NASA CONTRACTOR REPORT 166299

(NASA-CR-166299) A STUDY OF PARAMETER IDENTIFICATION Final Report (Iowa State Univ. of Science and Technology) 118 p HC A06/MF A01 CSCL 12A

N82-20996

Unclas

G3/65 09281

A Study of Parameter Identification

C. J. Herget
R. E. Patterson, III

CONTRACT NCA2- 40501
March 1978

NASA

NASA CONTRACTOR REPORT 166299

A Study of Parameter Identification

C. J. Herget
R. E. Patterson, III
Department of Electrical Engineering
Engineering Research Institute
Iowa State University
Ames, Iowa 50010

Prepared for
Ames Research Center
under Contract NCA2-40501

NASA
National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

A STUDY ON PARAMETER IDENTIFIABILITY

System identification is the process of modeling a system structure, establishing a mathematical representation of that structure and determining the values of the unknown coefficients, or parameters, from experimental input and output data records. In the broad sense, parameter identifiability may be considered as the mathematical assurance of the recoverability of the unknown parameters. Deterministic parameter identifiability pertains to systems in which no corruptive noise processes are present, while stochastic parameter identifiability treats those systems in which noise processes are present, either in the dynamics of the system itself, in the output observation process, or in both.

A set of definitions for deterministic parameter identifiability is proposed based on the necessary injectivity of the mapping from the system composite input/initial condition/parameter space into the system output space. The equivalence of the proposed definitions and of various definitions previously developed is demonstrated. Deterministic parameter identifiability properties are presented based on four system characteristics: direct parameter recoverability, properties of the system transfer function, properties of output distinguishability, and uniqueness properties of a quadratic cost functional.

Stochastic parameter identifiability is defined in terms of the existence of an estimation sequence for the unknown parameters which is consistent in probability. Stochastic parameter identifiability properties are presented based on the following characteristics: convergence properties of the maximum likelihood estimate, properties of the joint probability density functions of the observations, and properties of the information matrix.

Specific parameter identifiability properties for a number of specific systems and classes of systems are presented as theorems and examples.

TABLE OF CONTENTS

1. INTRODUCTION

The problem of system identification has recently become one of the more intensively studied and active fields of engineering and applied mathematics research. The final objective of the system identification problem is the production of an "accurate" model, or mathematical representation, to facilitate the study, understanding and, ultimately, the implementation, modification or improvement of a given system. The lack of such adequate system models is perhaps the greatest limiting obstacle preventing the application of powerful techniques of modern control, estimation, and filtering theory to such diverse areas as biological and human systems, ecological systems, socio-economic systems, and many other complex and multifaceted fields of endeavor not previously associated with exact, mathematical analysis.

The system identification problem is generally considered to consist of three phases. In the first, or modeling phase, the basic mathematical structure of the system is determined. In determining this structure, varying degrees of _a priori_ knowledge of the system may exist. As a minimum, the system observables (input and output variables) must be identified. Data records of these observables must be available or must be obtainable through measurement experiments. If these data records constitute the total _a priori_ knowledge of the system, the analyst is faced with the "total ignorance" or "black box" system identification problem [2].

More common, however, is the "grey box" identification problem in which considerable knowledge of the system variables and internal structural properties is known. In this case, the form of the describing mathematical equations is known or may be readily deduced from the available defining, physical theory. Only this more common "grey box" identification problem will be considered in the following. The structure or form of the mathematical equations defining the system under consideration will be considered to be known.

Having determined the form of the defining system equations, it is then necessary to determine values for the unknown equation coefficients by an analysis of the available input and output data records. The determination of these system parameters is the second phase of the system identification problem. Parameter estimation problems have been extensively investigated in the past, yielding well known results in such areas as least-squares analysis and curve fitting. Inherent to the parameter estimation problem, however, is the preliminary question of whether the system parameters can indeed be found under the given conditions and with the data available. That is, for the system as defined, are the system parameters mathematically identifiable? It is to this question of system parameter identifiability that this paper is addressed.

The final step in the system identification problem is that of model verification. In this step, a final judgment is made of the model's ability to describe adequately the given system in terms of the objectives of the study. Such objectives might include the design of a control strategy for the system, the accurate simulation of the system, or the accurate prediction of the system response to varied inputs.

Obviously, the question of parameter identifiability is critical to the generation of a parameter estimation algorithm. If the parameters are not mathematically identifiable, it is senseless, at best, to attempt the generation of such an algorithm. At worst, the implementation of an estimation algorithm may generate spurious parameter values which may well lead to incorrect conclusions about the system properties.

The question of parameter identifiability may also have a direct bearing on the other two phases of the parameter identification problem. An understanding of the parameter identifiability properties of a given class of systems may guide the investigator in selecting an appropriate system model. Obviously, a model in which the system parameters are not identifiable must be rejected. If the form of the system is well defined by its physical properties, a knowledge of the parameter identifiability properties of that particular class of systems may lead to the proper choice of input signals or to the design of a proper output measurement scheme to insure system parameter identifiability. In like manner, the evaluation of the adequacy of the system model in terms of the ultimate investigation objectives must be considered in light of the limitations imposed by the parameter identifiability requirements.

PRECEDING PAGE BLANK NOT FILMED

2. BASIC CONCEPTS

In its broadest sense, parameter identifiability may be considered as the mathematical assurance of the capability of determining unique values for the unknown parameters of a system from some set of input and output data records. That such assurance can not always be readily established is evident in the following simple examples.

Example 2.1. Consider the simple linear system described by the equations

$$\dot{x} = ax + bu \tag{1}$$

$$y = cx \tag{2}$$

where the output y, the system state x, and the input u are each scalar-valued. The system parameters a, b, and c are to be determined from some set of input and output data. If the initial state of the system is assumed to be zero, the system input-output relation can be immediately written in the familiar form

$$y(t) = bc \int_0^t \exp[a(t-\tau)]u(\tau)d\tau. \tag{3}$$

The system parameters b and c appear only as the product bc and thus can not be separated and determined uniquely from the input-output measurements alone. Any parameter pair (b,c) which satisfies the relationship

$$bc = \text{constant} = k \tag{4}$$

will produce identical input-output records. If an attempt is made to determine the values of b and c using a parameter estimation scheme of the small variations type, such as the Newton or Gauss-Newton methods, the non-unique values generated by such a sequence, if indeed the parameter estimation sequence converges at all, will usually be dependent upon the values assumed for the parameters at the first iteration. To obtain a uniquely described system, it is necessary to specify either b or c or to establish a defining relationship between the two parameters. These properties are demonstrated in Figure 1.

<u>Example 2.2</u>. [3] Figure 2 shows a two-compartment model in which the concentrations of compartments 1 and 2 are designated x_1 and x_2, respectively, and the input and rate coefficients are designated as u and a_1, a_2, a_3, and a_4, respectively. Such a model may be used to represent

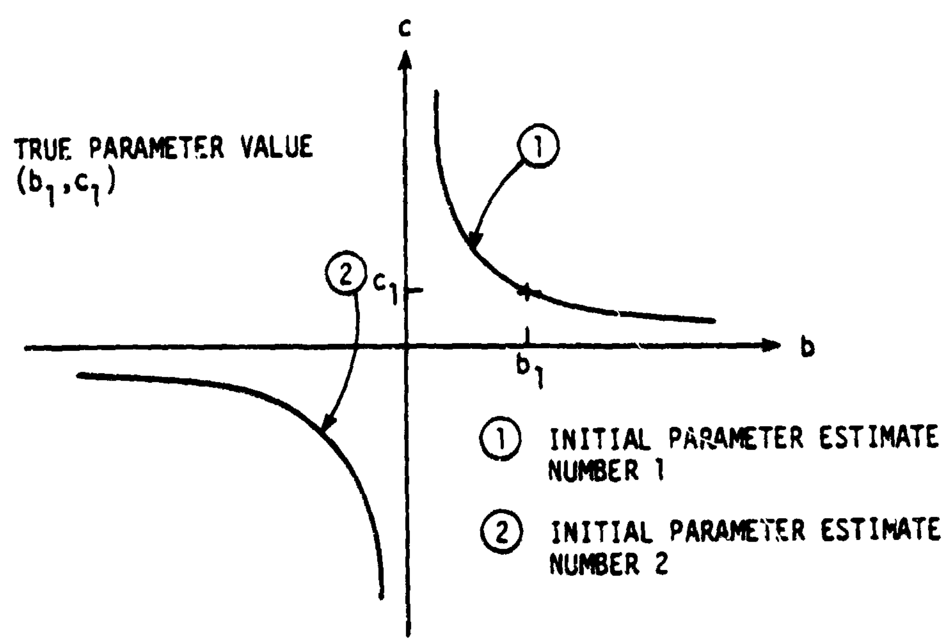

Figure 1. Parameter relationships for Example 2.1.

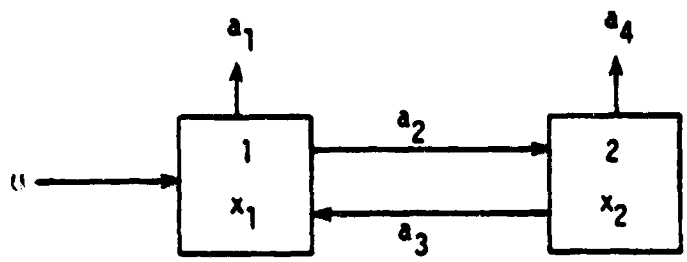

Figure 2. A two-compartment model.

the simplified oxygen transport characteristics of a pulmonary/cardiovascular system in which the pulmonary subsystem is represented by compartment 1 and the cardiovascular subsystem is represented by compartment 2. The oxygen levels or concentrations of the respective subsystems are denoted by x_1 and x_2, and the oxygen transport coefficients by a_1, a_2, a_3, and a_4.

For a given oxygen input, u, the two-compartment model may be analyzed by monitoring the concentrations of either one or both of the compartments. If only the concentration of compartment 1 is monitored (i.e., x_1 is the single output of the system), the equations describing the system may be written as

$$\dot{x}_1 = -(a_1 + a_2)x_1 + a_3x_2 + u \tag{5}$$

$$\dot{x}_2 = a_2x_1 - (a_3 + a_4)\, x_2 \tag{6}$$

$$y = x_1 \; . \tag{7}$$

It is desired to determine the rate coefficients, or system parameters a_1, a_2, a_3 and a_4, from a set of input and output data. Assuming the system has achieved steady state, the input data is directly related to the output data by the system transfer function which can be found to be

$$G(s) = \frac{s + a_3 + a_4}{s^2 + s(a_1 + a_2 + a_3 + a_4) + (a_1 + a_2)(a_3 + a_4) - a_2 a_3} . \quad (8)$$

It is obvious that G(s) will yield the same input-output relationship for any parameter combination for which

$$a_3 + a_4 = \text{constant} = c_1 \quad (9)$$

$$a_1 + a_2 + a_3 + a_4 = \text{constant} = c_2 \quad (10)$$

and

$$(a_1 + a_2)(a_3 + a_4) - a_2 a_3 = \text{constant} = c_3 . \quad (11)$$

As Eqs. (9), (10), and (11) constitute an underdetermined set of three equations in four unknowns, the system parameters can not be uniquely determined from any given set of input and output data. To uniquely describe the system, either the value of one of the system parameters must be specified or a defining relationship among the four system parameters must be specified.

It is desirable that the system property of parameter identifiability be defined independently of the method used to estimate parameter values. Parameter identifiability is considered to be a property of the system itself. It is therefore necessary to explicitly define what is meant by a system or, more exactly, the mathematical representation of a system and to enumerate the basic properties of such a system [12]. In consonance with commonly accepted practice, the mathematical representation of the system is itself referred to as the system in the sequel. A general system is represented schematically by Figure 3 and mathematically by Eq. (12).

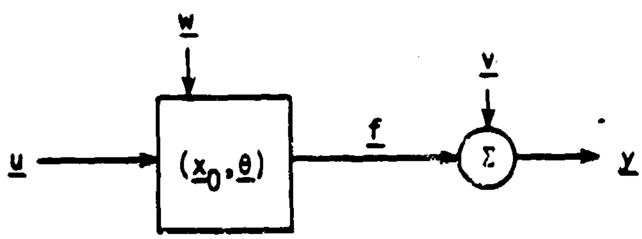

Figure 3. A general system.

$$\underline{y} = \underline{f}(\underline{x}_0, \underline{\theta}, \underline{u}, \underline{w}) + \underline{v} \tag{12}$$

The quantities $\underline{u}$, $\underline{y}$, $\underline{x}$, and $\underline{\theta}$ are the input, output, state, and parameters of the system, and $\underline{w}$ and $\underline{v}$ are noise processes. The system inputs and outputs and, indirectly, the noise processes are functions of time, t, normally defined on $[0,\infty)$ for the continuous case and on $\{k\tau | k \in I^+\}$, I^+ being the non-negative integers, for the discrete case. To encompass these two possibilities, the domain of the system inputs and outputs is designated as T where $T \subset R^+$ and R^+ is the set of non-negative real numbers. The system parameters, $\underline{\theta}$, are specifically time-independent; that is, they are constant.

The following sets and properties pertain to the system described by Eq. (12):

$\mathcal{U}$ is the space of allowable input functions. An element $\underline{u}(\cdot)$ of $\mathcal{U}$ is called a system input and, for any time $t \in T$, $\underline{u}(t)$ is called the value of the input $\underline{u}$ at time t. The input, $\underline{u}$, is of dimension r; specifically, $\underline{u}(t) \epsilon R^r$, where R^r is the space of ordered r-tuples over the field of real numbers.

$\mathcal{Y}$ is the space of output functions. An element of $\mathcal{Y}$ is designated as $\underline{y}(\cdot)$ and is called a system output. For any given time $t \epsilon T$, $\underline{y}(t)$ is called the value of the output $\underline{y}$ at time t. The dimension of the output $\underline{y}$ is m, and $\underline{y}(t) \epsilon R^m$.

X is the system state space for which an element $\underline{x} \epsilon X$ is known as a system state. X is n-dimensional and, in particular, $X \subset R^n$. For any time $t \epsilon T$, $\underline{x}(t) \epsilon X$ is called the system state at time t. In particular, $\underline{x}(0)$ is the initial state of the system and is designated $\underline{x}(0) = \underline{x}_0$. The initial state of the system may be stated more generally as $\underline{x}(t_0)$ for an arbitrary initial time t_0. It should be noted that $\underline{x}_0 \epsilon X$ and is therefore also n-dimensional, i.e., $\underline{x}_0 \epsilon R^n$.

Ω is the space of allowable system parameters, and an element of Ω is designated as $\underline{\theta}$ and is called a system parameter. At this point it should be noted that in generating properties of the general system or in manipulating system elements in generating parameter identifiability properties of the general system, arbitrary norms, designated by $||\cdot||$, may be defined on the subject spaces as required. Ω is a compact subset of R^p, and hence, $\underline{\theta} \epsilon R^p$ is p-dimensional. Arbitrary norms upon R^p and R^n may be considered to have been defined as required by the compactness property. Limited parameter identifiability results have been generated for more general parameter spaces. However, restricting Ω to be a subset of R^p is not limiting for systems that model physically realizable processes, and parameter identifiability in generalized parameter spaces will not be considered.

$\underline{w}$ and $\underline{v}$ are generally additive, white, Gaussian noise processes affecting the system plant or state and the system output, respectively. The output noise process is specifically limited to being additive in nature while the plant noise process may be other than strictly additive in nature. Though not generally the case, either process may exhibit properties other than those of white, Gaussian noise. Each process may be correlated or statistically independent, the latter being far more common.

A fundamental property of any given system in which the noise processes have been excluded is that, for any given initial time $t_0 \epsilon T$, for any given initial state $\underline{x}_0$, for any given parameter $\underline{\theta} \epsilon \Omega$, and for any given input $\underline{u}(\cdot) \epsilon \mathcal{U}$ defined on some interval $[t_0, t]$, both the resulting state $\underline{x}(t)$ and resulting output $\underline{y}(t)$ at some later time t are uniquely determined.

A minimal system or minimal realization is defined to be such that the dimension of the system state space X is less than or equal to that for any other equivalent system. Equivalent systems are defined to be those which generate identical outputs for any given input $\underline{u} \epsilon \mathcal{U}$.

A rich diversity of systems and options are encompassed by Eq. (12) and Figure 3. The system itself may be linear or nonlinear with respect to its state, discrete- or continuous-time, single-input single-output (SISO) or multiple-input multiple-output (MIMO). The form in which the parameters and state variables appear, the parameterization, may be canonical or noncanonical. The importance of a particular given parameterization is readily apparent by comparing the following Example

2.3 with Example 3.5 which involve two seemingly equivalent parameterizations of a linear, single-input single-output, second-order system.

Example 2.3. Consider the linear, SISO, second-order system characterized by the transfer function

$$H(s) = \frac{a_1}{(s - a_2)(s - a_3)} \tag{13}$$

with the corresponding state space formulation

$$\dot{x}_1 = a_3 x_1 + x_2 \tag{14}$$

$$\dot{x}_2 = a_2 x_2 + a_1 u \tag{15}$$

$$y = x_1 \ . \tag{16}$$

The parameter to be determined is $\underline{\theta} = [a_1, a_2, a_3]^T$ which is contained in the parameter space $\Omega = R^3$.

The system may be written in an equivalent state space form as

$$\underline{\dot{x}} = \underline{A}(\underline{\theta})\underline{x} + \underline{b}(\underline{\theta})u \tag{17}$$

$$y = \underline{c}^T \underline{x} \tag{18}$$

where the parameterized, constant, system matrices are

$$\underline{A}(\underline{\theta}) = \begin{bmatrix} a_3 & 1 \\ 0 & a_2 \end{bmatrix}, \quad \underline{b}(\underline{\theta}) = \begin{bmatrix} 0 \\ a_1 \end{bmatrix}, \quad \underline{c} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{19}$$

To show that the given parameterization is not unique, and hence not identifiable in $\Omega = R^3$, it is sufficient to show that there exists an invertible matrix $\underline{P}$ which creates an equivalently parameterized system through the similarity transformation

$$\underline{P}\,\underline{A}\,\underline{P}^{-1} = \underline{A}^* \tag{20}$$

$$\underline{P}\,\underline{b} = \underline{b}^* \tag{21}$$

$$\underline{c}\,\underline{P}^{-1} = \underline{c}^* \tag{22}$$

where the matrices $\underline{A}^*$, $\underline{b}^*$, and $\underline{c}^*$ define the equivalent system.

Consider the matrix

$$\underline{P} = \begin{bmatrix} 1 & 0 \\ a_3-a_2 & 1 \end{bmatrix} \tag{23}$$

which yields its inverse

$$\underline{P}^{-1} = \begin{bmatrix} 1 & 0 \\ a_2-a_3 & 1 \end{bmatrix} \tag{24}$$

Then

$$\underline{A}^* = \underline{P}\,\underline{A}\,\underline{P}^{-1} = \begin{bmatrix} 1 & 0 \\ a_3-a_2 & 1 \end{bmatrix} \begin{bmatrix} a_3 & 1 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ a_2-a_3 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_2 & 1 \\ 0 & a_3 \end{bmatrix} \neq \underline{A} \tag{25}$$

$$\underline{b}^* = \underline{P}\,\underline{b} = \begin{bmatrix} 1 & 0 \\ a_3-a_2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 0 \\ a_1 \end{bmatrix} = \underline{b} \tag{26}$$

$$\underline{c}^* = \underline{c}\,\underline{P}^{-1} = [1 \quad 0] \begin{bmatrix} 1 & 0 \\ a_2 - a_3 & 1 \end{bmatrix} = [1 \quad 0] = \underline{c}\,. \tag{27}$$

Note that the two parameterizations are identical except in the $\underline{A}$ and $\underline{A}^*$ matrices which indicate that the parameters a_2 and a_3 are not uniquely identifiable. The parameterizations characterized by the two triples (a_1, a_2, a_3) and (a_1, a_3, a_2) are indistinguishable, their only difference being the basis coordination of their respective state spaces.

In a system identification problem, the unknown quantities to be determined may include the system parameters $\underline{\theta}$, the initial state $\underline{x}_0$, the covariance matrices of the plant noise process and output noise process, or any combination of these quantities. While it is often desired to determine $\underline{x}_0$ from the system input-output data, a sizable body of results has been produced for systems operating in steady state in which the initial states are therefore unimportant and may be assumed to have been zero, or for systems actually having a zero, or otherwise known, initial state. Such results lead to the concept of zero-state parameter identifiability.

Consideration must also be given to the system input and its effect upon parameter identifiability. The input may be absent, present but unspecified, or present and designed to enhance the parameter identifiability properties of the system under study. Two commonly specified inputs are Gaussian white noise and the unit impulse, the latter being particularly attractive for linear, SISO systems as the impulse response defines the internal characteristics of such systems. While the characteristics of inputs are usually specified to enhance the convergence properties of parameter estimation schemes as opposed to the actual parameter identifiability properties of the system, the presence or absence of some input is often critical to system parameter identifiability properties as seen in the following example.

Example 2.4. Consider the unity feedback system of Figure 4.

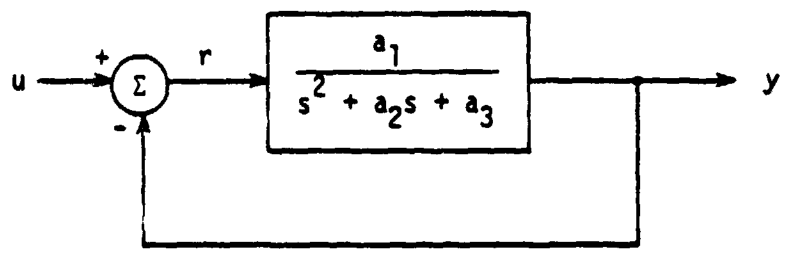

Figure 4. A unity feedback system.

The closed loop transfer function is readily found to be

$$H(s) = \frac{a_1}{s^2 + a_2 s + (a_1 + a_3)} = \frac{a_1}{s^2 + a_2 s + a_4} . \tag{28}$$

It will be explicitly demonstrated in Example 3.5 that the parameterization of a system characterized by the given transfer function is identifiable and that the parameter values a_1, a_2, and a_4 can be determined. Since $a_3 = a_4 - a_1$, the parameter values a_1, a_2, and a_3 are also identifiable.

However, if the input is set identically to zero, $u \equiv 0$, the following defining differential equation results

$$\ddot{y} + a_2\dot{y} + (a_1 + a_3)y = 0. \tag{29}$$

Obviously, any parameter pair (a_1, a_3) such that

$$a_1 + a_3 = \text{constant} = k_1 \tag{30}$$

yields an equivalent equation, and (a_1, a_3) is not uniquely identifiable in the absence of a system input.

While parameter identifiability is considered to be a system property and is therefore independent of the actual parameter values or the parameter estimation scheme used to recover these values, there exist derivative definitions based on the assumption of the existence of a scheme to generate a sequence of parameter estimates, which converges to the true parameter value. As these algorithms are usually of the small variations type, it is sufficient to consider parameter identifiability in terms of the uniqueness of the input/output relationship generated by the true set of parameters as compared to all other parameters within some neighborhood of the true parameters. If this uniqueness property holds for all parameter space, the system is considered to possess global parameter identifiability. However, if this uniqueness property holds only within some limited neighborhood of the true parameter value, the system is said to possess local parameter identifiability within the defined neighborhood. Consequently, consideration must be given to determining this neighborhood, or region of parameter identifiability, in which a given parameter value is unique. Initiation of a parameter estimation sequence within the region of parameter identifiability will ensure that the estimation sequence will converge to the true parameter values.

Example 2.5. The system of Example 2.3 was found to be not identifiable as the two parameterizations characterized by the ordered triples (a_1, a_2, a_3) and (a_1, a_3, a_2) were indistinguishable. If for a specific input-output data record the following parameter values hold

$$(a_1, a_2, a_3) = (k_1, k_2, k_3) \tag{31}$$

and

$$(a_1, a_2, a_3) = (k_1, k_3, k_2), \tag{32}$$

then the partial (two-dimensional) parameter space, Ω^*, of Figure 5 can be drawn.

