**Circuits and Systems**
Mekelweg 4,
2628 CD Delft
The Netherlands
http://ens.ewi.tudelft.nl/

**CAS-2011-00**

# M.Sc. Thesis

# Complex Factor Analysis

**Ahmad Mouri Sardarabadi B.Sc.**

**Abstract**

Many techniques for array processing assume either that the system has a calibrated array or that the noise covariance matrix is known. If the noise covariance matrix is unknown, training or other calibration techniques are used to find it.

In this thesis another approach to the problem of unknown noise covariance is presented. The factor analysis (FA) model is used to model the data. In order to make the theory applicable in telecommunication and radio astronomy, the model is extended to the case of complex numbers.

The necessary mathematical tools for estimation, detection and performance analysis are derived. The maximum likelihood estimator for the FA model in the case of proper complex Gaussian distributed noise and signals is given. Two different iterative algorithms for finding the MLE of the model parameters are presented. The necessary iteration steps for an alternating least squares algorithm are also presented.

The Cramér–Rao bound for the FA model is found and the convergence of the estimated parameters to this bound are illustrated using simulations. The improved performance of some popular algorithms like ESPRIT and spatial filtering have been demonstrated using computer simulations.

A general likelihood ratio test is presented as a tool to test the correctness of the estimated parameters. An especial case of this test is presented as a constant false alarm detector.

This thesis is an initial attempt for developing the factor analysis as a tool for signal processing. The multivariate nature of the theory makes it a good candidate for solving many problems in telecommunication and radio astronomy.

# Complex Factor Analysis
## Applications in Array Processing

---

Thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

in

Electrical Engineering

by

Ahmad Mouri Sardarabadi B.Sc.
born in Ahvaz, Iran

This work was performed in:

Circuits and Systems Group
Department of Microelectronics & Computer Engineering
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

**Delft University of Technology**

Delft University of Technology
Department of
Microelectronics & Computer Engineering

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled **"Complex Factor Analysis"** by **Ahmad Mouri Sardarabadi B.Sc.** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 27-06-2011

Chairman: _____
prof.dr.ir. Alle-Jan van der Veen

Advisor: _____
prof.dr.ir. Alle-Jan van der Veen

Committee Members: _____
dr. Richard Heusdens

_____
dr. Stefan Wijnholds

_____
dr.ir. Rufus Fraanje

# Abstract

Many techniques for array processing assume either that the system has a calibrated array or that the noise covariance matrix is known. If the noise covariance matrix is unknown, training or other calibration techniques are used to find it.

In this thesis another approach to the problem of unknown noise covariance is presented. The factor analysis (FA) model is used to model the data. In order to make the theory applicable in telecommunication and radio astronomy, the model is extended to the case of complex numbers.

The necessary mathematical tools for estimation, detection and performance analysis are derived. The maximum likelihood estimator for the FA model in the case of proper complex Gaussian distributed noise and signals is given. Two different iterative algorithms for finding the MLE of the model parameters are presented. The necessary iteration steps for an alternating least squares algorithm are also presented.

The Cramér–Rao bound for the FA model is found and the convergence of the estimated parameters to this bound are illustrated using simulations. The improved performance of some popular algorithms like ESPRIT and spatial filtering have been demonstrated using computer simulations.

A general likelihood ratio test is presented as a tool to test the correctness of the estimated parameters. An especial case of this test is presented as a constant false alarm detector.

This thesis is an initial attempt for developing the factor analysis as a tool for signal processing. The multivariate nature of the theory makes it a good candidate for solving many problems in telecommunication and radio astronomy.

# Acknowledgments

First of all I would like to thank my advisor prof.dr.ir. Alle-Jan van der Veen for his assistance during the writing of this thesis. I would also like to thank my family and friends, especially my parents for their support and encouragements. Especial thanks goes to my dear Shima for the confident she gave me.

Ahmad Mouri Sardarabadi B.Sc.
Delft, The Netherlands
27-06-2011

# Contents

# List of Figures

# List of Tables

# 1 Introduction

The electromagnetic spectrum is one of the biggest sources of information for most of the remote-sensing and modern telecommunication problems. Be it radar, radio astronomy, cellular communication or TV-broadcasting, the signal as received at the receiving elements needs to be processed to fulfill some desired function. Signal processing is the area of applied mathematics that deals with operations on or analysis of signals in order to do useful tasks [1]. In this thesis the theory of factor analysis is studied in relation with signal processing. This chapter addresses the motivation for the work done and outlines the main contributions.

## 1.1 Motivation

Detection and estimation form the core of many signal processing problems. In many cases the desired signal is not observed directly and must be extracted from the measured signals. In many algorithms for array processing, the array is either assumed calibrated or the full knowledge of noise statistics is available. If the noise statistics are not know or the array is not calibrated a more comprehensive model is needed. In multivariate analysis a signal that is not observed directly is called the latent variable and the observed signal is called the manifest variable. The statistical model that deals with the manifest and latent variables is called the latent variable model. Table **??** shows the classification given by Bartholomew [2, p.11] for latent variable models. Bartholomew uses the term metric for variables that are from the set of real numbers and categorical for those variables that are elements of a set of categories. Throughout this thesis metric variables are from the set of complex numbers.

Factor analysis is the class of latent variable models that deals with the metric variables. In [3] a very good overview of possible applications of factor analysis in various fields of natural sciences is given. This model will be used here in array processing for the telecommunication or radio-astronomy. Given the wireless nature of modern telecommunication and radio-astronomy the desired signal is always corrupted by noise and interference and seldom observed directly. Using the tools available from the theory of multivariate analysis is therefore very popular in these fields. It is common

Table 1.1: Latent Variable Models

| | | Manifest Variable | |
|---|---|---|---|
| | | Metric | Categorical |
| Latent Variable | Metric | Factor analysis | Latent trait analysis |
| | Categorical | Latent profile analysis | Latent class analysis |

to represent the measured electromagnetic field from a receiving element (usually an antenna) by complex numbers. As a result the variables have metric nature.

In radio-astronomy the signal power that are emitted from stars in different frequency bands are studied and if a frequency band is contaminated with interfering signal produced by man made devices, it is useless for the astronomers. Filtering the unwanted signal using subspaces has proven to be a very powerful and widely used method. Eigenvalue decomposition can be used on a data-set to find the so called signal and noise spaces. In multivariate analysis subspace techniques fall under the principal component analysis. The true principal component analysis does not have an error term and does not deal with the noise, in fact the principal component analysis does not pose any model on the data [3]. In practice, with the use of *a-priori* knowledge, training and calibration techniques noise contribution is mitigated. For example matched subspace detection, coherence detection, adaptive MSD, etc. can be used to detect the subspace even if it is contaminated with interference and noise, using training sets [4][5]. The factor analysis model that will be presented in the next chapter proposes an underlying structure for the measured data.

## 1.2 Contributions

As will be shown in Chapter 2 the factor analysis model is commonly used in the signal processing problems. However the mathematical tools needed for this model are not well-known in this field. One of the goals of this thesis is to give an overview of the theory in the concept of signal processing and develop it as signal processing tool.

In the literature on factor analysis the case of the complex numbers is not well studied. One of the main contribution of this paper is to extend much of the known theory to the complex numbers.

Fisher Scoring, alternating least squares and KullbackLeibler divergence (KLD) minimization algorithm are proposed and discussed as possible tools to estimate the model parameters.

In order to evaluate the performance of the algorithms, the Cramér-Rao Bound for the factor analysis model is derived.

## 1.3 Outline

In the following chapters the factor analysis will be extended to the case of complex-valued data and some new results are presented. In Chapter 2 the model is presented and its properties are discussed. In Chapter 3 the Cramér-Rao Bound for the factor analysis model is derived as a tool for performance evaluation. In Chapter 4 some algorithms for the estimation of the model parameters are developed. Chapter 5 deals with the detection of signals and correctness of the model. Chapter 6 summarizes the main results and discusses the possibilities for further research.

# 2 Model

The factor analysis (FA) model was first introduced in 1904 by Spearman to analyze human intelligence based on a series of tests [6]. Given test results on $p$ subjects (e.g. English, French, Math, etc) Spearman was looking for a common factor $f$ that would explain the performance of an individual. The contribution of this factor to each test is called *factor loading*, $\lambda$, and the random deviation from this contribution is the error in each test. The model for the results,

$$\mathbf{r} = \mathbf{\Lambda} f + \mathbf{e} + \boldsymbol{\mu} \tag{2.1}$$

is the "true" factor analysis model. Where $\mathbf{r}$ is a vector of $p \times 1$ which contains the results of subjects, $f$ is the *common factor*, the matrix $\mathbf{\Lambda}$ is a $p \times 1$ factor loading, $\mathbf{e}$ is a vector of $p \times 1$ error terms and $\boldsymbol{\mu}$ is the expected value of the results.

This model can easily be extended to $m < p$ common factors that might influence the results by replacing $f$ with a $m \times 1$ vector of common factors, $\mathbf{f}$. Then $\mathbf{\Lambda}$ becomes a $p \times m$ matrix. [7][8][3]

In the following section the signal processing model commonly used for array processing is presented and it is shown that the two models are the same. In the subsequent sections the main problem is defined and the model will be examined more closely.

