A THEORY OF ε-APPROXIMATION OF A CLASS OF SYSTEMS BASED ON ε-ENTROPY THEORY*

By

SH. KONDO

Department of Communication. Faculty of Engineering, Tokai University

Received October 22, 1978

1. Introduction

System identification is one of the most important problem not only in control engineering but also in information engineering. For example, let us consider a pattern recognition system such as man. We may know the inputs and the corresponding outputs of the system, but we cannot know how to recognize the visual system.

An input–output system can be represented by an operator from an input space into an output space. Let X, Y and A be input space, output

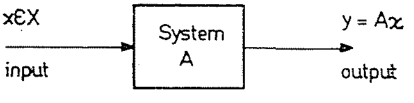

Fig. 1

space and operator from X into Y, respectively. Then the output y of the system A input x is denoted by $y = Ax$ (Fig. 1). If the given system is a communication system, then system A is called a "channel".

The fundamental problem of communication theory is to determine reliably the input x from the information about the channel A and its output $y = Ax$. On the contrary, the fundamental problem of system identification is to determine reliably the system from the information about some inputs $x_1, \dots x_n$ and the corresponding outputs $y_1 = Ax_1, \dots, y_n = Ax_n$. In the next section it will be seen that the system identification problem can be reduced to a communication problem.

* Submitted at the Joint Symposium Technical University, Budapest — Tokai University, 23—24 November, 1977.

2. Finite-shot channel

Here, we assume output space Y to be a vector space. Then the set $m(X,Y)$ of all operators from X into Y will be a vector space with the following addition and multiplication by scalar:

$$(A+B)(x) = A(x) + B(x) \tag{1}$$

$$(\lambda A)(x) = \lambda A(x), \quad A, B \in m(X, Y), \ x \in X, \tag{2}$$

$$\lambda\text{: scalar}$$

Let x be a fixed element of X. Define an operator Φ_x from $m(X,Y)$ into Y as:

$$\Phi_x(A) = A(x), \ A \in m(X, Y) \tag{3}$$

Regarding this operator as a communication channel, then $m(X,Y)$ will be an input space for the channel Φ_x, and an element of $m(X,Y)$, that is, a system will be an input signal. In other words, an input for an unknown system A is a channel to obtain the information about the system. Therefore, the problem of system identification is to determine reliably the input system

Fig. 2

from the information about the channel Φ_x and its output $\Phi_x(A) = A(x)$ (Fig. 2).

Let us call this channel Φ_x a "*one-shot channel*". Similarly, for some inputs $x_1, \ldots, x_N$, we can define "*N-shot channel*" an operator $\Phi_{x_1,\ldots,x_N}$ from $m(X,Y)$ into Y^N as:

$$\Phi_{x_1,\ldots,x_N}(A) = (\Phi_{x_1}(A), \ldots, \Phi_{x_N}(A)) \tag{4}$$

If N is finite, N-shot channel is called a finite-shot channel. Though each element of $m(X,Y)$ is not always a linear operator on X (here, we must assume X to be a vector space), a one-shot channel is always linear on $m(X,Y)$, since,

$$\Phi_{x_N}(A+B) = (A+B)(x) = (A(x) + B(x) = \Phi_x(A) + \Phi_x(B) \tag{5}$$

$$\Phi_x(\lambda A) = (\lambda A)(x) = \lambda A(x) = \lambda \Phi_x(A) \,. \tag{6}$$

Similarly, a finite-shot channel is also linear on $m(X,Y)$:

$$\begin{aligned}\Phi_{x_1,\ldots,x_N}((A+B) &= (\Phi x_1(A+B),\ldots,\ \Phi x_N(A+B)) \\ &= (\Phi x_1(A)+\Phi x_1(B),\ldots,\ \Phi x_N(A)+\Phi x_N(B) = \\ &= \Phi x_1,\ldots,\ x_N(A)+\Phi x_1,\ldots,\ x_N(B)\end{aligned} \quad (7)$$

$$\Phi x_1,\ldots,x_N(\lambda A) = (\lambda\Phi x_1(A),\ldots,\ \lambda\Phi_{x_N}(A) = \lambda\Phi x_1,\ldots,\ x_N(A)\ . \quad (8)$$

Therefore, the system identification theory is always a linear channel theory.

3. System space

Assume input space X to be a metric space and output space Y to be a complete normed space, that is, a Banach space.

