On the Achievable Accuracy in Identification Problems

Åström, Karl Johan

Published in:
IFAC symposium on identification in automatic control systems; Part 1

1967

ON THE ACHIEVABLE ACCURACY IN IDENTIFICATION PROBLEMS

K.J. Åström

Lund Institute of Technology, Lund, Sweden

INTRODUCTION

When solving an identification problem, two things are of major interest: to obtain a model of the system and to obtain an estimate of the accuracy of model. To obtain accuracy estimates it is necessary to put the problem in a statistical framework. The identification problem can then be stated as a statistical estimation problem. Having done this, there are many results from statistical estimation theory that can be applied to obtain results of practical value. In particular a lower bound on the accuracy of the model is given by the Cramér-Rao inequality (5). By exploiting this wellknown inequality we can thus obtain the accuracies that could possibly be achieved in any given situation. In this paper we present formulas for estimating the accuracy of an identification problem based on the Cramér-Rao theorem. The results are of interest when planning experiments on industrial processes.

The problem is stated and the solution is outlined in section 2 of the paper. In section 3 we give the solution for discrete time systems. The problem is discussed with reference to the maximum likelihood procedure. However, we also show how the results can be applied to generalized least squares and model adjustment procedures (7). Continuous time systems are covered in section 4 and in section 5 we summarize the results, outline various generalizations and give some aspects on the choice of model structures. This discussion will further illuminate the properties of maximum likelihood and least squares procedures.

STATEMENT OF THE PROBLEM AND AN OUTLINE OF ITS SOLUTION

Problem Statement

Consider a dynamical system whose inputs and outputs are observed. For the sake of simplicity in the presentation we will only consider single-input single-output systems. We will analyse the following.

Problem

Assume that the input u and the corresponding output y over a time interval are given. How accurate can the coefficients of a model relating the input u to the output y be determined?

To solve this problem we will use the following basic results of Cramér and Rao (5).

Theorem (Cramér-Rao)

Let x be a random vector whose density function $f(x;\theta)$ contains a set of parameters $\theta = \mathrm{col}(\theta_1, \ldots, \theta_n)$. Assume that f is continuous in θ and continuously differentiable with respect to $\theta \in \Omega$. Let $\hat{\theta}$ be an unbiased estimate of θ, then

$$E(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T \geq J^{-1} \qquad (2.1)$$

where J is the information matrix defined by

$$J = E\, L_\theta\, L_\theta^T \qquad (2.2)$$

and $L = \log f(x;\theta)$ and L_θ denotes the gradient of L with respect to θ. If $f(x;\theta)$ is twice continuously differentiable with respect to θ we have further

$$E\, L_\theta\, L_\theta^T = -E\, L_{\theta\theta} \qquad (2.3)$$

where $L_{\theta\theta}$ denotes the matrix of second order partial derivatives.

This theorem is usually stated for the case that the components x_i of x are independent samples of a random variable in which case the function $f(x;\theta)$ has a simple form. This does not cover our problem. The actual proof as given in (8) do however contain the results given above. The restriction that $\hat{\theta}$ is an unbiased estimate can be removed with a slight modification of the result (8).

To obtain a lower bound of the accuracy of an identification problem we will thus proceed as follows: By making suitable assumptions the problem is first formulated as a statistical parameter estimation problem. The logarithm of the likelihood function $L = \log f(x;\theta)$ associated with the estimation problem is then computed. The information matrix J is then evaluated by taking mathematical expectation of derivatives of the likelihood function. In the following sections we will show how this is carried out in detail for particular cases.

Minimum Variance Estimators

The Cramér-Rao theorem gives a lower bound for the covariance of an estimator. It is of course of great practical value to construct estimators for which the Cramér-Rao lower bound is achieved. In many cases the maximum likelihood estimators will, at least asymptotically, have this property. There is, however, no proof of this which is general enough to cover the situations that occur in connection with process identification. The (asymptotic) minimum variance property thus has to be established in each particular case.

DISCRETE TIME SYSTEMS

Main Result

Consider a linear single-input single-output system which is governed by the equation

$$A(z^{-1})\, y(t) = B(z^{-1})\, u(t-k) + \lambda\, C(z^{-1})\, e(t) \qquad (3.1)$$

where u is the input, y the output and {e(t), t = 0, ±1, ±2, ...} a sequence of independent normal (0,1) random variables. The variable e(t) is independent of e(s) for s ≠ t and e(t) is also independent of y(t) and u(t-k). Introduce z to denote the shift operator

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

and the polynomials A, B and C defined by

$$A(x) = 1 + a_1 x + \dots + a_n x^n$$
$$B(x) = b_0 + b_1 x + \dots + b_n x^n \qquad b_0 \neq 0$$
$$C(x) = 1 + c_1 x + \dots + c_n x^n$$

It is assumed that the functions $x^n A(x^{-1})$ and $x^n C(x^{-1})$ have all zeros inside the unit circle. We will now investigate how accurate the parameters of the model (3.1) can be determined from observations of inputs and outputs of the model. We introduce the symbol θ

$$\theta = (a_1, \dots, a_n, b_0, b_1, \dots, b_n, c_1, \dots, c_n)$$

as the vector containing all parameters to be determined except λ. If some parameters are known or related to other parameters they are simply deleted in the vector θ.

