Volume Title: Annals of Economic and Social Measurement, Volume 3, number 1

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm74-1

Publication Date: 1974

Chapter Title: Identification in Control and Econometrics: Similarities and Differences

Chapter Author: R. H. Mehra

Chapter URL: http://www.nber.org/chapters/c9993

Chapter pages in book: (p. 21 - 47)

Annals of Economic and Social Measurement, 3/1, 1974

TOPICS IN STOCHASTIC CONTROL THEORY

IDENTIFICATION IN CONTROL AND ECONOMETRICS; SIMILARITIES AND DIFFERENCES

BY R. K. MEHRA*

This report attempts to bridge the gap between the economic and the control literatures on the subject of system identification and parameter estimation. It is pointed out that the emphasis in the economic literature is on large simultaneous equation models and linear estimation techniques, whereas the emphasis in the control literature is on state vector and transfer function models, on problems due to partial state observations and nonlinear estimation techniques. Since a step in the direction of easier communication between researchers in the two fields would be the use of a common model, the state-vector model of control which has already been used in several economic studies is proposed as a unifying link. The relationship of the state-vector model to the simultaneous equation model and the role of process and measurement noise in the econometric context are discussed. Complete results on the identifiability of state-vector models along with a stepwise two-stage least squares method for model structure determination and a maximum likelihood method for parameter estimation are given. The problems of closed-loop system identification and input design are also briefly discussed.

1. INTRODUCTION

The purpose of this report is to make an attempt at bridging the gap between System Identification in the Control and Econometric literatures. The task is not simple due to a relatively long history of development of the area in both fields. Even the word "System Identification" has different connotations in the two fields, e.g. in control, the word generally denotes the complete three step iterative process of model specification, parameter estimation and model verification (see Figure 1). However, in the econometric literature, the term "Identification" refers mainly to identifiability questions which have to be settled before attempting parameter estimation. In this paper, we will use the word "System Identification" in the context of control systems.

There are perhaps more similarities than differences between the control and the econometric literature on the subject of system identification. Both rely heavily on the theories of probability and statistical inference, in particular least squares estimation, likelihood and Bayesian inference. The differences stem mainly from the models considered, availability of data, objectives of identification and the specific details of estimation algorithms. We elaborate on these points in the following sections and present a model which is general enough to include a large number of problems of interest to both econometricians and control engineers. This model is based on the state space concept and has been studied quite thoroughly in the control literature. We present identifiability results on this model and discuss the estimation techniques that have been employed. Finally, we give a few examples and mention other related problems. To help the readers with

* The research reported in this document was made possible through support extended the Division of Engineering and Applied Physics, Harvard University by the U.S. Army Research Office, the U.S. Air Force Office of Scientific Research and the U.S. Office of Naval Research under the Joint Services Electronics Program by Contracts N00014-67-A-0298-0006, 0005, and 0008. Technical Report No. 647. Reproduction in whole or in part is permitted by the U.S. government.

STEPS IN IDENTIFICATION

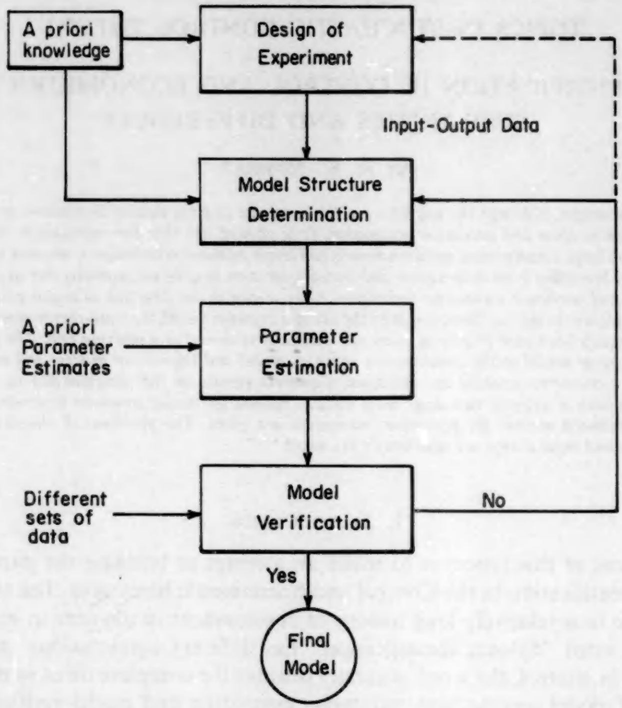


Figure 1

terminology, a table by Dhrymes, Klein and Steiglitz [1] has been expanded and presented below.

Because of the author's particular background, it has been very difficult to avoid an overemphasis on the contributions from control literature. Perhaps a similar attempt by an econometrician would help restore the balance by emphasizing the contributions from the econometric literature and showing their relevance to the control problems.

2. Comparison of Models and Estimation Techniques

Following the pioneering work of Weiner [2, 1930] on generalized harmonic analysis of random processes and his solution to the filtering problem in the spectral domain, the earlier work in control and communication used frequency domain concepts for system identification. Mostly single-input single-output systems with rational transfer functions were considered. Since most of the systems could be excited by deterministic sinusoidal inputs and relatively noise-free data obtained, little attention was paid to the statistical properties of the identified parameters and transfer functions. For illustration purposes, let us consider aircraft parameter identification. The equations of motion of a rigid aircraft are

TABLE 1

TERMINOLOGY OF SYSTEM IDENTIFICATION IN ECONOMETRICS AND CONTROL

Control	Econometrics
Noise	Error
White Gaussian noise	Nonautocorrelated normally distributed error
Colored noise	Autocorrelated error
Measurement noise	Error-in-variables
Process noise	Disturbance term
Record	Sample
Rational z-transform	Rational lag distribution
Identification	Specification and estimation of a model
Identifiable model	Justidentified or overidentified model
Unidentifiable model	Underidentified model
Input variable	Exogeneous variable
Output variable	Endogeneous variable
Equation error method	Ordinary least squares or linear regression
Output error method	Nonlinear regression
Impulse response model	Final form model
Impulse response function; Markov parameters; Weighting pattern	Impact, interim and total multipliers
Filtering	Exponential smoothing

written down easily using Newton's laws of motion in terms of the aerodynamic, gravitational and kinematic forces. The parameters relating the aerodynamic forces to the motion variables such as linear and angular velocities are called stability and control derivatives. For small deviations in velocities and angles from nominal values, the motion can be described by a set of linear differential equations of the type

$$A\dot{x} = Fx + Gu \tag{1}$$

$$y = Hx + Du + v \tag{2}$$

where A is an inertia matrix (nonsingular), x denotes the state vector consisting of displacements (angular) and velocities, u is the control input (elevator, rudder, aeleron), y is the measured output that is assumed to be contaminated with noise v. The matrices A, F, G, H and D are assumed constant and contain unknown parameters that are elements of a vector θ. All other variables viz. x, u, y and v are functions of time.

A brief survey of the methods used for estimating θ and refinements of the above model would now be presented. Greenberg [3, 1951] in an early survey paper describes following techniques.

(i) Sinusoidal response method

(ii) Inspection of the transient
(iii) Fourier Transform Method
(iv) Derivative method
(v) Prony's method

} Transient Response Methods.

Methods (i), (iii) and (iv) basically use the principle of least squares, whereas (ii) and (v) rely on the response of a linear system to a pulse-type of input. Prony's method is particularly interesting since it uses the fact that $y(t)$ for a step or impulse input $u(t)$ can be expressed as the sum of exponentials corresponding to the eigenvalues of the F-matrix, e.g.

$$y(t) = a_0 + \sum_{i=1}^{n} a_i e^{\lambda_i t} + v(t) \tag{3}$$

where it is assumed that F has distinct eigenvalues. It is easily seen that equation (3) is similar to the distributed lag model which has received increasing attention in the econometric literature in recent years [1, 4, 5]. Another technique used early on was called Equation–Error Method [6], which is identical to regression analysis. Shinbrot [7, 1951] proposed a "method–function" approach in which the equations of motion were multiplied by special functions to eliminate errors due to finite data lengths and unknown initial conditions in using the Fourier transform approach. The idea seems similar to the instrumental variable approach of Reiersøl [7] in which special matrices are used to obtain consistent estimates. The later work in aircraft parameter identification uses nonlinear regressions techniques variously called quasilinearization [8], modified Newton–Raphson [9] and differential correction. These techniques also apply to nonlinear models. Most recently [10, 11, 12] maximum likelihood and Bayesian methods have been used for parameter estimation in models of the type

$$A\dot{x} = Fx + Gu + w \tag{4}$$

$$y = Hx + Du + v \tag{5}$$

where $w(t)$ is an uncorrelated or "white noise" Gaussian process.