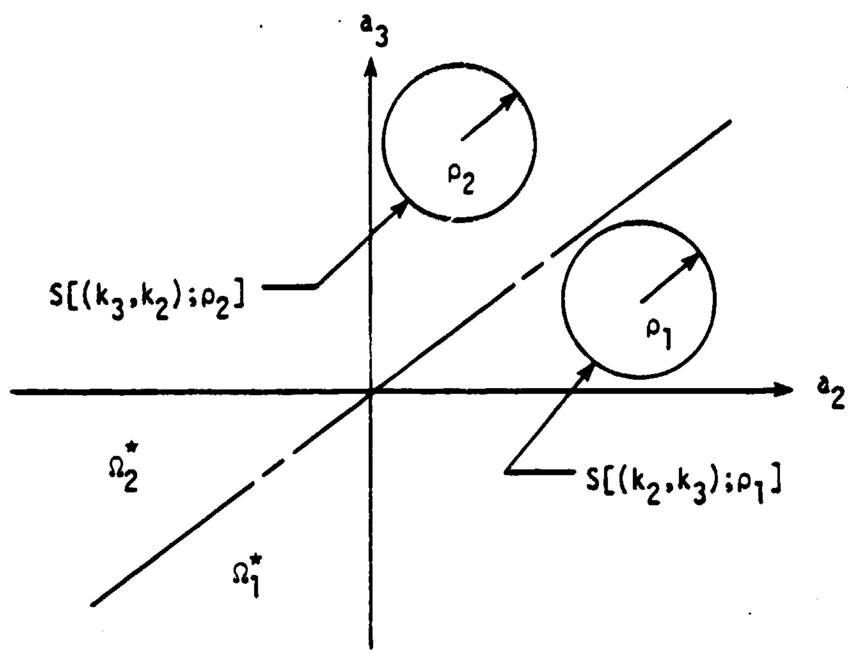

Figure 5. Partial parameter space, Ω^*.

If the partial parameter space is partitioned into Ω_1^* and Ω_2^* as shown, the system is locally identifiable in each of the two resulting regions. Determining the respective regions of parameter identifiability, designated in Figure 5 as $S[(k_2, k_3);\rho_1]$ and $S[(k_3, k_2);\rho_2]$ for a particular parameter estimation scheme, is a companion problem to the parameter identifiability problem (see [13]).

Consideration of the uniqueness properties of a given parameterization within a neighborhood of the true parameter value leads to the concept of parameter distinguishability, the property within the given local neighborhood by which the true parameter values can be isolated or distinquished from all other possible values in that neighborhood. This distinguishability property may be based upon different characteristics of the system under consideration, such as the ability of the true parameter value to generate a unique transfer function, a unique output, or a unique cost function for some parameter estimation scheme. This distinguishability property can then be related directly to the parameter identifiability property of the system.

Perhaps most significant in the development of parameter identifiability properties of systems is the dichotomy of definitions and methodology required by the consideration of deterministic or noise free systems versus stochastic or noisy systems. For this reason, each is considered as a major category with all other subcategories, as discussed above, developed under each.

It is generally desirable to develop parameter identifiability properties for an entire class of systems rather than approach each system separately. However, due to the large number of classes and options described by Figure 3 and Eq. (12), a compact, comprehensive treatment of the parameter identifiability question has not been accomplished to date. Instead, parameter identifiability definitions and developments have treated scattered, specific classes of systems, each seemingly with its own set of definitions and properties. It is the intent of this paper to consolidate these efforts into a unified and comprehensive overview. Therefore, basic definitions and approaches to the general spectrum of parameter identifiability problems are considered with the specific results for given classes of systems presented as examples of the basic philosophies. A chart is presented in Appendix A which relates the characteristics of the covered system classes to the applicable theorems.

3. DETERMINISTIC PARAMETER IDENTIFIABILITY

3.1 Concepts and Definitions

One of the two basic categories of systems in the study of parameter identifiability is that in which the noise processes $\underline{w}$ and $\underline{v}$ of Eq. (2) are absent. The general deterministic system is then defined as shown in Figure 6 as

$$\underline{y} = \underline{f}(\underline{x}_0, \underline{\theta}, \underline{u}) \tag{33}$$

where the definitions of the variables, arguments, and corresponding spaces remain the same as given in Section 2.

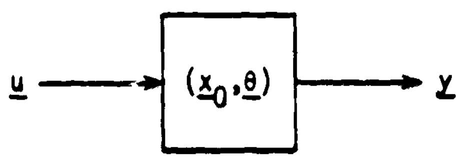

Figure 6. A general deterministic system.

For the given general deterministic system, the following general definitions are proposed:

Definition 3.1. A system is said to be Deterministically Identifiable if, for some $\underline{u} \in \mathcal{U}$, the mapping $\underline{f}(\cdot,\cdot,\underline{u}): X \times \Theta \to \mathcal{Y}$ is injective.

Definition 3.2. A system is said to be Locally Deterministically Identifiable at $(\underline{x}_0, \underline{\theta}_0)$ if there exists an open sphere $S(\underline{x}_0, \underline{\theta}_0; \epsilon)$ of radius $\rho > 0$ centered at $(\underline{x}_0, \underline{\theta}_0)$ such that, for some $\underline{u} \in \mathcal{U}$, the restriction of the mapping $\underline{f}(\cdot, \cdot, \underline{u}): X \times \mathcal{U} \rightarrow \mathcal{Y}$ to $S(\underline{x}_0, \underline{\theta}; \rho)$ is injective.

Definition 3.3. A system is said to be Zero State Deterministically Identifiable if, for some $\underline{u} \in \mathcal{U}$, the mapping $\underline{f}(0, \cdot, \underline{u}): X \times \mathcal{U} \rightarrow \mathcal{Y}$ is injective.

Definition 3.4. A system is said to be Locally Zero State Deterministically Identifiable at $(0, \underline{\theta}_0)$ if there exists an open sphere $S(0, \underline{\theta}_0; \rho)$ of radius $\rho > 0$ centered at $(0, \underline{\theta}_0)$ such that, for some $\underline{u} \in \mathcal{U}$, the restriction of the mapping $\underline{f}(0, \cdot, \underline{u}): X \times \mathcal{U} \rightarrow \mathcal{Y}$ to $S(0, \underline{\theta}_0; \rho)$ is injective.

It should be recalled that $\underline{f}$, a function from X into Y, is said to be injective if for every $\underline{x}_1, \underline{x}_2 \in X$, then $\underline{f}(\underline{x}_1) = \underline{f}(x_2)$ implies that $\underline{x}_1 = \underline{x}_2$. Equivalently, in terms of an arbitrary norm $\|\cdot\|$ defined on X and Y, $\underline{f}$ is said to be injective if for every $\underline{x}_1, \underline{x}_2 \in X$, then $\|\underline{f}(\underline{x}_1) - \underline{f}(\underline{x}_2)\| = 0$ implies that $\|\underline{x}_1 - \underline{x}_2\| = 0$ or $\underline{x}_1 = \underline{x}_2$.

It will be shown in the following that Definitions 3.1 through 3.4 may be considered as the basic, encompassing definitions for deterministic parameter identifiability and that other definitions previously proposed by other authors may be derived from, and hence are equivalent to, these basic definitions.

3.2 Deterministic Parameter Identifiability from Direct Parameter Recoverability

The injectivity of $\underline{f}$, as required by the basic definitions of deterministic parameter identifiability given in Section 3.1, directly implies the invertability of $\underline{f}$ and hence the direct recoverability of values for $\underline{\theta}_0$ and $\underline{x}_0$. Using such an approach Staley and Yue [20] have developed the deterministic identifiability properties for a class of linear, constant-coefficient, stable, single-input single-output, discrete-time systems as described by the scalar difference equation

$$x_j = \sum_{i=1}^{n} a_i x_{j-i} + \sum_{i=1}^{r} b_i u_{j-i}, \qquad j = 1, 2, \ldots, L \tag{34}$$

where the parameters a_i and b_i are unknown constants with $|a_i| < \infty$, $1 \le i \le n$, and $|b_i| < \infty$, $1 \le i \le r$. Although this system has a very particular structure, it represents a rather large class of realistic problems. In particular, it may also represent the discretized version of linear, constant-coefficient, stable, SISO, continuous-time systems.

The system of Eq. (34) may be stated in an equivalent input-state-output representation as

$$\underline{z}(j+1) = \underline{A}\,\underline{z}(j) + \underline{b}\,u_j, \qquad \underline{z}(0) = \underline{z}_0 \tag{35}$$

$$x_j = \underline{h}^T \underline{z}(j) \tag{36}$$

where

$$A = \begin{bmatrix} a_1 & 1 & 0 & 0 & \dots & 0 \\ a_2 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ a_{n-1} & 0 & 0 & 0 & \dots & 1 \end{bmatrix}, \; (n \times n);$$

$$\underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \end{bmatrix}, \; (r \times 1); \quad \underline{h} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; (n \times 1).$$

Staley and Yue [20] have proposed the following definition of identifiability for the given system which is readily seen to be a direct consequence of, and equivalent to, Definition 3.1.

<u>Definition 3.5 [20]</u>. The system parameter $\underline{\theta} = [a_1, \dots, a_n, b_1, \dots, b_r]^T$ and the initial state $\underline{z}_0$ are said to be <u>Identifiable in the Deterministic Sense</u>, <u>IDS</u>, if $\underline{\theta}$ and $\underline{z}_0$ are uniquely determined from the observed input and output sequences $\{u_j\}$ and $\{x_j\}$, $0 \le j \le L-1$, for some finite integer L.

The unknown parameter vector $\underline{\theta}$ and the initial state $\underline{z}_0$ may be expressed in terms of the input and output sequences using Toeplitz matrices as

$$\underline{x}_L = \underline{H}_L \underline{\theta} + \underline{E}_L \underline{z}_0 \qquad (37)$$

where

$$\underline{u}_L = [u_0, u_1, \ldots, u_{L-1}]^T \;, \quad (L \times 1);$$

$$\underline{x}_L = [x_0, x_1, \ldots, x_{L-1}]^T \;, \quad (L \times 1);$$

$$\underline{E}_L = \begin{bmatrix} \underline{I}_{nxn} \\ ---- \\ 0_{L-n,n} \end{bmatrix} , \; (L \times n);$$

and

$$\underline{H}_L = [\underline{S}_L\underline{x}_L, \underline{S}_L^2\underline{x}_L, \ldots, \underline{S}_L^n\underline{x}_L, \underline{S}_L\underline{u}_L, \ldots, \underline{S}_L^r\underline{u}_L], \quad L \times (n + r).$$

$\underline{S}_L$ is the LxL shift matrix defined such that $\underline{S}_L(i,j) = \delta_{i,j+1}$.

If $\underline{z}_0$ is known, such as when the system has an initial zero state, then Eq. (37) and Definition 3.5 together imply the following properties for Identifiability in the Deterministic Sense (IDS):

Property 1. $\underline{\theta}$ is identifiable, IDS, if and only if the matrix $\underline{H}_L^T\underline{H}_L$ is positive definite (or equivalently, $\underline{H}_L$ has rank n + r) such that $\underline{\theta} = (\underline{H}_L^T\underline{H}_L)^{-1}\underline{H}_L^T\underline{x}_L$ is uniquely determined. Note: $\underline{z}_0 = 0$.

Property 2. $\underline{\theta}$ is identifiable, IDS, only if $L \geq n + r$.

Property 3. $\underline{\theta}$ is identifiable, IDS, for all $L > L_0$ if it is identifiable for $L = L_0$.

By applying these three properties, the following theorems may be derived describing the identifiability properties of the class of systems described by Eqs. (34) or (35) and (36).

Theorem 3.1 [20]

In the autonomous version of the system described by Eq. (34) (i.e., $b_i \equiv 0$ or $u_i \equiv 0$, all i), $\underline{\theta} = [a_1, \ldots, a_n]^T$ is identifiable, IDS, if and only if $(\underline{A}, \underline{z}_0)$ is a completely controllable pair. Further, $\underline{\theta}$ is identifiable, IDS, if and only if it is identifiable in 2n steps; i.e., L = 2n.

In light of the universal Definition 3.3 for zero state deterministic parameter identifiability, it is proposed that the above theorem be modified to assert that $\underline{\theta}$ is simply "Zero State Deterministically Parameter Identifiable" as opposed to "identifiable, IDS".

It should be noted, in particular, that the initial state is assumed known, that the parameter vector $\underline{\theta}$ is limited to values of a_i, $1 \leq i \leq n$, and that the input $u_i \equiv 0$, all i or $b_i \equiv 0$, all i. It may be recalled that the pair $(\underline{A}, \underline{z}_0)$ is completely controllable if and only if the controllability matrix has rank n,

$$\operatorname{rank}[\underline{z}_0, \underline{A}\,\underline{z}_0, \underline{A}^2\underline{z}_0, \ldots, \underline{A}^{L-n-1}\underline{z}_0] = n. \tag{38}$$

Example 3.1 [20]. While it is easily shown that the initial state $\underline{z}_0 = [0, 0, \ldots, 1]^T$ always yields an identifiable system, consider the second-order system described by

$$x_j - (c + d)x_{j-1} + cdx_{j-2} = 0 \tag{39}$$

with an initial state $\underline{z}_0 = [1 - d]^T$. For this case,

$$\underline{A} = \begin{bmatrix} c+d & 1 \\ -cd & 0 \end{bmatrix}$$

and $(\underline{A}, \underline{z}_0)$ is not a controllable pair. The vector $\underline{z}_0$ is an eigenvector of $\underline{A}$ corresponding to the eigenvalue c. The state transition is confined to a proper subspace of R^n which is invariant under $\underline{A}$, and no information about d, the other eigenvalue of $\underline{A}$, is contained in the output.

For a non-autonomous, zero-state system the following theorem applies. A similar modification of the terminology "identifiable, IDS" is proposed to achieve conformity with Definition 3.3.

Theorem 3.2 [20]

If the system of Eq. (34) is stable and $\underline{z}_0 = 0$, then the parameter vector $\underline{\theta} = [a_1, \ldots, a_n, b_1 \ldots, b_r]^T$ is identifiable, IDS, from the input and output sequences $\{u_j\}, \{x_j\}$, $0 \leq j \leq L - 1$, if and only if

(0) $L \geq n + r$,

(i) b_i are not all zero,

(ii) the polynomials $A(z) = 1 - \sum_{j=1}^{n} a_j z^j$ and $B(z) = \sum_{j=1}^{m} b_j z^j$ do not have a common factor, and

(iii) u_j is not identically zero for $0 \leq j \leq L - n - r$.

When $\underline{z}_0$ is unknown, Theorem 3.2 can be generalized by the following modifications:

$$\underline{H}_L^T \underline{H}_L \rightarrow \begin{bmatrix} \underline{H}_L^T \\ \hdashline \underline{E}_L^T \end{bmatrix} [\underline{H}_L \vdots \underline{E}_L], \quad (2n + r) \times (2n + r)$$

$$\underline{\theta}^T \rightarrow [\underline{\theta}^T, \underline{z}_0^T].$$

The generalized result becomes:

Theorem 3.3 [20]

If the system of Eq. (34) is stable, then the parameter vector $\underline{\theta}$ and the initial state $\underline{z}_0$ are identifiable, IDS, from the input and output sequences $\{u_j\}$, $\{x_j\}$, $0 \le j \le L - 1$, if and only if

(0) $L \ge 2n + r$,

(i) b_i are not all zero,

(ii) $A(z)$ and $B(z)$, as defined in Theorem 3.2, do not have a common factor, and

(iii) the $(2n + r) \times (2n + r)$ matrix

$$\begin{bmatrix} \underline{U}_L^T \\ \hdashline \underline{E}_L^T \end{bmatrix} [\underline{U}_L \vdots \underline{E}_L]$$

is positive definite where $\underline{E}_L$ remains as defined previously and

$$\underline{U}_L = [\underline{S}_L \underline{u}_L,\ \underline{S}_L^2 \underline{u}_L,\ \ldots,\ \underline{S}_L^{n+r}\ \underline{u}_L], \quad L \times (n + r).$$

A change of terminology may also be made in Theorem 3.3 to conform with Definition 3.1.

It may be noted that if $b_i \equiv 0$, all i, Theorem 3.1 applies. The proof of Theorem 3.1 is presented in Appendix B. Theorems 3.2 and 3.3 are given without proof (see [20]).

3.3 Deterministic Parameter Identifiability from the Transfer Function

The injectivity of $\underline{f}$ required by the basic definitions of deterministic parameter identifiability may also be interpreted as requiring a unique relationship between the system input and output; that is, an injective mapping from $\mathcal{U}$ into $\mathcal{Y}$. Such a mapping is determined for two large classes of deterministic, linear, constant coefficient systems by their transfer functions. These two classes of systems may be defined for the continuous-time and discrete-time cases respectively by

$$\dot{\underline{x}}(t,\underline{\theta}) = \underline{A}(\underline{\theta})\ \underline{x}\ (t,\underline{\theta}) + \underline{B}(\underline{\theta})\ \underline{u}\ (t) \qquad (40)$$

$$\underline{y}(t,\theta) = \underline{C}(\underline{\theta})\ \underline{x}\ (t,\underline{\theta}) + \underline{D}(\underline{\theta})\ \underline{u}\ (t) \qquad (41)$$

and

$$\underline{x}(k+1,\underline{\theta}) = \underline{A}(\underline{\theta})\ \underline{x}\ (k,\underline{\theta}) + \underline{B}(\underline{\theta})\ \underline{u}\ (k) \qquad (42)$$

$$\underline{y}(k,\underline{\theta}) = \underline{C}(\underline{\theta})\ \underline{x}\ (k,\underline{\theta}) + \underline{D}(\underline{\theta})\ \underline{u}\ (k) \qquad (43)$$

where $\underline{x}(\cdot,\underline{\theta})\epsilon R^n$, $\underline{y}(\cdot,\underline{\theta})\epsilon R^m$, $\underline{u}(\cdot)\epsilon\mathcal{U}$ and $\underline{u}(\cdot)\epsilon R^r$, $\underline{\theta}\epsilon\Omega\subset R^p$, and $\underline{A}(\underline{\theta})$, $\underline{B}(\underline{\theta})$, $\underline{C}(\underline{\theta})$, and $\underline{D}(\underline{\theta})$ are appropriately dimensioned constant matrices parameterized by $\underline{\theta}$. For their respective transfer functions, $\underline{H}(s,\underline{\theta})$ and $\underline{H}(z,\underline{\theta})$, the systems must be operating in the steady state mode in order to relate the system input $\underline{u}(\cdot)$ to the system output $\underline{y}(\cdot,\underline{\theta})$. Hence, no information concerning the initial state $\underline{x}_0$ is available from the input-output data records. The systems may thus be considered to have a known initial state which may be taken without loss of generality as $\underline{x}_0 = 0$. The methodology employed and the results achieved are identical for both systems, except that the Z Transform is employed for discrete-time systems and the Laplace Transform is employed for continuous-time systems. Therefore, only the continuous-time system of Eqs. (40) and (41) will be considered in the following.

The given system may now be investigated for properties of Zero State Deterministic Parameter Identifiability, as required by Definition 3.3 and 3.4. For a given $\underline{u}\in U$, the system output may be written in the frequency domain as

$$\underline{Y}(s,\underline{\theta}) = \left\{\underline{C}(\underline{\theta})[s\underline{I} - \underline{A}(\underline{\theta})]^{-1}\, \underline{B}(\underline{\theta}) + \underline{D}(\underline{\theta})\right\}\, \underline{U}(s,\underline{\theta})$$

$$= \underline{H}(s,\underline{\theta})\ \underline{U}\ (s,\underline{\theta}) \tag{44}$$

It is seen that $\underline{H}(s,\underline{\theta})$ must be (locally) injective (equivalently, unique) as a function of $\underline{\theta}$. Equivalently, from the definition of injectivity, for every $\underline{\theta}_1$, $\underline{\theta}_2\in\Omega$, then $\underline{H}(s,\underline{\theta}_1) = \underline{H}(s,\underline{\theta}_2)$ implies that $\underline{\theta}_1 = \underline{\theta}_2$. That $\underline{H}(s,\underline{\theta})$ must be (locally) injective is evident from the uniqueness of the Laplace Transform/Inverse Laplace Transform pair and from the fact that

$$\mathcal{L}^{-1}\{\underline{H}(s,\underline{\theta})\} = \underline{f}[0,\ \theta,\ \underline{\delta}(t)] \tag{45}$$

where $\mathcal{L}^{-1}$ is the Inverse Laplace Transform operator and $\underline{\delta}(t)$ is the vector impulse function which generates the impulse response matrix. It should be noted that the vector impulse function $\underline{\delta}(t)$ implies that an impulsive input is applied to each of the m input ports in sequence and that the resulting outputs each form a column of the impulse response matrix. Thus, these two particular classes of systems may be considered to require an impulsive input for the determination of parameter identifiability properties and are so designated in the chart of system characteristics in Appendix A.

As implied in the sentence following Eq. (44), the simplest level of analysis may involve the direct inspection of $\underline{H}(s,\underline{\theta})$ to determine its uniqueness as a function of $\underline{\theta}$. Such an analysis was conducted in Examples 2.2 and 2.3.

Glover and Willems [3], using the general concepts above, have produced for the two classes of systems under consideration a set of definitions and results. However, it will be readily seen that their definitions and results are a direct consequence of, and thus equivalent to, the basic definitions given in Section 3.1. Although the actual definitions and results of Glover and Willems [39] are reproduced below for comparison, the obvious changes to produce conformity with the definitions of Section 3.1 are recommended.

Definition 3.6a [9]. The linear, dynamic system characterized by Eqs. (40) and (41) is said to be locally identifiable from the transfer function at the point $\underline{\theta}_0 \epsilon \underline{\Omega}$ if there exists a $\rho > 0$ such that

$$\text{(1)} \quad \| \underline{\theta}_1 - \underline{\theta}_0 \| < \rho, \quad \| \underline{\theta}_2 - \underline{\theta}_0 \| < \rho; \ \underline{\theta}_1, \underline{\theta}_2 \epsilon \Omega \tag{46}$$

and

$$\text{(2)} \quad \underline{C}(\underline{\theta}_1)\ [\underline{I}s - \underline{A}(\underline{\theta}_1)]^{-1}\ \underline{B}(\underline{\theta}_1) + \underline{D}(\underline{\theta}_1) =$$

$$\underline{C}(\underline{\theta}_2)\ [\underline{I}s - \underline{A}(\underline{\theta}_2)]^{-1}\ \underline{B}(\underline{\theta}_2) + \underline{D}(\underline{\theta}_2) \tag{47}$$

together imply $\underline{\theta}_1 = \underline{\theta}_2$, for all $s \epsilon C$ and $s \neq \left\{\lambda[\underline{A}(\underline{\theta}_2)],\ \lambda[\underline{A}(\underline{\theta}_1)]\right\}$ where $\lambda(\cdot)$ denotes the eigenvalues of the respective matrix, C is the field of complex numbers, and $\|\cdot\|$ is an arbitrary norm. In consonance with the concept above, Definition 3.6a equivalently states that, in a

ρ-neighborhood of the true parameter $\underline{\theta}_0$, there are no two systems with distinct parameters which have the same transfer function.

As the system matrices $(\underline{A}, \underline{B}, \underline{C}, \underline{D})(\underline{\theta})$ are normally continuously differentiable with respect to $\underline{\theta}$, the transfer function is meromorphic at $\underline{\theta}_0$ and Eq. (47) may be expanded in a power series to yield an equivalent definition.

Definition 3.6b [9]. The linear, dynamical system characterized by Eqs. (40) and (41) is said to be locally identifiable from the transfer function at the point $\underline{\theta}_0$ if there is an open sphere $S(0,\underline{\theta}_0;\rho) \subset \Omega$ with radius $\rho > 0$ and centered at $(0, \underline{\theta})$ such that

$$(1) \quad \underline{\theta} \in S(0,\underline{\theta}_0;\rho) \tag{48}$$

$$(2) \quad \underline{D}(\underline{\theta}) = \underline{D}(\underline{\theta}_0) \tag{49}$$

$$(3) \quad \underline{C}(\underline{\theta})\underline{A}^i(\underline{\theta})\underline{B}(\underline{\theta}) = \underline{C}(\underline{\theta}_0)\underline{A}^i(\underline{\theta}_0)\underline{B}(\underline{\theta}_0), \qquad i = 1, 2, \ldots \tag{50}$$

together imply $\underline{\theta} = \underline{\theta}_0$.

Definition 3.7. The Markov parameters for the system described by Eqs. (40) and (41) are defined in terms of the constant system matrices $\underline{A}(\underline{\theta})$, $\underline{B}(\underline{\theta})$, and $\underline{C}(\underline{\theta})$ as

$$Z_\ell = \underline{C}(\underline{\theta})\underline{A}^\ell(\underline{\theta})\underline{B}(\underline{\theta}), \qquad \ell = 0, 1, 2, \ldots \tag{51}$$

The Markov parameter matrix for the given system is defined as

$$\underline{G}(\underline{\theta}) = \begin{bmatrix} \underline{D}^T(\underline{\theta}) \\ [\underline{C}(\underline{\theta})\underline{B}(\underline{\theta})]^T \\ [\underline{C}(\underline{\theta})\underline{A}(\underline{\theta})\underline{B}(\underline{\theta})]^T \\ \vdots \\ [\underline{C}(\underline{\theta})\underline{A}^{2n-1}(\underline{\theta})\underline{B}(\underline{\theta})]^T \end{bmatrix} . \tag{52}$$

It should be noted that Definition 3.6b is equivalent to the requirement that the mapping from the parameter space Ω, or some subset thereof, into the Markov parameters also be injective. As a direct consequence of the constant rank theorem [17], it can be shown that for an open sphere $S(0,\underline{\theta}_0;\rho)$, centered at $(0,\underline{\theta}_0)$ with radius $\rho > 0$, contained in Ω and thus also an open subset of R^{ν}, then the mapping from $S(0,\underline{\theta}_0;\rho)\subset\Omega$ into the Markov parameters is locally injective if the rank of the Jacobian of the Markov parameter matrix equals p, the dimensionality of the unknown parameters,

$$\text{rank}\left(\frac{\partial\underline{G}(\underline{\theta})}{\partial\underline{\theta}}\right) = p. \tag{53}$$

As a direct consequence the following theorem is stated without proof

<u>Theorem 3.4 [9]</u>

The linear system characterized by Eqs. (40) and (41) is <u>locally identifiable from the transfer function at $\underline{\theta}_0\epsilon\Omega$</u> if the Jacobian of the Markov parameter matrix $\underline{G}(\underline{\theta})$ has constant rank p in an open sphere $S(0,\underline{\theta}_0;\rho)$ of radius $\rho > 0$ centered at $(0,\underline{\theta}_0)$; i.e., rank $[\partial\underline{G}(\underline{\theta})/\partial\underline{\theta}] = p$.

