*Automatica*, Vol. 8, pp. 621–636. Pergamon Press, 1972. Printed in Great Britain.

# System Identification via State Characterization*†

## Identification de système par caractérisation d'état

## Systemidentifikation über die Zustandscharakterisierung

## Идентификация системы путем характеризации состояния

E. MARK GOLD‡

*"State characterization" is a general, nonparametric method for designing experiments which will directly determine a state space representation for an unknown black box.*

**Summary**—Arbib and Zeiger's generalization of Ho's algorithm for system identification is presented from an alternative viewpoint called "state characterization". This is an essentially nonlinear, nonparametric method for designing experiments which will directly determine a state space representation for an unknown black box. This method assumes (1) a finite dimensional state space is possible, (2) the unknown black box can be reset to its initial state at will, so that an arbitrary set of experiments can be performed, and (3) there is no plant noise, so that applying given inputs to the initial state will always yield the same state, so various experiments can be performed on it.

In the initial, informal discussion the properties of this method of system identification are compared to the properties of the method of parameter estimation via loss function minimization, e.g. least square error or maximum likelihood, which is the only general approach to system identification presently available. Loss function minimization will utilize any data, but it is computationally difficult and essentially parametric. State characterization eliminates the computational requirement at the expense of requiring specific data, and it is essentially nonparametric. It is proposed that state characterization may have practical application in determining an approximate, low order description of a complex system about which we have little prior information, for instance in the social sciences and medicine.

The formal presentation is limited to finite state automata and discrete time linear systems. The situation is considered in which the order of the unknown black box, the minimum size state space required to represent it, is not known *a priori*; so it is necessary to continue experimenting indefinitely, using the results to obtain successively better descriptions of the unknown black box. Detailed algorithms are presented for obtaining either a Moore model or a Mealy model description. Branching rules for using past data to choose subsequent experiments in order to hasten convergence are presented. Known theorems are used to prove that after some finite time these algorithms will yield a correct description, and new theorems show that the subsequent representations will be invariant after a correct description is obtained—although there is no way for the experimenter to know when he has obtained a correct description.

* Received 19 November 1971; revised 4 April 1972. The original version of this paper was not presented at any IFAC meeting. It was recommended for publication in revised form by Associate Editor B. Gaines.

† Supported in part by the U.S. Air Force Office of Scientific Research, Contract F44620-C-0018, Information Sciences Directorate. Also, in part by, General Research Support Grant No. 05612, National Institute of Mental Health, Public Health Service.

‡ Departemenr d'Informatique, University of Montreal.

### 1. INTRODUCTION

ARBIB and ZEIGER [2] described a generalization of the system identification algorithm proposed by HO [14] in unpublished form and described in the open literature by ZEIGER [23]. This general approach to system identification, here called "state characterization", is described from a different viewpoint in this paper. The formal presentation is limited to the cases of identifying a finite automaton or a discrete time linear system, called a linear automaton, which were discussed by ARBIB and ZEIGER [2] in greatest detail, in order to facilitate comparison. The two viewpoints are compared in the Conclusions section.

Although the scope of the formal presentation is quite limited and abstract, I hope to make it clear in these initial two informal sections that I feel that the state characterization approach to system identification does have practical applications. Indeed, I originally proposed this approach independently in unpublished form as a means for determining the dynamics of memorization [11].

State characterization consists of choosing a sample of the states of the unknown system, choosing a set of experiments to be performed on each state, and using the results of the experiments to classify the states of the system. That is, a small number of experiments is sought, the results of which characterize the states of the system in the sense that any two states which can give different results for some experiment will give different results for at least one of the characterizing experiments. Once the states of the system are characterized in this manner they are essentially observable. Therefore, experiments can be performed which will determine the state transition and output functions directly.

In order to be able to apply this method of system identification, it is necessary to be able to perform a number of experiments on any reachable state, in order to characterize it and its transitions resulting from various possible inputs. This will be possible if we can repeatedly put the system into any reachable state; which, in turn, will be possible if (1) the system is "resettable", that is, either it can be reset to the initial state or we have available many copies of the system in the initial state, and (2) there is no plant noise, i.e. the state transitions are deterministic, so that every time we apply a given input sequence to the initial state the same state will result.

At present we have available only one general method of system identification, namely parameter estimation via loss function minimization. This method of system identification consists of determining the values of the parameters which best fit the available data, with "fit" defined, e.g. by least square, maximum likelihood, or Bayes' rule. Recent reviews of system identification, by ÅSTRÖM and EYKHOFF [3] and SAGE and MELSA [20], are concerned solely with loss function minimization, and what are called alternative methods of system identification usually turn out to be alternative choices of loss function or alternative computational techniques for finding the values of the parameters which minimize the loss function.

In order to determine the area of practical application of the state characterization method of system identification, it is necessary to compare its properties to those of loss function minimization: The method of loss function minimization can be applied to any type of system and can utilize any type of data. Its primary disadvantage is difficulty of computation, that is, determining the parameter values which minimize the chosen loss function. Also, it is a parametric approach, requiring a parametric model. State characterization makes it possible to design experiments which determine the state transition and output functions of the unknown system directly by using an appropriate representation for the states. Thus, state characterization eliminates the computational problem at the expense of requiring specific data. Also, state characterization is limited to resettable systems without plant noise. Given the order of the system, the state characterization approach is nonparametric, assuming that the problem of extrapolating the state transition and output functions from a finite number of observations of their values is considered nonparametric.

In both approaches the order of the system is a parameter which must be determined by sequential hypothesis testing. The notion of state characterization suggests experiments to be performed which are particularly informative concerning the order of the system, however this point will not be pursued in this paper.

Thus, it is only in cases in which there is little prior information about the structure of the system being studied that the state characterization approach may be preferable to loss function minimization. Another practical application for state characterization is in determining a low order approximation for a complex system, whether or not there is prior information concerning its structure. Namely, state characterization suggests that if the states of the system are of high dimensionality, but we wish to approximately characterize them by means of a small number of characteristics, then the problem is to find a small number of characteristics which reduce, as much as possible, the uncertainty concerning the outcomes of the various experiments which can be performed on an unknown state. This approach makes it possible to utilize the body of practical classificatory techniques which have been developed in the field of multivariate analysis.

These two considerations which favor the state characterization approach, namely little prior information and a complex state space, indicate that practical applications are likely to be found in the social and biological sciences. Indeed, it is just in these fields that classificatory techniques have been developed as means for studying stationary objects. I propose that these same techniques can be applied to the study of dynamic systems.

One of the referees has pointed out that there are precedents for the two basic notions of the state characterization approach in the automata theory literature, e.g. KOHAVI [16]. The notion of designing experiments to directly determine a state space representation for an unknown black box is called "machine identification" by Kohavi. The notion of a set of experiments which characterize the states corresponds to the notion of a "distinguishing experiment" in Kohavi's text. Whereas characterizing experiments are to be performed in parallel on copies of the black box in the same state, a "distinguishing experiment" consists of a sequence of experiments to be performed one after the other on a single copy of the black box. A sequential experiment of this type is certainly preferable if it exists, but Kohavi shows examples of simple automata for which a distinguishing experiment of this type does not exist.

## 2. BASIC IDEA OF STATE CHARACTERIZATION

The state characterization approach to system identification is based on 3 concepts: (1) the notion of a state space description inherent in abstract system theory and automata theory, (2) the well known notion of automata theory that a representation of a system can be constructed using the set of input–output functions of the reachable states for a state space, and (3) the notion of classifying objects according to their measurable properties which is studied in multivariate analysis.

A state space description of a system consists of the following: (1) an abstract set to serve as *state space*; (2) a *state transition function* which determines the next state as a function of the current state and input; (3) an *output function* which specifies the currently observed output as a function of the current, unobservable, state in the case of a Moore model, or as a function of the last state and input in the case of a Mealy model; and (4) an *initial state*, in which the system starts. Since the inputs and outputs are observable, if the states were observable we could design experiments which would determine the system. Namely, the initial state could be directly observed; by applying various input sequences to the system starting from its initial state we could determine the set of reachable states and use this as state space; and we could put the system into various states and observe the output and the next state arising from the application of various possible single input elements in the case of discrete time, or the rate of change of state in the case of continuous time, thereby determining the output and state transition functions.

The object of the state characterization approach is to define the states in terms of observable properties.

The second principle referred to at the beginning of this section is that state space can be taken to consist of the set of *achievable input–output functions*, that is, the input–output functions of the reachable states. These will be called the *black box states* in the following formal presentation, because the set of achievable input–output functions can be defined directly in terms of the initial input–output function of a black box without appealing to the intermediary abstract construct of a state.

Using the set of achievable input–output functions as state space, the states are almost observable: an input–output function is specified by determining the result of every possible experiment on the system when it is in this state, that is, the last output of the system which will result from the application of an arbitrary input string, or input function in the case of continuous time. Assuming that the means are available for putting the unknown system repeatedly into this same state, the result of any experiment of this type is observable. The only trouble is that it is necessary to perform an infinite number of such experiments to determine an input–output function.

Thus we are interested in studying a set of objects, namely the achievable input–output functions of the unknown black box, each of which is identified with the results of the experiments which can be performed on it. The main point I wish to make is that there is presently available a large body of research into this problem: the field of classification theory which has been developed as part of the multivariate analytic techniques used in the biological and social sciences.

For instance, suppose that a psychologist wishes to study the personality characteristics of a population. He will choose a large number of tests to be performed on the individuals of the population. He will perform this battery of tests on a sample of the population and then try to find a small number of tests or, more generally, functions of the test results, which characterize the individual. That is, any two individuals which give the same results on these characterizing tests will have the same characteristics, i.e. they will give the same results on all tests.

