Systems & Control Letters 20 (1993) 157–166
North-Holland

The sample complexity of worst-case identification of FIR linear systems *

Munther A. Dahleh, Theodore V. Theodosopoulos and John N. Tsitsiklis
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 1 May 1992
Revised 7 October 1992

Abstract: We consider the problem of identification of linear systems in the presence of measurement noise which is unknown but bounded in magnitude by some $\delta > 0$. We focus on the case of linear systems with a finite impulse response. It is known that the optimal identification error is related (within a factor of 2) to the diameter of a so-called uncertainty set and that the latter diameter is upper-bounded by 2δ, if a sufficiently long identification experiment is performed. We establish that, for any $K \geq 1$, the minimal length of an identification experiment that is guaranteed to lead to a diameter bounded by $2K\delta$ behaves like $2^{Nf(1/K)}$, when N is large, where N is the length of the impulse response and f is a positive function known in closed form. While the framework is entirely deterministic, our results are proved using probabilistic tools.

Keywords: Worst-case identification; sample complexity; bounded but unknown disturbance.

1. Introduction

Recently, there has been increasing interest in the problem of worst-case identification in the presence of bounded noise. In such a formulation, a plant is known to belong to a model set $\mathcal{M}$, and its measured output is subject to an unknown but bounded disturbance. The objective is to use input/output information to derive a plant estimate that approximates the true plant as closely as possible, in some induced norm. For frequency domain experiments, algorithms that guarantee accurate identification in the $\mathcal{H}_\infty$ setting were furnished in [4,5,6,7]. For general experiments, algorithms that guarantee accurate identification in the ℓ_1 sense were suggested in [17,18]. These algorithms are based on the Occam's Razor principle by which the simplest model is always used to explain the given data. The optimal asymptotic worst-case error is characterized in terms of the diameter of the 'uncertainty set': the set of all plants consistent with all the data and the noise model. Other related work on the worst-case identification problem can be found in [8,10,11,19]. In particular, [10] presents a specific experiment that uses a Galois sequence as an input, and shows that the standard Chebyshev algorithm results in an asymptotic error bounded by the worst-case diameter of the uncertainty set. A Galois sequence is constructed by concatenating a countable number of finite sequences, such that the k-th sequence contains all possible combinations of $\{-1, +1\}$ of length k, and so it is rich enough to accurately identify exactly k parameters of the impulse response. The length of each sequence is clearly exponential in k. Finally, identification problems with bounded but unknown noise were studied in the context of prediction (not worst-case) in [12,13]. Other related work, for nonlinear systems, can be found in [3].

An important result from the work of [17,18] states that for the model set of all stable plants, accurate identification in the ℓ_1 sense is possible if and only if the input excites all possible frequencies on the unit circle. This is due to two reasons: the first is that bounded noise is quite rich and the second is that

Correspondence to: Prof. M.A. Dahleh, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

* Research supported by the AFOSR under grant AFOSR-91-0368, and by the NSF under grants 9157306-ECS and ECS-8552419, and by a grant from Siemens AG.

minimizing an induced norm such as the ℓ_1 norm implies that the estimate has a very good predictive power. Inputs with such properties tend to be quite long, and this suggests that the sample complexity of this kind of identification problems tends to be quite high, as a function of the numbers of estimated parameters of the impulse response.

In this paper, we will study the sample complexity (required length) of the inputs for worst-case identification of FIR plants, under the ℓ_1 norm, in the presence of arbitrary bounded measurement noise. It will be shown that in order to guarantee that the diameter of the uncertainty set is bounded by $2K\delta$, where δ is the bound on the noise and K is a constant (larger than 1), the length of the input must increase like $2^{Nf(1/K)}$, where N is the length of the impulse response and f is a positive function. Since the worst-case error is at least half of the diameter, these results show that the sample complexity is exponential in N even if the allowable accuracy is far from optimal, and capture the limitations of accurate identification in the worst-case set-up. We also show that our sample complexity estimate is tight, in the sense that there exist inputs of length approximately equal to $2^{Nf(1/K)}$ that lead to a $2K\delta$ bound on the diameter. An interesting technical aspect of this paper is that the existence of such inputs is established by means of a probabilistic argument reminiscent of the methods commonly employed in information theory.

