda_i p_i)^T p_i \geq c_2 \nabla f(x_i)^T p_i \quad (2.23)$$

where $0 < c_1 < c_2 < 1$. For choosing search direction, it is usually chosen by the formula:

$$p_i = -Q_i^{-1} \nabla f(x_i) \quad (2.24)$$

where Q_i is symmetric and nonsingular matrix at i -th step. Q_i^{-1} is chosen as Hessian approximation. The more details of constructing Hessian approximation can be referred in [33]. Line search method has been used to find step size and search direction in a range of algorithms. In this thesis, we will use proximal gradient method in the section 2.3.4 and PDOM in the section 2.3.5 as two case studies.

2.3.4 Proximal Gradient Method

In the field of nonconvex optimisation, we consider the problem in the following form:

$$\min_x f(x) + g(x), \quad (2.25)$$

where $f(x)$ is differentiable and $g(x)$ is non-differentiable. The whole objective function $f(x) + g(x)$ is nonconvex and nonsmooth function. It is noted that the splitting is not unique, so the form of $f(x)$ and $g(x)$ can be different. Before we introduce proximal gradient method [34], we define the proximal operator $\text{prox}(\cdot)$ as:

$$\text{prox}_{\lambda, g}(v) := \arg \min_u g(u) + \frac{1}{2\lambda} \|u - v\|^2. \quad (2.26)$$

Then, at the $(i + 1)$ -th step, the proximal gradient method is defined as:

$$x^{i+1} := \text{prox}_{\lambda^i, g}(x^i - \lambda^i \nabla f(x^i)), \quad (2.27)$$

where λ^i is the step size at i -th step. Based on the idea of majorisation-minimisation, the value of λ^i is chosen by the line search. For i -th step, the algorithm of a typical line search is summarised as:

Algorithm 2: Proximal Gradient Method at i -th step

Define: The surrogate function: $\hat{f}_\lambda(u, v) = f(u) + \nabla f(u)^T(v - u) + \frac{1}{2\lambda}\|v - u\|_F^2$
Input: $x^i, \lambda^{i-1}, \beta \in (0, 1)$
while *true* ; /* Update step size λ */
 do
 $z = \text{prox}_{\lambda^{i-1}, g}(x^i - \lambda^{i-1}\nabla f(x^i))$;
 if $f(z) \leq \hat{f}_{\lambda^{i-1}}(z, x^i)$ **then**
 $\lambda^i = \lambda^{i-1}, x^{i+1} = z$;
 Break the loop;
 end
 $\lambda^i = \beta\lambda^{i-1}$;
end
Output: λ^i, x^{i+1} ;

In algorithm 2, it is noted that the typical value of parameter β is chosen as 0.5. For the stopping criteria, we define it as:

$$\|\partial g(x^{i+1}) + \nabla f(x^{i+1})\|_2 \leq \epsilon + \epsilon \max\{\|\nabla f(x^{i+1})\|_2, \frac{1}{\lambda^{i-1}}\|x^{i+1}\|_2, \frac{1}{\lambda^{i-1}}\|x^i\|_2\}, \quad (2.28)$$

$$i \leq i_{max}, \quad (2.29)$$

where $\partial g(x^{i+1})$ is the subgradient of non-differentiable function $g(x)$. The ϵ and i_{max} is a constant with small positive value and the maximum number of iterations.

Subgradient calculation: From (2.27), we can calculate the subgradient of $g(x)$: $\partial g(x^{i+1})$.

The point x^{i+1} is located at local or global optimal point, then:

$$0 \in \partial g(x^{i+1}) + \frac{1}{2\lambda^i} \nabla_{x^{i+1}} (\|x^{i+1} - x^i + \lambda^i \nabla f(x^i)\|^2). \quad (2.30)$$

Then the $\partial g(x^{i+1})$ can be derived as:

$$\partial g(x^{i+1}) \in \frac{-1}{\lambda^i} (x^{i+1} - x^i + \lambda^i \nabla f(x^i)). \quad (2.31)$$

Convergence rate: The convergence rate of proximal gradient method is $\mathcal{O}(1/i)$ at i -th iteration. More details can be referred in [34]. As the proximal gradient method can guarantee the convergence, we will use this algorithm to solve sensing problem in chapter 3 and 4.

2.3.5 PDOM

Based on the proximal gradient method, [35] develops a second-order proximal algorithm, named the Proximal Dogleg Opportunistic Majorisation (PDOM) algorithm. Different from proximal gradient method, the surrogate function is designed along the dogleg path, which include both gradient direction and Newton direction. Due to the Newton direction, PDOM achieve faster convergence compared with proximal gradient method.

The algorithm of PDOM is summarised as:

Algorithm 3: PDOM Algorithm

Define: $\mathbf{p}(\alpha) := \begin{cases} -\tau \mathbf{g}, & \alpha \in (0, 1] \\ -\tau \mathbf{g} + (\alpha - 1)(-\mathbf{Q}^{-1} \mathbf{g} + \tau \mathbf{g}), & \alpha \in (1, 2] \end{cases}$

where $\mathbf{g} := \nabla_x f$;

The surrogate function $\hat{f}_\lambda(u, v) = f(u) + \mathbf{g}_\alpha^T(v - u) + \frac{1}{2\lambda} \|v - u\|_F^2$;

where $\mathbf{g}_\alpha = \mathbf{g}^T \mathbf{p}^2(\alpha) / \|\mathbf{p}(\alpha)\|^2$

Input: $x^0, \mathbf{Q}^{-1}, i = 0$, fixed step size $\tau = 1/\lambda_{max}$, where λ_{max} is the largest eigenvalue of $\mathbf{Q}$. $\lambda = 0.98, \alpha \in [0, 2], \epsilon = 10^{-6}$, maximum number of iteration $i_{max} = 500$.

while *Stopping criteria* (2.33) or (2.34) *is not satisfied* **do**

$i = i + 1$;

$\tau_{\alpha^i} = -\|\mathbf{p}(\alpha)\|^2 / (\mathbf{g}^T \mathbf{p}(\alpha))$;

$x^{i+1} = \text{prox}_{\lambda \tau_{\alpha^i}, \mathbf{g}}(x^i + \mathbf{p}_\lambda(\alpha^i))$;

 ;

 /* Update step size τ_α */

 Find the largest value $\alpha^i \in \{1 + (1/2)^k | k \in \mathbb{N}\}$ such that $f(x^{i+1}) \leq \hat{f}_{\alpha^i}(x^{i+1}, x^i)$. ;

$x_v^{i+1} = \text{prox}_{\tau, \mathbf{g}}(x^i - \tau \nabla_x f(x^i))$;

if $f(x^{i+1}) > f(x_v^{i+1})$ **then**

$x^{i+1} = x_v^{i+1}$;

end

end

Output: x^{i+1}

In algorithm 3, the positive semi-definite Hessian matrix $\mathbf{Q}$ is calculated as:

$$\mathbf{Q} = \frac{\partial^2}{\partial x^2} f \quad (2.32)$$

Similar as proximal gradient method, the stopping criterion of PDOM is:

$$\|\partial g(x^{i+1}) + \nabla f(x^{i+1})\|_2 \leq \epsilon + \epsilon \max\{\|\nabla f(x^{i+1})\|_2, \frac{1}{\lambda\tau_{\alpha^i}}\|x^{i+1}\|_2, \frac{1}{\lambda\tau_{\alpha^i}}\|x^i\|_2\}, \quad (2.33)$$

$$i \leq i_{max} \quad (2.34)$$

where $\partial g(x^{i+1})$ is the subgradient of non-differentiable function $g(x)$. The ϵ and i_{max} is a constant with small positive value and the maximum number of iterations.

Chapter 3

Multi-band Active Array

3.1 Introduction

In this chapter, we use multi-band active array to sense both Direction-of-Arrival (DoA) and spectral response. In the section 2.1, the classical methods: MUSIC [6] has been reviewed. However, the classical methods assume that the transmitted signals are narrowband. However, to recover more information from received signals, narrowband signals cannot achieve it due to their limited information diversity.

To increase the information diversity of received signals, our array transmit the multi-band signals. Then there is a question: compared with narrowband, how can we efficiently use the multi-band to sense targets? In this chapter, we call the this problem as joint processing.

From previous studies, the point of joint processing is decomposing the received multi-band signals into multiple narrowband signals, and then apply narrowband DoA method. The common methods are based on the received signal's covariance matrix decomposition. One of the simplest method is incoherent signal subspace method (ISSM) [36]. It uses eigenvalue decomposition. However, the estimation is not satisfying at some frequency bins. To overcome it, the joint processing is proposed, as it can process each frequency bin simultaneously. For both narrowband and wideband signals, the joint processing is important [37–42].

Related Studies: For joint processing, inspired by ESPRIT, the test of orthogonality of

projected subspaces (TOPS) [43] and the study [44] processes the signal subspaces for each divided blocks of received signals. However, the computational burden suffers. If only one matrix needs to be processed, the popular method is finding the focusing frequency. The coherent signal subspace method (CSSM) [21] finds the focusing frequency by aligning each frequency bin. Inspired by it, the study [45] chooses the lower bound of multi-band signals as the focusing frequency. Without covariance matrix processing, the compressed sensing theory [46] is a new approach for the joint processing. The recent study [47] starts to use greatest common divisor (GCD) to fit all frequencies, but the frequency bins are assumed to be a multiple of the focusing frequency. The joint processing is still challenging.

For the number of resolvable targets in ULA, traditional study [20] show that M -element ULA can estimate up to $M - 1$ targets uniquely. The study [48] discuss another case: $2(M - 1)$ targets can be sensed by M -element ULA, when the received signals is quasi-stationary. Apart from ULA, sparse linear array (SLA) has been investigated recently. The study [23] [24] showed that the SLA can sense more targets than the number of physical antennas. Extended from it, the study [49] investigates the sparse linear array (SLA) with multi-band signals. To process their multi-band, they approximate all of frequency bands to the single band by using Jacobi-Anger approximation. The multi-band SLA can estimate more targets than the number of its virtual antennas. The recent paper [19] and our work [50] propose that multi-band array can resolve more targets than antennas in ULA.

In this chapter, we focus on active multi-band sensing for both uniform linear array (ULA) and sparse linear array (SLA). We use our new joint processing technique: choose GCD of all frequency bands as our reference frequency [51]. The benefits of this method is that the frequency bins are not limited to be equal. There is no restriction on frequency band division. In the future practice, the frequency division signal, such as OFDM, can be applied to cover the whole frequency band simultaneously. Inspired by the compressed sensing theory [46], we formulate this joint sensing problem as low rank matrix recovery problem. Rather than using nuclear norm minimisation (NNM), we use nonconvex problem formulation: rank projection minimisation. To guarantee its convergence, the proximal gradient method [34] is chosen to solve it. Based on the solution of the low rank matrix, the single-snapshot MUSIC algorithm [17]

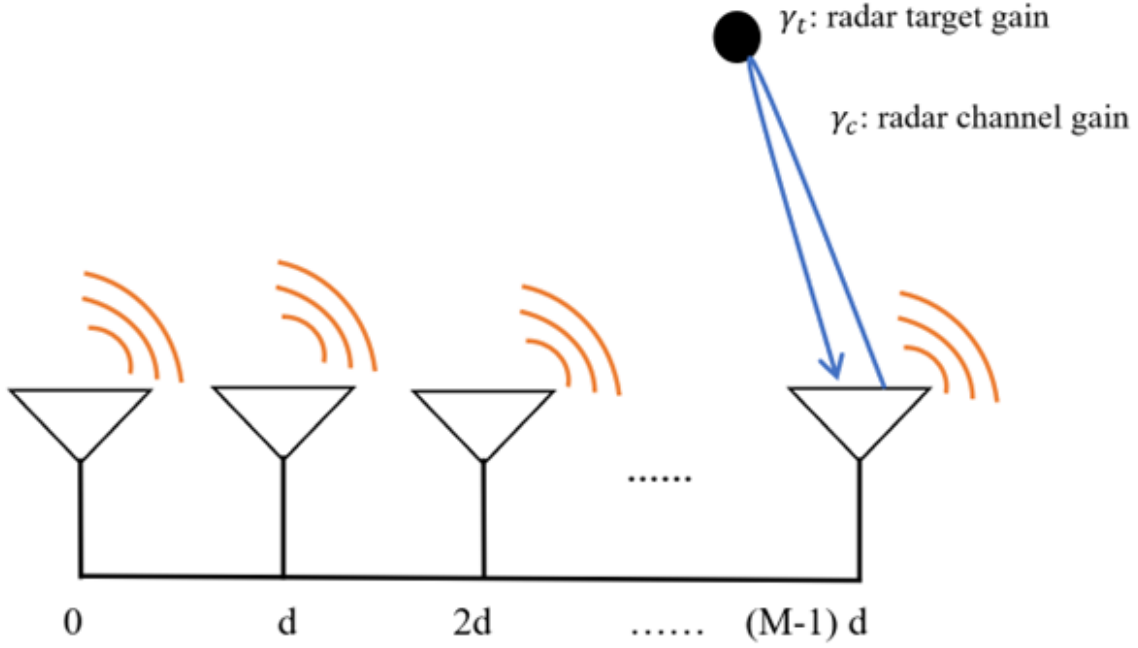


Figure 3.1: Active sensing

is applied to estimate the DoAs.

Apart from the sensed DoAs, we can also sense spectral responses of targets. As active sensing [52] means both transmitted and received signals are given to do sensing, the target spectral response can be sensed by comparing the difference between transmitted signals and received signals (Fig.3.1). Inspired by the geoscience and remote sensing [8] [9], we sketch both DoAs and spectral responses of targets in one plot.

Our Contributions: Our contributions are summarised as the following:

1. New spectral response model: The channel gain is always assumed to be constant in previous studies. However, these methods does not take the spectral response into account. To make our model of channel gain closer to the real environment, the study [53] has added propagation delay to its channel model in the large-scale antenna array. Inspired by this model, we make our channel gain to be dependent with frequency.

2. Joint sensing for DoAs and spectral responses: With our GCD-based joint processing, our multi-band array can estimate both DoAs and the target spectral response. Both of the two information is showed in one plot. This new capability can help in various applications,

such as, using spectral response information for contrast enhancement, object classification. It is noted that the multi-band can also provide the range resolution. This chapter focuses on DoAs and spectral response sensing only. The range sensing can be our future work.

3. Our own criteria of wideband to guarantee the resolution: Traditional criteria of wideband is the ratio of the bandwidth and central frequency. Based on the definition of resolution, we consider the effect of width of the array. When the width of the array increases, the threshold of wideband decreases. This new criteria means that the resolution can be guaranteed by both the width of the array and bandwidth.

4. Sensing more targets than the number of physical antennas: To sense more targets than the number of physical antennas, previous studies focused on improving the diversity of received signals, such as quasi-stationary signals [48]. Inspired by the idea of diversity, we take the advantage of multi-band to improve the signal diversity. In multi-band case, due to our GCD-based joint processing, the size of the low rank Hankel matrix is greater than the number of physical antennas in ULA. This Hankel matrix provides more degree of freedom to let our array to sense more targets than the number of physical antennas in ULA. Extended it to SLA, the more targets can be sensed than the number of virtual antennas.

3.2 System Model and Assumptions

This section presents system models for multi-band angular sensing for both uniform linear arrays (ULA) and sparse linear arrays (SLA).

3.2.1 The Standard Model

We start with the standard model for ULA and narrow-band signals. Consider any form of the MIMO system with M_T transmit antennas and M_R receive antennas. Suppose that the carrier frequency of the narrow-band signal is f . To avoid spatial aliasing at the carrier frequency, the inter-element spacing d is the half wavelength $\lambda/2 = v/2f$ where v is the signal speed in the medium [19]. Denote the transmitted signal symbols across the M_T antenna elements by

$\mathbf{c} \in \mathbb{C}^{M_T}$. Then the received measurement vector $\mathbf{y} \in \mathbb{C}^{M_R}$ can be written as [54]

$$\begin{aligned} \mathbf{y} &= [\mathbf{a}_{M_R}(f, \theta_1), \dots, \mathbf{a}_{M_R}(f, \theta_S)] \text{diag}([\gamma_1, \dots, \gamma_S]) [\mathbf{a}_{M_T}^T(f, \theta_1), \dots, \mathbf{a}_{M_T}^T(f, \theta_S)]^T \mathbf{c} + \mathbf{e}, \\ &= \sum_{s=1}^S \gamma_s \mathbf{a}_{M_R}(f, \theta_s) \mathbf{a}_{M_T}(f, \theta_s)^T \mathbf{c} + \mathbf{e}, \end{aligned} \quad (3.1)$$

where $S \in \mathbb{Z}^+$ is the number of the targets, $\gamma_s \in \mathbb{C}$ denotes the s -th target's coefficient including the effects of target reflection and signal propagation, $\theta_s \in [-90^\circ, 90^\circ)$ represents the angle of the s -th target, the steering vector $\mathbf{a}_M(f, \theta_s)$ is defined as

$$\mathbf{a}_M(f, \theta_s) := \left[1, e^{-j2\pi f \frac{d}{v} \sin \theta_s}, \dots, e^{-j(M-1)2\pi f \frac{d}{v} \sin \theta_s} \right]^T, \quad (3.2)$$

and $\mathbf{e} \in \mathbb{C}^{M_R}$ is additive white Gaussian noise.

3.2.2 Model for Multi-band ULA

Assume that the probing signal used in the system is composed of multiple narrow-band signals. Let $\mathcal{F} = \{f_1, f_2, \dots, f_N\}$ be the set of carrier frequencies of the multi-band signals. Let $\mathbf{y}(f) \in \mathbb{C}^M$ be the received measurement vector at the frequency f . Then one has

$$\mathbf{y}(f) = \sum_{s=1}^S \gamma_s(f) \mathbf{a}_{M_R}(f, \theta_s) \mathbf{a}_{M_T}(f, \theta_s)^T \mathbf{c}(f) + \mathbf{e}(f), \quad (3.3)$$

where $f \in \mathcal{F}$. The notation $\gamma_s(f)$ and $\mathbf{c}(f)$ emphasize that both the target's coefficients and the transmitted symbols vary across different narrow-bands. To avoid spatial aliasing, it is noted that the inter-element spacing $d = \lambda_{max}/2 = v/2f_1$ [19].

3.2.3 Model for Sparse Linear Arrays

Based on the subsection 2.1.4, the signal model for SLAs are similar to that in (3.3) except the steering vector in (3.2) is replaced with

$$\mathbf{a}_{\mathcal{M}}(f, \theta_s) := \left[e^{-j2\pi f \frac{m_1 d}{v} \sin \theta_s}, e^{-j2\pi f \frac{m_2 d}{v} \sin \theta_s}, \dots, e^{-j2\pi f \frac{m_P d}{v} \sin \theta_s} \right]^T,$$

To make the connection with the steering vector in (3.2), we use the idea of virtual antennas. The virtual antennas set $\mathcal{M}_v = \{1, 2, \dots, m_p - 1, m_p\}$, where $\mathcal{M} \subseteq \mathcal{M}_v$. Now the steering vector of SLA can be reformulated as:

$$\mathbf{a}_M(f, \theta_s) := \left[1, e^{-j2\pi f \frac{d}{v} \sin \theta_s}, \dots, e^{-j(m_p-1)2\pi f \frac{d}{v} \sin \theta_s} \right]^T, \quad (3.4)$$

Now the the number of virtual antennas is m_p . For the elements not in $\mathcal{M}$ but in $\mathcal{M}_v$, we treat these elements in sleep mode, which means their corresponding antennas transmit no signal, and cannot receive signals.

3.2.4 Multiple Snapshots

Most works in DoA estimation assume multiple snapshots, such as the studies [20] for ULA and [23] for SLA. Let L be the number of snapshots. We assume that both the locations and the spectral responses of targets remain constant during the snapshots. The system model is then

$$\begin{aligned} \mathbf{y}_l(f) &= \sum_{s=1}^S \gamma_s(f) \mathbf{a}_{M_R}(f, \theta_s) \mathbf{a}_{M_T}(f, \theta_s)^T \mathbf{c}_l(f) + \mathbf{e}_l(f), \\ f &\in \mathcal{F}, \quad l = 1, 2, \dots, L. \end{aligned} \quad (3.5)$$

where the subscript l denotes the l -th snapshot. To include all snapshots in one equation, we formulate it as:

$$\mathbf{Y}(f) = \sum_{s=1}^S \gamma_s(f) \mathbf{a}_{M_R}(f, \theta_s) \mathbf{a}_{M_T}(f, \theta_s)^T \mathbf{C}(f) + \mathbf{E}(f). \quad (3.6)$$

where $\mathbf{Y}(f) = \begin{bmatrix} \mathbf{y}_1(f), \mathbf{y}_2(f), \dots, \mathbf{y}_L(f) \end{bmatrix}$ and $\mathbf{C}(f) = \begin{bmatrix} \mathbf{c}_1(f), \mathbf{c}_2(f), \dots, \mathbf{c}_L(f) \end{bmatrix}$.

3.3 Some Useful Concepts and Assumptions

3.3.1 Spectral Response Model for Targets

In traditional studies, the channel gain γ_s is always assumed to be constants. These methods does not take the spectral response into account. In our method, we aims to make our model closer to the real environment: the channel gain $\gamma_s(f_k)$ is dependent with frequency f_k . To analyse channel gain for s -th target at n -th frequency, we aims to construct the channel gain matrix $\mathbf{\Gamma} \in \mathbb{C}^{S \times N}$ as:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1(f_1) & \gamma_1(f_2) & \cdots & \gamma_1(f_N) \\ \gamma_2(f_1) & \gamma_2(f_2) & \cdots & \gamma_2(f_N) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_S(f_1) & \gamma_S(f_2) & \cdots & \gamma_S(f_N) \end{bmatrix}. \quad (3.7)$$

In real environment, spectral response $\gamma_s(f_n)$ is not all various in the frequency set $\mathcal{F}$, so we denote K as the number of variety of $\gamma_s(f_n)$ in the frequency set. To show this property clearly, we denote $\mathbf{G} \in \mathbb{C}^{S \times K}$ as spectral response in spatial domain. $\mathbf{G}$ includes all S targets' spectral response information. We convert channel gain matrix in spatial domain $\mathbf{G} \in \mathbb{C}^{S \times K}$ to its frequency domain $\mathbf{\Gamma}$. We construct our own channel gain matrix by multiplying the N - point DFT matrix $\mathbf{F}$.

To reduce the number of unknown variables, we choose to formulate the matrix $\mathbf{\Gamma}$ as a low rank matrix. To explicitly formulate $\text{rank}(\mathbf{\Gamma}) = K$, the DFT matrix $\mathbf{F}_K \in \mathbb{C}^{K \times N}$ is the first K rows of N -point DFT matrix, which is:

$$\mathbf{\Gamma} = \mathbf{G}\mathbf{F}_K. \quad (3.8)$$

To guarantee the low rank structure, the K should satisfy $1 < K \ll \min(S, N)$. In traditional study, the channel gain γ_s is always assumed to be constants, which means $\text{rank}(\mathbf{\Gamma}) = 1$. As a result, there are $K \times S$ unknown variables in the channel gain matrix $\mathbf{\Gamma}$. We can estimate $\mathbf{\Gamma}$ by only estimating $\mathbf{G}$.

3.3.2 Hankel Matrix

We have reviewed the concept of Hankel matrix in the section 2.2.1. Square Hankel matrix have been used for spectral estimation as they admit Vandermonde decomposition. Furthermore, low-rank Hankel matrix recovery formulation is used for frequency estimation [17]. A Hankel matrix is a matrix of which all the entries in any given skew diagonal are the same. The general form of Hankel matrix $\mathbf{H} \in \mathbb{C}^{p \times q}$ can be generated from a vector $\mathbf{v} \in \mathbb{C}^{p+q-1}$ by using the first Hankel linear operator $\mathcal{H}_{p,q} : \mathbb{C}^{p+q-1} \rightarrow \mathbb{C}^{p \times q}$

$$\mathbf{v} \mapsto \mathcal{H}_{p,q}(\mathbf{v}) = \begin{bmatrix} v_1 & v_2 & \cdots & v_q \\ v_2 & v_3 & \cdots & v_{q+1} \\ \vdots & \vdots & \ddots & \vdots \\ v_p & v_{p+1} & \cdots & v_{p+q-1} \end{bmatrix}. \quad (3.9)$$

The value of p and q should be known when using $\mathcal{H}_{p,q}$. The Hankel matrix admits the Vandermonde decomposition [26]. It is straightforward to verify that:

$$\mathcal{H}_{M_R, M_T} \left(\sum_{s=1}^S \gamma_s \mathbf{a}_{M_T+M_R-1}(f, \theta_s) \right) = \sum_{s=1}^S \gamma_s \mathbf{a}_{M_R}(f, \theta_s) \mathbf{a}_{M_T}(f, \theta_s)^T. \quad (3.10)$$

When $M_R \approx M_T$, it is noted that the rank of $\mathcal{H}_{M_R, M_T} \left(\sum_{s=1}^S \gamma_s \mathbf{a}_{M_T+M_R-1}(f, \theta_s) \right)$ is no more than S . We can estimate up to the S DoAs when we apply Vandermonde decomposition and MUSIC algorithm to the Hankel matrix in (3.10). Based on this benefit, Hankel matrix naturally rises in our system model (3.2). We reformulate our noiseless system model (3.6) as:

$$\begin{aligned} \mathbf{Y}(f_n) &= \sum_{s=1}^S \gamma_s(f_n) \mathbf{a}_{M_R}(f_n, \theta_s) \mathbf{a}_{M_T}(f_n, \theta_s)^T \mathbf{C}(f_n) \\ &= \mathcal{H}_{M_R, M_T} \left(\sum_{s=1}^S \gamma_s(f_n) \mathbf{a}_{M_R+M_T-1}(f_n, \theta_s) \right) \mathbf{C}(f_n). \end{aligned} \quad (3.11)$$

More generally (M_R may not be equal to M_T), it is noted that the above Hankel matrix has a rank- S if $S < \min(M_T, M_R)$. Otherwise the rank is $\min(M_T, M_R)$. The weakness of non-square Hankel matrix is its rank is limited by $\min(M_T, M_R)$, such as $M_T = 1$.

