1992 ACC/WA8

Worst-Case Identification of Nonlinear Fading Memory Systems

Munther A. Dahleh Eduardo D. Sontag* David N.C. Tse John N. Tsitsiklis
Laboratory for Information and Decision Systems
M.I.T.
Cambridge, MA 02139
and
*Department of Mathematics
SYCON-Rutgers Center for Systems and Control
Rutgers University
New Brunswick, N.J. 08903

Abstract

In this paper, the problem of asymptotic identification for a class of nonlinear fading memory systems in the presence of bounded noise is studied . For any experiment, the worst-case error is characterized in terms of the diameter of the worst-case uncertainty set. Optimal inputs that minimize the radius of uncertainty are studied and characterized. Finally, a convergent algorithm that does not require knowledge of the noise upper bound is furnished. The methods as well as the results are quite general and are applicable to a larger variety of settings.

1 Introduction

Recently, there has been increasing interest in the control community in the problem of identifying plants for control purposes. This generally means that the identified model should approximate the plant in the operator topology, since this allows the immediate use of robust control tools for designing controllers [2, 3]. This problem is of special importance when the data is corrupted with bounded noise. This problem has been extensively studied in the literature when the objective is prediction with a fixed input [7]. The problem is more interesting when the objective is to approximate the original system as an operator, a problem extensively discussed in [13]. For linear time invariant plants, such approximation can be achieved by uniformly approximating the frequency response (H_∞-norm) or the impulse reponse (ℓ_1 norm). In H_∞ identification, it was shown that robustly convergent algorithms can be furnished, when the available data is in the form of corrupted frequency response, on a set of points which is dense in the unit circle. [5, 4]. When the topology is induced by the ℓ_1 norm, a complete study of asymptotic identification was furnished in [12] for arbitrary inputs, and the question of optimal input design was addressed. Related work on this problem was also reported in [8, 9, 6].

In this paper, the work in [12] is extended to general settings, which allows for immediate analysis of larger classes of systems, namely, nonlinear fading memory systems. The study is 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 analysis given in this paper is quite general, and can be applied to many classes of systems. In particular, simple topological conditions are furnished that guarantee the existence of an algorithm with a worst-case error within a factor of two from the lower bound. In the particular case of Fading Memory systems, a *near* optimal input is chosen to minimize the bound.

The rest of the paper is organized as follows: Section 2 gives a general formulation, Section 3 presents a characterization of the optimal error in terms of uncertainty sets, as well as the basic consistency results, Section 3 contains the discussion of fading memory systems and Section 4 contains the characterization of optimal inputs, as well as consistency results for such inputs. Conclusions are in Section 5.

2 General Formulation

In this section, we will give a general formulation of the identification problem.

Authorized licensed use limited to: MIT Libraries. Downloaded on July 30,2020 at 15:12:12 UTC from IEEE Xplore. Restrictions apply.

Denote by $\mathcal{L}$ the set of all possible *observations* (or observed, noise-corrupted, behaviors). Typically $\mathcal{L}$ is a fixed subset of $\Re^{\infty}$. Let $\mathcal{M}$ denote the model set: a set of possible models to explain the observed behavior, representing the *a priori* knowledge ; ρ is a metric on the set $\mathcal{M}$ used to measure identification errors. Given an observed behavior $y \in \mathcal{L}$, define the sets $\{\mathcal{S}_n(y), n \in \mathbb{Z}_+, y \in \mathcal{L}\}$, where each $\mathcal{S}_n(y) \subseteq \mathcal{M}$ denoting all *possible* models "consistent with the behavior observed up to time n". This set represents the uncertainty in the model up to time n, and captures exactly the information available at that time. These sets define an *uncertainty structure* if they satisfy two conditions:

1. $\mathcal{S}_n(y) \subseteq \mathcal{S}_m(y)$ if $n \geq m$, $\forall\, y$.
2. Each $\mathcal{S}_n(y)$ is a closed set.