Other researchers have also recently addressed the sample complexity of worst-case identification. In a personal discussion with Poolla (January 1992), he pointed out to us (specifically to Dahleh) that the optimal identification case had exponential complexity, as in the lower bound of our Theorem 2.1. We have recently received a preprint by Poolla and Tikku [14] which, among other results, contains exponential lower bounds for the sample complexity of suboptimal identification of FIR systems. These lower bounds are similar to, although somewhat weaker than, the lower bound in our Theorem 2.2. Chronologically, the results of [14] precede ours, although we didn't have knowledge of their results when writing our paper. Finally, [14] contains some upper bounds but, unlike our Theorem 2.2, they are far from being tight. Also, while writing our paper, we learned that Milanese [9] had arrived to results similar to the exponential lower bound in our Theorem 2.1. His report does not contain any discussion of the case where the error is within a factor of the optimal.

2. Problem definition

Let $\mathcal{M}_N$ be the set of all linear systems with a finite impulse response of length N. Any element h of $\mathcal{M}_N$ will be identified with a finite sequence $(h_1, \ldots, h_N) \in \mathbb{R}^N$. Let 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 U_n will be called an *input of length n*. Finally, for any positive number δ, let D_δ, called the *disturbance set*, be the set of all infinite sequences $d = \{d_i\}_{i=1}^{\infty}$ such that $|d_i| \leq \delta$ for all i.

We are interested in experiments of the following type: an input $u \in U_n$ is applied to an (unknown) system $h \in \mathcal{M}_N$, and we observe the noisy measurement

$$y = h * u + d, \tag{2.1}$$

where $*$ denotes convolution, and where $d \in D_\delta$ plays the role of an output disturbance or measurement noise. It is clear that, for $i > N + n$, we have $y_i = d_i$, and y_i carries no useful information on the unknown system h.

The set that contains all plants in the model set that are consistent with the input/output data and the noise model is called the uncertainty set and is given by

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

The diameter diam(S) of a subset S of ℓ_1 is defined by

$$\operatorname{diam}(S) = \sup_{x,y \in S} \| x - y \|_1.$$

We then define the worst case diameter for a given input $u \in U_n$ by

$$D_{N,n}(u) = \sup_{d \in D_\delta} \sup_{\phi \in \mathcal{M}_N} \operatorname{diam}\big(S_{N,n}(u * \phi + d, u)\big).$$

Any identification algorithm that lets its plant estimate be an element of the uncertainty set has an error upper-bounded by the diameter of the uncertainty set. Besides, it is shown in [15,16,17] that the error of any identification algorithm is lower-bounded by half the diameter of the uncertainty set. Define

$$D^*_{N,n} = \inf_{u \in U_n} D_{N,n}(u).$$

It is shown in [17] that

$$\lim_{n \to \infty} D^*_{N,n} = 2\delta. \tag{2.2}$$

Thus, as the length of the experiments increases, and with a suitable identification algorithm, the worst-case error can be made as small as twice the disturbance bound δ, but no smaller than δ. A question that immediately arises is how long should n be for the error to approach 2δ. We address this question by focusing on the behavior of the diameter of the uncertainty set, as the inputs are allowed to become longer.

Let us define

$$n^*(N) = \min\{n \mid D^*_{N,n} = 2\delta\}. \tag{2.3}$$

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 main theorem (Theorem 2.2) of this paper.

Theorem 2.1. [1] *For any $\delta > 0$ and N, we have $2^{N-1} + N - 1 \le n^*(N) \le 2^N + N - 1$.*

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 $\mathscr{A}$, $u \in U_m$, be such that $D_{N,m}(u) = 2\delta$. Let $v \in \{-1, 1\}^m$ be defined by $v_i = 1$ if $u_i \ge 0$, and $v_i = -1$ if $u_i < 0$. For notational convenience, we define $u_i = 0$ for $i \le 0$. We distinguish two cases:

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

(b) Suppose now that the assumption of case (a) fails to hold. Let $\phi \in \{-1, 1\}^N$ be such that both ϕ and $-\phi$ 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 $h = \delta\phi/(N-1)$. Then

$$|(h * u)_i| = \left| \sum_{k=1}^{N} h_k u_{i-k} \right| = \frac{\delta}{N-1} \left| \sum_{k=1}^{N} \phi_k u_{i-k} \right|. \tag{2.4}$$

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

$$\left| \sum_{k=1}^{N} \phi_k u_{i-k} \right| \le N - 1. \tag{2.5}$$

[1] We acknowledge Professor Poolla for pointing out an error in the previous version of this theorem.

