CHALMERS
UNIVERSITY OF TECHNOLOGY

Calibration of Array Antennas

A frequency domain approach based on realization theory and convex optimization

Master's thesis in Communication Engineering

HAN WU

Department of Electrical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020

Master's thesis

Calibration of Array Antennas

A frequency domain approach based on realization theory and convex optimization

HAN WU

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Signal Processing research Group
Chalmers University of Technology
Gothenburg, Sweden 2020

Calibration of Array Antennas
A frequency domain approach based on realization theory and convex optimization
HAN WU

Supervisor: Tomas McKelvey, Department of Electrical Engineering
Examiner: Tomas McKelvey, Department of Electrical Engineering

Master's Thesis
Department of Electrical Engineering
Signal Processing research group
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2020

Calibration of Array Antennas
A frequency domain approach based on realization theory and convex optimization
HAN WU
Department of Electrical Engineering
Chalmers University of Technology

Abstract

For array antennas, the signal from each antenna element is individually sampled and a digital radar yields vector valued measurements. When array antennas are manufactured, perturbations are inevitable due to various reasons like environmental effects, so the digital response from the array antenna will be subject to an unknown but deterministic multiplicative perturbation. In order to mitigate this impact, many methods from different perspectives have been proposed through these years. A new frequency domain approach in theis thesis is introduced based on system realization theory and convex optimization, which can help calibrate the estimated angle and unknown gains from radar. Moreover, some contributions to the realization theory in frequency domain are made.

Keywords: radar calibration, system theory, signal processing, convex optimization.

Acknowledgements

I want to thank Prof.Tomas McKelvey, who is my supervisor and gives me a lot of help and instructions when I am doing this thesis. Not only does he help me through difficulties, but he also teach me how to do research, which will equip me for life. Also, thanks to department of Electrical Engineering after I spent two most important years here and laid a solid foundation for future studying.

Han Wu, Gothenburg, June 22, 2020

Contents

1

Introduction

This thesis consists of five chapters. This first chapter gives an introduction to calibration of array antennas. Then, the second chapter will review some theoretical knowledge from system theory. Next, the third chapter is about the method we developed in this thesis. Finally, we will show our algorithms and results as well as conclusion in the forth and fifth chapter.

1.1 Array antenna

An array antenna is a set of $N(N > 1)$ spatially separated antennas that can receive multiple radar signals simultaneously. It has many advantages compared to traditional single antenna radar such as increasing the overall gain and SINR (Signal to Interference Noise Ratio). Here, we are only interested in the basics that how it can detect the angle of received signals. As array antenna is usually installed far away from targets, signals arriving at the spatially separated elementes of antenna can be considered parallel.

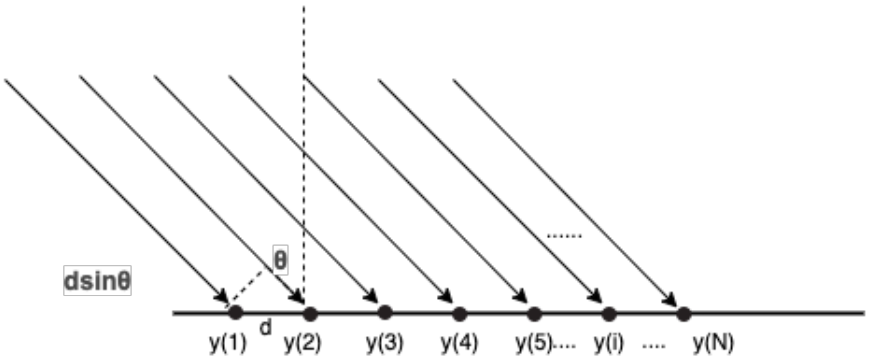

Figure 1.1: Example of array antenna

As the figure 1.1 shows, this array antenna has N elementes with signals from the same target arriving at different elementes simultaneously, and the distance between different elementes are the same, denoted as d. Assuming the signal received is a narrow band signal which during the aquisition time can be regarded as a sinusoidal signal, or a sum of signal if we have several incoming signal directions. After mixing and LP-fltering (demodultion), assuming the complex (IQ) signal we obtain in the first element is ϕ, that is, $y(1) = \phi$, because the signal arriving at $y(2)$ comes from

the same origin and is parallel to $y(1)$, it will have a phase delay ψ with

$$\psi = \frac{2\pi d sin\theta}{\lambda} \tag{1.1}$$

where d is the distance between two elementes, θ is the angle of the incoming signal with reference to the vector perpendicular to the extent of the array and λ is the wavelength of signal. Hence, we will get $y(2) = \phi e^{j\psi}$. Repeating this step, we will get the signal received at element k is

$$y(k) = \phi e^{j(k-1)\psi} \tag{1.2}$$

with $k = 1, 2, \cdots, N$. Therefore, we can easily get the angle of the target or other information when we receive the N signals with this relation.

1.2 Background

As we have seen, due to their advantages, array antennas play an important role in modern radar systems. But what we have talked about is only theoretical theory, when array antennas are manufactured, perturbations are inevitable due to various reasons like environmental effects, so the digital response from the array antenna will be subject to an unknown but deterministic multiplicative perturbation. In order to mitigate this impact, the process of calibration is employed by providing a compensation factor for all channels/antennas. A range of ways to perform calibration have been put forward through the years, see [1–3]. A classical approach is to measure the array in a perfectly known environment with known radar target geometry, e.g. an anechoic measurement chamber, while such method is very expensive.

1.3 Literature review and purpose

The problem of estimating the unknown gains in a linear array without knowing the directions to the active signal sources is called auto-calibration, which can be regarded as a joint estimation of the array gain and the directions to the signal sources. This problem has been researched over these years with various assumptions on the array and the target properties. For example, Paulraj and Kailath [4] first investigated the problem of estimating direction-of-arrival (DOA) with a approach that uses information in the observed covariarice matrix to mitigate the effects of noise, sensor gain error, etc. Then Friedlander and Weiss [5] studied the problem by simultaneously estimating the DOA's and the unknown gains and phases with an eigenstructure based method. In [6], they proposed a method that uses the special structure of mutual coupling matrix and the subspace principle. Next a new blind calibration method [7] was studied by imposing the condition of the minimum sample second moment of the array phase errors. Other examples of recent works are [8, 9]. The idea of this thesis comes from the research in [10], where the author proposed a method that uses linear system theory. In this thesis methods for autonomous calibration will be investigated where the exact locations of the radar targets are

unknown. First, we will investigate the system theory and develop the realization theory in frequency domain. Then, a signal model for linear array is built and we want to obtain the rank property if the signal comes from multiple radar targets. Finally, we will combine the realization theory and convex optimization theory to come up with a new algorithm to achieve the calibration.

2

Theory

Because the mathematical model we use in our method is the state-space model and we also use some knowledge in the system theory and convex optimization to do the calibration, this chapter is devoted to the system and control theory and convex optimization theory. The main part of this chapter is based on [11], [12], [13] and [14] .

2.1 System and control theory

Control means influencing an object's behavior to achieve a certain aim, and system is a mathematical model that is used to analyze the relation between the output and input and helps us to achieve control. Usually denoted by Σ, a system in time domain can be described by two equations

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{2.1}$$

Here, $u(t)$ is called the *input* of the system that is usually given outside. A, B, C, D are suitable maps and considered time invariant in this thesis, which means A, B, C, D do not change with time. The variable $x(t)$ is called the *state variable* with $y(t)$ called the *output* of the system. Obviously when $y(t)$ and $u(t)$ are determined, $x(t)$ can still change with different A, B, C, D choosed, which implies that there exist multiple systems that can have the same behavior. In many cases, D shows to be irrelevant to the output and can be considered as zero, so a system Σ is often connected with the triple (C, A, B). In this thesis, we only assume that $A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{n\times m}, C \in \mathbb{C}^{p\times n}$ and $x \in \mathcal{X}$ with $\mathcal{X}$ called *state space*. The *dimension* of the system is the dimension of $\mathcal{X}$.

Because this thesis is researched based on discrete time system, we will only discuss about the property of discrete time system from now on. In fact, the property of continuous time system can be derived easily corresponding to discrete time case. Now let's look at the discrete time system. For discrete time system Σ, it's described by

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned} \tag{2.2}$$

Therefore, when $x(0) = x_0$ we will get

$$\begin{aligned} x(k) &= A^k x_0 + \sum_{i=1}^{k-1} A^i B u(k) \\ y(k) &= CA^k x_0 + \sum_{i=1}^{k} CA^i B u(k-i) + Du(k) \end{aligned} \tag{2.3}$$

Also after applying discrete time Laplace transform, the system in frequency domain is

$$\begin{aligned} zX &= AX + BU \\ Y &= CX + DU \end{aligned} \tag{2.4}$$

Hence, $Y = [C(zI - A)^{-1}B + D]U$ and transfer function $H(z) = C(zI - A)^{-1}B + D$.

2.1.1 Controllability

In this section, we focus on the relation between input and the state.

Definition 1. *The system Σ is controllable in time T if every state can be reached from every state after a given time T.*
We introduce the Cayley-Hamilton theorem here.
Theorem 1. *Given $n \times n$ matrix A and identity matrix I_n, the characteristic polynomial of A is $p(\lambda) = \det(\lambda I_n - A)$. Then*

$$p(A) = 0 \tag{2.5}$$

Cayley-Hamilton theorem tells us there exists $\alpha_0, \alpha_1, \cdots, \alpha_{n-1}$ such that $A^n = \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1}A^{n-1}$. Denote

$$\mathbf{R}(A, B) = \mathbf{R}([B, AB, A^2B, \cdots, A^{n-1}B]) \tag{2.6}$$

$\mathbf{R}([B, AB, A^2B, \cdots, A^{n-1}B])$ is the range space of $[B, AB, A^2B, \cdots, A^{n-1}B]$. Given above,

$$x(k) = A^k x_0 + \sum_{i=1}^{k} A^i B u(k-i) \in \mathbf{R}(A, B) \tag{2.7}$$

When the system is controllable, every state can be reached from every state, which means $\mathcal{X} = \mathbf{R}(A, B)$ that happens iff rank$\mathbf{R}(A, B) = rank(\mathcal{X}) = n$.

Theorem 2. *The n-dimensional discrete time linear system Σ is controllable if and only if* rank $\mathbf{R}(A, B) = n$.
We will consider a pair(A, B) is controllable if rank $\mathbf{R}(A, B) = n$, and usually we will denote $\mathbf{R}(A, B)$ by $\mathcal{R}$. Corresponding result can also be obtained in continuous system.

2.1.2 Observability

In this section, we want to look at the the relation between state and output for a given input.

Definition 2. *For two distinct states x and z of a system Σ , they are indistinguishable in time T if they give the same outputs for any given input u within time T. Or they are distinguishable if there at least exist one input u with time $\tau \leq T$ leading to distinct outputs.*

Definition 3. *The system Σ is observable if any two distinct states are distinguishable.*
This suggests that there does not exist two distinct states such that

$$\begin{aligned} CA^k x + \sum_{i=1}^{k} CA^i Bu(k-i) + Du(k) = CA^k z + \sum_{i=1}^{k} CA^i Bu(k-i) + Du(k) \\ CA^k(x-z) = 0 \end{aligned} \tag{2.8}$$

Because this holds for any $k \geq 0$, according to Cayley-Hamilton thorem,

$$\mathcal{N}(\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}) = 0 \tag{2.9}$$

equivalently,

$$\text{rank}(\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}) = n \tag{2.10}$$

Define

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \tag{2.11}$$

Therefore, we can get the theorem about observability

Theorem 3. *The n-dimensional discrete time linear system Σ is observable if and only if* rank $\mathcal{O} = n$
We will consider a pair (C, A) observable if rank $\mathcal{O} = n$.

2.1.3 Transfer function

The transfer function of discrete time linear system is

$$H(z) = C(zI - A)^{-1}B + D \tag{2.12}$$

If z_i is the pole of $H(z)$, then $(z_iI - A)$ must be singular. It follows that z_i must be the eigenvalue of A. Usually we are more familiar with this form

$$H(z) = \frac{Y(z)}{U(z)} \tag{2.13}$$

Consequently, we will be interested to know when $H(z)$ can build up the relation with (C, A, B, D).

