INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600
On optimal algorithms for parameter set estimation

Cheung, Man-Fung, Ph.D.
The Ohio State University, 1991

Copyright ©1991 by Cheung, Man-Fung. All rights reserved.
ON OPTIMAL ALGORITHMS FOR PARAMETER SET ESTIMATION

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

by

Man-Fung Cheung, B.Sc., M.S.E.E.

The Ohio State University

1991

Dissertation Committee:
Professor Stephen Yurkovich
Professor Ümit Özgüner
Professor Kevin M. Passino

Approved by

Stephen Yurkovich
 Adviser
Department of Electrical Engineering
© Copyright by
Man-Fung Cheung
1991
To my dearest parents,

Chiu-Kwan Cheung and Hung Cheung
I am grateful to my advisor Professor Stephen Yurkovich for his guidance, sup­port and patience throughout my graduate studies. I have enjoyed the experience of working with him who has given me lots of opportunities to learn and grow. I would like to thank Professor Kevin Passino for his constructive criticism and valuable technical opinions. Thanks also to my dissertation committee members for their helpful suggestions, especially to Professor Ümit Özgüner for being my dissertation reader while he is on sabbatical leave.

In addition, I thank my fellow students in the Control Research Laboratory for their friendship and creative discussions. Special thanks go to Hubert Chan, Wilbur Dale, Jeff Donne, Dave Scheonwald, Ho-Cheung Wong, for their help during the preparation of this document.

I thank my family for their strong support and encouragement during my schooling thousands miles away from them. Without their endless love and patience, this work would not have been possible.

This work was supported in part by AF Wright Laboratories and in part by the Electrical Engineering Department of The Ohio State University.
VITA

January 22, 1964 ......................... Born - Hong Kong

1986 ................................. BSc. in E.E., University of Salford, Salford, England

1988 ................................. M.S. in E.E., The Ohio State University, Columbus, Ohio

1987-89 ............................... Graduate Teaching and Research Associate, Electrical Engineering Department, The Ohio State University, Columbus, Ohio

1990 - Present ........................ Graduate Teaching Associate, Electrical Engineering Department, The Ohio State University, Columbus, Ohio
PUBLICATIONS


FIELDS OF STUDIES

Major Field: Electrical Engineering .................. Professor S. Yurkovich
Systems and Control .................................. Professor R. E. Fenton

Minor Fields: Digital Systems ....................... Professor C. A. Klein
Mathematics ........................................... Professor U. H. Gerlach
TABLE OF CONTENTS

DEDICATION ................................................................. ii
ACKNOWLEDGEMENTS ........................................................ iii
VITA ...................................................................................... iv
LIST OF FIGURES ............................................................... viii
LIST OF TABLES ................................................................. x

CHAPTER PAGE

I INTRODUCTION ............................................................... 1

1.1 Overview of System Identification .................................... 2
1.2 Overview of Large Scale Systems ................................... 5
1.3 Motivation and Organization of this Thesis ..................... 7

II OPTIMAL PARAMETER SET ESTIMATION OF AN ARX MODEL
WITH BOUNDED NOISE ....................................................... 10-

2.1 Introduction ............................................................... 10
2.2 Problem Statement ..................................................... 14
2.3 The New Optimal Algorithm for Parameter Set Estimation . 15
2.4 Comparison with the OBE Algorithm ............................ 35
2.5 Computer Simulation .................................................. 38
2.5.1 Example 1 ............................................................. 38
2.5.2 Example 2 ............................................................. 39
2.6 Conclusion ................................................................. 45
III CONVERGENCE ANALYSIS OF THE OVE ALGORITHM

3.1 Introduction ..................................................................................... 47
3.2 Summary of the OVE Algorithm ..................................................... 49
3.3 The Main Convergence Properties of the OVE Algorithm .............. 51
3.4 Other Convergence Issues................................................................. 60
   3.4.1 The Feasible Parameter Set .................................................. 60
   3.4.2 The Optimal Bounding Set .................................................. 60
   3.4.3 Computational Issues ............................................................ 63
   3.4.4 The Noise Free Case ($\gamma = 0$) ........................................ 64
3.5 Conclusions . ................................................................................ 66

IV PARAMETER SET ESTIMATION OF AN ARMAX MODEL WITH
BOUNDING NOISE ............................................................................. 68

4.1 Introduction ...................................................................................... 68
4.2 Problem Definition ............................................................................. 72
4.3 Main Result: Two-Step Solution ................................................... 73
   4.3.1 Step 1: Estimating the Bound on $A(q^{-1}, \theta)$ ................. 73
   4.3.2 Step 2: Application of the OVE Algorithm ....................... 78
4.4 Computer Simulations ....................................................................... 80
   4.4.1 A Case study on Some Experiment Design Issues for the Two-
        Step Solution ........................................................................... 80
   4.4.2 Application of the OVE Algorithm ....................................... 84
4.5 Conclusions ......................................................................................... 85

V PARAMETER SET ESTIMATION IN INTERCONNECTED DYNAMIC
SYSTEMS ............................................................................................ 87

5.1 Introduction ...................................................................................... 87
5.2 Problem Definition ............................................................................. 91
5.3 Computation of the Instantaneous Bound for Subsystem Interactions 94
5.4 Simulation Example ........................................................................ 96
5.5 Conclusions ......................................................................................... 100

VI CONCLUSIONS AND FUTURE RESEARCH ..................................... 102

BIBLIOGRAPHY ............................................................................... 106
# List of Figures

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Schematic flow chart of system identification</td>
</tr>
<tr>
<td>2</td>
<td>Illustration of Khachian's bounding ellipsoid</td>
</tr>
<tr>
<td>3</td>
<td>Ellipsoid bounding the intersection of a hypersphere and two parallel hyperplanes</td>
</tr>
<tr>
<td>4</td>
<td>An illustration for $\alpha = \beta$</td>
</tr>
<tr>
<td>5</td>
<td>An illustration of the case $\sigma &lt; 0$ in 2-D</td>
</tr>
<tr>
<td>6</td>
<td>The geometry of ellipses at $k = 0, \cdots, 5$</td>
</tr>
<tr>
<td>7</td>
<td>The geometry of the ellipses at $k = 10, 15, 20, 30, 40, 50$</td>
</tr>
<tr>
<td>8</td>
<td>Volume of $E_k^{OVE}$, $E_k^{OBE}$ and $E_k^{MOBE}$</td>
</tr>
<tr>
<td>9</td>
<td>Parameter intervals for both algorithms</td>
</tr>
<tr>
<td>10</td>
<td>An illustration of $\theta_M$ being outside of $\mathcal{F}_{M+1}$</td>
</tr>
<tr>
<td>11</td>
<td>Behavior of $\det P_k$</td>
</tr>
<tr>
<td>12</td>
<td>Behavior of $(1 - \sigma_k) e_k$</td>
</tr>
<tr>
<td>13</td>
<td>Behavior of $e_k$</td>
</tr>
<tr>
<td>14</td>
<td>Behavior of $|\tilde{\theta}_k|_2$</td>
</tr>
<tr>
<td>15</td>
<td>An example of $\mathcal{F}^k$ converged to a point</td>
</tr>
<tr>
<td>16</td>
<td>An illustration of the conservatism of $E_k$ relative to $\mathcal{F}^k$</td>
</tr>
<tr>
<td>17</td>
<td>Non-convex parameter bound for a first order ARMAX model with bounded noise</td>
</tr>
<tr>
<td>18</td>
<td>Modification of a non-convex parameter set to a set bounded by two parallel hyperplanes</td>
</tr>
<tr>
<td>19</td>
<td>Bode gain plots of the 1st local and interconnected subsystems</td>
</tr>
<tr>
<td>20</td>
<td>The output signal and the noise sequences for the example</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Effect of $T_s$ on $g(H)$ ................................................. 81
2. Effect of $m_0$ on $g(H)$ .................................................. 82
3. Effect of $m_0$ on $\eta$ ...................................................... 82
4. Effect of $i$ on $g(H)$ ...................................................... 83
5. Effect of $i$ on $\eta$ ...................................................... 83
6. Estimation result of the example using Step 1 scheme ........ 84
7. Estimation result of the example using Step 2 scheme ........ 85
8. The estimates and the bounds on denominator coefficients for the first decoupled subsystem ..................................... 99
9. The parameter interval on denominator coefficients for the first decoupled subsystem ..................................... 99
10. The center estimate and the bounds on the first decoupled subsystem ......................................................... 100
CHAPTER I

INTRODUCTION

Since the first digital computer faced the world almost four decades ago, technology has been advancing at an amazing pace. The phenomenal increase that has taken place in the sophistication of integrated circuit technology, micro-electronics and communication technology has allowed us to use desk-top digital computers that two decades ago were just fictional. Building on the power of these emerging technologies, system and control researchers have created new ideas and applications in studying control systems.

One of the most important element in a control design problem is the modeling of a plant to be controlled. To control a physical plant is to design a feasible controller or an admissible control action, based on the available model of the plant, to accomplish the control objective. Normally, the \textit{a priori} knowledge of the plant for control is a mathematical model which can be obtained through derived physical laws or experimentation. However, uncertainties in an analytical model are usually unavoidable, arising from simplified models of natural laws, neglected dynamics and the desire to get a computationally tractable model. Model identification through experimentation seems more practical and realistic as idealized physical laws are avoided. Moreover, the aging and environmental effects which are difficult to model can also be accounted
for through experimentation.

1.1 Overview of System Identification

System identification is a branch of mathematical modeling of systems using experimental data. The model of a system can be used in a variety of engineering areas. In a control design problem, the system model is often used for compensator designs, signal prediction or simulation. Models can also be used in areas like signal processing for fault detection, pattern recognition, and adaptive filtering. In other fields such as pharmacy, environmental science, biology and econometrics, models can be used to increase scientific knowledge.

In the area of system identification, modeling methods can be generally classified into two distinct ways: the non-parametric and the parametric methods. Figure 1 shows a general picture of a system identification procedure. Once a priori assumptions are made, experimental data must be generated. In doing so, experiment conditions must be analyzed or chosen to make the experiment data as informative as possible. This step is usually called the experiment design. Upon collecting the measurement data, decisions have to be made to choose the appropriate model structure and identification method that can best identify the system. Typical parametric models for linear systems are ARX (Auto-Regressive with Exogenous input), ARMAX (Auto-Regressive-Moving-Average with Exogenous input) or state-space structures. There are various ways to identify system parameters, such as LS (Least Squares), ML (Maximum Likelihood), IV (Instrumental Variables) and so on. In the literature, some extensions or modifications to those classical methods are available. The identi-
fication procedure can be conducted in an off-line, on-line, non-recursive or recursive manner. In the last step of the identification procedure, regardless of what model structure, modeling method or approach one has used, the identified model must be verified against another set(s) of experimental data to qualify the accuracy of the estimate; this can be done by comparing the simulated data set with the actual data set. A detail discussion on different system identification issues can be found in [1, 2].

The ultimate goal of system identification is to reduce or remove the uncertainties in a mathematical model that represents the system in interest. Most available results for system identification have focused on linear systems with certain statistical assumptions. However, all real-world systems are non-linear by nature. It is inevitable that some practical uncertainties are inherent in the conventional stochastic identification methods. This problem was first recognized back in the 1960's by Schweppes [3] in state estimation. He proposed to use an ellipsoid to characterize the uncertainties of system states deterministically when the measurement noises are bounded; this type of problem is referred to as membership set estimation. This idea was further extended to parameter identification by Fogel and Haung[4, 5], and the jargon of parameter set estimation was used. There they considered the parameter set estimation of an ARX model with bounded noise. In the bounding ellipsoid algorithm developed in [5], ellipsoidal sets are recursively found to bound the feasible parameter set which is consistent with the measurements. Such a modeling algorithm is widely applied in adaptive signal processing. The set theoretic concept in
Figure 1: Schematic flow chart of system identification
the bounding algorithm was later interpreted as a quantification of model parameter uncertainties, the modeling information required in both adaptive and non-adaptive robust control design developed in the last few years[6, 7]. This latter deterministic approach is further investigated in this dissertation as being motivated by the recent advances in robust control theory.

1.2 Overview of Large Scale Systems

The availability of cheap and high-speed micro-computers and the advent of system engineering has made the utilization of computers in industrial control attractive and realistic. In particular, the use of multiple computers to control a large scale industrial process has received a lot attention[8, 9, 10]. One of the important problems of control theory is the analysis, design and control of large scale systems under decentralized information constraints. There are numerous control strategies introduced in the literature that handle the system size and the information constraints[11, 12, 13, 14, 10, 15]. Reasons for using decentralized control can be classified as physical and computational. Large systems may be physically separated into portions where information about the other portions is either not available or too costly to obtain. In many cases, the system is deliberately partitioned into subsystems for the sake of analyzing simpler and smaller subsystems and the complete system can be viewed as a composite of subsystems interacting with each other. In some cases, the system is physically separated and subsystems exist naturally.

In the past two decades, large scale system research has mainly focused on the control design aspects. Various decentralized control techniques and multi-level/hierarchical
control techniques are introduced[9, 11, 14, 10]. However, from the modeling point of view, little has been done to push forward the technology for identifying a large system under information constraints. This has been the case because the development of system identification in the framework of large scale systems is a difficult subject. Conventionally, models of large scale systems are usually constructed by combining models of the parts of the systems, but such models sometimes fail to describe the behavior of the whole system appropriately[14]. On the other hand, centralized system identification techniques are often prohibited because of the complexity of the problem, where numerical problems may arise when dealing with high order systems[16], or information constraints imposed in the system may cause difficulties. It is evident that some ways must be devised to solve the problem of system identification in large scale systems with information constraints.

Dual to the success of decentralized control methodologies, the parameter estimation problem in large scale systems may also be studied in a decentralized manner. The problem of state estimation or observation was a widely studied problem in the 1970’s. Both decentralized and hierarchical techniques have been reported in the literature[17, 18, 19, 20, 21, 22, 12]. Other work has focused on parameter estimation of local subsystems in an interconnected system[23, 24, 25, 26]. In this dissertation, the set theoretic robust estimation method is introduced to decentrally and robustly identify the local subsystems of an interconnected system.
1.3 Motivation and Organization of this Thesis

In the last two sections, three specific problems have been identified in the field of system identification:

1. It is known that models form the basis for system analysis and design; however, modeling errors are unavoidable in most analytical or experimental modeling techniques. The need to develop robust deterministic estimation technology is clear.

2. Most available identification techniques are generally developed for simulation, prediction and control design. In the past few years, tremendous progress has been made in both adaptive and non-adaptive robust control theory. There exists a gap between what a traditional system identification procedure can model and what a robust control design technique requires about the model. The need to develop identification procedures specifically for robust control designs is evident.

3. When centralized identification of large scale systems is not feasible due to numerical or computational complexities, a decentralized approach is justified particularly in the presence of information constraints. A better understanding of how decentralized identification can be conducted under decentralized information constraints is necessary.

The motivation for this dissertation lies in the above three problems and the contribution of this dissertation is to provide feasible solutions to these problems.
In Chapter II, the problem of parameter set estimation is addressed when the measurement noise or external disturbances are bounded. This problem is very practical because assumptions on stochastics (such as noise statistics) are often inaccurately obtained for finite data records. In the chapter, an optimal volume ellipsoid algorithm (OVE) is developed for an ARX model with bounded noise. This problem has been widely studied in the control and system literature, but the optimal parameter set has not been achieved and the goal of this algorithm is to give the true optimal solution. The OVE algorithm is essentially equivalent to ellipsoid algorithms with parallel cuts which have recently appeared in Operations Research literature. The new derivation given in this chapter provides geometric insight and offers correct development for the parameter set estimation problem.

In Chapter III, the convergence properties of the OVE algorithm are discussed. The importance of this convergence analysis lies in the fact that the convergence of an estimator has long been recognized as the key issue in establishing stable indirect adaptive control. In this chapter, the convergence of the size of the ellipsoids and their center are the main focus. To substantiate the discussion on convergence study, the issue on the true optimal bounding set for a feasible parameter set is raised, and the “small” noise and noise free cases are examined.

In Chapter IV, an extension of the optimal algorithm developed in Chapter II to a more general model structure, the ARMAX model is considered. As a first attempt to solve this problem, a two-step solution is proposed. The idea is to recast the problem in the ARMAX model setting to an ARX setting. The first step of the proposed
scheme is to find an instantaneous bound for the MA part so that the problem can be viewed as the one with an ARX model. The second step is just a direct application of the OVE algorithm which is developed solely for ARX models.

In Chapter V, the OVE algorithm is applied in the decentralized setting and parameter set estimation of local subsystems in an interconnected system is studied. Here, the interconnected system is viewed from an input-output setting and each subsystem is interconnected to other subsystems through some cross transfer functions. In the proposed solution, the worst case instantaneous interconnection strength is computed to define a new disturbance bound so that the set estimation problem of a local subsystem can be treated as an ARX system.

In Chapter VI, conclusions are drawn summarizing the achievement made in this dissertation. Further discussion and suggestion for improvement are provided for future research. Some final recommendations are given to solve other related parameter set estimation problems such as over-modeling, under-modeling and also the case to include a priori known modeling uncertainties.
CHAPTER II

OPTIMAL PARAMETER SET ESTIMATION
OF AN ARX MODEL WITH BOUNDED NOISE

2.1 Introduction

The concept of parameter set estimation in system identification has evolved over the past two decades. The motive in parameter set estimation is to identify a feasible set of parameters which is consistent with the measurement data and the model structure used. One can interpret the set estimate as some nominal parameter estimate accompanied by a quantification of the uncertainty parametrically around the nominal model. An important feature in the parameter set estimation is the guaranteed inclusion of the true plant which is not exactly known. The pioneering work of set estimation can be dated back to the late 1960s when the set estimation of system states with bounded noise[3, 27] was studied. The work was later extended to the area of system parameter identification[5] and a recursive Optimal Bounding Ellipsoid (OBE) algorithm was developed for the ARX model[5, 28] with bounded noise. Other researchers approach the parameter set estimation problem from a different perspective. Kosut et al. [29, 30] approach the problem from an adaptive control point of view in which some a priori known bounds on unstructured dynamics are
reflected into the parameter space and it is shown that the parameter estimator can always capture the true system. Goodwin et al. [31, 32] approach the robust estimation problem in the Bayesian framework in which Gaussian a priori distributions are assumed for the parameters to account for under-modeling and a posteriori distribution for the nominal parameters is proposed. Other identification work dealing with uncertainty in the frequency domain can be found in [33, 34, 35].

In this chapter, a new optimal recursive ellipsoid algorithm for parameter set estimation is developed. This new result is based on the Khachian ellipsoid algorithm [36] developed for solving the linear programming problem. At each recursion, the smallest volume ellipsoid bounding a convex polytope defined by the bounded noise is found. This new result is distinct from the innovative OBE algorithm due to Fogel and Huang [5] in that every new ellipsoid in the course of updating is optimal under no constraints. This is in contrast with the OBE algorithm in which the optimization is subjected to the constraint that the center of the ellipsoid is a "modified" recursive least squares estimate. Though both the new algorithm and the OBE algorithm have similar recursive equations for implementation, the new optimal bounding algorithm gives an appealing geometrical interpretation of the ellipsoid bounding the convex set in interest and ultimately results in the true minimum volume bounding ellipsoids.

As pointed out in [37, 38], the OBE is not optimal, that the final resulting volume is not minimal. In this chapter, it is shown that in the class of recursive ellipsoid algorithms, the new bounding algorithm is truly optimal, in the sense that the new algorithm gives the smallest volume ellipsoid at each recursive computation. Com-
paring to other optimal algorithms in [39, 37], the new algorithm presented in this paper has some distinct advantages in that it is recursive, easy to implement and the problem complexity does not increase with number of measurements.

In 1979 Leonid Khachian published a polynomial time algorithm for solving the linear programming problem. The ellipsoid algorithm developed by Khachian[36] was primarily used to study the separation problem[40], that is, for a given non-empty convex compact set $\mathcal{K} \subset \mathbb{R}^n$, $g_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ such that $\mathcal{K} = \{x : g_i^T x \leq b_i, i = 1, \cdots, n \text{ and } x \in \mathbb{R}^n\}$ and given a point $y \in \mathbb{R}^n$, decide if $y \in \mathcal{K}$, and if not, find a hyperplane which separates $y$ from $\mathcal{K}$.

