SIAM REVIEW
Vol. 24, No. 1, January 1982

0036-1445/82/2401-0002 $01.00/0

REVIEW AND UNIFICATION OF LINEAR IDENTIFIABILITY CONCEPTS*

V. V. NGUYEN† AND E. F. WOOD†

Abstract. To capture a unified perspective of the identification problem, we present a concise review of all contemporary concepts of identifiability. These concepts are axiomatically defined for a specific class of linear, deterministic and stochastic systems. From a derived set of criteria we demonstrate a basic correspondence between the principal concepts in discrete time. Such an equivalence statement constitutes the motif for the future development of a global identifiability concept applicable to nonlinear problems.

Introduction. The field of system identification is one of the most active and diverse branches of system theory. A multitude of papers dealing with various application aspects of the problem, appearing at an increasing rate, has complicated the current literature. Several attempts were conducted to survey and classify these computational works with the hope that they may be put into a systematic framework, and consequently a realizable unified theory. Two notable surveys are a paper by Åström and Eykhoff (1971) and the special issue on the subject in IEEE Transactions on Automatic Control (1974).

Earlier survey papers of related interest are those of Eykoff, van der Grinten, Kwakernaak and Veltman (1966), Cuenod and Sage (1968), Eykhoff (1968) and Balakrishnan and Peterka (1969).

More or less apart from the "engineering" works, the econometricians and statisticians have been working on the identification problem of economic models since the early 1940s. The basic results for linear simultaneous equation systems under linear parameter constraints were obtained in the 1950s, e.g., Koopmans and Reiersøl (1950). Extensions to nonlinear systems and nonlinear constraints were made by Fisher (1959), (1961), (1963), and (1966).

The word "system identification" has different connotations in the control (both system and time-series approaches) and econometric literatures. According to Mehra (1974a), in control theory the word generally denotes the complete three step iterative process of model specification, parameter estimation and model verification. In econometric literature, the word refers mainly to identifiability questions which have to be settled before attempting parameter estimation.

However, the models used in both areas are with few exceptions *parametric* models, i.e., input/output models which can be imbedded into finite dimensional state models. It is this that allows us to formulate a general concept of identifiability *using various equivalent definitions.*

Even though some of the definitions were introduced first for application purposes, they often possess considerable theoretical connotation and have withstood the test of time in spite of their differences. As these definitions are generalized to deal with more complex models than those they were originally designed for, especially those of higher dimensions, inconsistency and ambiguities began to arise (DiStefano and Cobelli (1980)). Recently, efforts have been made to construct similar definitions between stochastic and deterministic models, linear and nonlinear problems; see, for example, Grewal and Glover (1976), Ljung (1976) and DiStefano and Cobelli (1980) among others. However, up to now no formal proof has been presented on the

* Received by the editors August 25, 1980, and in final revised form March 25, 1981. This work was supported in part by the Wallace Fellowship of Princeton University and by the National Science Foundation under grant ENG-77-11841.

† Department of Civil Engineering, Princeton University, Princeton, New Jersey 08544.

interchangability of the current definitions of identifiability under certain assumptions. Such an equivalence statement would greatly consolidate the state-of-the-art on a unified theory of system identification and constitutes the principal motivation for this paper.

In multi-input/multi-output systems, there is not always a clear cut difference between the parametrization and the identification problems. As a consequence, a statement on identifiability is only precise if it is supported by a tractable canonical representation. This dilemma remains to be controversial in any extensive disussion of the matter. In this paper we attempt to review the identifiability concepts from an axiomatic point of view, and will not discuss their computational applications. After presenting the basic definitions in § 2, we proceed to prove an equivalence statement on these definitions for a chosen class of linear, time invariant, deterministic and stochastic systems in discrete time.

1. Exposition of the current identifiability concepts. Before beginning the exposition of identifiability concepts, we state two basic definitions. Extensive mathematical treatment relating to the implication of these definitions was expounded by Lee and Markus (1967).

DEFINITION 1. A linear system is said to be *controllable* at t_0 if it is possible to find some input function (or sequence in the discrete case), defined over some time interval, which will transfer the initial state $x(t_0)$ to the origin at some finite time. If this is true for all t_0 and all initial states $x(t_0)$, the system is completely controllable.

DEFINITION 2. A linear system is said to be *observable* at t_0 if $x(t_0)$ can be determined from the output function (or output sequence) during some interval $[t_0, t_1]$, where t_1 is some finite time point. If this is true for all t_0 and $x(t_0)$, the system is said to be *completely observable*.

Let us now consider the discrete state vector model of the form

$$x(t+1) = Ax(t) + Bu(t) + w(t), \tag{1.0.1}$$

$$z(t) = Cx(t) + v(t), \tag{1.0.2}$$

having the dimensions $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{r\times n}$, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^p$, $z \in \mathbb{R}^r$, $v \in \mathbb{R}^r$, $w \in \mathbb{R}^n$, and the covariance matrices

$$E[(w(t)w'(s)] = Q\delta(t, s), \qquad E[(v(t)v'(s)] = R\delta(t, s),$$

with w and v uncorrelated "white noise" Gaussian processes; $E[\cdot]$ and $\delta(t, s)$ are the expectation operator and the Kronecker delta, respectively. We assume that $r \leqq n$. When Definitions 1 and 2 are applicable to (1.0.1–2), we call (A,B) and (A,C) the controllable and the observable pairs, respectively.

From the works of Åström and Eykhoff (1971), Kalman et al. (1969), Mayne (1972), Popov (1972) among others, we see that the following conditions must be imposed on the model (1.0.1–2) to make its unknown parameters identifiable (Mehra (1974a)).

Conditions (i). (A,B) controllable and (A,C) observable. These conditions are generalizations of the no pole-zero cancellation conditions used in time-series analysis.

Condition (ii). All the elements of the process noise covariance matrix Q are specified independently from A. If $r < n$, only the proper structural representation of (1.01-2) given by a steady-state Kalman filter is amenable to quantitative analysis. The representation has the form

$$\hat{x}(t+1) = A\hat{x}(t) + Bu(t) + AK\nu(t), \tag{1.0.3}$$

(1.0.4) $$z(t) = C\hat{x}(t) + \nu(t),$$

where $E[\nu(t)\nu'(s)] = b\delta(t, s)$. The white noise process $\nu(t)$ represents one-step ahead prediction errors and is also known as the "innovation" process.

It is required that (A, K) be controllable.

Condition (iii). A and C possess certain algebraic structures or can be reduced to these structures by invertible transformations. Given these structures, the parameters can be uniquely specified from inputs and outputs. These structures are often referred to as canonical forms.

For discussion of these forms, the reader may see, for example, Popov (1972), Mayne (1972) or Rissanen (1974).

Condition (iv). The support of the spectral distribution function of the input $u(t)$, defined as

(1.0.5) $$\text{Support } S_{uu}(\omega) = \{\omega \mid -\pi \leqq \omega \leqq \pi, \forall \varepsilon > 0, (S_{uu}(\omega+\varepsilon) - S_{uu}(\omega-\varepsilon)) > 0\},$$

contains more than $k = [\mathrm{NP}/2p]$ points, where NP is the number of unknown parameters in A, B and C and $[\cdot]$ denotes the integer part of $(\cdot)$. This condition is derived by Mehra (1974b) and can be expressed in terms of the autocorrelation function of $u(t)$. If the input is sinusoidal, then it must contain more than k frequencies for the system to be identifiable. Such inputs have been called "persistently exciting" inputs (Åström and Eykhoff (1971)).

This is a necessary condition when $w(t) \equiv 0$ and $u(t)$ is assumed scalar. For sufficiency and for multi-input systems, further conditions are required.

Condition (v). A is a stable matrix (or the magnitudes of all the roots of $(z^{-1}I - A)$ are greater than 1, where z refers here to the backward shift operator).

