15th Triennial World Congress, Barcelona, Spain

MODEL-SET IDENTIFICATION BASED ON LEARNING-THEORETIC INEQUALITIES

Yasuaki Oishi[*] and Hidenori Kimura[†]

[*] *Department of Mathematical Informatics*
The University of Tokyo, Tokyo 113–0033, Japan

[†] *Department of Complexity Science and Engineering*
The University of Tokyo, Tokyo 113–0033, Japan

Abstract: A model-set identification algorithm is proposed in a probabilistic framework based on the leave-one-out technique. It provides a nominal model and a bound of its uncertainty for a provided plant assuming that the effect of the past inputs decays with a known bound. Since it does not require further assumptions on the true plant dynamics or on the noise, a risk to make inappropriate assumptions is small. The number of assumptions is shown to be minimum in the sense that identification is impossible after removing the assumption made here. An algorithm similar to the proposed one is constructed based on a mixing property. A simple plant is identified by means of the proposed algorithm for illustration.

Keywords: system identification, modelling errors, probabilistic models, linear programming, stochastic properties

1. INTRODUCTION

In order to use the robust control scheme, we need not only a nominal model of the plant to be controlled but also an upper bound of its uncertainty. Model-set identification (or worst-case identification, set-membership identification) is an identification method to obtain these two and has actively been investigated in the last decade by many authors including Zhu (1989), Helmicki *et al.* (1991), Milanese and Vicino (1991), Goodwin *et al.* (1992), Tse *et al.* (1993), De Vries and Van den Hof (1995), and Zhou and Kimura (1995).

In model-set identification, one makes prior assumptions on the true plant dynamics and on the noise; considers all the models that satisfy these assumptions and are consistent with the observed input-output data; gives a nominal model and a bound of its uncertainty so as to cover the worst case. In the existing algorithms for model-set identification, it is assumed for example that the true plant dynamics can be described by an ARX model of some order, that the noise has a known hard bound, or that the noise is subject to a Gaussian distribution. However, it does not seem to be often the case that the plant dynamics and the noise actually satisfy these strong assumptions. Indeed, many actual plants are infinite-dimensional and have nonlinearity; The noise usually contains a modeling error and is not Gaussian. Even if satisfaction of these assumptions is possible in principle, it is difficult to make appropriate assumptions using an appropriate bound or an appropriate distribution because neither the true plant dynamics nor the noise is directly observable. This problem is serious in model-set identification, which is for capturing the worst case. Suppose that the adopted assumptions are inappropriate and are not satisfied by the actual plant. Then we may miss a possible worst case that does not satisfy those assumptions.

In order to partially resolve this problem, we formulate model-set identification in a probabilistic framework and apply a technique called the leave-one-out estimation. The leave-one-out technique was used by Vapnik (1998) to prove statistical properties of a support vector machine. It enables us to make a worst-case estimation with little information on the underlying probability structure. In our approach, this technique is used so that one can perform model-set identification with only one assumption, which is on the decay rate of the past memory in the output. Consequently, a risk to make inappropriate assumptions is suppressed. It is possible to prove that identification is impossible after removing this remaining assumption. In this sense, the number of assumptions is reduced to the minimum. We also consider in this paper generalization of this approach in several directions.

There are other techniques to perform worst-case estimation with little information on the object especially in the field of learning theory. They have been applied to model-set identification by Oishi and Kimura (2001) and to conventional identification by Bartlett and Kulkarni (1998), Campi and Kumar (1998), Weyer *et al.* (1999), Weyer and Campi (1999, 2000), and Weyer (2000). See also Khargonekar and Tikku (1996), Tempo *et al.* (1997), and Vidyasagar (2001) for their application to analysis and synthesis of control systems. However, it has been often the case that the estimation given by learning theory is conservative and one has to collect millions of data to obtain meaningful information. In this regard, our approach seems to be advantageous because it gives a result from a reasonable number of data as is seen in Section 4.