Employment of this theorem yields a relatively simple test for local parameter identifiability for the given classes of linear systems. It is constructive to consider two such examples.

Example 3.2. Consider the single-input single-output, continuous, constant, second-order system characterizing the two compartment model of Figure 2 and Example 2.2

$$\dot{x}_1 = -(a_1 + a_2)x_1 + a_3x_2 + u \tag{54}$$

$$\dot{x}_2 = a_2x_1 - (a_3 + a_4)x_2 \tag{55}$$

$$y = x_1. \tag{56}$$

It is required to determine if the four unknown rate constants a_1, a_2, a_3, and a_4 are zero-state, deterministically identifiable. The four constant, system matrices defined in Eqs. (40) and (41) for the given systems are

$$\underline{A}(\underline{\theta}) = \begin{bmatrix} -(a_1 + a_2) & a_3 \\ a_2 & -(a_3 + a_4) \end{bmatrix}, \quad \underline{b}(\underline{\theta}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\underline{c}(\underline{\theta}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and } \underline{D}(\underline{\theta}) = [0] \tag{57}$$

where the unknown parameter vector $\underline{\theta}$ is defined as $\underline{\theta} = [a_1, a_2, a_3, a_4]^T$. The Markov parameter matrix for the second-order system ($n = 2$) is found to be

$$\underline{G}(\underline{\theta}) = \begin{bmatrix} \underline{D}(\underline{\theta}) \\ \underline{c}(\underline{\theta})\underline{b}(\underline{\theta}) \\ \underline{c}(\underline{\theta})\underline{A}(\underline{\theta})\underline{b}(\underline{\theta}) \\ \underline{c}(\underline{\theta})\underline{A}^2(\underline{\theta})\underline{b}(\underline{\theta}) \\ \underline{c}(\underline{\theta})\underline{A}^3(\underline{\theta})\underline{b}(\underline{\theta}) \end{bmatrix} = \begin{bmatrix} \underline{G}_1(\underline{\theta}) \\ \underline{G}_2(\underline{\theta}) \\ \underline{G}_3(\underline{\theta}) \\ \underline{G}_4(\underline{\theta}) \\ \underline{G}_5(\underline{\theta}) \end{bmatrix} \quad \text{(continued)} \tag{58}$$

$$
= \begin{bmatrix} 0 \\ 1 \\ -(a_1 + a_2) \\ a_1^2 + 2a_1a_2 + a_2^2 + a_2a_3 \\ -(a_1^3 + 3a_1^2a_2 + 3a_1a_2^2 + a_2^3 + 2a_1a_2a_3 + 2a_2^2a_3 \\ \quad + a_2a_3^2 + a_2a_3a_4) \end{bmatrix} \tag{58}
$$

In order to check the local injectivity of the mapping of the parameter space into the Markov parameters and apply the results of Theorem 3.4, the Jacobian of the Markov parameter matrix is calculated as

$$
\frac{\partial \underline{G}(\underline{\theta})}{\partial \underline{\theta}} = \begin{bmatrix} \partial\underline{G}_1/\partial a_1 & \partial\underline{G}_1/\partial a_2 & \partial\underline{G}_1/\partial a_3 & \partial\underline{G}_1/\partial a_4 \\ \partial\underline{G}_2/\partial a_1 & \partial\underline{G}_2/\partial a_2 & \partial\underline{G}_2/\partial a_3 & \partial\underline{G}_2/\partial a_4 \\ \partial\underline{G}_3/\partial a_1 & \partial\underline{G}_3/\partial a_2 & \partial\underline{G}_3/\partial a_3 & \partial\underline{G}_3/\partial a_4 \\ \partial\underline{G}_4/\partial a_1 & \partial\underline{G}_4/\partial a_2 & \partial\underline{G}_4/\partial a_3 & \partial\underline{G}_4/\partial a_4 \\ \partial\underline{G}_5/\partial a_1 & \partial\underline{G}_5/\partial a_2 & \partial\underline{G}_5/\partial a_3 & \partial\underline{G}_5/\partial a_4 \end{bmatrix}
$$

$$
= \left[\begin{array}{l|} 0 \\ 0 \\ -1 \\ 2a_1 + 2a_2 \\ -(3a_1^2 + 6a_2 + 3a_2^2 + 2a_2a_3) \end{array}\right.
$$

(continued) (59)

$$\left|\begin{matrix} 0 \\ 0 \\ -1 \\ 2a_1 + 2a_2 + a_3 \\ -(3a_1^2 + 6a_1a_2 + 3a_2^2 + 2a_1a_3 + 4a_2a_3 + a_3^2 + a_3a_4) \end{matrix}\right|$$

$$\left|\begin{matrix} 0 \\ 0 \\ 0 \\ a_2 \\ -(2a_1a_2 + 2a_2^2 + 2a_2a_3 + a_2a_4) \end{matrix}\right|\left.\begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ -a_2a_3 \end{matrix}\right] \tag{59}$$

Clearly the matrix $\frac{\partial G(\underline{\theta})}{\partial \underline{\theta}}$ is of rank 3, at most, while it must be of rank 4 for the mapping of the parameter space, $\Omega \subset R^4$, into the Markov parameters to be locally injective. By the application of Theorem 3.4, the system parameters are not identifiable. This prediction of non-identifiability of the system rate coefficients was confirmed by a direct analysis of the system transfer function in Example 2.2.

<u>Example 3.3</u>. Modify the system of Figure 2 and Examples 2.2 and 3.2 such that $a_4 = 0$, yielding the following linear, constant, second-order system

$$\dot{x}_1 = -(a_1 + a_2)x_1 + a_3x_2 + u \tag{60}$$

$$\dot{x}_2 = a_2x_1 - a_3x_2 \tag{61}$$

$$y = x_1. \tag{62}$$

The system matrices become

$$\underline{A}(\underline{\theta}) = \begin{bmatrix} -(a_1 + a_2) & a_3 \\ & \\ a_2 & -a_3 \end{bmatrix}, \quad \underline{b}(\underline{\theta}) = \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix}$$

$$\underline{c}(\underline{\theta}) = \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix}, \qquad \underline{D}(\underline{\theta}) = [0] \tag{63}$$

where the unknown parameter vector $\underline{\theta}$ to be identified now becomes $\underline{\theta} = [a_1, a_2, a_3]^T$. The Markov parameter matrix is calculated as

$$\underline{G}(\underline{\theta}) = \begin{bmatrix} 0 \\ 1 \\ -(a_1 + a_2) \\ a_1^2 + 2a_1a_2 + a_2a_3 \\ -\left(a_1^3 + 3a_1^2a_2 + 3a_1a_2^2 + a_2^3 + 2a_1a_2a_3 + 2a_2^2a_3 + a_2a_3^2\right) \end{bmatrix}. \tag{64}$$

The Jacobian of the Markov parameter matrix can then be found to be

$$\frac{\partial \underline{G}(\underline{\theta})}{\partial \underline{\theta}} = \left[\begin{array}{l|} 0 \\ 0 \\ -1 \\ \left(2a_1 + 2a_2\right) \\ -\left(3a_1^3 + 6a_1a_2 + 3a_2^2 + 2a_2a_3\right) \end{array}\right.$$

(continued) (65)

$$\left| \begin{matrix} 0 \\ 0 \\ -1 \\ \left(2a_1 + 2a_2 + a_3\right) \\ -\left(3a_1^2 + 6a_1a_2 + 3a_2^2 + 2a_1a_3 + 4a_2a_3 + a_3^2\right) \end{matrix} \right|$$

$$\left| \begin{matrix} 0 \\ 0 \\ 0 \\ a_2 \\ \left(2a_1a_2 + 2a_2^2 + 2a_2a_3\right) \end{matrix} \right] \tag{65}$$

which is clearly of rank 3 for all $a_2 \neq 0$. Thus, the mapping of the parameter space, $\Omega \subset R^3$, into the Markov parameters is injective for all $a_2 \neq 0$, and, by Theorem 3.4, the modified system is locally identifiable for all $\underline{\theta} \epsilon R^3$ such that $a_2 \neq 0$.

The prediction of the identifiability of the system parameters can be confirmed by direct analysis of the system transfer function which is found to be

$$H(s) = \frac{s + a_3}{s^2 + s(a_1 + a_2 + a_3) + a_1a_3} . \tag{66}$$

It is readily evident that, for a given triple of constants (c_1, c_2, c_3), an identical input/output relationship will result for any combination of parameters in which

$$a_3 = c_1 \tag{67}$$

$$a_1 + a_2 + a_3 = c_2 \tag{68}$$

$$a_1 a_3 = c_3 \, . \tag{69}$$

Since there are three equations in three unknowns, there exists a unique solution for $\underline{\theta} = [a_1, a_2, a_3]^T$; in particular,

$$a_1 = c_3/c_1 \, , \quad c_1 \neq 0 \tag{70}$$

$$a_2 = -c_1 + c_2 - c_3/c_1 \, , \quad c_1 \neq 0 \tag{71}$$

$$a_3 = c_3 \tag{72}$$

where the specific values of the triple (c_1, c_2, c_3) correspond to the specific input/output data record to be evaluated. Thus, as predicted by the application of Theorem 3.4, the parameters of the modified system are indeed identifiable.

In Example 2.3, the parameter identifiability properties of a linear system of the type characterized by Eqs. (40) and (41) were investigated by demonstrating the existence of a similarity transformation which would transform a parameterized system into an equivalent system with different parameter values but with the same parameterization. Glover and Willems [9] formalized the concept as follows.

For the vector of true parameters $\underline{\theta}_0$ and the space of invertible (nonsingular) n X n matrices $\underline{P} \epsilon GL(n)$, the solution $(\underline{P}, \underline{\theta})$ for the following similarity transformation equations must be investigated:

$$\underline{P} \, \underline{A} \, (\underline{\theta}) \, \underline{P}^{-1} = \underline{A} \, (\underline{\theta}_0)$$

$$\underline{P} \, \underline{B} \, (\underline{\theta}) = \underline{B} \, (\underline{\theta}_0) \qquad \text{(continued)} \qquad (73)$$

$$\underline{C}(\underline{\theta})\underline{P}^{-1} = \underline{C}(\underline{\theta}_0)$$

$$\underline{D}(\underline{\theta}) = \underline{D}(\underline{\theta}_0) \tag{73}$$

It follows that the given system is (locally) identifiable from the transfer function at $\underline{\theta}_0$ if there exist an open sphere $S(0,\underline{\theta}_0;\rho) \subset \Omega$ centered at $(0,\underline{\theta}_0)$ with radius $\rho > 0$ such that $(\underline{I}_{nxn},\underline{\theta}_0)$ is the unique solution of Eq. (73) in $GL(n) \times S(0,\underline{\theta}_0;\rho)$. Sufficient conditions for such a unique solution to exist are given in the following theorem.

<u>Theorem 3.5</u>

Let the linear system characterized by Eqs. (40) and (41) be a minimal realization and define

$$\underline{\mathcal{E}}(\underline{P},\underline{\theta}) \triangleq \begin{bmatrix} \underline{P}\ \underline{A}\ (\underline{\theta})\ \underline{P}^{-1} \\ \underline{P}\ \underline{B}\ (\underline{\theta}) \\ \underline{C}(\underline{\theta})\underline{P}^{-1} \\ \underline{D}(\underline{\theta}) \end{bmatrix} .$$

If there exists an open sphere $S(0,\underline{\theta}_0,\rho)$ centered at $(0,\underline{\theta}_0)$ with radius $\rho > 0$ such that $\nabla_{(\underline{P},\underline{\theta})}\underline{\mathcal{E}}(\underline{P},\underline{\theta})$ has constant rank $n^2 + p$ at $\underline{P} = \underline{I}_{nxn}$ for all $\underline{\theta} \epsilon S(0,\underline{\theta}_0;\rho)$, then the system is <u>locally identifiable from the transfer function at θ_0</u>.

The matrix $\nabla_{(\underline{P},\underline{\theta})}\underline{\mathcal{E}}(\underline{P},\underline{\theta})$ evaluated at the point $(\underline{I}_{nxn},\underline{\partial})$ is given by

$$\nabla_{(\underline{P},\underline{\theta})}\underline{e}(\underline{P},\underline{\theta})\Big|_{(\underline{I}_{nxn},\underline{\theta})} = \begin{bmatrix} \underline{I}_{nxn} \otimes \underline{A}'(\underline{\theta}) - \underline{A}(\underline{\theta}) \otimes \underline{I}_{nxn} & \Big| & \bar{\bar{\underline{A}}}'(\underline{\theta}) \\ \underline{I}_{nxn} \otimes \underline{B}'(\underline{\theta}) & \Big| & \bar{\bar{\underline{B}}}'(\underline{\theta}) \\ -\underline{C}(\theta) \otimes \underline{I}_{nxn} & \Big| & \bar{\bar{\underline{C}}}'(\underline{\theta}) \\ \underline{0}_{mrxn^2} & \Big| & \bar{\bar{\underline{D}}}'(\underline{\theta}) \end{bmatrix} \tag{74}$$

which is a $(n^2 + nm + rn + rm) \times (n^2 + p)$ matrix and where $\otimes$ represents the Kronecker product [19] and $\bar{\bar{\underline{X}}}$ represents the standard Kronecker product matrix ordering.

An application of Theorem 3.5 is given in the following example.

Example 3.4. Consider the linear, SISO, second-order system

$$\dot{\underline{x}} = \underline{A}(\underline{\theta})\underline{x} + \underline{b}(\underline{\theta})u \tag{75}$$

$$y = \underline{c}^T\underline{x} \tag{76}$$

where

$$\underline{A}(\underline{\theta}) = \begin{bmatrix} 0 & 1 \\ -a_3 & -a_2 \end{bmatrix}, \quad \underline{b}(\underline{\theta}) = \begin{bmatrix} 0 \\ a_1 \end{bmatrix}, \quad \underline{c} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$$\underline{D} = [0] \text{ and } \underline{\theta} = \left[a_1, a_2, a_3\right]^T.$$

Then

$$\nabla_{(\underline{P},\underline{\theta})}\underline{\ell}\,(\underline{P},\underline{\theta})\Big|_{(\underline{I}_{nxn},\underline{\theta})} =$$

$$\begin{bmatrix} \begin{bmatrix} 0 & -a_3 & 0 & 0 \\ 1 & -a_2 & 0 & 0 \\ 0 & 0 & 0 & -a_3 \\ 0 & 0 & 1 & -a_2 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -a_3 & 0 & -a_2 & 0 \\ 0 & -a_3 & 0 & -a_2 \end{bmatrix} & \begin{matrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{matrix} \\ \begin{matrix} 0 & a_1 & 0 & 0 \\ 0 & 0 & 0 & a_1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -a_3 & -1 & 0 & 0 & 0 & 0 \\ 1 & -a_2 & 0 & -1 & 1 & 0 & 0 \\ a_3 & 0 & a_2 & -a_3 & 0 & 0 & 0 \\ 0 & a_3 & 1 & 0 & 0 & 1 & 0 \\ 0 & a_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{77}$$

which is of rank $2^2 + 3 = 7$ for all $\underline{\theta} \epsilon R^3$. By Theorem 3.5, the system is identifiable from the transfer function (globally, since the rank of

$$\nabla_{(\underline{P},\underline{\theta})}\underline{\mathcal{E}}(\underline{P},\underline{\theta})\Big|_{(\underline{I}_{n\times n},\underline{\theta})} \quad \text{is 7 for all } \underline{\theta}\epsilon R^3).$$

It is constructive to consider a direct method of calculating a $\underline{P}\epsilon GL(n)$ satisfying Eq. (73).

Example 3.5. Consider the system of Example 3.4 and find a $\underline{P}\epsilon GL(n)$ satisfying the conditions of Eq. (73). Assume that such a matrix $\underline{P}$ exists, creating the equivalently parameterized system

$$\dot{\underline{x}}^* = \underline{A}^*(\underline{\theta})\ \underline{x}^* + \underline{b}^*(\underline{\theta})u \tag{78}$$

$$y^* = \underline{c}^{T*}\underline{x}^* \tag{79}$$

where

$$\underline{A}^*(\underline{\theta}) = \begin{bmatrix} 0 & 1 \\ \alpha_1 & \alpha_2 \end{bmatrix}, \quad \underline{b}^*(\underline{\theta}) = \begin{bmatrix} 0 \\ \alpha_3 \end{bmatrix} \quad \underline{c}^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{80}$$

and

$$\underline{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \tag{81}$$

From Eq. (73)

$$\underline{c}\ \underline{P}^{-1} = \underline{c}^* \rightarrow \underline{c} = \underline{c}^*\underline{P} \tag{82}$$

or

$$[1 \quad 0] = [1 \quad 0] \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \end{bmatrix} \tag{83}$$

which implies $p_{11} = 1$ and $p_{12} = 0$. (84)

Again from Eq. (73)

$$\underline{P}\,\underline{A}\,\underline{P}^{-1} = \underline{A}^{*} \rightarrow \underline{P}\,\underline{A} = \underline{A}^{*}\underline{P}$$

or

$$\begin{bmatrix} 1 & 0 \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -a_3 & -a_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_3 p_{22} & -a_2 p_{21} p_{22} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 \\ \alpha_1 & \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} p_{21} & p_{22} \\ \alpha_1 - \alpha_2 p_{21} & \alpha_2 p_{22} \end{bmatrix} \tag{85}$$

which implies $p_{21} = 0$ and $p_{22} = 1$. Therefore,

$$\underline{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{86}$$

the identity matrix, thus establishing the uniqueness of the parameterization of the system and confirming the results of Example 3.4.

It should be noted that the state space formulation of Eqs. (75) and (76) yield the transfer function

$$H(s) = \frac{a_1}{s^2 + a_2 s + a_3} \tag{87}$$

which appears essentially equivalent to that of Example 2.3. However, the system of Example 2.3 was found to be not identifiable, demonstrating the importance of a given parameterization to the parameter identifiability properties of a particular system.

3.4. Deterministic Parameter Identifiability from Output Distinguishability

In Section 3.3, the injectivity of $\underline{f}$ expressed in the basic definitions of deterministic parameter identifiability was interpreted as requiring an injective map from the system input function space U into the system output function space Y as delineated by a system transfer function in terms of the parameters $\underline{\theta}$. A very similar method is to analyze directly the system output properties in terms of their injectivity properties with respect to the parameters $\underline{\theta}$ as opposed to analyzing the system transfer function which generates the system output for several classes of systems. By analyzing the output directly rather than limiting the analysis to the output-generating transfer function, a large set of system classes may be considered.

The injectivity of $\underline{f}$ requires, by definition, for any given $\underline{u}\in U$ and $\underline{\theta}_1, \underline{\theta}_2\in\Omega$, that $\underline{y}(\cdot,\underline{\theta}_1) = \underline{y}(\cdot,\underline{\theta}_2)$ implies $\underline{\theta}_1 = \underline{\theta}_2$. This may be interpreted as requiring that two outputs of the given system, corresponding to two different parameters $\underline{\theta}_1$ and $\underline{\theta}_2$, $\underline{\theta}_1 \neq \theta_2$ but with a common input $\underline{u}\in U$, must be different or distinguishable from each other for all $\underline{\theta}_1$ and $\underline{\theta}_2$ in Ω, the parameter space for which the system is deterministically parameter identifiable.

Grewal et al. [10], [11] and [12] have developed a set of definitions and results based on the distinguishability properties of the system output. However, as before and as indicated above, it will readily be seen that their definitions and results can be considered as a direct consequence of, and therefore equivalent to, the basic definitions of deterministic parameter identifiability in Section 3.1.

As in previous sections, Grewal's definitions and results are reconstructed in the following for the purpose of comparison and, as previously, the obvious changes are recommended to establish conformity with the definitions of Section 3.1.

Grewal expanded the general deterministic system of Eq. (33) as follows:

$$\dot{\underline{x}}(t,\underline{\theta}) = \underline{g}[\underline{x}(t,\underline{\theta}),\ \underline{u}(t),\ t,\underline{\theta}]; \qquad \underline{x}(t_0) = \underline{x}_0 \tag{88}$$

$$\underline{y}(t,\underline{\theta}) = \underline{h}[\underline{x}(t,\underline{\theta}),\ \underline{u}(t),\ t,\underline{\theta}] \tag{89}$$

where $\underline{x}(t,\underline{\theta})\epsilon R^n$; $\underline{y}(t,\underline{\theta})\epsilon R^m$; $\underline{u}(t)\epsilon \mathcal{U}$; $\underline{\theta}\epsilon\Omega\subset R^p$; $t\epsilon R^+$; $\underline{g}:R^n \times \mathcal{U} \times \Omega \times R^+ \rightarrow R^n$; $\underline{h}:R^n \times \mathcal{U} \times \Omega \times R^+ \rightarrow R^m$. The function $\underline{g}$ is Lipschitz in $\underline{x}$, continuous in $\underline{u}$, and piecewise continuous in t; $\underline{h}$ is continuous in $\underline{x}$ and $\underline{u}$ and piecewise continuous in t. An equivalent formulation may be made for discrete-time systems.

For parameter identifiability based on the distinguishability of the system output, the problem remains that of ascertaining whether or not the parameter values can be uniquely determined from a knowledge of the system input and output. In terms of a specific initial state, $\underline{x}_0$, and a specific input, $\underline{u}(\cdot)\epsilon\mathcal{U}$, the output of the general system may be denoted as

$$\underline{y}(t,\underline{\theta}) = \underline{H}\left[\underline{x}_0,\ \underline{u}(\cdot),t,\underline{\theta}\right]. \tag{90}$$

A single experiment may be defined in terms of a specified initial condition and input pair, $\left[\underline{x}_0,\ \underline{u}(\cdot)\right]$, and the resulting output $\underline{H}\left[\underline{x}_0,\ \underline{u}(\cdot),t,\underline{\theta}\right]$. The collection of all such allowable pairs is denoted by

$$\mathcal{E} = \left\{\left[\underline{x}_0, \underline{u}(\cdot)\right] : \underline{x}_0 \epsilon R^n, \underline{u}(\cdot) \epsilon \mathcal{U}\right\} \tag{91}$$

where $\mathcal{U}$ is the space of piecewise continuous functions.

Having established the required background, the following definitions may be stated.

Definition 3.8 [10]. The pair of parameter values $(\underline{\theta}_1, \underline{\theta}_2)$, $\underline{\theta}_1 \epsilon \Omega$, $\underline{\theta}_2 \epsilon \Omega$, is said to be indistinguishable if $\underline{H}\left[\underline{x}_0, \underline{u}(\cdot), t, \underline{\theta}_1\right] \equiv \underline{H}\left[\underline{x}_0, \underline{u}(\cdot), t, \underline{\theta}_2\right]$ for all $\left[\underline{x}_0, \underline{u}(\cdot)\right] \epsilon \mathcal{E}$ and $0 \le t \le T$. Otherwise, the pair $(\underline{\theta}_1, \underline{\theta}_2)$ is said to be distinguishable.

Defining $S(\underline{x}_0, \underline{\theta}; \rho)$ as an open neighborhood centered at $(\underline{x}_0, \underline{\theta})$ with a radius $\rho > 0$, the definition of (local) parameter identifiability may be stated.

Definition 3.9 [10]. A parameter set Ω is identifiable at $\underline{\theta}_0$ if the pair $(\underline{\theta}_0, \underline{\theta})$ is distinguishable for all $\underline{\theta} \epsilon \Omega$, $\underline{\theta} \neq \underline{\theta}_0$. Further, a parameter set Ω is said to be locally identifiable if there exists a $\rho > 0$ such that the pair $(\underline{\theta}_0, \theta)$ is distinguishable for all $\underline{\theta} \epsilon S(\underline{x}_0, \theta; \rho)$, $\underline{\theta} \neq \underline{\theta}_0$.

It should be noted that the definition of identifiability is independent of whatever method might be used to extract the unknown parameter values from the input/output observation data. Further, although the class $\mathcal{E}$ of experiments is infinite, a finite number of experiments can be designed to distinguish between two systems, e.g., a zero initial state, impulse, or step response.

