Choice of Sampling Interval for Parametric Identification

Gustavsson, Ivar

1971

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Gustavsson, I. (1971). *Choice of Sampling Interval for Parametric Identification*. (Research Reports TFRT-3029). Department of Automatic Control, Lund Institute of Technology (LTH).

Total number of authors:
1

CHOICE OF SAMPLING INTERVAL FOR PARAMETRIC IDENTIFICATION

IVAR GUSTAVSSON

REPORT 7103 APRIL 1971
LUND INSTITUTE OF TECHNOLOGY
DIVISION OF AUTOMATIC CONTROL

CHOICE OF SAMPLING INTERVAL FOR PARAMETRIC IDENTIFICATION+

I.Gustavsson

ABSTRACT

The problem of choosing the sampling interval for identification experiments, analysed by parametric methods, has been studied. Optimal sampling rates have been determined under specific assumptions concerning the structure of the system, the input signal, the disturbances, the criterion of optimality and the method of sampling. For some cases optima exist, for others the best sampling rate is infinitely fast or the opposite. In those cases an economical sampling rate can be recommended. The results can be summarized as follows. A sampling rate corresponding to a Nyquist frequency twice to five times the highest breakpoint frequency or resonance frequency is reasonable. The higher value should be used for systems with poles close together or for systems with a high damping factor.

+This work was supported by the Swedish Board for Technical Development (Contract 70-337/U270)

TABLE OF CONTENTS | Page

1. INTRODUCTION.

One problem often arising in modern control theory is how to choose the sampling rate. Because of the increasing use of digital computers the problem will be even more important in the future. When digital equipment is used for process control the process variables have to be sampled, and discrete-time models must be used. The intuition says that the higher sampling rate, the better a discrete-time model represents a continuous-time system. However, this is not true in general.

Identification is one field of control theory where this problem exists. A simple problem formulation may be to determine the sampling rate for which one specific parameter can be estimated with minimum variance. Partial answers have been given, e.g. in [3], [7], [8], and [11]. In this paper the optimal sampling rate is determined for some cases when specific restrictions are given for the structure of the system, the input signal, the disturbances, the criterion of optimality and the method of sampling.

A canonical form to which any linear discrete-time system with one input and one output with a gaussian disturbance, which has a rational spectrum, can be reduced, is

$$A^{*}(q^{-1})y(t) = B^{*}(q^{-1})u(t) + \lambda C^{*}(q^{-1})e(t) \qquad (1.1)$$

where $y(t)$ is the output of the system,
$u(t)$ is the input to the system, and
$e(t)$ is a sequence of independent normal random variables of zero mean and variance one.

A^*, B^*, and C^* are polynomials of degree n of the backward shift operator q^{-1}

$$\begin{cases} A^*(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n} \\ B^*(q^{-1}) = b_0 + b_1 q^{-1} + \ldots + b_n q^{-n} \\ C^*(q^{-1}) = 1 + c_1 q^{-1} + \ldots + c_n q^{-n} \end{cases} \tag{1.2}$$

This general form is difficult to handle and in the following $C^*(q^{-1})$ is restricted to be 1 or to be equal to $A^*(q^{-1})$. These cases are referred to as the least squares structure case and the output white noise case respectively. This means that the results are restricted to cases when the disturbances can be modelled in one of these two ways in the discrete model.

The problem of choice of sampling rate is closely related to the problem of choosing input signal. A change of input signal may very likely cause a change of the optimal sampling rate. This problem is not treated. For simplicity the input signal is here chosen to be discrete-time white noise. It is probable that the results approximately hold also for white noise approximants like PRBS.

Other factors influencing on the optimal sampling rate are restrictions that may lie on the input or output signals, e.g. $E\,u^2(t) \leq$ const., $|u(t)| \leq$ const., $E\,y^2(t) \leq$ const. Again for simplicity the restriction $E\,u^2(t) = 1$ is chosen.

Another problem is what optimality criterion should really be used. When only one parameter is considered unknown, the criterion can be the minimization of the variance of the parameter estimate. When several parameters are unknown the problem is more difficult.

The solution, as always in identification problems, depends on the primary purpose of the identification. If the modelling is done for control purposes the influence on the regulated system should be studied. Another criterion would be to minimize the sum of variances, that is the trace of the covariance matrix, cf. [1]. The sum of relative errors of the estimates is another measure. These criteria are studied for the examples given.

Two cases are considered, N = const. and Nh = const., where N is the number of samples and h is the sampling interval. Thus Nh is the total experiment time, T_{exp}. In order to calculate the elements of the information matrix asymptotic theory most often has to be used. Thus N must be assumed to be a large number for the calculations to hold.

The method of sampling also influences on the optimal sampling rate. The sampling can be performed by integrating the signals during the sampling interval and averaging or by instantaneous reading at the sampling events. Integrating sampling will for instance change the variance of the noise of the sampled model when the sampling rate is changed. Therefore only instantaneous sampling is considered. In chapter 3 it is shown that even the signal to noise ratio, and not only the character of the disturbances, will influence on the optimal sampling rate.

The calculations performed are general and can principally be used for any system and any input, but they are often tedious and difficult to carry out. The most important drawback is, however, that the process has to be known in advance in order to calculate the optimal sampling rate. This, of course, diminishes the practical value of calculations like these. However, the minima, when existing, are not very sharp, and this fact can be used to obtain a sufficiently good estimate of the best sampling rate from estimated characteristics of the process like time constants, gain etc. For instance it is possible to use rules of thumb for certain types of systems. One rule is that a too fast sampling rate is not as dangerous as a too slow sampling rate. For systems with real poles only, a useful sampling interval seems to be 0.5 - - 1.5 times the shortest time constant of the system. For a second order system with complex poles a reasonable choice of the sampling interval is

$$h = \frac{\pi}{k\omega_r}$$

where ω_r is the resonance frequency and k = 2-5.

When calculations like these are too laborious to carry out or other model structures are necessary, simulations can be used to estimate the best sampling rate. However, these simulations may be very time-consuming if identification has to be performed by the maximum likelihood method, which is recommended because simulations have shown that the estimate of the covariance matrix obtained by this method is rather good. No other identification method known to the author gives an estimate of this matrix that has a comparable accuracy, when the C^* polynomial is of first or higher order.

Notice that sampling rates can be chosen from Bode diagrams rather well, just by choosing a sampling rate corresponding to a Nyquist frequency twice to five times the highest breakpoint frequency or resonance frequency. The higher value should be used for systems with poles close together or for systems with a high damping factor.