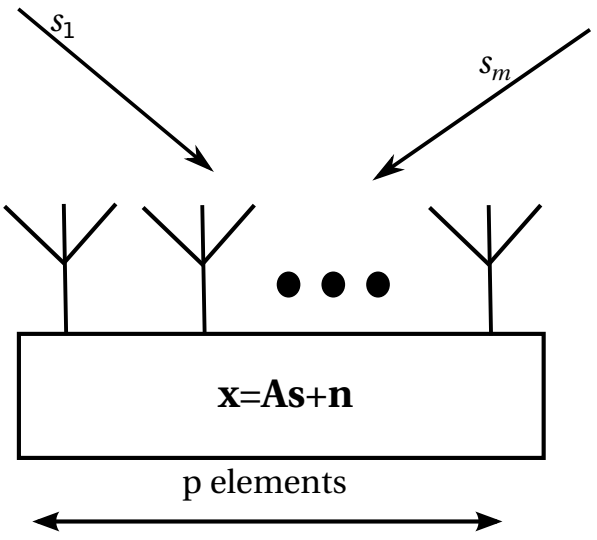



Figure 2.1: Schematic representation of the receiving elements

## 2.1 Signal Processing Model

The model used here is a simplified model commonly used in literature[9, p.58]. Consider a system with $p$ receiving elements that are exposed to $m$ sources, Figure 2.1. Stacking the $p$ received signals into one vector called $\mathbf{x}$ and assuming $m$ independent sources stacked into the vector $\tilde{\mathbf{s}}$ with a $\mathcal{CN}(0, \mathbf{\Sigma}_{\tilde{\mathbf{s}}})$ distribution, the following model is assumed for $\mathbf{x}$

$$\mathbf{x} = \tilde{\mathbf{A}}_{\mathbf{x}}\tilde{\mathbf{s}} + \mathbf{w} \tag{2.2}$$

where $\tilde{\mathbf{A}}_{\mathbf{x}}$ is a $p \times m$ array response and $\mathbf{w}$ is the noise term with a $\mathcal{CN}(0, \mathbf{\Sigma}_{\mathbf{w}})$ distribution.

From (2.2) it follows that $\mathbf{x}$ has a $\mathcal{CN}(0, \mathbf{\Sigma}_{\mathbf{x}})$ distribution with

$$\mathbf{\Sigma}_{\mathbf{x}} = \tilde{\mathbf{A}}_{\mathbf{x}}\mathbf{\Sigma}_{\tilde{\mathbf{s}}}\tilde{\mathbf{A}}_{\mathbf{x}}^{H} + \mathbf{\Sigma}_{\mathbf{w}} \tag{2.3}$$

where $\tilde{\mathbf{A}}_{\mathbf{x}}^{H}$ is the Hermitian transpose of $\tilde{\mathbf{A}}_{\mathbf{x}}$.

It is clear that the linear model "explaining" $\mathbf{x}$ is the same model that "explains" the test results in the previous section. The main differences are that all the parameters in (2.1) are real but the parameters in (2.2) could also be complex numbers and the fact that the expectation $\boldsymbol{\mu}$ is zero. Now that the equivalency of the two models is argued, some properties and issues with this model will be disscussed in the following sections.

## 2.2 Uniqueness

Keeping in mind that the problem at hand is a latent variable problem, one could ask if the relation between the latent and manifest variables (measured and desired signals) is unique? The answer as will be shown here is negative.

Given any invertible matrix $\mathbf{M}$, $\mathbf{x}$ can be rewritten as

$$\mathbf{x} = \tilde{\mathbf{A}}_{\mathbf{x}}\mathbf{M}\mathbf{M}^{-1}\tilde{\mathbf{s}} + \mathbf{w}.$$

Let $\mathbf{A}_{\mathbf{x}} = \tilde{\mathbf{A}}_{\mathbf{x}}\mathbf{M}$ and $\mathbf{s} = \mathbf{M}^{-1}\tilde{\mathbf{s}}$ then the factor analysis model (FA) for array processing becomes

$$\mathbf{x} = \mathbf{A}_{\mathbf{x}}\mathbf{s} + \mathbf{w}. \tag{2.4}$$

The covariance matrix of $\mathbf{x}$, $\mathbf{\Sigma}_{\mathbf{x}}$, is positive semi-definite so that the matrix $\mathbf{M}$ can always be chosen in such a way to make Cov($\mathbf{s}$)=$\mathbf{I}$. It is then straightforward to show that the covariance matrix of $\mathbf{x}$ can be written as

$$\mathbf{\Sigma}_{\mathbf{x}} = \mathbf{A}_{\mathbf{x}}\mathbf{A}_{\mathbf{x}}^{H} + \mathbf{\Sigma}_{\mathbf{w}}. \tag{2.5}$$

Given any unitary matrix $\mathbf{Q}$, let $\mathbf{B} = \mathbf{A}_{\mathbf{x}}\mathbf{Q}$, then $\mathbf{\Sigma}_{\mathbf{x}} = \mathbf{B}\mathbf{B}^{H} + \mathbf{\Sigma}_{\mathbf{w}}$ is also a valid model. This shows once more that the relationship between the measured signal and desired one is not unique. In fact the matrix $\mathbf{A}_{\mathbf{x}}$ can be seen as a basis for the signal space and as such the choice of a different basis is always available.

One way to choose a "unique" $\mathbf{A_x}$ is to rotate it in sucha way that its columns are orthogonal with respect to some weighting. One such a weighting that leads to easier mathematical results and is suggested in the literature [7][8] is

$$\mathbf{A_x}^H {\mathbf{\Sigma_w}}^{-1} \mathbf{A_x} = \mathbf{\Gamma} \tag{2.6}$$

where $\mathbf{\Gamma}$ is real and diagonal. It should be noted that $\mathbf{\Gamma}$is not chosen, but $\mathbf{A_x}$ is rotated till a diagonal matrix in formed.

The problem addressed here is a many-to-one problem that is similar with taking the square root. The choice of the weighting is a convention and not a real constraint on the results.

## 2.3 Scale Invariance

The original idea behind factor analysis as presented in the previous sections was to find a common factor that "explains" the results of different tests. The scaling for each of these exams can be different and also re-scaling them should not change the underlying factors.

Let $\mathbf{G}$ be any diagonal matrix, and let $\mathbf{y} = \mathbf{Gx}$. The factor analysis model still holds for $\mathbf{y}$ with the factor loading matrix $\mathbf{A} = \mathbf{GA_x}$ and noise vector $\mathbf{n} = \mathbf{Gw}$

$$\mathbf{y} = \mathbf{As} + \mathbf{n} \tag{2.7}$$

and the covariance matrix of $\mathbf{y}$ becomes

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^H + \mathbf{D} \tag{2.8}$$

where $\mathbf{D} = \mathbf{G}\mathbf{\Sigma_w}\mathbf{G}^H$. The noise is assumed to be spatially uncorrelated and as a result its covariance matrix is diagonal. Scaling the FA model with a diagonal matrix does not change this property.

An especial case for the scaling of the FA model is to choose $\mathbf{G} = \text{diag}(\mathbf{\Sigma_x})$, this will force the diagonal elements of $\mathbf{\Sigma}$ to unity. In array processing this is sometimes called gain compensation.

As shown above the FA model is unaffected by the scaling of the variables. (this does not hold for principal component analysis.[7])

## 2.4 Conclusion

In section 2.1 the equivalency of the signal model for array processing and the FA model is shown. The model suffers from many-to-one ambiguity and as a result the relation between the received signals and the desired ones is not unique. In 2.2 a method is suggested to choose one of the many possible solutions.

Another aspect of the FA model is its tolerance toward scaling of the variables. The model represented in section 2.3 is the most complete version of the FA model and is used for the rest of this thesis as the *factor analysis model*.

Thus the problem that is discussed in the following sections is:

Given $N$ samples from $\mathbf{y}$ as defined by (2.7) with a $\mathcal{CN}(0, \mathbf{\Sigma})$ distribution, where $\mathbf{\Sigma}$ is given by (2.8) and

$$\mathbf{A}^H \mathbf{D}^{-1} \mathbf{A} = \mathbf{\Gamma} \tag{2.9}$$

Where $\mathbf{\Gamma}$ is real and diagonal, how could the model parameters $\mathbf{A}$ and $\mathbf{D}$ be estimated? Also, given an estimation of these parameters, could the validity of the model be tested?

Some algorithms to estimate the model parameters are discussed in Chapter 4. In Chapter 5 the validity of the model will be presented in the concept of a detection problem.

# Cramér-Rao Bound

# 3

For an unbiased estimator the Cramér-Rao bound (CRB) is the lower bound on the covariance matrix, $\mathbf{C}$, of the estimated parameters, $\boldsymbol{\theta}$.

$$\mathbf{C}(\boldsymbol{\theta}) \geq \mathbf{F}^{-1} \tag{3.1}$$

where $\mathbf{F}$ is the Fisher information matrix. In this section the CRB for the FA model will be derived.

