Worst-Case Identification in $\ell_1$: Algorithms and Complexity

by

Theodore V. Theodosopoulos

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Master of Science in Operations Research
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 1993

© Massachusetts Institute of Technology 1993. All rights reserved.

Author .................................................................
Department of Electrical Engineering and Computer Science
January 26, 1993

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

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

Accepted by .........................................................
Richard C. Larson
Co-Chairman, Operations Research Center
Worst-Case Identification in $\ell_1$: Algorithms and Complexity

by

Theodore V. Theodosopoulos

Submitted to the Department of Electrical Engineering and Computer Science on January 26, 1993, in partial fulfillment of the requirements for the degree of Master of Science in Operations Research

Abstract

In this thesis we study the worst-case system identification of linear, shift invariant systems in the presence of unknown but bounded noise. Although the emphasis is on the $\ell_1$ error metric, certain results are related to other metrics (e.g. $\mathcal{H}^\infty$). The thesis contains negative results regarding the sample complexity of worst-case identification. In particular, it is shown that even suboptimal identification by a constant multiplicative factor requires experiments whose length grows exponentially in the number of parameters to be identified. On the positive side, a number of algorithms for worst-case identification is presented, together with bounds on their performance and implementation requirements.

Thesis Supervisor: Munther A. Dahleh
Title: Associate Professor

Thesis Supervisor: John N. Tsitsiklis
Title: Associate Professor


Contents

1 Introduction ................................................................. 7
  1.1 Problem Statement .................................................. 7
  1.2 Overview of Previous Work ........................................ 10
  1.3 Structure of Thesis ................................................ 12

2 Complexity Analysis of Worst-Case Identification of Linear Systems 13
  2.1 Introduction ......................................................... 13
  2.2 Optimal Identification in $\ell_1$ ................................ 15
  2.3 Relaxation of Optimality in $\ell_1$ ............................... 17
  2.4 Worst-Case Identification in $\ell_1$ with the $M(M, \rho)$ Model Set 25
  2.5 Near-Optimal Identification in $\mathcal{H}_\infty$ ................ 28
  2.6 Summary of Sample Complexities for Worst-Case Identification 33

3 Uncertainty Set Geometry and Off-Line Identification Algorithms 34
  3.1 Introduction ......................................................... 34
  3.2 Interpolatory Linear Programming Algorithm .................... 35
  3.3 Balanced Model Sets and Optimal Linear Algorithms ............ 38
  3.4 $\ell_1$ Geometry and Interior Point Identification Algorithms 45

4 On-Line Algorithms and Implementation Issues .......................... 53
  4.1 Introduction .......................................................... 53
  4.2 On-Line Interpolatory Algorithm .................................. 54
  4.3 Balanced Set Algorithm ............................................. 54
  4.4 Two-Stage Hierarchical Algorithm ................................ 57
5 Conclusion

5.1 Summary of Results ................................................. 60
5.2 Future Research ..................................................... 61
List of Figures

3-1 Example of set in $\mathbb{R}^2$ with $\text{rad}(C,\ell_1) > \frac{\text{diam}(C,\ell_1)}{2}$ .......................... 46
List of Tables

2.1 Sample Complexity Summary .......................... 33
Chapter 1

Introduction

1.1 Problem Statement

System identification has been a field traditionally linked to that of control. Over the years, the requirements for better control of various dynamic systems necessitated the development of system models. It is in this manner that the impressive developments on robust control during the 1980s generated the field of worst-case system identification.

More specifically, as opposed to the previously studied stochastic identification, worst-case identification set 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 are then to be used in the context of the Small Gain Theorem for the design of an $\ell_1$ or $\mathcal{H}_\infty$ optimal controllers.

We notice immediately that there are two aspects to worst-case identification. First of all, worst-case refers to disturbances that are completely unknown except for a bound on their magnitude. In particular, they could be persistent as is the case in a number of applications, e.g. wind loads on wings of aircraft, vibrations on automobile shock absorbers etc. Besides, no statistics are assumed to exist for the disturbances. So, if we call the disturbance signal $d$, then $d \in D_\delta = \{d \in \mathbb{R}^w | \|d\|_\infty \leq \delta\}$. 

7
Secondly, worst-case refers to the fact we are interested in worst-case error of the plant estimate, as opposed to average-case error and probabilistic error. That means we want to know how different can the true plant be from our estimate, irrespective of how unlikely it is that it be that far.

So, we can view worst-case identification as a game played against an omnipotent opponent. Our opponent chooses the true plant and the disturbances in order to maximize our error, and we pick our estimate and the inputs to the plant so as to minimize the maximum error achievable by our opponent.

One central concept in worst-case identification is that of a model set \( \mathcal{M} \). This concept captures our a priori knowledge about the true plant. For example, this knowledge could be that the plant is linear and has a finite impulse response of order \( N \). In this thesis we deal with linear, shift-invariant, discrete-time systems, whose impulse response, viewed as an element of \( \mathbb{R}^\omega \) belongs to one of two model sets:

\[
\mathcal{M}_N = \{ f \in \mathbb{R}^\omega | f_i = 0, \forall i \not\in [1, N] \}
\]

\[
\mathcal{M}(M, \rho) = \left\{ f \in \mathbb{R}^\omega | \sum_{i=0}^{\infty} |f_i| \rho^i \leq M, \rho > 1 \text{ and } \forall i < 0, f_i = 0 \right\}
\]

Another central concept in worst-case identification is that of an uncertainty set. This set contains all the plants in the model set that are consistent with the observations of the plant this far. So, in the worst-case sense, any of them could be the true plant according to the information we have gathered. In particular, let

\[
U_n = \{ u \in \mathbb{R}^\omega | ||u||_\infty \leq 1 \text{ and } u_i = 0 \text{, } \forall i \not\in [1, n] \}
\]

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

where \( \ast \) signifies convolution, and

\[
\mathcal{P}_n : \mathbb{R}^\omega \to \mathbb{R}^n
\]
such that
\[ \mathcal{P}_n(a_1, a_2, \ldots, a_n, a_{n+1}, \ldots) = (a_1, a_2, \ldots, a_n) \]

In words, \( \mathcal{P}_n \) is an operator that projects any vector form \( \mathbb{R}^\omega \) onto its first \( n \) components. Then, if we let \( u \in U_n \) be the input applied to the plant,

**Definition 1** We define the uncertainty set after \( k \) measurements to be the set

\[ S_k(\mathcal{M}, u, y, \delta) = \{ f \in \mathcal{M} | \| \mathcal{P}_k(f * u - y) \|_\infty \leq \delta \} . \]

We can easily check that

\[ f \in S_k(\mathcal{M}, u, y, \delta) \iff \exists d \in D_\delta \text{ such that } \mathcal{P}_k(f * u + d) = \mathcal{P}_k(y) \]

That is, for any plant \( f \) in the uncertainty set, there exists a legal disturbance that could make \( f \)'s output equal to the observed one. Note that

\[ S_k(\mathcal{M}_N, u, y, \delta) = S_{N+n}(\mathcal{M}_N, u, y, \delta) \text{ for } u \in U_n \text{ with } n \leq k - N. \]

Now, we need to quantify the uncertainty incorporated in the uncertainty set. For that reason, we develop the notion of radius and diameter for a set equipped with a metric.

**Definition 2** Let \( C \) be a subset of a metric space \( \mathcal{X} \). Let \( \rho_X \) denote the metric on \( \mathcal{X} \). Then, we define the radius and the diameter of \( C \), \( \text{rad}(C) \) and \( \text{diam}(C) \) respectively as

\[ \text{rad}(C, \mathcal{X}) = \inf_{x \in C} \sup_{y \in C} \rho_X(x, y) \]
\[ \text{diam}(C, \mathcal{X}) = \sup_{x \in C} \sup_{y \in C} \rho_X(x, y) \]

We can apply these concepts to measure the size of the uncertainty set and use that measure to quantify the level of uncertainty. We have already seen that the worst-case setting involves worst-case guarantees. In other words, we are interested in the maximum possible error our estimate could be making. And, naturally, we
want to pick an estimate that minimizes its maximum possible error. Let

\[ \mathcal{M} \subseteq \mathcal{X} \]

\[ D_k(\mathcal{M}, \mathcal{X}, u) = \sup_{d \in D_d} \sup_{h \in \mathcal{M}} \text{diam}(S_k(\mathcal{M}, u, h \ast u + d, \delta), \mathcal{X}) \]

\[ D_k^*(\mathcal{M}, \mathcal{X}, n) = \inf_{u \in U_n} D_k(\mathcal{M}, \mathcal{X}, u) \]

\[ R_k(\mathcal{M}, \mathcal{X}, u) = \sup_{d \in D_d} \sup_{h \in \mathcal{M}} \text{rad}(S_k(\mathcal{M}, u, h \ast u + d, \delta), \mathcal{X}) \]

\[ R_k^*(\mathcal{M}, \mathcal{X}, n) = \inf_{u \in U_n} R_k(\mathcal{M}, \mathcal{X}, u) \]

Let us consider, for each \( k \), a function \( \phi_k : \mathbb{R}^\omega \times \mathbb{R}^\omega \rightarrow S_k(\mathcal{M}, u, y, \delta) \) such that

\[ \phi_k(y, u) = h \in S_k(\mathcal{M}, u, y, \delta) \]

and consider the sequence \( \phi = (\phi_1, \phi_2, \ldots) \). We call \( \phi_n \) a consistent identification mapping and \( \phi \) an identification algorithm. Using the notation we developed, one problem studied in worst-case identification is to come up with an algorithm \( \phi \) and an input \( u \in U_n \) for some \( n \) so that there is a \( k_0 \) such that:

\[ E_k(\mathcal{M}, \mathcal{X}, u, \phi) \triangleq \sup_{d \in D_d} \sup_{g \in S_k(\mathcal{M}, u, y, \delta)} \rho_\mathcal{X}(g, \phi_k(g \ast u + d, u)) \leq D_k^*(\mathcal{M}, \mathcal{X}, n), \forall k \geq k_0 \]

### 1.2 Overview 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 creation of two main tracks in worst-case identification, one in \( \ell_1 \) based on time domain measurements ([22], [5], [6], [4], [15], [9], [2]) and one in \( \mathcal{H}_\infty \) based on frequency domain experiments([1], [7]).

In particular, [22] used the concepts from Information-Based Complexity developed in [21], [20] and [8] to analyze the infinite horizon case of worst-case identification.
in a $\sigma$-compact metric model space. It established the existence of an input\(^1\) $u^*$ which guarantees that, for all $C \subseteq \ell_1$, we have

$$D_{\infty}(C, \ell_1, u^*) = 2\delta = \lim_{n \to \infty} D_{\infty}^*(C, \ell_1, n).$$

It also concluded that infinite horizon worst-case identification is equivalent to stability testing. Makila in [15] 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. Poola and Tikku in [9] developed upper and lower bounds for the length of the input necessary to guarantee a specified level of accuracy in either $\mathcal{M}_N$ or $\mathcal{M}(M, \rho)$. Also, [2] established that in order for an input $u$ to guarantee $D_{\infty}(\mathcal{M}_N, \ell_1, u) = 2\delta$, its length has to be exponential in $N$.

Finally, in the $\mathcal{H}_\infty$ track, [1] developed a two-stage, nonlinear algorithm for plants in the $\mathcal{M}(M, \rho)$ model set. Their setup consists of solely frequency domain experiments. Their algorithm was shown to have the robust convergence property. This property requires that the plant error tends to zero as the number of frequency response measurements increases to infinity and the corruption level of each measurement decreases to zero.

Gu and Khargonekar in [7] generalized this idea into a class of nonlinear, robustly convergent algorithms, parametrized by a window function used to interpolate the inverse DFT of the frequency response measurements. In particular, the algorithm in [1] belongs to this class and uses spline window functions. Various other window functions are examined in [7] and the corresponding sample complexities are compared. In section 2.5 we discuss the sample complexity of this type of algorithms in $\mathcal{H}_\infty$.

\(^1\)The Galois sequence
1.3 Structure of Thesis

This thesis examines worst-case identification from the point of view of algorithms and complexity. Chapter 2 studies the sample complexity of any worst-case identification algorithm. In particular, for the model set $\mathcal{M}_N$ and the $\ell_1$ error metric, an asymptotically tight lower bound is developed for the minimum length of an input which guarantees suboptimal worst-case identification by a constant multiplicative factor. This lower bound is exponential in $N$. Section 2.4 extends this result to the case of the $\mathcal{M}(M, \rho)$ models set. Finally, section 2.5 examines the sample complexity of the two-stage, nonlinear algorithm from [1].

The remaining chapters present various algorithms for worst-case identification. These algorithms have been motivated from work on convex analysis and functional approximation ([16], [19], [14]). More specifically, Chapter 3 is devoted to off-line algorithms. Section 3.2 presents an interpolatory algorithm which is based on solving a linear program to find a member of the uncertainty set. Section 3.3 is devoted to balanced model sets. An algorithm is developed which uses the properties of these sets and tries to approximate the center of symmetry of the uncertainty sets using a sequence of linear programs\(^2\). Finally, section 3.4 studies the general case of nonbalanced model sets. A two-stage hierarchical algorithm is presented which estimates the incenter of the uncertainty set\(^3\).

Chapter 4 discusses the implementation of these algorithms. In particular, all three algorithms are shown to have on-line variants. The computational complexity and memory space requirements for the various algorithms are presented and compared. Finally, Chapter 5 offers a summary of the results in the thesis and suggestions for future research. It is my hope that reading this thesis will be as rewarding for you as its writing was for me.

---

\(^2\)The uncertainty sets are in general not center-symmetric, but the algorithm approximates the point closest to the center of symmetry.

\(^3\)See Definition 6 in section 3.4.
Chapter 2

Complexity Analysis of Worst-Case Identification of Linear Systems

2.1 Introduction

The worst-case identification problem as presented in Chapter 1 has mostly 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.

Eventually, the relationship between the quantity of information and the quality of the corresponding optimal approximation became central. The worst-case setup was recognized to be very demanding on the optimal input design. An important result from the work in [5] states that for the model set of all stable plants, accurate identification in the $\ell_1$ sense is possible if and only if the input excites all possible frequencies on the unit disc. This is due to two reasons: the first is that bounded noise is quite rich and the second is the fact that minimizing an induced norm such as the $\ell_1$ norm implies that the estimate has very good predictive power. Inputs with such properties tend to be quite long, and this suggests that the length of any optimal input for this kind of identification tends to increase very fast, as a function of the
number of estimated parameters of the impulse response.