It is also interesting to compare the results with the sampling interval recommended for correlation and spectral analysis, e.g. in [6]. Their choice is $0.25 - 0.4 \cdot f_c^{-1}$, where f_c^{-1} is the smallest "period" in the record, and this choice must be small enough so that aliasing will not be a problem. These rules give sampling intervals of the same magnitude as the results in this paper.

In chapter 2 the output noise case is discussed. The least squares structure case for a first order system is considered in chapter 3.

2. OUTPUT WHITE NOISE CASE.

In this chapter the optimal sampling rate is determined for different systems when they are disturbed by output white noise only, i.e. the model is

$$y(t) = \frac{B^{*}(q^{-1})}{A^{*}(q^{-1})} u(t) + \lambda e(t) \qquad (2.1)$$

A first order system is studied first. The results from this study are then extended to the case of several distinct real poles. At last calculations are done for a second order system with complex poles.

2.1. A First Order System.

For this case the system equation is

$$y(t) = \frac{bq^{-1}}{1 + aq^{-1}} u(t) + \lambda e(t) \qquad (2.1.1)$$

From measurements of the process it is possible to get unbiased estimates of the parameters a and b by different methods, e.g. the maximum likelihood method [5] and the tally principle [12]. It is also possible to derive the Cramér-Rao lower bounds for the variances of the estimates [2]. Assuming λ known we find that the inverse of the information matrix, i.e. the covariance matrix, is bounded by

$$J^{-1} = \frac{\lambda^2}{N} \begin{bmatrix} E\left(\frac{\partial \varepsilon}{\partial a}\right)^2 & E\left(\frac{\partial \varepsilon}{\partial a} \cdot \frac{\partial \varepsilon}{\partial b}\right) \\ E\left(\frac{\partial \varepsilon}{\partial a} \cdot \frac{\partial \varepsilon}{\partial b}\right) & E\left(\frac{\partial \varepsilon}{\partial b}\right)^2 \end{bmatrix}^{-1} \quad (2.1.2)$$

where

$$\varepsilon(t) = y(t) - \frac{bq^{-1}}{1 + aq^{-1}} u(t) \quad (2.1.3)$$

N is the number of samples and

$$\begin{cases} \dfrac{\partial \varepsilon(t)}{\partial a} = \dfrac{bq^{-2}}{(1 + aq^{-1})^2} u(t) \\ \dfrac{\partial \varepsilon(t)}{\partial b} = - \dfrac{q^{-1}}{1 + aq^{-1}} u(t) \end{cases} \quad (2.1.4)$$

It is now possible to calculate J^{-1} from (2.1.2) and (2.1.4) for different input signals. This is particularly easy if u(t) is white noise with variance one. In that case we get

$$\begin{cases} E\left(\dfrac{\partial \varepsilon}{\partial a}\right)^2 = \dfrac{1}{2\pi i} \oint \dfrac{bz^{-2}}{(1+az^{-1})^2} \cdot \dfrac{bz^2}{(1+az)^2} \dfrac{dz}{z} = \dfrac{b^2(1+a^2)}{(1-a^2)^3} \\ E\left(\dfrac{\partial \varepsilon}{\partial a} \cdot \dfrac{\partial \varepsilon}{\partial b}\right) = - \dfrac{1}{2\pi i} \oint \dfrac{z^{-1}}{(1+az^{-1})} \cdot \dfrac{bz^2}{(1+az)^2} \dfrac{dz}{z} = \dfrac{ab}{(1-a^2)^2} \\ E\left(\dfrac{\partial \varepsilon}{\partial b}\right)^2 = \dfrac{1}{2\pi i} \oint \dfrac{z^{-1}}{1+az^{-1}} \cdot \dfrac{z}{1+az} \dfrac{dz}{z} = \dfrac{1}{1-a^2} \end{cases}$$

(2.1.5)

Thus we have, cf. [7]:

$$J^{-1} = \frac{\lambda^2}{N}\begin{bmatrix} \frac{b^2(1+a^2)}{(1-a^2)^3} & \frac{ab}{(1-a^2)^2} \\ \frac{ab}{(1-a^2)^2} & \frac{1}{1-a^2} \end{bmatrix}^{-1} =$$

$$= \frac{\lambda^2}{N}\begin{bmatrix} \frac{(1-a^2)^3}{b^2} & -\frac{a(1-a^2)^2}{b} \\ -\frac{a(1-a^2)^2}{b} & (1-a^2)(1+a^2) \end{bmatrix} \tag{2.1.6}$$

and

$$\mathrm{Var}\left\{(\hat{\theta}-\theta_0)(\hat{\theta}-\theta_0)^T\right\} \geq J^{-1} \tag{2.1.7}$$

where $\hat{\theta}$ is the estimate and θ_0 is the correct value of the parameters. θ is a vector. Notice that these lower bounds for the variances can be obtained asymptotically by using the maximum likelihood method for the identification [5] assuming that e(t) is normally distributed.

The discrete-time system (2.1.1) can be considered as a discrete model of the continuous process

$$\begin{cases} \frac{dx}{dt} = -\alpha x + \beta u \\ y = x \end{cases} \tag{2.1.8}$$

sampled at uniform rate, sampling interval h, and with white measurement noise on the output signal.

Then we have:

$$\begin{cases} a = -e^{-\alpha h} \\ b = \frac{\beta}{\alpha}\left(1-e^{-\alpha h}\right) \end{cases} \qquad (2.1.9)$$

or

$$\begin{cases} \alpha = -\frac{1}{h}\ln(-a) \\ \beta = -\frac{b\ln(-a)}{h(1+a)} \end{cases} \qquad (2.1.10)$$

if u(t) is considered constant over the sampling interval.