I maintain that this approach which has been successfully applied to the study of populations of stationary objects can also be applied to the study of a dynamic system, or a population of dynamic systems, a topic which will not be pursued further in this paper. Namely, various input strings can be applied to an unknown black box, assuming that it can be reset to its initial state at will. This provides a population of objects to be studied, namely the achievable input–output functions. A battery of experiments can be chosen to be performed on each of a sample of achievable input–output functions. The results can be expressed in a "data matrix", the starting point of classification theory, which shows the result of each experiment on each object. Now the objective is to find a small number of experiments which characterize the achievable input–output functions, i.e. distinguish different achievable input–output functions. The number of experiments required will be the order of the resulting representation of the unknown black box, i.e. the dimensionality of the state space which will be used in constructing the representation.

Once the set of achievable input–output functions has been characterized in this experimentally determinable manner, the state transition and output functions can be observed directly: A sample of input strings is chosen to be applied to the unknown black box in its initial state in order to yield a sample of its achievable input–output functions. On each of these the following 3 sets of experiments is performed in order to determine values of the state transition and output functions: (1) The characterizing experiments are performed on it to determine what state it is according to the chosen method for naming states. (2) The last output resulting from the input string used to reach this state is observed in order to provide a value of the output function, that is if a Moore model is desired.

(3) Each input element of some chosen sample is applied in turn to the state being studied and the resulting state is characterized by performing the characterizing experiments, thereby yielding values of the state transition function.

Only the simplest notions of classification theory will be used to construct the automaton identification algorithms investigated in the formal sections of this paper. However, I propose that it would be worthwhile to consider each of the repertoire of techniques of classification theory for possible application to state characterization, for instance principal components analysis, factor analysis, conanical variables e.g. ANDERSON [1], KENDALL and STUART [17], CATTELL [4], and clustering analysis, e.g. SCOTT and SYMONS [21], WISHART [22], GOWER [12], FRIEDMAN and RUBIN [5]; in particular principal components analysis. For instance, the discussion of this paper is limited to systems with no plant noise, so that it is possible to repeatedly put the unknown system into the same state in order to be able to perform various experiments on that state. Classification theory has extensively considered the problem of characterizing objects by means of the results of noisy experiments. This may provide an approach to the problem of identifying an unknown system with plant noise, so that the operation of trying to repeatedly put the system into the same state is noisy. Also, techniques such as principal components analysis provide an approach to the problem of finding an approximate low order description, i.e. having a state space of low dimensionality, for a complex system, i.e. a system with a state space of high dimensionality. Namely, a small number of state characterizers, the functions of results of experiments, would be sought which would reduce the uncertainty in the outcome of experiments on unknown states as much as possible.

## 3. ALTERNATIVE SYSTEM IDENTIFICATION PROBLEMS

The formal presentation which comprises the remainder of this paper will be restricted to discrete time deterministic systems, referred to as "automata".

*Finite identification* vs *identification in the limit*

In the *finite identification of a black box* we are only allowed to make a finite number of experiments on the unknown black box, the choice of the experiments to be made and the number of experiments possibly depending on past results. After some finite number of experiments we are expected to say, "OK, that's enough information; here is a description of an automaton which has the same input–output function as the unknown black box". It was an early result of the theory of finite state automata that if the unknown black box is an arbitrary finite state automaton, then finite identification is impossible, MOORE [18] and GILL [7].

I have introduced elsewhere [9, 10], the notion of *black box identification in the limit:* We are allowed to go on experimenting on the unknown black box indefinitely. At each stage in the experimenting process we use the results of the experiments performed so far to guess the identity of the black box by constructing an automaton which is meant to represent it. The problem is to construct an algorithm for choosing the experiments and constructing the guesses so that the guesses will necessarily be correct after some finite time and will not change thereafter. I have shown that finite automata *are* identifiable in the limit. Obviously, we have no way of knowing when our guesses are correct, or finite identification would be possible.

*Resettable* vs *nonresettable black boxes*

If the problem is the *identification of a resettable black box*, we are allowed to reset the black box to its initial state before performing each successive experiment, so that any finite set of experiments can be performed. If the problem is the *identification of a nonresettable black box*, our experiments are to be performed sequentially. That is, the state of the black box at the beginning of each experiment is to be the state it was left in at the end of the last experiment.

*Results*

It is obvious that if we are presented with a *resettable* black box which is realizable by a finite automaton, then we can identify it in the limit by enumeration: Namely, we list all possible finite automata, since there are only a countable number of them; we perform every possible experiment in succession; after each experiment we cross out the finite automata from our list which give a different result for one of the experiments performed so far than was observed; and we guess the unknown black box at any time to be the first one of our list which has not been crossed off.

I have shown elsewhere, [9, 10], that *nonresettable* black boxes realizable by finite automata are identifiable in the limit by means of such enumerative techniques if the experiments are properly chosen.

ARBIB and ZEIGER [2] consider the problem of finite identification of a resettable black box which is realizable either by a finite automaton with the maximum number of states specified or by a linear automaton with a specified upper limit on the dimensionality of state space. Finite identification is now possible. For instance, in the case of finite automata, there are only a finite number of possible finite automata with a specified number of states, so they can be distinguished by means of a finite number of experiments. Arbib and Zeiger discuss Ho's algorithm for the finite identification of a resettable linear automaton with a specified maximum order, the state space dimensionality.

In this paper I use the notion of "state characterization" to construct algorithms for the identification in the limit of a resettable black box which is realizable either by a finite automaton or by a linear automaton in the case that the size of the state space is not specified *a priori.* I also describe the modification of these algorithms to give finite identification if the size of state space is specified *a priori.*

The question arises, if enumeration provides an identification in the limit algorithm for *nonresettable* black boxes, why introduce the state characterization algorithm for the identification in the limit of *resettable* black boxes? The answer is that the state characterization algorithm is much simpler computationally. The enumeration algorithm requires us to construct each possible realization of the black box and check to see if it would have given the same result on the experiments performed so far as were observed, until an automaton is found which does agree with our data. State characterization, on the other hand, specifies the experiments to be performed in such a way that data is obtained which allows us to construct an automaton directly. The enumeration algorithm, however, has the advantage that it does not specify the experiments to be performed; it will use any data it is given.

Note that "identification by enumeration" is the analogue of loss function minimization for discrete, deterministic systems. In particular, it is the analogue of the nonparametric use of Bayes' rule, which has had no practical application that I know of, presumably due to computational difficulties.

## 4. TERMINOLOGY AND NOTATION

### *Black boxes*

*Time* is taken to be discrete: $t=1, 2, \ldots$. A *black box* $B$ consists of an *input set* $U$, an *output set* $Y$, and an *input–output function* $b$ defined below. $\Sigma U$ will denote the set of finite strings of input elements and $U^*$ will denote $\Sigma U$ with the *null string* $\phi$ adjoined. $\bar{u}$ and $\bar{v}$ will denote an element of $\Sigma U$ or of $U^*$. The *input–output function* $b$ specifies the behavior of the black box by specifying the output $y_t=b(\bar{u})$ of the black box at time $t$ if the $t$ inputs $\bar{u}=u_1 \ldots u_t$ are applied to $B$. That is, $b$ is a function from $\Sigma U$ to $Y$.

If an input string $\bar{u}$ is first applied to $B$, it is defined what output will be obtained if any string $\bar{v}$ is then applied to it, namely, $b(\bar{u}\bar{v})$. That is, applying $\bar{u}$ to $B$ yields a new black box with the same input set $U$ and output set $Y$ and a new input–output function which will be denoted $\Delta_{\bar{u}}b$. The new input–output function is defined by

$$\Delta_{\bar{u}}b(\bar{v})=b(\bar{u}\bar{v}) \tag{1}$$

$\Delta_{\bar{u}}b$ will be called the *state of $B$ after applying $\bar{u}$.* The *initial state* of $B$ will signify $\Delta_{\phi}b=b$.

### *Experiments on black boxes*

An *experiment* $E_{\bar{v}}$ is specified by specifying an input string $\bar{v}\varepsilon\Sigma U$. The experiment consists of applying $\bar{v}$ to a black box and observing the last output. The *result of the experiment* will be designated $E_{\bar{v}}(b)=b(\bar{v})$. Note that $E_{\bar{v}}$ can be performed on any state $\Delta_{\bar{u}}b$ of $B$. The result will be

$$E_{\bar{v}}(\Delta_{\bar{u}}b)=b(\bar{u}\bar{v})\varepsilon Y. \tag{2}$$

### *State characterization*

A state of a black box is an input–output function, taking input strings into outputs. It is completely defined by the results of all experiments $E_{\bar{v}}$. The basic principle of state characterization is to find a finite set of experiments which distinguish the states of $B$. That is, any two states of $B$ which are different will give different results on at least one of this finite set of experiments.

A *set of state characteristics* will signify a set of $\bar{E}$ of experiments. Furthermore, a set of state characteristics is assumed to be finite and to be given a linear ordering $\bar{E}=\langle E_{\bar{v}_1}, \ldots, E_{\bar{v}_N}\rangle$, so that $\bar{E}$ assigns a *representation* to each state of $B$: State $\Delta_{\bar{u}}b$ is represented by the $N$-tuple $\langle y_1, \ldots, y_N\rangle$ where

$$y_i=E_{\bar{v}_i}(\Delta_{\bar{u}}b)\varepsilon Y. \tag{3}$$

A set of state characteristics will be said to *completely characterize* the states of $B$ if different states have different representations.

In practice, in order to find the representation of some state we must perform several experiments on the state. However, once one of the experiments is performed, the state is changed. Therefore it is assumed throughout this paper that the black box we are trying to identify is *resettable* to the initial state. State $\Delta_{\bar{u}}b$ is represented by $\langle y_1, \ldots, y_N\rangle$ where

$$y_n=b(\bar{u}\bar{v}_n)=E_{\bar{u}\bar{v}_n}(\Delta_{\phi}b). \tag{4}$$

Thus, if the unknown black box is resettable we can determine the representation of any state $\Delta_{\bar{u}}b$ by performing experiments $E_{\bar{u}\bar{v}_n}$, $n=1, \ldots, N$, on the initial state.