Now assume that a sequense of inputs (u(t), t = 1, 2, ..., N) and corresponding observations of the output (y(t), t = 1, 2, ..., N) are known. How accurate can the coefficients θ of the model (3.1) then be determined.

For this particular case (2) the logarithm of the likelihood function is given by

$$L = -\frac{1}{2\lambda^2}\sum_{t=1}^{N} \varepsilon^2(t) - N \log \lambda + \text{const} \qquad (3.2)$$

where

$$C(z^{-1})\,\varepsilon(t) = A(z^{-1})\,y(t) - B(z^{-1})\,u(t-k) \qquad (3.3)$$

Hence

$$\frac{\partial^2 L}{\partial \lambda^2} = -\frac{3}{\lambda^4}\sum_{t=1}^{N} \varepsilon^2(t) + \frac{N}{\lambda^2}$$

and

$$\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = -\frac{1}{\lambda^2}\sum_{t=1}^{N} \frac{\partial \varepsilon(t)}{\partial \theta_i}\cdot\frac{\partial \varepsilon(t)}{\partial \theta_j} - \frac{1}{\lambda^2}\sum_{t=1}^{N} \varepsilon(t)\cdot \frac{\partial^2 \varepsilon(t)}{\partial \theta_i \partial \theta_j}$$

$$\frac{\partial^2 L}{\partial \theta_i \partial \lambda} = +\frac{2}{\lambda^3}\sum_{t=1}^{N} \varepsilon(t)\,\frac{\partial \varepsilon(t)}{\partial \theta_i}$$

By taking derivatives of (3.3) with respect to the parameters we find

$$C\,\frac{\partial \varepsilon(t)}{\partial a_i} = z^{-i}\,y(t) = \frac{B}{A}\,u(t-k-i) + \lambda\,\frac{C}{A}\,e(t-i)$$

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

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

where the arguments of A, B and C are z^{-1}.

Compare reference (2). Now take mathematical expectation and we find

$$E\,\frac{\partial^2 L}{\partial \lambda^2} = -\frac{2N}{\lambda^2}$$

$$E\,\frac{\partial^2 L}{\partial \theta_i \partial \lambda} = 0$$

$$E\,\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = -\frac{1}{\lambda^2}\sum_{t=1}^{N} E\,\frac{\partial \varepsilon(t)}{\partial \theta_i}\cdot\frac{\partial \varepsilon(t)}{\partial \theta_j}$$

Introduce the variables

$$x_1(t) = \frac{B(z^{-1})}{A(z^{-1})\,C(z^{-1})}\,u(t)$$

$$x_2(t) = \frac{1}{C(z^{-1})}\,u(t)$$

$$x_3(t) = \frac{1}{A(z^{-1})}\,e(t)$$

$$x_4(t) = \frac{1}{C(z^{-1})}\,e(t)$$

As the functions $z^n A(z^{-1})$ and $z^n C(z^{-1})$ have all zeros inside the unit circle $x_3(t)$ and $x_4(t)$ are stationary random processes and we find for i,j = 3,4

$$\frac{1}{N}\sum_{t=1}^{N} E\,x_i(t)\,x_j(t+\tau) = E\,x_i(t)\,x_j(t+\tau) = r_{ij}(\tau)$$

where $r_{ij}(\tau)$ is the covariance function of $x_i(t)$ and $x_j(t)$ for i,j = 3,4. We further introduce the notations

$$r_{ij}^{N}(\tau) = \frac{1}{N}\sum_{t=1}^{N} x_i(t)\,x_j(t+\tau) \qquad i,j = 1,2$$

and we then find that the information matrix for the parameter estimation problem associated with the model (3.1) can be expressed as

$$J = \begin{bmatrix} J_{aa}^{N} & J_{ab}^{N} & J_{ac} & 0 \\ J_{ba}^{N} & J_{bb} & 0 & 0 \\ J_{ca} & 0 & J_{cc} & 0 \\ 0 & 0 & 0 & 2\lambda^{-2}N \end{bmatrix} \qquad (3.4)$$

where

$$(J_{aa}^{N})_{ij} = N(\lambda^{-2} r_{11}^{N}(i-j) + r_{33}(i-j))$$

$$(J_{ab}^{N})_{ij} = -N\lambda^{-2} r_{12}^{N}(i-j)$$

$$(J_{ac})_{ij} = -N r_{34}(i-j)$$

$$(J_{bb}^{N})_{ij} = N\lambda^{-2} r_{22}^{N}(i-j)$$

$$(J_{cc})_{ij} = r_{44}(i-j)$$

and

$$J_{ba} = J_{ab}^{T}$$

$$J_{ca} = J_{ac}^{T}$$

Now assume that the quantities u(t) and u(t) u(t + τ) are Cesaro summable i.e. that the following limits exist

