On Structural Identifiability

R. BELLMAN
Department of Mathematics, Electrical Engineering and Medicine
University of Southern California, Los Angeles, California
AND
K. J. ÅSTRÖM
Division of Automatic Control
Lund Institute of Technology
Lund, Sweden

ABSTRACT

In this article we introduce a new concept, *structural identifiability*, which plays a central role in identification problems. The concept is useful when answering questions such as: To what extent is it possible to get insight into the internal structure of a system from input–output measurements? What experiments are necessary in order to determine the internal couplings uniquely? The definition of the concept of an identifiable structure is given. Criteria as well as certain identifiable structures are discussed. Particular emphasis is given to compartmental models.

1. INTRODUCTION

A goal in many identification problems, in both industry and the biosciences, is to combine a priori knowledge with experimental data. By doing this we are frequently led to a model of the following type.

$$\frac{dx}{dt} = f(x, \theta, u), \qquad y = g(x, \theta, u), \tag{1}$$

where u is the input, y the output, and x the state of the system. The vector θ denotes a set of unknown parameters. The purpose of the identification is to determine these parameters. One way to achieve this is to select an input signal u, measure the corresponding output signal y_m, and

* Supported by the National Institute of Health under grant no. GM 16197.

determine the parameters θ in such a way that the criterion

$$V(\theta) = \int_0^1 (y(t) - y_m(t))^2 dt \tag{2}$$

is minimal. Identification of mathematical models for industrial processes [1, 2] and compartmental analysis [3] are typical examples of this approach.

In this article, we will investigate if it is at all possible to determine the parameters θ using such a procedure. To see that such questions are meaningful, we will first consider two examples.

Example 1

Consider the model

$$\frac{dx}{dt} = ax + bu, \qquad y = cx, \tag{3}$$

where x, u, and y are scalars. Assuming that the initial state of the system is zero, we find that the input–output relation of the system can be written as

$$y(t) = bc \int_0^t \exp[a(t - s)]u(s)\, ds. \tag{4}$$

Since the parameters b and c enter the input–output relation only as the product bc, it is immediately clear that it is not possible to determine all parameters a, b, and c from input–output measurements. In order to get a unique solution it is necessary to specify b or c, or to give a relation between b and c.

Example 2

Consider the compartmental model shown in Fig. 1. Assume that the system is analyzed by injecting a tracer into compartment 1 and taking

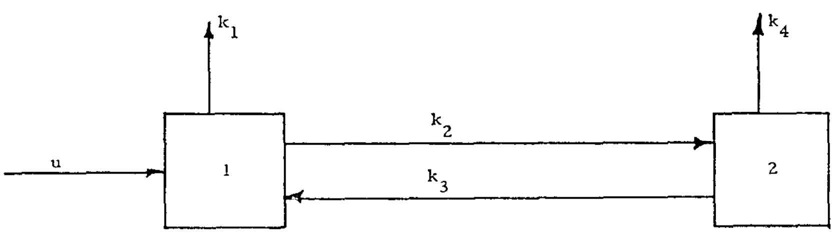

FIG. 1

samples from the same compartment. It will be shown that it is not possible to determine all rate coefficients k_1, k_2, k_3, and k_4 from data obtained in this way.

Let x_1 and x_2 be the concentrations of compartments 1 and 2. The equations describing the compartmental model are then

$$\begin{aligned} \frac{dx_1}{dt} &= -(k_1 + k_2)x_1 + k_3x_2 + u, \\ \frac{dx_2}{dt} &= k_2x_1 - (k_3 + k_4)x_2, \\ y &= x_1. \end{aligned} \tag{5}$$

By taking Laplace transforms we find that the input–output relation of the system (5) can be described by the transfer function

$$G(s) = \frac{[s + k_3 + k_4]}{[s^2 + s(k_1 + k_2 + k_3 + k_4) + (k_1 + k_2)(k_3 + k_4) - k_2k_3]}. \tag{6}$$

We thus find that any parameter combination for which

$$k_3 + k_4 = \text{const.} = b. \tag{7}$$

$$k_1 + k_2 + k_3 + k_4 = \text{const.} = a_1, \tag{8}$$

$$(k_1 + k_2)(k_3 + k_4) - k_2k_3 = \text{const.} = a_2 \tag{9}$$

will give the same input–output relation. Since (7), (8), and (9) give three equations to determine the four rate coefficients k_1, k_2, k_3, and k_4, we find immediately that the rate coefficients cannot be determined uniquely from input–output measurements. A one-parameter family of rate coefficients which has the same input–output relation can be constructed as follows. Select one coefficient arbitrarily. Solve (7)–(9) for the other parameters.