Other aspects of identifiability have also been mentioned in Balakrishnan (1969). These ideas will be explored further in the following sections.

We now proceed to present a sequence of identifiability concepts and some typical examples. *These concepts are built upon the above stated conditions of identifiability, and therefore we will assume the five conditions implicitly in the definitions to be formulated.*

The sequence of identifiability concepts is constructed mainly in a chronological order except the Fisher information matrix concept. The first and the last concepts in the sequence are of some historical interest.

More technical terminologies can be found in a variety of texts such as Aoki (1967), Brockett (1970), Jaswinski (1970), Gelb (1974), among others. Throughout this paper we adopt the notation y for true measurements (scalar or vector) and z for noise-corrupted measurements (scalar or vector).

1.1. Initial condition criteria of identifiability. In dynamical systems, the question of identification has concerned scientists for at least three decades. Besides the papers in statistics (Barankin (1960)), econometrics (Fisher (1966)) and time series analysis (Whittle (1952)) as mentioned previously, engineers have recorded their efforts of practical identification in scattered technical reports. These ideas finally surfaced with a mathematical setup in a few texts around 1964, cf. Lee (1964). In his book, Lee considered the linear discrete system

(1.1.1) $$x(t+1) = Ax(t),$$

where x is a state n-vector, A is a constant matrix description of the system, and the

measurement p-vector is

$$y(t) = Cx(t). \tag{1.1.2}$$

DEFINITION 1.1.1. A system represented by (1.1.1) and (1.1.2) is said to be (l or n)-identifiable if it is possible to determine A by a (scalar or all state components) measurement of the initial condition $x(0)$, i.e., if $x(0)$ has nontrivial projections on to all the eigenvectors of A.

1.2. Identifiability as the existence of variational extremum (late 1960s). The generality of the maximum likelihood (ML) method has made it a preferred technique to other methods such as linear least-squares, instrumental variable techniques, or stochastic approximation. Kashyap (1970) has observed that ML methods are applicable to multiple input-output systems corrupted by noise, where the statistics of the disturbances are unknown. Earlier analyses of dynamical systems using the ML approach have concentrated on parameter estimation only when the output is scalar. Kashyap has shown that the ML method can be designed to deal with higher dimensional noisy output measurements, with the emphasis being on the consistency of the estimates and their numerical computation. A technique related to the ML method which admits the measurement errors to be a stationary process with unknown rational spectral density has been proposed by Åström, Bohlin and Wensmark (1965).

Let an n-dimensional state space representation of a discrete, single input linear dynamic system be

$$x(t+1) = Ax(t) + Bw(t), \tag{1.2.1}$$

$$z(t) = Cx(t) + v(t), \tag{1.2.2}$$

where $z(t)$ is the single output or measured sequence, A is an unknown $n \times n$ transition matrix, B and C are $1 \times n$, $n \times 1$ constant vectors and $w(t)$ and $v(t)$ are Gaussian white noise sequences with zero mean and unknown variances.

The Lévy "proper canonical representation" for this system is provided by a Kalman filter (Kailath (1968)),

$$\hat{x}(t+1|t) = A\hat{x}(t|t-1) + AK\nu(t), \tag{1.2.3}$$

$$z(t) = C\hat{x}(t|t-1) + \nu(t), \tag{1.2.4}$$

with initial condition

$$\hat{x}(0|-1) \equiv \hat{x}(0).$$

Here $\hat{x}(t|t-1)$ denotes the expected value of $x(t)$ based on $z_j, j = 0, \cdots, t-1$ and $\nu(t)$ is the innovation sequence which for an optimal filter can be shown to be a Gaussian white noise process ($E\{\nu\} = 0$ and $E\{\nu_i \nu_j\} = b\delta_{ij}$). As mentioned previously, K is a constant $n \times 1$ Kalman gain vector and b is a constant scalar. That is to say, the initial transient portion in which K is time varying has been neglected. The system of equations (1.2.3) and (1.2.4) has $(3n+1)$ unknowns which we call $\theta = [A, k, x_0, b]'$ where $A = [A_1, \cdots, A_n]'$. To apply the ML method, an expression for the likelihood function $J(\theta)$ is obtained using Schweppe's formula (Schweppe (1965)),

$$J_N(\theta) = -\frac{1}{2b} \sum_{t=1}^{N} \nu_t^2 - \frac{N}{2} \ln(b), \tag{1.2.5}$$

where N is the number of time points.

By calling (1.2.3–5) the D-system, we propose the following notion of stochastic ML-identifiability for a noise corrupted system.

DEFINITION 1.2. A linear stochastic D-system is said to be *locally* (*or globally*) ML-*identifiable* if the related expectation of functional J processes a local (or global) maximum.

1.3. Least square (LS) identifiability (1970). It is often tempting to generalize some observability criteria for the identification of parameters when the dynamics of the state can be obtained in terms of differentiable mappings. For instance, let us take a typical model that appears regularly in modeling,

$$\begin{aligned} x(t+1) &= f(x(t), \theta, u(t)), \\ y(t) &= g(x(t), \theta, u(t)), \end{aligned} \tag{1.3.1}$$

where u is the input p-vector, y the output r-vector, and x the state of the system. The vector θ denotes a set of unknown parameters. Augmenting the state vector x by the parameter q-vector θ, we get

$$\begin{aligned} \begin{bmatrix} x \\ \theta \end{bmatrix}(t+1) &= \begin{bmatrix} f(x(t), \theta, u(t)) \\ \theta(t) \end{bmatrix}, \\ y &= g(x, \theta, u). \end{aligned} \tag{1.3.2}$$

The system (1.3.1) is then "locally identifiable" at θ_0 if the system (1.3.2) is locally observable at (x_0, θ_0). A criterion for observability can be obtained by linearizing around a reference trajectory (Åström 1972)). A sufficient criterion is that the matrix W defined by

$$W(t+1) = A'W(t) + W(t)A + C'C$$

has all eigenvalues of magnitude greater than 1;

$$A = \begin{bmatrix} f_x & f_\theta \\ 0 & 0 \end{bmatrix}_{x_0, \theta_0} \quad \text{and} \quad C = [g_x \quad g_\theta]_{x_0, \theta_0},$$

f_x, f_θ, g_x and g_θ denote the matrices of partial derivatives. We see immediately that this type of criterion cannot be applied in advance since it requires an estimate of θ_0 at some initial condition x_0. Bellman and Åström (1970) proposed a method-oriented definition to combine prior knowledge with experimental data. One way to achieve this is to select an input signal u, measure the corresponding output y_m and determine θ in such a way that the criterion over some chosen interval $[t_0, t_1]$,

$$J(\theta) = \frac{1}{2} \sum_{t=1}^{\infty} \|y(t, \theta) - y_m(t)\|^2, \tag{1.3.3}$$

where $\|\cdot\|$ is the Euclidean norm, is minimal.

Consider a class of systems S described by (1.3.1), where $\theta \in \Omega$, a compact subset of $\mathbb{R}^q$. The internal couplings of the system as well as the couplings between the state and the inputs/outputs via f and g give S a structure. The existence of a unique estimate for the identification problem depends not only on the structures but also on the input signal.

DEFINITION 1.3.1. The structure S is called *locally* LS-*identifiable* (LLS-*identifiable*) at θ_0 if the function $J(\theta)$ has a local minimum at $\theta = \theta_0$. If the minimum is global, the structure is said to be *globally* LS-identifiable.

Remark 1.3.1. Techniques of estimation related to LS-identifiability are given in Kailath (1975). We recall that the term "least square" of the definition has a

historical connotation from the Euclidean norm, and it does *not* indicate any particular method.