2. IDENTIFICATION WITH A SMALL NUMBER OF ASSUMPTIONS

2.1. Identification algorithm

In this paper, a plant to be identified is a single-input-single-output discrete-time time-invariant causal system with an input u_k and an output y_k, where k is an integer. It is supposed that the input-output relation of the plant is described by an unknown but fixed conditional probability distribution $P_{y_k|u_{k-1},u_{k-2},\ldots}$, which is invariant with k. This means that the output y_k is affected by a stochastic noise and its probability distribution varies depending on what inputs $u_{k-1}, u_{k-2}, \ldots$ are provided. Note that the plant is allowed to have quite a large class of dynamics including nonlinear one. We let the inputs $\{u_k\}$ be a sequence of random numbers, which distribute independently and identically in $[-1, 1]$. Here we make an assumption crucial for our purpose. Assume that the output y_k can be approximated by an auxiliary signal $\overline{y}_k$ with $|y_k - \overline{y}_k| \leq \phi$ for any k and that $\overline{y}_k$ is statistically independent of $\overline{y}_{k'}$ and $u_{k'}$ whenever $|k - k'| \geq K$. The value of $\overline{y}_k$ does not need to be known though those of ϕ and K should be known. Let us call this property *approximate K-independence* henceforth. The assumption of this property, which was used by Weyer *et al.* (1999), is acceptable because the effect of the past inputs decay exponentially fast in many practical plants and the amount of this effect may be estimated by preliminary experiments. Since the output y_k is directly observable, an assumption on y_k is considered to be easier to make than those on the true plant dynamics or the noise.

We approximate the plant output y_k using the one-step-ahead predictor $\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{h}$ associated with the d-th order FIR model, where $\boldsymbol{x}_k = [u_{k-1} \ \ \ldots \ \ u_{k-d}]^{\mathrm{T}}$ and $\boldsymbol{h} = [h_1 \ \ \ldots \ \ h_d]^{\mathrm{T}}$. Suppose that the input-output data $u_1, u_2, \ldots, u_N; y_1, y_2, \ldots, y_N$ are provided. From them we extract $n := \lfloor N/K \rfloor$ pairs $\{(y_{(j)}, \boldsymbol{x}_{(j)})\}_{j=1}^{n}$, where $y_{(j)} := y_{jK}$ and $\boldsymbol{x}_{(j)} := \boldsymbol{x}_{jK} = [u_{jK-1} \ \ \ldots \ \ u_{jK-d}]^{\mathrm{T}}$. Now obtain the pair $(\boldsymbol{h}, a)$ that minimizes a under the conditions $|y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\boldsymbol{h}| \leq a$, $j = 1, \ldots, n$, and write it as $(\boldsymbol{h}^*, a^*)$. Moreover, find $\boldsymbol{h}$ that maximizes $|h_i - h_i^*|$ for each $i = 1, \ldots, d$ under the conditions $|y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\boldsymbol{h}| \leq a^* + 2\phi$, $j = 1, \ldots, n$, and write the maximum value as e_i. The above optimization can be carried out by means of linear programming. Now sample a new input-output pair $(y_{k_0}, \boldsymbol{x}_{k_0})$ at $k_0 \geq N + K$. We then have the next theorem.

Theorem 1. *There holds* $|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}}\boldsymbol{h}^*| \leq a^* + 2\phi + \sum_{i=1}^{d} e_i$ *with probability greater than or equal to* $1 - (d+1)/(n+1)$. *Here the probability is measured with respect to the input-output data* $\{(y_k, \boldsymbol{x}_k)\}_{k=1}^{N}$ *and the newly sampled input-output pair* $(y_{k_0}, \boldsymbol{x}_{k_0})$.

Consider an identification algorithm that gives $y_k = \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{h}^*$ as a nominal model and $a^* + 2\phi + \sum_{i=1}^{d} e_i$ as its uncertainty bound. Theorem 1 gives a probabilistic guarantee to this algorithm. This guarantee is not asymptotic, that is, it is meaningful with a finite n. Note also that this algorithm captures the worst case since Theorem 1 holds irrespective of the probabilistic properties of the plant and the noise. One may notice that only a small part of input-output data is used in the proposed algorithm. This is because we suppose that two outputs closer in time than K may be dependent on each other to any level; we can expect fresh information to be obtained only at every K time instances.