Summing up, we find that it is not possible to determine all rate coefficients of the compartmental model of Fig. 1 from the data obtained by injecting a tracer in vessel 1 and taking samples from the same vessel.

Having realized that the problem is not trivial, let us proceed. The concept of identifiability is introduced in Section 2. The special case of identifiability of linear structures is discussed in Section 3. The following sections are devoted to special identifiable structures such as diagonal structures, companion structures, and structures where all states are directly measurable. Compartmental models are discussed in Section 7.

2. DEFINITIONS

Consider a class of systems $\mathcal{S}$ described by (1) where $\theta \in \tau$. By giving the functions f and g, we thus specify the internal couplings of the system as well as the couplings between the states and the inputs and outputs. For this reason we will call $\mathcal{S}$ a *structure*. A structure can typically arise from a compartmental model of the type illustrated in Fig. 2, where some rate coefficients are specified and others unknown.

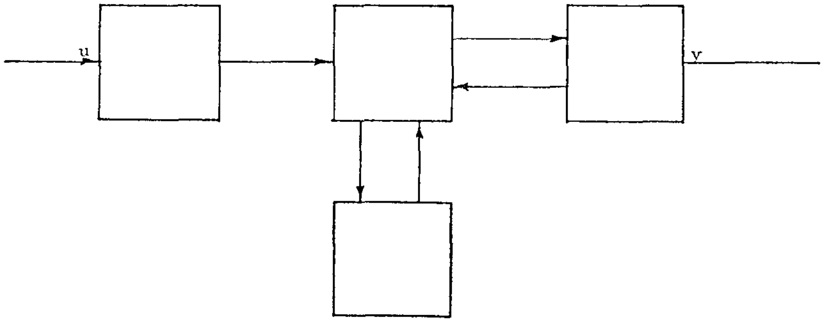

FIG. 2

Efficient numerical methods for minimizing loss functions, such as (2), are well known (see [1, 2]). In this article we pose the problem whether for a given structure $\mathcal{S}$ it is at all possible to find a unique minimum. If this were the case, we could, for example, in the case of linear systems let f and g be arbitrary linear functions. By making experiments and solving the identification problem we would then get the coefficients of the matrices characterizing the linear functions and thus also the internal couplings of the system. Unfortunately, this is not possible to do. Compare Examples 1 and 2. In order to get a unique solution it is necessary to introduce restrictions.

The possibility of finding a unique solution to the identification problem depends not only on the structure of the system but also on the input signal. It is, however, possible to choose universal input signals that can be used for large classes of structures (see [4, 5]). For linear systems we can choose a delta function. We will now specify what is meant by an identifiable structure.

Definition. Assume that the measured output is generated by a system $S_0 \in \mathcal{S}$ with parameters θ_0. The structure $\mathcal{S}$ is called (locally) *identifiable* if the function $V(\theta)$ has a local minimum at $\theta = \theta_0$. If the minimum is global, the structure is said to be *globally identifiable.* We immediately get the following criterion.

COROLLARY. *A sufficient condition for a structure $\mathcal{S}$ to be identifiable is that the matrix of second-order partial derivatives with respect to the parameters, $V_{\theta\theta}$, is positive definite for all $\theta \in \tau$.*

REMARK 1. Notice that with regard to the practical applications it is very important to define identifiability for a class of systems and not for a single system, since we will in practice never know the particular value of θ a priori. If we did, the identification problem would already be solved.

REMARK 2. It is possible to formulate the identification problem in a more complex setting using probabilistic arguments. In a typical case the loss function V can then be taken as the negative logarithm of the likelihood function. In such a situation the expected value of $V_{\theta\theta}$ is equal to the information matrix of the estimation problem and the condition that $V_{\theta\theta}$ is positive definite can be given nice probabilistic interpretations (see [4] and [6]).