In [15], Galois sequences were used to satisfy the input requirements mentioned above. The $N$th Galois sequence$^1$ is a shortest sequence that contains, as subsequences, all members of $\{-1,1\}^N$. Thus, such a sequence is rich enough to accurately identify exactly $N$ parameters of the impulse response. However, the length of this input is $2^N + N - 1$, which is exponential in $N$.

The next logical question is whether or not an experiment of length polynomial in $N$ could satisfy the richness requirements of optimal worst-case identification. The answer to this question could potentially depend on the norm used to judge the quality of approximation and on the amount of a priori knowledge of the plant's structure. For the case of the $\ell_1$ norm and model set $\mathcal{M}_N^2$, this question is addressed in section 2.2. There, we will see that the minimum length of an experiment that leads to optimal worst-case identification in $\ell_1$ is lower-bounded by an exponential in the number of impulse response coefficients to be estimated.

Further, it is interesting to investigate how the relaxation of optimality by a constant multiplicative factor affects the required length of experimentation. Section 2.3 deals with this issue in detail, for the $\ell_1$ norm. In particular, we use probabilistic arguments, reminiscent of those made in information theory, to establish an asymptotically tight lower bound to the minimum experiment length for suboptimal identification in the worst case. This bound is also exponential in the number of impulse response coefficients to be estimated. It should be mentioned here that sections 2.2 and 2.3 are included in [10] and are joint work.

In section 2.4 we extend our results for worst-case identification in $\ell_1$ to the $\mathcal{M}(M, \rho)$ model set. The parallelism between the FIR plants and $\mathcal{M}(M, \rho)$ is based on the definition of the effective length of a plant. This definition is then used in section 2.5 to prove an upper bound to the sample complexity of worst-case identification in $\mathcal{H}_\infty$. The analysis is based on the performance of the two-stage, nonlinear algorithm proposed in [1]. Surprisingly, for any fixed value of $M$, the upper bound

$^1$see [18]

$^2$plants with F.I.R. of length $N$ as defined in Chapter 1
achieved is only a polynomial function of the effective length of the plant to be identified. This result suggests that worst-case identification may be inherently easier in $\mathcal{H}_\infty$ than in $\ell_1$.

In comparison to the few other complexity results in worst-case identification using the $\ell_1$ norm that have recently been announced ([9],[2]), the results in this chapter provide bounds that are asymptotically tight. Besides, to the best of our knowledge, there are no other complexity results for worst-case identification in $\mathcal{H}_\infty$. The use of information-theoretic arguments, somewhat foreign to mainstream system identification distinguishes this work from others. The setup presented in section 2.3 can readily be extended to the probabilistic, rather than worst-case, scenario. In other words, one can use, in a straightforward manner, the probabilistic arguments presented in this chapter to investigate the sample complexity of identification when we only require a certain probability of success.

### 2.2 Optimal Identification in $\ell_1$

Recalling the problem definition in section 1.1, let us define

$$n^*(\mathcal{M}, \mathcal{X}, K) = \min \left\{ n \mid D_\infty^*(\mathcal{M}, \mathcal{X}, n) \leq K \inf_n D_\infty^*(\mathcal{M}, \mathcal{X}, n) \right\}.$$ 

That is the minimum length of an input that can guarantee an upper bound to the worst-case error that is a constant factor $K$ of the optimal infinite horizon uncertainty set diameter.

In this and the next section we will concentrate on $\mathcal{M}_N$. As we mentioned already in the Chapter 1, for this model set, the optimal worst-case diameter of the infinite horizon uncertainty set is equal to $2\delta$. For simplicity, let us denote $n^*(\mathcal{M}_N, \ell_1, 1)$ by $n^*(N)$. It is far from a priori clear whether $n^*(N)$ is finite. This is answered by the following theorem which also serves as motivation for the theorem (Theorem 2) of the next section.

**Theorem 1** For any $\delta > 0$ and $N$, we have $2^{N-1} + N - 1 \leq n^*(N) \leq 2^N + N - 1$. 

15
**Proof.** We start by proving the lower bound on $n^*(N)$. Fix $N$ and let us denote $n^*(N)$ by $m$. Suppose that $m < \infty$, and let $\phi, u \in U_m$, be such that $D_m(\mathcal{M}_N, \ell_1, u) = 2\delta$ and so $E_m(u, \phi, \mathcal{M}_N) \leq 2\delta$. Let $v \in \{-1, 1\}^m$ be defined by $v_i = 1$ if $u_i \geq 0$, and $v_i = -1$ if $u_i < 0$. For notational convenience, we define $u_i = 0$ for $i \leq 0$. We distinguish two cases:

(a) Suppose that for every $w \in \{-1, 1\}^N$, there exists some $i(w) \in \{1, \ldots, m - N + 1\}$ such that either $w$ or $-w$ is equal to $(v_{i(w)+N-1}, v_{i(w)+N-2}, \ldots, v_{i(w)})$. It is clear that $i(w)$ can be the same for at most two different values of $w$. Since the number of different choices for $w$ is $2^N$, it follows that $m - N + 1 \geq 2^{N-1}$, which proves that $m \geq 2^{N-1} + N - 1$.

(b) Suppose now that the assumption of case (a) fails to hold. Let $w \in \{-1, 1\}^N$ be such that both $w$ and $-w$ are different from $(v_{i+N-1}, v_{i+N-2}, \ldots, v_i)$, for all $i \in \{1, \ldots, m - N + 1\}$. Suppose that $h = \frac{\delta w}{N-1}$. Then,

$$|(h * u)_i| = \left| \sum_{k=1}^{N} h_kw_{i-k} \right| = \frac{\delta}{N-1} \left| \sum_{k=1}^{N} w_ku_{i-k} \right|$$

Since $|w_k| = 1$ and $|u_{i-k}| \leq 1$, we see that $|\sum_{k=1}^{N} w_ku_{i-k}| \leq N$. Let $i$ be such that $N < i \leq m$. By our assumption on $w$, the signs of $u_{i-k}$ cannot be the same as the signs of $w_k$ for all $k$, neither the same as the signs of $-w_k$ for all $k$, and this leads to the stronger inequality

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

We finally note that for $i \notin (N, m]$, at least one of the summands $w_ku_{i-k}$ is equal to zero, which implies that the above equation is valid for all $i$. Thus, we conclude that $|(h * u)_i| \leq \delta$ for all $i$. Therefore, there exists a choice for the disturbance sequence $d$ under which the observed output $h * u + d$ is equal to zero at all times. Using the same argument, we see that if $h = -\frac{\delta w}{N-1}$, there also exists another choice of the disturbance sequence for which the observed output is zero at all times.

We have thus shown that it is possible to observe an output sequence which is identically equal to zero while the true system can be either $\frac{\delta w}{N-1}$ or $-\frac{\delta w}{N-1}$. This
implies that the worst case diameter satisfies

\[ D_m(\mathcal{M}_N, \ell_1, u) \geq 2 \left\| \frac{\delta w}{N - 1} \right\|_1 > 2\delta \]

But this contradicts the definition of \( m = n^*(N) \) and shows that case (b) is not possible. Thus, case (a) is the only possible one, and the lower bound has already been established for that case. The upper bound follows easily by using the input sequence proposed in [5], [15]. Let \( u \) be a finite sequence whose entries belong to \( \{-1, 1\} \) and such that for every \( w \in \{-1, 1\}^N \) there exists some \( i(w) \) such that \( w = (u_{i(w)}, u_{i(w)+1}, \ldots, u_{i(w)+N-1}) \). Such a sequence, called a Galois sequence, can be chosen so that its length is equal to \( 2^N + N - 1 \) ([18]). With this input, the worst case diameter is equal to \( 2\delta \) \( \square \)

Theorem 1 has the disappointing conclusion that the worst-case error is guaranteed to become at most \( 2\delta \) only if a very long experiment is performed. In practice, values of \( N \) of the order of 20 or 30 often arise. For such cases, the required length of an experiment for optimal identification in \( \ell_1 \) is prohibitively long.

### 2.3 Relaxation of Optimality in \( \ell_1 \)

The disappointing result of section 2.2 motivates the problem studied in this section: if the objective is to obtain an identification error within a factor \( K \) of the optimal value, can this be accomplished with substantially smaller experiments?

Let us denote \( n^*(\mathcal{M}_N, \ell_1, K) \) by \( n^*(N, K) \). Theorem 2 below is equally disappointing with Theorem 1: it shows that experiments of length exponential in \( N \) are required to obtain such an error guarantee. The exponent depends of course on \( K \) and we are able to compute its asymptotic value (as \( N \) increases) exactly.

**Theorem 2** Fix some \( K > 1 \). Then, (a) \( n^*(N, K) \geq 2^{Nf(\frac{1}{K})-1} - N + 2 \left[ \frac{N}{K} \right] - 1 \).

(b) \( \lim_{N \to \infty} \frac{1}{N} \log n^*(N, K) = f \left( \frac{1}{K} \right) \).

Here, \( f : (0, 1) \to \mathbb{R} \) is the function defined by\(^3\)

\(^3\)In the definition of \( f \), and throughout the rest of the paper, all logarithms are taken with base
\[ f(\alpha) = 1 + \left( \frac{1 - \alpha}{2} \right) \log \left( \frac{1 - \alpha}{2} \right) + \left( \frac{1 + \alpha}{2} \right) \log \left( \frac{1 + \alpha}{2} \right). \]

Notice that the function \( f \) satisfies \( f(\alpha) = 1 - H(\frac{1-\alpha}{2}) \), where \( H \) is the binary entropy function. In particular, \( f \) is positive and continuous for \( \alpha \in (0,1) \). Before going ahead with the main part of the proof, we need to develop some lemmas that will be our main tools.

**Lemma 1** Let \( X_1, X_2, \ldots, X_N \) be independent binomial random variables with \( \Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2} \) for every \( i \).

(a) Let \( u_i \in [-1, 1], i = 1, \ldots, N \). Then, for every \( \alpha \in (0,1) \), we have

\[ \Pr \left( \frac{1}{N} \sum_{i=1}^{N} u_i X_i \geq \alpha \right) \leq 2^{-N f(\alpha)}. \]

(b)

\[ \lim_{N \to \infty} \frac{1}{N} \log \Pr \left( \frac{1}{N} \sum_{i=1}^{N} X_i \geq \alpha \right) = -f(\alpha). \]

**Proof.** Part (b) is obtained from the classical Chernoff bound ([17]). Part (a) also follows from the Chernoff bound, if \( u_i = 1 \) for all \( i \). It remains to prove part (a) for the general case of \( u_i \in [-1, 1] \).

We first note that because of the symmetry in the distribution of \( X_i \), we can assume, without any loss of generality that \( u_i \in [0, 1] \) for all \( i \). We then have

\[ \Pr \left( \frac{1}{N} \sum_{i=1}^{N} u_i X_i \geq \alpha \right) \leq \inf_{s > 0} \prod_{i=1}^{N} E \left[ e^{s(u_i X_i - \alpha)} \right] \leq \inf_{s > 0} \prod_{i=1}^{N} E \left[ e^{s(X_i - \alpha)} \right] = 2^{-N f(\alpha)}. \]

The first inequality is obtained by following the steps in the standard proof of the Chernoff bound; the second inequality is obtained by verifying that \( e^{su} + e^{-su} \leq e^s + e^{-s} \) for all \( u \in [0,1] \); finally, the final equality is a simple calculation which is also part of the classical proof of the Chernoff bound. \( \square \)

One consequence of Lemma 1 is that for any \( \epsilon > 0 \) and for any \( \alpha \in (0,1) \), there

---

2.
exists some \( N_0(\alpha, \epsilon) \) such that

\[
\Pr \left( \frac{1}{N} \sum_{i=1}^{N} X_i \geq \alpha \right) \geq 2^{-N(f(\alpha)+\epsilon)}, \quad \forall N \geq N_0(\alpha, \epsilon).
\]

The following lemma strengthens the previous equation and will be needed later in the proof.

**Lemma 2** Let \( X_1, \ldots, X_N \) be as in Lemma 1. Let \( \Theta_N = \left\{ (\theta_1, \ldots, \theta_N) \in \mathbb{R}^N \mid \sum_{i=1}^{N} |\theta_i| = N \right\} \).

Then, for any \( \epsilon_1 > 0 \) and any \( \alpha \in (0, 1) \), there exists some \( N_1(\alpha, \epsilon_1) \) such that

\[
\Pr \left( \frac{1}{N} \sum_{i=1}^{N} \theta_i X_i \geq \alpha \right) \geq 2^{-N(f(\alpha)+\epsilon_1)}, \quad \forall N \geq N_1(\alpha, \epsilon_1), \forall \theta \in \Theta_N.
\]

**Proof.** Note that the random variables \( \sum_{i=1}^{N} \theta_i X_i \) and \( \sum_{i=1}^{N} |\theta_i| X_i \) have the same probability distribution. Therefore, without loss of generality, we can and will assume that \( \theta_i \geq 0 \) for all \( i \). We have

\[
\Pr \left( \sum_{i=1}^{N} \theta_i X_i \geq \alpha N \right) = \Pr \left( \sum_{i=1}^{N} \theta_i X_i \geq \alpha N \mid \sum_{i=1}^{N} X_i \geq \alpha N \right) \cdot \Pr \left( \sum_{i=1}^{N} X_i \geq \alpha N \right) \\
\geq 2^{-N(f(\alpha)+\frac{\epsilon_1}{2})} \Pr \left( \sum_{i=1}^{N} \theta_i X_i \geq \alpha N \mid \sum_{i=1}^{N} X_i \geq \alpha N \right)
\]

where the last inequality holds for all \( N \) large enough.