We finally note that for $i \notin (N, m]$, at least one of the summands $\phi_k u_{i-k}$ is equal to zero, which implies that (2.5) is valid for all i. Combining (2.4) and (2.5), 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 $h = -\delta\phi/(N-1)$, 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 $\delta\phi/(N-1)$ or $-\delta\phi/(N-1)$. This implies that the worst case diameter satisfies

$$D_{N,m}(u) \geq 2 \,\| \delta\phi/(N-1) \|_1 > 2\delta. \tag{2.6}$$

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 [10,17]. Let u be a finite sequence whose entries belong to $\{-1, 1\}$ and such that for every $\phi \in \{-1, 1\}^N$ there exists some $i(\phi)$ such that $\phi = (u_{i(\phi)}, u_{i(\phi)+1}, \dots, u_{i(\phi)+N-1})$. Such a sequence, called a Galois sequence, can be chosen so that its length is equal to $2^N + N - 1$ [10]. With this input, the worst case diameter is equal to 2δ. □

Theorem 2.1 has the disappointing conclusion that the worst-case error is guaranteed to become at most 2δ 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 identification experiment is prohibitively long if an error guarantee as small as 2δ is desired. This motivates the problem studied in this paper: 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? Theorem 2.2 below is equally disappointing with Theorem 2.1: 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 2.2. *Fix some $K > 1$ and let*

$$n^*(N, K) = \min\{n \mid D^*_{N,n} \leq 2K\delta\}. \tag{2.7}$$

Then:

(a) $n^*(N, K) \geq 2^{Nf(1/K)-1} - N + 2\lceil N/K \rceil - 1$.

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

Here, $f : (0, 1) \mapsto \mathbb{R}$ is the function defined by [2]

$$f(\alpha) = 1 + \left(\frac{1-\alpha}{2}\right) \log\left(\frac{1-\alpha}{2}\right) + \left(\frac{1+\alpha}{2}\right) \log\left(\frac{1+\alpha}{2}\right). \tag{2.8}$$

Notice that the function f defined by (2.8) satisfies $f(\alpha) = 1 - H(\frac{1}{2}(1-\alpha))$, 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 2.1. *Let $X_1, X_2, \dots, X_N$ be independent binomial random variables with* $\Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2}$ *for every i.*

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

$$\Pr\left(\frac{1}{N}\sum_{i=1}^{N} u_i X_i \geq \alpha\right) \leq 2^{-Nf(\alpha)}. \tag{2.9}$$

[2] In the definition of f, and throughout the rest of the paper, all logarithms are taken with base 2.

(b)

$$\lim_{N\to\infty} \frac{1}{N} \log \Pr\left(\frac{1}{N} \sum_{i=1}^{N} X_i \geq \alpha \right) = -f(\alpha). \tag{2.10}$$

Proof. Part (b) is obtained from the classical Chernoff bound [1] or from counting arguments [2]. 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

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

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 $\mathrm{e}^{su} + \mathrm{e}^{-su} \leq \mathrm{e}^{s} + \mathrm{e}^{-s}$ 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. □

One consequence of Lemma 2.1 is that for any $\varepsilon > 0$, there exists some $N_0(\alpha, \varepsilon)$ such that

$$\Pr\left(\frac{1}{N} \sum_{i=1}^{N} X_i \geq \alpha \right) \geq 2^{-N(f(\alpha)+\varepsilon)}, \quad \forall N \geq N_0(\alpha, \varepsilon). \tag{2.11}$$

The following lemma strengthens (2.11) and will be needed later in the proof.