An operator A in $m(X,\ Y)$ is called continuous if for any $x \in X$, given $\varepsilon > 0$, there is $\delta > 0$ such that if $\varrho(x,x') < \delta$, $\|A(x) - A(x')\| < \varepsilon$. Let $c^\infty\ (X,Y)$ be subset of all continuous operators in $m(X,Y)$.

An operator A in $m(X,Y)$ is called bounded if there is a number $M > 0$ such that for any input x, $\|Ax\| \leqq M$. Let $b(X,Y)$ be a subset of all bounded operators in $m(X,Y)$, then $b(X,Y)$ is a Banach space with the norm:

$$\|A\| = \sup_{x\in X} \|Ax\| \quad (9)$$

Let $c(X,Y) = b(X,Y) \cap c^\infty(X,Y)$. Then $c(X,Y)$ is a closed subspace of $b(X,Y)$ and therefore it is also a Banach space. If X is compact, then $c(X,Y) = c^\infty(X,Y)$.

Since boundedness of the system means stability of the system, it is natural to assume the system to be bounded. Similarly, it is natural that the system is continuous.

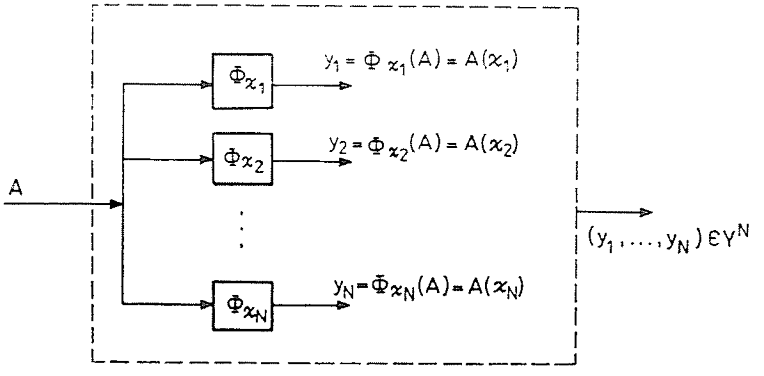

Fig. 3

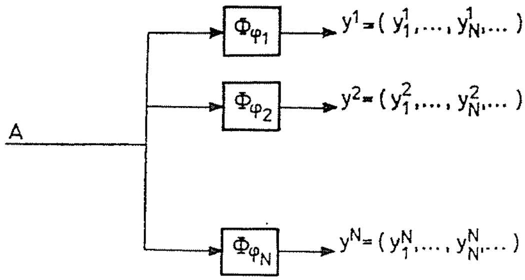

Fig. 4

Next, here, let us consider the one-shot channel Φ_x on $c(X,Y)$. Then, we have following inequality:

$$\| \Phi_x(A) \| = \| Ax \| \leq \| A \| \tag{10}$$

The inequality (10) implies that the one-shot channel Φ_x for any x is continuous. Thus we obtain the first theorem:

THEOREM 1

One-shot-channel Φ_x on $c(X,Y)$ is a continuous linear operator for any x in X.

4. ε-Decodable class of systems by finite-shot channel

Let D be a subset of the system in $c(X,Y)$. If, for some x in X, one-shot channel Φ_x is injective on D, that is, for any pair A, B in D, $Ax = Bx$ only when $A = B$, then Φ_x is invertible. Therefore, let $y = \Phi_x \cdot A$ be output of the channel, then we mathematically obtain a system $A = \Phi_x^{-1} y$.

Definition 1. Let D be a subset in $c(X,Y)$. D is *decodable class of systems* if there is some input x such that for any pair A, B in D, $Ax = Bx$ only when $A = B$.

If D is decodable, then the input system A can exactly be determined mathematically from the information $y = \Phi_x A$ for some x in X. But we dont know how to construct the inverse of Φ_x.

In practice, the following definition is more useful:

Definition 2. Subset D of systems in $c(X,Y)$ is called an *ε-decodable class of systems by finite-shot channel*, if for given $\varepsilon > 0$, there is a finite-shot channel such that approximate system $\tilde{A}$ can be constructed with $\|A - \tilde{A}\| < \varepsilon$, from the output of the finite-shot channel.

5. Construction of an approximate system

Definition 3. Subset D of systems in $c(X,Y)$ is called ***relatively compact (or totally bounded)***, if for given $\varepsilon > 0$, there are a finite number of systems $A_1, \ldots, A_N$ in D such that for any system A in D, there is A_j with $\|A - A_j\| < \varepsilon$. In this case, family of systems $\{A_1, \ldots, A_N\}$ is called ε-*net* of D.