Given any sequence \( X = (X_1, \ldots, X_N) \), let \( X^k \) be its cyclic shift by \( k \) positions; that is, \( X^k = (X_{k+1}, X_{k+2}, \ldots, X_N, X_1, \ldots, X_k) \). Let \( X_i^k \) be the \( i \)th component of \( X^k \). By symmetry, the conditional distribution of \( X \) and \( X^k \), conditioned on the event \( \sum_{i=1}^{N} X_i \geq \alpha N \), is the same. Therefore,

\[
\Pr \left( \sum_{i=1}^{N} \theta_i X_i \geq \alpha N \mid \sum_{i=1}^{N} X_i \geq \alpha N \right) = \frac{1}{N} \sum_{k=1}^{N} \Pr \left( \sum_{i=1}^{N} \theta_i X_i^k \geq \alpha N \mid \sum_{i=1}^{N} X_i \geq \alpha N \right) \\
\geq \frac{1}{N} \Pr \left( \exists k \text{ such that } \sum_{i=1}^{N} \theta_i X_i^k \geq \alpha N \mid \sum_{i=1}^{N} X_i \geq \alpha N \right) \\
= \frac{1}{N}.
\]
The last equality follows because if $\sum_{i=1}^{N} X_i \geq \alpha N$, then
\[
\sum_{k=1}^{N} \sum_{i=1}^{N} \theta_i X_i^k = \sum_{i=1}^{N} \theta_i \sum_{i=1}^{N} X_i \geq \alpha N^2,
\]
which immediately implies that there exists some $k$ for which $\sum_{i=1}^{N} \theta_i X_i^k \geq \alpha N$.

We conclude that
\[
Pr \left( \sum_{i=1}^{N} \theta_i X_i \geq \alpha N \right) \geq \frac{1}{N} 2^{-N(f(\alpha)+\frac{\delta}{N})} \geq 2^{-N(f(\alpha)+\epsilon_1)},
\]
where the last inequality follows if $N$ is large enough so that $\frac{1}{N} \geq 2^{-N \frac{\delta}{N}}$.

Having finished with the probabilistic preliminaries, we can now continue with the main part of the proof of Theorem 2. We will start with the proof of part (a).

**Lemma 3** Suppose that the length $n$ of an input sequence $u \in U_n$ is smaller than $2^{N f(\frac{k}{N})-1} - N + 2 \left\lfloor \frac{N}{K} \right\rfloor - 1$. Then, there exists some $h \in \left\{ \frac{K\delta}{N}, \frac{K\delta}{N} \right\}^N$ such that $\| u \ast h \|_\infty < \delta$.

**Proof.** Let $n$ be as in the statement of the lemma. We will show the existence of such an $h$ by showing that a random element of $\left\{ -\frac{K\delta}{N}, \frac{K\delta}{N} \right\}^N$ satisfies $\| u \ast h \|_\infty < \delta$ with positive probability. Indeed, let $h$ be such a random element, under the uniform distribution on $\left\{ -\frac{K\delta}{N}, \frac{K\delta}{N} \right\}^N$. Then,

\[
Pr(\| u \ast h \|_\infty \geq \delta) \leq \sum_{j=1}^{N+n} Pr(\| (u \ast h)_j \| \geq \delta) = \sum_{j=\left\lfloor \frac{N}{K} \right\rfloor + 1}^{N+n} Pr(\| (u \ast h)_j \| \geq \delta)
\]

\[
\leq (N+n-2 \left\lfloor \frac{N}{K} \right\rfloor + 1) \max_{1 \leq j \leq N+n} Pr(\| (u \ast h)_j \| \geq \delta).
\]

where the equality on the first line holds because for $j \leq \left\lfloor \frac{N}{K} \right\rfloor$, we have:

\[
\| (u \ast h)_j \| = \left| \sum_{i=1}^{N} h_i u_{j-i} \right| = \left| \sum_{i=1}^{j-1} h_i u_{j-i} \right| \leq (j-1) \frac{K\delta}{N} \leq \left( \frac{N}{K} - 1 \right) \frac{K\delta}{N} < \delta
\]
and for \( j \geq N + n - \left\lceil \frac{N}{K} \right\rceil + 2 \), we have:

\[
| (u * h)_j | = \left| \sum_{i=1}^{N} h_i u_{j-i} \right| = \left| \sum_{i=j-n}^{N} h_i u_{j-i} \right| \leq (N - j + n + 1) \frac{K \delta}{N} \leq \left( \left\lceil \frac{N}{K} \right\rceil - 1 \right) \frac{K \delta}{N} < \delta
\]

Furthermore,

\[
\Pr( | (u * h)_j | \geq \delta ) = \Pr \left( \left| \sum_{i=1}^{N} h_i u_{j-i} \right| \geq \delta \right) = \Pr \left( \frac{1}{N} \left| \sum_{i=1}^{N} \frac{N h_i u_{j-i}}{K \delta} \right| \geq \frac{1}{K} \right) \leq 2 \cdot 2^{-N f(\frac{1}{K})}.
\]

The last inequality follows from Lemma 1, because the random variables \( \frac{N h_i}{K \delta} \) are independent, take values in \( \{-1, 1\} \), and each value is equally likely. Therefore, we conclude that

\[
\Pr(||u * h||_{\infty} \geq \delta) \leq 2(N + n - 2 \left\lceil \frac{N}{K} \right\rceil + 1)2^{-N f(\frac{1}{K})}.
\]

If \( 2(N + n - 2 \left\lceil \frac{N}{K} \right\rceil + 1) < 2^{N f(\frac{1}{K})} \), then the right-hand side of the above inequality is smaller than 1. This implies that there exists some \( h \in \{-\frac{K \delta}{N}, \frac{K \delta}{N}\}^N \) for which \( ||h * u||_{\infty} < \delta \). \( \Box \)

Suppose now that the length \( n \) of the input sequence \( u \) is as in Lemma 3, and let the unknown system \( h \) have the properties described in that lemma. Since \( |(h * u)_i| < \delta \) for all \( i \), there is a choice of the disturbance sequence \( d \) that leads to zero output. Consider next the case where the unknown system is actually equal to \(-h\). We also have \(|(-h * u)_i| < \delta\), for all \( i \), and a zero output sequence is still possible. Thus, if the output sequence is equal to zero, both \( h \) and \(-h\) could be the true system. For any identification algorithm, the worst-case error will be at least equal to one half of the distance of these two systems, which is \( ||h||_1 = K \delta \). In fact, the same argument can be carried out if \( h \) is replaced by \((1 + \epsilon)h\), where \( \epsilon > 0 \) is small enough so that the property
\((1+\varepsilon)|(h*u)_i| < \delta\) holds. We can then conclude that the worst-case diameter will be at least \(2(1+\varepsilon)K\delta\). We have therefore shown that if \(n < 2^{Nf(1/K)^{-1}} - N + 2 \left\lceil \frac{N}{K} \right\rceil - 1\), then \(D_n(\mathcal{M}_N, \ell_1, u) > 2K\delta\). Equivalently, \(n^*(N, K) \geq 2^{Nf(1/K)^{-1}} - N + 2 \left\lceil \frac{N}{K} \right\rceil - 1\), which completes the proof of part (a).

We now turn to the proof of part (b) of the theorem. Part (a) implies that 
\[
\liminf_{N \to \infty} (1/N) \log n^*(N, K) \geq f(1/K).
\]
The proof will be completed by showing that
\[
\limsup_{N \to \infty} \frac{1}{N} \log n^*(N, K) \leq f \left( \frac{1}{K} \right).
\]
To show this, we have to show the existence of an input sequence \(u\) of length close to \(2^{Nf(1/K)}\) that results in an uncertainty set of diameter bounded by \(2K\delta\). Although we are not able to provide an explicit construction of such an input sequence, we will prove its existence using a probabilistic argument.

We now provide the details of the construction of the input sequence \(u\). Let us fix some \(\varepsilon > 0\). Let
\[
M(N) = \left\lceil 2^{Nf(\varepsilon + \frac{1}{K}) + 2\varepsilon} \right\rceil.
\]
For every \(k \in \{1, \ldots, M(N)\}\), we choose a vector \(u^k = (u^k_1, \ldots, u^k_N) \in \{-1, 1\}^N\).
The input \(u\) is then defined by
\[
u = (u^1, u^2, \ldots, u^{M(N)}),
\]
and has length \(NM(N)\).

**Lemma 4** Let the input \(u\) be constructed as in the preceding paragraph. Furthermore suppose that the entries of the vectors \(u^k\) are independent random variables, with each value in the set \(\{-1, 1\}\) being equally likely. Then, there exists some \(N_2(K, \varepsilon)\) such that
\[
Q_N \triangleq \Pr(\exists h \in \mathcal{M}_N \text{ such that } \|h\|_1 \geq K\delta, \|u * h\|_\infty \leq \delta) < 1, \forall N \geq N_2(K, \varepsilon).
\]
**Proof.** Notice that if \( i \) is an integer multiple of \( N \), say \( i = mN \), we have

\[
(u \ast h)_i = \sum_{j=1}^{N} u_j^m h_{N-j}, \quad i = mN.
\]

We then have

\[
Q_N = \Pr(\exists h \in \mathcal{M}_N \text{ such that } \|h\|_1 \geq K \delta, \|u \ast h\|_{\infty} \leq \delta) \\
= \Pr(\exists h \in \mathcal{M}_N \text{ such that } \|h\|_1 = K \delta, \|u \ast h\|_{\infty} \leq \delta) \\
= \Pr(\exists h \in \mathcal{M}_N \text{ such that } \|h\|_1 = N, \|u \ast h\|_{\infty} \leq \frac{N}{K}) \\
\leq \Pr(\exists h \in \mathcal{M}_N \text{ such that } \|h\|_1 = N, \left| \sum_{j=1}^{N} u_j^m h_{N-j} \right| \leq \frac{N}{K}, \ m = 1, \ldots, M(N)).
\]

Let us choose a finite subset \( \mathcal{M}_N^* \) of \( \mathcal{M}_N \) such that for every \( h \in \mathcal{M}_N \) with \( \|h\|_1 = N \), there exists some \( h' \in \mathcal{M}_N^* \) satisfying \( \|h'\|_1 = N \) and \( \|h - h'\|_{\infty} < \epsilon \). In particular, \( \mathcal{M}_N^* \) can be chosen as a subset of the set of all elements of \( \mathcal{M}_N \) for which each component is bounded by \( N \) and is an integer multiple of \( \frac{\epsilon}{N} \). It is then clear that \( \mathcal{M}_N^* \) can be assumed to have cardinality bounded by \( \left( \frac{2N+1}{\epsilon} \right)^N \). We then have

\[
\Pr(\exists h \in \mathcal{M}_N \text{ such that } \|h\|_1 = N, \left| \sum_{j=1}^{N} u_j^m h_{N-j} \right| \leq \frac{N}{K}, \ m = 1, \ldots, M(N)) \\
\leq \Pr(\exists h' \in \mathcal{M}_N^* \text{ such that } \left| \sum_{j=1}^{N} u_j^m h'_{N-j} \right| < N \left( \epsilon + \frac{1}{K} \right), \ m = 1, \ldots, M(N)) \\
\leq \left( \frac{2N+1}{\epsilon} \right)^N \max_{h' \in \mathcal{M}_N^*} \Pr(\left| \sum_{j=1}^{N} u_j^m h'_{N-j} \right| < N \left( \epsilon + \frac{1}{K} \right), \ m = 1, \ldots, M(N)).
\]

We provide an upper bound to the probability in the right-hand side of the above by applying Lemma 2. (Here, \( u_j^m \) and \( h'_{N-j} \) correspond to \( X_i \) and \( \theta_i \) in the notation of that lemma.) Indeed, Lemma 2 is applicable because \( \|h'\|_1 = N \) and the components of the input are i.i.d random variables, with the same distribution as the variables \( X_i \) of Lemma 1. We therefore conclude that there exists some \( N_2(K, \epsilon) \) such that

\[
\Pr(\left| \sum_{j=1}^{N} u_j^m h'_{N-j} \right| < N \left( \epsilon + \frac{1}{K} \right)) < 1 - 2^{-\Omega(f(\epsilon+\frac{1}{K}))+\epsilon}, \forall m, \forall N \geq N_2(K, \epsilon).
\]
Using the statistical independence of the vectors $u^m$, we obtain

$$Q_N \leq \left( \frac{2N + 1}{\epsilon} \right)^N \left( 1 - 2^{-N(\ell(\epsilon + \frac{1}{k}) + \epsilon)} \right)^{M(N)}$$

$$\leq \left( \frac{2N + 1}{\epsilon} \right)^N \exp \left\{ -M(N)2^{-N(\ell(\epsilon + \frac{1}{k}) + \epsilon)} \right\}$$

$$\leq \left( \frac{2N + 1}{\epsilon} \right)^N \exp \{-2^{\epsilon N}\},$$

where the second inequality follows from the fact $\left(1 - \frac{1}{x}\right)^x \leq \epsilon^{-1}$, for every $x > 0$, and the last inequality follows from the definition of $M(N)$. It is then easily seen that

$$\lim_{\epsilon \to 0} \lim_{N \to \infty} Q_N = 0$$

which establishes the desired result. □

Lemma 4 establishes that, if the input $u$ is constructed randomly as in the discussion preceding the lemma, then, with positive probability, $u$ will have the following property:

Property 1: if $h \in \mathcal{M}_N$ and $\|u * h\|_{\infty} \leq \delta$, then $\|h\|_1 \leq K\delta$.

In particular, there exists at least one\(^\dagger\) $u$ of length $NM(N)$ that satisfies Property 1.

**Lemma 5** If an input $u$ satisfies Property 1, then $D_n(\mathcal{M}_N, \ell_1, u) \leq 2K\delta$.

**Proof.** We apply the input $u$ and measure the output $y = h * u + d$, where $h$ is the unknown plant and $d$ is the disturbance sequence. Given the observed output $y$, we can infer that $h$ belongs to the uncertainty set

$$S_{\infty}(\mathcal{M}_N, u, y, \delta) = \{\phi \in \mathcal{M}_N \|y - \phi * u\|_{\infty} \leq \delta\}.$$

Let $\chi$ and $\psi$ be two elements of $S_{\infty}(\mathcal{M}_N, u, y, \delta)$. Then, $\|y - \chi * u\|_{\infty} \leq \delta$ and

\(^\dagger\) In fact, it is easily seen that $Q_N$ converges to zero very rapidly, which implies that most $u$'s will satisfy Property 1.
\| y - \psi * u \|_\infty \leq \delta. \text{ Using the triangle inequality, we obtain } \| u * (\chi - \psi) / 2 \|_\infty \leq \delta. \text{ Since } u \text{ satisfies Property 1, we conclude that } \| (\chi - \psi) / 2 \|_1 \leq K \delta \text{ or } \| \chi - \psi \|_1 \leq 2K \delta. \text{ Since this is true for any pair of elements of } S_\infty(\mathcal{M}_N, u, y, \delta), \text{ the diameter of } S_\infty(\mathcal{M}_N, u, y, \delta) \text{ is at most } 2K \delta. \quad \square

As discussed earlier, if \( N \) is large enough, there exists an input of length \( n = M(N)N \) that satisfies Property 1 and, by Lemma 5, leads to uncertainty sets whose diameter is bounded above by \( 2K \delta \). It follows that \( n^*(N, K) \leq M(N)N \). Using the definition of \( M(N) \), we see that

\[
\limsup_{N \to \infty} \frac{1}{N} \log n^*(N, K) \leq \limsup_{N \to \infty} \frac{1}{N} \log M(N)N \leq f \left( \epsilon + \frac{1}{K} \right) + 2\epsilon.
\]

Since the above equation is valid for all \( \epsilon > 0 \), and since \( f \) is continuous, we conclude that

\[
\limsup_{N \to \infty} \frac{1}{N} \log n^*(N, K) \leq f \left( \frac{1}{K} \right),
\]

which concludes the proof of Theorem 2.