Also as will be shown in Chapter 4, the Fisher information matrix could be used to numerically approximate the maximum likelihood estimates of the model parameters with the so-called scoring algorithm [10, p.180,187].

## 3.1 Partial Derivatives

In this section the partial derivatives needed to find the CRB are derived for the case of complex parameters, for the case of real parameters the second order derivatives can be found for example in [11]. [11] also gives the diagonal sub-matrices of the Fisher information matrix, but it does not mention the CRB. In order to simplify the derivation, the following relationships are going to be used.

Definition: $\boldsymbol{\Phi}^i_{\boldsymbol{\zeta}}$ is a (complex) matrix with all its columns equal zero, except for its $i$th column which equals $\boldsymbol{\zeta}$.

$$\boldsymbol{\Phi}^i_{\boldsymbol{\zeta}} = \boldsymbol{\zeta}\mathbf{e}_i^T \tag{3.2}$$

where $\mathbf{e}_i$ is a unit vector with entry $i$ equal to 1.

Let $\boldsymbol{\Phi}^i_{\boldsymbol{\zeta}}$ be a matrix as defined above and $\mathbf{A}$ be any complex matrix, then the following can easly be shown:

1. $\mathbf{A}\boldsymbol{\Phi}^i_{\boldsymbol{\zeta}} = \boldsymbol{\Phi}^i_{\mathbf{A}\boldsymbol{\zeta}}$,
2. $\mathbf{A}(\boldsymbol{\Phi}^i_{\boldsymbol{\zeta}})^H = \mathbf{a}_i\boldsymbol{\zeta}^H$, where $\mathbf{a}_i = (\mathbf{A})_i$ is the $i$th column of matrix $\mathbf{A}$,
3. $\mathbf{A}\boldsymbol{\delta}(i,j) = \boldsymbol{\Phi}^j_{\mathbf{a}_i}$, where $\boldsymbol{\delta}(i,j)$ is a matrix with all its entries equal to zero except for the element on $i$th row and $j$th column,
4. $\boldsymbol{\delta}(i,j)\mathbf{A} = (\Phi^i_{(\mathbf{A}^H)_j})^H$,
5. $\boldsymbol{\Phi}^i_{\boldsymbol{\zeta}}\boldsymbol{\Phi}^j_{\boldsymbol{\theta}} = \boldsymbol{\Phi}^j_{\theta_i\boldsymbol{\zeta}}$,
6. $$\boldsymbol{\Phi}^i_{\boldsymbol{\zeta}}(\Phi^j_{\boldsymbol{\theta}})^H = \begin{cases} \mathbf{0} & i \neq j \\ \boldsymbol{\zeta}\boldsymbol{\theta}^H & i = j \end{cases},$$

7. $\boldsymbol{\Phi}_{\boldsymbol{\zeta}}^{i\,H}\boldsymbol{\Phi}_{\boldsymbol{\theta}}^{j} = \boldsymbol{\zeta}^{H}\boldsymbol{\theta}\boldsymbol{\delta}(i,j)$,

8. $\mathrm{tr}\{\boldsymbol{\Phi}_{\boldsymbol{\zeta}}^{i}\} = \zeta_i$,

9. $\mathrm{tr}\left\{\boldsymbol{\Phi}_{\boldsymbol{\zeta}}^{i\,H}\right\} = \bar{\zeta}_i$.

[12, p.165] shows that for any proper multivariate Gaussian variable $\mathbf{y} \sim \mathcal{CN}_p(0, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, the Fisher information matrix with complex parameters is given by:

$$f_{ij} = N\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right)^{H}\right\} = \overline{f}_{ji} \tag{3.3}$$

where the partial derivatives are Wirtinger derivatives.

If the parameter vector $\boldsymbol{\theta}$ is partitioned into sub-vectors $\left[\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \cdots\right]^T$ then the Fisher information matrix can also be partitioned as

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} & \mathbf{F}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2} & \cdots \\ \mathbf{F}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_1} & \mathbf{F}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{3.4}$$

Using (2.8), (3.3) and (3.4), the Fisher information matrix for each sub-matrix is going to be derived.

The parameters that need to be estimated are the columns of $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_m]$ and diagonal entries of $\mathbf{D}$.

Let $\mathbf{d} = \mathrm{vecdiag}(\mathbf{D})$, then the parameter vector $\boldsymbol{\theta} = \left[\mathbf{a}_1^T, \cdots, \mathbf{a}_m^T, \mathbf{d}^T\right]^T$ and the Fisher information matrix becomes

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{\mathbf{a}_1\mathbf{a}_1} & \cdots & \mathbf{F}_{\mathbf{a}_1\mathbf{a}_m} & \mathbf{F}_{\mathbf{a}_1\mathbf{d}} \\ \mathbf{F}_{\mathbf{a}_2\mathbf{a}_1} & \cdots & \mathbf{F}_{\mathbf{a}_2\mathbf{a}_m} & \mathbf{F}_{\mathbf{a}_2\mathbf{d}} \\ \vdots & \cdots & \ddots & \vdots \\ \mathbf{F}_{\mathbf{d}\mathbf{a}_1} & \cdots & \mathbf{F}_{\mathbf{d}\mathbf{a}_m} & \mathbf{F}_{\mathbf{d}\mathbf{d}} \end{bmatrix}.$$

Each of these sub-matrices needs to be found. First the matrices with the form $\mathbf{F}_{\mathbf{a}_k\mathbf{a}_n}$, with $n, k = 1 \cdots m$, are derived. For the FA model

$$\begin{aligned} \frac{\partial\boldsymbol{\Sigma}}{\partial a_{ik}} &= \frac{\partial\left(\mathbf{A}\mathbf{A}^H + \mathbf{D}\right)}{\partial a_{ik}} \\ &= \frac{\partial\mathbf{A}\mathbf{A}^H}{\partial a_{ik}} \\ &= \begin{bmatrix} \frac{\partial(\mathbf{A}\mathbf{A}^H)_{11}}{\partial a_{ik}} & \cdots & \frac{\partial(\mathbf{A}\mathbf{A}^H)_{1p}}{\partial a_{ik}} \\ \vdots & \ddots & \vdots \\ \frac{\partial(\mathbf{A}\mathbf{A}^H)_{p1}}{\partial a_{ik}} & \cdots & \frac{\partial(\mathbf{A}\mathbf{A}^H)_{pp}}{\partial a_{ik}} \end{bmatrix} \end{aligned}$$

where $(\mathbf{YZ})_{ij}$ is the element of the matrix $\mathbf{YZ}$ at $i$th row and $j$th column.

Let $\mathbf{H} = \mathbf{A}\mathbf{A}^H$ then

$$\begin{aligned}
\frac{\partial h_{sr}}{\partial a_{ik}} &= \frac{\partial \sum_l a_{sl} a_{lr}^H}{\partial a_{ik}} \\
&= \sum_l \frac{\partial a_{sl}\overline{a_{rl}}}{\partial a_{ik}} \\
&= \begin{cases} \overline{a_{rk}} & s = i \\ 0 & \text{otherwise} \end{cases} \\
\frac{\partial \mathbf{H}}{\partial a_{ik}} &= \begin{bmatrix} & \mathbf{0} & \\ \overline{a_{1k}} & \cdots & \overline{a_{pk}} \\ & \mathbf{0} & \end{bmatrix}.
\end{aligned}$$

Thus

$$\frac{\partial \mathbf{\Sigma}}{\partial a_{ik}} = (\mathbf{\Phi}^i_{\mathbf{a}_k})^H. \tag{3.5}$$

Using (3.5) and (3.3) then the Fisher information matrix is

$$\begin{aligned}
f_{\mathbf{a}_k \mathbf{a}_{n\,ji}} &= N\mathrm{tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{\Phi}^i_{\mathbf{a}_k}\mathbf{\Sigma}^{-1}{\mathbf{\Phi}^j_{\mathbf{a}_n}}^H\right\} \\
&= N\mathrm{tr}\left\{{\mathbf{\Phi}^j_{\mathbf{a}_n}}^H\mathbf{\Sigma}^{-1}\mathbf{\Phi}^i_{\mathbf{a}_k}\mathbf{\Sigma}^{-1}\right\} \\
&= N\mathrm{tr}\left\{\mathbf{a}_n^H\mathbf{\Sigma}^{-1}\mathbf{a}_k\boldsymbol{\delta}(j,i)\mathbf{\Sigma}^{-1}\right\} \\
&= N\mathbf{a}_n^H\mathbf{\Sigma}^{-1}\mathbf{a}_k\mathrm{tr}\left\{\boldsymbol{\delta}(j,i)\mathbf{\Sigma}^{-1}\right\} \\
&= N\mathbf{a}_n^H\mathbf{\Sigma}^{-1}\mathbf{a}_k\mathrm{tr}\left\{{\mathbf{\Phi}^j_{(\mathbf{\Sigma}^{-H})_i}}^H\right\} \\
&= N\mathbf{a}_n^H\mathbf{\Sigma}^{-1}\mathbf{a}_k\mathrm{tr}\left\{{\mathbf{\Phi}^j_{\mathbf{\Sigma}_i^{-1}}}^H\right\} \\
&= N\mathbf{a}_n^H\mathbf{\Sigma}^{-1}\mathbf{a}_k\overline{\mathbf{\Sigma}^{-1}_{ji}} \\
&= N\mathbf{a}_n^H\mathbf{\Sigma}^{-1}\mathbf{a}_k\mathbf{\Sigma}^{-1}_{ij}
\end{aligned} \tag{3.6}$$

where the properties 1,7,4 and 9 were used.

