

Asymptotic Worst-Case Identification with Bounded Noise

Munther A. Dahleh*

Lab. for Information & Decision Systems, M.I.T., Cambridge MA 02139

1 Abstract

This paper presents an overview of the problem of asymptotic worst-case identification in the presence of bounded noise.

2 Introduction

Recently, there has been an increasing interest among the control community in the problem of identifying plants for control purposes. This generally means that the identified model should approximate the plant as it operates on a rich class of signals, namely signals with bounded norm, since this allows for the immediate use of robust control tools for designing controllers [2, 5]. This problem is of special importance when the data are corrupted with bounded noise. The case where the objective is to optimize prediction for a fixed input was analyzed by many researchers in [6, 15, 17, 18, 19, 22]. The problem is more interesting when the objective is to approximate the original system as an operator, a problem extensively discussed in [31]. For linear time invariant plants, such approximation can be achieved by uniformly approximating the frequency response (in the \mathcal{H}_∞ -norm) or the impulse response (in the ℓ_1 norm). In \mathcal{H}_∞ identification, it was shown that robustly convergent algorithms can be furnished, when the available data is in the form of a corrupted frequency response, at a set of points dense on the unit circle [10, 11, 12, 8, 9]. When the topology is induced by the ℓ_1 norm, a complete study of asymptotic identification was given in [27, 28, 30] for arbitrary inputs, and the question of optimal input design was addressed. Related work on this problem was also reported in [7, 13, 14, 16, 20, 21].

The work of Tse *et al* [27, 28, 30] and [3] allows for the analysis of large classes of systems including nonlinear fading memory systems. The study is

* Research Supported by AFOSR under grant AFOSR-91-0368 and by NSF under grant 9157306-ECS.

done in two steps. The first step is concerned with obtaining tight upper and lower bounds on the optimal achievable error, for a given fixed experiment. The second step is then to study these bounds and characterize the inputs that will minimize them. The upper and lower bounds are characterized under some mild topological assumptions on the model set through the diameter of the worst-case uncertainty set, a concept borrowed from Information Based Complexity [25, 26]. This characterization is valid for any input, and allows for closed loop identification. The papers include a detailed analysis of this diameter for different classes of systems, including LTI stable systems, unstable systems, and nonlinear time-invariant fading memory systems. Conditions on the input set are derived to guarantee finite (optimal) worst-case errors.

Another issue of importance in the context of worst-case identification is *Complexity*. It turns out that it is generally much harder to devise experiments that can guarantee small worst-case errors in the presence of bounded noise. This problem has been extensively analyzed in [4, 23].

It is important to caution at this point that the terminology “worst-case” does not mean that one can furnish guarantees on the worst-case error. Clearly, any result we obtain is a function of prior assumptions (which are not verifiable in general), and thus the results hold only when these assumptions are valid. This is no different from the traditional stochastic approach for system identification.

In this report, we will give an overview of the results we have developed in the context of this problem. Although we will not present proofs, the material is presented in a unified fashion. The objective is to study:

1. The fundamental limitations of worst-case identification.
2. The impact this has on input design.
3. The impact this has on Algorithms.

Details of these results can be found in [3, 4, 27].

3 Problem set-up

There are three basic ingredients for any identification problem. These are

1. **Prior knowledge and assumptions.** This includes assumptions/knowledge of the model set, disturbances, noise, etc.
2. **Data.** This includes specifications on input design, knowledge of open loop or closed loop data, and so on.
3. **Error criterion.** This dictates the sense in which the identified model approximates the actual process.

Worst case identification in the presence of bounded noise refers to a problem with specific ingredients. These are discussed below.

3.1 Prior Knowledge

Let \mathcal{X} be the class of all causal, single-input single-output, time-invariant, discrete-time systems. Let $\mathcal{M} \subset \mathcal{X}$ be the *model set* which is assumed to contain the unknown plant h to be identified. The set \mathcal{M} captures the experimenter's *a priori* knowledge about h . Some examples of \mathcal{M} are the set of all stable linear time-invariant systems, the set of stable systems with a bound on the decay rate of the impulse response, the set of all finite-dimensional systems with a bound on the order, space of stable nonlinear systems with fading memory, etc.