The infinite-horizon uncertainty set is given by:

$$\mathcal{S}_\infty(y) := \bigcap_{n \geq 0} \mathcal{S}_n(y) \tag{1}$$

Let an *estimator* be defined by a sequence of maps $\phi = \{\phi_n, n \geq 0\}$, each $\phi_n : \mathcal{L} \rightarrow \mathcal{S}_n(y) \subset \mathcal{M}$. An estimator is required to be *causal* in the following sense:

$$\forall\, n, y, y', \;\; \mathcal{S}_n(y) = \mathcal{S}_n(y') \;\Rightarrow\; \phi_n(y) = \phi_n(y')$$

Typically $\mathcal{S}_n(y)$ will depend only on the "restriction of y to interval $[0, n]$", so this will really mean that $\phi_n(y)$ depends only on past values.

Next, we will give precise definitions of the error functions. The largest possible asymptotic error if the true plant is h and the estimator is ϕ, is given by

$$e(\phi, h) := \sup_{\{y \mid h \in \mathcal{S}_\infty(y)\}} \limsup_{n \rightarrow \infty} \rho(\phi_n(y), h) \tag{2}$$

which is well defined for all estimators ϕ and all $h \in \mathcal{M}$. This is a pointwise error, which can be interpreted as follows: for a fixed model $h \in \mathcal{M}$, the estimate based on any observation (with which h is consistent), will eventually be within a radius of $e + \epsilon$ from h, for any $\epsilon > 0$.

The *worst case asymptotic error* is defined as

$$e(\phi) := \sup_{h \in \mathcal{M}} e(\phi, h) \tag{3}$$

This error can be interpreted as follows: for any model $h \in \mathcal{M}$, the estimate based on any observation (with which h is consistent), will eventually be within a radius of $e + \epsilon$ from h, for any small $\epsilon > 0$. The convergence rate may very well depend on the plant and noise, i.e. for a given ϵ there exists some $N(y, h, \epsilon)$ so that

$$\rho(\phi_n(y), h) \; < \; e(\phi) + \epsilon$$

whenever $n \geq N$. For the convergence to be uniform, we need $N := N(y, h, \epsilon)$ to be independent of y, h. only.

The main goal of this investigation is to characterize the smallest possible achievable worst-case error $e(\phi)$, over all possible estimators. In particular, the dependence of this error on the noise will be studied. The hope is that under suitable assumptions, the error will tend to zero as the noise level goes to zero. Two types of estimators will be studied, with two associated *Optimal Worst-Case Errors*

$$\begin{aligned} E^{\text{arb}} &:= \inf\{e(\phi) \mid \phi \text{ arbitrary }\} \\ E^{\text{cau}} &:= \inf\{e(\phi) \mid \phi \text{ causal }\} \end{aligned}$$

If causality is not required in the estimator, then the estimator can utilize the whole set of information $(\mathcal{S}_\infty(y))$ at any time n. This estimator is not desirable from a practical point of view,but has been studied extensively in the information-based complexity literature [10]. It is clear that: $E^{\text{arb}} \leq E^{\text{cau}}$.

3 Analysis of Optimal Worst-Case Error

For each nonempty subset $A \subset \mathcal{M}$ define the *Chebyshev radius* of A

$$R(A) := \inf_{c \in \mathcal{M}} \sup_{b \in A} \rho(c, b)$$

and the diameter of A

$$D(A) := \sup_{b_1, b_2 \in A} \rho(b_1, b_2) \,.$$

Let R and D denote the infinite-horizon radius and diameter of uncertainty, i.e.

$$R := \sup_{y \in \mathcal{L}} R(\mathcal{S}_\infty(y))$$

and

$$D := \sup_{y \in \mathcal{L}} D(\mathcal{S}_\infty(y))$$

It is a standard result from Information-Based Complexity (IBC) theory [10, 11] that

$$R \leq D \leq 2R \tag{4}$$

Also from IBC, the optimal worst-case asymptotic error over all arbitrary estimators satisfies:

$$E^{\text{arb}} = R \geq D/2 \tag{5}$$

Next, it will be shown that under mild assumptions, $E^{\text{cau}} \leq D$. This, combined with Equation(5), will give

$$D/2 \leq E^{\text{cau}} \leq D \tag{6}$$

It should be noted that the next result was proved in [12] for linear models and additive noise. The result here is more general and the proof is more elegant.

