From Circuit Theory to System Theory*

L. A. ZADEH†, FELLOW, IRE

Summary—The past two decades have witnessed profound changes in the composition, functions and the level of complexity of electrical as well as electronic systems which are employed in modern technology. As a result, classical RLC network theory, which was the mainstay of electrical engineering at a time when RLC networks were the bread and butter of the electrical engineer, has been and is being increasingly relegated to the status of a specialized branch of a much broader discipline—system theory—which is concerned with systems of all types regardless of their physical identity and purpose.

This paper presents a brief survey of the evolution of system theory, together with an exposition of some of its main concepts, techniques and problems. The discussion is centered on the notion of state and emphasizes the role played by state-space techniques. The paper concludes with a brief statement of some of the key problems of system theory.

I. Introduction

The past two decades have witnessed an evolution of classical circuit theory into a field of science whose domain of application far transcends the analysis and synthesis of RLC networks. The invention of the transistor, followed by the development of a variety of other solid-state devices, the trend toward microminiaturization and integrated electronics, the problems arising out of the analysis and design of large-scale communication networks, the increasingly important role played by time-varying, nonlinear and probabilistic circuits, the development of theories of neuroelectric networks, automata and finite state machines, the progress in our understanding of the processes of learning and adaptation, the advent of information theory, game theory and dynamic programming, and the formulation of the maximum principle by Pontryagin, have all combined to relegate classical circuit theory to the status of a specialized branch of a much broader scientific discipline—system theory—which, as the name implies, is concerned with all types of systems and not just electrical networks.

What is system theory? What are its principal problems and areas? What is its relationship to such relatively well-established fields as circuit theory, control theory, information theory, and operations research and systems engineering? These are some of the questions which are discussed in this paper, with no claim that the answers presented are in any way definitive. Technological and scientific progress is so rapid these days that hardly any assertion concerning the boundaries, content and directions of such a new field as system theory can be expected to have long-term validity.

The obvious inference that system theory deals with systems does not shed much light on it, since all branches of science are concerned with systems of one kind or another. The distinguishing characteristic of system theory is its generality and abstractness, its concern with the mathematical properties of systems and not their physical form. Thus, whether a system is electrical, mechanical or chemical in nature does not matter to a system theorist. What matters are the mathematical relations between the variables in terms of which the behavior of the system is described.†

To understand this point clearly, we have to examine more closely the concept of a system. According to Webster's dictionary, a system is "... an aggregation or assemblage of objects united by some form of interaction or interdependence." In this sense, a set of particles exerting attraction on one another is a system; so is a group of human beings forming a society or an organization; so is a complex of interrelated industries; so is an electrical network; so is a large-scale digital computer, which represents the most advanced and sophisticated system devised by man; and so is practically any conceivable collection of interrelated entities of any nature. Indeed, there are few concepts in the domain of human knowledge that are as broad and all-pervasive as that of a system.

It has long been known that systems of widely different physical forms may in fact be governed by identical differential equations. For example, an electrical network may be characterized by the same equations as a mechanical system, in which case the two constitute analogs of one another. This, of course, is the principle behind analog computation.

While the analogies between certain types of systems have been exploited quite extensively in the past, the recognition that the same abstract "systems" notions are operating in various guises in many unrelated fields of science is a relatively recent development. It has been brought about, largely within the past two decades, by the great progress in our understanding of the behavior of both inanimate and animate systems—progress which resulted on the one hand from a vast expansion in the scientific and technological activities directed toward the development of highly complex systems for such purposes as automatic control, pattern recognition, data-processing, communication, and machine computation, and, on the other hand, by the attempts at

* Received by the IRE, August 21, 1961; revised manuscript received, August 25, 1961. The preparation of this paper was supported in part by the National Science Foundation.
† Dept. of Elec. Engrg., University of California, Berkeley, Calif.

† This and the following four paragraphs are reproduced with minor changes from an earlier article by the author: "System theory," Columbia Engng. Quart., vol. 8, pp. 16-19, 34; November, 1954.
quantitative analyses of the extremely complex animate
and man-machine systems which are encountered in
biology, neurophysiology, econometrics, operations re-
search and other fields.

It is these converging developments that have led to
the conception of system theory—a scientific discipline
dedicated to the study of general properties of systems,
regardless of their physical nature. It is to its abstract-
ness that system theory owes its wide applicability, for
it is by the process of abstraction that we progress from
the special to the general, from mere collections of data
to general theories.²

If one were asked to name a single individual who
above anyone else is responsible for the conception of
system theory, the answer would undoubtedly be
"Norbert Wiener," even though Wiener has not con-
cerned himself with system theory as such, nor has he
been using the term "system theory" in the sense em-
ployed in this paper. For it was Wiener who, starting in
the twenties and thirties, has introduced a number of
concepts, ideas and theories which collectively consti-
tute a core of present-day system theory. Among these,
to mention a few, are his representation of nonlinear
systems in terms of a series of Laguerre polynomials and
Hermite functions, his theory of prediction and filtering,
his generalized harmonic analysis, his cinema-
integrable, the Paley-Wiener criterion, and the Wiener
process. It was Wiener who, in the late forties, laid the
foundation for cybernetics—the science of communi-
cation and control in the animal and the machine—of
which system theory is a part dealing specifically with
systems and their properties. It should be noted, how-
ever, that some of the more important recent develop-
ments in system theory no longer bear Wiener’s im-
print. This is particularly true of the theory of discrete-
state systems, the state-space techniques for continuous
systems, and the theory of optimal control, which is
associated mainly with the names of Bellman and Pon-
tryagin. We shall touch upon these developments later on in the paper.