Definition 4. *A state space model (C,A,B,D) is a realiztion of a transfer function H(z) if* $C(zI-A)^{-1}B+D = H(z)$. *The transfer function* $H(z)$ *is realizable if there exists a realization.*

Lemma 1. *Scalar transfer function* $H(z)$ *is realizable if and only if* $H(z)$ *is a proper scalar function.*

Proof. i) If $H(z)$ is realizable, there will exist at least one state space model (C, A, B, D) such that

$$\begin{aligned} H(z) &= C(zI - A)^{-1}B + D \\ &= \frac{C\text{adj}(zI - A)B}{\det(zI - A)} + D \end{aligned} \tag{2.14}$$

where $\text{adj}(A)$ is the adjugate matrix of matrix A. Because the degree of denominator is n while the degree of numerator is smaller than n, $H(z)$ is a proper function.
ii) If $H(z)$ is a strict proper function, it suggests that

$$y(n) + a_1y(n-1) + \cdots + a_ny(0) = c_nu(n-1) + \cdots + c_1u(0) \tag{2.15}$$

It's easy to see that we can rewrite the system into

$$\begin{aligned} x(n+1) &= Ax(n) + Bu(n) \\ y(n) &= Cx(n) \end{aligned} \tag{2.16}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & a_1 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, C = (c_1, c_2, \cdots, c_n) \tag{2.17}$$

When $H(z)$ is not strict proper, we can split $H(z)$ into the sum of a strict proper function and a constant D. Hence, we will have

$$\begin{aligned} x(n+1) &= Ax(n) + Bu(n) \\ y(n) &= Cx(n) + Du(n) \end{aligned} \tag{2.18}$$

□

System realized by (A, B, C) is called the *control canonical form* and A is called the companion matrix.
Here we only gives the realization lemma for a MIMO system without proof.

Lemma 2 (MIMO system)**.** *Transfer function $H(z)$ is realizable if and only if there eixsts a polynomial matrix pair $(N_R(z), D_R(z))$ or $(N_L(z), D_L(z))$ such that*

$$H(z) = N_R(z)D_R^{-1}(z) \text{ or } H(z) = D_L^{-1}(z)N_L(z) \tag{2.19}$$

Proof. See [15, Chapter 6] □

2.2 Realization theory for linear system

For given input and output, we will wonder whether we can rewrite the system into the equation 2.2 form, or more specifically, whether we can find the matrices (C, A, B, D) to build up the relation between the input and output. As we have assumed $A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{n\times m}, C \in \mathbb{C}^{p\times n}$. The dimension of the system is n, and m is the number of input with p suggesting the number of output.

Definition 5. *In frequency domain, for given output function Y(z) and input U(z), the linear time invariant system is realizable if there exists matrices (C,A,B) and D such that*

$$\begin{aligned} zX(z) &= AX(z) + BU(z) \\ Y(z) &= CX(z) + DU(z) \end{aligned} \tag{2.20}$$

(C,A,B) and D is called a realization of the system.

Definition 6. *The triple (C,A,B) is canonical if it is controllable and observable.*

Lemma 3. *For any non-canonical triple (C,A,B), there exist some nonsingular square matrices Q so that*

$$Q^{-1}AQ = \begin{bmatrix} A_{11} & A_{12} & 0 \\ 0 & A_{22} & 0 \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \tag{2.21}$$

as well as

$$Q^{-1}B = \begin{bmatrix} B_{11} \\ 0 \\ B_{31} \end{bmatrix} \quad \text{and} \quad CQ = (C_{11}, C_{12}, 0) \tag{2.22}$$

for some matrices $A_{11}, \cdots$. The triple (C_{11}, A_{11}, B_{11}) is controllable and observable with

$$CA^iB = C_{11}A_{11}^iB_{11} \tag{2.23}$$

Proof. See [12, lemma 6.5.1] □

Definition 7. *For two triples (C,A,B) and $(\tilde{C}, \tilde{A}, \tilde{B})$ where $\tilde{A} \in \mathbb{C}^{n\times n}, \tilde{B} \in \mathbb{C}^{n\times m}, \tilde{C} \in \mathbb{C}^{p\times m}$. Define (C, A, B) is similar to $(\tilde{C}, \tilde{A}, \tilde{B})$ if there exists a nonsingular square matrix T such that $\tilde{A} = T^{-1}AT, \tilde{B} = T^{-1}B$ and $\tilde{C} = CT$, which is denoted by (C,A,B) $\sim (\tilde{C}, \tilde{A}, \tilde{B})$*

Definition 8. *The n-dimensional triple (C,A,B) is minimal if any other* $(\tilde{C}, \tilde{A}, \tilde{B})$ *realizing the system must have dimension at least n.*
Now we can start to figure out these following theorems with these definitions. As we have mentioned, D is often irrelevant to the output. Here we assume $D = 0$.

Theorem 4. *The system has a canonical realization if there is a realization for the system.*

Proof. According to lemma 3, when the system has a non-canonical realization (C, A, B) with dimension n, we can find a canonical triple (C_{11}, A_{11}, B_{11}) with

$$CA^iB = C_{11}A_{11}^iB_{11} \tag{2.24}$$

Here we introduce

$$(zI - A)^{-1} = \sum_{i=1}^{\infty} A^{i-1}z^{-i} \tag{2.25}$$

It follows that

$$\begin{aligned} Y &= C(zI - A)^{-1}BU \\ &= \sum_{i=1}^{\infty} CA^{i-1}Bz^{-i}U \\ &= \sum_{i=1}^{\infty} C_{11}A_{11}^{i-1}B_{11}z^{-i}U \\ &= C_{11}(zI - A_{11})^{-1}B_{11}U \end{aligned} \tag{2.26}$$

Thus, the canonical triple (C_{11}, A_{11}, B_{11}) is also a realization of the system. □

Theorem 5. *A realizaiton of the system is minimal if and only if it is canonical.*

Proof. If the triple (C, A, B) is a canonical realizaiton of the system. The according to lemman 3, it is minimal because any other realization that is similar to (C, A, B). When the triple (C, A, B) is minimal, if (C, A, B) is not canonical, then we will be able to find a canonical triple (C_{11}, A_{11}, B_{11}) with a smaller dimension realizing the same system according to theorem 4 . Consequently, (C, A, B) will not be minimal. Therefore, (C, A, B) must also be canonical. □

Theorem 6. *Any two minimal realization must be similar.*

Proof. As a result from theorem 5, the two minimal triples are also canonical triple that are similar to each other. Therefore, the two minimal realization must be similar. □

2.3 Proximal algorithms

An *optimization problem* has this form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, 2, \cdots, m \end{array} \tag{2.27}$$

The vector $x = (x_1, x_2, \cdots, x_n)$ is the *optimization variable* of the problem, the function $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the *objective* function, the functions $f_i : \mathbb{R}^n \to \mathbb{R}, i =$

$1, 2, \cdots, m$ are the *constraint functions*, and the constants $b_1, b_2, \cdots, b_m$ are the limits or bounds for the constraints. A vector $x^\star$ is called the *optimal*, or a *solution* of the problem 2.27, if it has the smallest objective value among all vectors that satisfy the constraints, that is, for any y with $f_1(y) \leq b_1, \cdots, f_m(y) \leq b_m$, we have $f_0(y) \geq f_0(x^\star)$. A convex optimization problem is the problem whose objective and constraint functions are all convex, which means

$$f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y) \tag{2.28}$$

for all $x, y \in \mathbb{R}^n$ and all $\alpha, \beta \in \mathbf{R}$ with $\alpha \geq 0, \beta \geq 0$ as well as $\alpha + \beta = 1$.
A set is *convex* if the line segment between any two points from the set still lies in the set. More specifically, if for any $x_1, x_2 \in C$ and with any θ with $0 \leq \theta \leq 1$ we have $\theta x_1 + (1 - \theta)x_2 \in C$, the set C is a convex set.

2.3.1 Definition

In this thesis, we only talk about a class of algorithms, called proximal algorithms, for solving convex optimization problems. In proximal algorithms, the base operation is evaluating the *proximal operator* of a function, which involves solving a small convex optimization problem.
Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a closed proper convex function, which means that its epigraph

$$\mathbf{epi}f = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq t\} \tag{2.29}$$

is a nonempty closed convex set. The effecitive domain of f is

$$\mathbf{dom}f = \{x \in \mathbb{R}^n \mid f(x) < +\infty\} \tag{2.30}$$

The proximal operator $\mathbf{prox}_{\lambda f} : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\mathbf{prox}_{\lambda f}(v) = \arg\min_x \left(f(x) + \frac{1}{2\lambda} \|x - v\|_2^2 \right) \tag{2.31}$$

with $\lambda \geq 0$. The function minimized on the righthand side is strongly convex and not everywhere infinite, so it has a unique minimizer for every $v \in \mathbb{R}^n$. We can interpret it in this way that the proximal operator delivers a value x which is a compromise between making $f(x)$ small and being close to the value v and the parameter λ can be interpreted as a relative weight or trade-off parameter between these terms. Proximal algorithms are most useful when all the relevant proximal operators can be evaluated sufficiently quickly.

Lemma 4. *The point $x^\star$ miniminzes f if and only if $x^\star = \mathbf{prox}_{\lambda f} f(x^\star)$*

Proof. If $x^\star$ minimizes f, then $f(x) \geq f(x^\star)$ for any x. It follows that

$$f(x) + \frac{1}{2\lambda} \|x - x^\star\|_2^2 \geq f(x^\star) + \frac{1}{2\lambda} \|x^\star - x^\star\|_2^2 \tag{2.32}$$

for any x, so $x^\star = \mathbf{prox}_{\lambda f} f(x^\star)$. If $x^\star = \mathbf{prox}_{\lambda f} f(x^\star)$, here we assume $f(x)$ is subdifferentiable for convinence, where subdifferential is defined by

$$\partial f = \{y \mid f(z) \geq f(x) + y^T(x - z) \text{ for all } z \in \mathbf{dom}f\}. \tag{2.33}$$

When $x^\star = \mathbf{prox}_{\lambda f} f(x^\star)$, this means $x^\star$ minimizes $f(x) + \frac{1}{2\lambda}\|x - x^\star\|_2^2$, which only happens when $0 \in \partial f(x^\star) + \frac{1}{2\lambda}\|x^\star - x^\star\|_2^2$. Therefore, $0 \in \partial f(x^\star)$ and $x^\star$ minimizes f. □

Hence, many proximal algorithms for optimization can be interpreted as methods for finding fixed points $x^\star$ of appropriate operators.

2.3.2 Moreau envelop

Let $f : \mathbb{R}^n \to \mathbb{R}$. The function $f^* : \mathbb{R}^n \to \mathbb{R}$, defined as

$$f^*(y) = \sup_{x \in \mathbf{dom} f} (y^T x - f(x)) \tag{2.34}$$

is called the *conjugate* of the function f. The domain of the conjugate function consists of $y \in \mathbb{R}^n$ which the supremum is finite, *i.e.*, for which the difference $y^T x - f(x)$ is bounded above on $\mathbf{dom} f$. We see immediately that y^* is a convex function, since it is the pointwise supremum of a family of convex (indeed, affine) functions of y. This is true whether or not f is convex. We can understand conjugate in this way. For all x, there is a line $y^T x - \alpha$ such that

$$\begin{aligned} f(x) &\geq y^T x - \alpha \\ \alpha &\geq y^T x - f(x) \\ \alpha &\geq f^*(y) \end{aligned} \tag{2.35}$$

It follows that $f^*(y)$ is the best choice α and $f(x) \geq f^{**}(x)$.
Define the *lower convex envelope* $g(x)$ as

$$g(x) = \sup\{h(x) |\ h \text{ is convex and } h \leq f \text{ over } \mathbf{dom} f\} \tag{2.36}$$

Here these properties can be shown :

1. when f is convex, $f = f^{**}$.
2. $g(x) = f^{**}(x)$.
3. $g(x)$ have the same minimizer $x^\star$ as $f(x)$ when $g(x)$ and $f(x)$ has the same domain, which means we can find the minimizer of $f(x)$ by $g(x)$.