Authorized licensed use limited to: MIT Libraries. Downloaded on July 30,2020 at 15:12:12 UTC from IEEE Xplore. Restrictions apply.

Proposition 3.1 If $\mathcal{M}$ is σ-compact, then $D/2 \leq E^{cau} \leq D$.

The above proposition states the following: Given that the model set is σ-compact, then there exists a causal estimator such that the worst-case asymptotic error is bounded by D; the worst-case diameter of uncertainty.

As stated earlier, the convergence rate depends on the actual process h and possibly the noise. Uniform convergence can be obtained if the model set is compact. A proof of this result in the linear case is found in [12]; the proof of the general case is identical.

3.1 Separable Case

In many common application, the model set $\mathcal{M}$ is not σ-compact, however, it may be seperable, i.e. it has a countable dense set. In the sequel, the optimal worst case asymptotic error is analyzed for such model sets.

Given a subset $\mathcal{S} \subseteq \mathcal{M}$, for each $\varepsilon > 0$ we can define an ε-cover of the set S as:

$$B^{\epsilon}[S] := \{h' \in \mathcal{M} \,|\, \exists h \in \mathcal{S} \quad \rho(h,h') \leq \epsilon\}$$

Also denote

$$\mathcal{S}_n^{\epsilon}(y) := B^{\epsilon}[\mathcal{S}_n(y)]$$

and

$$\mathcal{S}^{\epsilon}(y) := \bigcap_{n \geq 0} \mathcal{S}_n^{\epsilon}(y)\,.$$

Note that the sets $\{\mathcal{S}_n^{\epsilon}(y) | n \geq 0\}$ give a valid uncertainty structure. For that, two conditions need to be verified. It is clear that $\mathcal{S}_n^{\epsilon}(y) \subseteq \mathcal{S}_m^{\epsilon}(y)$ if $m \leq n$, for all y, and that each set $\mathcal{S}_n^{\epsilon}(y)$ is closed. Note also that

$$B^{\epsilon}[\mathcal{S}_{\infty}(y)] \subseteq \mathcal{S}^{\epsilon}(y) \tag{7}$$

since for each n, $B^{\epsilon}[\mathcal{S}_{\infty}(y)] \subseteq B^{\epsilon}[\mathcal{S}_n(y)] = \mathcal{S}_n^{\epsilon}(y)$.

From

$$\mathcal{S}_n(y) = \mathcal{S}_n(y') \Rightarrow \mathcal{S}_n^{\epsilon}(y) = \mathcal{S}_n^{\epsilon}(y')$$

it follows that

$$\begin{array}{c} \phi \text{ causal estimator for } \{\mathcal{S}_n^{\epsilon}(y)\} \quad \Rightarrow \\ \text{also causal estimator for } \{\mathcal{S}_n(y)\} \end{array}$$

Let

$$D^{\epsilon} := \sup_{y \in \mathcal{L}} D(\mathcal{S}^{\epsilon}(y))$$

be the diameter computed relative to the ε-structure, for each fixed ε. Define

$$D^+ := \underline{\lim}_{\epsilon \to 0^+} D^{\epsilon}$$

Next, it is shown that if the model set is seperable, then the optimal worst case error satisfies $D/2 \leq E^{cau} \leq D^+$.

Theorem 1 *Assume that $\mathcal{M}$ is separable. Then, $D/2 \leq E^{cau} \leq D^+$.*

4 Fading Memory Systems

We now specialize and apply the above results to specific identification problems. In particular, model sets that contain operators (possibly nonlinear) with fading memory, with additive corrupting noise with a bound on its ℓ_∞ norm, will be analyzed in the context of worst-case identification.