Lemma 2.2. *Let $X_1, \ldots, X_N$ be as in Lemma 2.1. Let $\Theta_N = \{(\theta_1, \ldots, \theta_N) \in \mathbb{R}^N \mid \sum_{i=1}^N |\theta_i| = N\}$. Then, for any $\varepsilon_1 > 0$, there exists some $N_1(\alpha\varepsilon_1)$ such that*

$$\Pr\left(\frac{1}{N} \sum_{i=1}^{N} \theta_i X_i \geq \alpha \right) \geq 2^{-N(f(\alpha)+\varepsilon_1)}, \quad \forall N \geq N_1(\alpha\varepsilon_1), \forall \theta \in \Theta_N. \tag{2.12}$$

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

$$\begin{aligned} \Pr\left(\sum_{i=1}^{N} \theta_i X_i \geq \alpha N \right) &= \Pr\left(\sum_{i=1}^{N} \theta_i X_i \geq \alpha N \,\middle|\, \sum_{i=1}^{N} X_i \geq \alpha N \right) \cdot \Pr\left(\sum_{i=1}^{N} X_i \geq \alpha N \right) \\ &\geq 2^{-N(f(\alpha)+\varepsilon_1/2)} \Pr\left(\sum_{i=1}^{N} \theta_i X_i \geq \alpha N \,\middle|\, \sum_{i=1}^{N} X_i \geq \alpha N \right), \end{aligned} \tag{2.13}$$

where the last inequality holds for all N large enough, as a consequence of (2.11).

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,

$$\begin{aligned} \Pr\left(\sum_{i=1}^{N} \theta_i X_i \geq \alpha N \,\middle|\, \sum_{i=1}^{N} X_i \geq \alpha N \right) &= \frac{1}{N} \sum_{k=1}^{N} \Pr\left(\sum_{i=1}^{N} \theta_i X_i^k \geq \alpha N \,\middle|\, \sum_{i=1}^{N} X_i \geq \alpha N \right) \\ &\geq \frac{1}{N} \Pr\left(\exists k \text{ such that } \sum_{i=1}^{N} \theta_i X_i^k \geq \alpha N \,\middle|\, \sum_{i=1}^{N} X_i \geq \alpha N \right) \\ &= \frac{1}{N}. \end{aligned} \tag{2.14}$$

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 (2.13) becomes

$$\Pr\left(\sum_{i=1}^{N} \theta_i X_i \geq \alpha N \right) \geq \frac{1}{N} 2^{-N(f(\alpha)+\varepsilon_1/2)} \geq 2^{-N(f(\alpha)+\varepsilon_1)},$$

where the last inequality follows if N is large enough so that $1/N \geq 2^{-N\varepsilon_1/2}$. □

Having finished with the probabilistic preliminaries, we can now continue with the main part of the proof of Theorem 2.2. We will start with the proof of part (a).

Lemma 2.3. *Suppose that the length n of an input sequence $u \in U_n$ is smaller than $2^{Nf(1/K)-1} - N + 2\lceil N/K \rceil - 1$. Then, there exists some $h \in \{-K\delta/N, K\delta/N\}^N$ such that $\| u * h \|_\infty < \delta$.*

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 $\{-K\delta/N, K\delta/N\}^N$ satisfies $\| u * h \|_\infty < \delta$ with positive probability. Indeed, let h be such a random element, under the uniform distribution on $\{-K\delta/N, K\delta/N\}^N$. Then

$$\begin{aligned}\Pr(\| u * h \|_\infty \geq \delta) &\leq \sum_{j=1}^{N+n} \Pr(|(u * h)_j| \geq \delta) = \sum_{j=\lceil N/K \rceil+1}^{N+n-\lceil N/K \rceil+1} \Pr(|(u * h)_j| \geq \delta) \\ &\leq (N+n-2\lceil N/K \rceil + 1) \max_{1 \leq j \leq N+n} \Pr(|(u * h)_j| \geq \delta).\end{aligned} \tag{2.15}$$

where the equality on the first line holds because for $j \leq \lceil N/K \rceil$, we have

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

and for $j \geq N+n-\lceil N/K \rceil + 2$, 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,

$$\begin{aligned}\Pr(|(u * h)_j| \geq \delta) &= \Pr\left(\left| \sum_{i=1}^{N} h_i u_{j-1} \right| \geq \delta \right) \\ &= \Pr\left(\frac{1}{N} \left| \sum_{i=1}^{N} (Nh_i/K\delta) u_{j-i} \right| \geq \frac{1}{K} \right) \leq 2 \cdot 2^{-Nf(1/K)}.\end{aligned} \tag{2.16}$$