To solve the limitation of the rank, typically, we let the Hankel matrix close to square matrix [26]. Square Hankel matrix have been adopted for frequency estimation as they admit Vandermonde decomposition [5]. Now we use the second Hankel linear operator $\mathcal{H}_{h_1, h_2} : \mathbb{C}^{p+q-1} \rightarrow \mathbb{C}^{h_1 \times h_2}$ to construct the square (or close to square) Hankel matrix as:

$$\mathbf{v} \mapsto \mathcal{H}_{h_1, h_2}(\mathbf{v}) = \begin{bmatrix} v_1 & v_2 & \cdots & v_{h_2} \\ v_2 & v_3 & \cdots & v_{h_2+1} \\ \vdots & \vdots & \ddots & \vdots \\ v_{h_1} & v_{h_1+1} & \cdots & v_{p+q-1} \end{bmatrix}. \quad (3.12)$$

where $h_1 \approx h_2 \approx (p+q-1)/2$. In details, $h_1 = h_2 = (p+q)/2$ when $p+q$ is even number. $h_1 = (p+q+1)/2$ and $h_2 = (p+q-1)/2$ when $p+q$ is odd number. Specifically, the output of $\mathcal{H}_{p,q}$ and $\mathcal{H}_{h_1, h_2}$ are the same when $p=q$. In (3.10),

$$\text{rank}(\mathcal{H}_{h_1, h_2}(\sum_{s=1}^S \gamma_s \mathbf{a}_{M_T+M_R-1})) = S, \text{ if } S < \frac{M_T + M_R}{2}. \quad (3.13)$$

In the next section, $\mathcal{H}_{p,q}(\mathbf{v})$ will be used for system model reformulation. $\mathcal{H}_{h_1, h_2}(\mathbf{v})$ will be used for matrix decomposition.

3.3.3 Greatest Common Divisor (GCD) of Rational Numbers

In practice, although the value of frequency can be any real positive number, we can approximate any irrational number as its closest rational number. It is possible to map any real positive number to its closest rational number [51]. Based on this fact, we can assume that the values of all frequencies in $\mathcal{F}$ are rational numbers. To process all frequencies in $\mathcal{F}$ simultaneously, we use their GCD to represent them. We define the GCD of rational numbers as the following:

GCD definition: The greatest common divisor (GCD) of a set of positive rational numbers $f_1, f_2, \dots, f_N \in \mathbb{R}^+$ is the largest positive rational number $f_g \in \mathbb{R}^+$ such that $f_n/f_g \in \mathbb{Z}^+$ for all $n \in [N]$.

To calculate f_g , we firstly do approximation for all frequencies to its closest rational number.

The proof is based on elementary algebra, which can be referred in [55]. More details about calculating GCD for a set of rational numbers is the following [51] :

GCD calculation: Let $f_1, \dots, f_N \in \mathbb{Q}^+$ be positive rational numbers. Write $f_n = p_n/q_n$, $n \in [N]$, where p_n and q_n are co-prime. Denote the least common multiplier (lcm) of integers $q_1, \dots, q_N$ as q_π :

$$q_\pi := \text{lcm}(q_1, \dots, q_N). \quad (3.14)$$

Then the GCD of $f_1, \dots, f_N$ is given by

$$\begin{aligned} f_g &= \text{gcd}(f_1, \dots, f_N) \\ &= \text{gcd}(f_1 q_\pi, \dots, f_N q_\pi), \end{aligned} \quad (3.15)$$

where the function $\text{gcd}(\cdot)$ in the second line computes the GCD of integers. The GCD of multiple integers can be calculated by using the Euclidean algorithm pair by pair.

3.3.4 Assumptions

Before we discuss our proposed method, some assumptions in this chapter are summarised as follows:

A1) The noise $\mathbf{e}_l(f)$ in (3.5) is in Gaussian distribution $\mathcal{N}(0, \sigma^2)$, where σ^2 is the noise power in f -th frequency and l -th snapshot.

A2) For each frequency f_n , the channel response matrix $\sum_{s=1}^S \gamma_s(f) \mathbf{a}_{MR}(f, \theta_s) \mathbf{a}_{MT}(f, \theta_s)^T$ in (3.6) are independent with one another.

A3) The S targets are distinct to one another. The value of S is known before our sensing progress. Although this thesis does not aim to discuss the method that knowing the number of targets a prior, there are some practical methods in [21] [36].

A4) The value of K in (3.8) is known before our sensing progress.

3.4 DoAs and Spectral Responses Sensing

3.4.1 Linear Operator Notations

Apart from the linear operator $\mathcal{H}(\cdot)$ in (3.9), we list the following linear operator notations that will be used in this section: The linear operator $\mathcal{A}_n(\cdot)$ denotes as the sub-vector of $\mathbf{a}_{M_g}$ obtained by keeping the elements indexed by the set $\mathcal{I}_{n,M}$. The linear operator $\mathcal{B}_n(\cdot)$ is taking the n -th column of the input matrix. The linear operator $\mathcal{W}_n(\cdot)$ puts input matrix at $\mathbf{H} \in \mathbb{C}^{h_1 \times Nh_2}$ from column index $(n-1)h_2$ to nh_2 . All other elements are zero. We will compound the linear operators as: The linear operator $\mathcal{L}_n^1(\cdot) = \mathcal{W}_n \mathcal{H}_{h_1, h_2} \mathcal{B}_n(\cdot)$. The linear operator $\mathcal{L}_n^2(\cdot) = \mathcal{H}_{M_R, M_T} \mathcal{A}_n \mathcal{B}_n(\cdot)$.

3.4.2 One Vector Fits All

It has been widely recognized that estimating angles using data across different frequencies is a challenging task. This is mainly due to the highly nonlinear mapping between steering vectors at different frequencies, caused by the fact that the angle information $\sin \theta$ is present in the exponents of the vector entries. To overcome this difficulty, we propose constructing a single steering vector that includes the steering vectors of all frequencies in $\mathcal{F}$. This representation enables our later approach for joint estimation using data from multiple frequencies.

In particular, assume that all frequencies $f \in \mathcal{F}$ are positive rational numbers.¹ Define $f_g = \gcd(f_1, \dots, f_N)$ (Definition 1), $f_{\min} = \min(f_1, \dots, f_N)$, and $f_{\max} = \max(f_1, \dots, f_N)$. To shorten the notation, we use $M_l = M_T + M_R - 1$. We fit all frequencies by using the *template steering vector*:

$$\mathbf{a}_{M_g}(f_g, \theta_s) := \left[1, e^{-j2\pi f_g \frac{d}{v} \sin \theta_s}, \dots, e^{-j(M_g-1)2\pi f_g \frac{d}{v} \sin \theta_s} \right]^T, \quad (3.16)$$

where

$$M_g = (M_l - 1) \frac{f_{\max}}{f_g} + 1 \in \mathbb{Z}^+. \quad (3.17)$$

¹This is a rational assumption in practice.

It is clear by definition of GCD that $\mathbf{a}_{M_l}(f, \theta_s)$, $\forall f \in \mathcal{F}$, is a sub-vector of $\mathbf{a}_{M_g}(f_g, \theta_s)$. We denote it as the linear operator $\mathcal{A}_n$. It can be expressed as:

$$\begin{aligned}\mathbf{a}_{M_l}(f_n, \theta_s) &= \mathbf{a}_{M_g, \mathcal{I}_{n, M_l}}(f_g, \theta_s) \\ &= \mathcal{A}_n(\mathbf{a}_{M_g}(f_g, \theta_s)),\end{aligned}\tag{3.18}$$

where the index set $\mathcal{I}_{n, M_l}$ is

$$\mathcal{I}_{n, M_l} = \frac{f_n}{f_g} \{0, 1, \dots, M_l - 1\} + 1.\tag{3.19}$$

3.4.3 System Model Reformulation

In this subsection, we aim to show that the system model (3.6) can be reformulated by using two low-rank structured matrix: $\mathcal{H}(\cdot)$ with rank- S in (3.13) and $\mathbf{\Gamma}$ with rank- k in (3.8). This subsection is prepared for our low rank matrix recovery problem in the next subsection.

Before we express our new system model in matrix form, we use the template steering vector (3.16) with f_g to reformulate our noiseless system model (3.11).

$$\begin{aligned}\mathbf{Y}(f_n) &= \mathcal{H}_{M_R, M_T} \left(\sum_{s=1}^S \gamma_s(f_n) \mathbf{a}_{M_l}(f_n, \theta_s) \right) \mathbf{C}(f_n) \\ &= \mathcal{H}_{M_R, M_T} \left(\sum_{s=1}^S \gamma_s(f_n) \mathbf{a}_{M_g, \mathcal{I}_{n, M_l}}(f_g, \theta_s) \right) \mathbf{C}(f_n) \\ &= \mathcal{H}_{M_R, M_T} \mathcal{A}_n(\mathbf{x}(f_n)) \mathbf{C}(f_n), n = 1, \dots, N.\end{aligned}\tag{3.20}$$

In (3.20), $\mathbf{x}(f_n) := \sum_{s=1}^S \gamma_s(f_n) \mathbf{a}_{M_g}(f_g, \theta_s)$. To collect all $\mathbf{x}(f_n)$ in one matrix, we denote $\mathbf{X}$ include all frequencies in $\mathcal{F}$. It also means $\mathbf{x}(f_n) = \mathbf{X}_{:,n} = \mathcal{B}_n(\mathbf{X})$. The $\mathbf{X}$ defined as:

$$\begin{aligned}\mathbf{X} &:= [x(f_1), x(f_2), \dots, x(f_N)] \\ &= \mathbf{A}_{M_g} \mathbf{\Gamma},\end{aligned}\tag{3.21}$$

where $\mathbf{A}_{M_g} := [\mathbf{a}_{M_g}(f, \theta_1), \dots, \mathbf{a}_{M_g}(f, \theta_S)] \in \mathbb{C}^{M_g \times S}$. From (3.8), we introduce the new variable

$\mathbf{X}_G$ in (3.21) as the following:

$$\begin{aligned}\mathbf{X} &= \mathbf{A}_{M_g} \mathbf{G} \mathbf{F}_k \\ &= \mathbf{X}_G \mathbf{F}_k\end{aligned}\tag{3.22}$$

where $\mathbf{X}_G := \mathbf{A}_{M_g} \mathbf{G} \in \mathbb{C}^{M_g \times k}$. Given $\mathbf{F}_k$, both the value of DoA and spectral response are included in $\mathbf{X}_G$ now. The second dimension of $\mathbf{X}_G$ is k , which means the rank of $\mathbf{\Gamma}$ is also included in $\mathbf{X}_G$. Finally, we can express (3.20) by using $\mathbf{X}_G$ as:

$$\begin{aligned}\mathbf{Y}(f_n) &= \mathcal{H}_{M_R, M_T} \mathcal{A}_n \mathcal{B}_n(\mathbf{X}) \mathbf{C}(f_n) \\ &= \mathcal{H}_{M_R, M_T} \mathcal{A}_n \mathcal{B}_n(\mathbf{A} \mathbf{\Gamma}) \mathbf{C}(f_n) \\ &= \mathcal{L}_n^2(\mathbf{X}_G \mathbf{F}_k) \mathbf{C}(f_n).\end{aligned}\tag{3.23}$$

3.4.4 Low Rank Square Hankel Matrix Formulation

As shown in the section 3.3.2, we typically use square (or close to square) Hankel matrix to do Vandermonde decomposition. As $\mathcal{L}_n^2(\mathbf{X}_G \mathbf{F}_k)$ is not square matrix when $M_T \neq M_R$, we use the square Hankel linear operator $\mathcal{H}_{h_1, h_2}$ in (3.12) to formulate the square Hankel matrix. It is

$$\begin{aligned}\mathbf{H}_n &:= \mathcal{H}_{h_1, h_2} \left(\sum_{s=1}^S \gamma_s(f_n) \mathbf{a}_{M_g}(f_g, \theta_s) \right) \\ &= \mathcal{H}_{h_1, h_2}(\mathbf{x}(f_n)) \in \mathbb{C}^{h_1 \times h_2},\end{aligned}\tag{3.24}$$

where $h_1 = h_2 = (M_g + 1)/2$ if M_g is odd number. $h_1 = (M_g + 2)/2$ and $h_2 = M_g/2$ if M_g is even number. The $\mathbf{H}_n$ is rank- S Hankel matrix. If the value of M_g is sufficiently large, it will lead to $h_1 \approx h_2 > S$. This means target number S can be larger than the received antenna number M_R . To express $\mathbf{H}_n$ with our variable $\mathbf{X}_G$, from (3.22), $\mathbf{H}_n$ is written as:

$$\mathbf{H}_n = \mathcal{H}_{h_1, h_2} \mathcal{B}_n(\mathbf{X}_G \mathbf{F}_k), \forall n.\tag{3.25}$$

In (3.24), $\gamma_s(f_n)$ is the only coefficient makes differences for each Hankel matrix $\mathbf{H}_n$. It means that each $\mathbf{H}_n$ share the same signal subspace when do singular value decomposition (SVD). Based on this fact, we put all of $\mathbf{H}_n$ in one row. We define it as $\mathbf{H} \in \mathbb{C}^{h_1 \times h_2 N}$:

$$\begin{aligned}
 \mathbf{H} &:= [\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_N] \\
 &= \sum_{n=1}^N \mathcal{W}_n(\mathbf{H}_n) \\
 &= \sum_{n=1}^N \mathcal{W}_n \mathcal{H}_{h_1, h_2} \mathcal{B}_n(\mathbf{X}_G \mathbf{F}_k) \\
 &= \sum_{n=1}^N \mathcal{L}_n^1(\mathbf{X}_G \mathbf{F}_k).
 \end{aligned} \tag{3.26}$$

Now the collection of Hankel matrix $\mathbf{H}$ is still rank- S . From the section 2.2.1 and 3.3.2, we can still apply MUSIC algorithm to the Hankel matrix $\mathbf{H}_n$ to estimate DoAs. To efficiently use $\mathbf{H}_n, \forall n$, we can apply MUSIC algorithm to $\mathbf{H}$ to estimate DoAs.

3.4.5 Optimisation Formulation and Solving

In this subsection, the optimisation formulation with variable $\mathbf{X}_G$ will be derived. To recover the low rank matrix, we propose the rank projection minimisation (RPM). Due to the weakness of nuclear norm minimisation [17], we choose the indicator function $\delta(\cdot)$ as:

$$\delta(\text{rank}(\mathbf{H}) \leq S) := \begin{cases} 0, & \text{if } \text{rank}(\mathbf{H}) \leq S \\ +\infty, & \text{otherwise} \end{cases}$$

From (3.23) and (3.26), Given $\mathbf{Y}(f_n), \mathbf{C}(f_n), \mathbf{F}_k, \forall n$, we formulate the low rank matrix recovery problem as:

$$\begin{aligned}
 &\min_{\mathbf{X}_G, \mathbf{H}} \delta(\text{rank}(\mathbf{H}) \leq S) \\
 &\text{s.t. } \mathbf{H} = \sum_{n=1}^N \mathcal{L}_n^1(\mathbf{X}_G \mathbf{F}_k) \\
 &\mathbf{Y}(f_n) = \mathcal{L}_n^2(\mathbf{X}_G \mathbf{F}_k) \mathbf{C}(f_n), n = 1, \dots, N
 \end{aligned} \tag{3.27}$$

where the first constraint is to enforce the Hankel matrix collection structure. The second constraint is to ensure the data consistency. To solve (3.27) as unconstrained problem, the following relaxation with positive parameter $\alpha > 0$ is applied:

$$\begin{aligned} \min_{\mathbf{X}_G, \mathbf{H}} \quad & \delta(\text{rank}(\mathbf{H}) \leq S) + \frac{1}{2\alpha} \|\mathbf{H} - \sum_{n=1}^N \mathcal{L}_n^1(\mathbf{X}_G \mathbf{F}_k)\|_F^2 \\ & + \sum_{n=1}^N \frac{1}{2} \|\mathbf{Y}(f_n) - \mathcal{L}_n^2(\mathbf{X}_G \mathbf{F}_k) \mathbf{C}(f_n)\|_F^2. \end{aligned} \quad (3.28)$$

The value of α should be small to enforce the Hankel matrix collection structure. As (3.28) is nonconvex and nondifferentiable optimisation with variable pair $(\mathbf{X}_G, \mathbf{H})$. We choose standard proximal gradient method [34] to solve it. The standard proximal gradient method is to solve an optimization problem of the form: $\min f(x) + g(x)$, where $f(x)$ is differentiable function and $g(x)$ is nondifferentiable function. In order to apply the proximal gradient method to solve (3.28), our g and f are:

$$g(\mathbf{H}) = \delta(\text{rank}(\mathbf{H}) \leq S), \quad (3.29)$$

$$\begin{aligned} f(\mathbf{X}_G, \mathbf{H}) = & \frac{1}{2\alpha} \|\mathbf{H} - \sum_{n=1}^N \mathcal{L}_n^1(\mathbf{X}_G \mathbf{F}_k)\|_F^2 \\ & + \sum_{n=1}^N \frac{1}{2} \|\mathbf{Y}(f_n) - \mathcal{L}_n^2(\mathbf{X}_G \mathbf{F}_k) \mathbf{C}(f_n)\|_F^2. \end{aligned} \quad (3.30)$$

3.4.6 Target Sensing

Based on the solution of $\hat{\mathbf{X}}_G$ and $\hat{\mathbf{H}}$ in the optimisation problem (3.28), we can estimate the DoAs $\boldsymbol{\theta}$ and spectral response $\boldsymbol{\Gamma}$ as the following:

(1) DoAs $\boldsymbol{\theta}$:

From the estimated variable $\hat{\mathbf{H}}$, we use multiple signal classification (MUSIC) to find S largest local maxima in the imaging function. We denote the estimated DoA set as $\hat{\boldsymbol{\theta}}$. Although the input matrix for MUSIC algorithm is the collection of Hankel matrix rather than a covariance matrix, the MUSIC algorithm is still the same from the step 2 in algorithm 1.

(2) Spectral response $\mathbf{\Gamma}$:

From the estimated angle set $\hat{\boldsymbol{\theta}}$, the $\mathbf{A}_{Mg}$ can be recovered in (3.21). Then we calculate the $\hat{\mathbf{\Gamma}}$:

$$\begin{aligned}\hat{\mathbf{\Gamma}} &= (\mathbf{A}_{Mg}^H \mathbf{A}_{Mg})^{-1} \mathbf{A}_{Mg}^H \mathbf{X} \\ &= (\mathbf{A}_{Mg}^H \mathbf{A}_{Mg})^{-1} \mathbf{A}_{Mg}^H \hat{\mathbf{X}}_G \mathbf{F}_k.\end{aligned}\tag{3.31}$$

From both DoAs $\boldsymbol{\theta}$ and spectral response $\mathbf{\Gamma}$, we can sketch the HSI of targets. We use the standard in the reflective and transmissive cases of the CIE 1931 color space. More details can be found in appendix B.

3.5 Discussion

3.5.1 What Defines a Wideband for an Array

Traditionally, both Federal Communications Commission and [56] define the criteria of wideband depends on the ratio between the bandwidth Δf and the centre frequency f_c , where $\Delta f := f_{\max} - f_{\min}$ be the bandwidth and $f_c := (f_{\max} + f_{\min})/2$ be the centre frequency.

The paper [53] introduces the spatial-wideband effect in massive MIMO. The spatial-wideband effect is caused by the non-negligible propagation delay across a large-scale array. Inspired by this effect, we find that only small bandwidth can cause non-negligible phase delay in large size of array, so the wideband threshold can be lower than the traditional threshold. As a result, we need to consider the width of array when we define wideband. It is shown in definition 2:

Definition 2: we call a linear array with width $W := d(M - 1)$ and signal bandwidth Δf a *wideband array* if

$$\Delta f \geq \frac{v}{W},\tag{3.32}$$

This definition is based on the observation that when (3.32) holds, the variation in direction estimation caused by the different frequencies will exceed the resolution limit of the array operating with a narrow-band signal.

Justification: To motivate the definition, we start from a narrow-band M -element ULA operating at the frequency f . It is typically assumed the spacing $d = \lambda/2 = v/2f$ and hence $fd/v = 1/2$. As the angle resolution limit is $2/M$ (radians) in M -element ULA [57], we consider two different directions θ_1 and θ_2 , such that:

$$\Delta\theta = \theta_1 - \theta_2 = 2/M(\text{radians}) \quad (3.33)$$

As the value of $\Delta\theta$ is relatively small, then $\Delta_{\sin} = \sin \theta_1 - \sin \theta_2 \approx 2/M$. Under this condition, we prove this threshold as:

$$\begin{aligned} \mathbf{a}_M(f, \theta_1)^H \mathbf{a}_M(f, \theta_2) &= \sum_{m=1}^M e^{-j\pi(m-1)\Delta_{\sin}} \\ &= \sum_{m=1}^M e^{-j2\pi(m-1)/M} = 0. \end{aligned}$$

According to the phase term $2\pi(m-1)/M$ above, we can quantify the threshold of Δf for a wideband array to have a confusion now. The relationship is:

$$\mathbf{a}_M(f, \theta_2) = \mathbf{a}_M(f + \Delta f, \theta_1). \quad (3.34)$$

For the right hand side of (3.34), denote that the m -th element in the steering vector $\mathbf{a}_M(f + \Delta f, \theta_1)$ is:

$$\begin{aligned} \mathbf{a}_{M,m}(f + \Delta f, \theta_1) &= e^{-j(m-1)2\pi(f+\Delta f)\frac{d}{v}\sin\theta_1} \\ &= e^{-j(m-1)2\pi f\frac{d}{v}\sin\theta_1 - j(m-1)2\pi\Delta f\frac{d}{v}\sin\theta_1} \\ &= \mathbf{a}_{M,m}(f, \theta_1) e^{-j(m-1)2\pi\Delta f\frac{d}{v}\sin\theta_1} \end{aligned}$$

Similarly, the left hand side of (3.34) can be expressed as:

$$\begin{aligned}\mathbf{a}_M(f, \theta_2) &= \mathbf{a}_M(f, \theta_1 + \frac{2}{M}) \\ &= \mathbf{a}_{M,m}(f, \theta_1) e^{-j(m-1)2\pi f \frac{d}{v} \sin(\frac{2}{M})}\end{aligned}$$

This leads to the following, especially when the number of M is large:

$$\begin{aligned}|\Delta f| &= \frac{|f \sin \frac{2}{M}|}{|\sin \theta_1|} \\ &\geq |f \sin \frac{2}{M}| \\ &= \frac{v}{Md} \approx \frac{v}{W}.\end{aligned}$$

3.5.2 Why can we sense more targets than the number of physical antennas in ULA?

Traditionally, the M -element ULA can sense up to $M - 1$ targets in single frequency band. In previous work, the sparse linear array [24] [23] or quasi-stationary signal [48] can let more targets can be sensed by the number of physical antennas. In the multiple frequency bands, it is still possible to sense more targets than the number of physical antennas in ULA.

We start from single-input multiple-output (SIMO) array. For the single-band in (3.1), the size of virtual array is M [11]. We firstly assume the spectral response is independent with frequencies, which means $k = 1$ in (3.8). For SIMO case, we set the only first antenna is both transmitter and receiver. Now we reformulate (3.1) as:

$$\begin{aligned}\mathbf{Y}_M &= \sum_{s=1}^S \gamma_s \mathbf{a}_M(f, \theta_s) c \\ &= \mathbf{A}(f) \mathbf{S}\end{aligned}\tag{3.35}$$

where $\mathbf{A}(f) := [\mathbf{a}_M(f, \theta_1), \dots, \mathbf{a}_M(f, \theta_S)] \in \mathbb{C}^{M \times S}$ and $\mathbf{S} = ([\gamma_1, \dots, \gamma_S]^T c)$. As matrix $\mathbf{A}$ is a Vandermonde matrix with M independent rows, there are M independent equations in (3.35).

There are S DoAs and S spectral response to be sensed, the theoretical maximum resolvable targets S_{max} is:

$$S_{max} \leq \frac{M}{2}.$$

Now we extend this analysis to multi-band SIMO array. We also extend the spectral response is dependent with frequencies. From (3.3), we reformulate it for all frequencies as:

$$\begin{bmatrix} \mathbf{Y}_M(f_1) \\ \mathbf{Y}_M(f_2) \\ \dots \\ \mathbf{Y}_M(f_N) \end{bmatrix} = \begin{bmatrix} \mathbf{A}(f_1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}(f_2) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}(f_N) \end{bmatrix} \begin{bmatrix} \mathbf{S}(f_1) \\ \mathbf{S}(f_2) \\ \vdots \\ \mathbf{S}(f_N) \end{bmatrix} \quad (3.36)$$

where $\mathbf{S}(f_n) = ([\gamma_1(f_n), \dots, \gamma_S(f_n)]^T c(f_n))$. In (3.36), there are MN independent equations. Now there are still S DoAs but kS spectral response in (3.8) to be sensed, the theoretical maximum resolvable targets S_{max} is:

$$S_{max} \leq \frac{MN}{k+1}. \quad (3.37)$$

However, this upper bound (3.37) is theoretical. In practice, for the any frequency set $\mathcal{F}$, as not all M_g rows and columns can be recovered in the response array $\mathbf{H}_n$ in (3.20). For example, the frequency set $\mathcal{F} = \{1, 3\}$ GHz and $M = 2$. According to (3.16), the Hankel matrix is:

$$\mathbf{H} = \sum_{s=1}^S \begin{bmatrix} 1 & \omega_s & * & \omega_s^3 \\ \omega_s & * & \omega_s^3 & * \\ * & \omega_s^3 & * & * \\ \omega_s^3 & * & * & * \end{bmatrix}, \quad (3.38)$$

where $\omega_s = e^{-j2\pi f \frac{d}{v} \sin \theta_s}$ and the symbol $*$ denotes the missing elements. Now this is the low-rank matrix completion problem. The matrix $\mathbf{H}$ can be uniquely recovered if $S = 1$. However, it cannot be uniquely recovered if $S = 4$, which leads to the ambiguity. For more general

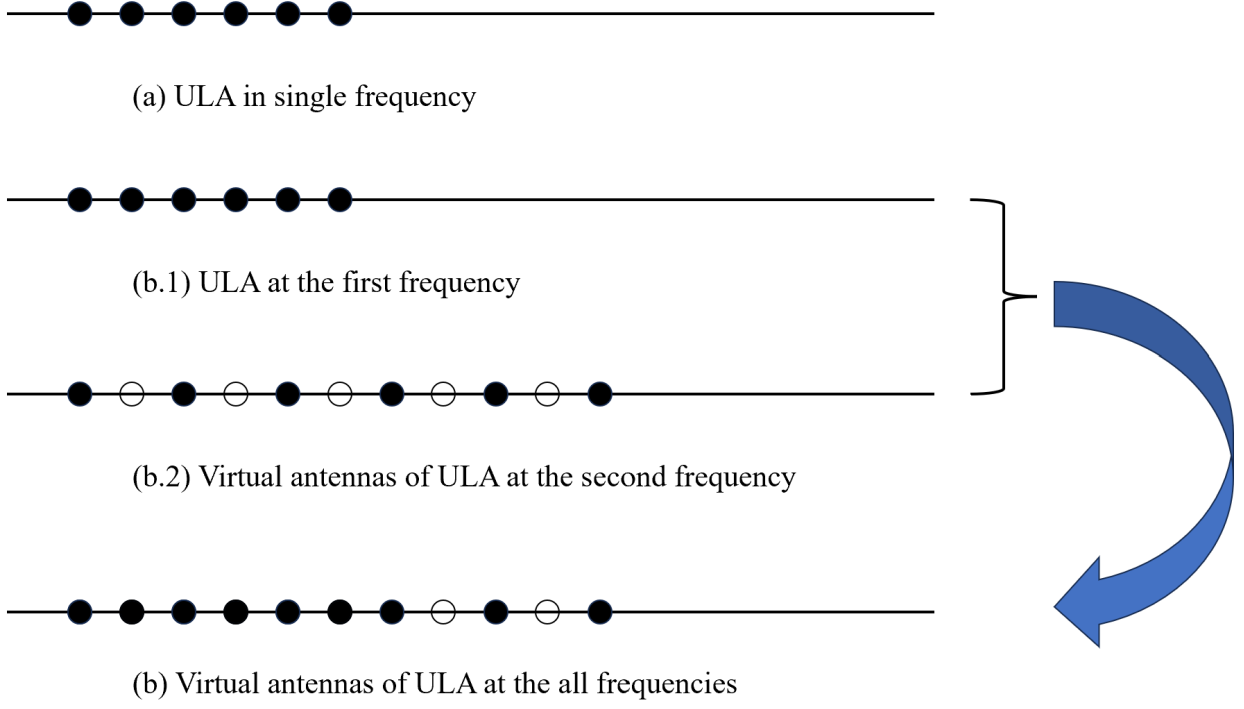


Figure 3.2: An example of virtual array for $(M_T, M_R) = (1, 6)$. White points are holes, which means no antenna in that position. The antenna spacing $d = \lambda_{max}/2$. (a) Single frequency case $\mathcal{F}_s = 1$ GHz. (b) Multiple frequency case $\mathcal{F}_m = \{1, 2\}$ GHz.

frequency set $\mathcal{F}$, the number of missing anti-diagonal can be more. It causes such matrix completion problem more challenging. More detailed discussion can be referred in [18] and [17]. As a result, the (3.37) provide the possibility that more targets can be sensed by the less number of physical antennas.