Given $\lambda > 0$, the *Moreau envelop* or $Moreau - Yosida\ regularization$ $M_{\lambda f}$ is defined as

$$M_{\lambda f}(v) = \inf_x \left(f(x) + \frac{1}{2\lambda}\|v - x\|_2^2 \right) \tag{2.37}$$

Here it can be shown that

$$M_{\lambda f}(v) = \left(f^*(x) + \frac{\lambda}{2}\|x\|_2^2 \right)^* (v) \tag{2.38}$$

It suggests that the Moreau envelope can be interpreted as obtaining a smooth approximation to a function by taking its conjugate, adding regularization, and then taking the conjugate again. With no regularization, this would simply give the original function; with the quadratic regularization, it gives a smooth approximation.

$\mathbf{prox}_{\lambda f}$ returns the unique point that actually achieves the infimum that defines $M_{\lambda f}$. Hence,

$$\begin{aligned} M_{\lambda f}(x) &= f\left(\mathbf{prox}_{\lambda f}(x)\right) + \frac{1}{2\lambda}\left\|x - \mathbf{prox}_{\lambda f}(x)\right\|_2^2 \\ \Rightarrow\ \nabla M_{\lambda f}(x) &= \frac{1}{\lambda}\left(x - \mathbf{prox}_{\lambda f}(x)\right) \\ \Rightarrow\ \mathbf{prox}_{\lambda f}(x) &= x - \lambda \nabla M_{\lambda f}(x) \end{aligned} \tag{2.39}$$

This means that $\mathbf{prox}_{\lambda f}(x)$ can be viewed as a gradient step, with step size λ, for minimizing $M_{\lambda f}(x)$ which has the same minimizers as f.

Lemma 5.

$$\mathbf{prox}_f(x) = \nabla M_{f^*}(x) \tag{2.40}$$

Proof.

$$\begin{aligned} M_f(v) &= \inf_x \left(f(x) + \frac{1}{2}\|v - x\|_2^2 \right) \\ &= \frac{1}{2}\|v\|_2^2 - \left(f(x) + \frac{1}{2}\|x\|_2^2 \right)^* (v) \end{aligned} \tag{2.41}$$

According to the definition,

$$M_{f^*}(v) = \left(f(x) + \frac{1}{2}\|x\|_2^2 \right)^* (v) \tag{2.42}$$

Hence,

$$\nabla M_f = v - \nabla M_{f^*} \tag{2.43}$$

It follows that

$$\mathbf{prox}_f(x) = \nabla M_{f^*}(x) \tag{2.44}$$

□

With this relation, we can develop the *Moreau decomposition*.

Lemma 6. *The following relation always holds that*

$$v = \mathbf{prox}_f(v) + \mathbf{prox}_{f^*}(v) \tag{2.45}$$

Proof.

$$\begin{aligned} M_f(v) &= \inf_x \left(f(x) + \frac{1}{2}\|v - x\|_2^2 \right) \\ &= \frac{1}{2}\|v\|_2^2 - \left(f(x) + \frac{1}{2}\|x\|_2^2 \right)^* (v) \end{aligned} \tag{2.46}$$

Hence,

$$\frac{1}{2}\|v\|_2^2 = M_f(v) + M_{f^*}(v) \tag{2.47}$$

After differential,

$$v = \mathbf{prox}_f(v) + \mathbf{prox}_{f^*}(v) \tag{2.48}$$

□

2.3.3 ADMM algorithm

The *proximal minimization algorithm* [14] , also called *proximal iteration* or the *proximal point algorithm*, is

$$x^{k+1} := \mathbf{prox}_{\lambda f}(x^k) \tag{2.49}$$

where $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is a closed proper convex function, k is the iteration counter, and x^k denotes the kth iterate of the algorithm. If f has a minimum, then x^kconverges to the set of minimizers of f and $f(x^k)$ converges to its optimal value. Consider the problem,

$$\text{minimize} \quad f(x) + g(x) \tag{2.50}$$

where $f, g : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ are closed proper convex functions. (In this splitting, both f and g can be nonsmooth.) Then the alternating direction method of multipliers (ADMM), also known as Douglas- Rachford splitting, is

$$\begin{aligned} x^{k+1} &:= \mathbf{prox}_{\lambda f}(z^k - u^k) \\ z^{k+1} &:= \mathbf{prox}_{\lambda g}(x^{k+1} + u^k) \\ u^{k+1} &:= u^k + x^{k+1} - z^{k+1} \end{aligned} \tag{2.51}$$

where k is an iteration counter. ADMM is most useful when the proximal operators of f and g can be efficiently evaluated but the proximal operator for $f + g$ is not easy to evaluate.

3

Method

This chapter is devoted to the theoretical result we developed based on chapter 2, and the approximation algorithm that can be applied to radar calibration.

3.1 Realization theorem in frequency domain

Given M matrices $Y_1, Y_2, \cdots, Y_M$, where $Y_i \in \mathbb{C}^{p\times m}$, $i = 1, 2, \cdots, M$, and M distinct complex number $z_1, z_2, \cdots, z_M$. Define

$$\mathbf{Y}_s = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_M \\ z_1Y_1 & z_2Y_2 & \cdots & z_MY_M \\ z_1^2Y_1 & z_2^2Y_2 & \cdots & z_M^2Y_M \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{s-1}Y_1 & z_2^{s-1}Y_2 & \cdots & z_M^{s-1}Y_M \end{bmatrix} \in \mathbb{C}^{sp\times mM} \tag{3.1}$$

$$\mathbf{U}_s = \begin{bmatrix} I_m & I_m & \cdots & I_m \\ z_1I_m & z_2I_m & \cdots & z_MI_m \\ z_1^2I_m & z_2^2I_m & \cdots & z_M^2I_m \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{s-1}I_m & z_2^{s-1}I_m & \cdots & z_M^{s-1}I_m \end{bmatrix} \in \mathbb{C}^{sm\times mM} \tag{3.2}$$

where I_m is identity mstrix with order m. Define the projection matrix

$$\mathbf{P}_s = I - \mathbf{U}_s^*(\mathbf{U}_s\mathbf{U}_s^*)^{-1}\mathbf{U}_s \tag{3.3}$$

which projects onto the nullspace of $\mathbf{U}_s$ and $\mathbf{U}_s^*$ is the conjugate transpose of $\mathbf{U}_s$. Define the matrix $\mathbf{Y}_s^{j-}$ as the matrix formed from the $M-1$ block columns of matrix $\mathbf{Y}_s$ in 3.1 with all indices except the index j and $j = 1, \cdots, M$. This will form a matrix where the block column j is removed from the original matrix $\mathbf{Y}_s$. Let $\mathbf{U}_s^{j-}$ be defined in the same manner based on matrix $\mathbf{U}_s$ with corresponding projection matrix $\mathbf{P}_s^{j-}$.

Theorem 7 (Scalar case with $p = 1$)**.** *There exists a minimal triple (C, A, B) and D where $C \in \mathbb{C}^{p\times n}, A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{n\times m}, D \in \mathbb{C}^{p\times m}$ such that*

$$Y_i = C(z_iI_n - A)^{-1}B + D \quad (i = 1, 2, \cdots, M) \tag{3.4}$$

if and only if

$$\text{rank}\mathbf{Y}_n\mathbf{P}_{n+1} = \text{rank}\mathbf{Y}_{n+1}\mathbf{P}_{n+1} = \text{rank}\mathbf{Y}_n^{j-}\mathbf{P}_n^{j-} = n, \text{ for } j = 1, 2\cdots, M \tag{3.5}$$

with $M \geq 2n + 1$.

3.2 Proof

i) When $\text{rank}\mathbf{Y}_n\mathbf{P}_{n+1} = \text{rank}\mathbf{Y}_{n+1}\mathbf{P}_{n+1} = \text{rank}\mathbf{Y}_n^{j-}\mathbf{P}_n^{j-} = n,$ for $j = 1, 2\cdots, M$.

Lemma 7. *Let $Y \in \mathbb{C}^{p\times M}, U \in \mathbb{C}^{q\times M}$ with $q < M$ and $\text{rank}U = q$. Define the projection matrix $P = I - U^*(UU^*)^{-1}U$, then*

$$\text{rank}(YP) = n \quad \text{if and only if} \quad \text{rank}\begin{bmatrix} U \\ Y \end{bmatrix} = q + n \tag{3.6}$$

Proof. Set U_c a matrix with $\text{rank}\begin{bmatrix} U \\ U_c \end{bmatrix} = M$, then there will exist two matrices L and R such that $Y = LU_c + RU$. Hence, $\text{rank}(YP) = \text{rank}(LU_cP) = n$. It follows that

$$\text{rank}\begin{bmatrix} U \\ Y \end{bmatrix} = \text{rank}\begin{bmatrix} U \\ LU_c \end{bmatrix} = \text{rank}U + \text{rank}(LU_cP) = q + n \tag{3.7}$$

□

According to this lemma, when $\text{rank}\mathbf{Y}_n\mathbf{P}_{n+1}$=$\text{rank}\mathbf{Y}_{n+1}\mathbf{P}_{n+1} = n$, we can get

$$\text{rank}\begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{Y}_n \end{bmatrix} = \text{rank}\begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{Y}_{n+1} \end{bmatrix} = m(n+1) + n \tag{3.8}$$

which implies that the last row in $\mathbf{Y}_{n+1}$ is a linear combination of the rows in $\begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{Y}_n \end{bmatrix}$. Thus, there will exist two non-zero vectors $v_1^T \in \mathbb{C}^{1\times(n+1)m}$ and $v_2^T = [a_0, \cdots, a_{n-1}, 1] \in \mathbb{C}^{1\times(n+1)}$ such that

$$[v_1^T, v_2^T]\begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{Y}_{n+1} \end{bmatrix} = 0 \tag{3.9}$$

It suggests that

$$E(z_i)Y_i = F(z_i) \quad \text{for } i = 1, 2\cdots, M \tag{3.10}$$

where $E(z_i) = z_i^n + a_{n-1}z_i^{n-1} + \cdots + a_0$, $F(z_i) = v_1^T\mathbf{U}_{n+1}^i$ and $\mathbf{U}_s^i$ denotes the i-th block column in $\mathbf{U}_s$. It can be easily checked that $F(z_i)$ is also a polynomial of z_i. If there exists a $j \in \{1, 2, \cdots, M\}$ such that $E(z_j) = 0$, according to equation 3.10, $F(z_j)$ must be equal to zero. Thus, $E(z_j)$ and $F(z_j)$ must have a common factor : $(z - z_j)$. Therefore,

$$\begin{aligned} (z - z_j)E^{'}(z_i)Y_i &= (z - z_j)F^{'}(z_i) \quad \text{for } i \neq j \\ E^{'}(z_i)Y_i &= F^{'}(z_i) \qquad\qquad\quad \text{for } i \neq j \end{aligned} \tag{3.11}$$

Note that the order of the the new polynomials $E^{'}(z_i)$ and $F^{'}(z_j)$ is reduced by one compared to the original ones. It implies that there exists a non-zero vector v^T such that

$$v^T\begin{bmatrix} \mathbf{U}_n^{j-} \\ \mathbf{Y}_n^{j-} \end{bmatrix} = 0 \tag{3.12}$$

which is contradictory to our assumption that rank$\mathbf{Y}_n\mathbf{P}_n^{j-} = n$ for $j = 1, 2, \cdots, M$. Hence, $E(z_i) \neq 0$. Consequently,

$$Y_i = \frac{F(z_i)}{E(z_i)} \tag{3.13}$$

Then, Y_i can be interpolated by a rational function. Hence, according to Lemma 1, there will exist matrices (C_0, A_0, B_0, D_0) such that $Y_i = C_0(z_iI - A_0)^{-1}B_0 + D_0$ for $i = 1, 2, \cdots, M$. According to Lemma 3, there will also exist a minimal triple (C, A, B) and D such that$Y_i = C(z_iI - A)^{-1}B + D$. Define $X_i = C(z_iI - A)^{-1}B$, and we can write it into a state-space model

$$\begin{aligned} z_iX_i &= AX_i + B \\ Y_i &= CX_i + D \end{aligned} \tag{3.14}$$