In summary, a finite, ordered, set of state characteristics $\bar{E}$ assigns a name, the representation, to each state of $B$. If $B$ is resettable, then the representation of any of its states is experimentally determinable by means of a finite set of experiments. If $\bar{E}$ completely characterizes the states of $B$, then different states are assigned different representations.

*Automata*

An automaton signifies a deterministic, discrete time system. An *automaton A* consists of an input set $U$, an output set $Y$, a state set $X$, an initial state $x_o$, a state transition function $f_{tr}$, and an output function $f_{out}$. The *state transition function* specifies the current state as a function of the last state and the current input:

$$x_t = f_{tr}(x_{t-1}, u_t). \tag{5}$$

In a *Mealy model automaton* the *output function* specifies the current output $y_t$ as a function of the current input $u_t$ and the last state $x_{t-1}$:

$$y_t = f_{out}(x_{t-1}, u_t) \tag{6}$$

in a *Moore model automaton* the *output function* specifies the current output as a function of the current state:

$$y_t = f_{out}(x_t). \tag{7}$$

Note that the notion of state is quite different in the case of a black box on the one hand, where it signifies an input–output function, and the case of an automaton on the other hand, where it is an element of an undefined set $X$.

*Black box representation*

An automaton $A$ will be said to be a *representation of a black box B* if $A$ has the same input and output set as $B$ and the input–output function of $A$ is the input–output function $b$ of $B$. The *input-output function of A* is defined to be the function which assigns to each input string $\bar{u} = u_1 \ldots u_t$ the last output $y_t$ produced by the automaton.

The problem considered in this paper is that of *black box identification:* An algorithm is sought which will specify a finite number of experiments to be performed on a resettable black box and will then use the results of these experiments to derive a representation of the black box.

Since any representation $A$ of $B$ must have the same input and output sets, the first problem in constructing $A$ is to choose a state set $X$. In the case of *black box identification by state characterization* this is done by choosing a finite string $\bar{E}$ of experiments to characterize the states of $B$. If $\bar{E}$ is of length $N$, then the state set of $A$ is taken to be the set of $N$-tuples of output elements. That is $X = Y^N$. Or, $X$ might be taken to be the subset of $Y^N$ consisting only of the representations of the states of $B$. Note that some elements of $Y^N$ may not be representations of any state of $B$.

## 5. BASIC IDEA OF STATE CHARACTERIZATION, CONTINUED

*Black box identification in the limit by state characterization*

It will be assumed that either the unknown black box $B$ is a *finite state black box*, that is it can be represented by means of a *finite automaton*, an automaton with a discrete state space containing a finite number of elements; or else $B$ is a *linear black box*, that is it can be represented by a *linear automaton*, a discrete time linear system with a finite dimensional state space. The size the state space required to represent $B$ is not specified *a priori*. By means of state characterization it is possible to *identify B in the limit* in the following sense: The results of successively larger finite sets of experiments are used to construct automata which are successively better approximations to $B$. After some finite time the derived automaton will be a correct representation for $B$ and will not change after that. That is, not only will $B$ be described correctly after a finite time, but it will also be described in the same way after that. However, it is not possible to know when the derived automaton is correct, so experimentation must be continued indefinitely in order to insure that a correct representation for $B$ is eventually obtained.

Let $U_1^* \subset U_2^* \subset \ldots$ and $V_1^* \subset V_2^* \subset \ldots$ be two infinite, ascending sequences of sets of input strings, each set of finite cardinality, such that each sequence contains all input strings in the limit. The elements of each $V_k^*$ are to be linearly ordered in some way which is consistent in the following sense: The initial elements of $V_{k+1}^*$ are the elements of $V_k^*$ in the same order.

At the $k$-th stage of the identification in the limit procedure, consider the finite set of states $\Delta_{\bar{u}} b$ with $\bar{u}$ ranging over $U_k^*$. Characterize the states i.e. find their representations, by means of the experiments $E_{\bar{v}}$ with $\bar{v}$ ranging over $V_k^*$. For each of the states considered, find its transitions and output, thereby constructing an automaton $A_k$ which approximately realizes $B$.

The ordering chosen for the elements of $V_k^*$ will only affect the method of representing the states of $B$. Eventually $U_k^*$ will be large enough so that the set of states $\Delta_{\bar{u}} b$, $\bar{u} \varepsilon U_k^*$, will include all its states if $B$ is a finite state black box or will span state space if $B$ is a linear black box, and $V_k^*$ will be large enough so that the states are completely characterized by the set of experiments $E_{\bar{v}}$, $\bar{v}\varepsilon V_k^*$. Thenceforth, the constructed automaton will correctly realize $B$.

However, even when $A_k$ correctly represents $B$ the representations $A_k$ will not be invariant with respect to $k$. This is because $V_k^*$ will continue to grow, so that the states will be represented by longer and longer strings of characteristics. An algorithm will be given for eliminating superflous state characteristics so that, when $V_k^*$ is large enough to completely characterize the states, the representations of the states will no longer change and, therefore, the representation of $B$ which is constructed will no longer change. For this purpose, it is necessary that the orderings chosen for the $V_k^*$ be consistent in the sense defined above.

At early stages of the identification in the limit procedure, before $U_k^*$ and/or $V_k^*$ are large enough to correctly realize $b$, it may be impossible to construct an automaton at some $k$ due to inconsistency in the obtained data. This can occur in two ways:

(1) When some single input $u$ is applied to some state $\Delta_{\bar{u}_1}b$ for $\bar{u}_1\varepsilon U_k^*$ it may be found that $\Delta_{\bar{u}_1u}b$ has characteristics different from those of the states $\Delta_{\bar{u}}b$, $\bar{u}\varepsilon U_k^*$, which have been studied at stage $k$. Therefore, the transition and output functions will not have been determined for $\Delta_{\bar{u}_1u}b$. In this case it is necessary to enlarge $U_k^*$ by adjoining $\bar{u}_1u$.

(2) Two states $\Delta_{\bar{u}_1}b$ and $\Delta_{\bar{u}_2}b$ which are considered, i.e. $\bar{u}_1\varepsilon U_k^*$ and $\bar{u}_2\varepsilon U_k^*$, may be found to have the same characteristics at stage $k$, but their transitions $\Delta_{\bar{u}_1u}b$ and $\Delta_{\bar{u}_2u}b$ may have different characteristics for some $u$ or $\Delta_{\bar{u}_1}b$ and $\Delta_{\bar{u}_2}b$ may have different outputs. In the first case the state transition function can't be defined and in the second case the output function can't be defined. This means that the set $V_k^*$ of state characteristics is not large enough to distinguish different states. An algorithm is given for using the inconsistency which is found to enlarge $V_k^*$ in order that $\Delta_{\bar{u}_1}b$ and $\Delta_{\bar{u}_2}b$ will have different representations.

At each stage $k$ it will only be necessary to perform the above operations of deleting superfluous state characteristics, adding states, and adding state characteristics to distinguish states which are found to be different, a finite number of times before a representation of $B$ can be constructed.

Actually, it is sufficient to increase only the set of state characteristics, or only the set of states considered, at each stage of identification in the limit. For instance, the set of state characteristics could be increased: $V_1^*\subset V_2^*\subset \ldots$; and the set $U_k^*$ of states to be considered could be taken to consist only of the initial state $x_\phi b$ at all stages $k$. The above operation of adding states would increase the set of states at each stage $k$. When $V_k^*$ is large enough to completely characterize the states of $b$, then the operation of adding states will necessarily increase the set $U_k^*$ of states considered to span all of state space. Alternatively, $V_k^*$ could be fixed at the set of experiments defined by the input strings of shortest length, i.e. 1 or 0, depending on exact definitions, while the set of states considered is increased: $U_1^*\subset U_2^*\subset \ldots$. Then, as the set of states considered grows, the above operation of adding characteristics to remove inconsistencies in the acquired data would increase the set $V_k^*$ of characteristics used to characterize the states until $V_k^*$ completely characterized all the states.

The first of these two approaches, namely increasing only $V_k^*$, will be used in constructing the following algorithms. Its validity is proven in Theorems 4 and 5 of Section 8.

*Finite identification of black boxes by state characterization*

Suppose that the unknown black box $B$ is realizable by a finite automaton and we are told the number $N$ of states required by the automaton, or $B$ is realizable by a linear automaton and we are told the dimensionality $N$ of its state space. Then it is possible to *finitely identify B*. That is, it is not necessary to perform the identification in the limit algorithm indefinitely. Namely, let $U_1^*$ and $V_1^*$ contain all input strings of length at most $N$. In the linear case, only a basis for $U$ need be used so that $U_1^*$ and $V_1^*$ are finite. Then it can be guaranteed that a correct representation for $B$ will be obtained in the first stage of the identification in the limit algorithm. This assertion is the content of Theorems 1 and 2 of Section 8.

This is the method described by Ho [14] and Arbib and Zeiger [2].

## 6. IDENTIFICATION OF FINITE AUTOMATA BY STATE CHARACTERIZATION

*Identification in the limit, Mealy model*

In the case of a black box realizable by a finite automaton, which will be called henceforth a *finite state black box*, $U$ and $Y$ each contain a finite number of elements. At the $k$-th stage of the identification in the limit procedure states $\Delta_{\bar{u}}b$, $\bar{u}\varepsilon U_k^*$ are considered, and the states of $B$ are characterized by experiments $E_{\bar{v}}$, $\bar{v}\varepsilon V_k^*$. It will be assumed that $\phi\varepsilon U_k^*$ and $U\subset V_k^*$. A finite automaton, the $k$-th guess of a representation for $B$, is constructed by performing the following five steps:

*Step* 1, *data collection*. Let

$$U_k^{*\prime}=U_k^*U-U_k^*. \tag{8}$$

That is, $U_k^{*\prime}$ consists of those strings $\bar{u}u$ such that $\bar{u}\varepsilon U_k^*$, $u\varepsilon U$, and $\bar{u}u\varepsilon U_k^*$. Each state $\Delta_{\bar{u}}b$, $\bar{u}\varepsilon U_k^*\cup U_k^{*\prime}$ is characterized by performing experiments $E_{\bar{v}}$, $\bar{v}\varepsilon V_k^*$ on it. That is, the values of $b(\bar{u}\bar{v})$ are determined for $\bar{u}\varepsilon U_k^*\cup U_k^{*\prime}$ and $\bar{v}\varepsilon V_k^*$.