The last inequality follows from Lemma 2.1 (a), because the random variables $Nh_i/K\delta$ are independent, take values in $\{-1, 1\}$, and each value is equally likely. Combining (2.15) and (2.16), we conclude that

$$\Pr(\| u * h \|_\infty \geq \delta) \leq 2\left(N+n-2\left\lceil \frac{N}{K} \right\rceil + 1 \right) 2^{-Nf(1/K)}. \tag{2.17}$$

If $2(N+n-2\lceil N/K \rceil + 1) < 2^{Nf(1/K)}$, then the right-hand side of (2.17) is smaller than 1. This implies that there exists some $h \in \{-K\delta/N, K\delta/N\}^N$ for which $\| h * u \|_\infty < \delta$. □

Suppose now that the length n of the input sequence u is as in Lemma 2.3, 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 $|(-h * u)| < \delta$, 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 $(1+\varepsilon)h$, where $\varepsilon > 0$ 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 $n < 2^{Nf(1/K)-1} - N + 2\lceil N/K \rceil - 1$, then $D_{N,n}(u) > 2K\delta$. Equivalently, $n^*(N, K) \geq 2^{Nf(1/K)-1} - N + 2\lceil N/K \rceil - 1$, 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

$$\limsup_{N\to\infty} (1/N) \log n^*(N, K) \leq f(1/K).$$

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 $\varphi > 0$. Let $M(N)$ be the smallest integer larger than

$$M(N) \geq 2^{N(f(\varepsilon+1/K)+2\varepsilon)}. \tag{2.18}$$

For every $k \in \{1, \ldots, M(N)\}$, we choose a vector $u^k = (u_1^k, \ldots, u_N^k) \in \{-1, 1\}^N$. The input u is then defined by

$$u = (u^1, u^2, \ldots, u^{M(N)}), \tag{2.19}$$

and has length $NM(N)$.

Lemma 2.4. *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 $N_2(\varepsilon)$ such that*

$$\Pr(\exists h \in \mathcal{M}_N \text{ such that } \| h \|_1 \geq K\delta,\ \| u * h \|_\infty \leq \delta) < 1, \quad \forall N \geq N_2(\varepsilon). \tag{2.20}$$

Proof. Let Q_N be the left-hand side of (2.20). Notice that if i is an integer multiple of N, with $i = mN$, we have

$$(u * h)_i = \sum_{j=1}^{N} u_j^m h_{N-j}, \quad i = mN. \tag{2.21}$$

We then have

$$\begin{aligned} Q_N &= \Pr(\exists h \in \mathcal{M}_N \text{ such that } \| h \|_1 \geq K\delta,\ \| u * h \|_\infty \leq \delta) \\ &= \Pr(\exists h \in \mathcal{M}_N \text{ such that } \| h \|_1 = K\delta,\ \| u * h \|_\infty \leq \delta) \\ &= \Pr(\exists h \in \mathcal{M}_N \text{ such that } \| h \|_1 = N,\ \| u * h \|_\infty \leq N/K) \\ &\leq \Pr\left(\exists h \in \mathcal{M}_N \text{ such that } \| h \|_1 = N,\ \left| \sum_{j=1}^{N} u_j^m h_{N-j} \right| \leq N/K,\ m = 1, \ldots, M(N)\right), \end{aligned} \tag{2.22}$$

where the last inequality follows from (2.21).

Let us choose a finite subset $\mathcal{M}_N^\varepsilon$ of $\mathcal{M}_N$ such that for every $h \in \mathcal{M}_N$ with $\|h\|_1 = N$, there exists some $h' \in \mathcal{M}_N^\varepsilon$ satisfying $\|h'\|_1 = N$ and $\|h - h'\|_\infty < \varepsilon$. In particular, $\mathcal{M}_N^\varepsilon$ 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 ε/N. It is then clear that $\mathcal{M}_N^\varepsilon$ can be assumed to have cardinality bounded by $((2N+1)/\varepsilon)^N$. We then have