The process generating the outputs is assumed to have the form:

$$y = h(u) + d$$

where u , y are the inputs and the outputs of the process respectively, and d is the disturbance. The disturbance is assumed to be unknown but bounded:

$$\|d\|_\infty = \sup_k |d(k)| \leq \delta$$

In this setting, there are no statistical assumptions made on the noise. Such assumptions have dominated the traditional approaches in system identification. Deterministic assumptions are not new in the estimation/identification literature, for instance see [24].

3.2 Data

For any experiment, only a finite set of data is available:

$$[(y_k, u_k) \mid k = 0, 1, \dots, n]$$

where $u \in \mathcal{U}$: the set of all input sequences that can be used in the identification experiments. Typically, \mathcal{U} is a norm-bounded set, to reflect physical limitations, power restrictions, safety, or to maintain the validity of the linear model of the plant. An *experiment* is conducted by either choosing or measuring an input sequence $u \in \mathcal{U}$ and measuring the output sequence y , related to u by

$$y = h(u) + d \tag{1}$$

for a length of time n . If a number of experiments are performed, we use a vector of inputs \mathbf{u} and outputs \mathbf{y} .

3.3 Error Criterion

An *identification algorithm* is a mapping ϕ which generates, at each time instant n , an estimate

$$\hat{h}^{(n)} \equiv \phi(P_n \mathbf{u}, P_n \mathbf{y}) \in \mathcal{X}$$

given the finite sequence of inputs and outputs. Here, P_n is the truncation operator, defined by $P_n \mathbf{x} = (x_0, x_1, \dots, x_n)$ for each infinite sequence \mathbf{x} . Its use signifies that the algorithm ϕ generates at each time instant an estimate based only on the input-output data it has seen so far. Generally, we will assume that the algorithm has access to what the model set \mathcal{M} is and also the value of δ , the bound on the disturbance. In the terminology of Helmicki et. al. [12], the algorithm is *tuned*. However, in some cases, we will be able to give stronger results using algorithms which are *untuned* to the value of δ .

The error criterion is defined through a metric $\rho_{\mathcal{X}}$. This metric can be chosen to be the induced operator norm over the spaces ℓ_{∞} , ℓ_2 and so on. We would like to evaluate the error $\rho_{\mathcal{X}}(h, \hat{h}_n)$. Of course this is difficult to evaluate since h is not known. The terminology *worst-case identification* is in fact motivated from looking at the maximum value this error criterion takes over all plants in the model set.

To explain the error criterion, consider the set of behavior \mathcal{L} :

$$\mathcal{L} = \{y = h(u) + d \mid d \in \|d\|_{\infty} \leq 1\}$$

So, for any plant $h \in \mathcal{M}$, $u \in \mathcal{U}$, this set consists of all possible outputs that can be generated by disturbances with $\|d\|_{\infty} \leq \delta$. The estimates based on these outputs can be different. The worst-case error criterion simply says that the distance between the estimate taken after a long experiment and the plant h , for any possible output in the behavior set, and for any $h \in \mathcal{M}$ should be minimized. The error is defined as:

$$e_{\infty}(\phi, \mathcal{M}, \mathbf{u}, \delta) \equiv \sup_{h \in \mathcal{M}} \sup_{\|d\|_{\infty} \leq \delta} \limsup_{n \rightarrow \infty} \rho_{\mathcal{X}}(\phi(P_n \mathbf{u}, P_n(\mathbf{u} * h + d)), h)$$

In the above definition of the worst-case asymptotic error, although convergence of the estimates to within $e_{\infty}(\phi, \mathcal{M}, \mathbf{u}, \delta)$ is guaranteed for all admissible plants and disturbance sequences, the *rate* of convergence may be arbitrarily slow for some plants and some disturbances. The worst-case asymptotic error is said to be *uniform* if the rate of convergence is uniform over all admissible plants and disturbance sequences. If the convergence is uniform, the worst-case asymptotic error defined above is the same as the limit of the worst-case error taken at each finite time n , i.e.

$$e_{\infty}(\phi, \mathcal{M}, \mathbf{u}, \delta) = \limsup_{n \rightarrow \infty} \sup_{h \in \mathcal{M}} \sup_{\|d\|_{\infty} \leq \delta} \rho_{\mathcal{X}}(\phi(P_n \mathbf{u}, P_n(\mathbf{u} * h + d)), h)$$

This allows one to *a priori* determine the experiment length required to guarantee that *any* plant in the model set can be identified to a prescribed accuracy. It is the notion of convergence considered by Helmicki et. al. in their framework [12].