2.2. Proof of Theorem 1

Lemma 1. *Let* $(\boldsymbol{h}^*, a^*)$ *minimize* a *subject to* $|y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\boldsymbol{h}| \leq a$, $j = 1, \ldots, n$, *and let* $(\overline{\boldsymbol{h}}^*, \overline{a}^*)$ *minimize* a *subject to* $|\overline{y}_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\boldsymbol{h}| \leq a$, $j = 1, \ldots, n$. *Then there hold* $|a^* - \overline{a}^*| \leq \phi$ *and* $|h_i^* - \overline{h}_i^*| \leq e_i$ *for each of* $i = 1, \ldots, d$.

Proof. Note that a^* is equal to $\inf_{\boldsymbol{h}} \max_{j=1,\ldots,n} |y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\boldsymbol{h}|$ and the infimum is attained at

$\boldsymbol{h} = \boldsymbol{h}^*$. The corresponding relation holds on $\overline{a}^*$ and $\overline{\boldsymbol{h}}^*$, too. Since the discrepancy between $\max_j |y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\boldsymbol{h}|$ and $\max_j |\overline{y}_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\boldsymbol{h}|$ is less than or equal to ϕ for any $\boldsymbol{h}$, the inequality $|a^* - \overline{a}^*| \leq \phi$ has to hold. It is also seen that $\max_j |y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\overline{\boldsymbol{h}}^*| - \phi \leq \max_j |\overline{y}_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\overline{\boldsymbol{h}}^*| = \overline{a}^*$, which is combined with $\overline{a}^* \leq a^* + \phi$ and implies $\max_j |y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\overline{\boldsymbol{h}}^*| \leq a^* + 2\phi$. By the definition of e_i, there has to hold $|h_i^* - \overline{h}_i^*| \leq e_i$. □

To the newly sampled input-output pair $(y_{k_0}, \boldsymbol{x}_{k_0})$, we can associate $\overline{y}_{k_0}$, which approximates y_{k_0}. Write $(\overline{y}_{k_0}, \boldsymbol{x}_{k_0})$ as $(\overline{y}_{(n+1)}, \boldsymbol{x}_{(n+1)})$ to see that $\{(\overline{y}_{(j)}, \boldsymbol{x}_{(j)})\}_{j=1}^{n+1}$ is a set of $n+1$ independently identically distributed input-output pairs. Now one can show the next lemma, which is based on the technique of *leave-one-out estimation* used by Vapnik (1998).

Lemma 2. *When $\overline{a}^*$ is chosen as is stated above, the probability to have $|\overline{y}_{(n+1)} - \boldsymbol{x}_{(n+1)}^{\mathrm{T}}\overline{\boldsymbol{h}}^*| > \overline{a}^*$ is less than or equal to $(d+1)/(n+1)$.*

Proof. Let us find $(\boldsymbol{h}, a)$ that minimizes a under the condition $|\overline{y}_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\boldsymbol{h}| \leq a$ for $j = 1, \ldots, n+1$ instead of $j = 1, \ldots, n$. We consider the probability with which the minimum value of a changes when arbitrary one of the $n+1$ conditions is removed. If this probability is shown to be less than or equal to $(d+1)/(n+1)$, the claim of the lemma is ensured.

The inequalities $|\overline{y}_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}}\boldsymbol{h}| \leq a$, $j = 1, \ldots, n+1$, are decomposed into $[\boldsymbol{x}_{(j)}^{\mathrm{T}} \quad 1]\begin{bmatrix}\boldsymbol{h}\\ a\end{bmatrix} \geq \overline{y}_{(j)}$ and $[-\boldsymbol{x}_{(j)}^{\mathrm{T}} \quad 1]\begin{bmatrix}\boldsymbol{h}\\ a\end{bmatrix} \geq -\overline{y}_{(j)}$. Among these $2(n+1)$ inequalities, we notice those satisfied with the equality at the minimizing point $(\overline{\boldsymbol{h}}^*, \overline{a}^*)$. By the Karush-Kuhn-Tucker necessary condition for optimality, the vector $[0 \ \ldots \ 0 \ \ 1]$ can be described as a linear combination with positive coefficients of the vectors $[\boldsymbol{x}_{(j)}^{\mathrm{T}} \quad 1]$ and $[-\boldsymbol{x}_{(j)}^{\mathrm{T}} \quad 1]$ corresponding to the inequalities noticed above. Since the description may not be unique, we notice the vectors included in common by all of such descriptions. The number of such vectors should be less than or equal to $d+1$.