This relation can be extended for the whole matrix in a straightforward manner.

$$\overline{\mathbf{F}}_{\mathbf{a}_k\mathbf{a}_n} = N\mathbf{a}_n^H\mathbf{\Sigma}^{-1}\mathbf{a}_k\mathbf{\Sigma}^{-1} \tag{3.7}$$

The remaining sub-matrices are $\mathbf{F}_{\mathbf{a}_k\mathbf{d}}$ and $\mathbf{F}_{\mathbf{dd}}$. To derive these first $\frac{\partial \mathbf{\Sigma}}{\partial d_i}$ is needed.

$$\begin{aligned}
\frac{\partial \mathbf{\Sigma}}{\partial d_i} &= \frac{\partial\left(\mathbf{A}\mathbf{A}^H + \mathbf{D}\right)}{\partial d_i} \\
&= \frac{\partial \mathbf{D}}{\partial d_i} \\
&= \boldsymbol{\delta}(i,i)
\end{aligned} \tag{3.8}$$

With this result and (3.3) the entries of $\mathbf{F}_{\mathbf{a}_k\mathbf{d}}$ can be derived.

$$\begin{aligned} f_{\mathbf{a}_k\mathbf{d}_{ji}} &= N\text{tr}\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}^i_{\mathbf{a}_k}\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}(j,j)\right\} \\ &= N\text{tr}\left\{\boldsymbol{\Phi}^i_{\boldsymbol{\Sigma}^{-1}\mathbf{a}_k}\boldsymbol{\Phi}^j_{\boldsymbol{\Sigma}_j^{-1}}\right\} \\ &= N\text{tr}\left\{\boldsymbol{\Phi}^j_{\boldsymbol{\Sigma}_{ij}^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{a}_k}\right\} \\ &= N\boldsymbol{\Sigma}_{ij}^{-1}(\boldsymbol{\Sigma}_j^{-1})^H\mathbf{a}_k \end{aligned} \tag{3.9}$$

in the matrix form it becomes

$$\overline{\mathbf{F}}_{\mathbf{a}_k\mathbf{d}} = N\boldsymbol{\Sigma}^{-1}\text{diag}(\boldsymbol{\Sigma}^{-1}\mathbf{a}_k) \tag{3.10}$$

where $\text{diag}(\mathbf{x})$ is a diagonal matrix with its entries equal the entries of the vector argument.

The last sub-matrix is $\mathbf{F}_{\mathbf{dd}}$

$$\begin{aligned} f_{\mathbf{dd}_{ij}} &= N\text{tr}\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}(i,i)\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}(j,j)\right\} \\ &= N\text{tr}\left\{\boldsymbol{\Phi}^i_{\boldsymbol{\Sigma}_i^{-1}}\boldsymbol{\Phi}^j_{\boldsymbol{\Sigma}_j^{-1}}\right\} \\ &= N\text{tr}\left\{\boldsymbol{\Phi}^j_{\boldsymbol{\Sigma}_{ij}^{-1}\boldsymbol{\Sigma}_i^{-1}}\right\} \\ &= N\boldsymbol{\Sigma}_{ij}^{-1}\boldsymbol{\Sigma}_{ji}^{-1} \\ &= N\boldsymbol{\Sigma}_{ij}^{-1}\overline{\boldsymbol{\Sigma}_{ij}^{-1}} \end{aligned} \tag{3.11}$$

or equivalently

$$\mathbf{F}_{\mathbf{dd}} = N(\boldsymbol{\Sigma}^{-1} \odot \overline{\boldsymbol{\Sigma}^{-1}}) \tag{3.12}$$

where $\odot$ is the Hadamard or element-wise product of two matrices.

The $\boldsymbol{\Sigma}^{-1}$ for the FA model can efficiently be calculated using Woodbury identity.

$$\boldsymbol{\Sigma}^{-1} = (\mathbf{D} + \mathbf{A}\mathbf{A}^H)^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{A}(\mathbf{I} + \mathbf{A}^H\mathbf{D}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{D}^{-1}$$

Using (2.9) calculating $\boldsymbol{\Sigma}^{-1}$ needs only the inversion of two diagonal matrices.

## 3.2 Conclusion

All the sub-matrices that are needed to construct the Fisher information matrix are derived and the CRB can be calculated at true values of $\mathbf{A}$ and $\mathbf{D}$.

It may happen that the Fisher information becomes singular. In such a situation (2.9) could be used to find the CRB for constrained estimations. For more details take a look at [13].

# 4 Estimation

In this chapter various techniques to estimate the model parameters as introduced in Chapter 2 will be discussed.

Given a series of measurements from $\mathbf{y}$ as defined by (2.7) with sample covariance matrix

$$\mathbf{S} = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{y}[i]\mathbf{y}[i]^H \tag{4.1}$$

where $N$ is the number of samples, the model parameters as defined by (2.7) and (2.8) will be estimated.

First the maximum likelihood estimator will be derived in section 4.1. The Alternating least squares (ALS) and the KLD minimization algorithms are then presented in Section 4.2 and Section 4.1.3.

## 4.1 Maximum Likelihood Estimator

In [9, p.50] a very interesting quantum physical reasoning is provided for the assumption of i.i.d. Gaussian distribution of celestial sources. Also in [7],[8] and [3] the common factors and the error terms have a Gaussian distribution. Therefor the aim is to find $\mathbf{A}$ and $\mathbf{D}$ that maximize the log-likelihood function

$$l(\mathbf{A},\mathbf{D},\mathbf{S}) = -N\log|\pi^p| + N\log|\mathbf{\Sigma}^{-1}| - N\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) \tag{4.2}$$

where

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^H + \mathbf{D}$$

as shown in (2.8).

### 4.1.1 Fisher Score

Finding the MLE is the same as setting the Fisher score equal to zero. The Fisher score for proper Gaussian distributed signal is given by [12, p.165]

$$\begin{aligned} t_{\theta_i} &= -N\mathrm{tr}\left[\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\bar{\theta}_i}\right] + N\mathrm{tr}\left[\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\bar{\theta}_i}\mathbf{\Sigma}^{-1}\mathbf{S}\right] \\ &= -N\mathrm{tr}\left[\mathbf{\Sigma}^{-1}\left(\frac{\partial\mathbf{\Sigma}}{\partial\theta_i}\right)^H\right] + N\mathrm{tr}\left[\mathbf{\Sigma}^{-1}\left(\frac{\partial\mathbf{\Sigma}}{\partial\theta_i}\right)^H\mathbf{\Sigma}^{-1}\mathbf{S}\right]. \end{aligned} \tag{4.3}$$

Using the results of Chapter 3 the Fisher score for the FA model is derived here:

$$
\begin{aligned}
t_{a_{ik}} &= -N\text{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial a_i k}\right)^H\right] + N\text{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial a_i k}\right)^H\boldsymbol{\Sigma}^{-1}\mathbf{S}\right] \\
&= -N\text{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}^i_{\mathbf{a}_k}\right] + N\text{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}^i_{\mathbf{a}_k}\boldsymbol{\Sigma}^{-1}\mathbf{S}\right] \\
&= -N\text{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}^i_{\mathbf{a}_k}\right] + N\text{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}^i_{\mathbf{a}_k}\right] \\
&= -N\text{tr}\left[\boldsymbol{\Phi}^i_{\boldsymbol{\Sigma}^{-1}\mathbf{a}_k}\right] + N\text{tr}\left[\boldsymbol{\Phi}^i_{\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{a}_k}\right] \\
&= -N(\boldsymbol{\Sigma}^{-1})^H_i\mathbf{a}_k + N(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1})^H_i\mathbf{a}_k
\end{aligned}
$$

In the Matrix form:

$$\mathbf{T_A} = -N\boldsymbol{\Sigma}^{-1}\mathbf{A} + N\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{A}. \tag{4.4}$$

Following the same procedure, $\mathbf{T_D}$ is found to be :

$$\mathbf{T_D} = -N\text{diag}(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}). \tag{4.5}$$

Thus the Fisher score vector for the FA model is then

$$\mathbf{t}_{\boldsymbol{\theta}} = [\text{vect}(\mathbf{T_A})^T, \text{vect}(\mathbf{T_D})^T]^T \tag{4.6}$$

where vect(.) is the vectorization function that transforms a matrix to a column vector by stacking its columns under each other. Solving for MLE directly from $\mathbf{t}$ or (4.4) and (4.5) is not possible. In the next section the scoring algorithm is used to find the MLE numerically.