The problem is now whether there is an optimal sampling rate or not. First choose the minimization of the variances of the estimated parameters in the continuous model as criterion of optimality. Given (2.1.6) and (2.1.10) it is possible to calculate the lower bounds of the variances of the estimated parameters $\hat{\alpha}$ and $\hat{\beta}$ [9]. We get

$$\begin{cases} \mathrm{Var}(\hat{\alpha}) = \left(\frac{\partial\alpha}{\partial a}, \frac{\partial\alpha}{\partial b}\right) J^{-1} \begin{pmatrix} \frac{\partial\alpha}{\partial a} \\ \frac{\partial\alpha}{\partial b} \end{pmatrix} \\ \mathrm{Var}(\hat{\beta}) = \left(\frac{\partial\beta}{\partial a}, \frac{\partial\beta}{\partial b}\right) J^{-1} \begin{pmatrix} \frac{\partial\beta}{\partial a} \\ \frac{\partial\beta}{\partial b} \end{pmatrix} \end{cases} \qquad (2.1.11)$$

where

$$\begin{cases} \dfrac{\partial\alpha}{\partial a} = -\dfrac{1}{ah} \\[2ex] \dfrac{\partial\alpha}{\partial b} = 0 \\[2ex] \dfrac{\partial\beta}{\partial a} = -\dfrac{b}{h}\left\{\dfrac{1 + a - a\cdot\ln(-a)}{a(1+a)^2}\right\} \\[2ex] \dfrac{\partial\beta}{\partial b} = -\dfrac{\ln(-a)}{h(1+a)} \end{cases} \qquad (2.1.12)$$

And now

$$\begin{cases} Var(\hat{\alpha}) = \dfrac{1}{a^2h^2} Var(\hat{a}) \\[3ex] Var(\hat{\beta}) = \dfrac{b^2}{h^2}\left\{\dfrac{1 + a - a\cdot\ln(-a)}{a(1+a)^2}\right\}^2 Var(\hat{a}) + \end{cases}$$

$$+ 2\,\frac{b\cdot\ln(-a)}{h^2(1+a)}\left\{\frac{1 + a - a\cdot\ln(-a)}{a(1+a)^2}\right\}Cov(\hat{a},\hat{b}) +$$

$$+ \frac{\ln^2(-a)}{h^2(1+a)^2} Var(\hat{b}) \qquad (2.1.13)$$

Inserting a and b from (2.1.9) the variances will be functions of λ, N, α, β, and h. For instance

$$\mathrm{Var}(\hat{\alpha}) = \frac{1}{e^{-2\alpha h}h^2} \cdot \frac{\lambda^2}{N} \cdot \frac{(1-e^{-2\alpha h})^3}{\frac{\beta^2}{\alpha^2}(1-e^{-\alpha h})^2} =$$

$$= \frac{\lambda^2}{N}\frac{\alpha^4}{\beta^2} \cdot \frac{1}{(\alpha h)^2} \cdot \frac{(1+e^{-\alpha h})^3(1-e^{-\alpha h})}{e^{-2\alpha h}} \qquad (2.1.14)$$

Assuming the number of observations, N, fixed we can compute the variances of $\hat{\alpha}$ and $\hat{\beta}$ according to the formulas of (2.1.13) for different sampling intervals h. Considering the variances of $\hat{\alpha}$ and $\hat{\beta}$ as functions of αh we find that they have minima for certain values of αh. The variances are plotted in Fig. 1. Notice that for these figures and the following the scaling factors

$$\frac{\lambda^2}{N} \quad \text{or} \quad \frac{\lambda^2}{Nh} = \frac{\lambda^2}{T_{exp}}$$

are often not considered. The minima do not occur exactly for the same value of αh, but fairly close, for $\alpha h = 1.15$ and $\alpha h = 1.08$ respectively. The functions of αh that appear can be minimized analytically and most often reduce to the problem of searching zeroes of a nonlinear equation. It is also easy to study what happens for small αh and when αh tends to infinity, and to investigate the characteristics of the function (if there are several stationary points, convexity etc.).

The criterion chosen was arbitrary. The parameters α and β are perhaps not the primary parameters of interest. The gain, $G = \beta/\alpha$, and the time constant, $T = 1/\alpha$, may be more interesting. However, it is easy to use the same method as outlined above for the calculation of the variances of those parameters.

We have

$$\begin{cases} G = \dfrac{\beta}{\alpha} = \dfrac{b}{1+a} \\[2ex] T = \dfrac{1}{\alpha} = -\dfrac{h}{\ln(-a)} \end{cases} \tag{2.1.15}$$

and

$$\begin{cases} \dfrac{\partial G}{\partial a} = -\dfrac{b}{(1+a)^2} \\[2ex] \dfrac{\partial G}{\partial b} = \dfrac{1}{1+a} \\[2ex] \dfrac{\partial T}{\partial a} = \dfrac{h}{a\ln^2(-a)} \\[2ex] \dfrac{\partial T}{\partial b} = 0 \end{cases} \tag{2.1.16}$$

The variances of these parameters are given in Fig. 1 as functions of αh. The variance of T has the same minimum as that of α. However, there is a scaling factor between them. The variance of G is decreasing for increasing αh. The reason for this is obvious, because N is fixed, and the best estimate of G will then occur when the sampling points are spread out as much as possible so that the output of the system has time to reach its stationary value. This means that a step response is the best way of estimating the gain. Notice the assumption that the input is constant during the sampling interval.

Now instead of N fixed, let $Nh = T_{exp}$, the total experiment length, be fixed. For this case the variances plotted to the right in Fig. 1 are obtained. We get the rule that the smaller sampling interval the smaller variances. However, also in this case the results obtained may be valuable because they indicate how fast we should sample out of economical reasons. A too fast sampling will give a tremendous computing time for the identification, but the increase of accuracy is small. Such a high sampling rate would not be worthwhile.

2.2. Higher Order Systems with Distinct Real Poles.

So far only systems of first order with one real pole have been considered. It is, however, easy to extend these results to systems with several distinct real poles. To do so we write the system as

$$y(t) = \sum_{i=1}^{n} \frac{b_i q^{-1}}{1 + a_i q^{-1}} u(t) + \lambda e(t) \qquad (2.2.1)$$

which is always possible in the case when the pulse transfer function of the system is given by

$$y(t) = \frac{b_1' q^{-1} + \ldots + b_n' q^{-n}}{1 + a_1' q^{-1} + \ldots + a_n' q^{-n}} u(t) + \lambda e(t) \qquad (2.2.2)$$

and all roots of the polynomial

$$z^n + a_1' z^{n-1} + \ldots + a_n' = 0 \qquad (2.2.3)$$

are real and distinct. Notice that the model (2.2.2) is often used for identification purposes. Calculations given in Appendix A show that the lower bound of the covariance matrix will be

$$J^{-1}=\frac{\lambda^2}{N}\begin{bmatrix} \frac{b_1^2(1+a_1^2)}{(1-a_1^2)^3} & \cdots & \frac{b_1 b_n(1+a_1 a_n)}{(1-a_1 a_n)^3} & \frac{a_1 b_1}{(1-a_1^2)^2} & \cdots & \frac{a_n b_1}{(1-a_1 a_n)^2} \\ & \ddots & \vdots & \vdots & & \vdots \\ & & \frac{b_n^2(1+a_n^2)}{(1-a_n^2)^3} & \frac{a_1 b_n}{(1-a_1 a_n)^2} & \cdots & \frac{a_n b_n}{(1-a_n^2)^2} \\ & & & \frac{1}{1-a_1^2} & \cdots & \frac{1}{1-a_1 a_n} \\ & & & & \ddots & \vdots \\ & & & & & \frac{1}{1-a_n^2} \end{bmatrix}^{-1} \qquad (2.2.4)$$

The matrix is symmetric.