We can form a vector relation by repeatedly using the second equation. Thus,

$$\begin{bmatrix} Y_i \\ z_iY_i \\ \vdots \\ z_i^{s-1}Y_i \end{bmatrix} = O_sX_i + \Gamma_s \begin{bmatrix} I_m \\ z_iI_m \\ \vdots \\ z_i^{s-1}I_m \end{bmatrix} \tag{3.15}$$

where

$$O_s = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{s-1} \end{bmatrix}, \Gamma_s = \begin{bmatrix} D & & & & \\ CB & D & & & \\ CAB & CB & D & & \\ \vdots & \ddots & \ddots & \ddots & \\ CA^{s-2}B & CA^{s-3}B & \cdots & CB & D \end{bmatrix} \tag{3.16}$$

Γ_s is the lower triangular Toeplitz matrix. Assembling the data vector we can get

$$\begin{aligned} \mathbf{Y}_s &= O_s[X_1, X_2, \cdots, X_M] + \Gamma_s\mathbf{U}_s \\ &= O_sX + \Gamma_s\mathbf{U}_s \end{aligned} \tag{3.17}$$

with $X = [X_1, X_2, \cdots, X_M], n \leq s \leq M - n$.
The next lemma is a well known result [15]
Lemma 8. rank$X$=n *if and only if* (A, B) *is controllable.*

Proof. X is rank deficient if and only if there exists a row vector C such that

$$C(z_iI - A)^{-1}B = 0, \text{for } i = 1, 2, \cdots, M \tag{3.18}$$

Because $C(zI - A)^{-1}B$ can only have at most $n - 1$ zeros and $M > n - 1$, it only holds when $C(zI - A)^{-1}B \equiv 0$, which means (A, B) is noncontrollable. □

Here we introduce the lemma given in [16],

Lemma 9. *When* $A \in \mathbb{C}^{n\times n}$, rank $\begin{bmatrix} \mathbf{U}_s \\ X \end{bmatrix} = sm + n$ *if and only if* (A, B) *is controllable.*

Proof. $\begin{bmatrix}\mathbf{U}_s \\ X\end{bmatrix}$ is rank deficient if and only if there exists a row vector

$$[D, E_1, E_2, \cdots, E_{s-1}, C] \neq 0 \tag{3.19}$$

such that

$$[D, E_1, E_2, \cdots, E_{s-1}, C]\begin{bmatrix}\mathbf{U}_s \\ X\end{bmatrix} = 0 \tag{3.20}$$

It follows that

$$H(z_i) = 0 \tag{3.21}$$

where $H(z) = D + C(zI - A)^{-1}B + \sum_{k=1}^{s-1} E_k z^k$. This only holds when i) $H(z)$ has M zeros that are exactly at z_i, but this is not true because $H(z)$ has at most $s + n - 1$ zeros that are less than M, or when ii) $H(z)$=0 for all z, this implies $C(zI - A)^{-1}B = \sum_{i=1}^{+\infty} CA^{i-1}Bz^{-i}$. Recall the Cayley-Hamilton theorem, case ii) is true if and only if (A, B) is noncontrollable. □

This lemma shows the row space of $\mathbf{U}_s$ and X do not have intersection and if $A \in \mathbb{C}^{r\times r}$, then rank$X\mathbf{P}_{n+1} = r$ when (A, B) is controllable. Also, rank$O_{n+1} = r$ when (C, A) is controllable and $A \in \mathbb{C}^{r\times r}$. Because

$$\text{rank}O_{n+1}X\mathbf{P}_{n+1} = \text{rank}\mathbf{Y}_{n+1}\mathbf{P}_{n+1} = n \tag{3.22}$$

this is true only when $r = n$. This implies that there exists a minimal triple (C, A, B) and D where $C \in \mathbb{C}^{p\times n}, A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{n\times m}, D \in \mathbb{C}^{p\times m}$ such that

$$Y_i = C(z_i I_n - A)^{-1}B + D \quad (i = 1, 2, \cdots, M) \tag{3.23}$$

ii) When $Y_i = C(z_i I_n - A)^{-1}B + D$ where (C, A, B) is minimal with order n, then

$$\begin{aligned} z_i X_i &= AX_i + B \\ Y_i &= CX_i + D \end{aligned} \tag{3.24}$$

Repeat the same method we used from equation 3.14 to equation 3.22, we will get

$$\text{rank}\mathbf{Y}_n\mathbf{P}_{n+1} = \text{rank}\mathbf{Y}_{n+1}\mathbf{P}_{n+1} = \text{rank}\mathbf{Y}_n^{j-}\mathbf{P}_n^{j-} = n, \text{ for } j = 1, 2\cdots, M \tag{3.25}$$

when $M \geq 2n + 1$

3.2.0.1 Another perspective

Now theorem 7 is proved, but here we offer another perspective to understand the theorem. If rank$\mathbf{Y}_n\mathbf{P}_{n+1}$=rank$\mathbf{Y}_{n+1}\mathbf{P}_{n+1} = n$, in the same way we will have

$$E(z_i)Y_i = F(z_i) \quad \text{for } i = 1, 2\cdots, M \tag{3.26}$$

If $E(z_i) = 0$ for some z_i, we assume $E(z_i) = 0$ for $z_i = z_1, z_2, \cdots, z_r$ without loss of generality. Note $r \leq n$ from the fact that the largest power of z in $E(z)$ is n. We

know from equation 3.10 that $F(z_i)$ must also be zero if $E_1(z_i)$=0. Hence, $E(z)$ and $F(z)$ must have at least r common roots. Thus

$$Y_i = \frac{F(z_i)}{E(z_i)} = \frac{\tilde{F}(z_i)(z_i - z_1)^{j_1} \cdots (z_i - z_r)^{j_r}}{\tilde{E}(z_i)(z_i - z_1)^{j_1} \cdots (z_i - z_r)^{j_r}} = \frac{\tilde{F}(z_i)}{\tilde{E}(z_i)} \quad \text{for } i = r+1, \cdots, M. \tag{3.27}$$

where $j_1, \cdots, j_r$ represent the multiplicity. Next we show that $j_1, \cdots, j_r$ must be equal to one and $\tilde{F}(z)$ must be relative prime to $\tilde{E}(z)$. If $j_1 \geq 2$, then we can easily check that

$$\tilde{E}(z_i)(z_i - z_1)^{j_1 - 1} \cdots (z_i - z_r)^{j_r} Y_i = \tilde{F}(z_i)(z_i - z_1)^{j_1 - 1} \cdots (z_i - z_r)^{j_r} \tag{3.28}$$

for all $i = 1, \cdots, M$. Note that the order of the the new polynomials in both sides of the equation is reduced by one compared to the original ones. This shows that there will exist a vector v^T such that

$$v^T \begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{Y}_n \end{bmatrix} = 0 \tag{3.29}$$

which is contradictory to our assumption. Therefore, $j_1 = j_2 = \cdots = j_r = 1$.
If $\tilde{F}(z)$ is not relative prime to $\tilde{E}(z)$, then there will exist at least a common root, denoted by z_0, after manipulation we can get

$$\begin{aligned} \tilde{E}(z_i)(z_i - z_1) \cdots (z_i - z_r) Y_i &= \tilde{F}(z_i)(z_i - z_1) \cdots (z_i - z_r) \\ (z - z_0)\tilde{E}'(z_i)(z_i - z_1) \cdots (z_i - z_r) Y_i &= (z - z_0)\tilde{F}'(z_i)(z_i - z_1) \cdots (z_i - z_r) \\ \tilde{E}'(z_i)(z_i - z_1) \cdots (z_i - z_r) Y_i &= \tilde{F}'(z_i)(z_i - z_1) \cdots (z_i - z_r) \end{aligned} \tag{3.30}$$

Note that the order of the the new polynomials in both sides of the equation is reduced by one compared to the original ones. This also shows that there will exist a vector v^T such that

$$v^T \begin{bmatrix} \mathbf{U}_{n+1} \\ \mathbf{Y}_n \end{bmatrix} = 0 \tag{3.31}$$

which is also contradictory to our assumption.
Therefore, the degree of $\tilde{E}(z)$ will be $n - r$, that is to say, $Y_{r+1}, \cdots, Y_M$ can be interpolated by a minimal triple (C, A, B) and D with order $n - r$. By adding $\text{rank}\mathbf{Y}_n^{j-}\mathbf{P}_n^{j-} = n$, we can make sure that $r = 0$.

3.3 Signal model

This part is based on [17]. For linear array antenna with N elementes, when there is only one target, if the signal received at the first element is $y(1) = \alpha$, the signal detected at element k will be $y(k) = \alpha e^{j\omega(k-1)} + n(k),\ k = 1, 2 \cdots, N$, where $n(k)$ represents the noise at element k. When there are n distinct targets, the signal at element k will be

$$y(k) = \sum_{i=1}^{n} \alpha_i e^{j\omega_i(k-1)} + n(k) \tag{3.32}$$

where $\omega_1, \omega_2, \cdots, \omega_n$ are all distinct.

First consider the noise free scenario, define

$$\begin{aligned} x_0 &= [\alpha_1, \alpha_2, \cdots, \alpha_n]^T, \quad c = [1, 1, \cdots, 1] \\ A &= \text{diag}\left(e^{j\omega_1}, e^{j\omega_2}, \cdots, e^{j\omega_n}\right) \end{aligned} \tag{3.33}$$

where A is a diagonal matrix. Then,

$$y(k) = cA^{k-1}x_0 \tag{3.34}$$

The representation of $y(k)$ given by the triple (c, A, x_0) is called a realization and not unique. For any nonsingular matrix $T \in \mathbb{C}^{n\times n}$, denote $\tilde{A} = T^{-1}AT, \tilde{c} = cT, \tilde{x}_0 = T^{-1}x_0$, , and then the triple $(\tilde{A}, \tilde{c}, \tilde{x}_0)$ also realizes $y(k)$. Note that $\tilde{A}$ has the same eigenvalues as A.
Define *state* vector $x(k) = A^{k-1}x_0, k = 1, 2, \cdots, N$, and hence we can get a $state - space\ model$

$$\begin{aligned} x(k+1) &= Ax(k) \\ y(k) &= cx(k) \end{aligned} \tag{3.35}$$

Consider the modified model, for $t = 1, 2, 3, \cdots$

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= cx(t) \end{aligned} \tag{3.36}$$

where the introduced *input* signal $u(t)$ is defined as

$$u(t) = \begin{cases} 1, & t = mN, \quad m = 1, 2, \cdots \\ 0, & \text{otherwise} \end{cases} \tag{3.37}$$

which is an N-periodic signal. Set $B = (I - A^N)x_0$, it follows that $x(mN + 1) = x(1), m = 1, 2, \cdots$, and then $y(t)$ will also be N-periodic signal. Because the modified output $y(t)$ is the same as the original output $y(k)$ in the observed interval $t = 1, 2, \cdots, N$, two models are both valid for the observed signal.
Apply the N-point Discrete Fourier Transform to the modified model, the modified signal model in frequency domain will be

$$\begin{aligned} z_i X_i &= AX_i + BU_i \\ Y_i &= CX_i \end{aligned} \tag{3.38}$$

where $z_i = U_i = e^{j\frac{2\pi i}{N}}$. We can form a vector relation by repeatedly using the second equation. Thus,

$$\begin{bmatrix} Y_i \\ z_i Y_i \\ \vdots \\ z_i^{s-1} Y_i \end{bmatrix} = O_s X_i + \Gamma_s \begin{bmatrix} U_i \\ z_i U_i \\ \vdots \\ z_i^{s-1} U_i \end{bmatrix} \tag{3.39}$$

where

$$O_s = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{s-1} \end{bmatrix}, \Gamma_s = \begin{bmatrix} 0 & & & & \\ CB & 0 & & & \\ CAB & CB & 0 & & \\ \vdots & \ddots & \ddots & \ddots & \\ CA^{s-2}B & CA^{s-3}B & \cdots & CB & 0 \end{bmatrix} \tag{3.40}$$