We can also demonstrate it in one example in Fig. 3.2. Assume there are $M_r = 6$ receiver antennas and $M_t = 1$ transmitter antenna. In single frequency case $\mathcal{F}_s = 1$ GHz, there are 6 antennas, which is shown in Fig. 3.2 (a). It can sense up to $M_r - 1 = 5$ targets. However, in multiple frequency case $\mathcal{F}_m = \{1, 2\}$ GHz. According to (3.19), we can find the virtual antennas location is $\mathcal{I}_{1,6} = \{1, 2, 3, 4, 5, 6\}$ for the first frequency, and $\mathcal{I}_{2,6} = \{1, 3, 5, 7, 9, 11\}$ for the second frequency. (b.1) and (b.2) shows the virtual antennas distribution for the first frequency and second frequency respectively. As shown in 3.2 (b), the union of $\mathcal{I}_{1,6}$ and $\mathcal{I}_{2,6}$ is $\{1, 2, 3, 4, 5, 6, 7, 9, 10, 11\}$. As there are 11 virtual antennas, it can sense up to $11 - 1 = 10$ targets.

3.6 Numerical Tests

In this section, we use numerical experiments to validate the benefits of our multi-band active array. Both ULA and SLA cases are tested. As a popular array in SLA, we test the performance of two-level nested array. To show our sensing ability, we use the SIMO array ($M_T = 1, M_R = 6$) in section 3.6.2 and 3.6.3. For the channel, the speed of the signal propagation $v = 3 \times 10^8$ m/s for electromagnetic signals. To avoid spatial aliasing, the inter-element spacing $d = (v/2f_{min})$ [47]. The noise is modeled as Gaussian processes. The transmitted signal $\mathbf{C}$ is randomly generated. The number of targets S and the number of different frequencies K in (3.8) are known before our signal processing. The following frequency set $\mathcal{F}$ will be used:

- $\mathcal{F}_1 = \{1, 2, 3, 4\}$ GHz, L and S band
- $\mathcal{F}_2 = \{1, 2, 3\}$ GHz

We define the Mean Square Error (MSE) as: $MSE_\theta = \mathbb{E}\{\frac{1}{S} \sum_{s=1}^S (\theta_s - \hat{\theta}_s)^2\}$, $MSE_\Gamma = \mathbb{E}\{\frac{1}{SN} \sum_{i=1}^{SN} (\gamma_i - \bar{\gamma}_i)^2\}$.

3.6.1 Comparison with NNM

In this test, we compare our formulation RPM with its convex relaxation formulation: nuclear norm minimisation (NNM). We only test the DoAs sensing in ULA with $M_T = M_R = 5$ antennas in both single and multiple frequency bands. The parameter $\alpha = 1.0$. The convex relaxation of our formulation (3.28) $g(\mathbf{H}) = \delta(\text{rank}(\mathbf{H}) \leq S)$ is the nuclear norm $\|\mathbf{H}\|_*$. The smooth part f is the same. The study [17] shows the NNM can fail to recover the Hankel matrix if targets are not well separated, even if the measurements are noiseless. In contrast, from our simulations, we observe that the indicator function of rank projection is better at estimating closely targets with higher resolution. To show this weakness of NNM and benefit of RPM, we test the 2 and 4 closed targets DoA sensing.

We replace the indicator function by nuclear norm function: $\|\mathbf{H}\|_*$ as our benchmark in (3.28). According to the traditional beamforming resolution limit [57], the noiseless resolution limit of M -element array is about $2/(M - 1) = 0.5$ (radians) ≈ 28.6 (degrees). One example is two

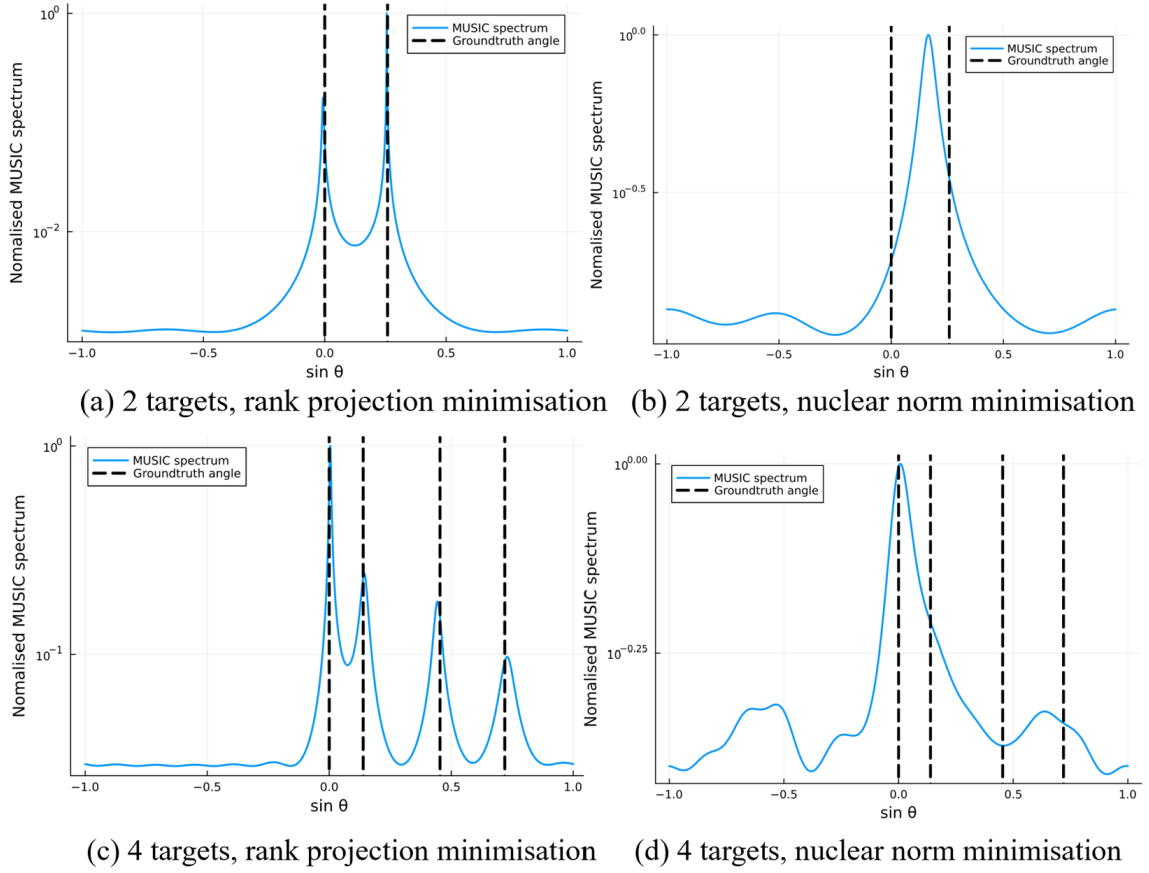


Figure 3.3: Resolution comparison in terms of formulation, with one snapshot, $M_T = M_R = 5$, $K = 1$, SNR = 50 dB: (a) (b) $\mathcal{F} = \{1\}$ kHz, the DoA $\theta = \{0^\circ, 15^\circ\}$. (c) (d) $\mathcal{F} = \{1, 3\}$ kHz, the DoA $\theta = \{0^\circ, 8^\circ, 27^\circ, 46^\circ\}$.

targets estimation. In Fig. 3.3 (a) and (b), we set the angle set $\boldsymbol{\theta} = \{0^\circ, 15^\circ\}$, it is clearly to find our nonconvex formulation RPM can estimate two closely targets accurately, whose angle difference can be less than the resolution limit 28.6° .

To gain more insight, we also test the rank of numerically resulted matrix $\hat{\mathbf{H}}$. In NNM, $\text{rank}(\hat{\mathbf{H}}) = 1 < 2$, which means the poor Hankel matrix recovery. It leads poor DoA estimation. For RPM, the successful matrix recovery is due to $\text{rank}(\hat{\mathbf{H}}) = 2$.

The other example is four closely targets DoA estimation. In Fig. 3.3 (c) and (d), we still set the every angle difference is less than the resolution limit 28.6° . The angle set $\boldsymbol{\theta} = \{0^\circ, 8^\circ, 27^\circ, 46^\circ\}$. Similarly, the RPM can estimate four targets accurately but NNM fails, as $\text{rank}(\hat{\mathbf{H}}) = 2 < 4$ in NNM. These two individual examples can verify the claim in [17] that the NNM fails to recover the Hankel matrix with closely targets. In contrast, RPM can estimate them well. The comparison with other benchmarks can be referred in the chapter 5.

3.6.2 Type of Target Localisation and Identification

This test aims to show our multi-band active array can localise and identify types of targets. With a limited number of snapshots, the array can still sense the DoAs and spectral response accurately. To show the result, the frequencies, DoAs and the norm value of spectral response of each targets are visualised in only one 3D plot. Due to the unique spectral response of targets, the shape of line curve will be different if the corresponding type of targets are different. We choose the frequency set as $\mathcal{F}_1$. The matrix $\mathbf{G} \in \mathbb{C}^{S \times k}$ in (3.8) is randomly generated. Each element is between 0 and 1. In (3.28), we set the parameter $\alpha = 0.5$.

We firstly show the case about the spectral response is independent with frequencies. Now $K = 1$ in (3.8). The groundtruth and its tested plot are shown in the Fig. 3.4.(a.1) and (a.2). The 3 targets are spatially and spectral scanned by our array. It can be found that 3 targets are shown in the same shape for each frequency bands. Only the height of each curve are different due to the distance difference. Both angle and spectral response of the 3 targets are estimated accurately. For the 3 targets with two types, now the $K = 2$. The spectral response is dependent with frequencies. After our sensing progress, the plot is sketched in Fig. 3.4.(b.1)

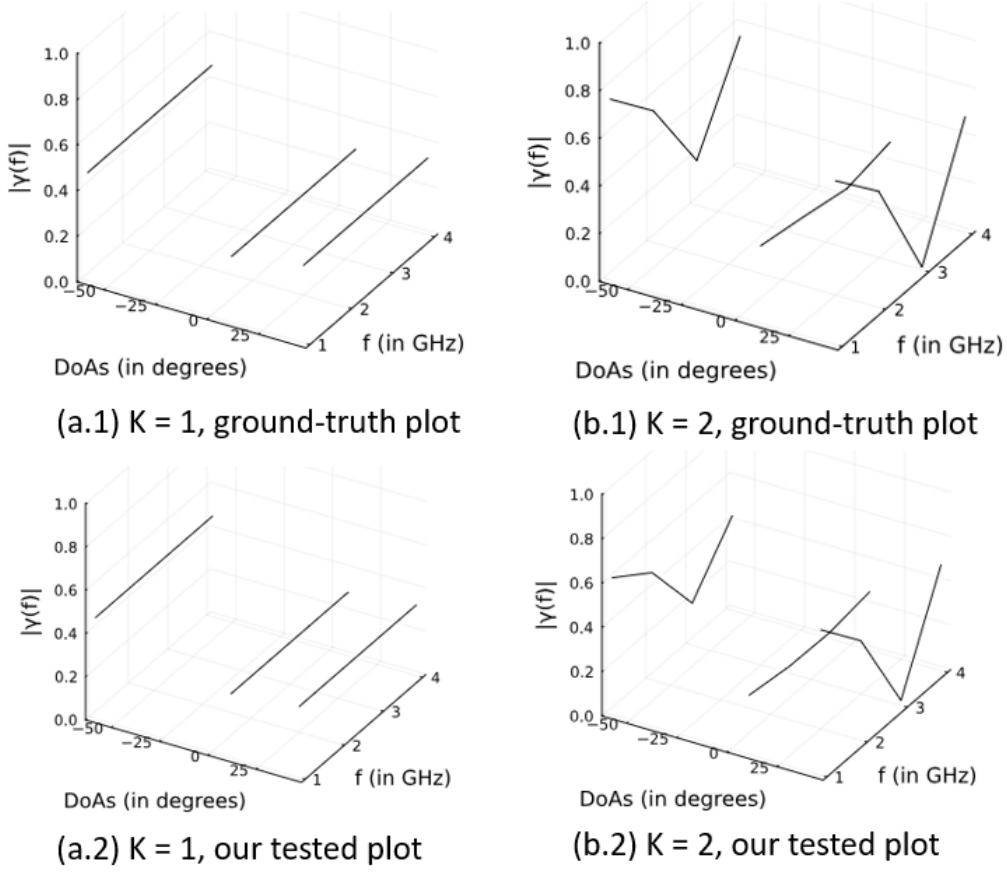


Figure 3.4: Target localisation plot for $(M_T, M_R, S) = (1, 6, 3)$ with only $L = 10$ snapshots and frequency set $\mathcal{F}_1$. DoA set $\boldsymbol{\theta} = \{-60^\circ, 10^\circ, 45^\circ\}$. SNR= 10 dB. (a) $k = 1$. $MSE_\theta = 0.04$, $MSE_\Gamma = 2.57 \times 10^{-4}$. (b) $k = 2$. $MSE_\theta = 0.39$, $MSE_\Gamma = 5.97 \times 10^{-3}$.

and (b.2).

Compared with other methods, we show the benefits of our proposed method in terms of the sensing accuracy of both DoAs and spectral response in Table 3.1. The same setup is used.

3.6.3 More Targets Sensed by the Less Number of Antennas

This test aims to show that SIMO system can sense more targets than the number of antennas, and by using a limited number of snapshots. For ULA, Fig. 3.5 shows a case that 7 targets can be sensed accurately by 6-element ULA. In (3.28), we set the parameter $\alpha_1 = 0.5$.

We extend the ULA case to SLA case. The SLA can sense more targets than the number of not only physical antennas M , but also virtual antennas m_p in (3.4). We start from the 2 level nested array [23]. The in our ULA with 6 antennas and 2 antennas in each level. The antenna

Table 3.1: MSE of both DoAs and spectral response, $(M_T, M_R, S) = (1, 6, 3)$, $SNR = 10dB$, $L = 10$, $S = 3$, Frequency set is $\mathcal{F}_1$. DoA set $\boldsymbol{\theta} = \{-60^\circ, 10^\circ, 45^\circ\}$.

	$K = 1$	$K = 1$	$K = 2$	$K = 2$
	MSE_θ	MSE_γ	MSE_θ	MSE_γ
RPM	0.04	2.57e-4	0.39	5.97e-3
NNM	0.05	0.01	6.36	0.07
CSSM [21]	887.72	NA	899.74	NA
TOPS [43]	1391.33	NA	444.87	NA

position set $\mathcal{M}$ is:

$$\mathcal{M} = \{1, 2, 3, 6\}. \quad (3.39)$$

In this case, the 4-th and 5-th antennas are in sleep mode in our ULA with $m_p = 6$ virtual antennas. In (3.28), we set the parameter $\alpha_1 = 2$. As shown in the figure 3.6, the 6 targets can be estimated accurately with only 4 physical antennas. Compared with the usage of hundreds or thousands of snapshots in [23], the benefit is we can only use $L = 40$ snapshots to sense still more targets than the less number of physical antennas. Now the number of target is the same as the number of virtual antennas $S = m_p$.

3.6.4 Wideband Versus Narrowband in terms of the number of antennas in ULA

This test aims to show our own criteria of wideband in (3.32) by simulations. We use MIMO array with $(M_T, M_R) = (6, 6)$ to improve its RMSE performance. We sense the two closely target, whose DoA set $\boldsymbol{\theta}_1 = \{0^\circ, 5^\circ\}$ and $\boldsymbol{\theta}_2 = \{0^\circ, 20^\circ\}$. We choose the frequency set $\mathcal{F}_3 = \{1.0, 1.5\}$ GHz. In traditional criteria, $\mathcal{F}_3$ can be seen as narrowband. However, from our wideband criteria:

$$\Delta f = 0.5 \text{ GHz} \geq \frac{v}{W} = \frac{2f_1}{M-1} = \frac{2}{M-1} \text{ GHz}. \quad (3.40)$$

(3.40) means that $\mathcal{F}_3$ is wideband when $M \geq 4$. To validate this wideband criteria, we use different number of antennas $M = M_T = M_R$ to test the MSE performance of two closely

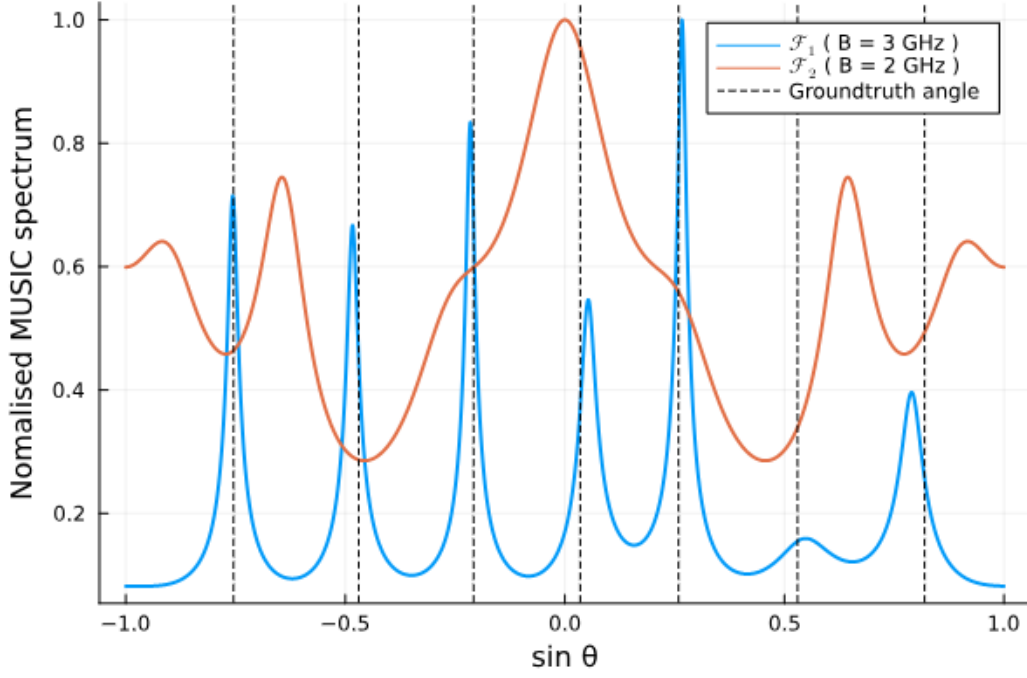


Figure 3.5: MUSIC spectrum for more targets ($S = 7$) sensed by the less number of antennas $(M_T, M_R) = (1, 6)$ in ULA with only 30 snapshots. (Blue: $\mathcal{F}_1$, $(MSE_\theta, MSE_\Gamma) = (1.73, 7.21 \times 10^{-3})$. Orange: $\mathcal{F}_2$, $(MSE_\theta, MSE_\Gamma) = (146.44, 1.91)$). DoA set $\boldsymbol{\theta} = \{-49^\circ, -28^\circ, -12^\circ, 2^\circ, 15^\circ, 32^\circ, 55^\circ\}$. The spectral response is independent with frequencies ($k = 1$). SNR = 10 dB.

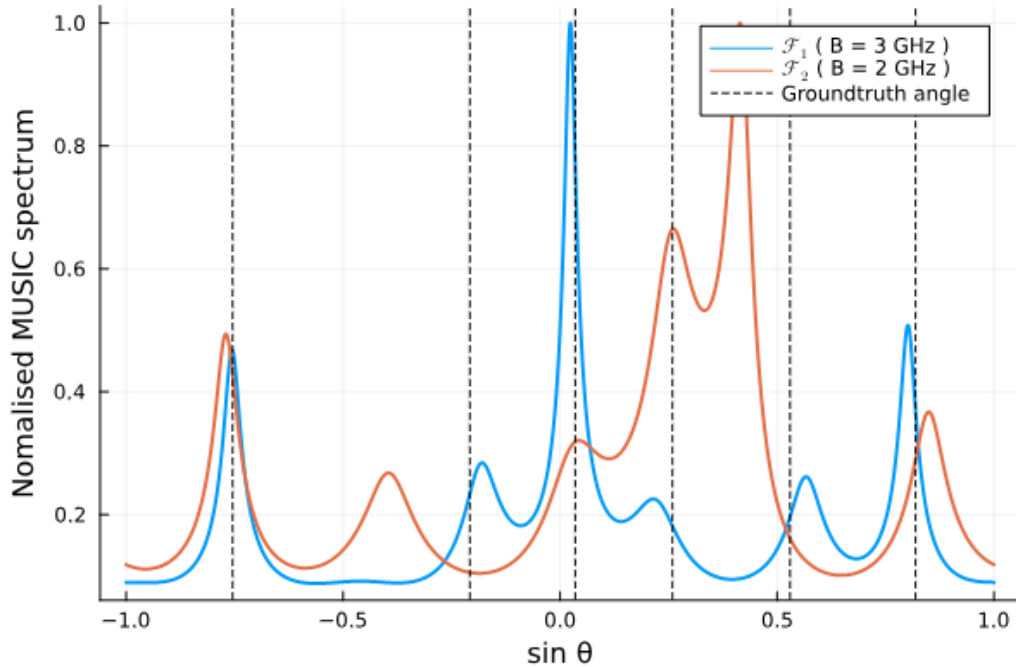


Figure 3.6: MUSIC spectrum for more targets ($S = 6$) sensed by the less number of antennas $(M_T, M_R) = (1, 6)$ in SLA with only 40 snapshots. (Blue: $\mathcal{F}_1$, $(MSE_\theta, MSE_\Gamma) = (3.21, 0.02)$. Orange: $\mathcal{F}_2$, $(MSE_\theta, MSE_\Gamma) = (32.58, 0.02)$). DoA set $\boldsymbol{\theta} = \{-49^\circ, -12^\circ, 2^\circ, 15^\circ, 32^\circ, 55^\circ\}$. The spectral response is independent with frequencies ($k = 1$). SNR = 10 dB.

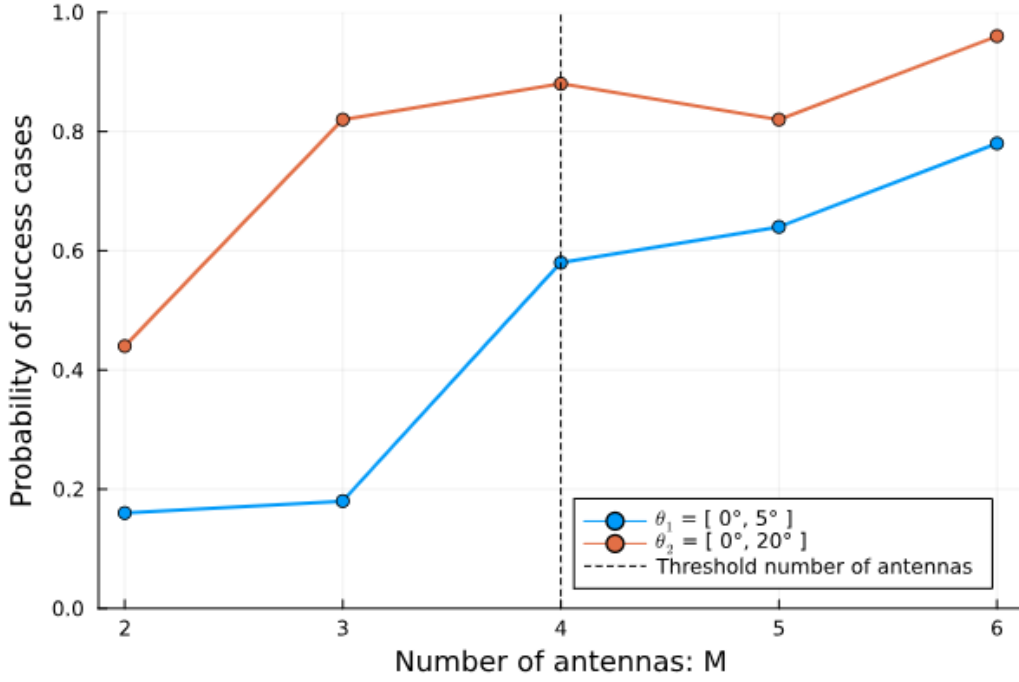


Figure 3.7: The probability of success cases ($RMSE_{\theta} < 1$) in terms of the number of antennas in ULA with $M_T = M_R = 6$. $L = 10$. $\mathcal{F}_3 = \{1.0, 1.5\}$ GHz. $K = 1$. SNR = 30 dB. From (3.40). From our wideband criteria (3.32), $\mathcal{F}_3$ is wideband when $M \geq 4$. The wideband case can detect two closely targets with higher probability.

targets. We define $RMSE_{\theta} < 1$ as the successful trail. The total number of trails is 50. The result is shown in Fig.3.7. Fig.3.7 shows that the $\mathcal{F}_3$ is wideband when the number of antennas $M \geq 4$.

