On Reduced Rank Linear Regression and Direction Estimation in Unknown Colored Noise Fields

KARL WERNER

Licentiate Thesis
Stockholm, Sweden 2005

TRITA–S3–SB–0563
ISSN 1103–8039
ISRN KTH/SB/R - - 05/63 - - SE
On Reduced Rank Linear Regression and Direction Estimation in Unknown Colored Noise Fields

KARL WERNER

Licentiate Thesis
Stockholm, Sweden, 2005
Abstract

Two estimation problems are treated in this thesis. Estimators are suggested and the asymptotical properties of the estimates are investigated analytically. Numerical simulations are used to assess small-sample performance. In addition, performance bounds are calculated.

The first problem treated is parameter estimation for the reduced rank linear regression. A new method based on instrumental variable principles is proposed and its asymptotical performance analyzed. In addition, the Cramér-Rao bound for the problem is derived for a general Gaussian noise model. The new method is asymptotically efficient (it has the smallest possible covariance) if the noise is temporally white, and outperforms previously suggested algorithms when the noise is temporally correlated. The approximation of a matrix with one of lower rank under a weighted norm is needed as part of the estimation algorithm. Two new, computationally efficient, methods are suggested. While the general matrix approximation problem has no known closed form solution, the proposed methods are asymptotically optimal as part of the estimation procedure in question. A new algorithm is also suggested for the related rank detection problem.

The second part of this thesis treats direction of arrival estimation for narrowband signals using an array of sensors. Most algorithms require the noise covariance matrix to be known (up to a scaling factor) or to possess a known structure. In many cases the noise covariance is in fact estimated from a separate batch of signal-free samples. This work addresses the combined effects of finite sample sizes both in the estimated noise covariance matrix and in the data with signals present. No assumption is made on the structure of the noise covariance. The asymptotical covariance of weighted subspace fitting (WSF) is derived for the case in which the data are whitened using the noise covariance estimate. The obtained expression suggests an optimal weighting that improves performance compared to the standard choice. In addition, a new method based on covariance matching is proposed. Both methods are asymptotically statistically efficient. The Cramér-Rao bound for the problem is derived, and the expression becomes surprisingly simple.
Acknowledgements

First of all, I would like to thank my supervisor Magnus Jansson for all his patient support for a from time to time quite slowly progressing student. He has been a never ending source of ideas and encouragement.

I would also like to thank prof. Per-Ola Börjesson and prof. Per Ödling at Lund University for introducing me to research in signal processing, and also for providing the first contact with KTH.

I would like to thank prof. Mats Viberg for acting opponent to this thesis.

I am very lucky to be part of the research group at floor four. I think such relaxed and cheerful, yet seriously professional, colleagues and coworkers provide the best possible environment for research (and for having a good time). One on them, Mats Bengtsson, has been very helpful when trying to make all the technicalities of writing a thesis right. My fellow muddy hill running connoisseur Niklas Wernersson has proofread parts of the thesis.

My running mates in FK Studenterna have done a tremendous job in keeping my thoughts far away from signal processing during many hours of sweating and panting and have thus helped in providing the necessary mental recovery.

I would also like to thank my family for always being loving, encouraging and supportive to me, and for all those lively discussions about everything from moths to intellectual property law. Last, but definitely not least, I would like to thank Helena for all her love. When I'm with you, I know that there is more to life than just x, y and ∞.
Contents

Acknowledgements v

Contents vi

1 Introduction 1
   1.1 The topics of the thesis .............................. 1
   1.2 Contributions ......................................... 6
   1.3 Notation ............................................. 8
   1.4 Abbreviations ....................................... 10

I Reduced rank linear regression and weighted low rank approximations 11

2 Introduction to reduced rank linear regression 13
   2.1 Background ........................................ 14
   2.2 Outline of Part I .................................. 15

3 An instrumental variable method for the reduced rank linear regression 17
   3.1 The proposed method ................................ 17
   3.2 Estimation of the weighting matrix ................. 19
   3.3 Weighted low rank approximation .................... 20
   3.4 Detecting the rank of the regression ............... 24
   3.5 The proposed algorithm step-by-step ............... 26
   3.6 A useful gradient and Hessian ..................... 27

4 Asymptotical analysis and performance bounds for the proposed IV method 31
   4.1 Asymptotical covariance of the estimate ............ 31
   4.2 The Cramer-Rao bound ................................ 33
   4.3 Relation to previous methods ...................... 37
   4.4 Equivalence of the proposed estimator to Maximum Likelihood . 38
5 Numerical study 41

6 Introduction to DOA estimation in the presence of colored noise 51
  6.1 Outline of Part II ........................................... 52
  6.2 Data model ...................................................... 53

7 Bounds on the performance of the DOA estimate 55
  7.1 The Cramér-Rao bound with signal-free samples available ....... 56
  7.A Proof of Theorem 1 ............................................. 58

8 Algorithms for DOA estimation with asymptotical performance analysis 65
  8.1 Whitened weighted subspace fitting .......................... 66
  8.2 A covariance matching approach .............................. 69
  8.A Proof of Theorem 2 ............................................ 70

9 Simulations 75
  9.1 Simulation setup .............................................. 75
  9.2 Results .......................................................... 76

10 Conclusions 85
  10.1 Part I- The reduced rank linear regression .................. 85
  10.2 Part II- DOA estimation in unknown colored noise fields .... 86
  10.3 Future work ................................................... 86

Bibliography 89
Chapter 1

Introduction

The chapter begins with a brief description of the subject of estimation theory. Two examples, partly related to the two main topics of the thesis, are then presented. Then the contributions and publications of the author are highlighted. The chapter is concluded with lists of the notation and abbreviations used.

1.1 The topics of the thesis

Estimation theory is about extracting information from measurements. In order to make the task tractable it is necessary to know, or at least postulate, in what way the measurements depend on the information of interest. In the technical literature, such a dependence is called a model. In estimation theory, the sought information is in the form of a set of scalars, called parameters. A device that processes measurements, data, in order to produce an estimate of the parameters in a particular model is called an estimator. According to a common convention, let the vector of the true values of the parameters be denoted by $\theta_0$, and the vector of estimated parameters be denoted by $\hat{\theta}$.

It is common to assume a random (stochastic) model for the measurements. When adopting such a model one assumes that the measurements are realizations of a random process. The model is then the probability density function of the measurements, and the parameters of interest affect the shape of this function. In this thesis, the measurements are random but the parameters to be estimated are fixed, deterministic.

This “stochastic” approach has several advantages. The most important one is that it allows us to answer questions about how the estimator will behave on the average if we repeat the measurement and estimation procedures a large number of times. If the measured data are stochastic, then, since the estimator is a function of the data, the estimate is also a stochastic variable. A natural and commonly used performance measure is the mean squared error (MSE). It is the average squared
deviation from the true parameter values,
\[ E \left[ (\hat{\theta} - \theta_0)^* (\hat{\theta} - \theta_0) \right]. \] (1.1)

A straightforward design procedure would then be to simply select the estimator that minimizes the MSE. However, typically, such an estimator depends on \( \theta_0 \) and would therefore not be realizable.

Often only estimators that on the average produce the correct parameter value (for all possible parameter values) are considered. Such estimators are \emph{unbiased}. For unbiased estimators, the MSE is the trace of the covariance of the estimate. Thus, when comparing the performance of unbiased estimators, the covariance is a quantity of great interest.

The stochastic data model is useful also when designing estimators: It provides a theoretical foundation for constructing measures of goodness, criterion- or loss-functions, that the estimator can be designed to optimize.

As we will see later on in this thesis, a stochastic model for the data allows us to make statements about limits on the attainable performance of an estimator. The Cramér-Rao bound (CRB) provides such a limit. As it is usually formulated, the CRB is a lower bound on the covariance of \emph{unbiased} estimators. There are, however, extensions to estimators with known bias gradient, see, e.g., [Eld04] and the references therein.

The following example uses a data model that is common in signal processing. The example is simple, but it illustrates some of the concepts mentioned above. It is also related to the results given in Part I of this thesis. Similar expositions are given in virtually any book on estimation theory, see, e.g., [SS89a], [Lju99], [Kay93].

\textbf{Example-linear regression}

Let the measurement at time \( t \) be denoted by \( y(t) \). The measured output at time \( t \) depends on the input \( x(t) \) and a disturbance term, \( n(t) \). Assume the model
\[ y(t) = Cx(t) + n(t). \] (1.2)

The vector \( n(t) \) accounts for uncertainties in the model and disturbances in the measurements. Assume that measurements are observed for \( t = 1, \ldots, N \). An example in which \( x(t) \) and \( y(t) \) both have dimension two is shown in Figure 1.1.

In that example
\[ C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}. \] (1.3)

The matrix \( C \) is unknown, and the task is to estimate it. It is natural to choose \( \hat{C} \), the estimate of \( C \), so that as little as possible of the measurements need to be attributed to the \( n(t) \)-term. Before returning to what is meant by this, note that the model can be rewritten as
\[ y = \Phi \theta + n, \] (1.4)
1.1. THE TOPICS OF THE THESIS

where the matrix $\Phi$ contains the elements of $x(t)$ for $t = 1, \ldots, N$ arranged in a suitable way and the vector $\theta$ contains the elements of $\mathbf{C}$. The vector $y$ contains the elements of $y(t)$ for $t = 1, \ldots, N$, and $\mathbf{n}$ contains $n(t)$ in the same way. Gauss suggested minimizing the “squares of the errors” in 1809 when faced with a similar problem. The same approach will be taken here. The estimate $\hat{\theta}$, or equivalently $\hat{\theta}$, will be chosen so that the criterion

$$n^T W n = (y - \Phi \theta)^T W (y - \Phi \theta)$$

(1.5)

is minimized. The matrix $W$ is any positive definite matrix. It should be considered a design parameter. If the matrix $\Phi$ has full column rank this gives

$$\hat{\theta} = (\Phi^T W \Phi)^{-1} \Phi^T W y.$$  

(1.6)

This is the well-known least-squares (LS) estimate.

Now, how do we know if this is a good or a bad estimator? One way would be to test it! But, when testing, how do we generate the measurements, and in particular the disturbance term? Obviously, without characterizing the disturbances, it is not possible to make any statement about the performance of the estimator. This leads to introducing a stochastic model for the measurements. Let

$$y \sim N(\Phi \theta, Q),$$

(1.7)

which means that each measured vector is a realization of a multidimensional Gaussian stochastic variable with expected value $\Phi \theta$ and covariance $Q$. Often, it is also assumed that each observation is statistically independent of the other. The noise model is then “temporally white”. It is easy to show that the LS-estimator is unbiased,

$$\mathbb{E} \left[ \hat{\theta} \right] = \theta$$

(1.8)

and that it has covariance

$$\text{Cov} \left[ \hat{\theta} \right] = (\Phi^T W \Phi)^{-1} \Phi^T W Q W \Phi (\Phi^T W \Phi)^{-1}.$$  

(1.9)

The latter expression allows us to select the weighting matrix $W$ so that the trace of the covariance is minimized. If $W = Q^{-1}$ then

$$\text{Cov} \left[ \hat{\theta} \right] = (\Phi^T Q^{-1} \Phi)^{-1}.$$  

(1.10)

In fact, the last expression equals the CRB for the problem, and thus no unbiased estimator can have a smaller covariance. Note however that, among biased estimators, it is possible to find estimators that yield a lower MSE (at least in a specific range of parameter values). Such estimators have been the topic of much research, see e.g. [EM05].
Figure 1.1: Standard linear regression. The output is a linear combination of the inputs and the noise.

What would happen if we knew something more about the model in the previous example? Assume that it is known that the “unknown device” described by the matrix $\mathbf{C}$ in the last example has the structure of Figure 1.2? In that model, the output $y(t)$ depends on the noise $\mathbf{n}(t)$ and some internal signal $\mathbf{z}(t)$ that is of dimension one! The internal signal $\mathbf{z}(t)$ cannot be measured, but it depends linearly on the known input $\mathbf{x}(t)$. This is an example of a reduced rank linear regression. With the notation of the last example it is possible to write

$$\mathbf{C} = \mathbf{A}\mathbf{B}^T,$$

with suitable dimensions of the matrices. If this kind of structure can be assumed, the accuracy when performing parameter estimation can be improved. This is the topic of Part I in this thesis.

**Narrowband direction finding**

Array signal processing, and in particular direction of arrival (DOA) estimation, has applications in diverse areas such as radar, sonar, biomedicine, communications systems and seismic exploration. The example below illustrates a data model that is commonly used in the array signal processing field. For an overview of array signal processing, and in particular direction of arrival estimation, see [KV96].

Consider the antenna array in Figure 1.3. The objective is to estimate the direction of arrival $[\theta]_1$ (the reason for the cumbersome notation will be clear later on) of the radio signal that is transmitted from the mobile phone. In the picture there is only one transmitter. The system has $m$ antennas. Each antenna is equipped with a mixer/filter and an analog-to-digital converter (ADC) that generates a complex,
1.1. THE TOPICS OF THE THESIS

![Diagram of a reduced rank linear regression model]

Figure 1.2: Reduced rank linear regression. The output is still a linear combination of the input and the noise, but there is extra structure in the model.

The sampled baseband equivalent of the received signal. The output from ADC $i$ at time $t$ is modelled

$$[x(t)]_i = [a(\theta)]_i[s(t)]_1 + [n(t)]_i,$$

where $[a(\theta)]_i$ is the response of antenna $i$ to a transmitter with DOA $\theta$. The signal $[s(t)]_i$ is the baseband equivalent of the signal that is transmitted from transmitter 1 at time $t$. The signal $[n(t)]_i$ is the noise at receiver $i$ at time $t$. The functional form of $[a(\theta)]_i$ depends on the array geometry and other array-specific parameters. It is assumed to be known.

It is easy to see that the output from the array at time $t$ can be collected in a vector $x(t)$ that can be written

$$x(t) = a(\theta)[s(t)]_1 + n(t).$$

The vector $a(\theta)$ is called the array steering vector. If there are $d$ transmitters, with DOAs at $[\theta]_1, \ldots, [\theta]_d$ collected in $\theta$ then linearity gives that

$$x(t) = A(\theta)s(t) + n(t).$$

Here, the columns of $A(\theta)$ are the array steering vectors. The vector $s(t)$ collects the signals transmitted from each source. It is possible to estimate more than one parameter per source.

The problem of estimating $\theta$ from a batch of measurements, $\{x(t)\}_{i=0}^{N-1}$ in the model (1.14) has been the topic of intensive research. A stochastic model for the noise, $n(t)$, is generally assumed, and often the signals are assumed to be stochastic as well. Typically, the noise covariance is assumed to be known or to have a
particular structure. In Part II of this thesis it is assumed that the noise covariance is completely unknown. Instead the noise is observed for a period of time when there are no signals. Estimation methods and performance bounds are developed.

![Diagram](image)

Figure 1.3: A simple model of an antenna array system for direction finding. The picture shows antennas, mixers and analog-to-digital converters.

### 1.2 Contributions

This thesis is divided into two parts, corresponding to two estimation problems. A technical description of each of the two estimation problems, together with a literature overview will be given in the introductory chapter of each part. In the following section, focus will be on highlighting the contributions of the author in each part.

**Part I**

This part deals with the problem of reduced rank linear regression. In Chapter 2, the estimation problem is introduced with a brief motivation and a literature overview.

In Chapter 3, the proposed estimator is presented and motivated. The proposed algorithm has an advantage over previously suggested algorithms when the noise is temporally colored. In addition, two novel methods for solving the related *weighted low rank approximation* (WLRA) problem are presented. The new methods are...
1.2. CONTRIBUTIONS

optimal when part of the estimation procedure proposed. Except for the estimation procedure, Chapter 3 includes a rank detection scheme that is based on the loss function of the proposed method.

Chapter 4 includes a novel asymptotical analysis of the performance of the proposed method. The CRB for the problem is also derived. The chapter is concluded with a comparison of the proposed method to other methods in the literature. Part I is concluded with Monte-Carlo type simulations that illustrate and support the results.

Most of the results of the first part of the thesis have previously been published in


The main contributions in this part of the thesis are:

- The use of WLRA together with an Instrumental Variable (IV) method for the reduced rank linear regression problem.

- The two computationally efficient single-update step based algorithms for the WLRA (that lacks a closed form solution in the general case). The algorithms are asymptotically optimal as parts of the estimation procedure proposed.

- The asymptotical analysis of the performance of the proposed algorithm.

- The exact equivalence to ML in a special case.

- The rank detection algorithm.

- The derivation of the CRB for a general noise model.

Part II

This part deals with direction of arrival estimation for narrowband signals impinging on an array of sensors. The sensor noise is assumed to be spatially correlated with an unknown correlation matrix. It is assumed that two batches of samples are available, one with signals plus noise, and one with only noise.

Chapter 6 introduces the problem and the data model used. A brief literature study is also included.
In Chapter 7, the Cramér-Rao lower bound for the problem is derived. It is found to have a simple form that can be compared to well-known results.

In Chapter 8, two algorithms for the estimation problem are considered. The first one is Weighted Subspace Fitting (WSF) applied to a whitened version of the data. The whitening transform is computed using the signal-free samples. The asymptotical covariance of the estimate is derived, and it turns out that the method can be made asymptotically efficient if a modified weighting is used instead of the standard WSF weighting. The second algorithm is a novel algorithm based on covariance matching. It is asymptotically efficient by design.

In Chapter 9, computer simulations are presented that serve as an illustration for the theoretical asymptotical results and helps evaluating small-sample performance of the proposed methods.

The results presented in Part II are previously published in


The main contributions in this part of the thesis are:

- The CRB for the proposed data model. The expression has a remarkably simple form (Theorem 1).

- The asymptotical analysis of the whitened WSF algorithm. Again, the obtained covariance expression is surprisingly simple (Theorem 2).

- The observation that, with the modified weighting matrix, whitened WSF is in fact asymptotically efficient.

- The proof of the optimality of the new weighting (Theorem 3).

- The covariance matching method applied to the estimation problem in question.

1.3 Notation

Throughout this thesis, matrices are denoted by boldface, uppercase letters, e.g. X, and vectors are denoted by boldface, lowercase letters, e.g. x. Scalars are denoted by plain symbols, e.g. a. The following notation will also be used
1.3. NOTATION

Denotes statistical expectation.
Cov[\mathbf{x}] The covariance matrix of \mathbf{x}.
\mathbb{R}^{m \times n} The set of \( m \times n \) matrices with real-valued elements.
\( [\mathbf{X}]_{i,j} \) Element \( i,j \) of the matrix \( \mathbf{X} \).
\( \mathbf{X}^T \) The transpose of \( \mathbf{X} \).
\( \mathbf{X}^* \) The conjugate transpose of \( \mathbf{X} \).
\text{vec}\{\mathbf{X}\} A vector obtained by stacking the columns of \( \mathbf{X} \).
\( \mathbf{I}_m \) The \( m \times m \) identity matrix. The subscript is sometimes dropped and the size is then implicitly defined.
\( \mathbf{0}_{mn} \) An \( m \times n \) matrix of all zeros.
\( \mathbf{1}_{mn} \) An \( m \times n \) matrix of all ones.
\( f'(\theta_0) \) The gradient (a column vector) of the function \( f(\theta) \) evaluated at \( \theta_0 \).
\( \mathbf{X}_i \) The elementwise derivative of the matrix \( \mathbf{X} \) with respect to (w.r.t.) parameter \( i \). The interpretation of “parameter \( i \)” is clear from the context if this notation is used.
\( \mathbf{X} \odot \mathbf{Y} \) The Kronecker product, see, e.g., [HJ91].
\( \mathbf{X} \circ \mathbf{Y} \) The direct (Schur) product, see, e.g., [HJ91].
\text{tr}\{\mathbf{X}\} The trace of the matrix \( \mathbf{X} \) (the sum of the diagonal elements).
\( |\mathbf{X}| \) The determinant of \( \mathbf{X} \).
\( \mathbf{X}^\dagger \) The Moore-Penrose pseudo inverse of \( \mathbf{X} \), see, e.g., [HJ91].
\( \Pi_{\mathbf{X}} \) \( \mathbf{X}\mathbf{X}^\dagger \), the orthogonal projection matrix onto the range space of \( \mathbf{X} \).
\( \Pi_{\mathbf{X}}^\perp \) \( \mathbf{I} - \Pi_{\mathbf{X}} \), the orthogonal projection matrix onto the nullspace of \( \mathbf{X}^* \).
\( \mathbf{X} \preceq \mathbf{Y} \) The difference \( \mathbf{X} - \mathbf{Y} \) is positive semi definite.