2.4 Worst-Case Identification in \( \ell_1 \) with the \( \mathcal{M}(M, \rho) \) Model Set

In sections 2.2 and 2.3 we have seen that, for plants with FIR of length \( N \), the length of a time domain experiment necessary to guarantee a worst-case error less than a constant factor of the optimal is exponential in \( N \). In section 1.1 we defined another model set, namely \( \mathcal{M}(M, \rho) \), where \( \rho \) represents a bound to the decay rate and \( M \) a bound to the overshoot. One would like to extend the FIR results arrived at before to this more general model set.

In order to establish the parallelism between \( \mathcal{M}_N \) and \( \mathcal{M}(M, \rho) \) we need two definitions:

**Definition 3** \( \mathcal{M}_N' \) is the subset of \( \mathcal{M}_N \) which contains all the plants with FIR of
length $N$ and $\ell_\infty$ norm less than $1/N$. That is

$$\mathcal{M}'_N = \left\{ f \in \mathcal{M}_N \mid \|f\|_\infty \leq \frac{1}{N} \right\}$$

**Definition 4** Let

$$L(M, \rho) = \max \left\{ N \mid \mathcal{M}'_N \subseteq \mathcal{M}(M, \rho) \right\}.$$  

We call $L(M, \rho)$ the effective length of plants in $\mathcal{M}(M, \rho)$.

For the next two sections we will use the big O notation for the order of growth of various quantities. Definitions of this notation can be found in [13]. In particular, when we use this notation here we mean it as $M$ increases to $\infty$ and as $\rho$ decreases to $1^+$. We first obtain a limiting result about the effective length of the plants in $\mathcal{M}(M, \rho)$.

**Proposition 1**

$$\lim_{\rho \to 1^+} \frac{(\rho - 1)L(M, \rho)}{\log M} = 1.$$  

**Proof.** The definition of the effective length can be rewritten as:

$$L(M, \rho) = \max \left\{ N \mid \sum_{i=0}^{N} \frac{1}{N^i} \rho^i \leq M \right\}.$$  

We know that, for any finite $K$,

$$\lim_{\rho \to 1^+} \frac{\rho^{K+1} - 1}{\rho - 1} = \frac{\rho^K}{K} = 1.$$  

The above limit means that

$$\forall \epsilon > 0, \exists \rho_0 > 1 \text{ such that } \forall \rho \in (1, \rho_0], (1 - \epsilon)K\rho^K \leq \frac{\rho^{K+1} - 1}{\rho - 1} \leq (1 + \epsilon)K\rho^K.$$  

From the definition of $L(M, \rho)$ we have

$$\frac{\rho^{L(M, \rho) + 1} - 1}{L(M, \rho)(\rho - 1)} \leq M \text{ and } \frac{\rho^{L(M, \rho) + 2} - 1}{(L(M, \rho) + 1)(\rho - 1)} \geq M.$$  

26
That is

\[
\frac{\rho^{L(M,\rho)+1} - 1}{\rho - 1} \leq ML(M, \rho) \leq \frac{\rho^{L(M,\rho)+1} - 1}{\rho - 1} + \rho^{L(M,\rho)+1} - M.
\]

Therefore, \(\rho^{L(M,\rho)+1} \geq M\). So, for all \(\rho \in (1, \rho_0]\),

\[
(1 - \epsilon)\rho^{L(M,\rho)} \leq M \leq (1 + \epsilon)\rho^{L(M,\rho)} + \frac{\rho^{L(M,\rho)+1} - M}{L(M, \rho)}.
\]

From \((1 - \epsilon)\rho^{L(M,\rho)} \leq M\), we know that \(\lim_{\rho \to 1^+} \rho^{L(M,\rho)} < \infty\). Since, apparently, \(\lim_{\rho \to 1^+} L(M, \rho) = \infty\), we conclude that:

\[
\lim_{\rho \to 1^+} \frac{\rho^{L(M,\rho)} - M}{L(M, \rho)} = 0.
\]

So, for any \(\epsilon' > 0\), \(0 \leq \rho^{L(M,\rho)} - M \leq \epsilon' L(M, \rho)\) which implies that

\[
(1 - \epsilon)\rho^{L(M,\rho)} \leq M \leq (1 + \epsilon)\rho^{L(M,\rho)} + \epsilon' = (1 + \epsilon + \epsilon'')\rho^{L(M,\rho)}
\]

where \(\epsilon'' = \epsilon' \rho^{-L(M,\rho)}\). If we notice that \(\lim_{\rho \to 1^+} \frac{\log \rho}{\rho - 1} = 1\), we obtain that, for any \(\epsilon, \epsilon'', \epsilon''' > 0\), there exists a \(\rho_0 > 1\), such that, for all \(\rho \in (1, \rho_0]\)

\[
\frac{\log M - \log(1 + \epsilon + \epsilon'')}{(1 + \epsilon''')(\rho - 1)} \leq N \leq \frac{\log M - \log(1 - \epsilon)}{(1 - \epsilon''')(\rho - 1)}.
\]

From the above we easily conclude the desired result. \(\Box\)

We now obtain the main result of this section which provides a lower bound to the sample complexity of \(\ell_1\) worst-case identification for plants in \(\mathcal{M}(M, \rho)\).

**Theorem 3** For any \(K \geq 1\), we have

\[
n^*(\mathcal{M}(M, \rho), \ell_1, K) = \Omega \left( M^{\ell(\frac{K}{n-1})} \right).
\]

**Proof.** We notice that in the proofs of Theorems 1 and 2 in the previous sections, we only assumed that the true plant is in a subset of \(\mathcal{M}_N\) with bounded \(\ell_\infty\) norm.
Therefore we can conclude that
\[ n^*(\mathcal{M}'_N, \ell_1, K) = \Omega \left( 2^{N_1(k)} \right). \]

However, by the definition of the effective length we see that
\[ \mathcal{M}'_N \subseteq \mathcal{M}(M, \rho) \implies n^* \left( \mathcal{M}'_N(M, \rho), \ell_1, K \right) \leq n^* \left( \mathcal{M}(M, \rho), \ell_1, K \right). \]

Therefore
\[ n^* \left( \mathcal{M}(M, \rho), \ell_1, K \right) = \Omega \left( 2^{L(M, \rho)} f(k) \right) \]

from which we easily obtain the desired result by applying Proposition 1. □

As we saw, Theorem 3 provides another version of the exponential sample complexity of worst-case identification in \( \ell_1 \). This result is similar to the one in [9] for a slightly different model set.

### 2.5 Near-Optimal Identification in \( \mathcal{H}_\infty \)

In this section we shift our attention to worst-case identification using the \( \mathcal{H}_\infty \) error norm. In particular, we obtain an upper bound to the sample complexity of identification with frequency domain experiments using the two-stage, nonlinear algorithm presented in [1].

One major difference from the previous setup is that we now try to approximate the frequency response of the plant. The plant \( h \) is assumed to belong to \( \mathcal{M}(M, \rho) \) and the input is a sequence of sinusoids. Let \( n \) signify the number of frequency response points we will use for the identification. Then, there are \( n \) distinct complex exponential inputs\(^5\), each of which is applied to the plant for a long enough time to guarantee that the remaining tail of the impulse response sums to something less than our error bound.

\(^5\)These are equivalent to \( 2n \) real sinusoid inputs.
Specifically, let's fix an $\epsilon > 0$. Then the inputs are given by $u_m^{(k)} = \exp(i\omega_km)$ where $\omega_k = \frac{2\pi(k-1)}{n}$ for $k = 1, \ldots, n$. The measurements we observe are the frequency responses of the plant to the inputs above, corrupted by additive measurement noise which captures both the lost tail of the impulse response and some output disturbance, similar to the one in the time domain experiments studied in the previous sections.

Let $T(M, \rho)$ signify the amount of time we need to apply any one of the inputs so that the error level is less than or equal to $\epsilon$, i.e.

$$T(M, \rho, \epsilon) = \min \left\{ T \geq 0 \mid \sum_{m=T+1}^{\infty} |h_m| + \sqrt{2}\delta \leq \epsilon \right\}$$

where $\delta$ is the $\ell_\infty$ magnitude bound for the output disturbance as in the time domain experiments in the previous sections\(^6\). Therefore, if we call $y_k$ the $k$th measurement, then

$$y_k = \sum_{m=0}^{T(M, \rho)} h_m u_m^{(k)} + d_k, |d_k| \leq \sqrt{2}\delta \text{ for all } k$$

where $d_k$ is additive output noise. Thus, we see that

$$y_k = \sum_{m=0}^{\infty} h_m u_m^{(k)} - \sum_{m=T(M, \rho)\! + 1}^{\infty} h_m u_m^{(k)} + d_k = \sum_{m=0}^{\infty} h_m u_m^{(k)} + \eta_k$$

where $\eta_k$ is the total measurement error due to both the output disturbance and the remaining tail of the plant's impulse response. We can easily check that the choice of $T(M, \rho, \epsilon)$ guarantees that $|\eta_k| \leq \epsilon$ for all $k$.

The following Proposition quantifies the rate of growth of $T(M, \rho, \epsilon)$.

**Proposition 2** For any $M < \infty$, $\rho > 1$,

$$T(M, \rho, \epsilon) = O \left( \frac{\log M - \log(\epsilon - \sqrt{2}\delta)}{\rho - 1} \right)$$

where the big $O$ notation is meant as $M$ increases to $\infty$, $\rho$ decreases to $1^+$ and as $\epsilon$ decreases to $\sqrt{2}\delta$.

\(^6\)The factor of $\sqrt{2}$ comes about because in the frequency domain setup, the disturbance sequence is complex.
Proof. From the above experimental setup we see that

\[ |\eta_k| \leq \sum_{m=T(M,\rho)+1}^{\infty} |h_m| + \sqrt{2\delta} \]
\[ \leq \sum_{m=T(M,\rho)+1}^{\infty} |h_m| \rho^m \rho^{-m} + \sqrt{2\delta} \]
\[ \leq M \rho^{-T(M,\rho)-1} + \sqrt{2\delta}. \]

Noticing that \( \lim_{\rho \to 1^+} \frac{\log \rho}{\rho - 1} = 1 \), we get

\[ T(M, \rho, \epsilon) = O \left( \frac{\log M - \log(\epsilon - \sqrt{2\delta})}{\rho - 1} \right). \]

\[ \square \]

Let \( u \) denote the concatenation of the \( n \) complex exponential inputs discussed before, i.e. \( u_j = u_m^{(k)} \) where \( k = \left\lceil \frac{j}{T(M,\rho,\epsilon)} \right\rceil \) and \( m = j \mod T(M,\rho,\epsilon) \). So, \( u \in U_{nT(M,\rho,\epsilon)} \). Let \( \phi^{(\infty)} \) denote the algorithm proposed in [1]. This algorithm is based on interpolating between the frequency response points obtained from the experiments. This interpolation is shown in [7] to be equivalent to the following sequence of steps:

1. Obtain the inverse DFT of the measurements \( y_k \) and truncate it at \( \pm R \) for some \( R \).
2. Attenuate the time domain coefficients appropriately.
3. Take the causal part and transform it back to the frequency domain.

It is shown in [1] that

**Fact 1** If we do \( n \) frequency response experiments, each guaranteed to have an error less than or equal to \( \epsilon \), then

\[ \sup_{h \in \mathcal{H}} \sup_{d \in D_4} \| \hat{h} - \phi^{(\infty)}(y, u) \|_{\mathcal{H}} \leq 2 \min \left\{ \frac{4M\pi}{(\rho-1)n}, \frac{M\pi^2(\rho+1)}{(\rho-1)^2n^2} \right\} + \frac{4(M+\epsilon)n^2}{R\pi^2} + 2\epsilon \]

where \( \hat{h} \) is the Fourier transform of \( h \).

Now, we want to concatenate the experiments. This means we need to allow enough time in between the application of the different inputs for the plant state to
settle down to an acceptable level. Equivalently, we can do the following: Let the number of experiments be \( n \). Let's fix an error bound for each experiment independently equal to \( \epsilon/n \). This assumes that \( \epsilon \) is large enough that \( \epsilon/n > \sqrt{2\delta} \). Each experiment takes \( T(M, \rho, \epsilon/n) \) time units. At the end of the first experiment, the remaining tail of the impulse response has an \( \mathcal{H}_\infty \) norm less than or equal to \( \epsilon/n \). Now, we start the second experiment and so on. After the \( n \)th experiment, the total error that has accumulated is upper-bounded by \( \epsilon \). Using Proposition 2 we see that \( T(M, \rho, \epsilon/n) \) is \( O \left( \frac{\log M + \log n}{\rho^{-1}} \right) \).