3.7 Summary

In this chapter, our multi-band active array can jointly sense the gridless DoA and spectral response of targets. We make spectral response of targets can be different at different frequency bands. We use GCD to fit all frequency bands in one vector. We use only one variable $\mathbf{X}_G$ to include both DoAs and spectral responses. We formulate a low rank matrix $\mathbf{H}$ including multiple low rank Hankel matrix. This joint sensing problem is formulated as RPM, which is nonconvex optimisation. The degree of freedom have been increased in multi-band. As a result, with a small number of snapshots, both DoAs and spectral response can be jointly sensed, which can localise and identify different targets.

Starting from the resolution of DoA, we also propose our own criteria of wideband: the band-

width depends on the signal speed and the array width. Due to the diversity provided by multi-band, the more virtual antennas are created. It means more targets than the number of physical antennas in ULA and virtual antennas in SLA can be sensed. Our simulations have validated it in even SIMO cases. The number of targets can be even more than the antennas in ULA, and virtual antennas in SLA. It is also noted that our multi-band can also do range sensing. This chapter only focus on DoAs and spectral response sensing only. The range sensing can be our future work for multi-band active array.

Chapter 4

Frequency Diversity Array

4.1 Introduction

This chapter discusses the direction-of-arrival (DoA) and range sensing in Frequency Diverse Array (FDA). FDA is an array that use a small frequency increment from antenna-to-antenna when transmitting signals. Unlike the traditional phased array, FDA applies the frequency increment among the antennas of arrays [15]. Traditionally, most papers calculate the covariance matrix of the received signals, and decompose it by using 2D-MUSIC [58] [59]. The problem is called joint frequency estimation problem. Similar to the 1D-MUSIC, 2D-MUSIC finds the location of peaks of the frequency estimation function with two variables. The computational complexity of the searching peaks is $\mathcal{O}(N^2)$. Then we raise a question: can we have a lower computational complexity to sense DoAs and ranges?

In this chapter, we consider FDA in M -element ULA. The transmit antennas and receive antennas are co-located. we propose a new method without using covariance matrix and 2D-MUSIC. we use DoAs and range variables decoupling method. Then we formulate the first Hankel matrix including only DoAs. The second Hankel matrix including DoAs and ranges. Then we formulate the problem as the low rank Hankel matrix recovery problem. As we apply the MUSIC in DoA domain and range domain independently. The computational complexity of the searching peaks is reduced to $\mathcal{O}(M + N)$.

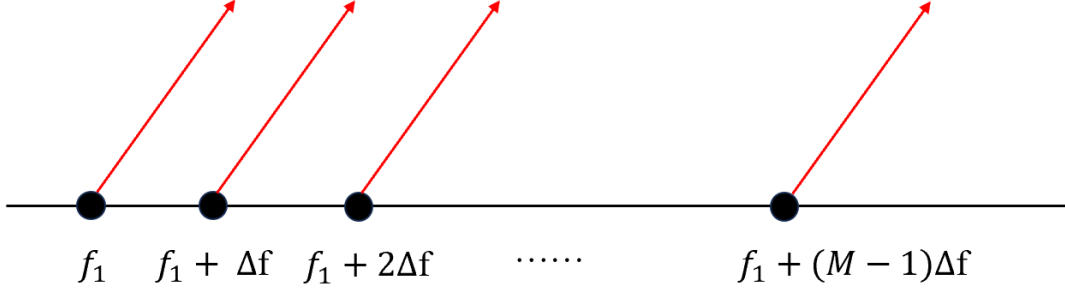


Figure 4.1: Co-located transmitters and receivers with M -element in FDA, and their corresponding transmitted carrier frequencies.

4.2 System Model

This section shows the system models for frequency diversity array (FDA) for ULA. We start from the M -element ULA, including M_T transmit antennas and M_R receive antennas. In this thesis, we assume the target's velocity is 0. The carrier frequency at m -th antenna f_m is:

$$f_m = f_1 + (m - 1)\Delta f, \text{ where } m = 1, 2, \dots, M_T \quad (4.1)$$

where the Δf is the frequency increment between each two adjacent antennas and f_1 is the carrier frequency at the first antenna. We assume that $\Delta f \ll f_1$ (narrowband). Fig. 4.1 shows the structure. The transmitted signal $C_m(t)$ at the m -th antenna is expressed as:

$$C_m(t) = \phi_m(t)e^{-j2\pi f_m t}, \quad 0 \leq t \leq T, \quad (4.2)$$

where E and T is the total transmitted energy and radar pulse duration. $\phi_m(t)$ is the normalised-energy baseband waveform. The waveform is assumed to be approximately orthogonal [60], such as:

$$\int_T \phi_{m_1}(t)\phi_{m_2}^*(t - \tau)dt \approx 0, \quad m_1 \neq m_2, \quad \forall \tau. \quad (4.3)$$

where τ is the time delay. To express the received signals $\mathbf{Y}$ after matched filtering, we define the transmit steering vector as [14]:

$$\mathbf{a}_{M_T}(\theta) \odot \mathbf{b}_{M_T}(r), \quad (4.4)$$

where the transmit range steering vector $\mathbf{b}_{M_T}(r) = [1, e^{-j4\pi\Delta fr/v}, \dots, e^{-j4(M_T-1)\pi\Delta fr/v}]^T$. r is the range of arrival. $\mathbf{a}_{M_T}(\theta) = [1, e^{j2\pi f_1 \frac{d}{v} \sin \theta}, \dots, e^{j(M_T-1)2\pi f_1 \frac{d}{v} \sin \theta}]$ is the transmit angle steering vector. The receive steering vector $\mathbf{a}_{M_R}(\theta)$ is:

$$\mathbf{a}_{M_R}(\theta) = [1, e^{j2\pi f_1 \frac{d}{v} \sin \theta}, \dots, e^{j(M_R-1)2\pi f_1 \frac{d}{v} \sin \theta}]^T. \quad (4.5)$$

Consider S targets with the DoA set $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_S\}$ and range set $\mathbf{r} = \{r_1, \dots, r_S\}$. For the l -th snapshot, after matched filter, the received signals $\mathbf{Y}_l \in \mathbb{C}^{M_T \times M_R}$ is:

$$\mathbf{Y}_l = \sum_{s=1}^S \gamma_s(l) (\mathbf{a}_{M_T}(\theta) \odot \mathbf{b}_{M_T}(r)) \mathbf{a}_{M_R}(\theta_s)^T + \mathbf{E}_l, \quad (4.6)$$

where $\gamma_s(l)$ is the spectral response for s -th target at l -th snapshot. $\mathbf{E}_l \in \mathbb{C}^{M_T \times M_R}$ is the noise at the l -th snapshot. We assume the noise is the additive white Gaussian noise. After vectorisation, the received signals at l -th snapshot: $\mathbf{y}_l = \text{vec}(\mathbf{Y}_l) \in \mathbb{C}^{M_R M_T}$ is:

$$\begin{aligned} \mathbf{y}_l &= \text{vec}(\mathbf{Y}_l) \\ &= \sum_{s=1}^S \gamma_s(l) \mathbf{a}_{M_R}(\theta_s) \otimes (\mathbf{a}_{M_T}(\theta) \odot \mathbf{b}_{M_T}(r)) + \text{vec}(\mathbf{E}_l). \end{aligned} \quad (4.7)$$

For multiple snapshots case, we include L snapshots of the received signals $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_L] \in \mathbb{C}^{M_R M_T \times L}$. To simplify our model, we assume $\gamma_s(l)$ is independent with various snapshots l .

The model is:

$$\mathbf{Y} = \sum_{s=1}^S \gamma_s \mathbf{a}_{M_R}(\theta_s) \otimes (\mathbf{a}_{M_T}(\theta) \odot \mathbf{b}_{M_T}(r)) \mathbf{1}_L^T + \mathbf{E}, \quad (4.8)$$

where $\mathbf{1}_L \in \mathbb{R}^L$ is the column vector whose all elements are 1. The noise $\mathbf{E} = [\text{vec}(\mathbf{E}_1), \dots, \text{vec}(\mathbf{E}_L)] \in \mathbb{C}^{M_R M_T \times L}$.

4.3 Low Rank Matrix Formulation

4.3.1 DoA and Range Variables Decoupling

As shown in (4.7), the DoA set $\boldsymbol{\theta}$ and range set $\mathbf{r}$ are coupled in the transmit steering vector $\mathbf{a}_{M_T}(\boldsymbol{\theta}) \odot \mathbf{b}_{M_T}(\mathbf{r})$, which will lead to ambiguous estimation. To overcome it, this subsection shows angle and range variables decoupling [61]. From (4.7), we reformulate $\mathbf{y}_l$ as:

$$\begin{aligned}
 \mathbf{Y} &= \sum_{s=1}^S \gamma_s \mathbf{a}_{M_R}(\theta_s) \otimes (\mathbf{a}_{M_T}(\boldsymbol{\theta}) \odot \mathbf{b}_{M_T}(\mathbf{r})) \mathbf{1}_L^T + \mathbf{E} \\
 &= \sum_{s=1}^S \gamma_s (\mathbf{1}_{M_R} \otimes \mathbf{b}_{M_T}(\mathbf{r}_s)) \odot (\mathbf{a}_{M_R}(\theta_s) \otimes \mathbf{a}_{M_T}(\theta_s)) \mathbf{1}_L^T + \mathbf{E} \\
 &= \sum_{s=1}^S \gamma_s \text{diag}\{(\mathbf{1}_{M_R} \otimes \mathbf{b}_{M_T}(\mathbf{r}_s))(\mathbf{a}_{M_R}(\theta_s) \otimes \mathbf{a}_{M_T}(\theta_s))^T\} \mathbf{1}_L^T + \mathbf{E} \\
 &= \sum_{s=1}^S \gamma_s \text{diag}\{\mathbf{1}_{M_R} \otimes [\mathbf{b}_{M_T}(\mathbf{r})(\mathbf{a}_{M_R}(\theta_s) \otimes \mathbf{a}_{M_T}(\theta_s))^T]\} \mathbf{1}_L^T + \mathbf{E}, \tag{4.9}
 \end{aligned}$$

where $\text{diag}\{\cdot\}$ extract the diagonal of input matrix as the column vector. $\mathbf{1}_{M_R} \in \mathbb{R}^{M_R}$ is the column vector whose all elements are 1. In (4.9), the angle θ_s has been separated in the vector $\mathbf{a}_{M_R}(\theta_s) \otimes \mathbf{a}_{M_T}(\theta_s)$. However, there are many redundant elements in it. We need to reformulate (4.9) by using the steering vector $\mathbf{a}_{M_T+M_R-1}(\boldsymbol{\theta}) = [1, e^{-j2\pi f_1 \frac{d}{v} \sin \theta}, \dots, e^{-j(M_T+M_R-2)2\pi f_1 \frac{d}{v} \sin \theta}]^T$. According to [61], $\mathbf{Y}$ can be expressed as:

$$\begin{aligned}
 \mathbf{Y} &= \sum_{s=1}^S \gamma_s \mathbf{P}(\mathbf{a}_{M_T+M_R-1}(\theta_s) \otimes \mathbf{b}_{M_T}(\mathbf{r}_s)) \mathbf{1}_L^T + \mathbf{E} \\
 &= \mathbf{P}(\mathbf{A}_{M_T+M_R-1}(\boldsymbol{\theta}) \star \mathbf{B}_{M_T}(\mathbf{r})) \boldsymbol{\gamma} \mathbf{1}_L^T + \mathbf{E}. \tag{4.10}
 \end{aligned}$$

$\mathbf{P} = [\mathbf{P}_1; \dots; \mathbf{P}_{M_R}] \in \mathbb{R}^{M_T M_R \times (M_T+M_R-1)}$ is the selection matrix. For each m , $\mathbf{P}_m \in \mathbb{R}^{M_T \times (M_T+M_R-1)}$ is the matrix with one at index $(i, 1+(m-1)M_T+(M_T+1)(i-1))$, $\forall i$. All other elements are zero. $\mathbf{A}_{M_T+M_R-1}(\boldsymbol{\theta}) = [\mathbf{a}_{M_T+M_R-1}(\theta_1), \dots, \mathbf{a}_{M_T+M_R-1}(\theta_S)]$ and $\mathbf{B}_{M_T}(\mathbf{r}) = [\mathbf{b}_{M_T}(\mathbf{r}_1), \dots, \mathbf{b}_{M_T}(\mathbf{r}_S)]$.

$\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_S]^T$. We denote $\star$ as Khatri-Rao product (or column-wise Kronecker product).

4.3.2 Hankel Matrix Formulation

In (4.10), the DoA set $\boldsymbol{\theta}$ and range set $\mathbf{r}$ are decoupled by using Khatri-Rao product. Based on this result, now we can use one variable to represent $\boldsymbol{\theta}$ without $\mathbf{r}$. To do that, we firstly define the new variable $\mathbf{x}$ as:

$$\begin{aligned}\mathbf{x} &= (\mathbf{A}_{M_T+M_R-1}(\boldsymbol{\theta}) \star \mathbf{B}_{M_T}(\mathbf{r}))\boldsymbol{\gamma} \\ &= \sum_{s=1}^S \gamma_s (\mathbf{a}_{M_T+M_R-1}(\theta_s) \otimes \mathbf{b}_{M_T}(r_s))\end{aligned}\quad (4.11)$$

Now the $\mathbf{Y}$ can be reformulated as:

$$\mathbf{Y} = \mathbf{P}\mathbf{x}\mathbf{1}_L^T + \mathbf{E}. \quad (4.12)$$

As the first element of $\mathbf{a}_{M_T+M_R-1}(\theta_s)$ and $\mathbf{b}_{M_T}(r_s)$ are 1, we can extract them from $\mathbf{x}$ by using linear operator $\mathcal{A}_1(\cdot)$ as:

$$\mathcal{A}_1(\mathbf{x}) = \sum_{s=1}^S \gamma_s \mathbf{a}_{M_T+M_R-1}(\theta_s). \quad (4.13)$$

The linear operator $\mathcal{A}_1(\cdot)$ sampling the input vector with the index set $\mathcal{I}_1 = \{1, M_T+1, \dots, mM_T+1, \dots, (M_T+M_R-2)M_T+1\}$. Now there is linear combinations with single variable in (4.13) and (4.43). We use the linear operator $\mathcal{H}(\cdot)$ to construct the Hankel matrix as the following:

$$\begin{aligned}\mathbf{H}_{\boldsymbol{\theta}} &= \sum_{s=1}^S \mathcal{H}(\gamma_s \mathbf{a}_{M_T+M_R-1}(\theta_s)) \\ &= \mathcal{H}\mathcal{A}_1(\mathbf{x}),\end{aligned}\quad (4.14)$$

$$\begin{aligned}\mathbf{H} &= \sum_{s=1}^S \mathcal{H}(\gamma_s (\mathbf{a}_{M_T+M_R-1}(\theta_s) \otimes \mathbf{b}_{M_T}(r_s))) \\ &= \mathcal{H}(\mathbf{x}).\end{aligned}\quad (4.15)$$

Now $\mathbf{H}_\theta$ is rank- S Hankel matrix. $\mathbf{H}$ is the rank- (SM_T) Hankel matrix, and it has a unique Vandermonde decomposition [62]. The number of rank can be shown as:

$$\begin{aligned}\mathbf{H} &= \sum_{s=1}^S \mathcal{H}(\gamma_s(\mathbf{a}_{M_T+M_R-1}(\theta_s) \otimes \mathbf{b}_{M_T}(r_s))) \\ &= \sum_{s=1}^S \sum_{m=1}^{M_T} \mathcal{H}(\gamma_s(\mathbf{a}_{M_T+M_R-1}(\theta_s) \otimes \mathcal{C}_m(\mathbf{b}_{M_T}(r_s))))).\end{aligned}\quad (4.16)$$

The linear operator $\mathcal{C}_n(\cdot)$ remain the n -th element of the input vector, but forces all other elements to zero. For example, consider $M_T = M_R = 2$, one has:

$$\mathbf{x} = \sum_{s=1}^S \gamma_s(\mathbf{a}_3(\theta_s) \otimes \mathbf{b}_2(r_s)) \quad (4.17)$$

$$= \sum_{s=1}^S \gamma_s([1, e^{j2\pi f_1 \frac{d}{v} \sin \theta_s}, e^{j2 \times 2\pi f_1 \frac{d}{v} \sin \theta_s}]^T \otimes [1, e^{-j4\pi \Delta f r_s/v}]^T) \quad (4.18)$$

$$\begin{aligned}&= \sum_{s=1}^S \gamma_s([1, e^{j2\pi f_1 \frac{d}{v} \sin \theta_s}, e^{j2 \times 2\pi f_1 \frac{d}{v} \sin \theta_s}]^T \otimes [1, 0]^T \\ &+ [1, e^{j2\pi f_1 \frac{d}{v} \sin \theta_s}, e^{j2 \times 2\pi f_1 \frac{d}{v} \sin \theta_s}]^T \otimes [0, e^{-j4\pi \Delta f r_s/v}]^T)\end{aligned}\quad (4.19)$$

$$= \sum_{s=1}^S \gamma_s(\mathbf{x}_{1,s} + \mathbf{x}_{2,s}), \quad (4.20)$$

where $\mathbf{x}_{1,s} = [1, e^{j2\pi f_1 \frac{d}{v} \sin \theta_s}, e^{j2 \times 2\pi f_1 \frac{d}{v} \sin \theta_s}]^T \otimes [1, 0]^T = [1, 0, e^{j2\pi f_1 \frac{d}{v} \sin \theta_s}, 0, e^{j4\pi f_1 \frac{d}{v} \sin \theta_s}]^T$. $\mathbf{x}_{2,s} = [1, e^{j2\pi f_1 \frac{d}{v} \sin \theta_s}, e^{j2 \times 2\pi f_1 \frac{d}{v} \sin \theta_s}]^T \otimes [0, e^{-j4\pi \Delta f r_s/v}]^T = [0, e^{-j4\pi \Delta f r_s/v}, 0, e^{j2\pi f_1 \frac{d}{v} \sin \theta_s - j4\pi \Delta f r_s/v}, 0, e^{j4\pi f_1 \frac{d}{v} \sin \theta_s - j4\pi \Delta f r_s/v}]^T$. Both $\sum_{s=1}^S \mathcal{H}(\mathbf{x}_{1,s})$ and $\sum_{s=1}^S \mathcal{H}(\mathbf{x}_{2,s})$ are rank- S . As these two Hankel matrix are totally distinct with each other, the rank of $\mathcal{H}(\mathbf{x})$ is $2S$.

Based on (4.12), (4.14) and (4.15), we formulate the low rank matrix recovery problem. Given $\mathbf{Y}$ and $\mathbf{P}$, the optimisation problem is:

$$\begin{aligned}\min_{\mathbf{H}_\theta, \mathbf{H}, \mathbf{x}} \quad & \delta(\text{rank}(\mathbf{H}_\theta) \leq S) + \delta(\text{rank}(\mathbf{H}) \leq SM_T) \\ \text{s.t.} \quad & \mathbf{H}_\theta = \mathcal{H}\mathcal{A}_1(\mathbf{x}), \\ & \mathbf{H} = \mathcal{H}(\mathbf{x}), \\ & \mathbf{Y} = \mathbf{P}\mathbf{x}\mathbf{1}_L^T.\end{aligned}\quad (4.21)$$

4.3.3 Problem Solving

To solve (4.21), we apply the following relaxation with positive parameters (α_1, α_2) as:

$$\begin{aligned} \min_{\mathbf{H}_\theta, \mathbf{H}, \mathbf{x}} & \delta(\text{rank}(\mathbf{H}_\theta) \leq S) + \delta(\text{rank}(\mathbf{H}) \leq SM_T) \\ & + \frac{1}{2\alpha_1} \|\mathbf{H}_\theta - \mathcal{H}\mathcal{A}_1(\mathbf{x})\|_F^2 \\ & + \frac{1}{2\alpha_2} \|\mathbf{H} - \mathcal{H}(\mathbf{x})\|_F^2 + \frac{1}{2} \|\mathbf{Y} - \mathbf{P}\mathbf{x}\mathbf{1}_L^T\|_F^2. \end{aligned} \quad (4.22)$$

We choose proximal gradient method to solve (4.22). The PDOM algorithm for this formulation can be referred in the Appendices C. To use proximal gradient method, we should separate the objective function into $g + f$. g is non-differentiable function and f is differentiable function.

We separate the (4.22) as:

$$g(\mathbf{H}_\theta, \mathbf{H}) = \delta(\text{rank}(\mathbf{H}_\theta) \leq S) + \delta(\text{rank}(\mathbf{H}) \leq SM_T) \quad (4.23)$$

$$f(\mathbf{H}_\theta, \mathbf{H}, \mathbf{x}) = \frac{1}{2\alpha_1} \|\mathbf{H}_\theta - \mathcal{H}\mathcal{A}_1(\mathbf{x})\|_F^2 + \frac{1}{2\alpha_2} \|\mathbf{H} - \mathcal{H}(\mathbf{x})\|_F^2 + \frac{1}{2} \|\mathbf{Y} - \mathbf{P}\mathbf{x}\mathbf{1}_L^T\|_F^2. \quad (4.24)$$

Then we derive the first-order derivative as the following:

First-order derivative: The closed form for the partial derivative $\nabla_{\mathbf{x}} f(\mathbf{H}_\theta^i, \mathbf{H}^i, \mathbf{x}^i)$ is:

$$\begin{aligned} \nabla_{\mathbf{x}} f(\mathbf{H}_\theta^i, \mathbf{H}^i, \mathbf{x}^i) &= \frac{1}{\alpha_1} \mathcal{A}_1^* \mathcal{H}^*(\mathbf{H}_\theta^i - \mathcal{H}\mathcal{A}_1(\mathbf{x}^i)) + \frac{1}{\alpha_2} \mathcal{H}^*(\mathbf{H}^i - \mathcal{H}(\mathbf{x}^i)) \\ &+ \mathbf{P}^T (\mathbf{Y} - \mathbf{P}\mathbf{x}^i \mathbf{1}_L^T) \mathbf{1}_L. \end{aligned} \quad (4.25)$$

The closed form for the partial derivative $\nabla_{\mathbf{H}} f(\mathbf{H}_\theta^i, \mathbf{H}^i, \mathbf{x}^i)$ is:

$$\nabla_{\mathbf{H}} f(\mathbf{H}_\theta^i, \mathbf{H}^i, \mathbf{x}^i) = \frac{1}{\alpha_2} (\mathbf{H}^i - \mathcal{H}(\mathbf{x}^i)). \quad (4.26)$$

The closed form for the partial derivative $\nabla_{\mathbf{H}_\theta} f(\mathbf{H}_\theta^i, \mathbf{H}^i, \mathbf{x}^i)$ is:

$$\nabla_{\mathbf{H}_\theta} f(\mathbf{H}_\theta^i, \mathbf{H}^i, \mathbf{x}^i) = \frac{1}{\alpha_1} (\mathbf{H}_\theta^i - \mathcal{H}\mathcal{A}_1(\mathbf{x}^i)). \quad (4.27)$$

$$\mathbf{H}^{i+1} = \mathbf{U}_{:,1:SM_T}^r \text{diag}(\boldsymbol{\sigma}_{1:SM_T}^r) \mathbf{V}_{:,1:SM_T}^{r,H}, \quad (4.30)$$

where $\mathbf{H}^i - \lambda \nabla_{\mathbf{H}} f(\mathbf{H}_{\theta}^i, \mathbf{H}^i, \mathbf{x}^i) = \mathbf{U}^r \text{diag}(\boldsymbol{\sigma}^r) \mathbf{V}^{r,H}$. For $\mathbf{x}^{i+1}$, only the differentiable function f is dependent with $\mathbf{x}$. As a result:

$$\begin{aligned} \mathbf{x}^{i+1} &= \arg \min_{\mathbf{x}} \frac{1}{2\lambda} \|\mathbf{x} - \mathbf{x}^i + \lambda \nabla_{\mathbf{x}} f(\mathbf{H}_{\theta}^i, \mathbf{H}^i, \mathbf{x}^i)\|_F^2 \\ &= \mathbf{x}^i - \lambda \nabla_{\mathbf{x}} f(\mathbf{H}_{\theta}^i, \mathbf{H}^i, \mathbf{x}^i). \end{aligned} \quad (4.31)$$

Subgradient calculation: Proximal gradient method terminate when the subgradient of objective function (4.22) in terms of every variables are sufficiently small. As $g(\mathbf{H}_{\theta}, \mathbf{H})$ is nonsmooth function in terms of both $\mathbf{H}_{\theta}$ and $\mathbf{H}$, we show their subgradient $\partial g(\mathbf{H}_{\theta}, \mathbf{H})$ at $(i+1)$ -th iteration as:

$$\frac{-1}{\lambda} (\mathbf{H}_{\theta}^{i+1} - \mathbf{H}_{\theta}^i + \lambda \nabla_{\mathbf{H}_{\theta}} f(\mathbf{H}_{\theta}^i, \mathbf{H}^i, \mathbf{x}^i)) \in \partial_{\mathbf{H}_{\theta}} g(\mathbf{H}_{\theta}^{i+1}, \mathbf{H}) \quad (4.32)$$

$$\frac{-1}{\lambda} (\mathbf{H}^{i+1} - \mathbf{H}^i + \lambda \nabla_{\mathbf{H}} f(\mathbf{H}_{\theta}^i, \mathbf{H}^i, \mathbf{x}^i)) \in \partial_{\mathbf{H}} g(\mathbf{H}_{\theta}, \mathbf{H}^{i+1}). \quad (4.33)$$