Among the scientists dealing with animate systems,
it was a biologist—Ludwig von Bertalanffy—who long
ago perceived the essential unity of systems concepts
and techniques in various fields of science and who in
writings and lectures sought to attain recognition for
"general systems theory" as a distinct scientific dis-
cipline.³ It is pertinent to note, however, that the
work of Bertalanffy and his school, being motivated
primarily by problems arising in the study of biological
systems, is much more empirical and qualitative in spirit
than the work of those system theorists who received
their training in the exact sciences. In fact, there is a

fairly wide gap between what might be regarded as
"animate" systems and "inanimate" systems theorists at the present time, and it is not at all certain
that this gap will be narrowed, much less closed, in the
near future. There are some who feel that this gap re-
flects the fundamental inadequacy of the conventional
mathematics—the mathematics of precisely-defined
points, functions, sets, probability measures, etc.—for
coping with the analysis of biological systems, and that
to deal effectively with such systems, which are gen-
erally orders of magnitude more complex than man-
made systems, we need a radically different kind of
mathematics, the mathematics of fuzzy or cloudy
quantities which are not describable in terms of prob-
ability distributions. Indeed, the need for such mathe-
ematics is becoming increasingly apparent even in the
realm of inanimate systems, for in most practical cases
the a priori data as well as the criteria by which the
performance of a man-made system is judged are far
from being precisely specified or having accurately-
known probability distributions.

System theory is not as yet a well-crystallized body of
concepts and techniques which set it sharply apart from
other better established fields of science. Indeed, there
is considerable overlap and interrelation between sys-
tem theory on the one hand and circuit theory, informa-
tion theory, control theory, signal theory, operations re-
search and systems engineering on the other. Yet, sys-
tem theory has a distinct identity of its own which per-
haps can be more clearly defined by listing its principal
problems and areas. Such a list is presented below,
without claims that it is definitive, complete and non-
controversial. (To avoid misunderstanding, brief ex-
planations of the meaning of various terms are given in
parentheses.)

Principal Problems of System Theory

1) System characterization (representation of input-
output relationships in mathematical form; tran-
sition from one mode of representation to
another).

2) System classification (determination, on the basis
of observation of input and output, of one among
a specified class of systems to which the system
under test belongs).

3) System identification (determination, on the basis
of observation of input and output, of a
system within a specified class of systems to
which the system under test is equivalent; de-
termination of the initial or terminal state of the
system under test).

4) Signal representation (mathematical representa-
tion of a signal as a combination of elementary
signals; mathematical description of signals).

5) Signal classification (determination of one among
a specified set of classes of signals or patterns to
which an observed signal belongs).

² To quote A. N. Whitehead, "To see what is general in what is
particular, and what is permanent in what is transitory is the aim of
modern science."

³ Dr. Bertalanffy is a founder of the Society for General Systems
Research, which publishes a yearbook and has headquarters at the
Menninger Foundation, Topeka, Kans.
6) System analysis (determination of input-output relationships of a system from the knowledge of input-output relationships of each of its components).

7) System synthesis (specification of a system which has prescribed input-output relationships).

8) System control and programming (determination of an input to a given system which results in a specified or optimal performance).

9) System optimization (determination of a system within a prescribed class of systems which is best in terms of a specified performance criterion).

10) Learning and adaptation (problem of designing systems which can adapt to changes in environment and learn from experience).

11) Reliability (problem of synthesizing reliable systems out of less reliable components).

12) Stability and controllability (determination of whether a given system is stable or unstable, controllable—subject to specified constraints—or not controllable).

**Principal Types of Systems**

1) Linear, nonlinear.

2) Time-varying, time-invariant.

3) Discrete-time (sampled-data), continuous-time.

4) Finite-state, discrete-state, continuous-state.

5) Deterministic (nonrandom), probabilistic (stochastic).

6) Differential (characterized by integro-differential equations), nondifferential.

7) Small-scale, large-scale (large number of components).

**Some Well-Established Fields Which May Be Regarded as Branches of System Theory**

1) Circuit theory (linear and nonlinear).

2) Control theory.

3) Signal theory.

4) Theory of finite-state machines and automata.

**Comment 1:** Note that approximation is not listed as a separate problem, as it is usually regarded in classical circuit theory. Rather, it is regarded as something that permeates all the other problems.

**Comment 2:** Note that information theory and communication theory are not regarded as branches of system theory. System theory makes extensive use of the concepts and results of information theory, but this does not imply that information theory is a branch of system theory, or vice versa. The same comment applies to such theories as decision theory (in statistics), dynamic programming, reliability theory, etc.