Demanding uniform convergence is too restrictive a formulation for a general theory of fundamental limitations of worst-case identification. Although such uniform convergence is certainly desirable, it is impossible to achieve for many interesting model sets. In fact, for many inherently infinite-dimensional model sets, the worst-case error at each finite time is always infinite, while the worst-case asymptotic error can be made small using an appropriate identification algorithm and inputs. Our formulation thus allows us to discuss optimal worst-case identification and optimal inputs for a much broader class of model sets. Besides, in some applications of identification, such as adaptive control, uniform convergence of estimates is not necessary to fulfill the desired objectives. However, because of the special importance of uniform convergence, we will give additional conditions on the model set for this to take place. It will be seen that these conditions are quite strong and essentially require the model set to be finite-dimensional. It is worthwhile to note that the model set considered in [9, 10] satisfies these conditions.

The optimal worst-case asymptotic error $E_\infty(\mathbf{u}, \mathcal{M}, \delta)$ is defined as the smallest error achievable by any algorithm:

$$E_\infty(\mathbf{u}, \mathcal{M}, \delta) \equiv \inf_{\phi} e_\infty(\phi, \mathcal{M}, \mathbf{u}, \delta)$$

4 General Results

Given the above set-up of the worst-case identification problem, there are general results that can be derived to characterize the optimal error $E_\infty(\mathbf{u}, \mathcal{M}, \delta)$. The following presentation is a summary of results reported in [3, 4, 27].

An important concept in such a characterization is the concept of an **Uncertainty Set**. Simply, the uncertainty set at time n is the set of all plants consistent with the data, i.e.,

$$S_n(\mathcal{M}, \mathbf{u}, \mathbf{y}, \delta) = \{g \in \mathcal{M} : \|P_n(g(\mathbf{u}) - \mathbf{y})\|_\infty \leq \delta\}$$

and the infinite-horizon uncertainty set is the set of all plants consistent with the infinite-horizon input and output, i.e.,

$$S_\infty(\mathcal{M}, \mathbf{u}, \mathbf{y}, \delta) = \{g \in \mathcal{M} : \|g(\mathbf{u}) - \mathbf{y}\|_\infty \leq \delta\}$$

4.1 Lower Bounds on $E_\infty(\mathbf{u}, \mathcal{M}, \delta)$

For a given input, and a plant in the set \mathcal{M} , any output in the set of behaviors \mathcal{L} is a possible output, which in turn results in a specific sequence of uncertainty sets. At time n , all the plants in S_n are indistinguishable, and

any estimator will result in an error which is closely related to the diameter of this set. Finally, since the actual process is unknown, we need to look at the maximum diameter generated for different choices of $h \in \mathcal{M}$. This motivates the next definitions.

For any set $A \subset \mathcal{X}$, define the diameter of the set A as

$$\text{diam}(A) = \sup_{g, h \in A} \rho_{\mathcal{X}}(g, h)$$

Next, we define a quantity known as the **Diameter of Information**.

Definition 1 *Given a choice of the inputs \mathbf{u} , define the infinite-horizon diameter of information $D(\mathbf{u}, \mathcal{M}, \delta)$ the diameter of the largest possible uncertainty set :*

$$D(\mathbf{u}, \mathcal{M}, \delta) \equiv \sup_{h \in \mathcal{M}} \sup_{\|\mathbf{d}\|_{\infty} \leq \delta} \text{diam}(S_{\infty}(\mathcal{M}, \mathbf{u}, \mathbf{u} * h + \mathbf{d}, \delta))$$

In information-based complexity terminology, these quantities correspond to the *diameter of information* for the infinite horizon problem where the information available is the entire infinite output sequence. The quantity $D(\mathbf{u}, \mathcal{M}, \delta)$ is the largest distance between two plants for which there are admissible disturbances such that the plants give exactly the same outputs. It turns out that it is precisely this quantity that characterizes the optimal worst-case asymptotic errors. First we show that half the infinite-horizon diameter of information is a lower bound to the optimal asymptotic error.

Proposition 2 *Let \mathcal{M} be any model set, \mathbf{u} be any vector of inputs and $\delta \geq 0$. Then*

$$e_{\infty}(\phi, \mathcal{M}, \mathbf{u}, \delta) \geq D(\mathbf{u}, \mathcal{M}, \delta)/2$$

for any algorithm ϕ .