The concept of parameter identifiability based on output distinguishability can be readily applied to linear, constant, dynamic systems such as those represented by the parameterized differential equations:

$$\dot{\underline{x}}(t,\underline{\theta}) = \underline{A}(\underline{\theta})\underline{x}(t,\underline{\theta}) + \underline{B}(\underline{\theta})\underline{u}(t), \qquad \underline{x}(t_0) = 0 \tag{92}$$

$$\underline{y}(t,\underline{\theta}) = \underline{C}(\underline{\theta})\underline{x}(t,\underline{\theta}) + \underline{D}(\underline{\theta})\underline{u}(t) \tag{93}$$

where $\underline{x}(t,\underline{\theta})\in R^n$, $\underline{u}(t)\in R^r$, $\underline{y}(t,\underline{\theta})\in R^m$, $\underline{A}:\Omega\rightarrow R^{n\times n}$, $\underline{B}:\Omega\rightarrow R^{n\times r}$, $\underline{C}:\Omega\rightarrow R^{m\times n}$, and $\underline{D}:\Omega\rightarrow R^{m\times r}$. The specifications for $\underline{x}$ and $\underline{u}$ cited for Eqs. (88) and (89) continue to hold. It should be noted that the total number of unknown parameters, the dimensionality of the parameter space Ω, equals the total number of elements in the matrices $\underline{A}$, $\underline{B}$, $\underline{C}$ and $\underline{D}$; i.e., $n(n + r + m) + mr = p$. It should also be particularly noted that the system is restricted to zero initial state analysis; i.e., $\underline{x}(t_0) = 0$.

The solution to the system equations may be readily obtained by state transition matrix techniques. Output distinguishability of the given linear system may then be expressed in terms of this solution as given in the following theorem.

<u>Theorem 3.6 [10]</u>

For the linear system described by Eqs. (92) and (93), the pair of parameter values $(\underline{\theta}_1, \underline{\theta}_2)$, $\underline{\theta}_1\in\Omega$, $\underline{\theta}_2\in\Omega$, is <u>indistinguishable</u> if and only if

$$\underline{C}(\underline{\theta}_1) \int_0^t e^{\underline{A}(\underline{\theta}_1)(t-\tau)} \underline{B}(\underline{\theta}_1)\underline{u}(\tau)d\tau + \underline{D}(\underline{\theta}_1)\underline{u}(t)$$
$$= \underline{C}(\underline{\theta}_2) \int_0^t e^{\underline{A}(\underline{\theta}_2)(t-\tau)} \underline{B}(\underline{\theta}_2)\underline{u}(\tau)d\tau + \underline{D}(\underline{\theta}_2)\underline{u}(t)$$

for all $\underline{u}(t)\in\mathcal{U}$ and for $0 \le t \le T$.

Output distinguishability of the given linear system may also be expressed in terms of the system Markov parameters as presented in the following corollary.

Corollary 3.6 [10]

For the linear system described by Eqs. (92) and (93), the pair of parameter vectors $(\underline{\theta}_1, \underline{\theta}_2)$, $\underline{\theta}_1 \epsilon \Omega$, $\underline{\theta}_2 \epsilon \Omega$, is indistinguishable if and only if

$$\underline{C}(\underline{\theta}_1)\underline{A}^{\ell}(\underline{\theta}_1)\underline{B}(\underline{\theta}_1) = \underline{C}(\underline{\theta}_2)\underline{A}^{\ell}(\underline{\theta}_2)\underline{B}(\underline{\theta}_2), \qquad \ell = 0, 1, 2, \ldots.$$

$$\underline{D}(\underline{\theta}_1) = \underline{D}(\underline{\theta}_2).$$

For the given linear system, it is evident that indistinguishability implies that the Markov parameters of the system will be identical at different values of the system parameters, $\underline{\theta}$.

Parameter identifiability criteria for the given linear system may now be obtained by relating Theorem 3.6 and Corollary 3.6 to Definition 3.10.

Theorem 3.7 [10]

For the linear system described by Eqs. (92) and (93), a parameter set Ω is identifiable at $\underline{\theta}_0$ if and only if

$$\underline{\theta} \epsilon \Omega$$

$$\underline{D}(\underline{\theta}_0) = \underline{D}(\underline{\theta}) \quad , \text{ and}$$

$$\underline{C}(\underline{\theta}_0)\underline{A}^{\ell}(\underline{\theta}_0)\underline{B}(\underline{\theta}_0) = \underline{C}(\underline{\theta})\underline{A}^{\ell}(\underline{\theta})\underline{B}(\underline{\theta}) \ , \qquad \ell = 0, 1, 2, \ldots$$

together imply $\underline{\theta}_0 = \underline{\theta}$.

It should be noted that this identical result was presented as Definition 3.6b in Section 3.3, while the result here is obtained as a Theorem. It likewise follows that Theorem 3.7 requires that the mapping from the parameter space Ω into the Markov parameters be injective, which gives rise to the following theorem which is equivalent to Theorem 3.4 of Section 3.3.

<u>Theorem 3.8 [10]</u>

For the linear system described by Eqs. (92) and (93), the parameter set Ω is identifiable at $\underline{\theta}_0$ if and only if the Jacobian of the Markov parameter matrix $\underline{G}(\underline{\theta})$ has constant rank p in an open sphere $S(\underline{x}_0,\underline{\theta};\rho)$ of radius $\rho > 0$ centered at $(\underline{x}_0,\underline{\theta})$; i.e., $\text{rank}[\partial\underline{G}(\underline{\theta})/\partial\underline{\theta}] = p$.

Examples 3.2 and 3.3 pertain equally well to Theorem 3.8 as well as to Theorem 3.4 and will not be repeated.

Distinguishability and identifiability results parallel to those of Theorems 3.6, 3.7, and 3.8 and Corollary 3.6 may also be obtained in the frequency domain for the linear system described by Eqs. (92) and (93). These results are stated in terms of the system transfer function (see Ref. [10]).

A commonly employed technique for the analysis of nonlinear systems is the linearization of the system about an equilibrium or operating point. As the identifiability of a parameter set Ω has been defined at and in terms of a nominal parameter value $\underline{\theta}_0$, the use of linearization techniques with this particular concept of parameter identifiability seems particularly appropriate. Sufficient conditions have been derived under which the local identifiability of the parameters of a linearized system will imply local identifiability of the original nonlinear system.

Consider the general system described by Eqs. (88) and (89) with the assumptions and restrictions as given. Let

$$\underline{x}(t,\underline{\theta}) = \underline{x}_e(t,\underline{\theta}) + \delta\underline{x}(t,\underline{\theta}) \tag{94}$$

$$\underline{y}(t,\underline{\theta}) = \underline{y}_e(t,\underline{\theta}) + \delta\underline{y}(t,\underline{\theta}) \tag{95}$$

and

$$\underline{\theta} = \underline{\theta}_0 + \delta\underline{\theta} \tag{96}$$

where $\underline{x}_e(t,\underline{\theta})$, $\underline{y}_e(t,\underline{\theta})$, and $\underline{\theta}_0$ are equilibrium or operating point values of the system state, system output, and system parameters, respectively, and $\delta\underline{x}(t,\underline{\theta})$, $\delta\underline{y}(t,\underline{\theta})$, and $\delta\underline{\theta}$ are perturbations on $\underline{x}_e(t,\underline{\theta})$, $\underline{y}_e(t,\underline{\theta})$, and $\underline{\theta}_0$, respectively. Equations (88) and (89) may now be rewritten as

$$\dot{\underline{x}}(t,\underline{\theta}) = \underline{g}\left[\underline{x}(t,\underline{\theta}_0 + \delta\underline{\theta}),\ \underline{u}(t),\ t,\ \underline{\theta}_0 + \delta\underline{\theta}\right],\ \underline{x}(t_0) = \underline{x}_0 \tag{97}$$

$$\underline{y}(t,\underline{\theta}) = \underline{h}\left[\underline{x}(t,\underline{\theta}_0 + \delta\underline{\theta})\ \underline{u}(t),\ t,\ \underline{\theta}_0 + \delta\underline{\theta}\right] . \tag{98}$$

With $\dot{\underline{x}}_e = \underline{g}\left[\underline{x}_e,\ \underline{u}(t),\ t,\underline{\theta}_0\right]$, Eqs. (97) and (98) become

$$\delta\dot{\underline{x}}(t,\ \underline{\theta}_0,\ \delta\underline{\theta}) = \frac{\partial\underline{g}}{\partial\underline{x}}\Bigg|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0} \delta\underline{x}(t,\underline{\theta}_0,\delta\underline{\theta}) + \frac{\partial\underline{g}}{\partial\theta}\Bigg|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0} \delta\underline{\theta}$$

$$+ \underline{\ell}_1(\delta\underline{x},\delta\underline{\theta},t) \quad , \qquad \delta\underline{x}(t_0,\underline{\theta}_0,\delta\underline{\theta}) = 0 \tag{99}$$

$$\delta\underline{y}(t,\ \underline{\theta}_0,\ \delta\underline{\theta}) = \left.\frac{\partial \underline{h}}{\partial \underline{x}}\right|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0} \delta\underline{x}(t,\underline{\theta}_0,\delta\underline{\theta}) + \left.\frac{\partial \underline{h}}{\partial \underline{\theta}}\right|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0} \delta\underline{\theta} + \underline{\mathcal{L}}_2(\delta\underline{x},\delta\underline{\theta},t) \tag{100}$$

where $\underline{\mathcal{L}}_1(\delta\underline{x},\delta\underline{\theta},t)$ and $\underline{\mathcal{L}}_2(\delta\underline{x},\delta\underline{\theta},t)$ represent higher order terms.

The linearized system equations then can be written as

$$\delta\dot{\underline{x}}_0(t,\underline{\theta}_0,\delta\underline{\theta}) = \left.\frac{\partial \underline{g}}{\partial \underline{x}}\right|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0} \delta\underline{x}_0(t,\underline{\theta}_0,\delta\underline{\theta}) + \left.\frac{\partial \underline{g}}{\partial \underline{\theta}}\right|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0} \delta\underline{\theta},\quad \delta\underline{x}_0(t_0,\underline{\theta}_0,\delta\underline{\theta}) = 0 \tag{101}$$

$$\delta\underline{y}_0(t,\underline{\theta}_0,\delta\underline{\theta}) = \left.\frac{\partial \underline{h}}{\partial \underline{x}}\right|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0} \delta\underline{x}_0(t,\underline{\theta}_0,\delta\underline{\theta}) + \left.\frac{\partial \underline{h}}{\partial \underline{\xi}}\right|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0} \delta\underline{\theta}\ . \tag{102}$$

For brevity, denote

$$\underline{B}(t,\underline{\theta}_0) = \left.\frac{\partial \underline{g}}{\partial \underline{\theta}}\right|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0} \tag{103}$$

$$\underline{C}(t,\underline{\theta}_0) = \left.\frac{\partial \underline{h}}{\partial \underline{x}}\right|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0} \tag{104}$$

$$\underline{D}(t,\underline{\theta}_0) = \left.\frac{\partial \underline{h}}{\partial \underline{\theta}}\right|_{\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0}\ . \tag{105}$$

Then the solution to the linearized system equations may be written directly as

$$\delta\underline{y}_0(t,\underline{\theta}_0,\delta\underline{\theta}) = \underline{C}(t,\underline{\theta}_0)\int_{t_0}^{t} \underline{\varphi}(t,\tau,\underline{\theta}_0)\underline{B}(\tau,\underline{\theta}_0)d\tau + \underline{D}(t,\underline{\theta}_0) \tag{106}$$

where $\underline{\varphi}(t,\tau,\underline{\theta}_0)$ is the system state transition matrix. It should be noted that $\delta\underline{\theta}$ represents a vector of parameters.

Denoting $\underline{C}(t,\underline{\theta}_0) \int_{t_0}^{t} \underline{\varphi}(t,\tau,\underline{\theta}_0)\underline{B}(\tau,\underline{\theta}_0)d\tau + \underline{D}(t,\underline{\theta}_0)$ as $\underline{N}(t,\underline{\theta}_0)$, an m X p time-varying matrix, Eq. (106) can be rewritten as

$$\delta\underline{y}_0(t,\underline{\theta}_0,\delta\underline{\theta}) = \underline{N}(t,\underline{\theta}_0)\delta\underline{\theta}. \tag{107}$$

It can be seen from Eq. (107) that the parameters $\delta\underline{\theta}$ are identifiable, that is, are uniquely recoverable from Eq. (107), only if the mapping of the parameter space $\delta\underline{\theta}\epsilon\Omega$ into the system outputs $\delta\underline{y}_0(t,\underline{\theta}_0,\delta\underline{\theta})$ is injective for a given input and initial condition. Such will be the case if and only if the columns of $\underline{N}(t,\underline{\theta}_0)$ are linearly independent. It can be shown [4, p. 75] that the columns of $\underline{N}(t,\underline{\theta}_0)$ are linearly independent if and only if the Grammian is nonsingular; i.e.,

$$\int_{t_0}^{t} \underline{N}^T(\tau,\underline{\theta}_0)\underline{N}(\tau,\underline{\theta}_0)d\tau > 0 \ . \tag{108}$$

The above results may be summarized in the following theorem.

Theorem 3.9 [10]

Consider a nonlinear system with a state differential equation

$$\dot{\underline{x}}(t,\underline{\theta}) = \underline{g}[\underline{x}(t,\underline{\theta}),\underline{u}(t),t,\underline{\theta}], \qquad \underline{x}(t_0) = \underline{x}_0 \tag{109}$$

and a linear output equation

$$\underline{y}(t,\underline{\theta}) = \underline{h}[\underline{x}(t,\underline{\theta}),\underline{u}(t),t,\underline{\theta}] \tag{110}$$

where the functions $\underline{g}$ and $\underline{h}$ possess continuous partial derivatives with respect to the components of $\underline{x}$ and $\underline{\theta}$. The linearized state differential equation about $(\underline{x}_e,\underline{\theta}_0)$ is

$$\delta\dot{\underline{x}}_0(t,\underline{\theta}_0,\delta\underline{\theta}) = \underline{A}(t,\underline{\theta}_0)\delta\underline{x}_0(t,\underline{\theta}_0,\delta\underline{\theta}) + \underline{B}(t,\underline{\theta}_0)\delta\underline{\theta}, \quad \delta\underline{x}_0(t_0) = 0 \tag{111}$$

and the output equation is

$$\delta\underline{y}_0(t,\underline{\theta}_0,\delta\underline{\theta}) = \underline{C}(t,\underline{\theta}_0)\delta\underline{x}_0(t,\underline{\theta}_0,\delta\underline{\theta}) + \underline{D}(t,\underline{\theta}_0)\delta\underline{\theta} \tag{112}$$

where the matrices $\underline{A}(t,\underline{\theta}_0)$ and $\underline{B}(t,\underline{\theta}_0)$ are the Jacobians of $\underline{g}$, and $\underline{C}(t,\underline{\theta}_0)$ and $\underline{D}(t,\underline{\theta}_0)$ are the Jacobians of $\underline{h}$ with respect to $\underline{x}$ and $\underline{\theta}$, respectively, each evaluated at $\left[\underline{x}_e,\underline{u}(t),t,\underline{\theta}_0\right]$. Let

$$N(t,\underline{\theta}_0) = \underline{C}(t,\underline{\theta}_0)\int_{t_0}^{t}\underline{\varphi}(t,\tau,\underline{\theta}_0)\underline{B}(\tau,\underline{\theta}_0)d\tau + \underline{D}(t,\underline{\theta}_0) \tag{113}$$

where $\underline{\varphi}(t,\tau,\underline{\theta}_0)$ is the transition matrix of the linearized system of Eqs. (111) and (112).

Then, for the given input $\underline{u}(t)\epsilon\,\mathcal{U}$, if

$$\int_{t_0}^{t}\underline{N}^T(\tau,\underline{\theta}_0)\underline{N}(\tau,\underline{\theta}_0)d\tau > 0 \tag{114}$$

the parameters $\underline{\theta}\epsilon\Omega$ of the nonlinear system can be locally identified.

Another sufficient condition has been derived for the parameter identifiability of a nonlinear system for any input "sufficiently close" to a specified input. Assume that the parameters, $\underline{\theta}\epsilon\Omega$, are fixed and consider only the effects of small perturbations in $\underline{x}$ and/or $\underline{u}$ on the system motion. Then, a sufficient condition for the identifiability of the parameters of a given nonlinear system is given in the following theorem.

Theorem 3.10 [10]

Consider a nonlinear system with a state differential equation

$$\dot{\underline{x}}(t,\underline{\theta}) = \underline{g}[\underline{x}(t,\underline{\theta}),\underline{u}(t),t,\underline{\theta}], \qquad \underline{x}(t_0) = \underline{x}_0 \tag{115}$$

and a linear output equation

$$\underline{y}(t,\underline{\theta}) = \underline{h}[\underline{x}(t,\underline{\theta}),\underline{u}(t),t,\underline{\theta}] \tag{116}$$

where the functions $\underline{g}$ and $\underline{h}$ possess continuous partial derivatives with respect to the components of $\underline{x}$ and $\underline{u}$ at $\underline{x}_e$ and $\underline{u}_0$, respectively. The linearized state differential equation about $(\underline{x}_e,\underline{u}_0)$ is

$$\delta\dot{\underline{x}}_0(t,\underline{\theta}) = \underline{A}(t,\underline{\theta})\delta\underline{x}_0(t,\underline{\theta}) + \underline{B}(t,\underline{\theta})\delta\underline{u}(t), \qquad \delta\underline{x}_0(t_0) = 0 \tag{117}$$

and the output equation is

$$\delta\underline{y}_0(t,\underline{\theta}) = \underline{C}(t,\underline{\theta})\delta\underline{x}_0(t,\underline{\theta}) + \underline{D}(t,\underline{\theta})\delta\underline{u}(t) \tag{118}$$

where the matrices $\underline{A}(t,\underline{\theta})$ and $\underline{B}(t,\underline{\theta})$ are the Jacobians of $\underline{g}$, and $\underline{C}(t,\underline{\theta})$ and $\underline{D}(t,\underline{\theta})$ are the Jacobians of $\underline{h}$ with respect to $\underline{x}$ and $\underline{u}$, respectively, each evaluated at $(\underline{x}_e,\underline{u}_0,t,\underline{\theta})$.

Then, if the parameter set Ω of the linearized system of Eqs. (117) and (118) is identifiable at $\underline{\theta}$, the parameter set Ω of the nonlinear system of Eqs. (115) and (116) is also identifiable at $\underline{\theta}$.

Example 3.5 [10]. Consider a single input/single output, nonlinear, second order system

$$\dot{x}_1(t) = -a_2x_2(t) \tag{119}$$

$$\dot{x}_2(t) = -a_1x_1^3(t) - a_2x_2(t) + u^2(t) \tag{120}$$

$$y(t) = x_1(t). \tag{121}$$

Note the parameterization $\underline{\theta}\epsilon\Omega = [(a_1, a_2)\epsilon R^2 : a_1 \neq 0, a_2 \neq 0]$. For a constant input $u_0 \neq 0$, the equilibrium states of the system are found to be

$$\begin{bmatrix} x_{e1} \\ x_{e2} \end{bmatrix} = \begin{bmatrix} \left(u_0^2/a_{1nom}\right)^{1/3} \\ 0 \end{bmatrix} . \tag{122}$$

Nominal values of the parameters a_1 and a_2 are denoted by a_{1nom} and a_{2nom}. Let δx_1, δx_2, and δu be perturbations on x_{e1}, x_{e2}, and u_0, respectively. The linearized equations are then found to be

$$\begin{bmatrix} \delta\dot{x}(t,\underline{\theta}) \\ \delta\dot{x}_2(t,\underline{\theta}) \end{bmatrix} = \begin{bmatrix} 0 & -a_2 \\ -3x_{e1}^2 & -a_2 \end{bmatrix} \begin{bmatrix} \delta x_1(t,\underline{\theta}) \\ \delta x_2(t,\underline{\theta}) \end{bmatrix} + \begin{bmatrix} 0 \\ 2u_0 \end{bmatrix} \delta u(t) \tag{123}$$

$$\delta y(t,\underline{\theta}) = [1 \quad 0] \begin{bmatrix} \delta x_1(t,\underline{\theta}) \\ \delta x_2(t,\underline{\theta}) \end{bmatrix} . \tag{124}$$

For this linearized set of equations the Markov parameter matrix is calculated as

$$\underline{G}(\underline{\theta}) = \begin{bmatrix} 0 \\ 0 \\ -2u_0 a_2 \\ 2u_0 a_2^2 \\ -12a_1 a_2^2 x_{e1}^2 - 2u_0 a_2^3 \end{bmatrix} \tag{125}$$

and its Jacobian as

$$\frac{\partial \underline{G}(\theta)}{\partial \underline{\theta}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & -2u_0 \\ 0 & 4a_2u_0 \\ -12a_2^2x_{e1}^2 & -24a_1a_2x_{e1}^2 - 6u_0a_2^2 \end{bmatrix} \tag{126}$$

The rank of the Jacobian is clearly equal to 2 for all $\underline{\theta}\epsilon R^2$, and, by Theorem 3.8, the parameterization of the linearized system is identifiable for all $\theta\epsilon R^2$. By Theorem 3.10, the parameterization of the original nonlinear system also is identifiable at any $\underline{\theta} = (a_1, a_2)\epsilon R^2$.

3.5. Least Square Deterministic Parameter Identifiability

The definitions of parameter identifiability stated above have been independent of the method used to recover the unknown parameter values. Bellman and Åström [3] and Mårtensson [15] have proposed an algorithm-oriented definition called locally least square identifiability in which experimental data are combined with a priori or assumed knowledge of the system structures.

In many estimation methods of the small variational type, the values of the unknown parameters are chosen to minimize a quadratic cost functional of the form

$$J_T(\underline{\theta}) = \int_{t_0}^{T} \| \underline{y}(\tau,\underline{\theta}) - \underline{y}_m(\tau)\|^2 d\tau, \qquad T > 0 \tag{127}$$

or

$$J_K(\underline{\theta}) = \sum_{k=0}^{K} \| \underline{y}(k,\underline{\theta}) - \underline{y}_m(k)\|^2, \qquad K = 0, 1, \ldots, \tag{128}$$

where $\underline{y}_m(\cdot)$ is the measured system output for some given input $\underline{u}$ and

$||\cdot||$ denotes the Euclidean norm imposed upon R^m. Local least square identifiability is then defined as follows.

Definition 3.10 [3]. The system parameters $\underline{\theta}$ are said to be locally least square identifiable if and only if $J(\underline{\theta})$ has an isolated local minimum at $\underline{\theta} = \underline{\theta}_0$ and $J(\underline{\theta}_0) = 0$. If the minimum is global, the parameters are said to be globally least square identifiable.

It should be noted that, contrary to later authors, Bellman and Åström [3] did not require $J(\underline{\theta}_0) = 0$. While this may be a valid omission for the extension of the concept of least square identifiability to the stochastic case, for deterministic systems the requirement that $J(\underline{\theta}_0) = 0$ must hold necessarily as

$$\underline{y}(\cdot,\underline{\theta}_0) \equiv \underline{y}_m(\cdot) \ . \tag{129}$$

In terms of Eq. (33), Eqs. (127) and (128) become

$$J_T(\underline{\theta}) = \int_{t_0}^{T} ||\underline{f}(\tau,\underline{x}_0,\underline{\theta},\underline{u}) - \underline{y}(\tau,\underline{\theta})||^2 d\tau, \qquad T > 0 \tag{130}$$

and

$$J_K(\underline{\theta}) = \sum_{k=0}^{K} ||\underline{f}(k,\underline{x}_0,\underline{\theta},\underline{u}) - \underline{y}(k,\underline{\theta})||^2, \qquad K = 0,\ 1,\ \ldots, \tag{131}$$

where the specific time dependencies have been added to the expressions. As equivalent results may be obtained from either formulation above, consider only Eq. (130). Since the integrand of Eq. (130) is non-negative, the requirement that $J(\underline{\theta}_0) = 0$ implies that

$$||\ \underline{f}(\tau,\underline{x}_0,\underline{\theta}_0,\underline{u}) - \underline{y}(\tau,\underline{\theta})\ || \equiv 0 \tag{132}$$

when

$$\| \underline{\theta}_0 - \underline{\theta} \| = 0 \tag{133}$$

which is an equivalent statement of the definition of injectivity of $\underline{f}$ with respect to $\underline{\theta}$. Clearly, then, the definition of least square identifiability is derivative upon, and equivalent to, the definitions of deterministic parameter identifiability given in Section 3.1.

A sufficient condition to insure least square deterministic parameter identifiability is found in a restatement of the implicit function theorem.

Theorem 3.11 [3]

A sufficient condition for the parameter $\underline{\theta}_0$ to be locally least square identifiable is that there exists an open sphere $S(\underline{\theta}_0,\rho)\subset\Omega\subset R^p$ with radius $\rho > 0$ centered at $\underline{\theta}_0$ such that $J_T(\underline{\theta}_0) = 0$ and that the $(p \times p)$ matrix of second-order partial derivatives with respect to the parameters, $J''_T(\underline{\theta})$, is positive definite for all $\underline{\theta}\epsilon S(\underline{\theta}_0,\rho)$.