The desired robust convergence property stipulates that

\[
\lim_{\epsilon \to 0} \lim_{n \to \infty} E_{nT(M, \rho, \epsilon/n)}(\mathcal{M}(M, \rho), \mathcal{H}_\infty, u, \varphi^{(\infty)}) = 0.
\]

One way to accomplish this is to have \( R = n^{2+\gamma} \) for some \( \gamma > 0 \). Consequently, we obtain the following upper bound to the sample complexity of worst-case identification in \( \mathcal{H}_\infty \):

**Theorem 4**

\[
n^*(\mathcal{M}(M, \rho), \mathcal{H}_\infty, K) = O \left( \frac{M \log M}{(\rho - 1)^3} \right).
\]

**Proof.** Pick \( \gamma = 1 \). We want to find \( n \) such that the right-hand-side of the inequality in Fact 1 is equal to \( 2K\epsilon \). That means

\[
n \leq \max \left\{ \frac{4M\pi}{(\rho - 1)(K - 1)\epsilon} + \frac{2(M + \epsilon)}{\pi^2(K - 1)\epsilon}, \frac{M(\rho + 1)}{(K - 1)\epsilon} \sqrt{\frac{M + \epsilon}{(K - 1)\epsilon}} \right\}
\]

\[
\leq \frac{4M\pi}{\rho - 1} \sqrt{\frac{\rho + 1}{(K - 1)\epsilon}} + \frac{4(M + \epsilon)}{\pi \sqrt{(K - 1)\epsilon}}
\]

So, \( n = O \left( \frac{M}{\rho^{-1}} \right) \). Obviously \( n^*(\mathcal{M}(M, \rho), \mathcal{H}_\infty, K) \leq nT(M, \rho, \epsilon/n) \). Using Proposition 2 and the fact that

\[
\lim_{\rho^{-1} \to \infty} \frac{\log \rho^{-1}}{\rho^{-1} - 1} = 0
\]

we obtain the desired upper bound. \( \square \)

The above theorem has three corollaries:
Corollary 1 The sample complexity of the two-stage, nonlinear algorithm proposed in [1], with $R = n^3$, is exponential in the effective length of the plant to be identified.

Proof. Comparing Theorem 4 with Proposition 1 of section 2.4, we easily see that the effective length of a plant in $\mathcal{M}(M, \rho)$ grows as $\log M$ whereas the required length of experimentation to be guaranteed a worst-case $\mathcal{H}_\infty$ error less than or equal to $2K\epsilon$ using $\phi^{(\infty)}$ grows as $M \log M$. □

Corollary 2 If we are given a fixed $M$, then the sample complexity of worst-case identification in $\mathcal{H}_\infty$ is polynomial in the effective length of the plant to be identified.

Proof. Indeed, for a fixed $M$, we can identify the plant using $\phi^{(\infty)}$ and an input of length that grows as the square of the effective length, $L(M, \rho)$, of the plant. □

The above Corollary is very interesting because it seems to differentiate between worst-case identification in $\ell_1$ and in $\mathcal{H}_\infty$. In particular, it says that, in the worst-case setup, if we have a bound on the overshoot of the plant's impulse response, then experiments in the frequency domain using the $\mathcal{H}_\infty$ error norm are significantly more effective than experiments in the time domain using the $\ell_1$ error norm.

Corollary 3

$$n^*(\mathcal{M}_N', \mathcal{H}_\infty, K) = O(N^3)$$

Proof. Fix an $M$. For every $N$, there is a $\rho > 1$ such that $L(M, \rho) = N$. Then, by combining Proposition 1 and Theorem 4 we get the desired corollary. □
<table>
<thead>
<tr>
<th>$M_N$</th>
<th>$M'_N$</th>
<th>$M(M, \rho)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_\infty$</td>
<td>$\mathcal{O}(N^3)$</td>
<td>$\mathcal{O}\left(\frac{M \log M}{(\rho - 1)^3}\right)$</td>
</tr>
<tr>
<td>$l_1$</td>
<td>$\mathcal{O}\left(2^{Nf\left(\frac{4}{K}\right)}\right)$</td>
<td>$\mathcal{O}\left(2^{Nf\left(\frac{1}{K}\right)}\right)$</td>
</tr>
</tbody>
</table>

Table 2.1: Sample Complexity Summary

### 2.6 Summary of Sample Complexities for Worst-Case Identification

In this chapter we have analyzed the sample complexity of worst-case identification for various model sets and error norms. Table 2.1 is a summary of the results that we obtained.
Chapter 3

Uncertainty Set Geometry and Off-Line Identification Algorithms

3.1 Introduction

Until now, we focused on the sample complexity of worst-case identification, mainly using the $\ell_1$ error norm. As we saw in Chapter 2, even suboptimal identification schemes require experiments whose length grows as an exponential of the number of parameters to be identified. Therefore, any identification algorithm is subject to this unfortunate fact. However, this is one of the few properties common to all worst-case identification algorithms. Now we turn our attention to these various algorithms in order to comprehend the tradeoffs involved in their use.

First we present an interpolatory algorithm. Such an algorithm provides estimates that are guaranteed to be in the uncertainty set at any given time. Therefore it is a natural type of algorithm, since any member of the uncertainty set could be the true plant. Our algorithm is implemented using linear programming. Its worst-case error is upper-bounded by $2\delta$, where $\delta$ is the bound on the magnitude of the output disturbance.

We then concentrate on the class of convex and balanced model sets. This case of identification has many interesting properties, which one can make use of in designing efficient algorithms. Besides, many model sets that arise in practical applications are
easily shown to be convex and balanced. For this class of model sets we present a specialized linear programming algorithm that calculates the diameter of the uncertainty set, provides a plant estimate and gives an upper-bound to its worst-case error.

Further, we discuss the general case where the model set is not required to be balanced\textsuperscript{1}. We formulate the problem and use various geometric arguments to develop upper bounds for the radius of the uncertainty set. These bounds are readily used to construct off-line approximation schemes by means of successive linear programs.

It should be pointed out that the algorithms in this chapter are presented in the context of the $\mathcal{M}_N$ model set and its subsets. In particular, they use the entire input-output history which is finite for the model sets considered in this Chapter. In Chapter 4 we discuss the on-line implementation of these algorithms that will make them applicable to other model sets as well.

### 3.2 Interpolatory Linear Programming Algorithm

As was mentioned in the introduction, an algorithm is called interpolatory if the estimate it provides is a member of the uncertainty set\textsuperscript{2}. In section 1.1 we defined the uncertainty set as the set of plants in the model set compatible with the input-output data so far. So, we want to be able to choose an $h \in S_n(\mathcal{M}_N, u, y, \delta)$. This can be readily accomplished if we notice that the uncertainty set is a polyhedron. We see that, for any $u \in U_n$,

\[
f \in S_n(\mathcal{M}, u, y, \delta) \iff \left| y_j - \sum_{i=1}^{N} f_i u_{j-i} \right| \leq \delta, \quad j = 1, \ldots, n
\]

\[
\iff y_j - \delta \leq f^T u^{(j)} \leq y_j + \delta, \quad j = 1, \ldots, n
\]

where $u^{(j)} = (u_{j-1}, \ldots, u_{j-N})^T$.

We would like to put the above set of inequalities in standard form\textsuperscript{3}. Thus we

---

\textsuperscript{1}The convexity requirement remains.

\textsuperscript{2}see for example [11]

\textsuperscript{3}Good references for the linear programming techniques used in this thesis is [13], [3].
write:

\[ f = f^+ - f^-, \text{ where } f^+, f^- \geq 0 \]

and we add slack variables \( z \geq 0 \) so:

\[ f \in S_n(M, u, y, \delta) \iff Ax = b, x \geq 0 \]

where

\[
A = \begin{pmatrix}
  u^{(1)T} & -u^{(1)T} \\
  -u^{(1)T} & u^{(1)T} \\
  \vdots & \vdots \\
  \end{pmatrix},
\quad
x = \begin{pmatrix} f^+ \\ f^- \\ z \end{pmatrix},
\quad
b = \begin{pmatrix} y_1 + \delta \\ -y_1 + \delta \\
  \vdots \end{pmatrix}
\]

\[ A \in \mathbb{R}^{2n \times 2(N+n)}, \quad x \in \mathbb{R}^{2(N+n)}, \quad b \in \mathbb{R}^{2n} \]

Consider the dummy cost function \( c(x) = 0x \). Then, we can think of the problem:

\[
\min_{Ax=b, \ x \geq 0} c(x)
\]

as the primal. We can run Phase I of the simplex algorithm with it and obtain a feasible solution, that is an element \( f \) of \( S_n(M, u, y, \delta) \). Actually, Phase I requires the introduction of one new variable per constraint. In particular, we need \( 2n \) more variables. The tableau size for this Phase I simplex is \( (2n+1) \times (2N+4n+1) \). Even using the revised simplex, the CARRY tableau size is \( (2n+1) \times (2n+1) \).

As we see, this tableau increases with the number of constraints. In other words, if we were to run our interpolatory algorithm after \( n-1 \) measurements, in incorporating the \( n \)th measurement, we would need to increase the size of the optimal tableau from the end of the previous run.
These results motivate us to formulate the dual LP, again in standard form:\(^4\):

\[
\min_{B\pi=0, \pi \geq 0} \pi^T b
\]

where

\[
B = \begin{pmatrix}
-u^{(1)} & u^{(1)} & \cdots \\
\end{pmatrix}
\]

\[
B \in \mathbb{R}^{N \times 2n}, \pi \in \mathbb{R}^{2n}, b \in \mathbb{R}^{2n}
\]

The dual is easy to initiate because \(\pi = 0\) is trivially feasible. If we run the dual simplex to optimality, we will either

1. have an optimal dual solution (of cost 0) or
2. see that the dual problem is unbounded.

It is not possible for the dual to be infeasible because \(\pi = 0\) is a feasible dual solution. Case 2 would imply that the primal is infeasible. Such a result can only be reached either because of a mistake in the calculations, or because of faulty data, since we normally know that, at least the true plant is in the uncertainty set.

In case 1, simplex on the dual terminates with a dual optimal vector \(\pi\) and a dual basis \(B_{\text{basis}}\). We can use the final dual tableau to obtain a primal optimal solution, which is of course also primal feasible:

\[
f = f^+ - f^- = -\left(B_{\text{basis}}^T\right)^{-1}b_{(B)}
\]

where \(b_{(B)}\) is a vector made up of the components of \(b\) corresponding to the basic dual variables. So, we have seen how we can construct a plant in the uncertainty set solving the dual problem.

The tableau size for the dual problem is \((N + 1) \times (2n + 1)\). If we use the revised simplex, then the CARRY tableau size is \((N + 1) \times (N + 1)\) which is independent of \(n\). In other words, the CARRY tableau of the dual problem has the desired property

\(^4\)In fact, this is the dual of the original primal, before we added the extra variables in order to put it in standard form, i.e. \(\min 0f\) subject to \(y_j - \delta \leq f^T u^{(j)} \leq y_j + \delta \) for \(j = 1, \ldots, n\).
that its size remains unchanged when more observations are incorporated.

Therefore, we see how to implement an interpolatory algorithm using the revised simplex on the dual problem. Any interpolatory algorithm is guaranteed to have an error less than the diameter of the uncertainty set. We know already from Chapter 2 that $D_\infty^*(\mathcal{M}_N, \ell_1, n^*(N)) = 2\delta$, for appropriately chosen $u \in U_{n^*(N)}$, where $\delta$ is the bound on the magnitude of the output disturbance. Thus, if we call $\phi^I_n(y, u)$ the estimate of the interpolatory algorithm after $n$ observations and if we choose $u \in U_{n^*(N)}$ as described in section 2.2, then:

$$E_{n^*(N)}(u, \phi^I_{n^*(N)}, \mathcal{M}_N, \ell_1) \leq 2\delta$$

### 3.3 Balanced Model Sets and Optimal Linear Algorithms

The discussion in the previous section was general. We showed constructively for the $\mathcal{M}_N$ model set that the uncertainty sets of worst-case identification in $\ell_1$ are polyhedra. So, they are convex. The interpolatory algorithm described in the previous section has an error guaranteed to be less than or equal to $2\delta$ if the input sequence is chosen appropriately, without any further assumptions on the prior information (model set). In this section, we show that under a certain condition on the model set we can construct a linear programming algorithm that pushes the worst-case error down towards $\delta$, which is a lower bound to the optimal worst-case error achievable by any algorithm as shown in [5].

One condition that may decrease the optimal worst-case error below $2\delta$ is that the model set be balanced.

**Definition 5** We say that a set $X$ is balanced if $f \in X \iff -f \in X$.

Another way to say that $X$ is balanced is to say that $X$ is center-symmetric about 0.

As was mentioned in the introduction, many practical applications give rise to
balanced model sets. For example, both the $\mathcal{M}_N$ and the $\mathcal{M}(M, \rho)$ model sets presented in Chapter 1 are balanced as one can easily check. The next proposition is a crucial property of balanced uncertainty sets. It has been presented in many forms before. The closest to the one presented here can be found in [5].

**Proposition 3** If $\mathcal{M}$ is a convex and balanced model set, then the worst-case diameter for any uncertainty set is attained when the true plant and the disturbance are both 0. That is,

$$D_n(\mathcal{M}, \ell_1, u) = 2 \sup_{g \in S_n(\mathcal{M}, u, 0, \delta)} \|g\|_1$$

**Proof.** For an arbitrary $y$, let $g_1, g_2 \in S_n(\mathcal{M}, u, y, \delta)$. Then $\|P_n(u \ast (g_1 - g_2))\|_\infty \leq 2\delta$. Because $\mathcal{M}$ is convex and balanced, $\frac{g_1 + g_2}{2} \in \mathcal{M}$. So

$$\text{diam}(S_n(\mathcal{M}, u, y, \delta), \ell_1) \triangleq \sup_{g_1, g_2 \in S_n(\mathcal{M}, u, y, \delta)} \|g_1 - g_2\|_1$$

$$\leq \sup_{\|P_n(u \ast (g_1 - g_2))\|_\infty \leq 2\delta} \|g_1 - g_2\|_1 = 2 \sup_{g \in S_n(\mathcal{M}, u, 0, \delta)} \|g\|_1$$

So $D_n(\mathcal{M}, \ell_1, u) \leq 2 \sup_{g \in S_n(\mathcal{M}, u, y, \delta)} \|g\|_1$. But

$$\text{diam}(S_n(\mathcal{M}, u, y, \delta), \ell_1) \geq 2 \sup_{g \in S_n(\mathcal{M}, u, y, \delta)} \|g\|_1.$$ 

So we obtain the required result. $\square$

Now, we construct the desired balanced-set algorithm. This algorithm tries to find a plant $f \in S_n(\mathcal{M}, u, y, \delta)$ that minimizes $\|P_n(f \ast u - y)\|_\infty$. If $\Lambda_{\text{max}}(n) = \infty$, we saw above that the uncertainty set is center-symmetric and the estimate of the algorithm is the center of symmetry. If on the other hand, $\Lambda_{\text{max}}(n) < \infty$, the uncertainty set isn’t center-symmetric. However, the algorithm tries to find an approximate center of symmetry.