Let $D(x)$ be the image of D by one-shot channel Φ_x. From theorem 1, we find that $D(x) = \Phi_x D$ is relatively compact. In fact, if $\{A_1, \ldots, A_N\}$ is ε-net of D, then $\{A_1 x, \ldots, A_N x\}$ is ε-net of $D(x)$, since,

$$\| Ax - A_j x \| \leq \| A - A_j \| < \varepsilon \tag{11}$$

However, inequality (11) does not imply that if $\| Ax - A_j x \| < \varepsilon$, then $\| A - A_j \| < \varepsilon$.

But this fact implies the following theorem.

THEOREM 2.

For every $\varepsilon > 0$, there is x in X such that for any pair A, B in D with $\| Ax - Bx \| < \varepsilon$, we have $\| A - B \| < \varepsilon$. Then subset D is an ε-decodable class of systems by one-shot channel. Proof: Given $\in > 0$, let x be an input satisfying the condition in the theorem. Let $\{A_1, \ldots, A_N\}$ be ε-net of D. Then $\{A_1 x, \ldots, A_N x\}$ is ε-net of $D(x)$. Now, we get the output $y = \Phi_x(A) = A(x)$ of one-shot channel Φ_x. There is A_j such that $\| y - A_j x \| = \| Ax - A_j x \| < \varepsilon$. Therefore, we can determine for input system to be A_j. Then, from the condition of the theorem, we have $\| A - A_j \| < \varepsilon$. (q.e.d.)

Next, let us consider the condition for subset D of systems to be relatively compact.

D is called *equicontinuous* on X if for any x in X, given $\varepsilon > 0$, there is $\delta > 0$ such that if $\| x - x' \| < \delta$, then we have $\| Ax - Ax' \| < \varepsilon$ for any A in D. The following lemma is called Ascoli's theorem:

Lemma 1. We assume X to be compact. Then subset D in $c(X,Y)$ is relatively compact if and only if D is equicontinuous on X and for any x in X, $D(x)$ is relatively compact in Y.

6. Schmidt class of linear systems

Now, let us confine our discussion to linear systems in $c(X,Y)$. Let $j(X,Y)$ be subset of all bounded linear operators in $c(X,Y)$. It is well-known that $j(X,Y)$ is also a Banach space with the norm:

$$\| A \| = \sup_{\| x \| = 1} \| Ax \| / \| x \| \tag{12}$$

From here, we assume that $X = Y = H$ is a Hilbert space, and we denote $j(X,Y) = J(H)$.

Let $\{\varphi_l\}_{l=1}^{\infty}$ is a complete orthonormal family of H.

Definition 4. An operator A in $j(H)$ is called Schmidt operator if $\sum_{k,l} |(A\varphi_l, \varphi_k)|^2 < \infty$ or $\sum_{l} ||A\varphi_l||^2 < \infty$.

We denote the set of all Schmidt operators by $s(H)$. $s(H)$ is called the Schmidt class of bounded linear operators. If A belongs to $s(H)$, then we have $\sum_{k,l} |(A\varphi_l, \varphi_k)|^2 = \sum_{l} ||A\varphi_l||^2$ and this value does not depend on the choice of complete orthonormal family $\{\varphi_l\}_{l=1}^{\infty}$. Let

$$(A, B) = \sum_{l} (A\varphi_l, B\varphi_l) \tag{13}$$

Then, $s(H)$ is a Hilbert space with this inner product and norm:

$$||A|| = (A, B)^{1/2} = \left(\sum_{l} ||A\varphi_l||^2\right)^{\frac{1}{2}} \tag{14}$$

Family of operators $\{\varphi_k \otimes \varphi_l\}_{k,l}$ is seen to be a complete orthonormal family of Hilbert space $s(H)$, where operator $\varphi_k \otimes \varphi_l$ is defined as:

$$(\varphi_k \otimes \varphi_l)(x) = (x, \varphi_l)\varphi_k, \quad x \in H. \tag{15}$$

Then, any system in $s(H)$ can be expressed as:

$$A = \sum_{k,l} (A\varphi_l, \varphi_x)\, \varphi_k \otimes \varphi_l \tag{16}$$

6. ε-Decoding of subset in s(H) by finite-shot channel

Let $\{\varphi_l\}_{l=1}^{\infty}$ be a complete orthonormal family of $s(H)$.