### 4.1.2 Scoring Algorithm

Scoring algorithm is a variant of Newton-Raphson algorithm with slight modification. If the gradient and the Hessian in the Newton-Raphson are replaced by the Fisher score and the inverse of the Fisher information matrix respectively, the result of the iterative method is called the *method of scoring* [10, p.180].

The flowing relation summarizes the method:

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \mathbf{F}(\hat{\boldsymbol{\theta}}_k)^{-1}\mathbf{t}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_k) \tag{4.7}$$

where $\mathbf{F}$ is the Fisher information matrix derived in Chapter 3 and $\mathbf{t}_{\boldsymbol{\theta}}$ is the Fisher score given by (4.6). To avoid inversion Kay [10, p.180] suggests to rewrite the equation as

$$\mathbf{F}(\hat{\boldsymbol{\theta}}_k)\hat{\boldsymbol{\theta}}_{k+1} = \mathbf{F}(\hat{\boldsymbol{\theta}}_k)\hat{\boldsymbol{\theta}}_k + \mathbf{t}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_k) \tag{4.8}$$

and solve the linear equations.

The Newton-Raphson suffers from the following difficulties [10, p.179]:

- There is no guarantee that the iterations will converge.
- There is no guarantee that in the case of convergence the global maximum is found.

- This method is very sensitive to the initial guess.

Using the Fisher information matrix instead of the Hessian adds some stability to this method but it still inherits the mentioned difficulties. To cope with these difficulties the following iteration is proposed. Solving the linear equations could be seen as multiplying (4.8) from left with the pseudo-inverse of $\mathbf{F}$.

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \mu \mathbf{F}(\hat{\boldsymbol{\theta}}_k)^{\dagger} \mathbf{t}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_k) \tag{4.9}$$

where $\mathbf{M}^{\dagger}$ is the pseudo-inverse of the matrix $\mathbf{M}$ and $\mu$ is a small constant. The constant $\mu$ will help the convergence, if the Fisher information matrix becomes very small. One good motivation for the use of the scoring method despite its difficulties is that the FA model has inherent limitation on the search domain of the parameters. This helps the convergence of the method. One of this limitation is that $\mathbf{D} > 0$. At maximum (4.5) becomes

$$\begin{aligned} \operatorname{diag}(\boldsymbol{\Sigma}^{-1}) &= \operatorname{diag}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}) \\ \operatorname{diag}(\mathbf{I}) &= \operatorname{diag}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) \\ \operatorname{diag}(\boldsymbol{\Sigma}) &= \operatorname{diag}(\mathbf{S}) \\ \operatorname{diag}(\mathbf{A}\mathbf{A}^H + \mathbf{D}) &= \operatorname{diag}(\mathbf{S}). \end{aligned} \tag{4.10}$$

Considering that $\operatorname{diag}(\mathbf{A}\mathbf{A}^H) \geq \mathbf{0}$, then $\mathbf{0} < \mathbf{D} \leq \operatorname{diag}(\mathbf{S})$. This limitation helps the convergence of the scoring method.

As mentioned earlier the choice of the starting point is very important. Reyment and Jöreskog [3, p.102] give the best possible lower bound on $\mathbf{D}$ to be

$$\mathbf{D}_0 = \operatorname{diag}(\mathbf{S}^{-1})^{-1}. \tag{4.11}$$

For the evaluation of MLE this value is proved to be a good starting point for all the iterative methods discussed in this thesis. However this only provides an initial guess for $\mathbf{D}$. The following theorem can be used to find an initial guess for $\mathbf{A}$.

**Theorem 1.** *Let the eigenvalue decomposition of* $\tilde{\mathbf{S}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{S}\mathbf{D}^{-\frac{1}{2}}$ *be*

$$\mathbf{Q}\boldsymbol{\Theta}\mathbf{Q}^H.$$

*The MLE of* $\mathbf{A}$*, is given by the columns of* $\mathbf{D}^{\frac{1}{2}}\tilde{\mathbf{A}}$ *that are given by* $\tilde{\boldsymbol{a}}_i = c_i \boldsymbol{q}_i$*, where* $c_i = [max(\theta_{ii} - 1, 0)]^{\frac{1}{2}}$ *for* $i = 1...m$*.*

Using (4.4) and (4.5) the prove given in [8] for real numbers can easily be extended to the case of complex numbers. [7] gives an alternating prove for this theorem that could be extended to complex numbers using (5.1) .

### 4.1.3 KLD Minimization

Another method to find the MLE is by the use of information geometric methods like KLD. Seghouane [14] uses this method to find the MLE of the FA model in the case of

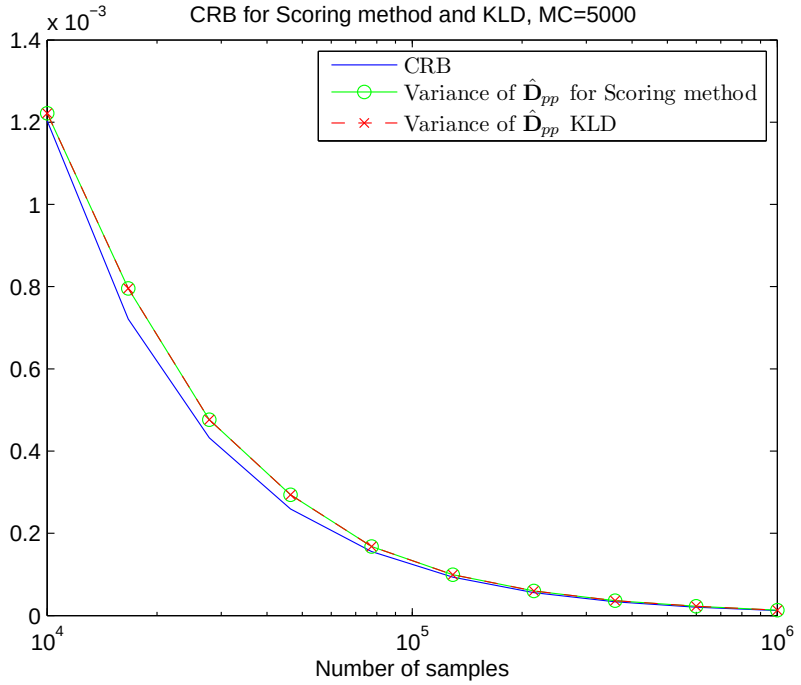



Figure 4.1: CRB for the Scoring and KLD algorithm

real numbers. Here this methods is extended for the case of proper complex Gaussian distribution.

Using the properties of complex Gaussian PDF as given by [10, p.582]. It is straightforward to follow the steps in [14]. The final iteration routine is given by

$$\hat{\mathbf{A}}_{j+1} = \mathbf{S}\hat{\mathbf{\Sigma}}_j^{-1}\hat{\mathbf{A}}_j\mathbf{\Phi}_j^{-1} \tag{4.12}$$

$$\hat{\mathbf{D}}_{j+1} = \mathrm{diag}(\mathbf{S} - \hat{\mathbf{A}}_{j+1}\hat{\mathbf{A}}_j^H\hat{\mathbf{\Sigma}}_j^{-1}\mathbf{S}) \tag{4.13}$$

where

$$\mathbf{\Phi} = \mathbf{I} - \mathbf{A}^H\mathbf{\Sigma}^{-1}\mathbf{A} + \mathbf{A}^H\mathbf{\Sigma}^{-1}\mathbf{S}\mathbf{\Sigma}^{-1}\mathbf{A};$$

This algorithm is a variation of EM algorithm for finding MLE. As such it shares the same convergence properties of the EM algorithm. This means that the algorithm will at least converge to a local maximum and that the convergence is guaranteed [10, p187][14].

### 4.1.4 Summery

The equations that are needed in order to find the MLE for the complex FA have been derived and two algorithms have been suggested to find a numerical approximation of MLE.

Both KLD and MLE need a reasonable initial guess to help them converge to global maximum. An initial guess based on the sample covariance matrix has been suggested.

The asymptotic distribution of the MLE under some "regularity" conditions, is given by [10, p.582]. Given the weak celestial sources, the assumption of large amount of data samples in concept of radio-astronomy is usually true. As a result the asymptotic covariance matrix of the parameter vectors as given in Chapter 3 is the corresponding sub-matrix in the inverse of the Fisher information matrix. Figure 4.1 illustrates the convergance of the MLE algorithm to the CRB.

## 4.2 Alternating Least Squares

In the previous section the MLE for the FA model was derived. The log-likelihood function was maximized with respect to the unknown parameters. In this section another function of the unknown parameters is considered.

For a given sample covariance matrix $\mathbf{S}$

$$\left\|\mathbf{S} - \mathbf{A}\mathbf{A}^H - \mathbf{D}\right\|_F^2 \tag{4.14}$$

is the Frobenius norm that is minimized with respect to model parameters $\mathbf{A}$ and $\mathbf{D}$.

This problem is approached as a two stage minimization problem. First for a given $\mathbf{A}$, (4.14) is minimized with respect to $\mathbf{D}$ and in the next stage, $\mathbf{D}$ is held constant and a new $\mathbf{A}$ is found.