At this stage, it is appropriate perhaps to say a few more things about the model of equations (4) and (5). This model, known as the "state-vector model" of the system, has assumed central importance in the control literature following the pioneering work of Kalman [13, 1960] on the filtering, prediction and control properties of this model. More recently, the structural and identifiability properties of this model have been studied [14, 15]. The discrete-time equivalent of this model bears close resemblance to the "simultaneous-equation" model of econometrics [16].

$$\begin{array}{lr} \text{(6)} & \\ & \\ \text{(7)} & \end{array} \quad \begin{array}{c}\text{Discrete-time}\\ \text{state-vector}\\ \text{model}\end{array} \left\{ \begin{array}{l} Ax(t+1) = Fx(t) + Gu(t) + w(t) \\ \\ y(t) = Hx(t) + Du(t) + v(t) \end{array} \right.$$

where $A(n \times n)$, $x(n \times 1)$, $F(n \times n)$, $G(n \times m)$, $D(p \times m)$, $u(m \times 1)$, $w(n \times 1)$, $y(p \times 1)$, $H(p \times n)$, $v(p \times 1)$ and

$$E[w(t)w^T(s)] = Q\delta_{t,s}$$

$$E[v(t)v^T(s)] = R\delta_{t,s}.$$

We assume that $p \leq n$, w and v uncorrelated.

$$\text{(8)} \quad \begin{array}{c}\text{Simultaneous}\\ \text{equations}\\ \text{model}\end{array} \left\{ \Gamma_0 y(t) = -\sum_{i=1}^{k} \Gamma_i y(t-i) + \sum_{i=0}^{l} B_i u(t-i) + e(t) \right.$$

where $y(p \times 1)$, $\Gamma_i(p \times p)$, $B_i(p \times m)$, $u(m \times 1)$, $e(p \times 1)$ and

$$E[e(t)e^T(s)] = E(t - s).$$

To see the similarities between these models, we use the lag operator z defined by

$$zy(t) = y(t - 1). \tag{9}$$

From equation (6),

$$(z^{-1}A - F)x(t) = Gu(t) + w(t).$$

From equation (7),

$$y(t) = H(z^{-1}A - F)^{-1}[Gu(t) + w(t)] + v(t) + Du(t). \tag{10}$$

From equation (8),

$$y(t) = \left(\sum_{i=0}^{k} \Gamma_i z^i\right)^{-1}\left[\left(\sum_{i=0}^{l} B_i z^i\right)u(t) + e(t)\right] \tag{11}$$

Equations (10) and (11) are same if we set

$$D + H(z^{-1}A - F)^{-1}G = \left(\sum_{i=0}^{k} \Gamma_i z^i\right)^{-1}\left(\sum_{i=0}^{l} B_i z^i\right) \tag{12}$$

and equate the spectral density functions or autocorrelation functions of

$$[H(z^{-1}A - F)^{-1}w(t) + v(t)]$$

and

$$\left(\sum_{i=0}^{k} \Gamma_0 z^i\right)^{-1} e(t).$$

The problem of obtaining (F, G, H, Q, R) from $\{\Gamma_i, B_i, E(t - s)\}$ has received attention in the control literature and is known as the stochastic realization problem [17, 18].[1]

Now let us consider some special cases which will bring out the similarity of the models (6)–(7) and (8) more clearly.

(i) *Complete State Vector Observed without Error (Perfect Measurements Case)*

In this case, $y(t) = x(t)$ so that equation (6) becomes

$$Ay(t + 1) = Fy(t) + Gu(t) + w(t). \tag{13}$$

This is a simultaneous equation model with "predetermined variables" consisting on one lag endogeneous variables $y(t)$, exogeneous variables $u(t)$ and uncorrelated

[1] In most of the control models, matrix A either turns out to be or can easily be reduced to an identity matrix. The "reduced form" state-vector model thus obtained can be given direct physical interpretation so that there is very little advantage in using the "simultaneous equation" state-vector model. The situation in econometrics is different since *a priori* information on parameters in the simultaneous equation model is not easily translated to the reduced form model.

stochastic errors $w(t)$. The time-index, t on u and w can be changed to $(t+1)$ without effecting the model.

(ii) *No Stochastic Disturbances in the State Equations* (*Zero Process Noise Case*)

With $w(t) = 0$, equation (10) can be written as

$$y(t) = H(z^{-1}A - F)^{-1}Gu(t) + v(t). \tag{14}$$

Assuming that the eigenvalues of $A^{-1}F$ lie inside the unit circle, we can perform a Laurent series expansion in z and obtain

$$y(t) = \sum_{i=0}^{\infty} M_i u(t-i) + v(t). \tag{15}$$

Equation (15) is a distributed lag model with lag coefficients M_i. In the control literature, M_i have been variously called Markov parameters, impulse response function, weighting pattern etc. The problem of obtaining matrices $\{F, G, H\}$ given $\{M_i,\ i = 0, \ldots, \infty\}$, assuming $A = I$ known as the minimal realization problem was solved by Ho and Kalman [19, 1966]. The concepts of controllability, observability and minimal realizations [20] play an important role in solving this problem. In general, one does not obtain unique $\{F, G, H\}$, but by imposing structural restrictions, it is possible to obtain unique $\{F, G, H\}$. In this way, one obtains unique canonical forms for the system which also have the property of containing the smallest numbers of unknown parameters. The extensions of these results to the process noise case are also available and will be discussed later.

(iii) *Role of Process and Measurement Noise*

It is seen from the above discussion that in econometric models only one noise term is present, which, however, can be correlated in time. The question arises: Is there any advantage of separating total noise into two parts? The significance of this in the control problem derives from the fact that in many situations one has sufficient *a priori* knowledge on the characteristics of $w(t)$ and $v(t)$ separately. For example, $v(t)$ being measurement noise, comes from the measuring instruments which can be separately calibrated. On the other hand, separate identification of $w(t)$ and $v(t)$ (whenever possible) provides much valuable information which is lost if only a combination of the two is identified. The situation is somewhat similar to the use of the simultaneous-equation model versus the reduced form model for parameter estimation in econometrics.

The use of both measurement noise and process noise in econometric models has certain applications. We mention two of these

(a) *Error in variables.* The state vector model of equations (6) and (7) allows one to consider errors in output or endogeneous variables. The errors in exogeneous variables $u(t)$ may be considered indirectly by adding them to process-noise $w(t)$. The questions of identifiability will be considered in Section 3.

(b) *Random coefficients in regression models.* Consider the scalar regression model

$$y(t) = \sum_{i=1}^{k} a_i y(t-i) + \sum_{i=0}^{l} b_i u(t-i) + v(t). \tag{16}$$

The usual assumption in regression models is that $(a_i, b_j, i = 1, -- k, j = 0, -- l)$ are constant. Suppose a_i, b_j are known to vary from one time-period to next and let the increments be random, e.g.

$$a_i(t + 1) = a_i(t) + w_i(t).$$

Denote by $x(t)$ the $(k + l + 1)$ vector of coefficients a_i, b_j at time t. Then

$$x(t + 1) = x(t) + w(t) \tag{17}$$

and

$$y(t) = H(t)x(t) + v(t) \tag{18}$$

where $H(t) = [y(t - 1), --, y(t - k), u(t), u(t - 1), -- u(t - l)]$. The case where $H(t)$ is deterministic[2] but time-varying has been considered extensively in the control literature. The case of $H(t)$ random has received less attention and needs to be further investigated [21].

(iv) *Correlated Errors and Colored Noise*

In equations (6) and (7), $w(t)$ and $v(t)$ were assumed to be uncorrelated in time and with each other. The correlation between $w(t)$ and $v(t)$ is easily handled by a transformation approach in which the system (6)–(7) is replaced by another system of equations having uncorrelated $w(t)$ and $v(t)$. If $E[w(t)v^T(s)] = C(t)\delta_{t,s}$ then the equivalent system is

$$\begin{aligned} Ax(t + 1) &= (F - CR^{-1}H)x(t) + Gu(t) + CR^{-1}y(t) + \eta(t) \\ y(t) &= Hx(t) + v(t) \end{aligned} \tag{19}$$

where

$$\begin{aligned} E[\eta(t)v^T(s)] &= 0, \\ E[\eta(t)\eta^T(s)] &= (Q - CR^{-1}C^T)\delta_{t,u}. \end{aligned} \tag{20}$$

The auto-correlation of $w(t)$ and $v(t)$ is handled by representing them as white noise through a linear system and augmenting the state vector.[3] For example, let $w(t)$ be represented as

$$\begin{aligned} x_w(t + 1) &= F_w x_w(t) + G_w \varepsilon(t) \\ w(t) &= H_w x_w(t) + \xi(t) \end{aligned} \tag{21}$$

where $x_w(t)$ is the state vector for representing $w(t)$; $\varepsilon(t)$ and $\xi(t)$ are white noise processes. Then equations (21) and (22) can be combined with (6) and (7) by using an augmented state vector

$$x_A = \begin{bmatrix} x \\ x_w \end{bmatrix}. \tag{22}$$

Notice that even if the state vector x is completely observed, the augmented state vector x_A is only partially observed.

[2] This would be the case if $a_i = 0, i = 1, -- k$.

[3] This can always be done for stationary processes with proper rational spectra.