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

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} u(t)\, u(t+\tau) \qquad (3.5)$$

The fact that the functions $z^{n}A(z^{-1})$ and $z^{n}C(z^{-1})$ have all zeros inside the unit circle then implies that the limits

$$\lim_{N\to\infty} r_{ij}^{N}(\tau) = r_{ij}(\tau) \qquad i,j = 1,2$$

exist.

Summarizing the results and applying the Cramér-Rao theorem we find.

Theorem 1

Assume that the limits (3.5) exist. The covariance matrix of the estimates of the parameters of the model (3.1) is then bounded by the inequality

$$E(\hat{\theta} - \theta)(\hat{\theta} - \theta)^{T} \geq J^{-1}$$

where the information matrix J is given by (3.4).

Notice that the theorem only gives a lower bound for the accuracy of the parameter estimates. Having obtained the information matrix, two problems then remains: to construct an estimator and to investigate whether the lower bound is actually achieved for that estimator. For the particular model (3.1) we have in an earlier paper (2) constructed an estimator which at least asymptotically has the minimum variance.

The accuracy (standard deviation) of the parameter θ_i is thus bounded by the square root of the i:th diagonal of the information matrix J i.e. $(J^{-1})_{ii}$.

An analysis of the information matrix gives a good insight into the identification problem. If the matrix is singular there are too many parameters in the problem and only linear combinations of the parameters can be estimated.

Notice that the elements of the information matrix are composed of two types of terms. One term depends only on the characteristics of the disturbance and the other term depends on the input signal. It is thus possible to investigate the influence of different input signals on the accuracy of the parameters.

The information matrix can be computed in several different ways. For simple problems involving systems of low order it can be computed by residue calculus. We observe e.g. that

$$r_{33}(\tau) = \frac{1}{2\pi i} \oint A^{-1}(\xi^{-1})\, A^{-1}(\xi)\, \xi^{\tau-1}\, d\xi$$

$$r_{34}(\tau) = \frac{1}{2\pi i} \oint A^{-1}(\xi^{-1})\, C^{-1}(\xi)\, \xi^{\tau-1}\, d\xi$$

$$r_{44}(\tau) = \frac{1}{2\pi i} \oint C^{-1}(\xi^{-1})\, C(\xi)\, \xi^{\tau-1}\, d\xi$$

For functions such that $\Sigma|x(t)| < \infty$ Parsevals theorem i.e.

$$\sum_{t=0}^{\infty} x(t)\, y(t) = \frac{1}{2\pi i} \oint X(\xi)\, Y(\xi^{-1})\, \xi^{-1}\, d\xi$$

where

$$X(\xi) = \sum_{t=0}^{\infty} \xi^{-t}\, x(t)$$

is sometimes useful to compute the terms r_{11}, r_{12} and r_{22}.

When analytical computation is not feasible the information matrix can be computed in a straight forward way if we observe that it follows by Parsevals theorem that

$$r_{33}(\tau) = \sum_{t=0}^{\infty} a(t)\, a(t+\tau)$$

where a(t) is the impulse response of the dynamical system

$$A(z^{-1})\, a(t) = 0$$

Example 1

Consider a stationary normal random process {y(t)} generated by

$$y(t) = \lambda \frac{z + c}{z + a}\, e(t) \qquad (3.6)$$

where {e(t)} is a sequence of independent normal N(0,1) random variables. How accurate can the parameters a, c and λ be estimated from N observations of y?

We have

$$\varepsilon(t) = \frac{z+a}{z+c}\, y(t)$$

$$\frac{\partial \varepsilon}{\partial a} = \frac{1}{z+c}\, y(t) = \frac{\lambda}{z+a}\, e(t)$$

$$\frac{\partial \varepsilon}{\partial c} = -\frac{1}{z+c}\, \varepsilon(t) = -\frac{\lambda}{z+c}\, e(t)$$