Proof. From (4.28), we denote the objective function $\phi_1(\mathbf{H}_{\theta})$ as:

$$\phi_1(\mathbf{H}_{\theta}) := \delta(\text{rank}(\mathbf{H}_{\theta}) \leq S) + \frac{1}{2\lambda} \|\mathbf{H}_{\theta} - \mathbf{H}_{\theta}^i + \lambda \nabla_{\mathbf{H}_{\theta}} f(\mathbf{H}_{\theta}^i, \mathbf{H}^i, \mathbf{x}^i)\|_F^2. \quad (4.34)$$

We set its subgradient at the global optima $\mathbf{H}_{\theta}^{i+1}$ as 0. One has:

$$0 \in \partial \phi(\mathbf{H}_{\theta}^{i+1}). \quad (4.35)$$

Then we can derive:

$$\nabla_{\mathbf{H}_{\theta}} \frac{1}{2\lambda} \|\mathbf{H}_{\theta}^{i+1} - \mathbf{H}_{\theta}^i + \lambda \nabla_{\mathbf{H}_{\theta}} f(\mathbf{H}_{\theta}^i, \mathbf{H}^i, \mathbf{x}^i)\|_F^2 \in \partial_{\mathbf{H}_{\theta}} g(\mathbf{H}_{\theta}^{i+1}, \mathbf{H}) \quad (4.36)$$

$$\frac{-1}{\lambda} (\mathbf{H}_{\theta}^{i+1} - \mathbf{H}_{\theta}^i + \lambda \nabla_{\mathbf{H}_{\theta}} f(\mathbf{H}_{\theta}^i, \mathbf{H}^i, \mathbf{x}^i)) \in \partial_{\mathbf{H}_{\theta}} g(\mathbf{H}_{\theta}^{i+1}, \mathbf{H}). \quad (4.37)$$

For the variable $\mathbf{H}$, the objective function $\phi_2(\mathbf{H})$ is:

$$\phi_2(\mathbf{H}) := \delta(\text{rank}(\mathbf{H}) \leq SM_T) + \frac{1}{2\lambda} \|\mathbf{H} - \mathbf{H}^i + \lambda \nabla_{\mathbf{H}} f(\mathbf{H}_{\theta}^i, \mathbf{H}^i, \mathbf{x}^i)\|_F^2. \quad (4.38)$$

We set its subgradient at the global optima $\mathbf{H}^{i+1}$ as 0. Similarly, we can derive:

$$\frac{-1}{\lambda}(\mathbf{H}^{i+1} - \mathbf{H}^i + \lambda \nabla_{\mathbf{H}} f(\mathbf{H}_{\theta}^i, \mathbf{H}^i, \mathbf{x}^i)) \in \partial_{\mathbf{H}} g(\mathbf{H}_{\theta}, \mathbf{H}^{i+1}). \quad (4.39)$$

□

Stopping Criteria: Then we conclude our stopping criterion as:

$$\partial g(\mathbf{H}_{\theta}^{i+1}) + \nabla f(\mathbf{H}_{\theta}^{i+1}) < 10^{-6}, \quad (4.40)$$

$$\partial g(\mathbf{H}^{i+1}) + \nabla f(\mathbf{H}^{i+1}) < 10^{-6}, \quad (4.41)$$

$$\nabla f(\mathbf{x}^{i+1}) < 10^{-6}. \quad (4.42)$$

4.3.4 MUSIC and identifiability

Extract DoAs and ranges by using MUSIC algorithm: After getting the estimated variables $(\hat{\mathbf{H}}_{\theta}, \hat{\mathbf{H}}, \hat{\mathbf{x}})$, we apply MUSIC algorithm in $\hat{\mathbf{H}}_{\theta}$ to get the angle θ . We define the linear operator $\mathcal{A}_2(\cdot)$ as sampling the input vector with the index set $\mathcal{I}_2 = \{1, 2, \dots, M_T\}$. From (4.11), the $\mathcal{A}_2(\mathbf{x})$ is

$$\mathcal{A}_2(\mathbf{x}) = \sum_{s=1}^S \gamma_s \mathbf{b}_{M_T}(r_s). \quad (4.43)$$

We extract the ranges from the estimated variable $\hat{\mathbf{x}}$. We construct the Hankel matrix $\hat{\mathbf{H}}_r$ as:

$$\hat{\mathbf{H}}_r = \mathcal{H} \mathcal{A}_2(\hat{\mathbf{x}}), \quad (4.44)$$

Now the Hankel matrix $\hat{\mathbf{H}}_r \in \mathbb{C}^{\frac{M_T+1}{2} \times \frac{M_T+1}{2}}$ is still rank- S . As there is only one variable: range in the Hankel matrix $\hat{\mathbf{H}}_r$, we can apply the MUSIC algorithm to estimate the range $\hat{r}$.

Identifiability: In our formulation (4.22), $\mathbf{H}_{\theta} \in \mathbb{C}^{\frac{M_T+M_R}{2} \times \frac{M_T+M_R}{2}}$. The Hankel matrix $\mathbf{H}_{\theta}$ can be uniquely recovered if the rank $S < \frac{M_T+M_R}{2}$. It means that The maximum resolvable DoAs is $(\frac{M_T+M_R}{2} - 1)$. For the range, the Hankel matrix $\hat{\mathbf{H}}_r \in \mathbb{C}^{\frac{M_T+1}{2} \times \frac{M_T+1}{2}}$ in (4.44) can be uniquely

recovered if the rank $S < \frac{M_T+1}{2}$. It means that The maximum resolvable DoAs is $(\frac{M_T+1}{2} - 1)$.

4.4 Numerical Tests

In this section, with a small number of snapshots, we show our FDA can sense DoAs and ranges. We consider an co-located ULA with $M_T = M_R = 10$ transmitters and receivers. The inter-element spacing $d = \lambda_1/2$ [47], where λ_1 is the wavelength of the carrier frequency f_1 . The transmitted signals are electromagnetic signal with the speed $v = 3 \times 10^8$ m/s. Both the transmitted signals $\mathbf{C}$ and the additive noise $\mathbf{E}$ are synthetically generated by Gaussian random processes. The signal-to-noise ratio (SNR) is defined as $10 \log_{10} \frac{\|\mathbf{C}\|_F^2}{\|\mathbf{E}\|_F^2}$ in dB. We calculate the Root Mean Square Error (RMSE) as $\sqrt{\frac{1}{S} \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_F^2}$ for DoAs, and $\sqrt{\frac{1}{S} \|\mathbf{r} - \hat{\mathbf{r}}\|_F^2}$ for spectral response. In (4.22), the parameter $\alpha_1 = \alpha_2 = 0.1$. From our identifiability analysis, the maximum number of resolvable DoAs is 9. The maximum number of resolvable ranges is 5.

4.4.1 Joint Sensing Performance

We firstly test the two targets sensing case with a limited number of snapshots $L = 10$. The two targets are located at $[-40^\circ, 0.2 \text{ km}]$ and $[20^\circ, 1.5 \text{ km}]$. SNR = 30 dB. Fig. 4.2 shows the normalised MUSIC spectrum of the the two targets sensing for both DoAs and ranges. Now the two targets are sensed accurately with low RMSE.

Then we sense one more targets. Now the number of targets is approaching to the maximum number of resolvable ranges. The three targets are located at $[-40^\circ, 0.4 \text{ km}]$, $[5^\circ, 1.1 \text{ km}]$ and $[20^\circ, 1.8 \text{ km}]$. SNR = 30 dB. Fig. 4.3 shows the three targets are sensed accurately.

4.4.2 RMSE versus SNR

In this subsection, we show how the RMSE performance in terms of SNR with a limited number of snapshots. We still test the two targets sensing case with snapshots $L = 10$ and $L = 5$. The two targets are located at $[-40^\circ, 0.2 \text{ km}]$ and $[20^\circ, 1.5 \text{ km}]$. SNR = 30 dB. The result is shown in Fig. 4.4.

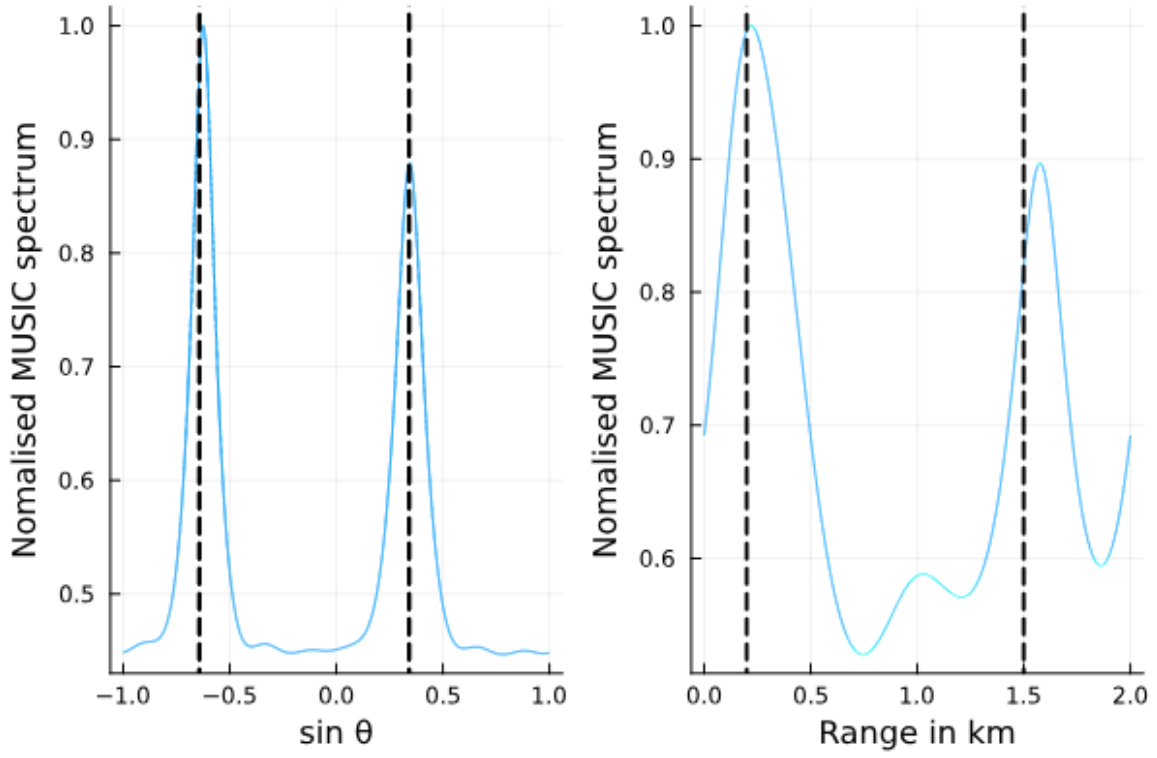


Figure 4.2: MUSIC spectrum for the targets sensed by 10-antenna FDA with only $L = 10$ snapshots ($M_T = M_R = 10$). The carrier frequency $f_1 = 1$ GHz. The frequency increment $\Delta f = 75$ MHz. The groundtruth DoAs $\boldsymbol{\theta} = \{-40^\circ, 20^\circ\}$ and range $\boldsymbol{r} = \{0.2, 1.5\}$ km. SNR = 30 dB. $(RMSE_\theta, RMSE_r) = (0.99, 5.59 \times 10^{-2})$.

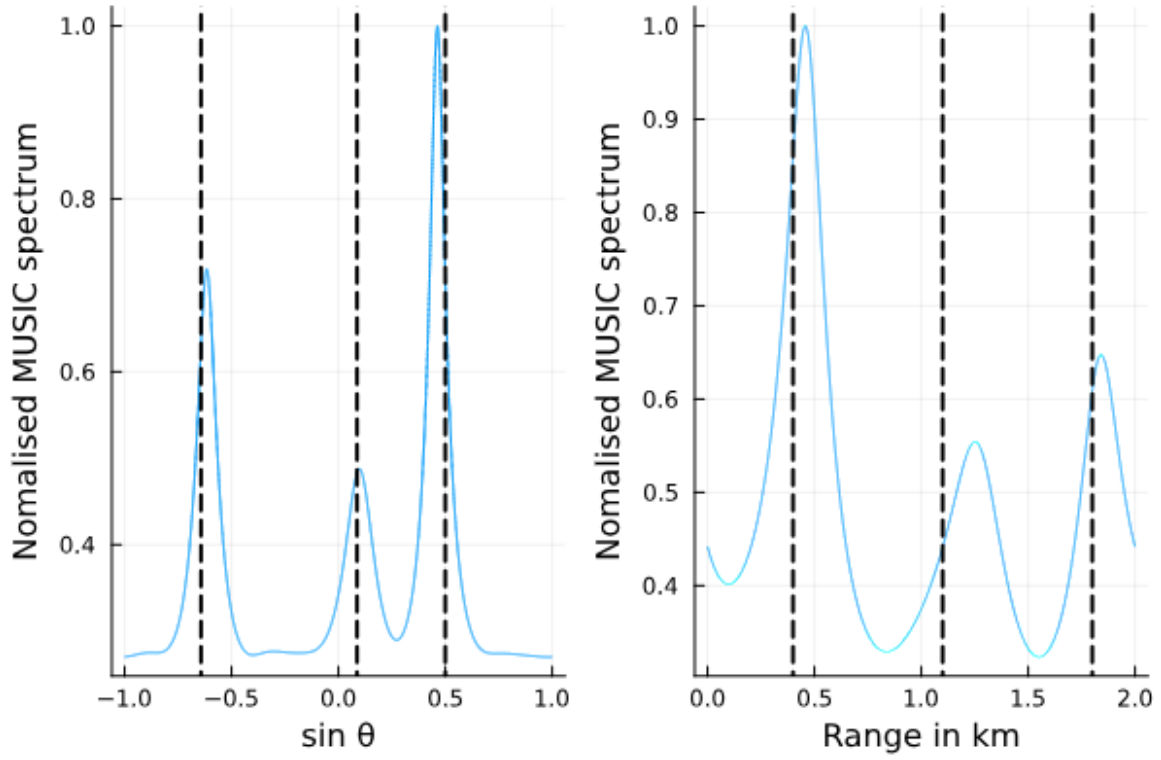


Figure 4.3: MUSIC spectrum for the targets sensed by 10-antenna FDA with only $L = 10$ snapshots ($M_T = M_R = 10$). The carrier frequency $f_1 = 1$ GHz. The frequency increment $\Delta f = 75$ MHz. The groundtruth DoAs $\boldsymbol{\theta} = \{-40^\circ, 5^\circ, 20^\circ\}$ and range $\boldsymbol{r} = \{0.4, 1.1, 1.8\}$ km. SNR = 30 dB. $(RMSE_\theta, RMSE_r) = (1.84, 5.59 \times 10^{-2})$.

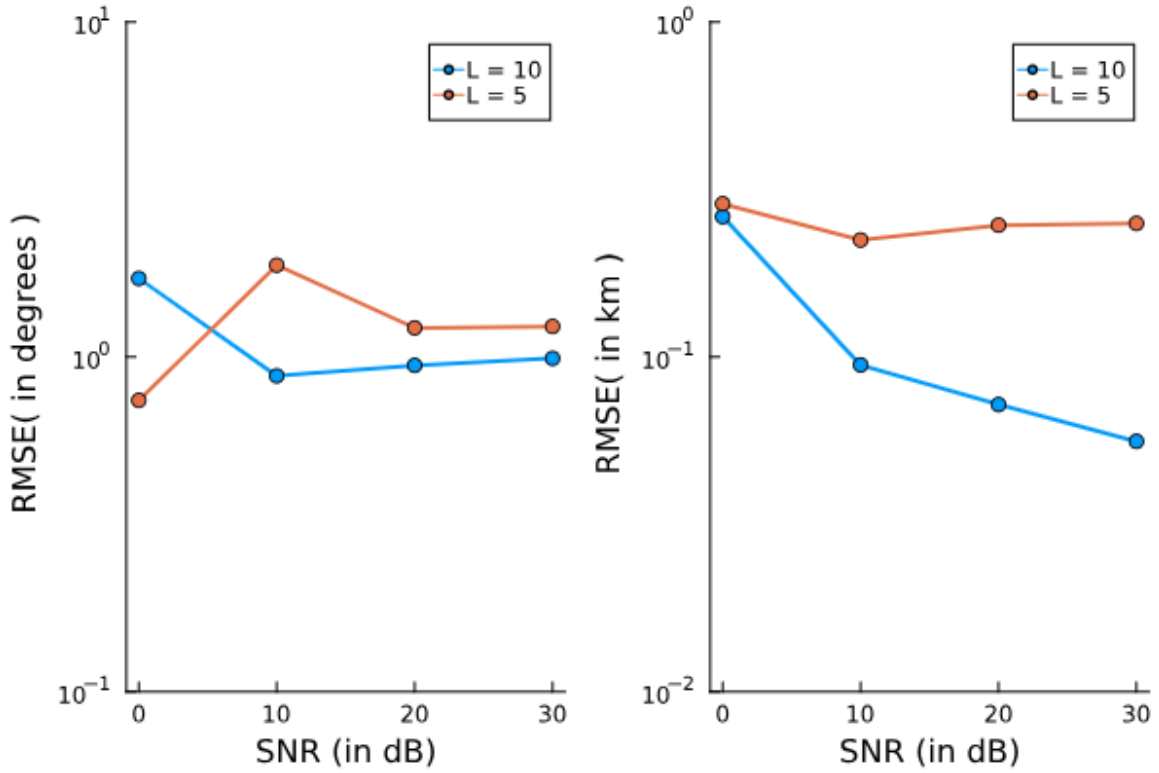


Figure 4.4: RMSE versus SNR in a limited snapshots $L = 10$ and $L = 5$. The carrier frequency $f_1 = 1$ GHz. The frequency increment $\Delta f = 75$ MHz. The groundtruth DoAs $\theta = \{-40^\circ, 20^\circ\}$ and range $r = \{0.2, 1.5\}$ km.

4.5 Summary

In this chapter, we use the angle and range variables decoupling from [61]. Without 2D-MUSIC from covariance matrix, we formulate two Hankel matrix from the signal to be estimated: one only contains DoAs and the other contains DoAs and ranges. Now we formulate the low rank Hankel matrix recovery problem, and solved it by using proximal gradient method. With a limited number of snapshots, we can sense the DoA and range of targets. Compared with 2D-MUSIC, our method has a lower computational complexity. This chapter only consider the narrow bandwidth scenario. The FDA with wider bandwidth scenario is an interesting topic for the future work.

Chapter 5

Nonconvex Algorithms Analysis

5.1 Introduction

In the last two chapters, we have used proximal gradient method and PDOM to solve our optimisation problems. These two algorithms can be classified as nonconvex algorithms. This chapter aims to show more details about these two algorithms. Furthermore, we compare the proximal gradient method and PDOM in terms of convergence rate. Then we evaluate the RMSE performance of our formulation. We use the optimisation problem proposed in the chapter 3. The first part of this chapter can be seen as algorithms discussion for the last two chapters.

In the second part of this chapter, we would like to discuss another nonconvex algorithm: Alternating Direction Method of Multipliers (ADMM). In this part, we use ADMM to solve short-and-sparse blind deconvolution problem. We show that ADMM can be applied to non-convex algorithms, and can achieve good results.

5.2 Proximal gradient method and PDOM

5.2.1 Introduction

In chapter 3 and 4, we use proximal gradient method to solve the formulations. Proximal gradient method can guarantee the convergence with the rate is $\mathcal{O}(1/k)$ at k -th iteration. Apart from accelerated proximal gradient method [34], is there another nonconvex algorithm with faster convergence rate? PDOM has a faster convergence rate $\mathcal{O}(1/k^2)$ at k -th iteration [35]. In this section, we show both performances of proximal gradient method and PDOM. We show the RMSE performance of proximal gradient method, and compare the convergence rate of proximal gradient method and PDOM. Followed by the chapter 3, we compare proximal gradient method with PDOM in the optimisation problem (3.28) in the chapter 3. The problem is:

$$\begin{aligned} \min_{\mathbf{X}_G, \mathbf{H}} \quad & \delta(\text{rank}(\mathbf{H}) \leq S) + \frac{1}{2\alpha} \left\| \mathbf{H} - \sum_{n=1}^N \mathcal{L}_n^1(\mathbf{X}_G \mathbf{F}_k) \right\|_F^2 \\ & + \sum_{n=1}^N \frac{1}{2} \left\| \mathbf{Y}(f_n) - \mathcal{L}_n^2(\mathbf{X}_G \mathbf{F}_k) \mathbf{C}(f_n) \right\|_F^2. \end{aligned} \quad (5.1)$$

The linear operator $\mathcal{L}_n^1(\cdot) = \mathcal{W}_n \mathcal{H} \mathcal{B}_n(\cdot)$. The linear operator $\mathcal{W}_n(\cdot)$ puts $\mathbf{H}_n$ at $\mathbf{H}$. $\mathcal{H}(\cdot)$ is the Hankel matrix operator, which is defined in (3.12). $\mathcal{B}_n(\cdot)$ is taking the n -th column of input matrix. The linear operator $\mathcal{L}_n^2(\cdot) = \mathcal{H}_{M_R, M_T} \mathcal{A}_n \mathcal{B}_n(\cdot)$. $\mathcal{A}_n(\cdot)$ is defined in (3.18) and (3.19). We calculate the Cramér–Rao Bound for this problem, and compare proximal gradient method.

5.2.2 Proximal Gradient Method details

Algorithm 5 summarises the steps of the standard proximal gradient method to (3.28).

The partial gradient of the smooth function f can be referred in Appendices A. Then we show the closed form solution for $(\mathbf{X}_G^{i+1}, \mathbf{H}^{i+1})$. For $\mathbf{X}_G^{i+1}$, only the differentiable function f is dependent with $\mathbf{X}_G$. As a result:

$$\begin{aligned} \mathbf{X}_G^{i+1} &= \arg \min_{\mathbf{X}_G} \left\| \mathbf{X}_G - \mathbf{X}_G^i + \lambda \nabla_{\mathbf{X}_G} f(\mathbf{X}_G^i, \mathbf{H}^i) \right\|_F^2 \\ &= \mathbf{X}_G^i - \lambda \nabla_{\mathbf{X}_G} f(\mathbf{X}_G^i, \mathbf{H}^i) \end{aligned} \quad (5.2)$$

Algorithm 5: Proximal Gradient Method for (3.28)

Define: $\hat{f}_\lambda(u, v) = f(u) + \nabla f(u)^T(v - u) + \frac{1}{2\lambda}\|v - u\|_F^2$
Initialise: $(\mathbf{X}_G^1, \mathbf{H}^1)$
Data: $\beta = 0.5$
while (5.9) and (5.10) are not satisfied or $i < 2000$ **do**
 $\lambda \leftarrow 1$;
 while *true* ; /* Update step size λ */
 do
 $\mathbf{X}_G^{i+1} = \mathbf{X}_G^i - \lambda \nabla_{\mathbf{X}_G} f(\mathbf{X}_G^i, \mathbf{H}^i)$;
 $\mathbf{H}^{i+1} = \text{prox}_{\lambda, g}(\mathbf{H}^i - \lambda \nabla_{\mathbf{H}} f(\mathbf{X}_G^i, \mathbf{H}^i))$;
 Update $\hat{f}_\lambda((\mathbf{X}_G^i, \mathbf{H}^i), (\mathbf{X}_G^{i+1}, \mathbf{H}^{i+1}))$;
 if $f(\mathbf{X}_G^{i+1}, \mathbf{H}^{i+1}) \leq \hat{f}_\lambda((\mathbf{X}_G^i, \mathbf{H}^i), (\mathbf{X}_G^{i+1}, \mathbf{H}^{i+1}))$ **then**
 end
 Output: $\lambda, (\mathbf{X}_G^{i+1}, \mathbf{H}^{i+1})$;
 Break the inner loop;
 Update $\lambda \leftarrow \beta \lambda$;
 end
 end
end
Output: $(\hat{\mathbf{X}}_G, \hat{\mathbf{H}}) \leftarrow (\mathbf{X}_G^{i+1}, \mathbf{H}^{i+1})$;

For $\mathbf{H}^{i+1}$, the nondifferentiable function g is dependent with $\mathbf{H}$. As a result:

$$\begin{aligned}
\mathbf{H}^{i+1} &= \arg \min_{\mathbf{H}} \delta(\text{rank}(\mathbf{H}) \leq S) \\
&\quad + \|\mathbf{H} - \mathbf{H}^i + \lambda \nabla_{\mathbf{H}} f(\mathbf{X}_G^i, \mathbf{H}^i)\|_F^2 \\
&= \text{prox}_{\lambda, g}(\mathbf{H}^i - \lambda \nabla_{\mathbf{H}} f(\mathbf{X}_G^i, \mathbf{H}^i))
\end{aligned} \tag{5.3}$$

where the proximal operator $\text{prox}(\cdot)$ is:

$$\text{prox}_{\lambda, g}(\mathbf{v}) = \arg \min_{\mathbf{u}} g(\mathbf{u}) + \frac{1}{2\lambda} \|\mathbf{u} - \mathbf{v}\|^2$$

To solve (5.3), we consider SVD: $\mathbf{H}^i - \lambda \nabla_{\mathbf{H}} f(\mathbf{X}_G^i, \mathbf{H}^i) = \mathbf{U} \text{diag}(\boldsymbol{\sigma}) \mathbf{V}^T$. The closed-form solution for (5.3) is:

$$\mathbf{H}^{i+1} = \mathbf{U}_{:,1:S} \text{diag}(\boldsymbol{\sigma}_{1:S}) \mathbf{V}_{:,1:S}^T \tag{5.4}$$

Proximal gradient method terminate when the subgradient of objective function (3.28) is suffi-

ciently small. As $g(\mathbf{H})$ is nonsmooth function, we also show its subgradient $\partial g(\mathbf{H})$ at $i + 1$ -th iteration as:

$$\partial g(\mathbf{H}) = \frac{-1}{\lambda^i}(\mathbf{H}^i - \mathbf{H}^i) - \frac{1}{\alpha}(\mathbf{H} - \sum_{n=1}^N \mathcal{L}_n^1(\mathbf{X}_G \mathbf{F}_k)). \quad (5.5)$$

Proof. From (5.3),

$$\mathbf{H}^{i+1} = \arg \min_{\mathbf{H}} g(\mathbf{H}) + \nabla_{\mathbf{H}} \hat{f}_{\lambda}(\mathbf{H}, \mathbf{H}^i) \quad (5.6)$$

Then we can derive:

$$\partial g(\mathbf{H}) + \nabla_{\mathbf{H}} \hat{f}_{\lambda}(\mathbf{H}, \mathbf{H}^i) = 0 \quad (5.7)$$

$$\partial g(\mathbf{H}) + \frac{1}{\lambda^i}(\mathbf{H} - \mathbf{H}^i) + \nabla f(\mathbf{H}^i) = 0 \quad (5.8)$$

□

Then we conclude our stopping criterion as:

$$\partial g(\mathbf{H}^{i+1}) + \nabla f(\mathbf{H}^{i+1}) < 10^{-6}, \quad (5.9)$$

$$\nabla f(\mathbf{X}_G^{i+1}) < 10^{-6}. \quad (5.10)$$

5.2.3 PDOM details

To calculate the positive semi-definite Hessian matrix $\mathbf{Q}$ in (eq.11, [35]), The derivation of $\mathbf{Q}$ can be referred in Appendices B. This part only discuss SIMO case ($M_T = 1$) and spectral response is independent with frequency ($K = 1$). Based on the Hessian matrix $\mathbf{Q}$ and [35], then we summarise the steps of PDOM to (3.28).