A few simple equalities will be used so frequently that they deserve an extra recollection.

\[
\text{vec}\{\mathbf{A}\mathbf{B}\mathbf{C}\} = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}\{\mathbf{B}\},
\]
\[
\text{tr}\{\mathbf{A}\mathbf{B}\} = \text{vec}^T\{\mathbf{A}^T\}\text{vec}\{\mathbf{B}\},
\]
\[
(\mathbf{X}_1 + \mathbf{Y}_1\mathbf{X}_2\mathbf{Y}_2)^{-1} = \mathbf{X}_1^{-1} - \mathbf{X}_1^{-1}\mathbf{Y}_1(\mathbf{X}_2^{-1} + \mathbf{Y}_2\mathbf{X}_1^{-1}\mathbf{Y}_1)^{-1}\mathbf{Y}_2\mathbf{X}_1^{-1}. \quad (1.15)
\]

The last equality is sometimes called the matrix inversion lemma (MIL). It holds when the indicated inverses exist.
1.4 Abbreviations

ADC  Analog-to-digital converter
AR   Auto Regressive
ARMA Auto Regressive Moving Average
CRB  Cramér-Rao (lower) Bound
DOA  Direction of Arrival (sometimes used in a wide sense meaning other signal parameters as well)
LS   Least Squares
MA   Moving Average
MIL  Matrix Inversion Lemma
ML   Maximum Likelihood
MSE  Mean Squared Error
MUSIC Multiple Signal Classification. An algorithm for DOA estimation.
p.d. positive definite
WLRA Weighted Low Rank Approximation
w.p.1 with probability one
w.r.t with respect to
Part I

Reduced rank linear regression and weighted low rank approximations
Chapter 2

Introduction to reduced rank linear regression

This part of the thesis deals with linear regressions of the form

\[ y(t) = C_0 x(t) + e(t), \quad t = 1, 2, \ldots, N, \quad (2.1) \]

where

\[ C_0 \in \mathbb{R}^{m \times n}, \]
\[ x(t) \in \mathbb{R}^{n \times 1}, \]
\[ y(t) \in \mathbb{R}^{m \times 1}, \]
\[ e(t) \in \mathbb{R}^{m \times 1}, \]
\[ \text{rank}(C_0) = r \leq \min(m, n). \]

The rank \( r \) of \( C_0 \) is known or estimated. The investigated problem is that of finding a rank \( r \) estimate \( \hat{C} \) of \( C_0 \) based on the observations \( \{x(t), y(t)\}_{t=1}^N \). The model allows the zero-mean, stationary noise \( e(t) \) to be correlated, both spatially and temporally. The following assumptions are made:

Assumption A1. The noise is assumed to be generated by a stable ARMA process driven by white noise. It has zero mean and autocorrelation function

\[ E \left[ e(t)e^T(t - \tau) \right] \equiv R_{ee}(\tau). \quad (2.2) \]

The statistics of the noise is unknown to the estimator. The input signal \( x(t) \) is assumed to be a quasi stationary deterministic signal [Lju99]. The autocorrelation of the input is

\[ R_{xx}(\tau) \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{t=\tau+1}^N x(t)x^T(t - \tau). \quad (2.3) \]
CHAPTER 2. INTRODUCTION TO REDUCED RANK LINEAR REGRESSION

It is assumed to be reminiscent of a stable ARMA process.

Assumption A2. The noise and the input signal are uncorrelated in the sense that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{t=r+1}^{N} x(t) e^T(t - \tau) = 0, \forall \tau$$

holds with probability one (w.p.1). Furthermore it is required that $R_{xx}(0)$ is positive definite.

Both these assumptions are fairly weak, allowing for a quite general class of noise and input models.

2.1 Background

Since $C_0$ in (2.1) has rank $r$, it can be factored as $C_0 = AB^T$, where $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{n \times r}$. Then

$$y(t) = A(B^Tx(t)) + e(t).$$

Note that the dimension of the vector $B^Tx(t)$ is smaller than that of $x(t)$. Thus, (2.5) states that the observations $y(t)$ can be equally well explained by the vector $B^Tx(t)$ that contains certain linear combinations of the elements of $x(t)$ as they can be by the full vector $x(t)$ (see Figure 1.2).

Let, for a moment, the input signal be a stochastic process. It is then intuitively quite clear that the reduced rank linear regression is related to canonical variate analysis and also principal components analysis, a connection formally established in [Ike75].

There are numerous examples of the use of the reduced rank regression model in the econometrics literature, for example in the modelling of macroeconomic time series [RV98]. Problems in econometrics were also the motivation for the pioneering work on the problem in [And51].

In signal processing, problems such as the one stated above appear in, e.g., filter design [MH01], state-space modelling [GR02, SJ00, JW94], array signal processing [VS09], space-time equalization [LT99] and channel estimation [NS03]. Other applications can be found in, e.g., [Sch91].

Most of the literature on the subject assume temporally white noise, a stronger assumption than A1. The maximum likelihood (ML) estimator for the problem under a white gaussian noise assumption is derived in [SV96]. The problem is also treated for a more general noise model in [GR02] and [WJ03]. The connection of the new method proposed in the following to these methods is elaborated on in Section 4.3. In [RV98] the case in which the noise stems from an AR(1) process is treated. That analysis relies heavily on the AR(1) noise model and does not extend to other noise models.
2.2 Outline of Part I

The present work is intended to address the estimation problem for a more general noise model. The proposed method performs significantly better than other methods the authors are aware of when the noise is temporally autocorrelated, and it can be shown to be asymptotically efficient when the noise is temporally white.

An important part of the new estimation method is the approximation of a full rank matrix with one of lower rank under a weighted norm. Except for special cases, this problem has no closed-form solution, [MMH03]. In order to address this issue, two novel non-iterative algorithms are derived that are asymptotically optimal for the particular problem encountered in the present work. The two methods rely on single Newton-steps in two different parameterizations.

The proposed estimation method (estimator) is presented in Chapter 3. In order to assess the performance of the estimator, the Cramér-Rao lower bound under a temporally and spatially colored gaussian noise assumption is derived in Section 4.2 and the asymptotic covariance of the estimator is derived in Section 4.1. A rank detection procedure is proposed in Section 3.4. An analytical comparison to previously published methods is presented in Section 4.3 and numerical comparisons are given in Chapter 5.
Chapter 3

An instrumental variable method for the reduced rank linear regression

In this chapter a novel algorithm for parameter estimation in the reduced rank linear regression data model will be described. The general algorithm is described in Section 3.1. A crucial part of the proposed estimation method is the problem of approximating one matrix with another matrix of lower rank, under a general quadratic norm. This problem is addressed in Section 3.3. The optimal norm used in the matrix approximation is weighted by the inverse of a certain covariance matrix, and the algorithm requires the consistent estimation of that matrix. This problem is addressed in Section 3.2.

The second last section in this chapter deals with the related problem of detecting the rank of $C_0$. The concluding section is a step-by-step guide through the suggested algorithm.

3.1 The proposed method

In this section, the proposed algorithm will be briefly outlined. The algorithm relies on instrumental variable principles (see e.g. [SS80a]). In order to simplify notation, consider rewriting (2.1) as

$$Y = C_0X + E,$$  \hspace{1cm} (3.1)

where

$$Y \equiv [y(1) \ldots y(N)],$$
$$X \equiv [x(1) \ldots x(N)],$$
$$E \equiv [e(1) \ldots e(N)].$$  \hspace{1cm} (3.2)
Proceed by defining the $(p + f + 1)n \times 1$ vector of instrumental variables

$$\mathbf{z}(t) \equiv \begin{bmatrix} x(t - p) \\ \vdots \\ x(t) \\ \vdots \\ x(t + f) \end{bmatrix}. \quad (3.3)$$

The lag lengths $p$ and $f$ are design variables. Analogous to (3.2), define the $(p + f + 1)n \times N$ matrix

$$\mathbf{Z} \equiv [\mathbf{z}(1) \ldots \mathbf{z}(N)] \quad (3.4)$$

(using some suitable convention for $x(t)$ if $t \leq 0$ or $t > N$). As an initial step in the algorithm, use $\mathbf{Z}$ together with the data matrices $\mathbf{X}$ and $\mathbf{Y}$ to form the quantities

$$\mathbf{K} \equiv \frac{1}{N} \mathbf{YZ}^T \in \mathbb{R}^{(m \times (p + f + 1)n)} \quad (3.5)$$

and

$$\mathbf{L} \equiv \frac{1}{N} \mathbf{XZ}^T \in \mathbb{R}^{(n \times (p + f + 1)n)}. \quad (3.6)$$

Then the expression

$$\mathbf{K} = \mathbf{C}_0 \mathbf{L} + \frac{1}{N} \mathbf{E} \mathbf{Z}^T \quad (3.7)$$

holds.

As a second step, an estimate $\hat{\mathbf{Q}}_\mathbf{K}$ of the normalized covariance matrix

$$\mathbf{Q}_\mathbf{K} \equiv \text{NCov}[\text{vec}(\mathbf{K})] \quad (3.8)$$

is calculated. This estimation problem is treated in Section 3.2.

Finally, an estimate $\hat{\mathbf{C}}$ of $\mathbf{C}_0$ is found by approximating the solution to the weighted low rank approximation problem (WLRA)

$$\hat{\mathbf{C}} \equiv \arg \min_{\text{rank}(\mathbf{C})=r} V(\mathbf{C}),$$

$$V(\mathbf{C}) \equiv \|\mathbf{K} - \mathbf{C}\mathbf{L}\|_{\mathbf{Q}_\mathbf{K}^{-1}} \equiv \text{vec}^T(\mathbf{K} - \mathbf{C}\mathbf{L})\mathbf{Q}_\mathbf{K}^{-1}\text{vec}(\mathbf{K} - \mathbf{C}\mathbf{L}). \quad (3.9)$$

It can be shown that, given the statistics $\mathbf{L}$ and $\mathbf{K}$, this final step is asymptotically optimal, as long as the estimated covariance matrix $\mathbf{Q}_\mathbf{K}$ is root-$N$ consistent (see [SS80a], complement C4.4). Problems such as (3.9) also appear in, e.g., the context of two-dimensional filter design [LPW97, Shp90], factor analysis of tabulated data.
3.2. Estimation of the weighting matrix

In the sequel, the definition

$$\hat{R}_{ab}(\tau) = \frac{1}{N} \sum_{t=\tau+1}^{N} a(t)b^T(t-\tau)$$

(3.10)

is used for the correlation matrix of two vector valued signals \(a(t)\) and \(b(t)\). Also define (assuming quasi stationarity)

$$R_{ab}(\tau) \equiv \lim_{N \to \infty} E\left[ \hat{R}_{ab}(\tau) \right].$$

(3.11)

A crucial step in the estimation procedure is to find an estimate of the matrix

$$Q_K = N^{-1}E \left[ \text{vec}(EZ^T) \text{vec}^T(EZ^T) \right]$$

(3.12)

that is defined in (3.8). It follows immediately from (3.2) and (3.4) that

$$\text{vec}(EZ^T) = \sum_{k=1}^{N} (z(k) \otimes e(k)).$$

(3.13)

Thus,

$$Q_K = \frac{1}{N} \sum_{k=1}^{N} \sum_{l=1}^{N} (z(k)z^T(l) \otimes E[e(k)e^T(l)])$$

$$= \frac{1}{N} \sum_{k=1}^{N} \sum_{l=1}^{N} (z(k)z^T(l) \otimes R_{ee}(k-l))$$

(3.14)

holds. An approximation of this is given in [S300]. It is...
\textbf{CHAPTER 3. AN INSTRUMENTAL VARIABLE METHOD FOR THE REDUCED RANK LINEAR REGRESSION}

\[ \hat{Q}_K = \frac{1}{N} \sum_{\tau=-N}^{N} (N - |\tau|)(\hat{R}_{zz}(\tau) \otimes \hat{R}_{\hat{w}\hat{w}}(\tau)). \] \hspace{1cm} (3.15)

The estimate (3.15) is always p.s.d. [SJ01]. Furthermore, due to assumption A1, the estimate is also mean-square consistent,

\[ \mathbb{E} \left[ \text{vec}^T \{ \hat{Q}_K - Q_K \} \text{vec} \{ \hat{Q}_K - Q_K \} \right] = O(N^{-1}), \] \hspace{1cm} (3.16)

for large $N$. See [SJ01] for a proof.

The quantity $\hat{R}_{\hat{w}}(\tau)$ for $\tau = -N, \ldots, N$ has to be estimated. This can be done using the sequence

\[ \hat{e}(t) \equiv y(t) - \hat{\Psi}x(t), \] \hspace{1cm} (3.17)

where $\hat{\Psi}$ is an initial consistent estimate of $C_0$, e.g., the unconstrained least squares estimate

\[ \hat{\Psi} = KL^T. \] \hspace{1cm} (3.18)

\subsection*{3.3 Weighted low rank approximation}

In the estimation procedure described above, the minimization problem (3.9) plays a central role. The general problem has no known closed-form solution. In this section, two new, non-iterative, algorithms will be derived in order to solve the problem. The new algorithms are asymptotically optimal for the estimation procedure proposed in this thesis. The algorithms are based on single approximative Newton steps (Gauss-Newton steps) in two different parameterizations.

If the rank constraint was ignored, the optimal solution to (3.9) would be

\[ \Psi = \arg \min_{C \in \mathbb{R}^{m \times m}} V(C), \] \hspace{1cm} (3.19)

which can be explicitly written as

\[ \text{vec} \{ \Psi \} = (L \otimes I_m)^{-1} (L^T \otimes I_m)^{-1} \hat{Q}_K^{-1} \text{vec} \{ K \}. \] \hspace{1cm} (3.20)

Since $V(C)$ is a quadratic function, its Taylor series expansion around $\Psi$ is

\[ V(C) = V(\Psi) + V'(\Psi) \text{vec} \{ C - \Psi \} + \frac{1}{2} \text{vec}^T \{ C - \Psi \} V''(\Psi) \text{vec} \{ C - \Psi \}. \] \hspace{1cm} (3.21)

The first term does not depend on $C$ and the second term is zero since $\Psi$ is a stationary point. Thus, the minimization problem

\[ \min_{C \in \mathbb{R}^{m \times m}} V(C), \] \hspace{1cm} (3.22)

\[ \tilde{V}(C) \equiv \| \Psi - C \|_Q, \] \hspace{1cm} (3.22)

\[ Q = V''(\Psi) = (L \otimes I_m) \hat{Q}_K^{-1} (L^T \otimes I_m) \]
3.3. WEIGHTED LOW RANK APPROXIMATION

has the same solution as the original problem (3.9). While this second formulation still lacks a closed form solution, it has the advantage of a reduced dimension.

For the reduced rank linear regression problem discussed in this work, note that, as a consequence of assumption A2, $\Psi$ (defined in (3.20)) is a consistent estimate of $C_0$

$$\lim_{N \to \infty} \Psi = C_0 \text{ (w.p.1).} \quad (3.23)$$

This means that the probability density function (pdf) of $\Psi$ is concentrated around a rank $r$ matrix for a large enough $N$. It follows that

$$\hat{C}^0 \equiv \arg \min_{C:\text{rank}(C) = r} \| \Psi - C \|_{1,m} \quad (3.24)$$

will be arbitrarily close to the solution to the WLRA (3.22) for large $N$. The optimal solution to the unweighted problem can therefore be used as an initial value. It can be concluded that higher order terms of a Taylor series expansion of the criterion function around this initial estimate can be ignored. One Newton step will asymptotically in $N$ give an estimator covariance that is equivalent to the covariance that would be achieved with a (hypothetical) exact solution to (3.22). In particular, there is asymptotically no benefit in performing multiple Newton steps. The solutions to the WLRA presented below are thus optimal in that particular sense. Two different parameterizations will be tested, yielding two variants of the Gauss-Newton step.

A first parameterization

A key issue is how to introduce the rank constraint into (3.22). A standard way is to use the factorization

$$C = AB^T, \ A \in \mathbb{R}^{m \times r}, \ B \in \mathbb{R}^{n \times r}. \quad (3.25)$$

Inspired by (3.25), define the parameter vector

$$\theta_{AB} \equiv \begin{pmatrix} \text{vec}(A) \\ \text{vec}(B^T) \end{pmatrix} \equiv \begin{pmatrix} a \\ b \end{pmatrix}. \quad (3.26)$$

This parameterization, together with (3.22), gives the new criterion function

$$\tilde{V}(\theta_{AB}) \equiv \text{vec}^T \{ \Psi - AB^T \} Q \text{vec} \{ \Psi - AB^T \}. \quad (3.27)$$

As discussed before, the minimizer of $\tilde{V}(\theta_{AB})$ is also the minimizer of (3.9).

Let $\theta_{AB}^0$ be a parameterization of an initial estimate $\hat{C}^0 = \hat{A}^0 \hat{B}^{0T}$. A correction vector $\delta_{AB}$ is sought in order to minimize $\tilde{V}(\theta_{AB}^0 + \delta_{AB})$. Performing a series
expansion of the gradient of the criterion function around $\hat{\theta}_A^0$ gives
\[
0 = \hat{V}'(\hat{\theta}_A^0 + \hat{\theta}_A^0) = \hat{V}'(\hat{\theta}_A^0) + \hat{V}''(\hat{\theta}_A^0)\hat{\theta}_A + \epsilon(\hat{\theta}_A^0, \hat{\theta}_A^0).
\]
Using a standard statistical linearization argument, the term $\epsilon(\hat{\theta}_A^0, \hat{\theta}_A^0)$ can be neglected. Thus the optimal update step must minimize (in a least squares sense)
\[
\hat{H}\hat{\theta}_A + \hat{V}'(\hat{\theta}_A^0),
\]
where $\hat{H}$ is the asymptotic Hessian of $\hat{V}(\hat{\theta}_A^0)$. Due to the ambiguous parameterization of $C$ (more about this in Section 4.2), $\hat{H}$ will be rank deficient. Thus there are infinitely many possible such Newton steps. It is natural to use the one with minimum 2-norm. It is uniquely given by
\[
\hat{\theta}_A = -\hat{H}^{-1}\hat{V}'(\hat{\theta}_A^0).
\]
The Hessian and gradient of (3.27) are derived in Appendix 3.A (setting $\Phi = I_{mn}$, $K = \Psi$ and $L = I_m$). Asymptotically in $N$ they simplify to
\[
\hat{V}''(\hat{\theta}_A^0) \simeq \hat{H} = 2\left( \hat{B}^0 I_m \otimes \hat{A}^0 I_n \otimes \hat{A}^0 \right) Q(\hat{B}^0 \otimes I_m I_n \otimes \hat{A}^0),
\]
\[
\hat{V}'(\hat{\theta}_A^0) = 2\left( \hat{B}^0 I_m \otimes \hat{A}^0 I_n \otimes \hat{A}^0 \right) \text{vec} (\hat{A}^0 \hat{B}^0 - \Psi),
\]
where $\simeq$ denotes equality in the dominating terms (based on the consistency of $\Psi$).

**Parameterizing the null-space of $C$**

The main motivation for the next approach is to reduce the number of parameters to be updated and thereby the computational complexity. The idea of using one Newton step with an asymptotically valid approximation of the Hessian is the same as in the previous section. The difference is that a new parameterization is used.

The parameterization used here is the “natural” parameterization suggested in [MMH03]. In that work an exact Newton method is derived that is designed to solve the general WLRA problem in an iterative fashion. The algorithm derived here is designed using the same loss function, but the statistical properties of the problem are used to derive an asymptotically valid approximation of the Hessian. Also, just as with the previous parameterization, the algorithm proposed here is optimal even if only one update step is performed.
3.3. **WEIGHTED LOW RANK APPROXIMATION**

It is clear that minimizing (3.27) w.r.t. \( A \) (for a fixed \( B \)) gives

\[
a_m \equiv \arg \min_{a=\text{vec}\{A\}} V(\theta_{AB}) = \arg \min_a \|Q^{1/2}\psi - Q^{1/2}(B \otimes I_m)a\|_2^2
\]

\[
= \left((B^T \otimes I_m)Q(B \otimes I_m)\right)^{-1}(B^T \otimes I_m)Q\psi,
\]

where \( \psi = \text{vec}\{\Psi\} \). Inserting (3.32) into (3.27) and defining \( B \equiv Q^{1/2}(B \otimes I_m) \) yields

\[
\hat{V}(B) = \frac{1}{2} \|I_{mn} - B(B^T B)^{-1}B^T\|Q^{1/2}\psi\|_2^2.
\]

Now, define \( M \in \mathbb{R}^{n \times (n-r)} \) to be a full column-rank matrix such that \( M^T B = 0 \) \((\Rightarrow \text{CM} = A B^T M = 0)\) and \( \hat{M} \equiv (M^T \otimes I_m)Q^{-1/2} \). Then \( \hat{M}^T B = 0 \) and \( \Pi_B^I = \Pi_M \). Thus,

\[
\hat{V}(B) = \frac{1}{2} \|I_{m\theta} Q^{1/2}\psi\|_2^2 = \|\hat{M}(\hat{M}^T \hat{M})^{-1}\hat{M}^T Q^{1/2}\psi\|_2^2
\]

\[
= \psi^T Q^{1/2} \hat{M}(\hat{M}^T \hat{M})^{-1} \hat{M}^T Q^{1/2} \psi = f(M).
\]

Given an \( M = \hat{M} \) minimizing \( f(M) \), it is easy to find the corresponding \( \hat{C} \) minimizing \( \hat{V}(C) \). To that end, reconsider (3.32) and note that

\[
\text{vec}\{C\} = \text{vec}\{A B^T\} = Q^{-1/2} \hat{B} a_m
\]

\[
= Q^{-1/2} \hat{B} (B^T B)^{-1} B^T Q^{1/2} \psi = Q^{-1/2} \Pi_B^I Q^{1/2} \psi.
\]

This means that minimizing \( \hat{V}(C) \) w.r.t. \( C \) is equivalent to minimizing \( f(M) \) w.r.t. \( M \). By using the same reasoning as with the previous parameterization, it can be concluded that the update step

\[
\hat{M} = -H_M^I f'(M)
\]

will be asymptotically optimal. The Hessian and gradient of the criterion function (3.34) are derived and simplified using the asymptotical properties of \( \Psi \) in Appendix 3.A. They are given by

\[
f''(M) \simeq H_M = 2(I_{n-r} \otimes \hat{C}^T)(\hat{M}^T \hat{M})^{-1}(I_{n-r} \otimes \hat{C}),
\]

\[
f'(\hat{M}) \simeq 2(I_{n-r} \otimes \hat{C}^T)(\hat{M}^T \hat{M})^{-1} \hat{M}^T Q^{1/2} \psi.
\]

With this second parameterization, the number of parameters to update is changed from \( nr + nr \) to \( n(n - r) \) compared to the first approach. Note that by transposing \( \Psi \) and \( C \) and by appropriately rearranging \( Q \) the WLRA can always be formulated so that \( n \leq m \). Depending on the application (and \( r \) in particular) either of the two parameterizations might be more natural to use. The analysis shows that the asymptotical covariance will be the same for the two suggested parameterizations. For small sample sizes however, second order terms might not be negligible, and so the performance might differ. This will be investigated numerically in Chapter 5.
CHAPTER 3. AN INSTRUMENTAL VARIABLE METHOD FOR THE REDUCED RANK LINEAR REGRESSION

Alternating projections

A well-known possibility for iteratively solving the WLRA problem is the alternating projections procedure. It starts by minimizing (3.27) w.r.t. \( \mathbf{B} \) using some fixed initial value of \( \mathbf{A} \). Next, \( \mathbf{B} \) is fixed to the minimizer in the previous step and a minimization is carried out w.r.t. \( \mathbf{A} \). The procedure can be iterated. Each minimization problem is quadratic and has a simple closed-form solution. In Chapter 5, a single iteration variant of this algorithm will be compared to the two new methods proposed in this work.