Consider the following linear program:

$$\max_{\tilde{A} \tilde{x} = b, \tilde{x} \geq 0} \Lambda$$

39
where

\[
\tilde{A} = \begin{pmatrix}
    u^{(1)^T} & -u^{(1)^T} & -y_1 \\
    -u^{(1)^T} & u^{(1)^T} & y_1 \\
    \vdots & \vdots & \vdots \\
    f^+ & f^- & z \\
    \Lambda
\end{pmatrix}, \quad \tilde{x} = \begin{pmatrix}
    y_1 + \delta \\
    -y_1 + \delta \\
    \vdots
\end{pmatrix}, \quad b = \begin{pmatrix}
    y_1 \\
    -y_1
\end{pmatrix}
\]

\[
\tilde{A} \in \mathbb{R}^{2n \times (2N+2n+1)}, \quad \tilde{x} \in \mathbb{R}^{2N+2n+1}, \quad b \in \mathbb{R}^{2n}
\]

The above linear program finds the largest \( \Lambda \) for which \( S_n(\mathcal{M}, u, (1+\Lambda)y, \delta) \neq \emptyset \). Let \( \tilde{x}^* \) be the optimal solution of the above LP and let \( \Lambda_{\text{max}}(n) = \tilde{x}^*_{2N+2n+1} \) and \( f_n^* = ((\tilde{x}_1^*)^T, \ldots, (\tilde{x}_N^*)^T)^T - ((\tilde{x}_{N+1}^*)^T, \ldots, (\tilde{x}_{2N}^*)^T)^T \). Also, \( \forall f \in S_n(\mathcal{M}, u, y, \delta) \), let \( z(f, y) = b + B^T f \), where \( B = (-u^{(1)}, u^{(1)}, \ldots) \) was defined in section 3.2. The following lemma provides an alternative way to obtain \( \Lambda_{\text{max}}(n) \). This lemma is then used to construct a new algorithm and obtain its performance guarantees.

**Lemma 6** Let \( \mathcal{M} \) be convex and balanced model set (still a subset of \( \mathcal{M}_N \)) and \( f \in S_n(\mathcal{M}, u, y, \delta) \). Let \( f = f_1 + f_2 \) be the unique decomposition of \( f \) into components in \( \text{span}(S_n(\mathcal{M}, u, 0, \delta)) \) and \( \text{span}(S_n(\mathcal{M}, u, 0, \delta))^\perp \), where \( \text{span}(C) \) for a subset \( C \) of \( \mathbb{R}^N \) is the set of all points in \( \mathbb{R}^N \) that can be written as linear combinations of points in \( C \). Define \( J(f) = \{ i | (z(f, y))_i < (z(f_1, 0))_i \} \). Then:

\[
\Lambda_{\text{max}}(n) = \max_{f \in S_n(\mathcal{M}, u, y, \delta)} \min_{i \in J(f)} \frac{(z(f, y))_i - (z(f_1, 0))_i}{(z(f, y))_i}
\]

**Proof.** We know that \( f_1 \in S_n(\mathcal{M}, u, 0, \delta) \Rightarrow \exists x_1 \geq 0 \) such that \( Ax_1 = (\delta, \delta, \ldots)^T \), where \( A \) is the matrix defined in section 3.2. Further, we know that \( f_2 \in S_n(\mathcal{M}, u, 0, \delta)^\perp \). Since \( f \in S_n(\mathcal{M}, u, y, \delta) \), we conclude that \( \exists x_2 \geq 0 \) such that \( Ax_2 = (y_1, -y_1, \ldots)^T + Ax_1 \). So, there exists \( x_3 = x_2 - x_1 \) such that \( Ax_3 = (y_1, -y_1, \ldots)^T \). Therefore:

\[
\tilde{A} = \begin{pmatrix}
    A & -Ax_3
\end{pmatrix}
\]
Let
\[
\tilde{x}_1 = \begin{pmatrix} f_1^+ \\ f_1^- \\ z(f_1,0) \\ 0 \end{pmatrix}
\] and
\[
\tilde{x}_2 = \begin{pmatrix} (1 + \Lambda) f_2^+ \\ (1 + \Lambda) f_2^- \\ (1 + \Lambda)(z(f,y) - z(f_1,0)) \\ \Lambda \end{pmatrix}
\]

Then, indeed
\[
\tilde{A}(\tilde{x}_1 + \tilde{x}_2) = b.
\]

Therefore, \( h = f + \Lambda f_2 \in S_n(\mathcal{M}, u, (1 + \Lambda)y, \delta) \) as long as
\[
z(f_1,0) + (z(f,y) - z(f_1,0))(1 + \Lambda) \geq 0
\]

which is true for any \( \Lambda \geq 0 \) if \( z(f,y) \geq z(f_1,0) \) and implies that
\[
\Lambda = \min_{i \in J(f)} \frac{(z(f,y))_i}{(z(f_1,0))_i - (z(f,y))_i}
\]

Since this is true for any \( f \in S_n(\mathcal{M}, u, y, \delta) \), we conclude the desired result. \( \Box \)

In order to define the estimate of the algorithm, we need to prove the following proposition.

**Proposition 4** If after \( n \) measurements we have \( \Lambda_{\max}(n) = \infty \), then there exists an \( f \) in \( S_n(\mathcal{M}, u, y, \delta) \) such that \( P_n(f * u) = P_n(y) \).

**Proof.** \( \Lambda_{\max}(n) = \infty \) means that, for any \( M \geq 1 \), there exists a plant \( f(M) \) in \( S_n(\mathcal{M}, u, My, \delta) \). So
\[
\|P_n(f(M) * u - My)\|_{\infty} \leq \delta \implies \|P_n(\frac{f(M)}{M} * u - y)\|_{\infty} \leq \frac{\delta}{M}.
\]

So, there exists a plant \( h \) in \( S_n(\mathcal{M}, u, y, \delta) \) such that \( \|P_n(f * u - y)\|_{\infty} \leq \frac{\delta}{M} \). Since this is true for arbitrary \( M \), we conclude the required relation. \( \Box \)

In other words, if \( \Lambda_{\max}(n) = \infty \), the \( n \) equations in \( P_n(f * u) = P_n(y) \) are consistent. Let \( \text{rank}(u^{(1)}, u^{(2)}, \ldots) = R \leq N \). Construct the matrix \( A_R \) so that it has as
its rows the first $R$ row vectors $u^{(i)}$ that are linearly independent. Let $y_R$ be the part of $y$ that corresponds to the $u^{(i)}$ in $A_R$. Finally, let $\bar{A}_R$ be the matrix that consists of only the first $R$ columns of $A_R$. Then, we define our estimate to be

$$
\varphi_n^{\text{bal}}(y, u) = \begin{cases}
\frac{f_n}{1 + \Lambda_{\max}(n)} & \text{if } \Lambda_{\max}(n) < \infty \\
(y_R^T \bar{A}_R - T, 0, \ldots, 0)_{N-R}^T & \text{if } \Lambda_{\max}(n) = \infty
\end{cases}
$$

Now we prove a lemma about center-symmetric sets and then we proceed to the main theorem of this section, which establishes bounds on the error of the estimate described above.

**Lemma 7** Let $C$ be a center-symmetric subset of a metric space $\mathcal{X}$ and let $c$ denote its center of symmetry. Then

$$
\text{rad}(C, \mathcal{X}) = \sup_{x \in C} \rho_{\mathcal{X}}(c - x) = \frac{\text{diam}(C, \mathcal{X})}{2}
$$

**Proof.** We know that $g \in C \iff 2c - g \in C$. Then

$$
\forall x \in C, \quad \rho_{\mathcal{X}}(c - x) \leq \frac{\text{diam}(C, \mathcal{X})}{2}
$$

because $\rho_{\mathcal{X}}(c - x) > \frac{\text{diam}(C, \mathcal{X})}{2} \Rightarrow \rho_{\mathcal{X}}(2c - x - x) = 2\rho_{\mathcal{X}}(c - x) > \text{diam}(C, \mathcal{X})$ which is a contradiction. From the definition of the radius of a set though we see that $\text{rad}(C, \mathcal{X}) \leq \sup_{x \in C} \rho_{\mathcal{X}}(c - x)$. But we also know that, for any set $C$, $\text{rad}(C, \mathcal{X}) \geq \frac{\text{diam}(C, \mathcal{X})}{2}$. So, we have proved the required equalities. □

**Theorem 5** i. For any length $n$ of experimentation and for any $u \in U_n$, we have the following lower bound for the worst-case error of the balanced-set algorithm:

$$
\max_{\delta \in \delta_{\bar{e}_m}, \bar{e}_n} K \leq E_n(u, \varphi_n^{\text{bal}}, \mathcal{M}_N, \ell_1)
$$

where $\bar{e}_m = (1, 1, \ldots)^T \in \mathbb{R}^m$, $\bar{e} = (g^+, g^-, K, z)^T$ with $g^+, g^- \in \mathbb{R}^N$, $K \in \mathcal{R}$,
$z \in \mathbb{R}^{2n+2N}$ and

$$D = \begin{pmatrix}
  u^{(1)T} & -u^{(1)T} \\
  -u^{(1)T} & u^{(1)T} \\
  \vdots & \vdots \\
  u^{(n)T} & -u^{(n)T} \\
  -u^{(n)T} & u^{(n)T} \\
  1 & 1 & \cdots & -1 & -1 & \cdots & -1 \\
  -1 & 1 & \cdots & 1 & -1 & \cdots & \vdots \\
  1 & -1 & \cdots & -1 & 1 & \cdots & \vdots \\
  \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots 
\end{pmatrix}$$

$D \in \mathbb{R}^{(2n+2N)\times(2N+2n+1+2N)}, \bar{x} \in \mathbb{R}^{2N+2n+1+2N}$

ii. If there exist an $f$ in $S_n(\mathcal{M}, u, y, \delta)$ such that there is at most one $i \leq n$ for which $|\langle f, u - y \rangle_i| = \delta$, then $\Lambda_{\text{max}}(u) > 0$.

iii. If $\Lambda_{\text{max}}(n) = \infty$, then, $E_n(u, \phi_n^{bal}, \mathcal{M}_N, \ell_1) = \delta$.

iv. Let $\mathcal{A}_n = \{u \in U_n | \forall v \in \mathbb{R}^N, \|u * v\|_\infty = \|v\|_1\}$. Let $\hat{n}(N) = \min \{n | \mathcal{A}_n \neq \emptyset\}$. Then, for any $n \geq \hat{n}(N)$ and for any $u \in \mathcal{A}_n$, if $\Lambda_{\text{max}}(n) < \infty$, we have

$$E_n(u, \phi_n^{bal}, \mathcal{M}_N, \ell_1) \leq \left(1 + \frac{1}{1 + \Lambda_{\text{max}}(n)}\right) \delta.$$

**Proof.** Part i: Since $\mathcal{M}$ is convex and balanced, from Proposition 3 we conclude that $D_n(\mathcal{M}, \ell_1, u) = 2 \max_{g \in S_n(\mathcal{M}, u, 0, \delta)} \|g\|_1$. But we know that

$$g \in S_n(\mathcal{M}, u, 0, \delta) \Rightarrow A(g^+, g^-, z)^T = \delta \bar{e}_{2n}, g = g^+ - g^- \text{ with } g^+, g^-, z \geq 0$$

Therefore, we see that

$$D_n(\mathcal{M}, \ell_1, u) =$$

$$2 \max \left\{K | \exists g^+, g^-, z \geq 0, A(g^+, g^-, z)^T = \delta \bar{e}_{2n}, \|g^+ - g^-\|_1 \leq K\right\}$$
Writing out the set over which we are maximizing $K$ as a polyhedron\(^8\), and observing that for any convex subset $C$ of $\ell_1$, $\forall x \in C$, $\max_{y \in C} \|x - y\|_1 \geq \frac{1}{2} \text{diam}(C, \ell_1)$, we obtain the required lower bound. Notice that for $n \geq n^*(N)$ the lower bound becomes $\delta$.

**Part ii:** Let $f$ be as in the statement of part ii. Let $i = \arg \max_{j \leq n} \| (f \ast u - y)_j \|$. We distinguish two cases:

*Case a:* This is the case $\| (f \ast u - y)_i \| < \delta$. Then there exists a small enough $\Lambda > 0$ such that $\| P_n(f \ast u - y - \Lambda y) \|_\infty \leq \delta$, so we are done.

*Case b:* This is the case $\| (f \ast u - y)_i \| = \delta$. From the statement of part ii we know that for all $j \neq i$, $j \leq n$, we have $\| (f \ast u - y)_j \| < \delta$. In this case we can find a small enough $\epsilon > 0$ such that, if we define $f'_m = f_m + \epsilon \text{sgn}(u_{i-m})\text{sgn}(y_i)$ for all $m = 1, \ldots, N$, \(\| P_n(f' \ast u - y) \|_\infty < \delta\). Therefore, we put ourselves in *Case a* once again and we are done.

**Part iii:** We know that, in this case, $P_n(\phi_n^{bal}(y, u) \ast u) = P_n(y)$. Thus, $g \in S_n(M, u, y, \delta) \iff 2\phi_n^{bal}(y, u) - g \in S_n(M, u, y, \delta)$ since $\| P_n(2\phi_n^{bal}(y, u) \ast u - g \ast u - y) \|_\infty = \| P_n(y - g \ast u) \|_\infty \leq \delta$. This means that $S_n(M, u, y, \delta)$ is center-symmetric about $\phi_n^{bal}(y, u)$. Using Lemma 7 we conclude that $\forall g \in S_n(M, u, y, \delta)$, $\| P_n(\phi_n^{bal}(y, u) - g) \|_1 = \frac{\text{diam}(S_n(M, u, y, \delta), \ell_1)}{2} \leq \frac{D_n(M, \ell_1, u)}{2} = \delta$.