Grewal and Glover [12] have established the equivalency of least square identifiability and of identifiability based on output distinguishability. Such equivalency will not be demonstrated here since mutual equivalency of all deterministic identifiability definitions has been established through the definitions of Section 3.1.

3.6. Comments on Deterministic Parameter Identifiability

It should be noted that although a number of seemingly different definitions for deterministic parameter identifiability have been presented by different authors, it has been shown that they may be derived from, and hence are equivalent to, the definitions presented in Section 3.1 which require the injectivity of the function $\underline{f}(\underline{x}_0,\underline{\theta}_0,\underline{u})$ of Eq. (33) as the criteria for deterministic parameter identifiability. It is thus recommended that the four definitions of Section 3.1 be considered as the general definitions for determinıstic parameter identifiability and that the resulting theorems and system properties derived in Section 3 be directly related to and derived from them.

It may be further noted that although certain system-specific properties and theorems have been generated, such as Theorem 3.2, there exist only three general methods of establishing deterministic parameter identifiability. These three methods essentially require the certification

(1) that the parameters are uniquely recoverable from the system mathematics (e.g., recoverable by Cramer's Rule, uniquely recoverable by observation of the transfer function, etc.);

(2) that the Jacobian of the Markov parameter matrix is of constant rank equal to the dimensionality of the parameter space in some open neighborhood of the true parameter value; or

(3) that the second partial derivative of a quadratic cost functional with respect to the parameters is positive definite in some open neighborhood of the true parameter value.

The first two methods are particularly suited for analysis of linear systems while the third may be employed with either linear or non-linear systems.

4. STOCHASTIC PARAMETER IDENTIFIABILITY

4.1. Concepts and Definitions

The second of the two basic categories in the study of parameter identifiability has been developed for the class of stochastic systems in which one or both of the noise process, $\underline{w}$ and $\underline{v}$, of Figure 3 and Eq. (12) are present. The general system as it was given in Section 2 is to be considered in the following.

In Section 2, parameter identifiability was considered in the broadest sense as the mathematical assurance of the capability of determining unique values for the unknown parameters of a system from some set of input and output data records. In Section 3, this definition was restated for the deterministic (or noiseless) category of systems in terms of the injectivity of the mapping, for some acceptable input, of the system composite parameter and initial state space into the system output space. The injectivity of this mapping insured, among other properties, the existence of the functional inverse and hence the recoverability of the parameter values. Stochastic parameter identifiability may be considered as the stochastic analog to deterministic parameter identifiability. That is, stochastic parameter identifiability is the mathematical assurance of recovering from noisy observation data the system parameter and/or initial state values of interest in some probabilistic sense. This may be further interpreted as assuring the existence of a sequence of estimates of the unknown quantities which converges in some probabilistic sense to the true values of the quantities.

In the presence of noise, the output data of the given system becomes a sequence of random variables

$$\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_k; \quad k = 1, 2, \ldots \text{ or } [\underline{y}_k]_{k=1}^{\infty} \quad . \tag{134}$$

Based on this observation sequence, in conjunction with an actual or assumed knowledge of the structural properties of the system, the parameter identification problem becomes that of generating a sequence of estimates of the unknown system parameters which will converge in a stochastic sense to the true parameter value. Such an estimation sequence is a measurable function of the observation or output data and is denoted by

$$\hat{\underline{\theta}}_k(\underline{y}_1, \ldots, \underline{y}_k); \quad k = 1, 2, \ldots \quad \text{or} \quad [\hat{\underline{\theta}}_k]_{k=1}^{\infty} \quad . \tag{135}$$

The true parameter value is denoted by $\underline{\theta}_0$. The vectors $\underline{\theta}_0$ and $\hat{\underline{\theta}}_k$ belong to Ω, the space of allowable parameters, which may be considered to be a subset of R^p, the space of ordered p-tuples. Although results have been obtained for more general spaces, the restriction of Ω to R^p is not generally limiting for realizable, physical systems.

Definition 4.1. A sequence of random variables $[\underline{z}_k]_{k=1}^{\infty}$ is said to converge to z in probability (converge stochastically to z) if for every $\epsilon > 0$

$$\lim_{k \to \infty} \Pr\left[|\underline{z}_k - \underline{z}| \geq \epsilon \right] = 0$$

where Pr is a probability measure defined on R^p.

Definition 4.2. An estimation sequence $[\hat{\underline{\theta}}_k]_{k=1}^{\infty}$ of $\underline{\theta}_0 \epsilon \Omega$ which converges stochastically to $\underline{\theta}_0$ is said to be consistent in probability; i.e., $[\hat{\underline{\theta}}_k]_{k=1}^{\infty}$ is a consistent estimate for $\underline{\theta}_0$. This property may be denoted by $\hat{\underline{\theta}}_k \xrightarrow{P} \underline{\theta}_0$.

Stochastic parameter identifiability of the initial state $\underline{x}_0$ and system parameter $\underline{\theta}_0$ can now be defined in terms of the consistency of the sequences of their estimates.

Definition 4.3. The initial state, $\underline{x}_0$, and the system parameters, $\underline{\theta}$, are said to be stochastically identifiable if there exist sequences of estimates $[\hat{\underline{x}}_k]_{k=1}^{\infty}$ and $[\hat{\underline{\theta}}_k]_{k=1}^{\infty}$ which are consistent in probability; i.e., $\hat{\underline{x}}_k \xrightarrow{P} \underline{x}_0$ and $\hat{\underline{\theta}}_k \xrightarrow{P} \underline{\theta}_0$.

It may be noted that, as with the definition of deterministic parameter identifiability, the definition of stochastic parameter identifiability is independent of the method chosen to generate the estimation sequences. However, the standard method normally chosen for generating these sequences has been the maximum likelihood estimate method. For simplicity, consider the case when only $\underline{\theta}_0$ is unknown.

Let $[\underline{y}_k]_{k=1}^{\infty}$ be a sequence of random variables with given joint probability density functions $p_k(\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_k; \underline{\theta})$, $k = 1, 2, \ldots$, which are of known functional form but which depend upon the unknown parameter vector $\underline{\theta} \epsilon \Omega$, the allowable parameter space. Thus, there exists a family of joint probability density functions denoted by $[p_k(\underline{y}_1, \ldots, \underline{y}_k; \underline{\theta}) : \underline{\theta} \epsilon \Omega, k = 1, 2, \ldots]$. For each value of $\underline{\theta} \epsilon \Omega$, there corresponds one member of the family, specifically $[p_k(\underline{y}_1, \ldots, \underline{y}_k; \underline{\theta}) : k = 1, 2, \ldots]$, which is a sequence of joint probability density functions indexed by $k = 1, 2, \ldots$, and parameterized by $\underline{\theta}$. The member of the family corresponding to the true parameter vector $\underline{\theta}_0$ is denoted by $[p_k(\underline{y}_1, \ldots, \underline{y}_k; \underline{\theta}_0): k = 1, 2, \ldots]$. The sequence of maximum likelihood estimates, then, is obtained by selecting $\hat{\underline{\theta}}_k$ such that

$$p_k(\underline{y}_1, \ldots, \underline{y}_k; \hat{\underline{\theta}}_k) = \max_{\underline{\theta} \in \Omega} p_k(\underline{y}_1, \ldots, \underline{y}_k; \underline{\theta}); \qquad k = 1, 2, \ldots \tag{136}$$

The estimates may be expressed more explicitly in terms of the maximum likelihood equation

$$\frac{\partial \log p_k(\underline{y}_1, \ldots, \underline{y}_k; \underline{\theta})}{\partial \underline{\theta}} = 0 \quad , \qquad k = 1, 2, \ldots \tag{137}$$

Under certain restrictions on the joint probability densities of the observations, Wald [24, 25] has shown that the maximum likelihood equation has at least one root which is a consistent estimate of the parameter $\underline{\theta}$ to be estimated. That is, if a given system satisfies the restrictions such that a consistent sequence of estimates for an unknown parameter exists, the method of maximum likelihood estimation will surely produce that sequence.

Let $[\underline{y}_k]_{k=1}^{\infty}$ be a sequence of independent, identically distributed random variables with joint probability density function $p(\underline{y}_1, \ldots, \underline{y}_k; \underline{\theta})$, $k = 1, 2, \ldots$, parameterized by the unknown parameter $\underline{\theta} \epsilon \Omega \subset R^p$, where Ω is the allowable parameter space. The probability density function is denoted by $p(\underline{y}; \underline{\theta})$ and the corresponding cumulative distribution function is denoted by $F(\underline{y}; \underline{\theta})$; i.e., $F(\underline{y}; \underline{\theta}) = \Pr[\underline{y}_k \leq \underline{y}]$. An arbitrary norm on R^p is denoted by $\|\cdot\|$. The following notation is used.

$$p(\underline{y};\underline{\theta},\rho) = \sup_{\|\underline{\theta} - \underline{\theta}'\| < \rho} p(\underline{y};\underline{\theta}'), \qquad \underline{\theta}\epsilon\Omega \text{ and } \rho > 0 \tag{138}$$

$$\psi(\underline{y},r) = \sup_{\|\underline{\theta}\| > r} p(\underline{y};\underline{\theta}), \qquad r > 0 \tag{139}$$

$$p^*(\underline{y};\underline{\theta},\rho) = \begin{cases} p(\underline{y};\underline{\theta},\rho) & , \ p(\underline{y};\underline{\theta},\rho) > 1 \\ 1 & , \ \text{otherwise} \end{cases} \tag{140}$$

$$\psi^*(\underline{y},r) = \begin{cases} \psi(\underline{y},r) & , \ \psi(\underline{y},r) > 1 \\ 1 & , \ \text{otherwise} \end{cases} \tag{141}$$

Wald's restrictions may now be expressed as the following eight assumptions.

<u>Assumption 1</u>. $F(\underline{y};\underline{\theta})$ is either discrete or is absolutely continuous for all $\underline{\theta}\epsilon\Omega$.

<u>Assumption 2</u>. For sufficiently small ρ and for sufficiently large r,

$$\int_{-\infty}^{\infty} \log p^*(\underline{y};\underline{\theta},\rho)\, dF\, (\underline{y};\underline{\theta}_0) < \infty$$

and

$$\int_{-\infty}^{\infty} \log \psi^*(\underline{y};r)\, dF\, (\underline{y};\underline{\theta}) < \infty \qquad \text{for all } \underline{\theta}\epsilon\Omega.$$

<u>Assumption 3</u>. If $\lim_{k\to\infty} \underline{\theta}_k = \underline{\theta}$, then $\lim_{k\to\infty} p(\underline{y};\underline{\theta}_k) = p(\underline{y};\underline{\theta})$ for all $\underline{y}$ except perhaps on a set whose probability measure is zero according to the probability distribution corresponding to $\underline{\theta}_0$.

<u>Assumption 4</u>. If $\underline{\theta}_1 \neq \underline{\theta}_0$, then $F(\underline{y};\underline{\theta}_1) = F(\underline{y};\underline{\theta}_0)$ for at least one value of $\underline{y}$.

Assumption 5. If $\lim_{k\to\infty} ||\underline{\theta}_k|| = \infty$, then $\lim_{k\to\infty} p(\underline{y};\underline{\theta}_k) = 0$ for every $\underline{y}$ except perhaps on a fixed set whose probability measure is zero according to the true parameter $\underline{\theta}_0$.

Assumption 6. $\int_{-\infty}^{\infty} | \log p(\underline{y};\underline{\theta}_0)| dF(\underline{y};\underline{\theta}_0) < \infty$.

Assumption 7. Ω is a closed subset of R^p.

Assumption 8. $p(\underline{y};\underline{\theta},\rho)$ is a measurable function of $\underline{y}$ for $\underline{\theta}\epsilon\Omega$ and $\rho > 0$.

Succeeding work in stochastic parameter identifiability has been primarily accomplished in applying or re-interpreting in more useful, system-oriented terms Wald's restrictions on the joint probability densities of the system observations. A very simplified application of the maximum likelihood estimation method is given in the next example.

Example 4.1. Consider a single observation y of a parameter α corrupted by additive, zero-mean, Gaussian noise with variance σ_n^2:

$$y = \alpha + n. \tag{142}$$

The probability density function of the noise is

$$p(n) = \frac{1}{(2\pi)^{1/2}\sigma_n} \exp\left[\frac{-n^2}{2\sigma_n^2}\right]. \tag{143}$$

Since $y = n - \alpha$, the probability density function of y conditioned upon α is

$$p(y;\alpha) = \frac{1}{(2\pi)^{1/2}\sigma_n} \exp\left[\frac{-(y - \alpha)^2}{2\sigma_n^2}\right] \tag{144}$$

and

$$\log p(y;\alpha) = \text{Constant}_1 \cdot \left[\frac{(y-\alpha)^2}{2\sigma_n^2}\right] \tag{145}$$

The maximum likelihood equation is

$$\frac{\partial \log p(y;\alpha)}{\partial\alpha} = \text{Constant}_2 \cdot \left[\frac{(y-\alpha)}{\sigma_n^2}\right] = 0 \tag{146}$$

and the best (maximum likelihood) estimate $\hat{\alpha}$ for the parameter α is the observed value y; i.e.,

$$\hat{\alpha} = y. \tag{147}$$

4.2. Stochastic Parameter Identifiability from the Properties of the Maximum Likelihood Estimate

Stochastic parameter identifiability results may be derived directly from the convergence properties of the maximum likelihood estimate. Though methods and results developed at later times are easier to apply to a given system, Aoki and Yue [1], using this direct analysis approach, established results for a class of linear, stable, constant-coefficient, discrete-time dynamic systems with both plant and observation noise present. Because of the importance of these results and as an aid to understanding the properties of the maximum likelihood estimate, the major parts of their results are reproduced in the following.

A class of systems nearly identical to that of Section 3.2, except with the addition of noise sources, was investigated. As with the previous formulation, this class of systems represents a large class of realistic problems. The system class is characterized by the equations

$$\underline{z}(j+1) = \underline{A}\,\underline{z}(j) + \underline{b}\,u_j, \qquad \underline{z}(0) = \underline{z}_0 \tag{148}$$

$$x_j = \underline{h}^T \underline{z}(j) \tag{149}$$

where

$$\underline{A} = \begin{bmatrix} -a_1 & 1 & 0 & 0 & \dots & 0 \\ -a_2 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ -a_{n-1} & 0 & 0 & 0 & \dots & 1 \\ -a_n & 0 & 0 & 0 & \dots & 0 \end{bmatrix}, \qquad (n \times n);$$

$$\underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad (n \times 1); \quad \underline{h} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (n \times 1);$$

and the parameters a_i and b_i are unknown constants with $|a_i| < \infty$ and $|b_i| < \infty$, $1 \le i \le n$.

Output observations, y_j, are made with additive noise

$$y_j = x_j + \eta_j. \tag{150}$$

The output noise process $\{\eta_j\}$ is restricted to be independent and identically distributed as zero-mean, normal with variance σ^2.

As in Section 3.2, the system may be alternately represented by

$$x_j + \sum_{i=1}^{n} a_i x_{j-1} = \sum_{i=1}^{n} b_i u_{j-i} \tag{151}$$

$$y_j = \sum_{i=1}^{j} (x_i + \eta_i) \quad , \qquad j = 0, 1, 2, \dots, L-1 \tag{152}$$

or with Toeplitz matrices as

$$\underline{A}_L\underline{x}_L = \underline{B}_L\underline{u}_L + \underline{E}_L\underline{z}_0 \tag{153}$$

$$\underline{y}_L = \underline{x}_L + \underline{\eta}_L \tag{154}$$

or

$$\underline{x}_L = \underline{H}_L\underline{\theta} + \underline{E}_L\underline{z}_0 \tag{155}$$

$$\underline{y}_L = \underline{x}_L + \underline{\eta}_L \tag{156}$$

where

$$\underline{x}_L = [x_0, x_1, \ldots, x_{L-1}]^T, \quad (L \times 1) ;$$

$$\underline{y}_L = [y_0, y_1, \ldots, y_{L-1}]^T, \quad (L \times 1) ;$$

$$\underline{u}_L = [u_0, u_1, \ldots, u_{L-1}]^T, \quad (L \times 1) ;$$

$$\underline{\eta}_L = [\eta_0, \eta_1, \ldots, \eta_{L-1}]^T, \quad (L \times 1) ;$$

$$\underline{\theta} = [a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n]^T, \quad (2n \times 1);$$

$$\underline{A}_L = \underline{I}_L + \sum_{i=1}^{n} a_i\underline{S}_L^i, \quad (L \times L);$$

$$\underline{B}_L = \sum_{i=1}^{n} b_i\underline{S}_L^i, \quad (L \times L) ;$$

$$\underline{E}_L = \left[\begin{array}{c} \underline{I}_{n\times n} \\ \hline 0_{L-n,n} \end{array}\right], \quad (L \times n) ;$$

$$\underline{H}_L = [\underline{S}_L\underline{x}_L, \underline{S}_L^2\underline{x}_L, \ldots, \underline{S}_L^n\underline{x}_L, \underline{S}_L\underline{u}_L, \ldots, \underline{S}_L^n\underline{u}_L], \quad (L \times 2n);$$

and $\underline{S}_L$ is the L x L shift matrix defined such that $\underline{S}_L(i,j) = \delta_{i,j+1}$

For the deterministic portion of the system [Eqs. (148) and (149) or Eqs. (151), (153) or (155)] results may be stated which are identical to Property 1, Property 2, Property 3, Theorem 3.1, Theorem 3.2, and Theorem 3.3 of Section 3.2 with the "r" of Section 3.2 replaced by "n".

In the following development, it is assumed that the true parameter $\underline{\theta}_0$ lies in the interior of a given compact subset Θ_S of R^{2n} and that all systems with $\underline{\theta}\in\Theta_S$ are stable. The assumption of compactness is not truly limiting as parameter values, from a priori knowledge of the system, usually fall within a limitable range thus permitting the parameter set to be contained in a compact region. System stability is assumed to permit the investigation of asymptotic properties of the system.

Under the conditions given, a characterization of the maximum likelihood estimates follows. From Eqs. (153) and (154),

$$\underline{y}_L = \underline{\eta}_L + \underline{A}_L^{-1}(\underline{B}_L\underline{u}_L + \underline{E}_L\underline{z}_0). \tag{157}$$

Noting that $\underline{\eta}_L = \underline{y}_L - \underline{x}_L$, the probability density function of the output $\underline{y}_L$ as parameterized by $\underline{\theta}$ and $\underline{z}_0$ is

$$p(\underline{y}_L;\underline{\theta},\underline{z}_0) = \text{Constant}\left\{\exp\left[-\frac{1}{2\sigma^2}\left\|\underline{y}_L - \underline{A}_L^{-1}(\underline{B}_L\underline{u}_L + \underline{E}_L\underline{z}_0)\right\|^2\right]\right\}. \tag{158}$$

Let $\hat{\underline{\theta}}_L$, $\hat{\underline{z}}_{0L}$ denote the maximum likelihood estimates of $\underline{\theta}$ and $\underline{z}_0$ from the observed data, $\underline{u}_L$ and $\underline{y}_L$; that is,

$$\log p\,(\underline{y}_L;\hat{\underline{\theta}}_L,\hat{\underline{z}}_{0L}) = \max_{\substack{\underline{\theta}\in\Theta_S \\ \underline{z}_0\in R^n}} \log p(\underline{y}_L;\underline{\theta},\underline{z}_0)\ . \tag{159}$$

For any $\underline{\theta}\in\Theta_S$, $\max_{\underline{z}_0\in R^n} \log p(\underline{y}_L;\underline{\theta},\underline{z}_0)$ is achieved by

$$\tilde{\underline{z}}_{0L}(\underline{\theta}) = \left(\underline{E}_L^T\underline{A}_L^{T-1}\underline{A}_L^{-1}\underline{E}_L\right)^{-1}\underline{E}_L^T\underline{A}_L^{T-1}\left(\underline{y}_L-\underline{A}_L^{-1}\underline{B}_L\underline{u}_L\right). \tag{160}$$

Thus, $\hat{\underline{\theta}}_L$ is obtained by

$$\min_{\underline{\theta}\in\Theta_S} J_L(\underline{\theta}) = J_L(\hat{\underline{\theta}}_L) \tag{161}$$

where the likelihood function $J_L(\underline{\theta})$ is

$$\begin{aligned} J_L(\underline{\theta}) &= \|\underline{y}_L - \underline{A}_L^{-1}\left[\underline{B}_L\underline{u}_L + \underline{E}_L\tilde{\underline{z}}_{0L}(\underline{\theta})\right]\|^2 \\ &= \|\underline{A}_L\underline{y}_L - \underline{B}_L\underline{u}_L - \underline{E}_L\tilde{\underline{z}}_{0L}(\underline{\theta})\|^2_{\left(\underline{A}_L\underline{A}_L^T\right)^{-1}} \end{aligned} \tag{162}$$

and

$$\hat{\underline{z}}_{0L} = \tilde{\underline{z}}_{0L}\left(\hat{\underline{\theta}}_L\right). \tag{163}$$

It has been shown that the consistency of the maximum likelihood estimates holds if the almost sure (a.s.) convergence of the likelihood function is established. This property leads to the following Propositions.

<u>Proposition 1</u>.

For all $\underline{\theta}\in\Theta_S$, $J_L(\underline{\theta})/L \to J(\underline{\theta})$ with probability one, where

$$J(\underline{\theta}) = \lim_{L\to\infty}\frac{1}{L}\,E\,J_L(\underline{\theta}) = \sigma^2 + \lim_{L\to\infty}\frac{1}{L}\|\underline{A}_L\underline{x}_L - \underline{B}_L\underline{u}_L^T\|^2_{\left(\underline{A}_L\underline{A}_L^T\right)^{-1}}$$

and E denotes the expectation operator.

Note that $J(\underline{\theta}_0) = \sigma^2 = \min J(\underline{\theta})$ which satisfies Eq. (153), namely, the true parameter vector, if $\underline{\theta}_0$, is unique in the representation of Eq.

(153). In the following, the subscript "0" denotes the particular element at the true parameter value $\underline{\theta}_0$; i.e., $\underline{A}_{L,0} = \underline{A}_L(\underline{\theta}_0)$, $\underline{B}_{L,0} = \underline{B}_L(\underline{\theta}_0)$, etc. As seen in the next proposition, only those $\underline{\theta}$ which give rise to $J(\underline{\theta}) = J(\underline{\theta}_0) = \sigma^2$ are of interest.

Proposition 2.

With probability one, $\hat{\underline{\theta}}_L$ converges to $\hat{\underline{\theta}}_L \in \Theta_0 \cap \Theta_S$, where

$$\Theta_0 = [\underline{\theta}: J(\underline{\theta}) = J(\underline{\theta}_0)].$$

Proposition 3.

$J(\underline{\theta}) = J(\underline{\theta}_0)$ if and only if

$$\lim_{L \to \infty} \frac{1}{L} \|(\underline{A}_L \underline{B}_{L,0} - \underline{A}_{L,0} \underline{B}_L)\underline{u}_L\|^2 = 0 .$$

Proposition 2a.

(An immediate consequence of Proposition 2.) Given that $\underline{\theta}_0$ is unique, the maximum likelihood estimate $\hat{\underline{\theta}}_L$ converges to $\underline{\theta}_0$ with probability one if and only if $\Theta_0 \cup \Theta_S$ is a singleton.

A necessary and sufficient condition to insure that the condition of Proposition 2a is always satisfied is contained in the following theorem.

Theorem 4.1 [1]

Given the linear dynamic system in any of its equivalent representations above, such that $\underline{b} \neq 0$ and $(\underline{A}, \underline{b})$ is completely controllable, the maximum likelihood estimate $\hat{\underline{\theta}}_L$ converges to $\underline{\theta}_0$ with probability one if and only if

$$\lim_{L \to \infty} \frac{1}{L} \underline{U}^T_{L,2n} \underline{U}_{L,2n} > 0$$

where

$$\underline{U}_{L,2n} = \left[\underline{S}_L\underline{u}_L,\ \underline{S}_L^2\underline{u}_L,\ \ldots,\ \underline{S}_L^{2n}\underline{u}_L\right].$$

The necessary and sufficient condition of Theorem 4.1 can also be stated in various forms for the purpose of different applications.