(v) *Recursive versus Simultaneous Equation Models*

In the econometric literature, there has been a lot of discussion about the use of recursive versus simultaneous equation models. Wold [28] has maintained that the real world economic systems are recursive in time (or causal chains) and that simultaneous equation models are approximations based on neglecting fast time-constant phenomenon that are unobservable due to the large sampling interval. Liu [29] has emphasized the interdependence of economic variables and has questioned whether the complete interdependence and simultaneity is properly considered in the proposed econometric models such as the Klein–Goldberger model [30]. Fisher [31] has taken an intermediate position and shown that the econometric models in use may be thought of as approximations to reality as conjectured by Wold [28] or Liu [29].

In the control literature, questions of this type have not received much attention. A control engineer, by training, is accustomed to thinking of the world as recursive or causal. The concepts of state and Markov models of a system are partly based on this notion of reality. The recent interest in large scale systems and model-simplification techniques has led to the use of aggregation concepts [32] and asymptotic expansion methods [33]. Some of these methods lead to a set of simultaneous equations corresponding to small time-constants in the system. The use of simultaneous equation models in control is an area for further research.

(vi) *Estimation Techniques*

A large amount of the work in control is concerned with estimation techniques as is evidenced by several survey papers including Astrom and Eykhoff [22]. The lack of emphasis on model structure determination and identifiability may be due to the fact that engineering models are fairly well understood and in single-input single-output models, the identifiability conditions are not very complicated. By and large, the models that have been considered in control applications are identifiable. The main concern in control has been in devising efficient computational methods for parameter estimation. In many applications, the estimation has to be done on-line and this rules out iterative methods like the full-information maximum likelihood method. Other differences in estimation methods arise from the fact that only part of the state vector is observed and the observations contain measurement noise. Under these conditions, a direct application of ordinary least squares (OLS) leads to biased estimates. The approach taken in control is to go to nonlinear least-squares rather than modify OLS as is often done in econometrics. It is fair to say that the control literature has not made full use of linear least squares techniques. On the other hand, much valuable experience has been gathered on nonlinear optimization techniques applied to least squares and maximum likelihood criteria. A combination of the experience gained in the two fields should certainly be fruitful. Two examples of this are refs. [1] and [23].

3. Identification and Estimation of State Vector Models

In this section, we present some known and some new results on the identification and maximum likelihood estimation of state-vector models. It is hoped that these results would find applications in the econometric literature.

For simplicity, we would consider the reduced form state-vector model by assuming that $A = I$ and $D = 0$. The extensions of the identifiability results given below to the case $A \neq I$ need to be worked out. From here on, we consider the simplified model

$$x(t + 1) = Fx(t) + Gu(t) + w(t) \tag{23}$$

$$y(t) = Hx(t) + v(t). \tag{24}$$

3.1. *Identifiability Results*

From the work of Aström [22], Kalman [20], Mehra [25, 26], Kailath [27], Mayne [14], Popov [15] and Tse Weinert, *et al.* [29], it is known that the following conditions must be imposed on the model (23)–(24) to make it identifiable.

Condition (*i*). $[F, G]$ controllable and $[F, H]$ observable, i.e.

$$\text{Rank } [G, FG, F^2G, --, F^{n-1}G] = n \tag{25}$$

$$\text{Rank } [H^T, F^TH^T, \ldots, (F^T)^{n-1}H^T] = n. \tag{26}$$

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 identifiable iff all the state variables are measured, i.e. $p = n$. If $p < n$, only the proper canonical representation of (23)–(24) given by a steady-state Kalman filter is identifiable [25–27]. This representation has the form

$$\hat{x}(t + 1) = F\hat{x}(t) + Gu(t) + FKv(t) \tag{27}$$

$$y(t) = Hx(t) + v(t) \tag{28}$$

where $E[v(t)v^T(s)] = \Sigma\delta_{t,s}$. The white noise process $v(t)$ represents one-step ahead prediction errors and is also known as the "innovation" process.

It is further required that $[F, K]$ be controllable.

Condition (*iii*). F and H are in the following canonical forms or their parameters are consistently solvable in terms of the parameters of this canonical form.

$$F_c = \begin{bmatrix} F_{11} & & \\ F_{21} & F_{22} & 0 \\ & & \ddots \\ F_{p1} & ------- & F_{pp} \end{bmatrix}, \quad \begin{matrix} F_{ij} = \\ (i \neq j) \end{matrix} \begin{bmatrix} 0 \\ ----------- \\ \beta_{1ij} \quad --- \quad \beta_{n_jij} \end{bmatrix}$$

$$F_{ii} = \left[\begin{array}{c|c} & I \\ 0 & \\ \hline \beta_{1ii} \; --- & \beta_{n_iii} \end{array}\right]$$

where F_{ij} is $n_i \times n_j$

(30) $$n_i = \text{Rank}\begin{bmatrix} H_i \\ H_iF \\ \vdots \\ H_iF^{n-1} \end{bmatrix} - \text{Rank}\begin{bmatrix} H_{i-1} \\ H_{i-1}F \\ \vdots \\ H_{i-1}F^{n-1} \end{bmatrix}$$

where H_i is $i \times n$ matrix consisting of the first i rows of H. n_1 represents the additional part of the state space observed by the ith output over the first $(i - 1)$ outputs. The set of indices $(n_1, n_2, -- n_p)$ have been called output numbers of the system by Mayne [14] and are invariant under coordinate transformations. From the observability condition,

$$\sum_{i=1}^{p} n_i = n.$$

The canonical form of H assuming none of the output numbers are zero is

(31) $$H_C = \left.\begin{bmatrix} \underbrace{\begin{matrix} 0 & 0 & . & 1 \\ 0 & 0 & . & . \\ 0 & 0 & . & . \\ . & . & . & . \end{matrix}}_{n_1 \text{ columns}} & \vdots & \underbrace{\begin{matrix} 0 & 0 & . & 0 \\ 0 & . & 0 & 1 \\ 0 & . & . & . \\ . & . & . & . \end{matrix}}_{n_2 \text{ columns}} & \vdots & \underbrace{\begin{matrix} . & . & . \\ . & . & . \\ . & . & 1 \\ . & . & . \end{matrix}}_{n_3} & \vdots & \begin{matrix} . \\ . \\ . \\ . \end{matrix} \end{bmatrix}\right\} p \text{ rows.}$$

If $n_i = 0$, then the ith row of H has non-zero entries in the first $(n_1 + -- + n_{i-1})$ columns.

The total number of unknown parameters in F_C and H_C is $s = n_1 + (n_1 + n_2) + -- + (n_1 + -- + n_p)$ which is less than np. An observable system can be put into the above canonical form by using the state transformation matrix

(32) $$T = \begin{bmatrix} h_1 \\ h_1F \\ \vdots \\ h_1F^{n_1-1} \\ h_2 \\ \vdots \\ h_2F^{n_2-1} \\ \vdots \\ h_pF^{n_p-1} \end{bmatrix}, \qquad H = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}$$

Let $x' = T\hat{x}$.

Using equations (27)–(28),

$$x'(t+1) = TFT^{-1}x'(t) + TGu(t) + TFKv(t) \tag{33}$$

$$y(t) = HT^{-1}x'(t) + v(t). \tag{34}$$

It can be shown that [39]

$$TFT^{-1} = F_c \tag{35}$$

$$HT^{-1} = H_c. \tag{36}$$

Also let

$$TG = G_c, \qquad TK = K_c. \tag{37}$$

Equations (33)–(37) represent the canonical model of the system. By using equation (32) for transformation T, we can relate the parameters of F, G, H, K to the parameters of F_c, H_c, G_c and K_c which are identifiable. The canonical set has a total of $(s + nm + np) \leq n(m + 2p)$ parameters. In addition, the covariance matrix of the innovation process $v(t)$ viz Σ is identifiable. The matrices Q and R are related to Σ, H, F, G, K by the following matrix equations.

$$\Sigma = H_c P_c H_c^T + R \tag{38}$$

$$K_c = P_c H_c^T \Sigma^{-1} \tag{39}$$

$$\dot{P}_c = F_c(I - K_c H_c)P_c F_c^T + Q_c \tag{40}$$

where P is an $n \times n$ positive definite matrix. Methods for solving equations (38)–(40) have been discussed by Faurre [17] and Mehra [34, 35].

Condition (*iv*).[4] The support of the spectral distribution function $S_{uu}(\omega)$ of the input $u(t)$ contains more than $k = [NP/2p]$ points where NP is the number of unknown parameters in F, G, and H and $[a]$ denotes the integer part of a. The support of $S_{uu}(\omega)$ is defined as

$$\text{Support } S_{uu}(\omega) = \{\omega \mid -\pi < \omega \leq \pi, \forall\varepsilon > 0, [S_{uu}(\omega + \varepsilon) - S_{uu}(\omega - \varepsilon)] > 0\}. \tag{41}$$

This condition is derived in Ref. [26] 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 [22].

Condition (*v*). F is a stable matrix or all the roots of $(z^{-1}I - F)$ lie outside the unit circle.