4.1 Definitions

Let $\mathcal{U}$ be the set of infinite sequences whose ℓ_∞ norm is bounded by 1. This can be viewed as the input set which contains the possible inputs that can be used for performing the identification experiments. Let $\mathcal{L} \subset \Re^{\infty}$ be the observation space which contain the output sequences that can be observed in the experiments. We view our models as causal functions from $\mathcal{U}$ to $\Re^{\infty}$; these are discrete-time systems, possibly non-linear, which take as input a sequence in $\mathcal{U}$ to give an output sequence in $\Re^{\infty}$. The input and the output at time n will be denoted by u_n and $h_n(u)$. The model sets $\mathcal{M}$ we shall look at will be subsets of this class of functions.

We consider now that the output observed y is corrupted by some additive disturbance d which is unknown but magnitude-bounded, $||d||_\infty \leq \delta$, i.e. if h is the system, then

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

With this, the set $\mathcal{L}$ is precisely given by:

$$\mathcal{L} = \{y | y = h(u) + d,\ h \in \mathcal{M},\ ||d||_\infty \leq \delta\}$$

Fix the input sequence u. The uncertainty set $\mathcal{S}_n(y)$ at time n contains all the systems in the given model set $\mathcal{M}$ which are consistent with the observed data until time n:

$$\mathcal{S}_n(y) = \{h \in \mathcal{M} : ||P_n(y - h(u))||_\infty \leq \delta\}$$

where P_n is the truncation operator. These are the plants which can give rise to the observed output for some valid disturbance sequence. The infinite-horizon uncertainty set is

$$\mathcal{S}_\infty(y) = \{h \in \mathcal{M} : ||y - h(u)||_\infty \leq \delta\}$$

To measure the identification error, we shall define a metric ρ to be the one corresponding to the operator-induced norm:

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

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

Authorized licensed use limited to: MIT Libraries. Downloaded on July 30,2020 at 15:12:12 UTC from IEEE Xplore. Restrictions apply.

4.2 Definition of Fading Memory Systems

We shall now specialise further in defining the systems we are looking. Assume that they further satisfy the following properties

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

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

where $0u$ is the input $0, u_0, u_1, \ldots$.

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.

Definition 4.1 An operator h has *fading memory* (FM) if for each $\epsilon > 0$ there is some $T = T(\epsilon)$ such that: for every k, every $t \geq T$ and every finite sequences $v \in [-1,1]^k$, $w \in [-1,1]^t$,

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

It can easily be seen that fading memory systems satisfying properties (1) and (2) have bounded operator-induced norms. The resulting topological set is in fact seperable.

Proposition 4.2 The class of all fading memory systems is separable.

This means that when we look at fading memory system, we can apply Theorem 1, and reduce the analysis of asymptotic optimal error to the analysis of infinite-horizon diameter.

4.3 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 $g : \Re \to \Re$. 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.

5 Optimal Inputs

Let V be a finite subset of the allowable input set $\mathcal{U}$; V contains the input sequences to be used for the identification experiments, the number of such inputs is equal to $|V|$. With a slight change of the notation, we define the finite and infinite horison uncertainty sets for multiple experiments as:

$$S_n(V, \mathbf{y}, \delta) = \cap_{i=1}^{|V|} S_n(y^{(i)})$$

and

$$S_\infty(V, \mathbf{y}, \delta) = \cap_{i=1}^{|V|} S_\infty(y^{(i)})$$

where

$$\mathbf{y} = [y^{(1)}, \ldots, y^{(|V|)}]$$

are the outputs observed for all the experiments. We define $D(V,\delta)$ to be the diameter of this set (δ is the bound on the disturbance). The goal is to find the input set V such that $D(V,\delta)$ is as small as possible for any given δ.

Definition 5.1 A subset $V \subseteq \mathcal{U}^\infty$ is *persistently exciting* (PE) if

$$\max_{u \in V} \|h(u)\| = \|h\| \;\forall\, h \in \mathcal{M}$$

The next theorem gives the main result in charactering the optimal worst-case error. Recall that $\mathcal{M}$ is called *balanced* if $h \in \mathcal{M}$ implies that $-h \in \mathcal{M}$.