**Comment 3:** Note that there is no mention in the list of systems engineering and operations research. We regard these fields as being concerned specifically with the operation and management of large-scale man-machine systems, whereas system theory deals on an abstract level with general properties of systems, regardless of their physical form or the domain of application. In this sense, system theory contributes an important source of concepts and mathematical techniques to systems engineering and operations research, but is not a part of these fields, nor does it have them as its own branches.

It would be futile to attempt to say something (of necessity brief and superficial) about each item in the above list. Instead, we shall confine our attention to just a few concepts and problems which play particularly important roles in system theory and, in a way, account for its distinct identity. Chiefly because of limitations of space, we shall not even touch upon a number of important topics such as the design of learning and adaptive systems, the analysis of large-scale and probabilistic systems, the notion of feedback and signal flow graph techniques, etc. In effect, the remainder of this paper is devoted to a discussion of the concept of state and state-space techniques, along with a brief exposition of systems characterization, classification and identification. We have singled out the concept of state for special emphasis largely because one can hardly acquire any sort of understanding of system theory without having a clear understanding of the notion of state and some of its main implications.

**II. STATE AND STATE-SPACE TECHNIQUES**

It is beyond the scope of this presentation to trace the long history of the evolution of the concept of state in the physical sciences. For our purposes, it will suffice to observe that the notion of state in essentially the same form it is used today was employed by Turing [1] in his classical paper, "On computable numbers, with an application to the Entscheidungs problem," in which he introduced what is known today as the Turing machine.\(^4\)

Roughly speaking, a Turing machine is a discrete-time (\(t=0, 1, 2, \cdots\)) system with a finite number of states or internal configurations, which is subjected to an input having the form of a sequence of symbols (drawn from a finite alphabet) printed on a tape which can move in both directions along its length. The output of the machine at time \(t\) is an instruction to print a particular symbol in the square scanned by the machine at time \(t\) and to move in one or the other direction by one square. A key feature of the machine is that the output at time \(t\) and the state at time \(t+1\) are determined by the state and the input at time \(t\). Thus, if the state, the input and the output at time \(t\) are denoted by \(s_t, u_t, \)

and $y_t$, respectively, then the operation of the machine is characterized by:
\begin{align}
  s_{t+1} &= f(s_t, u_t), \quad t = 0, 1, 2, \ldots, \tag{1} \\
  y_t &= g(s_t, u_t) \tag{2}
\end{align}
where $f$ and $g$ are functions defined on pairs of values of $s_t$ and $u_t$. Note that (1) and (2) imply that the output symbols from any initial time $t_0$ on are determined by the state at time $t_0$ and the input symbols from time $t_0$ on.

An important point about this representation, which was not pointed out by Turing, is that it is applicable not only to the Turing machine, but more generally to any discrete-time system. Furthermore, we shall presently see that it is a simple matter to extend (1) and (2) to systems having a continuum of states (i.e., continuous-state systems).

The characterization of a system by means of equations of the form (1) and (2) (to which we will refer as the Turing representation or, alternatively, as the state equations of a system) was subsequently employed by Shannon [2] in his epoch-making paper on the mathematical theory of communication. Specifically, Shannon used (1) and (2) to characterize finite-state noisy channels, which are probabilistic systems in the sense that $s_t$ and $u_t$ determine not $s_{t+1}$ and $y_t$, but their joint probability distribution. This implies that the system is characterized by (1) and (2), with $f$ and $g$ being random functions, or, equivalently, by the conditional distribution function $p(s_{t+1}, y_t|s_t, u_t)$, where for simplicity of notation the same letter is used to denote both a random variable and a particular value of the random variable (e.g., instead of writing $S_t$ for the random variable and $s_t$ for its value, we use the same symbol, $s_t$ for both).

It was primarily Shannon’s use of the Turing representation that triggered its wide application in the analysis and synthesis of discrete-state systems. Worthy of note in this connection is the important work of von Neumann [3] on probabilistic logics, which demonstrated that it is possible, at least in principle, to build systems of arbitrarily high degree of reliability from unreliable (probabilistic) components. Also worthy of note is the not-too-well-known work of Singleton [4] on the theory of nonlinear transducers, in which techniques for optimizing the performance of a system with quantized state space are developed. It should be remarked that the problem of approximating to a system having a continuum of states with a system having a finite or countable number of states is a basic and as yet unsolved problem in system theory. Among the few papers which touch upon this problem, those by Kaplan [5], [6] contain significant results for the special case of a differential system subjected to zero input. A qualitative discussion of the related problem of e-approximation may be found in a paper by Stebakov [7].

There are two important notions that are missing or play minor roles in the papers cited above. These are the notions of equivalent states and equivalent machines which were introduced by Moore [8] and, independently and in a somewhat restricted form, by Huffman [9]. The theory developed by Moore constitutes a contribution of basic importance to the theory of discrete-state systems and, more particularly, the identification problem.