The LS-identifiability can also be defined in a mean square sense. Let us consider the finite time points version of (1.3.3) with N the number of time points in $[t_0, t_1]$,

$$J_N(\theta) = \frac{1}{2} \sum_{t=1}^{N} \|y(t, \theta) - y_m(t)\|^2. \tag{1.3.4}$$

DEFINITION 1.3.2. The structure S is called LLS-*identifiable at* $\theta_0 \in \Omega$ *in the mean square sense*

(i) if θ_0 *is the local minimum of* $J(\theta)$, and

(ii) when θ_{0_N} *is the local minimum of* $J_N(\theta)$, then

$$\lim_{N \to \infty} \frac{1}{\Delta t_N} \|\theta_0 - \theta_{0_N}\|^2 = 0 \text{ in } \Omega$$

except on a subset of length zero of $[t, t_1]$. Δt_N is the time interval of partition N.

Remark 1.3.2. In stochastic problems, the augmenting state vector method corresponds to the extended Kalman filter (EKF) technique of identification. Readers may refer to Ljung (1979) for additional discussion concerning the properties of such a parameter estimator.

Remark 1.3.3. Usually LLS-identifiability in the mean square sense is implemented in computational procedures as a substitute for theoretical LLS-identifiability. When the partition N is indexed by a random sample variable ω, condition (ii) may be written as $\lim_{N \to \infty} E \|\theta_0(\omega) - \theta_{0_N}(\omega)\|^2 = 0$.

Remark 1.3.4. Equivalent to the summation formula for the cost functional of the LS-identifiability concept, one can define the functional from the prediction errors where the functional is set equal to a scalar criterion function $l(t, \theta, \varepsilon)$ with $\varepsilon \triangleq y_m(t) = g(\theta; t, y_m(t-1), u(t-1))$, $g_M(\cdot, \cdot, \cdot, \cdot)$ is the predictor function for a set $\mathcal{M}$ of models (Ljung and Glad (1980)).

1.4. Identifiability in a consistency-in-probability (CP) sense (1967–72). Perhaps we can go back a little further in time to recognize the root of the stochastic concept. In 1967, Aoki (1967) presented the following definition: a linear system with noisy measurements and no process noise is said to be *stochastically observable* if and only if the covariance matrix associated with the conditional probability density function of the state given the observation goes to zero as time increases.

The asymptotic and convergence aspects of the definition were then extended to the identification problem and examined more closely by a theoretical analysis a few years later (Aoki and Yue (1970), Balakrishnan (1969)).

Tse and Anton (1972) came up with a more efficient presentation for the basic idea when the unknown parameter belongs to a certain compact set.

Rigorously, let $\{x(t)\}_{t=1}^{\infty}$ be a sequence of observation statistics with joint probability density function $p_n(x_1, \cdots, x_m; \theta)$, $n = 1, 2, \cdots$, parameterized by the unknown parameter $\theta \in M$, where M is a separable metric space with distance function d. The true parameter θ_0 is known to lie in Ω, where Ω is a compact subset of M. Denote $X^n = \{x_1, x_2, \cdots x_n\}$; the true joint probability density function for observation sequence X^n is $p_n(X^n; \theta_0)$. An estimate of θ_0 based on the observation sequence X^n is a measurable function of X^n, and is denoted by $\theta_n = \theta_n(X^n)$.

DEFINITION 1.4.1. A sequence of estimates $\{\theta_n\}_{n=1}^{\infty}$ is said to be *consistent in probability*, if for any δ, ε real, positive and arbitrarily small, there exists an $N(\varepsilon, \delta) < \infty$

such that for $n > N(\varepsilon, \delta)$

$$\Pr\{d(\theta_n, \theta) > \delta\} < \varepsilon. \tag{1.4.1}$$

We write $\theta_n \to^p \theta_0$.

DEFINITION 1.4.2. The parameter θ_0 is said to be CP-*identifiable* if there exists a sequence of estimates $\{\tilde{\theta}_n\}_{n=1}^{\infty}$ which is consistent in probability, i.e., $\tilde{\theta}_n \to^p \theta_0$. The analogue of this definition is the definition of stochastic observability, and their locality is clear. Denote the joint density $p_n(x_1, \cdots, x_n; \theta)$ by $p(X^n; \theta)$. New information obtained at the nth sampling can be contained in a conditional density defined by

$$p(x_n | X^{n-1}; \theta = p(X^n; \theta) | p(X^{n-1}; \theta). \tag{1.4.2}$$

DEFINITION 1.4.3. Two parameters $\theta_1, \theta_2 \in \Omega$, $\theta_1 \neq \theta_2$ are said to be *unresolvable* if the equality

$$p(x_n | X^{n-1}; \theta_1) = p(x_n | X^{n-1}; \theta_2) \tag{1.4.3}$$

holds with probability 1 with respect to θ_1 and θ_2 for all except a finite number of integers $n > 0$. Let X^n be measurable; the constrained maximum likelihood (CML) estimate of θ_0 is defined as $\hat{\theta}_n$, which satisfies the equation

$$p(X^n; \hat{\theta}_n) = \max_{\theta \in \Omega} p(X^n; \theta). \tag{1.4.4}$$

Two parameters $\theta_1, \theta_2 \in \Omega$, $\theta_1 \neq \theta_2$, are said to be CML-unresolvable if (1.4.3) holds with probability 1 with respect to θ_1 and θ_2 as $n \to \infty$.

DEFINITION 1.4.4. The set Ω is said to be CP-*identifiable* if no two elements in it are CML-unresolvable.

Practical implications of the above definitions concerning Kalman filtering theory are discussed in Anderson and Moore (1979). A computational procedure for finding the domain Ω of CML-resolvability has been proposed by Lee and Herget (1976).

Remark 1.4.1. Before the development of the CP-identifiability notion, the local problem of minimal dimensionality associated with the parameter space Ω has been investigated rigorously by Barankin (1960). In his work, he constructed the definition of *parameter sufficiency* dual to that of sufficient statistics for the family of probability measures on Lebesgue measurable subsets of the state space (Silvey (1975)). Other definitions similar to CP-identifiability have also been proposed recently, especially for recursive estimation (Ljung (1976), (1979)).

1.5. Parameterization identifiability (1973). The study of algebraic structures of the transfer-function matrix of a linear system has illuminated a deterministic viewpoint toward identifiability as early as 1966. Kalman (1966) described the difficulty of identification in a notion called "degeneracy":

Consider a multivariable, constant, linear system of the form

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) \end{aligned} \tag{1.5.1}$$

with $x \in \mathbb{R}^n$, $u \in R^p$, $y \in \mathbb{R}^r$, $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{r \times n}$. Let the discrete transfer function be

$$G(z) = C(z^{-1}I - A)^{-1}B, \tag{1.5.2}$$

where z is the backward shift (or lag) operator.

We call any (1.5.1) description corresponding to a (1.5.2) description a realization of the system. A realization of G is *minimal* if the size $n \times n$ of A is as small as possible. A minimal realization is essentially unique, and is denoted by $(A, B, C)_{\min}$ for the single input-single output case. For multi-input/multi-output cases, unique parametrization of minimal realization must be assumed for identification analysis (Glover and Willems (1973)).

A transfer-function matrix G or a triple $(A, B, C)_{\min}$ specified by q parameters $\{\theta_{0,i}\}$, $i = 1, \cdots, q$, is a degenerate description of a system if in every neighborhood of $\{\theta_{0,i}\}$ there is a pair of parameter sets $\{\theta_{1,i}\}, \{\theta_{2,i}\}$ which correspond to the same system.

Let the unknown parameters be denoted by $\theta \in \Omega \subset \mathbb{R}^q$, then the system matrices may be written as $A(\theta)$, $B(\theta)$ and $C(\theta)$, where $A : \mathbb{R}^q \to \mathbb{R}^{n \times n}$, $B : \mathbb{R}^q \to \mathbb{R}^{n \times p}$ and $C : \mathbb{R}^q \to \mathbb{R}^{n \times r}$.

