MEMORANDUM
RM-4769-PR
NOVEMBER 1966

ON THE CONSTRUCTION OF A MATHEMATICAL THEORY OF THE IDENTIFICATION OF SYSTEMS

Richard Bellman

PREPARED FOR:

UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA • CALIFORNIA

MEMORANDUM
RM-4769-PR
NOVEMBER 1966

ON THE CONSTRUCTION OF A MATHEMATICAL THEORY OF THE IDENTIFICATION OF SYSTEMS

Richard Bellman

This research is supported by the United States Air Force under Project RAND—Contract No. AF 49(638)-1700—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA • 90406

PREFACE

This Memorandum discusses some of the problems involved in the formulation of a mathematical theory for system identification.

Professor Bellman, of the University of Southern California, is a consultant to The RAND Corporation.

SUMMARY

Let us consider a complex system consisting of a set of interacting subsystems. Let $x_i(t)$ represent the state vector for the i-th subsystem, and suppose that the behavior over time of these vectors is determined by a set of coupled functional equations

$$\frac{dx_i}{dt} = g_i(x_1, x_2, \ldots, x_5, a_i), \quad i = 1,2,3,4,5,$$

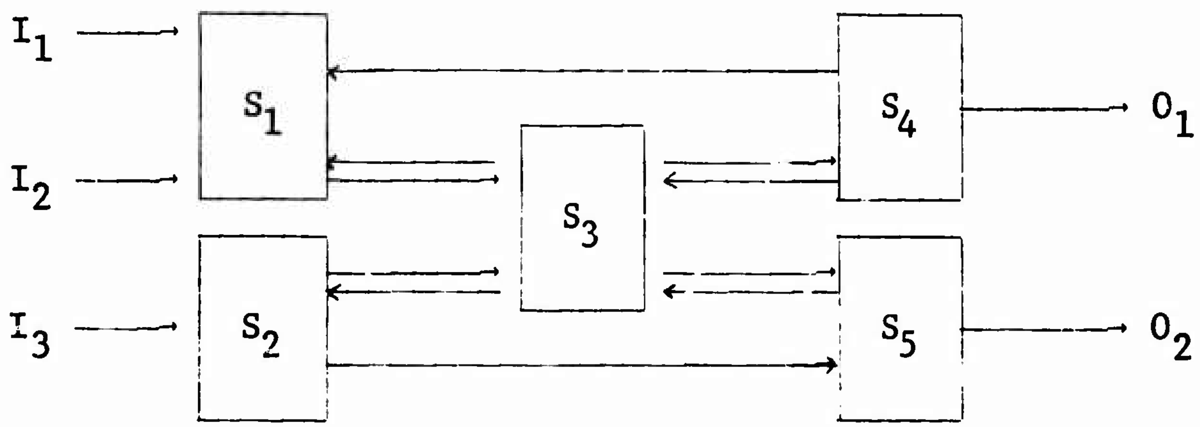
where the a_i are parameters determining both the structure of the subsystems and the linkage between these subsystems.

Much of classical analysis is devoted to the qualitative and quantitative analysis of the $x_i(t)$ as functions of t and of the structural parameters. Important as this effort is, it represents only a part of the principal objective, which is that of the identification of physical systems. By the "identification of systems," we mean the task of determining the structural parameters on the basis of observations over time and position of the inputs and outputs. This is an essential part of the validations of hypotheses and theories.

ON THE CONSTRUCTION OF A MATHEMATICAL THEORY OF THE IDENTIFICATION OF SYSTEMS

1. INTRODUCTION

Let us consider a complex system consisting of a set of interacting subsystems:



Let $x_i(t)$ represent the state vector for the i-th subsystem, and suppose that the behavior over time of these vectors is determined by a set of coupled functional equations

$$(1.1) \qquad \frac{dx_i}{dt} = g_i(x_1,x_2,\ldots,x_5,a_i), \quad i = 1,2,3,4,5,$$

where the a_i are parameters determining both the structure of the subsystems and the linkage between these subsystems.