Remarks

(a) The Kalman Filter representation (27)–(28) of the system is both causal and causally invertible with respect to the input–output pair $v(t)$ and $y(t)$. One can write using the lag operator z, and using equations (27)–(28),

$$y(t) = H(z^{-1}I - F)^{-1}Gu(t) + [H(z^{-1}I - F)^{-1}K + I]v(t) \tag{42}$$

[4] 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.

and

$$(43)\, v(t) = \{I - H[z^{-1}I - F(I - KH)]^{-1}FK\}y(t) - H[z^{-1}I - F(I - KH)]^{-1}Gu(t).$$

From (43), it is clear that $F(I - KH)$ must also be a stable matrix. Kalman [13] has shown that under the conditions of complete controllability and observability and K given by equations (38)–(40), $F(I - KH)$ is a stable matrix. If K is identified directly, this condition must be imposed separately.

(b) In the terminology of econometrics [16], if the number of unknown parameters θ in the original system $\{F, G, H, Q, R\}$ is exactly the same as the number of parameters ϕ in the canonical representation $\{F_c, G_c, H_c, K_c, \Sigma\}$, and the mapping from the sets Φ to Θ is one-to-one onto ($\theta \in \Theta$, $\phi \in \Phi$), the system is *just-identified*. If the dimension of θ is less than that of ϕ, and the mapping from Φ to Θ is onto, the system is *over-identified*. Finally if the dimension of θ is larger than that of ϕ or the mapping from Φ to Θ is not onto, the system is *under-identified* or *unidentifiable*.

(c) Rothenberg [40] has shown that a necessary and sufficient condition for local identifiability of θ to θ_0, under certain regularity conditions, is that the Fisher Information matrix be nonsingular at θ_0. However, this condition is not easy to verify in practice, except numerically. The author has not as yet related other conditions of identifiability given by Rothenberg [40] to the above conditions.

3.2. *Consistent Least Squares Estimation of Canonical Parameters and Determination of $n_1, \ldots, n_p$*

Equations (27)–(28) along with canonical forms F_c and H_c can be written as a set of p difference equations in terms of the input–output variables (y, u, v). The resulting form is known as the external model of the system. Once the system equations are written in the external form, the applicability of regression methods can be easily examined. We demonstrate this by considering the equation for the first output variable $y_1(t)$. (The elements of vectors and matrices will be denoted by subscripts.)

$$(44) \qquad y_1(t) = x_{n_1}(t) + v_1(t)$$

$$= \sum_{j=1}^{n_1} \beta_{11_j} x_j(t-1) + \sum_{j=1}^{m} (G_c)_{n_{1j}} u_j(t-1)$$

$$+ \sum_{j=1}^{p} (FK)_{n_{1j}} v_j(t-1) + v_1(t)$$

$$(45)\quad x_j(t-1) = x_{j+1}(t-2) + \sum_{i=1}^{m} (G_c)_{ji} u_i(t-2) + \sum_{i=1}^{b} (FK)_{ji} v_i(t-2).$$

Solving equation (45) recursively in terms of $x_{n_i}(t-2), --- x_{n_1}(t - n_1)$ and substituting in terms of $y_1(t-2), -- y_1(t-n)$, we get an equation of the type

$$(46) \qquad y_1(t) = \sum_{j=1}^{n_1} \beta_{11_j} y_1(t-j) + \sum_{j=1}^{n_1} \sum_{j=1}^{m} g_{j_i} u_i(t-j) + \sum_{j=0}^{n_1} \sum_{i=1}^{p} c_{ji} v_i(t-j)$$

where g_{ji} and c_{ji} are defined by correspondence. Equation (46) represents an autoregressive moving average (ARMA) model of order n_1 with m deterministic inputs and p white noise inputs.

Proceeding in the same fashion, $y_2(t)$ can be written as an autoregressive model of order n_2 with $y_1(t)$, $u(t)$ and $v(t)$ as inputs and so on. In other words, this system of equations is recursive and we can estimate the parameters by solving p regression problems in sequence. However, since the error terms are correlated with some of the independent variables, OLS would give biased and inconsistent estimates. We use the two-stage least-squares procedure of Theil [16] to obtain consistent estimates of β and g parameters. The c parameters can then be obtained fitting a moving average model to the residuals. It is also possible to use the three-stage least squares procedure [16], but we would instead use the maximum likelihood procedure of Appendix A for obtaining efficient estimates of all the parameters.

Two-Stage Least-Squares Estimation of Parameters

Define

$$\mathbf{y} = \begin{bmatrix} y_1(n_1) \\ \vdots \\ y_1(N) \end{bmatrix}$$

where N is the total number of sample points. Equation (46) may now be written for $t = n_1, --- N$ as

$$\mathbf{y} = [M_1 \vdots M_2]\begin{bmatrix} \theta_1 \\ \hline \theta_2 \end{bmatrix} + \mathbf{v} \tag{47}$$

where

$$\underset{(N-n_1)xn_1m}{M_1 =} \left[\begin{array}{c|c|c|c} u_1(n_1-1), \ldots, u_1(0) & u_2(n-1), \ldots, u_2(0) & \cdots & u_m(n_1-1), \ldots, u(0) \\ u_1(n_1), \ldots, u_1(1) & u_2(n_1), \ldots, u_2(0) & \cdots & u_m(n_1), \ldots, u_m(1) \\ \vdots & & & \\ u_1(N-1), \ldots, u_1(N-n_1) & u_2(N-1), \ldots, u_2(N-n) & \cdots & u_m(N-1), \ldots, u_m(N-n_1) \end{array}\right]$$

$$\underset{(N-n_1)xn_1}{M_2 =} \begin{bmatrix} y(n_1-1), \ldots, y(0) \\ y(n_1), \ldots, y(1) \\ \vdots \\ y(N-1), \ldots, y(N-n_1) \end{bmatrix}$$

$$\underset{mn_1 x1}{\theta_1 =} \begin{bmatrix} g_{11} \\ \vdots \\ g_{n_1 12} \\ \vdots \\ g_{12} \\ \vdots \\ g_{n_1 2} \\ \vdots \\ g_{n_1 m} \end{bmatrix}$$

$$\underset{(N - n_1)x1}{\mathbf{v} =} \begin{bmatrix} \sum_{j=0}^{n_1} \sum_{i=1}^{p} c_{ji} v_i(n_1 - j) \\ \vdots \\ \sum_{j=0}^{n_1} \sum_{i=1}^{p} c_{ji} v_i(N - j) \end{bmatrix}, \quad \underset{n_1 x1}{\theta_2 =} \begin{bmatrix} \beta_{111} \\ \vdots \\ \beta_{11} n_1 \end{bmatrix}.$$

Equation (47) can also be written as

$$\mathbf{y} = M\theta + \mathbf{v} \tag{48}$$

where $M = [M_1 \vdots M_2]$ and $\theta = \begin{bmatrix} \theta_1 \\ \hline \theta_2 \end{bmatrix}$. Notice that only M_2 and $\mathbf{v}$ are correlated in equation (47). In two stage least squares, one replaces M_2 by $\hat{M}_2$ such that $\hat{M}_2$ is uncorrelated with the new error term. The calculation of $\hat{M}_2$ is done by using the final form of the model (or the impulse response model) of equation (46) viz.

$$y_1(t) = \sum_{j=1}^{\infty} \sum_{i=1}^{m} \gamma_{ji} u_i(t - j) + \eta_1(t) \tag{49}$$

where γ_{ji} is obtained from

$$\left(1 + \sum_{\gamma=1}^{n_1} \beta_{11j} z^j\right)^{-1} \left(\sum_{j=1}^{n_1} \sum_{i=1}^{m} g_{ji} z^j\right) = \sum_{\gamma=1}^{\infty} \sum_{i=1}^{m} \gamma_{ji} z^j.$$

Using the stability property of the system, we will truncate the sum in equation (49) at an appropriately large value $q < (N - n_1)$. Then equation (49) for $t = n_1, \ldots, N$, can be written

$$\mathbf{y} = U\gamma + \boldsymbol{\eta} \tag{51}$$

where

$$\underset{(N - n_1) \times qm}{U =} \begin{bmatrix} u_1(n_1 - 1) & \ldots & u(0), 0, \ldots, 0 & \vdots \ldots \vdots & u_m(n_1 - 1) & \ldots & u_m(0), 0, \ldots, 0 \\ u_1(N - 1) & \ldots & u_1(N - q) & \vdots \ldots \vdots & u_m(N - 1) & \ldots & u_m(N - q) \end{bmatrix}$$

γ and $\boldsymbol{\eta}$ are easily defined by correspondence.