4.2 Upper Bound

Under some mild assumptions on the model set, the *Diameter of Information* is an upper bound on the best achievable error. Such assumptions are related to σ -compactness of the model set. A model set \mathcal{M} is σ -compact if it is a countable union of compact, nested subsets, i.e., $\mathcal{M} = \cup_i \mathcal{M}_i$. A model set is separable if it is the closure of a σ -compact set. The following results gives the upper bound. In here $\rho_{\mathcal{X}}$ indicates some induced norm.

Theorem 3. *Suppose that the model set \mathcal{M} is σ -compact, or separable in the $\rho_{\mathcal{X}}$ topology. there is an identification algorithm ϕ^* such that*

$$e_{\infty}(\phi^*, \mathcal{M}, \mathbf{u}, \delta) \leq D(\mathbf{u}, \mathcal{M}, \delta)$$

for all \mathbf{u} and $\delta \geq 0$.

This theorem can be interpreted as follows: For model sets with such topological structure, and for any input, there exists an algorithm such that the asymptotic worst-case error is bounded by the diameter of information. Such an algorithm is based on the Occam's Razor principle: Pick the estimate in the smallest set consistent with the data. This algorithm requires the knowledge of δ , and thus it is a *tuned* algorithm.

It should be noted that by an elementary result in information-based complexity theory, the optimal worst-case error achievable when the algorithm has *full* access to the entire *infinite* input-output sequences is also bounded between the infinite-horizon diameter of information and half the diameter of information. Our two results (Proposition 2 and Theorem 3) are of an entirely different nature: they assert that the optimal worst-case asymptotic error achievable when the algorithm has access to *finite* but arbitrarily long data records also satisfies the same bounds. The assumed topological conditions are crucial for the validity of Theorem 3.

4.3 Input Design

From the above discussion, the inputs should be designed to minimize the diameter of information, i.e.,

$$\inf_{\mathbf{u} \in \mathcal{U}} D(\mathbf{u}, \mathcal{M}, \delta)$$

If this quantity is infinite, then accurate worst-case identification is not possible. For certain model sets, such a minimization can be performed and optimal inputs can be furnished. In the sequel, we will present a few such examples.

5 Application to Specific model Sets

The above general results, can be applied to specific model sets, to derive optimal inputs that guarantee accurate identification.

5.1 Stable LTI Systems

Here \mathcal{X} is the space ℓ_1 . This is the space of BIBO linear time-invariant, causal operators on ℓ_∞ . The metric $\rho_{\mathcal{X}}$ can be either the ℓ_1 norm, or the \mathcal{H}_∞ . The set \mathcal{M} will be any balanced (i.e. if $h \in \mathcal{M}$ then $-h \in \mathcal{M}$) and convex closed subset of \mathcal{X} (with diameter larger than 2δ).

Since the space ℓ_1 is separable (with respect to both the ℓ_1 topology and \mathcal{H}_∞ topology), then there is an identification algorithm ϕ^* such that

$$\frac{D(\mathbf{u}, \mathcal{M}, \delta)}{2} \leq e_\infty(\phi^*, \mathcal{M}, \mathbf{u}, \delta) \leq D(\mathbf{u}, \mathcal{M}, \delta)$$

for all \mathbf{u} and $\delta \geq 0$.

To estimate the diameter of information, we first notice that

$$D(\mathbf{u}, \mathcal{M}, \delta) \geq 2\delta$$

In fact, one can show that there exists an input u^* , such that the diameter of information is equal to 2δ .

Theorem 4. *Assume \mathcal{M} is balanced and convex and contains only stable plants. If u^* contains all finite sequences of 1's and -1 's, then*

$$D(u^*, \mathcal{M}, \delta) \leq 2\delta$$

In the above theorem, the diameter is computed with respect to either the l_1 norm or \mathcal{H}_∞ norm. The results are the same. This, combined with the earlier result, shows that there exists a single input u^* , such that

$$E_\infty(u^*, \mathcal{M}, \delta) \leq 2\delta$$

Hence, to identify a plant accurately in the limit, it is enough to know *a priori* that it is stable; no additional information, such as bounds on decay rate and gain, is necessary. The achievable accuracy varies continuously with the noise bound δ for small δ ; thus, identification can be performed robust to measurement noise.