DEFINITION 1.5.1. Let $(A, B, C)(\theta) : \Omega \subset \mathbb{R}^q \to \mathbb{R}^{n(n+p+r)}$ be a parametrization of the system matrices (A, B, C). This parametrization is said to be *locally P-identifiable at* $\hat{\theta} \in \Omega$ if there exists an $\varepsilon > 0$ such that

$$\text{(i)} \qquad \|\theta_1 - \hat{\theta}\| < \varepsilon, \qquad \|\theta_2 - \hat{\theta}\| < \varepsilon,$$

(1.5.3)

$$\text{(ii)} \qquad C(\theta_1)(z^{-1}I - A(\theta_1))^{-1}B(\theta_1) = C(\theta_2)(z^{-1}I - A(\theta_2))^{-1}B(\theta_2),$$

for all $z \in \mathbb{C}$ (= complex plane), implying $\theta_1 = \theta_2$.

One possible way to test this definition is to consider the Markov parameters $M(\theta) = (C(\theta)B(\theta), \cdots, C(\theta)A^{2n-1}(\theta)B(\theta))'$ and to check whether, as a function of θ, it is one-to-one at $\hat{\theta}$. A local sufficient condition for this to be the case is that the Jacobian of $M(\theta)$ should be of full rank. This condition is in principle not difficult to verify, but from a computational point of view is inefficient and tedious.

Example 1.5.1 (Grewal and Glover (1976)). Consider the parametrization of a single input/single output, second order system;

$$\begin{aligned} x_1(t+1) &= -(\theta_1 + \theta_2)x_1 + \theta_2 x_2 + u, \\ x_2(t+1) &= \theta_2 x_1 - \theta_3 x_2 \quad \text{and} \quad \theta_2 \neq 0, \\ y &= x_1, \end{aligned}$$

(1.5.4)

$$M(\theta) = \begin{bmatrix} 1 \\ -(\theta_1+\theta_2) \\ \theta_1^2 + 2\theta_1\theta_2 + 2\theta_2^2 \\ -(\theta_1+\theta_2)(\theta_1^2+2\theta_1\theta_2+3\theta_2^2) - \theta_2^2\theta_3 \end{bmatrix},$$

$$\frac{\partial M}{\partial \theta} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & 0 \\ 2(\theta_1+\theta_2) & 2\theta_1+4\theta_2 & 0 \\ -3\theta_1^2-5\theta_2^2-6\theta_1\theta_2 & (-2\theta_2(\theta_1+\theta_3)-9\theta_2^2-3\theta_1^2-8\theta_1\theta_2) & -\theta_2^2 \end{bmatrix},$$

which is full rank for all $\theta \in \mathbb{R}^3$ such that $\theta_2 \neq 0$. Hence, the parametrization is locally P-identifiable (LP-identifiable) for all $\theta \in \mathbb{R}^3$ such that $\theta_2 \neq 0$.

The identifiability of this system can be checked by looking at its characteristic equation, i.e., the discrete transfer function which is given by

$$G(z)=\frac{(z^{-1}+\theta_3)}{[z^{-2}+z^{-1}(\theta_1+\theta_2+\theta_3)+(\theta_1+\theta_2)\theta_3-\theta_2^2]}.$$

Thus, any parameter combination for which

$$\theta_3=a_1, \qquad \theta_1+\theta_2=a_2-a_1,$$
$$\theta_1+\theta_2+\theta_3=a_2, \qquad \theta_2=\pm\sqrt{(a_2-a_1)a_1-a_3},$$
$$(\theta_1+\theta_2)\theta_3-\theta_2^2=a_3$$

will give the same input-output relations. There are three equations and three unknown parameters. Hence, the parametrization is LP-identifiable (or LS-identifiable) if there is no pole zero cancellation, i.e., $\theta_2\neq 0$.

DEFINITION 1.5.2. Let $(A, B, C)\ (\theta)\Omega\subset\mathbb{R}^q\to\mathbb{R}^{n(n+p+r)}$ be a parametrization of the matrices (A, B, C). This paramerization is said to be *globally* P-*identifiable* (GP-identifiable) if

(1.5.5) (i) $$C(\theta_1)(z^{-1}I-A(\theta_1))^{-1}B(\theta_1) = C(\theta_2)(z^{-1}I-A(\theta_2))^{-1}B(\theta_2) \quad \text{for all } z\in\mathbb{C},$$

(1.5.6) (ii) $(A, B, C)\ (\theta)$ being minimal implies $\theta_1=\theta_2$.

Condition (ii) could be deleted from the above definition to give a more restrictive test, but the most useful parametrizations admit multiple representations of non-minimal systems.

Remark 1.5.1. LP-identifiability for all $\theta\in\Omega$ does not, in general, imply GP-identifiability.

Remark 1.5.2. In the literature of mathematical biosciences, the identification of the state matrix $A(\theta)$ of a compartmental model from the transfer function is known as *structural identifiability*. This concept was explicitly proposed by Bellman and Åström (1970). Since the matrix $G(z)$ is made up of ratios of polynomials in z^{-1} as its elements, the question of identifiability becomes the question on the unique relationship between the coefficients of these polynomials and the components of parameter vector θ. Delforge (1977) observed that this uniqueness cannot be established through the number of coefficients and the dimension of θ alone because their relationship is essentially nonlinear. In other words, in regard to Definition 1.5.2, condition (1.5.5) without (1.5.6) is only necessary but not sufficient for both structural identifiability and GP-identifiability. In practice, the higher the dimensions of the system the more difficult it is to test condition (1.5.6). Simpler criteria equivalent to (1.5.6) have been needed for quite some time.

Example 1.5.2. Let

$$A(\theta)=\begin{pmatrix}1 & 0\\ 1+\theta & 2\end{pmatrix}, \quad B(\theta)=\begin{pmatrix}1+\theta\\ 1+3\theta\end{pmatrix}, \quad C(\theta)=(2-\theta, 4-3\theta).$$

$(A, B, C)(\theta)$ is LP-identifiable for all $\theta\in\mathbb{R}$, but is not GP-identifiable since $\theta=0$ and $\theta=1$ lead to distinct minimal representations of the same transfer function.

DEFINITION 1.5.3. A pair of parameters (θ_1, θ_2), $\theta_1, \theta_2\in\Omega$, $\theta_1\neq\theta_2$, is *locally indistinguishable* if $C(\theta_1)A^l(\theta_1)B(\theta_1)=C(\theta_2)A^l(\theta_2)B(\theta_2)$, $l=0, 1, 2\ldots$.

It is important to notice that complete controllability and observability are *insufficient* for identifiability (DiStefano (1977)).

1.6. Local identifiability via the information matrix (early 1950s). Parallel to the axiomatic development of identifiability concepts in system theory, dynamical economic models have been investigated within the context of identification problems using classical mathematical statistics and time-series analysis since the early 1950s. Often implicitly, the works during this period contained a considerable number of basic ideas on the identifiability question. Original contributions should be attributed to Cramér (1946), Koopmans and Reiersøl (1950), Wald (1950), Reiersøl (1950), Aitchinson and Silvey (1958), Whittle (1952), (1957), and Fisher (1959)–(1966) among others.

From the above works came the local identifiability concept, which is based on the rank test of the so-called *Fisher information matrix*. This concept is equivalent to the CP-identifiability concept in a local sense as discussed in the next section. We present here the basic elements of this concept, adapting from Wegge (1965) and Rothenberg (1971).

Let x be a vector-valued random variable $\mathbb{R}^n$ representing the outcome of some random experiment. The probability distribution function for x is known to belong to a family $\mathcal{F}$ of distribution functions on $\mathbb{R}^n$. A structure is a set of hypotheses which implies a unique distribution function in $\mathcal{F}$. The set of a priori possible structures to be investigated is called a model. In order to say more about the identification problem, we must be more specific about the models and $\mathcal{F}$. This is usually done by assuming that x is generated by a set of linear equations with an additive latent error term.