Lemma 2. Let D be a subset in $s(H)$. D is relatively compact if and only if given $\varepsilon > 0$, there is an integer number $N = N(\varepsilon)$ such that for any system A in D,

$$||A - A_N|| < \varepsilon \quad \text{or,} \quad \sum_{k \geq N+1,\, l \geq N+1} |(A\varphi_l, \varphi_k)|^2 < \varepsilon$$

where, $A_N = \sum_{k=1,\, l=1}^{N} (A\varphi_l, \varphi_k)\, \varphi_k \otimes \varphi_l$.

From this lemma, we have immediately the next theorem:

THEOREM 3.

If D in $s(H)$ is relatively compact, then D is an ε-decodable class of systems by finite-shot channel.

PROOF: Given $\varepsilon > 0$, there is $N=N(\varepsilon)$ such that $\| A - A_N \| < \varepsilon$, where $A_N = \Sigma_{k=1,\, l=1}^{N} (A\varphi_l, \varphi_k)\, \varphi_k \otimes \varphi_l$. Therefore it is sufficient to prove that the coefficients $\{(A\varphi_l, \varphi_k)\}_{k=1\, l=1}^{N}$ can be determined from the output data of the finite-shot channel. Let $x_i = \varphi_i$ $(i = 1, \ldots, N)$. Then,

$$\Phi\varphi_i(A) = A\varphi_i = (\Sigma_{k,l}(A\varphi_l, \varphi_k)\, \varphi_k \otimes \varphi_l)\, \varphi_i = \sum_k (A\varphi_i, \varphi_k)\, \varphi_k .$$

Let $y^i = \sum_k y_k^i \cdot \varphi_k$, $y_k^i = (y^i, \varphi_k)$ be output of $\Phi\varphi_i$. Then, N-components $y_1^i, \ldots, y_N^i$ are equal to $(A\varphi_i, \varphi_1), \ldots, (A\varphi_i, \varphi_N)$ respectively. Therefore from the N-shot channel $\Phi\varphi_1, \ldots, \varphi_N$, we obtain A_N (Fig. 4) (q.e.d.)

7. Example

We consider now as input and output space H, Hilbert space $L_2[-\pi,\pi]$, which consists of all square-integrable functions on $[-\pi,\pi]$ and inner product:

$$(x,y) = \int_{-\pi}^{\pi} x(t)\, y(t)\, dt . \tag{17}$$

It is well-known that $L_2[-\pi,\pi]$ has a complete orthonormal family:

$$\varphi_0(t) = 1/\sqrt{2\pi}, \quad \varphi_{2k}(t) = \frac{1}{\sqrt{\pi}} \cos k \cdot t, \quad \varphi_{2k+1}(t) = \frac{1}{\sqrt{\pi}} \sin kt$$

$$(k = 1, 2, \ldots) \tag{18}$$

and Schmidt class on $L_2[-\pi,\pi]$ equals the set of all integral operators as:

$$y(s) = (Ax)(s) = \int_{-\pi}^{\pi} k(s,t) x(t) dt \tag{19}$$

where integral kernel $k(s,t)$ satisfies the condition:

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} | k(s,t) |^2 ds dt < \infty . \tag{20}$$

Complete orthonormal family of $s(L_2[-\pi,\pi])$ is family of integral operators with kernel

$$k(s,t) = \varphi_k(s)\, \varphi_l(t) . \tag{21}$$

Summary

An input–output system can be mathematically determined by a triad (U, Y, F) where U, Y, and F are input space, output space and mapping from U to Y, respectively. Identification problem is to find out inputs and the corresponding outputs, if F is unknown — black box.

This paper presents methods to identify F within a tolerance ε, from knowledge of a finite set of input–output relations in the case where F belongs to a specified subset of the space consisting of all mappings from U to Y. These methods can be obtained from ε-approximation of the subset of mappings involving F.

References

1. DIEUDONNÉ, J.: Foundations of Modern Analysis, Academic Press, New York 1960
2. PROSSER, R. T.—ROOT, W. L.: Determinable class of channals, Jour. of Math. and Mech., Vol. 16, No. 4 (1966), pp. 365—397
3. RINGROSE, J. R.: Compact Non-self-adjoint Operators, Van Nostrand, Princeton 1971
4. KATO, T.: Perturbation Theory for Linear Operators, Springer Verlag, Berlin 1966

Shozo KONDO, Dept. of Communication, Faculty of Eng. Tokai University, Kyushu, Japan.