In Khachian’s algorithm, a minimal volume ellipsoid $\bar{E}$ is found to contain one-half of another ellipsoid, $E$, which includes the convex set $\mathcal{K}$ as illustrated in Figure 2. The ellipsoids $\bar{E}$ and $E$ are defined as

\begin{align*}
E &= \{x : (x - x_0)^T A^{-1} (x - x_0) \leq 1\} \\
\bar{E} &= \{x : (x - \bar{x}_0)^T \bar{A}^{-1} (x - \bar{x}_0) \leq 1\}
\end{align*}

where $A$ and $\bar{A}$ are symmetric p.d. matrices.

The Khachian ellipsoid algorithm essentially uses $g_i^T (x - x_0) = 0$ as the hyperplane to separate the ellipsoid $E$ into two halves, and uses a minimal volume $\bar{E}$ among a special class of ellipsoids[36] to enclose the half that contains the convex set $\mathcal{K}$. The process is recursive and continues until the center of an ellipsoid satisfies all the constraints. Modifications are suggested in [41, 42] to find an ellipsoid to bound only the intersecting region between the half space defined by $g_i^T x \leq b_i$ and the ellipse $E$ rather than one half of the ellipsoid $E$ cut with the constraint $g_i^T (x - x_0) = 0$. 
The main contribution of this chapter is the extension of the single hyperplane cut in Khachian's ellipsoid algorithm to the parameter set estimation problem where the feasible set of estimates are constrained between two parallel hyperplanes. Khachian's ellipsoid algorithm handles multiple constraints sequentially whereas here, multiple constraints are handled pairwise. In fact, our algorithm can be reduced to Khachian's ellipsoid algorithm. The case of ellipsoid algorithm with parallel cuts has also been studied in the Operation Research literature area[43, 44] for solving linear programming problems. The algorithm derived in this chapter offers

1. careful and correct development for the parameter set estimation problem, with examples and convergence proof;
2. an entirely different proof which offers more geometrical insight into the problem.

Moreover, the similarities between the new algorithm and the OBE algorithm in [5] are noted, and a qualitative comparison is included.

In the next section, we define the problem as finding the minimal volume ellipsoid to contain the intersecting region between another ellipsoid and two parallel hyperplanes. The derivation of the algorithm is given in Section 2.3 where the main result, Theorem 2.2 and Theorem 2.3 are proven. A comparison between the new ellipsoid algorithm and the OBE algorithm due to Fogel and Huang is discussed in Section 2.4, whereas Section 2.5 gives a comparison between the two algorithms through computer simulation. Some conclusions are drawn in the last section.

2.2 Problem Statement

Consider a SISO ARX model,

\[ y_k = - \sum_{i=1}^{n} a_i y_{k-i} + \sum_{j=0}^{m} b_j u_{k-j} + v_k \]

(2.3)

\[ = \theta^T \phi_k + v_k \]

(2.4)

where \( \theta^T = [a_1, \ldots, a_n, b_0, \ldots, b_m] \) is the parameter vector to be estimated; \( \phi_k = [-y_{k-1}, \ldots, -y_{k-n}, u_k, \ldots, u_{k-m}] \) is the regression vector containing the past inputs, \( u(\cdot) \), and outputs, \( y(\cdot) \); \( n \) and \( m \) are the number of system poles and zeroes, respectively; \( v_k \) is a sequence of bounded disturbances/noise corrupting the system output with \( |v_k| \leq \gamma \) for all \( k \geq 0 \). It is assumed that \( n, m \) and \( \gamma \) are known \textit{a priori}. 
Let $\mathcal{F} \subseteq \mathbb{R}^{n+m+1}$ be a set such that all $\theta \in \mathcal{F}$ are feasible parameter estimates of the plant which are consistent with the measurements. That is,

$$\mathcal{F} = \{ \theta : |y_k - \theta^T \phi_k| \leq \gamma, k = 0, \ldots, N \}. \tag{2.5}$$

The problem of parameter set estimation is to find $\mathcal{F}$ explicitly in the parameter space. In general, $\mathcal{F}$ is an irregular convex set, so we wish to find a more manageable convex set to over-bound $\mathcal{F}$ for the purpose of system analysis and control. Ellipsoids are commonly used to bound $\mathcal{F}$ for their simplicity in mathematical representation and manipulation in computation. It is therefore desired to find the smallest ellipsoid to contain the set $\mathcal{F}$ where the hyper-volume of an ellipsoid is used to measure "smallness".

### 2.3 The New Optimal Algorithm for Parameter Set Estimation

In this section, Khachian's ellipsoid algorithm is extended to the problem of parameter set estimation. For convenience, we will refer to the new algorithm as the Optimal Volume Ellipsoid (OVE) algorithm. For the set of inequality constraints in $\mathcal{F}$ defined in the last section, consider a pair of constraints

$$|y_{k+1} - \theta^T \phi_{k+1}| \leq \gamma, \tag{2.6}$$

and let the set $\mathcal{F}_{k+1}$ be defined as

$$\mathcal{F}_{k+1} = \{ \theta : |y_{k+1} - \theta^T \phi_{k+1}| \leq \gamma \}. \tag{2.7}$$

Geometrically, $\mathcal{F}_{k+1}$ is the region between the two parallel hyperplanes defined in (2.6). The set estimation problem is then stated as: Given an ellipsoid $E_k$, find
another ellipsoid $E_{k+1}$ with minimal volume, such that $E_{k+1}$ contains $E_k \cap \mathcal{F}_{k+1}$, for $k = 0, \ldots, N$, where $N$ is the number of data records. Mathematically, the optimization problem becomes

$$\min \{ \text{vol}(E_{k+1}) : E_{k+1} \supset E_k \cap \mathcal{F}_{k+1} \}.$$ 

Define $E_k$ and $E_{k+1}$ as

$$E_k = \{ \theta : (\theta - \theta_k)^T P_k^{-1} (\theta - \theta_k) \leq 1 ; \ \theta \in \mathbb{R}^r \}$$

and

$$E_{k+1} = \{ \theta : (\theta - \theta_{k+1})^T P_{k+1}^{-1} (\theta - \theta_{k+1}) \leq 1 ; \ \theta \in \mathbb{R}^r \}, \tag{2.8}$$

where $r = n + m + 1$ and $\theta_k$ is the center estimate of the ellipsoid at time $k$.

In the derivation of the OVE algorithm, an affine transformation [36, 40],

$$\theta = \theta_k + J \hat{\theta} \tag{2.10}$$

is used to simplify the analysis where $\theta \in \mathbb{R}^r$ is any vector in the parameter space, $\hat{\theta}$ is the parameter vector in the affine transformed coordinate and $P_k = JJ^T$. Through this transformation, the ellipsoid $E_k$ is mapped to the unit radius hypersphere centered at the origin. The set estimation problem for a specific value of $k$ reduces to finding the minimal volume ellipsoid containing the intersection between a unit radius hypersphere and two parallel hyperplanes defined by $\hat{\mathcal{F}}$, the affine transformation of $\mathcal{F}_{k+1}$. Let

$$\hat{S} = \{ \hat{\theta} : \hat{\theta}^T \hat{\theta} \leq 1 \} \tag{2.11}$$
and

\[ \hat{\mathcal{K}} = \{ \hat{\theta} : \frac{\hat{\phi}^T \hat{\phi}}{(\hat{\phi}^T \hat{\phi})^{1/2}} \leq \alpha \text{ and } \frac{\hat{\phi}^T \hat{\phi}}{(\hat{\phi}^T \hat{\phi})^{1/2}} \geq \alpha - 2\beta \} \] (2.12)

where \( \hat{\phi} \) is the transformed \( \phi_{k+1} \), and \( \alpha, \beta \) are parameters defining the location of the two parallel hyperplanes defined in \( \hat{\mathcal{K}} \) which are \( 2\beta (\beta > 0) \) apart (we will define these parameters later as related to the original parameter set estimation problem).

In what follows, we wish to find an ellipsoid

\[ \hat{E} = \{ \hat{\theta} : (\hat{\theta} - \hat{\theta}_0)^T \hat{\Lambda}^{-1}(\hat{\theta} - \hat{\theta}_0) \leq 1 \} \] (2.13)

such that \( \hat{S} \cap \hat{\mathcal{K}} \subset \hat{E} \) and the volume of \( \hat{E}, \text{vol}(\hat{E}) \), is minimized.

For the purpose of analysis, define:

\[ \hat{H}_1^* = \{ \hat{\theta} : \frac{\hat{\phi}^T \hat{\phi}}{(\hat{\phi}^T \hat{\phi})^{1/2}} = \alpha \} \]

\[ \hat{H}_1 = \{ \hat{\theta} : \frac{\hat{\phi}^T \hat{\phi}}{(\hat{\phi}^T \hat{\phi})^{1/2}} \leq \alpha \} \]

\[ \hat{H}_2^* = \{ \hat{\theta} : \frac{\hat{\phi}^T \hat{\phi}}{(\hat{\phi}^T \hat{\phi})^{1/2}} = \alpha - 2\beta \} \]

\[ \hat{H}_2 = \{ \hat{\theta} : \frac{\hat{\phi}^T \hat{\phi}}{(\hat{\phi}^T \hat{\phi})^{1/2}} \geq \alpha - 2\beta \} \]

\[ \hat{E}^* = \{ \hat{\theta} : (\hat{\theta} - \hat{\theta}_0)^T \hat{\Lambda}^{-1}(\hat{\theta} - \hat{\theta}_0) = 1 \} \]

\[ \hat{S}^* = \{ \hat{\theta} : \hat{\theta}^T \hat{\theta} = 1 \} \].

From these definitions, \( \hat{\mathcal{K}} \) is the region between the two hyperplanes, \( \hat{H}_1^* \) and \( \hat{H}_2^* \) are the two hyperplanes \( 2\beta \) apart with \( \hat{H}_1^* \) at a distance \( \alpha \) from the origin of \( \hat{S} \), and the vector \( \hat{\phi} \) is orthogonal to both the hyperplanes. Before we proceed, we need the
following theorem as a foundation for the main result.

**Theorem 2.1:** Given \( \hat{S} \) and \( \hat{\mathcal{C}} \), the minimal volume ellipsoid \( \hat{E} \) bounding \( \hat{S} \cap \hat{\mathcal{C}} \) must satisfy the following conditions:

\[
\begin{align*}
\hat{H}_1^* \cap \hat{S} & = \hat{H}_1^* \cap \hat{E} \\
\hat{H}_2^* \cap \hat{S} & = \hat{H}_2^* \cap \hat{E}.
\end{align*}
\] (2.14) (2.15)

Essentially, these conditions imply that \( \hat{E}^* \), the surface of \( \hat{E} \), must pass through the intersecting points between \( \hat{S}^* \) (the surface of \( \hat{S} \)) and the two parallel hyper-planes, \( \hat{H}_1^* \) and \( \hat{H}_2^* \). Qualitatively, since \( \hat{S} \cap \hat{\mathcal{C}} \) is symmetrical about \( \hat{\phi} \), the smallest volume ellipsoid must also be symmetrical about \( \hat{\phi} \), and given any ellipsoid symmetrical about \( \hat{\phi} \) and containing \( \hat{S} \cap \hat{\mathcal{C}} \) but not satisfying (2.14) and (2.15), a smaller ellipsoid can be constructed such that it not only contains \( \hat{S} \cap \hat{\mathcal{C}} \), but also satisfies both (2.14) and (2.15).

**Proof of Theorem 2.1:** Let \( \hat{\theta}_i \) be the \( i^{th} \) parameter coordinate axis of the Euclidean space \( \mathbb{R}^r \) and let \( \alpha_1 = \alpha - 2\beta, \alpha_2 = \alpha \). Define, without loss of generality,

\[
\begin{align*}
\hat{S} & = \{ \hat{\theta} : \hat{\theta}^T \hat{\theta} \leq 1; \hat{\theta} \in \mathbb{R}^r \} \\
\hat{\mathcal{C}} & = \{ \hat{\theta} : \alpha_1 \leq \hat{\phi}^T \hat{\theta} \leq \alpha_2; -1 \leq \alpha_1 < \alpha_2 \leq 1; \hat{\theta} \in \mathbb{R}^r \} \\
\hat{\phi} & = [1 \ 0 \ldots 0]^T_{(r-1)\text{terms}} \\
\hat{\mathcal{C}}^* & = \hat{H}_1^* \cup \hat{H}_2^*
\end{align*}
\]
Essentially, $\hat{S}$ is a $r$-dimensional hypersphere and $\hat{K}$ is an unbounded region between two parallel hyper-planes which are orthogonal to the $\hat{\theta}_1$ axis. The intersection $\hat{S} \cap \hat{K}$ can be parameterized as

$$\hat{S} \cap \hat{K} = \{ \hat{\theta} : \alpha_1 \leq \hat{\theta} \leq \alpha_2; \hat{\theta}_1^2 + \hat{\theta}_2^2 + \cdots + \hat{\theta}_r^2 \leq 1 \}$$

Consider any given ellipsoid $\bar{E}$ centered at $\bar{\theta}_c \in \mathbb{R}^r$

$$\bar{E} = \{ \hat{\theta} : (\hat{\theta} - \bar{\theta}_c)^T P^{-1} (\hat{\theta} - \bar{\theta}_c) \leq 1 \} \quad (2.16)$$

where $P \in \mathbb{R}^{r \times r}$ is p.d., with

$$\bar{E} \supset \hat{S} \cap \hat{K} \quad (2.17)$$

meaning $\bar{E}$ contains $\hat{S} \cap \hat{K}$ where $\hat{S} \cap \hat{K}^* \neq \bar{E} \cap \hat{K}^*$. The proof amounts to showing that there exists another smaller volume ellipsoid, $\bar{E}$, such that

$$\bar{E} \supset \hat{S} \cap \hat{K} \quad (2.18)$$

and

$$\hat{S} \cap \hat{K}^* = \bar{E} \cap \hat{K}^* \quad (2.19)$$

where (2.19) is equivalent to the conditions in (2.14) and (2.15). Intuitively, $\hat{S} \cap \hat{K}^* = \bar{E} \cap \hat{K}^*$ means that the surface of $\bar{E}$ contains the intersection between the surface of $\hat{S}$ and the two bounding hyperplanes associated with $\hat{K}$. The following lemma allows us to restrict our attention to the class of ellipsoids (2.16) that are orthogonal to the $\hat{\theta}_1$ axis.
Lemma 2.1: Given $\hat{S} \cap \hat{K}$ which is symmetrical about the $\hat{\theta}_1$ axis, the minimal volume ellipsoid containing $\hat{S} \cap \hat{K}$ must also be symmetrical about the $\hat{\theta}_1$ axis.

Proof of Lemma 2.1: Consider the $r = 2$ case and any $\hat{S} \cap \hat{K}$. Suppose there exists an oblique ellipse $\bar{E}$ such that $\bar{E} \supset \hat{S} \cap \hat{K}$ having also the smallest area among ellipses containing $\hat{S} \cap \hat{K}$. Since $\hat{S} \cap \hat{K}$ is invariant if the $\hat{\theta}_2$ axis is interchanged with the $-\hat{\theta}_2$ axis, essentially a rotation action about the $\hat{\theta}_1$ axis for 180 deg, a new ellipse $\bar{E}'$ which has the same area as $E$ will be formed which also contains $\hat{S} \cap \hat{K}$. This contradicts the result in [21, Theorem 3.1.9] that for every convex body in $\mathbb{R}^r$ there exists a unique ellipsoid of minimal volume containing that convex body. In other words, the minimal area ellipse must be symmetrical about the $\hat{\theta}_1$ axis, and therefore it must have one ellipsoid axis along the $\hat{\theta}_1$ axis and the other orthogonal to $\hat{\theta}_1$. In the general $r$-dimensional case, the same argument can be used that interchanging $\hat{\theta}_i$ with $-\hat{\theta}_i$ axis, $i \neq 1$, will not change $\hat{S} \cap \hat{K}$ and therefore the minimal volume ellipsoid must have one ellipsoid axis along $\hat{\theta}_1$ and all others in a hyperplane orthogonal to $\hat{\theta}_1$. This completes the proof of Lemma 2.1.

Due to this lemma, it is sufficient to consider the class of ellipsoids

$$\bar{E} = \{\hat{\theta} : \frac{(\hat{\theta}_1 - \hat{\theta}_c)^2}{d_1^2} + \frac{\hat{\theta}_2^2}{d_2^2} + \cdots + \frac{\hat{\theta}_r^2}{d_r^2} \leq 1\}$$

where an ellipsoid centered at $\hat{\theta} = (\hat{\theta}_c, 0, \cdots, 0)$ in $\bar{E}$ has one axis aligned with $\hat{\theta}_1$ and other axes parallel to the $\hat{\theta}_i$. 
The proof for the case where a smaller volume ellipsoid $\tilde{E}$ than $\tilde{E}$ can be found to contain $\hat{S} \cap \hat{K}$ is by construction. Assume $\tilde{E} \cap \hat{K}^* \neq \hat{S} \cap \hat{K}^*$. First in the proof is the construction of a smaller ellipsoid $\tilde{E}$ which is obtained by contracting $\tilde{E}$ in all directions except along the $\hat{\theta}_1$ axis, and then contracting further in only the $\hat{\theta}_1$ direction to obtain the desired $\tilde{E}$.

A. Construction of $\tilde{E}$

Let

$$\tilde{E} = \{ \hat{\theta} : \frac{(\hat{\theta}_1 - \tilde{\theta}_c)^2}{d_1^2} + \frac{\hat{\theta}_2^2}{\eta^2} + \cdots + \frac{\hat{\theta}_r^2}{\eta^2} \leq 1 \}$$

where

$$\eta^2 = \min\{d_i^2 : 2 \leq i \leq r\} \quad (2.20)$$

The ellipsoid $\tilde{E}$ is essentially a contracted $\tilde{E}$ where contraction is made in all directions except along the $\hat{\theta}_1$ axis. Both $\tilde{E}$ and $\tilde{E}$ have the same center and all the ellipsoid axes (except the one along $\hat{\theta}_1$) of $\tilde{E}$ are made to have the same length equal to the minimum among those in $\tilde{E}$, while the length of the axis along $\hat{\theta}_1$ remains the same.

Note that we can always find a point $\hat{\theta} = (\hat{\theta}_1, 0, \cdots, 0, \hat{\theta}_m, 0, \cdots, 0)$ in $\hat{S} \cap \hat{K}$ such that it is also on the surface of $\hat{S}$, where $\hat{\theta}_m$ is the coordinate where the minimum occurs in (2.20). Since this point is in $\tilde{E}$, then

$$\frac{(\hat{\theta}_1 - \tilde{\theta}_c)^2}{d_1^2} + \frac{\hat{\theta}_m^2}{\eta^2} \leq 1$$

and

$$\frac{(\hat{\theta}_1 - \tilde{\theta}_c)^2}{d_1^2} + \frac{1 - \hat{\theta}_1^2}{\eta^2} \leq 1 \quad (2.21)$$

In this context, by "contraction" we mean reducing the volume of an ellipsoid.
follows when \( \hat{\theta}_1^2 + \hat{\theta}_m^2 = 1 \). Clearly, (2.21) holds for any \( \hat{\theta}_1 \) in the interval \([\alpha_1, \alpha_2]\). Thus, considering any \( \bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_r) \), in \( \hat{\mathcal{S}} \cap \hat{\mathcal{K}} \), since \( \varepsilon^T \bar{\theta} \leq 1 \), then

\[
\frac{(\bar{\theta}_1 - \bar{\theta}_c)^2}{d_1^2} + \frac{\bar{\theta}_2^2}{\eta^2} + \cdots + \frac{\bar{\theta}_r^2}{\eta^2} \leq \frac{(\hat{\theta}_1 - \hat{\theta}_c)^2}{d_1^2} + \frac{1 - \hat{\theta}_1^2}{\eta^2} \leq 1
\] (2.22)

again because \( \alpha_1 \leq \hat{\theta}_1 \leq \alpha_2 \). This implies that \( \bar{\theta} \) is also in \( \bar{\mathcal{E}} \) and consequently \( \bar{\mathcal{E}} \supset \hat{\mathcal{S}} \cap \hat{\mathcal{K}} \).

**B. Construction of \( \bar{\mathcal{E}} \)**

Let

\[ \bar{\mathcal{E}} = \{ \hat{\theta} : \frac{(\hat{\theta}_1 - \hat{\theta}_c)^2}{d_1^2} + \frac{\hat{\theta}_2^2}{\eta^2} + \cdots + \frac{\hat{\theta}_r^2}{\eta^2} \leq 1 \} \]

Ellipsoid \( \bar{\mathcal{E}} \) has the same orientation as \( \bar{\mathcal{E}} \); both \( \bar{\mathcal{E}} \) and \( \bar{\mathcal{E}} \) have identical lengths in their corresponding ellipsoid axes except along the \( \hat{\theta}_1 \) axis. Essentially, \( \bar{\mathcal{E}} \) is a contracted \( \bar{\mathcal{E}} \) where the contraction is made only along the \( \hat{\theta}_1 \) axis.