Notice that in this case we get the variances of α_i, β_i, G_i and T_i, $i = 1, \ldots, n$, (notations according to Appendix A) easily but that it is a bit more tedious to calculate the variances of the parameters of the continuous model

$$y(t) = \frac{\beta_1' s^{n-1} + \ldots + \beta_n'}{s^n + \alpha_1' s^{n-1} + \ldots + \alpha_n'}\, u(t) \qquad (2.2.5)$$

but it is possible, since α_i' and β_i' can be expressed as functions of α_i and β_i. It is also easy to calculate the variance of the total gain G.

In Fig. 2 the variances of the parameters of the second order system

$$\frac{1.2(s+5)}{(s+1)(s+25)} = \frac{0.2}{s+1} + \frac{1}{s+25} \qquad (2.2.6)$$

are plotted as functions of the sampling interval h. The number of sampling events, N, is assumed fixed. For the parameters α_1, α_2, β_1, β_2, T_1, T_2, and G the curves for constant Nh will have the same characteristics as the curves to the right in Fig. 1. Notice that the shortest time constant of the system is 0.04 sec., and that the minimum of the sum of relative errors of T_1, T_2, and G occurs for approximately 0.06 sec. The relative error for a parameter θ is defined as $\sqrt{Var(\hat{\theta})}/|\theta|$. From this we conclude that it is the shortest time constant of the system, that has the greatest effect on the choice of sampling rate, if all parameters are wanted with good accuracy.

As another example a difference approximation model of a one dimensional heat diffusion process is used. The exact transfer function from temperature change in one end point of a heat rod to temperature change in the middle point of the rod is, e.g. [10],

$$G(s) = \frac{\sinh \frac{\sqrt{s\tau}}{2}}{\sinh \sqrt{s\tau}} \qquad (2.2.7)$$

where τ is a factor depending on the length of the rod and on the material it is made of.

A difference approximation of this with 8 steps will yield the transfer function

$$G(s) = \frac{64^4\tau^4}{(s\tau+9.74)(s\tau+79.01)(s\tau+176.90)(s\tau+246.26)} \tag{2.2.8}$$

The numerical value of τ, that is used, 1857.8, comes from a laboratory process, constructed at our institute by B. Leden [10]. (2.2.8) can then be fractioned as

$$G(s) = \frac{3.314\cdot10^{-3}}{s+5.243\cdot10^{-3}} - \frac{8.004\cdot10^{-3}}{s+4.253\cdot10^{-2}} + \frac{7.998\cdot10^{-3}}{s+9.522\cdot10^{-2}} -$$

$$- \frac{3.308\cdot10^{-3}}{s+1.326\cdot10^{-1}} \tag{2.2.9}$$

The time constants are 190.7, 23.5, 10.5, and 7.5.seconds. Now assuming that white measurement noise is the only disturbance on the process, the minimum variances of different parameters can be computed. In Fig. 3 the sum of relative errors are given for the cases N = const. and Nh = const. and for different parameter combinations. From these figures we conclude that 5 seconds is a useful sampling interval irrespective of the parameter combination chosen. Again the sampling interval chosen is of the same size as the shortest time constant. As an example the optimum sampling interval is 25 seconds for the longest time constant. But with such a sampling rate the variance of the shortest time constant is roughly 2500 times the minimum value at 5 seconds. This means in practice that it cannot be found at the identification. Generally speaking, the fact that identification by parametric methods often gives low order models, is at least partly explained by this phenomenon. Modes corresponding to time constants which are several times shorter than the sampling interval cannot be found at the identification. This is analogous to the aliasing effect when spectral analysis is used.

2.3. Second Order Systems with Complex Poles.

In order to handle general systems of the form (2.1), we also have to compute optimal sampling rates for second order systems with a transfer function of the form

$$\frac{cs + d}{s^2 + 2\zeta\omega s + \omega^2} \tag{2.3.1}$$

or

$$\frac{cs + d}{s^2 + 2as + a^2 + b^2} = \frac{cs + d}{(s+a)^2 + b^2} \tag{2.3.2}$$

The second parameter structure is used in first hand because the transformation between the continuous system and the sampled form of the system is simpler. Afterwards it is easy to transform the covariance matrix for the parameter combination a, b, c, and d to the covariance matrix for the combination ζ, ω, c, and d.

In appendix B it is shown that the sampled form of (2.3.2) is

$$\frac{b_1 q^{-1} + b_2 q^{-2}}{1 + a_1 q^{-1} + a_2 q^{-2}} \tag{2.3.3}$$

with

$$\begin{cases} a_1 = -\ 2e^{-ah}\cos(bh) \\ a_2 = e^{-2ah} \\ b_1 = b^{-1}(a^2+b^2)^{-1}\{bd + e^{-ah}[c(a^2+b^2) - ad]\sin(bh) - \\ \qquad - e^{-ah}bd\ \cos(bh)\} \\ b_2 = b^{-1}(a^2+b^2)^{-1}\{bde^{-2ah} - e^{-ah}[c(a^2+b^2) - ad] \cdot \\ \qquad \cdot \sin(bh) - e^{-ah}bd\ \cos(bh)\} \end{cases} \tag{2.3.4}$$

For

$$y(t) = \frac{b_1q^{-1} + b_2q^{-2}}{1 + a_1q^{-1} + a_2q^{-2}}\ u(t) + \lambda e(t) \tag{2.3.5}$$

it is now possible to use the same technique as before to compute the Cramér-Rao lower bound for the covariance matrix.

