Numerical Identification of Linear Dynamic Systems from Normal Operating Records

Åström, Karl Johan; Bohlin, Torsten

Published in:
Theory of Self-Adaptive Control Systems

1966

NUMERICAL IDENTIFICATION OF LINEAR DYNAMIC SYSTEMS FROM NORMAL OPERATING RECORDS

K. J. Åström T. Bohlin

IBM
SYSTEMS DEVELOPMENT DIVISION
NORDIC LABORATORY SWEDEN
TECHNICAL PAPER 15 JULY 1967 TP18.159

This report first appeared in Proceedings of the Second IFAC Symposium on the Theory of Self-Adaptive Control Systems, September 14-17, 1965, and was published by the Instrument Society of America, Plenum Press, New York, 1966.

NUMERICAL IDENTIFICATION OF LINEAR DYNAMIC SYSTEMS FROM NORMAL OPERATING RECORDS

K. J. Åström* T. Bohlin

SYNOPSIS

A technique for numerical identification of a discrete time system from input/output samples is described. The purpose of the identification is to design strategies for control of the system. The strategies are obtained using linear stochastic control theory.

The parameters of the system are estimated by Maximum Likelihood. An algorithm for solving the M.L. equations is given. The estimates are in general consistent, asymptotically normal and efficient for increasing sample lengths. These properties and also the parameter accuracy are determined by the information matrix. An estimate of this matrix is given.

The technique has been applied to simulated data and to plant data.

TP 18.159
Technical Paper
15 July 1967
IBM NORDIC LABORATORY
LIDINGÖ, SWEDEN

* Now at Lund's Institute of Technology, Sweden

table of contents

1. Introduction. 1

2. Statement of the Problem. 3

2.1. Model of the System to Be Identified 3
2.2. Problem Statement 4
2.3. Minimum Variance Prediction and Control Algorithms . . . 5

3. Solution. 8

3.1. The Likelihood Function 8
3.2. Maximizing the Likelihood Function. 9
3.3. Numerical Algorithm. 9
3.4. Starting Value. 12

4. Large Sample Properties of the Estimates. 13

4.1. Consistency. 16
4.2. Identifiability 20
4.3. Asymptotic Normality 22
4.4. Tests of the Order of Model 23

5. Example. 24

6. Comparison with Model Adjustment Techniques 26

7. Extensions 29

8. References 36

1. INTRODUCTION

In order to apply linear stochastic control theory, the process to be controlled should be described in terms of linear differential or difference equations driven by the input signals and disturbances in the form of weakly stationary stochastic processes with rational spectra. In this paper we describe a technique for determining such models from samples of the process inputs and outputs.

We restrict ourselves to the case of

- difference equations
- single input and single output
- time invariant models

As will be discussed in section 7, these restrictions are not essential for the application of the technique.

If it is also assumed that the disturbances are normal there is a canonical form (2.1) containing a finite number of unknown parameters, for the class of models of interest. Each set of parameters together with the input sequence determine uniquely the distribution of the output sequence. Conversely, given a sample of the observed output we have a statistical parameter estimation problem. We will solve it using the method of Maximum Likelihood. For alternative approaches to the identification problem we refer to [3], [4], [5], [8], [15], [18], [19], [21], [24].

The problem is stated and commented on in section 2. In section 3, an algorithm is given for maximizing the likelihood function, and in section 4 the statistical properties of the estimates are investigated. Section 5 contains an example. An alternative interpretation of the technique is given in section 6.

The algorithm has been tested on artifically generated data and has generally been able to identify the parameters with the theoretical

accuracy. It has also been applied to design control strategies for regulating a paper machine [2]. A large amount of data has been analyzed. In general, it has in this application been possible to fit the data with models of low order (first or second order with delay). As a rule, we encountered very few numerical difficulties with the standard algorithm. So far, our practical experience with the technique is, however, limited to systems of comparatively low order. The results of the practical experiments have been very satisfactory.

2. STATEMENT OF THE PROBLEM

2.1. Model of the System to be Identified

Consider a discrete time single-input single-output dynamical system whose input/output relation can be described by the equation

$$A(z^{-1})\, y(t) = B(z^{-1})\, u(t) + \lambda\, C(z^{-1})\, e(t) \qquad (2.1)$$

where $\{u(t)\}$ is the input $\{y(t)\}$ the output and $\{e(t)\}$ a sequence of independent normal $(0, 1)$ random variables. Furthermore, z denotes the shift operator [22].

$$z\, x(t) = x(t+1) \qquad (2.2)$$

and $A(z)$, $B(z)$ and $C(z)$ are polynomials

$$A(z) = 1 + a_1 z + \ldots + a_n z^n$$

$$B(z) = b_o + b_1 z + \ldots + b_n z^n$$

$$C(z) = 1 + c_1 z + \ldots + c_n z^n \qquad (2.3)$$

We also introduce the row-vectors a, b and c whose components are a_i, b_i and c_i respectively.

The following assumptions are made

- the functions $A(z^{-1})$ and $C(z^{-1})$ have all their zeros inside the unit circle
- there are no factors common to all three polynomials $A(z)$, $B(z)$ and $C(z)$

The assumption that the function $A(z^{-1})$ has all its zeros inside the unit circle implies that the homogeneous equation corresponding to (2.1) is

asymptotically stable. The assumption that there are no factors common to A(z), B(z) and C(z) implies that every state of the system (2.1) is controllable either from u or from e. [13]. This is no loss in generality. Neither is there any loss in generality to assume that the leading coefficient of the polynomials A(z) and C(z) are unity, we can, however, not make this assumption for the polynomial B(z). Also, notice that in (2.1) the degrees of all the polynomials A(z), B(z) and C(z) formally are the same. If this is not desired we can put some of the coefficients equal to zero.

The system represented by the equation (2.1) is in fact the general, single-input single-output linear discrete-time dynamical system, with normal disturbances having rational power spectra. Notice in particular that systems with time delays also can be represented by the model (2.1) if the time delay is an integer multiple of the sampling interval.

The system model (2.1) contains $4n + 2$ parameters, the $3n + 2$ coefficients of the equation (2.1) $a_1, a_2, \ldots, a_n, b_o, b_1, b_2, \ldots, b_n, c_1, c_2, \ldots, c_n, \lambda$ and n initial conditions for the equation (2.1). The initial conditions add little to the problem and are assumed zero. In practice, it is often necessary to include a constant level as an additional parameter. This adds nothing of interest to the identification problem and is therefore neglected. The complete problem including the initial conditions and the constant level is considered in [1].

2.2. Problem Statement

We now formulate the identification problem as follows

PROBLEM

Given the input $\{u(t),\ t = 1, 2, \ldots, N\}$ and observations of the output $\{y(t),\ t = 1, 2, \ldots, N\}$ find an estimate of the parameters of the model (2.1).