Assume that the distribution of x has a parametric representation, i.e., that every structure is described by a q-dimensional real vector θ and that the model is described by a set $\Omega \subset \mathbb{R}^q$. We associate with each θ in Ω a continuous probability density function $p(x, \theta)$ which, except for the parameter θ, is known. $p(x, \theta)$ is also referred to as the likelihood function and sometimes written as $L(x, \theta)$.

Thus the problem of distinguishing between structures is reduced to the problem of distinguishing between parameter points. Similar to § 1.4, we have

DEFINITION 1.6.1. Two parameter points (structures) θ_1 and θ_2 in Ω are said to be *observationally equivalent* if $p(x, \theta_1) = p(x, \theta_2)$ for all $x \in \mathbb{R}^n$.

DEFINITION 1.6.2. A parameter point θ_0 is said to be *locally* (information matrix) IM-*identifiable* if there exists an open neighborhood of θ_0 containing no other $\theta \in \Omega$ which is observationally equivalent.

DEFINITION 1.6.3. A parameter point θ_0 in Ω is said to be *globally* IM-*identifiable* if there is no other θ in Ω which is observationally equivalent.

The identification problem is to find conditions on $p(x, \theta)$ and Ω that are necessary and sufficient for the identification of the parameters in Ω. This question has been partially answered by Cramér (1946, p. 479). For LIM-identifiability, we need the following assumptions:

(1) The structural parameter space Ω is an open set in $\mathbb{R}^q$.

(2) The function p is a proper density function for every θ in Ω. In particular, p is nonnegative and the equation $\int p(x, \theta)\, dx = 1$ holds for all θ in Ω.

(3) The set of x values for which p is strictly positive is the same for all θ in Ω. We call this set the sample space of x.

(4) The function p is smooth in θ. Specifically, we assume that for all θ in a convex set containing Ω and for all x in the sample space the functions $p(x, \theta)$ and $\log p(x, \theta)$ are continuously differentiable with respect to θ.

(5) The elements of the information matrix $J_{n,n}(\theta) = E(\partial \ln p/\partial\theta_i) \cdot (\partial \ln p/\partial\theta_j)$ exist and are continuous functions of θ everywhere in Ω.

DEFINITION 1.6.4. Let $M(\theta)$ be a matrix whose elements are continuous functions of θ everywhere in Ω. The point θ_0 is said to be a *regular point* of the matrix if there exists an open neighborhood of θ_0 in which $M(\theta)$ has constant rank.

Let $\psi_i(\theta)=0$, $i=1,\cdots,k$ be a set of constraint equations where the Jacobian matrix for the ψ_i is well defined as

$$J_{k,q}(\theta)=\left(\frac{\partial\psi_i}{\partial\theta_j}\right)$$

Define the $(q+k)\times q$ partitioned matrix

$$J_{n,k}(\theta)=\left[\frac{J_{n,n}(\theta)}{J_{k,q}(\theta)}\right];$$

we then have the following criteria for LIM-identifiability.

Criterion I. Let θ_0 be a regular point of the matrix $J_{n,n}(\theta)$. Then θ_0 is LIM-identifiable if and only if $J_{n,n}(\theta_0)$ is nonsingular.

Criterion II. For a constrained model, let the new parameter space Ω' be the intersection of Ω and the solution set of the constraint equations. Suppose that θ_0 in Ω' is a regular point of both $J_{k,q}(\theta)$ and $J_{n,k}(\theta)$. Then θ_0 is CLIM-identifiable if and only if $J_{n,k}(\theta)$ has rank q.

Proofs of these criteria are shown in Rothenberg (1971). In the next section, we will show the linkage between CP-identifiability and LIM-identifiability when we construct the proof of the weak unification proposition.

Methods to obtain criteria for GIM-identifiability and IM-identifiability for reduced-form parameters of canonical models, and the identification of simultaneous equations models, are discussed in Wegge (1965), Fisher (1966) and Rothenberg (1971).

2. Unification of linear local identifiability criteria. In the preceding section, we have explored the current concepts of identifiability for a given class of linear systems in discrete time. On the basis of the identifiability criteria presented, we will prove next an equivalence statement among the main definitions. Such an equivalence statement means that one definition of local identifiability would imply another when the related criteria are enforced in a consistent manner with respect to the chosen representation of a deterministic or stochastic system. Without loss of generality, let us consider the two following simplified versions of system (1.0.1–2) on an interval $[t_0, t_1]$.

Deterministic model.

$$\begin{aligned} x(t+1,\theta)&=A(\theta)x(t,\theta)+B(\theta)u(t),\\ y(t,\theta)&=C(\theta)x(t,\theta). \end{aligned} \tag{2.0.1}$$

We recall that A, B and C are $n\times n$, $n\times p$ and $r\times n$ matrices, respectively. Let the dimension of $\theta\in\Omega$ be q; let (A, B, C) be controllable, observable, and minimal. The Markov parameter matrix $M(\theta)=(C(\theta)B(\theta),\cdots,C(\theta)A^{2n-1}(\theta)B(\theta))$.

Stochastic model.

$$\begin{aligned} x(t+1,\theta)&=Ax(t)+w(t),\\ z(t)&=Cx(t)+v(t), \end{aligned} \tag{2.0.2}$$

where $x(t)$ in $\mathbb{R}^n$, $z(t)$ in $\mathbb{R}^r$, and $w(t)$ and $v(t)$ are zero-mean Gaussian noises with

covariances

$$E\{w_i w_j'\} = R\delta_{ij}, \quad E\{v_i v_j'\} = Q\delta_{ij}, \quad E\{w_i v_j'\} = D\delta_{ij}.$$

Let $\theta = \{x_0, A, C, R, Q, D\}$. Suppose A is stable and (A, C) is an observable pair. We define the information matrix associated with (2.0.2) to be:

$$J_{n,n}(\theta_0) = E_{\theta_0}\left(\frac{\partial \ln p(x_n|X^{n-1}; \theta_0)}{\partial \theta_0}\right)\left(\frac{\partial \ln p(x_n|X^{n-1}; \theta_0)}{\partial \theta_0}\right)'.$$

The main result of this section can be precisely stated as follows.

WEAK UNIFICATION PROPOSITION. *For a class of linear, time invariant, deterministic or stochastic dynamics of the types* (2.0.1) *and* (2.0.2), *the definitions of* LLS-*identifiability*, LP-*identifiability*, CP-*identifiability and* LML-*identifiability can be used interchangeably with respect to the Lévy representation of the stochastic model.*

Remark 2.0.1. The term "weak" indicates the fact that the statement holds only for the class of systems specified. The proof of our proposition will begin with the implication of LP-identifiability by LLS-identifiability and proceed according to the interchange diagram:

$$\begin{array}{ccc} \text{LLS-Identifiability (m.s.)} & \longrightarrow & \text{LP-Identifiability} \\ \uparrow & & \downarrow \\ \text{LML-Identifiability} & \longleftarrow & \text{CP-Identifiability} \end{array}$$

where m.s. denotes the mean square sense. The arrow means "can be used in place of".

2.1. Assertion 1. LLS-identifiability (m.s.) → LP-identifiability. The proof of the interchangeability of LP-identifiability with LLS-identifiability consists of two steps.

PROPOSITION 2.1.1. *For $q \leqq n(n+p+r)$, the system* (2.1.1) *is* LP-*identifiable at $\hat{\theta}$ if $\partial M/\partial\theta$ has rank q at $\hat{\theta}$. For $q > n(n+p+r)$, the system is unidentifiable.*

Proof. It is clear that if $q > n(n+p+r)$ there are too many degrees of freedom and no identifiability is possible.

For $q \leqq n(n+p+r)$, $\partial M/\partial\theta$ of rank q implies that the mapping from the parameter space into the Markov parameters is locally one-to-one at $\hat{\theta}$. In other words, $C(\theta_1)B(\theta_1) = C(\theta_2)B(\theta_2), \cdots, C(\theta_1)A^{2n-1}(\theta_1)B(\theta_1) = C(\theta_2)A^{2n-1}(\theta_2)B(\theta_2)$ imply $\theta_1 = \theta_2$ for any θ_1, θ_2 in the neighborhood of θ.