Γ_s is the lower triangular Toeplitz matrix. Assembling the data vector we can get

$$\begin{aligned}\mathbf{Y}_s &= O_s[X_1, X_2, \cdots, X_N] + \Gamma_s \mathbf{U}_s \\ &= O_s X + \Gamma_s \mathbf{U}_s\end{aligned} \tag{3.41}$$

with $X = [X_1, X_2, \cdots, X_N], n \leq s \leq N - n$, and

$$\mathbf{Y}_s = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_N \\ z_1 Y_1 & z_2 Y_2 & \cdots & z_N Y_N \\ z_1^2 Y_1 & z_2^2 Y_2 & \cdots & z_N^2 Y_N \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{s-1} Y_1 & z_2^{s-1} Y_2 & \cdots & z_N^{s-1} Y_N \end{bmatrix} \in \mathbb{C}^{s \times N} \tag{3.42}$$

$$\mathbf{U}_s = \begin{bmatrix} U_1 & U_2 & \cdots & U_N \\ z_1 U_1 & z_2 U_2 & \cdots & z_N U_N \\ z_1^2 I U_1 & z_2^2 U_2 & \cdots & z_N^2 U_N \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{s-1} U_1 & z_2^{s-1} U_2 & \cdots & z_N^{s-1} U_N \end{bmatrix} \in \mathbb{C}^{s \times N} \tag{3.43}$$

Define the projection matrix

$$\Pi_s^{\perp} = I - \mathbf{U}_s^*(\mathbf{U}_s \mathbf{U}_s^*)^{-1} \mathbf{U}_s \tag{3.44}$$

which projects onto the nullspace of $\mathbf{U}_s$. Then it follows from lemma 9 that,

$$\operatorname{rank}(\mathbf{Y}_s \Pi_s^{\perp}) = n \text{ or } \operatorname{rank} \begin{bmatrix} \mathbf{U}_s \\ \mathbf{Y}_s \end{bmatrix} = n \tag{3.45}$$

3.4 Subspace algorithm

In the noisy case, $\mathbf{Y}_s \Pi_s^{\perp}$ will have full rank, so in the paper [17], an approximation method to estimate the diagonal matrix A through subspace algorithm was put forward.

Because $\mathbf{Y}_s \Pi_s^{\perp} = O_s X \Pi_s^{\perp}$, the range space of O_s is determined by factoring $\mathbf{Y}_s \Pi_s^{\perp}$ into two rank n matrices corresponding to O_s and $X\Pi_s^{\perp}$ respectively. In the noisy case $\mathbf{Y}_s \Pi_s^{\perp}$ will have full rank and a matrix is still be determined from $\mathbf{Y}_s \Pi_s^{\perp}$ to estimate O_s.The Singular Value Decomposition (SVD) gives the desired factorization

$$\mathbf{Y}_s \Pi_s^{\perp} = \begin{bmatrix} Z_s & Z_o \end{bmatrix} \begin{bmatrix} \Sigma_s & 0 \\ 0 & \Sigma_o \end{bmatrix} \begin{bmatrix} V_s^* \\ V_o^* \end{bmatrix} \tag{3.46}$$

where Σ_s denotes the n largest singular values. As an estimation of the range space of the observability matrix O_s we simply take Z_s. In the noise free case, the range space is exactly recovered. The extended-observability matrix Z_s has a special block row structure, where each new block-row is the previous block-row multiplied by A.

Using the shift structure of the observability matrix, a matrix $\tilde{A}$ is determined by minimizing the Frobenius norm

$$\min_A \left\| \underline{Z_s} A - \overline{Z_s} \right\|_F^2 \tag{3.47}$$

where $\underline{Z_s}$ and $\overline{Z_s}$ denotes the $(s-1)$ first rows ans last rows respectively. After we get $\tilde{A}$ we can simply recover the estimated angles by computing the eigenvalues of $\tilde{A}$. In the noise-free case, the rank of Z_s will be n and the minimum of equation 3.47 will be unique.

3.5 ADMM algorithm

In the noisy case $\mathbf{Y}_s\Pi_s^{\perp}$ has full rank , and then our idea is to find a matrix $\tilde{\mathbf{Y}}_s$ that is as close as possible to $\mathbf{Y}_s$ with the same structure as $\mathbf{Y}_s$ while rank($\tilde{\mathbf{Y}}_s\Pi_s^{\perp}$)=$n$, so we can recover the estimation of data from $\tilde{\mathbf{Y}}_s$. To be more specific, we consider the problem where we look for a *structured* matrix of given rank as close as possible to a given matrix, or the low rank structured matrix approximation problem.

Here we define structured to mean that matrix is an affine combination of a set of basis matrices

$$A(x) = A_0 + \sum_{i=1}^{n_A} x_i A_i \tag{3.48}$$

where x is a vector which parametrizes the structured matrix. The expression above can be used to define the set of admissible structured matrices $\mathcal{A}$ in the following way

$$\mathcal{A} = \{A(x) \mid x \in \mathbb{R}^{n_A}\} \tag{3.49}$$

Examples of structured matrices of this form are e.g. Hankel and Toeplitz matrices. Note that the set $\mathcal{A}$ is convex by this construction.
The low rank structured matrix approximation problem can be defined as

$$\min_A \|A - A_0\|_F^2 \quad \text{subject to} \quad \text{rank}A = k \text{ and } A \in \mathcal{A} \tag{3.50}$$

Let the function h encode the structural constraint as

$$h(A) = \begin{cases} 0, & A \in \mathcal{A} \\ \infty, & A \notin \mathcal{A} \end{cases} \tag{3.51}$$

The proximal operator for h is then

$$\mathbf{prox}_{\lambda h}(Z) = \arg\min_A h(A) + \frac{1}{2\lambda} \|A - Z\|_F^2 \tag{3.52}$$

Define

$$x^* = \arg\min_x \|A(x) - Z\|_F^2 \tag{3.53}$$

Then $\mathbf{prox}_{\lambda h} = A(x^*)$. The proximal operator is the orthogonal projection of Z on the set $\mathcal{A}$ and can be obtained by Least-square problems.
Now we need to find a convex function f that can encode the left information that

$$\min_A \|A - A_0\|_F^2 \quad \text{subject to} \quad \text{rank}A = k \tag{3.54}$$

In order to find f, first we introduce the Von Neumann trace inequality.

3.5.1 Von Neumann trace inequality

The following result is known as Von Neumanns trace inequality.

Lemma 10. *Let $A,B \in \mathbb{C}^{n\times m}$ and $q= \min(m,n)$. Assume $A=\sum_{i=1}^{q}\sigma_i(A)u_iv_i^H$, where $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_q(A) \geq 0$. Let $\sigma_i(B)$ be the corresponding ordered singular values of B. Then,*

$$\mathrm{tr}(A^HB) \leq \sum_{i=1}^{q}\sigma_i(A)\sigma_i(B) \tag{3.55}$$

with the equality if $B =\sum_{i=1}^{q}\sigma_i(B)u_iv_i^H$.

Proof. According to Theorem 3.3.13 and Theorem 3.3.14 in [18], we know that

$$\mathrm{tr}(A^HB) \leq \sum_{i=1}^{q}\sigma_i(A^HB) \leq \sum_{i=1}^{q}\sigma_i(A)\sigma_i(B) \tag{3.56}$$

□

From this result we can derive the following theorem

Theorem 8. *Let $A = U\Sigma_AV^T$ be the singular value decomposition. Denote Σ_A and Σ_B as diagonal matrices with the singular values of A and B respectively ordered in non-increasing order and let f be any real valued functions. Then,*

$$\min_B f(\Sigma_B) + \|A-B\|_F^2 = \min_{\Sigma_B} f(\Sigma_B) + \mathrm{tr}(\Sigma_A-\Sigma_B)^2 \tag{3.57}$$

Denote the miniminzing argument to the right hand expression by Σ_{B^*}. It follows that the minimizing argument of the left hand expression is $B^* = U\Sigma_{B^*}V^T$

Proof. Since

$$\begin{aligned}
\|A-B\|_F^2 &= \mathrm{tr}(A^TA) - 2(A^TB) + \mathrm{tr}(B^TB)\\
&= \mathrm{tr}\Sigma_A^2 - 2\mathrm{tr}(A^TB) + \mathrm{tr}\Sigma_B^2\\
&\geq \mathrm{tr}\Sigma_A^2 - 2\mathrm{tr}(\Sigma_A\Sigma_B) + \mathrm{tr}\Sigma_B^2\\
&= \mathrm{tr}(\Sigma_A-\Sigma_B)^2\\
&\geq \mathrm{tr}(\Sigma_A-\Sigma_{B^*})^2\\
&= \left\|U(\Sigma_A-\Sigma_{B^*})V^T\right\|_F^2\\
&= \|A-B^*\|_F^2
\end{aligned} \tag{3.58}$$

this yields that

$$\|A-B\|_F^2 \geq \|A-B^*\|_F^2 = \mathrm{tr}(\Sigma_A-\Sigma_{B^*})^2 \tag{3.59}$$

□

3.5.2 Nuclear norm approximation

First we obtain f by using a Lagrange formulation

$$\min_A \|A - A_0\|_F^2 + \gamma \text{rank} A \tag{3.60}$$

where γ is a parameter that control the rank of the solution, and we can get a desired rank k by choosing a suitable γ^*. Since a rank function is not convex, we need to seek a convex function as a substitute. We can find it by deriving the Fenchel bi-conjugate to the rank function restricted to the domain $\mathcal{A} = \{A|\sigma_1(A) \leq 1\}$. Set $g(A) = \text{rank}A$. The conjugate is

$$\begin{aligned} g^*(Z) &= \max_{A\in\mathcal{A}} < A, Z > -\text{rank}A \\ &= \max_{A\in\mathcal{A}} \text{tr}(A^T Z) - \text{rank}A \\ &\leq \max_{A\in\mathcal{A}} \sum_i (\sigma_i(A)\sigma_i(Z)) - \text{rank}A \\ &= \sum_i (\sigma_i(Z) - 1)_+ \end{aligned} \tag{3.61}$$

where $(x)_+ = \max(x, 0)$. Then the bi-conjugate follows as

$$\begin{aligned} g^{**}(A) &= \max_Z \text{tr}(A^T Z) - \sum_i (\sigma_i(Z) - 1)_+ \\ &= \max_Z - \sum_i (\sigma_i(A)\sigma_i(Z) - (\sigma_i(Z) - 1)_+) \\ &= \sum_i \sigma_i(A) \end{aligned} \tag{3.62}$$

where the sum of singular values are called the nuclear norm denoted by $\|A\|_*$. In conclusion, the nuclear norm $\|A\|_*$ is the convex envelope to the rank$A$. If the domain is selected to $\sigma_i(A) \leq 1/\alpha$, the convex envelope changes to $\alpha \|A\|_*$. Therefore, we can define the function $f(A)$ as

$$f(A) = \|A - A_0\|_F^2 + \gamma \|A\|_* \tag{3.63}$$

Then the proximal operator for $f(A)$ is determined as

$$\begin{aligned} \textbf{prox}_{\lambda f}(Z) &= \arg\min_A \|A - A_0\|_F^2 + \gamma \|A\|_* + \frac{1}{2\lambda} \|A - Z\|_F^2 \\ &= \arg\min_A (1 + \frac{1}{2\lambda}) \|A\|_F^2 - 2\text{tr}(A^H(A_0 + \frac{1}{2\lambda}Z)) + \gamma \|A\|_* \\ &= \arg\min_A \|A\|_F^2 - 2\text{tr}(A^H \frac{2\lambda}{1+\lambda}(A_0 + \frac{1}{2\lambda}Z)) + \frac{2\lambda\gamma}{1+2\lambda} \|A\|_* \\ &= \arg\min_A \left\|A - \tilde{Z}\right\|_F^2 + 2\mathcal{K} \|A\|_* \end{aligned} \tag{3.64}$$

where $\tilde{Z} = \frac{1}{1+2\lambda}(2\lambda A_0 + Z)$ and $\mathcal{K} = \frac{\lambda\gamma}{1+2\lambda}$. Let $\tilde{Z} = U\Sigma_{\tilde{Z}}V^H$ be the SVD. According to Theorem 8, the minimizing A has the structure $A = U\Sigma_A V^H$ where u and V are

the left and right singular vectors to $\tilde{Z}$. Hence, the minimization can be recast as

$$\begin{aligned}&\min_{\Sigma_A \geq 0} \left\| U\Sigma_A V^H - U\Sigma_{\tilde{Z}} V^H \right\|_F^2 + 2\mathcal{K}\mathrm{tr}\Sigma_A \\ &= \min_{\sigma(A)\geq 0} \sum_i \Big((\sigma_i(A) - \sigma_i(\tilde{Z}))^2 + 2\mathcal{K}\sigma_i(A) \Big)\end{aligned} \tag{3.65}$$