Let

$$\varepsilon(t) = y(t) - \frac{b_1q^{-1} + b_2q^{-2}}{1 + a_1q^{-1} + a_2q^{-2}}\ u(t) \tag{2.3.6}$$

and we get

$$\begin{cases} \dfrac{\partial\varepsilon(t)}{\partial a_i} = \dfrac{q^{-i}(b_1q^{-1} + b_2q^{-2})}{(1 + a_1q^{-1} + a_2q^{-2})^2}\ u(t) \qquad i = 1,\ 2 \\ \dfrac{\partial\varepsilon(t)}{\partial b_i} = -\ \dfrac{q^{-i}}{1 + a_1q^{-1} + a_2q^{-2}}\ u(t) \qquad i = 1,\ 2 \end{cases} \tag{2.3.7}$$

The covariance matrix is

$$J^{-1} = \begin{bmatrix} E\left(\frac{\partial \varepsilon}{\partial a_1}\right)^2 & E\left(\frac{\partial \varepsilon}{\partial a_1}\frac{\partial \varepsilon}{\partial a_2}\right) & E\left(\frac{\partial \varepsilon}{\partial a_1}\frac{\partial \varepsilon}{\partial b_1}\right) & E\left(\frac{\partial \varepsilon}{\partial a_1}\frac{\partial \varepsilon}{\partial b_2}\right) \\ & \ddots & & \vdots \\ & & \ddots & \vdots \\ & & & E\left(\frac{\partial \varepsilon}{\partial b_2}\right)^2 \end{bmatrix}^{-1} \quad (2.3.8)$$

For u(t) white noise it is possible to compute the elements of this matrix, by using techniques described in [4].

The problem is now to express the variances of a, b, c, and d as functions of the sampling interval. The transformation from a, b, c, and d to a_1, a_2, b_1, and b_2 is given by (2.3.4). The inverse of this transformation is difficult to handle and we use the following:

$$\begin{bmatrix} \frac{\partial a_1}{\partial a} & \frac{\partial a_1}{\partial b} & \frac{\partial a_1}{\partial c} & \frac{\partial a_1}{\partial d} \\ \frac{\partial a_2}{\partial a} & \frac{\partial a_2}{\partial b} & \frac{\partial a_2}{\partial c} & \frac{\partial a_2}{\partial d} \\ \frac{\partial b_1}{\partial a} & \frac{\partial b_1}{\partial b} & \frac{\partial b_1}{\partial c} & \frac{\partial b_1}{\partial d} \\ \frac{\partial b_2}{\partial a} & \frac{\partial b_2}{\partial b} & \frac{\partial b_2}{\partial c} & \frac{\partial b_2}{\partial d} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial a}{\partial a_1} & \frac{\partial a}{\partial a_2} & \frac{\partial a}{\partial b_1} & \frac{\partial a}{\partial b_2} \\ \frac{\partial b}{\partial a_1} & \frac{\partial b}{\partial a_2} & \frac{\partial b}{\partial b_1} & \frac{\partial b}{\partial b_2} \\ \frac{\partial c}{\partial a_1} & \frac{\partial c}{\partial a_2} & \frac{\partial c}{\partial b_1} & \frac{\partial c}{\partial b_2} \\ \frac{\partial d}{\partial a_1} & \frac{\partial d}{\partial a_2} & \frac{\partial d}{\partial b_1} & \frac{\partial d}{\partial b_2} \end{bmatrix} = I \quad (2.3.9)$$

or $WZ = I$, i.e. $Z = W^{-1}$, and W^{-1} exists if the functions are sufficiently smooth. The derivatives of a_1, a_2, b_1, and b_2 with respect to a, b, c, and d respectively, that is the elements of the matrix W, are given in Appendix B.

We get

$$\mathrm{Cov}(\hat{a},\hat{b},\hat{c},\hat{d}) = ZJ^{-1}Z^T = W^{-1}J^{-1}(W^{-1})^T \tag{2.3.10}$$

The variances of a, b, c, and d are now expressed as functions of h, but it is difficult to study these functions analytically. However, it is relatively easy to compute them digitally for different values of h. Examples of the results are shown in Fig. 4-7 for three different systems and for fixed N.

As test systems the following ones have been chosen:

I $\quad \dfrac{1}{s^2 + 0.4s + 4}$ Fig. 4

II $\quad \dfrac{1}{s^2 + 2s + 4}$ Fig. 5

III $\quad \dfrac{1}{s^2 + 3.6s + 4}$ Fig. 6

IV $\quad \dfrac{1}{s^2 + 0.6s + 1}$

V $\quad \dfrac{1}{s^2 + 1.2s + 4}$

VI $\quad \dfrac{1}{s^2 + 2.4s + 16}$

which can be characterized by the damping factor, ζ, and the resonance frequency, ω, according to Table 1.

System	ζ	ω	h_{max}
I	0.1	2	1.58
II	0.5	2	1.81
III	0.9	2	3.60
IV	0.3	1	3.29
V	0.3	2	1.64
VI	0.3	4	0.82

Table 1 - Characteristics of test systems I-VI.

h_{max} is the maximum sampling interval, for which the transformation of the poles of the continuous model (the left half plane) to the discrete model (the inner of the unit circle) can be done without overlapping.

In Fig. 4-6 the variances of $\hat{a}$, $\hat{b}$, $\hat{c}$, and $\hat{d}$ as functions of h can be studied for fixed ω, N = const., and varying ζ. The result is that the bigger ζ, the smaller h required. Notice that if the experiment time instead of the samples is fixed, the variances of Fig. 4-6 should be multiplied by h. This means that for small h the variances are approximately constant (cf. the top diagrams of Fig. 1). Notice that for constant N, even the variance of $\hat{G}$ has a minimum. It is also remarkable that for some of the parameters (a and b) several local minima sometimes exist. The variances of $\hat{b}$, $\hat{c}$, and $\hat{d}$ have absolute minima for approximately the same h, but the absolute minimum for the variance of $\hat{a}$ may occur for a somewhat different sampling interval (Table 2).

System	ω	$Var(\hat{a})$	$Var(\hat{b})$	$Var(\hat{c})$	$Var(\hat{d})$
IV	1	2.46	1.27	1.17	1.28
V	2	1.23	0.63	0.58	0.64
VI	4	0.62	0.32	0.29	0.32

Table 2 - The sampling interval for which the respective variances have minima.

The consistency of the computations has been tested in the following way. For $\zeta = 1$ the second order system has two equal real poles. The method of chapter 2.2 has been used to compute the optimal sampling interval for a second order system with two poles close to the double pole of

$$\frac{1}{s^2 + 4s + 4}$$

The result was that an optimal sampling interval was found near 1/3 sec., which can be compared to Fig. 6 and Table 3, where we can see that the optimal sampling intervals are of the same size for the case $\zeta = 0.9$.