Further

$$E\left(\frac{\partial \varepsilon}{\partial a}\right)^2 = \frac{1}{2\pi i}\oint \frac{\lambda^2}{(z+a)(1+az)}\, dz = \frac{\lambda^2}{1-a^2}$$

$$E\left(\frac{\partial \varepsilon}{\partial a}\right)\left(\frac{\partial \varepsilon}{\partial c}\right) = -\frac{1}{2\pi i}\oint \frac{\lambda^2}{(z+a)(1+cz)} = -\frac{\lambda^2}{1-ac}$$

$$E\left(\frac{\partial \varepsilon}{\partial c}\right)^2 = \frac{1}{2\pi i}\oint \frac{\lambda^2}{(z+c)(1+cz)}\, dz = \frac{\lambda^2}{1-c^2}$$

and the information matrix becomes

$$J = N \begin{bmatrix} \frac{1}{1-a^2} & \frac{-1}{1-ac} & 0 \\ \frac{-1}{1-ac} & \frac{1}{1-c^2} & 0 \\ 0 & 0 & 2\lambda^{-2} \end{bmatrix}$$

The covariance matrix of the estimates is thus bounded by

$$J^{-1} \geq \frac{1}{N} \begin{bmatrix} \frac{(1-a^2)(1-ac)^2}{(a-c)^2} & \frac{(1-a^2)(1-c^2)(1-ac)}{(a-c)^2} & 0 \\ \frac{(1-a^2)(1-c^2)(1-ac)}{(a-c)^2} & \frac{(1-c^2)(1-ac)^2}{(a-c)^2} & 0 \\ 0 & 0 & \lambda^2/2 \end{bmatrix}$$

The standard deviations of the parameters are thus bounded by

$$\sigma_a \geq \frac{1-ac}{a-c}\sqrt{\frac{1-a^2}{N}}$$

$$\sigma_b \geq \frac{1-ac}{a-c}\sqrt{\frac{1-c^2}{N}}$$

$$\sigma_\lambda \geq \lambda\sqrt{\frac{2}{N}}$$

Introducing the numerical values $a = -0.5$, $c = 0.5$ and $\lambda = 1$ we find

$$J = \frac{N}{100} \begin{bmatrix} 133 & -80 & 0 \\ -80 & 133 & 0 \\ 0 & 0 & 200 \end{bmatrix}$$

$$J^{-1} = \frac{100}{N} \begin{bmatrix} 0.012 & 0.007 & 0 \\ 0.007 & 0.012 & 0 \\ 0 & 0 & 0.005 \end{bmatrix}$$

For $N = 100$ and $N = 500$ we thus get the following standard deviations of the minimum variance estimates of the parameters

$\sigma_a = 0.11 \quad (0.05)$

$\sigma_c = 0.11 \quad (0.05)$

$\sigma_\lambda = 0.07 \quad (0.03)$

In table I we give the results of some numerical experiments.

Table I

Minimum variance estimates of the parameters of model (3.6)

Sample	N	$\hat{a}$	$\hat{c}$	$\hat{\lambda}$
1	100	-0.153	0.669	0.950
2	100	-0.448	0.561	1.036
3	100	-0.353	0.571	1.038
4	100	-0.474	0.463	0.870
5	100	-0.473	0.512	1.065
6	100	-0.387	0.571	1.058
7	100	-0.618	0.461	0.889
8	100	-0.512	0.546	0.918
9	100	-0.483	0.452	0.953
10	100	-0.614	0.491	0.972
11	500	-0.511	0.516	0.953

The results of the numerical experiment are also illustrated in Fig. 1 where the estimates are plotted together with the concentration ellipsoid.

The information matrix J can be diagonalized in the following way

$$J = \frac{N}{200} \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 53 & 0 & 0 \\ 0 & 213 & 0 \\ 0 & 0 & 400 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This implies that the linear combinations $a + c$ and $a - c$ can be estimated with standard deviations bounded by $\sqrt{200/53N}$ and $\sqrt{200/213N}$ respectively. Compare Fig. 1.

We have thus seen how theorem 1 can be applied to a lower bound of the accuracy of a parameter estimation problem. We will now demonstrate how the results can be applied to other identification procedures.

Application to Generalized Least Squares

In the generalized least squares procedure (7), (9), (12) a linear model is fitted to filtered input-output data in such a way that the deviation between the filtered process output and the output of the model is as small as possible in the sense of least squares. Let the input signal be u, the process output be y, and let the filter be characterized by the operator G. The generalized least squares procedures can then be formulated as follows. Find the coefficients of the polynomials A and B such that

$$\sum_{t=1}^{N} \varepsilon^2(t)$$

is minimum, where

$$\varepsilon(t) = A(z^{-1})\,\{G(z^{-1})\,y(t)\} - B(z^{-1})\,\{G(z^{-1})\,u(t)\} \qquad (3.7)$$

Hence

$$G^{-1}(z^{-1})\,\varepsilon(t) = A(z^{-1})\,y(t) - B(z^{-1})\,u(t) \qquad (3.7)$$

If it is now assumed that the process is actually governed by a linear model and that the residuals are a sequence of independent normal $N(0,\lambda)$ random variables we find that the generalized least squares procedures is identical to the parameter estimation problem stated in section 3.1. We can further identify G^{-1} with C. Hence estimation of the parameters of the model (3.1) is equivalent to a generalized least squares method where we simultaneously estimate the best filter to be used in the generalized least squares procedure. With the above assumption the results of section 3.1 can thus be applied to the generalized least squares procedure.