It is naturally highly desirable to use identifiable structures when formulating and solving identification problems. If this is not the case, the parameters cannot be determined uniquely. If we attempt to identify a structure that is not identifiable, we will also encounter numerical problems. This is particularly the case in connection with algorithms like the Newton–Raphson method, which make use of the inverse of the matrix $V_{\theta\theta}$.

3. LINEAR STRUCTURES

We now consider linear structures, that is, a class of systems $\mathcal{S}$ whose elements $S = S(A, B, C)$ are described by the equations

$$\frac{dx}{dt} = Ax + Bu, \qquad y = Cx, \tag{10}$$

where u is the input, y the output, and x the state. It is assumed that the state vector is of dimension n. The matrices A, B, and C are assumed to have constant coefficients and $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $C \in \mathcal{C}$, where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are classes of matrices. The class $\mathcal{A}$ may be all 2×2 matrices with a_{11} positive, $a_{12} = 0$, and a_{21} and a_{22} arbitrary. By defining the sets $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ we are thus specifying all internal couplings of the system.

We will now discuss the identifiability of the linear structure $\mathcal{S}$ characterized by (10). We first observe that since the input–output relation of (10) is uniquely determined by the controllable and observable parts of (10), it is necessary to assume that the system (10) is observable and controllable in the sense of Kalman [7] for all $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $C \in \mathcal{C}$.

Also assuming that the system is initially at rest, that is, $x(0) = 0$, the input–output relation is uniquely given by the impulse response of the system. From this we can immediately conclude that if it is at all possible to identify the linear structure, the identification can be achieved from an impulse response measurement. For linear systems it is thus possible to find at least one input signal, a delta function, that can be used universally. Admittedly, if there are disturbances, there might be other signals that are better, but we will not be concerned with that at this stage. We can also make another observation, namely, that all structures whose parameters are uniquely given by the impulse response are identifiable. This proves the heuristic arguments used in the Introduction. After these general observations, we now discuss some specific linear structures.

4. THE DIAGONAL STRUCTURE

Consider a system in diagonal form, that is,

$$\frac{dx}{dt} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} x + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix} u, \tag{11}$$

The transfer function characterizing the input–output relation of the system is

$$G(s) = \sum_{i=1}^{n} \frac{\beta_i}{s + \lambda_i}. \tag{12}$$

If all λ_i are distinct, we find that the parameters β_i and λ_i can be uniquely determined from the impulse response. We can thus conclude that a diagonal structure with distinct eigenvalues is identifiable. We can also prove this directly by showing that the matrix $V_{\theta\theta}$ is positive definite.

5. THE COMPANION STRUCTURE

Consider a system in companion form, that is,

$$\frac{dx}{dt} = \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & \cdots & 0 \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ -a_{n-1} & 0 & 0 & \cdots & 1 \\ -a_n & 0 & 0 & \cdots & 0 \end{bmatrix} x + \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_{n-1} \\ b_n \end{bmatrix} u, \tag{13}$$

$$y = [1 \quad 0 \quad 0 \quad \cdots \quad 0]x.$$

Assuming that the initial state is zero, we find that the input–output relation of the system is given by

$$\frac{Y(s)}{U(s)} = \frac{b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n}{s^n + a_1 s^{n-1} + \cdots + a_n}, \tag{14}$$

where Y and U are the Laplace transforms of y and u. If the polynomials

$$A(s) = s^n + a_1 s^{n-1} + \cdots + a_n \tag{15}$$

and

$$B(s) = b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n \tag{16}$$

do not have any common factors, we find that the parameters a_i and b_i are uniquely given by the transfer function and thus also by the impulse response. We can thus conclude that *the companion structure is identifiable.*

Since the impulse response is invariant under the transformations

$$A^* = TAT^{-1}, \qquad B^* = TB, \qquad C^* = CT^{-1}, \tag{17}$$

we can also conclude that a structure $\mathcal{S} = \{(A^*, B^*, C^*)\}$, where A^*, B^*, C^* are given by (17), T is given, and A, B, and C correspond to a controllable system in companion form, is identifiable.