The minimum is attained for $\sigma_i(A) = (\sigma_i(\tilde{Z}) - \mathcal{K})_+$ and we obtain

$$\mathbf{prox}_{\lambda f}(Z) = U(\Sigma_{\tilde{Z}} - \mathcal{K}I)_+ V^H \tag{3.66}$$

3.5.2.1 Alternative formulation

If we use the alternative formulation

$$f(A) = \frac{\rho}{2} \|A - A_0\|_F^2 + \|A\|_* \tag{3.67}$$

the proximal operator is then given by

$$\mathbf{prox}_{\lambda f}(Z) = U(\Sigma_{\tilde{Z}} - \mathcal{K}I)_+ V^H \tag{3.68}$$

where $\tilde{Z} = \frac{1}{1+\lambda\rho}(\lambda\rho A_0 + Z) = U\tilde{\Sigma}V^H$ and $\mathcal{K} = \frac{\lambda}{1+\lambda\rho}$

Proof.

$$\begin{aligned}\mathbf{prox}_{\lambda f}(Z) &= \arg\min_A \frac{\rho}{2} \|A - A_0\|_F^2 + \frac{1}{2\lambda} \|A - Z\|_F^2 + \|A\|_* \\ &= \arg\min_A (\frac{\rho}{2} + \frac{1}{2\lambda}) \|A\|_F^2 - 2\mathrm{tr}(A^H(\frac{\rho}{2}A_0 + \frac{1}{2\lambda}Z)) + \|A\|_* \\ &= \arg\min_A \frac{1+\lambda\rho}{2\lambda} \|A\|_F^2 - 2\mathrm{tr}(A^H(\frac{\rho}{2}A_0 + \frac{1}{2\lambda}Z)) + \|A\|_* \\ &= \arg\min_A \left\| A - \frac{2\lambda}{1+\lambda\rho}(\frac{\rho}{2}A_0 + \frac{1}{2\lambda}Z) \right\|_F^2 + \frac{2\lambda}{1+\lambda\rho} \|A\|_*\end{aligned} \tag{3.69}$$

□

3.5.2.2 Drawbacks of nuclear norm approximation

As derived above, the nuclear norm is the convex envelop to the rank function and together with the Lagrange relaxation this could be one method to approximation the original problem. However, two issues exists with this approach

1. The nuclear norm will put a penalty on all singular values of the solution and hence the solution will be biased towards zero for any non-zero γ.
2. The Lagrange multiplier γ is unknow and must be searched for in an external loop which adds to the computation complexity.

In what follows we will discuss an alternative formulation which has been proposed in [19] and [20] .

3.5.3 Convex envelope

This section is based on [20]. Revisit the original optimization problem

$$\min_A \|A - A_0\|_F^2 \quad \text{s.t.} \quad \text{rank}A = k, \text{and } A \in \mathcal{A} \tag{3.70}$$

and reformulate the hard constraint on the rank by introducing the function

$$g(A) = \begin{cases} 0, & \text{rank}A \leq k \\ \infty, & \text{rank}A > k \end{cases} \tag{3.71}$$

and add it to the objective function

$$\min_A \|A - A_0\|_F^2 + g(A) \quad \text{s.t.} \quad A \in \mathcal{A} \tag{3.72}$$

As this objective function is not convex, we seek a convex approximation through conjugate. Consider

$$\min_A \|A - A_0\|_F^2 + g(A) \triangleq \min_A f(A) \tag{3.73}$$

The conjugate is

$$\begin{aligned} f^*(Y) &= \sup_A \text{tr}(A^T Y) - f(A) \\ &= \sup_A \text{tr}(A^T Y) - \|A - A_0\|_F^2 - g(A) \\ &= \sup_A \text{tr}(A^T Y) - \|A\|_F^2 + 2\text{tr}(A^T A_0) - \|A_0\|_F^2 - g(A) \\ &= \sup_A - \|A\|_F^2 + 2\text{tr}(A^T(\frac{1}{2}Y + A_0)) - \|A_0\|_F^2 - g(A) \\ &= \sup_A - \left\|A - (\frac{1}{2}Y + A_0)\right\|_F^2 + \left\|\frac{1}{2}Y\right\|_F^2 + \text{tr}(Y^T A_0) - g(A) \\ &= \sup_A - \left\|A - (\frac{1}{2}Y + A_0)\right\|_F^2 + \left\|\frac{1}{2}Y + A_0\right\|_F^2 - \|A_0\|_F^2 - g(A) \\ &= \sup_A - \|A - Z\|_F^2 + \|Z\|_F^2 - \|A_0\|_F^2 - g(A) \end{aligned} \tag{3.74}$$

where $Z = \frac{1}{2}Y + A_0$, By Von Neumann we have

$$f^*(Y) = \sup_{\Sigma_A} - \sum_{i=1}^n (\sigma_i(A) - \sigma_i(Z))^2 + \sum_{i=1}^n \sigma_i^2(Z) - \|A_0\|_F^2 - g(\Sigma_A) \tag{3.75}$$

Due to the term $g(\Sigma_A)$ the matrix A is forced to have rank k and the optimal A is given by the rank k truncated singular value decomposition of Z. This yields $\sigma_i(A) = \sigma_i(Z)$ for $i = 1, 2, \cdots, k$.

$$\begin{aligned} f^*(Y) &= -\sum_{i=k+1}^n \sigma_i^2(Z) + \sum_{i=1}^n \sigma_i^2(Z) - \|A_0\|_F^2 \\ &= \sum_{i=1}^k \sigma_i^2(Z) - \|A_0\|_F^2 \\ &= -\sum_{i=k+1}^n \sigma_i^2(Z) + \|Z\|_F^2 - \|A_0\|_F^2 \end{aligned} \tag{3.76}$$

The bi-conjugate is derived as

$$\begin{aligned} f^{**}(A) &= \sup_Y \operatorname{tr}(A^T Y) + \sum_{i=k+1}^{n} \sigma_i^2(Z) - \|Z\|_F^2 + \|A_0\|_F^2 \\ &= \sup_Z 2\operatorname{tr}(A^T(Z - A_0)) + \sum_{i=k+1}^{n} \sigma_i^2(Z) - \|Z\|_F^2 + \|A_0\|_F^2 \\ &= \sup_Z \sum_{i=k+1}^{n} \sigma_i^2(Z) - \|Z - A\|_F^2 + \|A - A_0\|_F^2 \\ &= \sup_Z \left(\sum_{i=k+1}^{n} \sigma_i^2(Z) - \sum_{i=1}^{n} (\sigma_i(Z) - \sigma_i(Z))^2 \right) + \|A - A_0\|_F^2 \end{aligned} \tag{3.77}$$

3.5.3.1 The proximal operator

The Fenchel bi-conjugate function was shown to be,

$$f^{**}(A) = \sup_Z \left(\sum_{i=k+1}^{n} \sigma_i^2(Z) - \|Z - A\|_F^2 \right) + \|A - A_0\|_F^2 \tag{3.78}$$

We will use ADMM to minimize objective functions containing f^{**} subject to convex constraints. Hence, we need to derive the proximal operator for f^{**}. The proximal operator of f^{**} is given by

$$\begin{aligned} \mathbf{prox}_{\lambda f^{**}}(M) = \arg\min_A \min_Z \left(\sum_{i=k+1}^{n} \sigma_i^2(Z) - \|Z - A\|_F^2 \right) + \\ \|A - A_0\|_F^2 + \frac{1}{2\lambda} \|A - M\|_F^2 \end{aligned} \tag{3.79}$$

If we swap the minimization and maximation, we obtain that the optimal A is given by $A_* = M + 2\lambda(A_0 - Z)$. To find the final form of the proximal operator, the maximization w.r.t. Z needs to be completed. Inserting the optimal A into the

$\mathbf{prox}_{\lambda f^{**}}(M)$ yields

$$\begin{aligned}
&\max_Z \sum_{i=k+1}^{n} \sigma_i^2(Z) - \|M + 2\lambda(A_0 - Z) - Z\|_F^2 + \|M + 2\lambda(A_0 - Z) - A_0\|_F^2 \\
&\qquad + \frac{1}{2\lambda}\|M - 2\lambda(A_0 - Z) - M\|_F^2 \\
&= \max_Z \sum_{i=k+1}^{n} \sigma_i^2(Z) - \|M + 2\lambda(A_0 - Z) - Z\|_F^2 + \|M + 2\lambda(A_0 - Z) - A_0\|_F^2 \\
&\qquad + 2\lambda\|A_0 - Z\|_F^2 \\
&= \max_Z \sum_{i=k+1}^{n} \sigma_i^2(Z) - 2\mathrm{tr}[M^T(2\lambda(A_0 - Z) - Z)] + 2\mathrm{tr}[M^T(2\lambda(A_0 - Z) - A_0)] \\
&\qquad - \|2\lambda(A_0 - Z) - Z\|_F^2 + \|2\lambda(A_0 - Z) - A_0\|_F^2 + 2\lambda\|A_0 - Z\|_F^2 \\
&= \max_Z \sum_{i=k+1}^{n} \sigma_i^2(Z) + 2\mathrm{tr}[M^T(Z - A_0)] - 4\lambda\mathrm{tr}[A_0^T(A_0 - Z)] + 4\lambda\mathrm{tr}[Z^T(A_0 - Z) \\
&\qquad + \|A_0\|_F^2 - \|Z\|_F^2 + 2\lambda\|A_0 - Z\|_F^2 \\
&= \max_Z \sum_{i=k+1}^{n} \sigma_i^2(Z) + 2\mathrm{tr}(M^T Z) - 4\lambda\mathrm{tr}[(A_0 - Z)^T(A_0 - Z)] + 2\lambda\|A_0 - Z\|_F^2 \\
&\qquad - \|Z\|_F^2 \\
&= \max_Z \sum_{i=k+1}^{n} \sigma_i^2(Z) + 2\mathrm{tr}(M^T Z) - 2\lambda\|A_0 - Z\|_F^2 - \|Z\|_F^2 \\
&= \max_Z \sum_{i=k+1}^{n} \sigma_i^2(Z) - (1 + 2\lambda)\|Z\|_F^2 + 4\lambda\mathrm{tr}(A_0^T Z) + 2\mathrm{tr}(M^T Z) \\
&= \max_Z \sum_{i=k+1}^{n} \sigma_i^2(Z) - (1 + 2\lambda)\left\|Z - \frac{M + 2\lambda A_0}{1 + 2\lambda}\right\|_F^2
\end{aligned} \tag{3.80}$$

Define $Y \triangleq \frac{M+2\lambda A_0}{1+2\lambda}$ with SVD $Y \triangleq U\Sigma_Y V^T$. Then by Theorem 8 any optimal Z has the form $Z = U\Sigma_Z V^T$ and we need to solve

$$\max_{\Sigma_Z} \sum_{i=k+1}^{n} \sigma_i^2(Z) - (1 + 2\lambda)\sum_{i=1}^{n}(\sigma_i(Z) - \sigma_i(Y))^2 \quad \text{s.t. } \sigma_i(Z) \geq 0\ , \sigma_i(Z) \geq \sigma_{i+1}(Z) \tag{3.81}$$

Define

$$\begin{aligned}
f_i(s) &= -(1 + 2\lambda)(s - \sigma_i(Y))^2 \quad \text{for } i = 1, 2, \cdots, k. \\
f_i(s) &= s^2 - (1 + 2\lambda)(s - \sigma_i(Y))^2 \quad \text{for } i = k + 1, \cdots, n. \\
\boldsymbol{\sigma} &= [\ \sigma_1, \sigma_2, \cdots, \sigma_n\ ]
\end{aligned} \tag{3.82}$$

Then, what we need to solve will be

$$\max_{\boldsymbol{\sigma}} \sum_{i=1}^{n} f_i(\sigma_i) \quad \text{s.t. } \sigma_i \geq 0\ , \sigma_i \geq \sigma_{i+1} \tag{3.83}$$

Note the maximizer s_i of $f_i(s)$ is $s_i = \sigma_i(Y)$ for $i = 1, \cdots, k$ and $s_i = \frac{1+2\lambda}{2\lambda}\sigma_i(Y)$ for $i = k + 1, \cdots, n$. It follows that

$$s_1 \geq s_2 \geq \cdots \geq s_k \quad \text{and} \quad s_{k+1} \geq \cdots \geq s_n \tag{3.84}$$

Hence, if $s_k \geq s_{k+1}$, the problem will be solved with $\sigma_i(Z) = s_i$.
When $s_k \leq s_{k+1}$, here we introduce the theorem which is given in [20, Appendix 2].