The minimum value of a changes only when the inequality to be removed corresponds to one of those vectors. Hence its probability is no greater than $(d+1)/(n+1)$. □

Now the theorem is proven. Assume that $|\overline{y}_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}}\overline{\boldsymbol{h}}^*| \leq \overline{a}^*$, which holds with probability greater than or equal to $1 - (d+1)/(n+1)$ by Lemma 2. Then we have

$$\begin{aligned}&|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}}\boldsymbol{h}^*|\\ \leq &|y_{k_0} - \overline{y}_{k_0}| + |\overline{y}_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}}\overline{\boldsymbol{h}}^*| + |\boldsymbol{x}_{k_0}^{\mathrm{T}}(\overline{\boldsymbol{h}}^* - \boldsymbol{h}^*)|\\ \leq &\phi + \overline{a}^* + \sum_{i=1}^{d} e_i \leq a^* + 2\phi + \sum_{i=1}^{d} e_i.\end{aligned}$$

3. DISCUSSION

The proposed identification algorithm is investigated in several aspects.

3.1. Minimality of the number of assumptions

The proposed identification algorithm relies on the assumption of approximate K-independence. It is shown here that identification is impossible after removing this assumption. In this sense, the number of assumptions is set minimum possible in the proposed identification algorithm.

The plant and the inputs are considered as before except that approximate K-independence is not assumed. This means that the class of considered plants is taken larger than before. Supposing that the input-output data $\{(y_k, \boldsymbol{x}_k)\}_{k=1}^N$ are provided, we consider two functions of them, $\widehat{\boldsymbol{h}}$ and $\widehat{a}$. Each pair of these functions can be regarded as an identification algorithm; The desired algorithm is the one that guarantees $|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}}\widehat{\boldsymbol{h}}| \leq \widehat{a}$ with high probability for a newly sampled input-output pair $(y_{k_0}, \boldsymbol{x}_{k_0})$. However, it is not possible to guarantee this for all the considered plants.

Theorem 2. *For any algorithm $(\widehat{\boldsymbol{h}}, \widehat{a})$ and any time instant $k_0 > N$, there exists a plant that establishes*

$$\left|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}}\widehat{\boldsymbol{h}}\big(\{(y_k, \boldsymbol{x}_k)\}_{k=1}^N\big)\right| \geq \widehat{a}\big(\{(y_k, \boldsymbol{x}_k)\}_{k=1}^N\big) + 1$$

with probability one.

To prove this theorem, just consider the plant having the deterministic dynamics

$$\begin{aligned}y_k = &\boldsymbol{x}_k^{\mathrm{T}}\widehat{\boldsymbol{h}}\big(\{(y_\ell, \boldsymbol{x}_\ell)\}_{\ell=1+k-k_0}^{N+k-k_0}\big)\\ &+ \widehat{a}\big(\{(y_\ell, \boldsymbol{x}_\ell)\}_{\ell=1+k-k_0}^{N+k-k_0}\big) + 1,\end{aligned}$$

which is nonlinear but time-invariant. The inequality in the theorem obviously holds for this plant.

3.2. Relaxation of the assumption

Theorem 2 states that the assumption of approximate K-independence cannot be removed. However, it does not mean that this assumption cannot be relaxed or replaced by some other one.

Let us relax this assumption by supposing that the inequality $|y_k - \overline{y}_k| \le \phi$ holds only in a probabilistic sense, that is, it holds with probability greater than or equal to $1-p$. This relaxation is practically reasonable because it is sometimes difficult to guarantee $|y_k - \overline{y}_k| \le \phi$ with 100% confidence. In order that the statement of Theorem 1 continues valid with this relaxed assumption, it suffices that there hold the $n+1$ inequalities, $|y_{(j)} - \overline{y}_{(j)}| \le \phi$, $j = 1, \dots, n$, and $|y_{k_0} - \overline{y}_{k_0}| \le \phi$. The probability with which at least one of these $n+1$ inequalities fails to hold is $(n+1)p$ at most. Hence the next theorem can be derived from Theorem 1.

Theorem 3. *There holds* $|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}} \boldsymbol{h}^*| \le a^* + 2\phi + \sum_{i=1}^{d} e_i$ *with probability greater than or equal to* $1 - (d+1)/(n+1) - (n+1)p$.