If instead of model (2.3.2) the model (2.3.1) was used and the variances of $\hat{\zeta}$ and $\hat{\omega}$ were computed only one minimum was found for the examples tested. Results are shown in Fig. 7. The optimal sampling intervals are given in Table 3.

System	ζ	ω	$Var(\hat{\zeta})$	$Var(\hat{\omega})$
I	0.1	2	1.32	0.90
II	0.5	2	1.08	0.46
III	0.9	2	0.40	0.29
IV	0.3	1	2.21	1.23
V	0.3	2	1.10	0.62
VI	0.3	4	0.55	0.31

Table 3 - Optimal sampling intervals for respective variances.

From Tables 2 and 3 we conclude that the optimal sampling interval for fixed ζ is proportional to $1/\omega$ where ω is the resonance frequency.

The calculations have been performed on a UNIVAC 1108 in double precision. Even though numerical difficulties occured in the neighbourhood of h_{max}. These are explained by the use of the inverses J^{-1} and W^{-1} in (2.3.10). Both J and W have tendencies of being ill-conditioned when h approaches h_{max}.

3. LEAST SQUARES STRUCTURE CASE.

With the least squares structure means the representation

$$y(t) = \frac{B^*(q^{-1})}{A^*(q^{-1})} u(t) + \lambda \frac{1}{A^*(q^{-1})} e(t) \tag{3.1}$$

or

$$A^*(q^{-1})y(t) = B^*(q^{-1})u(t) + \lambda e(t) \tag{3.2}$$

Because it is a bit more cumbersome to do computations with this structure, only a first order case is considered, that is

$$y(t) = \frac{bq^{-1}}{1 + aq^{-1}} u(t) + \lambda \frac{1}{1 + aq^{-1}} e(t) \tag{3.3}$$

Let u(t) be discrete white noise with variance one. The Cramér-Rao lower bound for the covariance matrix will be (Appendix C)

$$J^{-1} = \frac{\lambda^2}{N} \begin{bmatrix} \frac{b^2 + \lambda^2}{1 - a^2} & 0 \\ 0 & 1 \end{bmatrix}^{-1} \tag{3.4}$$

which gives

$$
\begin{cases}
\mathrm{Var}(\hat{a}) = \dfrac{\lambda^2}{N} \dfrac{1 - a^2}{b^2 + \lambda^2} \\[2ex]
\mathrm{Var}(\hat{b}) = \dfrac{\lambda^2}{N} \\[2ex]
\mathrm{Cov}(\hat{a},\hat{b}) = 0
\end{cases}
\tag{3.5}
$$

Now we can proceed as in 2.1 and we get

$$
\mathrm{Var}(\hat{\alpha}) = \frac{\lambda^2}{N} \frac{1}{a^2h^2} \frac{1-a^2}{b^2+\lambda^2} = \frac{\lambda^2}{N} \cdot \frac{1}{e^{-2\alpha h}h^2} \frac{1 - e^{-2\alpha h}}{\frac{\beta^2}{\alpha^2}(1-e^{-\alpha h})^2 + \lambda^2}
\tag{3.6}
$$

etc. These variances are not only depending on λ^2 as a multiplicative factor, but λ^2 is also involved in the function of h. In Fig. 8 - 9 variances of $\hat{T}$ and $\hat{G}$ are plotted for three different values of λ and for N = const. and Nh = const. α and β have both been chosen to one. Notice that the characteristics of the curves may change remarkably with λ. E.g. the variance of $\hat{T}$ with Nh = constant is ascending for $\lambda = 1.0$, has a minimum for $h \approx 0.92$ for $\lambda = 0.1$, but a maximum for $h \approx 0.21$ and a minimum for $h \approx 0.51$ for $\lambda = 0.5$. It is also remarkable that for the case of least squares structure it often happens that even if Nh = const., the variance is increasing when h approaches zero. This is not so for the output white noise case.

In this case it is also interesting to study the variation of the output of the system with a minimum variance controller, synthesized from the estimated model.

Assume the system is given by (3.1). This system is modelled by the model

$$y(t) = \frac{\hat{b}\,q^{-1}}{1 + \hat{a}q^{-1}}\,u(t) + \lambda\,\frac{1}{1 + \hat{a}q^{-1}}\,e(t) \qquad (3.7)$$

The minimum variance strategy is given by [4]

$$u(t) = \frac{\hat{a}}{\hat{b}}\,y(t) \qquad (3.8)$$

If this strategy is used for the system (3.3) the pulse transfer function for the closed loop system will be

$$y(t) = \lambda\,\frac{1}{1 + \left(a - b\frac{\hat{a}}{\hat{b}}\right)q^{-1}}\,e(t) \qquad (3.9)$$

that is the expected value of the output variance is given by

$$E\,y^2 = \lambda^2\,\frac{1}{1 - \left(a - b\frac{\hat{a}}{\hat{b}}\right)^2} \qquad (3.10)$$

Now if the modelling is performed many times, different estimates, $\hat{a}$ and $\hat{b}$, will be obtained. Assuming the mean values of $\hat{a}$ and $\hat{b}$ the correct ones, and the covariance matrix for the estimates known, it is possible to calculate the expected value of the output variance considering the estimates, $\hat{a}$ and $\hat{b}$, as stochastic variables.

By Taylor series expansion and averaging it is possible to write the expression (3.10) as

$$E\{E\ y^2\} = \lambda^2\left[1 + \{Var(\hat{a}) - 2\frac{a}{b}\ Cov(\hat{a},\hat{b}) + \frac{a^2}{b^2}\ Var(\hat{b})\}\right] \quad (3.11)$$

if terms of higher order than two are neglected.

This means that it is possible to study the variation of the expected output variance as a function of the sampling interval. The relative variation of $E\ y^2$ is

$$Var(\hat{a}) - 2\frac{a}{b}\ Cov(\hat{a},\hat{b}) + \frac{a^2}{b^2}\ Var(\hat{b}) \quad (3.12)$$

and is depending on λ as well. The function is plotted as a function of h for different values of λ and for N = const. and Nh = const. in Fig. 10. For Nh = const. an optimal sampling rate exists. For N = const. a rather slow sampling rate can be used. Notice that λ has been held constant for different sampling intervals. But even if we let λ vary so that the signal to noise ratio is constant for all sampling rates for this least squares structure case, the variances increase when h approaches zero. In this case, however, not all variances vary with the parameter λ (Table 4). For these cases only one stationary point existed, when any.