The model (51) has the property that U and $\boldsymbol{\eta}$ are uncorrelated and a consistent estimate of γ is obtained by OLS.

$$\hat{\gamma} = (U^T U)^{-1} U^T \mathbf{y}.$$

But $\hat{\gamma}$ is not BLUE (best linear unbiased estimator) since $E[\boldsymbol{\eta}\boldsymbol{\eta}^T] \neq \sigma^2 I$. We use it to obtain $\hat{\mathbf{y}}$ which denotes the part of $\mathbf{y}$ that is linearly correlated with U, i.e.

$$\hat{\mathbf{y}} = U\hat{\gamma}. \tag{52}$$

We now write

$$\mathbf{y} = [M_1 \vdots \hat{M}_2]\begin{bmatrix} \theta_1 \\ \hline \theta_2 \end{bmatrix} + \mathbf{v} + \boldsymbol{\varepsilon} \tag{53}$$

where

$$\hat{M}_2 = \begin{bmatrix} \hat{y}_1(n_1 - 1) & \ldots & \hat{y}(0) \\ \hat{y}_1(n_1) & \ldots & \hat{y}(1) \\ & \vdots & \\ \hat{y}(N-1) & \ldots & \hat{y}(N - n_1) \end{bmatrix}$$

and $\boldsymbol{\varepsilon} = (M_2 - \hat{M}_2)\theta_2$. It is easily shown that $(\mathbf{v} + \boldsymbol{\varepsilon})$ is uncorrelated with $\hat{M}_2$ so that one can use OLS to obtain consistent estimates of θ.[5]

$$\begin{bmatrix} \theta_2 \\ \hline \hat{\theta}_2 \end{bmatrix} = \begin{bmatrix} M_1^T M_1 & M_1^T \hat{M}_2 \\ \hat{M}_2^T M_1 & \hat{M}_2^T M_2 \end{bmatrix}^{-1} \begin{bmatrix} M_1 \\ \hat{M}_2 \end{bmatrix} \mathbf{y}. \tag{54}$$

Estimator (54) can also be written as

$$\begin{bmatrix} \theta_1 \\ \hline \hat{\theta}_2 \end{bmatrix} = \begin{bmatrix} M_1^T M_1 & M_1^T M_2 \\ M_2^T M_1 & M_2^T M_2 - kW^T W \end{bmatrix} \begin{bmatrix} M_1^T \\ (M_2 - W)^T \end{bmatrix} \mathbf{y} \tag{55}$$

with $k = 1$ and $W = (M_2 - \hat{M}_2)$.

The estimator (55) is called the k-class estimator. It reduces to OLS for $k = 0$ and can be shown to be related to the maximum-likelihood estimator [16].

Estimation of $n_1 \ldots n_p$

Model (46) after replacement of $y_1(t - j)$ on the right-hand side by $\hat{y}_1(t - j)$ is in a form suitable for using step-wise regression [36] as proposed by Parzen [37] in a somewhat different context. In this procedure, the significance of various regression terms is tested by using partial correlations and partial F-tests. Other statistical criteria, such as Akaike's FPE (final prediction error) [38] can also be used depending on the objective of identification (i.e. prediction, control, etc.).

Remarks

(a) The case in which there is no deterministic input $u(t)$, the two-stage least squares procedure cannot be used. However, in that case, modified Yule–Walker

[5] The estimates are efficient only if $c_{ji} = 0$. $\forall\, j, i$.

equations can be used to obtain consistent estimates of β parameters [25]. More efficient estimates can be obtained by using Durbin's method [51] or the Hannan–Parzen approach [52, 37].

(b) The consistent estimates of c_{ji} or matrix K can be obtained in several ways. See, e.g. Refs. [34, 23, 39].

(c) The two-stage least-squares approach can be made on-line as has already been demonstrated by Pandya [23]. In fact, most of the linear least-squares procedures can be made recursive by expressing the inverse of the information matrix for $(N + 1)$ measurements in terms of the inverse for N measurements using the matrix inversion lemma

$$(M^{-1} + H^T R^{-1} H)^{-1} \equiv M - MH^T(HMH^T + R)^{-1}HM.$$

4. Further Comments

In this section, we discuss two problems which have received considerable attention in the control literature.

4.1. *Closed-Loop or Feedback Systems*

Closed-loop systems are of interest to both control engineers and econometricians. A typical closed-loop system is shown in Figure 2. In particular situations, only some of the external inputs shown may be present. A number of interesting results are available in the control literature on closed-loop system identification. It was shown by E. Fisher [41, 1965] that in the deterministic case (i.e., no stochastic inputs), no external input and linear system dynamics, the system is unidentifiable with linear feedback, but is identifiable with nonlinear feedback. It is interesting to contrast this result with that of Reiersøl [42, 1950] who showed that the error-in-variables simultaneous equation model of econometrics is unidentifiable with normal errors, but is identifiable with nonnormal errors.

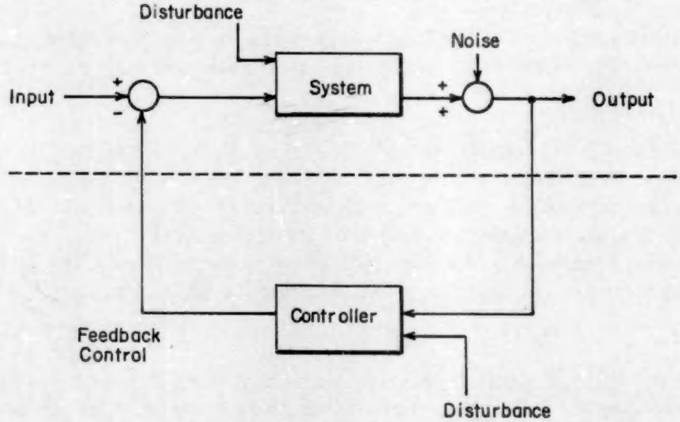


Figure 2 Block diagram of a feedback control system

In stochastic systems, interesting results on identifiability have been given by Wingrove [43], Box and MacGregor [44] and Phadke and Wu [45] for the case where both the system and controller dynamics are unknown. They have shown that without the external input or disturbances into the system, the closed-loop system (i.e., the system and the controller) may be unidentifiable. If one simply cross-correlates the input sequence and the output sequence of the system (after pre-whitening), one obtains the inverse of the controller transfer function rather than the system transfer function. Consequently, it is necessary to have an external input or disturbance into the system that is uncorrelated with output noise in order to identify the system. In addition, it is also necessary to have some time-delay or dead-time in the loop to make the system completely identifiable [43, 44, 45].

An area of active interest where closed-loop identification is essential and has been extensively used for the last 15 years is human operator modeling [46]. Most of the work has used spectral methods, but recently maximum likelihood estimation and parametric models have been used with good success [47].

4.2. *Input Design*

The problem of input design has received considerable attention in the control literature due to the fact that inputs can often be selected and they can have considerable influence on the accuracy of parameter estimation. This problem has been formulated in a number of different ways which include

(i) an optimal control formulation [48, 49], and

(ii) a minimax approached based on the theory of optimal experiments in regression [26, 50].

The latter approach has given very general results on the design of optimal inputs. The two approaches have been applied to the design of control inputs for aircraft parameter identification.

5. Conclusions

In this report, we have pointed out certain similarities and differences between system identification in the control and econometric literatures. In particular, the state-vector model commonly used in control is compared with the simultaneous equations model used in econometrics. An approach to the identification and estimation of parameters in state vector models is presented based on canonical forms, stepwise two-stage least squares and maximum likelihood estimation using a Kalman filter. The problems of closed-loop identification and input design are also briefly discussed.

Harvard University

References

[1] P. J. Dhrymes, L. R. Klein, and K. Steiglitz, "Estimation of Distributed Lags," *International Economic Review*, Vol. II, No. 2, June 1970.

[2] N. Weiner, "Generalized Harmonic Analysis," *Acta Math.*, 55, 1930, pp. 117–258

[3] H. Greenberg, "A Survey of Methods for Determining Stability Parameters of an Airplane from Dynamic Flight Measurements," Tech. Note 2340, NACA, 1951.