A major difference from Theorem 1 is that the confidence achieves the maximum when n is equal to $\lfloor \sqrt{(d+1)/p} \rfloor$ or $\lfloor \sqrt{(d+1)/p} \rfloor - 1$. When n is larger than this value, increase of the input-output data does not imply increase of the confidence.

3.3. Analysis of a given model

Suppose that a linear model $y_k = \boldsymbol{x}_k^{\mathrm{T}} \boldsymbol{h}^0$ has already been obtained for a plant at hand and we want to analyze the quality of this model. The leave-one-out technique used in Section 2 is applicable to this problem, too. In fact, one can obtain a better result for the present problem than for the previous identification problem since the existence of the model makes the problem easier.

With the notation of Section 2, compute $\max_{j=1,\dots,n} |y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}} \boldsymbol{h}^0|$ and regard it as a quality index of the model $y_k = \boldsymbol{x}_k^{\mathrm{T}} \boldsymbol{h}^0$. Let $(y_{k_0}, \boldsymbol{x}_{k_0})$ be a newly sampled input-output pair at $k_0 \ge N + K$. Then the next property holds.

Theorem 4. *There holds the inequality*

$$|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}} \boldsymbol{h}^0| \le \max_{j=1,\dots,n} |y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}} \boldsymbol{h}^0| + 2\phi$$

with probability greater than or equal to $1 - 1/(n+1)$. *Here the probability is measured with respect to the input-output data* $\{(y_k, \boldsymbol{x}_k)\}_{k=1}^{N}$ *and the newly sampled pair* $(y_{k_0}, \boldsymbol{x}_{k_0})$.

Proof. Computing $\max_{j=1,\dots,n} |y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}} \boldsymbol{h}^0|$ is equivalent to finding the minimum a under the conditions $|y_{(j)} - \boldsymbol{x}_{(j)}^{\mathrm{T}} \boldsymbol{h}^0| \le a$, $j = 1, \dots, n$. The difference between this minimization problem and the one considered in Section 2 is that the minimization variable is a here but $(\boldsymbol{h}, a)$ there. Hence the leave-one-out technique can be used here but the variable dimension is set to unity in place of $d+1$. This proves the above theorem. □

Comparing this theorem to Theorem 1, one can see that the result obtained here is better than the previous one. Namely, the upper bound of $|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}} \boldsymbol{h}^0|$ does not include the term $\sum_{i=1}^{d} e_i$; The result holds with more confidence $1 - 1/(n+1)$; The computational load to obtain e_i is eliminated. Hence the application here is considered to be more useful than the previous one in many cases. Note also that the relaxation of Section 3.2 is possible in this application, too.

3.4. Assumption of a mixing property

We assumed approximate K-independence in order to apply the leave-one-out technique to a dependent stochastic process. In the literature, a mixing property is often assumed for the same purpose. See Campi and Kumar (1998), Weyer and Campi (1999), and Weyer (2000) for its use in system identification; See Karandikar and Vidyasagar (2001) and Vidyasagar and Karandikar (2001) for its application in learning theory. In this subsection, we derive a result similar to the one in Section 2 by assuming a mixing property instead of approximate K-independence.

Let Q be the probability measure on the stochastic process $\{(y_k, \boldsymbol{x}_k)\}_{k=-\infty}^{\infty}$ and $Q_{-\infty}^{0}$ and Q_1^{∞} be its semi-infinite marginals. With σ_K being the σ-algebra generated by the random variables $(y_k, \boldsymbol{x}_k)$, $k \le 0$ or $k \ge K$, define the function

$$\beta(K) := \sup_{A \in \sigma_K} \left| Q(A) - (Q_{-\infty}^{0} \times Q_1^{\infty})(A) \right|.$$

If $\beta(K)$ converges to zero as $K \to \infty$, the process $\{(y_k, \boldsymbol{x}_k)\}_{k=-\infty}^{\infty}$ is said to be *β-mixing* or *absolutely regular* (Nobel and Dembo, 1993; Karandikar and Vidyasagar, 2001; Vidyasagar and Karandikar, 2001). Let us assume the β-mixing property instead of approximate K-independence. The rest of the setup and the identification procedure are the same as in Section 2. Here we have the following.