λ	N = const.			Nh = const.		
	$Var(\hat{T})$	$Var(\hat{G})$	Rel.var (Ey^2)	$Var(\hat{T})$	$Var(\hat{G})$	Rel.var (Ey^2)
0.1	1.37	↘	↘	0.93	1.90	1.26
0.5	1.37	↘	↘	0.93	1.85	1.33
1.0	1.37	↘	↘	0.93	1.74	1.51

Table 4 - Optimal sampling interval for the least squares structure case, signal/noise ratio = const. (↘ denotes descending with increasing h)

It must be emphasized that for this regulation problem only the relative variation of the expected output variance due to the uncertainty of the parameters has been studied. The effect of smaller sampling interval on the absolute value of the prediction error has not been involved.

4. CONCLUSIONS.

The calculations have shown that an optimal sampling rate often exists for the kind of problem studied here or at least that an economical sampling rate exists. The value of the results is reduced by the fact that the system has to be known in advance in order to be able to determine the optimal sampling interval. Furthermore, the optimality criterion is critical and has to be chosen appropriately and in accordance with the purpose of the identification. It has also been shown that the disturbances influence on the optimal sampling rate, not only the character of the disturbances but even the noise level.

It is hoped that the results give a feeling for the problem and a relatively good estimate of the sampling rate can be obtained from simple tests of the system, like step responses. Another tool can be the Bode diagram for the system. Rules of thumb like sampling rate equal to the shortest time constant may also be given. Notice, however, that all factors involved in the problem of choosing the sampling rate have not been discussed. For instance the effects of different input signals and of disturbances of other types than can be described by the least squares structure or as output white noise have not been considered. A change from limited input signal to limited output signal will also cause significant deviations from the results given.

5. REFERENCES.

[1] Aoki, M., and Staley, R.M.: On Input Signal Synthesis in Parameter Identification, IFAC, Warsaw, 1969.

[2] Åström, K.J.: On the Achievable Accuracy in Identification Problems, IFAC, Prague, 1967.

[3] Åström, K.J.: On the Choice of Sampling Rates in Parameter Identification of Time Series, Information Sciences, July, 1969.

[4] Åström, K.J.: Introduction to Stochastic Control Theory, Academic Press, 1970.

[5] Åström, K.J., and Bohlin, T.: Numerical Identification of Linear Systems from Normal Operating Records, in Hammond, P.H.: Theory of Self-Adaptive Control Systems, Plenum Press, 1966.

[6] Bendat, J.S., and Piersol, A.G.: Measurement and Analysis of Random Data, Wiley, New York, 1966.

[7] Clarke, D.W.: The Accuracy of Identification Techniques, Conference on Industrial Applications of Dynamic Modelling, 16-18 Sept., 1969, IEE Conference Publications, No. 57.

[8] Fantauzzi, G.: The Relation Between the Sampling Time and Stochastic Error for the Impulsive Response of Linear Time-Independent Systems, IEEE AC, August, 1968.

[9] Kendall, M.G., and Stuart, A.: The Advanced Theory of Statistics, Griffin, London, 1961.

[10] Leden, B.: The Design of a One-Dimensional Heat Diffusion Process, Report 7010, Dec., 1970, Lund Institute of Technology, Division of Automatic Control.

[11] Liff, A.I., and Wolf, J.K.: On the Optimum Sampling Rate for Discrete-Time Modelling of Continuous-Time Systems, IEEE AC, April, 1966.

[12] Peterka, V., and Smuk, K.: On-Line Estimation of Dynamic Model Parameters from Input-Output Data, IFAC, Warsaw, 1969.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to professor K.J. Åström, for initiating the problem and for valuable suggestions and discussions. He also thanks Mrs. G. Christensen, who typed the manuscript and Mrs. K. Bolinder, who drew the figures.

APPENDIX A

The system is given by

$$y(t) = \sum_{i=1}^{n} \frac{b_i q^{-1}}{1 + a_i q^{-1}} u(t) + \lambda e(t) \qquad (A.1)$$

With the same assumptions as for the first order case (2.1) we get

$$\varepsilon(t) = y(t) - \sum_{i=1}^{n} \frac{b_i q^{-1}}{1 + a_i q^{-1}} u(t) \qquad (A.2)$$

and

$$\begin{cases} \dfrac{\partial \varepsilon(t)}{\partial a_i} = \dfrac{b_i q^{-2}}{(1 + a_i q^{-1})^2} u(t) \\ \\ \dfrac{\partial \varepsilon(t)}{\partial b_i} = - \dfrac{q^{-1}}{1 + a_i q^{-1}} u(t) \end{cases} \qquad (A.3)$$

White noise as input gives

$$\begin{cases} E\left(\dfrac{\partial \varepsilon}{\partial a_i} \dfrac{\partial \varepsilon}{\partial a_j}\right) = \dfrac{1}{2\pi i} \oint \dfrac{b_i z^{-2}}{(1+a_i z^{-1})^2} \dfrac{b_j z^2}{(1+a_j z)^2} \dfrac{dz}{z} = \dfrac{b_i b_j (1+a_i a_j)}{(1-a_i a_j)^3} \\ \\ E\left(\dfrac{\partial \varepsilon}{\partial a_i} \dfrac{\partial \varepsilon}{\partial b_j}\right) = - \dfrac{1}{2\pi i} \oint \dfrac{z^{-1}}{1+a_j z^{-1}} \dfrac{b_i z^2}{(1+a_i z)^2} \dfrac{dz}{z} = \dfrac{a_j b_i}{(1-a_i a_j)^2} \\ \\ E\left(\dfrac{\partial \varepsilon}{\partial b_i} \dfrac{\partial \varepsilon}{\partial b_j}\right) = \dfrac{1}{2\pi i} \oint \dfrac{z^{-1}}{1+a_i z^{-1}} \dfrac{z}{1+a_j z} \dfrac{dz}{z} = \dfrac{1}{1-a_i a_j} \end{cases} \qquad (A.4)$$

The corresponding continuous system is

$$\sum_{i=1}^{n} \frac{\beta_i}{s + \alpha_i} \tag{A.5}$$

where

$$\begin{cases} \alpha_i = -\frac{1}{h} \ln(-a_i) \\ \beta_i = -\frac{b_i \ln(-a_i)}{h(1+a_i)} \end{cases} \tag{A.6}$$

and

$$\begin{cases} \frac{\partial \alpha_i}{\partial a_j} = \begin{cases} -\frac{1}{ha_i} & i = j \\ 0 & i \neq j \end{cases} \\ \frac{\partial \alpha_i}{\partial b_j} = 0 \\ \frac{\partial \beta_i}{\partial a_j} = \begin{cases} -\frac{b_i}{h}\left\{\frac{1 + a_i - a_i \ln(-a_i)}{a_i(1+a_i)^2}\right\} & i=j \\ 0 & i \neq j \end{cases} \\ \frac{\partial \beta_i}{\partial b_j} = \begin{cases} -\frac{1}{h} \frac{\ln(-a_i)}{1 + a_i} & i=j \\ 0 & i \neq j \end{cases} \end{cases} \tag{A.7}$$

From this it is easy to see that if the covariance matrix is known, the variances of $\hat{\alpha}_i$ and $\hat{\beta}_i$ only depend on a_i and b_i and no other a:s or b:s, and are easy to compute. This is one benefit of using the structure (A.1) and splitting the system into its modes. However, due to the inversion of the information matrix, there will still be a connection between the modes.