Before we show that \( \bar{\mathcal{E}} \supset \hat{\mathcal{S}} \cap \hat{\mathcal{K}} \) and \( \bar{\mathcal{E}} \cap \hat{\mathcal{K}}^* = \hat{\mathcal{S}} \cap \hat{\mathcal{K}}^* \), we first consider a 2-D ellipse obtained from the \( r \)-dimensional ellipsoid \( \bar{\mathcal{E}} \) by setting all \( \hat{\theta}_i \) to zero except \( \hat{\theta}_1 \) and some \( \hat{\theta}_c \). Without loss of generality, let \( \hat{\theta}_c = \hat{\theta}_2 \), and denote all quantities in such a 2-D case with a subscript "2". Therefore, from the expression for \( \bar{\mathcal{E}} \),

\[ \bar{\mathcal{E}}_2 = \{ \hat{\theta} : \frac{(\hat{\theta}_1 - \hat{\theta}_c)^2}{d_1^2} + \frac{\hat{\theta}_2^2}{\eta^2} \leq 1 \} \]

and from the expression for \( \bar{\mathcal{E}} \),

\[ \bar{\mathcal{E}}_2 = \{ \hat{\theta} : \frac{(\hat{\theta}_1 - \hat{\theta}_c)^2}{d_1^2} + \frac{\hat{\theta}_2^2}{\eta^2} \leq 1 \} \]

and

\[ \hat{\mathcal{S}}_2 \cap \hat{\mathcal{K}}_2 = \{ \hat{\theta} : \alpha_1 \leq \hat{\theta}_1 \leq \alpha_2; \hat{\theta}_1^2 + \hat{\theta}_2^2 \leq 1 \} \].
Given \( \mathcal{E}_2 \supset \hat{S}_2 \cap \hat{K}_2 \), the existence of \( \bar{\theta}_c \) and \( \bar{d}_1 \) such that \( \mathcal{E}_2 \supset \hat{S}_2 \cap \hat{K}_2 \) and \( \mathcal{E}_2 \cap \hat{K}_2^* = \hat{S}_2 \cap \hat{K}_2^* \) is easily proven since any (2-D) ellipse orthogonal to the \( \hat{\theta}_1 \) axis containing \( \hat{S}_2 \cap \hat{K}_2 \) can be shifted along the \( \hat{\theta}_1 \) axis to touch \( \hat{S}_2 \cap \hat{K}_2^* \), then contracted (along \( \hat{\theta}_1 \)) to obtain the desired result, an ellipse with center \( \bar{\theta}_c \) and axial length \( \bar{d}_1 \).

Let \( \hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2) \), and \( \hat{\theta} \in \mathcal{E}_2 \):

(i) if \( \hat{\theta} \in \hat{S}_2 \cap \hat{K}_2 \), then

\[
\frac{(\hat{\theta}_1 - \bar{\theta}_c)^2}{d_1^2} + \frac{\hat{\theta}_2^2}{\eta^2} \leq 1
\]

since \( \hat{\theta} \in \mathcal{E}_2 \). In the particular case where \( \alpha_1 \leq \hat{\theta}_1 \leq \alpha_2 \) and \( \hat{\theta}_1 + \hat{\theta}_2 = 1 \), it follows that

\[
\frac{(\hat{\theta}_1 - \bar{\theta}_c)^2}{d_1^2} + \frac{1 - \hat{\theta}_1}{\eta^2} \leq 1
\]  

(2.23)

(ii) if \( \hat{\theta} \in \hat{S}_2 \cap \hat{K}_2 \) such that \( \hat{\theta}_1 = \alpha_i, i = 1, 2 \), then

\[
\frac{(\alpha_i - \bar{\theta}_c)^2}{d_1^2} + \frac{\hat{\theta}_2^2}{\eta^2} \leq 1.
\]

In the particular case where \( \alpha_i^2 + \hat{\theta}_2 = 1 \), it follows that

\[
\frac{(\alpha_i - \bar{\theta}_c)^2}{d_1^2} + \frac{1 - \alpha_i^2}{\eta^2} = 1
\]  

(2.24)

Now, use these results to show that, in the general case for \( r > 2 \), \( \mathcal{E} \supset \hat{S} \cap \hat{K} \) and \( \mathcal{E} \cap \hat{K}^* = \hat{S} \cap \hat{K}^* \). To do so, it is required to show that for any \( \bar{\theta} = (\bar{\theta}_1, \cdots, \bar{\theta}_r) \), (a) \( \bar{\theta} \in \hat{S} \cap \hat{K} \) implies \( \bar{\theta} \) is also in \( \mathcal{E} \); and (b) \( \mathcal{E} \cap \hat{K}^* = \hat{S} \cap \hat{K}^* \).

(a) If \( \bar{\theta} \in \hat{S} \cap \hat{K} \), then

\[
\frac{(\bar{\theta}_1 - \bar{\theta}_c)^2}{d_1^2} + \frac{\bar{\theta}_2^2}{\eta^2} + \cdots + \frac{\bar{\theta}_r^2}{\eta^2} \leq \frac{(\bar{\theta}_1 - \bar{\theta}_c)^2}{d_1^2} + \frac{1 - \bar{\theta}_1^2}{\eta^2} \leq 1
\]  

(2.25)
using the result of equation (2.23) with \( \alpha_1 \leq \bar{\theta}_1 \leq \alpha_2 \). This implies \( \bar{\theta} \in \bar{E} \).

(b) If \( \bar{\theta} \in \hat{K}^* \) and \( \bar{\theta} \in \hat{S} \), then

\[
\frac{\left(\alpha_i - \bar{\theta}_c\right)^2}{d_i^2} + \frac{\bar{\theta}_m^2}{\eta^2} + \cdots + \frac{\bar{\theta}_n^2}{\eta^2} \leq \frac{\left(\alpha_i - \bar{\theta}_c\right)^2}{d_i^2} + \frac{1 - \alpha_i^2}{\eta^2} = 1
\] (2.26)

using the result of equation (2.24). This implies \( \bar{\theta} \) is also in \( \bar{E} \) and therefore \( \bar{\theta} \in \bar{E} \cap \hat{K}^* \), that is \( \bar{E} \cap \hat{K}^* \supset \hat{S} \cap \hat{K}^* \). Note that if \( \bar{\theta} \) is on the surface of \( \hat{S} \) (\( \hat{S}^* \)), that is \( \bar{\theta}^T \bar{\theta} = 1 \), then the "\( \leq \)" sign in (2.26) becomes "\( = \)" which implies \( \bar{\theta} \) is also on the surface of \( \bar{E} \) (\( \bar{E}^* \)). In other words, the boundaries of \( \bar{E} \cap \hat{H}^*_1 \) and \( \bar{E} \cap \hat{H}^*_2 \) are identical to that of \( \hat{S} \cap \hat{H}^*_1 \) and \( \hat{S} \cap \hat{H}^*_2 \), respectively. Since \( \bar{E} \cap \hat{H}^*_1 \), \( \bar{E} \cap \hat{H}^*_2 \), \( \hat{S} \cap \hat{H}^*_1 \) and \( \hat{S} \cap \hat{H}^*_2 \) are compact convex sets \( (r > 2) \), it follows that

\[
\bar{E} \cap \hat{H}^*_1 = \hat{S} \cap \hat{H}^*_1
\]

and

\[
\bar{E} \cap \hat{H}^*_2 = \hat{S} \cap \hat{H}^*_2,
\]

which is equivalent to \( \bar{E} \cap \hat{K}^* = \hat{S} \cap \hat{K}^* \).

Therefore, we have shown that for any given ellipsoid \( \Bar{E} \) containing \( \hat{S} \cap \hat{K} \), we can find a smaller volume ellipsoid \( \bar{E} \) than \( \Bar{E} \) such that \( \bar{E} \supset \hat{S} \cap \hat{K} \) and \( \bar{E} \cap \hat{K}^* = \hat{S} \cap \hat{K}^* \). This completes the proof of Theorem 2.1. \( \square \)

We now state the main result for the OVE algorithm.

**Theorem 2.2:** For the sets \( \hat{S} \) and \( \hat{K} \), if \( |\alpha| \leq 1 \) and \( |2\beta - \alpha| \leq 1 \), then the following parameters will result in a minimal volume \( \hat{E} \) that contains \( \hat{S} \cap \hat{K} \):

\[
\hat{\theta}_0 = r \hat{\phi}/(\hat{\phi}^T \hat{\phi})^{1/2}
\]
\[ \hat{A} = \delta(I - \frac{\sigma \hat{\phi}^T}{\hat{\phi}^T \hat{\phi}}) \]

provided \( \sigma \geq 0 \) where

(i) if \( \alpha \neq \beta \):

\[ \begin{align*}
\delta &= \frac{(\tau + 1)^2(\beta - \alpha) - \tau(1 + \alpha)(2\beta - \alpha - 1)}{\tau + \beta - \alpha} \\
\sigma &= \frac{-\tau}{\beta - \alpha}
\end{align*} \tag{2.27} \]

and \( \tau \) is the real solution of

\[ (r + 1)r^2 + \left\{ \frac{(1 + \alpha)(\alpha - 2\beta + 1)}{\beta - \alpha} + 2[r(\beta - \alpha) + 1]\right\} \tau + r\alpha(\alpha - 2\beta) + 1 = 0 \tag{2.29} \]

such that \( \alpha - 2\beta < \tau < \alpha \);

(ii) if \( \alpha = \beta \):

\[ \begin{align*}
\delta &= \frac{r}{r - 1}(1 - \beta^2) \\
\sigma &= \frac{1 - r\beta^2}{1 - \beta^2} \\
\tau &= 0
\end{align*} \tag{2.30, 2.31, 2.32} \]

In the above, note that \( I \) is an \( r \times r \) identity matrix and \( r = n + m + 1 \); moreover, if \( \sigma < 0 \), the minimal volume \( \hat{E} \) is \( \hat{S} \) itself.

Proof of Theorem 2.2: The proof is divided into two parts: (I) the parameters required for a minimal volume \( \hat{E} \) to contain \( \hat{H}_1^* \cap \hat{S} \) and \( \hat{H}_2^* \cap \hat{S} \), (II) the condition for \( \hat{E} \) to actually contain \( \hat{S} \cap \hat{K} \).

(I)(i) \( \alpha \neq \beta \):

Because of symmetry and the conditions in (2.14) and (2.15), the center \( \hat{\theta}_0 \) of the
ellipsoid \( \hat{E} \) must lie on \( \hat{\phi} \) at some distance \( \tau \) from the origin as depicted in Figure 3 for the two dimensional case, or

\[
\hat{\theta}_0 = \tau \frac{\hat{\phi}}{\left(\phi^T \hat{\phi}\right)^{\frac{1}{2}}} .
\] (2.33)

Since \( \hat{S} \cap \hat{C} \) is symmetrical about \( \hat{\phi} \) and \( \hat{\phi} \) is orthogonal to both of the two hyperplanes \( \hat{H}_1 \) and \( \hat{H}_2 \), the ellipsoid \( \hat{E} \) that has minimal volume must have one of the axes aligned with \( \hat{\phi} \) and all the other axes on a hyper-plane parallel to \( \hat{H}_1 \) or \( \hat{H}_2 \). Thus, to ensure that \( \hat{\phi} \) is an eigenvector of \( \hat{A} \) (i.e., ensuring that one of ellipsoid axes is along \( \hat{\phi} \)), \( \hat{A} \) can be parameterized as

\[
\hat{A} = \delta(I - \frac{\sigma \phi \phi^T}{\phi^T \phi}) .
\] (2.34)

Equation (2.14) implies

\[
\begin{align*}
(\hat{\theta} - \hat{\theta}_0)^T(\hat{\theta} - \hat{\theta}_0) &= \tau^2 - 2\alpha\tau + 1 \\
\frac{(\hat{\theta} - \hat{\theta}_0)^T \hat{\phi}}{(\phi^T \phi)^{1/2}} &= \alpha - \tau 
\end{align*}
\] (2.35)

while Equation (2.15) implies

\[
\begin{align*}
(\hat{\theta} - \hat{\theta}_0)^T(\hat{\theta} - \hat{\theta}_0) &= 1 - (\alpha - 2\beta)^2 + (\tau - \alpha + 2\beta)^2 \\
\frac{(\hat{\theta} - \hat{\theta}_0)^T \hat{\phi}}{(\phi^T \phi)^{1/2}} &= -\tau + \alpha - 2\beta 
\end{align*}
\] (2.36)

Substituting Equation (2.34) into the expression for \( \hat{E}^* \), one gets

\[
(1 - \sigma)(\hat{\theta} - \hat{\theta}_0)^T(\hat{\theta} - \hat{\theta}_0) + \frac{\sigma(\hat{\theta} - \hat{\theta}_0)\phi \phi^T(\hat{\theta} - \hat{\theta}_0)}{\phi^T \phi} = \delta(1 - \sigma) .
\] (2.37)

Upon using Equations (2.35) and (2.36), one obtains

\[
(1 - \sigma)(\tau^2 + 1 - 2\alpha\tau) + \sigma(\alpha - \tau)^2 = \delta(1 - \sigma) .
\] (2.38)
Figure 3: Ellipsoid bounding the intersection of a hypersphere and two parallel hyperplanes
and

\[(1 - \sigma)(\tau^2 + 1 - 2\alpha \tau + 4\beta \tau) + \sigma(2\beta - \alpha + \tau)^2 = \delta(1 - \sigma) \quad (2.39)\]

and solving for \(\delta\) and \(\sigma\) (2.28) and (2.27) result.

Since \(\alpha, \beta\) are fixed for a given \(\hat{H}_1^*\) and \(\hat{H}_2^*\), \(\tau\) is the only free parameter that can be used to minimize the volume of \(\hat{E}\). Let \(f(\tau) = \sqrt{\det(\hat{A})}\), then

\[f(\tau) = (\delta(1 - \sigma))^{1/2} \quad (2.40)\]

Since \(\text{vol}(\hat{E})\) is proportional to \(\sqrt{\det(\hat{A})}\) \((> 0)\), minimizing \(\text{vol}(\hat{E})\) is thus equivalent to minimizing \(f(\tau)\) with respect to \(\tau\). Upon setting \(\frac{df(\tau)}{d\tau} = 0\), one obtains

\[r\delta^{-1}(1 - \sigma)\frac{d\delta}{d\tau} = \delta\frac{d\sigma}{d\tau} \quad (2.41)\]

From (2.28) and (2.27),

\[
\begin{align*}
\frac{d\sigma}{d\tau} &= \frac{-1}{\beta - \alpha} \\
\frac{d\delta}{d\tau} &= \frac{(\beta - \alpha)(\tau - \alpha)(2\beta - \alpha + \tau)}{(\tau + \beta - \alpha)^2}
\end{align*}
\]

which, upon substituting into (2.41), gives Equation (2.29). Examination of the discriminant \(\Delta\) in (2.29),

\[
\Delta = r^2(4\beta^4 - 8\alpha\beta^3 + 4\alpha^2\beta^2) + (4\alpha^2 - 4)\beta^2 - 4(\alpha^3 - 4\alpha)\beta + \alpha^4 - 2\alpha^2 + 1
\]

\[= 4r^2\beta^2(\beta - \alpha)^2 + (\alpha^2 - 1)[(2\beta - \alpha)^2 - 1],\]

reveals that \(\Delta \geq 0\) if \(|\alpha| \leq 1\) and \(|2\beta - \alpha| \leq 1\), and in such a case, Equation (2.29) has two real solutions. Only one such solution satisfies \(\alpha - 2\beta < \tau < \alpha\) because,
otherwise, the ellipsoid center will be located outside the intersecting region which
cannot occur if we wish to find the smallest ellipsoid bounding the region.

(I)(ii) \( \alpha = \beta \):

In this case, \( \hat{S} \cap \hat{K} \) is symmetrical not only about \( \hat{\phi} \), but also about any other vector
perpendicular to \( \hat{\phi} \), denoted as \( \hat{\phi}^\perp \), which passes through the origin. It is trivial to
see that \( \tau = 0 \) due to the symmetry of \( \hat{S} \cap \hat{K} \) about all vectors \( \hat{\phi} \) and \( \hat{\phi}^\perp \). To find
the parameters for a minimal volume ellipsoid, \( \hat{E} \), to contain \( \hat{H}_f \cap \hat{S} \) and \( \hat{H}_i \cap \hat{S} \),
we consider first the 2-D case in which one of the axes is \( \hat{\phi} \) and the other axis is \( \hat{\phi}^\perp \).

With the circle in this case centered at the origin, we essentially consider a rectangle
with width \( 2\beta \) and length \( 2\sqrt{1-\beta^2} \) as shown in Figure 4. The minimal area ellipse
containing the rectangle has the following parameters:

\[
\begin{align*}
\tau &= 0 & (2.42) \\
\delta &= 2(1-\beta^2) & (2.43) \\
\delta(1-\sigma) &= 2\beta^2 & (2.44)
\end{align*}
\]

To see this, transform the rectangle to a square and then use the fact that the mini­
mal area ellipse circumscribing a square is a circle. The result then follows by trans­
forming the circle back to the original co-ordinates giving the minimal area ellipse
circumscribing the rectangle. Note that \( \sqrt{\delta} \) and \( \sqrt{\delta(1-\sigma)} \) are the lengths of the
major and minor axes of the ellipse, respectively. The result can be carried through
for the \( r \)-dimensional case, because all axes of the ellipsoid other than the one along
\( \hat{\phi} \) have the same lengths \( \sqrt{\delta} \).
(II) We have proved the results for a minimal volume ellipsoid $\hat{E}$ to contain $\hat{S} \cap \hat{H}_1^*$ and $\hat{S} \cap \hat{H}_2^*$ only. To ensure that $\hat{E}$ contains $\hat{S} \cap \hat{K}$, we first note that the ellipsoid $\hat{E}$ obtained has $r - 1$ axes with length $\sqrt{\delta}$ and one axis with length $\sqrt{\delta(1 - \sigma)}$ where the latter is along $\hat{\phi}$. In what follows, the proof is based on the fact that the size of the ellipsoid in the set we are considering can be parameterized by two different axial lengths, $\sqrt{\delta}$ and $\sqrt{\delta(1 - \sigma)}$. It is quite obvious that if $\sigma = 0$, $\hat{E}$ is essentially $\hat{S}$, the original hypersphere. Further if $\sigma > 0$, the containment of $\hat{S} \cap \hat{K}$ by $\hat{E}$ follows immediately. This can be seen if we consider the 2-D case when $\sigma > 0$, the extreme points on $\hat{E}$ along each of the ellipsoid axes are always located outside $\hat{S}$. That is, between the two parallel hyper-planes, the hyper-surface of the the hyper-sphere is
always included inside the ellipsoid, and therefore $\hat{E}$ always contains $\hat{S} \cap \hat{K}$. As a result, parts (i) and (ii) in the theorem hold only for $\sigma \geq 0$.

However, if $\sigma < 0$, then $\hat{E}$ will not contain the entire set of $\hat{S} \cap \hat{K}$; a 2-D case is illustrated in Figure 5. The minimal volume ellipsoid that contains $\hat{S} \cap \hat{K}$ can be found by "squeezing" $\hat{E}$ along $\hat{\phi}$, by decreasing $\delta(1 - \sigma)$, while expanding along all other axes in $\hat{\phi}^\perp$ of the ellipsoid to increase $\delta$. It is easy to see that as $\delta(1 - \sigma)$ decreases and $\delta$ increases, $\delta$ and $\delta(1 - \sigma)$ will eventually be identical, in which case $\hat{E}$ becomes $\hat{S}$ and $\hat{E}$ will certainly contain $\hat{S} \cap \hat{K}$. In the other words, if $\sigma < 0$, then the minimal volume ellipsoid is just the original hypersphere, $\hat{S}$, itself. This concludes the proof of Theorem 2.2.

\[\mathbf{□}\]

Figure 5: An illustration of the case $\sigma < 0$ in 2-D
Remarks:

1. The intuitive meaning of $|a| \leq 1$ is that $\hat{H}_1^*$ must cut, or at least touch, the hyper-sphere $\hat{S}$, whereas $|2\beta - a| \leq 1$ means that $\hat{H}_2^*$ must cut, or at least touch the hyper-sphere $\hat{S}$.