Uniform Convergence

Next, we look at the issue of uniform convergence. For the model set l_1 , it can at once be seen that although convergence to a small asymptotic error is possible, such convergence cannot be uniform. To guarantee uniform convergence, we need to look at compact model sets.

Proposition 5 *Let $\mathcal{M} \subset l_1$ be a compact set (in the l_1 -topology) or a subset of a compact set in l_1 . For the single input u^* which contains all finite sequences of 1's and -1 's, there is an algorithm the estimates of which converge, uniformly for all $h \in \mathcal{M}$ and all $\|\mathbf{d}\|_\infty \leq \delta$, to an l_1 ball of radius 2δ around the true plant. Moreover the algorithm does not require the knowledge of the value of δ to compute its estimates.*

Common examples of such compact model sets are the uniformly stable ones, of the form $M_s(g) \equiv \{h : |h_i| \leq |g_i| \text{ for all } i\}$ where g is any stable plant. The specific model sets considered in [9] and [10] belong to this class. In the particular case when g_i is taken to be 0 for all i larger than some given M , we get the model set of finite-impulse- response of length M . For this model set the near-optimal algorithm ϕ^* is given by

$$\phi^*(P_n \mathbf{u}, P_n \mathbf{y}) = \arg \min_{|h_i| \leq |g_i|, i=0,1,\dots,M} \|P_n(\mathbf{y} - \mathbf{u} * h)\|_\infty$$

which is computable by linear programming.

Sample Complexity

The input designed to achieve a radius of information equal to 2δ is quite a rich input. This suggests that achieving accurate identification when only stability is assumed can be quite difficult in practice. To study this problem in a more precise fashion, let \mathcal{M}_N denote the subset of ℓ_1 that contains systems with finite impulse response of length N . Let \mathcal{U}_n be the set of all infinite real sequences $\{u_i\}_{i=1}^{\infty}$ such that $|u_i| \leq 1$ for all i , and $u_i = 0$ for $i > n$. Any element of \mathcal{U}_n will be called an *input of length n* . Let $D_{N,n}^*$ be the smallest diameter of information over the class of inputs in \mathcal{U}_n , i.e.,

$$D_{N,n}^* = \inf_{u \in \mathcal{U}_n} D(\mathcal{M}_N, u, \delta)$$

We have shown earlier that

$$\lim_{n \rightarrow \infty} D_{N,n}^* = 2\delta$$

The sample (time) complexity of such problems is captured in the length of n necessary to get within a factor of 2δ . The next result shows that this length is exponential in N , and thus can be quite unrealistic for large N .

Theorem 6. Fix some $K > 1$ and let

$$n^*(N, K) = \min\{n \mid D_{N,n}^* \leq 2K\delta\}$$

Then

1. $n^*(N, K) \geq 2^{Nf(1/K)-1} - N$
2. $\lim_{N \rightarrow \infty} \frac{1}{N} \log n^*(N, K) = f(1/K)$.

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

$$f(\alpha) = 1 + \left(\frac{1-\alpha}{2}\right) \log\left(\frac{1-\alpha}{2}\right) + \left(\frac{1+\alpha}{2}\right) \log\left(\frac{1+\alpha}{2}\right). \quad (2.11)$$

Notice that the function f defined by (2.11) satisfies $f(\alpha) = 1 - H((1-\alpha)/2)$, where H is the binary entropy function. In particular, f is positive and continuous for $\alpha \in (0, 1)$.

This theorem furnishes a lower bound on the minimum time required to get within a factor of the optimum solution, which is exponential in N . This lower bound is asymptotically tight.

5.2 Nonlinear Systems with Fading Memory

The set \mathcal{X} contains causal functions from \mathcal{U} to \mathfrak{R}^∞ ; these are discrete-time systems, possibly non-linear, which take as input a sequence in \mathcal{U} to give an output sequence in \mathfrak{R}^∞ . The input and the output at time n will be denoted by u_n and $h_n(u)$. Assume that they further satisfy the following properties

1. $h_n(u)$ depends continuously on u_0, \dots, u_{n-1} .
2. h has equilibrium-initial behavior:

$$h_{n+1}(0u) = h_n(u) \text{ for all } n$$

where $0u$ is the input $0, u_0, u_1, \dots$

The model sets \mathcal{M} we shall consider will be *balanced and convex* subsets of this class of functions \mathcal{X} . In general, we will use the notation vw for concatenation, i.e. first apply the finite sequence v , then w . Since we are dealing with causal systems, we shall slightly abuse the notation and write $h_n(w)$ to mean $h_n(u)$, where u is any infinite sequence the first n elements of which is the finite sequence w .