So far, our discussion of the notion of state has been conducted in the context of discrete-state systems. In the case of a differential system, the state equations (1) and (2) assume the form
\begin{align}
  \dot{s}(t) &= f(s(t), u(t)) \tag{3} \\
  y(t) &= g(s(t), u(t)) \tag{4}
\end{align}
where $\dot{s}(t) = \frac{d}{dt} s(t)$, and $s(t)$, $y(t)$ and $u(t)$ are vectors representing the state, the input and output of the system at time $t$. Under various guises, such equations [particularly for the case where $u(t) = 0$] have long been used in the theory of ordinary differential equations, analytical dynamics, celestial mechanics, quantum mechanics, econometrics, and other fields. Their wide use in the field of automatic control was initiated largely in the late forties and early fifties by Soviet control theorists, notably A. I. Lur’ë, M. A. Aizerman, Ya. Z. Tsypkin, A. A. Feldbaum, A. Ya. Lerner, A. M. Letov, N. N. Krasovskii, I. G. Malkin, L. S. Pontryagin, and others. In the United States, the introduction of the notion of state and related techniques into the theory of optimization of linear as well as nonlinear systems is due primarily to Bellman, whose invention of dynamic programming [10] has contributed by far the most powerful tool since the inception of the variational calculus to the solution of a whole gamut of maximization and minimization problems. Effective use of and or important contributions to the state-space techniques in the field of automatic control have also been made by Kalman [11], Kalman and Bertram [12], LaSalle [13], Laning and Battin [14] (in connection with analog simulation), Friedland [15], and others. It is of interest to note, however, that it was not until 1957 that a general method for setting up the state equations of an RLC network was described by Bashkow [16]. An extension of Bashkow’s technique to time-varying networks was recently presented by Kinawawala [17].

Despite the extensive use of the notion of state in the current literature, one would be hard put to find a satisfactory definition of it in textbooks or papers. A reason for this is that the notion of state is essentially a primitive concept, and as such is not susceptible to exact

\[\text{\footnote{By a differential system, we mean a system which is characterized by one or more ordinary differential equations.}}\]
definition. However, it is possible, as sketched below, to define it indirectly by starting with the notion of complete characterization of a system. Specifically, consider a black box $B$ and some initial time $t_0$. We assume that $B$ is associated with three types of time-functions:

1) a controllable variable $u$ \( [\text{i.e., a time-function whose value can be chosen at will from a specified set (input space) for all } t \geq t_0] \);

2) an initially controllable variable $s$ \( [\text{i.e., a time-function whose value can be chosen at will at } t = t_0 \text{ from a specified set (state space), but not thereafter}] \); and

3) an observable variable $y$ \( [\text{i.e., a time-function whose values can be observed for } t \geq t_0, \text{ but over which no direct control can be exercised for } t \geq t_0] \).

Furthermore, we assume that this holds for all values of $t_0$.

If these assumptions are satisfied, then we shall say that $B$ is completely characterized if for every $t \geq t_0$ the value of the output at time $t$, $y(t)$, is uniquely determined by the value of $s$ at time $t_0$ and the values of $u$ over the closed interval $[t_0, t]$. Symbolically, this is expressed by

$$y(t) = B(s(t_0); u(t_0, t))$$

where $u(t_0, t)$ denotes the segment of the time-function $u$ extending between and including the end points $t_0$ and $t$; $s(t_0)$ is the value assumed by $s(t)$ at time $t_0$, and $B(\cdot; \cdot; \cdot)$ is a single-valued function of its arguments.

[Note that $B$ is a functional of $u(t_0, t)$ and an ordinary function of $s(t_0); s(t)$ is usually a vector with a finite number of components.] It is understood that (5) must hold for all possible values of $s(t_0)$ and $u(t_0, t)$ and that to every possible input-output pair $u(t_0, t), y(t_0, t)$ there should correspond a state $s(t)$ in the state space of $B$.

If $B$ is completely characterized in the sense defined above, then $u(t)$, $y(t)$ and $s(t)$ are, respectively, the values of the input, the output and the state of $B$ at time $t$. [The range of values of $s(t)$ constitutes the state-space of $B$. A particular value of $s(t)$, i.e., a particular state of $B$, will be denoted by $q$.] In this way, the input, the output and the state of $B$ are defined simultaneously as byproducts of the definition of complete characterization of $B$.

The intuitive significance of the concept of state is hardly made clear by the somewhat artificial definition sketched above. Essentially, $s(t)$ constitutes a description of the internal condition in $B$ at time $t$. Eq. (5), then, signifies that, given the initial conditions at time $t_0$, and given the values of the input between and including $t_0$ and $t$, we should be able to find the output of $B$ at time $t$ if the system is completely characterized.

With (5) as the starting point, it is a simple matter to demonstrate that (5) can be replaced by an equivalent pair of equations:

$$s(t) = f(s(t_0); u(t_0, t)), \quad t \geq t_0$$

$$y(t) = g(s(t); u(t)),$$

the first of which expresses the state at time $t$ in terms of the state at time $t_0$ and the values of the input between and including $t_0$ and $t$, while the second gives the output at time $t$ in terms of the state at time $t$ and the input at time $t$. Note that these relations are in effect continuous analogs of the Turing representation

$$s_{t+k} = f(s_t, u_t, \cdots, u_{t+k-1})$$

$$y_t = g(s_t, u_t).$$

It will be helpful at this point to consider a simple illustrative example. Let $B$ be the network shown in Fig. 1, in which $u$ is the input voltage and $y$ is the output voltage.