The iteration steps are as follow

$$\hat{\mathbf{D}}_{k+1} = \mathrm{diag}(\mathbf{S} - \hat{\mathbf{A}}_k \hat{\mathbf{A}}_k^H) \tag{4.15}$$

$$\hat{\mathbf{A}}_{k+1} = \mathbf{U}_m \mathbf{L}_m^{\frac{1}{2}} \tag{4.16}$$

where $\mathbf{L}_m$ and $\mathbf{U}_m$ are $m$ largest eigenvalues and the corresponding eigenvectors of the matrix $\mathbf{S} - \hat{\mathbf{D}}_{k+1}$ respectively

$$(\mathbf{S} - \hat{\mathbf{D}}_{k+1}) = \mathbf{U}\mathbf{L}\mathbf{U}^{-1}. \tag{4.17}$$

This method will monotonically approach a local minimum. Its rate of convergence is slower than that of MLE.

## 4.3 Simulation Results

In this section, simulation results for some possible applications of the FA model will be presented. Two well-known subjects, direction of arrival estimation using ESPRIT and spatial filtering of interfering signals for radio astronomy will be illustrated using these simulations.

In Section 4.3.1 the FA model is used as a calibration step or pre-processing for ESPRIT algorithm. Section 4.3.2 illustrates the possibility of filtering interfering signals on an uncalibrated array.

### 4.3.1 DOA Estimation with ESPRIT

One of the widely used subspace algorithms for the direction of arrival (DOA) and time delay estimation is ESPRIT. The noise model usually used for this algorithm assumes independent and identically distributed (i.i.d) noise sources for the receiving elements. In this section the performance of the ESPRIT algorithm is evaluated both with and without pre-processing by FA in the case that noise sources on each channel are no longer identically distributed. The evaluation is done by the means of simulations that are programmed using MATLAB.

One of the requirements for the use of ESPRIT is that the array is linear and uniform. This guarantees the shift-invariant structure in $\mathbf{A}$.

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{D} \tag{4.18}$$

For a uniform and linear array $\mathbf{A}$ can be written as:

$$\mathbf{A} = \begin{pmatrix} 1 \\ e^{-j2\pi\Delta sin(\theta)} \\ \vdots \\ e^{-j2\pi(p-1)\Delta sin(\theta)} \end{pmatrix} \tag{4.19}$$

where $j = \sqrt{-1}$, $\Delta$ is the distance between the elements per wavelength and $\theta$ is the angle from broadside.

If $\mathbf{D}$ in (4.18) is equal to $\sigma^2\mathbf{I}$ then an identical performance is expected for both ESPRIT with and without pre-processing by FA. In order to make the difference in performance visible $\mathbf{D}$ must deviate from $\sigma^2\mathbf{I}$. In these simulation the DOA of different sources is estimated. the parameters that are changed are number of samples that are used for the estimation and the deviation from $\mathbf{I}$. Power of both sources is chosen to be 0dB in this way the different variances for the noises can easily be seen as different signal-to-noise ratios (SNR) for each receiving element.

The deviation from $\mathbf{I}$ is realized by choosing a minimum and a maximum SNR in dB for two of the receiving elements and then calculating the SNR for the other elements by linearly interpolating between these two extremes.

For this series of simulations there are two sources at $-20$ and 30 degrees. For the first simulation all the receiving elements have a SNR of 0dB. For the other simulations deviations of 5,10 and 15dB around 0dB are used (e.g. $\mathrm{SNR}_{\mathrm{min}} = -20\mathrm{dB}$ and $\mathrm{SNR}_{\mathrm{max}} = 20\mathrm{dB}$).

The results that are presented in Figure 4.2 show the standard deviation , $\sigma_{\hat{\theta}}$, of the estimated angles in logarithmic scale. Also the CRB for the angles is shown.

As expected the performance in the case of $\mathbf{D} = \mathbf{I}$ is almost identical. However as the deviation increases, the FA model gives a lesser standard deviation than without it.

### 4.3.2 Spatial Filtering

To illustrate the advantage of the FA model in subspace estimation, and provide a good example for radio astronomy, spatial filtering of interfering signals is chosen. The

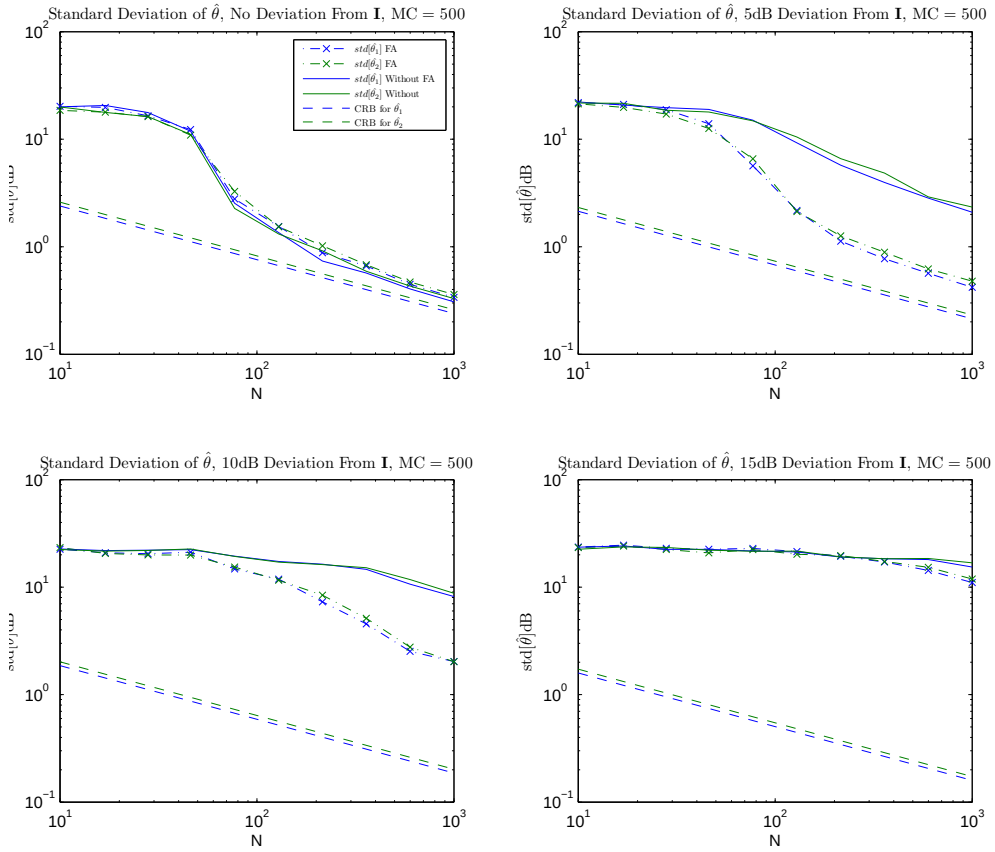



Figure 4.2: Standard Deviation of $\hat{\theta}$ with and without FA

algorithm that is presented here is the extension of the one used in [15] to the case that the covariance matrix of the noise can no longer be modeled as $\sigma^2\mathbf{I}$ and is unknown.

The data model give in [15] is

$$\mathbf{x}(t) = \mathbf{v}(t) + \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{4.20}$$

where $\mathbf{v}$ is the vector of sky sources, $\mathbf{s}$ is the vector of intefereing sources with a spatial signiture equlal to $\mathbf{A}$ and $\mathbf{n}$ is the noise vector. However power of the sky sources (about -20dB) is much smaller than the noise and interference power. This could be exploited to filter the interfering signals. For a short enough integration time (1-100ms [9, 15]) power of the sky sources could be neglected. Given $L = NM$ samples from $\mathbf{x}$ the aim is to filter the interfering signals and retrieve $\mathbf{R}_\mathbf{v}$. The short term sample covariance matrices are given by

$$\hat{\mathbf{R}}_k = \frac{1}{N}\sum_{n=kN}^{(k+1)N-1} \mathbf{x}[n]\mathbf{x}[n]^H, k = 0, \ldots, M-1. \tag{4.21}$$

With the assumption that the spatial signature of the interfering signal is stationary

during the short-term integration and not stationary between them, an estimate of $\mathbf{R}_0 = \mathbf{R_v} + \mathbf{\Sigma_n}$ could be found from the following procedure.

The short-term projection matrix

$$\mathbf{P}_k = \mathbf{I} - \mathbf{A}_k(\mathbf{A}_k^H \mathbf{A}_k)^{-1}\mathbf{A}_k^H \tag{4.22}$$

is used to filter any signal lying in the subspace of the interfering signal. The filtered short-term sample covariance becomes

$$\tilde{\mathbf{R}}_k = \mathbf{P}_k \hat{\mathbf{R}}_k \mathbf{P}_k. \tag{4.23}$$

Thus the long-term estimate of covariance matrix after filtering can be written as

$$\tilde{\mathbf{R}} = \frac{1}{M}\sum_{k=0}^{M-1} \tilde{\mathbf{R}}_k = \frac{1}{M}\sum_{k=0}^{M-1} \mathbf{P}_k \hat{\mathbf{R}}_k \mathbf{P}_k. \tag{4.24}$$

[15] shows that an unbiased estimate of $\mathbf{R}_0$ is given by

$$\mathrm{vec}(\hat{\mathbf{R}}) = \mathbf{C}^{-1}\mathrm{vec}(\tilde{\mathbf{R}}) \tag{4.25}$$

where $\mathbf{C} = \frac{1}{M}\sum_{k=0}^{(M-1)}(\mathbf{P}_k^T \otimes \mathbf{P}_k)$ and $\otimes$ is the Kronecker product. However in order to calculate this estimate first the subspace of the interfering signals needs to be estimated. For the case that $\mathbf{\Sigma_n} = \sigma^2\mathbf{I}$ with known $\sigma$ the eigenvalue decomposition could be used to find the needed subspace.