2. Since $\beta$ and $\alpha$ are not arbitrary and are defined in the given constraints, it is possible that $\alpha > 1$ or $2\beta - \alpha > 1$ which means that $\hat{H}_1^*$ or $\hat{H}_2^*$ does not cut $\hat{S}$, respectively. In the first case, $\beta$ can be reset to $\beta = \frac{a-1}{2}$ and then reset $\alpha$ to 1 if $\alpha > 1$ to make up a new hyper-plane parallel to $\hat{H}_1^*$ but touching $\hat{S}$. In the second case $\beta$ is set to $\frac{1+\alpha}{2}$ if $2\beta - \alpha > 1$ to make up a hyper-plane parallel to $\hat{H}_2^*$ but touching $\hat{S}$. Such a resetting of $\alpha$ or $\beta$ is essential to ensure that Equation (2.29) has real solutions and to ensure the bounding ellipsoid is appropriately oriented tightly circumscribing $\hat{S} \cap \hat{\mathcal{S}}$. Note that the set $\hat{S} \cap \hat{\mathcal{S}}$ is non-empty for a given physical system implying that there is always at least one hyperplane cutting $\hat{S}$ (either $\hat{H}_1^*$ or $\hat{H}_2^*$), and that only one of the two cases noted above can occur at any given time.

3. If $2\beta - \alpha = 1$, then $\tau = 1 - \frac{2r}{r+1}$, and if $\alpha = 1$, then $\tau = \frac{2r}{r+1} - 1$ for a minimal volume ellipsoid. These results are equivalent to those using the Khachian algorithm in [41]; that is, the OVE algorithm reduces to the Khachian algorithm for the case of one hyperplane (constraint).

4. Define the parameter $\rho$ as

$$\rho = \frac{\text{vol}(\hat{E})}{\text{vol}(\hat{S})}$$

(2.45)
\[ \Delta_{k+1} = \sqrt{(1 - \sigma)\delta} \] 

(2.46)

If \( \rho \geq 1 \), then \( \hat{E} \) has a larger (or the same) volume than \( \hat{S} \); that means we may not be justified to use the ellipsoid \( \hat{E} \) to replace \( \hat{S} \) to bound the set in interest. That is, the ellipsoid is updated only if \( \rho < 1 \). In this sense, the volume of the recursive ellipsoids are non-increasing, and because volume is a positive quantity, the OVE algorithm in fact converges weakly.

5. There are two versions of ellipsoid algorithm for parallel cuts in the Operation Research literature, one of them is described in [44] where the proof has not been published, the other one appeared in [43] where an algebraic proof is given. For these two algorithms, the derivation is essentially focused on linear programming problems. In this chapter, motivated by parameter set estimation of systems with bounded noise, the ellipsoid algorithm with parallel cuts is derived again using a geometric approach. In addition, the OVE algorithm discusses also the case when any one of the two parallel hyperplanes does not cut an ellipsoid, while the algorithms in [44] and in [43] do not. The proof of convergence of the OVE algorithm will be given in the next chapter.

\textbf{Theorem 2.3 (The OVE Algorithm):} For the system in (2.4) with the bounded noise constraint (2.5), the equations for \( \theta_{k+1} \) and \( P_{k+1} \) that result in the minimal volume ellipsoid \( E_{k+1} \) bounding the intersection between the given \( \mathcal{F}_{k+1} \) and
$E_k$ defined in Equations (2.7) and (2.8), respectively, are as follows:

\begin{align*}
\theta_{k+1} &= \theta_k + \frac{\tau_k P_k \phi_{k+1}}{(\phi_{k+1}^T P_k \phi_{k+1})^{1/2}} \\

P_{k+1} &= \delta_k \left( P_k - \sigma_k \frac{P_k \phi_{k+1} \phi_{k+1}^T P_k}{\phi_{k+1}^T P_k \phi_{k+1}} \right) ,
\end{align*}

where $\tau_k$, $\delta_k$, $\sigma_k$ are, respectively, equivalent to $\tau$, $\delta$ and $\sigma$ in Equations (2.27)-(2.29) or (2.30)-(2.32) depending on the values of $\alpha$ and $\beta$ in Theorem 2. These values of $\alpha$ and $\beta$ are given by

\begin{align*}
\beta &= \frac{\gamma}{\sqrt{\phi_{k+1}^T P_k \phi_{k+1}}} \\
\alpha &= \frac{y_{k+1} + \gamma - \phi_{k+1}^T \theta_k}{\sqrt{\phi_{k+1}^T P_k \phi_{k+1}}} .
\end{align*}

If $\alpha > 1$, reset $\beta$ to $\beta - \frac{\alpha - 1}{2}$ and then reset $\alpha$ to 1; on the other hand, if $2\beta - \alpha > 1$, reset $\beta$ to $\frac{1 + \alpha}{2}$. The algorithm can be initiated with a sufficiently large $E_0$ containing the feasible parameter set, where $\theta_0 = 0$ and $P_0 = \frac{1}{\epsilon} I$ (with $0 < \epsilon \ll 1$) are typical starting values.

**Proof of Theorem 2.3:** Consider an affine transformation, $\theta = \theta_k + J\hat{\theta}$ where $P_k = JJ^T$. Let $P_{k+1} = J\hat{A}J^T$ and $\phi_{k+1} = J^{-T}\hat{\phi}$. For this affine transformation, $E_k$ is mapped to $\hat{E}$ and $E_{k+1}$ is mapped to $\hat{E}$. From Theorem 2, $\hat{\theta}_0 = \tau \hat{\phi}/(\hat{\phi}^T \hat{\phi})^{1/2}$ is transformed using $\hat{\theta} = J^{-1}(\theta - \theta_k)$ to

\begin{equation}
J^{-1}(\theta_{k+1} - \theta_k) = \tau_k J^T \phi_{k+1}/(\phi_{k+1}^T P_k \phi_{k+1})^{1/2} .
\end{equation}
or

\[ \theta_{k+1} = \theta_k + \frac{\tau_k P_k \phi_{k+1}}{(\phi_{k+1}^T P_k \phi_{k+1})^{1/2}} \]  

and \( \hat{A} = \delta(I - \sigma \hat{\phi}^T \hat{\phi}^T) \) is transformed to

\[ J \hat{A} J^T = \delta_k(J J^T - \frac{\sigma_k J J^T \phi_{k+1} \phi_{k+1}^T J J^T}{\phi_{k+1}^T P_k \phi_{k+1}}) \]

or

\[ P_{k+1} = \delta_k(P_k - \frac{\sigma_k P_k \phi_{k+1} \phi_{k+1}^T P_k}{\phi_{k+1}^T P_k \phi_{k+1}}) \]  

Since \( \theta^T \phi_{k+1} \leq y_{k+1} + \gamma \), the transformation results in \( \hat{\theta}^T \hat{\phi} \leq y_{k+1} + \gamma - \theta_k \phi_{k+1} \) and therefore \( \alpha = \frac{y_{k+1} + \gamma - \theta_k \phi_{k+1}}{\sqrt{\phi_{k+1}^T P_k \phi_{k+1}}} \). Similarly, \( \alpha - 2\beta = \frac{y_{k+1} - \gamma - \theta_k \phi_{k+1}}{\sqrt{\phi_{k+1}^T P_k \phi_{k+1}}} \) gives \( \beta = \frac{\gamma}{\sqrt{\phi_{k+1}^T P_k \phi_{k+1}}} \).

This ends the proof for Theorem 2.3, which completes the construction for the OVE algorithm.

**Remark:** Both the OBE and the OVE algorithms can handle the case of a time varying noise bound \( \gamma_k \).

### 2.4 Comparison with the OBE Algorithm

The problem considered in this chapter was first studied in 1982 by Fogel and Huang[5], and was further refined in [28]. The idea taken here is similar to that in [5, 28], that is to find a smallest volume ellipsoid \( E_{k+1} \) to contain the set \( E_k \cap F_{k+1} \). The work in [5, 28] is based on the result in [46] to find an ellipsoid to contain the intersection of another two ellipsoids, and a modified RLS type update of the center of ellipsoids is
adopted. The derivation of the OVE algorithm takes on a different approach based on a geometrical point of view.

The OBE algorithm can be summarized as follows, using the notation from the last section,

\[
\hat{\delta}_{k+1} = y_{k+1} - \theta_k^T \phi_{k+1}
\]

\[
\hat{\sigma}_{k+1}^2 = 1 + \lambda_{k+1} \gamma^2 - \frac{\lambda_{k+1} \hat{\sigma}_{k+1}^2}{1 + \lambda_{k+1} \phi_{k+1}^T P_k \phi_{k+1} \hat{\sigma}_{k+1}^2}
\]

\[
P_{k+1} = \hat{\sigma}_{k+1}^2 (P_k - \frac{\lambda_{k+1} P_k \phi_{k+1} \phi_{k+1}^T P_k}{1 + \lambda_{k+1} \phi_{k+1}^T P_k \phi_{k+1}})
\]

\[
\theta_{k+1} = \theta_k + \frac{\lambda_{k+1}}{\hat{\sigma}_{k+1}^2} P_{k+1} \phi_{k+1} \hat{\delta}_{k+1}
\]

(2.51)

where \(0 < \lambda_{k+1} < \infty\) is a design parameter used to minimize the volume of \(E_{k+1}\) subject to the constraint that \(\theta_k\) is a modified RLS estimate.

Compared to the \(\theta_{k+1}\) and \(P_{k+1}\) in the last section, it is seen that \(P_{k+1}\) takes on a similar form in both algorithms, that is the update of \(P_{k+1}\) is the same if

\[
\sigma_k = \hat{\sigma}_k^2
\]

\[
\frac{\sigma_k}{\phi_{k+1}^T P_k \phi_{k+1}} = \frac{\lambda_{k+1}}{1 + \lambda_{k+1} \phi_{k+1}^T P_k \phi_{k+1}}
\]

The similarities in the form of the expression for \(P_{k+1}\) indicate that \(E_{k+1}\) in both the OVE and OBE algorithms have the same orientation with one of the axes parallel to \(\hat{\phi}\) in the affine transformed coordinate, given the same \(E_k\). However, \(\theta_{k+1}\) in the OBE algorithm depends on \(P_{k+1}\), \(\phi_{k+1}\) and the prediction error \(\hat{\delta}_{k+1}\) (essentially a modified RLS estimate), while \(\theta_{k+1}\) in the OVE algorithm depends explicitly on \(P_k\) and \(\phi_{k+1}\) only. From the geometrical viewpoint, if \(E_k\) is mapped to a unit radius hypersphere
\( \hat{S} \) and \( E_{k+1} \) to \( \hat{E} \) through an affine transformation, it is seen that the center of ellipsoid \( \hat{E} \) in the OBE algorithm does not necessarily lie on the vector \( \hat{\phi} \) because of the dependence of \( \theta_{k+1} \) on \( P_{k+1} \), whereas in the OVE algorithm the center of \( \hat{E} \) always lies on \( \hat{\phi} \). The main difference, and this is of paramount importance, is the location of the center of the ellipsoid. To see this, consider a two dimensional case for simplicity.

The smallest area ellipse that contains the intersection between a unit radius circle and two parallel lines must pass through the intersecting points. In addition, because of inherent symmetry, it is easy to see that the center of the bounding ellipse must lie on the vector which passes through the origin and is orthogonal to the two parallel lines. It is therefore obvious that the OBE algorithm will surely result in a larger volume ellipsoid than the OVE algorithm. This is because of the extra RLS type constraint that is imposed on the center estimate of the ellipsoids in the OBE algorithm; this constraint essentially precludes the satisfaction of the necessary condition for a minimal volume ellipsoid \( E_{k+1} \) to contain \( E_k \cap F_{k+1} \), where \( E_k \cap F_{k+1}^* = E_{k+1} \cap F_{k+1}^* \) in which \( F_{k+1}^* \) is the boundary of \( F_{k+1} \).

The OVE algorithm also applies to the case when one of the hyperplanes does not cut the recursive ellipsoid which may be crucial when the goal is to find the smallest set. Such case is also considered in the Modified OBE (MOBE) algorithm in [37, 38]. Essentially, OBE and MOBE are equivalent except when one of the hyperplanes does not cut the recursive ellipsoid. The updating process in the OBE algorithm is explicitly governed by the parameter \( \lambda_k \) which is dependent on the prediction error and the noise bound. In the OVE algorithm, the updating decision is totally governed
by the ratio \( p \). If \( p < 1 \), the ellipsoid is updated and the volume of the ellipsoids are guaranteed to be monotonic non-increasing. In other words, the OBE algorithm has the RLS features while the OVE algorithm is derived entirely from a geometrical point of view. Both the OVE and OBE algorithms update estimates selectively according to the received data.

As a final note on comparing the OBE and the OVE algorithms, the numbers of multiplication and addition operations required in the information evaluation are on the order of \( r^2 \) for both algorithms where \( r \) is the number of unknown system parameters. However, for updating estimates the OBE algorithm requires \( 6r^2 \) multiplications whereas in the OVE algorithm, only \( 5r^2 \) multiplications are required (and no more additions are required for the OVE algorithm). The extra computation for the OBE algorithm is basically due to the need for \( P_{k+1} \) in (2.51). The performance of both algorithms is compared by way of examples in the next section.

2.5 Computer Simulation

2.5.1 Example 1

Consider a simple second order system to illustrate the geometrical properties of both the OVE and OBE algorithms,

\[
Y(z) = \frac{0.3z^{-1}}{1 - 0.5z^{-1}} U(z) + \frac{1}{1 - 0.5z^{-1}} V(z)
\]

with the true parameter \( \theta_{true} = [-0.5 \quad 0.3]^T \). In the simulation, \(|v(k)| \leq 0.1\) with \( S/N=24 \) dB and 50 data records are simulated. The same data set is used in both the OVE and OBE algorithms. Figure 6 shows six pairs of ellipses for \( k = 0, \cdots, 5 \) for
both algorithms; note that the axis scaling differs only for the $k = 0$ and $k = 1$ cases when the change in volume is most dramatic. Pictorially, it is seen that the OVE algorithm gives smaller area ellipses in bounding the parameter set. More recursive ellipses are shown in Figure 7 to illustrate the progress of both algorithms, up to the $k = 50$ case.

2.5.2 Example 2

Consider a second order system which represents a flexible structure truss model containing only the first x-bending mode[47] with the following discretized transfer function:

$$Y(z) = \frac{0.1156(z^{-1} - z^{-2})}{1 - 1.55z^{-1} + 0.8267z^{-2}} U(z) + \frac{1}{1 - 1.55z^{-1} + 0.8267z^{-2}} V(z). \quad (2.53)$$

In the simulation, $|v_k| \leq 0.05$ and the S/N is 20dB, $N = 100$ data points are taken. To compare the OVE and OBE algorithms, the same input sequence and noise sequence are used in the two simulations. The following notation is adopted: $\theta_{true}$ is the true estimate, $\theta^{OBE}_k$ is the center estimate of the ellipsoid associated with the OBE algorithm and $\theta^{OVE}_k$ is the center estimate of the ellipsoid associated with the OVE algorithm at time $k$; the parameter interval associated with the ellipsoids are denoted as $I^{OBE}$ and $I^{OVE}$ accordingly for the final ellipsoids. The results of the two algorithms are as follows:

$$\begin{align*}
\theta_{true} &= \begin{bmatrix} -1.55 \\ 0.8267 \\ 0.1156 \\ -0.1156 \end{bmatrix}, \\
\theta^{OBE}_N &= \begin{bmatrix} -1.538 \\ 0.8226 \\ 0.1155 \\ -0.1158 \end{bmatrix}, \\
\theta^{OVE}_N &= \begin{bmatrix} -1.545 \\ 0.8244 \\ 0.1157 \\ -0.1182 \end{bmatrix} \quad (2.54)
\end{align*}$$
Figure 6: The geometry of ellipses at $k = 0, \ldots, 5$
Figure 7: The geometry of the ellipses at $k = 10, 15, 20, 30, 40, 50$
and the parameter intervals are

\[
\delta_{\text{OBE}} = \begin{bmatrix}
-1.72 & \leftrightarrow & -1.356 \\
0.6702 & \leftrightarrow & 0.9751 \\
0.1004 & \leftrightarrow & 0.1311 \\
-0.1405 & \leftrightarrow & -0.0905
\end{bmatrix},
\delta_{\text{OVE}} = \begin{bmatrix}
-1.657 & \leftrightarrow & -1.434 \\
0.7213 & \leftrightarrow & 0.927 \\
0.1062 & \leftrightarrow & 0.1253 \\
-0.1347 & \leftrightarrow & -0.1016
\end{bmatrix}.
\]  

(2.55)

Note that both of the intervals contain the true parameters. Figure 8 shows the volume of ellipsoids \(E_{k}^{\text{OVE}}, E_{k}^{\text{OBE}}\) and \(E_{k}^{\text{MOBE}}\) due to the OVE algorithm, the OBE algorithm, and the Modified OBE algorithm, respectively, over 100 data points. It is clear from the figure that ellipsoids from the OVE algorithm are always smaller than that from the OBE algorithm (they are always less than half of the volume of the OBE ellipsoids). Also, the MOBE algorithm improves over the OBE algorithm only slightly, and the size of ellipsoids due to the MOBE algorithm are still much bigger than those due to the OVE algorithm. Nevertheless, all three algorithms guarantee monotonic non-increasing volume of the ellipsoids, consistent with the theory.

Moreover, from the viewpoint of the parameter interval by putting an orthotope which is orthogonal to the parameter axes and tightly overbound the ellipsoid, Figure 9 shows the interval for each parameter at each iteration for the two algorithms. In the plot, the upper and lower intervals of each parameter are indicated by the arrows for each algorithm. It can be seen that the OVE algorithm almost always gives tighter parameter bounds than does the OBE algorithm. Exceptions to this occur when the size of some of the axes of an ellipsoid are reduced, while other axes are expanded to contain a certain convex set, possibly causing a larger bound in some of the parameters at early iterations. Eventually, tighter bounds are noted in the recursion as in Figure 9 and Equation (2.55).
Figure 8: Volume of $E_k^{OVE}$, $E_k^{OBE}$ and $E_k^{MOBE}$
Figure 9: Parameter intervals for both algorithms
2.6 Conclusion

In this chapter, the Khachian ellipsoid algorithm for the linear programming problem is extended to the problem of parameter set estimation with bounded noise. We first showed that a minimal volume ellipsoid bounding the intersection between a hypersphere and two parallel hyperplanes must tightly contain the intersecting points between the surface of the hypersphere and the hyperplanes. Using this result, we derived the minimal volume ellipsoid containing the intersection between a hypersphere and two parallel hyperplanes, resulting in the new recursive algorithm, OVE, for parameter set estimation. It is noted that the OVE algorithm has similar form to the well known OBE algorithm. However, the OVE algorithm possesses several attractive features in comparison to the OBE algorithm:

1. the OVE algorithm is rich in geometrical interpretations;

2. the OVE algorithm is flexible enough to accommodate the case when one of the hyperplanes defined in $\mathcal{F}_{k+1}$ does not cut the ellipsoid $E_k$;

3. the OVE algorithm requires no additional computational complexity than the OBE algorithm;

4. The OVE algorithm results in the smallest volume ellipsoid $E_{k+1}$ bounding $E_k \cap \mathcal{F}_{k+1}$ without any constraints.

The OVE algorithm derived in this chapter is equivalent to the results in [43] and [44] except that the OVE algorithm
1. discusses also the case when any one of the hyperplanes does not cut an ellipsoid;

2. offers correct development for the parameter set estimation problem with examples and convergence proof;

3. offers geometrical insight into the problem.
CHAPTER III

CONVERGENCE ANALYSIS OF THE OVE ALGORITHM

3.1 Introduction

In the last chapter, the OVE algorithm for parameter set estimation has been developed. In this chapter, the convergence properties of the OVE algorithm will be discussed. Before doing so, in the following the importance of estimation convergence from a control standpoint is motivated.

The demands of a rapidly growing technology for faster and more accurate control of systems with increasing complexity have had a strong influence on the progress of automatic control theory development. An adaptive system is one feasible solution which provides a means of continuously monitoring its own performance and adjusting its own parameters to give optimal performance. A feedback controller can be incorporated and made intelligent enough to modify its characteristics in a changing environment so as to operate in an optimum manner according to some specified criterion.