Much of classical analysis is devoted to the qualitative and quantitative analysis of the $x_i(t)$ as functions of t and of the structural parameters.

Important as this effort is, it represents only a part of the principal objective, which is that of the identification of physical systems. By the "identification of systems," we mean the task of determining the structural parameters on the basis of observations over time and position of the inputs and outputs. This is an essential part of the validations of hypotheses and theories.

There are many levels to this identification process. Let us cite some of the problems encountered at the lowest level:

1. Given a set of observations, $\{(x_i(t_j), c_{ij})\}$, $i = 1,2,\ldots,M$, $j = 1,2,\ldots,N$, to what extent are the a_{ij} determined, and how do we construct algorithms for their determination?

2. Where should the inputs to the system be applied and the outputs measured? How does one do this so as to minimize the disturbance both to behavior of the system and to the calculated values caused by the testing process?

3. At what times should the observations be made?

4. What accuracy is required in measurement to obtain a desired accuracy for the a_i?

5. What kinds of measuring instruments should be used and what types of stimuli applied? Analytically, this may be considered to be a determination of the functions φ_k in the observations $\varphi_k(x_i(t_j))$.

6. What data are desired? Analytically, this may be considered to be a determination of functions $\psi_j(a_i)$.

These questions, difficult as they are, are only preliminary to the study of enveloping problems of identification and control. These, in turn, introduce interesting and recondite stochastic and adaptive aspects; see [1].

All of this in its turn is prologomena to the construction of a meaningful theory of information, and, finally, an embracing theory of instrumentation and experimentation.

In what follows, we wish to describe briefly some mathematical techniques we have employed to treat the first of the problems posed above, and to indicate some of the applications we have made. As the reader will see, the mathematical level is not very high. What we have been particularly concerned with is an examination of how far one can penetrate into this domain with relatively simple analytic techniques tailored to the use of a contemporary digital computer. Essentially, we have attempted to construct a handbook of soluble problems [2], and also to carry out systematic mathematical experimentation in this most important scientific area. What is remarkable is that many significant questions can be resolved using the ideas sketched below.

2. A FUNDAMENTAL PROBLEM

Consider a vector differential equation

$$(2.1) \qquad \dot{x} = g(x,a), \quad x(0) = c,$$

where x is N-dimensional and a is M-dimensional. Suppose that we are given a set of observations $\{(x(t_i),b_i)\}$, $i = 1,2,\ldots,R$, and asked to determine a and c on this basis.

There are several distinct classes of problems that may be distinguished here. The first is that where we suppose that $x(t)$ does indeed satisfy an equation such as (2.1) and that a and c exist. The second is a design problem where a and c are to be determined so that $(x(t_i),b_i) \cong d_i$, $i = 1,2,\ldots,N$. The third is where we know in advance that $x(t)$ does not satisfy an equation of as simple a type as (2.1), but where we want to find the best approximate equation of this structure. This gets us into the area of differential approximation [2,3,4].

Let us write d_i for the actual observation at time i and determine a and c by the condition that the quantity

$$(2.2) \qquad \sum_{i=1}^{R} ((x(t_i),b_i) - d_i)^2$$

be minimized. This analytic problem, which can be handled in various ways as we shall indicate below, is common to all three classes of questions.

3. SEARCH TECHNIQUES

As an indication of a simple and direct approach, let us consider the problem of determining λ in the Van der Pol equation

$$(3.1) \qquad u'' + \lambda(u^2 - 1)u' + u = 0,$$

on the basis of observations, $u(t_i) = d_i$, $i = 1,2,\ldots,R$. Letting $u(t,\lambda)$ denote the unique periodic solution of (3.1), $\lambda > 0$, we wish to minimize

$$(3.2) \qquad f(\lambda) = \sum_{i=1}^{R} (u(t_i,\lambda) - d_i)^2$$

over $\lambda > 0$.