Application to Model Adjustment Technique

Consider a model adjustment procedure (4), (7) where a linear model

$$y_m(t) = \frac{B(z^{-1})}{A(z^{-1})}\,u(t) \qquad (3.8)$$

is adjusted to measured input-output data. The coefficients of the model A, B are adjusted in such a way that the error

$$\varepsilon(t) = y(t) - y_m(t)$$

is as small as possible in the sense of least squares. If it is now assumed that the input output data is generated by a process of the type (3.8) and that the deviations between the outputs of the process and the model are independent normal random variables we find that the model adjustment procedure can be interpreted as the following parameter estimation problem.

Consider a process

$$y(t) = \frac{B}{A}\,u(t) + \lambda\,e(t)$$

where $\{e(t)\}$ are normal (0,1) random variables Determined an estimate of the coefficients of A, B from an observation of the inputs $\{u(t), t = 1,2,\ldots,N\}$ and the corresponding outputs $\{y(t), t = 1,2,\ldots,N\}$. This is however the parameter estimation problem that was discussed in section 3.1 with $C \equiv A$, and the results of section 3.1 can thus be applied. If the deviations between the output of the process and that of the model are not independent it is possible to introduce a filter in the same way as was done in the generalized least squares procedure.

CONTINUOUS TIME SYSTEMS

We will now extend the results of section 3 to continuous time systems. The model (3.1) then has to be substituted by a stochastic differential equation. In the analysis we then have the usual difficulties associated with those equations. Space does not permit a completely rigorous analysis, so we will proceed heuristically using delta functions. Consider a linear single input single output system governed by

$$A(p)\,y(t) = B(p)\,u(t) + \lambda\,C(p)\,e(t) \qquad (4.1)$$

where u is the input, y the output and e white gaussian noise

$$E\,e(t)\,e(t+\tau) = \delta(\tau)$$

The symbol p denotes the derivation operator

$$px = \frac{dx}{dt}$$

and A, B and C are polynomials

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

$$B(s) = b_0 s^n + b_1 s^{n-1} + \ldots + b_n$$

$$C(s) = s^n + c_1 s^{n-1} + \ldots + c_n$$

It is assumed that functions A(s) and C(s) have all zeros in the left half plane. We will now investigate how accurate the parameters of the model (4.1) can be determined from observations of inputs $\{u(t), 0 \le t \le T\}$ and outputs $\{y(t), 0 \le t \le T\}$. Following the procedure outlined in section 2 we first have to determine the likelihood function. According to Arato (1) and Striebel (11) we find that the logarithm of the likelihood function is given by

$$L = -\frac{1}{2\lambda^2}\int_0^T \varepsilon^2(t)\,dt - T\log\lambda + \text{const.} \qquad (4.2)$$

where

$$C(p)\,\varepsilon(t) = A(p)\,y(t) - B(p)\,u(t) \qquad (4.3)$$

Now proceeding in the same way as we did in section 3 we get

$$\frac{\partial^2 L}{\partial \lambda^2} = -\frac{3}{\lambda^2}\int_0^T \varepsilon^2(t)\,dt + \frac{I}{\lambda^2}$$

$$\frac{\partial^2 L}{\partial \lambda \partial \theta_i} = \frac{2}{\lambda^3}\int_0^T \varepsilon(t)\,\frac{\partial \varepsilon(t)}{\partial \theta_i}\,dt \qquad (4.4)$$

$$\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = -\frac{1}{\lambda^2}\int_0^T \frac{\partial \varepsilon(t)}{\partial \theta_i}\cdot\frac{\partial \varepsilon(t)}{\partial \theta_j}\,dt - \frac{1}{\lambda^2}\int_0^T \varepsilon(t)\,\frac{\partial^2 \varepsilon(t)}{\partial \theta_i \partial \theta_j}\,dt \qquad (4.5)$$

where θ denotes all parameters of (4.1) except λ.

$$C(p)\,\frac{\partial \varepsilon}{\partial a_i} = p^{n-i}y$$

$$C(p)\,\frac{\partial \varepsilon}{\partial b_i} = -p^{n-i}u$$

$$C(p)\,\frac{\partial \varepsilon}{\partial c_i} = -p^{n-i}\varepsilon$$

where $\theta = (a_1,\ldots,a_n,b_0,b_1,\ldots b_n,c_1,\ldots c_n)$. If the stochastic differential equation is defined properly the mathematical expectation of equation (4.4) and of the last term of (4.5) will vanish and we get

$$E\,\frac{\partial^2 L}{\partial \lambda^2} = -\frac{2T}{\lambda^2}$$

$$E\,\frac{\partial^2 L}{\partial \lambda \partial \theta_i} = 0$$

$$E\,\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = -\frac{1}{\lambda^2}\int_0^T E\,\frac{\partial \varepsilon(t)}{\partial \theta_i}\,\frac{\partial \varepsilon(t)}{\partial \theta_j}$$