If the $\mathbf{\Sigma_n}$ is an unknown diagonal matrix, the short-term covariance could be modeled as

$$\hat{\mathbf{R}}_k \approx \mathbf{A}_k \hat{\mathbf{R}}_{\mathbf{s}_k} \mathbf{A}_k^H + \mathbf{D} \tag{4.26}$$

where $\mathbf{\Sigma_n} = \mathbf{D}$ is a real diagonal matrix.

A simulation is programmed in MATLAB to illustrate the use of the FA model for spatial filtering. It is assumed that the system has 5 receiving elements and there are 2 interfering signals. Figure **??** shows the results of the simulations. The Frobenius norm is used to calculate the mean square error between the estimated covariance matrix $\tilde{\mathbf{R}}$ and the true covariance matrix $\mathbf{R}_0$. The MSE is calculated as a function of interference-to-noise-ratio (INR). For these simulations a deviation of 5 dB from $\mathbf{I}$ has been chosen.

The constant false alarm test that will be introduced in Chapter 5 is used to see if it is necessary to estimate the model parameters. Also the true spatial signature of the interfering signals is used as a reference. To show the improvement made by using the FA model, the eigenvalue method is also used.

In [15] using a standard likelihood test (white-noise test with known $\sigma$) the number of interfering signals and their subspace is detected. This technique needs the full knowledge of noise covariance matrix. The FA model gives similar performance without needing this knowledge.

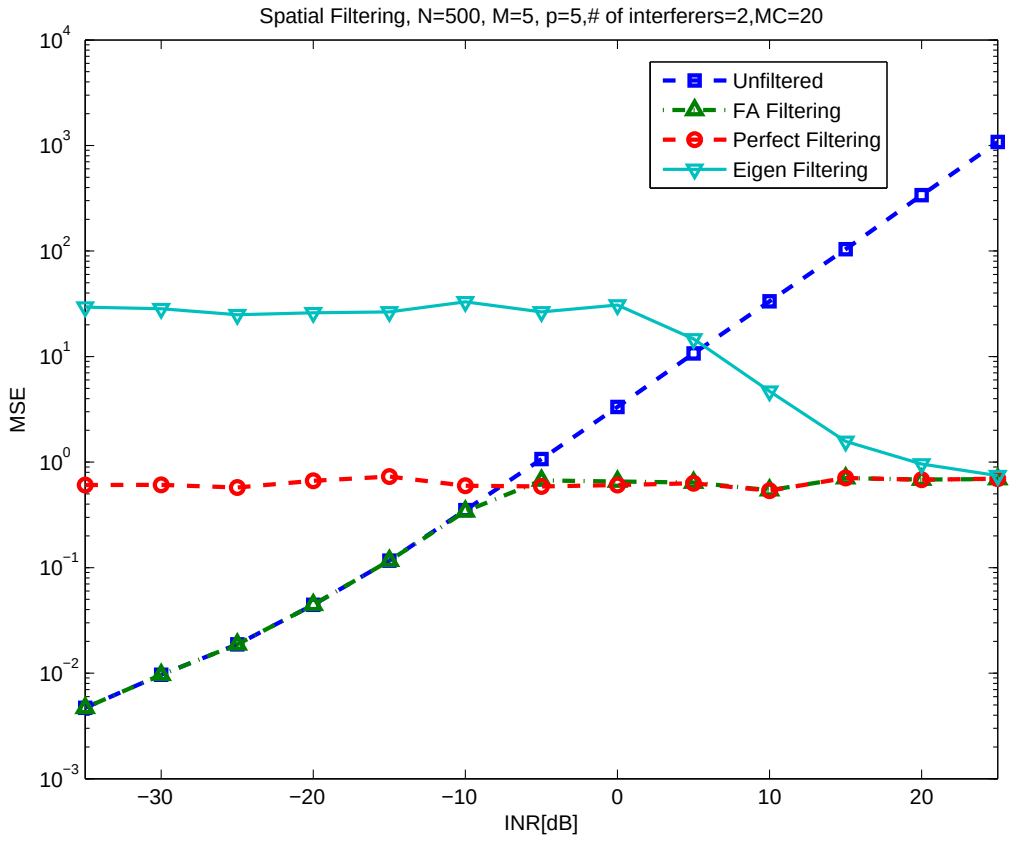



Figure 4.3: Spatial Filtering Using FA

## 4.4 Conclusions

It is shown that the model parameters for the complex valued data could indeed be estimated. Two different algorithms have been suggested to find the MLE for unknown parameters.

The Fisher score is derived and used to find the iterations needed for scoring method. Also the KLD algorithm was extended to the case of complex numbers. The convergence of the estimator to CRB is shown with the help of a simulation.

The alternating least squares for the FA model was also presented. This method illustrates the close relationship between the FA model PCA. This method gives constant estimates but is (asymptoticly) not efficient [8].

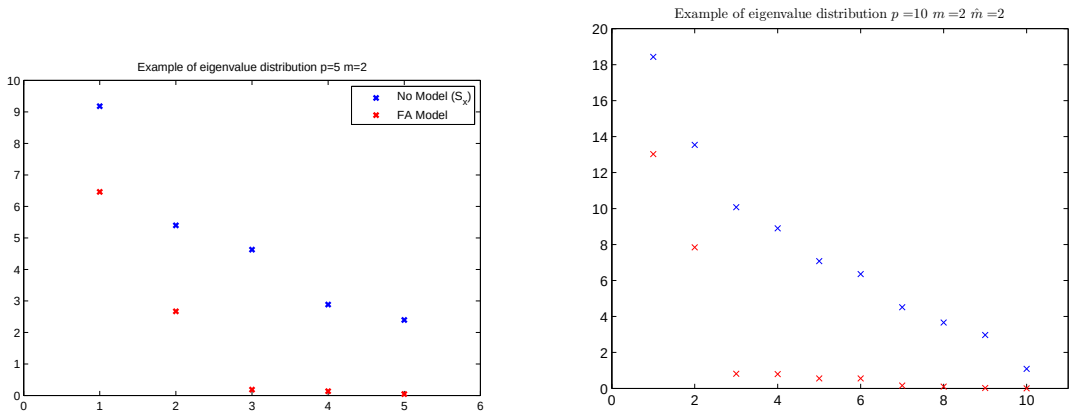
All of the discussion in this chapter assumes proper complex Gaussian distribution. Generalization to other distributions for both real and complex case of FA model is still an open question. Also the number of sources is assumed to be known in advanced. In Chapter 5 a GLRT is proposed to see if the chosen value for $m$ is acceptable.

# 5 Detection

One of the parameters that needs to be found for the FA model is the number of underlying common factors. In the concept of array-processing it is the same as the number of sources that the array is exposed to.

In this section a general likelihood ratio test (GLRT) is used to decide whether the FA model fits a given sample covariance matrix. If the model fits, it can be seen as detecting $m$ sources.

Because the noise is not identically distributed, using the spreading of eigenvalues for detecting the number of sources will be very difficult. Figure 5.1 illustrates two examples and also shows the improvement made by using FA model. In Figure 5.1b $\hat{m}$ is the the dimension for which the FA model has been calculated.