![Network for illustrative example](insert_image)

It can easily be shown that the state of this network is a 2-vector, $s(t)$, whose components can be taken to be $v(t)$ (voltage across $C$) and $i(t)$ (current through $L$). With this choice of $s(t)$, the state equations can readily be set up by inspection. In matrix form, they read:

$$s = As + Bu$$

$$y = as$$

where

$$A = \begin{bmatrix}
1 & -1 \\
CR & C \\
1 & -R \\
L & -L
\end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ RC \end{bmatrix}, \quad a = [0 \ R].$$

On integrating (10), we obtain

$$s(t) = G(t - t_0)s(t_0) + \int_{t_0}^{t} G(t - \xi)Bu(\xi)d\xi$$

$$y(t) = as$$

where the matrix $G$, called the transition matrix of $B$, is given by the inverse Laplace transform of the matrix $(s I - A)^{-1}$. Specifically,

$$G(t) = \mathcal{L}^{-1}\{ (s I - A)^{-1} \}$$

$$= \mathcal{L}^{-1}\begin{bmatrix}
s + \frac{R}{L} & -\frac{1}{C} \\
\det A & \det A \\
\frac{1}{L} & s + \frac{RC}{\det A} \\
\det A & \det A
\end{bmatrix}$$

For simplicity, it is assumed that $B$ is (1) deterministic (i.e., nonrandom), (2) time-invariant, (3) nonanticipative (not acting on the future values of the input), and (4) continuous-time (i.e., with $t$ ranging over the real time-axis).
where
\[
\det A = \left( s + \frac{1}{RC} \right) \left( s + \frac{R}{L} \right) + \frac{1}{LC}.
\]
Note that (12) expresses \( s(t) \) as a function of \( s(t_0) \) and \( u(t_{i-1}, t) \). Thus, (12) and (13) represent, respectively, the state equations (6) and (7) for \( B \).

III. State and System Equivalence

One cannot proceed much further with the discussion of state-space techniques without introducing the twin notions of equivalent states and equivalent systems.

Suppose that we have two systems \( B_1 \) and \( B_2 \), with \( q_1 \) being a state of \( B_1 \) and \( q_2 \) being a state of \( B_2 \). As the term implies, \( q_1 \) and \( q_2 \) are equivalent states if, for all possible input time-functions \( u \), the response of \( B_1 \) to \( u \) starting in state \( q_1 \) is the same as the response of \( B_2 \) to \( u \) starting in state \( q_2 \). Following Moore, \( B_1 \) and \( B_2 \) are said to be equivalent systems if, and only if, to every state in \( B_1 \) there is an equivalent state in \( B_2 \), and vice versa. What is the significance of this definition? Roughly speaking, it means that if \( B_1 \) and \( B_2 \) are equivalent, then it is impossible to distinguish \( B_1 \) from \( B_2 \) by observing the responses of \( B_1 \) and \( B_2 \) to all possible inputs \( u \), if the initial states of \( B_1 \) and \( B_2 \) are not known to the experimenter.

To illustrate, let us consider the two simple networks shown in Fig. 2.

The state of \( B_1 \) is a vector with two components (which may be identified with the currents flowing through the two inductances); the state of \( B_2 \) is a scalar (the current through \( 2L \)). Nevertheless, by writing the state equations for \( B_1 \) and \( B_2 \), it is easy to verify that \( B_1 \) and \( B_2 \) are equivalent in the sense defined above, as well as in the more conventional sense of circuit theory.

On the other hand, consider the constant-resistance network shown in Fig. 3. Its input impedance is equal to unity at all frequencies. Does this mean that \( B \) is equivalent to a unit resistor?

The circuit theorist's answer to this question would be "yes," since in circuit theory two networks are defined to be equivalent if their respective terminal impedances (or admittances) have the same values at all frequencies. By contrast, the system theorist would say "no," since there are states in \( B_1 \), say the state \((0, 1) \) (where the first component is the voltage across \( C \) and the second component is the current through \( L \)), to which no equivalent state in the unit resistor can be found.\(^8\) However, we note that the unexcited state (ground state) of \( B_1 \), which is \((0, 0)\), is equivalent to the ground state of the resistor. Thus, we can assert that, although \( B_1 \) is not equivalent to the unit resistor, it is ground-state equivalent to the unit resistor. In plain words, this means that if \( B \) is initially unexcited, then it will behave like a unit resistor. On the other hand, if \( B \) is initially excited, then it may not behave like a unit resistor.

This simple example shows that the notion of equivalence in circuit theory has a narrower meaning than it does in system theory. More specifically, the equivalence of two networks in the sense of system theory implies, but is not implied by, their equivalence in the sense of circuit theory. In effect, the notion of equivalence in circuit theory corresponds to the notion of ground-state equivalence in system theory.

