Worst-Case Identification and Complexity under Bounded but Low Correlated Noise

by

Kuang Han Chen

Submitted to the Department of Aeronautics and Astronautics in partial fulfillment of the requirements for the degree of Master of Science in Engineering at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 1995

© Massachusetts Institute of Technology 1995. All rights reserved.

Author ..........................................................
Department of Aeronautics and Astronautics
January 17, 1995

Certified by ...........................................
John Tsitsiklis
Professor
Thesis Supervisor

Certified by ...........................................
Munther Dahleh
Associate Professor
Thesis Supervisor

Accepted by ...........................................
Professor Harold Y. Wachman
Chairman, Departmental Committee on Graduate Students
Worst-Case Identification and Complexity under Bounded but Low Correlated Noise

by

Kuang Han Chen

Submitted to the Department of Aeronautics and Astronautics on January 17, 1995, in partial fulfillment of the requirements for the degree of Master of Science in Engineering

Abstract

This thesis investigates the worst-case system identification problem. In the first part we study a certain class of linear systems in the presence of unknown but bounded noise. This class of linear systems is proposed to see if the sample complexity still grows exponentially in the number of parameter to be identified. Unfortunately, the sample complexity for this class turns out to be the same as the class of all linear, shift invariant systems. We also study the sample complexity for multi-input-single-output systems. The second part of the thesis study the worst-case identification problem in the presence of unknown but bounded noise with low correlation. This set of disturbance includes uniformly bounded sequences that satisfy a time-averaged correlation. The motivation for using such a disturbance set is that a white noise signal belongs to the set with high probability if the bound on the correlation approaches zero at certain rate. Also the constraint on the time-averaged correlation forces this noise set to be not as rich as the bounded noise set. Thus result of sample complexity can not be worse than growing exponential. The main results shows a lower bound on the sample complexity that depends on the allowable error and the correlation bound.

Thesis Supervisor: John Tsitsiklis
Title: Professor

Thesis Supervisor: Munther Dahleh
Title: Associate Professor
Acknowledgments

First I would like to thank my advisors, Professor John Tsitsiklis and Munther Dahleh, for their guidance. I greatly appreciate their support and patience for me as an individual. Their mathematical insight and understanding of the field have deeply influenced this thesis.

I also like to thank all my friends at LIDS who make this learning experience an enlightening one. In particular, I like to thank Marcos, Jorge, Venkatesh, Ted, Nicola.

I would like to thank Dan Kuan Huang for her support throughout the progress of the thesis.

Last but not least, I like to thank my parents and family members for their unconditional love and support. This thesis is dedicated to them.

This research was partially supported by the Air Force Office of Scientific Research under grant AFOSR-91-0368.
Contents

1 Introduction ................................................................. 5
  1.1 Problem Statement .................................................. 5
  1.2 Summary of Previous Work .......................................... 8
  1.3 Summary of Thesis .................................................. 9

2 Complexity Analysis of Worst-case Identification of Linear Systems 11
  2.1 Introduction ........................................................... 11
  2.2 Optimal Identification in $\ell_1$ for Increasing F.I.R .............. 12
  2.3 Complexity Analysis for Multi-input Single-output Systems .......... 15
  2.4 Problem Definition ................................................ 17
  2.5 Sample Complexity Analysis ....................................... 18

3 An Algorithm for $H_\infty$ identification for F.I.R. systems .......... 27

4 Optimal Identification Under Low-Correlated Noise .................. 31
  4.1 Introduction ........................................................ 31
  4.2 Preliminaries .................................................... 32
  4.3 A Bound on the worst-case diameter ................................ 33
  4.4 Some Properties on $D_{\delta_T}$ .................................. 39
  4.5 Sample Complexity ................................................ 41

5 Conclusions .............................................................. 45
  5.1 Summary of Results ................................................. 45
  5.2 Future Research .................................................. 46
Chapter 1

Introduction

1.1 Problem Statement

System identification is the field of modeling dynamic systems from experimental data. System identification has been closely linked to control engineering since it is used to get appropriate models for synthesis of a regulator, design of a prediction algorithm, or simulation. Moreover, better control of a dynamic system can only be achieved if we have accurate models for the system. The field of worst-case system identification has received special attention in the 1980s to parallel the development in robust control.

More specifically, as opposed to the previously studied stochastic identification, worst-case identification sets as its goal to capture the inherent uncertainty in our knowledge of any dynamic system. Specifically, given magnitude-bounded but otherwise unknown, additive, output disturbances, we need to obtain bounds on the maximum error in the plant estimation process. These bounds will then be used in the context of the Small Gain Theorem for the design of $\ell_1$ or $H_\infty$ optimal controllers.

In worst-case identification, we assume the disturbance to be bounded on magnitude but otherwise unknown. This assumption on the noise is less restrictive than the usual stochastic noise assumptions. In fact, this kind of noise description can be found in a number of applications, e.g. wind loads on wings of aircraft, vibrations on automobile shock absorbers, etc. Since no statistics are assumed to exist for the dis-
turbance, we are forced to consider worst-case error of the plant estimate, as opposed to average-case error and probabilistic error.

Thus, we can view the worst-case identification as a game played between the experimenter and an omnipotent adversary. Our adversary selects the true plant and the disturbance in order to ruin the accuracy of the estimate. We look for results on minimizing the worst-case error achievable when we pick the input to the plant and the estimate of the plant using some identification algorithm.

Our goal is to investigate the optimal error achievable when we are given a model set $\mathcal{M}$. This model set $\mathcal{M}$ provides us with the a priori knowledge about the true plant. For example, this knowledge could be that the system has an impulse response of length $N$. Furthermore, we would also like to find out the length of the experiments needed in order to achieve optimal or near-optimal error. In this thesis, we investigate the following model set:

$$M_N = \left\{ f \in \mathbb{R}^N \mid f_i = 0, \forall i \notin [1, N] \right\}$$

One of the key concepts in worst-case identification is that of an uncertainty set. This set contains all the plants within the model set that are consistent with the input-output relation. Therefore, every member of the uncertainty set could be the true plant in the worst-case setup. In particular, let

$$U_n = \{ u \in \mathbb{R}^n \mid \|u\|_\infty \leq 1 \text{ and } u_i = 0, \forall i \notin [1, n] \}$$

be the set of allowed inputs. We assume that we are dealing with plants of the form

$$y = u * f + d$$

where $y$ is the output, $u$ is the input, $d$ is the disturbance and $*$ signifies convolution. Let the truncation operator be an operator truncates from a vector of dimension $\omega$ to a vector of dimension $n$

$$P_n : \mathbb{R}^\omega \rightarrow \mathbb{R}^n, \ \omega > n$$
be defined by

\[ P_n(a_1, a_2, \ldots, a_n, a_{n+1}, \ldots) = (a_1, a_2, \ldots, a_n) \]

**Definition 1** We introduce two different norms for signals.

**1-norm.** The 1-norm of a signal \( u \) is the summation of its absolute value:

\[ \|u\|_1 := \sum_{k=-\infty}^{\infty} |u[k]| \]

Note that the 1-norm of a system is defined in the same.

**\( \infty \)-norm.** The \( \infty \)-norm of a signal \( u \) is the least upper bound of its absolute value:

\[ \|u\|_\infty := \sup_k |u[k]| \]

**Definition 2** We define the disturbance set \( D_\delta \) to be the set

\[ D_\delta = \{ d \in \mathbb{R}^\omega \mid \|d\|_\infty \leq \delta \} \]

Then, with the definition of \( D_\delta \) and \( P_n \) we can go ahead and define the following

**Definition 3** Let \( u \in U_n \) and \( y \) be the input applied to the plant and the measured output respectively. The uncertainty set at time \( n \) is defined as

\[ S_n(M_N, u, y, \delta) = \{ f \in M_N \mid \|P_n(f \ast u - y)\|_\infty \leq \delta \} \]

We can easily check that

\[ f \in S_n(M_N, u, y, \delta) \iff \exists d \in D_\delta \text{ such that } P_n(f \ast u + d) = P_n(y) \]

That is, for any member \( f \) in the uncertainty set, there exists a disturbance \( d \) in \( D_\delta \) that would make \( u \ast f + d \) equal to the observed output \( y \) and therefore \( f \) is a possible plant consistent with the observed output. Note that

\[ S_n(M_N, u, y, \delta) = S_{N+k}(M_N, u, y, \delta) \text{, for } u \in U_n \text{ with } k \geq n - N. \]
To provide a measure on this uncertainty set, we define the radius and diameter of a set $A \subset M_N$ as

$$rad(A) = \inf_{x \in A} \sup_{y \in A} \| x - y \|_1$$

$$diam(A) = \sup_{x \in A} \sup_{y \in A} \| x - y \|_1$$

We will use these notions to measure the size of the uncertainty set. We have already seen that in the worst-case setup, our opponent is trying to maximize our error by choosing the true plant and the disturbance. We shall now define an important quantity

**Definition 4** Given a choice of the inputs $u$, define the worst-case diameter $D_k(M_N, u, \delta)$ to be the diameter of the largest possible uncertainty set:

$$D_k(M_N, u, \delta) = \sup_{d \in D_k} \sup_{h \in M_N} diam(S_k(M_N, u, y, \delta))$$

The quantity $D_k(M_N, u, \delta)$ is a measure of the worst-case uncertainty over all the possible output sequences that can be generated by any plant in the model set and any permissible disturbance, for identification experiments of length $k$.

### 1.2 Summary of Previous Work

As we mentioned before, during the last few years, a number of researchers turned their attention to worst-case identification, mainly because of its importance for the modern, robust control methodologies. Robust control has also caused the formulation of two main directions in worst-case identification, one in $\ell_1$ based on time domain measurements ([1, 2, 14, 15]) and one in $H_\infty$ based on frequency domain experiments ([5, 8]).

In particular, [1] used the concepts from Information-Based complexity developed in [10], [11] and [9] to analyze the infinite horizon case of worst-case identification in
a $\sigma$–compact metric model space. It established the existence of an input $u^*$ which guarantees that for all $M \subseteq \ell_1$, we have

$$D_{\infty}(M, u^*, \delta) = 2\delta = \lim_{n \to \infty} \inf_{u \in U_n} D^*_n(M, u, \delta)$$

It also concluded that infinite horizon worst-case identification is equivalent to stability testing. Makila in [14] used the Galois inputs as well in a finite horizon context and showed the corresponding optimality of that input design.

Only very recently have researchers addressed the question of sample complexity for the worst-case identification setup. Sample complexity is the length of experiments required to achieve the optimal worst-case diameter. Dahleh et al in [2] and Poolla et al in [15] developed upper and lower bounds for the length of the input necessary to guarantee a specified level of accuracy in either $M_N$. Also [2] established that in order for an input $u$ to guarantee $D_k(M_N, u, \delta) = 2\delta$, its length has to be exponential in $N$.

### 1.3 Summary of Thesis

This thesis examines worst-case identification from the point of view of algorithms and complexity. Chapter 2, in particular, studies the sample complexity of any worst-case identification algorithm for some finite impulse response multivariable systems. The result is comparable to that obtained for the model set $M_N$ in [2] and [16]. In section 2.2, we focus the worst-case identification on a model set which is a proper subset of $M_N$. Such set is proposed to see whether or not the sample complexity can be reduced. The sample complexity turns out to be comparable because the optimal worst-case diameter is the same for both set. Section 2.3 derives the optimal worst-case diameter for multivariable system. Then we study the sample complexity for the multivariable systems. We provide a lower bound for the minimum length of any input which guarantees optimal worst-case identification of the whole system.