Theorem 9. *The maximizer* $\boldsymbol{\sigma}$ *can be written as*

$$\begin{aligned} \sigma_i &= max(s_i, s^*), \quad i = 1, \cdots, k \\ \sigma_i &= min(s_i, s^*), \quad i = k+1, \cdots, n \end{aligned} \tag{3.85}$$

where s^* *fulfills* $s_k \leq s^* \leq s_{k+1}$.

Proof. See [20, Appendix 2] □

As for how to determine s^*, we input σ_i to $f(\sigma_i)$ according to this theorem. Then,

$$\begin{aligned} &f_i(s^*) = -(1+2\lambda)[s^* - \sigma_i(Y)]_+, \text{ for } i = 1, \cdots, k. \\ &f_i(s^*) = -2\lambda[(1+\frac{1}{2\lambda})\sigma_i(Y) - s^*]_+^2 + (1+\frac{1}{2\lambda})\sigma_i^2(Y), \text{ for } i = k+1, \cdots, n. \end{aligned} \tag{3.86}$$

That is to say, the original problem 3.83 becomes

$$\max_{s^*} \sum_{i=1}^{n} f_i(s^*) \quad \text{s.t.} \quad s_k \leq s^* \leq s_{k+1} \tag{3.87}$$

Note that $f_i(s^*)$ is concave and differentiable everywhere for all $i = 1, 2 \cdots, n$. Therefore, this maximization problem can be easily solved through differential. Denote the maximizer of this problem is s_0. If $s_k \leq s_0 \leq s_{k+1}$, then the problem is solved with $s^* = s_0$. Because the functions are concave, and then $s^* = s_k$ if $s_0 \leq s_k$ with $s^* = s_{k+1}$ if $s_{k+1} \leq s_0$.
When the optimal Σ_{Z_*} is found we have $Z_* = U\Sigma_{Z_*}V^T$ and

$$\mathbf{prox}_{\lambda f^{**}}(M) = A_* = M + 2\lambda(A_0 - Z_*) \tag{3.88}$$

In conclusion, this ADMM algorithm we obtained for low rank structured matrix approximation is:

3.5.3.2 Algorithm implementation

Data: $A_0, n, \mathcal{A}$
Initialize **X, Z, U**
while 1 do
  $\mathbf{X} = \mathbf{prox}_{\lambda f^{**}}(\mathbf{Z} - \mathbf{U})$;
  $\mathbf{Z} = \mathbf{prox}_{\lambda h}(\mathbf{X} + \mathbf{U})$;
  $\mathbf{U} = \mathbf{U} + \mathbf{Z} - \mathbf{X}$;
  if $\|\mathbf{X} - \mathbf{Z}\|_F^2 < threshold$ **then**
    break ;
  end
end

Algorithm 1: ADMM algorithm

4

The calibration algorithm and result analysis

This chapter is devoted to the analysis of calibration algorithm and results.

4.1 The calibration algorithm

Assume we have one linear array antenna with N elements and the number of unknown targets is known as n. Define time domain data $\mathbf{y} = [y_1, y_2, \cdots, y_N]$. With the same notation we used in section 3.3 and we collect m snapshots in the same element, then the signal received at element k with multiple snapshots will be

$$\begin{aligned} y_k &= cA^{k-1}[x_0, x_1, \cdots, x_{m-1}] + n_k \\ &= cA^{k-1}\mathbf{x} + n_k \end{aligned} \tag{4.1}$$

where $\mathbf{x} = [x_0, x_1, \cdots, x_{m-1}] \in \mathbb{C}^{1\times m}$. Define the non-zero unknown gain at the k-th antenna element as g_k, for $k = 1, 2\cdots, N$. Then, the received signal is

$$y_k = g_k cA^{k-1}\mathbf{x} + n_k \tag{4.2}$$

Our aim is to estimate the angles of targets as well as the unknown gains g_k simultaneously. Calibration means we calibrate the signal $y(k)$ with a scalar to mitigate the affect of unknown gains. The calibrated output in time domain is defined by

$$y_k^c = h_k y_k \tag{4.3}$$

where $h_k \in \mathbb{C}$ compensates the deviations from ideal unit gain. We want to find a $\mathbf{h} = [h_1, h_2, \cdots, h_N]$ such that $\text{rank}\mathbf{Y}_s^c(\mathbf{h})\mathbf{P}_s = n$ and minimize $\|\mathbf{Y}_s^c(\mathbf{h}) - \mathbf{Y}_s\|_F^2$, where $\mathbf{Y}_s^c(\mathbf{h})$ denotes the matrix constructed from the frequency domain calibration output $\mathbf{Y}^c = [Y_1^c, Y_2^c, \cdots, Y_N^c]$ as shown in equation 3.1 . In the noise free case, accroding to [21, Theorem 3], the relation between g_k and h_k is

$$h_k = g_k^{-1}\beta^{k-1}, \quad \beta \in \mathbb{C} \text{ and } \beta \neq 0 \tag{4.4}$$

β is introduced due to the fact that

$$\begin{aligned} y_k^c &= h_k y_k \\ &= g_k^{-1}\beta^k g_k cA^{k-1}\mathbf{x} \\ &= c(\beta A)^{k-1}\mathbf{x} \\ &= c\tilde{A}^{k-1}\mathbf{x} \end{aligned} \tag{4.5}$$

Hence, the eigenvalues of $\tilde{A}$ will be

$$\lambda_i = \beta e^{j\omega_i} \tag{4.6}$$

where ω_i is the angle of i-th target. Without loss of generality, the common assumption is that $g_1 = 1, \gamma = g_2/g_1 = g_2$, seen in [4]. Then, $\beta = g_2 h_2 = \gamma h_2$.
Now the problem is to solve this optimaztion problem

$$\min_{\mathbf{h}} \|\mathbf{Y}_s^c(\mathbf{h}) - \mathbf{Y}_s\|_F^2 \quad \text{s.t.} \quad \text{rank}\mathbf{Y}_s^c(\mathbf{h})\mathbf{P}_s = n,\ h_1 = 1 \tag{4.7}$$

Note that $\text{rank}\mathbf{Y}_s^c(\mathbf{h})\mathbf{P}_s = n$ is equivalent to $\text{rank}\begin{bmatrix}\mathbf{U}_s \\ \mathbf{Y}_s^c(\mathbf{h})\end{bmatrix} = sm + n$. We construct the $\mathbf{Y}_s$ according to the frequency domain raw data $\mathbf{Y} = [Y_1, Y_2, \cdots, Y_N]$ as shown in equation 3.1, so here we define the set $\mathcal{A} = \{\begin{bmatrix}\mathbf{U}_s \\ \mathbf{Y}_s\end{bmatrix}, \text{for } \mathbf{Y} \in \mathbb{C}^{1\times mN}\}$. It can easily be check that this set $\mathcal{A}$ is a convex set.

4.1.0.1 Algorithm

1. Initialize $\mathbf{h} = [1, 1, \cdots, 1]$.
2. Construct the corresponding matrix $\mathbf{G}(\mathbf{h}) = \begin{bmatrix}\mathbf{U}_s \\ \mathbf{Y}_s^c(\mathbf{h})\end{bmatrix}$ with $\mathbf{h}$ and raw data $\mathbf{y} = [y_1, y_2 \cdots, y_N]$
3. Run the algorithm 1 to solve the problem:

$$\min_{\tilde{\mathbf{G}}} \left\|\tilde{\mathbf{G}} - \mathbf{G}(\mathbf{h})\right\|_F^2 \quad \text{s.t. } \tilde{\mathbf{G}} \in \mathcal{A},\ \text{rank}\tilde{\mathbf{G}} = sm + n \tag{4.8}$$

4. Denote the $(sm+1)$-th row in $\tilde{\mathbf{G}}$ by $\tilde{\mathbf{Y}}$, and convert $\tilde{\mathbf{Y}}$ to time domain $\tilde{\mathbf{y}}$. Estimate $\mathbf{h}^*$ by comparing $\tilde{\mathbf{y}}$ with $\mathbf{y}$. Set $\mathbf{h} = \mathbf{h}^*$, $\mathbf{h} = \mathbf{h}/h_1$ and repeat step 2-3 until $\|\mathbf{h}^* - \mathbf{h}\|_F^2 < threshold$.
5. Set $\beta = \gamma h_2$ and adjust the calibration vector $\hat{h}_i = h_i \beta^{1-i}$, $i = 2, 3, \cdots, N$.

4.2 Numerical illustration

We choose one linear array with $N = 16$ elements and two signal resources with relative frequencies $\omega_1 = -2\pi * 0.122, \omega_2 = 2\pi * 0.22$. Two cases $m = 5$ and $m = 10$ snapshots are generated for each element. A zero mean complex circularly symmetric Gaussian distributed noise with variance ranging from 10^{-4} to 10^{-1} is added to the noise free array responses. We simulate the unknown gain to each element by adding a zero mean complex circularly symmetric uniformly distributed random variable with variance 0.2 to the unit gain every time we run this algorithm. The two thresholds can be chosen freely, and theoretically smaller thresholds will give better outputs. 50 independent reliaztions of the source signals, noise, and the antenna gains are generated to evaluate the performance.
The results of numerical evaluation when we have $m = 10$ snapshots are shown in Figure 4.1 and Figure 4.2. In Figure 4.1 , the subspace algorithm with uncalibrated data from [17] and the ADMM algorithm are compared by RMS error of spatial

frequency. In Figure 4.2, the ADMM algorithm is compared to the eigenstructure method from [4] as well as the uncalibrated outputs. We can see from the first figure that the ADMM algorithm has clearly improved the performance compared to the subspace algorithm, but when the noise variance is large at 10^{-1} the subspace algorithm is slightly better. This is because in the ADMM algorithm we use β to adjust the calibration vector $\mathbf{h}$. If the noise is large, the estimated h_2 is not as accurate as before, which introduces more errors to the estimated $\mathbf{h}$ after the β adjustment step. This also explains why in Figure 4.2 the result of the ADMM algorithm is worse than uncalibrated case when the noise variance is 10^{-1}. In the Figure 4.2, the uncalibrated outputs mean that we ignore the unknows gains and just regard them as unit gains. The performance of eigenstructure method is even worse than the uncalibrated outputs, and the reason is that we only use 10 snapshots here, which leads to the poor estimatation of covariance matrix in [4, Equation 8]. In the Figure 4.1, the errors of subspace method do not vary with different noise variances. This is because there is no calibration in the subspace method, and then the unknown gains in the elements are equivalent to adding very large noise to the received signal. In both figures we see that the errors of ADMM algorithm decrease with improved SNR which suggests that this method is consistent in SNR.