By the Cayley–Hamilton theorem (Hoffman and Kunze (1971)), these equations can be immediately extended to: $C(\theta_1)B(\theta_1) = C(\theta_2)B(\theta_2), \cdots, C(\theta_1)A^N(\theta_1)B(\theta_1) = C(\theta_2)A^N(\theta_2)B(\theta_2)$ with $N \to \infty$ imply $\theta_1 = \theta_2$ near $\hat{\theta}$. From condition (v), i.e., the stability of A, we can expand $(z^{-1}I - A)^{-1}$ in a Laurent series of z with coefficients of the form A^n (Mehra (1974, p. 26)), and the above result is equivalent to (1.5.3):

$$C(\theta_1)(z^{-1}I - A(\theta_1))^{-1}B(\theta_1) = C(\theta_2)(z^{-1}I - A(\theta_2))^{-1}B(\theta_2)$$

implies $\theta_1 = \theta_2$ for θ_1, θ_2 near $\hat{\theta}$. Therefore by definition 1.5.1, we conclude that (2.0.1) is LP-identifiable at $\hat{\theta}$ in Ω. Here, condition (i) has been used in obtaining (1.5.3). □

Remark 2.1.1. Using Definition 1.5.3, we say that a system is LP-identifiable if it is locally distinguishable.

PROPOSITION 2.1.2. *The system* (2.0.1) *is locally distinguishable if it is* LLS-*identifiable at $\hat{\theta}$.*

Proof. Define for θ_1, θ_2 in Ω, $\theta_1 \neq \theta_2$, two following functionals (cf. (1.3.3)):

$$J(\theta_i) = \frac{1}{2} \sum_{t=1}^{\infty} \|y(t, \theta_i) - y_m(t)\|^2, \qquad i = 1, 2. \tag{2.1.1}$$

Subtracting these two functionals, we have

$$J(\theta_1) - J(\theta_2) = \frac{1}{2} \sum_{t=1}^{\infty} [\|y(t, \theta_1) - y_m(t)\|^2 - \|y(t, \theta_2) - y_m(t)\|^2]. \tag{2.1.2}$$

From the definition of LLS-identifiability, (2.1.4) can be rewritten as

$$J(\theta_1) - J(\hat{\theta}) = \frac{1}{2} \sum_{t=1}^{\infty} [\|y(t, \theta_1) - y_m(t)\|^2 - \|y(t, \hat{\theta}) - y_m(t)\|^2], \tag{2.1.3}$$

and $J(\hat{\theta}) = 0$ implies $y(t, \hat{\theta}) = y_m(t)$. Then

$$J(\theta_1) = \frac{1}{2} \sum_{t=1}^{\infty} \|y(t, \theta_1) - y(t, \hat{\theta})\|^2. \tag{2.1.4}$$

If θ_1 is near $\hat{\theta}$ and if J has a local minimum at $\theta = \hat{\theta}$, then $J(\theta_1) > 0$, which implies $y(t, \theta_1) \neq y(t, \hat{\theta})$ from (2.1.4). Therefore LLS-identifiability implies local distinguishability.

Propositions 2.1.1 and 2.1.2 verify Assertion 1. (See Grewal and Glover (1976) for a similar proof.) □

Remark 2.1.2. If there is no input ($u(t) = 0$) in the system (2.0.1), we set $B(\theta)$ to be the identity matrix and the proof is unchanged.

Remark 2.1.3. Glover and Willems (1973) used the implicit function theorem to obtain the general test for LP-identifiability. This test is in fact equivalent to the criterion on the rank of the Jacobian of the Markov parameter mapping.

2.2. Assertion 2. LP-identifiability → CP identifiability. Using the Lévy representation, we have, for model (2.0.2),

$$\begin{aligned} \hat{x}(t|t-1) &= A\hat{x}(t|t-2) + K\nu(t-1), \\ \hat{z}(t|t-1) &= C\hat{x}(t|t-1), \\ \nu(t) &= z(t) - C\hat{x}(t|t-1), \end{aligned} \tag{2.2.1}$$

$$\begin{aligned} K &= (APC' + D)(CPC' + Q)^{-1}, \\ P &= APA' + R - K(CPC' + Q)K'. \end{aligned} \tag{2.2.2}$$

The prime denotes the matrix transpose operation. Assume that (A, K) is a controllable pair, i.e., condition (ii). $\nu(t)$ acts as an input sequence. The observability assures the steady state of the Kalman gain K. Stability of A, i.e., condition (v), gives meaning to the results as $t \to \infty$.

Assume that the system (2.2.1) is LP-identifiable in the steady state, with the Markov parameter test verified at some point $\hat{\theta}$ in Ω. Let θ_1, θ_2 be near $\hat{\theta}$. From results of § 2.1, if $\theta_1 \neq \theta_2$, then

$$C_1 A_1^N K_1 \neq C_2 A_2^N K_2, \qquad N = 1, 2, \cdots. \tag{2.2.3}$$

Together with (2.2.2), we have with probability 1

$$\hat{z}_1(t|t-1) \neq \hat{z}_2(t|t-1) \quad \text{for large } t. \tag{2.2.4}$$

With similar arguments as in § 2.1, (2.2.1), (2.2.3) and (2.2.4) indicate, for an arbitrary nonsingular matrix T,

$$A_1 \neq TA_2T^{-1}, \quad K_1 \neq TK_2, \quad C_1 \neq C_2T^{-1} \tag{2.2.5}$$

and additionally for sufficiently large time, with probability 1,

$$C_1P_1C_1' + Q_1 \neq C_2P_2C_2' + Q_2. \tag{2.2.6}$$

Consequently, $p(z(t)|Z^{t-1}; \theta_1) \neq p(z(t)|Z^{t-1}; \theta_2)$ holds with probability 1 as $t \to \infty$ because $p(x(t)|Z^{t-1}; \theta)$ is Gaussian with mean $\hat{z}(t|t-1)$ and covariance $CPC' + Q$ as $t \to \infty$; i.e., θ_1, θ_2 *cannot* be CML-unresolvable near $\hat{\theta}$. In other words, with appropriate assumptions, LP-identifiability of system (2.2.1) leads to CML-resolvability of Ω.

By Definition 1.4.4, the proof of Assertion 2 is finished. □

See Tse and Weinert (1973) for comparison. It is also possible to prove the same result on the framework of Akaike's stochastic theory of minimal realization (Akaike 1974)).

2.3. Assertion 3. CP-identifiability → LML identifiability. To prove this assertion, we examine first the following lemma of Tse (1973).

LEMMA 2.3.1. *The parameter θ_0 is locally* CP-*identifiable if, for all $n = 1, 2, \cdots$, there exists $\lambda^2 > 0$ such that*

$$J_{n,n}(\theta_0) > \lambda^2 I, \tag{2.3.1}$$

where I is the $n \times n$ identity matrix.