Special cases of this problem are well-known

1. $n = 0$, regression analysis [6].
2. $b_o = b_1 = \ldots = b_n = c_1 = c_2 = \ldots = c_n = 0$,
   estimation of parameters in autoregressive processes [10], [16].
3. $b_o = b_1 = \ldots = b_n = a_1 = a_2 = \ldots = a_n = 0$,
   estimation of parameters in a moving average [10], [23], [26].
4. $b_o = b_1 = \ldots = b_n = 0$, parametric estimation of rational power spectra [2], [7].
5. $c_1 = c_2 = \ldots = c_n = 0$, least squares model building [12].
6. $a_1 = c_1$, $i = 1, 2 \ldots, n$, identification of noisefree process with measurement errors [15].

The general case has been considered by Galtieri [9].

2.3. Minimum Variance Prediction and Control Algorithms

Before proceeding to the solution of the stated problem we will demonstrate that once a model of type (2.1) is obtained it is a very simple matter to derive the minimum mean square control algorithm. This will be discussed in detail elsewhere, let it therefore suffice to give an example. For the sake of simplicity, we assume that $b_o = 0$ and $b_1 \neq 0$. Consider the situation at time t. The data $\ldots, y(t-1), y(t), \ldots, u(t-1), \ldots$ have been observed. The crucial step in the derivation of minimum mean square control algorithms is to find the minimum mean square prediction. It is well known that this problem is solved if we express $y(t+1)$ as a function of the data $\ldots, y(t-1), y(t), \ldots, u(t-1), u(t)$ and a residual which is independent of the data. From the equation (2.1) we can immediately obtain such an expression. Solving (2.1) in terms of $y(t+1)$ we get

$$y(t+1) = \lambda e(t+1) + A^{-1}(z^{-1})B(z^{-1})u(t+1)$$

$$+ A^{-1}(z^{-1})\,[C(z^{-1}) - A(z^{-1})]\lambda e(t+1) \qquad (2.4)$$

Eliminating $e(t+1)$ using (2.1) we get

$$y(t+1) = \lambda e(t+1) + C^{-1}(z^{-1})B(z^{-1})\,zu(t)$$

$$+ C^{-1}(z^{-1})\,[C(z^{-1}) - A(z^{-1})]\,zy(t) \qquad (2.5)$$

Due to the assumptions made we find that the series expansion in powers of z^{-1} of the operators $C^{-1}(z^{-1})B(z^{-1})$ and $C^{-1}(z^{-1})\,[C(z^{-1}) - A(z^{-1})]$ have no constant terms. The right member of (2.5) depends only on the data $y(t)$, $y(t-1)$, ..., $u(t)$, $u(t-1)$, ... and $e(t+1)$. As $e(t+1)$ is independent of the other terms of the right member we have obtained the desired expression. The last two terms of the equation (2.5) can thus be interpreted as the minimum mean square prediction of $y(t+1)$ based on the data $y(t)$, $y(t-1)$, ..., $u(t)$, $u(t-1)$, ... The prediction error is $\lambda e(t+1)$. As $e(t)$ is normal $(0, 1)$ the number λ has physical interpretation as the standard deviation of the prediction error.

Having obtained the minimum mean square predictor we will now derive the minimum mean square control law. We observe that

$$Ey^2(t+1) \geq \lambda^2 \qquad (2.6)$$

where equality holds if

$$u(t) = -B^{-1}(z^{-1})[C(z^{-1}) - A(z^{-1})]y(t) \qquad (2.7)$$

The equation (2.7) is thus the minimum mean square control law. As $b_o = 0$ and $b_1 \neq 0$, the series expansion of the operator of the right member does only contain non-negative powers of z^{-1} and (2.7) is thus a physically realizable control law.

Thus we have demonstrated that under the particular assumptions $b_1 \neq 0$, $b_o = 0$ the minimum mean square control algorithm is easily obtained from the model (2.1). Hence once the identification problem is solved we have in fact also a solution to the minimum mean square control problem, and the statement made in the introduction is proven.

3. SOLUTION

The problem as stated in section 2 is a statistical parameter estimation problem. We will solve it by the method of Maximum Likelihood. We first give an algorithm for the Maximum Likelihood estimator and we will later show that the estimates have desirable properties as the number of observations increase.

3.1. The Likelihood Function

Let $p[\{y(t)\} \mid \{u(t)\}, a, b, c, \lambda]$ be the probability density function of the outputs $\{y(t)\}$ given the inputs $\{u(t)\}$ and the parameters a, b, c, λ. The likelihood function is defined as the function p regarded as a function of the parameters and with the observed values $\{y(t)\}$ and $\{u(t)\}$ inserted [6], [14], [27]. The function is thus a stochastic variable. We will now derive an expression for the likelihood function. It follows from (2.1) that the numbers $\varepsilon(t)$ defined by

$$C(z^{-1})\,\varepsilon(t) = A(z^{-1})\,y(t) - B(z^{-1})\,u(t) \tag{2.8}$$

are independent and normal (0, λ). The logarithm of the probability density function of $\{e(t)\}$ now becomes

$$L = \frac{1}{2\lambda^2} \sum_{t=1}^{N} \varepsilon^2(t) - N \log \lambda + \text{const.} \tag{2.9}$$

Since $\{y(t)\}$ is a one-to-one transformation of $\{\varepsilon(t)\}$ and the Jacobian is 1,

L is also the logarithm of the likelihood function. The logarithm of the likelihood function is thus obtained from (2.9) where the "errors" ε are computed from the input $\{u(t)\}$ and the output $\{y(t)\}$ by (2.8). The likelihood function is thus a function of the parameters a, b, c, λ and of n initial conditions of (2.8). For simplicity we will here assume that the initial conditions of (2.8) are zero. This is not essential. For an analysis of the complete case we refer to [1]

3.2. Maximizing the Likelihood Function

We observe that the function L can be maximized with respect to the parameters a, b and c separately. To do this we introduce the function $V(\theta)$ defined by

$$V(\theta) = \frac{1}{2} \sum_{t=1}^{N} \varepsilon^2(t) \tag{2.10}$$

where $\theta = \text{col}\,(a, b, c)$. Maximizing L is equivalent to minimizing the loss function V. When we have found $\hat{\theta}$ such that $V(\hat{\theta})$ is minimal we get the Maximum Likelihood estimate of λ from

$$\hat{\lambda}^2 = \frac{2}{N} V(\hat{\theta}) \tag{2.11}$$

and all parameters are estimated. We observe that $V(\theta)$ is a quadratic function of a and b but that the dependence on c is more complex. Thus we cannot obtain an analytical solution.

3.3. Numerical Algorithm

To maximize the likelihood function/minimize the loss function $V(\theta)$ we use the following Newton-Raphson algorithm

$$\theta^{k+1} = \theta^k - [V_{\theta\theta}(\theta^k)]^{-1} V_\theta(\theta^k) \tag{2.12}$$

where V_θ denotes the gradient and $V_{\theta\theta}$ the matrix of second partial derivatives of $V(\theta)$. For a discussion of the algorithm (2.12) and related ones see [1], [11].