Chapter 3 is devoted to $H_\infty$ identification for finite impulse response system. An algorithm is developed for this type of problem.
Chapter 4 studies worst-case identification for a different kind of noise. This noise is referred to as deterministic white noise. It is assumed to be bounded and low-correlated. This set of disturbance includes uniformly bounded sequences that satisfy a time-averages correlation condition. The noise described is contained in $D_\delta$, as result we hope by shrinking the size of the noise set, we can reduce the size of the worst-case diameter and sample complexity.

The motivation for using such a disturbance set is that the disturbances resemble stochastic white noise processes. In particular, white noise belongs to the set with high probability if the bound on the correlations approaches zero at a certain rate. In this chapter, the lower bound of worst-case diameter for low-correlated noise is derived. We also present an upper bound of worst-case diameter for the same noise given a specific input. The input chosen is to take the full advantage of the time-averages condition, thus minimizing the worst-case diameter. From the previous work we know that the worst-case error is related to the worst-case diameter. The remaining chapters present analysis of sample complexity for the model set $M_N$ corrupted by low-correlated noise. Finally, Chapter 5 offers a summary of the results in the thesis and suggestions for future research.
Chapter 2

Complexity Analysis of
Worst-case Identification of Linear Systems

2.1 Introduction

As was stated in the previous chapter, the worst-case setup for the model set $M_N$ requires inputs whose length grows exponentially in the number of parameters identified. This is a consequence of a very important result in [13], [17]. That particular result states that for all linear stable plants, accurate identification in the $\ell_1$ sense can only be achieved if and only if the input excites all possible frequencies on the unit circle. This is due to the richness of the bounded noise and the inherited difficulty in identification problem under the $\ell_1$ norm. The worst-case identification problem has been studied in the context of asymptotic behavior. In other words, researchers have focused on how well any algorithm could approximate an unknown linear plant in some induced norm sense, given an infinite horizon information stream. The goal is to come up with a plant estimate that approximates the true plant as closely as possible using available input/output information. All the algorithms that guarantee accurate identification in either $\mathcal{H}_\infty$ or $\ell_1$ are based on the Occam’s Razor principle.
by which the simplest model is always used to explain the given data. In the work of [14], it is shown that the standard Chebyshev algorithm, using a Galois sequence as input, would give an asymptotic error bounded by the optimal worst-case diameter of the uncertainty set. The following section presents some results regarding the sample complexity of worst-case identification for a class of F.I.R. systems. The set of the F.I.R. systems chosen here is a proper subset of the model set $M_N$. First, it is shown that the optimal worst-case diameter is $2\delta$ which is the same as that of the model set $M_N$. Then since the optimal worst-case diameter is the same for both sets, one can compare their sample complexity. And because the model set studied here is smaller than $M_N$, the sample complexity could not be longer than that obtained for $M_N$. By reducing the size of the model set, the sample complexity provide some qualitative information about the disturbance set in the direction which the model set get smaller. In general, we hope to get shorter sample complexity results than that in [2] since our set is smaller than that studied in [2]. We also investigate the sample complexity for the case of certain multivariable system. Although we did not investigate the minimum experiment length for suboptimal identification in the worst-case, the analysis can be followed from [2].

2.2 Optimal Identification in $\ell_1$ for Increasing F.I.R

In this section, we concentrate on the following model sets:

$$M_N^+ = \{ f \in \mathbb{R}^N \mid f_i = 0, \forall i \notin [1, N] \text{ and } f_{i+1} \geq f_i \}$$

$$M_N^- = \{ f \in \mathbb{R}^N \mid f_i = 0, \forall i \notin [1, N] \text{ and } f_{i+1} \leq f_i \}$$

Both model sets $M_N^+$ and $M_N^-$ are contained in $M_N$, and we know that the optimal worst-case diameter for both model sets $M_N^+$ and $M_N^-$ are $2\delta$. Thus the sample complexity for $M_N^+$ can not be worse than exponential in $N$. The sample complexity of these two model sets serves as a measure of the richness of the disturbance set in the direction on which we have reduced the size of the set from $M_N$ to $M_N^+$. If the sample complexity of $M_N^+$ remains the same, then one can say that the noise is rich
in the direction where we have reduced the model set from \( M_N \) to \( M_N^+ \).

Here we just need to investigate the sample complexity of worst-case identification in the presence of bounded measurement noise for the model set \( M_N^+ \), under the \( \ell_1 \) norm. The same result should apply to \( M_N^- \). Any element of \( M_N^+ \) will be identified with a finite sequence \( (h_1, \ldots, h_N) \in \mathbb{R}^N \). We have the following experiment setup: an input \( u \in U_n \) is applied to an (unknown) system \( h \in M_N^+ \), and we observe the noisy measurement

\[
y_n = \sum_{i=1}^{N} h_i u_{n-i} + d_n
\]

where \( d \in D_\delta \) plays the role of an output disturbance or measurement noise. It is obvious that for \( i > N + n \) the output \( y_i \) does not contain useful information on the unknown system \( h \). In [1], it is shown that the optimal worst-case diameter \( D_\infty(M_N^+, u^*, \delta) \) of the infinite horizon uncertainty set is equal to \( 2\delta \). Moreover, any algorithm that provides an estimate within the uncertainty set has an error upper-bounded by the diameter of the uncertainty set. Therefore, with a feasible algorithm, as the length of the experiments increases, the worst-case error can be made as small as twice the disturbance bound \( \delta \), but no smaller than \( \delta \). Now we want to determine how long should the experiment be for the error to approach \( 2\delta \). Let us define

**Definition 5**

\[
n^*(N) = \min \{ n \mid D_n(M_N^+, u^*, \delta) = 2\delta \}
\]

This definition gives us the minimum length of the optimal input such that the optimal error is achieved.

**Theorem 6** For any \( \delta > 0 \) and \( N \) being the length of the increasing F.I.R., we have \( 2^{N-1} + N - 1 \leq n^*(N) \leq 2^N + N - 1 \).

**Proof** The upper bound follows easily by using the input sequence proposed [1]. The Galois sequence is shown in [14] to provide most satisfactory worst-case error bounds under \( \ell_1 \) norm for F.I.R. models. The Galois input \( u \) is a binary sequence of length \( n \) such that \( |u_i| = 1, i = 1, \ldots, n \). Furthermore, let \( u \) have the property that it
is a generating sequence for all $N$-tuples of the binary numbers $\{-1, 1\}$. Thus if $b_N$ denotes an arbitrary $N$-tuples of the numbers in $\{-1, 1\}$ then we can find an integer $m \ (1 \leq m \leq n - N + 1)$ such that

$$(u_m, u_{m+1}, \ldots, u_{m+N-1}) = b_N$$

And the minimal length of a generating sequence for all $N$-tuples is $2^N + N - 1$. With this input, the worst case diameter is equal to $2\delta$ since all the possible combinations of $\phi \in \{-1, 1\}^N$ are contained in this input of length $n$. Now for the lower bound, we will show that if $n^*(N) < 2^{N-1} + N - 1$ then we can construct a counterexample such that $\|h - g\|_1 > 2\delta$ but $\|u * (h - g)\|_1 < 2\delta$. Here both $g$ and $h$ belong to $M_N^+$. This is to say that there exists a choice for the disturbance sequences $d' \neq d$ under which the observed output $y = h * u + d$ can not be distinguished from $g * u + d'$.

Let $h_i$ be either $\frac{2\delta(i+1)}{N-\frac{1}{2}}$ or $\frac{-2\delta}{N-\frac{1}{2}}$ for each $i$. Then choose $g_i$ to be the other choice, i.e. different from $h_i$. Both $h$ and $g$ belong to the model set $M_N^+$ since $h_{i+1} \geq h_i$ and $g_{i+1} \geq g_i$.

Now note that $h_i - g_i$ can be either $\frac{2\delta}{N-\frac{1}{2}}$ or $-\frac{2\delta}{N-\frac{1}{2}}$. Let $\phi \in \{-1, 1\}^N$, it is clear that the number of different choice for $\phi$ is $2^N$. From these $2^N$ choices for $\phi$, half of them are the negative of the other half. That is, we can group $2^{N-1}$ choices of $\phi$ and the remaining choices of $\phi$ are obtained from $-\phi$. Thus if $n - N + 1 < 2^{N-1}$, we can pick a $\phi$ such that both $\phi$ and $-\phi$ are different from $(u_{i+N-1}, u_{i+N-2}, \ldots, u_i)$ for all $i \in \{1, \ldots, n - N + 1\}$. Suppose that

$$h - g = \frac{\phi 2\delta}{N - \frac{1}{2}}$$

We then have

$$\|h - g\|_1 = \sum_{i=1}^{N} |h_i - g_i| = \sum_{i=1}^{N} \frac{2\delta}{N - \frac{1}{2}} > 2\delta$$

(2.3)
Now let us examine $\|u \ast (h - g)\|_\infty$. We have

$$\|u \ast (h - g)\|_\infty = \sup_i \left| \sum_{k=1}^N u_{i-k} \phi_k \frac{2\delta}{N - \frac{1}{2}} \right|$$

(2.4)

By our choice of $\phi$, the signs of $u_{i-k}$ can not be the same as the signs of $\phi_k$ for all possible values of $k$ and $i$, neither the same as the signs of $-\phi_k$. Also we have $|\phi_k| = 1$ and $|u_{i-k}| \leq 1$. Thus we have the following inequality

$$\left| \sum_{k=1}^N u_{i-k} \phi_k \right| \leq N - 1$$

(2.5)

plug this back to Eq. (2.4), we have

$$\|u \ast (h - g)\|_\infty = \frac{2\delta(N - 1)}{N - \frac{1}{2}} < 2\delta.$$  

Therefore, for $u$ to be rich enough, a segment of $u$ must match either $\phi$ or $-\phi$.

2.3 Complexity Analysis for Multi-input Single-output Systems

This section is devoted to the sample complexity of multivariable systems. All the plants considered in this section belong to the model set $M_N$ defined in chapter 1, namely F.I.R. plants with length $N$. Let $n$ be the number of the F.I.R. plants to be identified and $u_i$ be the input to the $i$th plant ($1 \leq i \leq n$). And each $u_i$ belongs to $\{-1, 0, 1\}^m$ instead of $\{-1, 1\}^m$, where $m$ is again the length of the experiments. The input-output relation is given by $y = \sum_{i=1}^n u_i \ast h_i + d$, here each $h_i \in M_N$ and we will use $h_i[j]$ to denote the $j$th component of $h_i$. We need to modify some of the definitions given for single-input-single-output analysis.

**Definition 7** Let each $u_i \in \{-1, 0, 1\}^m$ and $y$ be the input applied to the $i$th plant and the measured output respectively. The uncertainty set at time $k$ is defined to be
\[ S_k(M_N, u_1, \ldots, u_n, y, \delta) = \{ f_1, \ldots, f_n \in M_N \mid \left\| P_k(\sum_{i=1}^{n} f_i * u_i - y) \right\|_\infty \leq \delta \}. \]

Now to simplify some writing, let \( f = (f_1, \ldots, f_n) \), where \( f \in M_N \) means that each element of \( f \) belongs to \( M_N \). Also let \( u = (u_1, \ldots, u_n) \) and \( u \in \{-1, 0, 1\}^m \) means that each element of \( u \) belongs to \( \{-1, 0, 1\}^m \). We can easily check that

\[ f \in S_k(M_N, u, y, \delta) \iff \exists d \in D_\delta \text{ such that } P_n(\sum_{i=1}^{n} f_i * u_i + d) = P_n(y) \]

That is, for any member \( f \) in the uncertainty set, there exists a disturbance \( d \) in \( D_\delta \) that would make \( \sum_{i=1}^{n} u_i * f_i + d \) equal to the observed output \( y \).