Theorem 5. *There holds* $|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}} \boldsymbol{h}^*| \le a^*$ *with probability greater than or equal to* $1 - (d+1)/(n+1) - (n+1)\beta(K)$.

Proof. Due to Lemma 2 of Nobel and Dembo (1993) (also see Karandikar and Vidyasagar (2001) and Vidyasagar and Karandikar (2001)), there holds the inequality

$$\Big| Q\big(|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}} \boldsymbol{h}^*| \le a^*\big) - (Q_0)^{n+1}\big(|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}} \boldsymbol{h}^*| \le a^*\big) \Big| \le (n+1)\beta(K),$$

where Q_0 is the marginal distribution of one input-output pair $(y_k, \boldsymbol{x}_k)$ and $(Q_0)^{n+1}$ is its $(n+1)$-fold product. With the measure $(Q_0)^{n+1}$, the input-

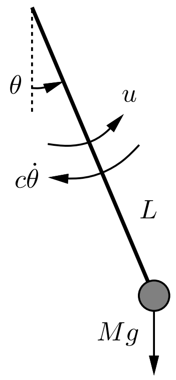

Figure 1. Identified system

output pairs $\{(y_{(j)}, \boldsymbol{x}_{(j)})\}_{j=1}^{n}$ and $(y_{k_0}, \boldsymbol{x}_{k_0})$ are statistically independent. Hence one can use the technique of Lemma 2 and obtains

$$(Q_0)^{n+1}\big(|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}}\boldsymbol{h}^*| \leq a^*\big) \geq 1 - (d+1)/(n+1).$$

This ensures the theorem. □

The present result is advantageous to the one in Section 2 because the bound of modeling error does not include the terms $2\phi + \sum_{i=1}^{d} e_i$. The confidence seems to have a term proportional to n just like the result in Section 3.2. However, since $\beta(K) \to 0$ as $K \to \infty$, one can make the confidence arbitrarily close to unity as the number of input-output data, N, grows. A disadvantage of the present approach is that the definition of $\beta(K)$ is abstract and it is not obvious how one can estimate its value for a provided actual plant. This is the reason why we mainly propose the approach of Section 2.

4. EXAMPLE

The proposed identification algorithm is applied to a simple example plant.

Consider a pendulum shown in Figure 1, whose dynamics is described by $ML^2\ddot{\theta} = -MgL\sin\theta - c\dot{\theta} + u$. Here, the mass is taken as $M = 1\,\mathrm{kg}$, the length of the arm as $L = 1\,\mathrm{m}$, the coefficient of the friction as $c = 2\,\mathrm{Nms/rad}$, and the acceleration of the gravity as $g = 9.8\,\mathrm{m/s^2}$. The torque $u(t)$[Nm] is provided as $u(t) = u_k$ for $k \leq t < k+1$, where the input sequence $\{u_k\}$ is a sequence of independent random numbers that distribute uniformly in $[-1, 1]$. The output is the angle $\theta(t)$[rad] sampled at a discrete time $t = k$ and is denoted by y_k. Based on a preliminary experiment we assume approximate K-independence with $K = 6$ and $\phi = 0.001$. In fact, one can theoretically confirm that the true plant dynamics satisfies this assumption.

We use the one-step-ahead predictor

$$\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{h} := [u_{k-1} \;\; u_{k-2} \;\; u_{k-3}][h_1 \;\; h_2 \;\; h_3]^{\mathrm{T}}$$

associated with the third-order FIR model in order to approximate y_k. The number of input-output data, N, is chosen as 2400, which implies $n = 400$. The computational time to solve the associated linear programming problems is 1.17 s with Pentium II 450 MHz and 128 MByte memory. As a result, we have $\boldsymbol{h}^* = [0.137 \;\; -0.0462 \;\; 0.0145]^{\mathrm{T}}$, $a^* = 0.00556$, $e_1 = 0.00349$, $e_2 = 0.00375$, and $e_3 = 0.00387$. By Theorem 1, the inequality $|y_{k_0} - \boldsymbol{x}_{k_0}^{\mathrm{T}}\boldsymbol{h}^*| \leq 0.0187$ holds with probability greater than 99.0%.