The partial derivatives of the loss function are obtained by straightforward differentiation. We get

$$\frac{\partial V}{\partial \theta_i} = \sum_{t=1}^{N} \varepsilon(t) \frac{\partial \varepsilon(t)}{\partial \theta_i} \tag{2.13}$$

$$\frac{\partial^2 V}{\partial \theta_i \partial \theta_j} = \sum_{t=1}^{N} \frac{\partial \varepsilon(t)}{\partial \theta_i} \cdot \frac{\partial \varepsilon(t)}{\partial \theta_j} + \sum_{t=1}^{N} \varepsilon(t) \frac{\partial^2 \varepsilon(t)}{\partial \theta_i \partial \theta_j} \tag{2.14}$$

The derivatives of $\varepsilon(t)$ are obtained by differentiating the difference equation (2.8).

$$C(z^{-1}) \frac{\partial \varepsilon(t)}{\partial a_i} = z^{-i} y(t)$$

$$C(z^{-1}) \frac{\partial \varepsilon(t)}{\partial b_i} = -z^{-i} u(t) \tag{2.15}$$

$$C(z^{-1}) \frac{\partial \varepsilon(t)}{\partial c_i} = -z^{-i} \varepsilon(t)$$

$$C(z^{-1}) \frac{\partial^2 \varepsilon(t)}{\partial a_i \partial c_j} = -z^{-i-j} \frac{\partial \varepsilon(t)}{\partial a_1}$$

$$C(z^{-1}) \frac{\partial^2 \varepsilon(t)}{\partial b_i \partial c_j} = -z^{-i-j} \frac{\partial \varepsilon(t)}{\partial b_1} \tag{2.16}$$

$$C(z^{-1}) \frac{\partial^2 \varepsilon(t)}{\partial c_1 \partial c_j} = -2z^{-i-j} \frac{\partial \varepsilon(t)}{\partial c_1}$$

The initial values for the difference equations (2.15) and (2.16) are zero.

Notice that the second order partial derivatives of $\varepsilon(t)$ that do not involve the coefficients c_i are all identically zero. Notice that the derivatives of the residuals are equivalent to Meissinger's [20] sensitivity coefficients.

The equations (2.8), (2.10), (2.15) and (2.16) immediately suggest a recursive scheme for computating the loss-function $V(\theta)$ and its partial derivatives. Alternatively these functions can be obtained as outputs of linear dynamical systems. Notice that in (2.15) and (2.16) the derivatives with respect to different parameters in the same group (a, b or c) can be obtained by shifting. We get e.g. from (2.15)

$$\frac{\partial \varepsilon(t)}{\partial a_i} = \frac{\partial \varepsilon(t-i+1)}{\partial a_1} \qquad i \leq t+1 \tag{2.17}$$

This leads to considerable simplifications of the computations as it is only necessary to solve the equations (2.15), (2.16) for $i = j = 1$.

Also notice that by utilizing (2.17) it is also possible to simplify the computation of the first term of the right member of (2.14). We have e.g.

$$\sum_{t=1}^{N} \frac{\partial \varepsilon(t)}{\partial a_i} \cdot \frac{\partial \varepsilon(t)}{\partial b_j} = \sum_{t=1}^{N} \frac{\partial \varepsilon(t-i+1)}{\partial a_1} \cdot \frac{\partial \varepsilon(t-j+1)}{\partial b_1}$$

$$= \sum_{t'=1}^{N-j} \frac{\partial \varepsilon(t'+j-i)}{\partial a_1} \cdot \frac{\partial \varepsilon(t)}{\partial b_1} \qquad j \leq i \tag{2.18}$$

Similar formulas also hold for the other derivatives.

The algorithm for maximizing the likelihood function is thus

1. Put $\theta^k = \theta^o$ (starting value of θ)

2. Evaluate $V_\theta(\theta^k)$ and $V_{\theta\theta}(\theta^k)$ using (2.13), (2.14), (2.15) and (2.16)

3. Calculate θ^{k+1} from (2.12) and repeat from 2.

3.4. Starting Value

The algorithm (2.12) requires a starting value. We observe that if the parameters c are given, then $V(\theta)$ is a quadratic function of a and b and the second partial derivatives of $\epsilon(t)$ are all zero. The iteration (2.12) then converges in one step from any initial value for the parameters a and b. In particular if we put $c_i = 0$, $i > 0$ we obtain in one step with the approximative second partial derivatives, the least squares estimate a^o and b^o of a and b. The initial value for the iteration (2.12) is then taken as $\theta^o = \text{col}\,(a^o, b^o, 0)$.

4. LARGE SAMPLE PROPERTIES OF THE ESTIMATES

When the identification scheme is applied there are many problems which naturally arise. Typical examples are the following:

- What happens to the estimate as the number of observations (N) increases?
- How accurate is the estimate?
- Are there "better" ways of estimating the parameters?
- What systems are possible to identify?
- In case we can choose the input signal, how should it be chosen?
- What order should be chosen for the model?
- Does the solution obtained correspond to the absolute maximum?

In this section we develop some means for dealing with such questions. Many of these are essentially answered by an investigation of the large sample properties of the estimates. It means that asymptotic expressions can be utilized. The problem of several local maxima, however, cannot be solved by the results of this section. In the sequel, we disregard it, and hence assume that $\hat{\theta}^N$ for all sufficiently large N is the parameter value corresponding to the absolute maximum of the likelihood function.

The investigation of the large sample properties is a purely statistical problem, and we will use concepts and methods from mathematical statistics to find them. The complete investigation is uncomfortably involved and detailed, and we have chosen to omit it from this paper. However, the results are presented below in the form of mathematical theorems, which are commented on with regard to their application to some of the above problems. The proofs are published in [1].

To facilitate the reading we restate and define more closely some of the problems and introduce a few statistical concepts.

- Consistency, i.e. $\hat{\theta}^N$ converges to θ_o when N increases.

- Asymptotic normality, i.e. convergence in distribution of the quantity $\sqrt{N}\ (\hat{\theta}^N - \theta_o)$ to a normal variable.

- Asymptotic efficiency, i.e. equality of the covariance matrix of the limiting distribution to the Cramér-Rao lower bound for regular estimators.

For further definitions of terms see [27].

We want to have conditions that guarantee these properties. In the present case the properties depend on the input and the parameters θ_o. Since θ_o is not known, the conditions should preferably be expressed in terms of the input u and the output y which are both known quantities. If this is possible we are then able to resolve whether a certain sequence of estimates $\{\hat{\theta}^N\}$ actually has one of the desired properties in each case. It is evident that this can be done with certainty only if the sample is infinitely long, and we will confine ourselves to this case when defining the required conditions. This means that we will not consider here the problem of constructing statistical tests for finite sample lengths N.