6. A LINEAR STRUCTURE WHERE ALL STATES ARE DIRECTLY OBSERVABLE

We now consider the particular case of the linear system (10) when the matrix C is of full rank. Without losing generality, we can assume that C equals the identity matrix. If this were not the case we could always introduce another output $y' = Cy$. The impulse response of the system is

$$h(t) = Ce^{At}B = e^{At}B. \tag{18}$$

We now show that if the system is completely controllable in Kalman's sense [7], that is, the matrix $[B, AB, \ldots, A^{n-1}B]$ has rank n, then the corresponding structure is also identifiable. To do this we simply have to show that the matrices A and B can be determined uniquely from the impulse response. We have

$$B = h(0). \tag{19}$$

To determine A we observe that

$$\begin{aligned} h'(0) &= AB = Ah(0) \\ h''(0) &= A^2B = Ah'(0) \\ &\cdot \\ &\cdot \\ &\cdot \\ h^{(n)}(0) &= A^nB = Ah^{(n-1)}(0). \end{aligned} \tag{20}$$

This follows directly by differentiating (18). Hence,

$$A[h(0), h'(0), \ldots, h^{(n-1)}(0)] = [h'(0), h''(0), \ldots, h^{(n)}(0)].$$

Since the system was assumed to be controllable, the matrix

$$[h(0), h'(0), \ldots, h^{(n-1)}(0)] = [B, AB, \ldots, A^{(n-1)}B]$$

is of rank n. Hence,

$$A = [h'(0), h''(0), \ldots, h^{(n)}(0)][h(0), h'(0), \ldots, h^{(n-1)}(0)]^{-1},$$

and we have thus shown how to determine A and B uniquely from the impulse response.

7. SOME COMPARTMENTAL STRUCTURES

We now investigate the identifiability of some compartmental structures. Consider the system with n cascaded compartments shown in Fig. 3.

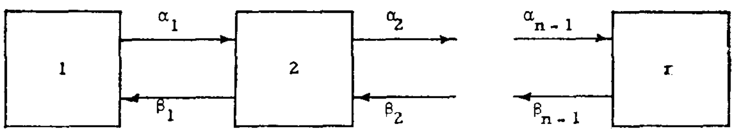

FIG. 3

Assume that the system is analyzed by injecting a tracer into compartment n and that the concentration of the nth compartment is considered as the output. The system can be described by the following equations.

$$\frac{dx}{dt} = \begin{bmatrix} -\alpha_1 & \beta_1 & 0 & \cdots & 0 & 0 \\ \alpha_1 & -(\alpha_2+\beta_1) & \beta_2 & \cdots & 0 & 0 \\ 0 & \alpha_2 & -(\alpha_3+\beta_2) & \cdots & 0 & 0 \\ \cdot & & & & & \\ \cdot & & & \cdots & & \\ \cdot & & & & & \\ 0 & 0 & 0 & \cdots & -(\alpha_{n-1}+\beta_{n-2}) & \beta_{n-1} \\ 0 & 0 & 0 & \cdots & \alpha_{n-1} & -\beta_{n-1} \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{bmatrix} u,$$

$$y = [\,0 \quad 0 \quad 0 \quad \cdots \quad 0 \quad 1\,]x. \tag{21}$$

The transfer function of this system is given by

$$G(s) = \frac{d_{n-1}(s)}{d_n(s)}, \tag{22}$$

where

$$d_n(s) = \det \begin{bmatrix} s+\alpha_1 & -\beta_1 & \dots & 0 & 0 \\ \alpha_1 & s+\alpha_2+\beta_1 & \dots & 0 & 0 \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ 0 & 0 & \dots & s+\alpha_{n-1}+\beta_{n-2} & -\beta_{n-1} \\ 0 & 0 & \dots & -\alpha_{n-1} & s+\alpha_n+\beta_{n-1} \end{bmatrix}. \tag{23}$$

To obtain a symmetrical form, we have introduced $\alpha_n = 0$. Using elementary rules for determinants, we get

$$\begin{aligned} d_1(s) &= s + \alpha_1 \\ d_2(s) &= (s + \alpha_2 + \beta_1)(s + \alpha_1) - \alpha_1\beta_1, \\ &\cdot \\ &\cdot \\ &\cdot \\ d_k(s) &= (s + \alpha_k + \beta_{k-1})\, d_{k-1}(s) - d_{k-1}\beta_{k-1}\, d_{k-2}(s). \end{aligned} \tag{24}$$

The determinants can thus be evaluated recursively. It follows from the equations (22) and (24) that the parameters α_i and β_i can be determined uniquely from the transfer function. Put $k = n$ and we find from (24) that β_{n-1} and $\alpha_{n-1}\beta_{n-1}$ can be determined. Proceeding in this way we can then determine all the other parameters. Notice in particular that we are not using the equation

$$d_1(s) = s + \alpha_1,$$

which means that it would in fact be possible to have an additional parameter corresponding to the compartmental model of Fig. 4.