[4] D. W. Jorgensen, "Rational Distributed Lag Functions," *Econometrica*, XXXIV, January 1966.
[5] P. J. Dhrymes, *Distributed Lags, Problems of Estimation and Formulation*, Holden-Day, Inc., 1971.
[6] M. Shinbrot, "A Least Squares Curve Fitting Method with Applications to the Calculation of Stability Coefficients from Transient Response Data," Tech. Note 2341, NACA, April 1951.
[7] O. Reiersøl, "Confluence Analysis by Means of Instrumental Sets of Variables," *Arkiv for Mathematik, Astronimi och Fysik*, 32A (4), 1941.
[8] R. Bellman and R. Kalaba, *Quasilinearization and Nonlinear Boundary-value Problems*, American Elsevier Publishing Co., New York, 1965.
[9] L. W. Taylor, K. W. Iliff, and B. G. Powers, "A Comparison of Newton–Raphson and Other Methods for Determining Stability Derivatives from Flight Data," AIAA Third Flight Test, Simulation and Support Conf., March 1969.
[10] R. K. Mehra, "Maximum Likelihood Identification of Aircraft Parameters," Preprints 1970 Joint Aut. Cont. Conf., Atlanta, Georgia, June 1970.
[11] R. K. Mehra, D. E. Stepner, and J. S. Tyler, "Generalized Method for the Determination of Aircraft Stability and Control Derivatives from Flight Test Data," Preprints 1972 Joint Aut. Cont. Conf., Stanford, California, 1972.
[12] A. V. Balakrishnan, —'Identification and Adaptive Control: An Application to Flight Control Systems," *Journal of Optimization Theory and Applications*, March 1972.
[13] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Trans. ASME, J. Basic Eng.*, Series D, Vol. 82, March 1960.
[14] D. Mayne, "A Canonical Model for Identification of Multivariable Linear Systems," *IEEE Trans. Auto. Cont.*, Vol. AC-17, pp. 728–729, 1972.
[15] V. Popov, "Invariant Description of Linear Time-invariant Controllable Systems," *SIAM J. Control*, Vol. 10, pp. 252–264, 1972.
[16] H. Theil, *Principles of Econometrics*, John Wiley & Sons, Inc. 1971.
[17] P. Faurre and J. P. Marmorat, "Une Algorithm de Realization Stochastique," *C.R. Acad. Sci.*, Vol. 268, April 1969.
[18] R. Rissanen and T. Kailath, "Partial Realization of Random Systems," *Automatic*, Vol. 8, pp. 389–396, 1972.
[19] B. L. Ho and R. E. Kalman, "Effective Construction of Linear State-Variable Models from Input/Output Functions," Proc. Third Allerton Conf., October 1965.
[20] R. Kalman, P. Falb, and M. Arbib, *Topics on Mathematical System Theory*, McGraw-Hill, New York, 1969.
[21] S. S. L. Chang, "Optimum Filtering and Control of Randomly Sampled Systems," *IEE T-AC*, Vol. AC-12, October 1967.
[22] K. J. Aström and P. Eykhoff, "System Identification—A Survey," *Automatica* 7, pp. 123–162, 1971.
[23] R. N. Pandya, "A Class of Bootstrap Estimates for Identification of Linear Discrete Time Models," TR No. SE72-3, Carleton Univ., Ottawa, Canada, August 1972.
[24] A. E. Bryson and Y. C. Ho, *Applied Optimal Control*, Blaisdell, Mass., 1969.
[25] R. K. Mehra, "On-Line Identification of Linear Dynamic Systems with Applications to Kalman Filtering," IEEE T-AC, Vol. AC-16, No. 1, February 1971.
[26] R. K. Mehra, "Frequency Domain Synthesis of Optimal Inputs for Estimating Parameters in Linear Dynamic Systems," Harvard Tech. Report, July 1973 (to appear in *IEEE T-AC*, Dec. 1974).
[27] R. Geesey and T. Kailath, "Applications of the Canonical Representation to Estimation and Detection in Colored Noise," Symp on Computer Processing in Communications, Polytechnic Inst. of Brooklyn, 1969.
[28] H. O. Wold, Ed., *Econometric Model Building*, Chapter 1, North Holland Publ. Co., 1964.
[29] T. Liu, "Underidentification Structural Estimation and Forecasting," *Econometrica*, Vol. 28, 4, October 1960.
[30] L. R. Klein and A. S. Goldberger, *An Econometric Model of the United States, 1929–1952*, North-Holland Publ. Co., 1955.
[31] F. M. Fisher, "On the Cost of Approximate Specification in Simultaneous Equation Estimation," *Econometrica*, Vol. 29, April 1961.
[32] M. Aoki, "Control of Large-Scale Dynamic Systems by Aggregation," *IEEE T-AC*, June 1968.
[33] P. Sannuti and P. Kokotovic, "Near Optimum Design of Linear Systems by a Singular Perturbation Method," *Ibid*, February 1969.
[34] R. K. Mehra, "On the Identification of Variances and Adaptive Kalman Filtering," *Ibid.*, April 1970.
[35] R. K. Mehra, "An Algorithm to Solve Matrix Equation $PH^T = G$ and $P = \Phi P \Phi^T + \Gamma\Gamma^T$," *Ibid.*, October 1970.

[36] N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, Inc., 1966.
[37] E. Parzen, Time Series Analysis and System Identification, Lecture Notes, Regional Conf., Univ. of N. Carolina, January 1973.
[38] H. Akaike, "Statistical Predictor Identification," *Ann. Inst. Statistical Math.*, Vol. 22, 1970.
[39] E. Tse, H. Weinert, J. Anton, and R. Mehra, "Model Structure Determination and Identifiability Problems in System Identification," Final Report, Systems Control, Inc., February 1973.
[40] T. J. Rothenberg, "Identification in Parametric Models," *Econometrica*, Vol. 39, May 1971.
[41] E. Fisher, "The Identification of Linear Systems," *Proc. 1965 JACC*, pp. 473–475.
[42] O. Reiersøl, "Identifiability of a Linear Relation Between Variables which are Subject to Error," *Econometrica*, 18, 1950.
[43] R. Wingrove, "Comparison of Methods for Identifying Pilot Transfer Functions from Closed-Loop Operating Records," NASA TN D-6235, March 1971.
[44] C. E. P. Box and J. F. MacGregor, "The Analysis of Closed-Loop Dynamic–Stochastic Systems," TR 309, Dept. of Statistics, Univ. of Wisconsin, July 1972.
[45] M. S. Phadke and S. M. Wu, "Identification of Process Dynamics from Feedback Control Data—Its Application to Papermaking Process," Regional Conference on Time Series Analysis, Univ. of N. Carolina, January 1973.
[46] D. T. McRuer and H. R. Jex, "A Review of Quasi-Linear Pilot Models," *IEEE T-Human Factors Electronics*, HFE-8, No. 3, Sept. 1967.
[47] R. K. Mehra and J. S. Tyler, "Modeling the Human Operator Under Stress Conditions Using System Identification," Preprints 1972 JACC, Stanford, Calif., August 1972.
[48] M. Aoki and R. Staley, "On Input Signal Synthesis in Parameter Identification," *Automatica*, Vol. 6, 1970.
[49] R. K. Mehra, "Optimal Inputs for Linear System Identification," 1972 JACC, Stanford, Calif.
[50] B. Viort, "D-Optimal Designs for Dynamic Models," TR 314, Dept. of Statistics, Univ. of Wisconsin, October 1972.
[51] J. Durbin, "Efficient Estimators of Parameters in Moving Average Models," *Biometrika*, 46, 1959.
[52] E. J. Hannan, *Multiple Time Series Analysis*, John Wiley and Sons, Inc., 1970.

Appendix A: An Innovations Approach to Maximum Likelihood Identification of Linear and Nonlinear Dynamic Systems

This appendix presents an approach to maximum likelihood identification of multi-input multi-output linear and nonlinear dynamic systems with arbitrary inputs. The approach is based on state vector formulation and uses the innovation properties of optimal filters for these systems. Application to the identification of the transfer function of a chemical reactor is considered.

1. *Introduction*

The maximum likelihood estimation of autoregressive and moving average parameters in time series analysis has been considered by several investigators [1, 2].[6] The related problem of linear system identification can often be cast in this framework, though the parameter transformations involved may be nonlinear and nonunique. Special difficulties are encountered in handling multi-input multi-output linear models and nonlinear models using the time-series approach. The author [3, 4] has tried to circumvent these difficulties by working directly with the physical models and using the innovations approach of Kailath [5, 6]. A schematic diagram of this method is shown in Figure A.1.

[6] References for Appendix A are given separately at the end.

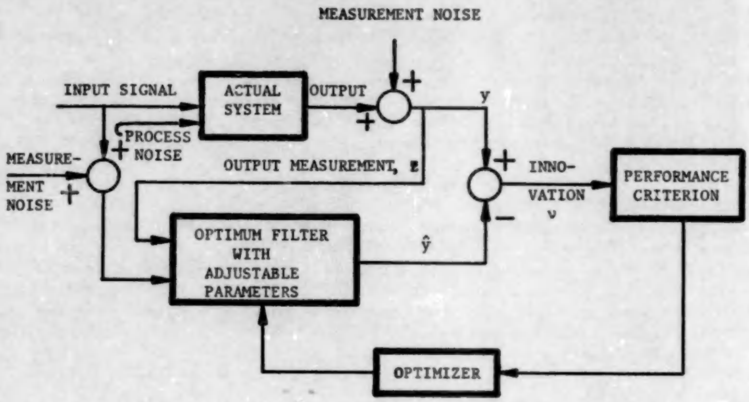


Figure A.1 Implementation of maximum likelihood estimator

2. *Linear Systems*

Consider a discrete-time linear system[7]

$$x(t+1) = Fx(t) + Gu(t) + \Gamma w(t) \tag{A.1}$$

$$y(t) = Hx(t) + v(t) \tag{A.2}$$

where

$x(t) = n \times 1$ state vector; $u(t) = p \times 1$ input vector;

$w(t) = q \times 1$ vector of random forcing functions;

$y(t) = r \times 1$ output vectors; and $v(t) = r \times 1$ vector of output errors

and

$$E\{w(t)\} = 0, \qquad E\{w(t)w^T(\tau)\} = Q\delta_{t,\tau}$$

where δ, is the Kronecker delta function.

$$E\{w(t)v^T(\tau)\} = 0$$

$$E\{v(t)\} = 0, \qquad E\{v(t)v^T(\tau)\} = R\delta_{t,\tau}$$

It is assumed that the structure of the model is known. The vector of unknown parameters from F, G, Γ, H, Q and R is denoted by θ. It is assumed that θ is identifiable.