The consistency conditions can be used to resolve the following important problem: to be able to design an experiment in order to estimate θ_o we must know what class of input sequences u that are able to excite the system sufficiently enough to yield consistent estimates of the system parameter θ_o. Since at this stage we do not know θ_o (or y) we can utilize only u such that the estimates $\{\hat{\theta}^N\}$ are consistent irrespective of the value θ_o. We are interested in such (smaller) classes of u.

The results of this section solve the following problems:

1. What set in the parameter space does $\hat{\theta}^N$ converge into (Theorem 1 + Lemma 3).

2. In cases where $\{\hat{\theta}^N\}$ is not consistent, find (singular) function of $\hat{\theta}^N$ that is consistent (Theorem 2).

3. Find a class of inputs u and of system parameters θ_o such that $\{\hat{\theta}^N\}$ is consistent (Theorem 3).

4. Find conditions on u and y for $\{\hat{\theta}^N\}$ to be consistent (Corrolary, theorem 2).

5. Find conditions such that $\{\hat{\theta}^N\}$ is asympotically normal and asymptotically efficient (Theorem 4).

6. When $\{\hat{\theta}^N\}$ is asymptotically normal and efficient find an estimate of the covariance matrix of the limiting distribution (Lemma 4).

The theorems are in a sense ergodic theorems, since they all deal with asymptotic properties of functions of a single sample y. The ergodic property establishes that a single realization of the process output may be used in place of an ensemble of realizations.

The following is a general regularity condition on the input sequence $\{u(t)\}$

Condition A

$u(t)$ and the crossproducts $u(t)\ u(t+\tau)$ be bounded and Cesaro summable, i.e. the following limits

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} u(t)$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} u(t)\, u(t+T)$$

exist for all finite T.

In the sequel it is necessary to distinguish between θ = an arbitrary point in the parameter space and θ_o = the true parameter point, i.e. the parameter point defining the observed output according to (2.1).

Denote the logarithm of the likelihood function by $L^N(y \mid \theta)$, its gradient vector $L^N_\theta(y \mid \theta)$, and its second derivative matrix $L^N_{\theta\theta}(y \mid \theta)$. Further, denote the vector of length N with components y(i) by y. Analogously for u.

4.1. Consistency

Lemma 1

Let R be a region in 3n+2 dimensional Euclidian space E^{3n+2} defined by

$R = \{\theta \mid \lambda>0$, and all zeros of $A(z^{-1})$ and $C(z^{-1})$ lie strictly inside the unit circle$\}$

Assume that u satisfies the condition A. Then

$$\lim_{N\to\infty} \frac{1}{N} L^N(y \mid \theta) = \lim_{N\to\infty} \frac{1}{N} EL^N(y \mid \theta) = L(\theta, \theta_o)$$

with probability one if $\theta \in R$ and $\theta_o \in R$.

Lemma 2

Let $R' \subset R$ be a closed set.

Assume u satisfies condition A.

Then $L(\theta, \theta_o)$ is an analytic function in R', and we have

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{grad}_\theta L^N(y \mid \theta) = \operatorname{grad}_\theta \lim_{N\to\infty} \frac{L^N(y \mid \theta)}{N} = L_\theta(\theta, \theta_o)$$

with probability one. The relation also holds for higher derivatives.

The lemmas establish that the time average of the residuals $\varepsilon^2(t)$ i.e. $\frac{1}{N} L^N(y \mid \theta)$ converges to its ensemble average, which is a differentiable function in the parameters θ. The conditions are mild and natural, i.e. the system, the model, and the optimal predictor of the noise component $C(z^{-1})e$ be asymptotically stable.

The lemmas are fundamental for

Theorem 1

Let S_o be a set in E^{3n+1} defined by

$$S_o = \{\theta \mid L(\theta, \theta_o) = L(\theta_o, \theta_o)\}$$

Assume u satisfies condition A, and that for all sufficiently large N, $\hat{\theta}^N \in R'$, where $R' \subseteq R$ is a closed set. Then

$$\lim_{N\to\infty} \| \hat{\theta}^N - P\hat{\theta}^N \| = 0$$

with probability one, where $P\vartheta$ is the projection on $S_o \cap R'$, i.e. the nearest point $\in S_o \cap R'$.

This theorem replaces a consistency theorem.

It asserts that the estimates $\hat{\theta}^N$ converge into the set S_o, though not necessarily to a point. It gives the consistency conditions in terms of conditions on (u, θ_o) through the

Corrolary

If the set S_o contains only one point (θ_o), $\{\hat{\theta}^N\}$ is strongly consistent.

The set S_o can be interpreted as the set of parameters θ that are equivalent to θ_o in the sense that any model with $\theta \in S_o$ generates realizations that for long samples have the same statistical behaviour (same likelihood function) as the system output y. The condition is then natural since there is no way to judge from the output only which of the parameters $\theta \in S_o$ that generated the observed output.

The purpose of the following theorems is to characterize S_o and find conditions for S_o to contain only the point $\theta = \theta_o$.

Lemma 3

The set S_o, as defined in theorem 1 has the following property

$$S_o = R \cap S'_o$$

where S'_o is a linear set.

Hence, we know that in all cases $\hat{\theta}^N$ will at least converge into a hyperplane S_o. This suggests that components of $\hat{\theta}^N$ orthogonal to this hyperplane will be consistent. We want to be able to calculate such components. This can be done with the aid of

Theorem 2

Let $\Lambda^N(y \mid \hat{\theta}^N)$ be the diagonal matrix of eigenvalues of $\frac{1}{N} L^N(y \mid \hat{\theta}^N)$ and let $P^N(y \mid \hat{\theta}^N)$ be a matrix of corresponding (orthogonal) eigenvectors. Then

$$\lim_{N\to\infty} \| \Lambda^N(y \mid \hat{\theta}^N)\, P^{N\,T}(y \mid \hat{\theta}^N)\, (\hat{\theta}^N - \theta_o) \| = 0$$

with probability one.

The rather complicated form of this theorem is due to our desire to express the projections in computable terms. The main difficulty arises from the fact that it cannot be shown that $\frac{1}{N} L^N_{\theta\theta}(y \mid \hat{\theta}^N)$ converges. Even if this is the case, the limit may have no unique set of eigen vectors so that it is difficult or impossible to define a convergent sequence of orthogonal transformations $\{P^N(y \mid \hat{\theta}^N)\}$. This is the case particularly if the limit is singular with at least two eigenvalues zero. The singular case is of practical interest, since it arises from choosing a too high order model or from the fact that the system is degenerate (not controllable) or not excited (see theorem 3). In such cases we can then use theorem 2 to find a more reasonable model by taking the projections orthogonal to S'_o as new parameters. The estimates of these new parameters are then consistent.