The metric $\rho_{\mathcal{X}}$ will be taken to be the operator-induced norm:

$$\|h\| = \sup_{u \in \mathcal{U}} \|h(u)\|_\infty$$

This is the natural norm to consider for robust control applications.

Definition of Fading Memory Systems

Definition 7 An operator h has fading memory (FM) if for each $\varepsilon > 0$ there is some $T = T(\varepsilon)$ such that: for every k , every $t \geq T$ and every finite sequences $v \in [-1, 1]^k$, $w \in [-1, 1]^t$,

$$\|h_{t+k}(vw) - h_t(w)\| < \varepsilon$$

It can easily be seen that fading memory systems satisfying properties (1) and (2) have bounded operator-induced norms.

Examples of FM Systems

Example 1: *stable LTI systems.*

For each $h \in \ell_1$ define the input/output map $u \mapsto u * h$ by convolution. It is clear that these systems satisfy the above conditions. The operator-induced norm in this case is just the ℓ_1 norm.

Example 2: *Hammerstein Systems.*

These are systems which are formed by composition of a LTI system followed by a memoryless nonlinear element:

$$y_n = g((u * h)_n)$$

for some $h \in \ell_1$ and some continuous function $g : \mathfrak{R} \rightarrow \mathfrak{R}$. It is easy to verify that these systems satisfy the first two conditions above. Since $|(h * u)_n| \leq \|h\|_1$ then g is uniformly continuous on $[-\|h\|_1, \|h\|_1]$. Hence h has fading memory. Any stable system with fading memory can be approximated arbitrarily closely by a Hammerstein system.

Proposition 8 *The class of all fading memory systems is separable.*

A σ -compact set is constructed as follows: for every $n \geq 0$, consider all Hammerstein systems, in which h is FIR of length n . The set of all stable systems with fading memory is the closure of the countable union of such systems.

This means that when we consider fading memory system, we can apply the general results, and reduce the analysis of asymptotic optimal error to the analysis of infinite-horizon diameter, i.e.,

$$\frac{D(\mathbf{u}, \mathcal{M}, \delta)}{2} \leq e_\infty(\phi^*, \mathcal{M}, \mathbf{u}, \delta) \leq D(\mathbf{u}, \mathcal{M}, \delta)$$

for all \mathbf{u} and $\delta \geq 0$.

It is evident that for any \mathcal{M} in \mathcal{X} , balanced and convex (with diameter larger than 2δ) satisfies:

$$D(u, \mathcal{M}, \delta) \geq 2\delta$$

The existence of an experiment that results in equality is established below.

Theorem 9. *Let the model set \mathcal{M} be some subset of the set of fading memory systems. Let W be any countable dense subset of $[-1, 1]$ and consider any input $u^* \in [-1, 1]^\infty$ which contains all possible finite sequences of elements of W . Then*

$$D(u^*, \mathcal{M}, \delta) \leq 2\delta$$

Complexity

Consider the class of p memory systems. These are systems that operate on the last p component of the input, and are given by continuous functions on $[-1, 1]^p$. Let g be such a function. It can be easily seen that in general, the time needed to identify a system to a prescribed accuracy grows exponentially as the order of the system, even when there is no noise. For example, if we assume a certain Lipschitz condition on the order p memory function g , such as $|g(x) - g(y)| < M\|x - y\|$, then to identify the function up to accuracy ϵ (in the $\|\cdot\|_\infty$ norm), the number of data points needed is at least the minimum number of ϵ -balls to cover $[-1, 1]^p$. Since the volume of an ϵ -ball is $O(\epsilon^p)$, it is clear that this minimum number is $\Omega((\frac{1}{\epsilon})^p)$, and hence so is the experiment length. This means that if p is large, the experiment length will be very long if we make no further assumption on the unknown plant.