5. CONCLUSION

In this paper, we saw how one can identify a provided plant in the worst-case sense using the leave-one-out technique. The proposed identification algorithm gives a nominal model and its uncertainty bound with a non-asymptotic probabilistic guarantee. The number of assumptions is kept small in order to reduce a risk to make too optimistic assumptions. Generalization of the algorithm is possible in several directions. The example in Section 4 suggests that the proposed algorithm gives a meaningful result with a reasonable number of input-output data. This is important because many existing learning-theoretic identification algorithms often require a large number of data beyond the practical level.

ACKNOWLEDGEMENT

The authors are thankful to Dr. M. Vidyasagar, who acquainted them with the β-mixing property.

REFERENCES

Bartlett, P. L., and S. R. Kulkarni (1998). The complexity of model classes, and smoothing noisy data. *Systems & Control Letters*, **34**, 133–140.

Campi, M. C., and P. R. Kumar (1998). Learning dynamical systems in a stationary environment. *Systems & Control Letters*, **34**, 125–132.

De Vries, D. K., and P. M. J. Van den Hof (1995). Quantification of uncertainty in transfer function estimation: a mixed probabilistic-worst-case approach. *Automatica*, **31**, 543–557.

Goodwin, G. C., M. Gevers, and B. Ninness (1992). Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Transactions on Auto-*

matic Control, **37**, 913–928.

Helmicki, A. J., C. A. Jacobson, and C. N. Nett (1991). Control oriented system identification: a worst-case/deterministic approach in H^∞. *IEEE Transactions on Automatic Control*, **36**, 1163–1176.

Karandikar, R. L., and M. Vidyasagar (2001). Rates of uniform convergence of empirical means with mixing processes. Submitted for publication.

Khargonekar, P., and A. Tikku (1996). Randomized algorithms for robust control analysis and synthesis have polynomial complexity. *Proceedings of the 35th IEEE Conference on Decision and Control*, Kobe, Japan (pp. 3470–3475).

Milanese, M., and A. Vicino (1991). Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, **27**, 997–1009.

Nobel, A., and A. Dembo (1993). A note on uniform laws of averages for dependent processes. *Statistics & Probability Letters*, **17**, 169–172.

Oishi, Y., and H. Kimura (2001). A learning-theoretic approach to model-set identification. *Asian Journal of Control*, **3**, 10–20.

Tempo, R., E. W. Bai, and F. Dabbene (1997). Probabilistic robustness analysis: explicit bounds for the minimum number of samples. *Systems & Control Letters*, **30**, 237–242.

Tse, D. N. C., M. A. Dahleh, and J. N. Tsitsiklis (1993). Optimal asymptotic identification under bounded disturbances. *IEEE Transactions on Automatic Control*, **38**, 1176–1190.

Vapnik, V. N. (1998). *Statistical Learning Theory*. John Wiley & Sons, New York, NY.

Vidyasagar, M. (2001). Randomized algorithms for robust controller synthesis using statistical learning theory. *Automatica*, **37**, 1515–1528.

Vidyasagar, M., and R. L. Karandikar (2001). On the behaviour of learning algorithms with β-mixing process inputs. Submitted for publication.

Weyer, E. (2000). Finite sample properties of system identification of ARX models under mixing conditions. *Automatica*, **36**, 1291–1299.

Weyer, E., and M. C. Campi (1999). Finite sample properties of system identification methods. *Proceedings of the 38th IEEE Conference on Decision and Control*, Phoenix, AZ (pp. 510–515).

Weyer, E., and M. C. Campi (2000). Non-asymptotic confidence ellipsoids for the least squares estimate. *Proceedings of the 39th IEEE Conference on Decision and Control*, Sydney, Australia (pp. 2688–2693).

Weyer, E., R. C. Williamson, and I. M. Y. Mareels (1999). Finite sample properties of linear model identification. *IEEE Transactions on Automatic Control*, **44**, 1370–1383.

Zhou, T., and H. Kimura (1995). Structure of model uncertainty for a weakly corrupted plant. *IEEE Transactions on Automatic Control*, **40**, 639–655.

Zhu, Y. C. (1989). Estimation of transfer functions: asymptotic theory and a bound of model uncertainty. *International Journal of Control*, **49**, 2241–2258.