Theorem 2 *Assume $\mathcal{M}$ is balanced and convex and take any input set V. Then:*

1. *V is PE $\Rightarrow$ $D(V,\delta) \leq 2\delta$ for all $\delta > 0$. Also, $D^+(V,\delta) \leq 2\delta$ for all $\delta > 0$.*
2. *V is PE $\Rightarrow$ $D(V,\delta) = 2\delta$ for each $0 < \delta \leq \frac{D(\mathcal{M})}{2}$.*

Thus, if the set of inputs used in the experiments is PE, then accurate identification can be achieved asymptotically. An interesting question arises: Does there exist a single input that is PE? In the stable linear shift invariant case, this was shown to be the case [12]. In the next theorem, it is shown that such an input will also exist when the model set consists of all nonlinear fading memory systems.

Proposition 5.2 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 $\omega_0 \in [-1,1]^\infty$ which contains all possible finite sequences of elements of W. Then, $V = \{\omega_0\}$ is PE.

Authorized licensed use limited to: MIT Libraries. Downloaded on July 30,2020 at 15:12:12 UTC from IEEE Xplore. Restrictions apply.

5.1 Consistency without Knowing δ

The general consistency results Proposition 3.1 and Theorem 1 assume that the identification algorithm has knowledge of δ, the bound on the disturbance. This enables it to compute the uncertainty sets S_n and hence compute the simpliest plant consistent with the data. It will now be shown that, for the special case of fading memory systems, one can achieve consistency without knowing δ, provided that we use a persistently exciting input. This is in fact a generalisation of a result by Makila [8], which was developed in the context of stable LTI systems.

Theorem 3 *If u is persistently exciting then*

$$E^{cau}(\{u\}, \delta) \leq 2\delta$$

and there is an algorithm which attains the bound and does not require the knowledge of δ.

6 Conclusions

A general set-up for worst-case identification has been introduced and lower bounds on the optimal achievable errors were derived. For model sets that are either σ-compact or seperable, and for any experiment, the optimal worst-case error is always bounded by twice the lower bound. For the class of nonlinear fading memory system, optimal inputs are characterized. It is shown that accurate asymptotic identification can be achieved by one input, using an algorithm that does not require the bound on the disturbance when additive noise is considered.

Acknowledgement This research was supported by Grants AFOSR-91-0368, NSF 9157306-ECS and AFOSR-91-0346.

References

[1] A. Blumer, A. Ehrenfeucht. D. Haussler, M. Warmuth, "Occam's Rasor", *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] J. C. Doyle, "Analysis of Feedback Systems with Structured Uncertainty," *IEEE Proceedings* 129, 242-250,1982.

[4] G. Gu and P.P. Khargonekar, "A Class of Algorithms for Identification in H_∞", preprint.

[5] A.J. Helmicki, C.A. Jacobson and C.N. Nett, "Control-oriented System Identification: A Worst-case/deterministic Approach in H_∞'," to appear in *IEEE Trans. Automatic Control.*

[6] 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.

[7] 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.

[8] P.M. Makila, "Robust Identification and Galois Sequences", Technical Report 91-1, Process Control Laboratory, Swedish University of Abo, January, 1991.

[9] P.M. Makila and J.R. Partington, "Robust Approximation and Identification in H_∞", Proc. 1991 American Control Conference, June, 1991.

[10] J.F. Traub and H. Wosniakowski, *A General Theory of Optimal Algorithms*, Academic Press, New York, 1980.

[11] J.F. Traub, G. Wasilkowski and H. Wazniakowski, *Information-Based Complexity*, Academic Press, 1988.

[12] D. Tse, M.A. Dahleh and J.N. Tsitsiklis. Optimal Asymptotic Identification under bounded disturbances. Submitted to the *IEEE Trans. Automat. Contr.*

[13] 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.

Authorized licensed use limited to: MIT Libraries. Downloaded on July 30,2020 at 15:12:12 UTC from IEEE Xplore. Restrictions apply.