Let the number of elements in $V_k^*$ be designated $N$. Then the results of these experiments can be arranged in a matrix with each row representing a state of $U_k^* \cup U_k^{*\prime}$ and the elements of the row consisting of the results of the experiments performed on that state. Thus, each row is the $N$-tuple of output elements which represents that state according to the characteristics used at this stage.

*Step* 2, *add characteristics.* Check the rows of the data matrix which represent the states of $U_k^*$. Consider each pair $(\bar{u}_1, \bar{u}_2)$ such that their rows are the same. In order to be able to construct an automaton which represents the data matrix, the following condition needs to be satisfied: for each $u \varepsilon U$ the rows corresponding to $\bar{u}_1 u$ and $\bar{u}_2 u$ need to be the same. If they are not the same, this implies that the states $\Delta_{\bar{u}_1} b$ and $\Delta_{\bar{u}_2} b$ are different states of $B$, but the set of state characteristics defined by $V_k^*$ is not large enough to distinguish them. So $V_k^*$ must be enlarged. This can be done as follows: Look for the first element of the $\bar{u}_1 u$ and $\bar{u}_2 u$ rows of the data matrix which differs in the two rows. Let this pair of elements correspond to the experiment $E_{\bar{v}}$. Then adjoin the element $u\bar{v}$ to $V_k^*$. Now repeat the data collection, step 1, to obtain a data matrix with one more column. The rows corresponding to $\bar{u}_1$ and $\bar{u}_2$ will now differ in the last column.

Continue this procedure of adding characteristics to $V_k^*$ until it is no longer necessary.

*Step* 3, *add states.* Check the rows of the data matrix corresponding to $\bar{u} \varepsilon U_k^{*\prime}$. In order to be able to construct an automaton which will realize the data matrix it is necessary that each of these rows be the same as at least one of the rows corresponding to $U_k^*$. If this is not the case, then choose one representative $\bar{u}$ from each of the set of equal rows corresponding to $U_k^{*\prime}$ which are not equal to any of the rows corresponding to $U_k^*$, and adjoin $\bar{u}$ to $U_k^*$. Now it is necessary to perform the data collection again, since $U_k^{*\prime}$ must be enlarged.

After this step is performed it will not be necesary to return to step 2, adding characteristics. This is because, of the rows of the data matrix corresponding to $U_k^*$, there can be no new equal pairs, since each new row that was added was different from all the others.

*Step* 4, *delete superfluous characteristics.* Check each column of the data matrix in succession, from last to first, deleting each characteristic which adds no information to the others: If all entries in the last column are the same, delete it. In the case of the $i$-th column, if there is no pair of rows which have the same entries in each of the other columns but different entries in the $i$-th column, delete it.

When $U_k^*$ and $V_k^*$ are large enough so that a correct representation of the unknown black box $B$ is obtained, successive stages of the identification in the limit procedure will yield correct representations, but the representations will differ if the states of $B$ are represented by different characteristics. This is why it is necessary to delete superfluous characteristics and why it is necessary to do it in a consistent manner. This is why it was assumed in the second paragraph of Section 5 that the orderings chosen for the $V_k^*$ are consistent.

*Step* 5, *construct automaton.* Now an automaton can be constructed which represents the data matrix: If there are $N$ characteristics in $V_k^*$, then each state of $U_k^*$ is represented by an $N$-tuple of output elements, namely the corresponding row of the data matrix. Let the input and output sets of the constructed automaton be the same as those of $B$, and let its state set $X$ consist of the set of $N$-tuples $\langle y_1, \ldots, y_N \rangle$ which occur as rows of the data matrix. For each $x \varepsilon X$ choose a $\bar{u} \varepsilon U_k^*$ such that its row in the data matrix is $x$. In order to define the state transition function $f_{tr}(u, x)$, look at the row of the data matrix corresponding to $\bar{u}u$. If this row is $x'$, then define $f_{tr}(u, x) = x'$. The information required to construct the output function $f_{out}$ was in the data matrix before superfluous characteristics were deleted. Namely, define $f_{out}(u, x) = E_u(\Delta_{\bar{u}} b)$, where $E_u$ was one of the original characteristics since it was assumed that $U \subset V_k^*$. For the initial state $x_0$ of the automaton, use the row of the final data matrix corresponding to $\phi$, which was assumed to be one of the elements of $U_k^*$.

*Simplification*

It was stated at the end of Section 5, and it will be proved in Section 8, that it is sufficient to have $U_k^*$ always consist of only the single element $\phi$ and let $V_k^*$ grow at successive stages of the identification in the limit procedure so that every $\bar{v} \varepsilon \Sigma U$ is eventually included. In this case step 2 of adding characteristics can be deleted. This is because at each stage $k$ the original data matrix will contain only one row corresponding to $U_k^*$, and so can have no equal pair of rows.

*Example*

Suppose that $U$ and $Y$ are both binary sets:

$$U = \{I, J\} \qquad Y = \{P, Q\}.$$

Table 1 gives an example of input–output function values.

For the first stage of the identification in the limit procedure, let us take $U_1^*$ to contain only $\phi$ and $V_1^*$ to consist of the input strings of length 1, namely $I$ and $J$. Then the original data matrix derived from Table 1 at stage 1 is shown on the left side of Table 2. Since the 2 rows corresponding to

$U_1^{*\prime}$ are the same as the row corresponding to $U_1^*$, it is not necessary to add states. In order to obtain the final data matrix on the right side of Table 2, it is only necessary to delete superfluous characteristics. Starting from the last column of the original data matrix, we see that this column is constant, and so can carry no information, and is deleted. This leaves us with the states characterized by a single experiment, $E_I$ and only one state. The automaton derived from the final data matrix is shown in Fig. 1. Since the original data matrix contained the first set of values in Table 1, the automaton of Fig. 1 will yield these values.

TABLE 1. EXAMPLE OF BLACK BOX FUNCTION VALUES. THE INPUT AND OUTPUT SETS ARE
$U=\{I, J\}$ $\quad Y=\{P, Q\}$

| | Input string $\bar{u}$ | Output $y=b(\bar{u})$ | Input string $\bar{u}$ | Output $y=b(\bar{u})$ | Input string $\bar{u}$ | Output $y=b(\bar{u})$ |
|---|---|---|---|---|---|---|
| First set of values | $I \rightarrow$ | $P$ | $II \rightarrow$ | $P$ | $JI \rightarrow$ | $P$ |
| | $J \rightarrow$ | $P$ | $IJ \rightarrow$ | $P$ | $JJ \rightarrow$ | $P$ |
| Second set of values | $III \rightarrow$ | $Q$ | $IIII \rightarrow$ | $Q$ | $IIIII \rightarrow$ | $Q$ |
| | $IIJ \rightarrow$ | $P$ | $IIIJ \rightarrow$ | $P$ | $IIIIJ \rightarrow$ | $P$ |
| | $IJI \rightarrow$ | $Q$ | $IIJI \rightarrow$ | $P$ | $IIIJI \rightarrow$ | $P$ |
| | $IJJ \rightarrow$ | $P$ | $IIJJ \rightarrow$ | $P$ | $IIIJJ \rightarrow$ | $P$ |
| | $JII \rightarrow$ | $Q$ | $IJII \rightarrow$ | $Q$ | $IIJII \rightarrow$ | $P$ |
| | $JIJ \rightarrow$ | $P$ | $IJIJ \rightarrow$ | $P$ | $IIJIJ \rightarrow$ | $P$ |
| | $JJI \rightarrow$ | $Q$ | $IJJI \rightarrow$ | $P$ | $IIJJI \rightarrow$ | $P$ |
| | $JJJ \rightarrow$ | $P$ | $IJJJ \rightarrow$ | $P$ | $IIJJJ \rightarrow$ | $P$ |

TABLE 2. ON THE LEFT IS ORIGINAL DATA MATRIX FOR BLACK BOX OF TABLE 1 AT THE BEGINNING OF STAGE 1:

$$U_1^*=\{\phi\} \qquad V_1^*=\{I, J\}.$$

THE ENTRIES ARE THE $\bar{v}$-th CHARACTERISTIC OF THE $\bar{u}$-th STATE:

$$E_{\bar{v}}(\Delta_{\bar{u}}b)=b(\bar{u}\bar{v}).$$

THIS DATA MATRIX CONTAINS THE FIRST SET OF VALUES IN TABLE 1. ON THE RIGHT IS THE FINAL DATA MATRIX, AFTER DELETING SUPERFLUOUS CHARACTERISTICS

Original

| States | | Characteristics $\bar{v}\varepsilon V_1^*$: $I$ | $J$ |
|---|---|---|---|
| $\bar{u}\varepsilon U_1^*$ | $\phi$ | $P$ | $P$ |
| $\bar{u}\varepsilon U_1^{*\prime}$ | $I$ | $P$ | $P$ |
| | $J$ | $P$ | $P$ |

Final

| States | | Characteristic $\bar{v}\varepsilon V_1^*$: $I$ |
|---|---|---|
| $u\varepsilon U_1^*$ | $\phi$ | $P$ |
| $\bar{u}\varepsilon U_1^{*\prime}$ | $I$ | $P$ |
| | $J$ | $P$ |

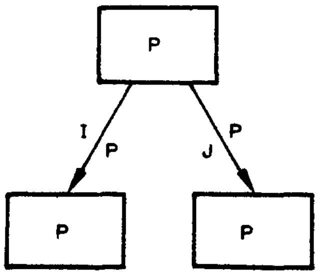


FIG. 1. Automaton representing the data matrix on the right side of Table 2. Agrees with the first set of values in Table 2. The boxes are the states (there is only one) labelled by their characteristics (there is only one). The arrows represent transitions, with the input element causing the transition on the left side of the arrow and the output from the transition shown on the right side of the arrow.