IV. The Notion of Policy

Another important difference between circuit theory and system theory manifests itself in the way in which the input to a system (circuit) is represented. Thus, in circuit theory it is customary to specify the desired input to a network as a function of time. By contrast, in system theory it is a common practice—particularly in dealing with control problems—to express the input as a function of the state of the system rather than as a function of time.\(^9\) In many ways, this is a more effective representation, since it is natural to base the decision on what input to apply at time \( t \) on the knowledge of the state of the system at time \( t \). Furthermore, in the latter representation (input in terms of state) we have feedback, whereas in the former (input in terms of time) we have not.

To say that the input depends on the state of the system means, more precisely, that the input at time \( t \) is a function of the state at time \( t \), i.e.,

\[
u(t) = \pi(s(t))
\]

where \( \pi \) is a function defined on the state space with values in the input space. This function is referred to as a policy function, or simply a policy. In effect, a policy is

\(^8\) The resistor represents a degenerate case in which the output (say voltage) depends only on the input (say current) and not on the state of the system. In such cases, the choice of state is immaterial. For example, it may be taken to be the current through \( R \).

\(^9\) For simplicity, it is tacitly assumed here that the system as well as its performance criteria are time-invariant; otherwise, the input would be expressed as a function of both the state of the system and time.
a function which associates a particular input with each state of the system.

The notion of policy plays a key role in system theory and, especially, in control theory. Thus, a typical problem in control theory involves the determination of a policy for a given system $B$ which is optimal in terms of a specified performance criterion for $B$. More specifically, the performance criterion associates with each policy $\pi$ a number $Q(\pi)$, which is a measure of the “goodness” of $\pi$. The problem, then, is to find a policy $\pi$ which maximizes $Q(\pi)$. Such a policy is said to be optimal.

As was stated previously, the most effective general technique for solving problems of this nature is provided by Bellman’s dynamic programming. The basis for dynamic programming is the so-called principle of optimality which in Bellman’s words reads: “An optimal policy has the property that whatever the initial state and the initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.”

Needless to say, one can always resort to brute force methods to find a policy $\pi$ which maximizes $Q(\pi)$. The great advantage of dynamic programming over direct methods is that it reduces the determination of optimal $\pi$ to the solution of a succession of relatively simple maximization or minimization problems. In mathematical terms, if the payoff resulting from the use of an optimal policy when the system is initially (say at $t=0$) in state $s(0)$ is denoted by $R(s(0))$, then by employing the principle of optimality one can derive a functional equation satisfied by $R$. In general, such equations cannot be solved in closed form. However, if the dimension of the state vector is fairly small, say less than four or five, then it is usually feasible to obtain the solution through the use of a moderately-sized digital computer. The main limitation on the applicability of dynamic programming is imposed by the inability of even large-scale computers to handle problems in which the dimensionality of the state vector is fairly high, say of the order of 20. A number of special techniques for getting around the problem of dimensionality have recently been described by Bellman [18].

A very basic problem in system theory which has been attacked both by the techniques of dynamic programming [19], [20] and by extensions of classical methods of the calculus of variations [21] is the so-called minimal-time or optimal-time problem. This problem has attracted a great deal of attention since the formulation of the so-called maximum principle by Pontryagin [20] in 1956, and at present is the object of numerous investigations both in the United States and the Soviet Union. Stated in general terms (for a time-invariant, single-input, continuous-time system) the problem reads as follows.

Given: 1) A system $B$ characterized by the vector differential equation

$$\dot{x}(t) = f(x(t), u(t))$$

where $x(t)$ and $u(t)$ represent, respectively, the state and the input of $B$ at time $t$. ($x$ is a vector; $u$ is a scalar, and $f$ is a function satisfying certain smoothness conditions.)

2) A set of constraints on $u$, e.g., $|u(t)| \leq 1$ for all $t$ or $|u(t)| \leq 1$ and $|\dot{u}(t)| \leq 1$ for all $t$.

3) A specified initial state $x(0) = q_0$ and a specified terminal state $q_1$.

Find an input $u$ (satisfying the prescribed constraints) which would take $B$ from $q_0$ to $q_1$ in the shortest possible time. This, in essence, is the minimal-time problem.

In a slightly more general formulation of the problem, the quantity to be minimized is taken to be the cost of taking the system from $q_0$ to $q_1$, where the cost is expressed by an integral of the form

$$C(u; q_0, q_1) = \int_{t_0}^{t_1} f_0(x(t), u(t))dt.$$  \hspace{1cm} (17)

In this expression, $f_0$ is a prescribed function, $t_1$ is the time at which $B$ reaches the state $q_1$, and $C(u; q_0, q_1)$ denotes the cost of taking $B$ from $q_0$ to $q_1$, when the input $u$ is used.

It is not hard to see why the minimal-time (or, more generally, the minimal-cost) problem plays such an important role in system and, more particularly, control theory. Almost every control problem encountered in practice involves taking a given system from one specified state to another. The minimal time-problem merely poses the question of how this can be done in an optimal fashion.

Various special cases of the minimal time problem were considered by many investigators in the late forties and early fifties. What was lacking was a general theory. Such a theory was developed in a series of papers by Pontryagin, Boltyanskii, and Gamkrelidze [21], [22].