It can be shown that in indirect adaptive control, convergence of the parameter estimator is essential in establishing a stable closed loop system[48]. Lozano-Leal and Goodwin[49] provide a proof of global stability for an adaptive pole-placement
scheme without requiring persistent excitation on plant signals; the convergence of the estimator is the foundation of the proof. For the bounded noise case, robust adaptive control has been studied by several authors; see [50] and the references therein. Their methodologies can be summarized as those formulated with a dead zone in the parameter estimation. Essentially, update of parameter estimate is ceased if the prediction error falls within the dead zone. From a different perspective, ellipsoid algorithms analogous to the recursive least square method are formulated in a set theoretic framework. An ellipsoid is used to bound the feasible parameter set consistent with the measurement corrupted by bounded noise, and the ellipsoid is updated recursively. Fogel and Huang[5] give an ellipsoid algorithm which continuously reduces the geometric size of the set, but no solid convergence proof is provided for the center estimates of the ellipsoids. Dasgupta and Huang[51], and Lozano-Leal and Ortega[52] give similar, but convergent ellipsoid algorithms for the estimation; however, the geometrical size of the ellipsoidal set bounding the feasible set is not reduced optimally at each recursion. In [53], Lozano-Leal and Collado give an improved and convergent algorithm; the traces of the matrices or the sum of the semi-axes associated with the recursive ellipsoids are non-increasing and convergent.

In the proceeding chapter, a new ellipsoid algorithm is derived from a geometrical view point, where the developed algorithm is optimal in the sense that at each recursion, the smallest volume ellipsoid set is found which bounds the parameter set consistent with the observations. However, no proof of any convergence properties is given there, and it is the purpose of this chapter to discuss the convergence properties
of the OVE algorithm. In the next section, a summary of the OVE algorithm will be given. The reason for repeating the OVE algorithm in this chapter is to group the problem and solution together for easy reference in the following sections. Also, the parameters for an ellipsoid are indexed for easy discussion. The main convergence results of the OVE algorithm will be given in Section 3.3; the behavior of some OVE parameters are demonstrated through simulations. In Section 3.4, an extended discussion on convergence issues regarding the feasible parameter set will be given. Some conclusions and remarks will be made in Section 5 to conclude this chapter.

3.2 Summary of the OVE Algorithm

Consider a SISO ARX model,

\[ y_k = -\sum_{i=1}^{n} a_i y_{k-i} + \sum_{j=0}^{m} b_j u_{k-j} + v_k \]  

(3.1)

\[ = \theta^T \phi_k + v_k \]  

(3.2)

where \( \theta^T = [a_1, \ldots, a_n, b_0, \ldots, b_m] \) is the unknown parameter vector;

\[ \phi_k = [-y_{k-1}, \ldots, -y_{k-n}, u_k, \ldots, u_{k-m}]^T \]

is the regression vector containing the past inputs \( u(\cdot) \) and past outputs \( y(\cdot) \); \( n, m \) are the number of system poles and zeroes known a priori; \( v_k \) is a sequence of bounded disturbances/noise corrupting the system output with \( |v_k| \leq \gamma \) for all \( k \geq 0 \).

Let \( \mathcal{F}^k \subset \mathbb{R}^r \) be the set of feasible parameter estimates of the plant which are consistent up to the \( k^{th} \) measurement and \( r := n + m + 1 \). That is,

\[ \mathcal{F}^k = \{ \theta : |y_i - \theta^T \phi_i| \leq \gamma, i = 0, \ldots, k \} . \]  

(3.3)
Define also
\[ \mathcal{F}_k = \{ \theta : |y_k - \theta^T \phi_k| \leq \gamma_k \} \] (3.4)
and
\[ E_k = \{ \theta : (\theta - \theta_k)^T P_k^{-1} (\theta - \theta_k); \ \theta \in \mathbb{R}^r \} \] (3.5)
where \( \theta_k \) is the center of an ellipsoid \( E_k \) containing \( \mathcal{F}_k \). The OVE algorithm finds the smallest volume ellipsoid \( E_{k+1} \) containing the intersection of \( \mathcal{F}_{k+1} \) and \( E_k \). The OVE algorithm has the following recursive equations:
\[ \theta_{k+1} = \theta_k + \frac{\tau_k P_k \phi_{k+1}}{(\phi_{k+1}^T P_k \phi_{k+1})^{1/2}} \] (3.6)
\[ P_{k+1} = \delta_k (P_k - \sigma_k \frac{P_k \phi_{k+1} \phi_{k+1}^T P_k}{\phi_{k+1}^T P_k \phi_{k+1}}) \] (3.7)
where if
(a) \( \alpha_k \neq \beta_k \), then
\[ \delta_k = \frac{(\tau_k + 1)^2 (\beta_k - \alpha_k) - \tau_k (1 + \alpha_k) (2 \beta_k - \alpha_k - 1)}{\tau_k + \beta_k - \alpha_k} \] (3.8)
\[ \sigma_k = \frac{-\tau_k}{\beta_k - \alpha_k} \] (3.9)
and \( \tau_k \) is the real solution of
\[ (r+1) \tau_k^2 + \left\{ \frac{(1 + \alpha_k)(\alpha_k - 2 \beta_k + 1)}{\beta_k - \alpha_k} + 2[r(\beta_k - \alpha_k) + 1] \right\} \tau_k + r \alpha_k (\alpha_k - 2 \beta_k) + 1 = 0 \] (3.10)
such that \( \alpha_k - 2 \beta_k < \tau_k < \alpha_k \);
(b) \( \alpha_k = \beta_k \), then
\[ \delta_k = \frac{r}{r-1} (1 - \beta_k^2) \] (3.11)
\[ \sigma_k = \frac{1 - r \beta_k^2}{1 - \beta_k^2} \] (3.12)
\[ \tau_k = 0 \] (3.13)
where $\alpha_k$ and $\beta_k$ are defined as

$$
\begin{align*}
\beta_k &= \frac{\gamma}{\sqrt{\phi_{k+1}^T P_k \phi_{k+1}}} \\
\alpha_k &= \frac{y_{k+1} + \gamma - \phi_{k+1}^T \theta_k}{\sqrt{\phi_{k+1}^T P_k \phi_{k+1}}}
\end{align*}
$$

(3.14) \hspace{1cm} (3.15)

If $\alpha_k > 1$, reset $\beta_k$ to $\beta_k - \frac{\alpha_k - 1}{2}$ and then reset $\alpha_k$ to one; on the other hand, if $2\beta - \alpha_k > 1$, reset $\beta_k$ to $\frac{1 + \alpha_k}{2}$. Further, if $\sigma_k < 0$, set $E_{k+1}$ to be $E_k$, that is the computed ellipsoid, $E_{k+1}$ will replace $E_k$ only if it has a smaller volume that of $E_k$. The algorithm can be initialized with a sufficiently large $E_0$ containing the feasible parameter set, where $\theta_0 = 0$ and $P_0 = \frac{1}{\epsilon}I$ (with $0 < \epsilon \ll 1$) are typical starting values. Note that, if $E_0$ contains the parameter set, so does all other $E_k$. It is therefore important to make sure that $E_0$ is large enough so that the new $E_k$ computed are meaningful.

3.3 The Main Convergence Properties of the OVE Algorithm

Theorem 3.1: Consider the Optimal Volume Ellipsoid Algorithm summarized in the last section, and define the Lyapunov function

$$
V_k \equiv \det P_k
$$

and

$$
\tilde{\theta}_k \equiv \theta^* - \theta_k
$$

where $\theta^*$ is the true parameter vector. Then
(i) $V_{k+1} \leq V_k$

(ii) $(1 - \sigma_{k-1})e_k \leq \gamma$ if $\theta_k \neq \theta_{k-1}$ where $e_k = y_k - \theta_{k-1}^T \phi_k$

(iii) $|e_k| \leq \gamma$ as $k \to \infty$

(iv) $\|\tilde{\theta}_k\|_2^2 \leq \tilde{\sigma}(P_k)$ where $\tilde{\sigma}(P_k)$ denotes the largest singular value of $P_k$

(v) $\delta_k \geq 1$ as $k \to \infty$

(vi) If there exists an integer $N_1$ ($\geq r$, the number of unknown parameters), and positive real numbers $c_1$ and $c_2$ such that

$$c_1 I \leq \sum_{i=k+1}^{k+N_1} \phi_i \phi_i^T \leq c_2 I$$

for all $k$, then

$$\|\hat{\theta}_* - \theta_k\|_2^2 \leq \frac{4\gamma^2 N_1}{c_1}$$

as $k \to \infty$.

Property (i) essentially means that the volume of the ellipsoids are non-increasing. Property (ii) says that if there is a parameter update, then the square of the ratio between the lengths of the ellipsoidal semi-axis of $E_k$ along and orthogonal to the data vector $\phi_k$, in the affine transformed coordinate with respect to $E_{k-1}$, multiplied with the prediction error should be less than or equal to the noise bound. Statement (iii) implies that, in the limit, the prediction error will be no bigger than the noise bound. Property (iv) says that the norm of the error between the $k^{th}$ center estimate
and the true parameter vector is no greater than the square of the longest semi-axis
of $E_k$. Property (v) states that the square of the length of the ellipsoidal semi-axes
in $E_{k+1}$ orthogonal to the data vector $\phi_{k+1}$ must be greater than 1 when the ellipsoid
$E_k$ is transformed to a unit hypersphere. The last statement (vi) in the theorem
essentially says that the center estimate will converge to a "ball" centered at the true
parameter vector with a radius $2\gamma \sqrt{\frac{N}{c_1}}$.

Proof of Theorem 3.1:

(i) This is a direct result from the OVE algorithm in which the ellipsoid $E_{k+1}$ has
a smaller volume than $E_k$. Since the volume of an ellipsoid $E_k$ is proportional to
$\sqrt{\det P_k}$, the result follows immediately. That is, $V_k$ is non-increasing. Since the
volume of an ellipsoid is a non-negative number, this implies $V_k$ will also converge to
some non-negative number.

(ii) Define $G_k = \phi_k^T P_{k-1} \phi_k (> 0)$ and use the definition for $\alpha_k$ and $\beta_k$
to get $\tau_{k-1} = \frac{\sigma_{k-1} e_k}{\sqrt{\sigma_k}}$. Then by transposing the recursive equation for $\theta_k$ and post-multiply it by $\phi_k$
on both sides, we get

$$\theta_k^T \phi_k = \theta_{k-1}^T \phi_k + \frac{\sigma_{k-1} e_k}{G_k} \phi_k^T P_{k-1} \phi_k = \theta_{k-1}^T \phi_k + \sigma_{k-1} e_k$$

$$\Rightarrow y_k - \theta_k^T \phi_k = y_k - \theta_{k-1}^T \phi_k - \sigma_{k-1} e_k$$

$$\Rightarrow y_k - \theta_k^T \phi_k = (1 - \sigma_{k-1}) e_k.$$ 

Since $\theta_k$ must lie inside the hyperplanes defined by $y_k$ and $\phi_k$ if $\theta_k$ is a new update of
$\theta_{k-1}$ that $\theta_k \neq \theta_{k-1}$, this implies $|y_k - \theta_k^T \phi_k| \leq \gamma$ and the result follows immediately.

(iii) Since $V_k$ is non-increasing and convergent, there exists a sufficiently large integer
\( M \) such that

\[
det P_{M+1} = det P_M - \epsilon
\]

where \( \epsilon \) is an arbitrarily small positive real number. Now assume that the center of \( E_M, \theta_M \), does not lie inside the two hyperplanes defined by \( y_{M+1} \) and \( \phi_{M+1} \), then

\[
|e_{M+1}| = |y_{M+1} - \theta_M^T \phi_{M+1}| > \gamma.\]

This implies that \( \theta_k \) is located outside the intersecting region between \( E_M \) and \( F_{M+1} \) where \( F_{M+1} \) defines the region between the two hyperplanes consistent with \( y_{M+1} \) and \( \phi_{M+1} \); a two dimensional case is shown in Figure 10. As \( |y_{M+1} - \theta_M^T \phi_{M+1}| > \gamma \), geometrically the intersecting region \( F_{M+1} \cap E_M \) is located inside one half of the ellipsoid \( E_M \). However, according to the Khachian Ellipsoid Algorithm[36], the smallest volume ellipsoid containing one half of \( E_M \) must have a straightly smaller volume than \( E_M \). Therefore, we have \( vol(E_{M+1}) < vol(E_M) \)

where \( vol(\cdot) \) denotes the volume and \( E_{M+1} \) is the smallest volume ellipsoid containing \( F_{M+1} \cap E_M \). In other words, we have shown that, if \( \theta_M \) does not lie in \( F_{M+1} \cap E_M \), then \( vol(E_{M+1}) \) and \( vol(E_M) \) can not be arbitrarily close to each other. As a result, if \( det P_{M+1} \) and \( det P_M \) are arbitrarily close, \( \theta_M \) must necessarily be located inside of \( F_{M+1} \). Therefore, \( |e_{M+1}| \leq \gamma \).

(iv) Since for all \( \theta \in E_k \), \( (\theta - \theta_k)^T P_k^{-1} (\theta - \theta_k) \leq 1 \) and that \( \theta^* \in E_k \), this implies

\[
(\theta^* - \theta_k)^T P_k^{-1} (\theta^* - \theta_k) \leq 1
\]

\[
\Rightarrow \sigma(P_k)^2 \leq \tilde{\theta}_k^T P_k^{-1} \tilde{\theta}_k \leq 1
\]

\[
\Rightarrow \frac{1}{2} \leq \sigma(P_k)
\]

(v) By putting \( \beta_k = \frac{\gamma}{\sqrt{G_{k+1}}} \), \( \alpha_k = \frac{\tau_{k+1}}{\sqrt{G_{k+1}}} \) and \( \tau_k = \frac{\sigma_k \tilde{e}_{k+1}}{\sqrt{G_{k+1}}} \) into the expression for \( \delta_k \),
we get

\[ \delta_k = 1 + \frac{\sigma_k^2 e_{k+1}^2 + \sigma_k (\gamma^2 - e_{k+1}^2)}{(1 - \sigma_k) G_{k+1}}. \]

Since from the OVE algorithm, \( 1 > \sigma_k \geq 0 \) and from (iii), \( |e_k| \leq \gamma \), result (v) is obtained when \( k \to \infty \). For the \( \alpha_k = \beta_k \) case, since \( \sigma_k \geq 0 \), then from the equation for \( \sigma_k \), one gets

\[ \frac{1 - \beta_k^2}{r - 1} \geq \beta_k^2, \]

and from the equation for \( \delta_k \), one gets

\[ \delta_k = 1 - \beta_k^2 + \frac{1 - \beta_k^2}{r - 1} \geq 1 - \beta_k^2 + \beta_k^2 \]
\[ = 1 \]

for all \( k \).
(vi) To show the property that \( \theta_k \) will converge to some region around the true parameters, we first see that when \( k \) is sufficiently large, there exists a positive integer \( N_2(\geq N_1) \), such that

\[
|y_{k+i} - \theta_k^T \phi_{k+i}| \leq \gamma ,
\]

where \( 1 \leq i \leq N_2 \). Such \( N_2 \) exists due to the convergence of the volume of ellipsoids, which means that, for some \( k_1 \) sufficiently large such that ellipsoids converge (arbitrarily close), no new information can be obtained for \( k \in [k_1, k_1 + N_2] \) in regards to the parameter set. Furthermore, from the assumption on the noise bound, we have

\[
|y_{k+i} - \theta^T \phi_{k+i}| \leq \gamma ,
\]

and we can combine these two properties to give

\[
|(\theta^* - \theta_k)^T \phi_{k+i}| \leq 2\gamma .
\]

Define

\[
\Phi \equiv \begin{bmatrix}
\phi_{k+1} & \phi_{k+2} & \cdots & \phi_{k+N_1}
\end{bmatrix} ,
\]

then we get

\[
(\theta^* - \theta_k)^T \Phi \Phi^T (\theta^* - \theta_k) = \begin{bmatrix}
2\gamma & 2\gamma & \cdots & 2\gamma \\
2\gamma & 2\gamma & \cdots & 2\gamma \\
\vdots & \vdots & \ddots & \vdots \\
2\gamma & 2\gamma & \cdots & 2\gamma
\end{bmatrix} .
\]

Since

\[
\Phi \Phi^T = \sum_{i=k+1}^{k+N_1} \phi_i \phi_i^T ,
\]
it follows that

\[ \|\theta^* - \theta_k\|_2^2 \leq \frac{4\gamma^2 N_1}{\sigma(\Phi \Phi^T)} \leq \frac{4\gamma^2 N_1}{c_1} \]

This completes the proof of Theorem 3.1. □

As an illustration, Example 2 with \( \gamma = 0.05 \) in Chapter 2 is simulated to demonstrate the convergence behavior. The monotonic decreasing and convergent behavior of \( \det P_k \) is shown in Figure 11. Figure 12 shows the property (ii) in Theorem 3.1. In Figure 13, it is seen that the prediction error is actually inside the noise bound for large \( k(> 600) \). Figure 14 shows the behavior of the sequence \( \|\hat{\theta}_k\|_2 \).

Comments:

1. The convergence properties of the OVE algorithm remain when the noise bound \( \gamma \) is time varying. This is true because the crucial element in the convergence of the OVE algorithm for the constant noise bound case is the decreasing volume of the recursive ellipsoids, and this property carries over even if the noise bound is time varying.

2. The parameter set estimator using the OVE algorithm can be applied to robust control design with parametric uncertainty, and can also be used in an adaptive control setting to establish a stable closed loop adaptive system.
Figure 11: Behavior of $\text{det} P_k$

Figure 12: Behavior of $(1 - \sigma_k) e_k$
Figure 13: Behavior of $e_k$

Figure 14: Behavior of $\|\tilde{\theta}_k\|_2$
3.4 Other Convergence Issues

3.4.1 The Feasible Parameter Set

In this section, we show via an example that \( \mathcal{F}^k \) can reduce to a point for some \( k \) provided that the unknown noise reaches the limit frequently enough. In other words, it is possible that \( \mathcal{F}^k \), the intersection of the pairs of hyperplanes, converges to or in fact is a point provided that the noise behavior is such that it hits the bound frequently enough, that is, \( |v_k| = \gamma \) for a sufficient number of times. This behavior will be illustrated by means of an example.

Consider the following system,

\[
y_k = b_1 u_k + a_1 y_{k-1} + v_k
\]  

(3.16)

with \( |v_k| \leq 0.1 \), \( b_1 = 1 \) and \( a_1 = -0.5 \); let \( y_0 = 0 \), \( u_0 = 0 \), \( u_1 = 1 \), \( u_2 = 0.5 \), \( u_3 = -0.5 \), \( u_4 = -1 \) and \( v_0 = 0 \), \( v_1 = 0.05 \), \( v_2 = 0.1 \), \( v_3 = -0.1 \), \( v_4 = 0.1 \). Note that \( |v_k| = 0.1 \) for \( k=2,3,4 \). Figure 15 illustrates the feasible region of \( \mathcal{F}_k \) for \( k = 1,2,3,4 \), and it is clear that \( \mathcal{F}^4 \) is just a point, the true estimate.

3.4.2 The Optimal Bounding Set

In the last section, the convergence of the OVE algorithm is discussed. However, the ellipsoid \( E_k \), resulting from the algorithm may not be the smallest one enclosing \( \mathcal{F}^k \). In other words, the OVE algorithm is optimal only in a certain sense, that is \( E_k \) is the smallest volume ellipsoid enclosing \( E_{k-1} \cap \mathcal{F}_k \). A simple illustration is given in Figure 16 and it is seen that the OVE algorithm will not give an optimal bounding set. In fact, the ellipsoids will never converge to a point even though \( \mathcal{F}^k \) is actually
Figure 15: An example of $T^k$ converged to a point

a point as illustrated in the last example. This could be the case because as long as the system is sufficiently excited, the two parallel hyperplanes consistent with the observation at any time instance are always at some finite non-zero distance apart. This can be seen from the fact that the two parallel hyperplanes are $2\beta_k$ apart where

$$\beta_k = \frac{\gamma}{\sqrt{\phi_{k+1}P_k\phi_{k+1}}}$$

As motivated in the last chapter, the main reason for using ellipsoids to bound the feasible parameter set is that it is easy to manipulate ellipsoids in implementation. There are other alternatives to bound the feasible set, such as the use of “boxes”. However, the kind of conservatism illustrated in Figure 16 still exists if “boxes” are used. In fact, there is no recursive solution available that is optimal in the sense that the feasible set is tightly bounded. From the viewpoint of batch implementation,
Figure 16: An illustration of the conservatism of $E_k$ relative to $\mathcal{F}_k$
the non-recursive solution in [39] does give an optimal box that bounds the feasible parameter set "tightly". In that case, if the feasible set is a point, the "box" reduces to a point as well. The disadvantages of the scheme in [39] are that the algorithm is off-line and cumbersome to implement and that all measurements must be available prior to the implementation of the algorithm.