Algorithm 6: PDOM for (3.28)

Define: $\mathbf{p}(\alpha) := \begin{cases} -\tau \mathbf{g}, & \alpha \in (0, 1] \\ -\tau \mathbf{g} + (\alpha - 1)(-\mathbf{Q}^{-1} \mathbf{g} + \tau \mathbf{g}), & \alpha \in (1, 2] \end{cases}$

where $\mathbf{g} := [\text{vec}(\nabla_{\mathbf{X}_G} f)^T, \text{vec}(\nabla_{\mathbf{H}} f)^T]^T$;

The surrogate function $\hat{f}_\lambda(u, v) = f(u) + \mathbf{g}_\alpha^T(v - u) + \frac{1}{2\lambda} \|v - u\|_F^2$;

where $\mathbf{g}_\alpha = \mathbf{g}^T \mathbf{p}^2(\alpha) / \|\mathbf{p}(\alpha)\|^2$

Input: $(\mathbf{X}_G^1, \mathbf{H}^1), \mathbf{Q}^{-1}$

Data: fixed step size $\tau = 1/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{Q}$.

$\lambda = 0.98, \alpha \in [0, 2]$

while (5.11) and (5.12) are not satisfied or $i < 2000$ **do**

```

; /* PDOM */
vec( $\mathbf{X}_G$ )i+1 = vec( $\mathbf{X}_G$ )i +  $\mathbf{p}_\lambda(\alpha^i)$ ;
 $\tau_{\alpha^i} = -\|\mathbf{p}(\alpha)\|^2 / (\mathbf{g}^T \mathbf{p}(\alpha))$  ;
vec( $\mathbf{H}$ )i+1 = vec(prox $\lambda \tau_{\alpha^i} \mathbf{g}$ ( $\mathbf{H}^i + \mathbf{p}_\lambda(\alpha^i)$ ));
; /* Update step size  $\tau_\alpha$  */
Find the largest value  $\alpha^i \in \{1 + (1/2)^k | k \in \mathbb{N}\}$  such that
 $f(\mathbf{X}_G^{i+1}, \mathbf{H}^{i+1}) \leq \hat{f}_{\alpha^i}((\mathbf{X}_G^{i+1}, \mathbf{H}^{i+1}), (\mathbf{X}_G^i, \mathbf{H}^i))$ . ;
/* Proximal gradient method */
 $\mathbf{X}_{G_v}^{i+1} = \mathbf{X}_G^i - \tau \nabla_{\mathbf{X}_G} f(\mathbf{X}_G^i, \mathbf{H}^i)$ ;
 $\mathbf{H}_v^{i+1} = \text{prox}_{\tau, \mathbf{g}}(\mathbf{H}^i - \tau \nabla_{\mathbf{H}} f(\mathbf{X}_G^i, \mathbf{H}^i))$ ;
; /* Comparison and selection */
if  $f(\mathbf{X}_G^{i+1}, \mathbf{H}^{i+1}) > f(\mathbf{X}_{G_v}^{i+1}, \mathbf{H}_v^{i+1})$  then
|  $\mathbf{X}_G^{i+1} = \mathbf{X}_{G_v}^{i+1}$  ;
|  $\mathbf{H}^{i+1} = \mathbf{H}_v^{i+1}$  ;
end

```

end

Output: $(\hat{\mathbf{X}}_G, \hat{\mathbf{H}}) \leftarrow (\mathbf{X}_G^{i+1}, \mathbf{H}^{i+1})$;

Similar as PG, the stopping criterion of PDOM is:

$$\begin{aligned} \|\partial g(\mathbf{H}^{i+1}) + \nabla f(\mathbf{H}^{i+1})\| &< 10^{-6} \sqrt{h_1 h_2 N} \\ &+ 10^{-6} \max\{\|\nabla f(\mathbf{H}^{i+1})\|, \frac{1}{\gamma \tau_{\alpha^k}} \|\mathbf{H}^{i+1}\|, \frac{1}{\gamma \tau_{\alpha^k}} \|\mathbf{H}^i\|\} \end{aligned} \quad (5.11)$$

$$\begin{aligned} \|\nabla f(\mathbf{X}_{\mathbf{G}}^{i+1})\| &< 10^{-6} \sqrt{M_g K} + 10^{-6} \max\{\frac{1}{\gamma \tau_{\alpha^k}} \|\mathbf{X}_{\mathbf{G}}^{i+1}\|, \\ &\frac{1}{\gamma \tau_{\alpha^k}} \|\mathbf{X}_{\mathbf{G}}^i\|\}. \end{aligned} \quad (5.12)$$

The subgradient $\partial g(\mathbf{H}^{i+1})$ is:

$$\partial g(\mathbf{H}^{i+1}) = \frac{-1}{\lambda \tau_{\alpha^i}} (\mathbf{H}^{i+1} - \mathbf{H}^i - \mathbf{P}_{\lambda}(\alpha^i)). \quad (5.13)$$

5.2.4 Cramér–Rao Bound Derivation

This section aims to derive the CRB on unbiased estimator of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$. CRB provides lower bound for variance of any estimator and lower bound for mean square error for unbiased estimator. CRB is useful to evaluate the estimator performance [63] [64], which means the estimator is the best if an unbiased estimate attains CRB. The unknown parameter vector $\mathbf{u} = [\boldsymbol{\theta}, \bar{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\gamma}}]^T$, where $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_S\}$, $\bar{\boldsymbol{\gamma}} = \{\bar{\gamma}_1, \dots, \bar{\gamma}_S\}$ and $\tilde{\boldsymbol{\gamma}} = \{\tilde{\gamma}_1, \dots, \tilde{\gamma}_S\}$, where $\bar{\gamma} = \mathcal{R}(\gamma)$ and $\tilde{\gamma} = \mathcal{I}(\gamma)$. The Fisher Information Matrix (FIM) is defined as:

$$\mathcal{F}(\mathbf{u})_{i,j} = \mathbb{E}\left\{\frac{\partial \ln L(\mathbf{y}, c|\mathbf{u})}{\partial \mathbf{u}_i} \frac{\partial \ln L(\mathbf{y}, c|\mathbf{u})}{\partial \mathbf{u}_j}\right\}, \quad (5.14)$$

where $L(\cdot)$ is likelihood function. To simplify our analysis, we only consider SIMO case only ($M_T = 1$). Based on our assumptions from A1) to A4), we use fisher information to calculate the CRB. We start from the single-band case with f at l -th snapshot in (3.1). We rewrite it as:

$$\mathbf{y}_l = \mathbf{A} \boldsymbol{\gamma} c_l + \mathbf{e}_l, \quad (5.15)$$

where $\mathbf{A} = [\mathbf{a}_{M_R}(f, \theta_1), \dots, \mathbf{a}_{M_R}(f, \theta_S)]$ and $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_S]^T$. For $\mathbf{e}_l$, we assume $\mathbb{E}\{\mathbf{e}_l \mathbf{e}_l^H\} = \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{M_R}$ and $\mathbb{E}\{\mathbf{e}_l \mathbf{e}_l^T\} = 0$. σ^2 is the noise power at the frequency f_n and l -th snapshot. As

the noise of every snapshot $\mathbf{e}_l \sim \mathcal{CN}(0, \mathbf{\Sigma})$, the likelihood function of the data is given by:

$$L(\mathbf{y}_1, c_1, \dots, \mathbf{y}_L, c_L | \mathbf{u}) = \frac{e^{-\sum_{l=1}^L \mathbf{e}_l^H \mathbf{\Sigma}^{-1} \mathbf{e}_l}}{\pi^{M_R L} \det(\mathbf{\Sigma})^L}, \quad (5.16)$$

where $\det(\cdot)$ denotes the determinant of a matrix. Then the log-likelihood function is:

$$\begin{aligned} \ln L(\mathbf{y}_1, c_1, \dots, \mathbf{y}_L, c_L | \mathbf{u}) &= \text{const} - M_R L \ln \sigma \\ &\quad - \frac{1}{\sigma^2} \sum_{l=1}^L \mathbf{e}_l^H \mathbf{e}_l. \end{aligned} \quad (5.17)$$

Based on the signal model (5.15), we take a derivative with respect to θ_s :

$$\frac{\partial}{\partial \theta_s} \ln L = \frac{2}{\sigma^2} \sum_{l=1}^L \mathcal{R}(c_l^H \mathbf{d}_s^H \mathbf{e}_l), \quad (5.18)$$

where we use $\omega_m = -j(m-1)2\pi f \frac{d}{v}$ to express $\mathbf{d}_s$ as:

$$\begin{aligned} \mathbf{d}_s &= \frac{\partial \mathbf{a}_{M_R}(f, \theta_s)}{\partial \theta_s} \gamma_s \\ &= \left[0, \omega_2 \cos \theta_s e^{j\omega_2 \sin \theta_s}, \dots, \omega_{M_R} \cos \theta_s e^{j\omega_{M_R} \sin \theta_s} \right]^T \gamma_s. \end{aligned} \quad (5.19)$$

We define $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_S]$. Similarly we have

$$\frac{\partial}{\partial \tilde{\gamma}_s} \ln L = \frac{2}{\sigma^2} \sum_{l=1}^L \mathcal{R}(c_l^H \mathbf{a}_{M_R}(f, \theta_s)^H \mathbf{e}_l), \quad (5.20)$$

$$\frac{\partial}{\partial \tilde{\gamma}_s} \ln L = \frac{2}{\sigma^2} \sum_{l=1}^L \mathcal{I}(c_l^H \mathbf{a}_{M_R}(f, \theta_s)^H \mathbf{e}_l). \quad (5.21)$$

Inspired by [58], we define $\mathbf{\Delta}_l = \mathbf{D} c_l \mathbf{I}_S = [\mathbf{d}_1 c_l, \dots, \mathbf{d}_S c_l]$. $\mathbf{V}_l = \mathbf{A} c_l \mathbf{I}_S = [\mathbf{a}_{M_R}(f, \theta_1) c_l, \dots, \mathbf{a}_{M_R}(f, \theta_S) c_l]$.

Then we introduce the properties [64]. For two complex numbers x and y , we have:

$$\mathcal{R}(x) \mathcal{R}(y^T) = \frac{1}{2} [\mathcal{R}(xy^T) + \mathcal{R}(xy^H)] \quad (5.22)$$

$$\mathcal{I}(x) \mathcal{I}(y^T) = \frac{-1}{2} [\mathcal{R}(xy^T) - \mathcal{R}(xy^H)] \quad (5.23)$$

$$\mathcal{R}(x) \mathcal{I}(y^T) = \frac{1}{2} [\mathcal{I}(xy^T) - \mathcal{I}(xy^H)]. \quad (5.24)$$

Based on our assumption on $\mathbf{e}_l$, $\mathbb{E}\{\mathbf{e}_l \mathbf{e}_l^H\} = \sigma^2 \mathbf{I}$ and $\mathbb{E}\{\mathbf{e}_l \mathbf{e}_l^T\} = 0$. We define $p_s = \sum_{l=1}^L c_l^H c_l$ is the total transmitted signal power at the frequency f_n for all L snapshots. Then we can derivative the following block in FIM:

$$\begin{aligned}
\mathbb{E}\left\{\frac{\partial \ln L}{\partial \theta_i} \frac{\partial \ln L}{\partial \theta_j}\right\} &= \frac{4}{\sigma^4} \sum_{l=1}^L \frac{1}{2} \mathbb{E}\{\mathcal{R}(c_l^H \mathbf{d}_i^H \mathbf{e}_l \mathbf{e}_l^H \mathbf{d}_j c_l)\} \\
&= \frac{2}{\sigma^2} \sum_{l=1}^L \mathcal{R}(\{\Delta_l^H \Delta_l\})_{i,j} \\
&= \frac{2}{\sigma^2} \mathcal{R}\left(\left(\sum_{l=1}^L c_l^H c_l \mathbf{I}_s\right) \mathbf{D}^H \mathbf{D}\right)_{i,j} \\
&= \frac{2p_s}{\sigma^2} \mathcal{R}(\mathbf{D}^H \mathbf{D})_{i,j}.
\end{aligned} \tag{5.25}$$

Similarly,

$$\begin{aligned}
\mathbb{E}\left\{\frac{\partial \ln L}{\partial \tilde{\gamma}_i} \frac{\partial \ln L}{\partial \tilde{\gamma}_j}\right\} &= \frac{4}{\sigma^4} \sum_{l=1}^L \frac{1}{2} \mathbb{E}\{\mathcal{R}(c_l^H \mathbf{a}_{MR}(f, \theta_i)^H \mathbf{e}_l \mathbf{e}_l^H \mathbf{a}_{MR}(f, \theta_j) c_l)\} \\
&= \frac{2}{\sigma^2} \sum_{l=1}^L \mathcal{R}(\{\mathbf{V}_l^H \mathbf{V}_l\})_{i,j} \\
&= \frac{2p_s}{\sigma^2} \mathcal{R}(\mathbf{A}^H \mathbf{A})_{i,j}.
\end{aligned} \tag{5.26}$$

$$\begin{aligned}
\mathbb{E}\left\{\frac{\partial \ln L}{\partial \tilde{\gamma}_i} \frac{\partial \ln L}{\partial \tilde{\gamma}_j}\right\} &= \frac{4}{\sigma^4} \sum_{l=1}^L \frac{1}{2} \mathbb{E}\{\mathcal{R}(c_l^H \mathbf{a}_{MR}(f, \theta_i)^H \mathbf{e}_l \mathbf{e}_l^H \mathbf{a}_{MR}(f, \theta_j) c_l)\} \\
&= \frac{2}{\sigma^2} \sum_{l=1}^L \mathcal{R}(\{\mathbf{V}_l^H \mathbf{V}_l\})_{i,j} \\
&= \frac{2p_s}{\sigma^2} \mathcal{R}(\mathbf{A}^H \mathbf{A})_{i,j}.
\end{aligned} \tag{5.27}$$

$$\begin{aligned}
\mathbb{E}\left\{\frac{\partial \ln L}{\partial \tilde{\gamma}_i} \frac{\partial \ln L}{\partial \tilde{\gamma}_j}\right\} &= \frac{-4}{\sigma^4} \sum_{l=1}^L \frac{1}{2} \mathbb{E}\{\mathcal{I}(c_l^H \mathbf{a}_{MR}(f, \theta_i)^H \mathbf{e}_l \mathbf{e}_l^H \mathbf{a}_{MR}(f, \theta_j) c_l)\} \\
&= \frac{-2}{\sigma^2} \sum_{l=1}^L \mathcal{I}(\{\mathbf{V}_l^H \mathbf{V}_l\})_{i,j} \\
&= \frac{-2p_s}{\sigma^2} \mathcal{I}(\mathbf{A}^H \mathbf{A})_{i,j}.
\end{aligned} \tag{5.28}$$

$$\begin{aligned}
\mathbb{E}\left\{\frac{\partial \ln L}{\partial \tilde{\gamma}_i} \frac{\partial \ln L}{\partial \theta_j}\right\} &= \frac{4}{\sigma^4} \sum_{l=1}^L \frac{1}{2} \mathbb{E}\{\mathcal{R}(c_l^H \mathbf{a}_{M_R}(f, \theta_i)^H \mathbf{e}_l \mathbf{e}_l^H \mathbf{d}_j c_l)\} \\
&= \frac{2}{\sigma^2} \sum_{l=1}^L \mathcal{R}(\{\mathbf{V}_l^H \Delta_l\})_{i,j} \\
&= \frac{2p_s}{\sigma^2} \mathcal{R}(\mathbf{A}^H \mathbf{D})_{i,j}.
\end{aligned} \tag{5.29}$$

$$\begin{aligned}
\mathbb{E}\left\{\frac{\partial \ln L}{\partial \tilde{\gamma}_i} \frac{\partial \ln L}{\partial \theta_j}\right\} &= \frac{-4}{\sigma^4} \sum_{l=1}^L \frac{1}{2} \mathbb{E}\{\mathcal{I}(c_l^H \mathbf{a}_{M_R}(f, \theta_i)^H \mathbf{e}_l \mathbf{e}_l^H \mathbf{d}_j c_l)\} \\
&= \frac{-2}{\sigma^2} \sum_{l=1}^L \mathcal{I}(\{\mathbf{V}_l^H \Delta_l\})_{i,j} \\
&= \frac{-2p_s}{\sigma^2} \mathcal{I}(\mathbf{A}^H \mathbf{D})_{i,j}.
\end{aligned} \tag{5.30}$$

Now the FIM for the frequency f_n is derived as:

$$\begin{aligned}
\mathbf{F}_n(\mathbf{u}) &= \begin{bmatrix} \mathbf{J}_{\theta\theta_n} & \mathbf{J}_{\theta\bar{\Gamma}_n} & \mathbf{J}_{\theta\tilde{\Gamma}_n} \\ \mathbf{J}_{\bar{\Gamma}\theta_n} & \mathbf{J}_{\bar{\Gamma}\bar{\Gamma}_n} & \mathbf{J}_{\bar{\Gamma}\tilde{\Gamma}_n} \\ \mathbf{J}_{\tilde{\Gamma}\theta_n} & \mathbf{J}_{\tilde{\Gamma}\bar{\Gamma}_n} & \mathbf{J}_{\tilde{\Gamma}\tilde{\Gamma}_n} \end{bmatrix} \in \mathbb{C}^{3S \times 3S}. \\
&= p \begin{bmatrix} \mathcal{R}\{\mathbf{D}^H \mathbf{D}\} & \mathcal{R}\{\mathbf{D}^H \mathbf{A}\} & -\mathcal{I}\{\mathbf{D}^H \mathbf{A}\} \\ \mathcal{R}\{\mathbf{A}^H \mathbf{D}\} & \mathcal{R}\{\mathbf{A}^H \mathbf{A}\} & -\mathcal{I}\{\mathbf{A}^H \mathbf{A}\} \\ -\mathcal{I}\{\mathbf{A}^H \mathbf{D}\} & -\mathcal{I}\{\mathbf{A}^H \mathbf{A}\} & \mathcal{R}\{\mathbf{A}^H \mathbf{A}\} \end{bmatrix}
\end{aligned} \tag{5.31}$$

where $p = 2p_s/\sigma^2$. As SNR is defined as $10 \log_{10} p_s/L\sigma^2$, so $p = 2L \times 10^{(SNR/10)}$. Based on the result on single-band CRB analysis, we can now extend it to multi-band CRB analysis. To simplify our derivation, we only show the case about $K = 1$, which means the spectral response is independent with frequencies. As (5.15) is independent in each frequency f_n [65], FIM for multiple frequency is:

$$\mathbf{F}(\mathbf{u}) = \sum_{n=1}^N \mathbf{F}_n(\mathbf{u}). \tag{5.32}$$

Finally the CRB for (θ_s, γ_s) is:

$$CRB(\theta_s) = [\mathbf{F}^{-1}]_{s,s} \quad (5.33)$$

$$CRB(\gamma_s) = [\mathbf{F}^{-1}]_{S+s,S+s} + [\mathbf{F}^{-1}]_{2S+s,2S+s}. \quad (5.34)$$

5.2.5 Numerical tests

Convergence rate comparison:

This test aims to compare proximal gradient method with PDOM. we have found that the RMSE performance of the two algorithms are the same. To gain more insights about the two algorithms, we test their convergence rates. The setup is still SIMO system. $(M_T, M_R) = (1, 20)$, $L = 5$, $N = 1$, $K = 1$, $\boldsymbol{\theta} = \{-20^\circ, 45^\circ, 60^\circ\}$. SNR = 30 dB.

We test the value of objective function in (3.28) in terms of the number of iterations. As shown in Fig.5.1, we can find the convergence rate of PDOM is much faster than proximal gradient method when the penalty parameter $\alpha = 1.0$ in (3.28). PDOM can also converge to the global optima for all of α . All of their RMSE are 2.95×10^{-3} . In contrast, proximal gradient method can converge to the local optima when the value of α is smaller. The RMSE is $(2.95 \times 10^{-3}, 0.97, 1.04)$ when α is $(1.0, 0.5, 0.01)$.

To gain more insights, we also observe the value of two functions f_1 and f_2 in the objective function. They are defined as:

$$f_1 := \frac{1}{2\alpha} \left\| \mathbf{H} - \sum_{n=1}^N \mathcal{L}_n^1(\mathbf{X}_G \mathbf{F}_k) \right\|_F^2 \quad (5.35)$$

$$f_2 := \sum_{n=1}^N \frac{1}{2} \left\| \mathbf{Y}(f_n) - \mathcal{L}_n^2(\mathbf{X}_G \mathbf{F}_k) \mathbf{C}(f_n) \right\|_F^2. \quad (5.36)$$

To eliminate the affect of scaling from the penalty parameter α , we observe the scaled smooth function αf_1 . As shown in Fig.5.2, αf_1 converge to over 1 when $\alpha = 0.01$ in proximal gradient method. It means proximal gradient method converge to the local optima when $\alpha = 0.01$. In contrast, PDOM can converge to lower than 10^{-3} for every α . Fig.5.3 also shows this fact. We

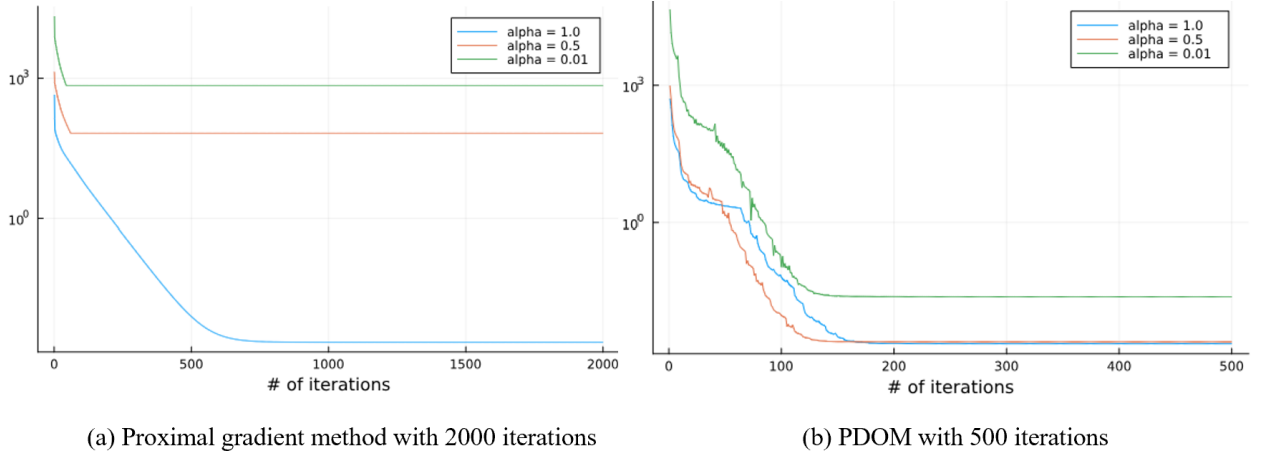


Figure 5.1: The value of objective function versus the number of iterations for various penalty parameter α in (3.28). $(M_T, M_R) = (1, 20)$, $L = 5$, $N = 1$, $K = 1$, $\boldsymbol{\theta} = \{-20^\circ, 45^\circ, 60^\circ\}$. SNR = 30 dB.

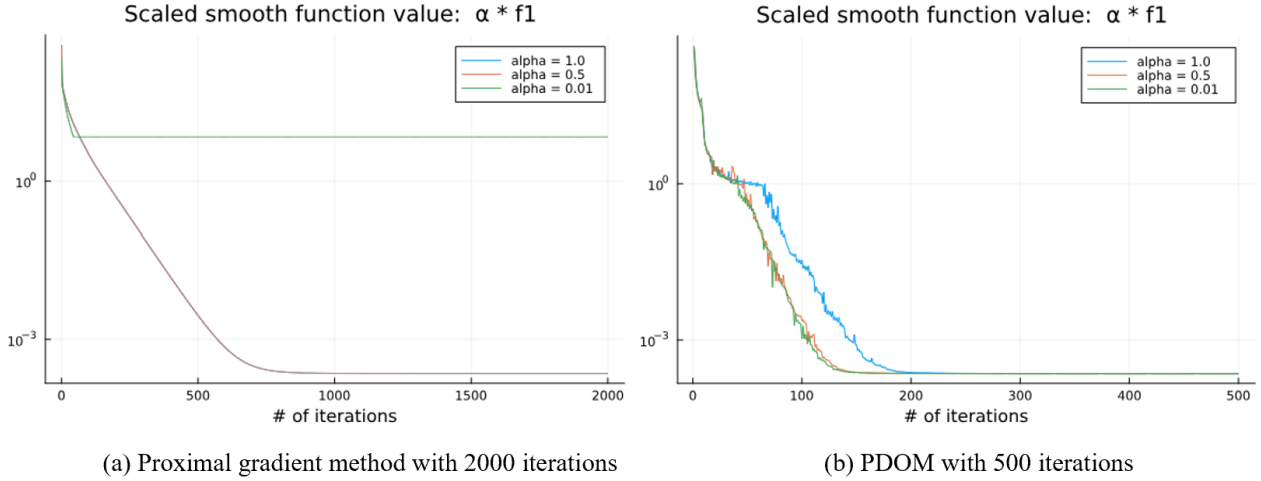


Figure 5.2: The value of scaled f_1 (5.35): αf_1 versus the number of iterations for various penalty parameter α . $(M_T, M_R) = (1, 20)$, $L = 5$, $N = 1$, $K = 1$, $\boldsymbol{\theta} = \{-20^\circ, 45^\circ, 60^\circ\}$. SNR = 30 dB.

can find that it is hard for proximal gradient method to converge to the global optima when the value of α is too small. PDOM can overcome this weakness.