The maximum principle of Pontryagin is essentially a set of necessary conditions satisfied by an optimal input $u$. Briefly stated, let $\psi$ be a solution of the variational system

$$\dot{\psi} = - \left( \frac{\partial f}{\partial x} \right)^T \psi$$ \hspace{1cm} (18)

where $[\partial f/\partial x]^T$ is the transpose of the matrix $[\partial f_i/\partial x_j]$, in which $f_i$ and $x_j$ are, respectively, the $i$th and $j$th components of $f$ and $x$ in the equation $\dot{x} = f(x, u)$. Construct a Hamiltonian function $H(x, \psi, u) = \psi \cdot \dot{x}$ (dot product of $\psi$ and $\dot{x}$), with the initial values of $\psi$ in (18) constrained by the inequality $H(x(0), \psi(0), u(0)) \geq 0$. The

10 It is tacitly assumed that $B$ is a deterministic system. Otherwise, $Q(\pi)$ would be a random variable, and the performance of $B$ would generally be measured by the expected value of $Q(\pi)$.

11 More generally, there might be additional constraints imposed on $x$. This case is not considered here.
maximum principle asserts that if \( \hat{u}(t) \) is an optimal input, then \( \hat{u}(t) \) maximizes the Hamiltonian \( H(x, \psi, u) \), with \( x \) and \( \psi \) held fixed for exact \( t \).

An application of the maximum principle to a linear system characterized by the vector equation

\[
\dot{x} = Ax + Bu,
\]

(19)

where \( A \) is a constant matrix and \( B \) is a constant vector, leads to the result that an optimal input is "bang-bang," that is, at all times the input is as large amplitude-wise as the limits permit. More specifically, an optimal input is of the form

\[
\hat{u}(t) = \text{sgn} (\dot{\psi}(t) \cdot B)
\]

(20)

where \( \text{sgn} \) stands for the function \( \text{sgn} x = 1 \) if \( x > 0 \), \( \text{sgn} x = -1 \) if \( x < 0 \), and \( \text{sgn} x = 0 \) if \( x = 0 \), and \( \dot{\psi} \) is a solution of the adjoint equation

\[
\dot{\psi} = -A^* \psi
\]

(21)

satisfying the inequality

\[
\psi(0) \cdot [Ax(0) + Bu(0)] \geq 0.
\]

(22)

This and other general results for the linear case were first obtained by Pontryagin and his co-workers. Some what more specialized results had been derived independently by Bellman, Glicksberg, and Gross [25].

The main trouble with the maximum principle is that it yields only necessary conditions. The expression for \( \hat{u}(t) \) given by (20) is deceptively simple; in fact, in order to determine \( \hat{u}(t) \), one must first find a \( \psi(t) \) which satisfies the differential equation (21), the inequality (22), and, furthermore, is such that \( B \) reaches \( q_1 \) when subjected to the \( \hat{u} \) given by (20). Even then, there is no guarantee that \( \hat{u} \) is optimal, except when either the initial state or the terminal state coincides with the origin. Still another shortcoming of the method is that the optimal input is obtained as a function of time rather than the state of the system.

One could hardly expect the maximum or any other principle to yield complete and simple solutions to a problem as difficult as the minimal-time problem for nonlinear, continuous-state, continuous-time systems. Actually, complete solutions can be and have been obtained for simpler types of systems. Particularly worthy of note is the solution for the case of a linear discrete-time system which was recently obtained by Desoer and Wing [26]. Quite promising for linear continuous-time systems is the successive approximation technique of Bryson and Ho [27]. In the case of systems having a finite number of states, the minimal-time problem can be solved quite easily even when the system is probabilistic and the terminal state changes in a random (Markovian) fashion [28].

Closely related to the minimal-time problem are the problems of reachability and controllability, which involve the existence and construction of inputs which take a given system for one specified state to another, not necessarily in minimum time. Important contributions to the formulation and solution of these problems for unconstrained linear systems were made by Kalman [29]. It appears difficult, however, to obtain explicit necessary and sufficient conditions for reachability in the case of constrained, linear, much less nonlinear, systems.

V. Characterization, Classification and Identification

Our placement of system characterization, classification and identification at the top of the list of principal problems of system theory (see Section I) reflects their intrinsic importance rather than the extent of the research effort that has been or is being devoted to their solution. Indeed, it is only within the past decade that significant contributions to the formulation as well as the solution of these problems, particularly in the context of finite-state systems, have been made. Nevertheless, it is certain that problems centering on the characterization, classification and, especially, identification of systems as well as signals and patterns, will play an increasingly important role in system theory in the years to come.