Choosing a value of λ, the determination of the set of values $\{u(t_i,\lambda)\}$ for R of the order of magnitude of 10, 100, 1000, consumes a few seconds of computer time—at most. Consequently, we can begin by evaluating $f(0.1), f(0.2), \ldots, f(1), f(2), \ldots, f(10)$, and thereby determine an interval within which the correct value of λ lies. Having obtained this interval, we can subdivide it into ten parts, and so on. In this way, by means of a few minutes of computer time, we can determine λ accurately to several significant figures.

In more complex, multidimensional minimization problems, it will pay to use sophisticated search methods of the Kiefer-Johnson variety [5].

The point we wish to emphasize is that the capabilities of modern computers frequently permit us to approach problems of this nature in a simple and direct fashion.

4. QUASILINEARIZATION

A systematic approach to the multidimensional minimization problem described in (2.2) is supplied by the theory of quasilinearization [2], which in a number of cases yields techniques which can be considered to be extensions of the Newton-Raphson-Kantorovich approximation method [6].

The basic idea in this case is quite simple. Let $a^{(0)}$, $c^{(0)}$ be initial guesses as to the values of a and c and let $x^{(0)}$ be computed as the solution of

$$(4.1) \qquad \dot{x}^{(0)} = g(x^{(0)}, a^{(0)}), \quad x^{(0)}(0) = c^{(0)}.$$

These values are obtained either on the basis of experience and intuition, through the use of simpler models, or by way of search techniques of the type described above, or a combination of all three. The next approximations are obtained in the following fashion. Expand $g(x,a)$ around $x^{(0)}$, $a^{(0)}$, keeping zero and first-order terms, and let $x^{(1)}$ be determined as the solution of

$$(4.2)\qquad \dot{x}^{(1)} = g(x^{(0)},a^{(0)}) + J_1(x^{(1)} - x^{(0)}) + J_2(a^{(1)} - a^{(0)}),$$

$x^{(1)}(0) = c^{(1)}$. Here J_1 and J_2 are Jacobian matrices. The vectors $a^{(1)}$ and $c^{(1)}$ are now obtained by way of the minimization of the quadratic form

$$(4.3)\qquad \sum_{i=1}^{R} ((x^{(1)}(t_i),b_i) - d_i)^2.$$

We have thus reduced the original optimization problem to a succession of operations involving the numerical solution of linear differential equations and the numerical solution of linear algebraic equations. These operations can be carried out accurately and quickly for systems of quite high dimension. A detailed discussion of this method, with numerous applications, will be found in [2].

5. USE OF TRANSFORM TECHNIQUES

Functional equations of more complex type, such as the heat equation,

$$(5.1)\qquad k(x)u_t = u_{xx},$$

where $k(x) = k(x,a)$, and a is an unknown parameter, or a differential-difference equation

(5.2) $$u'(t) = au(t - \tau),$$

where both a and the time-lag τ are unknown, may be handled directly by means of the techniques described above, or by use of Laplace transform techniques. In this latter case, we use the foregoing method as applied to the resulting differential equations in the transform space [7], [8].