For the second stage let us take $U_2^*$ to consist solely of $\phi$ and $V_2^*$ to consist of the input strings of length less than or equal to 2. The original data matrix in stage 2 is shown in Table 3. The two rows

TABLE 3. ORIGINAL DATA MATRIX FOR BLACK BOX OF TABLE 1, STAGE 2:

$$U_2^*=\{\phi\} \qquad V_2^*=\{I, J, II, IJ, JI, JJ\}$$

| States | | Characteristics $\bar{v}\varepsilon V_2^*$: $I$ | $J$ | $II$ | $IJ$ | $JI$ | $JJ$ |
|---|---|---|---|---|---|---|---|
| $\bar{u}\varepsilon U_2^*$ | $\phi$ | $P$ | $P$ | $P$ | $P$ | $P$ | $P$ |
| $\bar{u}\varepsilon U_2^{*\prime}$ | $I$ | $P$ | $P$ | $Q$ | $P$ | $Q$ | $P$ |
| | $J$ | $P$ | $P$ | $Q$ | $P$ | $Q$ | $P$ |

representing the states of $U_2^{*\prime}$ differ from the row representing the state of $U_2^*$. Since the two rows of $U_2^{*\prime}$ are the same, only one of them is to be added to $U_2^*$. In Table 4 the row representing state $\Delta_I b$ is

TABLE 4. STATES ADDED TO DATA MATRIX OF TABLE 3

| States | | Characteristics $\bar{v}\varepsilon V_2^*$: $I$ | $J$ | $II$ | $IJ$ | $JI$ | $JJ$ |
|---|---|---|---|---|---|---|---|
| $\bar{u}\varepsilon U_2^*$ | $\phi$ | $P$ | $P$ | $P$ | $P$ | $P$ | $P$ |
| | $I$ | $P$ | $P$ | $Q$ | $P$ | $Q$ | $P$ |
| $\bar{u}\varepsilon U_2^{*\prime}$ | $J$ | $P$ | $P$ | $Q$ | $P$ | $Q$ | $P$ |
| | $II$ | $Q$ | $P$ | $Q$ | $P$ | $P$ | $P$ |
| | $IJ$ | $Q$ | $P$ | $Q$ | $P$ | $P$ | $P$ |

added to $U_2^*$ and the necessary additions are made to $U_2^{*\prime}$. Again two identical rows of $U_2^{*\prime}$ differ from the rows of $U_2^*$, and one of them is added to $U_2^*$ in

Table 5. Now each row of $U_2^{*\prime}$ is the same as one of the rows of $U_2^*$, so it is no longer necessary to add states. The data matrix now contains all the entries in Table 1, so the derived automaton will agree with these values. In order to delete superfluous characteristics, first note that columns 2, 4, and 6 of Table 5 are constants, can carry no information, and so can be eliminated immediately. This leaves us with columns $I$, $II$ and $JI$. In order to see if the last of these carries information in addition to the first 2, note that we only need to check the states of $U_2^*$ since the states of $U_2^{*\prime}$ have the same representations. Of the 3 states of $U_2^*$, the column $JI$ can distinguish between the first and second, and between the second and third. Since the $II$ column distinguishes between the first and second rows and the $I$ column distinguishes between the second and third rows, the $JI$ column can be deleted. This leaves us with the final data matrix of Table 6, containing 3 different states characterized by the two experiments $E_I$ and $E_{II}$. The derived automaton is shown in Fig. 2.

TABLE 5. MORE STATES ADDED TO TABLE 4. THIS DATA MATRIX CONTAINS ALL BLACK BOX FUNCTION VALUES IN TABLE 1.

| States | | Characteristics $\bar{v}\varepsilon V_2^*$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | $I$ | $J$ | $II$ | $IJ$ | $JI$ | $JJ$ |
| $\bar{u}\varepsilon U_2^*$ | $\phi$ | $P$ | $P$ | $P$ | $P$ | $P$ | $P$ |
| | $I$ | $P$ | $P$ | $Q$ | $P$ | $Q$ | $P$ |
| | $II$ | $Q$ | $P$ | $Q$ | $P$ | $P$ | $P$ |
| $\bar{u}\varepsilon U_2^{*\prime}$ | $J$ | $P$ | $P$ | $Q$ | $P$ | $Q$ | $P$ |
| | $IJ$ | $Q$ | $P$ | $Q$ | $P$ | $P$ | $P$ |
| | $III$ | $Q$ | $P$ | $Q$ | $P$ | $P$ | $P$ |
| | $IIJ$ | $P$ | $P$ | $P$ | $P$ | $P$ | $P$ |

TABLE 6. FINAL DATA MATRIX FOR TABLE 5, OBTAINED BY DELETING SUPERFLUOUS CHARACTERISTICS. THE OUTPUTS, THE FIRST 2 COLUMNS OF TABLE 5, ARE ADDENDED ON RIGHT SIDE OF THIS DATA MATRIX

| States | | Characteristics $\bar{v}\varepsilon V_2^*$ | | | |
|---|---|---|---|---|---|
| | | $I$ | $II$ | $I$ | $J$ |
| $\bar{u}\varepsilon U_2^*$ | $\phi$ | $P$ | $P$ | $P$ | $P$ |
| | $I$ | $P$ | $Q$ | $P$ | $P$ |
| | $II$ | $Q$ | $Q$ | $Q$ | $P$ |
| $\bar{u}\varepsilon U_2^{*\prime}$ | $J$ | $P$ | $Q$ | | |
| | $IJ$ | $Q$ | $Q$ | | |
| | $III$ | $Q$ | $Q$ | | |
| | $IIJ$ | $P$ | $P$ | | |

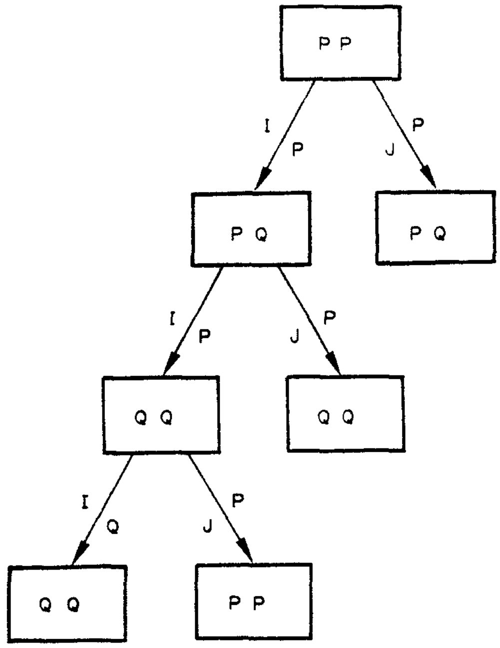


FIG. 2. Mealy model representation for data matrix of Table 6. The boxes signify states with the representation of state inside the box. Initial state on top, transitions shown with input on left side and output on right side of arrow. This automaton agrees with both sets of values in Table 1.

*Identification in the limit, Moore model*

State characterization can be used to identify a finite state black box in the limit by constructing successively better Moore model automata in much the same way as was described above for constructing Mealy model automata. The difference is that in the Mealy model algorithm it was required that the successive sets $V_k^*$ of state characteristics contain the input strings of length 1 so that the data matrix would contain the Mealy model output function. The Moore model output function, on the other hand, corresponds to the experiment $E_\phi$, so now the requirement is that $\phi\varepsilon V_k^*$. However, it should be noted that the result of this experiment on the initial state is undefined.

$$E_\phi(\Delta_\phi b) = b(\phi). \tag{9}$$

Since $b(\phi)$ is undefined, any output element can be put in this position in the data matrix and the derived automaton will agree with $b$. This element of the data matrix can be chosen to give the simplest representation.

## 7. IDENTIFICATION OF LINEAR AUTOMATA BY STATE CHARACTERIZATION

*Linear automata*

Whereas a finite automaton has finite input, output, and state sets, each containing a finite number of elements, a linear automaton has an input, output, and state space which are finite dimensional vector spaces over some field. The initial state of a linear automaton is required to be 0 and its state transition and output functions are required to be linear:

$$x_t = f_{tr}(u_t, x_{t-1}) = A(u_t) + B(x_{t-1}). \qquad (10)$$

Mealy model:

$$y_t = f_{out}(u_t, x_{t-1}) = C(u_t) + D(x_{t-1}). \qquad (11)$$

Moore model:

$$y_t = f_{out}(x_t) = D(x_t). \qquad (12)$$

where $A$ is a linear transformation from $U$ to $X$, $B$ is a linear transformation from $X$ to $X$, $C$ is a linear transformation from $U$ to $Y$ and $D$ is a linear transformation from $X$ to $Y$.

*Linear input–output functions*

The following properties of the input–output functions of linear automata, that is of *linear input–output functions*, have been demonstrated by Arbib and Zeiger in their paper referenced in the Introduction:

$b(\bar{u})$ is linear in the following sense: If $\bar{u}$ and $\bar{v}$ are input strings of the same length, and $\alpha$ and $\beta$ are elements of the underlying field, then

$$b(\alpha\bar{u} + \beta\bar{v}) = \alpha b(\bar{u}) + \beta b(\bar{v}). \qquad (13)$$

In particular, if $\bar{0}$ is a string of 0 inputs of any length, then $b(\bar{0}) = 0$. Furthermore,

$$b(\bar{0}\bar{u}) = b(\bar{u}). \qquad (14)$$

Let $u_1, \ldots, u_I$ be a basis for $U$. Due to (13) and (14), $b(\bar{u})$ is completely determined by the values of $b(u_i 0^{t-1})$, $t=1, 2, \ldots$, $i=1, \ldots, I$, and $0^n$ signifies a string of input 0's of length $n$.