It is interesting to compare this situation with the problem of identifying *linear* finite impulse response systems. For nonlinear systems the time complexity is exponential of the order, whether or not there is noise. For the linear case, while it takes only linear time to identify a FIR system *exactly* when there is no noise, we have shown that the time complexity immediately becomes exponential once we introduce any unknown but bounded noise. Moreover, it has been demonstrated that if we are willing to put a probability distribution on the noise, polynomial time complexity can often be obtained [29]. These facts show that while in the nonlinear case, the plant uncertainty determines the time complexity of the identification, in the linear case, the complexity is sensitive to how the noise is modeled.

5.3 Untuned algorithms

So far, all the algorithms devised to deliver accurate identification are *tuned*. As we discussed, such algorithms are based on Occam's Razor in which the simplest model that explains the data is picked. That is, if $\mathcal{M} = \cup_i \mathcal{M}_i$, where the \mathcal{M}_i 's are nested compact sets, and S_n is the uncertainty set at time n , then the estimate \hat{h}_n can be any element in the intersection of S_n and \mathcal{M}_j , where j is the smallest integer such that this intersection is not empty. To implement such an algorithm, the sets S_n need to be computed. This requires the knowledge of the bound on the disturbance, δ .

Notice that these results are derived for any choice of inputs in \mathcal{U} . It is possible, however, to derive *untuned* algorithms for specific experiments. In particular, this is possible when the input used has the property that it minimizes the diameter of information. Such results can be found in [20, 3] and will not be discussed here.

6 Summary

Finally, I would like to discuss some of the issues that this line of research has highlighted.

1. Choose the simplest estimate to explain the data. This is a consequence of the Occam's Razor principle [1]. This implies that the algorithm should have enough information to make such a decision, which typically translates into the knowledge of an upper bound on the magnitude of the noise. Estimates that fit the data very well generally fit the noise as well, and result in non-convergent estimates.
2. For nonlinear fading memory systems, we need local interpolants to fit the nonlinearities. This can be accomplished by basis functions such as "Linear Splines", or what is known as "Gaussian radial" functions, or any other local interpolant. Polynomial interpolants generally are not appropriate. This is due to the fact that if the interpolant in some neighborhood depends on far away points, noise can be amplified and the algorithm may not be convergent.

3. Inputs have to be quite rich. Of course this depends on the amount of prior information. This can have major impact on closed loop identification, where the inputs are not arbitrary. On the other hand, for fairly understood models (small model sets), worst-case identification can be quite easy, see for example [13].
4. It may be too "hard" in practice to give guarantees under such noise assumptions. This follows from the sample complexity results discussed earlier. These results capture the fundamental limitations of worst-case identification in the presence of bounded noise. Also, this suggests alternate formulations in which the error criterion is a worst-case criterion, however, noise is assumed to be stochastic. It is hinted in [23, 29] that experiments can be much shorter in such formulations.
5. More work is needed in the area of algorithms and particularly recursive algorithms.

References

1. A. Blumer, A. Ehrenfeucht, D. Haussler, M. Warmuth, "Occam's Razor". *Information Processing Letters* 24, pp.377-380, 1987.
2. M.A. Dahleh and M.H. Khammash, "Controller Design in the Presence of Structured Uncertainty," to appear in *Automatica* special issue on robust control.
3. M.A. Dahleh, E.D. Sontag, D.N. Tse and J.N. Tsitsiklis, "Worst-Case identification of nonlinear fading memory systems," to appear in *ACC Proc.*, 1992.
4. M.A. Dahleh, T. Theodosopoulos, and J.N. Tsitsiklis, "The sample complexity of worst-case identification of F.I.R. Linear systems." To appear in *Systems and Control Letters*.
5. J. C. Doyle, "Analysis of feedback systems with structured uncertainty," *IEEE Proceedings* 129, 242-250, 1982.
6. E. Fogel and Y. F. Huang, "On the value of information in system identification—bounded noise case". *Automatica*, vol.18, no.2, pp.229-238, 1982.
7. G.C. Goodwin, M. Gevers and B. Ninness, "Quantifying the error in estimated transfer functions with application to model order selection." *IEEE Trans. A-C*, Vol 37, No. 7, July 1992.
8. G. Gu, P.P. Khargonekar and Y. Li, "Robust convergence of two stage nonlinear algorithms for identification in \mathcal{H}_∞ ." *Systems and Control Letters*, Vol 18, No. 4, April 1992.
9. G. Gu and P.P. Khargonekar, "Linear and nonlinear algorithms for identification in \mathcal{H}_∞ with error bounds." *IEEE Trans. A-C*, Vol 37, No. 7, July 1992.
10. A.J. Helmicki, C.A. Jacobson and C.N. Nett, "Identification in H_∞ : A robust convergent nonlinear algorithm". Proceedings of the 1989 International Symposium on the Mathematical Theory of Networks and System, 1989.
11. A.J. Helmicki, C.A. Jacobson and C.N. Nett, "Identification in H_∞ : Linear Algorithms". Proceedings of the 1990 American Control Conference, pp 2418-2423.