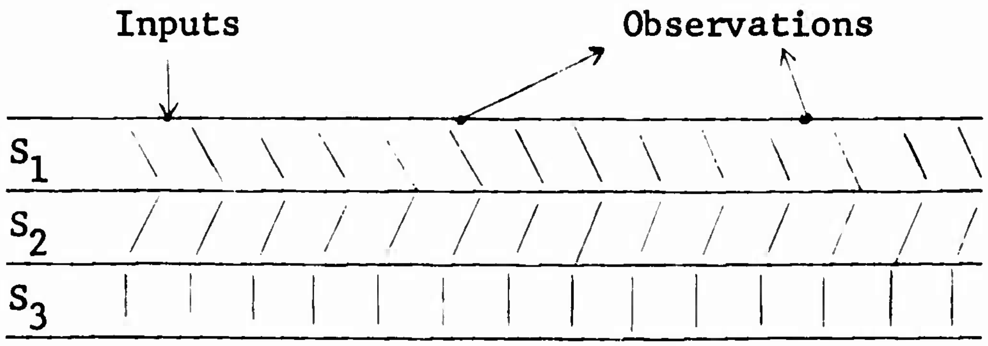
6. BASIC UNCERTAINTIES OF ON-LINE IDENTIFICATION AND CONTROL

Let us consider briefly some measurement and data processing aspects of large systems. Systems with stochastic effects exhibit similar properties. Suppose we are attempting to implement a policy at time t which depends upon a knowledge of the components of the state vector $x(t)$. If the dimension of x is large, we must accept the fact that an appreciable time will be involved in the measurement of all of the components and the transmission of the required information to the decision-maker. During this time the system is uncontrolled, which means that the system is not operating in its optimal fashion. It follows then that there are two extremes: Act instantaneously on the basis of a minimum of information; act on the basis of a maximum information at a later time. Both extreme policies will introduce errors and associated costs. Various simple models of control and decision processes of this type will yield analytic estimates which are similar in structure to the classical uncertainty principles of quantum mechanics.

The multistage nature of the identification and control process enables us to diminish considerably the effect described above. One might expect that similar multistage measurement techniques applied in physics would reduce the uncertainties in quantum mechanics. In any case, there are a number of fascinating questions in these areas associated with the allocation of time, energy, and other resources to most effective identification, which we shall discuss at another time.

7. DECOMPOSITION INTO SUBSYSTEMS

In many cases of interest, the problem is that of determining the structural properties of a system in position or time, or both. Thus, for example, both in geophysics and in astrophysics, we wish to determine the stratification of a medium on the basis of observations made at one interface. The theory of invariant imbedding can be applied in these cases; see [2], [9].



In the study of metabolic processes, and in certain areas of neurophysiology, we encounter a stratification in time. We know that the time history of the system is governed by equations of the following form,

$$\frac{dx}{dt} = g(x,a_i), \quad t_i \leq t \leq t_{i+1}, \quad i = 0,1,2,\ldots, \tag{7.1}$$

and the problem is to determine the structural vectors a_i and the transition times t_i on the basis of a set of observations $\{(x(t_j),b_j)\}$. A discussion of how dynamic programming may be applied to problems of this and the foregoing type will be found in [2,10,11]. Discussions of the application of a combination of dynamic programming and quasilinearization will be found in [12,13,14,15].

8. PATTERN RECOGNITION

It is clear that the problem of decomposition into subsystems is a special case of the general problem of pattern recognition. In turn, pattern recognition can be viewed as a dynamic programming process and particular problems can be handled by means of the functional equation technique; see [16,17,18,19].

9. CONCLUDING REMARKS

We hope that the foregoing has made it clear that the ge l problem of the identification of systems is one of the most important and challenging in modern science. Many new types of mathematical problems arise. Some we have stated explicitly, some we have hinted at, but most have been passed over in silence for want of precise statement or formulation. With practically no past, and a glorious future, this is an ideal field for the young mathematician.