RMSE performance comparison:

In the last test, we find that PDOM and proximal gradient method can converge to the same point after the sufficient number of iterations. Due to this fact, we choose proximal gradient method as our algorithm in this test. We aim to evaluate our formulation by comparing it with benchmarks and CRB. The benchmark is NNM formulation [17], CSSM [21] and TOPS [43]. The ULA has 6 elements with $(M_T, M_R) = (1, 6)$. In (3.28), we set the parameter $\alpha = 1.0$. The number of Monte Carlo trials is 50.

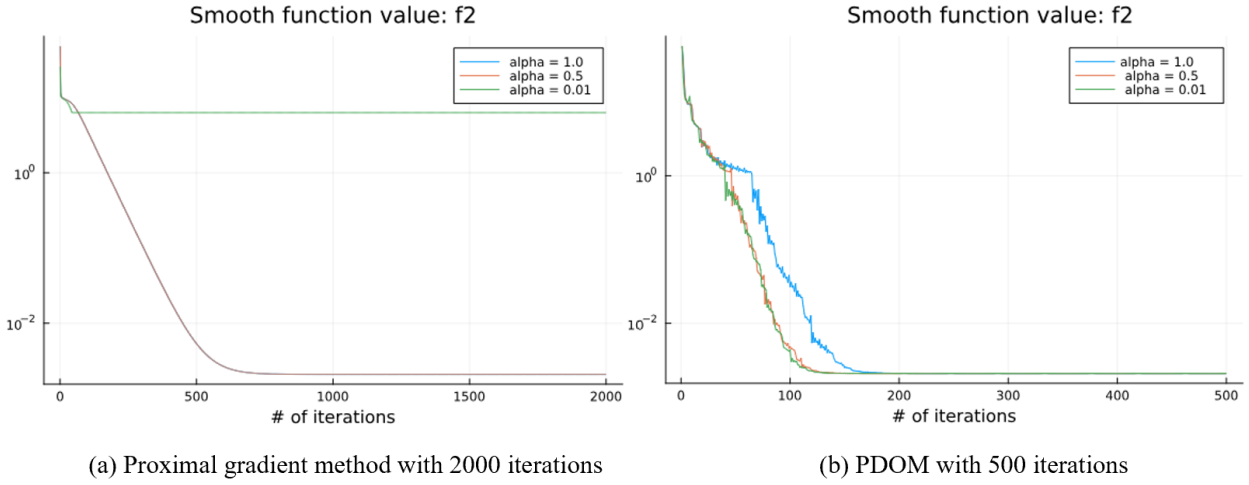


Figure 5.3: The value of f_2 (5.36) versus the number of iterations for various penalty parameter α . $(M_T, M_R) = (1, 20)$, $L = 5$, $N = 1$, $K = 1$, $\theta = \{-20^\circ, 45^\circ, 60^\circ\}$. SNR = 30 dB.

As shown in Fig. 5.4, our proposed nonconvex algorithm RPM performs better than benchmark algorithms. The advantage of our proposed nonconvex algorithm is from our nonconvex formulation. The covariance-based method - CSSM performs unwell, because the limited number of snapshot provide inadequate information for covariance matrix construction. Compared with convex relaxation - NNM, our proposed method RPM have the most advantage at high-SNR case. This is because the global optima is easier to be found in this nonconvex problem. When in low-SNR case, our proposed algorithm performs unstable, as the local optima is easier to be found. For the multi-frequency case ($\mathcal{F} = \{1, 2\}\text{GHz}$), the result is shown in Fig. 5.5. However, there are still gaps between the RMSE of our formulation and CRB. Narrowing this gap is necessary for the future work.

5.3 ADMM in nonconvex optimisation

5.3.1 Introduction

Apart from proximal gradient method and PDOM, ADMM is another popular algorithm. Traditionally, most papers use ADMM to solve convex optimisation problems [27] [28]. In this section, we analyse ADMM for solving nonconvex optimisation problems with linear constraints. [29] discusses the convergence in ADMM in nonconvex and nonsmooth optimisation problems. Followed by this work, we apply ADMM to solve nonconvex optimisation problem. The problem

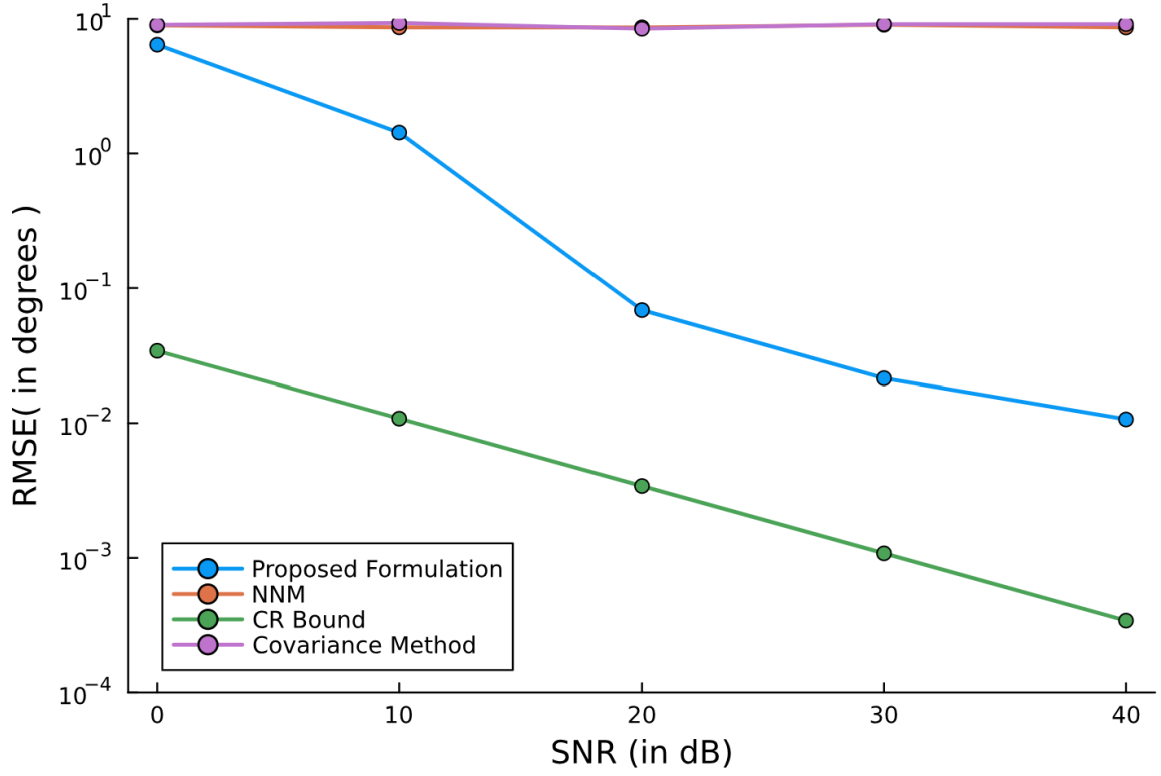


Figure 5.4: RMSE (in degrees) of estimating algorithms versus SNR for the single frequency. $\mathcal{F} = 1\text{GHz}$, $(M_T, M_R, S) = (1, 10, 3)$, $K = 1$, $L = 3$. DoA set $\boldsymbol{\theta} = \{-30^\circ, 5^\circ, 45^\circ\}$.

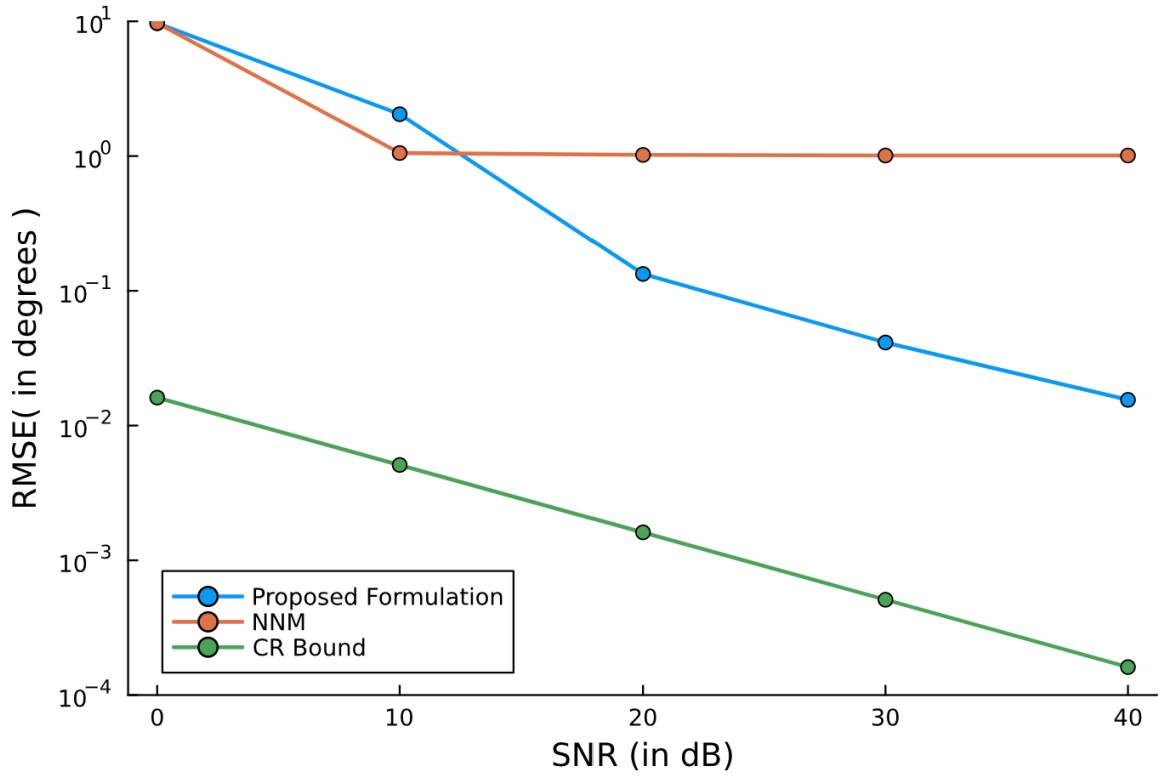


Figure 5.5: RMSE (in degrees) of estimating algorithms versus SNR for the two frequencies. $\mathcal{F} = \{1, 2\}\text{GHz}$, $(M_T, M_R, S) = (1, 5, 2)$, $K = 1$, $L = 10$. DoA set $\boldsymbol{\theta} = \{-20^\circ, 25^\circ\}$.

we choose is Short-and-Sparse Blind Deconvolution.

Blind deconvolution is one of bilinear inverse problems. It aims to separate two unknown signals: $\mathbf{h}$ and $\mathbf{x}$, from the circular convolution of them. The circular convolution is defined as [66]:

$$\mathbf{y}[n] = \mathbf{x}[n] * \mathbf{h}[n] = \sum_{m=0}^{N-1} \mathbf{x}[m] \mathbf{h}[n - m]. \quad (5.37)$$

Most studies formulate the blind deconvolution problem by using convex relaxation [67]. To formulate blind deconvolution, the paper [67] and [68] shows convex relaxation for the formulation, such as nuclear norm minimisation. The paper [69] formulate blind deconvolution in the form of bilinear Lasso problem, and then solve it by using Alternating Descent Method (ADM).

However, it is closer to the problem itself if we formulate it as nonconvex and nonsmooth optimisation, so the nonconvex and nonsmooth optimisation needs to be researched. We formulate the blind deconvolution problem as the nonconvex and nonsmooth optimisation with linear constraint. The problem is in the form:

$$\begin{aligned} & \min_x f(x) \\ & \text{s.t. } Ax = b \end{aligned} \quad (5.38)$$

where the objective function $f(x)$ is nonconvex and nonsmooth [70]. Then we apply the ADMM to solve it.

5.3.2 Short-and-sparse blind deconvolution formulation

Compared with blind deconvolution, the short-and-sparse blind deconvolution problem is under the following extra conditions: $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ is long and sparse signal. $\mathbf{h} \in \mathbb{R}^M$, ($M \ll N$) is short and normalised kernel [69]. One of formulations for blind deconvolution is low rank recovery

problem. For this formulation, we define matrix $\mathbf{X}$ and its relationship with observed signal $\mathbf{y}$:

$$\mathbf{X} := \mathbf{x}\mathbf{h}^T, \quad (5.39)$$

$$\mathbf{y} = \mathcal{A}(\mathbf{X}), \quad (5.40)$$

where matrix $\mathbf{X}$ is rank one matrix and $\mathcal{A}(\cdot)$ is linear operator corresponding to circular convolution. More details can be referred in the section 3 of [66]. To promote the sparsity of sparse signal $\mathbf{x}$, the blind deconvolution can be formulated as:

$$\min_{\mathbf{X}} \|\mathbf{X}\|_{r,2,0} \quad (5.41)$$

$$\text{s.t. } \mathbf{y} = \mathcal{A}(\mathbf{X}), \mathbf{X} \in \mathcal{R}1$$

where $\|\cdot\|_{r,2,0}$ calculates the row sparsity for the input matrix. It counts the sparsity of $\downarrow_2$ norm for each row. It also means counts the number of non-all-zero rows. To solve the problem (5.41) by projecting the matrix $\mathbf{X}$ to rank one matrix, it can be reformulated as:

$$\min_{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3} \|\mathbf{X}_1\|_{r,2,0} + \delta(\text{rank}(\mathbf{X}_2) \leq 1) + \frac{\lambda}{2} \|\mathbf{y} - \mathcal{A}(\mathbf{X}_3)\|_2^2 \quad (5.42)$$

$$\text{s.t. } \mathbf{X}_1 = \mathbf{X}_3, \mathbf{X}_2 = \mathbf{X}_3$$

where $\delta(\cdot)$ is the indicator function, which means the rank of matrix $\mathbf{X}_2$ is at most one.

5.3.3 Problem solving

To solve the problem (5.42) by using ADMM method, the corresponding augmented Lagrangian is:

$$\begin{aligned} \mathcal{L}_\rho(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2) = & \|\mathbf{X}_1\|_{r,2,0} + \frac{\rho}{2} \|\mathbf{X}_3 - \mathbf{X}_1 + \mathbf{\Lambda}_1\|_F^2 - \frac{\rho}{2} \|\mathbf{\Lambda}_1\|_F^2 \\ & + \delta(\text{rank}(\mathbf{X}_2) \leq 1) + \frac{\rho}{2} \|\mathbf{X}_3 - \mathbf{X}_2 + \mathbf{\Lambda}_2\|_F^2 - \frac{\rho}{2} \|\mathbf{\Lambda}_2\|_F^2 + \frac{\lambda}{2} \|\mathbf{y} - \mathcal{A}(\mathbf{X}_3)\|_2^2. \end{aligned} \quad (5.43)$$

where $\rho > 0$, then the ADMM iterations are given by:

$$\mathbf{X}_1^{k+1} = \arg \min_{\mathbf{X}_1} \|\mathbf{X}_1\|_{r,2,0} + \frac{\rho}{2} \|\mathbf{X}_3^k - \mathbf{X}_1 + \mathbf{\Lambda}_1^k\|_F^2 \quad (5.44)$$

$$\mathbf{X}_2^{k+1} = \arg \min_{\mathbf{X}_2} \delta(\text{rank}(\mathbf{X}_2) \leq 1) + \frac{\rho}{2} \|\mathbf{X}_3^k - \mathbf{X}_2 + \mathbf{\Lambda}_2^k\|_F^2 \quad (5.45)$$

$$\mathbf{X}_3^{k+1} = \arg \min_{\mathbf{X}_3} \frac{\lambda}{2} \|\mathbf{y} - \mathcal{A}(\mathbf{X}_3)\|_2^2 + \frac{\rho}{2} \|\mathbf{X}_3 - \mathbf{X}_1^{k+1} + \mathbf{\Lambda}_1^k\|_F^2 + \frac{\rho}{2} \|\mathbf{X}_3 - \mathbf{X}_2^{k+1} + \mathbf{\Lambda}_2^k\|_F^2 \quad (5.46)$$

$$\mathbf{\Lambda}_1^{k+1} = \mathbf{\Lambda}_1^k + \mathbf{X}_3^{k+1} - \mathbf{X}_1^{k+1} \quad (5.47)$$

$$\mathbf{\Lambda}_2^{k+1} = \mathbf{\Lambda}_2^k + \mathbf{X}_3^{k+1} - \mathbf{X}_2^{k+1} \quad (5.48)$$

The closed form solution for $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ are the following:

$$\mathbf{X}_{1,(:,i)}^{k+1} = \text{sign}(\|\mathbf{C}_{:,i}^k\|_F^2 - \frac{2}{\rho}) \mathbf{C}_{:,i}^k \quad (5.49)$$

$$\mathbf{X}_2^{k+1} = \boldsymbol{\sigma}_1 \mathbf{u}_1 \mathbf{v}_1^T \quad (5.50)$$

$$\mathbf{X}_3^{k+1} = \frac{\lambda}{2\rho} S_{-CL} \left[\frac{\rho}{\lambda} S_{CL}(\mathbf{C}_1^k + \mathbf{C}_2^k) + \mathbf{C}_3 \right] \left(\mathbf{I}_n - \frac{1}{n + \frac{2\rho}{\lambda}} \mathbf{1}_{n \times n} \right) \quad (5.51)$$

where $\mathbf{C}^k = \mathbf{X}_3^k + \mathbf{\Lambda}_1^k$, $\mathbf{X}_3^k + \mathbf{\Lambda}_2^k = \sum_{i=1}^n \boldsymbol{\sigma}_i \mathbf{u}_i \mathbf{v}_i^T$. $\mathbf{C}_1^k = \mathbf{X}_1^{k+1} - \mathbf{\Lambda}_1^k$, $\mathbf{C}_2^k = \mathbf{X}_2^{k+1} - \mathbf{\Lambda}_2^k$, $\mathbf{C}_3 = \mathbf{y} \mathbf{1}_{1 \times n}$. The $\text{sign}(\cdot)$ function is defined as:

$$\text{sign}(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (5.52)$$

The linear operator $S_{CL}(\cdot)$ is defined as:

$$S_{CL}(X) = [X_{:,1}, S_1(X_{:,2}), \dots, S_{i-1}(X_{:,i}), \dots]. \quad (5.53)$$

The linear operator $S_i(\cdot)$ is to move the input vector circularly downward by i positions.

5.3.4 Numerical Test: image deblurring

This section demonstrates the numerical test results of short-and-sparse deconvolution. We compare ADMM and ADM [69]. For real applications for short-and-sparse deconvolution,

	ADM	ADMM ($\rho = 90, \lambda = 1$)
PSNR for digit '0' (dB)	14.7688	25.6404
PSNR for digit '3' (dB)	17.9029	29.5361
PSNR for digit '5' (dB)	16.3345	32.4049
PSNR for digit '9' (dB)	17.3260	26.1316
Average running time (second)	16.534	0.818

Table 5.1: PSNR and average running time for image deblurring

I made tests regarding to image deblurring. The data set we chose is MNIST database [71]. MNIST (Modified National Institute of Standards and Technology) database is a large database that contains handwritten digits. To choose ground truth images, 4 images are randomly chosen from test set images of this database. To make blurry images, the kernel $h_0 \in \mathbb{R}^9$ is used to convolute with vectorised form of the ground truth images. The kernel generation is the same as synthetic data tests. The penalty parameter $\rho = 90$. $\lambda = 1$. The performance of image deblurring is measured by average running time and Peak Signal-to-Noise Ratio (PSNR), which is defined as:

$$\text{PSNR} = \max[10 \log_{10} \frac{N}{\|s_{\tau}(\hat{I}) - I_0\|_F^2} : \tau \in [0, \sqrt{N} - 1]], \quad (5.54)$$

where N is 28×28 . I_0 is ground truth images. $\hat{I}$ is the output of algorithms. The maximum PSNR is selected from 50 trails. The results are shown in the table 5.1. The result show that ADMM has higher PSNR than ADM, and the avarage running time is much less than ADM.

5.4 Summary

In this chapter, we have analysed proximal gradient method, PDOM and ADMM in nonconvex and nonsmooth optimisation problem. PDOM has a faster convergence rate $\mathcal{O}(1/k^2)$ at k -th iteration. We choose the formulation from the chapter 3 to show these two algorithms. Both of them can convergence to the same point after the sufficient number of iterations, but PDOM reach the point faster. Compared with benchmarks, our nonconvex and nonsmooth optimisation formulation has a lower RMSE performance. However, there are still gaps between the RMSE of our formulation and CRB. Narrowing this gap is still in the progress.

For ADMM, we use ADMM to solve the blind deconvolution problem. Compared with ADM, we show that ADMM have a better deblurring ability with less time. Apart from proximal gradient method and PDOM, the potential of ADMM can be a future research direction in the field of nonconvex and nonsmooth optimisation problem.

Chapter 6

Conclusion

6.1 Summary of Thesis Achievements

As the title suggests, this thesis discusses the gridless processing for multi-band MIMO sensing. We firstly discuss the multi-band active array in the chapter 3. This multi-band array can sense the DoAs and spectral responses of targets. As a special case of multi-band active array, we discuss the FDA in the chapter 4. This multi-band array can sense the DoAs and ranges of targets. We finally analyse nonconvex algorithms: proximal gradient method, PDOM and ADMM in the chapter 5. The detailed achievements of each chapter are summarised in the following.

In the chapter 3, we propose a new multi-band active array to sense DoAs and spectral responses. We make spectral response of targets can be different at different frequency bands. The degree of freedom can be increased in multi-band. To take advantage of multiple frequency bands efficiently, we have developed the GCD-based joint processing to construct the corresponding frequency grid. We use GCD to fit all frequency bands in one vector. The RPM is formulated to numerically recover the Hankel matrix for high resolution. With a limited number of snapshots, our simulations shows the GCD-based joint processing can estimate more accurately than covariance-based method. Both DoAs and spectral response can be sensed with high accuracy. The number of targets can be even more than the antennas in ULA.

In the chapter 4, we use range-angle decoupling model in FDA, and formulate the problem as the low rank Hankel matrix recovery problem. To recover DoAs and ranges, traditional methods calculate the covariance matrix of received signals, and search the location of peaks in 2D (DoA and range) domain by using 2D-MUSIC. The computational complexity is $\mathcal{O}(N^2)$. Our method apply 1D-MUSIC to DoA and range domain independently, so our computational complexity can be reduced to $\mathcal{O}(M + N)$. With a limited number of snapshots, we can sense both DoAs and ranges with high accuracy.

In the chapter 5, we discuss the algorithms for nonconvex and nonsmooth optimisation problems: proximal gradient method, PDOM and ADMM. In the last two chapters, we use proximal gradient method to solve the formulation. Proximal gradient method has the convergence rate $\mathcal{O}(1/k)$. PDOM have a faster convergence rate than proximal gradient method $\mathcal{O}(1/k^2)$. We apply PDOM to solve our formulation in the chapter 3. Compared with proximal gradient method, we find PDOM can converge to the same point with less number of iterations, which verify the existence of fast convergence. For ADMM, ADMM is always used to solve convex optimisation problems with linear constraints. Followed by [29], we use ADMM to solve the image deblurring problem. It means that ADMM can solve the nonconvex problem.

6.2 Applications

The thesis achievements can be used in the field of MIMO sensing, such as autonomous driving and medical imaging. For both our proposed multi-band active array and FDA, only the limited number of snapshots are required. This benefit can be used for the real MIMO systems that cannot sample sufficient snapshots, or need to sense the targets in a short time.

The chapter 3 should be seen as an introduction of the array which can sense both DoAs and spectral responses. In this chapter, we propose our spectral response model, which is closer to the real environment. In real applications, the sensed targets can be sketched with different colors. Apart from the target localisation, the colors can help people to identify different targets. Considering the effect of width of array, our new criteria of wideband can be applied to other multi-band arrays. In the chapter 5, we have shown that both PDOM and ADMM

can be applied to the nonconvex and nonsmooth optimisation problem. More nonconvex and nonsmooth formulation can consider using PDOM and ADMM.

6.3 Future Work

For our multi-band active array, the design of frequency set is attractive, including the distribution of each frequency bands. As it is an important factor of the maximum resolvable targets in our array. Apart from that, in reality, both number of targets S and rank of spectral response model K cannot be known before our sensing progress. The more general system model is to be proposed. For our estimation method, there are still gaps between our proposed formulation and CRB. Narrowing the gaps is necessary in the future. Apart from that, the multi-band active array for the near-field scenario is an interesting topic for the future work.

For FDA, the wideband FDA is an interesting topic for the future work. In the chapter 4, we only discuss the narrowband case. As we discussed in the chapter 3, the wider bandwidth can increase the number of antennas in the virtual array. It means the maximum of resolvable targets can be improved in wideband case. It is possible that the number of targets can be more than antennas in ULA. Although we achieve the fast sensing in FDA, the estimation accuracy is to be improved.

Bibliography

- [1] NASA, “Nasa applied remote sensing training program,” tech. rep., <https://www.earthdata.nasa.gov/learn/backgrounders/active-sensors>.
- [2] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, “Joint radar and communication design: Applications, state-of-the-art, and the road ahead,” *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3834–3862, 2020.
- [3] 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.
- [4] L. Wan, Y. Sun, L. Sun, Z. Ning, and J. J. P. C. Rodrigues, “Deep learning based autonomous vehicle super resolution DOA estimation for safety driving,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4301–4315, 2021.
- [5] Z. Yang, J. Li, P. Stoica, and L. Xie, “Sparse methods for direction-of-arrival estimation,” *arXiv preprint arXiv:1609.09596*, 2016.
- [6] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
- [7] R. Roy and T. Kailath, “Esprit-estimation of signal parameters via rotational invariance techniques,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.
- [8] D. Manolakis and G. Shaw, “Detection algorithms for hyperspectral imaging applications,” *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, 2002.