Now, to provide a measure on this uncertainty set, we extend the definition of radius and diameter given in chapter 1. If \( x = (x_1, \ldots, x_n) \) and \( y = (y_1, \ldots, y_n) \)

\[ rad(A) = \inf_{x \in A} \sup_{y \in A} \sum_{i=1}^{n} \|x_i - y_i\|_1 \]

\[ diam(A) = \sup_{x \in A} \sup_{y \in A} \sum_{i=1}^{n} \|x_i - y_i\|_1 \]

We can then use these notions to measure the size of the uncertainty set. We have already seen that in the worst-case setup, our opponent is trying maximize our error by choosing the true plant and the disturbance.

**Definition 8** Given a choice of the inputs \( u \), define the worst-case diameter \( D_k(u, M_N, \delta) \) to be the diameter of the largest possible uncertainty set:

\[ D_k(u, M_N, \delta) = \sup_{d \in D_\delta} \sup_{f \in M_N} diam(S_k(u, M_N, y, \delta)) \]

The quantity \( D_k(u, M_N, \delta) \) is a measure of the worst uncertainty over all the possible output sequences that can be generated by any plant in the model set and any permissible disturbance.
2.4 Problem Definition

From the problem setup it makes more sense to talk about the collective uncertainty size \( \sum_{i=1}^{n} \|x_i - y_i\|_1 \) rather than the individual uncertainty since we have just one common noise. We also know from the previous work that the individual uncertainty size is greater than \( 2\delta \) (in the worst-case), we can not make it smaller than \( 2\delta \) for any input and any algorithm. From the work in [2], the main result for single-input-single-output states that the minimal length of an input that give the optimal worst-case diameter is \( 2^{N-1} + N - 1 \). We shall first establish lower bound for the quantity \( D_k(u, M_N, \delta) \) which holds for all inputs \( u \) and for any algorithm.

For two set of plants \( \{h_i\}_{i=1}^{n}, \{g_i\}_{i=1}^{n} \in M_N \) such that \( \sum_{i=1}^{n} \|h_i - g_i\|_1 = 2\delta \). Then \n
\[
D_k(u, M_N, \delta) \geq 2\delta
\]

(2.6)

This is a direct consequence of the work in [1] since for each individual plant the uncertainty size is greater than \( 2\delta \). However, this quantity provides some interesting points. If the experiment can be allowed to go on forever, is it possible for \( D_k(u, M_N, \delta) \) to be \( 2\delta \)?

The answer to the question is yes. Assume after 2 independent experiment we have observed 2 identical outputs \( y \) and \( y' \), where \( y = \sum_{i=1}^{n} u_i \star h_i + d \) and \( y' = \sum_{i=1}^{n} u_i \star g_i + d' \) (\( d, d' \in D_\delta \)). Then we observe the following

\[
\sum_{i=1}^{n} u_i \star (h_i - g_i) = d' - d \Rightarrow \left\| \sum_{i=1}^{n} u_i \star (h_i - g_i) \right\|_\infty = \|d' - d\|_\infty \leq 2\delta
\]

(2.7)

Thus if the experiment is long enough, we can have each input \( u_i \) matching the sign of \( h_i - g_i \) and \( \left\| \sum_{i=1}^{n} u_i \star (h_i - g_i) \right\|_\infty = \sum_{i=1}^{n} \|(h_i - g_i)\|_1 \leq 2\delta \). This is equivalent to say that if the experiment is sufficiently long, then for \( \sum_{i=1}^{n} \|(h_i - g_i)\|_1 > 2\delta \) we can not have \( y \) and \( y' \) to be identical.

There seems to be a contradiction since we know for each plant the optimal worst-case diameter is \( 2\delta \), and from the above result it seems that we can do bet-
ter collectively. If we let $D_k^i(u, M_N, \delta)$ be the associated worst-case diameter for each $i$th plant, then the above result says that for long experiments we can have $\sum_{i=1}^n D_k^i(u, M_N, \delta) = 2\delta$. However, we only know that we shrink the total sum of the diameter to $2\delta$, and this means that each diameter can not be greater $2\delta$. Thus there is no discrepancy between the previous result and our result. The reason behind this is that for the SISO case, we can have two plants ($h$ and $g$) $2\delta$ apart in $\ell_1$ norm giving the same output, but this does not mean that $|h[i] - g[i]|$ has a magnitude equal to $\frac{2\delta}{N}$ for $1 \leq i \leq N$. That is, we can have two plants $h$ and $g$ such that the difference between the two vanishes everywhere except at a particular $i$, $|h[i] - g[i]| = 2\delta$ but $|h[j] - g[j]| = 0$ for $j \neq i$. So the multivariable case can be transformed to the SISO identification problem with the model set $M_{nN}$ (i.e. the length of the F.I.R. plant is $nN$). So the term $\sum_{i=1}^n \|(h_i - g_i)\|_1$ can be thought to be $\|H - G\|_1$, where $H$ and $G \in M_{nN}$.

### 2.5 Sample Complexity Analysis

The question addressed in this section is how long should each input be, so the worst-case diameter of the uncertainty set $D_k(u, M_N, \delta)$ can be reduced to $2\delta$.

We need the following preliminary definition for the analysis. For an input of length $m$, we can expand the input into segments of length $N$. Let the input be

$$u = (u(1), u(2), \ldots, u(m)) \in \{-1, 0, 1\}^m$$

By expansion of length $N$ we mean that $v = (u^1, u^2, \ldots, u^q)$, where each $u^k = (u(k), u(k + 1), \ldots, u(k + N - 1))$ for $k \in \{1, \ldots, q\}$ and $q = m - N + 1$. Note each $u^k$ is of length $N$. For example if $N = 3$ and $u = (1, -1, -1, 1, 1)$, then we have $q = 3$ and $v$ to be

$$v = ((1, -1, -1), (-1, -1, 1), (-1, 1, 1))$$

(2.8)

Let $v$ be the expansion of the input $u$. From the expanded input we can see that
we can easily change the order of the segments. And the sample complexity results should not change if we change the order of the segments from the expanded input. Now instead of studying the problem \( m^*(N, n) = \min \{ m \mid D_k(u, M_N, \delta) = 2\delta \} \), we can analyze

\[
q^*(N, n) = \min \{ q \mid D_k(u, M_N, \delta) = 2\delta \}
\]

Here \( m^*(N, n) \) is approximately the required length of the input (if \( n \) is much greater than \( N \)) and \( q^*(N, n) \) is the required number of segments. Note that \( m^*(N, n) \) and \( q^*(N, n) \) are closely related.

In the following analysis, we assume that all the inputs have equal length \( m \) and each \( u_i \in \{-1, 0, 1\}^m \). Thus all the inputs can be expanded into the same number of segments. With the definition of expanded input and segment, the convolution \( u_1 * h_1 + \cdots + u_n * h_n \) at time \( t \) can be thought of as the convolution between \( u^k_1 * h_1 + \cdots + u^k_n * h_n \) where \( u^k_i \) is the \( k \)th expanded input segment to the \( i \)th plant. Now let us define an alignment of the expanded input at time \( k \) to be

\[
\begin{bmatrix}
u^k_1 \\
\vdots \\
u^k_n
\end{bmatrix}
\]

Note that the output at time \( k \) is closely related to this alignment, i.e. \( y[k] = u^k_1 * h_1 + \cdots + u^k_n * h_n \).

**Proposition 9** Given \( n = 2 \), for any \( \delta > 0 \) and \( N \) we have \( q^*(N, n) \geq 2 \cdot 2^{N-1} \).

**Proof** Our final goal is to prove that for general \( n \), the sample complexity increases by a factor of \( n \). For the moment just consider the case when \( n = 2 \). We will show that if \( q < 2 \cdot 2^{N-1} \) then we can construct a counter example with two sets of plant estimates \( (h_1, h_2) \) and \( (g_1, g_2) \in M_N \), and a \( d' \in D_N \) such that the output \( y = u_1 * h_1 + u_2 * h_2 + d \) can not be distinguished from \( u_2 * g_2 + u_2 * g_2 + d' \) while \( \sum_{i=1}^{2} \| (h_i - g_i) \|_1 > 2\delta \). Of course, we assume here that \( g_1 \neq h_1, g_2 \neq h_2 \) and \( d' \neq d \). Instead of showing that a counter example can be constructed for \( q < 2 \cdot 2^{N-1} \), we will just show that when \( q = 2 \cdot 2^{N-1} - 1 \) a counter example can be constructed. Note that for any expanded input of length shorter than \( 2 \cdot 2^{N-1} - 1 \) the same counter example can be used.

From the SISO case it is shown in [2] that each \( u_i \) must have \( 2^{N-1} \) segments of \( \{-1, 1\}^N \), i.e. \( n(N) \geq 2^{N-1} + N - 1 \iff q \geq 2^{N-1} \). As a result of this limitation,
we as the experimenter can only choose \( q - 2^{N-1} = 2^{N-1} - 1 \) segments freely from \( \{-1, 0, 1\}^N \), i.e. to identify each plant accurately within \( 2\delta \) we need \( 2^{N-1} \) segments from the Galois sequence and the total number of segments in each input is \( q \). We now define that a zero segment is a segment of \((0, 0, \ldots, 0)\) of length \( N \) and let \( \alpha_N \) denotes an arbitrary \( N \)-tuple of the number \( \{-1, 1\} \). Note that \( v_1 \) and \( v_2 \) can have at most \( 2^{N-1} - 1 \) zero segments. The proof uses the pigeon hole principle and the reason behind it is that we know that there are \( 2^{N-1} \) segments of \( \{-1, 0, 1\}^N \) from \( v_1 \) and \( v_2 \) can only have \( 2^{N-1} - 1 \) zero segments. Thus \( 2^{N-1} \) segments of \( \{-1, 0, 1\}^N \) can be thought to be \( 2^{N-1} \) pigeons and we only have \( 2^{N-1} - 1 \) holes, so there is at least one pigeon without a house.