3.4 Detecting the rank of the regression

The purpose of this section is to derive a statistical test for detecting the rank of \( \mathbf{C}_0 \) using the known data \( \{x(t), y(t)\}_{t=1}^N \).

As a test quantity for the statistical test, the minimum value of the criterion function \( \hat{V}(\mathbf{C}) \), defined in (3.22), is a natural choice. Note that this value is independent of the parameterization of the matrix \( \mathbf{C} \). The statistical distribution of

\[
\xi_{\hat{r}} = \min_{\text{rank}(\mathbf{C}) = \hat{r}} N \hat{V}(\mathbf{C})
\]  

(3.38)

will now be derived under the hypothesis

\[
H_0 : \hat{r} = \text{rank}(\mathbf{C}_0) = r.
\]  

(3.39)

Examining a Taylor series expansion around the estimate \( \hat{\theta}_{AB} \) gives

\[
\hat{V}(\theta_0) \simeq \hat{V}(\hat{\theta}_{AB}) + \hat{V}'(\hat{\theta}_{AB})(\theta_0 - \hat{\theta}_{AB})
\]

\[
+ \frac{1}{2}(\hat{\theta}_{AB} - \theta_0)^T \hat{V}''(\hat{\theta}_{AB})(\hat{\theta}_{AB} - \theta_0)
\]

\[
\simeq \hat{V}(\hat{\theta}_{AB}) + \frac{1}{2}(\hat{\theta}_{AB} - \theta_0)^T \hat{\mathbf{H}}_0 \hat{\mathbf{H}}_0^T (\hat{\theta}_{AB} - \theta_0)
\]

\[
\simeq \hat{V}(\hat{\theta}_{AB}) + \frac{1}{2} \hat{V}'(\theta_0)^T \hat{\mathbf{H}}_0 \hat{\mathbf{H}}_0^T \hat{V}'(\theta_0),
\]  

(3.40)

where the equality \( \hat{V}'(\hat{\theta}_{AB}) = 0 \) was used and \( \theta_0 \) and \( \hat{\theta}_{AB} \) are defined previously. The matrix \( \hat{\mathbf{H}}_0 \) is the limit of \( \hat{\mathbf{H}} \) in (3.31) as \( N \to \infty \). The last equality follows since

\[
0 = \hat{V}'(\hat{\theta}_{AB}) \simeq \hat{V}'(\theta_0) + \hat{\mathbf{H}}_0 (\hat{\theta}_{AB} - \theta_0).
\]  

(3.41)

The desired result is now

\[
\hat{V}(\hat{\theta}_{AB}) \simeq \hat{V}(\theta_0) - \frac{1}{2} \hat{V}'(\theta_0)^T \hat{\mathbf{H}}_0 \hat{\mathbf{H}}_0^T \hat{V}'(\theta_0).
\]  

(3.42)
3.4. DETECTING THE RANK OF THE REGRESSION

Making use of (3.27), (3.31) and (3.42) one arrives at the expression

$$\xi = N \bar{V}(\hat{\theta}_{AB}) \approx N \text{vec}^T \{A_0^{-T} B_0^{-T} - \Psi\} Q \text{vec} \{A_0^{-T} B_0^{-T} - \Psi\}$$

$$- 2 N \text{vec}^T \{A_0^{-T} B_0^{-T} - \Psi\} Q (B_0 \otimes I_n \otimes A_0) \hat{H}_0^\dagger$$

$$\times \left( B_0^T \otimes I_n \otimes A_0^T \right) \text{vec} \{A_0^{-T} B_0^{-T} - \Psi\}$$

$$= N \text{vec}^T \{A_0^{-T} B_0^{-T} - \Psi\} Q^{1/2}$$

$$\times \left[ I_{mn} - 2 Q^{1/2} (B_0 \otimes I_n \otimes I_n) \hat{H}_0^\dagger \left( B_0^T \otimes I_n \otimes A_0^T \right) Q^{1/2} \right]$$

$$\times Q^{1/2} \text{vec} \{A_0^{-T} B_0^{-T} - \Psi\}. \quad (3.43)$$

By defining

$$D \equiv (B_0 \otimes I_n \otimes I_n) \otimes A_0)^T, \quad (3.44)$$

the matrix within square brackets in (3.43) can be written

$$G \equiv (I_{mn} - Q^{1/2} D^T (DQD^T)^T DQ^{1/2}). \quad (3.45)$$

In order to follow the procedure used in [SV96], note that this matrix is idempotent. Since all eigenvalues of an idempotent matrix are either zero or one, the relation

$$\text{rank}(G) = \text{tr}(I_{mn}) - \text{tr}\left(Q^{1/2} D^T (DQD^T)^T DQ^{1/2}\right)$$

$$= mn - \text{tr}\left((DQD^T)^T (DQD^T)\right) = mn - \text{tr}(\hat{H}_0^\dagger \hat{H}_0^\dagger)$$

$$= mn - \text{rank}(\hat{H}_0^\dagger \hat{H}_0^\dagger) = mn - \text{rank}(\hat{H}_0) \quad (3.46)$$

holds. It can be verified that [SV96]

$$\text{rank}(\hat{H}_0) = \text{rank}\{(B_0 \otimes I_n \otimes I_n) \otimes A_0\} = mn + rn - r^2. \quad (3.47)$$

Thus,

$$\text{rank}(G) = (m - r)(n - r). \quad (3.48)$$

The central limit theorem (see, e.g., [SS89a] Lemma B.3) and assumption A1 give that

$$\sqrt{N} Q^{1/2} \text{vec} \{A_0^{-T} B_0^{-T} - \Psi\} \sim N(0, I_{mn}) \quad (3.49)$$

for large $N$. It follows from Lemma B.13 in [SS89a] that

$$\xi = N \bar{V}(\hat{\theta}_{AB}) \sim \chi^2((m - r)(n - r)) \quad (3.50)$$

holds asymptotically in $N$ under $H_0$.

Using the above result it is a simple task to devise an algorithm for estimating the rank of $C_0$ in (2.1). An outline of an algorithm performing both rank detection and parameter estimation is given in the next section.
3.5 The proposed algorithm step-by-step

This section is intended as an overview of the algorithm, and also as a step-by-step guide to implementation.

1. The input is $x(t)$ and $y(t)$ for $t = 1, \ldots, N$.
2. Select $p$ and $f$.
3. Select $\alpha$, the probability of falsely rejecting $H_0 : r = \bar{r}$.
4. Form $z(t)$, $t = 1, \ldots, N$ according to (3.3). Assume, e.g., that $x(t)$ for $t < 1$ and $t > N$ is zero.

5. Calculate

$$K = \sum_{t=1}^{N} y(t)z^T(t),$$
$$L = \sum_{t=1}^{N} x(t)z^T(t).$$

6. Calculate

$$\hat{\Psi} = KL^+. $$

7. Calculate

$$\hat{e}(t) = y(t) - \hat{\Psi}x(t).$$

8. Calculate

$$\hat{R}_{zz}(\tau) = \sum_{t=1}^{N} z(t)z^T(t - \tau),$$
$$\hat{R}_{ee}(\tau) = \sum_{t=1}^{N} \hat{e}(t)\hat{e}^T(t - \tau).$$

Symmetry can be exploited in the computation.

9. Form the matrix

$$\hat{Q}_K = \frac{1}{N} \sum_{\tau = -N}^{N} (N - |\tau|)(\hat{R}_{zz}(\tau) \otimes \hat{R}_{ee}(\tau)).$$

10. Compute $\Psi$ according to (3.20).
11. Compute $Q$ according to (3.22).
12. Set $\bar{r} = 1$.
13. Compute the SVD
\[ \Psi = U\Sigma V^T. \]
14. Use the SVD to form an initial, rank $\bar{r}$, value for the WLRA problem. Use either of the $A^0$, $B^0$ or the $M^0$ parameterizations.
15. Perform one update step according to either (3.30) or (3.36), depending on the choice of parameterization. Compute the estimate and denote it by $\hat{C}$.
16. Find $\xi_r = N\hat{V}(\hat{C})$. The function $\hat{V}(C)$ is given by (3.22).
17. If $\Pr(\xi < \xi_r | \xi \sim \chi^2((m-\bar{r})(n-\bar{r}))) > (1 - \alpha)$, then set $\bar{r} = \bar{r} + 1$ and proceed to 14. Otherwise go to 18.
18. $\hat{r} = \bar{r}$, $\hat{C} = C$.

Typically the choice of $\alpha$ should depend on $N$ and the signal to noise ratio.

For large $N$, step 8) will dominate the computations. For each value of $\tau$ $Nn(p + f + 1) + Nm$ multiplications and $Nn^2(p + f + 1)^2 + Nm^2$ additions are required. The total complexity of step 8) is then
\[
\frac{N^2}{2}(n(p + f + 1) + m)
\]
multiplications and
\[
\frac{N^2}{2}(n^2(p + f + 1)^2 + m^2)
\]
additions.

### 3.A A useful gradient and Hessian

First, the gradient and Hessian of the function
\[
V(\theta) = (\text{vec}\{K\} - \Phi \text{vec}\{AB^T L\})^T Q (\text{vec}\{K\} - \Phi \text{vec}\{AB^T L\})
\]
will be calculated w.r.t. the parameter vector
\[
\theta = \begin{pmatrix} \text{vec}\{A\} \\ \text{vec}\{B^T\} \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}.
\]
CHAPTER 3. AN INSTRUMENTAL VARIABLE METHOD FOR THE REDUCED RANK LINEAR REGRESSION

The results will be useful in several parts of the thesis, e.g., Section 4.2, Section 4.1 and Section 3.3. The matrices \( K, L \) and \( \Phi \) will have different meanings depending on the context. Using \( k \equiv \text{vec}(K) \), the function can be written

\[
V(\theta) = k^T Q k - a^T (B^T L \otimes I_m) \Phi^T Q k - k^T Q \Phi (L^T B \otimes I_m) a
+ a^T (B^T L \otimes I_m) \Phi^T Q \Phi (L^T B \otimes I_m) a
\]

or, similarly,

\[
V(\theta) = \begin{align*}
k^T Q k &- b^T (L \otimes A^T) \Phi^T Q k - k^T Q \Phi (L^T \otimes A) b \\
&+ b^T (L \otimes A^T) \Phi^T Q \Phi (L^T \otimes A) b.
\end{align*}
\]

Using (3.53) and (3.54), it is easy to verify that

\[
\frac{\partial V(\theta)}{\partial a} = 2(B^T L \otimes I_m) \Phi^T Q \Phi (L^T B \otimes I_m) a - 2(B^T L \otimes I_m) \Phi^T Q k - (B^T L \otimes I_m) \Phi^T Q \Phi (L^T B \otimes I_m) a \tag{3.55}
\]

and

\[
\frac{\partial V(\theta)}{\partial b} = 2(L \otimes A^T) \Phi^T Q \Phi (L^T \otimes A) b - 2(L \otimes A^T) \Phi^T Q k \tag{3.56}
\]

Thus, the gradient is

\[
\frac{\partial V(\theta)}{\partial \theta} = 2 \begin{pmatrix} B^T L \otimes I_m \\ L \otimes A^T \end{pmatrix} \Phi^T Q \left( \Phi \text{vec} \{AB^T L\} - k \right) \tag{3.57}
\]

In order to find the Hessian, note that

\[
\frac{\partial^2 V(\theta)}{\partial a \partial a^T} = 2(B^T L \otimes I_m) \Phi^T Q \Phi (L^T B \otimes I_m) \tag{3.58}
\]

and

\[
\frac{\partial^2 V(\theta)}{\partial b \partial b^T} = 2(L \otimes A^T) \Phi^T Q \Phi (L^T \otimes A) \tag{3.59}
\]

Furthermore, applying the product rule gives

\[
\frac{\partial^2 V(\theta)}{\partial b \partial a^T} = 2 \frac{\partial}{\partial a} (L \otimes A^T) \Phi^T Q \left( \Phi \text{vec} \{AB^T L\} - k \right)
+ 2(L \otimes A^T) \Phi^T Q \Phi (L^T B \otimes I_m) \tag{3.60}
\]

which concludes the derivation of the Hessian.

Next, the gradient and Hessian of (3.34) will be derived. In order to differentiate \( f(M) \), first investigate the differential of \( M(M^T M)^{-1} M^T \). It is

\[
dM(M^T M)^{-1} M^T + M(M^T M)^{-1} dM^T
- M(M^T M)^{-1} (dM^T M + M^T dM)(M^T M)^{-1} M^T
\]

\[
\tag{3.61}
\]

where
3.A. A USEFUL GRADIENT AND HESSIAN

(see [MN88] for the necessary relations). Then

\[
df(M; dM) = \psi^T Q^{1/2} ((I_{mn} - ML^{-1}M^T) dML^{-1}M^T
+ ML^{-1}dM^T (I_{mn} - ML^{-1}M^T)) Q^{1/2} \psi,
\]

(3.62)

where \( J \equiv (M^T M) \). Since the differential is a scalar function it is clear that (3.62) can be written

\[
df(M; dM) = 2\psi^T Q^{1/2} dMJ^{-1}M^T Q^{1/2} \psi - 2\psi^T Q^{1/2} MJ^{-1}M^T dMJ^{-1}M^T Q^{1/2} \psi.
\]

(3.63)

In order to proceed, define \( \text{vec}(D) \equiv J^{-1}M^T Q^{1/2} \psi \), \( D \in \mathbb{R}^{m \times (n-r)} \) and \( \text{vec}(F) \equiv Q^{-1}\text{vec}(DM^T) \), \( F \in \mathbb{R}^{m \times n} \). Also note that \( dM = Q^{-1/2}(dM \otimes I_m) \). This gives

\[
df(M; dM) = 2\psi^T (dM \otimes I_m) \text{vec}(D) - 2\psi^T (M^T \otimes I_m) Q^{-1} (dM \otimes I_m) \text{vec}(D)
= 2 (\text{vec}(\Psi^T D) - \text{vec}(\Psi^T F)) \text{vec}(dM).
\]

(3.64)

The identification theorem gives the gradient

\[
\frac{\partial f}{\partial M} = 2\text{vec}(\{\Psi - F\}^T D) \equiv \nabla f.
\]

(3.65)

Proceeding to find the Hessian it is convenient to first calculate the differentials of \( D \) and \( F \). They are

\[
d\text{vec}(D) = J^{-1}dM^T Q^{1/2} \psi - J^{-1}(dM^T M + M^T dM) \text{vec}(D)
= J^{-1}\{\text{vec}(\Psi dM) - \text{vec}(FdM) - (M^T \otimes I_m)Q^{-1} \text{vec}(DdM^T)\}
\]

and

\[
d\text{vec}(F) = Q^{-1}\text{vec}(DdM^T) + Q^{-1}\text{vec}(dDM^T).
\]

(3.66)

The differential of the gradient (3.65) is then

\[
\frac{1}{2}d\nabla f = (I_{n-r} \otimes (\Psi^T - F^T)) \text{vec}(dD) - (D^T \otimes I_n) \text{vec}(dF^T)
= (I_{n-r} \otimes (\Psi^T - F^T))
- (D^T \otimes I_n)K_{mn,n} Q^{-1} (M \otimes I_m) \text{vec}(dD)
- (D^T \otimes I_n)K_{mn,n} Q^{-1} (I_n \otimes D)K_{n,(n-r)}dM,
\]

(3.68)

where \( K_{x,y} \) is the matrix satisfying \( K_{x,y} \text{vec}(X) = \text{vec}(X^T) \) for any \( X \in \mathbb{R}^{x \times y} \). From (3.68) the Hessian can be found to be
\[ \frac{1}{2} f''(M) = (I_{n-r} \otimes (\Psi^T - F^T))J^{-1}(I_{n-r} \otimes (\Psi - F)) + (D^T \otimes I_n)K_{m,n}Q^{-1}(M \otimes I_m)J^{-1} \times (M^T \otimes I_m)Q^{-1}(I_n \otimes D)K_{n,(n-r)} - (I_{n-r} \otimes (\Psi^T - F^T))J^{-1} \times (M^T \otimes I_m)Q^{-1}(I_n \otimes D)K_{n,(n-r)} - (D^T \otimes I_n)K_{m,n}Q^{-1} \times (M \otimes I_m)J^{-1}(I_{n-r} \otimes (\Psi - F)) - (D^T \otimes I_n)K_{m,n}Q^{-1}(I_n \otimes D)K_{n,(n-r)}. \] (3.69)

Finally, using the consistency of \( \Psi \) and (3.35) it can be noted that
\[ \text{vec}\{\Psi - F\} = \psi - \text{vec}\{F\} \simeq \text{vec}\{C\}. \] (3.70)

In addition, since \( \tilde{C} \hat{M} \approx 0 \), it can be concluded that \( D \approx 0 \). Thus, the asymptotic Hessian and gradient are given by (3.37).
Chapter 4

Asymptotical analysis and performance bounds for the proposed IV method

While the previous chapter introduced the estimation algorithm, this chapter will focus on its performance. In Section 4.1, the asymptotical (for large $N$) covariance of the estimate is derived. Then a bound for the covariance of any unbiased estimator, given the data model, is derived in Section 4.2. Finally, a comparison to related estimation algorithms is presented in Section 4.3.

4.1 Asymptotical covariance of the estimate

In this section, the asymptotical covariance of $\hat{C}$ will be derived. The derivations below follow the general ideas of [SV96] on how to handle ambiguous parameterizations, but the underlying criterion function is different. Again the parameterization (3.25) is useful. First note that the consistency of the estimator $\hat{C}$ follows immediately from the fact that $\Psi$ in (3.20) is a consistent estimate of $C_0$ and that $Q$ is positive definite. Fix a factorization $\hat{C} = \hat{A}\hat{B}^T$ by defining

$$\hat{A} = U(\Sigma_{r \times m}^{1/2})^T,$$
$$\hat{B} = V(\Sigma_{r \times n}^{1/2T})^T,$$  \hspace{1cm} (4.1)

where

$$\Sigma_{r \times n}^{1/2} = [\Sigma_{r \times r}^{1/2} \ 0_{r \times (n-r)}]$$  \hspace{1cm} (4.2)

and $\Sigma_r$ is the $r \times r$ upper left block of $\Sigma$ in the singular value decomposition (SVD)

$$\hat{C} = U\Sigma V^T.$$  \hspace{1cm} (4.3)
It follows that \( \hat{A} \) and \( \hat{B} \) will converge to some true values \( A_0 \) and \( B_0 \) that are defined using a similar factorization of \( C_0 \). It should be noted that this requires some suitable convention for making the SVD unique.