**Part iv:** Let $f$ be an arbitrary element of $S_n(M_N, u, y, \delta)$. Then

$$\| f \ast u - (1 + \Lambda_{\max}(n))y \|_\infty \leq \delta \iff \| \phi_n^{bal}(y, u) \ast u - y \|_\infty \leq \frac{\delta}{1 + \Lambda_{\max}(n)}$$

Now, let's assume that there exists a plant $g \in S_n(M_N, u, y, \delta)$ such that

$$\| \phi_n^{bal}(y, u) - g \|_1 > \left(1 + \frac{1}{1 + \Lambda_{\max}(n)}\right) \delta$$

Then

\(^8\)of course with exponentially many constraints
\[ \| (\phi_n^{bal}(y, u) - g) * u + d \|_\infty \geq \left(1 + \frac{1}{1 + \Lambda_{\text{max}}(n)}\right) \delta - \delta = \frac{\delta}{1 + \Lambda_{\text{max}}(n)} \]

So we have a contradiction and we obtain the required upper bound. \( \square \)

To conclude, in this section we have seen how we may decrease the achievable worst-case error below 2\( \delta \) using the balanced-set algorithm. Next we discuss another algorithm which applies even when the model set isn't balanced.
\[ \text{diam}^{(1)}(ACE) = ||C-D||_1 = 2a \]
\[ \text{rad}^{(1)}(ACE) = ||A-B||_1 = \frac{4a}{3} \]

Figure 3-1: Example of set in \( \mathbb{R}^2 \) with \( \text{rad}(C, \ell_1) > \frac{\text{diam}(C, \ell_1)}{2} \)

### 3.4 \( \ell_1 \) Geometry and Interior Point Identification Algorithms

The case of balanced models sets analyzed in the previous section is only a special case. In this section we will present some results for general convex model sets\(^6\). For an example of a nonbalanced set with the \( \ell_1 \) topology, that is a set whose radius is more than half its diameter, see Figure 3-1.

First, we will present a number of geometric inequalities for polyhedra in \( \mathbb{R}^N \) under the Euclidean metric (\( \ell_2 \)) and the rectilinear metric (\( \ell_1 \)). In particular, we generalize the concept of incenter and inradius from plane geometry. The incenter is a concept in some sense dual to the Chebyshhev center. More specifically, while the Chebyshhev center is the point that minimizes the maximum distance to any point in the set, the incenter is the point that maximizes the minimum distance to any point on the boundary of the set.

**Definition 6** Let \( C \) be a nonempty, closed and bounded subset of \( \mathbb{R}^N \). Define the

---

\(^6\)We still only consider the \( \mathcal{M}_N \) model set and its subsets.
inradius \( r(C, \ell_p) \) and incenter \( I(C, \ell_p) \) as

\[
r(C, \ell_p) = \max_{x \in C} \min_{y \in \partial C} \|x - y\|_p
\]

\[
I(C, \ell_p) = \arg \max_{x \in C} \min_{y \in \partial C} \|x - y\|_p
\]

**Lemma 8** Let \( C \) be a nonempty, closed, bounded and convex subset of \( \mathbb{R}^N \). If there is a point \( x \) and a positive, finite scalar \( a \) such that \( \forall y \in \partial C, \|x - y\|_p \geq a \), then

\[
\forall z \in C, \|x - z\|_p \leq \text{diam}(C, \ell_p) - a
\]

**Proof.** Say that there exists a \( \hat{z} \) such that \( \|\hat{z} - x\|_p > \text{diam}(C, \ell_p) - a \). Since \( C \) is closed and bounded,

\[
L = \partial C \cap \{v \in C | v = (1 - \beta)\hat{z} + \beta x, \beta \in \mathbb{R} \} \neq \emptyset
\]

Pick a \( v \in L \). Then

\[
\|\hat{z} - v\|_p = \|\hat{z} - x\|_p + \|x - v\|_p > \text{diam}(C, \ell_p)
\]

This contradicts the definition of the diameter. So we have proven the required inequality. \( \square \)

The following two propositions derive a lower bound for the inradius of a set in \( \mathbb{R}^N \):

**Proposition 5** Let \( P = A_1, A_2, \ldots, A_{N+1} \) be a nonempty, bounded simplex in \( \mathbb{R}^N \) with vertices \( A_1, A_2, \ldots, A_{N+1} \). Let \( \text{vol}(P) \) denote its \( N \)-dimensional volume and \( F_i(P) \) denote the \( (N-1) \)-dimensional volume of the face \( \hat{A}_{(i)} = A_1, A_2, \ldots, A_{i-1}, A_{i+1}, \ldots, A_{N+1} \). Then

\[
\text{vol}(P) \geq \frac{F(P)h_{\min}(P, \ell_2)}{N(N + 1)}
\]

where \( F(P) = \sum_{i=1}^{N+1} F_i(P) \) is the total \( (N - 1) \)-dimensional volume of \( \partial P \) and \( h_{\min}(P, \ell_p) = \min_{1 \leq i \leq N+1} \min_{x \in A_{(i)}} \|A_i - x\|_p \) is the minimum generalized height of
the set $P$.

**Proof.** The polyhedron $P$ has $N + 1$ faces. If $h_{\text{min}}$ corresponds to face $\hat{A}_{(j)}$, then $F_{j}(P) \geq \frac{F(P)}{N + 1}$. Besides, we know that$^7$:

$$N\text{vol}(P) = F_i(P) h_i(P, \ell_2), \quad \forall i \in [1, N + 1]$$

where $h_i(P, \ell_2) = \min_{x \in \hat{A}_{(i)}} \|A_i - x\|_2$ is the height corresponding to face $\hat{A}_{(i)}$. So, we conclude the required inequality. $\square$

**Proposition 6** For the set $P$ of Proposition 5 we have

$$r(P, \ell_1) \geq \frac{h_{\text{min}}(P, \ell_2)}{N + 1}$$

**Proof.** In [12] we find that $r(P, \ell_2) = \frac{N\text{vol}(P)}{F(P)}$. Using the inequality from Proposition 5 we see that $r(P, \ell_2) \geq \frac{h_{\text{min}}(P, \ell_2)}{N + 1}$. Noticing that $\forall a \in \mathbb{R}^N, \|a\|_2 \leq \|a\|_1$, we conclude the required lower bound for the inradius of $P$ in the $\ell_1$ topology. $\square$

Using the concept of inradius and the inequalities derived above we now proceed to a theorem that sets an upper bound to the radius of the optimal worst-case uncertainty set.

**Theorem 6** Let $V = \{V_1, V_2, \ldots\}$ be the set of the vertices$^8$ of $S_n(M, u, y, \delta)$. Let $M = \text{rank}\{u^{(1)}, u^{(2)}, \ldots, u^{(n)}\} \leq N$. Finally, let $V(k) = \{V' \subset V | \text{card}(V') = k\}$. Then

i. There exists a $V' \in V(M + 1)$ such that $h_{\text{min}}(\text{conv}(V'), \ell_2) > 0$. in particular, by Proposition 6 this implies that $r(\text{conv}(V'), \ell_1) > 0$.

ii.

$$R_n(M, \ell_1, u) \leq \text{diam}(S_n(M, u, y, \delta), \ell_1) - \max_{V' \in V(M + 1)} r(\text{conv}(V'), \ell_1)$$

where $\text{conv}(V')$ is the convex hull of $V'$.

$^7$see [12]

$^8$The uncertainty sets are polyhedra since we only consider the $M_N$ model set and its subsets.
Proof. Part i: By the definition of \( M \), we know that there exist a set \( V' \in V(M+1) \) such that the \( M + 1 \) points in \( V' \) are not on the same \( M - 1 \) dimensional hyperplane. So, for any \( V_i \in V' \), \( h_i(\text{conv}(V'), \ell - 2) > 0 \).

Part ii: Pick any \( V' \in V(M+1) \). By the definition of the incenter and by the convexity of the uncertainty set we have

\[
\forall x \in \partial S_n(M, u, y, \delta), \| I(S_n(M, u, y, \delta), \ell_1) - x \|_1 \geq r(S_n(M, u, y, \delta), \ell_1) \geq r(\text{conv}(V'), \ell_1).
\]

Using Lemma 8 we see that

\[
\exists c \in S_n(M, u, y, \delta) \text{ such that } \forall g \in S_n(M, u, y, \delta), \| c - g \|_1 \leq \text{diam}(S_n(M, u, y, \delta), \ell_1) - \max_{V' \in V(M+1)} r(\text{conv}(V'), \ell_1).
\]

From the definition of the radius we see that \( \forall c \in S_n(M, u, y, \delta), R_n(M, \ell_1, u) \leq \sup_{g \in S_n(M, u, y, \delta)} \| c - g \|_1 \) and therefore we conclude the desired upper bound to the worst-case radius. \( \square \)

The above theorem gives rise to our identification algorithm for the general (non-balanced) case. First we need a series of definitions:

Definition 7 For any \( V' \in V(M+1) \) define

\[
I(V') = \{ i | V_i \in V' \}
\]

\[
\forall i \in I(V'), \hat{V}'(i) = V' \setminus \{ V_i \}
\]

\[
\forall x \in \text{conv}(V') \text{ and } \forall i \in I(V'), \rho_{V', i}(x) = \min_{y \in \text{conv}(\hat{V}'(i))} \| x - y \|_1
\]

Definition 8 By Carathéodory's Theorem\(^9\) we know that, if \( C \) is a polyhedron in \( \mathbb{R}^N \), any point \( x \in C \) can be written as a convex combination of at most \( N + 1 \) of the

\(^9\)see [16]
vertices of $C$. Let $\lambda_i(x)$ denote the coefficients of the convex combination, where $i$ is an index for the $N + 1$ necessary vertices.

The following is a recently proven fact that we use in the algorithm:

**Fact 2 (Tsintsifas, Klamkin)**\(^{10}\) Let $P = A_1, A_2, \ldots, A_{N+1}$ be a simplex of $\mathbb{R}^N$. Then

$$\min_{\pi \in P} \sum_{i=1}^{N+1} \lambda_i(x)\rho_{P,i}(x) = r(P, \ell_1).$$

**Algorithm 1** The two-stage hierarchical algorithm is:

1. Pick a nonempty $V' \in V(M + 1)$.

2. Calculate $r(\text{conv}(V'), \ell_1)$ as in Fact 2. This calculation involves the solution of the following two-stage hierarchical linear program:

$$\min \sum_{i \in I(V')} \lambda_i \rho_{V',i} \left( \sum_{i \in I(V')} \lambda_i V_i \right)$$

subject to

$$\sum_{i \in I(V')} \lambda_i = 1$$

$$\lambda_i \geq 0 \text{ for all } i \in I(V').$$

For each choice of the $\lambda$s we have to solve $M+1$ LPs to calculate $\rho_{V',i} \left( \sum_{i \in I(V')} \lambda_i V_i \right)$ (one for each $i \in I(V')$). Let the solution of the two-stage hierarchical LP be $\lambda_i^*$. Then define: $c(V') = \sum_{i \in I(V')} \lambda_i^* V_i$.

3. Using Theorem 6, the solution $c(V')$ of step 2 is guaranteed to have

$$\forall z \in S_n(\mathcal{M}, u, y, \delta), \|z - c(V')\|_1 \leq \text{diam}(S_n(\mathcal{M}, u, y, \delta), \ell_1) - r(\text{conv}(V'), \ell_1).$$

4. Pick a different nonempty $V'' \in V(M + 1)$ and go to step 2.

\(^{10}\)The original theorem in [12] is for the case of $\ell_2$ geometry but the result for any metric.
5. At the end, let \( \phi_n^{\text{hier}}(y, u) = \arg \max_{V' \in V(M+1)} r(\text{conv}(V'), \ell_1) \). Then:

\[
\inf_n \inf_{u \in U_n} E_n(u, \phi_n^{\text{hier}}, M, \ell_1) = 2\delta - \inf_n \inf_{u \in U_n} \max_{V' \in V(M+1)} r(\text{conv}(V'), \ell_1)
\]

where the infimums refer to the uncertainty set.

Now, we present a simplification of the above algorithm that provides a lower bound to the optimal worst-case error of this hierarchical algorithm.

If in Fact 2, instead of \( \rho_{P_i}(x) \) we used \( h_i(P, \ell_1) \) we would have an upper bound to the inradius since \( \rho_{P_i}(x) \leq h_i(P, \ell_1) \) for all \( x \in P \). Then we define the following approximating algorithm:

**Algorithm 2**

1. Pick a \( V' \in V(M+1) \).

2. Calculate an upper bound to \( r(\text{conv}(V'), \ell_1) \) by solving the following LP:

\[
\min \sum_{i \in I(V')} \lambda_i h_i(\text{conv}(V'), \ell_1)
\]

subject to

\[
\sum_{i \in I(V')} \lambda_i = 1
\]

\[
\lambda_i \geq 0
\]

We still have to solve \( M + 1 \) LPs of the form \( \min_{y \in \text{conv}(\mathcal{V}, \ell_0)} \|V_i - y\|_1 \) in order to get \( h_i(\text{conv}(V'), \ell_1) \) for each \( i \in I(V') \) but that is only done once, rather than for every iteration of the master program in the original algorithm. So, this algorithm is considerably easier. Let \( \lambda_i^{\text{appr}} \) be the solution of the the master program in the approximating algorithm. Let also \( e^{\text{appr}}(V') = \sum_{i \in I(V')} \lambda_i^{\text{appr}} V_i \).

3. The approximation tells us that

\[
r(\text{conv}(V'), \ell_1) \leq \sum_{i \in I(V')} \lambda_i^{\text{appr}} h_i(\text{conv}(V'), \ell_1).
\]

4. Pick a different \( V'' \in V(M + 1) \) and go to step 2.
5. At the end, let $\hat{V} = \arg \max_{V' \in V(M+1)} e^\text{opt}(V')$ and $\hat{\lambda}_i$ the corresponding solution of the LP. Then define

$$B = \sum_{i \in I(V')} \hat{\lambda}_i \rho_{V',i} \left( \sum_{i \in I(V')} \hat{\lambda}_i V_i \right)$$

by solving the $M + 1$ LPs to calculate $\rho_{V',i}$. Finally we have

$$\inf_{n} \inf_{u \in U_n} E_n(u, \phi_n^{\text{hier}}, M, \ell_1) \geq 2\delta - B.$$ 

On a final note we quote a fact from convex geometry that provides an upper bound to the optimal worst-case error by any algorithm.

**Fact 3** Let $C$ be a subset of $\mathbb{R}^N$. Then\textsuperscript{11}:

$$\text{rad}(C, \ell_p) \leq \frac{N}{N + 1} \text{diam}(C, \ell_p).$$

In the $\ell_1$ case, this fact implies that there is an algorithm whose worst-case error is less than or equal to $\frac{2N}{N+1}$. Certainly the $\ell_1$ Chebyshev center of the worst-case uncertainty set provides such an algorithm.

In this section we have discussed the case of general worst-case identification in $\ell_1$, when the model set isn't necessarily balanced. We studied some geometric inequalities about polyhedra in $\mathbb{R}^N$ that led to the development of an identification algorithm. Finally, we estimated upper and lower bounds on the performance of this algorithm.