Let \( a_i \) be the \( i \)th segment from \( v_1 \) and \( b_i \) be the \( i \)th segment from \( v_2 \) \((1 \leq i \leq 2 \cdot 2^{N-1})\). We then break the problem into two cases and provide a counter example for each case. Case 1: When \( v_1 \) has more or the same number of zero segments than \( v_2 \). Let’s say there are \( k \) zero segments in \( v_1 \) \((0 \leq k \leq 2^{N-1} - 1)\).

\[
\begin{bmatrix}
v_1 \\
v_2
\end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_q \\
b_1 & \cdots & b_q
\end{bmatrix}
\]

(2.9)

If \( a_i \) and \( b_i \) are segments of length \( N \), \( a_i \) sits on top of \( b_i \) means that they are aligned, i.e. \( \begin{bmatrix} a_i \\
b_i
\end{bmatrix} \). Thus for \( \begin{bmatrix} v_1 \\
v_2
\end{bmatrix} \) there are \( q \) alignments. Note that no matter how we change the order of the segments we always have one unique segment of \( \{-1, 1\}^N \) (there are \( 2^{N-1} \) of them) from \( v_1 \) sitting right on top of a nonzero segment from \( v_2 \). Now this is true because there are at most \( 2^{N-1} - 1 \) zero segments in \( v_2 \). Let \( \phi \) be that segment of \( \{-1, 1\}^N \) from \( v_1 \) and let \( \psi \) be the nonzero segment sitting right below \( \phi \) from \( v_2 \). First \( \phi \) can be chosen so that it is unique because there are \( 2^{N-1} \) choices of them, and there are only \( 2^{N-1} - 1 - k \) segments free to be determined from \( v_1 \) (Note we have \( k \) zero segments). In fact there are at least \( k + 1 \) unique \( \{-1, 1\}^N \) in \( v_1 \). Because \( v_2 \) has at most \( k \) zero segments, this implies the existence of unique \( \phi \) aligned with \( \psi \) is guaranteed because of the pigeon hole principle, i.e. \( k + 1 \) choices for \( \phi \) but only \( k \) zero segments. There could be more than one choice for \( \phi \) if \( v_2 \) has
even fewer zero segments.

Now we need to prove that if \( q^* < 2 \cdot 2^{N-1} \) the following is not empty

\[
\{(h_1, h_2), (g_1, g_2) \in M_N | \sum_{i=1}^{2} \|h_i - g_i\|_1 > 2\delta \text{ and } \left\| \sum_{i=1}^{2} u_i * (h_i - g_i) \right\|_\infty \leq 2\delta \}
\]

This is equivalent to showing that if \( q^* < 2 \cdot 2^{N-1} \) then the set of \( \{(h_1, h_2) \in M_N | \sum_{i=1}^{2} \|h_i\|_1 > \delta \text{ and } \left\| \sum_{i=1}^{2} u_i * h_i \right\|_\infty \leq \delta \} \) is not empty since we can pick \( g_i = -h_i \in M_N \) and this will give us back the original condition. The counter example in this case can be chosen as the following:

\[
h_1 = \frac{\delta \phi}{N - \frac{1}{2}}, \quad \|h_1\|_1 > \delta \quad (2.10)
\]

again \( \phi \) is a unique segment of the \( \{-1, 1\}^N \) from \( v_1 \). Also choose \( h_2 \) as the following:

\[
h_2 = -\frac{\delta \psi}{2c(N - \frac{1}{2})}, \quad \|h_2\|_1 < \delta \quad (2.11)
\]

where \( \psi \) is a nonzero segment and \( c = \sum_{i=1}^{N} |\psi[i]| \). Note that \( \sum_{i=1}^{2} \|h_i\|_1 > \delta \) since \( \|h_1\|_1 > \delta \). The multiplication between segments is defined to be

\[
a_i \times a_j = \sum_{k=1}^{N} a_i[k] \cdot a_j[k].
\]

We note that \( |\psi \times \psi| = c \) since \( \psi \times \psi = \sum_{k=1}^{N} \psi[k]^2 = \sum_{i=1}^{N} |\psi[i]| \) (i.e. each \( \psi[k] \in \{-1, 0, 1\} \)). Then for a segment \( b_j \) different from \( \psi \) we have

\[
|\psi \times b_j| = d \leq c
\]

(i.e. \( \psi \) and \( b_j \) are two different segments). This is true since \( b_j \) might not match all the signs of \( \psi \). With the definition of multiplication, the convolution in the \( y \) expression can be converted to multiplication between segments.

Now we need to show that the quantity \( \|u_1 * h_1 + u_2 * h_2\|_\infty \) is less than \( \delta \). The possible maximum values for \( |u_1 * h_1 + u_2 * h_2| \) are:
1. When the segment $\phi$ of the input $v_1$ is multiplied by the plant $h_1$, and the segment $\psi$ of the input $v^2$ is multiplied by the plant $h_2$. Note that $\phi$ is sitting on top of $\psi$, so when $\phi$ is multiplied by the first plant, $\psi$ must be multiplied to the second plant.

$$|\phi \times h_1 + \psi \times h_2| = \frac{\delta N}{N - \frac{1}{2}} \frac{\delta}{2(N - \frac{1}{2})} = \delta \leq \delta. \quad (2.12)$$

2. When any other $a_j$ segment different from $\phi$ in $v_1$ is multiplied by $h_1$ and $b_j$ in $v_2$ is multiplied by $h_2$.

First $a_j$ can not match all the signs of $\phi$, so the best $|a_j \times h_1|$ can do is to be $N - 1$.

$$|a_j \times h_1 + b_j \times h_2| \leq |a_j \times h_1| + |b_j \times h_2| = \frac{\delta(N-1)}{N-\frac{1}{2}} + \frac{\delta b}{2a(N-\frac{1}{2})}$$

$$\leq \frac{\delta(N-\frac{1}{2})}{N-\frac{1}{2}} = \delta. \quad (2.13)$$

This can also be seen that we picked a unique alignment, so the rest of the alignments must be differed at least by one element. This argument will be used to prove the general case. We will provide a second proof for this case later in the chapter. Therefore $\|u_1 \ast h_1 + u_2 \ast h_2\|_\infty = \delta \leq \delta$.

To summarize, the existence of $\phi$ and $\psi$ is guaranteed due to the following argument:

- $\phi$ exists in $v_1$ since there are $2^{N-1}$ choices for $\phi$ but only $2^{N-1} - 1$ segments free to be determined (i.e. if we try to put $2^{N-1} - 1$ toys into $2^{N-1}$ boxes and each box can contain more than one toy, at least one box will be empty.). So there is at least one choice for $\phi$ which appears only once in $v_1$.

- If there are $k$ zero-segments in $v^1$, there would be $k + 1$ choices for $\phi$. The analogy would be if there are $2^{N-1}$ boxed but only $2^{N-1} - 1 - k$ toys ($-k$ because we have $k$ zero segments), so there would be at least $k + 1$ boxes empty. Since it is assumed that $v^2$ have at most $k$ zero segments and we have $k + 1$
choices for \( \phi \), there is at least one choice for \( \phi \) that is sitting on top of a nonzero segment. As a result, \( \psi \) is chosen.

Case 2: \( \nu_2 \) has more zero-segments than \( \nu_2 \). The proof follows as in case 1. We just interchange the choices for \( h_1 \) and \( h_2 \). The rest of the proofs follows as in case 1.

**Corollary 10** If \( q = 2 \cdot 2^{N-1} \), then each alignment of
\[
\begin{bmatrix}
a_i \\
b_i
\end{bmatrix}
\]
must be unique. That is, only for \( q \) greater than \( 2 \cdot 2^{N-1} \), there might be repeated alignments.

**Proof** Assume we have two alignment of the same, e.g.
\[
\begin{bmatrix}
\ldots & a_i & \ldots & a_j & \ldots \\
\ldots & b_i & \ldots & b_j & \ldots
\end{bmatrix}
\]
where \( a_i = a_j \) and \( b_i = b_j \). Then we can take one alignment out since they both give us the same output (with some choice of \( d \)). Since we can provide a counter example when \( q = 2 \cdot 2^{N-1} - 1 \), we can provide the same counter example when \( q = 2 \cdot 2^{N-1} \) with one or more repeated alignment. In fact we can say more about each alignment. Let \( \alpha_N \) denotes an arbitrary \( N \)-tuple of the number \( \{-1, 1\} \), then each alignment is of the form
\[
\begin{bmatrix}
\alpha_N \\
b_i
\end{bmatrix}
\]
or
\[
\begin{bmatrix}
a_j \\
\alpha_N
\end{bmatrix}
\]
(where \( b_i \) and \( a_j \) are \( \{-1, 0, 1\}^N \)). If one of the alignment is of the form
\[
\begin{bmatrix}
a \\
b
\end{bmatrix}
\]
where \( a \) and \( b \) are not \( N \)-tuple of the number \( \{-1, 1\} \) (i.e. \( a \) and \( b \) contains at least one zero in the segment), then we can construct a counter example by ignoring this alignment (i.e. \( q = 2 \cdot 2^{N-1} - 1 \)). And this counter example would still work even if we throw in this extra alignment since \( a \) and \( b \) can not match the signs of the counter example.

**Corollary 11** For \( n = 2 \), let \( \alpha \) and \( \beta \) be segments of length \( N \) of a Galois sequence. If \( q < 2 \cdot 2^{N-1} \), then there exists a unique alignment of
\[
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}
\]
in the expanded sequence.

**Proof** As we mentioned at the beginning, the reason of using the expanded input is that we can easily interchange the order of the segments. Note that once the inputs are chosen each alignment is fixed, though we can interchange the order of the alignment since we just interchange the value of the output. However, we can
not change only the order of segments in $v_1$ without changing the order of segments in $v_2$ since doing so will produce different output. Thus let us group the first $2^{N-1}$ alignments that contains all possible segments of Galois sequence from the expanded input $v_1$ (Note there is total $2 \cdot 2^{N-1} - 1$ alignments). Let $\alpha_i$ be the $i$th segment of a Galois sequence. Then the first $2^{N-1}$ alignments looks like the following

$$
\begin{bmatrix}
\alpha_1 & \ldots & \alpha_i & \ldots & \alpha_{2^{N-1}} \\
 b_1 & b_i & b_{2^{N-1}} 
\end{bmatrix}
$$

(2.14)

Now let us call each $\alpha_i$ a Galois segment. Assume there are $k$ distinct Galois segments from $v_2$ in the first $2^{N-1}$ alignments. We know that there is at least one Galois segment since $v_2$ contains $2^{N-1}$ Galois segments and there are only $2 \cdot 2^{N-1} - 1$ total alignments. Let us define a zero alignment is an alignment containing a zero segment from $v_1$ or $v_2$. Thus there are $k$ ($1 \leq k \leq 2^{N-1}$) unique nonzero alignments in the first $2^{N-1}$ alignments since there are $k$ distinct Galois segments from $v_2$ and each $\alpha_i$ from $v_1$ is nonzero segment. Now for the rest $2^{N-1} - 1$ alignments, there are only $k - 1$ alignments free to be chosen (When we say an alignment is free to be chosen we mean that we can choose freely both $a$ and $b$ of $\begin{bmatrix} a \\ b \end{bmatrix}$). This is true since we have $2^{N-1} - k$ Galois segments from $v_2$ in the last $2^{N-1} - 1$ alignments, i.e. we have grouped the first $2^{N-1}$ alignment, so only segments from $v_1$ are free to be chosen. Thus by pigeon hole principle, there is one unique nonzero alignment from the first $2^{N-1}$ ($k$ pigeons and $k - 1$ holes). Now assume this unique alignment is $\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ where $\alpha$ and $\beta$ are both Galois segments. That is why a counter example such that $\sum_{i=1}^{2} \|h_i\|_1 > \delta$ but $\|u_1 \ast h_1 + u_2 \ast h_2\|_\infty$ is less than $\delta$ can be constructed if $q < 2 \cdot 2^{N-1}$. The following counter example is different from the previous proof.

Choose $h_1 = \frac{\delta \alpha}{2^{N-\frac{1}{2}}}$ and $h_2 = -\frac{\delta \beta}{2^{N-\frac{1}{2}}}$ (Note that $\|h_1\|_1 + \|h_2\|_1 = \frac{\delta (2N-1)}{2^{N-\frac{1}{2}}}$). It can be shown that $\|u_1 \ast h_1 + u_2 \ast h_2\|_\infty = \frac{\delta (2N-1)}{2^{N-\frac{1}{2}}} < \delta$ since either $u_1$ misses a sign of $\alpha$ or $u_2$ misses a sign of $\beta$.