The same conclusions can be drawn about the states of a linear automaton: If $x(\bar{u})$ is the state of the automaton after application of input string $\bar{u}$, then $x(\bar{u})$ satisfies (13) and (14) with $b$ replaced by $x$.

*Application of state characterization to linear black boxes*

In constructing a linear automaton to represent a linear black box, the initial state of the automaton must be the 0 of state space by definition. Therefore, it is not necessary to characterize the initial state of $B$.

From the above considerations, we need only characterize the black box states

$$\Delta_{\bar{u}} b,\ \bar{u} = u_i 0^{t-1}, \qquad i=1, \ldots, I \qquad t=1, 2, \ldots \qquad (15)$$

and for characteristics we need only consider experiments

$$E_{\bar{v}},\ \bar{v} = 0^t,\ t=0, 1, \ldots \qquad (16)$$

where $0^0$ signifies $\phi$.

In characterizing the states of $B$ by the outcomes of specified experiments we may be forced to represent the states by a space of higher dimension than necessary. This is because the output space $Y$ may have dimension greater than 1. Then the outcome of an experiment is a vector of dimension higher than 1. It may occur that, if we express the elements of $Y$ in terms of some basis $y_1, \ldots, y_J$, that an experiment is useful in characterizing states of $B$ because some of the components of its outcome distinguish between states, while others do not. Therefore, it is useful to break up the outcome of an experiment and assign its components to subexperiments:

If $y = \alpha_1 y_1 + \ldots + \alpha_J y_J$, where the $\alpha_j$ are elements of the underlying field, then $\alpha_j$ will be called the $j$-th *component* of $y$. Now we can break up experiment $E_{\bar{v}}$ into subexperiments $E_{\bar{v},j}$, where the result of $E_{\bar{v},j}$ is defined to be the $j$-th component of the result of $E_{\bar{v}}$.

In order to apply the method of state characterization to the identification in the limit of a linear black box, at stage $k$ the set of states of $B$ to be characterized is specified by a set $U_k^*$ of input strings of the form (15). These states are characterized by an ordered set of experiments specified by an ordered set $V_k^*$ of pairs $(\bar{v}, j)$, where $\bar{v}$ is an input string of the form (16) and $j=1, \ldots, J$.

*Identification in the limit, Moore model*

In order to identify a linear black box in the limit by means of state characterization with a Moore model representation, the following 5 steps are performed at each stage. It is assumed that at each stage $k$, $U_k^*$ is taken initially to consist of a set of input strings of length 1 which are a basis for $U$:

$$U_k^* = \{u_1, \ldots, u_I\} \qquad (17)$$

$V_k^*$ is increased at successive stages. It is assumed that $(\phi, j) \varepsilon V_k^*$ for all $j=1, \ldots, J$.

*Step* 1, *Data collection.*

Let

$$U_k^{*\prime} = U_k^* 0 - U_k^* \qquad (18)$$

where $U_k^* 0$ designates the set of input strings consisting of the elements of $U_k^*$ with a 0 added to the end of each. Each state $\Delta_{\bar{u}} b$, $\bar{u} \varepsilon U_k^* \cup U_k^{*\prime}$ is characterized by performing experiments $E_{\bar{v},j}$, where $(\bar{v}, j)$ ranges over the successive pairs of $V_k^*$. The result of such an experiment is the $j$-th component of $b(u_i 0^a)$ where $u_i$ is the first element of $\bar{u}$ and $a$ is the sum of the number of 0's in $\bar{u}$ and the length of $\bar{v}$ (which is all 0's).

As in the case of finite state black box identification, the results of these experiments can be arranged in a matrix, each row representing one of states $\Delta_{\bar{u}}b$ which has been characterized and the elements of the row consisting of the results of the experiments on this state. Let the number of elements of $V_k^*$ be $N$. Each state $\Delta_{\bar{u}}b$ is represented by its row of the data matrix which is an $N$-tuple of elements from the underlying field. This representation will be considered to be an element of an $N$-dimensional vector space and will be designated $x(\bar{u})$.

*Step* 2, *add characteristics*. Since $U_k^*$ consists of more than one element, this step is necessary. In the case of linear black boxes it is performed as follows:

Consider the representations $x(\bar{u})$ of the $\Delta_{\bar{u}}b$, $\bar{u}\varepsilon U_k^*$. If they are linearly independent nothing remains to be done. If not, then in order to be able to construct a representation for this data matrix the following must hold: Every linear relation satisfied by the $x(\bar{u})$ for $\bar{u}\varepsilon U_k^*$ must also hold for the $x(\bar{u}0)$ with $\bar{u}\varepsilon U^*$. If not, then this means that there are states of $B$ which are linearly independent, but appear to be linearly dependent because we are not using enough characteristics to represent them. The fact that the $x(\bar{u}0)$ didn't satisfy the required linear relationship means that one of the components of the $x(\bar{u}0)$ didn't satisfy this relationship. Then it is sufficient to add $(0\bar{v}, j)$ to $V_k^*$. Now the $x(\bar{u})$, $\bar{u}\varepsilon U_k^*$, will no longer satisfy that relationship.

It is necessary to repeat step 1, data collection, each time a characteristic is added or else more characteristics may be added than are necessary.

*Step* 3, *add states*. Each $x(\bar{u})$ for $\bar{u}\varepsilon U_k^{*\prime}$ must be a linear combination of the $x(\bar{u})$ for $\bar{u}\varepsilon U_k^*$. Otherwise, delete the offensive $\bar{u}$ from $U_k^{*\prime}$ and add it to $U_k^*$. If this is done one $\bar{u}$ at a time so that the $\bar{u}$ which are added to $U_k^*$ don't add any new linear relationships among the $x(\bar{u})$ for $\bar{u}\varepsilon U_k^*$, then it will not be necessary to repeat step 2, add characteristics, only step 1, data collection.

*Step* 4, *delete superfluous characteristics*. Now think of the columns of the data matrix as elements of a linear vector space. Starting from the last column, delete each column which is a linear combbination of the others. Note that it is only necessary to consider the rows of the data matrix corresponding to $U_k^*$, since the rows corresponding to $U_k^{*\prime}$ are linear combinations of these.

*Step* 5, *construct automaton*. If the data matrix now has $N$ columns, then it has exactly $N$ independent rows, due to the operation of deleting superfluous characteristics. Let these rows be designated $x_1, \ldots, x_N$. I assume that they are chosen from the $U_k^*$ rows. It is convenient to use these as a basis for the state space of the automaton.

We know that the initial state of the automaton is the 0 of state space. So, the first problem is to define the state transition function $f_{tr}$. Since

$$f_{tr}(u, x) = A(u) + B(x) \tag{19}$$

where $A$ and $B$ are linear transformations, it follows that

$$A(u) = f_{tr}(u, 0) \tag{20}$$

$$B(x) = f_{tr}(0, x). \tag{21}$$

Let $u_1, \ldots, u_I$ be a basis for $U$. Then, due to the linearity of $A$ and $B$ it is sufficient to define $f_{tr}(u_i, 0)$ for $i=1, \ldots, I$ and $f_{tr}(0, x_n)$ for $n=1, \ldots, N$. Since the initial state of the black box corresponds to the 0 of state space,

$$f_{tr}(u_i, 0) = x(u_i) \tag{22}$$

which is one of the rows of our data matrix by assumption (17). Let $x_n$ be the row of our data matrix corresponding to $u_i0^{t-1}\varepsilon U_k^*$. Then

$$f_{tr}(0, x_n) = x(u_i0^t) \tag{23}$$

which is also a row of our data matrix since $\bar{u}\varepsilon U_k^*$ implies $\bar{u}0\varepsilon U_{k_1}^* \cup U_k^{*\prime}$.

All that $U_k^* \cup U_k^{*\prime}$ is to define the output function $f_{out}$. If $x_n$ is the $\bar{u}$-th row of our data matrix, then, for a Moore model,

$$f_{out}(x_n) = b(\bar{u}) = E_\phi(\Delta_{\bar{u}}b) \tag{24}$$

which is specified by columns $(\phi, j)$, $j=1, \ldots, J$, of the data matrix before deleting superfluous characteristics since it was assumed that $V_k^*$ initially included these experiments.

*Identification in the limit, Mealy model*

The principal difference between constructing a Moore model for the black box and a Mealy model is this: In constructing a Mealy model, 2 states $\Delta_{\bar{u}_1}b$ and $\Delta_{\bar{u}_2}b$ of the black box will be given the same representation if they are the same function from $\Sigma U$ to $Y$. In the case of constructing a Moore model, on the other hand, they must also give the same result for the experiment $E_\phi$, that is,

$$b(\bar{u}_1) = b(\bar{u}_2). \tag{25}$$

Therefore, in order to construct a Mealy model the experiments $E_{\phi,j}$ are not included in $V_k^*$ for the purpose of characterizing the states. However, it is necessary to perform the experiment $E_\phi$ on each state $\Delta_{\bar{u}}b$ for $\bar{u}\varepsilon U_k^* \cup U_k^{*\prime}$ in order to be able to define the Mealy model output function:

$$\begin{aligned} f_{out}(x(\bar{u}), u_i) &= b(\bar{u}u_i) \\ &= b(\bar{u}0) + b(0^t u_i) \\ &= b(\bar{u}0) + b(u_i) \\ &= E_\phi(\Delta_{\bar{u}0}b) + E_\phi(\Delta_{u_i}b). \end{aligned} \tag{26}$$

*Finite identification of linear black boxes*

It is shown in Section 8 that if a linear automaton has a state space of dimension $N$, then the states reachable by input strings of length at most $N$ span the reachable subspace of state space and the states of its input–output function can be characterized by experiments $E_{\bar{v}}$ with $\bar{v}$ of length at most $N$. Therefore, a finite data matrix can be specified *a priori* which will yield a correct representation. This is Ho's algorithm.