The *ML* estimate of θ is given by

$$\hat{\theta} = \text{Arg}\,\{\max_{\theta} \log p(Y_N/\theta)\} \tag{A.3}$$

where

$$Y_N = \{y(1), \ldots, y(N)\}$$

[7] Continuous-time systems are handled in the same fashion. See Ref. 3.

and

$$p(Y_N/\theta) = \text{conditional probability density of } Y_N \text{ given } \theta.$$

An expression for $p(Y_N/\theta)$ is derived as

$$p(Y_N/\theta) = p(y(1), \ldots, y(N)/\theta)$$
$$= p(y(N)\,|\,Y_{N-1}, \theta)p(Y_{N-1}|\theta)$$
$$= p(y(N)\,|\,Y_{N-1}, \theta)p(y(N-1)\,|\,Y_{N-2}, \theta)p(Y_{N-2}|\theta)$$

$$= \sum_{j=1}^{N} p(y(j)\,|\,Y_{j-1}, \theta).$$

Therefore

$$\log p(Y_N|\theta) = \sum_{j=1}^{N} \log p(y(j)\,|\,Y_{j-1}, \theta) \tag{A.4}$$

Consider the case in which $x(0)$, $w(t)$ and $v(t)$ are normally distributed. Then $p(y(j)\,|\,Y_{j-1}, \theta)$ by a well-known property of normal distributions is also normal.

Let

$$E\{y(j)|\,Y_{j-1}, \theta\} = \hat{y}(j|j-1) \tag{A.5}$$

and

$$\text{Cov}\,\{y(j)\,|\,Y_{j-1}, \theta\} = B(j|j-1). \tag{A.6}$$

It is known that $\hat{y}(j|j-1)$ and $B(j|j-1)$ can be obtained from a Kalman filter [7] of the following form:

$$\hat{x}(t+1/t) = F\hat{x}(t/t) + Gu(t) \tag{A.7}$$

$$\hat{x}(t/t) = \hat{x}(t/t-1) + K(t)v(t) \tag{A.8}$$

$$v(t) = y(t) - H\hat{x}(t/t-1) \tag{A.9}$$

$$K(t) = P(t/t-1)H^TB^{-1}(t/t-1) \tag{A.10}$$

$$B(t) = HP(t/t-1)H^T + R \tag{A.11}$$

$$P(t/t) = (I - K(t)H)P(t/t-1) \tag{A.12}$$

$$P(t+1/t) = FP(t/t)F^T + \Gamma Q \Gamma^T. \tag{A.13}$$

The likelihood function (A.4) can now be written as

$$\log p(Y_N|\theta) = -\tfrac{1}{2} \sum_{j=1}^{N} [v^T(j)B^{-1}(j/j-1)v(j) + \log|B(j/j-1)|]. \tag{A.14}$$

Here $v(t)$ denotes the innovation sequence which is zero mean, Gaussian and white [5]. ML estimate $\hat{\theta}$ is obtained by maximizing (A.14) with respect to θ subject to the constraints (A.7)–(A.13). This is a very difficult optimization problem. An approximation suggested in Ref. (3) simplifies the problem tremendously. It is assumed that the filter gain $K(t)$ and covariance $B(t/t-1)$ have reached constant values K and B and the vector θ consists of unknown parameters from F, G, K and B only. Then

$$\log p(Y_N|\theta) = -\tfrac{1}{2}\sum_{j=1}^{N}[v^T(j)B^{-1}v(j) + \log|B|]. \tag{A.15}$$

Maximizing (A.15) over B, produces

$$\hat{B} = \frac{1}{N}\sum_{j=1}^{N} v(j|\hat{\alpha})v^T(j|\hat{\alpha}) \tag{A.16}$$

where α is the ML estimate of unknowns in F, G and K. It is given by the root of the equation

$$\sum_{j=1}^{N} v^T(j)B^{-1}\frac{\partial v(j)}{\partial\alpha} = 0 \tag{A.17}$$

where $(\partial v(j))/\partial\alpha$ is calculated from equations (A.7)–(A.9). The root of equation (A17) is found by a Newton–Raphson or Gauss–Newton iteration. Once $\hat{\alpha}$ is obtained, Γ, Q and R are obtained from equations (10)–(13). In this way, the nonlinear constraints of equations (10)–(13) are avoided during optimization. The above method is no more complicated than the well-known output error method. In fact, it reduces to the output error method when there is no process noise, i.e., $w(t) = 0$. In that case, $Q = 0$, $K = 0$ and $v(t) = y(t) - Hx(t)$ is the output error. A flow chart of the method is shown in Figure A.2.

3. *Nonlinear Systems*

Consider a nonlinear dynamic system

$$x(t+1) = f(x(t), \theta, u(t)) + \Gamma w(t) \tag{A.18}$$

$$y(t) = h(x(t)) + v(t) \tag{A.19}$$

where $f(\cdot)$ and $h(\cdot)$ are $n \times 1$ and $r \times 1$ vectors of nonlinear functions. Also, $w(t)$ and $v(t)$ are Gaussian white noise sequences with zero mean and covariances Q and R.

The evaluation of the true ML estimate would require the calculation of $p(y(j)|Y_{j-1}, \theta)$ using an optimal nonlinear filter. Since this is computationally infeasible, we approximate $p((y(j)|Y_{j-1}, \theta)$ by a Gaussian density with mean and covariance obtained from an Extended Kalman Filter [8] of the following form:

$$\hat{x}(t+1/t) = f(\hat{x}(t/t)), \theta, u(t)) \tag{A.20}$$

$$\hat{x}(t/t) = \hat{x}(t/t-1) + K(t)v(t) \tag{A.21}$$

$$v(t) = y(t) - h(\hat{x}(t/t-1)) \tag{A.22}$$

$K(t)$ is calculated from equations (A.10)–(A.13) by using time-varying matrices $F(t)$ and $H(t)$.

$$H(t) = \left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}(t/t-1)} \tag{A.23}$$

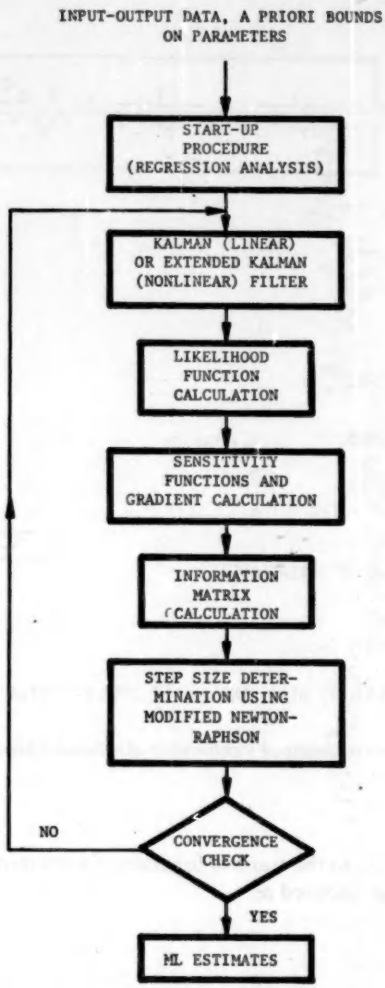


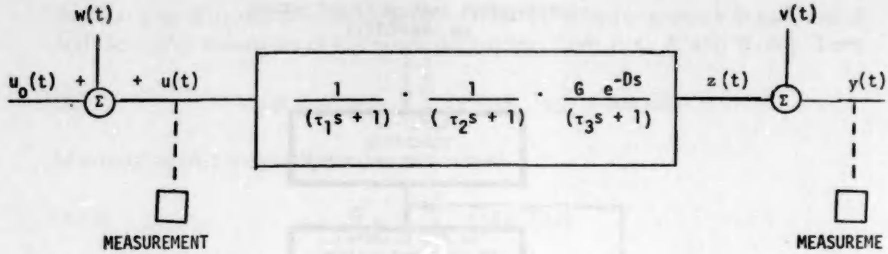
Figure A.2 Flow chart of the maximum likelihood algorithm

$$F(t) = \left.\frac{\partial f}{\partial x}\right|_{x=\hat{x}(t/t)}. \tag{A.24}$$

Kailath [6] has shown that the density of the innovation $\nu(t)$ tends to a Gaussian density as the sampling rate is increased. Thus the above approximation is quite good for high sampling rates.

4. *Applications*

The above method has been applied to two bench-mark problems. The first problem involves estimation of 3 time constants, gain, dead time and variances of



UNKNOWN PARAMETERS:

TIME CONSTANTS: τ_1, τ_2, τ_3

TIME DELAY: D

GAIN: G

WHITE NOISE STANDARD DEVIATIONS:

INPUT: q

OUTPUT: r

DATA: 480 SAMPLES OF u(t) AND y(t); SAMPLING INTERVAL: 1 MIN.