3.4.3 Computational Issues

The Khachian's ellipsoid algorithm that received most attention in the operation research literature was shown to exhibit numerical problems[45]. It is not the intent here to repeat the numerical analysis, but to provide strategies to overcome the numerical problems that are also anticipated in using the OVE algorithm. To make sure that ellipsoids actually contain the feasible parameter set when there are numerical approximations, it is suggested in [45] that at each recursion, the computed ellipsoid should be blown up to compensate for the possible translation of the actual ellipsoid. On the other hand, one can expand the distance between the two parallel hyperplanes to counteract the effect of numerical inaccuracies.

In the case when noise $\gamma$ is small, the OVE algorithm behaves nicely as long as $\gamma$ is sufficiently larger than the machine precision of the digital computer. When the first example in Section 2.5 is considered with $a_1 = -0.5$, $b_1 = 0.3$, and with $\gamma = 0.0005$, the simulation result gives $\theta_{10} = [-0.5 \ 0.3]^T$ which is equal to the true estimate up to the fourth decimal place, and the corresponding $P_{10}$ is given as

$$P_{10} = \begin{bmatrix} 2.52 \times 10^{-7} & 1.03 \times 10^{-8} \\ 1.03 \times 10^{-8} & 2.19 \times 10^{-7} \end{bmatrix}.$$
This example clearly shows that if $\gamma$ is small, the result of the OVE algorithm will be close to the true estimate, as expected.

### 3.4.4 The Noise Free Case ($\gamma = 0$)

If $\gamma = 0$, then the two parallel hyperplanes consistent with the observation are degenerated to one. The OVE algorithm can be modified slightly to fit this situation and the OVE algorithm will result in parameter estimates which are exactly the true parameters. Note that the algorithm is equivalent to a special case of the well known Least Square estimation.

**Theorem 3.2:** When $\gamma = 0$, $\delta_k$, $\sigma_k$ and $\tau_k$ will take on the following values for the smallest volume ellipsoid in the OVE algorithm:

\[
\delta_k = 1 - \alpha_k^2 \tag{3.17}
\]

\[
\sigma_k = 1 \tag{3.18}
\]

\[
\tau_k = \alpha_k . \tag{3.19}
\]

**Proof of Theorem 3.2:**

In the affine transformed coordinate, $E_k$ can be represented as a hypersphere,

\[
E_k = \{ \theta : \theta_1^2 + \theta_2^2 + \cdots + \theta_n^2 \leq 1, \theta = [\theta_1, \theta_2, \cdots, \theta_n]^T \in \mathbb{R}^n \} . \tag{3.20}
\]

Since $\gamma = 0$ and $E_k$ is invariant to all axes, by putting $\theta_1 = \alpha_k$ into (3.20), one gets the surface of the ellipsoid $E_{k+1}$ as

\[
\theta_2^2 + \theta_3^2 + \cdots + \theta_n^2 = 1 - \alpha_k^2
\]
or

\[ \frac{\theta_2^2}{1 - \alpha_k^2} + \frac{\theta_3^2}{1 - \alpha_k^2} + \cdots + \frac{\theta_n^2}{1 - \alpha_k^2} = 1, \]

then the ellipsoid \( E_{k+1} \), in the affine transformed coordinate, is a hypersphere with radius \( \sqrt{1 - \alpha_k^2} \). As a result, \( \delta_k = 1 - \alpha_k^2 \). The result \( \sigma_k = 1 \) and \( \tau_k = \alpha_k \) are obvious because the hypersphere \( E_{k+1} \) has one dimension less than \( E_k \) and it lies on the degenerate hyperplane which is located at \( \theta_1 = \alpha_1 \). This completes the proof of Theorem 3.2.

\[ \square \]

**Corollary 3.2:** For \( \gamma = 0 \), the modified OVE algorithm given in Theorem 3.2 requires only \( r \) sets of measurement data to compute the true system parameters. Moreover, \( P_k \) becomes positive semi-definite in this case.

**Proof of Corollary 3.2:**

Since \( \gamma = 0 \), the feasible parameter set \( F_k \) consistent with the \( k^{th} \) set of measurement data is reduced to one hyperplane. In other words, one set of measurement data defines one linear constraint for the feasible parameters. As one of these linear constraints is added to an ellipsoid each time, the newly formed constrained ellipsoid will have one dimension less. This can be easily seen when an ellipsoid is transformed to a hypersphere, say \( E_{k-1} \), then a new ellipsoid, \( E_k \), will also be a hypersphere which is formed by constraining the hypersphere \( E_{k-1} \) on a hyperplane. However, ellipsoid \( E_k \) must have a one dimension less than that of \( E_{k-1} \). As there are only \( r \) unknown parameters, one needs only \( r \) independent linear constraints to reduce the parameter
set to a point which is the true system parameter vector. The fact that $P_k$ is positive semi-definite is obvious since an ellipsoid constrained on a hyperplane is another ellipsoid with one dimension less, which is equivalent to the claim that the matrix $P_k$ is rank deficient or positive semi-definite in the ellipsoid case. This completes the proof of Corollary 3.2. □

To avoid any singularities in the computation when using the OVE algorithm, $\phi_{k+1}^T P_k \phi_{k+1} > 0$ must be enforced; this should not be a problem in general if the input is sufficiently exciting[1]. As before, the first example in Section 2.5 was simulated and the exact estimate, to the machine precision, $\theta_2 = [-0.5 \ 0.3]^T$ was computed to give a null volume ellipsoid, where the associated matrix has entries on the order of machine precision.

Remark: In the case when the time varying noise bound $\gamma_k$ goes to zero in a finite number of steps, then $r$ steps later, the true parameters can be computed.

3.5 Conclusions

In this chapter, the OVE algorithm has been summarized and the convergence properties of the OVE algorithm have been analyzed. It is found that the OVE estimator algorithm possesses the convergence properties essential for stable indirect adaptive control application[52]. The most important of all are the convergence of the output prediction error to within the noise bound ((iii) of Theorem 3.1) and the convergence of the center estimate to a region around the true parameter ((vi) of Theorem 3.1). The convergence properties are also substantiated via an example. Further discussion on the behavior of the feasible parameter set is given via an example. It is found that
the feasible parameter set defined by the noise bound and the measurements can be just a single point. However, due to the conservatism of the OVE algorithm, the ellipsoids which over-bound the feasible set will never converge to a point even though the feasible set does converge to a point. Nevertheless, the OVE algorithm gives a convergent center estimate which can be useful in adaptive control designs.

We have further investigated the convergence of the OVE algorithm when the noise bound approaches zero. The results from the OVE algorithm are very close to the true values as the noise bound tends to zero. In the noise free case, some modifications on the OVE algorithm are needed to avoid ill-conditioning in finding the true estimate. In the latter case, the OVE algorithm requires exactly \( r \) iterations where \( r \) is the number of unknown parameters.
CHAPTER IV
PARAMETER SET ESTIMATION OF AN ARMAX MODEL WITH BOUNDED NOISE

4.1 Introduction

Parameter set estimation of an ARX model with bounded noise has been studied in [5, 39, 51, 52]. The algorithms therein use either boxes or ellipsoids to bound the feasible parameter set consistent with the observations in which the noise or disturbances are bounded. Parameter set estimation has been recently extended to ARMA[54] and ARMAX[55] model structures. In [54], the OBE algorithm developed in [51] is extended to identify the moving average coefficients associated with noise. The main result of the analysis is that all the bounding ellipsoids will contain the true parameter, provided that the true moving average coefficients satisfy a condition which is analogous to the strictly positive real condition required in the extended least square algorithm. The model structure studied there has a practical limitation because no exogenous input is allowed, and is therefore not practical from the control point of view. In [55], an ARMAX model is considered in which the moving average coefficients are the same as the auto-regressive coefficients or equivalently the coefficients of the denominator of the transfer function; the disturbances are assumed to be quasi-stationary. The main result of the analysis in [55] is that, from a stochastic
viewpoint, the parameter set estimate obtained from finite data records is guaranteed to contain the true plant parameter set as the data length tends to infinity. The usefulness of the result in [55] is also limited, however, because practically only finite data records are available for identification purposes.

In this chapter, the same ARMAX model structure as in [55] will be examined for the bounded noise or disturbances case. That is, from an input-output point of view, a general dynamic system with output additive disturbances is considered. An interesting note by Norton[56] finds that the parameter bound given by the observation $y_k$ in such an ARMAX model may not be convex. As an example[56], consider the following model,

$$y_k + a_1 y_{k-1} = b_1 u_{k-1} + v_k + a_1 v_{k-1}$$

with $|v_k| \leq \gamma$ for all $k$. The system can also be represented as

$$y_k - \gamma \leq -a_1 (y_{k-1} - v_{k-1}) + b_1 u_{k-1} \leq y_k + \gamma$$

or

$$y_k - \gamma \leq -a_1 (y_{k-1} \pm \gamma) + b_1 u_{k-1} \leq y_k + \gamma .$$

As shown from Figure 17, the parameter bound given by the observation $y_k$ for this first order ARMAX model is non-convex. This is not surprising because the past noises are correlated with the coefficients of the denominator of the transfer function. Further, the OVE or OBE algorithm cannot be used directly to solve this problem because of the non-convexity of the feasible parameter set. Up to the point of writing this dissertation, no result has been published that solves this parameter.
set estimation problem with ARMAX model structure in a deterministic manner. Though it is inherently conservative to use a convex set to bound a non-convex set, we adopt this strategy here due to the ease in manipulation afforded by the convex set structure.

![Non-convex parameter bound for a first order ARMAX model with bounded noise](image)

Figure 17: Non-convex parameter bound for a first order ARMAX model with bounded noise

With this motivation, the goal of this chapter is to extend the application of the OVE algorithm developed in Chapter 2 to a more general class of systems, namely the ARMAX model with bounded disturbances. In the sequel, a two-step approach is proposed to solve the parameter set estimation problem. In the first step of the proposed scheme, the least square method is used to solve a linear prediction problem
which estimates the denominator coefficients. Then a perturbation analysis is used to find a bound on the estimated denominator coefficients. A new noise bound will then be defined that bounds the moving average part of the model so that the OVE algorithm can be used. Essentially, the type of non-convex parameter set illustrated in Figure 17 is modified to a set bounded by two parallel hyperplanes as shown in Figure 18. This is done by using the bounds on the moving average part, $a_1$ in the figure, computed \textit{a priori}. The details of the procedure is discussed in the sequel.

Figure 18: Modification of a non-convex parameter set to a set bounded by two parallel hyperplanes

In Section 4.2, the parameter set estimation problem is formulated. The first step in the proposed scheme is to compute an absolute bound for the denominator
coefficients; this is analyzed in Section 4.3. In the same section, a modified noise term is then defined and its bound is computed for uses in the OVE algorithm. In Section 4.4, a case study is conducted to examine different experiment design issues to reduce the bound found in Section 4.3; the capability of the proposed scheme is demonstrated via an example. A summary of the results will be given in Section 4.5 to conclude this chapter.

4.2 Problem Definition

Consider the following special SISO ARMAX model,

\[ y_k = \frac{B(q^{-1}, \theta)}{A(q^{-1}, \theta)} u_k + v_k \quad , \]

or equivalently

\[ A(q^{-1}, \theta)y_k = B(q^{-1}, \theta)u_k + A(q^{-1}, \theta)v_k \]

where \( v_k \) is the bounded noise or disturbances added directly to the output \( y_k \); \( \theta \) is the unknown parameter vector. The polynomials \( A(q^{-1}, \theta) \) and \( B(q^{-1}, \theta) \) are defined as:

\[ A(q^{-1}, \theta) = 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n} \quad (4.3) \]

\[ B(q^{-1}, \theta) = b_0 + b_1 q^{-1} + b_2 q^{-2} + \cdots + b_m q^{-m} \quad (4.4) \]

where \( n \) and \( m \) are the number of poles and zeroes in the model which are known \textit{a priori}. The problem here is exactly the same as the parameter set estimation problem solved in Chapter 2, except for the different model structure. One salient feature of
this problem is that no other information of the system is assumed except the system order. The goal is to identify the coefficients in the polynomials $A(q^{-1})$ and $B(q^{-1})$ consistent with the measurements given that $|v(k)| \leq \gamma$ where $\gamma$ is some known bound.

As mentioned previously, the OVE algorithm is not suitable for an ARMAX model since the algorithm is developed only for a system with an ARX model structure. As a result, before applying the OVE algorithm, an a priori bound on the coefficients of $A(q^{-1})$ must be found so that the worst case bound on the $A(q^{-1}, \theta)v_k$ term of (4.2) can be computed.

4.3 Main Result: Two-Step Solution

4.3.1 Step 1: Estimating the Bound on $A(q^{-1}, \theta)$

The Linear Prediction Method for Estimating $A(q^{-1}, \theta)$

The system in (4.1) can be represented in the state-space form as

$$x_{k+1} = \tilde{A}x_k + \tilde{B}u_k \quad (4.5)$$

$$y_k = \tilde{C}x_k + v_k \quad (4.6)$$

where $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are real matrices of appropriate dimensions, $x_k$, $y_k$ and $u_k$ are the state vector, output and input of the system. Assume the system is of order $n$ which is known a priori. Let the characteristic equation of the system be

$$z^n + \sum_{j=0}^{n-1} a_{n-j}z^j = 0 \quad (4.7)$$
Without loss of generality, let the initial condition $x_0 \neq 0$ and let $u_k = 0$ for all $k \geq 0$.

Upon using the Cayley Hamilton Theorem for the matrix $A$, one gets

$$A^n + a_1 A^{n-1} + \cdots + a_{n-1} A + a_n = 0$$

and further pre-multiplying the vectors $C$ and post-multiplying $x_0$ to (4.8) gives

$$C A^n x_0 + a_1 C A^{n-1} x_0 + \cdots + a_{n-1} C A x_0 + a_n C x_0 = 0 .$$

Define the Markov parameters $M(k) = C A^k x_0$, and upon using equation (4.9), one can construct the following matrix equation,

$$H a = h .$$

Because the entries of the Hankel matrix $H$ and the vector $h$ are exact or noise-free Markov parameters, the least square solution of the problem is also exact. In other words, the characteristic coefficients can be identified exactly in the noise-free case.

In the sequel, the noisy case will be discussed.

Perturbation Analysis on the Linear Prediction Problem

If the data entries in the matrix $H$ and $h$ are perturbed or corrupted by noise, it is easy to see that equation (4.9) is no longer valid and therefore no exact solution exists in (4.10). As a matter of fact, if it is known a priori that $M(k)$ will be perturbed by the
bounded noise \( v_k \) with \( |v_k| \leq \gamma \), one can quantitatively over-bound the least square solution of the over-determined equation (4.11). In the analysis below, the objective is to find a bound on the estimated coefficients, \( \hat{a} \), such that the true coefficients, \( a^* \), will be inside the region defined by \( \hat{a} \) and its associated bound.

Suppose the data entries in the matrix \( H \) and \( h \) are noisy and the least square problem \( \hat{a} = \text{arg}(\min \| Ha - h \|_2) \) is solved such that

\[
\rho_{LS} = \| H \hat{a} - h \|_2 
\]  

(4.12)

Since both \( H \) and \( h \) are corrupted with noise, there exists some \( \delta H \) and \( \delta h \) such that \( H + \delta H \) and \( h + \delta h \) are noise-free. Therefore analogous to (4.10), one obtains

\[
(H + \delta H)(\hat{a} + \delta a) = h + \delta h 
\]  

(4.13)

where \( \delta a \) is the perturbation of \( \hat{a} \) in (4.12) such that \( a^* = \hat{a} + \delta a \), where \( a^* \) is the true parameter vector.

Note that since the system order is \( n \), \( \text{rank}(H + \delta H) = n \). Assume that (1) \( \text{rank}(H) = n \) and (2) \( \sigma(H) > \tilde{\sigma}(\delta H) \) where \( \sigma(\cdot) \) and \( \tilde{\sigma}(\cdot) \) represent the smallest and the largest singular values of a matrix, respectively. Assumption (1) says that the noise added to the data will not reduce the system order while assumption (2) says that the output data set used to construct the Hankel matrix is sufficiently excited in the energy sense, which can be interpreted as that the output signal has a sufficiently high SNR. The second assumption may exclude systems with low gains, low input excitations or noisy measurements.

On expanding (4.13), one gets

\[
(H + \delta H)\delta a = (h - H\hat{a} - \delta H\hat{a} + \delta h) 
\]  

(4.14)
To proceed, the following Lemma is required.

**Lemma 4.1:** For any real matrix $A \in \mathbb{R}^{p \times q}$ and vector $x \in \mathbb{R}^q$ where $p \geq q$, then

$$\|Ax\|_2 \geq \sigma(A)\|x\|_2.$$

**Proof of Lemma 4.1:** Using Singular Value Decomposition,

$$A = USV^T$$

where $U \in \mathbb{R}^{p \times p}$, $S \in \mathbb{R}^{p \times q}$, $V \in \mathbb{R}^{q \times q}$, $U$ and $V$ are unitary, and

$$S = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & s_q \end{bmatrix}$$

with $s_i \geq s_j \geq 0$ if $i \leq j$. Then

$$\|Ax\|_2 = \|USV^Tx\|_2$$

$$= \|US\bar{x}\|_2$$

$$= \|U\bar{x}_s\|_2$$

$$= \|\bar{x}_s\|_2$$

where $\bar{x} = V^Tx$, $\bar{x}_s = S\bar{x}$ and the last equality holds since $U$ is unitary. Let $\bar{x}^T = [\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_q]$, then

$$\bar{x}_s^T = [s_1\bar{x}_1, s_2\bar{x}_2, \cdots, s_q\bar{x}_q, 0, \cdots, 0]$$
and

\[ \| \mathbf{x}_s \|_2 = \sqrt{s_1^2 x_1^2 + \cdots + s_q^2 x_q^2} \]
\[ \geq \sqrt{s_q^2 x_1^2 + \cdots + s_q^2 x_q^2} \]
\[ = s_q \sqrt{x_1^2 + \cdots + x_q^2} \]
\[ = \sigma(A) \| \mathbf{x}_s \|_2 \]
\[ = \sigma(A) \| V^T x \|_2 \]
\[ = \sigma(A) \| x \|_2 \]

as \( V^T \) is also unitary. This completes the proof of Lemma 4.1. \( \square \)

On applying the 2-norm to both sides of (4.14) and using Lemma 4.1, one obtains the worst case bound for \( \hat{a} \) from

\[ \| \delta a \|_2 \leq \frac{1}{\sigma(H + \delta H)} (\rho_{LS} + \| \delta H \|_2 \| \hat{a} \|_2 + \| \delta h \|_2) \]  (4.15)
\[ \leq \frac{1}{\sigma(H) - \sigma(\delta H)} (\rho_{LS} + \| \delta H \|_2 \| \hat{a} \|_2 + \| \delta h \|_2) \]  (4.16)

The last inequality is due to a theorem in [57] concerning the perturbation of singular values of a perturbed matrix, that

\[ |\sigma_k(H + \delta H) - \sigma_k(H)| \leq \tilde{\sigma}(\delta H) \]  (4.17)

where \( \sigma_k(\cdot) \) denotes the \( k \)th singular value of a matrix arranged in descending order.

Since |\( v_k \)\| \( \leq \gamma \), then

\[ \| \delta H \|_2 = \sigma_1(\delta H) \]
\[ \leq \sqrt{\sum_{i=1}^{n} \sigma_i^2(\delta H)} \]
\[
\begin{align*}
\|\delta H\|_F &= \sqrt{\sum_{i=1}^{m_0} \sum_{j=1}^{n} (\delta H)_{ij}} \\
&\leq \gamma \sqrt{m_0 n}
\end{align*}
\]

where \((\delta H)_{ij}\) denotes the \((i,j)^{th}\) element of \(\delta H\). Similarly

\[
\|\delta h\|_2 \leq \gamma \sqrt{m_0},
\]

where \(\| \cdot \|_F\) is the Frobenius matrix norm. Using these bounds on \(\|\delta H\|_2\) and \(\|\delta h\|_2\), we can compute the worst case bound on \(\hat{a}\) in (4.16) as

\[
\|\delta a\|_2 \leq \frac{1}{\sigma(H) - \gamma \sqrt{m_0 n}} (\rho_{LS} + \gamma \sqrt{m_0 n}) \|\hat{a}\|_2 + \gamma \sqrt{m_0}) \quad (4.18)
\]

\[
\equiv \eta. \quad (4.19)
\]

Note that, \(m_0\) is a design parameter, but not \(n\). That means if the system order is high\(^1\), \(\eta\) will be large and the bound can be conservative. Nevertheless, through using the bound defined in (4.18), one obtains a norm bound \(\|a - \hat{a}\|_2\) which guarantees the inclusion of the true parameter vector, \(a^*\). Essentially, a ball around \(\hat{a}\) is defined which contains the true parameters.