## 8. THEOREMS

The following two theorems are of fundamental importance, because they show that the identification in the limit algorithms of the previous sections will correctly identify a finite state black box or a linear black box after a finite number of stages. Also, they show that finite identification is possible if we know the size of state space.

*Theorem* 1

If a finite state black box has $N$ states, then they are reachable by input strings of length at most $N$. If a linear automaton has a state space of dimensionality $N$, then the reachable subspace is reachable by input strings of length at most $N$.

*Theorem* 2

If a finite state black box has $N$ states, then they are completely characterized by experiments $E_{\bar{v}}$ with $\bar{v}$ of length at most $N$. If a linear automaton has a state space of dimensionality $N$, then the states of its input–output function are completely characterized by experiments of length at most $N$.

These two theorems are demonstrated, using different terminology, in Arbib and Zeiger's paper referenced in the Introduction. Results of this type were introduced in the theory of finite automata by GINSBURG [8] and HIBBARD [13] and in the theory of linear systems by GILBERT [6] and KALMAN [15].

Theorem 1 is rather trivial: In the case of a finite state black box, consider state $\Delta_{\bar{u}}b$. If $\bar{u}$ is of length greater than $N$, then 2 of the states which $b$ passed through on the way to $\Delta_{\bar{u}}b$ must have been the same, so the corresponding segment of $\bar{u}$ can be deleted. In the case of a finite linear automaton, suppose that state $x$ is reached by input string $\bar{u}$. If the length of $\bar{u}$ exceeds $N$, then $x$ must be a linear combination of the previous states passed through along the way. So take the initial $t-1$ substrings of $\bar{u}$, add 0's to their beginnings to make them all of length $t-1$, and take the same linear combination of them. An input string of length $t-1$ results which, when applied to the automaton, puts it in state $x$ since the initial 0's which were added to the substrings of $\bar{u}$ take the 0 state into itself.

Theorem 2 follows from Arbib and Zeiger's first lemma on page 595 of their paper. Using the notion of state characterization, I will now derive a generalization of that lemma without the use of abstract algebra.

The following theorem is of importance because it reduces the amount of experimentation we have to do to determine if a proposed set of state characteristics $\bar{E}$ completely characterizes the states of $B$, compared to the original definition (25). Namely, according to the original definition, $\bar{E}$ completely characterizes the states of $B$ if the following condition is satisfied: If $\Delta_{\bar{u}_1}b$ and $\Delta_{\bar{u}_2}b$ are two states of $B$ which give the same result for every experiment of $\bar{E}$, then they must give the same result for every experiment $E$. Theorem 3 says that we do not have to consider every possible experiment $E$. It is sufficient to limit our attention to (1) the immediate outputs of a state, defined by $E_u$ where $u$ ranges over $U$, or over a basis for $U$ in the case of linear black boxes; and (2) the characteristics of the immediate transitions of the state, defined by $E_{u\bar{v}}$ where $u$ ranges over $U$ and $\bar{v}$ over the input strings which define the characteristics $\bar{E}$. Thus, the set of experiments $E$ which must be considered is reduced from an infinite set to a finite set. However, the set of pairs of states to be considered remains infinite.

*Theorem* 3

$\bar{E}$ completely characterizes the states of $B$ if the following conditions hold: If $\Delta_{\bar{u}_1}b$ and $\Delta_{\bar{u}_2}b$ are 2 states of $B$ which give the same result for each experiment of $\bar{E}$ then for, all $u\varepsilon U$, 1)$b(\bar{u}_1u)=b(\bar{u}_2u)$ and (2) $\Delta_{\bar{u}1u}b$ and $\Delta_{\bar{u}2u}b$ give the same result for each experiment of $\bar{E}$.

*Proof.* In order to show that $\bar{E}$ completely characterizes the states of $B$ it must be shown that if each experiment of $\bar{E}$ gives the same result on the 2 states $\Delta_{\bar{u}_1}b$ and $\Delta_{\bar{u}_2}b$, then

$$E_{\bar{v}}(\Delta_{\bar{u}_1}b)=E_{\bar{v}}(\Delta_{\bar{u}_2}b) \quad \text{for all } \bar{v}\varepsilon\Sigma U. \tag{27}$$

That is,

$$b(\bar{u}_1\bar{v})=b(\bar{u}_2\bar{v}) \qquad \text{for all } \bar{v}\varepsilon\Sigma U. \tag{28}$$

Let $\bar{v}=u_1 \dots u_t$. Let $\bar{v}'=u_1 \dots u_{t-1}$. Then the use of hypothesis 2) $t-1$ times implies that $\Delta_{\bar{u}_1\bar{v}'}b$ and $\Delta_{\bar{u}_2\bar{v}'}b$ give the same result for each experiment of $\bar{E}$. Hypothesis (1) now implies (28) by setting $u=u_t$ and replacing $\bar{u}_1$ by $\bar{u}_1v'$ and $\bar{u}_2$ by $\bar{u}_2v'$.

To see that Theorem 3 is a generalization of Arbib and Zeiger's first lemma on page 595, note that it immediately implies the following corollary which is a restatement of Arbib and Zeiger's lemma.

Let $\bar{E}_n$ denote the set of experiments $E_{\bar{v}}$ with $\bar{v}$ ranging over input strings of length less than or equal to $n$.

*Corollary.* Suppose that every pair of states of $B$ which give the same result on each experiment of $\bar{E}_n$ also give the same result on each experiment of $\bar{E}_{n+1}$. Then $\bar{E}_n$ completely characterizes the states of $B$.

Theorem 2 can be proved from this corollary by the method of Arbib and Zeiger: In the case of a finite state black box, consider the number of different representations given to the states of $B$ by $\bar{E}_n$. The corollary shows that if this number is no greater for $\bar{E}_{n+1}$ than for $\bar{E}_n$, then $\bar{E}_n$ completely characterizes the states of $B$. If $B$ only has $N$ different states, then this number can increase at most $N$ times in succession. This gives Theorem 2. In the case of a linear black box, consider the number of linearly independent representations given to the states of $B$ by $\bar{E}_n$ and use the same reasoning.

The following 2 theorems will only be stated and proved for finite state black boxes, although analogous results hold for linear black boxes.

*Theorem* 4

In the identification in the limit of a finite state black box, if $U_k^* = \{\phi\}$ and $V_k^*$ completely characterizes the states of $B$, then the operation of adding states will increase $U_k^*$ until it includes a representative of every state of $B$.

*Proof.* When the operation of adding states ends, each state of $U_k^{*\prime}$ will have the same representation as one of the states of $U_k^*$. Since $V_k^*$ is assumed to completely characterize the states of $B$, this means that the $U_k^{*\prime}$ states are included in the $U_k^*$ states. That is, the state transition function takes the $U_k^*$ states into themselves. Therefore, the $U_k^*$ states includes all reachable states of $B$. Note that all states of $B$ are reachable.

*Theorem* 5

If $V_k^*$ completely characterizes the states of $B$ and the elements of $V_k^*$ are the initial elements of $V_{k'}^*$, then after the operation of deleting superflous characteristics the states of $B$ will be given the same representations in both cases.

This theorem shows that the representations for $B$ which are derived by the identification in the limit algorithms are not only correct when $V_k^*$ is big enough but are also invariant after that.

*Proof.* Since $V_k^*$ already distinguishes between every pair of different states of $B$, the experiments which come after $V_k^*$ in $V_k^{*\prime}$ can't add any information, and so will be deleted first, since the deletion starts from the end. When they are deleted, $V_k^{*\prime}$ will be reduced exactly to $V_k^*$ and the deletions will be the same from then on.

## CONCLUSIONS

System identification algorithms are proposed using the following two concepts: (1) A set of experiments can be found which characterizes the states of an unknown black box. (2) When the states have been characterized in this experimentally determinable manner they are observable, so the dynamics of the black box can be determined directly. Precedents for these two concepts are cited at the end of the Introduction.

Arbib and Zeiger, in their paper referenced in the Introduction, have discussed the construction of the same type of algorithm motivated by a somewhat different viewpoint: Whereas I propose choosing a finite set of experiments to "characterize the states of the black box" by defining observable properties of the black box states, Arbib and Zeiger think of the experiments as partitioning the set of possible input strings into equivalence classes. Their viewpoint arose in the application of automata theory to formal linguistics by RABIN and SCOTT [19], where the primary objective was the classification of input strings. Such a viewpoint becomes unwieldy when we consider continuous inputs, outputs and time, so that we are led to consider a continuum of equivalence classes of continuous input functions of continuous time.

In the past, the literature of automata theory has been concerned exclusively with finite identification. That is, an upper bound on the size of state space is specified *a priori* and it is required that the unknown black box be identified in a finite time. In this paper I have emphasized the identification in the limit problem. An upper bound on the size of state space is not specified *a priori*. Instead, we are allowed to collect information indefinitely, updating our guess as to the identity of the unknown black box when necessary. It is required that our guesses be correct after some finite time, and our representations invariant, but we are not required to know when our guess is correct.

The use of the notion of state characterization to construct identification in the limit algorithms introduces problems which do not occur in the case of finite identification: because we are to try to construct representations for the unknown black box without knowing when a sufficient amount of data has been collected, there will be times when the data is insufficient to use the state characterization approach. I have shown how to determine if the data is sufficient to construct a representation and, if not, what data should be obtained. This introduces branching into the data collection in order to hasten convergence.

## REFERENCES

[1] T. W. ANDERSON: *An Introduction to Multivariate Statistical Analysis.* Chapters 11, 12 Wiley, New York (1958).

[2] M. A. ARBIB and H. P. ZEIGER: On the relevance of abstract algebra to control theory. *Automatica* 5, 589–606 (1969).

[3] K. J. ÅSTRÖM and P. EYKOFF: System identification—a survey. *Automatica* 7, 123–162 (1971).