Corollary 4.1 [1]

Given the linear dynamic system in any of its equivalent representations above, such that $\underline{b} \neq 0$ and $(\underline{A},\underline{b})$ is completely controllable, the maximum likelihood estimate $\hat{\underline{\theta}}_L$ converges to $\underline{\theta}_0$ with probability one if and only if

$$\lim_{L\to\infty} \frac{1}{L}\,\underline{H}_L^T\underline{H}_L > 0$$

or, equivalently,

$$\lim_{L\to\infty} \frac{1}{L}\,M_L > 0$$

where

$$M_L = \frac{1}{\sigma^2}\,\underline{H}_L^T\left(\underline{A}_L\underline{A}_L^T\right)^{-1}\underline{H}_L\ .$$

Theorem 4.1 can be viewed as the stochastic version of Theorem 3.2, and Corollary 4.1 can be viewed as the stochastic version of Property 1 of Section 3.2 modified such that $\underline{z}_0 \neq 0$; i.e., $\underline{\theta} = \left(\underline{H}_L^T\underline{H}_L\right)^{-1}\underline{H}_L^T\left(\underline{x}_L - \underline{E}_L\underline{z}_0\right)$ if and only if $\underline{H}_L^T$ is positive definite.

At this point, a link may be established between the deterministic parameter identifiability properties of the system and the stochastic parameter identifiability properties of the system in terms of y_j, the noise-corrupted output, and M_L, the Fisher information matrix.

Theorem 3.2a [20]

If the system described by Eqs. (148), (149) and (150) is stable and $\underline{z}_0 = 0$, then $\underline{\theta} = [a_1, \ldots, a_n, b_1, \ldots, b_n]^T$ is identifiable, I.D.S. (equivalently, zero-state deterministically parameter identifiable per Section 3.1), if and only if the Fisher information matrix, M_L, is positive definite for some finite L, where

$$M_L = \int p(\underline{y}_L;\underline{\theta},\underline{z}_0)\ [\nabla_{\underline{\theta}} \log p(\underline{y}_L;\underline{\theta},\underline{z}_0)]$$

$$[\nabla_{\theta} \log p(\underline{y}_L;\underline{\theta},\underline{z}_0)]^T d\underline{y}_L,$$

and

$$\nabla_{\underline{\theta}} s = \left(\frac{\partial s}{\partial a_1}, \ldots, \frac{\partial s}{\partial a_n}, \frac{\partial s}{\partial b_1}, \ldots, \frac{\partial s}{\partial b_n}\right)^T$$

for any scalar s.

By direct calculation, it may be shown that $M_L = \underline{H}_L^T \left(\underline{A}_L\underline{A}_L^T\right)^{-1}\underline{H}_L$. It may also be shown (Appendix B, [1]) that $\left(\underline{A}_L\underline{A}_L^T\right)$ is bounded such that

$$\rho_1\underline{I}_L \le \underline{A}_L\underline{A}_L^T \le \rho_2\underline{I}_L$$

where

$$0 < \rho_1 < \rho_2 < \infty .$$

Thus, $\underline{H}_L^T\underline{H}_L$ is positive definite if and only if M_L is positive definite and, by Property 1 of Section 3.2, the system is identifiable, I.D.S., if and only if $M_L > 0$ for some finite L.

Note the marked similarity between Corollary 4.1 and Theorem 3.2a with the distinct difference that M_L must be positive definite for some finite L in the deterministic case while the limit of $\frac{1}{L} M_L$ must remain positive definite in the stochastic case.

The initial state does not effect the convergence properties of the parameter estimates $\hat{\underline{\theta}}_L$. However, the initial state estimate, as a function of $\hat{\underline{\theta}}_L$ and $\underline{y}_L$, is given uniquely by Eqs. (160) and (163); i.e.,

$$\hat{\underline{z}}_{0L} = \left(\underline{E}_L^T \hat{\underline{A}}_L^{T-1} \underline{A}_L^{-1} \underline{E}_L\right)^{-1} \underline{E}_L^T \hat{\underline{A}}_L^{T-1}\left(\underline{y}_L - \hat{\underline{A}}_L^{-1} \hat{\underline{B}}_L \underline{u}_L\right) . \tag{164}$$

Theorem 4.2 [1]

If the conditions of Theorem 4.1 are satisfied, then $E\ \hat{\underline{z}}_{0L} \to \underline{z}_0$ as $L \to \infty$, where E denotes the expectation operator.

The system under consideration may be extended to include an additive plant noise process which may be represented by the following Gauss-Markov model

$$\underline{z}(j+1) = \underline{A}\ \underline{z}(j) + \underline{b}\ u_j + \underline{d}\ \xi_j \tag{165}$$

$$x_j = \underline{h}^T \underline{z}(j) + b_0 u_j + \xi_j \tag{166}$$

where $\{\xi_j\}$ is a Gaussian white noise process identically, normally distributed with zero-mean and variance λ^2; $\underline{d} = [d_1, d_2, \ldots, d_n]^T$; $\underline{A}$, $\underline{h}$, and $\underline{b}$ are defined as before. The unknown system parameters are a_i, b_i, d_i for $1 \le i \le n$, b_0 and λ^2; i.e., $\underline{\theta} = [\underline{a}, \underline{b}, \underline{d}, b_0, \lambda^2]^T$. The output is observed with additive noise

$$y_j = x_j + \eta_j \tag{167}$$

as before with $\{\eta_j\}$ independent and identically distributed, without loss of generality, as zero-mean, normal with unit variance. The noise processes $\{\xi_j\}$ and $\{\eta_j\}$ are independent.

Equivalent representations of the expanded system may be made as with the initial system but will not be presented at this time (see Ref. [1]). The following results were obtained.

Theorem 4.3 [1]

Given the system described by Eqs. (165), (166), and (167) with $\underline{d} \neq 0$ and $(\underline{A}, \underline{d})$ completely controllable, $\hat{\underline{\theta}}_L$ $\left[\text{where } \underline{\theta} = (\underline{a},\underline{b},\underline{d},b_0,\lambda^2)^T\right]$ converges with probability one to $\underline{\theta}_0$ if and only if

$$V = \lim_{L\to\infty} \frac{1}{L} \tilde{\underline{U}}^T_{L,n} \tilde{\underline{U}}_{L,n} > 0$$

where

$$\tilde{\underline{U}}_{L,n} = \left[\underline{u}_L, \underline{S}_L\underline{u}_L, \ldots, \underline{S}^n_L\underline{u}_L\right].$$

Theorem 4.3a [1]

Given the system described by Eqs. (165), (166) and (167) such that $[\underline{A},(\underline{b},\underline{d})]$ is completely controllable, $\hat{\underline{\theta}}_L$ converges with probability one if

$$\lim_{L\to\infty} \frac{1}{L} \tilde{\underline{U}}^T_{L,2n} \tilde{\underline{U}}_{L,2n} > 0$$

where

$$\tilde{\underline{U}}_{L,2n} = \left[\underline{u}_L, \underline{S}_L\underline{u}_L, \ldots, \underline{S}^{2n}_L\underline{u}_L\right].$$

Selected proofs are presented in Appendix B.

4.3 Stochastic Parameter Identifiability Generalized from Constrained Maximum Likelihood (CML) Estimate Properties

Using the convergence properties of the maximum likelihood estimate, Tse and Anton [22] developed stochastic parameter identifiability criteria expressed in terms of conditional probability densities for the sequence of system observation statistics. Their definition of stochastic parameter identifiability remained the existence of consistent estimates as in Definitions 4.2 and 4.3.

As before, $\{\underline{y}_k\}_{k=1}^{\infty}$ denotes a sequence of observation statistics with a joint probability density function $p_k(\underline{y}_1, \ldots, \underline{y}_k; \underline{\theta})$, k = 1, 2, ..., parameterized by the unknown parameter $\underline{\theta} \in \Omega \subset R^P$. Although the development by Tse and Anton was set in a more general separable metric space, in consonance with previous remarks, Ω is taken as a compact subset of R^P. The true parameter, $\underline{\theta}_0$, is assumed to lie in the interior of Ω. An arbitrary norm on R^P is denoted by $\|\cdot\|$. Denoting the observation sequence

$$Y_k = \{\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_k\}, \qquad [168]$$

the sequence of true joint probability density functions may be denoted as

$$p_k(\underline{y}_1, \ldots, \underline{y}_k; \underline{\theta}_0) = p_k(Y_k; \underline{\theta}_0), \qquad k = 1, 2, \ldots \qquad (169)$$

Since the context clearly indicates which density is indicated, the "k" subscript is routinely deleted, yielding $p(Y_k; \underline{\theta}_0)$. By Bayes rule, a conditional probability density function may be defined as

$$p\left(\underline{y}_k | Y_{k-1}; \underline{\theta}\right) = p\left(Y_k; \underline{\theta}\right) / p\left(Y_{k-1}; \underline{\theta}\right), \qquad k = 1, 2, \ldots \qquad (170)$$

Any new information obtained from the kth sampling will be contained in the conditional probability density of Eq. (170). For $\underline{\theta}\epsilon\Omega$ and $\rho > 0$ define a regional conditional probability density

$$p(\underline{y}_k, \rho | Y_{k-1}; \underline{\theta}) = \sup_{\|\underline{\theta} - \underline{\theta}'\| \le \rho} p(\underline{y}_k | Y_{k-1}; \underline{\theta}') . \tag{171}$$

The following assumptions are made.

<u>Assumption 1</u>. The probability density function $p(Y_k; \underline{\theta})$ is measurable in Y_k with respect to $p(Y_k; \underline{\theta}_0)dY_k$ and is continuous in $\underline{\theta}\epsilon\Omega$ for Y_k almost everywhere; i.e., for any $\epsilon > 0$ and $\underline{\theta}\epsilon\Omega$, there exists a $\delta(\epsilon) > 0$ such that for all $\underline{\theta}'\epsilon\Omega$ with $\|\underline{\theta} - \underline{\theta}'\| < \delta$ we have $|p(Y_k; \underline{\theta}) - p(Y_k; \underline{\theta}')| < \epsilon$ for Y_k almost everywhere.

<u>Assumption 2</u>.

$$\int_{R^k} \log p\,(\underline{y}_k, \rho | Y_{k-1}; \underline{\theta})\; p(Y_k; \underline{\theta}_0) dY_k < \infty \tag{172}$$

for each $\underline{\theta}\epsilon\Omega$, for some $\rho > 0$ and for all $k = 1, 2, \ldots;$ and

$$\int_{R^k} \log p(\underline{y}_k | Y_{k-1}; \underline{\theta}_0)\; p(Y_k; \underline{\theta}_0) dY_k < \infty \tag{173}$$

for all $k = 1, 2, \ldots,$.

<u>Assumption 3</u>.

$$\mathrm{Var}\left\{ \sum_{i=1}^{k} \log p(\underline{y}_k, \rho | Y_{k-1}; \underline{\theta}) \right\} = 0(k^2) \tag{174}$$

for all $\underline{\theta}\epsilon\Omega$ and some $\rho_0 > 0$ where $0 \le \rho \le \rho_0$ and where $0(k^2)$ is defined such that

$$\lim_{k \to \infty} \frac{0(k^2)}{k^2} = 0 \; .$$

<u>Assumption 4</u>. Defining the set $\mathcal{B}_k(\underline{\theta}) = \left[Y_k; p(Y_k; \underline{\theta}) = 0\right]$,

$$\mathcal{B}_k(\underline{\theta}) = \mathcal{B}_k(\underline{\theta}') \tag{175}$$

for $\underline{\theta}$, $\underline{\theta}' \epsilon \Omega$ and for all k = 1, 2, ..., .

Conceptually, the second and third assumptions restrict the growth rate of accumulated information about the unknown parameter relative to the accumulated uncertainty. The fourth assumption implies that, for two different parameters, the corresponding density functions must have all the impulses located at the same points in the observation space.

Since the only information about $\underline{\theta}_0$ is contained in the observation statistics $\{\underline{y}_k\}_{k=1}^{\infty}$ with their corresponding joint density function $p(Y_k; \underline{\theta})$, k = 1, 2, ..., if there exist two parameter vectors $\underline{\theta}_1$, $\underline{\theta}_2 \epsilon \Omega$, $\underline{\theta}_1 \neq \underline{\theta}_2$ such that

$$p(Y_k; \underline{\theta}_1) = p(Y_k; \underline{\theta}_0) \tag{176}$$

or

$$p\left(\underline{y}_k \middle| Y_{k-1}; \underline{\theta}_1\right) = p\left(\underline{y}_k \middle| Y_{k-1}; \underline{\theta}_2\right), \quad \text{all } k = 1, 2, \ldots, \tag{177}$$

the two parameters are indistinguishable in Ω.

<u>Definition 4.4 [22]</u>. Two parameters $\underline{\theta}_1$, $\underline{\theta}_2 \epsilon \Omega$, $\underline{\theta}_1 \neq \underline{\theta}_2$ are said to be <u>unresolvable</u> if the equality

$$p(\underline{y}_k | Y_{k-1}; \underline{\theta}_1) = p(\underline{y}_k | Y_{k-1}; \underline{\theta}_2) \tag{178}$$

holds with probability one for all except a finite number of integers k > 0; i.e., if Eq. (178) holds with respect to the measure $p(Y_k;\ \underline{\theta}_1)dY_k$ as well as $p(Y_k;\ \underline{\theta}_2)dY_k$.

Definition 4.5[22]. The set Ω is said to be identifiable if no two elements of Ω are unresolvable.

Definition 4.6 [22]. For the observation sequence, Y_k, the constrained maximum likelihood (CML) estimate of $\underline{\theta}_0$ is defined as $\hat{\underline{\theta}}_k$, which satisfies

$$p(Y_k;\ \hat{\underline{\theta}}_k) = \max_{\underline{\theta}\in\Omega} p(Y_k;\ \underline{\theta}). \tag{179}$$

Essentially, this is the maximum likelihood estimate of $\underline{\theta}_0$ but takes into account the a priori knowledge that $\underline{\theta}_0$ is constrained to be a member of Ω; i.e., $\underline{\theta}_0\in\Omega$. Since Ω is compact and $p(Y_k;\ \underline{\theta})$ is continuous almost surely by Assumption 1, at least one solution to Eq. (179) exists almost surely. Thus, the CML generated estimate sequence $[\hat{\underline{\theta}}_k]_{k=1}^{\infty}$ is a consistent estimate for $\underline{\theta}_0$ if $\underline{\theta}_0$ is unique, a consequence of the properties of the maximum likelihood estimate. However, if there exist two parameter vectors $\underline{\theta}_1$, $\underline{\theta}_2\in\Omega$, $\underline{\theta}_1 \neq \underline{\theta}_2$ such that

$$\lim_{k\to\infty} p(Y_k;\ \underline{\theta}_1) = \lim_{k\to\infty} p(Y_k;\ \underline{\theta}_2) \tag{180}$$

then, obviously, $[\hat{\underline{\theta}}_k]_{k=1}^{\infty}$ will fail to converge.

Definition 4.7 [22]. Two parameters, $\underline{\theta}_1$, $\underline{\theta}_2\in\Omega$, $\underline{\theta}_1 \neq \underline{\theta}_2$, are said to be CML unresolvable if

$$\lim_{k\to\infty} p(\underline{y}_k|Y_{k-1};\ \underline{\theta}_1) = \lim_{k\to\infty} p(\underline{y}_k|Y_{k-1};\ \underline{\theta}_2) \tag{181}$$

with probability one.

Definition 4.8 [22]. The set Ω is said to be CML identifiable if no two elements in Ω are CML unresolvable.

Theorem 4.4 [22].

A sufficient condition for Ω to be CML identifiable is that, for all $\underline{\theta}_1$, $\underline{\theta}_2 \epsilon \Omega$, $\underline{\theta}_1 \neq \underline{\theta}_2$, there exists an infinite set $L \subset I^+$, I^+ the set of positive integers, such that the inequality

$$p(\underline{y}_k | Y_{k-1}; \underline{\theta}_1) \neq p(\underline{y}_k | Y_{k-1}; \underline{\theta}_2)$$

holds with nonzero probability with respect to $\underline{\theta}_1$ and $\underline{\theta}_2$ uniformly in $k \epsilon L$.

It should be noted that Theorem 4.4 provides sufficient conditions to insure CML identifiability. If the conditions of the Theorem are not met, it does not necessarily imply that the CML estimation method will fail; rather, it implies there exists no guarantee of consistency of the estimate. In point of fact, certain studies have indicated that the CML estimate is "fairly" consistent even though some of the required assumptions are violated.

Example 4.2 [23]. Consider the linear, time-invariant system given by

$$\underline{x}_{j+1} = \underline{A}\, \underline{x}_j + \underline{w}_j \tag{182}$$

$$\underline{y}_j = \underline{C}\, \underline{x}_j + \underline{v}_k \tag{183}$$

where $\underline{x}_j \epsilon R^n$, $\underline{y}_j \epsilon R^m$ and $\{\underline{w}_j\}$ and $\{\underline{v}_j\}$ are zero-mean Gaussian noise processes with covariances

$$E\left[\underline{w}_k\underline{w}_j^T\right] = \underline{R}\ \delta_{kj},$$

$$E\left[\underline{v}_k\underline{v}_j^T\right] = \underline{Q}\ \delta_{kj},$$

and

$$E\left[\underline{w}_k\underline{v}_j^T\right] = \underline{D}\ \delta_{kj}.$$

Let $\underline{\theta} = [\underline{x}_0, \underline{A}, \underline{C}, \underline{R}, \underline{Q}, \underline{D}]^T$ and assume that $\underline{A}$ is stable, $(\underline{A},\underline{C})$ is an observable pair and $(\underline{A},\underline{B})$ is a controllable pair, where $\underline{B}$ is the steady-state Kalman filter gain given by Eq. (188) (see below).

Two parameters $\underline{\theta}_1$, $\underline{\theta}_2 \epsilon R^p$, $\underline{\theta}_1 \neq \underline{\theta}_2$ are defined to be CML unresolvable if the equality

$$p(\underline{y}_k|Y_{k-1}, \underline{\theta}_1) = p(\underline{y}_k|Y_{k-1},\underline{\theta}_2) \tag{184}$$

holds with probability 1 with respect to $\underline{\theta}_1$ and $\underline{\theta}_2$ as $k\rightarrow\infty$. Since the system is linear and the noises are Gaussian, $p(\underline{y}_k|Y_{k-1}, \underline{\theta})$ is Gaussian with mean $\hat{\underline{y}}_{k|k-1}$ and covariance $\underline{C}\ \underline{P}\ \underline{C}^T + \underline{Q}$ as $k\rightarrow\infty$ (steady state). $\underline{P}$ and $\hat{\underline{y}}_{k|k-1}$ are given by the usual steady-state Kalman filter equations.

$$\hat{\underline{y}}_{k|k-1} = \underline{C}\ \hat{\underline{x}}_{k|k-1} \tag{185}$$

$$\hat{\underline{x}}_{k|k-1} = \underline{A}\ \hat{\underline{x}}_{k-1|k-2} + \underline{B}\ \underline{\nu}_{k-1} \tag{186}$$

$$\underline{\nu}_k = \underline{y}_k - \underline{C}\ \hat{\underline{x}}_{k|k-1} \tag{187}$$

$$\underline{B} = (\underline{A}\ \underline{P}\ \underline{C}^T + \underline{D})\ (\underline{C}\ \underline{P}\ \underline{C}^T + \underline{Q})^{-1} \tag{188}$$

$$\underline{P} = \underline{A}\ \underline{P}\ \underline{A}^T + \underline{R} - \underline{B}\ (\underline{C}\ \underline{P}\ \underline{C}^T + \underline{Q})\underline{B}^T \tag{189}$$

It can then be shown that $\underline{\theta}_1$, $\underline{\theta}_2 \epsilon R^p$, $\underline{\theta}_1 \neq \underline{\theta}_2$ are CML unresolvable if and only if there exists a nonsingular matrix $\underline{T}$ such that

$$\underline{A}_1 = \underline{T}\,\underline{A}_2\,\underline{T}^{-1} \tag{190}$$

$$\underline{B}_1 = \underline{T}\,\underline{B}_2 \tag{191}$$

$$\underline{C}_1 = \underline{C}_2\,\underline{T}^{-1} \tag{192}$$

$$\underline{C}_1\,\underline{P}_1\,\underline{C}_1^T + \underline{Q}_1 = \underline{C}_2\,\underline{P}_2\,\underline{C}_2^T + \underline{Q}_2\ . \tag{193}$$

If Eq. (193) is satisfied, it is clear that Eq. (184) holds with probability 1 as $k \to \infty$, and thus, $\underline{\theta}_1$ and $\underline{\theta}_2$ are CML unresolvable. Then in the steady state,

$$\hat{\underline{y}}_{1,k|k-1} = \hat{\underline{y}}_{2,k|k-1} \tag{194}$$

with probability 1, for all k

$$\underline{C}_1\,\underline{P}_1\,\underline{C}_1^T + \underline{Q}_1 = \underline{C}_2\,\underline{P}_2\,\underline{C}_2^T + \underline{Q}_2. \tag{195}$$

Equations (185) through (188) and (191) imply that

$$\underline{C}_1\,\underline{A}_1^{\ell}\,\underline{B}_1 = \underline{C}_2\,\underline{A}_2^{\ell}\,\underline{B}_2, \qquad \ell = 0,\ 1,\ 2,\ \ldots \tag{196}$$

But Eq. (196) implies that the two steady-state Kalman filters [Eqs. (185) through (188)] have the same impulse response. Since $(\underline{A}_i, \underline{C}_i)$ is an observable pair and $(\underline{A}_i, \underline{B}_i)$ is a controllable pair, Eq. (193) results.

Additionally, Glover and Willems [9] provide an example of a deterministic system driven by a white, Gaussian noise input.

4.4. Stochastic Parameter Identifiability from the Information Matrix

Under essentially equivalent assumptions on the joint probability density function of the system observation sequence as found in Section 4.3, Tse [21] has developed conditions for local stochastic parameter identifiability in terms of the information matrix. The system setting and terminology are also parallel to that of Section 4.3

Definition 4.9 [21]. A parameter $\underline{\theta}_0 \epsilon \Omega \subset R^p$ is said to be locally identifiable if

(1) There exists an open set S_0 such that $\underline{\theta}_0$ is an interior point of S_0; and

(2) There exists a consistent, local estimation sequence, $\left\{ \hat{\underline{\theta}}_k (Y_k; \bar{S}_0) \right\}_{k=1}^{\infty}$ in $\bar{S}_0$ where $\bar{S}_0$ is the closure of S_0. The set S_0 is said to be the region of parameter identifiability.

Considering the concept of resolvability of Definition 4.4, the Definition 4.9 is equivalent to stating that $\underline{\theta}_0$ is locally identifiable if there exists a neighborhood about $\underline{\theta}_0$, denoted by S_0, such that $\underline{\theta}_0$ is resolvable from its neighboring elements $\underline{\theta} \epsilon S_0$.

Theorem 4.5 [21].

If for all k = 1, 2, ... there exists $\lambda^2 > 0$ such that

$$J_{k,k}(\underline{\theta}_0) = E_{\underline{\theta}_0}\left\{\left[\frac{\partial \log p(\underline{y}_k | Y_{k-1}; \underline{\theta}_0)}{\partial \underline{\theta}_0}\right]\left[\frac{\partial \log p(\underline{y}_k | Y_{k-1}; \underline{\theta}_0)}{\partial \underline{\theta}_0}\right]^T\right\}$$

$$> \lambda^2 \underline{I}_{p \times p}$$

where $E_{\underline{\theta}_0}$ represents the expectation operator with respect to the probability density function $p(Y_k;\ \underline{\theta}_0)$, then $\underline{\theta}_0$ is locally identifiable.

The function $J_{k,k}(\underline{\theta}_0)$ is a conditional information matrix. Vaguely, Theorem 4.5 implies that if there exists "positive information" about the unknown parameter in each new observation, then that parameter may be recovered asymptotically provided that the region of uncertainty for the unknown parameter is small.

Unfortunately, the condition of Theorem 4.5 is rather difficult to verify as it involves checking the positive definiteness properties of a countably infinite number of matrices. Further, it must be demonstrated that these matrices are uniformly bounded below by $\lambda^2\underline{I}$, $\lambda^2 > 0$. Thus, it is desirable to establish a weaker sufficient condition by considering an additive, and eventually total, information matrix

$$J_{m,n}(\underline{\theta}) = E_{\underline{\theta}}\left\{\left[\frac{\partial \log p_{m,n}(\underline{\theta})}{\partial\underline{\theta}}\right]\left[\frac{\partial \log p_{m,n}(\underline{\theta})}{\partial\underline{\theta}}\right]^T\right\} \tag{197}$$

where

$$p_{m,n}(\underline{\theta}) = p(\underline{y}_m,\ \underline{y}_{m+1},\ \ldots,\ \underline{y}_n | Y_{m-1};\ \underline{\theta}). \tag{198}$$

Noting that

$$E_{\underline{\theta}}\left\{\frac{\partial^2 \log p_{m,n}(\underline{\theta})}{\partial\underline{\theta}^2}\right\} = E_{\underline{\theta}}\left\{\left[\frac{\partial \log p_{m,n}(\underline{\theta})}{\partial\underline{\theta}}\right]\left[\frac{\partial \log p_{m,n}(\underline{\theta})}{\partial\underline{\theta}}\right]^T\right\} \tag{199}$$

and

$$E_{\underline{\theta}}\left\{\frac{\partial^2 \log p_{m,n}(\underline{\theta})}{\partial\underline{\theta}^2}\right\} = E_{\underline{\theta}}\left\{\sum_{i=m}^{n} \frac{\partial^2 \log p_{i,i}(\underline{\theta})}{\partial\underline{\theta}^2}\right\}$$

$$= \sum_{i=m}^{n} E_{\underline{\theta}} \left[\frac{\partial^2 \log p_{i,i}(\underline{\theta})}{\partial \underline{\theta}^2} \right]$$

$$= \sum_{i=m}^{n} J_{i,i}(\underline{\theta}) \tag{200}$$

we have

$$J_{m,n}(\underline{\theta}) = \sum_{i=m}^{n} J_{i,i}(\underline{\theta}) \tag{201}$$

and

$$J_{1,k}(\underline{\theta}_0) = \sum_{i=1}^{k} J_{i,i}(\underline{\theta}_0) \tag{202}$$

If the condition on $J_{k,k}(\underline{\theta}_0)$ of Theorem 4.5 is satisfied then

$$\lim_{k \to \infty} \frac{1}{k} J_{1,k}(\underline{\theta}_0) \triangleq J^*(\underline{\theta}_0) \geq \lambda^2 \underline{I}_{pxp};\ \lambda^2 > 0 \tag{203}$$

where $J^*(\underline{\theta}_0)$ is also known as the average Fisher's information matrix. While Eq. (203) does not necessarily imply that the condition on $J_{k,k}(\underline{\theta}_0)$ of Theorem 4.5 holds, this weaker condition is sufficient for local identifiability.