**4.3.2 Step 2: Application of the OVE Algorithm**

We now address the problem of how the results obtained in Step 1 can be utilized and how the OVE algorithm can be used to estimate both \(A(q^{-1}, \theta)\) and \(B(q^{-1}, \theta)\) will be examined. Note that the parameter set for \(A(q^{-1})\) has already been estimated in Step 1. Using the OVE algorithm in this next step will result in another parameter

\(^1\)Recall that in this context the variable \(n\) denotes the order of the SISO subsystem in an interconnected framework.
set, which now describes the uncertainty of the entire parameter space of $A(q^{-1})$ and $B(q^{-1})$, rather than just that of $A(q^{-1})$.

Let

$$
\tilde{v}_k = A(q^{-1}, \theta)v_k \quad ,
$$

(4.20)

so that

$$
\tilde{v}_k = [1 \ a_1 \ \cdots \ a_n][v_k \ v_{k-1} \ \cdots \ v_{k-n}]^T .
$$

(4.21)

From the last subsection, the bound $||a - \hat{a}||_2 \leq \eta$ is obtained. Since the true value $a^*$ is inside that ball centered at $\hat{a}$ with radius $\eta$, we get $a_i \leq a_i^* \leq \bar{a}_i$ with $a_i = \hat{a}_i - \eta$ and $\bar{a}_i = \hat{a}_i + \eta$, where the index $i$ indicates the $i^{th}$ coefficient in a parameter vector.

Define $a_i^M \equiv \max(|a_i|, |\bar{a}_i|)$. Then the worst case bound for $\tilde{v}_k$ is such that

$$
|\tilde{v}_k| \leq (1 + \sum_{i=1}^{n} a_i^M)\gamma
$$

$$
\equiv \tilde{\gamma} .
$$

(4.22)

(4.23)

As a result, equation (4.2) becomes

$$
A(q^{-1}, \theta)y_k = B(q^{-1}, \theta)u_k + \tilde{v}_k .
$$

Using (4.23) to define the new noise bound, one can readily use the OVE algorithm to identify the parameter set consistent with the measurements. That is, the new noise bound $\tilde{\gamma}$ should replace $\gamma$ in the OVE algorithm when an ARMAX model is considered. As a guarantee for consistency, the initial ellipsoid must be chosen big enough such that it contains the feasible (non-convex) parameter set.
4.4 Computer Simulations

4.4.1 A Case study on Some Experiment Design Issues for the Two-Step Solution

The purpose of experiment design in system identification is to maximize the information return from the input-output data; the design is often cast as an optimization problem and the solution is very much dependent on the chosen identification algorithm, the criterion and the given system constraints. In general, an experiment design for the identification of a dynamical system may include the choice of input excitation signals, sampling time, where, what and when to take measurements, and pre-treatment of noisy data. Each variable has a significant bearing upon the information provided by an experiment. The effects of these variables are, in general, closely interrelated and a joint design is often preferred. Experiment design can be analyzed either in the time or frequency domain. For more involved treatment of experiment design in system identification, readers are referred to [58, 59, 1].

In this section, we take the philosophy of experiment design to minimize the bound for the denominator coefficients given in the last section. From equation (4.18), the bound \( \eta \) on \( \hat{a} \), is given as

\[
\eta = \frac{1}{\sigma(H) - \gamma \sqrt{m_0 n}} (\rho_{LS} + \gamma \sqrt{m_0 n} \| \hat{a} \|_2 + \gamma \sqrt{m_0})
\]

Upon careful examination it is seen that the equation for \( \eta \) contains some free parameters that can be used to minimize \( \eta \). The parameter \( \eta \) in fact depends on \( \sqrt{m_0} \) and \( \sigma(H) \), while \( \sigma(H) \) depends on the sampling time, the size of the matrix \( H \) and the first data entry, \( i \), in the Hankel matrix \( H \) of equation (4.10). In this
case study, we will show how these parameters affect the size of the bound $\eta$ via an example of computer simulation. These variables are considered separately here to give some insight into their individual effects upon the information content of the data.

Consider the following second order transfer function,

$$G(s) = \frac{1.7s}{s^2 + 0.5s + 4}.$$ 

**Case 1: Effect of choice of sampling time**

In this case, the continuous system $G(s)$ is discretized with different sampling periods, $T_s$, and the noise-free pulse response is simulated. To study only the effect of sampling time on $\sigma(H)$, we pick $m_0 = 10$ and $i = 1$. The system $G(s)$ is known to have a resonant frequency at 0.32 Hz and a minimum sampling frequency of 2 Hz is assumed. Table 1 shows how $\sigma(H)$ can be affected by the different choices of sampling times. From this example, it is seen that $\sigma(H)$ increases with the sampling period indicating that a smaller sampling frequency can reduce $\eta$.

<table>
<thead>
<tr>
<th>$T_s$</th>
<th>0.01</th>
<th>0.02</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma(H)$</td>
<td>$9.8 \times 10^{-4}$</td>
<td>$9.4 \times 10^{-4}$</td>
<td>$6.04 \times 10^{-3}$</td>
<td>$4.86 \times 10^{-2}$</td>
<td>$1.63 \times 10^{-1}$</td>
<td>$7.34 \times 10^{-1}$</td>
</tr>
</tbody>
</table>

**Case 2: Effect of choice of $m_0$**

In this case, we choose $^2 T_s = 0.5s$, $i = 1$. Table 2 shows how $\sigma(H)$ can be affected

$^2$Although for many applications, such a sampling frequency may be quite low, here we are interested in illustrating the effect of the design parameter $m_0$ on the information content of the data.
by the different choices of $m_0$ in the noise-free case. In another simulation, a bounded random noise of $\gamma = 0.05$ is now introduced to the pulse responses to give a signal to noise ratio of 36dB. Table 3 shows the effect on the bound $\eta$ for different choices of $m_0$. It is shown that for this example, smaller $m_0$ would be a good choice; this in fact is what one should expect because of the linear dependence of $\eta$ on $\sqrt{m_0}$. Apparently from these two tables, $\sigma(H)$ does not seem to be a good indicator for $\eta$ as $\sigma(H)$ is not very sensitive with respect to $m_0$. However, the fact that $\sigma(H)$ does not change very much with different $m_0$ has actually indicated to us that it is not justifiable to select a larger $m_0$ because the bound $\eta$ increases linearly with $\sqrt{m_0}$.

Table 2: Effect of $m_0$ on $\sigma(H)$

<table>
<thead>
<tr>
<th>$m_0$</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma(H)$</td>
<td>0.711</td>
<td>0.675</td>
<td>0.742</td>
<td>0.734</td>
<td>0.765</td>
<td>0.768</td>
</tr>
</tbody>
</table>

Table 3: Effect of $m_0$ on $\eta$

<table>
<thead>
<tr>
<th>$m_0$</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta$</td>
<td>0.087</td>
<td>0.106</td>
<td>0.129</td>
<td>0.139</td>
<td>0.193</td>
<td>0.238</td>
</tr>
</tbody>
</table>

Case 3: Effect of choice of $i$

As in the last case, we choose $T_s = 0.5s$. For the noise free case, Table 4 shows how $\sigma(H)$ can be affected by different choice of $i$ for $m_0 = 4$. It indicates that the choice of $i$ has a profound effect on $\sigma(H)$. If the same bounded noise as in the last case is introduced, the effect of different choices of $i$ on $\eta$ is shown in Table 5. The
general trend is that the bound $\eta$ increases with $i$ for larger $i$. For this example with a fixed $m_0$, $g(H)$ is clearly a good indicator for $\eta$ and a smaller $i$ should be chosen.

Table 4: Effect of $i$ on $g(H)$

<table>
<thead>
<tr>
<th>$i$</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g(H)$</td>
<td>0.620</td>
<td>0.473</td>
<td>0.251</td>
<td>0.214</td>
<td>0.068</td>
<td>0.020</td>
</tr>
</tbody>
</table>

Table 5: Effect of $i$ on $\eta$

<table>
<thead>
<tr>
<th>$i$</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta$</td>
<td>0.061</td>
<td>0.092</td>
<td>0.128</td>
<td>0.229</td>
<td>0.867</td>
<td>5.35</td>
</tr>
</tbody>
</table>

The three different cases in the simulation study conducted above has indicated that the step 1 approach to finding the bound for the denominator coefficients can generally be affected by different choices of sampling time, $m_0$ and $i$. Through the example, one is capable of getting some insight into the importance of maximizing the information content of the data to minimize the uncertainty in the estimation. It is intuitive that an optimal experiment can be carried out with an information criterion related to the design parameters studied here. Nevertheless, the study in this section only serves to indicate the significance of experiment design in parameter set estimation, but it is not our purpose here to carry out a full investigation on optimal experiment design for parameter set estimation.
### 4.4.2 Application of the OVE Algorithm

In this section, we choose, based on the experiment design study given above, a sampling time of 0.5 seconds to discretize \( G(s) \) given previously for simulation. The corresponding system in discrete time is

\[
Y(z) = G(z)u(z) + V(z)
\]

where

\[
G(z) = \frac{0.633(z^{-1} - 1z^{-2})}{1 - 0.9652z^{-1} + 0.7788z^{-2}}.
\]

A \( 4 \times 2 \) Hankel matrix \( H \) with the starting data \( i = 2 \) is constructed; the noise bound is given as \(|v(k)| \leq 0.05\). Upon using the step 1 solution described in the last section, \( \eta \) is found to be 0.061. Table 6 shows the estimate of \( a \) and its associated parameter interval. We therefore obtain \( a_1^M = 1.0284 \) and \( a_2^M = 0.8428 \), and the new noise bound \( \bar{\gamma} \) defined in Section 4.3 is given as

\[
\bar{\gamma} = 1 + a_1^M + a_2^M = 2.8712.
\]

At this point, \( \bar{\gamma} \) may be used in the OVE algorithm. A random excitation signal is used to generate the data records; the SNR is 28 dB. For 100 iterations, Table 7

<table>
<thead>
<tr>
<th>True value</th>
<th>Estimated Value</th>
<th>Parameter interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>-0.9652</td>
<td>-0.9678</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>0.7788</td>
<td>0.7818</td>
</tr>
</tbody>
</table>
shows the estimation result. The result indicates that the proposed two-step solution to parameter set estimation of an ARMAX model with bounded noise is feasible, and the true values of $\theta$ are inside the region.

Table 7: Estimation result of the example using Step 2 scheme

<table>
<thead>
<tr>
<th></th>
<th>True value</th>
<th>Estimated Value</th>
<th>Parameter interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>-0.9652</td>
<td>-0.9574</td>
<td>-1.0548 $\leftrightarrow$ -0.86</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.7788</td>
<td>0.7702</td>
<td>0.6792 $\leftrightarrow$ 0.8612</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.633</td>
<td>0.6339</td>
<td>0.5763 $\leftrightarrow$ 0.6915</td>
</tr>
<tr>
<td>$b_2$</td>
<td>-0.633</td>
<td>-0.6368</td>
<td>-0.6903 $\leftrightarrow$ -0.5833</td>
</tr>
</tbody>
</table>

4.5 Conclusions

In this chapter, a two-step approach is proposed to identify the parameter set of an ARMAX model with bounded noise. Specifically, the model structure is such that the moving average part is identical to the auto-regressive part; such model structure covers a large class of systems for control purposes. That is, from the input-output point of view, the noise is directly added to the output. In the first step of the proposed scheme, the least square method is used to solve the linear prediction problem which identifies the denominator coefficients and a perturbation analysis is used to compute an $a$ priori bound on each of the denominator coefficients. The bound on the estimate can be characterized as a ball centered at the estimate with a radius $\eta$. Though using a ball to characterize the system parameter uncertainty is conservative, the gain here is that this method is fast and easy to implement to compute a bound for the moving average part in the ARMAX model. Essentially, the scheme helps to find a bound
for the "modified" noise term which results by casting the ARMAX structure into an
ARX structure so that the OVE algorithm can be used.

In the last section, some of the experiment design issues have been brought up as
a case study to illustrate the importance of experiment design in reducing the size of
the parameter set. For the example studied, sampling time, the size of the Hankel
matrix and the starting entries in $H$ and $h$, all play significant roles in reducing the
bound on the denominator coefficients. An example is used to illustrate the capability
of the proposed scheme to solve the parameter set estimation problem of an ARMAX
model with bounded noise. Note, however, that due to parameter uncertainty in
the AR coefficients computed from Step 1, a certain degree of conservatism will be
introduced into the computation of $\tilde{\gamma}$, and such conservatism will be added onto the
conservatism of the OVE algorithm. This can be seen from Table 6 and Table 7 that
the parameter intervals for $a_i$ are expanded after Step 2.
CHAPTER V

PARAMETER SET ESTIMATION IN INTERCONNECTED DYNAMIC SYSTEMS

5.1 Introduction

The study of Large Scale Systems (LSS) has been a very active research topic in the past two decades [11, 10]. LSS, in general, are complex systems of high dimension involving a large number of variables. Usually, systems categorized as LSS may contain a number of interconnected (weakly coupled or strongly coupled) subsystems which may be geographically separated. Typical examples of LSS are the national economy, traffic routing, computer networks, electric power systems, water resource systems, industrial robotic systems, large flexible space structures and so on.

Because of the complexity and/or high dimensionality normally implicated in the LSS, centralized control design techniques are often avoided. Numerically intractability in controller designs for high dimensional systems and/or the overhead time incurred in the design process is a burden to real time application. It may also be very costly or cumbersome to collect information from a geographically separated complex system and then send it to the central information center for analysis and design. The reliability of the communication links between subsystems which are physically far apart is also a concern. There is therefore a need to study how an interconnected
system can be modeled and controlled under decentralized information constraints.

In general, certain model structures such as decentralized systems and interconnected systems are assumed in the study of large scale problems for they are simpler to design and analyze. The classical approach to study large scale problems is to decompose the system into smaller subsystems for simpler local designs and then study the overall effect when these subsystems are interconnected. A commonly used representation of an interconnected system, $S$, in the discrete time domain is

$$S : \quad x_i(k + 1) = A_i x_i(k) + B_i u_i(k) + d_i(k) + \sum_{j=1}^{N_s} f_{ij}(k, x_j)$$

$$y_i(k) = C_i x_i(k), \quad i = 1, \ldots, N_s$$

where $A_i$, $B_i$ and $C_i$ are the system matrices of the $i^{th}$ subsystem; $y_i$, $u_i$ and $d_i$ are the outputs, control inputs and disturbance inputs of the $i^{th}$ subsystem; $f_{ij}$ is the interconnection between the $i^{th}$ subsystem and the $j^{th}$ subsystem; $N_s$ is the total number of subsystems in the composite model, and $k$ is the time index in discrete time. In the frequency domain, a multi-input multi-output system transfer function matrix, $G(z)$, is partitioned as,

$$G(z) = \begin{bmatrix}
G_{11}(z) & G_{12}(z) & \cdots & G_{1N_s}(z) \\
G_{21}(z) & G_{22}(z) & \cdots & G_{2N_s}(z) \\
\vdots & \vdots & \ddots & \vdots \\
G_{N_s1}(z) & G_{N_s2}(z) & \cdots & G_{N_sN_s}(z)
\end{bmatrix}$$

to form an interconnected system where $G_{ii}(z)$ is the $i^{th}$ subsystem transfer function matrix and $G_{ij}(z), \quad i \neq j$ is the interconnecting transfer function between the outputs of the $i^{th}$ subsystem and the inputs of the $j^{th}$ subsystem.
Over the past decades, research in large scale systems has mainly focused on the control aspects, and little or no effort has been put into the area of system identification for those special model structures. There is no doubt that accurate mathematical representation of a dynamical system forms the foundation for reliable and cost-effective control synthesis. The tremendous progress in the control theory development and system identification techniques for modeling during the past two decades has sparked special attention to interconnected systems. In the realm of linear quadratic feedback control, abundant results have been developed for decentralized systems which are analogous to the centralized design[11, 10]. In the frequency domain, Diagonal Dominance or Block Diagonal Dominance concept[60, 61], Block Relative Gain concept[62], Interactive Measures[63, 64], $H_\infty$ optimization concept[65] and so on are developed for robust decentralized control of interconnected systems. Decentralized adaptive control for interconnected systems has also been given our attention recently[23, 24, 32, 25]. The application of indirect adaptive control to interconnected systems presents one of the most challenging research areas in adaptive control. Two fundamental problems exist in decentralized control of interconnected systems. First is how a subsystem can be identified in the presence of unknown or partially known interconnections from other interconnected subsystems and second is how the identified subsystem can be used to design local controllers to stabilize the overall system.

In this chapter, the first fundamental problem is investigated. The concept of parameter set estimation is introduced to an interconnected dynamic system for the
purpose of identifying local subsystems given that only local input-output information is available. The approach taken in this chapter is different from the analysis in [24, 32, 25], where the authors used state-space structure to represent an interconnected system. In this chapter, an interconnected system is viewed from the input-output point of view, and only linear interconnections between subsystems are considered. As in all existing results, we consider that the interconnections to a subsystem are bounded disturbances or perturbations.

The goal of this research is to identify a local parameter set for each local subsystem in the presence of interconnections from other subsystems, with bounded external additive output disturbances. The OVE algorithm, developed in Chapter 2 and 3, will be used to find an ellipsoid to characterize the parameter set which is consistent with the measurements. The motivation is that once a set estimate is obtained for each subsystem, it remains to develop robust control design and performance analysis decentrally for each subsystem set estimate independently. This idea will be discussed further in the next section. The parameter set estimate can also be used in indirect adaptive control design since the OVE algorithm provides a convergent estimate as shown in Chapter 3.

In the sequel, only the effects of subsystem interconnections and bounded disturbances are considered, and they are reflected into the local subsystem parameter space. As mentioned previously, the OVE algorithm will be the core step in the parameter set estimation of an interconnected system; there are several reasons in using the ellipsoidal OVE algorithm:
(1) the parameterization of a set as an ellipsoid is easy to implement and manipulate in computation\cite{46};

(2) the algorithm is recursive and the parametric set can be refined recursively as additional measurement data is available;

(3) the algorithm is capable of saving computation time as parameter updates are information dependent;

(4) the OVE estimator possesses some nice convergence properties as examined in Chapter 3;

(5) the subsystem parameter set estimates can be used in either adaptive or non-adaptive robust control designs.

In Section 5.2, the set estimation problem in interconnected systems is defined and the assumptions are given for the problem. In Section 5.3, a time-varying instantaneous bound will be derived which accounts for the interconnection, and disturbance effects for use in the OVE algorithm will be given. Simulation results of set estimation in interconnected systems will be presented in Section 5.4. Discussion appears in the last section to conclude this chapter.

5.2 Problem Definition

Consider the following SISO interconnected subsystem,

\[ S_i : \quad y_i(z) = \sum_{j=1}^{N} G_{ij}(z)u_j(z) + d_i(z) \]  

(5.1)
where the subscript $i$ denotes that the entity is associated with the $i^{th}$ subsystem, $d_i(z)$ is a bounded external disturbance or measurement noise and $N_s$ is the number of SISO subsystems in the composite system. The problem is to robustly estimate parameters in the $i^{th}$ subsystem in the presence of subsystem interconnections and bounded external disturbances.

It is well known that multiple non-robust local control designs which stabilize a decoupled system may not be adequate for an overall closed loop system[11]; certain interconnection constraints must be satisfied for global stability. Moreover, non-robust control design in interconnected systems requires accurate identification of the local subsystems and their interconnections. However, a centralized strategy for estimating system parameters may be impossible in a large scale system due to either numerical difficulties or the growing complexities involved in the identification algorithms for high order systems. To tackle this problem, we propose to break down the single identification problem of the entire interconnected system into smaller ones of identifying the subsystems. The problem amounts to identifying a subsystem set estimate given some partial knowledge of the interconnecting subsystems.