$$\Pr\left(\exists h \in \mathcal{M}_N \text{ such that } \|h\|_1 = N, \left|\sum_{j=1}^N u_j^m h_{N-j}\right| \le N/K,\ m = 1,\dots,M(N)\right)$$
$$\le \Pr\left(\exists h' \in \mathcal{M}_N^\varepsilon \text{ such that } \left|\sum_{j=1}^N u_j^m h'_{N-j}\right| < N(\varepsilon + 1/K),\ m = 1,\dots,M(N)\right)$$
$$\le \left(\frac{2N+1}{\varepsilon}\right)^N \max_{h' \in \mathcal{M}_N^\varepsilon} \Pr\left(\left|\sum_{j=1}^N u_j^m h'_{N-j}\right| < N(\varepsilon + 1/K),\ m = 1,\dots,M(N)\right). \tag{2.23}$$

We provide an upper bound to the probability in the right-hand side of (2.23) by applying Lemma 2.2. (Here, u_j^m and h'_{N-j} correspond to X_i and θ_i in the notation of that lemma.) Indeed, Lemma 2.2 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 2.1. A minor difference is that the components of h' could be negative, while in Lemma 2.2 we assumed that the components of θ are nonnegative. Nevertheless, if we replace each component of h' with its absolute value, the distribution of the random variables $\sum_{j=1}^N u_j^m h'_{N-j}$ remains the same. We therefore conclude that there exists some $N_2(K, \varepsilon)$ such that

$$\Pr\left(\left|\sum_{j=1}^N u_j^m h'_{N-1}\right| < N(\varepsilon + 1/K)\right) \le 1 - 2^{-N(f(\varepsilon+1/K)+\varepsilon)}, \quad \forall m, \forall N \ge N_2(K, \varepsilon). \tag{2.24}$$

By combining (2.22), (2.23), (2.24), and using the statistical independence of the vectors u^m, we obtain

$$Q_N \le ((2N+1)/\varepsilon)^N \left(1 - 2^{-N(f(\varepsilon+1/K)+\varepsilon)}\right)^{M(N)}$$
$$\le ((2N+1)/\varepsilon)^N \exp\{-M(N)2^{-N(f(\varepsilon+1/K)+\varepsilon)}\} \le ((2N+1)/\varepsilon)^N \exp\{-2^{\varepsilon N}\}, \tag{2.25}$$

where the second inequality follows from the fact $(1 - 1/x)^x \le e^{-1}$, for every $x > 0$, and the last inequality follows from the definition of $M(N)$ [cf. (2.18)]. It is then easily seen that Q_N converges to zero as N increases, which establishes the desired result. □

Lemma 2.4 establishes that, if the input u is constructed randomly as in the discussion preceding the lemma, then, with positive probability, u will have property P below:

$$\text{P:} \quad \text{if } h \in \mathcal{M}_N \text{ and } \|u * h\|_\infty \le \delta, \text{ then } \|h\|_1 \le K\delta. \tag{2.26}$$

In particular, there exists at least one u, of length $n = M(N)N$ that has property P. [3]

Lemma 2.5. *If an input u has property* P *of* (2.26), *then* $D_{N,n}(u) \le 2K\delta$.

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 set of uncertainty

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

Let χ and ψ be two elements of $S_{N,n}(y, u)$. Then, $\|y - \chi * u\|_\infty \le \delta$ and $\|y - \psi * u\|_\infty \le \delta$. Using the triangle inequality, we obtain $\|u * (\chi - \psi)/2\|_\infty \le \delta$. Since u has property P, we conclude that

[3] In fact, it is easily seen that Q_N converges to zero very rapidly, which implies that most u's will have property P.

$\|(\chi - \psi)/2\|_1 \le K\delta$ or $\|\chi - \psi\|_1 \le 2K\delta$. Since this is true for all elements of $S_{N,n}(y, u)$, the diameter of $S_{N,n}(y, u)$ is at most $2K\delta$. □

As discussed earlier, if N is large enough, there exists an input of length $n = M(N)N$ that has property P and, by Lemma 2.5, leads to uncertainty sets whose diameter is bounded above by $2K\delta$. It follows that $n^*(N, K) \le M(N)N$. Using the definition of $M(N)$ [cf. (2.18)], we see that

$$\limsup_{N\to\infty}(1/N)\log n^*(N, K) \le \limsup_{N\to\infty}(1/N)\log M(N)N \le f\left(\varepsilon + \frac{1}{K}\right) + 2\varepsilon. \tag{2.27}$$