12. A.J. Helmicki, C.A. Jacobson and C.N. Nett, "Control-oriented System Identification: A Worst-case/deterministic Approach in H_∞ ," *IEEE Trans. A-C*, Vol 36, No. 10, October 1991.
13. C.A. Jacobson and C.N. Nett, "Worst-case system identification in ℓ_1 : Optimal algorithms and error bounds," in *Proc. of the 1991 American Control Conference*, June 1991.
14. J.M. Krause, G. Stein, P.P. Khargonekar, "Robust Performance of Adaptive Controllers with General Uncertainty Structure", Proceedings of the 29th Conference on Decision and Control, pp. 3168-3175, 1990.
15. R. Lozano-Leal and R. Ortega, "Reformulation of the parameter identification problem for systems with bounded disturbances", *Automatica*, vol.23, no.2, pp.247-251, 1987.
16. M.K. Lau, R.L. Kosut, S. Boyd, "Parameter Set Estimation of Systems with Uncertain Nonparametric Dynamics and Disturbances", Proceedings of the 29th Conference on Decision and Control, pp. 3162-3167, 1990.
17. M. Milanese and G. Belforte, "Estimation theory and uncertainty intervals evaluation in the presence of unknown but bounded errors: Linear families of models and estimators", *IEEE Trans. Automatic Control*, AC-27, pp.408-414, 1982.
18. M. Milanese and R. Tempo, "Optimal algorithm theory for robust estimation and prediction", *IEEE Trans. Automatic Control*, AC-30, pp. 730-738, 1985.
19. M. Milanese, "Estimation theory and prediction in the presence of unknown and bounded uncertainty: a survey", in *Robustness in Identification and Control*, M. Milanese, R. Tempo, A. Vicino Eds. Plenum Press, 1989.
20. P.M. Makila, "Robust Identification and Galois Sequences", Technical Report 91-1, Process Control Laboratory, Swedish University of Abo, January, 1991.
21. P.M. Makila and J.R. Partington, "Robust Approximation and Identification in H_∞ ", Proc. 1991 American Control Conference, June, 1991.
22. J.P. Norton, "Identification and application of bounded-parameter models", *Automatica*, vol.23, no.4, pp.497-507, 1987.
23. K. Poolla and A. Tikku, "On the time complexity of worst-case system identification," preprint, 1992.
24. F.C. Schweppe, *Uncertain dynamic systems*, Prentice-Hall, Englewood cliffs, N.J., 1973.
25. J.F. Traub and H. Wozniakowski, *A General Theory of Optimal Algorithms*, Academic Press, New York, 1980.
26. J.F. Traub, G. Wasilkowski and H. Wozniakowski, *Information-Based Complexity*, Academic Press, 1988.
27. D. Tse, M.A. Dahleh and J.N. Tsitsiklis, Optimal Asymptotic Identification under bounded disturbances. To appear in *IEEE Trans. Automat. Contr.*.
28. D.N.C. Tse, M.A. Dahleh, J.N. Tsitsiklis, "Optimal and Robust Identification in the ℓ_1 norm", in *Proc. of the 1991 American Control Conference*, June 1991.
29. D.N.C. Tse and J.N. Tsitsiklis, "Sample complexity of system identification under a high probability formulation", in preparation.
30. D.N.C. Tse, "Optimal and robust identification under bounded disturbances", Master's thesis, Dept. of Elec. Eng. and Comp. Sci., M.I.T., February, 1991.
31. G. Zames, "On the metric complexity of casual linear systems: ϵ -entropy and ϵ -dimension for continuous-time", *IEEE Trans. on Automatic Control*, Vol. 24, April 1979.

This article was processed using the \LaTeX macro package with L^AM^EX style