**System A Priori Assumptions**

$A1: |u_i(k)| \leq \bar{u}_i$ and $|d_i(k)| \leq \bar{d}_i$ for all $i \in [1, N_s]$, $k \geq 0$ in both the identification and feedback control configurations, where $\bar{u}_i$ and $\bar{d}_i$ are some positive known constants.

$A2: G_{ij}(z)$ is BIBO stable for all $i, j \in [1, N_s]$. 
A3: The pulse response function $g_{ij}(k)$ of $G_{ij}(z)$, $i \neq j$, is exponentially stable and

$$|g_{ij}(k)| \leq \rho_{ij}e^{-\tau_{ij}kT}$$

for some known and positive $\rho_{ij}, \tau_{ij}$; $T$ is the sampling time.

A4: The order of the polynomials in $G_n(z)$ are known a priori, and the bounds on

the coefficients of the denominator polynomial of $G_n(z)$ are known.

A5: Only $y_i(\cdot)$ and $u_i(\cdot)$ are available at the $i^{th}$ station for identification purposes.

Assumption A1 guarantees that the parameter set estimate obtained from the
identification procedure must be compatible with the situation when the system is
under feedback control. Generally speaking, assumptions A1-A4 are all intended to
make decentralized set estimation of each subsystem feasible under the information
constraints A5.

The estimation approach taken here is set theoretic in nature and the motiva-
tion for doing so is that, with respect to each subsystem parameter set, a robust
local controller design is sufficient to guarantee robust stability and performance in
the composite system despite the presence of interactions from other subsystems and
the additive external output disturbances. This reduces the global system analysis
to local analyses in which only decentralized control design and control performance
analysis is required on each subsystem parameter set regardless of the existence of
other subsystems, and while achieving the global objectives. Note that such a control
conjecture makes sense only under the above assumptions that essentially all subsys-
tem interconnection effects, even under closed-loop control, can be reflected into each
bounding parameter set for each subsystem. In the sequel, an instantaneous bound
on the "total" disturbance will be found which includes subsystem interactions and
external disturbances. As will be shown, the "total" disturbance bound is monotonically increasing and assumption A2 is needed to guarantee that that such a bound does converge.

The problem can now be formally defined as, for each \( i \in [1, N_x] \), finding a parameter set for each \( G_{ii}(z) \) satisfying the assumptions A1 to A5 using the measurement sequences \( u_i(\cdot) \) and \( y_i(\cdot) \).

### 5.3 Computation of the Instantaneous Bound for Subsystem Interactions

In this section, we will formulate the subsystem parameter set estimation problem in an interconnected system as one with an ARMAX model structure as discussed in Chapter 4. In the sequel, an instantaneous bound for the "total" disturbance term, which includes subsystem interactions and external output disturbances, is derived.

The system in (5.1) can be rewritten in the time domain as

\[
y_i(k) = \frac{B_i(q^{-1}, \theta_i)}{A_i(q^{-1}, \theta_i)} u_i(k) + \sum_{j=1, j \neq i}^{N_x} \sum_{l=0}^{k} g_{ij}(l) u_j(k - l) + d_i(k) \tag{5.2}
\]

or

\[
y_i(k) - \phi_i^T(k) \theta_i = A_i(q^{-1}, \theta_i) \bar{y}_i(k) + B_i(q^{-1}, \theta_i) \bar{u}_i(k) \tag{5.3}
\]

where \( g_{ij}(k) \) is the pulse response of \( G_{ij}(z) \), \( q^{-1} \) is the delay operator and

\[
A_i(q^{-1}, \theta_i) = 1 + a_{1,i} q^{-1} + a_{2,i} q^{-2} + \cdots + a_{n_i,i} q^{-n_i}
\]

\[
B_i(q^{-1}, \theta_i) = b_{0,i} + b_{1,i} q^{-1} + b_{2,i} q^{-2} + \cdots + b_{m_i,i} q^{-m_i}
\]
\[ \bar{v}_i(k) = \sum_{j=1, j \neq i}^{N_i} \sum_{l=0}^{k} g_{ij}(l) u_j(k - l) + d_i(k) \]

\[ \theta_i^T = [-a_{1,i}, \ldots, -a_{n_i,i}, b_{0,i}, b_{1,i}, \ldots, b_{m_i,i}] \]

\[ \phi_i^T(k) = [y_i(k - 1), \ldots, y_i(k - n_i), u_i(k), \ldots, u_i(k - m_i)] . \]

Define \( v_i(k) = A_i(q, \theta_i) \bar{v}_i(k) \) which is essentially the "total" disturbance term for the \( i^{th} \) subsystem. Note that the right-hand side of equation (5.4) is a function of \( \theta_i \), causing some difficulties in obtaining a bound for \( v_i(k) \) since \( \theta_i \) is unknown. By assumptions A3 and A4, using the step 1 technique discussed in the Chapter 4, an instantaneous bound for \( v_i(k) \) can be easily derived. To do so, we essentially need to find the bounds on \( \bar{v}_i(k) \) and \( a_{i,j} \), respectively. From the assumptions A1 and A3, the following bound for \( \bar{v}_i(k) \) is obtained:

\[ |\bar{v}_i(k)| \leq \sum_{j=1, j \neq i}^{N_i} (\bar{a}_j \sum_{i=0}^{k} \rho_{ij} e^{-\tau_{ij}kT}) + \bar{d}_i = \hat{\nu}_i(k) . \] (5.5)

Now from assumption A4, assume that \( a_{i,i} \leq a_{i,i} \leq \bar{a}_{i,i} \), and define \( \alpha_{i,i} = \max(|a_{i,i}|, |\bar{a}_{i,i}|) \). On expanding \( v_i(k) \), one gets

\[ v_i(k) = [1 \ a_{1,i} \ \cdots \ a_{n_i,i}][\bar{v}_i(k) \ \bar{v}_i(k - 1) \ \cdots \ \bar{v}_i(k - n_i)]^T , \] (5.6)

providing a bound for \( v_i(k) \) as,

\[ |v_i(k)| = |y_i(k) - \phi_i^T(k)\theta_i(k)| \] (5.7)

\[ \leq (1 + \sum_{j=1}^{n_i} \alpha_{i,j}^M)\hat{\nu}_i(k) \] (5.8)

\[ = \gamma_i(k) \] (5.9)

as \( \hat{\nu}_i(j) \geq \hat{\nu}_i(k) \) for \( j \geq k \). In fact, one can use a less conservative bound for \( \gamma_i(k) \) as

\[ \gamma_i(k) = [1 \ \alpha_{i,1}^M \ \cdots \ \alpha_{i,n}^M][\hat{\nu}_i(k) \ \hat{\nu}_i(k - 1) \ \cdots \ \hat{\nu}_i(k - n_i)]^T . \] (5.10)
Note that $\gamma_i(k)$ is increasing with time, as is $\beta_i(k)$. Since $\gamma_i(k) \leq \gamma_i(k + 1)$ and $\gamma_i(k)$ converges as a result of assumption A2, the convergence result of the OVE algorithm discussed in Chapter 3 still holds. Direct application of the OVE algorithm at this point using $\gamma_i(k)$ and also $\{y_i(\cdot), u_i(\cdot)\}$ will give the $i^{th}$ subsystem set estimate.

5.4 Simulation Example

In this section, an example of an interconnected dynamic system is simulated and the OVE algorithm is implemented to estimate the subsystem parameters in the presence of subsystem interactions.

Consider a large space structure example in the form of a 4th order truss model [47] which contains only the 1st $x$-bending and $y$-bending modes:

$$\begin{align*}
\dot{x} &= \left[ \begin{array}{cccc}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
-120.95 & 0 & -0.3806 & 0 \\
0 & -120.93 & 0 & -0.3805 \\
\end{array} \right] x + \left[ \begin{array}{c}
0 \\
0 \\
-2.0079 \\
-0.2604 \\
\end{array} \right] u \\
\dot{y} &= \left[ \begin{array}{cc}
0 & -1.3283 \\
0 & -0.1725 \\
\end{array} \right] \begin{array}{c}
x \\
v_1 \\
\end{array} + \left[ \begin{array}{c}
-0.1722 \\
1.3281 \\
\end{array} \right] \begin{array}{c}
x \\
v_2 \\
\end{array} \\
\end{align*}$$

(5.11) (5.12)

Note that this model has been modified from the actual model in [47]; the modal damping has been increased and the interaction between the two modes has also been increased. Figure 19 shows the bode gain plots of the $x$-bending mode subsystem and its interconnecting subsystem. A sampling time of 0.05 seconds is used to discretize the model for the purpose of data generation. As a result, we have the following $G_{11}(z)$ and $G_{12}(z)$:

$$G_{11}(z) = \frac{0.1276(z^{-1} - z^{-2})}{1 - 1.689z^{-1} + 0.9811^{-2}}$$

(5.13)
\[ G_{12}(z) = \frac{10^{-4}(0.192z^{-1} - 0.871z^{-2} + 0.868z^{-3} - 0.1894z^{-4})}{1 - 3.38z^{-1} + 4.819z^{-2} - 3.317z^{-3} + 0.963z^{-4}}. \]  

It is known \textit{a priori} that both subsystems, \( G_{11}(z) \) and \( G_{22}(z) \), are of second order and both \( g_{12}(k), g_{21}(k) \leq 0.0022e^{-0.065kT} \) where \( T \) is the sampling time. A bounded random noise sequence with \( |v_{1}(k)| \leq 0.05 \) is used and the corresponding \( S/N \) ratio is 30.5dB. In the sequel, only the first subsystem is considered.

Before applying the OVE algorithm, the \textit{a priori} bound for the local subsystem denominator coefficients must be found in compliance with assumption A4 in Section 5.2. To facilitate this, only the first subsystem is excited with an external signal and the other subsystem is assumed at rest in the simulation. A bounded zero-mean random signal \( u_{1}(k) \) with \( |u_{1}(i)| \leq 5 \) for \( i = 0, \cdots, 20 \) and \( u(i) = 0 \) for \( i > 20 \) is used and the output \( y_{1}(k) \) is measured. Output values for \( k \geq 25 \) are used to construct the matrix \( H \) and the vector \( h \) defined in Chapter 4. The noise sequence and the noise-free signal sequence are shown in Figure 20. Tables 8 and 9 show the results of using the algorithm discussed in Section 4 for computing a bound on the denominator.
coefficients; $a_{i,j}^*$ indicates the true value while $\hat{a}_{i,j}$ indicates the estimated values. In the tables, different bounds are computed and listed for different choices of the size of the Hankel matrix $H$. Clearly, $m_0 = 10$ is the best choice for it gives the smallest bound on the denominator coefficients. Correspondingly, one gets $a_{1,1}^M = 2.26$ and $a_{2,1}^M = 1.557$.

![Figure 20: The output signal and the noise sequences for the example](image)

Now, the result discussed in the last section can be used to compute the bound for the "total" disturbance term that accounts for the effect of interconnection and the OVE algorithm can be used to estimate the $1^{st}$ subsystem. For this example, the system is simulated and 200 data points are collected. From the simulation, it is found that the ellipsoids cease to update after $k = 30$. This is due to the fact that the "noise" bound is growing with time as mentioned before and when this "noise"
Table 8: The estimates and the bounds on denominator coefficients for the first decoupled subsystem

<table>
<thead>
<tr>
<th>$m_0$</th>
<th>$a_{2,1}^*$</th>
<th>$a_{3,1}^*$</th>
<th>$\hat{a}_{1,1}$</th>
<th>$\hat{a}_{2,1}$</th>
<th>$|a_{j,1} - \hat{a}_{j,1}|_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-1.689</td>
<td>0.981</td>
<td>-1.711</td>
<td>0.962</td>
<td>0.642</td>
</tr>
<tr>
<td>10</td>
<td>-1.689</td>
<td>0.981</td>
<td>-1.686</td>
<td>0.984</td>
<td>0.574</td>
</tr>
<tr>
<td>20</td>
<td>-1.689</td>
<td>0.981</td>
<td>-1.692</td>
<td>0.981</td>
<td>0.642</td>
</tr>
</tbody>
</table>

Table 9: The parameter interval on denominator coefficients for the first decoupled subsystem

<table>
<thead>
<tr>
<th>$m_0$</th>
<th>Parameter Interval ($a_{1,1}$)</th>
<th>Parameter Interval ($a_{2,1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-2.353 $\leftrightarrow$ -1.069</td>
<td>0.32 $\leftrightarrow$ 1.604</td>
</tr>
<tr>
<td>10</td>
<td>-2.26 $\leftrightarrow$ -1.112</td>
<td>0.41 $\leftrightarrow$ 1.558</td>
</tr>
<tr>
<td>20</td>
<td>-2.334 $\leftrightarrow$ -1.05</td>
<td>0.339 $\leftrightarrow$ 1.623</td>
</tr>
</tbody>
</table>

bound has reached a certain level, it is impossible to find a smaller volume ellipsoid to bound the feasible parameter set consistent with the future observations. The center estimate and parameter interval defined by the ellipsoid are given below in Table 10, for $k = 30$.

The overall result indicates that the proposed scheme discussed in Chapter 4 is a viable method to obtain the a priori bound on the denominator coefficients of a local subsystem. By finding an instantaneous bound on the "total" disturbance term, one is able to identify subsystem parameters in a set theoretic framework via the OVE algorithm. Since the worst-case conservative bound computed in Section 5.3 was used to account for the "total" disturbance term, the parameter set identified for the
Table 10: The center estimate and the bounds on the first decoupled subsystem

<table>
<thead>
<tr>
<th></th>
<th>True Value</th>
<th>Estimated Value</th>
<th>Parameter Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_{1.1}$</td>
<td>0.1276</td>
<td>0.1273</td>
<td>$-0.09 \leftrightarrow 0.3447$</td>
</tr>
<tr>
<td>$b_{2.1}$</td>
<td>-0.1276</td>
<td>-0.1275</td>
<td>$-0.4214 \leftrightarrow 0.1664$</td>
</tr>
<tr>
<td>$a_{1.1}$</td>
<td>-1.689</td>
<td>-1.6342</td>
<td>$-3.3974 \leftrightarrow 0.129$</td>
</tr>
<tr>
<td>$a_{2.1}$</td>
<td>0.9811</td>
<td>0.9234</td>
<td>$-0.8466 \leftrightarrow 2.6934$</td>
</tr>
</tbody>
</table>

$x$-bending mode subsystem in the example is most likely, also conservative; this can be seen by comparing the parameter intervals in Table 9 and Table 10. Using the current methodology, we see an increase in the interval for $a_{j.1}$ when interconnection effects are included. The conservatism of the result is due to the increasing noise bound derived in the last section which leads to an early cessation of parameter set update.

5.5 Conclusions

In this chapter, the set theoretic approach to parameter estimation is introduced to robustly identify local subsystems of an interconnected system. Due to the nature of the influence of subsystem interconnections and output additive external disturbances, the set estimation problem has an ARMAX model structure, requiring additional assumptions for problem formulation. The results discussed in Chapter 4 can be used to identify an \textit{a priori} bound on the denominator coefficients of each local subsystem. With this \textit{a priori} bound, the set estimation problem of interconnected systems can be recast in the ARX structure and the OVE algorithm can be readily
applied to identify the uncertainty in the parameter space of each local subsystem in
the presence of interconnections and disturbances.

An example problem is studied, and the two-step procedure in Chapter 4 is imple­
mented to find an \textit{a priori} bound on the denominator coefficients of the $1^{st}$ subsystem.
Further, the result obtained in the identification of local subsystem parameters using
the OVE algorithm has suggested that this decentralized subsystem parameter set
estimation strategy is feasible. This estimator can be used in decentralized indirect
adaptive control design because the OVE estimator possesses adequate convergence
properties as discussed in Chapter 3. For application in non-adaptive robust control,
the conservatism of the estimated parameter set may limit the degree of control ro­
bustness one can achieve. Further research is required to reduce the conservatism of
subsystem parameter uncertainty sets.
CHAPTER VI

CONCLUSIONS AND FUTURE RESEARCH

In this dissertation, solutions are given for three fundamental problems in system identification, namely the robustness in parameter estimation, the deterministic quantification of estimation uncertainty for robust control design and the estimation strategy in a large interconnected system with information constraints. The deterministic problem of estimating an ARX model with bounded noise is examined. Chapter II provides the OVE algorithm which is a recursive ellipsoid algorithm similar to the commonly known RLS estimation method. The OVE algorithm has a salient feature that at each recursion, an optimal solution volume ellipsoid is found to bound the intersection between the feasible parameter set consistent with the current measurement and the previous ellipsoid. Additionally, the previous ellipsoid is updated only if the current optimal ellipsoid has a smaller volume, making the recursive OVE algorithm information dependent. This is in contrast to the RLS estimation method where parameter estimates are updated every iteration. From the parameter robust control point of view, the OVE algorithm provides a quantification of the parameter uncertainty region within which the true parameter lies.

The convergence of the OVE set estimator is analyzed in Chapter 3. The convergence of the estimator is the key issue in establishing stable indirect adaptive control.
It has been shown that, the OVE algorithm gives not only non-increasing and convergent volume ellipsoids, but also convergent center estimate to some region around the true estimate. More important, the output prediction error using the center estimate is also converging to within the noise bound which corresponds closely to the robust estimation strategy using a dead zone. The convergence of the ellipsoid parameter set is further studied from the viewpoint of feasible parameter set degeneracy. It is shown that the recursive ellipsoid does not converge to a point even though the feasible parameter set does. Further, if the noise bound is small, it is found that the OVE algorithm will result in small ellipsoids around the true parameter via an example. In the noise-free case, the OVE algorithm must be modified slightly in order to compute the true estimate exactly. These convergence results and analysis are lacking in [43] and [44].

In Chapter IV, the main result is that a two-step approach is proposed to solve the parameter set estimation problem of an ARMAX model with bounded noise using convex ellipsoids; this problem has been unsolved prior to this dissertation. This work has a significant bearing on control systems as most physical systems can be modeled in ARMAX structure rather than ARX structure. In the first step of the proposed scheme, a bound on the MA part is found by using the linear prediction method followed by a perturbation analysis. Essentially, the ARMAX model structure is cast into an ARX structure so that the OVE algorithm can be applied directly. Simulations show that this proposed scheme is feasible in using a convex ellipsoid to bound a non-convex parameter region associated with an ARMAX model. The issue
of experiment design has also been highlighted in the chapter. It is found that the choice of sampling time and the size of Hankel matrix used in the first step of the proposed scheme play significant roles in shrinking the size of ellipsoid which defines the parameter set.

In Chapter V, a scheme is proposed to handle the estimation problem in interconnected systems with information constraints. The proposed scheme is the application of the OVE algorithm to robustly estimate the parameter set of each local subsystem under the influence of bounded subsystem interactions and bounded external disturbances. The idea is similar to the case for the ARMAX model, where a "new" disturbance bound is found to bound the disturbance contributions from subsystem interactions and external disturbances so that the OVE algorithm can be used. Under this scheme, the estimation problem for an interconnected system becomes several subsystems parameter set estimation problems with each one cast in the ARX structure. As subsystem interactions have been accounted for in each subsystem parameter set, then robust decentralized controllers, if they exist, with respect to each subsystem parameter set can be designed which are sufficient to achieve control and performance robustness in the composite system.

**Recommendation for Future Research**

The following areas are proposed for further research:

1. Extension of the OVE algorithm to account for some *a priori* or partially known under-modeling uncertainty which exists for practical reasons. Such uncertainty knowledge can be obtained through experience or engineering intuition.
The motivation behind this is the quantification of the partially known non-parametric uncertainty in terms of parametric uncertainty for parametric robust control designs.

(2) The research in integrating parameter set estimation and robust control design together is still in an infant stage. A recent paper by Lau et al.[66] has begun discussing the integration of robust control design with ellipsoidal parameter set uncertainty, and there remains much opportunity and need for future research. In [66], only FIR plants are considered, research on robust control design for ARX and ARMAX systems with ellipsoidal parameter set uncertainties is still open.

(3) Investigate different ways, such as experiment designs and choice of model order, to reduce the conservatism of the ellipsoid algorithm. If system stability is \textit{a priori} known, one can also investigate and eliminate subsets in the parameter set which correspond to unstable systems to reduce the size of the feasible parameter set. This uncertainty reduction has a strong bearing on the trade-off in robust control design.

(4) Revisit the interconnected system case, compare the conservatism of the ellipsoidal sets obtained through an instantaneous bound method and an energy bound method.

(5) Modify the algorithm developed in Chapter 4 to make the first step of the proposed two-step approach recursive for real-time implementation.