Introduce the variables

$$x_1 = \frac{B(p)}{A(p)\,C(p)}\,u$$

$$x_2 = \frac{1}{C(p)}\,u$$

$$x_3 = \frac{1}{A(p)}\,e$$

$$x_4 = \frac{1}{C(p)}\,e$$

As the functions A(s) and C(s) have zeros only in the left half plane x_3 and x_4 are thus stationary gaussion processes and we find

$$E\int_0^T (p^{n-i}x_k)(p^{n-j}x_\ell)\,dt = (-1)^{n-i}\,T\,p^{2n-i-j}r_{k\ell}(0)$$

$$k,\ell = 3,4 \qquad (4.6)$$

where

$$r_{k\ell}(\tau) = E\,x_k(t)\,x_\ell(t+\tau)$$

If we further introduce

$$\frac{1}{T}\int_0^T x_k(t)\,x_\ell(t+\tau)\,dt = r_{k\ell}{}^T(\tau)$$

we find that the information matrix can be expressed as

$$J = \begin{bmatrix} J_{aa}{}^T & J_{ab}{}^T & J_{ac} & 0 \\ J_{ba}{}^T & J_{bb} & 0 & 0 \\ J_{ca} & 0 & J_{cc} & 0 \\ 0 & 0 & 0 & 2T\,\lambda^{-2} \end{bmatrix}$$

where the elements of submatrices such as J_{aa} are easily computed from (4.6) in full analogy with the discrete time case. This result cures the obscurity of (10) for multiparameter systems. Analogous to the discrete time case we can again use Parsevals theorem and complex function theory to evaluate the elements of the information matrix.

CONCLUSIONS

Given a particular model of a dynamical system we have shown how accurate the parameters of the model can possibly be determined by an identification procedure. It was a consequence of the Cramér-Rao inequality that a lower limit to the covariance of the parameter estimates was given by the information matrix. We have shown in detail how to evaluate the information matrix in typical cases. The results obtained can also be extended to nonlinear systems with a known structure. Notice that the results are given in terms of lower limits. It then remains to construct the estimators and to investigate whether or not the lower limits are actually achieved for a particular estimator. Also notice that the results are based on the assumption that the model is known e.g. from a priori knowledge or from results of process measurements. A sensitivity study often reveals that the results do not depend critically on the model parameters. The results have been successfully applied to the determination of

process dynamics in cases where approximative models have been available from pilot experiments. In particular the results have proven useful to analyse the relative merits of different input signals in situations where experiments must be performed during normal operation, and it is of extreme importance to keep process variables within strict limits during the experiment.

It should also be noticed that we assume that the structure of the system is given. This is probably the most restricting assumption at least in the case where the model is determined by measurements on the process. Hence in order to apply the results we must first ensure that the system has the correct structure. If this is not done properly serious mistakes can be made. We will not discuss this in general but we will illustrate what can happen in a particular case.

Example 2

Consider a sampled linear single-input single-output system with a time delay that is an integer of the sampling interval. The input output relation can be described as follows

$$y(t) = \frac{B'(z^{-1})}{A'(z^{-1})}\, u(t-k) \qquad (5.1)$$

Because of linearity the disturbances can be represented as an equivalent disturbance d(t) in the output

$$y(t) = \frac{B'(z^{-1})}{A'(z^{-1})}\, u(t-k) + d(t) \qquad (5.2)$$

If the disturbance d(t) is a stationary random process with a rational power spectral density it can always be represented as

$$d(t) = \lambda\, \frac{C'(z^{-1})}{D'(z^{-1})}\, e(t) \qquad (5.3)$$

where $(e(t),\ t = 0, \pm1, \pm2, \ldots)$ is a sequence of independent equally distributed random variables. The polynomials $C'(z^{-1})$ and $D'(z^{-1})$ can always be chosen so that the functions $z^m C'(z^{-1})$ and $z^m D'(z^{-1})$ have no zeros outside the unit circle.

Introducing (5.3) into (5.2) and writing the two terms on common denominators we thus find that the input output relation can be described by the following model

$$A(z^{-1})\, y(t) = B(z^{-1})\, u(t-k) + \lambda\, C(z^{-1})\, e(t) \qquad (5.4)$$

The special case C = 1 i.e.

$$A(z^{-1})\, y(t) = B(z^{-1})\, u(t-k) + \lambda\, e(t) \qquad (5.5)$$

can be interpreted as the special case of the disturbance being an autoregression.