REFERENCES

1. Bellman, R., Adaptive Control Processes: A Guided Tour, Princeton University Press, Princeton, New Jersey, 1961.
2. Bellman, R., and R. Kalaba, Quasilinearization and Nonlinear Boundary-value Problems, American Elsevier Publishing Company, Inc., New York, 1965.
3. Bellman, R., and B. Kotkin, "Differential Approximation Applied to the Solution of Convolution Equations," Math. of Comp., Vol. 18, 1965, pp. 487-491.
4. Bellman, R., and J. M. Richardson, "On Some Questions Arising in the Approximate Solution of Nonlinear Differential Equations," Q. Appl. Math., Vol. 20, 1963, pp. 333-339.
5. Bellman, R., and S. Dreyfus, Applied Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1962.
6. Bellman, R., M. Juncosa, and R. Kalaba, "Some Numerical Experiments Using Newton's Method for Nonlinear Parabolic and Elliptic Boundary-value Problems," Comm. Assoc. Comput. Machinery, Vol. 4, 1961, pp. 187-191.
7. Bellman, R., H. Kagiwada, R. Kalaba, and M. Prestrud, Invariant Imbedding and Time-dependent Processes, American Elsevier Publishing Company, Inc., New York, 1964.
8. Bellman, R., R. Kalaba, and J. Lockett, Numerical Inversion of the Laplace Transform: Applications to Biology, Economics, Engineering, and Physics, American Elsevier Publishing Company, Inc., New York, to appear.
9. Bellman, R., R. Kalaba, and M. Prestrud, Invariant Imbedding and Radiative Transfer in Slabs of Finite Thickness, American Elsevier Publishing Company, Inc., New York, 1963.
10. Bellman, R., "On the Approximation of Curves by Line Segments Using Dynamic Programming," Comm. Assoc. Comput. Machinery, Vol. 4, 1961, p. 284.
11. Bellman, R., B. Gluss, and R. Roth, "Segmental Differential Approximation and the 'Black Box' Problem," J. Math. Anal. Appl., to appear.
12. Bellman, R., "Dynamic Programming, System Identification, and Suboptimization," SIAM J. Control, to appear.

13. Bellman, R., and R. Roth, Segmental Differential Approximation and Biological Systems: An Analysis of a Metabolic Process, The RAND Corporation, RM-4716-NIH, 1965.

14. Roth, R., The Unscrambling of Data: Studies in Segmental Differential Approximation, Avco Corporation, Wilmington, Delaware, TM-65-22, 1965.

15. Bellman, R., and B. Gluss, Adaptive Segmental Differential Approximation, The RAND Corporation, RM-4314-PR, 1965.

16. Bellman, R., "Dynamic Programming, Pattern Recognition, and Location of Faults in Complex Systems," J. Appl. Probability, to appear.

17. ———, "Dynamic Programming, Generalized States, and Switching Systems," J. Math. Anal. Appl., to appear.

18. ———, "On the Application of Dynamic Programming to the Determination of Optimal Play in Chess and Checkers," Proc. Nat. Acad. Sci. USA, Vol. 53, 1965, pp. 244-247.

19. Bellman, R., R. Kalaba, and L. Zadeh, "Abstraction and Pattern Classification," J. Math. Anal. Appl., to appear.

DOCUMENT CONTROL DATA

1. ORIGINATING ACTIVITY	2a. REPORT SECURITY CLASSIFICATION
THE RAND CORPORATION	UNCLASSIFIED
	2b. GROUP

3. REPORT TITLE

ON THE CONSTRUCTION OF A MATHEMATICAL THEORY OF THE IDENTIFICATION OF SYSTEMS

4. AUTHOR(S) (Last name, first name, initial)

Bellman, Richard

5. REPORT DATE	6a. TOTAL No. OF PAGES	6b. No. OF REFS.
November 1966	18	19

7. CONTRACT OR GRANT No.	8. ORIGINATOR'S REPORT No.
AF 49(638)-1700	RM-4769-PR

9a. AVAILABILITY / LIMITATION NOTICES	9b. SPONSORING AGENCY
DDC-1	United States Air Force Project RAND

10. ABSTRACT

A brief discussion of some of the mathematical techniques used in formulating a theory for system identification. Particular attention is given to the question of how far this domain can be penetrated with relatively simple analytic techniques tailored to the use of a contemporary digital computer.

11. KEY WORDS

Mathematics
Numerical methods and processes
Functions
Quasilinearization
Approximation
Computer programming
Control theory
Laplace transform