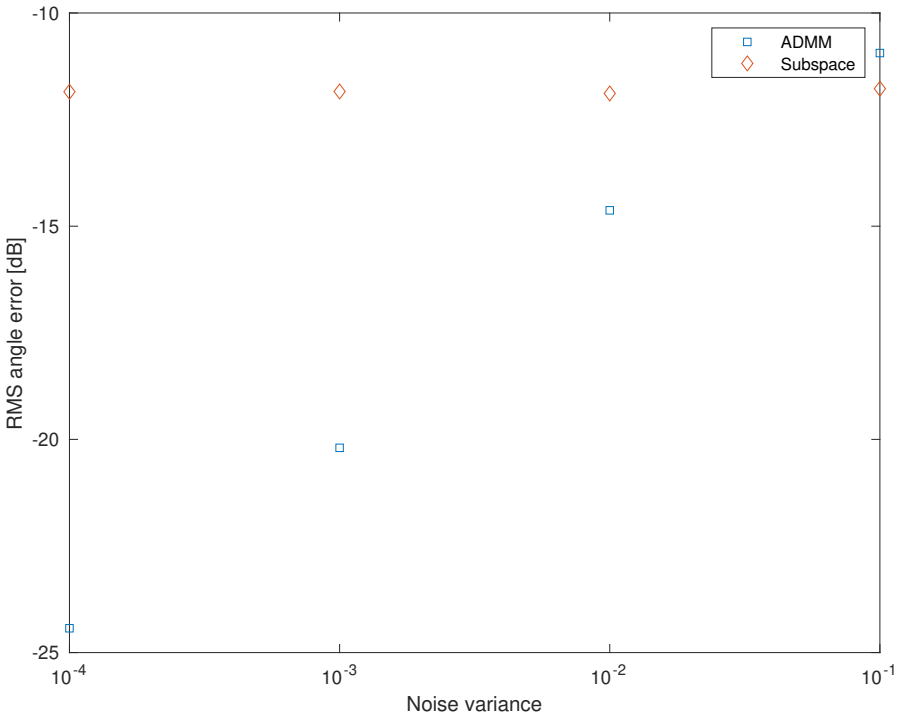

Figure 4.1: RMS error for estimated frequencies versus variance of noise

Next, we show the result comparison with different snapshots. According to Figure 4.3 and Figure 4.4, the implementation with more snapshots will have better performances.

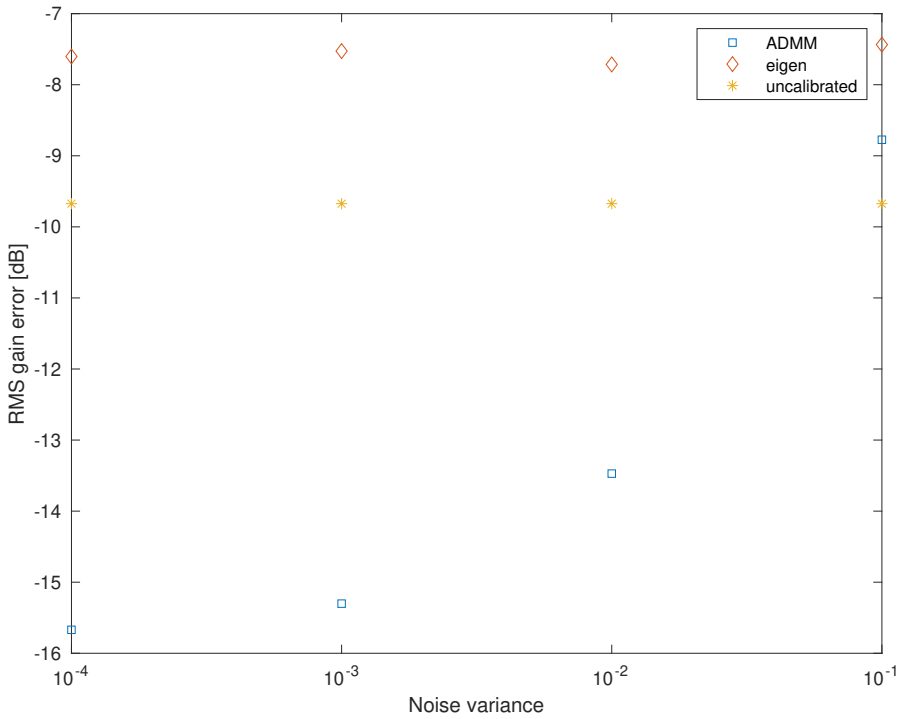

Figure 4.2: RMS error for calibration versus variance of noise

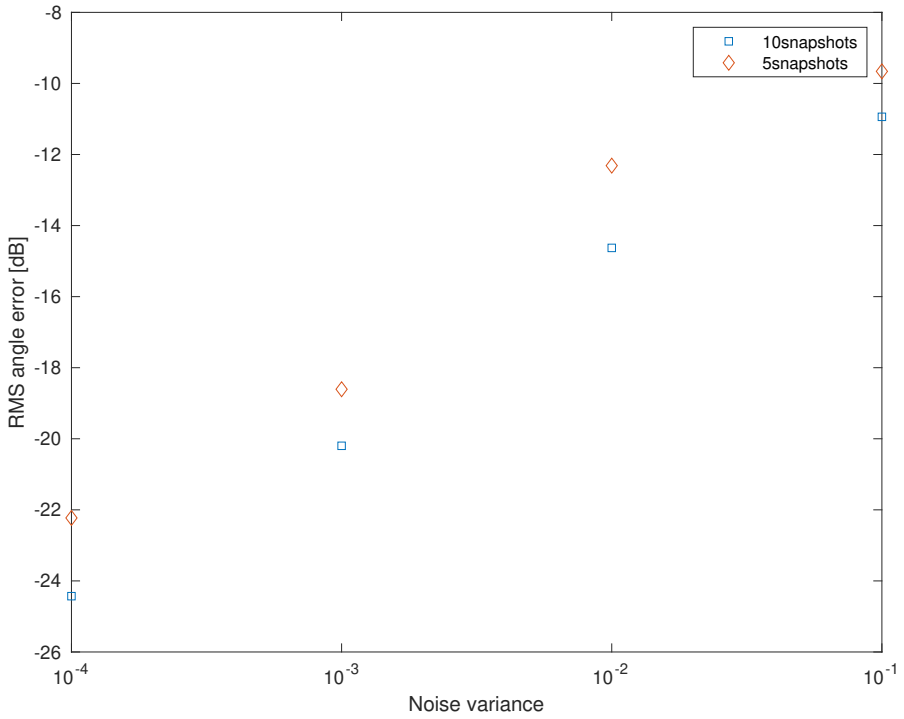

Figure 4.3: RMS error for estimated frequencies versus variance of noise

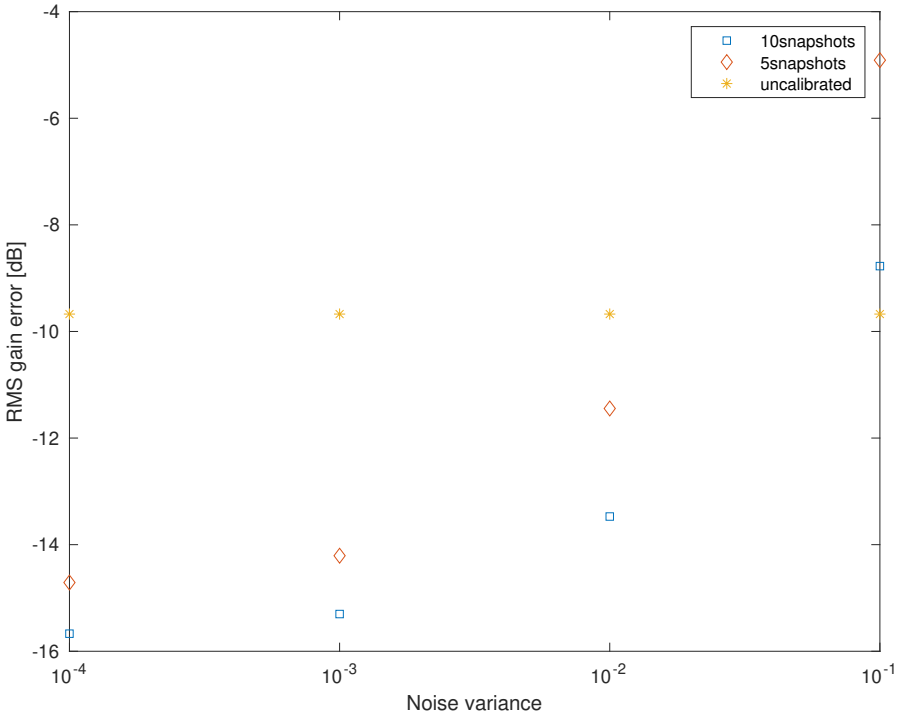

Figure 4.4: RMS error for calibration versus variance of noise

5

Conclusion

In this Master thesis, we developed the realization theory in frequency domain and proved that in which case the time domain data can be realized by a minimal system. With the help of the theoretic knowledge, we applied the ADMM algorithm from convex optimization theory to achieve the calibration of array antenna. Especially if we know the the gain ratio between two consecutive antenna elements, we can determine the unknown gain effectively. The result shows the new algorithm has improved the performance compared to the old one and is also consistent with SNR. As for the future work, because the number of targets are the same for multiple arrays, we can do the calibration together with the signals received from different arrays, which is supposed to have better performance than doing the calibration individually for each array.

Bibliography

[1] W. L. Stutzman and G. A. Thiele, *Antenna theory and design*. Wiley, 2013.

[2] R. L. Haupt, *Antenna arrays: a computational approach*. John Wiley & Sons, 2010.

[3] R. J. Mailloux, *Phased array antenna handbook*. Artech House., 2018.

[4] A. Paulraj and T. Kailath, "Direction of arrival estimation by eigenstructure methods with unknown sensor gain and phase," in *ICASSP '85. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 10, pp. 640–643, 1985.

[5] B. Friedlander and A. J. Weiss, "Eigenstructure methods for direction finding with sensor gain and phase uncertainties," in *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, pp. 2681–2684 vol.5, 1988.

[6] M. Lin and L. Yang, "Blind calibration and doa estimation with uniform circular arrays in the presence of mutual coupling," *IEEE Antennas and Wireless Propagation Letters*, vol. 5, pp. 315–318, 2006.

[7] J. Kim, H. J. Yang, B. W. Jung, and J. Chun, "Blind calibration for a linear array with gain and phase error using independent component analysis," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 1259–1262, 2010.

[8] K. Nambur Ramamohan, S. Prabhakar Chepuri, D. F. Comesaña, G. Carrillo Pousa, and G. Leus, "Blind calibration for acoustic vector sensor arrays," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3544–3548, 2018.

[9] K. N. Ramamohan, S. Prabhakar Chepuri, D. F. Comesaña, and G. Leus, "Blind calibration of sparse arrays for doa estimation with analog and one-bit measurements," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4185–4189, 2019.

[10] T. McKelvey, "Auto-calibration of co-located uniform linear array antennas," in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 56–60, 2019.

[11] H. L. Trentelman, M. L. J. Hautus, and A. Stoorvogel, *Control theory for linear systems*. Springer, 2012.

[12] E. D. Sontag, *Mathematical control theory: deterministic finite dimensional systems*. Springer, 1990.

[13] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2018.

[14] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends¬Æ in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[15] T. Kailath, *Linear systems*. Prentice Hall International, 1998.

[16] T. McKelvey, H. Akcay, and L. Ljung, "Subspace-based multivariable system identification from frequency response data," *IEEE Transactions on Automatic Control*, vol. 41, no. 7, pp. 960–979, 1996.

[17] T. McKelvey and M. Viberg, "A robust frequency domain subspace algorithm for multi-component harmonic retrieval," *Conference Record of Thirty-Fifth Asilomar Conference on Signals, Systems and Computers (Cat.No.01CH37256)*, 2001.

[18] R. A. Horn and C. R. Johnson, *Topics in matrix analysis*. Cambridge Univ. Press, 2010.

[19] V. Larsson, C. Olsson, E. Bylow, and F. Kahl, "Rank minimization with structured data patterns," *Computer Vision – ECCV 2014 Lecture Notes in Computer Science*, p. 250–265, 2014.

[20] V. Larsson and C. Olsson, "Convex low rank approximation," *International Journal of Computer Vision*, vol. 120, no. 2, p. 194–214, 2016.

[21] T. McKelvey, "Auto-calibration of uniform linear array antennas," in *2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5, 2019.