Due to the convergence established above, the estimation error can, asymptotically in \( N \), be written

\[
C_0 - \hat{C} = A_0 B_0^T - \hat{A} \hat{B}^T = (A_0 - \hat{A}) \hat{B}^T + A_0 (B_0 - \hat{B})^T \\
\simeq (A_0 - \hat{A}) B_0^T + A_0 (B_0 - \hat{B})^T.
\] (4.4)

Next, let \( \theta_0 \) be a parameterization of \( C_0 = A_0 B_0^T \), corresponding to (3.26). The vector \( \hat{\theta}_{AB} \) is defined similarly. Then the vectorized estimation error can be written

\[
\text{vec}\{C_0 - \hat{C}\} \simeq [B_0 \otimes I_m \ I_n \otimes A_0] (\theta_0 - \hat{\theta}_{AB}).
\] (4.5)

Note that \( \hat{\theta}_{AB} \) is the minimizer of the criterion function

\[
V(\theta_{AB}) = \text{vec}^T \{K - AB^T L\} \hat{Q}_K^{-1} \text{vec}\{K - AB^T L\}.
\] (4.6)

A Taylor series expansion of the gradient around \( \theta_0 \) gives

\[
0 = V'(\hat{\theta}_{AB}) \simeq V'(\theta_0) - V''(\theta_0)(\theta_0 - \hat{\theta}_{AB}) \\
\simeq V'(\theta_0) - H(\theta_0 - \hat{\theta}_{AB}).
\] (4.7)

The limiting Hessian, \( H = \lim_{N \to \infty} V''(\theta_0) \), and the gradient follow from the derivations in Appendix 3.A with \( \Phi = I \) and the consistency

\[
\lim_{N \to \infty} K = C_0 L_\infty,
\]

\[
L_\infty = [R_{xx}(-p) \ldots R_{xx}(f)].
\] (4.8)

They are

\[
H = 2 \left( \begin{array}{c} B_0^T L_\infty \otimes I_m \\ L_\infty \otimes A_0^T \end{array} \right) \hat{Q}_K^{-1} \left( L_\infty^T B_0 \otimes I_m L_\infty^T \otimes A_0 \right),
\]

\[
V'(\theta_0) = 2 \left( \begin{array}{c} B_0^T L \otimes I_m \\ L \otimes A_0^T \end{array} \right) \hat{Q}_K^{-1} \text{vec}\{A_0 B_0^T L - K\}.
\] (4.9)

It follows from (4.7) that

\[
(\theta_0 - \hat{\theta}_{AB}) - (I_{nr + nr} - H^H)(\theta_0 - \hat{\theta}_{AB}) \simeq H^H V'(\theta_0).
\] (4.10)

Consequently it holds that

\[
((B_0 \otimes I_m) (I_n \otimes A_0)) (\theta_0 - \hat{\theta}_{AB}) \\
- ((B_0 \otimes I_m) (I_n \otimes A_0)) (I_{nr + nr} - H^H)(\theta_0 - \hat{\theta}_{AB}) \\
\simeq ((B_0 \otimes I_m) (I_n \otimes A_0)) H^H V'(\theta_0).
\] (4.11)
4.2. THE CRAMÉR-RAO BOUND

By noting that $(I_{m r + r} - HH')$ is the orthogonal projector onto the null-space of $H$ and that the range-space of $(B_0 \otimes I_m, I_n \otimes A_0)$ coincides with that of $H$ it can be concluded that the second term on the left hand side of (4.11) is zero. By making use of (4.5), it follows that

$$\text{vec}(C_0 - \hat{C}) \simeq ((B_0 \otimes I_m) (I_n \otimes A_0)) H'V'(\theta_0). \quad (4.12)$$

Before the final asymptotic covariance can be calculated, it remains to evaluate the covariance of $V'(\theta_0)$. It is

$$\lim_{N \to \infty} \text{NE} \left[ V'(\theta_0)V'^T(\theta_0) \right] = \lim_{N \to \infty} 4N \left( B_0^T L \otimes I_m \right) \hat{Q}_K^{-1} \text{E} \left[ N^{-2} \text{vec}(EZ^T) \text{vec}^T(EZ^T) \right] \times \hat{Q}_K^{-1} (L^T B_0 \otimes I_m L^T \otimes A_0) \approx 4 \left( B_0^T L \otimes I_m \right) Q_K^{-1} (L^T B_0 \otimes I_m L^T \otimes A_0) = 2H, \quad (4.13)$$

where use was made of (3.16) in the second last equality. Using (4.12) and (4.13) the asymptotic covariance of the estimation error can be calculated as

$$\lim_{N \to \infty} \text{NE} \left[ \text{vec}(C_0 - \hat{C}) \text{vec}^T(C_0 - \hat{C}) \right] = \left( (B_0 \otimes I_m) (I_n \otimes A_0) \right) H^T H \left( B_0^T \otimes I_m \right) \left( I_n \otimes A_0^T \right) \left( B_0^T \otimes I_m \right) \left( I_n \otimes A_0^T \right). \quad (4.14)$$

The obtained expression can easily be compared to that of [SV96]. The expressions coincide if the noise is temporally white. For some numerical examples, see Chapter 5.

4.2 The Cramér-Rao bound

In this section, the CRB for the reduced rank linear regression problem will be derived. The CRB is a lower bound on the achievable estimation covariance for unbiased estimators for a particular probability density function of the observations. The general theory of the CRB is given a detailed treatment in for example [Kay93].

For the reduced rank linear regression data model, the CRB has previously been derived in [SV96]. The bound that will be presented below is a generalization that allows for a larger class of noise models, in particular temporally colored noise. The derivations follow a different, and very direct, path.
Parameterizing the problem

In order to derive the Cramér-Rao lower bound (CRB) for $C_0$, assume that the noise $e(t)$ is Gaussian. With the data model of (3.1) in mind, write

$$e \equiv \text{vec}\{E\} \sim N(0, Q_e).$$ \hfill (4.15)

The noise covariance $Q_e$ is assumed to be unknown to the estimator. Note that this is a very general noise covariance structure, allowing for virtually any temporal and spatial noise correlation, including the ARMA model assumed when deriving the algorithm proposed in this part of the thesis. It is however necessary to assume that $Q_e$ is invertible. Furthermore, assume that the noise model is parameterized independently of $C$. Let $\theta_{AB}$ parameterize the $C$ matrix and let $\theta_Q \in \mathbb{R}^{n_q}$ parameterize the noise covariance matrix $Q_e$. The full $(mr + nr + n_q)$ parameter vector is $\theta = (\theta_{AB}^T, \theta_Q^T)^T$. The rank constraint is introduced via the factorization $C = AB^T$. Define the function

$$f(\theta) = C = AB^T.$$ \hfill (4.16)

An important note has to be made concerning this parameterization: It is not possible to estimate the elements of $A$ and $B$ using only the model (2.1), even if $C_0$ itself can be perfectly well estimated from the data. This is because the mapping between $\theta_{AB}$ and $C$ is many-to-one. To see that, consider the equality $C = (AT)(BT^{-1})^T$ which holds for any nonsingular $r \times r$ matrix $T$. This ambiguity will typically lead to a singular information matrix when deriving the CRB. However, since the parameters of interest in the estimation are functions of $A$ and $B$, not the matrices themselves, this problem can be handled.

Before proceeding, note that the input signal is modelled as a known deterministic signal. This means that the CRB derived below in the general case will depend explicitly on the input signal.

The Fischer information matrix

By introducing

$$\Phi \equiv \begin{pmatrix} x^T(1) \otimes I_m \\ \vdots \\ x^T(N) \otimes I_m \end{pmatrix}$$ \hfill (4.17)

and $y \equiv \text{vec}\{Y\}$ the negative log-likelihood function for the data can be written

$$L(\theta) = \text{const.} + \frac{1}{2} \log |Q_e| + \frac{1}{2} (y - \Phi \text{vec}\{C\})^T Q_e^{-1} (y - \Phi \text{vec}\{C\}).$$ \hfill (4.18)

The Fischer information matrix is defined as [Kay93]

$$[I(\theta)]_{i,j} \equiv E \left[ \frac{\partial^2 L(\theta)}{\partial \theta_i \partial \theta_j} \right].$$ \hfill (4.19)
With obvious notation, write
\[
I(\theta_0) = \begin{pmatrix}
I_{AA} & I_{AB} & I_{AQ} \\
I_{BA} & I_{BB} & I_{BQ} \\
I_{QA} & I_{QB} & I_{QQ}
\end{pmatrix}.
\] (4.20)

The submatrices \(I_{AA}, I_{BB}, I_{BA}\) and \(I_{AB}\) are readily found using \(L = I_n, \ K = Y\) and \(Q = Q_e^{-1}\) together with the equality \(E[\kappa] = \Phi \text{vec}(AB^T)\) at the true parameter value in the derivations in Appendix 3A. The blocks \(I_{BQ}\) and \(I_{AQ}\) can be determined by differentiating (3.55) and (3.56) to get
\[
\frac{\partial^2 L(\theta)}{\partial \text{vec}\{A\} \partial \theta_Q} = \frac{\partial}{\partial \theta_Q} \left( (B^T \otimes I_m) \Phi T Q_e^{-1} \right) (\Phi c - y)
\] (4.21)

and
\[
\frac{\partial^2 L(\theta)}{\partial \text{vec}\{B^T\} \partial \theta_Q} = \frac{\partial}{\partial \theta_Q} \left( (I_n \otimes A^T) \Phi T Q_e^{-1} \right) (\Phi c - y).
\] (4.22)

This is possible since the second factor is a constant (w.r.t. \(\theta_Q\)) in both equations. When taking expectation, the first factor can be treated as a constant independent of \(y\). It can be concluded that
\[
I_{AQ} = I_{QA} = 0_{nr \times n_q},
\]
\[
I_{BQ} = I_{QB} = 0_{nr \times n_q}.
\] (4.23)

Hence, the sought information matrix can be written
\[
I(\theta_0) = \begin{pmatrix} U & 0_{(nr+mr) \times n_q} \\
0_{n_q \times (nr+mr)} & I_{QQ}
\end{pmatrix},
\] (4.24)
where \(U\) is given by
\[
U = \begin{bmatrix} B_0^T \otimes I_m \\
I_n \otimes A_0^T
\end{bmatrix} \Phi^T Q_e^{-1} \Phi [B_0 \otimes I_m \ I_n \otimes A_0].
\] (4.25)

Since \(Q_e^{-1}\) is positive definite, it follows that
\[
\text{rank}(U) = \text{rank}([B_0 \otimes I_m \ I_n \otimes A_0]) = rm + rn - r^2,
\] (4.26)
(cf. (3.47)). The singularity of \(U\) is a consequence of the over-parameterization mentioned above.

Finding the CRB

The parameters of interest are the elements of \(C\). They are mapped from \(\theta\) by the function \(c = \text{vec}(f(\theta))\) where \(f(\theta)\) is defined in (4.16). Standard theory gives that
CHAPTER 4. ASYMPTOTICAL ANALYSIS AND PERFORMANCE
BOUNDS FOR THE PROPOSED IV METHOD

the inequality

\[
\text{Cov} [\hat{\varepsilon}] \succeq \Delta \Gamma(\theta_0) \Delta^T,
\]

\[
\Delta = \frac{\partial f}{\partial \theta} |_{\theta = \theta_0}
\]  

(4.27)

holds if \( \hat{\varepsilon} \) is an unbiased estimator of \( \text{vec}\{C_0\} \). In the present problem the information matrix is singular, a case that is treated in [SM01]. While mathematically correct, (4.27) is then only useful under certain conditions on the model. These conditions will be verified below.

The block diagonal structure of \( \Gamma(\theta_0) \) can be exploited when evaluating (4.27). Clearly

\[
\Gamma(\theta_0) = \begin{pmatrix}
U^T & 0_{(nr+mr) \times n_q} \\
0_{n_q \times (nr+mr)} & I_{n \otimes (nr+mr)}
\end{pmatrix}
\]  

(4.28)

and

\[
\Delta = [B_0 \otimes I_m \ I_n \otimes A_0 \ 0_{mn \times n_q}].
\]  

(4.29)

Insertion into (4.27) gives the CRB for the parameters of \( C_0 \)

\[
\text{Cov} [\hat{\varepsilon}] \succeq [B_0 \otimes I_m \ I_n \otimes A_0] U^T \left[ B_0^T \otimes I_m \ I_n \otimes A_0^T \right] [B_0 \otimes I_m \ I_n \otimes A_0].
\]

(4.30)

If the noise is temporally white, the matrix \( Q_e \) will be block diagonal with equal diagonal matrices and the matrix \( \Phi^T Q_e^{-1} \Phi \) can for large \( N \) be written \( N (R_{xx}(0) \otimes R_{ee}^{-1}(0)) \). The asymptotic (in \( N \) ) CRB then coincides with the asymptotical covariance (4.14). If the noise is not temporally white, then the asymptotic CRB cannot be easily expressed directly in terms of \( R_{xx}(\tau) \).

As stated and proved in [SM01] the condition

\[
\Delta = \Delta \Gamma^T
\]  

(4.31)

must hold for (4.30) to be a meaningful bound. This means that the rows of \( \Delta \) must be in the range space of \( \Gamma(\theta_0) \). Considering the structure of (4.30), it is easy to see that this holds under weak assumptions on the input signal \( x(t) \). Furthermore, for the Cramér-Rao bound to be valid, the regularity condition

\[
E \left[ \frac{\partial L(\theta)}{\partial \theta_k} \right] = 0
\]  

(4.32)

must hold. For \( k \leq (mr + nr) \) this can easily be verified by (with proper definitions of the matrices) taking expectation of (3.55) and (3.56) in Appendix 3.A. For the remaining parameters, the condition can be checked in a straightforward manner.
4.3 Relation to previous methods

In [WJ03], an estimation method for the reduced rank linear regression was derived that relies on a two step procedure similar to the one treated in this part of the thesis. The method proposed in this work specializes to the one of [WJ03] if \( p \) and \( f \) are set to zero. In that case \( L = \hat{R}_{xx} \) and \( K = \hat{R}_{xy} \) and the criterion function in (3.9) becomes

\[
V(C) = ||(\hat{R}_{yx} \hat{R}_{xx}^{-1} - C)\hat{R}_{xx}||_2\hat{Q}_n^{-1}
= ||\hat{R}_{yx} \hat{R}_{xx}^{-1} - C||_2(\hat{R}_{xx} \otimes I_m)\hat{Q}_n^{-1}(\hat{R}_{xx} \otimes I_m)
\equiv ||\hat{\Psi} - C||_2\hat{Q}_n^{-1},
\]

(4.33)

where \( \hat{\Psi} \) is the least squares fit to the data (cf. (3.18)).

In [SV96], the maximum likelihood (ML) solution to the regression problem under the assumption that the noise \( e(t) \) is temporally white and gaussian is derived. Compared to A1 (2.2) this is a much stronger assumption. The ML solution of [SV96] can conveniently be stated [HNS01]

\[
\hat{C}_{ML} = \hat{R}_{yx} \hat{R}_{xx}^{-1/2} \hat{V}_r \hat{V}_r^T \hat{R}_{xx}^{-1/2},
\]

(4.34)

where the \( n \times r \) matrix \( \hat{V}_r \) is obtained from the SVD

\[
\hat{U} \hat{\Sigma} \hat{V}_r^T = \hat{R}_{yy}^{-1/2} \hat{R}_{yx} \hat{R}_{xx}^{-1/2}
\]

(4.35)

by taking the columns of \( \hat{V} \) corresponding to the \( r \) largest singular values. The ML estimate \( \hat{C}_{ML} \) is identical to the estimate \( \hat{C} \) that results from the method derived here with \( p = f = 0 \) if the noise \( e(t) \) is known to be white and the exact minimum of (4.33) is found (for that special case there is an analytical solution). See Appendix 4.A for a proof.

The regression problem (2.1) is also treated in [GR02]. The algorithm in that work uses a two step procedure similar to the one presented here if \( p = f = 0 \) but with a different weighting. The estimator proposed in [GR02] is given by

\[
\hat{C}_W = \hat{\Psi} W_R^{-1} \hat{V}_r \hat{V}_r^T W_R^{-1},
\]

(4.36)

where \( \hat{V}_r \) is obtained from the SVD

\[
\hat{U} \hat{\Sigma} \hat{V}_r^T = W_L \hat{\Psi} W_R
\]

(4.37)

analogously to \( \hat{V}_r \). Three different alternatives for the weighting matrices \( W_L \) and \( W_R \) are considered except for the trivial case \( W_L = I_m, W_R = I_n \) which corresponds to \( Q_\psi = I_{mn} \) in (4.33).

The first alternative, referred to as the CVA method, is exactly the ML method of [SV96]. The second choice, referred to as the CVA2 method is based on the
CHAPTER 4. ASYMPTOTICAL ANALYSIS AND PERFORMANCE BOUNDS FOR THE PROPOSED IV METHOD

matrices

\[ W_L = \left( \sum_{\tau = -N}^{N} (\hat{R}_{xx}^{-1}(\tau)\hat{R}_{xx}^{-1}T(\tau)) \right)^{-1/2} \]

and

\[ W_R = \left( \sum_{\tau = -N}^{N} (\hat{R}_{ee}(\tau)\hat{R}_{xx}(\tau)) \right)^{-1/2}. \tag{4.38} \]

In the third method, CVA3, the weights are \( W_R = L_1^{-1/2} \) and \( W_L = L_2^{-1/2} \) where \( L_1 \otimes L_2 \) is an approximation (optimal in the Frobenius norm sense) of \( Q_\psi \). This approximation problem is addressed in \([\text{Str95}]\) and \([\text{vLP93}]\). These algorithms and the ML method for the white noise case are compared numerically to the proposed algorithm in Chapter 5.

4.A Equivalence of the proposed estimator to Maximum Likelihood

The claim in Section 4.3 that, under a temporally white noise assumption, \( \hat{C}_{ML} = \hat{C} \) (defined in (4.34) and (4.33)) will be proven here. Under the assumptions, estimation of the covariance matrix \( \hat{Q}_\psi \) in (4.33) reduces to estimating the covariances of the noise and input signal and forming

\[ \hat{Q}_\psi = (\hat{R}_{xx}^{-1} \otimes \hat{R}_{ee}). \tag{4.39} \]

The noise correlation matrix can be estimated as

\[ \hat{R}_{ee} = \frac{1}{N} \sum_{i=1}^{N} (y(t) - \hat{Ψ}x(t))(y(t) - \hat{Ψ}x(t))^T \]

\[ = \hat{R}_{yy} - \hat{R}_{yx}\hat{R}_{xx}^{-1}\hat{R}_{yx}^T. \tag{4.40} \]

Since the weighting matrix \( \hat{Q}_\psi^{-1} = (\hat{R}_{xx} \otimes \hat{R}_{ee}^{-1}) \) in this special case is of the form \( (Q_1 \otimes Q_2) \) it can be shown (see, e.g., \([\text{MMH03}]\), Theorem 3) that the matrix \( \hat{C} \) that minimizes (4.33) is

\[ \hat{C} = \hat{R}_{ee}^{-1/2}U\Sigma V^T\hat{R}_{xx}^{-1/2}. \tag{4.41} \]

The matrices \( U \) and \( V \) are obtained from the SVD

\[ U\Sigma V^T = \hat{R}_{ee}^{-1/2}\hat{R}_{yx}\hat{R}_{xx}^{-1}\hat{R}_{xx}^{1/2}. \tag{4.42} \]
and $\Sigma_r$ is obtained from $\Sigma$ in (4.42) by setting all but the largest $r$ singular values to zero. It is easy to verify that this is equivalent to

$$\hat{C} = \hat{R}_{xx}^{-1/2} \hat{R}_{yx} \hat{R}_{xy} \hat{R}_{xx}^{-1/2} V_r V_r^T \hat{R}_{xx}^{-1/2},$$

where $V_r$ is defined analogously to $\tilde{V}_r$. In order to show that $\tilde{V}_r \tilde{V}_r^T = V_r V_r^T$ as defined in (4.35) and (4.42) consider

$$\hat{R}_{yy} = \hat{R}_{ee} + \hat{R}_{xy} \hat{R}_{xx}^{-1/2} \hat{R}_{yx} \hat{R}_{xx}^{-1/2} \hat{R}_{ee}$$

$$= \hat{R}_{ee}^{1/2} (I_m + \hat{R}_{ee}^{-1/2} \hat{R}_{xy} \hat{R}_{xx}^{-1/2} \hat{R}_{yx} \hat{R}_{ee}^{-1/2}) \hat{R}_{ee}^{1/2}$$

$$= \hat{R}_{ee}^{1/2} U (I_m + \Sigma^2)^{-1} U^T \hat{R}_{ee}^{1/2},$$

where use was made of (4.42) in the last equality. Next, consider the matrix

$$\tilde{V} \Sigma^2 \tilde{V}^T = \hat{R}_{xx}^{-1/2} \hat{R}_{xy}^{-1/2} \hat{R}_{yx} \hat{R}_{xx}^{-1/2}$$

$$= \hat{R}_{xx}^{-1/2} \hat{R}_{yy} \hat{R}_{xx}^{-1/2} U$$

$$\times (I_m + \Sigma^2)^{-1} U^T \hat{R}_{ee}^{-1/2} \hat{R}_{yx} \hat{R}_{xx}^{-1/2}$$

$$= V \Sigma U^T U (I_m + \Sigma^2)^{-1} U^T \Sigma U^T$$

$$= V \Sigma (I_m + \Sigma^2)^{-1} U^T \Sigma U^T \equiv VFV^T.$$

(4.45)

Since $F$ defined above is diagonal, it can be concluded that both $V$ and $\tilde{V}$ contain the orthonormal eigenvectors of $\hat{R}_{xx}^{-1/2} \hat{R}_{yx} \hat{R}_{xy} \hat{R}_{xx}^{-1/2}$. Thus, the columns of $V$ must be the columns of $\tilde{V}$, except for a factor $\pm 1$. It can also be concluded that the diagonal elements of $F$ are the diagonal elements of $\Sigma^2$ since they are the corresponding eigenvalues. It remains to show that the ordering of the diagonal elements of $F$ and $\Sigma$, and thus the columns of $V$ and $\tilde{V}$, are the same.