**Corollary 12** Assume we have two F.I.R. systems with different length (i.e. $h_1 \in$
\[ M_{N_1} \text{ and } h_2 \in M_{N_2} \). Then \( q^* \geq 2^{N_1-1} + 2^{N_2-1} \).

**Proof** Here we will assume that \( N_1 \geq N_2 \). Now we need to expand \( v_1 \) into segments of length \( N_1 \) and \( v_2 \) into segments of length \( N_2 \). Then again from the equation \( u_1 \ast h_1 + u_2 \ast h_2 \) at each time \( t \), we will look at one alignment. For \( D_k(u, M_N, \delta) = 2\delta \), the diameter of the uncertainty of each F.I.R. system must be less or equal to \( 2\delta \). Thus we need \( 2^{N_1-1} \) Galois segments for \( h_1 \) and \( 2^{N_2-1} \) Galois segments for \( h_2 \). As in the previous proof, when \( q = 2^{N_1-1} + 2^{N_2-1} - 1 \) we just need to provide a counter example such that \( \sum_{i=1}^{2} ||h_i||_1 > \delta \) but \( ||u_1 \ast h_1 + u_2 \ast h_2||_\infty < \delta \). If we rearrange the order of the alignments of \( \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \) so that the first \( 2^{N_1-1} \) alignments contains all the Galois segments from \( v_1 \) and assuming that there are \( k \) distinct Galois segments from \( v_2 \) in the first \( 2^{N_1-1} \) alignments. Since \( q = 2^{N_1-1} + 2^{N_2-1} - 1 \), we know that \( k \geq 1 \). Then from the corollary we can say that there is one unique alignment \( \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \) where \( \alpha \) and \( \beta \) are both Galois segments of length \( N_1 \) and \( N_2 \). Thus a counter example is chosen as in the corollary.

Now we can go ahead and prove the case for \( n = 3 \). We will again expand the inputs \( \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \) and we need to show that \( q^* \geq 3 \cdot 2^{N-1} \). First we know that \( q^* \geq 2 \cdot 2^{N-1} \), this is a consequence of \( n = 2 \). Let \( v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \), then we can consider \( \begin{bmatrix} v \\ v_3 \end{bmatrix} \) where we know \( v \) has at least \( 2 \cdot 2^{N-1} \) segments and \( v_3 \) needs at least \( 2^{N-1} \) segments. Then from the corollary we can show that if \( q = 3 \cdot 2^{N-1} - 1 \), then there exists a unique alignments \( \begin{bmatrix} \gamma \\ \beta \end{bmatrix} \) where \( \gamma \) is of the form \( \begin{bmatrix} \alpha \\ b \end{bmatrix} \) or \( \begin{bmatrix} a \\ \alpha \end{bmatrix} \) (\( \alpha \) is a Galois segment of length \( N \), and \( a, b \in \{-1, 0, 1\}^N \)) and \( \beta \) is a Galois segment. Since the alignments is unique, a counter example can be constructed. If the alignment is \( \begin{bmatrix} \alpha \\ b \\ \beta \end{bmatrix} \) and assume \( \|b\|_1 = \sum_{i=1}^{N} |b_i| = c \), choose \( h_1 = \frac{\delta}{2N+c-\frac{1}{2}}, h_2 = \frac{\delta}{2N+c-\frac{1}{2}} \) and \( h_3 = -\frac{\delta}{2N+c-\frac{1}{2}} \). First we
see that $\sum_{i=1}^{3} \| h_i \|_1 = \frac{(2N+c)\delta}{2N+c-\frac{1}{2}} > \delta$, and we verify that

$$\| u_1 \ast h_1 + u_2 \ast h_2 + u_3 \ast h_3 \|_\infty = \frac{(2N + c - 1)\delta}{2N + c - \frac{1}{2}} < \delta.$$  \hspace{1cm} (2.15)

Now assume that this is true for $n$, and we try to prove for $n+1$. Again follows the steps used for the case when $n = 3$. We consider $[v \quad \vdots \quad v_{n+1}]$ where $v = \begin{bmatrix} v_1 \\
v_2 \\
\vdots \\
v_n \end{bmatrix}$. It can be shown if $q = (n + 1) \cdot 2^{N-1} - 1$, then there exists an unique alignment. Thus a counter example can be constructed. \hspace{1cm} \blacksquare
Chapter 3

An Algorithm for $\mathcal{H}_\infty$

identification for F.I.R. systems

The algorithm proposed in [5] for system identification problem in $\mathcal{H}_\infty$ is not quite efficient when the model set consists of F.I.R. systems. In the first step of the algorithm in [5], $n$ noisy point samples of the frequency response of the unknown stable plant are used to construct an $\mathcal{L}_\infty$ approximation to the plant. In the second step of the algorithm, this $\mathcal{L}_\infty$ approximation is mapped into a stable and causal real-rational $\mathcal{H}_\infty$ approximation to the unknown stable plant. Then it is this $\mathcal{H}_\infty$ approximation which serves as the identified plant model. Moreover, if $n$ represents the number of data points, the order of the identified model is roughly $n^3$.

Thus, if we are interested in identifying a F.I.R. system of a small order but use many samples of the frequency response, the algorithm in [5] will produce estimates of too high order. This is because the algorithm in [5] is not based on F.I.R. systems. Thus we propose the following algorithm for worst-case identification in $\mathcal{H}_\infty$ for F.I.R. systems.

Let $y$ be the observed output corrupted by bounded noise.

$$y[n] = (u * h)[n] + d[n] = \sum_{k=0}^{N-1} h[k]u[n - k] + d[n], \text{ where } |d[n]| \leq \delta \forall n,$$
where the length of the F.I.R. system is assumed to be \( N \). In order to get a sample of the frequency response at frequency \( \omega \), the input is chosen to be

\[
u[n] = \begin{cases} 
e^{j\omega n} & \text{if } n \geq 0 \\ 0 & \text{otherwise} \end{cases}
\]

Define \( \tilde{H}(e^{j\omega}) \) as the standard discrete-time Fourier transform

\[
\tilde{H}(e^{j\omega}) = \sum_{k=0}^{N-1} h[k]e^{-j\omega k}
\]

then the following is true

\[
y[n] - d[n] = e^{j\omega} \tilde{H}(e^{j\omega}), \tag{3.1}
\]

for \( n \geq N \). Now the disturbance is varying in a unit ball of \( L_{\infty} \). Let \( y[n] = y_r[n] + jy_i[n] \), where \( y_r[n] \) is the real part of \( y[n] \) and \( y_i[n] \) is the imaginary part of the \( y[n] \). By doing so and combining the fact that \( |v[n]| \leq \delta \ \forall n \) we have the following linear programming (LP) problem:

\[
y_i[n] - \delta \leq [\sin(\omega(n-0)) \sin(\omega(n-1)) \cdots \sin(\omega(n-N))] \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[N] \end{bmatrix} \leq y_i[n] + \delta
\]

\[
y_r[n] - \delta \leq [\cos(\omega(n-0)) \cos(\omega(n-1)) \cdots \cos(\omega(n-N))] \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[N] \end{bmatrix} \leq y_r[n] + \delta
\]

We then introduce a vector \( z \) of slack variables and let \( h = h^+ - h^- \), where \( h^+ \), \( h^- \geq 0 \). We then write the problem in the standard form and use the simplex method to solve the problem. Suppose that we do experiments at \( m \) different frequencies \( \omega_i \),
\( i=1, \ldots, m \). We then have the following constraints

\[
\begin{bmatrix}
\sin(\omega_1(n - 0)) & \cdots & -\sin(\omega_1(n - N)) \\
-\sin(\omega_1(n - 0)) & \cdots & \sin(\omega_1(n - N)) \\
cos(\omega_1(n - 0)) & \cdots & -\cos(\omega_1(n - N)) \\
-\cos(\omega_1(n - 0)) & \cdots & \cos(\omega_1(n - N)) \\
\vdots & & \vdots \\
\sin(\omega_m(n - 0)) & \cdots & -\sin(\omega_m(n - N)) \\
-\sin(\omega_m(n - 0)) & \cdots & \sin(\omega_m(n - N)) \\
cos(\omega_m(n - 0)) & \cdots & -\cos(\omega_m(n - N)) \\
-\cos(\omega_m(n - 0)) & \cdots & \cos(\omega_m(n - N)) \\
\end{bmatrix}
\begin{bmatrix}
h^+ \\
h^- \\
z
\end{bmatrix} =
\begin{bmatrix}
y_i[n](\omega_1) + \delta \\
y_i[n](\omega_1) - \delta \\
y_r[n](\omega_1) + \delta \\
y_r[n](\omega_1) - \delta \\
y_i[n](\omega_m) + \delta \\
y_i[n](\omega_m) - \delta \\
y_r[n](\omega_m) + \delta \\
y_r[n](\omega_m) - \delta
\end{bmatrix}
\]

and \( h^+, h^-, z \geq 0 \). If \( m \) is the number of the frequency sample points and \( N \) equals to the length of the F.I.R. then we have \( 4m \) constraints with \( 2N \) variables. This problem can be solved by the linear programming algorithm proposed in [16], where the estimate provided is a member of the uncertainty set. The dual LP of the problem is easy to initialize [16] and the CARRY tableau size is reduced in size from \((2n+1)\) by \((2n+1)\) to \((N+1)\) by \((N+1)\) if we use the revised simplex method. Here we have explored several sample identification problems using different software related to LP. First, all the softwares are asked to solve the zero cost LP. By that we mean that we are solving the following LP problem with a dummy cost function \( cx = 0x \):

\[
\min cx \\
\text{subject to } Ax = b, x \geq 0
\]

The zero cost should not give us any problem since all we are interested is finding a member in the uncertainty set. We have the following observations for different software. It is needed to perturb the zero cost function in “mathematica” to get a answer. Also if the system of linear constraints gets too large, more than 80 constraints, “mathematica” could not handle the problem. Secondly, we tried the same samples with “LINDO”. “LINDO” works with most of the samples we tried, but since

29
it is a small program, it can only handle LP problems with 120 constraints with 300 variables.
Chapter 4

Optimal Identification Under Low-Correlated Noise

4.1 Introduction

In this chapter, we focus on the same worst-case identification problem as in previous chapters but with a slightly different disturbance set. The true plant \( h \) again belongs to \( M_N \), and it is identified using one input-output experiment. The input can be chosen freely and the observed output is corrupted by an additive disturbance in the following set

\[
D_{\delta\tau} = \{d \mid \|d(t)\|_\infty \leq \delta \text{ and } \left| \sum_{t=1}^{m} d(t)d(t+\tau) \right| \leq \gamma m \delta^2, \forall t, 1 \leq \tau \leq m - 1 \} \quad (4.1)
\]

Here, \( \gamma \) is a constant between 0 and 1. Let us define the autocorrelation of a sequence to be

\[
r_d(\tau) = \sum_{t=1}^{m} d(t)d(t+\tau) \quad (4.2)
\]

If \( \gamma \) tends to zero, the auto correlation function looks like a delta function.