From Tse's proof (1973), we learn that (2.3.1) indicates that, around the true parameter θ_0, there is a local minimum in the likelihood function which, in probability, will be located arbitrarily close to θ_0 asymptotically. Therefore, if we assume that the function $\ln p(x_n|X^{n-1}; \theta_0)$ possesses an extremum at θ_0, and using the formula for a likelihood function L of θ (Eykhoff (1974)), we obtain

$$E\left\{\left(\frac{\partial \ln L}{\partial \theta}\right)\left(\frac{\partial \ln L}{\partial \theta}\right)'\right\} = -E\left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'}\right) \tag{2.3.2}$$

PROPOSITION 2.3.1. *Provided that $\ln p(x_n|X^{n-1}; \theta_0)$ possesses an extremum at θ_0, for all $n = 1, 2, \cdots$ we have*

$$J_{n,n}(\theta_0) = -E_{\theta_0}\left\{\frac{\partial^2 \ln p(x_n|X^{n-1}; \theta_0)}{\partial \theta_0 \partial \theta_0'}\right\} > 0^n,$$

if θ_0 is locally CP-*identifiable (0^n is a zero $n \times n$ matrix).*

Proof. Local resolvability near θ_0 implies that the extremum is isolated. Thus,

$$E_{\theta_0}\left\{\frac{\partial \ln p(x_n|X^{n-1}; \theta_0)}{\partial \theta_0}\right\} \neq 0$$

for finite n because the true minimum has not been reached. Consequently,

$$E_{\theta_0}\left\{\left(\frac{\partial \ln p(x_n|X^{n-1}; \theta_0)}{\partial \theta_0}\right)\left(\frac{\partial \ln p(x_n|X^{n-1}; \theta_0)}{\partial \theta_0}\right)'\right\} > 0^n,$$

and the proposition follows. □

For discrete stochastic systems with Gaussian assumptions, simple forms of the maximum likelihood function can be obtained using Schweppe's formula (Mehra

(1969), Kashyap (1970)); cf. § 1.2. Proposition 2.3.2 means

$$E_{\theta_0}\left\{\frac{\partial^2 \ln p(x_n|X^{n-1};\theta_0)}{\partial\theta_0\partial\theta_0'}\right\}<0^n \quad \text{for all } n=1,2\cdots.$$

Therefore, θ_0 is actually a local maximum, and thus by definition Assertion 3 is justified. □

Remark 2.3.1. When the stochastic systems are continuous in the mean square, the associated likelihood functions take on the stochastic integral forms and should be understood in Ito's sense (Mehra (1969)). To date the field of stochastic calculus of variations has not been completely established; therefore we exclude such systems from the present work.

Remark 2.3.2. Necessary and sufficient conditions have also been recently derived for LML-identifiability in terms of the limit matrix of a sequence of $J_{n,n}(\theta_0)$ from sequential log likelihood functions L_k (k is the serial index), and these results are similar to Lemma 2.3 and Proposition 2.3.1 (Goodrich and Caines (1979)). $J_{n,n}(\theta_0)$ is sometimes referred to as the Fisher information matrix, and its inverse is the Cramér-Rao lower bound (Silvey (1975)). For the computation of $J_{n,n}(\theta_0)$, see Mehra (1974b) and Goodwin et al. (1974).

2.4. Assertion 4. LML identifiability → LLS identifiability (m.s.).

Proof. For this proof, we use the concept of LLS-identifiability in the mean square sense instead of LLS-identifiability. By strict definition, given the system (2.0.2), or equivalently given its Lévy representation (2.2.1), the discrete least square functional for the steady state may be expressed as

$$J_N[\theta]=\frac{1}{2}\sum_{t=1}^{N}\|\nu(t,\theta)\|^2 \tag{2.4.1}$$

for N measurements. We notice that such an expression is possible due to the orthogonality of the innovation sequence (indirectly from condition (iii), see Mehra (1970)).

The minimization of $J[\theta]$ for the zero-mean Gaussian white noise process $\nu(t)$, i.e., optimal filtering, means that the parameter vector $\hat{\theta}_{0_N}$ must be chosen such that the state estimate $\hat{x}(t|t-1)$ at $i=N$ is closest to $z(t)$. If we call the distribution density of $\{z(t), t=1,\cdots,N\}$ for a given θ: $L(\nu_1,\cdots,\nu_N;\theta)=p(z_1,\cdots,z_N;\theta)$, the proper choice of θ would in turn maximize L. As the logarithmic function is monotonic, the local maximum of L and the maximum of $\ln(L)$ occur at the same value of $\hat{\theta}_{0_N}$. Hence the sufficient condition to minimize $J_N[\theta]$ is to maximize $\ln L(\nu_1,\cdots,\nu_N;\theta)$. Using Schweppe's formula, LML-identifiability means the positive definiteness of $J_{n,n}[\theta]$, where

$$J_{n,n}[\theta]=-E\left(\frac{\partial^2 L_N}{\partial\theta\partial\theta'}\right). \tag{2.4.2}$$

In other words, $\hat{\theta}_N$ is the local minimum of $J_N[\theta]$. As $N\to\infty$, the local maximum $\hat{\theta}_0$ of $\lim_{N\to\infty}\ln L(\nu_1,\cdots,\nu_N;\theta)$ will be the local minimum of

$$J[\theta]=\frac{1}{2}\sum_{t=1}^{\infty}\|\nu(t,\theta)\|^2,$$

except over a set of probability measure P zero, i.e.,

$$\lim_{N\to\infty}E\|\hat{\theta}_0-\hat{\theta}_{0_N}\|^2=\lim_{N\to\infty}\int_{\mathscr{P}}\|\theta_0(\omega)-\hat{\theta}_{0_N}(\omega)\|^2P(d\omega)=0, \tag{2.4.3}$$

where ω is the sample index and $\mathcal{S}$ is the sample space. By Definition 1.3.2, we prove the fourth assertion. With this last assertion, the interchange diagram commutes. □

3. Summary and conclusions. We have presented a combination of ideas underlying the meaning of identifiability of a class of linear dynamical systems. These ideas have been established, most of them recently, under the forms of definitions and intrinsic criteria. In spite of a number of survey papers on identification techniques and the state-of-the-art, the basic elements concerning the concept itself have not been treated as one definite entity. By bringing them together in preceding sections, we now see the emergence of a general concept called "identifiability".

The existence of this concept may be regarded as the veracity of the following statements:

(1) A uniqueness property justified in the parameter space.
(2) The set of criteria for obtaining a specific parameter vector consists of:
   (a) conditions on controllability and observability,
   (b) conditions on dimensions of natural and canonical representations of state dynamics,
   (c) conditions on the canonical forms and minimal realizations,
   (d) conditions on types of noise and distribution functions of state,
   (e) condition on persistent excitation from inputs,
   (f) condition on stability of state.

So far, these statements have been applied mostly to linear, deterministic and stochastic systems; they are necessary and sufficient for local identification of single input/single output models. For multi-input/multi-output systems, further conditions are needed and these are not completely resolved at this time.

The present knowledge of linear identifiability can be extended to a general definition applicable to nonlinear systems with degenerate cost functionals. By degenerate functionals we mean the appearance of polynomials of order higher than quadratic in the transition of linearized state dynamics from one time interval to another. These higher order polynomials, known as normal forms in the literature of singularity theory, would preserve the continuity of identifiability in the global quest of identification. The construction of that definition will appear in a forthcoming paper.

Acknowledgment. The authors would like to thank Raman Mehra, Bradley Dickinson and John Casti for many of their suggestions, and the referee for his comments on improving the presentation.

REFERENCES

J. AITCHINSON AND S. D. SILVEY (1958), *Maximum-likelihood estimation of parameters subject to restraints*, Ann. of Math. Statist., 29, pp. 813–828.

H. AKAIKE (1974), *Stochastic theory of minimal realization*, IEEE, Trans. Automat. Control, AC-19, pp. 667–674.

B. D. O. ANDERSON AND J. B. MOORE (1979), *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ

M. AOKI (1976), *Optimization of Stochastic Systems*, Academic Press, New York.

M. AOKI AND P. C. YUE (1970) *On certain convergence questions in system identification*. SIAM J. Control, 8, pp. 239–256.

K. J. ÅSTRÖM (1972) *System identification*, Lecture Notes in Mathematics 294, R. Curtain ed., Springer, New York, pp. 35–55.

K. J. ÅSTRÖM, T. BOHLIN AND S. WENSMANK (1965), *Automatic Construction of Linear Stochastic Dynamic Models for Industrial Processes with Random Disturbances Using Operating Records*, Report, IBM, Nordic Lab., Stockholm, Sweden.