Next, denote the diagonal elements of $\Sigma$ by $a_i$. From the properties of the SVD (as defined and used in this work) it follows that $a_{i+1} \leq a_i$. The ith diagonal element of $F$, defined in (4.45), is

$$\frac{a_i^2}{1 + a_i^2} = 1 - \frac{1}{1 + a_i^2}.$$  (4.46)

By noting that this function is monotonically increasing (for positive $a_i$), it follows that $F$ has positive diagonal elements of falling magnitude. This proves that $F = \Sigma^2$ and it can be concluded that $\tilde{V}_r \tilde{V}_r^T = V_r V_r^T$. Inserting this result into (4.34) and comparing to (4.43) show that $C = \hat{C}_{ML}$ under the assumptions stated above.
Chapter 5

Numerical study

Monte Carlosimulations have been performed in order to illustrate the performance of the new method. Samples were generated according to (2.1). The noise was generated according to the generalized MA process

$$
e(t) = \sum_{k=0}^{T_e - 1} M_k^e w(t - k), \quad (5.1)$$

where $M_k^e \in \mathbb{R}^{m \times m}$ for $k = 0, \ldots, T_e - 1$ are matrices with randomly chosen elements, drawn from a gaussian zero-mean, unit variance distribution and squared. The driving noise $w(t)$ is white gaussian with covariance $\sigma_w^2 I_m$. The coefficient matrices $M_k^e$ were kept constant while the noise $w(t)$ was regenerated each Monte Carlo experiment. The input signal was generated similarly, based on

$$x(t) = \sum_{k=0}^{T_x - 1} M_k^x u(t - k). \quad (5.2)$$

Each element of the matrices $M_k^x \in \mathbb{R}^{n \times n}$ for $k = 0, \ldots, T_x - 1$ is randomly chosen and positive similar to the $M_k^e$ coefficients. The driving noise $u(t)$ is white gaussian with covariance $\sigma_u^2 I_n$. It was regenerated each Monte Carlo experiment while the $M_k^x$ matrices were kept constant.

When calculating the asymptotic covariance expression (4.14), theoretical $R_{ee}(\tau)$ and $R_{xx}(\tau)$ were used. When evaluating the Cramér-Rao lower bound this was not possible since it depends on the input sequence directly. In order to make a comparison possible, the CRB shown is the average over several realizations of a random input sequence with temporal and spatial correlation given by $R_{xx}(\tau)$.

The SNR was defined to $\frac{\sigma_u^2}{\sigma_w^2}$. The “true regressor” $C_0$ was fixed to the rank two
matrix
\[
\begin{pmatrix}
0.0276 & 0.1847 \\
0.1990 & 0.3721 \\
0.2628 & 0.1943 \\
0.5100 & 0.1732 \\
0.6007 & -0.5705
\end{pmatrix}
\begin{pmatrix}
0.6225 & -0.2954 \\
0.1560 & 0.0023 \\
-0.3122 & 0.2979 \\
0.1626 & -0.1787 \\
-0.9209 & -0.8700
\end{pmatrix}^T.
\]

For the WLRA part of the algorithm, three alternatives were suggested in Section 3.3. Two single step Newton-methods were suggested in addition to a single iteration alternating projections algorithm. In the examples studied, the two Newton step based algorithms gave almost exactly the same result. Thus they are represented as a single entity in the plots. Note that all included algorithms are non-iterative, and thus no stopping criteria is needed.

In Figure 5.1, 5.2 and 5.3, the experimental mean of the quantity \( \frac{1}{mn} \| \hat{C}_0 - \hat{C} \|_F^2 \), forming an estimate of the MSE\(^1\), is presented. It is compared to the theoretical asymptotical covariances derived in Section 4.1.

In Figure 5.1 and 5.2 the methods for the reduced rank linear regression referred to in Section 4.3 are compared to the method in this work for a few choices of the parameters \( p \) and \( f \). In Figure 5.1 the sample set size \( N \) was varied at a fixed SNR while in Figure 5.2 the SNR was varied at a fixed \( N \). For comparison, the result of a straightforward least squares fit to the data (3.18) is included together with the result for unweighted rank reduction of \( \Psi \) using the SVD. Clearly, the proposed method has an advantage over existing methods when both the noise and the input signals are temporally auto correlated. The ML, CVA2 and CVA3 methods seem to perform very similarly. The performance of the new method using \( p = f = 0 \) is slightly better than that of those methods, since the optimal weighting exploits some of the noise correlation. The performance improves significantly as \( p \) and \( f \) increase. The gain in performance with the proposed method depends on the noise correlation. The asymptotical covariance expressions are valid for fairly small \( N \) if \( p \) and \( f \) are small but the convergence becomes slower as they increase. The proposed method is quite robust to poor SNR and small sample set sizes. The CRB for the problem, (4.30), is also included in the plots. It is significantly lower than the MSE of all included methods.

Simulations were also performed in which the noise was temporally white (\( T_e = 1 \)). In that case the CVA2, CVA3, ML and the method proposed in this work had very similar performance, also for relatively small sample set sizes. In the white noise case the method of this thesis is asymptotically efficient (as is the ML method, of course). Recall that the algebraic equivalence result from Section 4.3 is valid if the noise is known to be white and that knowledge is exploited when forming the weighting matrix.

In Figure 5.3, the new approach to the WLRA problem presented in Section 3.3 is compared to unweighted rank reduction (setting \( Q = I \) in (3.22)) and one step

\(^1\)MSE = \( E \left\{ vec^T (C_0 - \hat{C}) vec (C_0 - \hat{C}) \right\} \frac{1}{mn} \)
of the well-known alternating projections method described in Section 3.3. The minimizer of (3.9) when the rank constraint is disregarded, $\Psi$, is also included for comparison. The new method performs better than the alternating projections method but the difference is small in this example. The most important observation is the importance of the weighting matrix when performing the rank reduction.

In Figure 5.4, the expression (4.14) for the asymptotical covariance of the new method is compared to the CRB (4.30). The design variables $p$ and $f$ in (3.3) are varied. The plot for the CRB shows the result of Monte Carlo simulations over the input signal $x(t)$. The input was generated using the MA model described above and the noise covariance structure was chosen to agree with that model as well. The asymptotical covariance for the ML method (which is derived assuming white noise and thus not true ML with this signal model) and the CRB without rank constraint are also included. Clearly the asymptotical covariance of the proposed method approaches the CRB as the lag-lengths $p$ and $f$ increase. The rate of this convergence depends on the correlation lengths of the noise and the input. The exact correspondence is unclear. It is interesting to note that the proposed method performs better asymptotically than the ML method also if $p$ and $f$ are zero. Naturally this does not hold if the noise is white, in which case the two methods are asymptotically equivalent.

In Figure 5.5, the performance of the rank detection algorithm is investigated. The probability of a correctly detected rank is estimated. The probability seems to stabilize close to the 95% correct detections that would correspond to the selected probability ($\alpha = 0.05$) of falsely rejecting the null hypothesis.
Figure 5.1: Experimental MSE (over 500 realizations) of estimated linear regression matrices and asymptotical theoretical covariances. The data set size was varied between 100 and 3162 while the SNR was fixed to 10 dB. The noise and input correlation lengths $T_x$ and $T_z$ were both set to 10. The squares correspond to a least squares fit to the data. The hollow stars correspond to unweighted rank reduction of the least squares estimate. The ‘*’s, ‘x’ s and ‘+’ s correspond to the method presented in this work with $p$ and $f$ set to 0, 2, and 4, respectively. The ‘o’ s correspond to the ML method derived for temporally white noise from [SV96]. The ‘<’ s correspond to the CVA2 method from [GR02] (the CVA3 method gave almost exactly the same result in this scenario). The asymptotical covariance for the ML method is the dotted line. The asymptotical covariances corresponding to $p = f = 0, 2, 4$ for the method of this work are the dash-dotted, dashed, and the solid lines, respectively. The solid line marked with dots is the CRB (4.30). The CRB is evaluated using Monte Carlo simulations over $x(t)$.
Figure 5.2: Experimental MSE (over 500 realizations) of estimated linear regression matrices and asymptotical theoretical covariances. The SNR was varied between -10 and 20 dB while the data set size $N$ was fixed to 700. The noise and input correlation lengths $T_x$ and $T_z$ were both set to 10. The squares correspond to a least squares fit to the data. The hollow stars correspond to unweighted rank reduction of the least squares estimate. The ‘*’s, ‘x’ s and ‘+’ s correspond to the method presented in this work with $p$ and $f$ set to 0, 2, and 4, respectively. The ‘o’ s correspond to the ML method derived for temporally white noise from [SV96]. The <s> correspond to the CVA2 method from [GR02] (the CVA3 method gave almost exactly the same result in this scenario). The asymptotical covariance for the ML method is the dotted line. The asymptotical covariances corresponding to $p = f = 0, 2, 4$ for the method of this work are the dash-dotted, dashed, and the solid lines, respectively. The solid line marked with dots is the CRB (4.30). The CRB is evaluated using Monte Carlo simulations over $x(t)$. 
Figure 5.3: Experimental MSE (over 500 realizations) of estimated linear regression matrices and asymptotical theoretical covariances. The IV-method ($p = f = 0$) presented in this work was used together with different methods for solving the WLRA subproblem (3.22). The data set size was varied between 100 and 3162 while the SNR was fixed to 10 dB. The noise and input correlation lengths $T_x$ and $T_z$ were both set to 10. The squares correspond to the method presented here without the rank reduction ($\Psi$ in (3.20)). The 'o':s correspond to unweighted rank reduction ($Q = I$ in (3.22)). The hollow stars correspond to the result of using the alternating projections method to solve the WLRA. The 'x':s correspond to the method of Section 3.3 with the $\theta_{AB}$ parameterization. The other parameterization gives the same result. The solid line is the asymptotical covariance expression.
Figure 5.4: The data set size $N$ was 700 and the SNR was 10dB. The covariance lengths $T_x$ and $T_c$ are both fixed to 10. The matrix $Q$ represents different covariance matrices corresponding to the different expressions plotted in the graph. The stars show the asymptotic covariance expression for the proposed method for different values of $p, f$ ($Q$ is given by (4.14)). The solid line shows the Cramèr-Rao lower bound ($Q$ is given by (4.30) averaged over 30 realizations of $x(t)$). The dash-dotted line shows the asymptotic covariance of the ML method. Note that this is not true ML in this case since it was derived for another noise model (the corresponding covariance, $Q$, can be derived similar to Section 4.1). The dashed line shows the CRB for linear regression without rank constraints.
Figure 5.5: Experimental probability of correct rank detection (over 500 realizations) using the algorithm outlined in Section 3.4. The parameter $\alpha$ was set to 0.05. The data set size was varied between 100 and 3500 while the SNR was fixed to 10 dB. The rank 2 regression matrix is given in (5.3). The parameters $p$ and $f$ were both set to 2. The input signal and the noise were generated according to the model described in the text with $T_x = T_c = 10$. 
Part II

Optimal utilization of signal-free samples for array processing in unknown colored noise-fields
Chapter 6

Introduction to DOA estimation in the presence of colored noise

The problem of estimating parameters of narrowband signals impinging on an array of sensors is important in a wide range of applications. Examples include radar, sonar, seismic exploration, microwave and radio communications and biomedicine. Consider the data model that was developed in an example in Chapter 1. It is

\[ x(t) = A(\theta)s(t) + n(t), \]

where, at time \( t \), \( x(t) \) is the \( m \)-vector of sensor outputs, \( s(t) \) is the \( d \)-vector of impinging signals and \( n(t) \) is an \( m \)-vector of noise. The \( m \times d \)-matrix \( A(\theta) \) is composed of the steering vectors of the array, parameterized by the \( n_p \)-vector \( \theta \).

The direction of arrival (DOA) estimation problem is that of estimating \( \theta \) from the sequence \( \{x(t)\}_{t=0}^{N-1} \). In this work, the vectors \( s(t) \) and \( n(t) \) are assumed to be mutually independent, temporally white circularly symmetric complex Gaussian vectors with covariance matrices \( P \) and \( Q \), respectively. A number of estimation techniques have been proposed for this data model in the literature, see, e.g., [Sch79, RK89, SS90a, VO91]. Most algorithms are based on the assumption that the noise is spatially white and uniform, i.e., \( Q = \sigma^2 I_m \). If this assumption cannot be made, a few approaches are suggested in the literature:

The first one is to parameterize the covariance matrix \( Q \) and then perform a joint estimation of \( \theta \) and the noise parameters [FW95, GO99, SWW96, PG01, NK96, YD95]. A similar approach to the problem of signal power estimation is taken in [GMB95].

Another possibility is to assume and exploit certain structures in the noise covariance matrix (without parameterizing it), see e.g., [ZW93] and [SWW96].

Alternatively, some structure can be imposed on the signals. In [VS97], the signals are assumed to be the superposition of a known set of basis functions. The data model of that work can then be written

\[ x(t) = A(\theta)B^Ts(t) + n(t), \]

(6.2)
where $s(t)$ is known and $B$ is unknown. Note the similarity to the estimation problem treated in Part I of this thesis.

While the above approaches have the distinct advantage of not requiring prior knowledge of the elements of $Q$, they do require prior knowledge of either the structure of the noise covariance or the signals. Fundamental identifiability issues pose a limit on how general the imposed structure can be. An interesting Bayesian approach, based on a non-informative prior distribution of the noise covariance parameters, has also been suggested [WRWQ92, RW92].

The second standard approach is to estimate $Q$ using a separate batch of signal-free samples and then use the so-obtained estimate to whiten the noise part of the signals $x(t)$. After the whitening, any of the standard methods for DOA estimation can be used.

While this approach is frequently suggested or implicitly assumed in the literature, there are few results published on the effects on the quality of the DOA estimates due to errors in the whitening. An exception to this is given in [Vil93], where the effect of small deterministic deviations from the uniform spatially white noise assumption is analyzed for some common algorithms. In [SK92] and [SK93] the effect of subspace estimation algorithms of stochastic perturbations of the noise covariance and array manifold are investigated. The derivations in those two papers do however not assume that the noise covariance errors stem from an imperfectly estimated whitening matrix. Thus, they are not directly applicable to the problem at hand. Also, unlike the present work, [SK92] and [SK93] do not treat finite sample effects in the same framework as perturbations in the noise covariance matrix.

6.1 Outline of Part II

In this work a statistical approach is taken to assess the impact of estimation errors in the whitening matrix and to devise algorithms with improved performance. It is assumed that $M$ samples $\{n(t-M)\}_{t=0}^{M-1}$ drawn from the same distribution as $n(t)$ are available in addition to the data $\{x(t)\}_{t=0}^{N-1}$. A more formal description of the data model is given in the next section.

In Section 7.1, the Cramér-Rao lower bound (CRB) on the covariance of any unbiased estimator of $\theta$, given the proposed data model, is derived. The expression is relatively compact, and allows for a comparison to well-known results.

In Section 8.1, the asymptotical covariance of the Weighted Subspace Fitting (WSF) [VO91] estimate based on whitened data is derived. The whitening transformation is estimated from the $M$ signal-free samples that are assumed to be available. Moreover, an optimal weighting is introduced that improves performance compared to the standard weighting at no additional computational cost. With the new weighting, whitened WSF is asymptotically efficient for the given data model. The expression for the asymptotical covariance allows for a useful and simple comparison to the standard (assuming known noise covariance, and thus perfect whitening) WSF asymptotical covariance derived in, e.g., [OVK92].
6.2. DATA MODEL

In Section 8.2, an asymptotically optimal algorithm based on covariance matching is derived for the estimation problem at hand. This new algorithm does not explicitly rely on pre-whitening of the data.

Some numerical results are given in Chapter 9.

6.2 Data model

The estimation problem that is the subject of the next few chapters is that of estimating $\theta$ from the data

$$\{n(-M), \ldots, n(-1), x(0), \ldots, x(N-1)\}. $$

Both $n(t)$ and $x(t)$ are temporally white complex Gaussian circularly symmetric random vectors with covariance matrices

$$E[n(t)n^*(t)] = Q,$$

$$E[x(t)x^*(t)] = A(\theta)PA^*(\theta) + Q = R. \quad (6.3)$$

The (known) functional form of the matrix

$$A(\theta) = [a(\theta_1), \ldots, a(\theta_{(d+1)n_\theta})] \quad (6.4)$$

depends, among other things, on the array geometry and sensor characteristics. When there is no risk of confusion, the argument will be dropped. The vector $[\theta]_{(l-1)t+1:t}$ is the subvector of $\theta$ containing the parameters associated with source $i$. If not otherwise stated, the results of this work hold when the steering vectors, $a(\theta)$, are parameterized by multiple parameters per source ($l > 1$). The steering vectors are assumed to be twice continuously differentiable w.r.t. $\theta$, and to possess a bounded third order derivative. This assumption is necessary for the asymptotic analysis that will follow. The array must also be unambiguous, which means that $a(\theta_1)$ and $a(\theta_2)$ are linearly independent for any $\theta_1 \neq \theta_2$. It is assumed that the unknown matrix $Q$ is a p.d. matrix. Furthermore, identifiability requires that $\frac{(m+d')}{2} > d$, where $d'$ is the rank of the signal correlation matrix, $P$ (see [WZ89] for details). If not otherwise stated, the correlation matrix of the signals, $P$, is p.s.d. meaning that $d' \leq d$. 
Chapter 7

Bounds on the performance of the DOA estimate

The stochastic CRB for the case of uniform white noise \( Q = \sigma^2 I \) was derived in [SN90b, OVK92]. I should be noted that in the references, \( \sigma^2 \) is assumed to be an unknown scalar, however the derived CRB is unchanged if \( \sigma^2 \) is known. Obviously, since this assumption implies a known noise covariance (up to a scaling factor), any extra signal-free samples will not affect the attainable estimation accuracy. For easy reference, this bound will be generalized (in a very straightforward way) to the case of colored noise with a general known noise covariance matrix. To that end, consider the transformed data

\[
\bar{x}(t) = Zx(t), \\
ZZ = Q^{-1},
\]

and note that the transformation is invertible. The covariance matrix is then

\[
\text{Cov} [ \bar{x}(t) ] = \bar{R} = \bar{A} \bar{A}^* + I,
\]

with

\[
\bar{A} = ZA, \\
\bar{R} = ZRZ.
\]

Since (7.2) is equivalent to the covariance in the data model of [SN90b, OVK92], it can be concluded that the corresponding CRB result can be used. Thus, the sought CRB is (after a minor rearrangement)

\[
NCRB_\theta = \left( 2\text{Re} \, D_A^* ( [P\bar{A}^*E_xA_x^{-1}E_x^*\bar{A}P]^T \otimes Z\bar{f}^*Z) D_A \right)^{-1},
\]

55
CHAPTER 7. BOUNDS ON THE PERFORMANCE OF THE DOA ESTIMATE

where

\[ D_A = \{\text{vec}(A_1), \ldots, \text{vec}(A_n)\}, \]

\[ \Pi = \Pi_A, \]

\[ \hat{R} = E_n A_s E_n^* + E_n E_n^*. \]  \hspace{1cm} (7.5)

The diagonal matrix \( A_s \) contains the \( d' \) largest eigenvalues of the whitened array covariance \( \hat{R} \). The corresponding eigenvectors (that span the signal subspace) are the columns of \( E_n \). The remaining eigenvectors are the columns of \( E_n \). The matrix \( A_s \) is the elementwise derivative of \( A \) w.r.t. \( [\theta]^i \).

The uniform noise CRB was generalized to unknown colored noise fields (\( Q \) is parameterized with unknown parameters) in [GSPL02]. This bound will not be stated here. It suffices to note that the bound that will be derived below is for a different data model, and it can not be viewed as a specialization of the CRB mentioned above.

7.1 The Cramér-Rao bound with signal-free samples available

The CRB that will be presented in this section is conceptually different from the ones presented above due to the extra observed signal-free samples in the data model. It should be noted that the bound is for an estimator with no knowledge of the rank \( d' \) of \( P \), see [SOVM96] for more details about this assumption.