This disturbance set besides being bounded by \( \delta \), it is described in terms of time
averaged correlation. Thus the noise is restricted to a smaller set. This set of disturbance includes uniformly bounded sequences that satisfy a time-averages correlation condition. The motivation of using this disturbance set is that in some stochastic control problems this disturbance set serves as a measure whether or not a signal is white noise. In particular, a white noise signal belongs to this set with high probability if the bound on the correlations approaches zero at a certain rate [4]. Also, given any stable linear time-invariant system, the variance of the output for a white noise input is given by the $H_2$ norm of the system. It turns out that the $\ell_2$ induced norm of any stable LTI system over the above set of inputs approximates asymptotically the $H_2$ norm of the system [4]. To be able to use this kind of result we need to do a statistical hypothesis test from the experimental data. Thus if a signal belongs to this set, we have reasonable confidence that the signal is white.

We like to point out that the set $D_{\delta r}$ is closely related though somewhat different from the set described in [4]. In fact $D_{\delta r}$ contains the disturbance set in [4]. The reason for the differences is that they simplify the analysis.

In [2] and [15], it is shown that the length of input grows exponentially in the number of estimated parameters even for suboptimal identification. That is, if we relax the worst-case diameter by a constant multiplicative, it would still require a long experiment before the estimated plant is within the uncertainty. This stimulates us to study a new description of noise (almost white) to see and compare its worst-case diameter and sample complexity with that of bounded but otherwise unknown noise. The result is still negative in the sense that even with this new description of the noise, we still get exponential sample complexity; though, as the noise correlation become smaller, the sample complexity gets much smaller.

4.2 Preliminaries

Let $m$ be the number of observations, $m$ is usually related to both the length of the input $n$ and F.I.R. response $N (m = N + n)$. We still use the inputs proposed in the previous chapters. Thus every element of $U_n$ will be an input sequence with length
The input-output relation when an input \( u \in U_n \) is applied to an unknown system \( h \in M_N \), and with noisy measurement is given by

\[
y = h \ast u + d
\]  

(4.3)

Here, it is obvious that the number of the measurements should be \( N + n \) since \( y(i) \) does not contain useful information for \( i > N + n \). The following definition is a modification from previous chapters.

**Definition 13** Let \( S_{N,n}(y, u) \) be the uncertainty set that contains all the plants in the model set consistent with an input \( u \in U_n \) and a measurement vector \( y \).

\[
S_{N,n}(y, u) = \{ \phi \in M_N \mid \| y - \phi \ast u \|_\infty \leq \delta \text{ and } |r_{y-\phi \ast u}(\tau)| \leq \gamma m \delta^2 \},
\]

where \( m = N + n \). The definition of diameter and worst-case diameter is the same as in chapter 1.

### 4.3 A Bound on the worst-case diameter

It is shown in [1] that the worst-case error of any identification algorithm is lower-bounded by half the diameter of the uncertainty set and upper-bounded by the diameter of the uncertainty set. Therefore, we first investigate bounds for worst-case diameter which holds for any input \( u \). Note that the optimal identification error is related to this worst-case diameter.

**Proposition 14** For all \( N \) and \( n \), and all \( u \in U_n \), we have \( D_{N,n}(u) \geq 2 \sqrt{\gamma \delta} \)

**Proof** Let \( g, h \in M_N \) satisfy \( \| g - h \|_1 = 2 \sqrt{\gamma \delta} \). Let \( u \in U_n \) and suppose that \( h \) is the true plant. Let the disturbance be \( d = u \ast (g - h) / 2 \). Note that \( d \) is in the disturbance set \( D_\delta \tau \) (i.e. \( \| d \|_\infty \leq \| u \|_\infty \| \frac{1}{2} (g - h) \|_1 = \sqrt{\gamma \delta} \leq \delta \) and \( |r_d(\tau)| = |\sum_{t=1}^{m} d(t) d(t + \tau)| \leq \gamma m \delta^2 \)). Thus the observed output is

\[
y = u \ast h + d = u \ast h + u \ast \frac{(g - h)}{2} = u \ast \frac{(g + h)}{2}
\]
We want to show that the same output can be generated by a different plant \( g \neq h \) with an admissible disturbance \( d' \) that is different from \( d \). Choose

\[
d' = y - u \ast g = u \ast \frac{(h - g)}{2} \neq d.
\]

We need to verify that \( d' \) is an admissible disturbance. With \( y = u \ast h + d = u \ast (g + h)/2 \), We have

\[
\|d'\|_\infty = \left\| \frac{1}{2} u \ast (h - g) \right\|_\infty \leq \frac{1}{2} \|u\|_\infty \|h - g\|_1 \leq \sqrt{\gamma} \delta \leq \delta. \tag{4.4}
\]

\[
|r_{d'}(\tau)| = \left| \sum_{\tau=1}^{m} \left\{ \frac{1}{2} u \ast (h - g)(t) \right\} \cdot \left\{ \frac{1}{2} u \ast (h - g)(t + \tau) \right\} \right|
\leq \sum_{\tau=1}^{m} \frac{1}{2} |u \ast (h - g)(t)| \frac{1}{2} |u \ast (h - g)(t + \tau)| \leq \sum_{\tau=1}^{m} \left\| \frac{1}{2} u \ast (h - g) \right\|_\infty^2 \leq \gamma m \delta^2. \tag{4.5}
\]

Therefore, \( g \) belongs to the uncertainty set \( S_{N,n}(y, u) \) and \( d' \) is in the disturbance set \( D_{\delta \tau} \). Since \( h \) is also in \( S_{N,n}(y, u) \), it follows that

\[
diam(S_{N,n}(y, u)) \geq \|g - h\|_1 = 2\sqrt{\gamma} \delta. \tag{4.6}
\]

To compute the worst case diameter \( D_{N,n}(u) \) for a given input, we must solve the following problem:

\[
\max_{g, h \in M_N} \|g - h\|_1
\text{ s.t. } u \ast g + d_1 = u \ast h + d_2, \text{ where } d_1, d_2 \in D_{\delta \tau}
\tag{4.7}
\]

If we let \( G = g - h \) and \( d_3 = d_1 - d_2 \), this problem is equivalent to

\[
\max \|G\|_1
\text{ subject to } u \ast G + d_3 = 0
\tag{4.8}
\]

We need to check some properties of \( d_3 \). First \( \|d_3\|_\infty \) is upper bounded by \( 2\delta \), since both \( d_1 \) and \( d_2 \) are in \( D_{\delta} \) \( (\|d_3\|_\infty \leq \|d_1\|_\infty + \|d_2\|_\infty \leq 2\delta) \). Then the autocorrelation term for \( d_3 \) is
\[ r_d(t) = \sum_{i=1}^{d_3}(t_1 - d_2)(t_1 - d_2)(t + \tau) \]
\[ = \sum \{ d_1(t_1) - d_2(t_1) + d_2(t_1 - d_2(t_1) - d_1(t_1) - d_2(t_1 - d_1(t_1)) \} \]
Thus, \[ |r_d(t)| \leq |\sum d_1(t_1) - d_2(t_1)| + |\sum d_2(t_1) - d_2(t_1)| + |\sum d_2(t_1) - d_1(t_1)| \]
\[ \leq 2\gamma m \delta^2 + 2m \delta^2 = 2m(\gamma + 1) \delta^2 \]

Let us focus on a particular input, a repeated Galois sequence. A Galois sequence as defined in previous chapters is a finite sequence that with elements in the set \{+1, -1\}. The reason for selecting Galois sequence as input would become apparent later on. The input used is a repeated Galois sequence in order to take full advantage of the auto correlation constraints.

The auto correlation condition gives us \(m - 1\) constraints on \(d_3\) (i.e. \(|r_d(t)| \leq 2m(\gamma + 1) \delta^2, 1 \leq \tau \leq m - 1\)). Here we want to look at one particular constraint that gives us an upper bound on the worst-case diameter when a repeated Galois input sequence is used. Let us prove the following lemma that will be used in the main theorem.

**Lemma 15** Let \(\overrightarrow{x}_N\) be some real vector of length \(N\). For each \(i \in \{1, 2, \ldots, 2^N\}\), let \(\overrightarrow{a}_N(i)\) be a distinct unique vector of length \(N\) with elements in \{+1, -1\}. Define

\[ \overrightarrow{a}_N(i) \cdot \overrightarrow{x}_N = b_N(i) \]

where \(\cdot\) stands for inner product. Then

\[ \sum_{i=1}^{2^N} (\overrightarrow{a}_N(i) \cdot \overrightarrow{x}_N)^2 = 2^N(x_1^2 + x_2^2 + \cdots + x_N^2) \quad (4.9) \]

**Proof** We will provide 2 ways to prove the lemma. The first way is by induction.

First consider the case when \(N = 2\), we have the following:

\[ \sum_{i=1}^{4} (\overrightarrow{a}_2(i) \cdot \overrightarrow{x}_2)^2 = \{(x_1 + x_2)^2 + (x_1 - x_2)^2 + (-x_1 + x_2)^2 + (-x_1 - x_2)^2\} \]
\[ = 4(x_1^2 + x_2^2) = 2^2(x_1^2 + x_2^2) \]

35
Now assuming this is true for $N$, i.e. $\sum_{i=1}^{2^N} (\overrightarrow{a_N(i)} \cdot \overrightarrow{x_N})^2 = \sum_{i=1}^{2^N} (b_N(i))^2 = 2^N(x_1^2 + x_2^2 + \cdots + x_N^2)$, for $N + 1$ we can have the following condition:

$$\overrightarrow{a_{N+1}(i)} = (\overrightarrow{a_N(i)}, 1), \ i = 1, 2, \cdots, 2^N.$$  

$$\overrightarrow{a_{N+1}(i+2^N)} = (\overrightarrow{a_N(i)}, -1), \ i = 1, 2, \cdots, 2^N.$$  

Note that $\overrightarrow{a_{N+1}}$ has $2^{N+1}$ distinctive vectors. If we let $\overrightarrow{a_{N+1}(i)}$ to be defined as above we can find a relation between $b_{N+1}(i)$ and $b_N(i)$. Since the order in the summation is not important, we can rearrange the order and let the first $2^N$ terms of $b_{N+1}(i)$ be the following:

$$b_{N+1}(i) = b_N(i) + x_{N+1}, \ i = 1, 2, \cdots, 2^N.$$  

and the next $2^N$ terms of $b_{N+1}(i)$ be

$$b_{N+1}(i) = b_N(i) - x_{N+1}, \ i = 1, 2, \cdots, 2^N.$$  

By some algebraic simplification, we can combine $b_{N+1}^2(i)$ and $b_{N+1}^2(2^N + i)$ from the summation, i.e.

$$b_{N+1}^2(i) + b_{N+1}^2(2^N + i) = (b_N(i) + x_{N+1})^2 + (b_N(i) - x_{N+1})^2 = 2(b_N^2(i) + x_{N+1}^2)$$

Thus,

$$\sum_{i=1}^{2^{N+1}} (\overrightarrow{a_{N+1}(i)} \cdot \overrightarrow{x_{N+1}})^2 = \sum_{i=1}^{2^{N+1}} (b_{N+1}(i))^2$$

$$= \sum_{i=1}^{2^N} (b_{N+1}(i))^2 + \left(b_{N+1}(2^N + i)\right)^2$$

$$= \sum_{i=1}^{2^N} 2(b_N^2(i) + x_{N+1}^2)$$

$$= 2 \sum_{i=1}^{2^N} b_N^2(i) + 2 \sum_{i=1}^{2^N} x_{N+1}^2$$

$$= 2 \cdot 2^N(x_1^2 + x_2^2 + \cdots + x_N^2) + 2 \cdot 2^N \cdot x_{N+1}^2$$

$$= 2^{N+1}(x_1^2 + x_2^2 + \cdots + x_N^2 + x_{N+1}^2)$$

(4.10)