The problem of characterization is concerned primarily with the representation of input-output relationships. More specifically, it is concerned both with the alternative ways in which the input-output relationship of a particular system can be represented (e.g., in terms of differential equations, integral operators, frequency response functions, characteristic functions, state equations, etc.), and the forms which these representations assume for various types of systems (e.g., continuous-time, discrete-time, finite-state, probabilistic, finite-memory, nonanticipative, etc.). Generally, the input-output relationship is expressed in terms of a finite or at most countable set of linear operations (both with and without memory) and nonlinear operations (without memory). For example, Cameron and Martin [30] and Wiener [31] have shown that a broad class of nonlinear systems can be characterized by (ground-state) input-output relationships of the form

\[
y(t) = \sum_{n=0}^\infty A_n X_n(t)
\]

(23)

where the \( X_n(t) \) represent products of Hermite functions of various orders in the variables \( z_1, z_2, \ldots \), which in turn are linearly related to \( u \) (input) through Laguerre functions. Note that the operations involved in this representation are 1) linear with memory, \( \text{viz.} \), the rela-

---

\[13\] A system has finite memory of length \( T \) if its output at time \( t \) is determined by \( u(t-T) \).

\[14\] A system is nonanticipative if, roughly speaking, its output depends only on the past and present, but not future, values of the input.
tions between the \( z' \)s and \( u' \); 2) memoryless nonlinear, viz., the relations between the \( X_a \) and \( z_1, z_2, \ldots \); and 3) linear with no memory, viz., the summations. In this connection, it should be pointed out that the basic idea of representing a nonlinear input-output relationship as a composition of an infinite number of 1) memoryless nonlinear operations and 2) linear operations with memory, is by no means a new one. It had been employed quite extensively by Volterra [32] and Frechet [33] near the turn of the century.14

A key feature of the Wiener representation is its orthogonality [meaning uncorrelatedness of distinct terms in (23)] for white noise inputs. This implies that a coefficient \( A_k \) in (23) can be equated to the average (expected) value of the product of \( X_s(t) \) and the response of the system to white noise. In this way, a nonlinear system can be identified by subjecting it to a white noise input, generating the \( X_s(t) \) functions, and measuring the average values of the products \( y(t)X_s(t) \). However, for a variety of technical reasons this method of identification is not of much practical value.

The problem of system classification may be stated as follows. Given a black box \( B \) and a family (not necessarily discrete) of classes of systems \( C_1, C_2, \ldots \), such that \( B \) belongs to one of these classes, say \( C_k \), the problem is to determine \( C_k \) by observing the responses of \( B \) to various inputs. Generally, the inputs in question are assumed to be controllable by the experimenter. Needless to say, it is more difficult to classify a system when this is not the case.

A rather important special problem in classification is the following. Suppose it is known that \( B \) is characterized by a differential equation, and the question is: What is its order? Here, \( C \) can be taken to represent the class of systems characterized by a single differential equation of order \( n \). An interesting solution to this problem was described by Bellman [34].

Another practical problem arises in the experimental study of propagation media. Suppose that \( B \) is a randomly-varying stationary channel, and the problem is to determine if \( B \) is linear or nonlinear. Here, we have but two classes: \( C_1 = \) class of linear systems, and \( C_2 = \) class of nonlinear systems. No systematic procedures for the solution of problems of this type have been developed so far.

Finally, the problem of system identification is one of the most basic and, paradoxically, least-studied problems in system theory. Broadly speaking, the identification of a system \( B \) involves the determination of its characteristics through the observation of responses of \( B \) to test inputs. More precisely, given a class of systems \( C \) (with each member of \( C \) completely characterized), the problem is to determine a system in \( C \) which is equivalent to \( B \). Clearly, the identification problem may be regarded as a special case of the classification problem in which each of the classes \( C_1, C_2, \ldots \), has just one member. This, however, is not a very useful viewpoint.

It is obvious that such commonplace problems as the measurement of a transfer function, impulse response, the \( A_k \) coefficients in the Wiener representation (23), etc., may be regarded as special instances of the identification problem. So is the problem of location of malfunctioning components in a given system \( B \), in which case \( C \) is the class of all malfunctioning versions of \( B \).

A complicating feature of many identification problems is the lack of knowledge of the initial state of the system under test. Another source of difficulties is the presence of noise in observations of the input and output. For obvious reasons, the identification of continuous-state continuous-time systems is a far less tractable problem than the identification of finite-state discrete-time systems. For the latter, the basic theory developed by Moore [8] provides very effective algorithms in the case of small-scale systems, that is, systems in which the number of states as well as the number of input and output levels is fairly small. The identification of large-scale systems calls for, among other things, the development of algorithms which minimize the duration of (or the number of steps in) the identifying input sequence. With the exception of an interesting method suggested by Bellman [35], which combines dynamic programming with the minimax principle, little work along these lines has been done so far.

Another important area which is beginning to draw increasing attention is that of the identification of randomly-varying systems. Of particular interest in this connection is the work of Kailath [36] on randomly-varying linear systems, the work of Hofstetter [37] on finite-state channels, the work of Gilbert [38] on functions of a Markov process, and the generalization by Carlyle [39] of some aspects of Moore's theory to probabilistic machines. All in all, however, the sum total of what we know about the identification problem is far from constituting a body of effective techniques for the solution of realistic identification problems for deterministic, much less probabilistic, systems.

VI. Concluding Remarks

It is difficult to do justice to a subject as complex as system theory in a compass of a few printed pages. It should be emphasized that our discussion was concerned with just a few of the many facets of this rapidly-developing scientific discipline. Will it grow and acquire a distinct identity, or will it fragment and become submerged in other better-established branches of science? This writer believes that system theory is here to stay, and that the coming years will witness its evolution into a respectable and active area of scientific endeavor.
REFERENCES