Theorem 4.6 [21]

Let $\underline{h}$ be any unit vector in R^p. If there exists $\lambda^2 > 0$ such that

$$\lim_{k \to \infty} \frac{1}{k} \underline{h}^T J_{1,k}(\underline{\theta}_0)\underline{h} \geq \lambda^2(\underline{h}) > 0,$$

then $\underline{\theta}_0$ is resolvable from $\underline{\theta} = \underline{\theta}_0 \pm \epsilon\underline{h}$; $0 < \epsilon \leq \lambda(\underline{h})/\tilde{c}(\underline{h})$ for some constant $\tilde{c}(\underline{h}) < \infty$. Note that both λ and $\tilde{c}$ may be dependent upon $\underline{h}$.

Definition 4.10 [21]. A subspace $\Omega \subset R^p$ is said to be locally identifiable if all elements $\underline{\theta} \epsilon \Omega$ are locally identifiable.

Theorem 4.7 [21]

A sufficient condition for a subset $\Omega \subset R^p$ to be identifiable is that

$$J^*(\underline{\theta}) \geq \lambda^2(\underline{\theta})\underline{I}_{pxp}; \qquad \lambda^2(\underline{\theta}) > 0 \text{ for all } \underline{\theta}\epsilon\Omega.$$

While the results presented above appear to be particularly applicable to the analysis of a given system rather than to a class of systems, Tse [21] has applied the results to a linear, discrete-time, autonomous system presented in the following example.

Example 4.3 [21]. Consider the system

$$\underline{x}_{k+1} = \underline{A}_k \underline{x}_k \tag{204}$$

$$\underline{y}_k = \underline{C}_k \underline{x}_k + \underline{v}_k \tag{205}$$

where $\underline{A}_k$, $\underline{C}_k$ are known matrices and $[\underline{v}_k]_{k=1}^{\infty}$ is a sequence of zero-mean, independent, Gaussian random vectors with covariance $\sigma_k^2\underline{I}$. The only unknown parameter is the initial state, $\underline{\theta} = \underline{x}_0 \epsilon R^p$. It can be shown that all required assumptions are satisfied [21] if the system is stable; i.e., $\|\underline{x}_k\| < c_1 < \infty$, k = 1, 2, ..., and if $\|\underline{x}_0\| < c_2 < \infty$.

The conditional information matrix is given by

$$J_{k,k}(\underline{\theta}) = \frac{1}{4\sigma_k^2} \underline{\varphi}_{k,0}^T \underline{C}_k^T \underline{C}_k \underline{\varphi}_{k,0} \tag{206}$$

where

$$\varphi_{k,j} = \underline{A}_k \cdot \underline{A}_{k-1}, \ldots, \underline{A}_j. \tag{207}$$

Therefore, the total information matrix is given by

$$J_{1,n}(\underline{\theta}) = \frac{1}{4} \sum_{k=1}^{n} \sigma_k^{-2} \underline{\varphi}_{k,0}^T \underline{C}_k^T \underline{\varphi}_{k,0} \quad \text{for all } \underline{\theta} \in R^p. \tag{208}$$

Suppose that the system is uniformly observable,

$$\sum_{j=k}^{k+n-1} \underline{\varphi}_{j,k}^T \underline{C}_j^T \underline{\varphi}_{j,k} \geq \beta \underline{I}, \quad \beta > 0; \qquad j = 1, 2, \ldots, \tag{209}$$

and $[\sigma_k]_{k=1}^{\infty}$ is bounded. Define

$$\sigma_{j+1}^{*} \overset{\Delta}{=} \max \left(\sigma_{jn}, \sigma_{jn+1}, \ldots, \sigma_{(j+1)n} \right). \tag{210}$$

Then, from Eqs. (208) through (210),

$$J_{1,jn}(\underline{\theta}) \geq \frac{\beta}{4} \sum_{i=1}^{j} \sigma_{in,0}^{*} \underline{\varphi}_{in,0} \qquad \text{for all } \underline{\theta} \in R^p. \tag{211}$$

Therefore, from Theorem 4.7, a sufficient condition for $\underline{x}_0$ to be identifiable is

$$\lim_{j \to \infty} \frac{1}{j} \sum_{i=1}^{j} \sigma_i^{*-2} \underline{\varphi}_{in,0}^T \underline{\varphi}_{in,0} \geq \lambda^2 \underline{I}; \qquad \lambda^2 > 0. \tag{212}$$

Note that, since the system is linear, the least square estimate has error equal to $J_{1,n}(\underline{\theta})$ exactly. Therefore, a less restrictive sufficient condition will require

$$\lim_{j \to \infty} \sum_{i=1}^{j} \sigma_i^{*-2} \underline{\varphi}_{in,0}^T \underline{\varphi}_{in,0} \to \infty. \tag{213}$$

In an attempt to more directly relate the above results to systems notation, consider three systems in which the parameters $\underline{\theta}$ are linear, nonlinear, and dynamic-nonlinear functions of the observations $\underline{y}$ in the presence of independent, Gaussian noise $\underline{v}$ with covariance matrix $\underline{Q}$. The three representations are, respectively:

$$\underline{y} = \underline{H}\,\underline{\theta} + \underline{v} \qquad \text{(Linear)} \qquad (214)$$

$$\underline{y} = \underline{H}\,\underline{z}\,(\underline{\theta}) + \underline{v} \qquad \text{(Nonlinear)} \qquad (215)$$

$$\underline{y}(t_i) = \underline{H}\,\underline{z}\,(t_i,\underline{\theta}) + \underline{v}(t_i) \qquad \text{(Dynamic-Nonlinear)} \qquad (216)$$

The form of the information matrix may then be determined for each case. Consider first the linear case of Eq. (214) which implies that $\underline{y} - \underline{H}\,\underline{\theta}$ is distributed as $\underline{v}$. Then:

$$\log p(\underline{y}|\underline{\theta}) = \text{Constant} + \tfrac{1}{2}[\underline{y} - \underline{H}\,\underline{\theta}]^T\underline{Q}^{-1}[\underline{y} - \underline{H}\,\underline{\theta}] \qquad (217)$$

$$\frac{\partial \log p(\underline{y}|\underline{\theta})}{\partial\underline{\theta}} = -[\underline{y} - \underline{H}\,\underline{\theta}]^T\underline{Q}^{-1}\underline{H} \qquad (218)$$

$$E\left\{\left[\frac{\partial \log p(\underline{y}|\underline{\theta})}{\partial\underline{\theta}}\right]\left[\frac{\partial \log p(\underline{y}|\underline{\theta})}{\partial\theta}\right]\Big|\underline{\theta}\right\}$$

$$= E\left\{\underline{H}^T\underline{Q}^{T-1}\,[\underline{y} - \underline{H}\,\underline{\theta}][\underline{y} - \underline{H}\,\underline{\theta}]^T\underline{Q}^{-1}\,\underline{H}\right\}$$

$$= \underline{H}^T\underline{Q}^{-1}\underline{H}. \qquad (219)$$

For the nonlinear case:

$$\log p(\underline{y}|\underline{\theta}) = \text{Constant} + \tfrac{1}{2}[\underline{y} - \underline{H}\,\underline{z}(\underline{\theta})]^T\underline{Q}^{-1}[\underline{y} - \underline{H}\,\underline{z}(\underline{\theta})] \qquad (220)$$

$$\frac{\partial \log p(\underline{y}|\underline{\theta})}{\partial\underline{\theta}} = -[\underline{y} - \underline{H}\,\underline{z}(\underline{\theta})]^T\underline{Q}^{-1}\underline{H}\,\frac{\partial\,\underline{z}(\underline{\theta})}{\partial\underline{\theta}} \qquad (221)$$

$$E\left\{\left[\frac{\partial \log p(\underline{y}|\underline{\theta})}{\partial\underline{\theta}}\right]\left[\frac{\partial \log p(\underline{y}|\underline{\theta})}{\partial\underline{\theta}}\right]\Big|\underline{\theta}\right\}$$

$$= \left[\frac{\partial\,\underline{z}(\underline{\theta})}{\partial\underline{\theta}}\right]^T \underline{H}^T\,\underline{Q}^{-1}\,\underline{H}\left[\frac{\partial\,\underline{z}(\underline{\theta})}{\partial\underline{\theta}}\right]. \qquad (222)$$

For the dynamic-nonlinear case:

$$\log p(\underline{y}|\underline{\theta}) = \text{Constant} + \frac{1}{2} \sum_{i=1}^{K} \left\{ [\underline{y}(t_i) - \underline{H}\, \underline{z}(t_i,\underline{\theta})]^T \right\}$$

$$\left\{ \underline{Q}^{-1} [\underline{y}(t_i) - \underline{H}\, \underline{z}\, (t_i,\underline{\theta})] \right\} \tag{223}$$

$$\frac{\partial \log p(\underline{y}|\underline{\theta})}{\partial \underline{\theta}} = - \sum_{i=1}^{K} [\underline{y}(t_i) - \underline{H}\, \underline{z}(t_i,\underline{\theta})]^T \underline{Q}^{-1} \underline{H} \frac{\partial \underline{z}(t_i,\underline{\theta})}{\partial \underline{\theta}} \tag{224}$$

$$E\left\{\left[\frac{\partial \log p(\underline{y}|\underline{\theta})}{\partial \underline{\theta}}\right]\left[\frac{\partial \log p(\underline{y}|\underline{\theta})}{\partial \underline{\theta}}\right] | \underline{\theta}\right\}$$
$$= \sum_{i=1}^{K} \left[\frac{\partial \underline{z}(t_i,\underline{\theta})}{\partial \underline{\theta}}\right]^T \underline{H}^T \underline{Q}^{-1} \underline{H} \left[\frac{\partial \underline{z}(t_i,\underline{\theta})}{\partial \underline{\theta}}\right] \tag{225}$$

It should be noted that these exact expressions appear in small variational parameter estimation methods as the gradients of the quadratic cost functionals. Likewise, a minor link may also be established to deterministic parameter identifiability by noting that the cost functional for least square identifiability may be taken in the stochastic case to be the negative logarithm of the likelihood equation. Then, the expected value of the second partial derivative of the cost functional with respect to the parameters is the information matrix discussed above.

4.5. Comments on Stochastic Parameter Identifiability

While the injectivity of the function $\underline{f}$ provides a unifying set of concepts and definitions for deterministic parameter identifiability, no corresponding unifying concept or definition has yet been determined for stochastic parameter identifiability. A loosely unifying concept for stochastic parameter identifiability appears to be the existence of a consistent estimate for the unknown parameters. However, the requirements for the existence of a consistent estimate is expressed mathematically in a number of different ways and thus no mathematically uniform concept now exists which is equivalent to the functional injectivity requirement for deterministic parameter identifiability.

As shown in the previous material, several tentative links between the forms found in deterministic and stochastic parameter identifiability have been established. However, no mathematically explicit, consistent relationships have been established to relate the two concepts together.

It should be noted, however, that deterministic parameter identifiability is a prerequisite for stochastic parameter identifiability. Indeed, if a system is not deterministically identifiable, then certainly no consistent estimate for the unknown parameters can exist. On the other hand, the assurance of deterministic parameter identifiability is not sufficient to insure stochastic parameter identifiability since the stochastic properties of any given system may supercede the deterministic properties.

It should be noted that the final two stochastic parameter identifiability concepts presented in Section 4, specifically, those predicated upon the properties of the conditional probability density functions of the system observations and those predicated upon the properties of the conditional and total information matrices, are oriented towards the analysis of a given specific system rather than of a total class of systems. The application of these concepts to some classes of systems of interest seems to be an opportune field for study.

5. ACKNOWLEDGMENT

This research was supported by the Engineering Research Institute at Iowa State University through funds provided by the NASA-Ames Research Center, Moffett Field, California.

PRECEDING PAGE BLANK NOT FILMED

6. BIBLIOGRAPHY

1. M. Aoki and P. C. Yue, "On Certain Convergence Questions in System Identification," SIAM Journal on Control, 8, 239-256, (1970).

2. G. A. Bekey, "System Identification - An Introduction and a Survey," Simulation, 9, 151-166, (1970).

3. R. Bellman and K. J. Åström, "On Structural Identifiability," Mathematical Biosciences, 7, 329-339, (1970).

4. R. W. Brockett, Finite Dimensional Linear Systems, John Wiley and Sons, New York, (1970).

5. C. A. Desoer, Notes for a Second Course on Linear Systems, Van Nostrand Reinhold Company, New York, 1970.

6. C. A. Desoer and K. K. Wong, "Small Signal Behavior of Nonlinear Lumped Networks," Proc. IEEE, 56:1, 14-22, (1968).

7. P. Eykhoff, System Identification: Parameter and State Estimation, John Wiley and Sons, New York, (1974).

8. M. F. Gardner and J. L. Barnes, Transients in Linear Systems, John Wiley, New York, (1956).

9. K. Glover and J. C. Willems, "Parameterization of Linear Dynamical Systems: Canonical Forms and Identifiability," IEEE Trans. on Automatic Control, AC-19, 640-645, (1974).

10. M. S. Grewal, "Modeling and Identification of Freeway Traffic Systems," Ph.D. dissertation, Department of Electrical Engineering, University of Southern California, Los Angeles, (1974).

11. M. S. Grewal, G. A. Bekey, and H. J. Payne, "Parameter Identifiability of Dynamical Systems," Proc. of the 1974 IEEE Conference on Decision and Control, Phoenix, November 20-22, 1974.

12. M. S. Grewal and K. Glover, "Identifiability of Linear and Nonlinear Dynamical Systems," IEEE Transactions on Automatic Control, AC-21, 833-837, (1976).

13. C. J. Herget, "Regions of Parameter Identifiability," Proc. of the 1974 IEEE Conference on Decision and Control, Phoenix, November 20-22, 1974.

14. B. L. Ho and R. E. Kalman, "Effective Construction of Linear State-Variable Models from Input/Output Data," Proc. of the Third Allerton Conference, Urbana-Champaign, Illinois, (1965).

15. K. Mårtensson, "Least-Square Identifiability of Dynamic Systems," University of Southern California, USCEE Report No. RB 7344, (1973).

16. R. Mehra and D. Lainiotis, eds., Advances in System Identification, Dekker, New York, (1975).

17. R. Norasimham, Analysis on Real and Complex Manifolds, Amsterdam, The Netherlands: North-Holland, (1968).

18. A. P. Sage and J. Melsa, System Identification, Academic Press, New York, (1971).

19. V. I. Smirnoff, Linear Algebra and Group Theory. New York: McGraw-Hill, 1961.

20. R. M. Staley and P. C. Yue, "On System Parameter Identifiability," Information Sciences, 2, 127-138, (1970).

21. E. Tse, "Information Matrix and Local Identifiability of Parameters," Proc. of the 1973 Joint Automatic Control Conference, Columbus, Ohio, 1973.

22. E. Tse and J. Anton, "On the Identifiability of Parameters," IEEE Trans. on Automatic Control, AC-17, 637-646, (1972).

23. E. Tse and H. Weinert, "Correction and Extension of 'On the Identifiability of Parameters'," IEEE Trans. on Automatic Control, AC-17, 687-688, (1973).

24. A. Wald, "Asymptotic Properties of the Maximum Likelihood Estimate of an Unknown Parameter of a Discrete Stochastic Process," Ann. Math. Statist., 19, 40-46, (1948).

25. A. Wald, "Note on the Consistency of the Maximum Likelihood Estimate," Ann. Math. Statist., 20, 595-601, (1949).

APPENDIX A. PARAMETER IDENTIFIABILITY RESULTS VERSUS SYSTEMS' CHARACTERISTICS.

Characteristic	Theorem 3.1	Theorem 3.2	Theorem 3.2a	Theorem 3.3	Theorem 3.4	Theorem 3.5	Theorem 3.7	Theorem 3.8	Theorem 3.9	Theorem 3.10	Theorem 3.11	Theorem 4.1	Corollary 4.1	Theorem 4.2	Theorem 4.3	Theorem 4.3a	Theorem 4.4	Theorem 4.5	Theorem 4.6
General, Not System Specific																	x	x	x
Deterministic	x	x	x	x	x	x	x	x	x	x	x								
Stochastic												x	x	x	x	x			
Linear	x	x	x	x	x	x	x	x			x	x	x	x	x	x			
Nonlinear									x	x	x								
Continuous					x	x	x	x	x	x	x								
Discrete	x	x	x	x							x	x	x	x	x	x			
SISO	x	x	x	x							x								
MIMO					x	x	x	x	x	x	x	x	x	x	x	x			
$\underline{x}_o$ Known	x	x	x		x	x	x	x	x	x	x	x	x		x	x			
$\underline{x}_o$ Unknown				x										x					
Unspecified Input		x	x	x			x	x	x	x	x	x	x	x	x	x			
Zero Input	x																		
Impulsive Input					x	x													
Local					x	x	x	x	x	x	x								
Global	x	x	x	x								x	x	x	x	x			

PRECEDING PAGE BLANK NOT FILMED

APPENDIX B. SELECTED PROOFS

The proofs selected for presentation in this Appendix are not totally comprehensive but are presented as samples of the methods and procedures involved. Generally, proofs which require the establishment of preliminary lemmas or proofs have been omitted. Other proofs which are similar in content and nature to those chosen have also been omitted.

Theorem 3.i [20]

From the defining Eqs. (35) through (37), direct computation yields

$$\left[x_n,\ x_{n+1},\ \ldots,\ x_{L-1}\right] = \left[a_n,\ a_{n-1},\ \ldots,\ a_1\right]$$

$$\times \begin{bmatrix} x_o & x_1 & \cdots & x_{L-n-1} \\ x_1 & x_2 & \cdots & x_{L-n} \\ \vdots & & & \\ x_{n-1} & x_n & \cdots & x_{L-2} \end{bmatrix} = \left[a_n,\ a_{n-1},\ \ldots,\ a_1\right]$$

$$\times \begin{bmatrix} \underline{h}^T \\ \underline{h}^T\underline{A} \\ \vdots \\ \underline{h}^T\underline{A}^{(n-1)} \end{bmatrix} \left[\underline{z}_o,\ \underline{A}\,\underline{z}_o,\ \ldots,\ \underline{A}^{(L-n-1)}\,\underline{z}_o\right]. \qquad \text{(B.1)}$$

The first matrix on the right is non-singular by Eq. (36). Therefore, $[a_n,\ a_{n-1},\ \ldots,\ a_1]$ is uniquely determined if and only if the second matrix, the controllability matrix for the pair $(\underline{A},\underline{z}_o)$, has rank n and the theorem follows immediately.

Theorem 3.2 [20]

Necessity.

(i) $\underline{H}_L$ is L X (n + r). $\underline{H}_L^T\underline{H}_L > 0$ implies $L \geq n + r$.

(ii) If $b_i \equiv 0$, then $\underline{B}_L = 0$, $\underline{x}_L = \underline{A}_L^{-1}\underline{B}_L\underline{u}_L = 0$

where Eq. (37) is restated as

$$\underline{A}_L\underline{x}_L = \underline{B}_L\underline{u}_L + \underline{E}_L\underline{z}_o \tag{B.2}$$

and

$$\underline{A}_L = \underline{I}_L + \sum_{j=1}^{n} a_j\underline{S}_L^j \tag{B.3}$$

and

$$\underline{B}_L = \sum_{j=1}^{r} b_j\underline{S}_L^j \ . \tag{B.4}$$

Consequently, any $\underline{A}_L$ satisfies Eq. (B.2).

(iii) If $u_i \equiv 0$ for $0 \leq i \leq L - n - r$, then by Eq. (35), $x_i \equiv 0$, also, for $0 \leq i \leq L - n - r$. Thus, the first $L - n - r + 1$ rows of $\underline{H}_L$ become zero and rank $\underline{H}_L < n + r$, and $\underline{H}_L^T H_L$ is not positive definite.

(iv) If A(z) and B(z) have a common divisor, then

$$\underline{A}_L = \underline{\tilde{A}}_L\underline{D}_L, \quad \underline{B}_L = \underline{\tilde{B}}_L\underline{D}_L \tag{B.5}$$

where

$$\underline{\tilde{A}}_L = \underline{I}_L - \sum_{j=1}^{n_1} \tilde{a}_j\underline{S}_L^j, \quad \underline{\tilde{B}}_L = \sum_{j=1}^{n_1} \tilde{b}_j\underline{S}_L^j, \; n_1 < n$$

and

$$\underline{D}_L = d_o \underline{I}_L + \sum_{j=1}^{n-n_1} d_j \underline{S}_L^j, \quad d_o \neq 0.$$

Substituting Eq. (B.5) into Eq. (B.2) and multiplying both sides by $\underline{D}_L^{-1}$, since $|\underline{D}_L| \neq 0$, we obtain

$$\underline{\tilde{A}}_L \underline{x}_L = \underline{\tilde{B}}_L \underline{u}_L;$$

then $\underline{\tilde{\theta}} = \left[\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_{n_1}, 0, \ldots, 0, \tilde{b}_1, \ldots, \tilde{b}_{n_1}, 0, \ldots, 0\right]^T$ would satisfy Eq. (B.2) and contradict the uniqueness assumption.

Sufficiency. Let $\underline{\tilde{\theta}} = \left[\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n, \tilde{b}_1, \tilde{b}_2 \ldots, \tilde{b}_r\right]^T$ be any vector such that the corresponding matrices $\underline{\tilde{A}}_L$, $\underline{\tilde{B}}_L$ satisfy Eq. (B.2). Then $\underline{\tilde{A}}_L \underline{x}_L = \underline{\tilde{B}}_L \underline{u}_L$, $\underline{A}_L \underline{x}_L = \underline{B}_L \underline{u}_L$. Therefore,

$$\begin{aligned} \underline{A}_L \underline{\tilde{B}}_L \underline{u}_L &= \underline{A}_L \underline{\tilde{A}}_L \underline{x}_L \\ &= \underline{\tilde{A}}_L \underline{A}_L \underline{x}_L = \underline{\tilde{A}}_L \underline{B}_L \underline{u}_L . \end{aligned} \tag{B.6}$$

Let

$$\underline{C}_L = \underline{A}_L \underline{\tilde{B}}_L - \underline{\tilde{A}}_L \underline{B}_L = \sum_{j=1}^{n+r} c_j \underline{S}_L^j . \tag{B.7}$$

Let

$$\underline{U}_{L,n+r} = \left[\underline{S}_L \underline{u}_L, \underline{S}_L^2 \underline{u}_L, \ldots, \underline{S}_L^{n+r} \underline{u}_L\right] \tag{B.8}$$

and

$$\underline{c} = \left[c_1, c_2, \ldots, c_{n+r}\right]^T. \tag{B.9}$$

Combining Eqs. (B.6) through (B.9), we obtain

$$\underline{U}_{L,n+r}\underline{c} = \underline{C}_L\underline{u}_L = 0. \tag{B.10}$$

By (i) and (iii), $\underline{U}_{L,n+r}$ has rank n + r. Hence $\underline{c} = 0$ and, by Eq. (B.7),

$$\underline{A}_L\underline{\widetilde{B}}_L = \underline{\widetilde{A}}_L\underline{B}_L.$$

This implies $B(z)/A(z) = \widetilde{B}(z)/\widetilde{A}(z)$ and, by (ii) and (iv), $\underline{a} = \underline{\widetilde{a}}$ and $\underline{b} = \widetilde{b}$. This completes the proof.

Theorem 3.4 [9] is an immediate consequence of the definitions given and is an application of the constant rank theorem for injective maps [17].