The second approach is by means of probabilistic argument. Let $\overrightarrow{a_N}$ be a random vector which is equal to any one of the vector $\overrightarrow{a_N(i)}$, with probability $2^{-N}$. Note that $E[\overrightarrow{a_N(i)}\overrightarrow{a_N(i)}^T] = I$. The diagonal entry is 1 because $E[(\pm1)^2] = 1$, and off diagonal entry is 0 because the off diagonal entry of the product $\overrightarrow{a_N(i)}\overrightarrow{a_N(i)}^T$ can be either $-1$ or 1 both with probability $\frac{1}{2}$. Thus,
\[ \sum_{i=1}^{2^N} \left( \overrightarrow{a_N}(i) \overrightarrow{x_N} \right)^2 = 2^N E \left[ \left( \overrightarrow{a_N}(i) \overrightarrow{x_N} \right)^2 \right] \\
= 2^N E \left[ \overrightarrow{x_N} \overrightarrow{a_N}(i) \overrightarrow{a_N}(i) \overrightarrow{x_N} \right] = 2^N \overrightarrow{x_N} \overrightarrow{I} \overrightarrow{x_N} \] (4.11)

Recall the equation \( u \ast G = -d_3 \) and let \( F = u \ast G \). Then the autocorrelation of \( F \) must satisfy the autocorrelation constraints on \( d_3 \). If this condition is not satisfied, it is not possible for \( h \) and \( g \) to lead to identical outputs.

**Theorem 16** Suppose that we repeat the Galois input \( k \) times (so \( n = k2^N + N - 1, m = k2^N + 2N - 1 \)). Then the constraint at \( \tau = 2^N \) becomes \( |r_F(2^N)| \leq 2m(\gamma + 1)\delta^2 \) for \( k \) large.

**Proof** Knowing that \( G \) is a F.I.R. of length \( N \), we first obtain an expression for \( |r_F(2^N)| \). We have

\[ |r_F(2^N)| = \left| \sum_{t=1}^{k2^N+2N-1} F(t)F(t + 2^N) \right| \\
= \left| \sum_{t=1}^{N-1} F(t)F(t + 2^N) + \sum_{t=N}^{(k-1)2^N+N-1} F(t)F(t + 2^N) \right| \\
+ \sum_{t=(k-1)2^N+N}^{k2^N+2N-1} F(t)F(t + 2^N) \] (4.12)

In last Eq. (4.12), by convolution since \( u_i = 0 \) for \( i < 0 \) and \( i > n = k2^N + N - 1 \), the first \( N - 1 \) terms of \( F(t) \) and the last \( 2^N + N \) terms of \( F(t - 2^N) \) do not contain all elements of \( G_i's \). Furthermore, the second summation \( F(t) \) is equal to \( F(t + 2^N) \) because of the repeated input. Since each \( F(t) \) is the convolution of \( u \ast G \), from Lemma 13 it follows that \( u \) can be thought as the vector \( \overrightarrow{a_N}(i) \) and \( G \) as \( \overrightarrow{x_N} \). Now using the triangle inequality in Eq. (4.12) (i.e. \(|a + b + c| \geq |c| - |a + b|\),

37
\[
|r_F(2^N)| \geq \left| \sum_{t=N}^{(k-1)2^N+N-1} F(t)F(t+2^N) \right| - \left| \sum_{t=1}^{N-1} F(t)F(t+2^N) + \sum_{t=(k-1)2^N+N}^{2^N+2N-1} F(t)F(t+2^N) \right| \tag{4.13}
\]

Since each \( |F(t)| \) is bounded by \( 2\delta \) (this is due to \( F + d_3 = 0 \), and \( \|d_3\|_\infty \leq 2\delta \)), the second term in Eq. (4.13) is bounded by a function of \( \delta \). And Lemma 3 gives us the first term of Eq. (4.13). Thus,

\[
|r_F(2^N)| \geq (k-1)2^N \|G\|_2^2 - (2^N + 2N - 2)(2\delta)^2. \tag{4.14}
\]

Since \( |r_F(2^N)| = |r_{d_3}(2^N)| \leq 2m(\gamma + 1)\delta^2 = 2(k2^N + 2N - 1)(\gamma + 1)\delta \), it follows that for large \( k \) and fixed \( \gamma \)

\[
\|G\|_2^2 \leq 2(\gamma + 1)\delta^2, \quad \text{(Note that } \frac{2^N}{k2^N} \rightarrow 0, \frac{(k-1)2^N}{k2^N} \rightarrow 1) \tag{4.15}\]

Thus we have the maximization problem of the \( \ell_1 \)-norm of \( G \) with an extra \( 2-norm \) constraint obtained from the autocorrelation condition (neglecting the constraint of the other values of \( \tau \)). Let us define this simpler problem to be problem \( A \):

\[
\max \quad \|G\|_1 \tag{4.16}
\]

\[
\text{s.t.} \quad \|G\|_1 \leq 2\delta, \text{ and } \|G\|_2 \leq 2(\gamma + 1)\delta^2
\]

without loss of generality each \( G_i \) can be assumed to be positive. Then we can define problem \( B \) to be the following

\[
\max \quad \|G\|_1 \tag{4.17}
\]

\[
\text{s.t.} \quad \|G\|_2 \leq 2(\gamma + 1)\delta^2
\]
The constraint in $\|G\|_1 \leq 2\delta$ is dropped in problem B since it only imposes an upper bound for $\|G\|_1$. This problem can be solved using Lagrange multiplier, thus the solution to the problem A is the minimum between the solution of problem B and $2\delta$.

Since the original problem (4.7) contains many more constrains (particularly, more autocorrelation constraints), the max $\|G\|_1$ is upper bounded by the answer given in problem A (i.e. $D_{N,n}(u) \leq$ solution of problem A). Solving the problem given by Eq. (4.17) we obtain the following

$$D_{N,n}(u) \leq \min\{\sqrt{2N(\gamma + 1)} \delta, \ 2\delta\}$$ (4.18)

From this last Eq., for $N > 2$, even if $\gamma$ goes to zero, $\|G\|_1$ is still $2\delta$. However, if we had assumed that $|r_{d_1d_2}(\tau)| \leq \gamma m\delta^2$, then

$$D_{N,n}(u) = \min\{2\sqrt{N\gamma} \delta, \ 2\delta\}$$

For fix $N$, if $\gamma$ is less than $\frac{1}{N}$, then the size of the diameter can be further reduced to $2\sqrt{N\gamma} \delta$. This assumption can be interpreted as two random variables being independent in the stochastic setup. From the work in [4], if we let each $d(i)$ be Independent Identically Distributed random variable, then $\gamma$ must decay in certain way so that the autocorrelation constraints are satisfied. Now if $\gamma$ decays in such way then one can show that $|r_{d_1d_2}(\tau)| \leq \gamma m\delta^2$ given that $d_1 = (d_1(1), \ldots, d_1(m))$ and $d_2 = (d_2(1), \ldots, d_2(m))$ are independent. However, this assumption might be too much to ask.

### 4.4 Some Properties on $D_{\delta\tau}$

In this section we provide some properties on the disturbance set $D_{\delta\tau}$, and from these properties one might be able to design an instrument to uncorrelate the noise from the output.
Lemma 17 Let \( d \in D_{s_t} \), then

\[
\frac{1}{m} \left| \sum_{i=1}^{m} d(i) \right| \leq \delta \sqrt{\frac{1 + (m - 1)\gamma}{m}} \tag{4.19}
\]

Proof The trick of the proof is taking the square of the sum. Thus

\[
\frac{1}{m^2} \left| \sum_{i=1}^{m} d(i) \right|^2 = \frac{1}{m^2} \left[ d(1) + \ldots + d(m) \right] \left[ d(1) + \ldots + d(m) \right]
\]

\[
= \frac{1}{m^2} \left[ (\sum_{i=1}^{m} d(i)^2) + 2 \left( d(1)d(2) + \ldots + d(m)d(1) \right) + \ldots \right.
\]

\[
+ 2 \left( d(1)d(\frac{m-1}{2}) + \ldots + d(m)d(\frac{m-1}{2}) \right) \left( \sum_{i=1}^{m} d(i) \right)
\]

\[
\leq \frac{1}{m^2} \left( m\delta^2 + (m - 1)\gamma m\delta^2 \right) \leq \frac{\delta^2}{m} (1 + (m - 1)\gamma)
\]

then taking square roots of the result we have the lemma.

\[
\text{Corollary 18} \text{ If } d \in D_{s_t} \text{ with length of } 2m \text{ (i.e. } d = (d(1), \ldots, d(2m))) \text{, then}
\]

\[
\frac{1}{2m} \left[ \left| \sum_{i=1}^{m} d(2i - 1) \right| + \left| \sum_{i=1}^{m} d(2i) \right| \right] \leq \delta \sqrt{\frac{2(1 + (2m - 1)\gamma)}{2m}} \tag{4.21}
\]

Proof The proof is also done by taking the square, just like the proof of the lemma. However, the corollary is a stronger result than the lemma since by partitioning \( d \) into \( d_{\text{odd}} \) and \( d_{\text{even}} \), \( |\sum_{i=1}^{m} d(2i - 1)| + |\sum_{i=1}^{m} d(2i)| \) can only be as big as \( |\sum_{i=1}^{2m} d(i)| \). The first property (Lemma 17) shows that if the plant is F.I.R. of length 1, then one can use

\[
u(t) = \begin{cases} 
1 & \text{if } 0 \leq t \leq m \\
0 & \text{otherwise}
\end{cases}
\]

as input. And the average of the observed outputs gives an estimate of the plant (in this case \( h_0 \)). As the length \( m \) of the input increases, the error goes to zero. And the second property suggests that if the plant is F.I.R. of length 2, then one can use the following sequence as the input

\[
u(t) = \begin{cases} 
1 & \text{if } t \text{ is even and } t \leq 2m \\
0 & \text{otherwise}
\end{cases}
\]

40
as input (note that the length of the input is $2m$). The average of the even observation gives an estimate in $h_0$, and the average of the odd observation estimates $h_1$. Again, as $m$ increases, the error goes to zero.

4.5 Sample Complexity

In this section, we will study the minimum length of an experiment that leads to optimal worst-case identification in $\ell_1$, and this problem is referred to as the Sample Complexity problem. To study the sample complexity, one need to find the required length of the input so that the diameter of the uncertainty set is guaranteed to be bounded by a function of $\delta$. A result from previous work [2] shows that for accurate identification in the $\ell_1$ sense, inputs tend to be quite long, and it is shown that the experiment length grows exponentially in the number of parameters to be identified. In this section we study the sample complexity for the noise model described previously, and we compare the results with those for bounded but unknown noise.