Define the sample covariances

\[ \hat{R}_1 = \frac{1}{M} \sum_{k=-M}^{-1} n(k)n^*(k) \]

and

\[ \hat{R}_2 = \frac{1}{N} \sum_{k=0}^{N-1} x(k)x^*(k). \]  \hspace{1cm} (7.6)

By exploiting the statistical independence of the observations, the negative log-likelihood function can be written (neglecting constant terms)

\[ l(\eta) = M \log |Q| + M \text{tr}\{Q^{-1}\hat{R}_1\} + N \log |R| + N \text{tr}\{R^{-1}\hat{R}_2\}, \]  \hspace{1cm} (7.7)

where

\[ \eta = [\theta^T, \phi^T, \xi^T]^T. \]  \hspace{1cm} (7.8)

The \( d^2 \)-vector \( \phi \) and the \( m^2 \)-vector \( \xi \) contain the real parameters required to parameterize the hermitian matrices \( P \) and \( Q \), respectively. In the parameterization,
the fact that the matrices are p.d. or p.s.d. is not exploited. This will not affect the final result, see, e.g., [SOVM96] or [GSPL02]. The so-called Bang’s formula, see, e.g., [OVK92], gives

\[
[I(\eta)]_{ij} = \mathbb{E} \left[ \frac{\delta^2 (\eta)}{\delta [\eta]_i \delta [\eta]_j} \right] = M \operatorname{tr} \{ Q^{-1} Q_i Q^{-1} Q_j \} + N \operatorname{tr} \{ R^{-1} R_i R^{-1} R_j \},
\] (7.9)

where \( X_j \) denotes the element-wise derivative of the matrix \( X \) w.r.t. \( [\eta]_j \). This gives the Fisher information matrix

\[
[I(\eta)]_{ij} = M \operatorname{tr} \{ Q^{-1} Q_i Q^{-1} Q_j \}
+ N \operatorname{tr} \{ (APA^* + Q)^{-1} (A_i PA^* + AP_i A^* + APA^*_i + Q_i) \}
\times (APA^* + Q)^{-1} (A_j PA^* + AP_j A^* + APA^*_j + Q_j) \}.
\] (7.10)

The CRB for the signal parameters is given by the upper-left \( n_\theta \times n_\theta \) submatrix of \( I^{-1}(\eta_0) \), with \( \eta_0 \) being the true values of the parameters. A simple expression can be found by exploiting the independent parameterization of the matrices \( P, Q \) and \( A(\theta) \). To that end, consider partitioning the information matrix as

\[
I(\eta) = \begin{pmatrix}
I_{\theta \theta} & I_{\theta P} & I_{\theta Q} \\
I_{P \theta} & I_{PP} & I_{P Q} \\
I_{Q \theta} & I_{QP} & I_{QQ}
\end{pmatrix}
\].
\] (7.11)

with obvious definitions of the submatrices. Then the CRB for the parameters in \( \theta \) can be shown to be

\[
\text{CRB}_\theta = (H_1 + H_2)^{-1},
\]

\[
H_1 = I_{PP} (I_{PP} - I_{PQ} I_{QP}^{-1} I_{QQ})^{-1} (I_{PQ} I_{QP}^{-1} I_{Q\theta} - I_{P\theta}),
\]

\[
H_2 = I_{\theta \theta} + I_{\theta Q} (I_{QQ} - I_{QP} I_{PP}^{-1} I_{PQ})^{-1} (I_{QP} I_{PP}^{-1} I_{P\theta} - I_{Q\theta}).
\] (7.12)

In order to give the CRB in a more compact form, the definition

\[
\alpha = \frac{N}{M}
\] (7.13)

is natural. By simplifying (7.12), and by using the definitions introduced in the beginning of this chapter, it is possible to state:

**Theorem 1.** The CRB for the parameters of \( \theta \) in the data model presented in Section 6.2 is given by

\[
N \text{CRB}_\theta = \left( 2 \operatorname{Re} D_A^* ( [P \tilde{A}^* E_s (A_s + \alpha I)^{-1} E_s \tilde{A} P]^T \otimes Z P^T Z ) D_A \right)^{-1}.
\] (7.14)
CHAPTER 7. BOUNDS ON THE PERFORMANCE OF THE DOA ESTIMATE

The proof is given in Appendix 7.A. Since it holds that

\[ \hat{A}^*E_s(A_s + \alpha I)^{-1}E_s^*\hat{A} = \hat{A}^*(\hat{R} + \alpha I)^{-1}\hat{A}, \]  \hspace{1cm} (7.15)

the result

\[ [\text{CRB}_\theta^{-1}]_{ij} = N2\text{Re} \{\text{tr}}\{A_s^*Z\Pi^+Z\alpha^*A^*(\hat{R} + \alpha I)^{-1}A\} \]  \hspace{1cm} (7.16)

is easily obtained. As \( \alpha \) approaches zero, this bound coincides with the Cramér-Rao bound for the case with perfectly known noise covariance matrix. This is seen by comparing (7.14) with (7.4).

7.A Proof of Theorem 1

Begin by defining

\[ D_Q = \{\text{vec}\{Q_1\}, \ldots, \text{vec}\{Q_m^2\}\} \]  \hspace{1cm} (7.17)

and \( D_P \) and \( D_{A^*} \) similarly. It is not necessary to precisely specify how the matrices \( P \) and \( Q \) are parameterized. It is sufficient to note that the (hermitian) matrices have \( d^2 \) and \( m^2 \) degrees of freedom respectively. If extra structure is imposed, i.e., fewer degrees of freedom, then the derivations to follow are invalid. Note also that \( D_Q \) and \( D_P \) are invertible matrices. In order to simplify (7.12), consider

\[ [I_{QQ}]_{ij} = \text{tr}\{MQ^{-1}Q_iQ^{-1}Q_j + NR^{-1}Q_iR^{-1}Q_j\} \]
\[ = M\text{vec}^*\{Q_i\}(Q^{-T}Q^{-1})\text{vec}\{Q_j\} + N\text{vec}^*\{Q_i\}(R^{-T}R^{-1})\text{vec}\{Q_j\}. \]  \hspace{1cm} (7.18)

It follows immediately that

\[ I_{QQ} = D_Q^*[M(Q^{-T}\otimes Q^{-1}) + N(R^{-T}\otimes R^{-1})]D_Q. \]  \hspace{1cm} (7.19)

In the same way it is easy to show that

\[ I_{PP} = ND_P^*(A^TR^{-T}A^T\otimes A^*R^{-1}A)D_P, \]
\[ I_{QP} = I_{PQ} = ND_Q^*(R^{-T}A^T\otimes R^{-1}A)D_P, \]
\[ I_{Q\theta} = I_{\theta Q} = ND_Q^*(R^{-T}\otimes R^{-1})M_\theta, \]
\[ I_{P\theta} = I_{\theta P} = ND_P^*(A^TR^{-T}\otimes A^*R^{-1})M_\theta, \]
\[ I_{\theta \theta} = NM_\theta^*(R^{-T}\otimes R^{-1})M_\theta, \]  \hspace{1cm} (7.20)

where, for convenience, the matrix

\[ M_\theta = (A^TP^T\otimes I)D_A + (I \otimes AP)D_{A^*}. \]  \hspace{1cm} (7.21)
was introduced. It follows that

\[
I_{PQ}^{-1}I_{QQ}^{-1}I_{P\theta} = D_p(A^T \otimes A^*) \left[ N(R^{-T} \otimes R^{-1}) \right.
\]

\[
\times \left( \frac{M}{N} Q^{-T} \otimes Q^{-1} + R^{-T} \otimes R^{-1} \right)^{-1} (R^{-T} \otimes R^{-1})^{-1}
\]

\[
- N(R^{-T} \otimes R^{-1}) \right] M_\theta
\]

\[
= -ND_p(A^T \otimes A^*) \left[ R^T \otimes R + \frac{N}{M} Q^T \otimes Q \right]^{-1} M_\theta. \tag{7.22}
\]

The matrix inversion lemma (MIL) (see (1.15)) was used to obtain the last equality. In a similar manner it holds that

\[
I_{PP}^{-1}I_{PQ}^{-1}I_{QQ}^{-1}I_{QP} = ND_p(A^T \otimes A^*) \left[ R^T \otimes R + \frac{N}{M} Q^T \otimes Q \right]^{-1} \left( A^T \otimes A \right) D_p \tag{7.23}
\]

and that

\[
I_{QQ}^{-1}I_{QP}^{-1}I_{PP}^{-1}I_{PQ} = ND_q \left[ \frac{M}{N} Q^{-T} \otimes Q^{-1} \right.
\]

\[
+ R^{-T} \otimes R^{-1} - \left( R^{-1} A(A^* R^{-1} A)^{-1} A^* R^{-1} \right)^T
\]

\[
\left. \otimes R^{-1} A(A^* R^{-1} A)^{-1} A^* R^{-1} \right] D_q. \tag{7.24}
\]

Finally, it is straightforward to show that

\[
I_{QP}^{-1}I_{PP}^{-1}I_{P\theta} - I_{P\theta} = ND_q \left[ -R^{-T} \otimes R^{-1} \right.
\]

\[
+ \left( R^{-1} A(A^* R^{-1} A)^{-1} A^* R^{-1} \right)^T
\]

\[
\left. \otimes R^{-1} A(A^* R^{-1} A)^{-1} A^* R^{-1} \right] M_\theta. \tag{7.25}
\]

Consequently it holds that

\[
H_1 = -NM_\theta \left( R^{-T} A^T R^{-1} A \right) (A^T \otimes A^*)
\]

\[
\times \left[ R^T \otimes R + \frac{N}{M} Q^T \otimes Q \right]^{-1} (A^T \otimes A)^{-1}
\]

\[
\times (A^T \otimes A^*) \left[ R^T \otimes R + \frac{N}{M} Q^T \otimes Q \right]^{-1} M_\theta, \tag{7.26}
\]

where \(H_1\) was defined in (7.12). In order to simplify notation, introduce

\[
M_\theta = \sqrt{N\left(Z^T \otimes Z\right)} M_\theta, \tag{7.27}
\]
and consider (7.5) and (7.13). Then

\[
H_1 = -M_\theta^*(R^{-T} \otimes \tilde{R}^{-1})(A^*T \otimes \tilde{A})
\]

\[
\times \left[ (A^T \otimes A^*) [R^T \otimes \tilde{R} + \alpha I]^{-1}(A^*T \otimes \tilde{A}) \right]^{-1}
\]

\[
\times (A^T \otimes A^*) [R^T \otimes \tilde{R} + \alpha I]^{-1}M_\theta.
\] (7.28)

This expression can be simplified using the relations

\[
\tilde{R}^{-1} = I - AP(I + \tilde{A}^* \tilde{A})^{-1} \tilde{A}^*,
\]

\[
\tilde{R}^{-1} \tilde{A} = \tilde{A}(I + P(\tilde{A}^* \tilde{A}))^{-1},
\]

\[
\tilde{A}^* \tilde{R}^{-1} \tilde{A} = ((\tilde{A}^* \tilde{A})^{-1} + P)^{-1}
\] (7.29)

that can be obtained by applying the MII repeatedly. By again applying the MII, the result

\[
(A^T \otimes A^*) [R^T \otimes R + \alpha I]^{-1}(A^*T \otimes A)
\]

\[
= (A^*R^{-1} \tilde{A})^T \otimes A^*R^{-1} \tilde{A} - (A^T \tilde{R}^{-T} \otimes A^* \tilde{R}^{-1})
\]

\[
\times [\alpha^{-1} I + \tilde{R}^{-T} \otimes R^{-1}]^{-1}
\]

\[
\times (R^{-T} \tilde{A}^*T \otimes \tilde{R}^{-1} \tilde{A}).
\] (7.30)

is obtained. By combining (7.29) and (7.30) with the MII, a few calculations give

\[
\left[ (A^T \otimes A^*) [R^T \otimes R + \alpha I]^{-1}(A^*T \otimes A) \right]^{-1} =
\]

\[
(A^*R^{-1} \tilde{A})^{-T} \otimes (A^*R^{-1} \tilde{A})^{-1} + (A^*T \otimes A^*)
\]

\[
\times [\alpha^{-1} I + R^{-T} \otimes R^{-1} - [\Pi R^{-1} \Pi^T \otimes \Pi R^{-1} \Pi]^{-1}
\]

\[
\times (A^*T \otimes A^*).
\] (7.31)

Next the equation

\[
\tilde{A}(A^*R^{-1} \tilde{A})^{-1} \tilde{A}^* = \Pi \tilde{R} \Pi = \tilde{R} \Pi \tilde{R} \Pi
\] (7.32)

is needed. It follows quite immediately from (7.29). Simple calculations give

\[
(A^*T \otimes A)[(A^T \otimes A^*) [R^T \otimes R + \alpha I]^{-1}(A^*T \otimes A)]^{-1}
\]

\[
\times (A^T \otimes A^*) = \Pi^T R^T \otimes \Pi R + (\Pi^T \otimes \Pi)
\]

\[
\times [\alpha^{-1} I + (R^{-T} \otimes R^{-1})(I - \Pi^T \otimes \Pi)]^{-1}(\Pi^T \otimes \Pi).
\] (7.33)
7.A. PROOF OF THEOREM 1

Using the above in (7.28)
\[
H_1 = -M_\theta^*(\Pi^T \otimes \Pi)(\mathbf{R}^T \otimes \mathbf{R} + \alpha \mathbf{I})^{-1} \\
+ \alpha(\Pi^T \otimes \Pi)[\mathbf{R}^T \otimes \mathbf{R} + \alpha(\mathbf{I} - \Pi^T \otimes \Pi)]^{-1} \\
\times (\Pi^T \otimes \Pi)(\mathbf{R}^T \otimes \mathbf{R} + \alpha \mathbf{I})^{-1})M_\theta. \tag{7.34}
\]

Before turning our attention to $H_2$ in (7.12), consider the relation
\[
R^{-1/2}\Pi_{R^{-1/2}}^A R^{-1/2} = R^{-1} - R^{-1/2}A A^* R^{-1} \\
= Z R^{-1} Z - Z R^{-1} \dot{A} (A^* R^{-1} \dot{A})^{-1} \dot{A}^* R^{-1} Z \\
= Z R^{-1} Z - Z R^{-1} \Pi R^{-1} Z \\
= Z R^{-1} (\mathbf{I} - \Pi) Z = Z \Pi^T Z. \tag{7.35}
\]

Use was made of (7.32) in the third equality. The expression
\[
R^{-1} A (A^* R^{-1} A)^{-1} A^* R^{-1} \\
= R^{-1} - R^{-1/2} \Pi_{R^{-1/2} A} R^{-1/2} \\
= Z (R^{-1} - \Pi^T) Z \tag{7.36}
\]
will also be useful. Combining (7.12), (7.24) and (7.25) gives
\[
H_2 = NM_\theta^*(\mathbf{R}^T \otimes \mathbf{R}^{-1} \\
+ (\mathbf{R}^T \otimes \mathbf{R}^{-1}) \left[ \frac{M}{N} \mathbf{Q}^T \otimes \mathbf{Q}^{-1} + \mathbf{R}^{-T} \otimes \mathbf{R}^{-1} \\
- (\mathbf{Z}(\mathbf{R}^{-1} - \Pi^T) \mathbf{Z})^T \otimes \mathbf{Z}(\mathbf{R}^{-1} - \Pi^T) \mathbf{Z} \right]^{-1} \\
\times \left[ \mathbf{Z}(\mathbf{R}^{-1} - \Pi^T) \mathbf{Z}^T \otimes \mathbf{Z}(\mathbf{R}^{-1} - \Pi^T) \mathbf{Z} - \mathbf{R}^{-T} \otimes \mathbf{R}^{-T} \right] \right)M_\theta. \tag{7.37}
\]

After some manipulations this can be written
\[
H_2 = NM_\theta^*(\mathbf{R}^{-T} \otimes \mathbf{R}^{-1}) \left[ \frac{M}{N} \mathbf{Q}^T \otimes \mathbf{Q}^{-1} + \mathbf{R}^{-T} \otimes \mathbf{R}^{-1} \\
- (\mathbf{Z}(\mathbf{R}^{-1} - \Pi^T) \mathbf{Z})^T \otimes \mathbf{Z}(\mathbf{R}^{-1} - \Pi^T) \mathbf{Z} \right]^{-1} \left( \frac{M}{N} \mathbf{Q}^T \otimes \mathbf{Q}^{-1} \right) M_\theta \\
= NM_\theta^*(\mathbf{R}^{-T} \otimes \mathbf{R}^{-1}) \left[ \mathbf{I} + \alpha(\mathbf{Z}^T \otimes \mathbf{Z}^{-1}) \\
\times (\mathbf{R}^{-T} \otimes \mathbf{R}^{-1} - [\mathbf{R}^{-1} - \Pi^T]^T \otimes [\mathbf{R}^{-1} - \Pi^T])(\mathbf{Z}^T \otimes \mathbf{Z}) \right]^{-1} M_\theta. \tag{7.38}
\]
CHAPTER 7. BOUNDS ON THE PERFORMANCE OF THE DOA ESTIMATE

By using \( \bar{\Pi}^+ \bar{R} = \bar{\Pi}^+ \) one obtains

\[
H_2 = M_\theta [\bar{R}^T \otimes \bar{R} + \alpha (I - \bar{\Pi}^T \otimes \bar{\Pi})]^{-1} M_\theta. \tag{7.39}
\]

Applying the MIL to the bracketed expression gives

\[
(\bar{R}^T \otimes \bar{R} + \alpha I)^{-1} + \alpha (\bar{R}^T \otimes \bar{R} + \alpha I)^{-1} (\bar{\Pi}^T \otimes \bar{\Pi}) \times [I - \alpha (\bar{\Pi}^T \otimes \bar{\Pi}) (\bar{R}^T \otimes \bar{R} + \alpha I)^{-1} (\bar{\Pi}^T \otimes \bar{\Pi})]^{-1} \\times (\bar{\Pi}^T \otimes \bar{\Pi}) (\bar{R}^T \otimes \bar{R} + \alpha I)^{-1}. \tag{7.40}
\]

Combining (7.12), (7.34), (7.39) and (7.40) yields

\[
\text{CRB}_\theta = M_\theta (\bar{R}^T \otimes \bar{R} + \alpha I)^{-1} (I - \bar{\Pi}^T \otimes \bar{\Pi}) M_\theta. \tag{7.41}
\]

In order to obtain the last equality it was used that

\[
(\bar{R}^T \otimes \bar{R} + \alpha I)^{-1} (\bar{\Pi}^T \otimes \bar{\Pi}) = (\bar{\Pi}^T \otimes \bar{\Pi}) (\bar{\Pi}^T \otimes \bar{R} + \alpha I)^{-1}. \tag{7.42}
\]

Using

\[
(P^T \bar{A}^T \otimes Z) (I - \bar{\Pi}^T \otimes \bar{\Pi}) = (P^T \bar{A}^T \otimes Z \bar{\Pi}^+ \otimes Z) \tag{7.43}
\]

and

\[
(I \otimes \bar{\Pi}^+) (\bar{R}^T \otimes \bar{R} + \alpha I)^{-1} = (\bar{R}^T \otimes \bar{R} + \alpha I)^{-1} (I \otimes \bar{\Pi}^+) \tag{7.44}
\]

together with symmetry arguments give

\[
\text{CRB}_\theta = N 2 \text{Re} \{D_\Lambda (P^T \bar{A}^T \otimes Z \bar{\Pi}^+) (\bar{R}^T \otimes \bar{R} + \alpha I)^{-1} (\bar{A}^T \bar{P}^T \otimes \bar{\Pi}^+ Z) D_\Lambda \}. \tag{7.45}
\]

In order to complete the proof, note that

\[
\bar{R}^T \otimes \bar{R} + \alpha I = [E_a (A_s - I) E_a^* + I]^T \otimes E_a (A_s - I) E_a^* \\
+ [E_a (A_s - I) E_a^* + (\alpha + 1) I]^T \otimes I. \tag{7.46}
\]

Applying the MIL gives

\[
(\bar{R}^T \otimes \bar{R} + \alpha I)^{-1} = [E_a (A_s - I) E_a^* + (\alpha + 1) I]^{-T} \otimes I \\
- ([\ldots]^T \otimes E_a [\ldots] E_a^*) ([\ldots])^{-1} [\ldots], \tag{7.47}
\]

where \([\ldots]\) denotes a matrix that is of no immediate interest in the derivations (note however that the indicated inverses do exist). Thus,

\[
(P^T \bar{A}^T \otimes Z \bar{\Pi}^+) (\bar{R}^T \otimes \bar{R} + \alpha I)^{-1} (\bar{A}^T \bar{P}^T \otimes \bar{\Pi}^+ Z) = \\
P^T \bar{A}^T [E_a (A_s - I) E_a^* + (\alpha + 1) I]^{-T} \bar{A}^T \bar{P}^T \otimes Z \bar{\Pi}^+ Z. \tag{7.48}
\]
7.A. PROOF OF THEOREM 1

Continue by writing

\[ \bar{A}^* \left[ \mathbf{E}_s (A_s - I) \mathbf{E}_s^* + (\alpha + 1)I \right]^{-1} \bar{A} \]

\[ = \bar{A}^* \left[ \frac{1}{\alpha + 1} I - \frac{1}{(\alpha + 1)^2} \mathbf{E}_s (A_s - I)^{-1} + \frac{1}{\alpha + 1} I \right]^{-1} \bar{A} \]

\[ = \bar{A}^* \mathbf{E}_s \left[ \frac{1}{\alpha + 1} I - \frac{1}{(\alpha + 1)^2} (A_s - I)^{-1} + \frac{1}{\alpha + 1} I \right]^{-1} \bar{A} \]

\[ = \bar{A}^* \mathbf{E}_s [(\alpha + 1)I + (A_s - I)]^{-1} \bar{A}. \]

Finally, the expression for the CRB is

\[ CRB_{\theta}^{-1} = N^2 \text{Re} \left\{ D_A^* [\mathbf{P} \bar{A}^* \mathbf{E}_s (A_s + \alpha I)^{-1} \mathbf{E}_s^* \bar{A} \mathbf{P}]^T \otimes \mathbf{Z} \mathbf{H}^\dagger \mathbf{Z} \} D_A \right\} \quad (7.49) \]

and the proof is complete.
Chapter 8

Algorithms for DOA estimation with asymptotical performance analysis

Two algorithms for the estimation of $\theta$, given the data model presented in Chapter 6, will be treated in this chapter. Each one is provided together with an expression for its asymptotical (in $M$ and $N$ as will be clear later on) covariance. The performance for finite sample sizes is then evaluated by computer simulations in Chapter 9.