From theorem 2 we immediately obtain a solution of the problem 4) posed above through the

Corrolary

If $[\frac{1}{N} L^N_{\theta\theta}(y \mid \hat{\theta}^N)]^{-1}$ is bounded then $\hat{\theta}^N$ is consistent.

It is not shown that the converse is necessarily true. When calculating $\hat{\theta}^N$ the matrix $L^N_{\theta\theta}(y \mid \hat{\theta}^N)$ is actually computed. Each estimate is thus accompanied by a quantity which can be used to judge its significance. A complete characterization of S_o in terms of (u, θ_o) is given in the proof of theorem 3 [1].

4.2. Identifiability

From the equations characterizing S_o it is possible, at least in principle, to resolve whether they possess a unique solution $\theta = \theta_o$ in which case $\hat{\theta}^N$ is consistent. We can also construct conditions for this being the case. These conditions are complicated and impractical for applications. It is desirable to find simpler conditions, possibly more restrictive, that are sufficient to ensure that S_o contains only one point. The corresponding theorems will be called identifiability theorems, because they give conditions for the system to be identifiable from the input/output record, in the sense that the parameter estimates are consistent. It is attractive to think of the conditions in the following terms:

A system at rest is defined by the parameters θ_o according to (2.1), $u = 0$ and regarded as a black box containing the unknown parameters θ_o. In order to draw conclusions about the contents of this black box it is necessary to excite the system by applying some $u \neq 0$ and observe the response. The input must excite all components of θ_o and must be sufficiently persistent, since the response is obscured by noise. It is evident that parameter components that cannot be reached by the input (or the noise) can never be estimated. Hence, some controllability requirements are also needed.

Definition

A process is said to be completely identifiable if the Maximum Likelihood

estimates of θ_o are consistent

Definition

A bounded signal u is said to be persistently exciting of order m if the limits

$$\bar{u} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} u(t) \text{ and } r_u(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} u(t)\, u(t+\tau)$$

exist and if the matrix

$$R_u = \{r_u(i-j) \mid i,j = 1,\ldots,m+1\}$$

is positive definite.

We can now state the main result.

Theorem 3

The process (2.1) is completely identifiable if the input is persistently exciting of order 2n and every state of the process is controllable either from u or e.

The first condition is easy to verify in practice, since u is known. The second condition is of less importance in a practical case, we can never verify it before the experiment and we can always verify it after by means of theorem 2. It is, however, useful as a means of diagnosis: Why are the estimates inconsistent?

4.3. Asymptotic Normality

Theorem 4

Assume that S_o contains only the point θ_o so that $\hat{\theta}^N$ is consistent. Then the stochastic variable $L_{\theta\theta}(\theta_o, \theta_o)\sqrt{N}\,(\hat{\theta}^N - \theta_o)$ is asymptotically normal $(0, -L_{\theta\theta})$.

If in addition $L_{\theta\theta}(\theta_o, \theta_o)$ is nonsingular, then $\hat{\theta}^N$ is asymptotically normal $(\theta_o, -\frac{1}{N} L^{-1}_{\theta\theta})$.

Since $NL_{\theta\theta}(\theta_o, \theta_o) \sim L^N_{\theta\theta}(y \mid \theta_o)$ = the information matrix, the estimates are also asymptotically efficient.

The fact that the estimates are asymptotically efficient means in practice that we cannot expect to find an estimator with a greater accuracy for long samples.

The asymptotic normality implies that the distribution of $\hat{\theta}^N$ is completely known, that confidence regions for the parameters can be determined, and that approximate significance tests can be performed. To perform the tests an estimate of the covariance matrix is required. This is obtained from

Lemma 4

Assume that lemma 2 and theorem 1 hold. Then

$$\| \tfrac{1}{N} L^N_{\theta\theta}(y \mid \hat{\theta}^N) - L_{\theta\theta}(P\,\hat{\theta}^N, \theta_o) \| \to 0$$

with probability one.

Besides solving the consistency problem, the matrix $\lim_{N\to\infty} L^N_{\theta\theta}(y \mid \hat{\theta}^N)$ then

also yields an estimate of the accuracy of the estimated parameters in cases where the parameter estimates are consistent.

In practice we do not have an infinite sample and then the result means that whenever we can invert the matrix $\frac{1}{N} L_{\theta\theta}^{N}(y \mid \hat{\theta}^{N})$ without difficulty, we may consider the obtained estimate $\hat{\theta}^{N}$ as consistent. The accuracy of the estimate is given approximately by the inverse of the matrix. If the matrix is nearly singular then the inverse will always contain some very large diagonal elements, and we may either say that the corresponding component of $\hat{\theta}^{N}$ is not consistent, or it is consistent but with a very large standard deviation. The practical result is the same, namely the conclusion that we have included too many parameters.

4.4. Tests of the Order of Model

According to Theorem 2, the second derivative matrix $L_{\theta\theta}^{N}(y \mid \hat{\theta}^{N})$ can be used as an indication that there are redundant parameters in the model and also to determine a new, non-redundant set of parameters. Hence, we may in practice determine the order of the model by repeating the identification with increasing order using some measure of singularity of the matrix as a test figure. However, the following alternative may sometimes be preferable:

We observe that if the model order is not less than the order of the system then the residuals $\{\varepsilon(t),\ t = 1, \ldots, N\}$ form a series of independent normal variables. We have then another test on the order:

If the residuals form a sequence of independent variables then the order of the model is equal to or greater than the system order. A simple test of independence is to compute the covariance function

$$\frac{1}{N-T} \sum_{t=1}^{N-T} \varepsilon(t)\, \varepsilon(t+\tau) \quad \text{for a few delays.}$$

$$\tau = 1, 2, 3, \ldots$$

A quick method is to count the sign changes, the number of which should be $\approx \frac{1}{2} N$ for a sequence of independent variables.

5. EXAMPLE

As an example we will consider the identification of the following system

$$y(t) = \frac{z^{-1} + 0.5z^{-2}}{1 - 1.5z^{-1} + 0.7z^{-2}} u(t)$$

$$+ \lambda \frac{1 - z^{-1} + 0.2z^{-2}}{1 - 1.5z^{-1} + 0.7z^{-2}} e(t) \qquad (5.1)$$

Three cases are considered

1. $\lambda = 0.4$
2. $\lambda = 1.8$
3. $\lambda = 7.2$

In the experiment 240 pairs of input/output variables (u, y) were generated using the equation (5.1). The random numbers $\{e(t)\}$ were obtained as suitably scaled sums of twelve rectangularly distributed pseudorandom numbers obtained from a modified Fibonacci series. The same sequence of pseudorandom numbers were used in all three cases. In Figure 1 we show the chosen input and the output y in the three cases. As a reference we have also in each case shown the output for $\lambda = 0$.

The identification scheme described in this paper was applied to the generated data. The estimates of the model parameters obtained are given in Table 1. In this table we also give the standard deviations of the estimates which are computed from the estimate of the covariances given by the matrix $\hat{\lambda}^2 V_{\theta\theta}^{-1}$.