K. J. ÅSTRÖM AND P. EYKHOFF (1971), *System identification—a survey*, Automatica, 7, pp. 123–162.
A. V. BALAKRISHNAN (1969), *Techniques of system identifications*, Lecture Notes, Institute Electrotecnio, University of Rome, Italy.
A. V. BALAKRISHNAN AND V. PETERKA (1969), *Identification in automatic control systems*, Survey Paper, Automatica, 5, pp. 817–829.
E. W. BARANKIN (1960) *Sufficient parameters: solution of the minimal dimensionality problem*, Inst. Stat. Math., 12, pp. 91–118.
R. BELLMAN AND K. J. ÅSTRÖM (1970), *On structural identifiability*, Math. Biosci, 7, pp. 329–339.
R. W. BROCKETT (1970), *Finite Dimensional Linear Systems*, John Wiley, New York.
H. CRAMÉR (1946), *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ.
M. CUENOD AND A. SAGE (1968) *Comparison of some methods used for process identification*, Automatica, 4, pp. 235–269.
J. DELFORGE (1977), *The problem of structural identifiability of a linear compartmental system: Solved or not?* Math. Biosci., 36, pp. 119–125.
J. J. DISTEFANO III (1977), *On the relationships between structural identifiability and the controllability, observability properties*, IEEE Trans. Automat. Control, AC-22, p. 625.
J. J. DISTEFANO AND C. COBELLI (1980), *On parameter and structural identifiability: Nonunique observability/reconstructibility for identifiable systems, other ambiguities, and new definitions*, IEEE Trans. Automat. Control, AC-25, pp. 830–833.
P. EYKHOFF (1968) *Process parameter and state estimation*, Automatica, 4, pp. 205–233.
——— (1974), *System Identification*, John Wiley, London.
P. EYKHOFF, P. M. VAN DER GRINTEN, H. KWAKERNAAK AND B. P. VELTMAN (1966), *System modeling and identification*. Proc. III IFAC Congress, London.
F. M. FISHER (1959), *Generalization of the rank and order conditions for identifiability*, Econometrica, 27, pp. 431–447.
——— (1961), *Identifiability criteria in nonlinear systems*, Econometrica, 29, pp. 574–590.
——— (1963), *Uncorrelated disturbances and identifiability criteria*, Intl. Econ. Rev., 4, pp. 134–152.
——— (1966), *The Identification Problem in Econometrics*, McGraw-Hill, New York.
A. GELB (1974) *Applied Optimal Estimation*, MIT Press, Cambridge, MA.
K. GLOVER AND J. C. WILLEMS (1973), *Parametrizations of linear dynamical systems* Proc., III IFAC Congress, the Hague, Netherlands, pp. 867–870.
R. L. GOODRICH AND P. E. CAINES (1979), *Necessary and sufficient conditions for local second-order identifiability*, IEEE Trans. Automat. Control, AC-24, pp. 125–127.
G. C. GOODWIN, M. B. ZARROP AND R. L. PAYNE (1974), *Coupled design of test signals, sampling intervals and filters for system identification*, IEEE Trans. Automat. Control, AC-19, pp. 748–749.
M. S. GREWAL AND K. GLOVER (1976), *Identifiability of linear and nonlinear dynamical systems*, IEEE Trans. Automat. Control, AC-21, pp. 833–837.
K. HOFFMAN AND R. KUNZE (1971), *Linear Algebra* (2nd ed.), Prentice-Hall, Englewood Cliffs, NJ.
IEEE Transactions on Automatic Control (1974) 19, 6, Special Issue on System Identification and Time-Series Analysis.
A. H. JASWINSKI (1970), *Stochastic Processes and Filtering Theory*, Academic Press, New York.
T. KAILATH (1968), *An innovation approach to least-square estimation* I, IEEE Trans. Automat. Control, AC-13, pp. 146–655.
——— (1975), *Lectures on linear least squares estimation*, Department of Electrical Engineering, Stanford University, Stanford, CA.
R. E. KALMAN (1966), *On structural properties of linear, constant, multivariate systems*, Proc. III IFAC Congress, London, pp. 6A.1–9.
R. E. KALMAN, P. FALB AND M. ARBIT (1969), *Topics in Mathematical System Theory*, McGraw-Hill, New York.
R. L. KASHYAP (1970), *Maximum likelihood identification of stochastic linear systems*, IEEE Trans. Automat. Control, AC-15, pp. 25–34.
T. C. KOOPMANS AND O. REIERSØL (1950), *Identification of structural characteristics*, Ann. of Math. Statist., 21, pp. 165–181.
C.-H. LEE AND C. J. HERGET (1976), *Regions of constrained maximum likelihood parameter identifiability* IEEE Trans. Automat. Control, AC-21, pp. 556–560.
R. C. K. LEE (1964), *Optimal Estimation, Identification and Control*, MIT Press, Cambridge, MA.
E. B. LEE AND L. MARKUS (1967), *Foundations of Optimal Control Theory*, John Wiley, New York.
L. LJUNG (1976), *On consistency and identifiability*, Math. Progr. Stud. 5, pp. 169–190.
——— (1979), *Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems*, IEEE Trans. Automat. Control, AC-24, pp. 36–50.

L. LJUNG AND T. GLAD (1980), *Division of Automatic Control Activity Report on System Identification*, Dept. of Electrical Engineering, Linkoping University, Linkoping, Sweden.

D. MAYNE (1972), *A canonical model for identification of multivariable linear systems* IEEE Trans. Automat. Control, AC-17, pp. 728–729.

R. K. MEHRA (1969), *Identification on stochastic linear dynamic systems*, Proc. IEEE Symposium on Adaptive Processes VIII, University Park, Pennsylvania, pp. 6.Fl-5.

——— (1970), *On identification of variances and adaptive Kalman filtering*, IEEE Trans. Automat. Control, AC-15, pp. 175–184.

——— (1974a), *Topics in stochastic control theory—identification in control and econometrics: similarities and differences*, Ann. Econ. Soc. Meas. 3, pp. 24–47.

——— (1974b), *Optimal input signals for parameter estimation in dynamical systems—Survey and new results*, IEEE, Trans. Automat. Control, AC-19, pp. 753–768.

V. POPOV (1972), *Invariant description of linear time-invariant controllable systems* SIAM J. Control, 10, pp. 252–264.

O. REIRSØL (1950), *Identifiability of a linear relation between variables which are subject to error*, Econometrica, 18, pp. 375–389.

J. RISSANEN (1974), *Basis of invariants and canonical forms for linear dynamical systems*, Automatica, 10, pp. 175–182.

T. J. ROTHENBERG (1971), *Identification in parametric models*, Econometrica, 39, pp. 577–591.

F. C. SCHWEPPE (1965), *Likelihood functions for Gaussian signals*, IEEE Trans. Inform. Theory, IT-11, pp. 61–70.

S. D. SILVEY (1975), *Statistical Inference*, Chapman and Hall, London.

E. TSE (1973), *Information matrix and local identifiability of parameters*, Proc. XIV JACC, Columbus, Ohio, pp. 611–619.

E. TSE AND J. J. ANTON (1972), *On the identifiability of parameters*, IEEE Trans. Automat. Control, AC-17, pp. 637–646.

E. TSE AND H. WEINERT (1973), *Correction and estimation of "On the identifiability of parameters"*, IEEE Trans. Automat. Control, AC-18, pp. 687–688.

A. WALD (1950), *Note on the identification of economic relations* in Statistical Inference in Dynamic Economic Models, T. C. Koopmans, ed., John Wiley, New York.

L. L. WEGGE (1965), *Identifiability criteria for a system of equations as a whole.*, Australian J. Statist., 7, pp. 67–77.

P. WHITTLE (1952), *Tests of fit in time-series* Biometrica, 39, pp. 309–318.

——— (1957), *On the use of normal approximation in the treatment of stochastic processes*, J. Roy. Statist. Soc., Ser. B, 20, pp. 334–343.