The first algorithm is based on \textit{weighted subspace fitting} (WSF), see, e.g., [VO91]. The original WSF method requires the noise to be uniform and spatially white,

$$Q = \sigma^2 I. \quad (8.1)$$

The noise power $\sigma^2$ can be unknown. In order to understand the algorithm, the notion of \textit{signal subspace} is important. The signal subspace is the span of the matrix

$$A(\theta)P. \quad (8.2)$$

Note that, in the noise-free case, the measurements $x(t)$ would be confined to the signal subspace. Next, consider the eigendecomposition of the signal covariance matrix

$$R = A(\theta)PA^*(\theta) + \sigma^2 I = E_sA_sE_s^* + \sigma^2 E_nE_n^*. \quad (8.3)$$

As in the previous chapter, the columns of the matrix $E_s$ are the orthogonal eigenvectors corresponding to the $d'$ largest eigenvalues. The remaining eigenvectors, that correspond to the eigenvalue $\sigma^2$, are the columns of $E_n$. The critical observation is that

\[
\text{span}\{E_s\} \subseteq \text{span}\{A(\theta)\}. \quad (8.4)
\]

Thus, there is a $d \times d'$ matrix, $T$ so that

$$E_s = A(\theta)T. \quad (8.5)$$
CHAPTER 8. ALGORITHMS FOR DOA ESTIMATION WITH ASYMPTOTICAL PERFORMANCE ANALYSIS

The first step in the WSF algorithm is to find an estimate \( \hat{E}_s \) of \( E_s \) by computing the eigenvectors of the sample covariance matrix. Since there is noise in the data, (8.5) will of course not hold exactly with \( E_s \) replaced by \( \hat{E}_s \). Instead the minimization

\[
\arg\min_{\theta, T} \operatorname{tr}\{ (\hat{E}_s - A(\theta)T)^* (\hat{E}_s - A(\theta)T)W \}
\]

(8.6)
gives the estimate. By concentrating w.r.t \( \theta \), the familiar expression

\[
\arg\min_{\theta} \operatorname{tr}\{ \Pi_{\Lambda} \hat{E}_s W \hat{E}_s^* \}
\]

(8.7)
is obtained. The matrix \( W \) is a p.d. weighting matrix, the choice of which affects performance. The choice

\[
W = W_{WSF} = (\Lambda_s - I)^2 \Lambda_s^{-1}
\]

(8.8)
gives the lowest possible asymptotical covariance.

WSF has several attractive properties. It is asymptotically efficient, i.e. it reaches the CRB for large enough data sizes. It can be modified to exploit certain array structures in order to decrease the computational complexity. It works well also for coherent signals (the signal covariance matrix \( P \) is rank deficient).

In the estimation problem treated in this thesis, the critical assumption of spatially white uniform noise in the WSF derivation is violated. If the noise covariance was known (perfectly estimated), then it would be possible to whiten the signals prior to applying WSF. This prewhitening procedure is often implicitly assumed in order to motivate the uniform spatially white noise assumption. In the following section it is shown how performance is affected if the covariance matrix is in fact estimated from a sample set with a size comparable to the size of the signal-containing sample set. This situation corresponds precisely to the data model given in Chapter 6.

The second algorithm presented in this chapter is based on covariance matching and does (at least conceptually) not rely on prewhitening of the data.

8.1 Whitened weighted subspace fitting

The asymptotic (in a sense made clear later on) covariance of the WSF estimate of \( \theta \) when whitened data are used will be derived in this section. The estimate is given by

\[
\hat{\theta}_W = \arg\min_{\theta} V_W(\theta),
\]

\[
V_W(\theta) = \operatorname{tr}\{ \Pi_{\tilde{Z}\Lambda} \hat{E}_s W \hat{E}_s^* \},
\]

\[
\tilde{Z}^* \hat{Z} = \hat{E}_s \tilde{\Lambda} \hat{E}_s^* + \hat{E}_n \tilde{\Lambda}_n \hat{E}_n^*,
\]

\[
\tilde{Z}^* \hat{Z} = \hat{\Lambda}_1^{-1}.
\]

(8.9)
The matrix \( \hat{A}_\alpha \) is diagonal and the diagonal elements are the \( d' \) largest eigenvalues of the whitened array sample covariance matrix, \( \hat{Z}\hat{R}\hat{Z} \). The columns of \( \hat{E}_\alpha \) are the corresponding orthogonal eigenvectors. In the same way the matrix \( \hat{A}_\alpha \) contains the remaining eigenvalues and the columns of \( \hat{E}_\alpha \) are the corresponding eigenvectors. The weighting matrix \( W \) is an arbitrary hermitian p.d. matrix. This whitened version of WSF seems to be the standard choice in the literature. The following theorem can be shown.

**Theorem 2.** Let \( \theta_0 \) be the true parameter vector in the data model of Section 6.2. Then the quantity \( \sqrt{N}(\hat{\theta}_W - \theta_0) \) has a limiting (as \( N \to \infty \) and \( N/M \to \alpha \), \( \alpha \) constant) zero-mean gaussian distribution

\[
\sqrt{N}(\hat{\theta}_W - \theta_0) \sim \text{AsN}(0, C(W)),
\]

\[
C(W) = H^{-1}GH^{-1},
\]

where

\[
[H]_{ij} = \text{Re \{ } A_s^{*}Z\Pi^\dagger ZA_{\alpha}A_{\alpha}^{*}E_{\alpha}W_{s}\hat{Z}A_{\alpha}^{*}\text{\}},
\]

\[
[G]_{ij} = \text{Re \{ } \Pi^{*}Z\hat{A}_{\alpha}^{*}E_{\alpha}W(A_{\alpha} - I)^{-2}(A_{\alpha} + \alpha I)W_{s}\hat{Z}A_{\alpha}^{*}\text{\}}. \tag{8.11}
\]

Furthermore, just as in standard WSF, the weighting matrix \( W \) can be replaced by a consistent estimate without affecting the asymptotical performance.

The proof is given in Appendix 8.A.

It is easy to see that the asymptotical covariance coincides with the standard expression for WSF (see, e.g., [OVK92]) as the ratio \( \alpha = \frac{N}{M} \) tends to zero.

If the array steering matrix \( A \) is parameterized with one parameter per source, then

\[
A_s = \begin{bmatrix} 0_{m \times (i-1)} & a_s & 0_{m \times (n_s - i)} \end{bmatrix} \tag{8.12}
\]

By defining the matrix

\[
\Delta_A = [a_1, \ldots, a_{n_s}] \tag{8.13}
\]

it is easy to show that (8.11) reduces to

\[
H = 2\text{Re \{ } \Delta_A^*Z\Pi^\dagger Z\Delta_A \odot \hat{A}_{\alpha}^{*}E_{\alpha}W_{s}\hat{Z}A_{\alpha}^{*}\text{\}},
\]

\[
G = 2\text{Re \{ } \Delta_A^*Z\Pi^\dagger Z\Delta_A \odot \hat{A}_{\alpha}^{*}E_{\alpha}W(A_{\alpha} - I)^{-2}(A_{\alpha} + \alpha I)W_{s}\hat{Z}A_{\alpha}^{*}\text{\}}. \tag{8.14}
\]

These expressions are easily compared to the well-known expressions for the covariance of standard WSF (obtained by setting \( \alpha = 0 \)). In the general case, with multiple parameters per source, the corresponding expressions are

\[
H = 2\text{Re \{ } D_A \left( \hat{A}_{\alpha}^{*}E_{\alpha}W_{s}\hat{Z}A_{\alpha}^{*} \odot \Pi^\dagger Z \right) D_A \text{\}},
\]

\[
G = 2\text{Re \{ } D_A \left( \hat{A}_{\alpha}^{*}E_{\alpha}W_{s}\hat{Z}A_{\alpha}^{*} \odot \Pi^\dagger Z \right) D_A \text{\}},
\]

\[
\Lambda_{\alpha} = (A_{\alpha} - I)^{-2}(A_{\alpha} + \alpha I). \tag{8.15}
\]
CHAPTER 8. ALGORITHMS FOR DOA ESTIMATION WITH ASYMPTOTICAL PERFORMANCE ANALYSIS

The matrix $D_A$ was defined in (7.5). By modifying and generalizing (to the case of multiple parameters per source) the derivations in [VO91] and [SN90a], the following theorem can be shown.

**Theorem 3.** Let $C(W)$ be given as in (8.10) and (8.15). Then

$$\begin{align*}
C(W) &\succeq C(W_{\text{new}}), \\
W_{\text{new}} &= (A_s - I)^2(A_s + \alpha I)^{-1}. 
\end{align*}$$

(8.16)

**Proof.** Begin by defining (for notational brevity)

$$\begin{align*}
Y &= \hat{A}^\dagger E_s, \\
X &= Z\Pi Z^\dagger.
\end{align*}$$

(8.17)

The claim of the theorem can then be written

$$\begin{align*}
[\Re D_A^*(Y W Y^* \otimes X) D_A]^{-1} &\times [\Re D_A^*(Y W A_\alpha W Y^* \otimes X) D_A] [\Re D_A^*(Y W Y^* \otimes X) D_A]^{-1} \\
&\succeq [\Re D_A^*(Y A_\alpha^{-1}Y^* \otimes X) D_A]^{-1} [\Re D_A^*(Y W Y^* \otimes X) D_A] \succeq 0.
\end{align*}$$

(8.18)

which is equivalent to

$$\begin{align*}
[\Re D_A^*(Y A_\alpha^{-1}Y^* \otimes X) D_A] - [\Re D_A^*(Y W Y^* \otimes X) D_A] \\
&\times [\Re D_A^*(Y W A_\alpha W Y^* \otimes X) D_A]^{-1} [\Re D_A^*(Y W Y^* \otimes X) D_A] \succeq 0.
\end{align*}$$

(8.19)

The latter matrix inequality can also be written

$$\begin{align*}
\Re \left( \begin{array}{cc}
D_A & 0 \\
0 & D_A
\end{array} \right) \left( \begin{array}{cc}
Y A_\alpha^{-1}Y^* & Y W Y^* \\
Y W Y^* & Y W A_\alpha W Y^*
\end{array} \right) \otimes X \left( \begin{array}{cc}
D_A & 0 \\
0 & D_A
\end{array} \right) \succeq 0.
\end{align*}$$

(8.20)

Since the real part of a p.s.d. matrix is p.s.d. it is sufficient to show that

$$\begin{align*}
\left( \begin{array}{cc}
D_A & 0 \\
0 & D_A
\end{array} \right) \left( \begin{array}{cc}
Y A_\alpha^{-1}Y^* & Y W Y^* \\
Y W Y^* & Y W A_\alpha W Y^*
\end{array} \right) \succeq 0.
\end{align*}$$

(8.21)

Furthermore, note that the matrix $X$ is p.s.d. Since all eigenvalues of a kronecker product are products of the eigenvalues of the factors (see e.g. [HJ91], Theorem 4.2.12) it remains to show that

$$\begin{align*}
\left( \begin{array}{cc}
Y A_\alpha^{-1}Y^* & Y W Y^* \\
Y W Y^* & Y W A_\alpha W Y^*
\end{array} \right) &= \left( \begin{array}{cc}
Y & Y W A_\alpha \\
Y W A_\alpha & A_\alpha^{-1} \left( Y^* A_\alpha W Y^* \right)
\end{array} \right)
\succeq 0,
\end{align*}$$

(8.22)

which is easily seen to hold. \(\square\)
8.2. A COVARIANCE MATCHING APPROACH

Inserting (8.16) into (8.11) and using (7.15) leads to the interesting result
\[
\text{C}(\text{W}_{\text{new}}) = N\text{CRB}_{\theta},
\]  
(8.23)
where \( \text{CRB}_{\theta} \) is given by (7.14) or (7.16). The conclusion is that if the optimal weighting is used, whitened WSF is asymptotically efficient for the data model. This observation does in fact serve as an indirect alternative proof of Theorem 3.

It is interesting to note that the standard method can be made efficient by a small modification of the weighting matrix. The change of weighting does not complicate the implementation of the algorithm in any way (note that the new weighting matrix is diagonal). Algorithms such as MODE [SS90a] can be used with only minor modifications, avoiding a multidimensional search by exploiting certain array geometries.

8.2 A covariance matching approach

Although conceptually appealing, it is not obvious that explicit whitening of the data using an estimated whitening matrix is the best way to solve the estimation problem at hand. In this section, a covariance matching approach that is based on joint estimation of \( \mathbf{Q}, \mathbf{P} \) and \( \theta \) will be pursued. The underlying framework for covariance matching is presented more closely in [OSR98]. The proposed method is by construction asymptotically (in \( N \) and \( M \) in the same way as in the previous section) efficient. This follows from the extended invariance principle, see [SS90b] and [OSR98]. In order to claim this, it is necessary to assume that the signal covariance matrix \( \mathbf{P} \) is p.d. (see [SOVM06] for a discussion on this issue).

The criterion function to be minimized is (since \( \hat{\mathbf{R}}_1 \) and \( \hat{\mathbf{R}}_2 \) are uncorrelated)
\[
V_C(\theta, \mathbf{P}, \mathbf{Q}) = \text{vec}^* \{ \hat{\mathbf{R}}_1 - \mathbf{Q} \} \mathbf{W}_1 \text{vec} \{ \hat{\mathbf{R}}_1 - \mathbf{Q} \} \\
+ \text{vec}^* \{ \hat{\mathbf{R}}_2 - \mathbf{Q} - \mathbf{AP}^* \} \mathbf{W}_2 \text{vec} \{ \hat{\mathbf{R}}_2 - \mathbf{Q} - \mathbf{AP}^* \}.
\]  
(8.24)
The weighting matrices should be chosen as [SS90b]
\[
\mathbf{W}_1 = N^{-1} \text{Cov} \left[ \text{vec} \{ \hat{\mathbf{R}}_1 \} \right]^{-1} = \alpha^{-1} (\mathbf{Q}^{-1} \mathbf{Q}^{-1}),
\]
\[
\mathbf{W}_2 = N^{-1} \text{Cov} \left[ \text{vec} \{ \hat{\mathbf{R}}_2 \} \right]^{-1} = (\mathbf{R}^{-1} \mathbf{R}^{-1}).
\]  
(8.25)
The necessary (asymptotical) covariances are derived in [OSR98]. The matrices \( \mathbf{Q} \) and \( \mathbf{R} \) can be replaced by the consistent estimates \( \hat{\mathbf{R}}_1 \) and \( \hat{\mathbf{R}}_2 \), respectively, without affecting the asymptotical properties of the estimate. The criterion function (8.24) can be concentrated w.r.t. \( \mathbf{P} \) and \( \mathbf{Q} \). To that end,
\[
\hat{\mathbf{P}}(\theta, \mathbf{Q}) = \arg\min_{\mathbf{P}} V_C(\theta, \mathbf{P}, \mathbf{Q}) \\
= (\mathbf{A}^* \hat{\mathbf{R}}_2^{-1} \mathbf{A})^{-1} A^* \hat{\mathbf{R}}_2^{-1} (\hat{\mathbf{R}}_2 - \mathbf{Q}) \hat{\mathbf{R}}_2^{-1} \mathbf{A} (\mathbf{A}^* \hat{\mathbf{R}}_2^{-1} \mathbf{A})^{-1} \\
= (\hat{\mathbf{R}}_2^{-\frac{1}{2}} \mathbf{A})^\dagger \hat{\mathbf{R}}_2^{-\frac{1}{2}} (\hat{\mathbf{R}}_2 - \mathbf{Q}) \hat{\mathbf{R}}_2^{-\frac{1}{2}} (\hat{\mathbf{R}}_2^{-\frac{1}{2}} \mathbf{A})^\dagger. 
\]  
(8.26)
In order to concentrate the criterion w.r.t. $Q$, write
\[
V_C(\theta, \hat{P}, Q) = \alpha^{-1} \text{vec}^*(\tilde{R}_1 - Q)(\tilde{R}_1^{-T} \otimes \tilde{R}_1^{-1})\text{vec}(\tilde{R}_1 - Q)
\]
\[
+ \text{vec}(\tilde{R}_2 - Q)\hat{W}_3\text{vec}(\hat{R}_2 - Q),
\]
where
\[
\hat{W}_3 = (\tilde{R}_2^{-T} \otimes \tilde{R}_2^{-1} - \tilde{R}_2^{-\frac{3}{2}} \Pi_{\tilde{R}_2^{-A}}^{-\frac{1}{2}} \tilde{R}_2^{-\frac{3}{2}} \otimes \tilde{R}_2^{-\frac{3}{2}} \Pi_{\tilde{R}_2^{-A}}^{-\frac{1}{2}} \tilde{R}_2^{-\frac{3}{2}}).
\]

As (8.27) is quadratic in $Q$, its minimizer is easily found to be
\[
\text{vec}(\hat{Q}) = (\alpha^{-1}(\tilde{R}_1^{-T} \otimes \tilde{R}_1^{-1}) + \hat{W}_3)^{-1}(\alpha^{-1}\text{vec}(\tilde{R}_1) + \hat{W}_3\text{vec}(\hat{R}_2)).
\]

It can be proved that $\hat{P}$ and $\hat{Q}$ are hermitian by construction. Furthermore, due to the consistency, both $\hat{P}$ and $\hat{Q}$ are p.d. for large enough $N$ and $M$. These are in fact necessary conditions for optimality (efficiency) of the approach. Finally, insertion gives
\[
\hat{\theta}_C = \underset{\theta}{\text{argmin}} \ V_C(\theta),
\]
\[
V_C(\theta) = V_C(\theta, \hat{P}, \hat{Q}) + \alpha^{-1}\theta - \text{indep. term}
\]
\[
= -\text{vec}(\alpha^{-1}(\tilde{R}_1^{-1} + \tilde{R}_2^{-\frac{3}{2}} \Pi_{\tilde{R}_2^{-A}}^{-\frac{1}{2}} \tilde{R}_2^{-\frac{3}{2}}) + \hat{W}_3)
\]
\[
\times \text{vec}(\alpha^{-1}(\tilde{R}_1^{-1} + \tilde{R}_2^{-\frac{3}{2}} \Pi_{\tilde{R}_2^{-A}}^{-\frac{1}{2}} \tilde{R}_2^{-\frac{3}{2}}),
\]
since
\[
\text{vec}(\tilde{R}_2)\hat{W}_3\text{vec}(\hat{R}_2) = (m - d).
\]

While having the same large-sample performance as the optimally weighted whitened WSF method, the two methods can be expected to behave differently in small samples. In Chapter 9, several numerical studies are presented that indicate that the covariance matching approach performs better for small sample sizes. However, existing techniques for avoiding the multidimensional minimization of the criterion function can be used for the WSF method while no such methods are known for the covariance matching technique.

8.A Proof of Theorem 2

In the remainder of the section, the subscript will be dropped from $\hat{\theta}_W$ and $V_W$. It follows from the consistency of standard WSF [VO91] and the consistency relation
\[
\lim_{N \to \infty} \hat{Z}\hat{R}_W\hat{Z} = \text{ZAP}_A\text{Z} + I
\]
that
\[
\lim_{N \to \infty} \dot{\theta} = \theta_0. \tag{8.33}
\]
Both results holds w.p.1. In order to see this, consider the fact that \(M\) and \(N\) has a fixed ratio asymptotically. Next, write
\[
V'(\dot{\theta}) = 0 \simeq V'(\theta_0) + V''(\theta_0)(\dot{\theta} - \theta_0), \tag{8.34}
\]
where \(\simeq\) denotes equality in the dominating terms (as \(M, N\) grow). It follows that, for large \(N\)
\[
(\dot{\theta} - \theta_0) \simeq -H^{-1}V'(\theta_0), \tag{8.35}
\]
where
\[
H = \lim_{N \to \infty} V''(\dot{\theta}). \tag{8.36}
\]
It is a well known and easily verifiable result that
\[
\{\Pi_A\}' = \Pi_A^* A_i A^* + \ldots, \quad \{\Pi_A\}'' = -\Pi_A^* A_j A_i A^* A_i A^* A_j A^* + \Pi_A^* A_j A^* A_i A^* A_j A^* \Pi_A^* - \Pi_A^* A_j A_i A^* A_j A^* + \ldots, \tag{8.37}
\]
where, with some abuse of notation, \(\{\Pi_A\}'\) and \(\{\Pi_A\}''\) denote the element-wise first and second derivatives of \(\Pi_A\) w.r.t. the elements of \(\theta\). With \(A\) replaced by \(A = ZA\), it is a direct consequence of the above expression and the relation \(\Pi_A ZA E_s = 0\) that \([H]_{ij}\) is given by (8.11). The \(i\)th element of the gradient, evaluated at \(\theta_0\), can be written
\[
V_i = -2\text{Re} \text{ tr} ([\Pi_{ZA}^*][\dot{Z}A (\dot{Z}A)^*][\dot{E}, W \dot{E}]^*]. \tag{8.38}
\]
In order to show that
\[
\sqrt{V'(\theta_0)} \simeq \text{AsN}(0, G), \tag{8.39}
\]
with \(G\) given by (8.11), proceed by finding a first-order expression for \(V_i\). Introduce
\[
\dot{R} = \dot{R}_2 = R + \ddot{R} \tag{8.40}
\]
and analogous definitions for \(\dot{R}_1, \dot{Z}\) and \(\ddot{E}\). Consider
\[
\Pi_{ZA} = I - (Z + \ddot{Z})A(A^* ZZA + A^* \ddot{Z}ZA)
+ A^* \ddot{Z}ZA + A^* ZZA)^{-1} A^* (Z + \ddot{Z})
\simeq \Pi_{ZA} - \ddot{Z}A(A^* ZZA)^{-1} A^* Z
- ZA(A^* ZZA)^{-1} A^* \ddot{Z} + ZA(A^* ZZA)^{-1}
\times A^* (Z \ddot{Z} + \ddot{Z}Z)A(A^* ZZA)^{-1} A^* Z
\]
\[
= \Pi_{ZA} - \Pi_{ZA} \ddot{Z} Z^{-1} \Pi_{ZA} - \Pi_{ZA} Z^{-1} \ddot{Z} \Pi_{ZA}. \tag{8.41}
\]
The matrix inversion lemma was used in the first step.