From this table we find that the estimates of the coefficients b_1 and b_2

are getting increasingly inaccurate in the experiments while the accuracy of the coefficients a and c are unaffected. This is very natural since the response of the deterministic part of the model is corrupted by an increasing amount of noise. The estimates of the parameters c_1 and c_2 should in general not be expected to depend on the noise amplitude.

To demonstrate the convergence of the algorithm we give in table II the successive iterates for Case 3 and in table III the gradients of the loss function in the various iteration steps. Notice in table II the large difference between the least squares estimate (step 1) of a_1, a_2, b_1 and b_2 and the maximum likelihood estimate.

To illustrate the test of the order of the system as discussed in section 4, the data of Case 2 was also identified as a first order system. The covariances of the residuals for the first and second order models are graphed in Figure 2.

6. COMPARISON WITH MODEL ADJUSTMENT TECHNIQUES

There are many ways to approach the identification problem. Two approaches are represented by

- statistical parameter estimation
- model adjustment

In the first case the problem is put in a probabilistic framework and sufficiently many assumptions are made in order to ensure that the methods of mathematical statistics can be applied. In the model adjustment technique [4], [17], [25] a model characterized by some parameters and a criterion is postulated. The problem is then to adjust the model parameters in such a way that the criterion is satisfied. The model adjustment technique is very general in the sense that complicated models and criteria can be used. The result of the model adjustment procedure is a set of parameter values. When the problem is approached as a parameter estimation problem many more assumptions must be made. Instead the results are much more far reaching. The parameters as well as their confidence intervals are obtained, questions related to significance of the estimates can be answered. In many situations it is rewarding to consider a particular problem from both points of view. Statistical considerations may suggest a suitable criterion for the model adjustment procedure. In a situation where the assumptions required by the statistical approach are not fulfilled we can still obtain a solution to a model adjustment problem. A typical case may be the situation when the assumption on normality of the residuals is not fulfilled.

Classically, the two techniques have been developed in parallel. Let it suffice to mention least squares fitting of linear models.

So far in this paper the identification problem has been discussed entirely from the point of view of statistical parameter estimation. We will now interpret our procedure as a model adjustment technique.

Consider the equation (2.4). We recall that the last two terms of the right member of this equation can be interpreted as the prediction of $y(t+1)$ based on $y(t)$, $y(t-1)$, ... and $u(t)$, $u(t-1)$, ... and that the quantity $\varepsilon(t+1)$ has physical interpretation as the error of the one-step ahead prediction of $y(t)$. Now consider the equation (2.8) which we rewrite as

$$\varepsilon(t+1) = y(t+1) - \{C^{-1}(z^{-1})\,[C(z^{-1}) - A(z^{-1})]y(t+1)$$
$$+ C^{-1}(z^{-1})\,B(z^{-1})\,u(t+1)\} \qquad (6.1)$$

A comparison with the equation (2.4) now shows that the last two terms of the right member can be interpreted as the prediction of $y(t+1)$ based on $y(t)$, $y(t-1)$, ... and $u(t)$, $u(t-1)$, ... In the algorithm (2.8) the number $\varepsilon(t+1)$ can thus be interpreted as the difference between $y(t+1)$ and its one step ahead prediction based on $y(t)$, $y(t-1)$, ..., and $u(t)$, $u(t-1)$, ...

Now consider the one step ahead predictor

$$\hat{y}(t) = C^{-1}(z^{-1})\,B(z^{-1})zu(t-1)$$
$$+ C^{-1}(z^{-1})\,[C(z^{-1}) - A(z^{-1})]zy(t-1) \qquad (6.2)$$

as the model, and let the criterion be the sum of the squares of the prediction errors i.e.

$$V = \frac{1}{2}\sum_{t=1}^{N}[y(t) - \hat{y}(t)]^2 = \frac{1}{2}\Sigma\,\varepsilon^2(t) \qquad (6.3)$$

Compare with equations (2.5) and (2.10). Notice that in the ordinary model reference techniques the model is a deterministic input/output relation while in our case the model is a predictor.

As we intend to use the results of the identification procedure to calculate the minimum variance control strategy i.e. a control strategy such that the one-step ahead prediction of the output is zero we find that even with

the model adjustment interpretation our identification procedure has the required properties. Also notice that the identification algorithm solves the prediction problem for a stationary process with unknown, but rational power spectrum.

It is also of interest to compare the algorithm (2.12) with the algorithms currently used in model reference techniques. Blandhol [4] only evaluates the function $V(\theta)$. Judging from our experience it is very difficult to get a reasonable convergence rate by probing techniques using the values of $V(\theta)$. Blandhol also confirms this.

Notice that $V(\theta)/V(\theta_o)$ is asymptotically independent of N. This implies that the loss function does not get "sharper" with an increasing number of observations and that the "sharpness" of the minimum alone does not determine the accuracy of the estimates.

The gradient $V_\theta(\theta)$ is evaluated in some model reference techniques that are implemented in adaptive systems e.g. [17], [20]. In these cases the parameter adjustment routine is chosen as

$$\theta^{k+1} = \theta^k - \alpha V_\theta(\theta^k)$$

Notice that a more effective algorithm is obtained with very little extra computational effort, using an approximate second partial derivative i.e. neglecting the last term of (2.14). We then conclude that it appears worthwhile to consider this modification in model reference adaptive systems currently in use. By including this feature we could also obtain an estimate of the information matrix and thus also of the accuracy of the estimated parameters.

7. EXTENSIONS

There are many ways in which the problem can be generalized. The identification scheme can be immediately generalized to continuous time. The convergence proofs are, however, more difficult in this case and some modifications might be necessary.

The extension to multiple inputs is trivial. Both the algorithm and the convergence proofs generalize immediately. The extension to multiple outputs is more difficult. The crucial problem is to find a suitable structure. Once the structure is given, the generalization is immediate.

The algorithm can also be extended directly to non-linear and/or time variable systems with known structure. Consider for example the following system:

$$x(t+1) = g[x(t),\ u(t),\ t]$$

where $g(x, u, t)$ is a function which contains some unknown parameters. Let u be the input(s) of the system and let the output y be given by

$$y(t) = x_1(t) + z_1(t) + c_o\ e(t) + \varkappa$$

where the vector $z(t)$ is given by

$$z(t+1) = Fz(t) + Ge(t)$$

The system described by these equations is an arbitrary non-linear system with a single output, with a random disturbance in the output that is stationary and a rational spectrum. The problem is to identify the unknown parameters of the function $g(x, u, t)$, the constants c_o and $\varkappa$, and the elements of the matrices F and G. This identification problem can be solved immediately using the technique described in the report. To obtain the likelihood function we first write $e(t)$ as a function

of the inputs and the observations. We get

$$x(t+1) = g[x(t),\ u(t),\ t]$$

$$z(t+1) = Fz(t) + \frac{1}{c_o}\, G\,[y(t) - x_1(t) - z_1(t) - \varkappa]$$

$$\varepsilon(t) = c_o\, e(t) = y(t) - x_1(t) - z_1(t) - \varkappa$$

and the logarithm of the likelihood function is

$$-L(y;\ \theta) = \frac{1}{2c_o^2} \sum_{t=1}^{N} \varepsilon^2(t) + N \log c_o + \frac{N}{2} \log 2\pi$$

We can now proceed in exactly the same way as was done in section 3 to obtain an algorithm to maximize $L(y|\theta)$.