\textsuperscript{11}see [23]
Chapter 4

On-Line Algorithms and Implementation Issues

4.1 Introduction

In the previous chapter, we presented three off-line algorithms for worst-case identification, for model sets that are subsets of $M_N$. In particular, we saw

1. the interpolatory linear programming algorithm which constructs a plant in the uncertainty set,
2. the balanced set algorithm which essentially approximates the center of symmetry of the uncertainty set\(^1\) and
3. the two-stage hierarchical algorithm which approximates the incenter of the uncertainty set.

We have also seen various bounds on the worst-case performance of these algorithms.

In this chapter, we concentrate on the implementation of the above algorithms. First we seek ways to transform the off-line algorithms into on-line variants and then we discuss the number of necessary computations and the memory required for these algorithms.

\(^1\)I general, the uncertainty sets aren't center-symmetric.
4.2 On-Line Interpolatory Algorithm

As we discussed in section 3.2, the interpolatory linear programming algorithm can be implemented using the revised dual simplex in a manner so that the number of operations per iteration is independent of the number of observations. In other words, we could run the revised dual simplex on the uncertainty set, say after \( n \) observations, obtain an estimate, incorporate the new observation and update the old tableau to get a new estimate, without changing the underlying structure.

Algorithm 3 The on-line version of the interpolatory linear programming algorithm is:

1. Start having incorporated the first, say \( n \), measurements. Calculate \( \phi_I^n(y,u) \).

2. Incorporate the next measurement \( (y_{n+1}) \). If \( \left| \left( u^{(n+1)} \right)^T \phi_I^n(y,u) - y_{n+1} \right| < \delta \), then incorporate the next measurement \( (y_{n+2}) \), maintaining the same estimate. Else goto 3.

3. Perform simplex pivots on the dual tableau until an optimal dual basis has been found. Obtain the corresponding primal optimal solution, call that \( \phi_{n+1}(y,u) \) and goto 2.

At worst, we will have to solve \( n \) dual linear programs, each with a revised simplex CARRY tableau of size \( (N+1) \times (N+1) \).

4.3 Balanced Set Algorithm

In section 3.3 we presented the balanced-set algorithm. Actually, this algorithm is similar to the interpolatory linear programming algorithm. It is more sophisticated in that it attempts to use the slack variables in order to bring the estimate in the interior of the uncertainty set.

As was the case for the interpolatory LP algorithm, the balanced set algorithm can also be implemented using the revised simplex on the dual problem. Such an implementation requires a constant size CARRY tableau. More specifically, we formulate
the dual to the maximization problem at the heart of the balanced set algorithm:

\[
\min_{\tilde{B}\pi = c, \pi \geq 0} \pi^T b
\]

where

\[
\tilde{B} = \begin{pmatrix}
-u(1) & u^{(1)} & \cdots \\
y_1 & -y_1 & \cdots
\end{pmatrix}, \quad c = \begin{pmatrix}
0 \\
\vdots \\
0 \\
1
\end{pmatrix}, \quad b = \begin{pmatrix}
y_1 + \delta \\
-y_1 + \delta \\
\vdots
\end{pmatrix}
\]

\[
\tilde{B} \in \mathbb{R}^{(N+1)\times 2n}, \quad \pi \in \mathbb{R}^{2n}, \quad c \in \mathbb{R}^{N+1}, \quad b \in \mathbb{R}^{2n}
\]

The dual simplex needs to be initialized. However, because of the special structure of \(c\), we only need to add one auxiliary variable for Phase I of the dual simplex:

\[
\min_{w \geq 0} w
\]

subject to

\[
\begin{pmatrix}
-u^{(1)} & u^{(1)} & \cdots & 0 \\
y_1 & -y_1 & \cdots & 1
\end{pmatrix} \begin{pmatrix}
\pi \\
w
\end{pmatrix} = \begin{pmatrix}
0 \\
0 \\
1
\end{pmatrix}
\]

We can initiate the above LP with \(\pi = (0, \ldots, 0)^T \in \mathbb{R}^{2n}\) and \(w = 1\). We can then use the revised simplex and solve the LP using a CARRY tableau of constant size equal to \((N+1) \times (N+1)\). If \(w_{\text{min}} > 0\), then the original dual problem is infeasible. This implies that the primal maximization problem at the heart of the balanced set algorithm is unbounded. In other words, we would then know that \(\Lambda_{\text{max}} = +\infty\) and, according to Theorem 5,

**Proposition 7** If, after \(n^*(N)\) measurements have been incorporated, \(w_{\text{min}} > 0\), then

\[
E_{n^*(N)}(u, \phi_{n^*(N)}^{\text{bal}}, \mathcal{M}, \ell_1) = \delta
\]

Of course, if \(w_{\text{min}} = 0\), then the corresponding \(\pi\) vector is feasible for the original
dual problem. We can then begin the revised dual simplex whose CARRY tableau is still \((N + 1) \times (N + 1)\). This original dual problem cannot be unbounded because the primal maximization problem is always feasible\(^2\). Therefore, when the revised dual simplex terminates, the optimal dual cost is equal to the optimal primal cost according to the strong duality theorem, that is \((\pi^T b)_{\text{min}} = \Lambda_{\text{max}}\). Using \(\pi_{\text{min}}\) and \(\tilde{B}_{\text{basis}}\), we duplicate the argument made in section 2 of Chapter 3 and conclude that a primal optimal bfs is given by:

\[
\tilde{x} = \begin{pmatrix}
-\tilde{B}_{\text{basis}}^T b(B) \\
0 \\
\vdots \\
0 \\
\Lambda_{\text{max}}
\end{pmatrix}
\]

**Algorithm 4** The on-line version of the balanced set algorithm is:

1. Start having incorporated the first, say \(n\), measurements. Run Phase I for the dual problem. If \(w_{\text{min}} > 0\), we know that \(\Lambda_{\text{max}} = \infty\). If this happens when \(n \geq n^*(N)\), then Proposition 7 tells us that the error of the balanced set algorithm is exactly equal to \(\delta\).

2. Assuming that at termination of Phase I for the dual problem we obtain \(w_{\text{min}} = 0\), the top \(2n\) components of the terminal solution of Phase I constitute a feasible solution for the original dual problem. We iterate the revised dual simplex and obtain the current \(\Lambda_{\text{max}}(n)\).

3. As described above, construct a \(\tilde{x}\) feasible for the primal maximization problem.

4. Using the definition in section 3.3, obtain \(\phi_{\text{bal}}(y, u)\).

\(^2\)\(\Lambda = 0\) gives us the original uncertainty set which we know contains at least the true plant.
5. Incorporate the next measurement \((y_{n+1})\). If

\[
\begin{pmatrix}
  u^{(1)^T} & -y_1 \\
  -u^{(1)^T} & y_1 \\
  \vdots & \vdots \\
  u^{(n+1)^T} & -y_{n+1} \\
  -u^{(n+1)^T} & y_{n+1}
\end{pmatrix}
\begin{pmatrix}
  \varphi_{n}^{bal}(y, u) \\
  \Lambda_{\text{max}}(n)
\end{pmatrix} \leq
\begin{pmatrix}
  y_1 + \delta \\
  -y_1 + \delta \\
  \vdots \\
  y_{n+1} + \delta \\
  -y_{n+1} + \delta
\end{pmatrix}
\]

then incorporate the next measurement \((y_{n+1})\) maintaining the same estimate.

Else goto 1.

Furthermore, the lower bound in Theorem 5 of section 3.3, which is equal to half the diameter of the uncertainty set after \(n\) measurements, is calculated using a linear program. As before, we construct the dual to that problem

\[
\max(\delta \varepsilon_{2n}^T, 0\varepsilon_{2n}^T) \pi
\]

subject to

\[
\begin{pmatrix}
  u^{(1)} & -u^{(1)} & \cdots & u^{(n)} & -u^{(n)} \\
  1 & -1 & 1 & \cdots & \\
  1 & 1 & -1 & \cdots & \\
  \vdots & \vdots & \vdots & \ddots & \\
  1 & 1 & 1 & \cdots & \\
  0 & 0 & \cdots & 0 & 0 & 1 & 1 & \cdots
\end{pmatrix}
\]

\[
\pi =
\begin{pmatrix}
  0 \\
  \vdots \\
  0 \\
  1
\end{pmatrix}
\]

### 4.4 Two-Stage Hierarchical Algorithm

In section 3.4 we developed the two-stage hierarchical algorithm for the case of non-balanced model sets. As we mentioned before, that algorithm is based on trying to estimate the incenter of the uncertainty sets. In particular, if after the \(n\)th measurement \(\text{rank}(u^{(1)}, \ldots, u^{(n)}) = M \leq N\), the two-stage hierarchical algorithm picks sets of \(M + 1\) vertices of \(S_n(M, u, y, \delta)\). Then, for any point in the convex hull of the chosen vertices, each one of the \(M + 1\) subproblems calculates the distance of that point to
one of the $M+1$ faces of the convex hull of the chosen vertices. Using these distances, the master level iterates to find the inradius of the convex hull of the chosen set of vertices.

Each subproblem minimizes some $\ell_1$ distance over a convex set which is a linear program. The master problem is also a linear program as argued in section 3.4. This algorithm can be converted to an on-line variant in a straightforward manner.

**Algorithm 5** The on-line version of the two-stage hierarchical algorithm is:

1. Start having incorporated the first, say $n$, measurements. Let $M(n) = \text{rank}(u^{(1)}, \ldots, u^{(n)})$. Pick a $V' \in V(M(n) + 1)$ and run the two-stage hierarchical algorithm. Define

$$\phi_{n,V'}^{\text{hier}}(y, u) = c(V')$$

where $c(V')$ was defined in step 2 of Algorithm 1.

2. Incorporate the next measurement $(y_{n+1})$. If

$$S_n(\mathcal{M}, u, y, \delta) \cap (C^+(n+1) \cup C^-(n+1)) = \emptyset$$

where

$$C^+(n+1) = \{ f \in \mathcal{M} | f^T u^{(n+1)} = y_{n+1} \}$$

$$C^-(n+1) = \{ f \in \mathcal{M} | f^T u^{(n+1)} = -y_{n+1} \}$$

then incorporate the next measurement $(y_{n+2})$ maintaining the same estimate. Else goto 3.

3. Calculate $M(n+1)$. If $M(n+1) = M(n)$ then goto 4. Else goto 5.

4. Pick a $w \in S_n(\mathcal{M}, u, y, \delta) \cap (C^+(n+1) \cup C^-(n+1))$ and $v \in V'$. Substitute $v$ with $w$ to create the new set of vertices to optimize over: $V'_{\text{new}} = V' \setminus \{v\} \cup \{w\}$. Run the two-stage hierarchical algorithm with $V'_{\text{new}}$ to obtain the new estimate:

$$\phi_{n+1,V'_{\text{new}}}^{\text{hier}}(y, u) = c(V'_{\text{new}})$$
Goto 2.

5. Pick a \( w \in S_n(M, u, y, \delta) \cap (C^+(n+1) \cup C^-(n+1)) \). Run the two-stage hierarchical algorithm on \( V'_{\text{new}} = V' \cup \{w\} \) to obtain the new estimate:

\[
\phi_{n+1, V'_{\text{new}}}^\text{hier} (y, u) = c(V'_{\text{new}})
\]

Goto 2.

The approximating algorithm can be run in parallel with this as follows: in step 1 above we also calculate \( c^{\text{app}}(V') \) and in steps 4 and 5 we calculate \( c^{\text{app}}(V'_{\text{new}}) \).
Chapter 5

Conclusion

5.1 Summary of Results

This thesis has studied issues of complexity and algorithms in the context of worst-case identification mainly using the \( \ell_1 \) error norm. First of all, we presented two results regarding the sample complexity of worst-case identification of FIR systems under the \( \ell_1 \) error metric. In particular, it was shown that, optimal worst-case identification as well as suboptimal identification, to a constant multiplicative factor, require experiments exponentially long in the length of the FIR. These results were also generalized to the \( \mathcal{M}(M, \rho) \) model set.

Moreover, an upper bound was obtained for the sample complexity of worst-case identification in \( \mathcal{H}_\infty \) using the two-stage, nonlinear algorithm from [1]. Although in general this upper bound is exponential in the effective length of the plant to be identified, for a fixed time domain bound on the overshoot of the impulse response, the upper bound is polynomial. This illustrates an unexpected difference between the fundamental difficulty of worst-case identification in time domain using the \( \ell_1 \) norm and in frequency domain using the \( \mathcal{H}_\infty \) norm.

At the end of Chapter 2 we analyzed the performance of a recursive least-squares algorithm for worst-case identification in \( \ell_2 \). We saw that its asymptotic worst-case plant error could not be guaranteed to be constant. So, that algorithm was judged unsuitable for worst-case identification.
Further, three algorithms are presented for worst-case identification, all implemented with the use of linear programs. The simplest one is the Interpolatory Linear Programming Algorithm which guarantees estimates in the uncertainty set. Then, for the case of balanced model sets we developed the Balanced Set Algorithm, which estimates the center of symmetry of the uncertainty set using a sequence of linear programs. For the general case in which the model sets are not balanced we have the Two-Stage Hierarchical Algorithm which uses two-stage linear programming to approximate the incenter of the uncertainty set. Finally, these algorithms are transformed to on-line variants, which can be implemented while the input-output data are being gathered.

5.2 Future Research

This thesis probably raises more questions than it answers. We now know that worst-case identification in \( \ell_1 \) requires exponentially long experiments. This of course raises the question of relaxing the worst-case setup. In particular, an interesting new setup involves plant errors for "most" rather than all possible disturbance sequences. In order to quantify this, one has to put a probability measure on the space of disturbance sequences and obtain bounds on the probability that the plant error exceed a certain value.

Furthermore, no general results are available for worst-case identification in any other induced norms. In particular, the information theoretic technique used in this thesis for the development of the lower bound in Theorem 1 doesn't seem readily extendible to the \( \mathcal{H}_\infty \) error metric. On the other hand, the unexpected polynomial upper bound that was shown for the case of fixed \( M \) raises the question of what are the fundamental differences between the \( \ell_1 \) and \( \mathcal{H}_\infty \) norm and between time and frequency domain experiments.

On the algorithmic side, a lot will be gained by understanding and utilizing the geometry of uncertainty sets more thoroughly. Many results that exist for polyhedra in the Euclidean metric need to find their counterparts in the \( \ell_1 \) topology. Specifically,
the central question of finding the Chebyshev center of a convex set in the $\ell_1$ topology remains mostly unanswered. We believe it is research in this area that can improve the existing algorithms for worst-case identification.
Bibliography