It will be useful to have a first order expression for \( \tilde{Z} \). Write

\[
(Z + \tilde{Z})(Q + \tilde{R}_1)(Z + \tilde{Z}) = I.
\]  

(8.42)

This gives

\[
\tilde{Z}Q + Z\tilde{Z} \approx -Z\tilde{R}_1 Z.
\]

(8.43)

Vectorizing both sides gives

\[
(Z^T Q^T \otimes I + I \otimes ZQ) \text{vec}(\tilde{Z}) \approx -(Z^T \otimes Z) \text{vec}(\tilde{R}_1)
\]

(8.44)

and, since \( ZQ = Z^{-1} \),

\[
\text{vec}(\tilde{Z}) \approx -(Z^{-T} \otimes I + I \otimes Z^{-1})^{-1} (Z^T \otimes Z) \text{vec}(\tilde{R}_1).
\]

(8.45)

Note that \( \tilde{Z} \) will be of the same order of magnitude as \( \tilde{R}_1 \) and \( \tilde{R} \) (since \( M \) and \( N \) have a constant ratio asymptotically). Next, by definition

\[
\tilde{Z}\tilde{R}Z\tilde{E}_s = \tilde{E}_s\tilde{A}_s.
\]

(8.46)

After some manipulations this gives

\[
\tilde{E}_sA_s + E_s\tilde{A}_s \cong \tilde{Z}\tilde{R}ZE_s + \tilde{Z}\tilde{R}\tilde{Z}E_s + \tilde{Z}\tilde{R}\tilde{E}_s + \tilde{Z}\tilde{R}\tilde{E}_s.
\]

(8.47)

By expanding \( ZRZ \) and by multiplying with \( \Pi_{ZA}^\perp \) from the left it follows that

\[
\Pi_{ZA}^\perp \tilde{E}_sA_s \cong \Pi_{ZA}^\perp (\tilde{Z}\tilde{Z}^{-1}(E_sA_sE_s^* + E_sE_n^*)E_s + \tilde{Z}\tilde{R}ZE_s + Z^{-1}\tilde{Z}E_s + \tilde{E}_s).
\]

(8.48)

In the last expression it was used that \( \Pi_{ZA}^\perp ZRZ = \Pi_{ZA}^\perp \). Now

\[
\Pi_{ZA}^\perp \tilde{E}_s(A_s - I) \approx \Pi_{ZA}^\perp (\tilde{Z}\tilde{Z}^{-1}E_sA_s + \tilde{Z}\tilde{R}ZE_s + Z^{-1}\tilde{Z}E_s).
\]

(8.49)

Note that \( \tilde{E}_s \) is of the same order as \( \tilde{Z} \) and \( \tilde{R} \). From (8.38) and (8.41) it now follows that

\[
V_i \approx -2\text{Re} \{ \text{tr}(\Pi_{ZA}^\perp [\tilde{Z}A_s(\tilde{Z}A)^\dagger]|E_sW^*_s)\} \\
- \text{tr}(\Pi_{ZA}^\perp \tilde{Z}\tilde{Z}^{-1}Z\Pi_{ZA}^\perp |ZA_s(ZA)^\dagger|E_sW^*_s) \\
- \text{tr}(\Pi_{ZA}^\perp Z^{-1}\tilde{Z}\Pi_{ZA}^\perp |ZA_s(ZA)^\dagger|E_sW^*_s) \\
+ \text{tr}(\Pi_{ZA}^\perp Z\Pi_{ZA}^\perp |ZA_s(ZA)^\dagger(E_sW\tilde{E}_s^* + \tilde{E}_sW^*_s))\}
= -2\text{Re} \left(-\text{tr}(\Pi_{ZA}^\perp Z^{-1}\tilde{Z}\Pi_{ZA}^\perp |Z\Pi_{ZA}^\perp |E_sW^*_s) \\
+ \text{tr}(\Pi_{ZA}^\perp Z\Pi_{ZA}^\perp |ZA_s(ZA)^\daggerE_sW^*_s)\right),
\]

(8.50)
8.A. PROOF OF THEOREM 2

where, again, the relation $E_s^*\Pi_{\text{ZA}}^\perp = 0$ was used. At this point it can be noted that replacing $W$ with the consistent estimate

$$\hat{W} = W + \tilde{W}$$

(8.51)

will only introduce higher order terms in the expression. This proves that the covariance of the estimate will asymptotically not be affected by the use of a consistent estimate of the weighting matrix. Upon returning to the main proof, insertion of (8.49) into (8.50) gives

$$V_i \approx 2\text{Re} \left\{ \text{tr} \{ E_s^*Z^{-1}\tilde{Z}\Pi_{\text{ZA}}^\perp ZA_s(ZA)^\dagger E_sW \} \right\} - \text{tr} \{ (A_s - I)^{-1}A_sE_s^*Z^{-1}\tilde{Z}\Pi_{\text{ZA}}^\perp ZA_s(ZA)^\dagger E_sW \} - \text{tr} \{ (A_s - I)^{-1}E_s^*\tilde{Z}Z\Pi_{\text{ZA}}^\perp ZA_s(ZA)^\dagger E_sW \} - \text{tr} \{ (A_s - I)^{-1}E_s^*\tilde{Z}Z^\perp \Pi_{\text{ZA}}^\perp ZA_s(ZA)^\dagger E_sW \} \right\}. $$

(8.52)

Making use of the equality

$$I - (A_s - I)^{-1}A_s = -(A_s - I)^{-1}$$

(8.53)

leads to

$$V_i \approx -2\text{Re} \left\{ \text{tr} \{ (Z^{-1}\tilde{Z} + \tilde{Z}Z^{-1} + Z\tilde{Z})M_i \} \right\} = \Pi_{\text{ZA}}^\perp ZA_s(ZA)^\dagger E_sW(A_s - I)^{-1}E_s^*. $$

(8.54)

Using the formula $\text{tr} \{ AB \} = \text{vec}^T \{ A^T \} \text{vec} \{ B \}$ this can be written

$$V_i \approx -2\text{Re} \left\{ \text{vec}^T \{ (M_iZ^{-1} + Z^{-1}M_i)^T \} \text{vec} \{ \tilde{Z} \} + \text{vec}^T \{ (ZM_iZ)^T \} \text{vec} \{ \tilde{R} \} \right\}. $$

(8.55)

The (asymptotical) covariance expressions

$$\text{Cov} \left[ \tilde{R} \right] = N^{-1}(R^T \otimes R),$$

$$\text{Cov} \left[ \tilde{R}_1 \right] = M^{-1}(Q^T \otimes Q)$$

(8.56)

will once again be useful. It now follows from (8.45) that

$$\text{Cov} \left[ \tilde{Z} \right] = M^{-1}(Z^{-T} \otimes I + I \otimes Z^{-1})^{-2}. $$

(8.57)

Before finding an expression for $G_{ij}$, note that

$$\text{vec} \{ (M_iZ^{-1} + Z^{-1}M_i)^T \} = (I \otimes Z^{-T} + Z^{-1} \otimes I) \text{vec} \{ M_i^T \},$$

and

$$\text{vec} \{ (ZM_iZ)^T \} = (Z \otimes Z^T) \text{vec} \{ M_i^T \}. $$

(8.58)
Combining (8.56), (8.55), (8.57) and (8.58) with the uncorrelatedness of \( \tilde{\mathbf{R}} \) and \( \tilde{\mathbf{Z}} \) now gives

\[
G_{ij} = \text{NE} [V_i V_j] = \\
2\text{Re} \left\{ \text{vec}^T \left( \mathbf{M}_i^T \right) (\mathbf{Z}\mathbf{R}\mathbf{Z})^T \otimes \mathbf{Z}\mathbf{R}\mathbf{Z} \text{vec}(\mathbf{M}_j^*) \right\} + 2\text{Re} \left\{ \frac{N}{M} \text{vec}^T \left( \mathbf{M}_i^T \right) \text{vec}(\mathbf{M}_j^*) \right\} \\
= 2\text{Re} \text{tr}(\mathbf{Z}\mathbf{R}\mathbf{M}_i \mathbf{Z}\mathbf{R}\mathbf{M}_j^* + \frac{N}{M} \mathbf{M}_i \mathbf{M}_j^*) \\
= 2\text{Re} \text{tr}(\mathbf{M}_i \mathbf{Z}\mathbf{R}\mathbf{M}_j^* + \frac{N}{M} \mathbf{M}_i \mathbf{M}_j^*),
\]

(8.59)

where the relation \( \mathbf{Z}\mathbf{R}\mathbf{Z}\mathbf{I}_{\mathbf{Z}\mathbf{A}}^\perp = \mathbf{I}_{\mathbf{Z}\mathbf{A}} \) was used in the last step. Continue by writing

\[
G_{ij} = 2\text{Re} \text{tr}(\mathbf{M}_i (\mathbf{E}_s \mathbf{A}_s \mathbf{E}_s^* + \mathbf{E}_n \mathbf{E}_n^* + \frac{N}{M} \mathbf{I}) \mathbf{M}_j^*) \\
= 2\text{Re} \text{tr}(\mathbf{M}_i (\mathbf{E}_s (\mathbf{A}_s + \frac{N}{M} \mathbf{I}) \mathbf{E}_s^*) \mathbf{M}_j^*),
\]

(8.60)

where it was used that \( \mathbf{E}_s^* \mathbf{E}_s = 0 \). Insertion of (8.54) gives (8.11). The asymptotic normality of \( \hat{\theta} \) follows from the asymptotic normality of \( V_i \) which is a consequence of the central limit theorem.
Chapter 9

Simulations

A number of computer simulations have been performed in order to illustrate the theoretical, asymptotical results derived in this work and to investigate small sample performance of the methods.

9.1 Simulation setup

The asymptotical covariance expression for whitened WSF as introduced in Theorem 2 (found in Chapter 8) is included in all plots. Two weighting matrices (W in (8.9)) were tested. The asymptotical covariance with $W_{new} = (\Lambda_s - I)^2(\Lambda_s + \alpha I)^{-1}$, which is the optimal weighting for the problem, is shown as a solid line. The asymptotical covariance with $W_{WSF} = (\Lambda_s - I)^2\Lambda_s^{-1}$, which would be the optimal weighting if $Q$ (and thereby $Z$) was known exactly, is shown as a dotted line.

In addition to the asymptotical covariances, three versions of the CRB are included. The CRB for the estimation problem treated in this part of the thesis coincides with the asymptotical covariance for optimally weighted WSF (solid line). The CRB for the estimation problem if $Q$ is known exactly, (7.4), is also included. It is shown as a dashed line in the plots. In one case the CRB with $N \gg M$ is included. It is shown as a dash-dotted line.

In addition to the theoretical quantities, a number of estimation methods were used and their performances were evaluated using Monte Carlo methods. Let $\hat{\theta}_k^i$ be the estimate of the $k$th DOA in Monte Carlo trial $i$ (using the method in question) and $\theta_k$ be the true DOA. Then the performance measure used is the RMS per source

$$\sqrt{\frac{1}{d} \sum_{i=1}^L \sum_{k=1}^d (\hat{\theta}_k^i - \theta_k)^2},$$

where $L$ is the number of Monte Carlo trials. In the simulations presented here, $L = 1000$. 

75
Two versions of whitened WSF were tested. The results for the version based on the optimal $W_{\text{new}}$ weighting are marked with stars in the plots. The squares are for the version that is based on $W_{\text{WSF}}$.

The covariance matching approach of Section 8.2 was also tested. The results using this method are shown as circles in the plots.

Both methods require a $n_d$-dimensional search for a minimum value of a loss function. Unlike in any practical case, the true value was actually known in the simulations, and it was used as an initial point in a crude search. To keep things simple, a Nelder-Mead simplex method was used to find a local minimum close to the true value. The implementation used was the one provided in the routine fminsearch with the software package matlab (version 7.0.1.24704).

In addition to the methods discussed in the present work, the well-known root-MUSIC method was also included in one of the plots (see Figure 9.4). It was implemented using whitening in the same way as the WSF method above. The results for the root-MUSIC method are shown as hollow stars.

In all simulations, the signal covariance matrix $P$ was of the form

$$P = \begin{pmatrix} 1 & \rho \\ \rho^* & 1 \end{pmatrix}, \quad 0 \leq |\rho| < 1. \quad (9.2)$$

This makes it possible to use the parameter $\rho$ to control the correlation of the signals. In the plots, $\rho$ is real. The noise covariance matrix $Q$ was designed to be reminiscent of a spatial AR(1)-process. The $i,j$th element of $Q$ is

$$[Q]_{ij} = \sigma^2 a^{|i-j|}, \quad (9.3)$$

where $a$ is a parameter used to control the noise covariance. The array steering vectors correspond to those of a uniform linear array (ULA) with sensors separated by a half wavelength. The DOAs of the sources are measured from array broadside.

9.2 Results

Results for a case with sources at 0 and 10 degrees and five sensors are shown in Figure 9.1. The two sample sizes $M$ and $N$ have a fixed ratio of five in the experiment, and they are both varied. The noise was spatially correlated with parameters $\sigma^2 = 0.5$ and the signals were correlated with $\rho = 0.7$. It can be seen from the plot that the suggested optimal weighting gives a small performance gain compared to the standard choice. This is evident both in the asymptotical expressions and in the experimental evaluations of the different methods. It can also be noted that the covariance matching approach performs better than optimally weighted WSF for small $M$ and $N$. Asymptotically, this difference disappears.

In Figure 9.2, the source spacing is made smaller and when comparing to Figure 9.1 it is evident that this affects performance of all methods. The gain obtained by weighting with $W_{\text{new}}$ instead of $W_{\text{WSF}}$ in WSF is also slightly larger for this case.
9.2. RESULTS

In Figure 9.3, the number of sensors is doubled. The performance gain for the covariance matching approach compared to the two WSF methods seems to increase.

In the next plot, Figure 9.4, the number of samples were kept constant at $N = 5000$ and $M = 1000$. Instead, the signal correlation, $\rho$, was varied. As could be suspected, the gain obtained by using $W_{\text{new}}$ instead of $W_{\text{WSF}}$ depends on $\rho$. This can be explained by the fact that $\Lambda_\rho$ will have fewer dominating eigenvalues when the signals are correlated, making $W_{\text{new}}$ differ more from $W_{\text{WSF}}$. In the experiment, $\rho$ was real. The phase of $\rho$ is however not unimportant. Further investigations also indicate that the source spacing has a similar significance.

It is quite clear that the advantage of the improved methods developed here will be larger when $M$ is small compared to $N$. This is illustrated in Figure 9.5. For the particular case studied there it can be seen that the effects of small $M$ (poor whitening matrix) will dominate until $M > \frac{N}{\sqrt{3}}$. Note that the new methods never will perform worse than the standard methods. The plot (and the other plots in the study) also indicates that the asymptotical expressions are valid for fairly small sample set sizes.

In Figure 9.6, the impact of the noise correlation is investigated. In terms of difference between the tested methods, the plot indicates that the noise correlation is of little importance. It is somewhat surprising to note that increasing noise correlation gives an improved performance. An explanation for this is that the noise is then partially filtered-out by the whitening transformation. Note that a performance gain can be obtained with the new methods also when the true noise covariance is identity, as is the case in Figure 9.4.
Figure 9.1: RMS per source in estimation of direction of arrival with two sources at 0 and 10 degrees. A ULA with \( m = 5 \) elements and half-wavelength spacing was used. The sources were correlated with \( \rho = 0.7 \). The noise covariance matrix had parameters \( \alpha = 0.5 \) and \( \sigma^2 = 0.5 \). The number of signal-free samples, \( M \), was varied and the number of signal-containing samples \( N \) was \( 5M \).
Figure 9.2: RMS per source in estimation of direction of arrival with two sources at 0 and 6 degrees. A ULA with $m = 5$ elements and half-wavelength spacing was used. The sources were correlated with $\rho = 0.7$. The noise covariance matrix had parameters $a = 0.5$ and $\sigma^2 = 0.5$. The number of signal-free samples, $M$, was varied and the number of signal-containing samples $N$ was $5M$. 
Figure 9.3: RMS per source in estimation of direction of arrival with two sources at 0 and 6 degrees. A ULA with $m = 10$ elements and half-wavelength spacing was used. The sources were correlated with $\rho = 0.7$. The noise covariance matrix had parameters $a = 0.5$ and $\sigma^2 = 0.5$. The number of signal-free samples, $M$, was varied and the number of signal-containing samples $N$ was $5M$. 
9.2. RESULTS

Figure 9.4: RMS per source in estimation of direction of arrival with two sources at 0 and 10 degrees. A ULA with $m = 5$ elements and half-wavelength spacing was used. The sources were correlated with $\rho$ varying from 0.3 to 0.99. The noise covariance matrix had parameters $a = 0$ and $\sigma^2 = 0.5$. The sample sizes were fixed at $M = 1000$, $N = 5000$. 
Figure 9.5: RMS per source in estimation of direction of arrival with two sources at 0 and 10 degrees. A ULA with \( m = 5 \) elements and half-wavelength spacing was used. The sources were correlated with \( \rho = 0.7 \). The noise covariance matrix had parameters \( a = 0.5 \) and \( \sigma^2 = 0.5 \). The number of signal-free samples varied from 10 to 1980 with \( N = 2000 - M \).
Figure 9.6: RMS per source in estimation of direction of arrival with two sources at $-8$ and $10$ degrees. A ULA with $m = 5$ elements and half-wavelength spacing was used. The sources were correlated with $\rho = 0.7$. The noise covariance matrix had parameters $\sigma$ varying from 0 to 0.99 and $\sigma^2 = 0.5$. The sample sizes were $M = 200$ and $N = 800$. 
Chapter 10

Conclusions

Two parameter estimation problems have been studied in this thesis. Estimation methods are suggested for each problem, and the asymptotical performances are investigated. In order to put performance into perspective, the Cramér-Rao lower bound has been derived for each data model. Computer simulations are used to investigate performance for small samples. Comparisons, numerical and analytical, are made to other estimators in the literature.

10.1 Part I- The reduced rank linear regression

A method for parameter estimation in the reduced rank linear regression is derived and analyzed. The CRB for the covariance of any unbiased estimator for the problem is also derived.

The new estimator relies upon instrumental variable principles. The first step in the estimation procedure is to form a set of statistics for the problem. In the second step these statistics are used to produce the estimate using a weighted low rank approximation. This second step is asymptotically optimal. The overall optimality of the approach depends only on the statistics calculated in the first step, and in particular on two parameters denoted \( p \) and \( f \). Some relations to previously known methods are also established, notably to the ML method derived for the white noise case.

The new method relies on solving a weighted low rank approximation problem (WLRA). As the WLRA has no closed form solution, two approximative but asymptotically optimal solutions are derived. The new algorithms are non-iterative and also less complex than previous methods. The WLRA problem appears in several technical areas, and the new methods are thus of interest in their own right.

In addition, a rank detection algorithm is devised, based on the statistical distribution of the minimum value of the criterion function for the parameter estimation algorithm.
Computer simulations support the theoretical findings and show that the new algorithms are useful also on relatively small sample set sizes. The simulations and analysis indicate that the new method has significantly better performance compared to previously known methods when the noise is temporally correlated, and equivalent performance for the white noise case.

10.2 Part II- DOA estimation in unknown colored noise fields

The problem of estimating parameters of narrowband signals impinging on an array of sensors has been studied. It was assumed that the sensor array output is the superposition of the response to the impinging signals and a spatially correlated noise (with an unknown correlation matrix). In addition to the samples containing both signals and noise, a batch of samples containing only noise was assumed to be available. One of the results presented in this part of the thesis was the CRB for the estimation problem. It was found to have a relatively simple structure.

A standard approach is to use the noise-only samples to estimate a transformation that is used to whiten the signal containing samples. Here, this procedure was used together with WSF. The asymptotical (in the sample sizes) covariance of the estimation error for the resulting estimate was derived for a general weighting matrix. It was further shown that a particular choice of weighting matrix minimizes the asymptotical covariance and, in fact, makes it coincide with the CRB. The optimal weighting matrix is diagonal and the new formulation does not increase the computational complexity compared to the standard choice.

A covariance matching approach to the estimation problem was also introduced, avoiding explicit whitening of the data. The method is statistically efficient. Computer simulations indicated that the asymptotical results are useful also for finite sample sizes, and that the new, optimal weighting for WSF gives a performance improvement. In the simulations, the covariance matching method had the best small sample behavior of all tested methods.

10.3 Future work

There are a few natural extensions to the work presented in this thesis

- Testing the proposed algorithm for the reduced rank linear regression problem on practical problems. An interesting example would be state-space system identification.
- In the case of temporally white noise, the reduced rank linear regression method proposed is optimal if \( p = f = 0 \). Is it possible to select the parameters \( p \) and \( f \) in an optimal way for other noise models as well? This would be of interest since the computational complexity of the method grows with \( p \) and \( f \).
10.3. FUTURE WORK

- Investigate ways to improve the complexity of the method.

- Depending on the noise correlation, the proposed method has a performance quite far from the CRB. Can the accuracy be improved? Perhaps a different choice of the instruments $z(t)$ would improve performance?

- In the simulations presented for the DOA estimation algorithm, the optimization problems that are parts of both the WSF and the covariance matching approach were solved very inefficiently using a crude search around the true value (which, of course, is unknown in practice). Better implementations are of course necessary.
  
  - For uniform linear arrays, it should be straightforward to implement MODE (see e.g. [SS90b]) with the new weighting, $W_{new}$.
  
  - The covariance matching method can be implemented much more efficiently (for example using a Newton-type algorithm)?

- Investigate how realistic the assumptions on the size of the available signal-free sample set are. It would, of course, be very useful to evaluate the proposed method with a more specific application in mind.

- If the noise covariance $Q$ is assumed to possess a known structure, how does that affect the performance bound. Also, can the algorithms be modified to take advantage of that structure?
Bibliography


BIBLIOGRAPHY