For linear systems we thus have at least two modelstructures represented by (5.4) and (5.5) respectively. As a spectral density can be approximated by a rational function as well as by a polynomial, there are no principal differences between the models (5.4) and (5.5). Consider e.g. the input output data shown in Fig. 2. This data was generated from the recursive equation

$$y(t) - 1.5y(t-1) + 0.7y(t-2) = u(t-1) + 0.5u(t-2) + e(t) - e(t-1) + 0.2e(t-2) \qquad (5.6)$$

If a second order model having the structure (5.5) is identified we get the following coefficients. The true parameters values are shown in brackets.

$a_1 = -1.25 \pm 0.07$ (-1.50)
$a_2 = 0.47 \pm 0.07$ (+0.70)
$b_1 = 0.83 \pm 0.16$ (1.00)
$b_2 = 0.80 \pm 0.17$ (0.50)

The estimates of the coefficients a_1 and a_2 are wrong and the accuracy estimates are much too low which is even more serious. In this case we know the reason, the least squares estimate is biased if the data was actually generated by (5.6). But suppose that we did not know the source of the data we could easily make the wrong judgement.

The difficulty can be resolved by varying the order of the model (5.5). To investigate the order required of a model with the structure given by the equation (5.5) we fit models of increasing orders to the input output data and test the significance of the results. Let V_i be the negative logarithm of the likelihood function defined by equation (3.2) when the model given by equation (5.5) contains n_i parameters. For the model (5.5) the statistic

$$\xi = \frac{V_1 - V_2}{V_2} \cdot \frac{N - n_2}{n_2 - n_1} \qquad (5.7)$$

then has an F-distribution, $F(n_2 - n_1, N - n_2)$. See reference (8). Evaluating ξ for systems of order 3, 4, 5 we get $\xi = 9.74$, $\xi = 9.47$ and $\xi = 1.46$ respectively. We have $n_2 - n_1 = 2$, $N - n_2 = 81$ and $F(2, 81) = 3.1$ at the 5% level. In this case we thus find that if the modelstructure (5.5) is used the system is of fourth order. The coefficients are given by

$a_1 = -0.76 \pm 0.10$ $\quad b_1 = 0.68 \pm 0.13$
$a_2 = 0.03 \pm 0.13$ $\quad b_2 = 1.25 \pm 0.17$
$a_3 = -0.12 \pm 0.13$ $\quad b_3 = 0.73 \pm 0.19$
$a_4 = 0.29 \pm 0.08$ $\quad b_4 = 0.45 \pm 0.17$

Analysing the results we find that within the parameter accuracy, the polynomials A(z) and B(z) have a common second order factor, $A''(z)$. This means that two states of the fourth order system are not controllable from u and that the system has the structure

$$A''(z^{-1})\, y(t) = \beta''(z^{-1})\, u(t) + \frac{\lambda}{A''(z^{-1})}\, e(t)$$

where $A'' = A/A''$ and $B'' = B/A''$ are of second order. We thus find that if the modelstructure (5.5) is used the order of the system is of fourth order to account for the colouring of the residuals.

If we identify a model with the structure (5.4) we find that the system is of second order ($\xi_2 = 21$, $\xi_3 = 0.12$) and the coefficients are given by

$a_1 = -1.54 \pm 0.04$ $\quad a_2 = 0.73 \pm 0.03$

$b_1 = \ \ 0.99 \pm 0.12$ $\quad b_2 = 0.45 \pm 0.15$

$c_1 = -0.94 \pm 0.11$ $\quad c_2 = 0.04 \pm 0.10$

Compare section 4 of reference (3).

REFERENCES

(1) Arato, M. "Estimation of the Parameters of a Stationary Gaussian Markov Process", Dokl. Akad. Nauk. SSSR 145 (1962), 13

(2) Åström, K.J. and Bohlin, T. "Numerical Identification of Linear Dynamic Systems from Normal Operating Records", Proc. IFAC Symposium on Self-Adaptive Control Systems, Teddington Sept. 1966

(3) Åström, K.J. "Computer Control of a Paper Machine - An Application of Linear Stochastic Control Theory", to appear

(4) Blandhol, E. and Balchen, J.G. "Determination of System Dynamics by the Use of Adjustable Models", Automatic and Remote Control (Proc. IFAC 1963) Butterworths, London 1964

(5) Cramér, H. "Mathematical Methods of Statistics", Princeton University Press, Princeton 1946

(6) Cuenod, M. and Sage, A. "Comparison of Some Methods Used for Process Identification", IFAC Symposium on the Problems of Identification in Automatic Control Systems, Prague 1967

(7) Eykhoff, P. "Process Parameter and State Estimation", IFAC Symposium on the Problems of Identification in Automatic Control Systems, Prague 1967

(8) Kendall, M.G. and Stuart, A. "The Advanced Theory of Statistics", Vol II, Griffin, London 1961