The results can also be generalized in a different direction. So far, we have assumed that the estimate should be calculated from a complete record of inputs and observations. Such a situation is referred to as off-line estimation. In certain applications, particularly in connection with adaptive control, the problem is different because the inputs and outputs are obtained recursively in time. This situation is referred to as on-line identification. Due to the recursive structure of the computation scheme only minor modifications are required to obtain an on-line identification. Some preliminary numerical experiments with very encouraging results have been performed.

The choice of model structure has not been discussed in this paper. There is, however, one point we would like to comment upon. The important feature of the model (2.1) is that it contains only one noise source e(t). This is essential, for the reason that it enables us to solve (2.1) for e(t) in terms of y(t) and u(t). According to the representation theorem for stationary random processes it is always possible to find a representation such as (2.1). However, in many cases a different representation

would appear more natural. Consider for example the case when there are independent measurement errors. In such a case we would obtain a model of the type

$$A(z^{-1})\, x(t) = B(z^{-1})\, u(t) + \lambda C(z^{-1})\, e(t)$$

$$y(t) = x(t) + \mu v(t) \qquad (7.1)$$

where $\{e(t)\}$ and $\{v(t)\}$ are sequences of independent equally distributed $(0, 1)$ random variables. The disturbances $e(t)$ and $v(t)$ represent process disturbances and measurement errors. In the model (7.1) we thus have two noise sources and one output. To solve the identification problem for the model (7.1) we can proceed by the method at Maximum Likelihood. The negative logarithm of the likelihood function is

$$-\log L = \frac{1}{2\lambda^2} \sum_{t=1}^{N} \varepsilon^2(t) + \frac{1}{2\mu^2} \sum_{t=1}^{N} [y(t) - x(t)]^2$$

$$-\frac{1}{2} N \log \lambda\mu + \text{const.}$$

where

$$C(z^{-1})\, \varepsilon(t) = A(z^{-1})\, x(t) - B(z^{-1})\, u(t)$$

Analyzing the details, we find that the problem of maximizing the likelihood leads to a two point boundary value problem for the equation (7.1) and its adjoint. The computational aspects of this have been investigated and tried. We have found that the computations are much more involved and time consuming than the corresponding computations for the model (2.1). It is also much easier to solve the minimum variance control problem for the model (2.1). The reason for this is that the necessary spectral factorization has already been carried out.

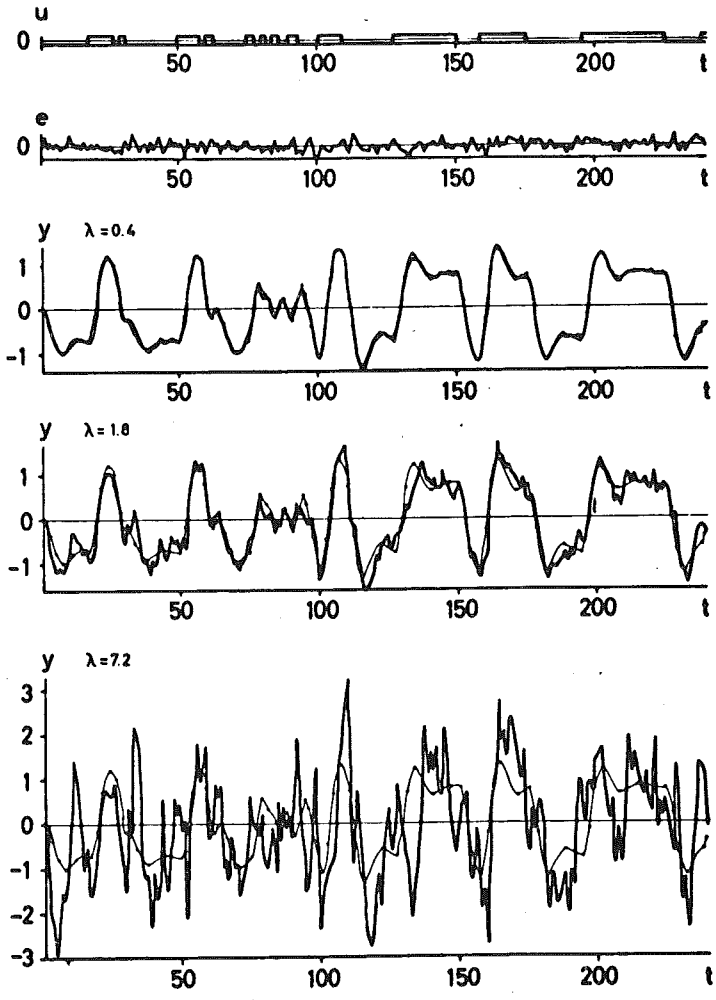

Figure 1 Input sequence, sequence of random numbers, and output sequences for test example.

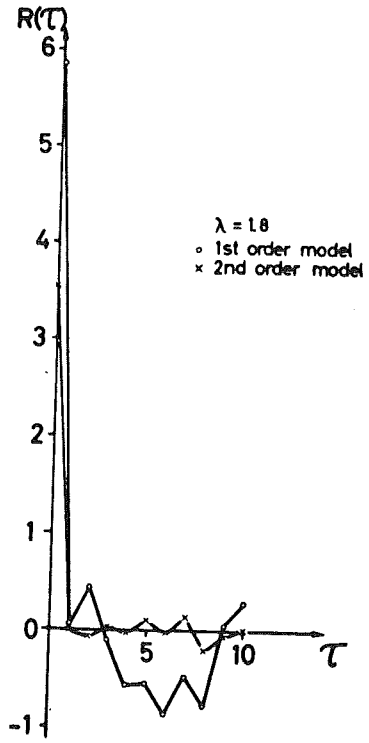

Figure 2 Covariance functions for residuals of first and second order models of second order system.

TABLE I

Estimated and true parameter values for cases 1, 2 and 3.