[4] R. B. CATTELL: Factor analysis: an introduction to essentials. *Biometrics* **21**, 190–215 (1965).
[5] H. P. FRIEDMAN and J. RUBIN: On some invariant criteria for grouping data. *J. Am. Statist. Ass.* **62**, 1159–1178 (1967).
[6] E. G. GILBERT: Controllability and observability in multivariable control systems. *J. SIAM Control* **1A**, 128–151 (1963).
[7] A. GILL: State identification experiments in finite automata. *Inform. Control* **4**, 132–154 (1961).
[8] S. GINSBURG: On the length of the smallest uniform experiment which distinguishes the terminal states of a machine. *J. Ass. Computing Machinery* **5**, 266–280 (1958).
[9] E. MARK GOLD: Limiting recursion. *J. Symb. Logic* **30**, 28–48 (1965).
[10] E. MARK GOLD: Language identification in the limit. *Inform. Control* **10**, 447–474 (1967).
[11] E. MARK GOLD: Memory Dynamics, unpublished proposal, Institute for Formal Studies, Los Angeles, California, available from author.
[12] J. C. GOWER: A comparison of some methods of cluster analysis. *Biometrics* **27**, 623–637 (1967).
[13] T. N. HIBBARD: Least upper bounds on minimal terminal state experiments for two classes of sequential machines. *J. Ass. Computing Machinerv* **8**, 601–612 (1961).
[14] B. L. HO: An Effective Construction of Realizations from Input/Output Descriptions, Ph.D. Thesis. Stanford University (1966).
[15] R. E. KALMAN: Mathematical description of linear dynamical systems. *J. SIAM Control* **1A**, 152–192 (196)3.
[16] Z. KOHAVI: *Switching and Finite Automata Theory.* McGraw-Hill, New York (1970).
[17] M. G. KENDALL and A. STUART: *The Advanced Theory of Statistics, vol. 3*, Chapter 43, Hafner (1966).
[18] E. F. MOORE: Gedanken experiments on sequential machines. *Automata Studies*, Princeton University Press (1956).
[19] M. O. RABIN and D. SCOTT: Finite automata and their decision problems. *IBM J. Res. Develop.* 3, 114–125 (1959).
[20] A. P. SAGE and J. L. MELSA: *System Identification.* Academic Press (1971).
[21] A. J. SCOTT and M. J. SYMONS: Clustering methods based on likelihood ratio criteria. *Biometrics* **27**, 387–397 (1971).
[22] D. WISHART: An algorithm for hierarchical classifications. *Biometrics* **25**, 165–170 (1969).
[23] H. P. ZEIGER: Ho's algorithm, commutative diagrams and the uniqueness of minimal linear systems. *Inform. Control* **11**, 71–79 (1967).

**Résumé**—La généralisation d'Arbib et de Zeiger de l'algorithme de Ho pour identification de système est présentée d'un point de vue alternatif appelé "caractérisation d'état". Essentiellement, ce'est une méthode non linéaire non paramétrique pour la conception d'expériences qui détermineront directement une représentation d'espace d'état pour une boite noir inconnue. Cette méthode suppose (1) un espace d'état dimensionel fini est possible, (2) la boite noire inconnue peut être réajustée à son état initial à volonté, de sorte qu'un ensemble arbitraire d'expériences puisse être mené et (3) il n'y a pas de bruit d'installation, de sorte que la boite noire puisse être réajustée à un état donné afin de mener diverses expériences sur l'état'

Dans la discussion intiale informelle, les propriétés de cette méthode d'identification de système sont comparées aux propriétés de la méthode d'estimation de paramètre par minimisation de fonction de perte, par exemple erreur carrée minimum ou possibilité maximum, qui est le seul abordage général à l'identification de système disponible à présent. La minimisation de fonction de perte utilisera toute donnée, mais elle est difficile à calculer et essentiellement paramétrique. La caractérisation d'état élimine la demande en calculs au prix de données spécifiques, et elle est essentiellement non paramétrique. Il est proposé que la caractérisation d'état pourrait avoir des applications pratiques dans la détermination d'une description approximative, d'ordre bas d'un système complexe sur lequel nous n'avons que très peu de renseignements, par exemple en science sociale et en médecine.

La présentation formelle est limitée à des automates d'état finis et à des systèmes linéaires à temps discret. On considerè la situation dans laquelle l'ordre de la boite noire inconnue, la grandeur minimum de l'espace d'état voulu pour le représenter, n'est pas connu à priori; il est par conséquent nécessaire de continuer les expériences indéfiniment, utilisant les résultats pour obtenir des descriptions successivement améliorées de la boite noire inconnue. On présente des algorithmes détaillés pour obtenir la description soit d'un modèle de Moore soit d'un modèle de Mealy. On présente des règles de branchement pour l'emploi de données déjà connues dans le choix des expériences futures afin d'accélerer la convergence. Des théorêmes connus sont utilisés pour prouver qu'après un temps défini ces algorithmes donneront une description correcte, et de nouveaux théorêmes montrent que les représentations subséquentes seront invariantes après qu'une description correcte est obtenue.—bien qu'il n'existe pas de moyen pour le chercheur de savoir quand une description correcte est obtenue.

**Zusammenfassung**—Arbib und Zeiger's Verallgemeinerung von Ho's Algorithmus für die Systemidentifikation wird von einem alternativen "Zustandscharakterisierung" genannten Gesichtspunkt aus dargestellt. Das ist eine wesentlich nichtlineare, nichtparametrische Methode für den Entwurf von Experimenten, die direkt eine Zustandsraumdarstellung für eine unbekannte black box bestimmen. Bei dieser Methode wird angenommen, (1) daß ein Zustandsraum endlicher Dimension möglich ist, (2) daß die unbekannte black box nach Belieben in ihrem Anfangszustand versetzt werden kann, um an dem Zustand verschiedene Experimente durchführen zu können.

Zunächst werden in einer formlosenn skussion die Eigenschaften dieser Methode der Systemidentifikation mit den Eigenschaften der Methode der Parameterschätzung über die Minimierung der Verlustfunktion, d.h. des kleinsten mittleren Quadrates oder der Maximum likelihood Methode, die der einzige gegenwärtig verfügbare Zugang zur Systemidentifikation ist. Zur Minimierung der Verlustfunktionen werden irgendwelche Daten verwendet, doch ist sie rechnerisch schwierig und wesentlich parametrisch. Die Zustandscharakterisierung behebt den Rechenaufwand bei der Ausgabe von angeforderten spezifischen Daten und ist wesentlich nichtparametrisch. Vorgeschlagen wird, daß die Zustandscharakterisierung bei der Bestimmung einer approximierten Besschreibung niederer Ordnung eines komplexen Systems angewandt wird, über das nur geringe vorausgehende Information vorliegt, zum Beispiel in den Sozialwissenschaften und in der Medizin.

Die formale Darstellung ist auf Automaten mit endlichen Zuständen und diskrete zeitlineare Systeme beschränkt. Betrachtet wird die Situation, in der die Ordnung der unbekannten black box (der erforderliche Zustandsraum minimaler Größe um ihn darzustellen) a priori nicht bekannt ist. Daher ist es nötig, mit dem Experimentieren in unbestimmter Weise fortzufahren und die Ergebnisse zu benutzen, um sukzessive bessere Beschreibungen der unbekannten black box zu erhalten. Angegeben werden ausführliche Algorithmen, um entweder die Beschreibung eines Moore-Modells oder eines Mealy-Modells zu erhalten. Verzweigungsregeln zur Benutzung vorheriger Daten und nachfolgender Wahl von Experimenten, um die Konvergenz zu beschleunigen, werden dargestellt. Bekannte Theoreme werden benutzt, um zu beweisen, daß diese Algorithmen nach endlicher Zeit eine korrekte Beschreibung ergeben und neue Theoreme zeigen, dass die folgenden Darstellungen nach Erlangung einer richtigen Beschreibung invariant sein werden (obgleich der Experimentator keine Möglichkeit besitzt zu wissen, wann er eine richtige Beschreibung erreicht hat).

Резюме—Обобщение Арбиба и Цейгера аляоритма Хо для идентификации системы приказано с альтернативной точки зрения, названной "характеризация состояния". Это предсталяет собой в сущности нелинейный, непараметрический метод расчета эспериментов для нечосредственного определения представления пространства состояний для неизвестной черной коробки. Метод предполагает 1. что ограниченное размерное пространство состояний возможно, 2. что неизвестную черную коробку можно произвольно вновь устанавливать в ее начальное состояние, позволяя провести произвольную группу экспериментов и 3. что отсутствует шум объекта, позволяя этим образом вновь устанавливать черную коробку в данное состояние с целью провести разные зкспериментьı в состоянии.

В начальной неформальной дискуссии характеристики этого метода идентификации системы сравнены с характеристиками метода оценки параметров путем минимизации функции ослабления, на пример путем среднеквадратической ошибки или максимальной вероятности, представляющей собой единственный, сейчас доступный общии подход к изентификации системы. Минимизация функции ослабления может пользоваться какими-либо даными но метод труден с вычислительной точки зрения и является в сущости параметрическим. Характеризация состояния исключает вычислительное условие на счет требования специфических данных и является в сущности непараметрической. Предложено что характеризадия состояния может найти практического прменеия в определении приблизительного описания низкой степени сложной системы о которой имеется мало предварительных данных, нанных, например в социальных науках и в медицине.

Формальное представление ограничено на автоматизацию ограниченного состояния и линейные системы с дискретным временем. Рассмотрена систуация в которой степень неизвестной перной коробки—минимальный размер пространства состояний нужный для ее представления—не известный *a приори*. Потому нужно продалжать эксперименты за неопределенное время, пользусь результатами для получения все лучших описаний неизвестной черной коробки. Приказаны детальные алгоритмы для получения описания модели Мора или модели Мили. Приведены правила разветвления для использования прошлых данных при выборе последующих экспериментов с цедью ускорять конвергенцию. Использованы знакомые теормы с целью досказать что после ограниченного времени эти алгоритмы дают точное описание; новые теоремы показывают что последующие представления будут неварнантиые после получения точного описания—хотя экспериментатору не возможно установить когда он получил точное описание.