(9) Levadi, V.S. "Parameter Estimation of Linear Systems in the Presence of Noise", International Conference on Microwaves, Circuit Theory and Information Theory, Tokyo 7-11 Sept. 1964

(10) Qvarnström, B. "An Uncertainty Relation for Linear Mathematical Models", Automatic and Remote Control (Proc. IFAC 1963) Butterworths, London 1964

(11) Striebel, C.T. "Densities for Stochastic Processes", Ann. Math. Stat. 30 (1959) 559-67

(12) Young, P.C. "Process Parameter Estimation and Self Adaptive Control", Proc. IFAC Symposium on Self-Adaptive Control Systems, Teddington, Sept. 1966

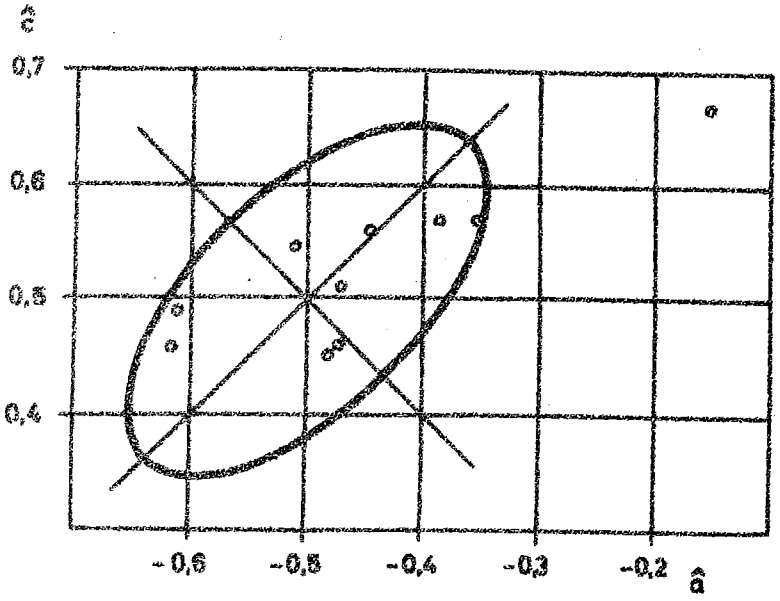

Fig. 1 - Results of numerical experiment with parameter identification described in example 1 and the concentration ellipsoid

$$133(\hat{a} + 0.5)^2 - 160(\hat{a} + 0.5)(\hat{c} - 0.5) + 133(\hat{a} + 0.5)^2 = 7$$

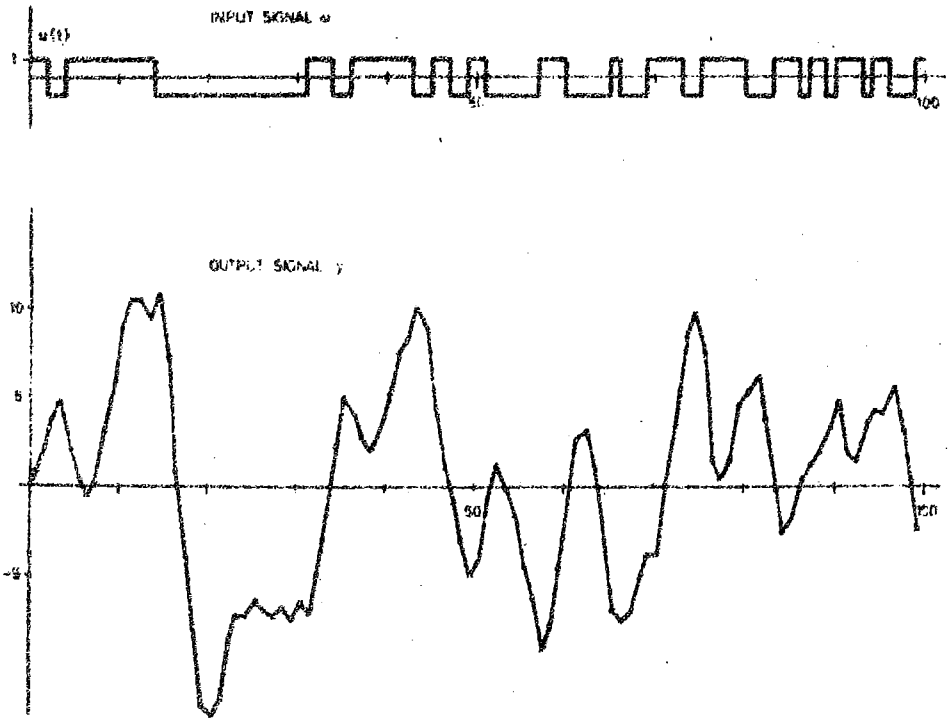

Fig. 2 - Input output signals for system of example 2.