- [9] S.Lee and G.Meister, “Modis aqua optical throughput degradation impact on relative spectral response and calibration of ocean color products,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5214–5219, 2017.
- [10] P. K. Mogensen and A. N. Riseth, “Optim: A mathematical optimization package for Julia,” *Journal of Open Source Software*, vol. 3, no. 24, p. 615, 2018.
- [11] D. Bliss and K. Forsythe, “Multiple-input multiple-output (MIMO) radar and imaging: degrees of freedom and resolution,” in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 1, pp. 54–59 Vol.1, 2003.
- [12] A. Ben and M. Ghavami, *Adaptive array systems: fundamentals and applications*. John Wiley & Sons, 2005.
- [13] F. C. Commission, *Ultra-Wideband Transmission Systems*. Federal Register, 2002.
- [14] B. Huang, P. Li, J. Jian, W.-Q. Wang, and A. Basit, “Fda-mimo radar target detection with limited test samples,” *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–17, 2024.
- [15] P. Antonik, M. Wicks, H. Griffiths, and C. Baker, “Frequency diverse array radars,” in *2006 IEEE Conference on Radar*, pp. 3 pp.–, 2006.
- [16] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, “Sparse sampling of signal innovations,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, 2008.
- [17] W. Xu, J. Yi, S. Dasgupta, J. F. Cai, M. Jacob, and M. Cho, “Separation-free super-resolution from compressed measurements is possible: an orthonormal atomic norm minimization approach,” in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 76–80, 2018.
- [18] S.Sun and A. Petropulu, “A sparse linear array approach in automotive radars using matrix completion,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8614–8618, 2020.

- [19] Y. Wu, M. B. Wakin, and P. Gerstoft, “Non-uniform array and frequency spacing for regularization-free gridless doa,” *IEEE Transactions on Signal Processing*, vol. 72, pp. 2006–2020, 2024.
- [20] Y. Bresler and A. Macovski, “On the number of signals resolvable by a uniform linear array,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 6, pp. 1361–1375, 1986.
- [21] H. Wang and M. Kaveh, “Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 823–831, 1985.
- [22] S. Rao, “MIMO radar,” *Texas Instruments*, 2018.
- [23] P. Pal and P. P. Vaidyanathan, “Nested arrays: A novel approach to array processing with enhanced degrees of freedom,” *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, 2010.
- [24] P. P. Vaidyanathan and P. Pal, “Sparse sensing with co-prime samplers and arrays,” *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 573–586, 2011.
- [25] A. Manikas and C. Proukakis, “Modeling and estimation of ambiguities in linear arrays,” *IEEE Transactions on Signal Processing*, vol. 46, no. 8, pp. 2166–2179, 1998.
- [26] X. Yao and W. Dai, “A preconditioned hessian proximal algorithm for spectral compressed sensing,” in *2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 476–480, 2023.
- [27] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “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.
- [28] D. Hajinezhad, T.-H. Chang, X. Wang, Q. Shi, and M. Hong, “Nonnegative matrix factorization using admm: Algorithm and convergence analysis,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4742–4746, 2016.

- [29] Y. Wang, W. Yin, and J. Zeng, “Global convergence of admm in nonconvex nonsmooth optimization,” *Journal of Scientific Computing*, vol. 78, pp. 29–63, 2019.
- [30] Y. Sun, P. Babu, and D. P. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2017.
- [31] A. Breloy, S. Kumar, Y. Sun, and D. P. Palomar, “Majorization-minimization on the stiefel manifold with application to robust sparse pca,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 1507–1520, 2021.
- [32] W. Dai and S. Vlaski, *EE4.66 Topics in Large Dimensional Data Processing*. Imperial College London, 2021.
- [33] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer New York, NY, 2006.
- [34] N. Parikh and S. Boyd, “Proximal algorithms,” *Foundations and trends (R) in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [35] Y. Zhou and W. Dai, “Proximal dogleg opportunistic majorization for nonconvex and nonsmooth optimization,” 2024.
- [36] M. Wax, T. Shan, and T. Kailath, “Spatio-temporal spectral analysis by eigenstructure methods,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 4, pp. 817–827, 1984.
- [37] Z. Tang, G. Blacchiere, and G. Leus, “Aliasing-free wideband beamforming using sparse signal representation,” *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3464–3469, 2011.
- [38] K. L. Gemba, S. Nannuru, P. Gerstoft, and W. S. Hodgkiss, “Multi-frequency sparse bayesian learning for robust matched field processing,” *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3411–3420, 2017.
- [39] N. Antonello, E. D. Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, “Joint acoustic localization and dereverberation through plane wave decomposition and sparse

- regularization,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 1893–1905, 2019.
- [40] S. Nannuru, K. L. Gemba, and P. Gerstoft, “Sparse bayesian learning with multiple dictionaries,” in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1190–1194, IEEE, 2017.
- [41] K. L. Gemba, S. Nannuru, and P. Gerstoft, “Robust ocean acoustic localization with sparse bayesian learning,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 49–60, 2019.
- [42] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, “An efficient maximum likelihood method for direction-of-arrival estimation via sparse bayesian learning,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 1–11, 2012.
- [43] Y.-S. Yoon, L. M. Kaplan, and J. H. McClellan, “TOPS: new DOA estimator for wideband signals,” *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 1977–1989, 2006.
- [44] F. Raimondi, P. Comon, and O. Michel, “Wideband multilinear array processing through tensor decomposition,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2951–2955, IEEE, 2016.
- [45] J. Krolik and D. Swingler, “Focused wide-band array processing by spatial resampling,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 2, pp. 356–360, 1990.
- [46] D. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [47] Y. Wu, M. B. Wakin, and P. Gerstoft, “Gridless DOA estimation with multiple frequencies,” *IEEE Transactions on Signal Processing*, vol. 71, pp. 417–432, 2023.
- [48] W.-K. Ma, T.-H. Hsieh, and C.-Y. Chi, “DOA estimation of quasi-stationary signals with less sensors than sources and unknown spatial noise covariance: A khatri-rao subspace approach,” *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2168–2180, 2010.

- [49] F. Wang, Z. Tian, G. Leus, and J. Fang, "Direction of arrival estimation of wideband sources using sparse linear arrays," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4444–4457, 2021.
- [50] N. Mao and W. Dai, "Multi-band active arrays for gridless doa and spectral response sensing," in *2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2024 (accepted).
- [51] N. Mao and W. Dai, "A multi-band active array for direction-of-arrival estimation," in *2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 341–345, 2023.
- [52] G. K. Carvajal *et al.*, "Comparison of automotive FMCW and OFDM radar under interference," in *2020 IEEE Radar Conference (RadarConf20)*, pp. 1–6, 2020.
- [53] B. W. et al, "Spatial-wideband effect in massive mimo with application in mmwave systems," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 134–141, 2018.
- [54] J. Sanson, A. Gameiro, D. Castanheira, and P. P. Monteiro, "Comparison of DoA algorithms for MIMO OFDM radar," in *2018 15th European Radar Conference (EuRAD)*, pp. 226–229, 2018.
- [55] P. Kulkarni and P. P. Vaidyanathan, "Non-integer arrays for array signal processing," *IEEE Transactions on Signal Processing*, vol. 70, pp. 5457–5472, 2022.
- [56] A. R. et al, "1 - introduction to RFID and chipless RFID," in *RFID and Wireless Sensors Using Ultra-Wideband Technology* (A. R. et al, ed.), pp. 1–18, Elsevier, 2016.
- [57] P. Stoica and R. Moses, *Spectral analysis of signals*, vol. 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.
- [58] J. Xiong, W.-Q. Wang, and K. Gao, "Fda-mimo radar range-angle estimation: Crlb, mse, and resolution analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 1, pp. 284–294, 2018.

- [59] M. Henninger, S. Mandelli, M. Arnold, and S. Ten Brink, “A computationally efficient 2d music approach for 5g and 6g sensing networks,” in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 210–215, 2022.
- [60] J. Xu, G. Liao, S. Zhu, L. Huang, and H. C. So, “Joint range and angle estimation using mimo radar with frequency diverse array,” *IEEE Transactions on Signal Processing*, vol. 63, no. 13, pp. 3396–3410, 2015.
- [61] W.-G. Tang, H. Jiang, and Q. Zhang, “Range-angle decoupling and estimation for fda-mimo radar via atomic norm minimization and accelerated proximal gradient,” *IEEE Signal Processing Letters*, vol. 27, pp. 366–370, 2020.
- [62] Z. Yang, L. Xie, and P. Stoica, “Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution,” *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3685–3701, 2016.
- [63] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall PTR, 1993.
- [64] P. Stoica and A. Nehorai, “Music, maximum likelihood, and cramer-rao bound,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, 1989.
- [65] Y. Liang, W. Liu, Q. Shen, W. Cui, and S. Wu, “A review of closed-form cramer-rao bounds for doa estimation in the presence of gaussian noise under a unified framework,” *IEEE Access*, vol. 8, pp. 175101–175124, 2020.
- [66] C. Cheng and W. Dai, “Short-and-sparse deconvolution via rank-one constrained optimization (roco),” 2021.
- [67] A. Ahmed, B. Recht, and J. Romberg, “Blind deconvolution using convex programming,” 2012.
- [68] M. A. Davenport and J. Romberg, “An overview of low-rank matrix recovery from incomplete observations,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 608–622, 6 2016.

- [69] Y. Lau, Q. Qu, H.-W. Kuo, P. Zhou, Y. Zhang, and J. Wright, “Short-and-sparse deconvolution – a geometric approach,” 2019.
- [70] J. Zeng, W. Yin, and D.-X. Zhou, “Moreau envelope augmented lagrangian method for nonconvex optimization with linear constraints,” 2021.
- [71] Y. LeCun, C. Cortes, and C. J. Burges, *The MNIST Database of handwritten digits*.

Appendices

A. Gradient calculation for MDA formulation in (3.28)

To use the proximal gradient method and PDOM [35], we need the gradient computation for differentiable function f in (3.28). We show the gradient $(\nabla_{\mathbf{X}_G} f, \nabla_{\mathbf{H}} f)$ calculation as:

$$\begin{aligned} \nabla_{\mathbf{X}_G} f = & -\frac{1}{\alpha} \sum_{n=1}^N \mathcal{L}_n^{1,*}(\mathbf{H} - \sum_{n=1}^N \mathcal{L}_n^1(\mathbf{X}_G \mathbf{F}_k)) \mathbf{F}_k^H, \\ & - \mathcal{L}_n^{2,*}((\mathbf{Y}(f_n) - \mathcal{L}_n^2(\mathbf{X}_G \mathbf{F}_k)) \mathbf{C}(f_n)) \mathbf{C}(f_n)^H) \mathbf{F}_k^H \end{aligned} \quad (6.1)$$

$$\nabla_{\mathbf{H}} f = \frac{1}{\alpha} (\mathbf{H} - \sum_{n=1}^N \mathcal{L}_n^1(\mathbf{X}_G \mathbf{F}_k)). \quad (6.2)$$

The linear operator $\mathcal{L}_n^{1,*}(\cdot) = \mathcal{B}_n^* \mathcal{H}_{h_1, h_2}^* \mathcal{W}_n^*(\cdot)$ and $\mathcal{L}_n^{2,*}(\cdot) = \mathcal{B}_n^* \mathcal{A}_n^* \mathcal{H}_{M_R, M_T}^*(\cdot)$ are the adjoint of the linear operator $\mathcal{L}_n^1(\cdot)$ and $\mathcal{L}_n^2(\cdot)$ respectively. The $\mathcal{B}_n^*(\cdot)$, $\mathcal{H}^*(\cdot)$, $\mathcal{W}_n^*(\cdot)$ and $\mathcal{A}_n^*(\cdot)$ are the adjoint of its corresponding linear operator $\mathcal{B}_n(\cdot)$, $\mathcal{H}(\cdot)$, $\mathcal{W}_n(\cdot)$ and $\mathcal{A}_n(\cdot)$.

B. Hessian matrix for MDA formulation in (3.28)

We firstly define the Hessian matrix $\mathbf{Q}$ in (3.28) as:

$$\mathbf{Q} = \begin{bmatrix} \frac{\partial^2}{\partial \mathbf{X}_G^2} f & \frac{\partial^2}{\partial \mathbf{X}_G \partial \mathbf{H}} f \\ \frac{\partial^2}{\partial \mathbf{H} \partial \mathbf{X}_G} f & \frac{\partial^2}{\partial \mathbf{H}^2} f \end{bmatrix}. \quad (6.3)$$

However, as the two variables ($\mathbf{X}_G, \mathbf{H}$) are matrix, we should vectorise them to make $\mathbf{Q}$ well-defined. Now it is:

$$\mathbf{Q} = \begin{bmatrix} \frac{\partial^2}{\partial \text{vec}(\mathbf{X}_G)^2} f & \frac{\partial^2}{\partial \text{vec}(\mathbf{X}_G) \partial \text{vec}(\mathbf{H})} f \\ \frac{\partial^2}{\partial \text{vec}(\mathbf{H}) \partial \text{vec}(\mathbf{X}_G)} f & \frac{\partial^2}{\partial \text{vec}(\mathbf{H})^2} f \end{bmatrix} \quad (6.4)$$

$$\in \mathbb{R}^{(h_1 h_2 N + M_g) \times (h_1 h_2 N + M_g)}.$$

To derive $\mathbf{Q}$, we need to rewrite (6.1) in matrix form. It is

$$\nabla_{\mathbf{X}_G} f = \frac{1}{\alpha} \mathbf{D} \mathbf{X}_G \mathbf{F}_K \mathbf{F}_K^H + \sum_{n=1}^N \mathbf{L}_n \mathbf{X}_G \mathbf{D}_n. \quad (6.5)$$

We rewrite (6.2) in matrix form as:

$$\nabla_{\mathbf{H}} f = \frac{1}{\alpha} (\mathbf{H} - \sum_{n_1=1}^N \sum_{n_2=1}^N \mathbf{L}_{2,n_2} \mathbf{X}_G \mathbf{F}_K \mathbf{R}_{2,n_1,n_2}). \quad (6.6)$$

In (6.5), $\mathbf{D}$ is a diagonal matrix as $\text{diag}(\mathbf{w})$, where $\mathbf{w} = [1, 2, \dots, q, \dots, q, q-1, \dots, 2, 1] \in \mathbb{R}^{M_g}$. $q \approx (M_g + 1)/2$. There is only one q if M_g is odd number, and two if M_g is even number. For each n , $\mathbf{D}_n = \mathbf{F}_K \mathbf{R}_n \mathbf{F}_K^T$. $\mathbf{R}_n \in \mathbb{R}^{N \times N}$ is all zero matrix, excepting $\mathbf{R}_n(n, n) = 1$. $\mathbf{L}_n = \text{diag}(\mathcal{I}_{n, M_l}) \in \mathbb{R}^{M_R \times M_R}$ in (3.19).

In (6.6), $\mathbf{L}_{2,n} = [\mathbf{0}, \mathbf{I}, \mathbf{0}] \in \mathbb{R}^{h_1 \times M_g}$. The identity matrix $\mathbf{I} \in \mathbb{R}^{h_1 \times h_1}$ is from the column index $n+1$ of $\mathbf{L}_{2,n}$. $\mathbf{R}_{2,n_1,n_2} \in \mathbb{R}^{N \times h_2 N}$ is all zero matrix, excepting $\mathbf{R}_{2,n_1,n_2}(n_1, (n_1 - 1)h_2 + n_2) = 1$.

According to the property of vectorisation: $\text{vec}(\mathbf{A}\mathbf{X}\mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{X})$, the closed-form solution of the second order derivative of f are the following:

$$\frac{\partial^2}{\partial \text{vec}(\mathbf{H})^2} f = \frac{1}{\alpha} \mathbf{I}_{h_1 h_2 N}, \quad (6.7)$$

$$\frac{\partial^2}{\partial \text{vec}(\mathbf{X}_G)^2} f = \frac{N}{\alpha} \otimes \mathbf{D} + \sum_{n=1}^N \mathbf{D}_n \otimes \mathbf{L}_n, \quad (6.8)$$

$$\frac{\partial^2}{\partial \text{vec}(\mathbf{X}_G) \partial \text{vec}(\mathbf{H})} f = -\frac{1}{\alpha} \sum_{n_1=1}^N \sum_{n_2=1}^N (\mathbf{F}_K \mathbf{R}_{2,n_1,n_2})^T \otimes \mathbf{L}_{2,n_2}. \quad (6.9)$$

C. PDOM details for FDA formulation in (4.22)

To calculate the positive semi-definite Hessian matrix $\mathbf{Q}$ in (eq.11, [35]), we derive $\mathbf{Q}$ including second-order derivative for each vectorised variables: $[\text{vec}(\mathbf{H}_\theta), \text{vec}(\mathbf{H}), \mathbf{x}]^T$. We assume $M_T = M_R = M$. The Hessian matrix $\mathbf{Q}$ in (4.22) is:

$$\mathbf{Q} = \begin{bmatrix} \frac{\partial^2}{\partial \text{vec}(\mathbf{H}_\theta)^2} f & \frac{\partial^2}{\partial \text{vec}(\mathbf{H}_\theta) \partial \text{vec}(\mathbf{H})} f & \frac{\partial^2}{\partial \text{vec}(\mathbf{H}_\theta) \partial \mathbf{x}} f \\ \frac{\partial^2}{\partial \text{vec}(\mathbf{H}) \partial \text{vec}(\mathbf{H}_\theta)} f & \frac{\partial^2}{\partial \text{vec}(\mathbf{H})^2} f & \frac{\partial^2}{\partial \text{vec}(\mathbf{H}) \partial \mathbf{x}} f \\ \frac{\partial^2}{\partial \mathbf{x} \partial \text{vec}(\mathbf{H}_\theta)} f & \frac{\partial^2}{\partial \mathbf{x} \partial \text{vec}(\mathbf{H})} f & \frac{\partial^2}{\partial \mathbf{x}^2} f \end{bmatrix} \quad (6.10)$$

$$\in \mathbb{R}^{(h_1+h_2+h_3) \times (h_1+h_2+h_3)},$$

where $h_1 = M^2$, $h_3 = (2M - 1)M$, $h_2 = (\frac{h_3+1}{2})^2$. From the first-order derivative (4.25), (4.26) and (4.27), we derive the elements in $\mathbf{Q}$ in the following:

$$\frac{\partial^2}{\partial \text{vec}(\mathbf{H}_\theta)^2} f = \frac{1}{\alpha_1} \mathbf{I}_{h_1}, \quad (6.11)$$

$$\frac{\partial^2}{\partial \text{vec}(\mathbf{H})^2} f = \frac{1}{\alpha_2} \mathbf{I}_{h_2}, \quad (6.12)$$

$$\frac{\partial^2}{\partial \text{vec}(\mathbf{H}) \partial \text{vec}(\mathbf{H}_\theta)} f = \mathbf{0}_{h_2 \times h_1} \quad (6.13)$$

To derive the rest of elements, we need to rewrite the linear operators in first-order derivative as the matrix form. The result are:

$$\begin{aligned} \frac{\partial^2}{\partial \mathbf{x}^2} f &= \frac{1}{\alpha_1} \mathcal{A}_1^* \mathcal{H}^* \mathcal{H} \mathcal{A}_1(\cdot) + \frac{1}{\alpha_2} \mathcal{H}^* \mathcal{H}(\cdot) - \mathbf{P}^T \mathbf{P} \\ &= \frac{1}{\alpha_1} \mathbf{D}_1 + \frac{1}{\alpha_2} \mathbf{D}_2 - \mathbf{P}^T \mathbf{P}. \end{aligned} \quad (6.14)$$

The diagonal matrix $\mathbf{D}_2 = \text{diag}(\mathbf{p}_1)$, where $\mathbf{p} = [1, \dots, q-1, q, \dots, q, q-1, \dots, 1]$ [26]. The length of $\mathbf{p}_1$ is the same as the $\mathbf{x}$. If the length is odd number, there are only 1 q in $\mathbf{p}_1$. Otherwise there are two q . The diagonal matrix $\mathbf{D}_2 = \text{diag}(\mathbf{p}_2) \mathbf{D}_1$, where $\mathbf{p}_2 = [1, 0, \dots, 1, 0, \dots, 0]$. The index for the nonzero set is $\mathcal{I}_1 = \{1, M_T + 1, \dots, m M_T + 1, \dots, (M_T + M_R - 2) M_T + 1\}$ defined

in $\mathcal{A}_1(\cdot)$. The other two elements in $\mathbf{Q}$ is:

$$\begin{aligned} \frac{\partial^2}{\partial \text{vec}(\mathbf{H}) \partial \mathbf{x}} f &= \frac{1}{\alpha_2} \text{vec } \mathcal{H}(\cdot) \\ &= \frac{1}{\alpha_2} \mathbf{A}_1, \end{aligned} \quad (6.15)$$

$$\begin{aligned} \frac{\partial^2}{\partial \text{vec}(\mathbf{H}_\theta) \partial \mathbf{x}} f &= \frac{1}{\alpha_1} \text{vec } \mathcal{H} \mathcal{A}_1(\cdot) \\ &= \frac{1}{\alpha_1} \mathbf{A}_2. \end{aligned} \quad (6.16)$$

The matrix $\mathbf{A}_1 \in \mathbb{R}^{h_3 \times h_2}$ is given by:

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{A}_{1,1} \\ \mathbf{A}_{1,2} \\ \vdots \\ \mathbf{A}_{1,N} \end{bmatrix}, \quad (6.17)$$

where $\mathbf{A}_{1,n} \in \mathbb{R}^{N \times (2N-1)}$. The identify matrix $\mathbf{I}_N$ is located from the n -th column. All other elements are zero. The matrix $\mathbf{A}_2 \in \mathbb{R}^{h_3 \times h_1} = \mathbf{A}_1 \mathbf{B}$. The matrix $\mathbf{B}$ represent the linear operator $\mathcal{A}_1(\cdot)$. It is all zero matrix, excluding $\mathcal{I}_1(n)$ -th column and n -th row are 1. $\forall n$.

PDOM algorithm summary

Algorithm 7: PDOM for (4.22)

Define: dogleg path: $\mathbf{p}(\alpha) := \begin{cases} -\tau \mathbf{g}, & \alpha \in (0, 1] \\ -\tau \mathbf{g} + (\alpha - 1)(-\mathbf{Q}^{-1} \mathbf{g} + \tau \mathbf{g}), & \alpha \in (1, 2] \end{cases}$

where $\mathbf{g} := [\text{vec}(\nabla_{\mathbf{H}_\theta} f)^T, \text{vec}(\nabla_{\mathbf{H}} f)^T, \nabla_{\mathbf{x}} f^T]^T$;

The surrogate function $\hat{f}_\lambda(u, v) = f(u) + \mathbf{g}_\alpha^T(v - u) + \frac{1}{2\lambda} \|v - u\|_F^2$;

where $\mathbf{g}_\alpha = \mathbf{g}^T \mathbf{p}^2(\alpha) / \|\mathbf{p}(\alpha)\|^2$

Input: $(\mathbf{H}_\theta^1, \mathbf{H}^1, \mathbf{x}^1), \mathbf{Q}^{-1}, i = 0$

Data: fixed step size $\tau = 1/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{Q}$.

$\lambda = 0.98, \alpha \in [0, 2]$

while (4.40), (4.41) and (4.42) are not satisfied or $i < 2000$ **do**

```

     $i = i + 1$  ;                                     /* Update step size  $\tau_{\alpha^i}$  */
    Find the largest value  $\alpha^i \in \{1 + (1/2)^k | k \in \mathbb{N}\}$  such that
     $f(\mathbf{H}_\theta^{i+1}, \mathbf{H}^{i+1}, \mathbf{x}^{i+1}) \leq \hat{f}_{\alpha^i}((\mathbf{H}_\theta^{i+1}, \mathbf{H}^{i+1}, \mathbf{x}^{i+1}), (\mathbf{H}_\theta^i, \mathbf{H}^i, \mathbf{x}^i))$ . ;
     $\tau_{\alpha^i} = -\|\mathbf{p}(\alpha)\|^2 / (\mathbf{g}^T \mathbf{p}(\alpha))$  ;
    ;                                               /* PDOM */
     $\mathbf{x}^{i+1} = \mathbf{x}^i + \mathbf{p}_\lambda(\alpha^i)$ ;
     $\text{vec}(\mathbf{H})^{i+1} = \text{vec}(\text{prox}_{\lambda \tau_{\alpha^i}, g}(\mathbf{H}^i + \mathbf{p}_\lambda(\alpha^i)))$ ;
     $\text{vec}(\mathbf{H}_\theta)^{i+1} = \text{vec}(\text{prox}_{\lambda \tau_{\alpha^i}, g}(\mathbf{H}_\theta^i + \mathbf{p}_\lambda(\alpha^i)))$ ;
    ;                                               /* Proximal gradient method */
     $\mathbf{x}_v^{i+1} = \mathbf{x}^i - \lambda \nabla_{\mathbf{x}} f(\mathbf{H}_\theta^i, \mathbf{H}^i, \mathbf{x}^i)$ ;
     $\mathbf{H}_v^{i+1} = \text{prox}_{\lambda, g}(\mathbf{H}^i - \lambda \nabla_{\mathbf{H}} f(\mathbf{H}_\theta^i, \mathbf{H}^i, \mathbf{x}^i))$ ;
     $\mathbf{H}_{\theta, v}^{i+1} = \text{prox}_{\lambda, g}(\mathbf{H}_\theta^i - \lambda \nabla_{\mathbf{H}_\theta} f(\mathbf{H}_\theta^i, \mathbf{H}^i, \mathbf{x}^i))$ ;
    ;                                               /* Compare and select */
    if  $f(\mathbf{H}_\theta^{i+1}, \mathbf{H}^{i+1}, \mathbf{x}^{i+1}) > f(\mathbf{H}_{\theta, v}^{i+1}, \mathbf{H}_v^{i+1}, \mathbf{x}_v^{i+1})$  then
         $\mathbf{x}^{i+1} = \mathbf{x}_v^{i+1}$  ;
         $\mathbf{H}^{i+1} = \mathbf{H}_v^{i+1}$  ;
         $\mathbf{H}_\theta^{i+1} = \mathbf{H}_{\theta, v}^{i+1}$  ;
    end

```

end

Output: $(\hat{\mathbf{x}}, \hat{\mathbf{H}}, \hat{\mathbf{H}}_\theta) \leftarrow (\mathbf{x}^{i+1}, \mathbf{H}^{i+1}, \mathbf{H}_\theta^{i+1})$;