(a) (b)

Figure 5.1: Eigenvalue distribution

## 5.1 General Likelihood Ratio Test

Given (2.7) two hypotheses are tested against each other. $H_0$ assumes that there is an FA model underlying the data, while $H_1$ assumes no structure. Consider the following test

$$\lambda = \frac{L_1^*}{L_0^*} > \gamma$$

where $L_1^*$ is the maximum value of the likelihood when $H_1$ is true, and $L_0^*$ is maximum value of the likelihood for a FA model.

$$\log(L_0^*) = l_0^* = -N \log |\pi^p| + N \log |\mathbf{\Sigma}^{-1}| - N\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{S})$$

$$\log(L_1^*) = l_1^* = -N \log|\pi^p| - N \log|\mathbf{S}| - Np$$

putting these back in for $\lambda$

$$\log(\lambda) = N\mathrm{F}(\mathbf{\Sigma}) = N\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) - N \log|\mathbf{\Sigma}^{-1}\mathbf{S}| - Np$$

$N$ is the number of samples. For a given $\mathbf{S}$, maximizing the likelihood for the FA model is the same as minimizing F.

$$\mathrm{F}(\mathbf{A}, \mathbf{D}; \mathbf{S}) = \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) - \log|\mathbf{\Sigma}^{-1}\mathbf{S}| - p \tag{5.1}$$

$$2\log(\lambda) = 2N\mathrm{F}(\mathbf{A}, \mathbf{D}; \mathbf{S})$$

The statistic $2\log(\lambda)$ has an asymptotic $\chi_s^2$ distribution under $H_0$ where $s$ is the error degree of freedom [7, p.267][8, p.281]. For the complex case, the degree of freedom is slightly different than for the real case. The total number of parameters that are needed to be fixed are the parameters of $\mathbf{\Sigma}$ covariance matrix. $\mathbf{\Sigma}$ is Hermitian so these are $p$ elements on its diagonal, and $p(p-1)$ off-diagonal elements. The number of free parameters in the FA model are $2mp$ for the elements of $\mathbf{A}$ matrix and $p$ for the elements of $\mathbf{D}$. However (2.9) imposes $m^2$ constraints on the FA model. So the free parameters in $\mathbf{\Sigma}$ that are not fixed by FA model are

$$\begin{aligned} s &= p + p(p-1) - [2pm + p - m^2] \\ &= (p-m)^2 - p > 0. \end{aligned} \tag{5.2}$$

This number of free parameters gives a constraint on the number of sources that the FA model can separate for a given number of receiving elementis.

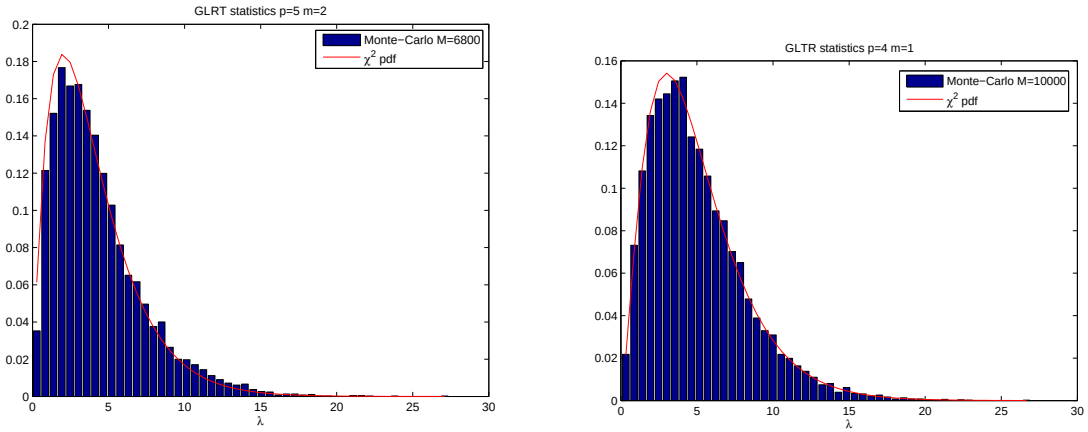
Bartlett has suggested [7, p.267][8, p.281] a modified version of this test to improve its $\chi^2$ statistics by replacing $N$ with

$$N^* = N - \frac{2p+11}{6} - \frac{2}{3}m$$

Figure 5.2 illustrates the results of Monte-Carlo simulations and its statistics. As it is shown in these figures the results follow the $\chi^2$-distribution very closely.

## 5.2 Constant False Alarm Detector

In the special case that $\hat{m} = 0$ this test becomes a constant false alarm detector. Figure 5.3 shows the results of Monte-Carlo simulations for various false alarm probabilities. The results match the predicted theoretical values closely and show the correctness of the thresholds based on $\chi^2$-distribution. This detector could be extended to detect the number of signals by increasing $\hat{m}$ each time the $\chi^2$-test fails.




(a) (b)

Figure 5.2: GLRT statistics for FA model

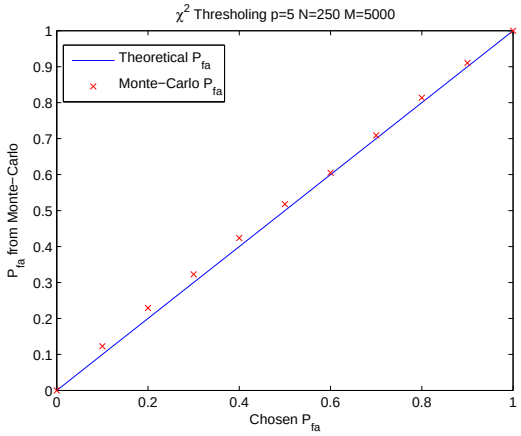



Figure 5.3: Constant False Alarm Simulations

## 5.3 Conclusions

It is shown that for a given number of sources the validity of the model can be test using a GLRT. If the test fails then either the estimated number of sources is wrong or the FA model is not appropriate for the given data collection. In the case that the test passes it can be concluded that $m \leq \hat{m}$.

With the help of simulations it is shown that the test has a asymptotic $\chi^2$ distribution. Also the number of free parameters that are found for the complex FA model are correct. The calculated number of free parameters puts a limit on the number of sources that could be separated for a given number of receiving elements.

# Conclusions and Suggestions for Further Research

# 6

In this chapter the main conclusions and results are presented. Also some recommendations for further study are given.

## 6.1 Conclusions

It has been shown that the complex factor analysis model is the same model which is commonly used in literature for narrow-band array processing. The CRB and the Fisher score for the case of proper complex Gaussian distribution are found using Wirtinger derivatives. Most of the techniques used in factor analysis for real numbers are extended to the case of the complex numbers. With the help of simulations it has been demonstrated that the factor analysis can be used to improve the performance of current algorithms when the array is not calibrated. DOA estimation and spatial filtering of the interfering signals for radio astronomy has been studied.

Various estimation techniques have been presented. The Fisher scoring and KLD algorithm has been given as iterative methods to find MLE of complex parameters. The alternating LS algorithm is discussed. The asymptotic convergence of the MLE techniques to the CRB has been shown using MC simulations.

The validity of the estimated parameters for a given number of sources is tested with the help of a GLRT. An especial case of this test is presented as a constant false alarm ratio detector. A threshold based on $\chi^2$ test is proposed. The maximum number of sources that can be detected for a given number of receiving elements is given.

## 6.2 Future Work

There are some aspects and applications of the FA that remain unexplored.

- The computational complexity of the algorithm needs to be evaluated thoroughly and more efficient algorithms are needed. This is especially important when the number of receiving elements becomes very large.
- Fisher scoring and MLE rely heavily on the proper Gaussian distribution. The performance of these algorithms needs to be evaluated if the underlying signal has a different distribution.
- The false alarm detector that is introduced in Chapter 5 only differentiates between the case of no signal and an unknown number of signals. New techniques to detect the number of signals could be studied. Threshold for sequential test for FA is an open question.

- Sample noise covariance matrix is almost never completely diagonal. By forcing the estimate of this covariance matrix to a perfectly diagonal matrix, the error in the estimated subspace becomes larger. How much imperfection could one allow in the estimate of the noise covariance matrix, and how much improvement this could have on the estimated subspaces needs to be answered.

# Bibliography


[1] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.

[2] D. J. Bartholomew, M. Knott, and I. Moustaki, *Latent Variable Models and Factor Analysis: A Unified Approach*. John Wiley and Sons, 2011.

[3] R. A. Reyment and K. G. Jöreskog, *Applied Factor Analysis in the Natural Sciences*. Cambridge University Press, 1996.

[4] L. L. Scharf, S. Kraut, and M. L. McCloud, "A review of matched and adaptive subspace detectors," in *Adaptive Systems for Signal Processing, Communications, and Control Symposium*, IEEE, 2000.

[5] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, vol. 42, AUGUST 1994.

[6] C. Spearman, "The proof and measurement of association between two things," *The American Journal of Psychology*, vol. 15, pp. 72–101, Jan 1904.

[7] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. ACADEMIC PRESS, 1997.

[8] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. Wiley, third edition ed., 2003.

[9] S. J. Wijnsholds, *Fish-Eye Observing with Phased Array Radio Telescope*. PhD thesis, Delft University of Technology, 2010.

[10] A. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation theory*, vol. Volume I. Prentice Hall, 1993.

[11] K. G. Jöreskog, "A general approach to confirmatory maximum likelihood factor analysis," *PSYCHOMETRIKA*, vol. 34, no. 2, 1969.

[12] P. J. Schreier, *Statistical Signal Processing of Complex-Valued Data*. Cambridge University Press, 2010.

[13] A. K. Jagannatham and B. D. Rao, "Cramerrao lower bound for constrained complex parameters," *IEEE SIGNAL PROCESSING LETTERS*, vol. 11, NOVEMBER 2004.

[14] A. K. Seghouane, "An iterative projections algorithm for ml factor analysis," in *IEEE Workshop on Machine Learning for Signal Processing*, pp. 333 –338, oct. 2008.

[15] J. Raza, A.-J. Boonstra, and A.-J. van der Veen, "Spatial filtering of rf interference in radio astronomy," *IEEE SIGNAL PROCESSING LETTERS*, vol. 9, FEBRUARY 2002.