Figure A.3 Maximum likelihood identification of a chemical reactor benchmark problem

input and output noise in the transfer function of a chemical reactor (Figure A.3). The state variables are defined as

$$x_1 = z$$
$$x_2 = \tau_1 \dot{x}_1 + x_1$$
$$x_3 = \tau_2 \dot{x}_2 + x_2.$$

The state equations are

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\theta_1, & \theta_1, & 0 \\ 0, & -\theta_2, & \theta_2 \\ 0, & 0, & -\theta_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \theta_4 \end{bmatrix} u(t - \theta_5)$$

where $\theta_1 = 1/\tau_1$, $\theta_2 = 1/\tau_2$, $\theta_3 = 1/t_3$, $\theta_4 = G/\tau_3$ and $\theta_5 = D$.

Table A.1 shows the values of true and estimated parameters based on input and output time histories of Figure A.4. The innovations pass the whiteness test at 95 per cent confidence level. A sample fit is shown in Figure A.5.

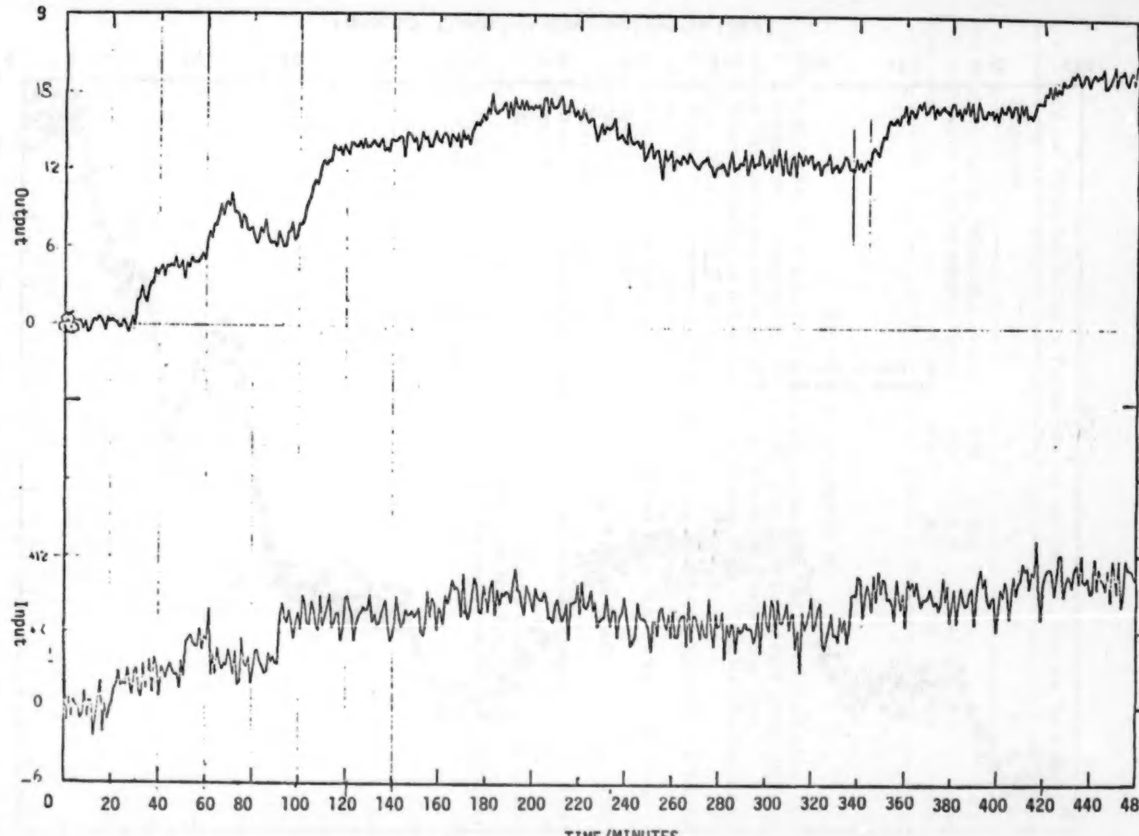


Figure A.4 Input–output data for the benchmark problem

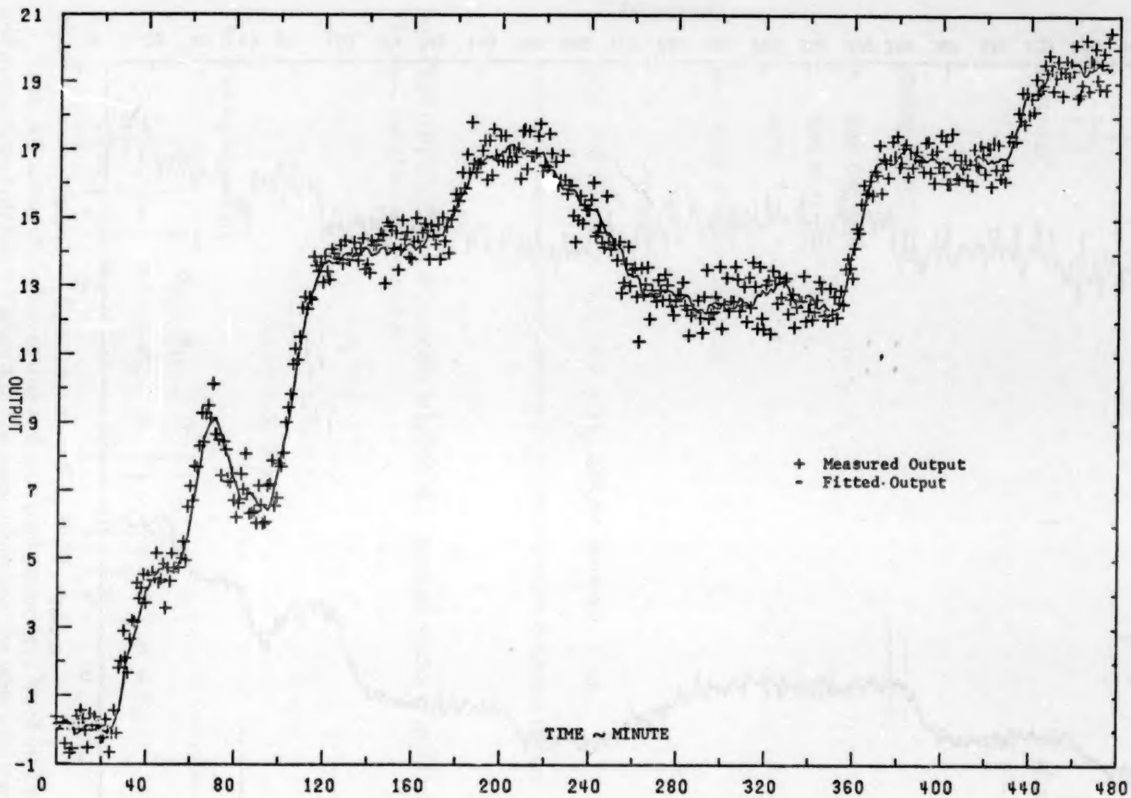


Figure A.5 Comparison of measured and fitted data

The second application concerns the determination of stability and control derivatives of an X-22 VTOL aircraft. The model is nonlinear and has flight disturbance (gusts, etc.) forcing functions with 23 unknown parameters. For this and other applications, see Ref. [9].

TABLE A.1

TRUE AND ESTIMATED VALUES OF PARAMETERS FOR THE CHEMICAL REACTOR PROBLEMS

Parameters	D	τ_1	τ_2	τ_3	G	r
ML estimates based on 480 samples	4 min	4.98	5.22	1.9	1.91	0.5
ML estimates based on 240 samples	5 min	3.63	4.13	1.63	1.94	0.46
True values	4.2 min	3.9	4.7	2.9	1.9	0.5

REFERENCES (Appendix A)

[1] E. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden Day, 1970.

[2] K. J. Astrom and S. Wennmark, "Numerical Identification of Stationary Time Series," 6th International Instruments and Measurement Congress, Stockholm, Sweden, Sept. 1964.

[3] R. K. Mehra, "Identification of Stochastic Linear Dynamic Systems Using Kalman Filter Representation," *AIAA Journal*, 9, No. 1, 28–31, January 1971.

[4] R. K. Mehra, "On-Line Identification of Linear Dynamic Systems with Applications to Kalman Filtering," *IEEE Trans. Automatic Control*, AC-16, No. 1, February 1971.

[5] T. Kailath, "An Innovations Approach to Least Squares Estimation: Part I," *IEEE Trans. on Automatic Control*, December 1968.

[6] T. Kailath, "A General Likelihood—Ratio Formula for Random Signals in Gaussian Noise," *IEEE Trans.* IT, May 1969.

[7] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Trans. ASME J. Basic Eng.* 82, 34–45, March 1960.

[8] R. E. Larson, R. M. Dressler, and R. S. Ratner, "Application of the Extended Kalman Filter to Ballistic Trajectory Estimation," Final Report SRI, Proj. 5188–103, January 1967.

[9] R. K. Mehra and J. S. Tyler, "Case Studies in Aircraft Parameter Identification," 1973 IFAC Sym. on Identification, The Hague, Netherlands.

th
se
in
pa
m
vo
tio

a
of
ar
ot
al
th
m

th
h
le
ea

h
n
n
pa
h

n
o
n
n

n
h
x