Theorem 3.6 [10] results immediately from the application of the variation of constants formula

$$\underline{y}(t,\underline{\theta}) = \underline{C}(\underline{\theta}) \int_0^t e^{\underline{A}(\underline{\theta})(t-\tau)}\, \underline{B}(\underline{\theta})\underline{u}(\tau)d\tau + \underline{D}(\underline{\theta})\underline{u}(t) \tag{B.11}$$

and Definition 3.8 of distinguishability.

Corollary 3.6 [10]

$$\underline{C}(\underline{\theta}_1)\underline{A}^{\ell}(\underline{\theta}_1)\underline{B}(\underline{\theta}_1) = \underline{C}(\underline{\theta}_2)\underline{A}^{\ell}(\underline{\theta}_2)\underline{B}(\underline{\theta}_2), \qquad \ell = 0, 1, 2, \ldots \tag{B.12}$$

$$\underline{D}(\underline{\theta}_1) = \underline{D}(\underline{\theta}_2) \tag{B.13}$$

Sufficiency. Equations (B.12) and (B.13) imply that the two systems corresponding to the pair of parameter values $(\underline{\theta}_1,\underline{\theta}_2)$ both have identical impulse responses and hence have identical transfer functions. By Definition 3.8 the pair of parameter values is indistinguishable.

Necessity. By Definition 3.8, the indistinguishability of $(\underline{\theta}_1,\underline{\theta}_2)$ implies for $\underline{x}_0 = 0$, for any given $\underline{u}(t)\epsilon\mathcal{U}$ and for all $0 \le t \le T$, that

$$\underline{y}(t,\underline{\theta}_1) = \underline{y}(t,\underline{\theta}_2)$$

which, by the variation of constants formula, becomes

$$\underline{C}(\underline{\theta}_1) \int_0^t e^{\underline{A}(\underline{\theta}_1)(t-\tau)}\underline{B}(\underline{\theta}_1)\underline{u}(\tau)d\tau + \underline{D}(\underline{\theta}_1)\underline{u}(t)$$
$$\equiv \underline{C}(\underline{\theta}_2) \int_0^t e^{\underline{A}(\underline{\theta}_2)(t-\tau)}\, \underline{B}(\underline{\theta}_2)\underline{u}(\tau)d\tau + \underline{D}(\underline{\theta}_2)\, \underline{u}(t). \quad \text{(B.14)}$$

Since Eq. (B.14) holds for all $0 \le t \le T$, and in particular for $t = 0$, then

$$\underline{D}(\underline{\theta}_1)\underline{u}(t) = \underline{D}(\underline{\theta}_2)\underline{u}(t)$$

for all $\underline{u}(t)\epsilon\mathcal{U}$ and for all $0 \le t \le T$. Clearly,

$$\underline{D}(\underline{\theta}_1) \equiv \underline{D}(\underline{\theta}_2) \quad \text{(B.15)}$$

and

$$\underline{C}(\underline{\theta}_1) \int_0^t e^{\underline{A}(\underline{\theta}_1)(t-\tau)}\, \underline{B}(\underline{\theta}_1)\underline{u}(\tau)d\tau$$
$$= \underline{C}(\underline{\theta}_2) \int_0^t e^{\underline{A}(\underline{\theta}_2)(t-\tau)}\, \underline{B}(\underline{\theta}_2)\underline{u}(\tau)d\tau. \quad \text{(B.16)}$$

Combining like terms and moving $\underline{C}(\underline{\theta}_1)$ and $\underline{C}(\underline{\theta}_2)$ under the integral, since they are independent of τ, yields

$$\int_0^t \left[\underline{C}(\underline{\theta}_1)e^{\underline{A}(\underline{\theta}_1)(t-\tau)}\underline{B}(\underline{\theta}_1) - \underline{C}(\underline{\theta}_2)e^{\underline{A}(\underline{\theta}_2)(t-\tau)}\underline{B}(\underline{\theta}_2)\right] \underline{u}(\tau)d\tau \equiv 0.$$

Since the functions in the equation above within the brackets and $\underline{u}(\tau)$ are continuous, by the standard theorem of the calculus of variations

$$\underline{C}(\underline{\theta}_1)e^{\underline{A}(\underline{\theta}_1)(t-\tau)}\underline{B}(\underline{\theta}_1) - \underline{C}(\underline{\theta}_2)e^{\underline{A}(\underline{\theta}_2)(t-\tau)}\underline{B}(\underline{\theta}_2) = 0.$$

By repeated differentiation with respect to t and evaluation at $t = \tau$

$$\begin{aligned}
\underline{C}(\underline{\theta}_1)\underline{B}(\underline{\theta}_1) &\equiv \underline{C}(\underline{\theta}_2)\underline{B}(\underline{\theta}_2) \\
\underline{C}(\underline{\theta}_1)\underline{A}(\underline{\theta}_1)\underline{B}(\underline{\theta}_1) &\equiv \underline{C}(\underline{\theta}_2)\underline{A}(\underline{\theta}_2)\underline{B}(\underline{\theta}_2) \\
&\vdots \\
\underline{C}(\underline{\theta}_1)\underline{A}^{\ell}(\underline{\theta}_1)\underline{B}(\underline{\theta}_1) &\equiv \underline{C}(\underline{\theta}_2)\underline{A}^{\ell}(\underline{\theta}_2)\underline{B}(\underline{\theta}_2), \qquad \ell = 0, 1, 2, \ldots
\end{aligned} \tag{B.17}$$

as required.

Theorem 3.7 [10] follows immediately from Corollary 3.6 and Definition 3.9 for (local) parameter identifiability.

Theorem 3.8 [10] is an immediate consequence of the given definitions, Theorem 3.7 and the constant rank theorem for injective mappings [17].

Theorem 3.9 [10]

It has been shown [6] that the solution to Eqs. (97) and (98) can be written as

$$\begin{aligned}
\|\delta\underline{y}(\cdot,\underline{\theta}_0,\delta\underline{\theta}\| &= \|\delta\underline{y}_0(\cdot,\underline{\theta}_0,\delta\underline{\theta}) + \underline{\chi}(\cdot,\delta\underline{\theta})\| \\
&\geq \|\delta\underline{y}_0(\cdot,\underline{\theta}_0,\delta\underline{\theta})\| - \|\underline{\chi}(\cdot,\delta\underline{\theta})\|
\end{aligned} \tag{B.18}$$

where $\underline{\chi}(\cdot,\delta\underline{\theta})$ are terms of $O(\|\delta(\underline{\theta})\|)$ and $\lim_{\delta\to 0} O(\delta)/\delta = 0$.

Defining the $L_2(t_0,T)$ norms of Eq. (B.18) yields

$$\left[\int_{t_0}^{T} \|\delta \underline{y}(\tau,\underline{\theta}_0,\delta\underline{\theta}\|^2 d\tau\right]^{1/2} \geq \left[\int_{t_0}^{T} \|\delta \underline{y}_0(\tau,\underline{\theta}_0,\delta\underline{\theta}\|^2 d\tau\right]^{1/2}$$

$$- \left[\int_{t_0}^{T} \|\underline{h}(\tau,\delta\underline{\theta})\|^2 d\tau\right]^{1/2}. \qquad (B.19)$$

Divide both sides of Eq. (B.19) by $\|\delta\underline{\theta}\|$. Assume that $\|\delta\underline{\theta}\|$ lies between $0 < \|\delta\underline{\theta}\| < \delta$. Then

$$\left[\int_{t_0}^{T} \left(\frac{\|\delta \underline{y}(\tau,\underline{\theta}_0,\delta\underline{\theta}\|}{\|\delta\underline{\theta}\|}\right)^2 d\tau\right]^{1/2} \geq \left[\int_{t_0}^{T} \left(\frac{\|\delta \underline{y}_0(\tau,\underline{\theta}_0,\delta\underline{\theta})\|}{\|\delta\underline{\theta}\|}\right)^2 d\tau\right]^{1/2}$$

$$- \left[\int_{t_0}^{T} \left(\frac{\|\underline{h}(\tau,\delta\underline{\theta})\|}{\|\delta\underline{\theta}\|}\right)^2 d\tau\right]^{1/2}. \qquad (B.20)$$

Now assume

$$\int_{t_0}^{T} \left[\underline{N}^T(\tau,\underline{\theta}_0)\underline{N}(\tau,\underline{\theta}_0)\right] d\tau > 0,$$

and

$$\lambda_{min}\left\{\int_{t_0}^{T} \left[\underline{N}^T(\tau,\underline{\theta}_0)\underline{N}(\tau,\underline{\theta}_0)\right] d\tau\right\} = \epsilon^2 > 0$$

where λ_{min} is the minimum eigenvalue.

The first term on the right-hand side of Eq. (B.20) can be written as

$$\left[\int_{t_0}^{T} \frac{\delta\underline{\theta}}{\|\delta\underline{\theta}\|} \underline{N}^T(\tau,\underline{\theta}_0)\underline{N}(\tau,\underline{\theta}_0) \frac{\delta\underline{\theta}^T}{\|\delta\underline{\theta}\|} d\tau\right]^{1/2} \geq \epsilon > 0,$$

for all δ. (B.21)

As $\delta \to 0$, the second term on the right-hand side of Eq. (B.20) tends to zero, and the total right-hand side is greater than zero for all δ sufficiently small, say $\delta \le \bar{\delta}$. Therefore, the left-hand side of Eq. (B.20) is

$$\left[\int_{t_0}^{T} \left(\frac{\|\delta \underline{y}(\tau,\underline{\theta}_0,\delta\underline{\theta}\|}{\|\delta\underline{\theta}\|}\right)^2 d\tau\right]^{1/2} > 0; \tag{B.22}$$

hence, $\|\delta\underline{y}(\cdot,\underline{\theta}_0,\delta\underline{\theta})\|_2 > 0$. Hence, for all $\underline{\theta} \neq \underline{\theta}_0$, $\underline{\theta}\epsilon S(\underline{\theta}_0,\bar{\delta})$, $\underline{y}(t,\underline{\theta}) \neq \underline{y}(t,\underline{\theta}_0)$ for some $t\epsilon[t_0,T]$, which implies that the parameters, $\underline{\theta}\epsilon\Omega$, of the nonlinear system are locally identifiable at $\underline{\theta}_0$.

Wald demonstrated the consistency of the maximum likelihood estimate by first establishing three lemmas which are presented below without proof (see Ref. [24]).

<u>Lemma B.1.</u> If $\underline{\theta} \neq \underline{\theta}_0$, then

$$E \log p(y;\underline{\theta}) < E \log p(y;\underline{\theta}_0) \tag{B.23}$$

<u>Lemma B.2.</u>

$$\lim_{\rho\to 0} E \log p(y;\underline{\theta},\rho) = E \log p(y;\underline{\theta}_0) \tag{B.24}$$

<u>Lemma B.3.</u>

$$\lim_{r\to\infty} E \log \psi(y;r) = -\infty \tag{B.25}$$

With the above lemmas and the law of large numbers, the following theorems, leading to the consistency of the maximum likelihood estimate, can be proven.

Theorem B.4

Let $W \subset \Omega$ be a closed subset of Ω. If $\underline{\theta}_0$ does not belong to W, then

$$\Pr\left\{ \lim_{k\to\infty} \frac{\sup_{\underline{\theta}\in W} p(y_1,y_2 \ldots, y_k;\underline{\theta})}{p(y_1,y_2, \ldots, y_k;\underline{\theta}_0)} = 0 \right\} = 1. \tag{B.26}$$

Proof: By Lemma B.3, we can choose $r_o > 0$ such that

$$E \log \psi(y,r_o) < E \log p(y;\underline{\theta}). \tag{B.27}$$

Let W_1 be the subset of W such that

$$W_1 = \left\{ \underline{\theta}: \|\theta\| \le r_o, \ \underline{\theta}\in W \right\}.$$

For each $\underline{\theta}\in W_1$, we can choose a $\rho_{\underline{\theta}} > 0$ such that

$$E \log p(y;\underline{\theta},\rho_{\underline{\theta}}) < E \log p(y;\underline{\theta}_0). \tag{B.28}$$

The existence of $\rho_{\underline{\theta}}$ is guaranteed by the law of large numbers and Lemma B.1. The set W_i is closed and bounded and, hence, is compact. Thus, there exists a finite number of points $\underline{\theta}_1, \ldots, \underline{\theta}_j$ in W_1 such that the union of the spheres with center $\underline{\theta}_i$ and radius $\rho_{\underline{\theta}_i}$, $i = 1, \ldots, j$, $\bigcup_{i=1}^{j} S\left(\underline{\theta}_i,\rho_{\underline{\theta}_i}\right)$ covers W_1.

It is seen that

$$0 \le \sup_{\underline{\theta}\in W} p(y_1, y_2, \ldots, y_k;\underline{\theta})$$

$$\le \sum_{i=1}^{j} p\left(y_1;\underline{\theta}_i,\rho_{\underline{\theta}_i}\right) \ldots p\left(y_k;\underline{\theta}_i,\rho_{\underline{\theta}_i}\right)$$

$$+ \psi(y_1,r_o) \ldots \psi(z_k,r_o).$$

It is then required to show that

$$\Pr\left\{\lim_{k\to\infty} \frac{p(y_1;\underline{\theta}_i,\rho_{\theta i}) \cdots p(y_k;\underline{\theta}_i,\rho_{\theta i})}{p(y_1;\underline{\theta}_0) \cdots p(y_k;\underline{\theta}_0)} = 0\right\} = 1, \qquad i = 1, \ldots, j; \tag{B.29}$$

and

$$\Pr\left\{\lim_{k\to\infty} \frac{\psi(y_1,r_o) \cdots \psi(y_k,r_o)}{p(y_1;\underline{\theta}_0) \cdots p(y_k;\underline{\theta}_0)} = 0\right\} = 1 \tag{B.30}$$

which is equivalent to showing that

$$\Pr\left\{\lim_{k\to\infty} \sum_{m=1}^{k} \left[\log p\left(y_m;\underline{\theta}_i,\rho_{\underline{\theta}_i}\right) - \log p\left(y_m;\underline{\theta}_0\right)\right] = -\infty\right\} = 1, \qquad i = 1, \ldots, j; \tag{B.31}$$

and

$$\Pr\left\{\lim_{k\to\infty} \sum_{m=1}^{k} \left[\log \psi(y_m;r_o) - \log p(y_m;\underline{\theta}_0)\right] = -\infty\right\} = 1. \tag{B.32}$$

But Eqs. (B.31) and (B.32) follow immediately from Eqs. (B.27) and (B.28) and from the (strong) law of large numbers.

<u>Theorem B.5</u>

Let $\hat{\underline{\theta}}_k(y_1, \ldots, y_k)$ be a function of the observations such that

$$\frac{p(y_1, \ldots, y_k;\hat{\underline{\theta}}_k)}{p(y_1, \ldots, y_k;\underline{\theta}_0)} \geq c > 0 \text{ for all } k \text{ and for all } y_1, \ldots, y_k. \tag{B.33}$$

Then

$$\Pr\left\{\lim_{k\to\infty} \hat{\underline{\theta}}_k = \underline{\theta}_0\right\} = 1.$$

Proof: Let Θ denote the set of limit points of $\{\hat{\underline{\theta}}_k\}_{k=1}^{\infty}$. Then it suffices to show that for any $\epsilon > 0$,

$$\sup \left\{ \|\hat{\underline{\theta}} - \underline{\theta}_0\| : \hat{\underline{\theta}} \in \Theta \right\} \leq \epsilon \text{ with probability one.} \tag{B.34}$$

Suppose that there exists a $\hat{\underline{\theta}} \in \Theta$ such that $\|\hat{\underline{\theta}} - \underline{\theta}_0\| > \epsilon$, then

$$\sup_{\|\underline{\theta} - \underline{\theta}_0\| \geq \epsilon} p(y_1, \ldots, y_k; \underline{\theta}) \geq p(y_1, \ldots, y_k; \hat{\underline{\theta}}_k),$$

for infinitely many k. But this implies

$$\frac{\sup_{\|\underline{\theta} - \underline{\theta}_0\| \geq \epsilon} p(y_1, \ldots, y_k; \underline{\theta})}{p(y_1, \ldots, y_k; \underline{\theta}_0)} \geq c > 0 \tag{B.35}$$

for infinitely many k by Eq. (B.33). By Theorem B.4, Eq. (B.35) is an event with probability zero; thus, Eq. (B.34) holds with probability one.

Recall that the maximum likelihood estimate $\hat{\underline{\theta}}_k$ is obtained by

$$p(y_1, \ldots, y_k; \hat{\underline{\theta}}_k) = \max_{\underline{\theta} \in \Omega} p(y_1, \ldots, y_k; \underline{\theta}), \; k = 1, 2, \ldots \tag{B.36}$$

If $\hat{\underline{\theta}}_k$ exists, then

$$\frac{p(y_1, \ldots, y_k; \hat{\underline{\theta}}_k)}{p(y_1, \ldots, y_k; \underline{\theta}_0)} \geq 1, \text{ for } y_1, \ldots, y_k, \quad k = 1, 2, \ldots$$

Clearly, by Theorem B.5, the maximum likelihood estimate is consistent.

<u>Proposition 1 (page 71)</u>. In Eq. (162), the vector $\underline{E}_L \tilde{\underline{z}}_{0L}$ has only a finite number, n, of nonzero elements. As $L \to \infty$, it contributes nothing to $J_L(\underline{\theta})/L$ in the limit. Therefore, $\tilde{\underline{z}}_{0L}$ can be dropped from Eq. (162)

without loss of generality. Then

$$\frac{1}{L} J_L(\underline{\theta}) = \frac{1}{L} \|\underline{A}_L \underline{\eta}_L + \underline{A}_L \underline{x}_L - \underline{B}_L \underline{u}_L\|^2_{\left(\underline{A}_L \underline{A}_L^T\right)^{-1}} \tag{B.37}$$

$$= \frac{1}{L} \|\underline{A}_L \underline{x}_L - \underline{B}_L \underline{u}_L\|_{\left(\underline{A}_L \underline{A}_L^T\right)^{-1}}$$

$$+ \frac{2}{L} \underline{\eta}_L^T \left[\underline{x}_L - \underline{A}_L^{-1} \underline{B}_L \underline{u}_L\right] + \frac{1}{L} \|\underline{\eta}_L\|^2. \tag{B.38}$$

The first term is deterministic. In the second term, both $\underline{x}_L$ and $\underline{A}_L^{-1} \underline{B}_L \underline{u}_L$ represent output sequences of some stable system with a bounded input sequence. Thus, $\underline{x}_L - \underline{A}_L^{-1} \underline{B}_L \underline{u}_L$ is uniformly bounded. The limit in the Proposition statement $\left[\lim_{L \to \infty} \frac{1}{L} E\, J_L(\underline{\theta})\right]$ exists, and

$$\frac{1}{L} \underline{\eta}_L^T \left[\underline{x}_L - \underline{A}_L^{-1} \underline{B}_L \underline{u}_L\right] = \frac{1}{L} \sum_{i=0}^{L-1} \alpha_i \eta_i, \tag{B.39}$$

where $\alpha_i \le \alpha < \infty$ for all i. Since $\{\eta_i\}$ are independent random variables with

$$E\eta_i = 0, \; E\eta_i^2 = \sigma^2 < \infty$$

and

$$\sum_i \frac{\alpha_i^2}{i^2} \le \alpha^2 \sum_i \frac{1}{i^2} < \infty,$$

the strong law of large numbers applies. Hence, with probability one,

$$\frac{1}{L} \sum_i \alpha_i \eta_i \to 0, \tag{B.40}$$

$$\frac{1}{L} \sum_i \eta_i^2 \to \sigma^2, \tag{B.41}$$

and substituting Eqs. (B.39) through (B.41) into Eq. (B.38), we have, with probability one,

$$\frac{1}{L} J_L(\underline{\theta}) \to J(\underline{\theta}).$$

Theorem 4.1 [1]

Sufficiency. Define the matrix $\underline{C}_L$ and the associated vector $\underline{c}$ by

$$\begin{aligned}\underline{C}_L &= \underline{A}_L\underline{B}_{L,0} - \underline{A}_{L,0}\underline{B}_L \\ &= \sum_{i=1}^{2n} c_i \underline{S}_L^i,\end{aligned} \tag{B.42}$$

and

$$\underline{c} = \left[c_1, c_2, \ldots, c_{2n}\right]^T. \tag{B.43}$$

Then,

$$0 = \lim_{L\to\infty} \|(\underline{A}_L\underline{B}_{L,0} - \underline{A}_{L,0}\underline{B}_L)\underline{u}_L\|^2$$

if and only if

$$\lim_{L\to\infty} \|\underline{C}_L\underline{u}_L\|^2 = \underline{c}^T\left[\lim_{L\to\infty} \frac{1}{L} \underline{U}^T_{L,2n}\underline{U}_{L,2n}\underline{c}\right] = 0. \tag{B.44}$$

The limit condition of the Theorem statement; i.e.

$$\lim_{L\to\infty} \frac{1}{L} \underline{U}^T_{L,2n}\underline{U}_{L,2n} > 0, \tag{B.45}$$

implies that $\underline{c} = 0$, or, equivalently,

$$C(z) = A(z)B_0(z) - A_0(z)B(z) = 0,$$

$$B(z)/A(z) = B_0(z)/A_0(z),$$

where

$$A(z) = 1 + \sum_{j=1}^{n} a_i z^i, \; B(z) = \sum_{j=1}^{n} b_i z^i.$$

Hence, by controlability, $\underline{b} = \underline{b}_0$, $\underline{a} = \underline{a}_0$, $\underline{\theta} = \underline{\theta}_0$, and $\Theta_s \cap \Theta_0$ is a singleton.

Necessity. It suffices to show that if the limit condition of the Theorem statement, Eq. (B.45), is not satisfied, then there exists $\underline{\theta} = (\underline{a}_0 + \delta\underline{a}, \underline{b}_0 + \delta\underline{b})$ such that $\delta\underline{a} \neq 0$, $\delta\underline{b} \neq 0$, the condition of Proposition 3 is satisfied, and $\underline{\theta} \in \Theta_0 \cap \Theta_s$.

Note that the vector $\underline{c}$ as defined in Eqs. (B.42) and (B.43) can be re-expressed as

$$\underline{c} = \underline{T}_{bo}\underline{a} + \underline{T}_{ao}\underline{b} + \underline{E}_{2n}\underline{b}_o, \tag{B.46}$$

where

$$\underline{T}_{ao} = -\underline{A}_{2n,0}\,\underline{E}_{2n}, \quad (2n \times n);$$

$$\underline{T}_{bo} = \underline{B}_{2n,0}\,\underline{E}_{2n}, \quad (2n \times n);$$

$$\underline{E}_{2n} = \begin{bmatrix} \underline{I}_n \\ \hdashline \underline{0}_{n,n} \end{bmatrix}, \quad (2n \times n).$$

Let $V = \lim\limits_{L\to\infty} \frac{1}{L}\, \underline{U}^T_{L,2n}\, \underline{U}_{L,2n}$. If the matrix V is not positive definite, then there exists a nontrivial solution to the equation

$$0 = V\left(\underline{T}_{bo}, \underline{T}_{ao}\right)\begin{bmatrix} \delta\underline{a} \\ \delta\underline{b} \end{bmatrix}. \tag{B.47}$$

It is immediate from the definition of $\underline{C}_L$ that $\underline{C}_{L,0} = 0$ and thus

$$0 = \underline{T}_{bo}\underline{a}_o + \underline{T}_{ao}\underline{b}_o + \underline{E}_{2n}\underline{b}_o. \tag{B.48}$$

Therefore, letting $\underline{a} = \underline{a}_0 + \alpha \cdot \delta\underline{a}$, $\underline{b} = \underline{b}_0 + \alpha \cdot \delta\underline{b}$ for any scalar α, we obtain by combining Eqs. (B.46) through (B.48) that $\underline{c}^T V \underline{c} = 0$ which by Eq. (B.44) implies that the condition of Proposition 3 is satisfied and $\underline{\theta} \in \Theta_0$. It only remains to show that $\underline{\theta} \in \Theta_s$.

Suppose $[\lambda_i(A)]$, $1 \leq i \leq n$; are the roots of $A(z)$. Then, $\lambda_i(A_o)$ are exterior points of the unit disc $D = \{z: |z| \leq 1\}$ on the complex plane by stability. Since the roots of $A(z)$ are continuous in $\underline{a}$ at $\underline{a}_o$ in the sense that there exists a neighborhood $\Theta_{\underline{a}}$ where $\Theta_{\underline{a}} = \{\underline{a}: \|\underline{a} - \underline{a}_o\| < \epsilon\}$ such that $\lambda_i(A) \in D$ for all $\underline{a} \in \Theta_{\underline{a}}$, clearly $\underline{a} = \underline{a}_o + \alpha \cdot \delta\underline{a}$ is stable if $\alpha < \epsilon / \|\delta\underline{a}\|$. Furthermore, $\underline{\theta}_0$ is an exterior point of Θ_s. Thus, α can be chosen to have

$$\underline{\theta} = (\underline{a}_o + \alpha \cdot \delta\underline{a},\ \underline{b}_o + \alpha \cdot \delta\underline{b}) \in \Theta_s. \tag{B.49}$$