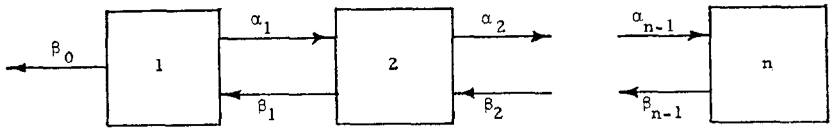

FIG. 4

Summing up, we find that the compartmental models of Fig. 3 and Fig. 4 are identifiable if a tracer is injected in compartment n and if measurements are taken from the same compartment.

It is thus possible to determine all the rate coefficients of the models of Figs. 3 and 4 by a simple experiment, which consists of injecting a tracer into the nth compartment and taking measurements from the same compartment.

Notice, however, that even for the compartmental model of Fig. 3 it is in general not possible to determine all the rate coefficients from an experiment arranged in a different way. This is illustrated by the following counterexample.

Example

Consider a compartmental model according to Fig. 3 with three compartments. Assume that a tracer is injected into the third compartment and that the system is analyzed by taking samples from the first compartment. The equations governing the system then become

$$\frac{dx}{dt} = \begin{bmatrix} -\alpha_1 & \beta_1 & 0 \\ \alpha_1 & -(\alpha_2 + \beta_1) & \beta_2 \\ 0 & \alpha_2 & -\beta_2 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u, \tag{25}$$

$$y = [\ 1 \quad 0 \quad 0\ ].$$

The input-output relation of this linear system is characterized by the transfer function

$$G(s) = \frac{\beta_1\beta_2}{s^3 + s^2(\alpha_1 + \alpha_2 + \beta_1 + \beta_2) + s(\alpha_1\alpha_2 + \alpha_1\beta_2 + \beta_1\beta_2)}. \tag{26}$$

The impulse response measurement or any other identification experiment with the same inputs and outputs gives the following equations.

$$\begin{aligned} \beta_1\beta_2 &= c_1, \\ \alpha_1 + \alpha_2 + \beta_1 + \beta_2 &= c_2, \\ \alpha_1\alpha_2 + \alpha_1\beta_2 + \beta_1\beta_2 &= c_3. \end{aligned} \tag{27}$$

It is obvious that it is not possible to determine the coefficients α_i and β_i uniquely from these equations.

Also notice that if we have a compartmental model according to Fig. 3 with additional couplings (say a direct coupling between compartment 1 and compartment $n - 1$), the corresponding structure will not be identifiable simply for the reason that there are more parameters in the system than the $2n$ parameters that characterize the impulse response.

Notice, however, that if measurements are taken from all compartments, it then follows from the result of Section 6 that if the tracers injection is chosen in such a way that the system is controllable, then the corresponding structure will be identifiable no matter how complex the interactions are.

The question of identifiability of a compartment structure when several but not all compartments are observed is an interesting and important problem that has not yet been resolved.

REFERENCES

1 R. Bellman and R. Kalaba, *Quasilinearization and nonlinear boundary value problems.* American Elsevier, New York, 1965.
2 K. J. Åström and T. Bohlin, Numerical identification of linear dynamic systems from normal operating records, in *Theory of self-adaptive control systems*, (P. H. Hammond, ed.). Plenum Press, New York, 1966.
3 A. Rapaport, Periodicities of open linear systems with positive steady states, *Bull. Math. Biophys.* **14**(1952), 171–183.
4 K. J. Åström, T. Bohlin, and S. Wensmark, *Automatic construction of linear stochastic dynamic models for stationary industrial processes with random disturbances using operating records*, Rep. TP18. 150, IBM Nordic Laboratory, Sweden (June, 1965.)
5 K. J. Åström, *Lectures on the identification problem—The least squares method*, Rep. 6806, Division of Automatic Control, Lund Inst. Technol. (September, 1968.)
6 A. V. Balakrishnan, *Techniques of system identification* (Lecture notes), Rome, September, 1968.
7 R. E. Kalman, Mathematical description of linear dynamical systems, *SIAM J. Control* **1**(1963), 152–192.