Since Eq. (2.27) is valid for all $\varepsilon > 0$, and since f is continuous, we conclude that

$$\limsup_{N\to\infty}(1/N)\log n^*(N, K) \le f(1/K),$$

which concludes the proof of Theorem 2.2. □

3. Conclusions

This paper addresses issues in the sample complexity of worst-case identification in the presence of unknown but bounded noise. Two main results are furnished: the first is a lower bound on the length of inputs necessary to approximate N steps of an impulse response to an accuracy within a factor K of the best possible achievable error. This bound has the form $2^{Nf(1/K)}$, and hence is exponential in N. The second result shows that this lower bound in asymptotically tight, i.e. for large enough N, there exists an input of length close to the lower bound that allows the identification of N steps of the impulse response.

References

[1] R.R. Bahadur, *Some Limit Theorems in Statistics* (SIAM, Philadelphia, PA, 1971).
[2] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Channels* (Academic Press, New York, 1981).
[3] M.A. Dahleh, E. Sontag, D.N. Tse and J.N. Tsitsiklis, Worst-case identification for a class of nonlinear fading memory systems, in *Proc. American Control Conference* (1992).
[4] G. Gu, P.P. Khargonekar and Y. Li, Robust convergence of two stage nonlinear algorithms for identification in H_∞, *Systems Control Lett.* **18** (4) (1992) 253–263.
[5] G. Gu and P.P. Khargonekar, Linear and nonlinear algorithms for identification in $\mathscr{H}_\infty$ with error bounds, *IEEE Trans. Automat. Control* **37** (7) (1992) 953–963.
[6] A.J. Helmicki, C.A. Jacobson and C.N. Nett, Control-oriented system identification: A worst-case/deterministic approach in H_∞, *IEEE Trans. Automat. Control* **36** (10) (1991) 1163–1176.
[7] A.J. Helmicki, C.A. Jacobson and C.N. Nett, Identification in H_∞: Linear algorithms, *Proc. American Control Conference* (1990) 2418–2423.
[8] C.A. Jacobson and C.N. Nett, Worst-case system identification in ℓ_1: Optimal algorithms and error bounds, *Proc. American Control Conference* (June 1991).
[9] B. Kacewicz and M. Milanese, On the optimal experiment design in the worst-case ℓ_1 system identification, *Proc. Conf. Dec. Control* (1992) 56–61.
[10] P.M. Makila, Robust Identification and Galois Sequences, Technical Report 91-1, Process Control Laboratory, Swedish University of Abo (Jan. 1991).
[11] P.M. Makila and J.R. Partington, Robust approximation and identification in H_∞, *Proc. American Control Conference* (June 1991).
[12] M. Milanese and R. Tempo, Optimal algorithm theory for robust estimation and prediction, *IEEE Trans. Automat. Control* **30** (1985) 730–738.
[13] M. Milanese, Estimation theory and prediction in the presence of unknown and bounded uncertainty: A survey, in: M. Milanese, R. Tempo and A. Vicino, Eds., *Robustness in Identification and Control* (Plenum Press, New York, 1989).
[14] K. Poolla and A. Tikku, On the time complexity of worst-case system identification, Preprint (1992).

[15] J.F. Traub and H. Wozniakowski, *A General Theory of Optimal Algorithms* (Academic Press, New York, 1980).
[16] J.F. Traub, G. Wasilkowski and H. Wozniakowski, *Information-Based Complexity* (Academic Press, New York, 1988).
[17] D. Tse, M.A. Dahleh and J.N. Tsitsiklis. Optimal asymptotic identification under bounded disturbances, *IEEE Trans. Automat. Control* (to appear).
[18] D.N.C. Tse, M.A. Dahleh, J.N. Tsitsiklis, Optimal and robust identification in the ℓ_1 norm, *Proc. American Control Conference* (June 1991).
[19] G. Zames, On the metric complexity of causal linear systems: ε-entropy and ε-dimension for continuous time, *IEEE Trans. Automat. Control* **24** (2) (1979) 222–230.