Let us define

$$D_{\delta,n}^* = \inf_{u \in U_n} D_{\delta,n}(u)$$

From the worst-case diameter analysis of the previous section and the results of [1], we see that

$$\lim_{n \to \infty} D_{\delta,n}^* \leq 2\delta \quad (4.22)$$

since the disturbance set is smaller than the bounded disturbance set. Then we have the following results

**Theorem 19** Fix $\gamma$ and let

$$n^*(N, \gamma) = \min\{n \mid D_{\delta,n}(u) = 2\delta\}. \quad (4.23)$$
Then for $N$ large, we have

$$n^*(N, \gamma) \geq \sqrt{2^N f(\sqrt{\gamma})^{-2}}. \quad (4.24)$$

Here, $f : (0, 1) \mapsto \mathbb{R}$ is the function defined by

$$f(\alpha) = 1 + \left(\frac{1 - \alpha}{2}\right) \log\left(\frac{1 - \alpha}{2}\right) + \left(\frac{1 + \alpha}{2}\right) \log\left(\frac{1 + \alpha}{2}\right). \quad (4.25)$$

Note that all logarithms are taken with base 2.

**Proof** To simplify the notation, let $g = u \ast h$ and $y = g + d$. Let us define $A$ to be the condition that $\|g\|_{\infty} < \delta$ and $B$ to be the condition that $|r_g(\tau)|_{\infty} < \gamma m \delta^2$ (for $1 \leq \tau \leq m - 1$). If $u$ is an input of length smaller than $n^*(N, \gamma)$, we will show that there exists some $h \in \{-\frac{\delta}{N}, \frac{\delta}{N}\}^N$ such that $\|g\|_{\infty} < \delta$ and $|r_g(\tau)|_{\infty} < \gamma m \delta^2$. This is to say that we can find a F.I.R. $h'$ whose distance (measured in $\ell_1$ norm) from $h$ is greater than $2\delta$ and a disturbance $d' \in D_\delta$ such that the output given by $u \ast h' + d'$ is identical to $y$ (Note that both boundedness and autocorrelation conditions on $g$ need to be satisfied).

In order to show $\Pr(A \text{ and } B) > 0$, it is sufficient to show that $\Pr(A^c)$ and $\Pr(B^c)$ are both less than $\frac{1}{2}$. This follows from the fact

$$\Pr(A \text{ and } B) = 1 - \Pr(A^c \text{ or } B^c) \geq 1 - \Pr(A^c) - \Pr(B^c)$$

Thus if $\Pr(A^c)$ and $\Pr(B^c)$ are both less than $\frac{1}{2}$, then $\Pr(A \text{ and } B) > 0$.

We break the problem into two parts. First, we will find some $n_1(N, \gamma)$ such that $\Pr(A^c) < \frac{1}{2}$ whenever $n < n_1(N, \gamma)$. Second, we will find some $n_2(N, \gamma)$ such that $\Pr(B^c) < \frac{1}{2}$ whenever $n < n_2(N, \gamma)$. After finding $n_1(N, \gamma)$ and $n_2(N, \gamma)$, we note that $\Pr(A^c)$ and $\Pr(B^c)$ are both less than $\frac{1}{2}$ if $n < \min\{n_1(N, \gamma), n_2(N, \gamma)\}$.

To find $n_1(N, \gamma)$, we bound $\Pr(A^c)$

$$\Pr(\|g\|_{\infty} \geq \delta) \leq \sum_{i=1}^{N+n} \Pr(|g(i)| \geq \delta) = \sum_{i=N+1}^{n+1} \Pr(|g(i)| \geq \delta)$$

42
\[
\leq (n - N + 1) \max_{1 \leq i \leq N+n} \Pr(|g(i)| \geq \delta)
\]  
(4.26)

We can bound \(\Pr(|g(i)| \geq \delta)\), then

\[
\Pr(|g(i)| \geq \delta) = \Pr\left(\left| \sum_{j=1}^{N} h(j)u(i - j) \right| \geq \delta \right)
\]
\[
= \Pr\left(\frac{1}{N} \left| \sum_{j=1}^{N} \left( \frac{Nh(i)}{\delta} \right)u(i - j) \right| \geq 1 \right) = \Pr\left(\frac{1}{N} \left| \sum_{j=1}^{N} \left( \frac{Nh(i)}{\delta} \right) \right| \geq 1 \right)
\leq 2 \cdot 2^{-N}
\]  
(4.27)

where the last inequality is a consequence of the Chernoff bound. Combining Eq. (4.26) and (4.27), we get

\[
\Pr(||g||_{\infty} \geq \delta) \leq (n - N + 1)2 \cdot 2^{-N}
\]  
(4.28)

Thus \(\Pr(||g||_{\infty} \geq \delta) < \frac{1}{2}\) if \(n < n_1(N, \gamma) = 2^{N-2} + N - 1\).

We will now find some \(n_2(N, \gamma)\) such that \(\Pr(|r_f(\tau)|_{\infty} > \gamma m \delta^2, \tau = 1, \cdots, m-1) < \frac{1}{2}\) whenever \(n < n_2(N, \gamma)\). Again we try to bound \(\Pr(B^c)\)

\[
\Pr(|r_g(\tau)|_{\infty} > \gamma m \delta^2, \forall \tau) \leq \sum_{\tau=1}^{m-1} \Pr(|r_g(\tau)| > \gamma m \delta^2)
\]
\[
\leq (m - 1) \max_{1 \leq \tau \leq m-1} \Pr(|r_g(\tau)| > \gamma m \delta^2)
\]  
(4.29)

For the event \(|r_g(\tau)|_{\infty} = |\sum_{i=1}^{m} g(i)g(i + \tau)| > \gamma m \delta^2\) to be true, there must be some \(i\) such that \(|g(i)| > \sqrt{\gamma} \delta\) (i.e. if \(|\sum_{i=1}^{m} g(i)g(i + \tau)| > \gamma m \delta^2\), then \(||g||_{\infty} > \sqrt{\gamma} \delta\)). Thus

\[
\Pr\left(\left| \sum_{i=1}^{m} g(i)g(i + \tau) \right| > \gamma m \delta^2 \right) < \frac{1}{2(m - 1)} \implies \Pr(||g(i)||_{\infty} > \sqrt{\gamma} \delta) < \frac{1}{2(m - 1)}
\]  
(4.30)

Now,
\[
\Pr(\|g(i)\|_\infty > \sqrt{\gamma} \delta) \leq \sum_{i=1}^m \Pr(|g(i)| > \sqrt{\gamma} \delta) = \sum_{i=\lceil N\sqrt{\gamma} \rceil + 1}^{m} \Pr(|g(i)| > \sqrt{\gamma} \delta) = (m - 2 \lceil N\sqrt{\gamma} \rceil + 1) \max_{1 \leq i \leq m} \Pr(|g(i)| > \sqrt{\gamma} \delta)
\] (4.31)

From [2], using the lemma \(\Pr(\frac{1}{N} \sum_{i=1}^N u_i X_i \geq \alpha) \leq 2^{-N \int f(\alpha)}\) for every \(\alpha \in (0, 1)\) and \(X_1, X_2, \cdots, X_N\) are independent random variables, we have the following

\[
\Pr\left(\sum_{j=1}^N u_j h_{i-j} > \sqrt{\gamma} \delta\right) = \Pr\left(\frac{1}{N} \sum_{j=1}^N u_{i-j} \frac{h_j N}{\delta} > \sqrt{\gamma}\right) \leq 2 \cdot 2^{-N \int f(\sqrt{\gamma})} \tag{4.32}
\]

Thus, \(\max_{1 \leq i \leq m} \Pr(|g(i)| > \sqrt{\gamma} \delta) \leq 2^2 \cdot 2^{-N \int f(\sqrt{\gamma})}\) and

\[
\Pr(|g|_\infty > \sqrt{\gamma} \delta) \leq 2 \cdot 2^{-N \int f(\sqrt{\gamma})}(m - 2 \lceil N\sqrt{\gamma} \rceil + 1) \tag{4.33}
\]

Using Eq. (4.33) and Eq. (4.30) we get,

\[
2 \cdot 2^{-N \int f(\sqrt{\gamma})}(m - 2 \lceil N\sqrt{\gamma} \rceil + 1) < \frac{1}{2(m - 1)} \tag{4.34}
\]

The above equation is quadratic in \(m\), so one can solve for \(m\) (Note \(m > 0\)). For \(N\) large, we have

\[
n_2(N, \gamma) \simeq \sqrt{2^N \int f(\sqrt{\gamma})^{-2}} \tag{4.35}
\]

Comparing \(n_1(N, \gamma)\) and \(n_2(N, \gamma)\), we can see that the dominant term will be \(2^N\). Therefore, \(n_2(N, \gamma)\) is smaller than \(n_1(N, \gamma)\) because of the square root. As a result if \(n^* < n_2(N, \gamma)\), then the probability of both condition \(A\) and condition \(B\) satisfied is larger than zero. This implies that there exists an element \(h \in \{-\frac{\delta}{N}, \frac{\delta}{N}\}^N\) for which \(\|g\|_\infty \leq \delta\) and \(|r_g(\tau)| \leq \gamma m \delta^2\) \((g = u * h, \text{ and } 1 \leq \tau \leq m - 1)\). Comparing \(n_1(N, \gamma)\) and \(n_2(N, \gamma)\), we see that \(n_2(N, \gamma) < n_1(N, \gamma)\). Therefore, it is necessary that \(n^* \geq n_2(N, \gamma)\) so that \(\Pr(A \text{ and } B)\) can be zero.
Chapter 5

Conclusions

5.1 Summary of Results

This thesis shows that the sample complexity for a certain subset of the model set $M_N$, namely monotonic impulse responses, still grows exponentially. Then we show that for multivariable systems the worst-case diameter can be reduced collectively to $2\delta$. This is possible because there is just one disturbance for the whole system. We also provide a lower bound for the sample complexity such that the optimal worst-case diameter is achieved.

This thesis also analyzes the problem of worst-case identification in the presence of disturbances with autocorrelation constraints. Two main results are shown. The first result provides a lower bound on the worst-case diameter of uncertainty which decreases to zero as the correlation bound approaches zero. This is similar to identification in the presence of white noise (unfortunately, no upper bound better than $2\delta$ has been found.) The second result analyzes the sample complexity of identification of FIR systems. In that regard, it is shown that for every finite non-zero correlation bound, the length of the input required to get within $2\delta$ error is exponential in the number of parameters to be identified. A lower bound on the sample complexity is derived that depends on both the allowable error and the correlation bound. However, as $\gamma$ goes to zero this lower bound degenerates (i.e. the lower bound also goes to zero). Thus, the behavior of the lower bound as the correlation bound approaches
zero remains an open question.

5.2 Future Research

In the robust control setting, mathematical convenience calls for deterministic disturbance rejection problems (e.g. $\mathcal{L}_1, \mathcal{H}_\infty$) where the disturbance is allowed to vary in a given set. Even though this leads to a successful robustness analysis methodology, some conservatism is involved in these deterministic classes, since the signal which give the worst-case performance are often very unlikely to be disturbances in practical situations. Some work must be done to bridge the gap between classical system identification and robust control, since the former relies entirely in the stochastic paradigm for disturbances. One might also investigate how the sample complexity lower bound degenerates as $\gamma$ goes to zero since both $\gamma$ and the lower bound are functions of $N$. That is, for $\gamma$ goes to zero $N$ must be large, and if $N$ is large then the lower bound grows exponentially. Also from the properties derived on the disturbance set $D_{\delta_T}$, one should look if there exists an input of polynomial length so that the optimal error goes to zero. If the plant is F.I.R. of length 1 or 2 then one can use impulse as input to identify the plant accurately. However, if $N > 3$, this remains an open question.
Bibliography