Parameter	CASE 1	CASE 2	CASE 3	TRUE
$-a_1$	1.512 ± 0.008	1.544 ± 0.03	1.586 ± 0.06	1.500
a_2	0.705 ± 0.005	0.720 ± 0.02	0.722 ± 0.06	0.700
b_1	1.025 ± 0.04	1.161 ± 0.16	1.338 ± 0.6	1.000
b_2	0.413 ± 0.05	0.076 ± 0.2	-0.313 ± 0.6	0.500
$-c_1$	0.978 ± 0.06	1.015 ± 0.07	1.039 ± 0.10	1.000
c_2	0.158 ± 0.06	0.151 ± 0.07	0.143 ± 0.07	0.200
λ	0.419 ± 0.019	1.880 ± 0.08	7.572 ± 0.3	

TABLE II

Successive estimates of the parameters

STEP	c_1	c_2	b_1	b_2	$a_1 - c_1$	$a_2 - c_2$	LOSS FUNCTION
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	17794.97
1	0.000000	0.000000	1.793699	1.215727	-0.669223	0.067462	7696.58
2	-0.953107	0.036294	1.939274	-1.258858	-0.658175	0.654775	7162.93
3	-0.992611	0.108536	1.370642	-0.282776	-0.558668	0.583354	6891.57
4	-1.038508	0.134742	1.389544	-0.396304	-0.549532	0.585416	6882.01
5	-1.035053	0.142406	1.332628	-0.294744	-0.548205	0.578284	6880.34
6	-1.038635	0.142974	1.337403	-0.313218	-0.547242	0.578912	6880.12
7	-1.038668	0.143086	1.337638	-0.313263	-0.547204	0.578810	6880.12
8	-1.038671	0.143088	1.337638	-0.313265	-0.547203	0.578809	

TABLE III

First order partial derivatives of the loss function

STEP	$\frac{\partial V}{\partial c_1}$	$\frac{\partial V}{\partial c_2}$	$\frac{\partial V}{\partial b_1}$	$\frac{\partial V}{\partial b_2}$	$\frac{\partial V}{\partial (a_1 - c_1)}$	$\frac{\partial V}{\partial (a_2 - c_2)}$
0	0.0000	0.0000	1339.5956	-1450.3118	——	——
1	197.1321	-704.2079	0.0001	0.0000	-0.0001	-0.0049
2	9352.8641	8950.7504	532.4016	513.5850	——	——
3	-479.7225	-960.5074	78.6591	97.8004	1007.8257	1459.4972
4	-167.4281	-163.6832	-20.8965	-27.5368	241.4333	300.0271
5	8.0683	-8.4731	14.5029	15.7974	-107.2980	-124.7181
6	3.0052	2.3351	-0.6982	0.6224	8.6651	9.0144
7	0.0131	-0.0039	-0.0010	0.0000	0.0030	0.0011

9. REFERENCES

[1] Åström, K.J., Bohlin, T.& Wensmark, S. Automatic construction of linear stochastic dynamic models for stationary industrial processes with random disturbances using operating records. IBM Nordic Laboratory, Sweden, June 1, 1965 (Report TP 18.150).

[2] Åström, K.J., Control Problems in Paper Making. IBM Scientific Computing Symposium: Control Theory and Applications, New York, Oct. 20, 1964.

[3] Bellman, R.E. & Kalaba, R.E., Quasilinearization and nonlinear boundary-value problems. New York, Elsevier, 1965.

[4] Blandhol, E., On the use of adjustable models for determination of system dynamics. Institutt for Reguleringsteknikk, NTH, Trondheim, Norway, March 1962. (Technical report no. 62-5-D).

[5] Box, G.E.P., Fitting empirical data. Annals New York Academy of Sciences. 86 (1960): p. 792-816.

[6] Cramér, H., Mathematical methods of statistics. Princeton Univ. Press, (1946) 1958.

[7] Durbin, J., Efficient fitting of linear models for continuous stationary time series from discrete data. Bulletin de l'Institut International de Statistique. Tokyo. 38 (1960): 4, p. 273-?8?.

[8] Eykhoff, P., Process parameter estimation. Progress in control engineering. Ed. R.H. MacMillan. Vol. ?. London, Heywood, 1964, p. 161-208.

[9] Galtieri, C.A., Problems of estimation in discrete-time processes. IBM Research Laboratory. San Jose, Calif. Aug. 26, 1964. (Report RJ-315).

[10] Grenander, U., Stochastic processes and statistical inference. Arkiv för Matematik. 1 (1950): 17, p. 195-277.

[11] Kale, B.K., On the solution of likelihood equations by iteration processes. The multiparametric case. Biometrica. 49 (1962): p. 479-486.

[12] Kalman, R.E., Design of a self-optimizing control system. ASME Trans. 80 (1958): 2, p. 468-478.

[13] Kalman, R.E., Mathematical description of linear dynamical systems. Journal on Control. 1 (1963): 2, p. 152-192.

[14] Kendall, M.G. & Stuart, A., The advanced theory of statistics. Vol. 2. London, Griffin, 1961.

[15] Levin, M.J., Estimation of the characteristics of linear systems in the presence of noise. D.Sc. thesis. Columbia Univ., New York, 1959.

[16] Mann, H.B. & Wald, A., On the statistical treatment of linear stochastic difference equations. Econometrica. 11 (1943): 3 & 4, p. 173-220.

[17] Margolis, M. & Leondes, C.T., On the theory of adaptive control systems: The learning model approach. 1st International Congress IFAC. Automatic and remote control. Moscow. 1960. Proceedings, Vo. 2, p. 556-563.

[18] Maslov, E.P., Application of the Theory of Statistical Decisions to the Estimation of Object Parameters, Automation and Remote Control, 24 (1964) 1214-1226.

[19] Mayne, D.Q., Estimation of system parameters. Symposium on optimal control. Imperial College of Science and Technology, London. 1964. Paper No. 9.

[20] Meissinger, H.F., Parameter influence coefficients and weighting functions applied to perturbation analysis of dynamic systems. 3rd International analogue computation meeting. Opatija, Yugoslavia. 1961. Proceedings, p. 207-216.

[21] Mishkin, E. & Brown, L. (editors) Adaptive Control Systems, McGraw Hill, New York, 1961.

[22] Ragazzini, J.R. & Franklin, G.F., Sampled data control systems, New York, McGraw Hill, 1958.

[23] Walker, A.M., Large-sample estimation of parameters for moving average models. Biometrica. 48 (1961): p. 343-357.

[24] Westcott, J.H., The parameter estimation problem. 1st International Congress IFAC. Automatic and remote control. Moscow. 1960. Proceedings. Vol. 2. p. 779-787.

[25] Whitaker, H.P. et al., Design of a model reference adaptive control system for aircraft. MIT Instrumentation Laboratory. Boston. 1958. (Report R-164).

[26] Whittle, P., Estimation and information in stationary time series. Arkiv för Matematik. 2 (1952): 23. p. 423-434

[27] Wilks, S.S., Mathematical statistics. New York, Wiley, 1962.

NUMERICAL IDENTIFICATION OF LINEAR DYNAMIC SYSTEMS FROM NORMAL OPERATING RECORDS

TP 18.159
15 July 1967

K. J. Åström T. Bohlin