We also have

$$\begin{cases} G_i = \dfrac{\beta_i}{\alpha_i} = \dfrac{b_i}{1 + a_i} \\ T_i = \dfrac{1}{\alpha_i} = - \dfrac{h}{\ln(-a_i)} \end{cases} \tag{A.8}$$

and

$$G = \sum_{i=1}^{n} \frac{b_i}{1 + a_i} \tag{A.9}$$

From these formulas we get

$$\begin{cases} \dfrac{\partial G_i}{\partial a_j} = \begin{cases} - \dfrac{b_i}{(1 + a_i)^2} & i=j \\ 0 & i \neq j \end{cases} \\ \dfrac{\partial G_i}{\partial b_j} = \begin{cases} \dfrac{1}{1 + a_i} & i=j \\ 0 & i \neq j \end{cases} \end{cases} \tag{A.10}$$

$$\begin{cases} \dfrac{\partial T_i}{\partial a_j} = \begin{cases} \dfrac{h}{a_i \ln^2(-a_i)} & i=j \\ 0 & i \neq j \end{cases} \\ \dfrac{\partial T_i}{\partial b_j} = 0 \end{cases} \qquad \text{(A.11)}$$

$$\begin{cases} \dfrac{\partial G}{\partial a_i} = -\dfrac{b_i}{(1 + a_i)^2} \\ \dfrac{\partial G}{\partial b_i} = \dfrac{1}{1 + a_i} \end{cases} \qquad \text{(A.12)}$$

APPENDIX B

Let

$$G(s) = \frac{cs + d}{s^2 + 2as + a^2 + b^2} \tag{B.1}$$

be the transfer function for a continuous system of second order. The system is sampled with a sampling interval of h and it is assumed that the input signal is constant during each sampling interval. The pulse transfer function, $H(q^{-1})$, for system (B.1) will be determined. One way is to use the formula

$$H(q^{-1}) = \sum \frac{q^{-1}(e^{sh}-1)}{1 - q^{-1}e^{sh}} \frac{1}{s} G(s) \tag{B.2}$$

where the sum should be taken over the residues of G(s). After tedious but trivial calculations we get

$$H(q^{-1}) = \frac{b_1 q^{-1} + b_2 q^{-2}}{1 + a_1 q^{-1} + a_2 q^{-2}} \tag{B.3}$$

with

$$\begin{cases} a_1 = -2e^{-ah}\cos(bh) \\ a_2 = e^{-2ah} \\ b_1 = b^{-1}(a^2+b^2)^{-1}\{bd + [c(a^2+b^2)-ad]e^{-ah}\sin(bh) - \\ \qquad - bde^{-ah}\cos(bh)\} \\ b_2 = b^{-1}(a^2+b^2)^{-1}\{bde^{-2ah} - [c(a^2+b^2)-ad]e^{-ah}\sin(bh) - \\ \qquad - bde^{-ah}\cos(bh)\} \end{cases} \tag{B.4}$$

Thus

$$\frac{\partial a_1}{\partial a} = 2he^{-ah}\cos(bh)$$

$$\frac{\partial a_1}{\partial b} = 2he^{-ah}\sin(bh)$$

$$\frac{\partial a_1}{\partial c} = 0$$

$$\frac{\partial a_1}{\partial d} = 0$$

$$\frac{\partial a_2}{\partial a} = -\ 2he^{-2ah}$$

$$\frac{\partial a_2}{\partial b} = 0$$

$$\frac{\partial a_2}{\partial c} = 0$$

$$\frac{\partial a_2}{\partial d} = 0$$

$$\frac{\partial b_1}{\partial a} = b^{-1}(a^2+b^2)^{-1}\{-\ h[c(a^2+b^2)-ad]e^{-ah}\sin(bh) +$$

$$+\ hbde^{-ah}\cos(bh) + (2ac-d)e^{-ah}\sin(bh)\} -$$

$$-\ 2ab^{-1}(a^2+b^2)^{-2}\{bd + [c(a^2+b^2)-ad]e^{-ah}\sin(bh) -$$

$$-\ bde^{-ah}\cos(bh)\}$$

$$\frac{\partial b_1}{\partial b} = b^{-1}(a^2+b^2)^{-1}\{d + h[c(a^2+b^2)-ad]e^{-ah}\cos(bh) -$$

$$- de^{-ah}\cos(bh) + hbde^{-ah}\sin(bh) + 2bce^{-ah}\sin(bh)\} -$$

$$- (a^2+3b^2)b^{-2}(a^2+b^2)^{-2}\{bd + [c(a^2+b^2)-ad]e^{-ah}\sin(bh) -$$

$$- bde^{-ah}\cos(bh)\}$$

$$\frac{\partial b_1}{\partial c} = b^{-1}e^{-ah}\sin(bh)$$

$$\frac{\partial b_1}{\partial d} = b^{-1}(a^2+b^2)^{-1}[b - ae^{-ah}\sin(bh) - be^{-ah}\cos(bh)]$$

$$\frac{\partial b_2}{\partial a} = b^{-1}(a^2+b^2)^{-1}\{-2hbde^{-2ah} + h[c(a^2+b^2)-ad]e^{-ah}\sin(bh) -$$

$$- (2ac-d)e^{-ah}\sin(bh) + hbde^{-ah}\cos(bh)\} -$$

$$- 2ab^{-1}(a^2+b^2)^{-2}\{bde^{-2ah} - [c(a^2+b^2)-ad]e^{-ah}\sin(bh) -$$

$$- bde^{-ah}\cos(bh)\}$$

$$\frac{\partial b_2}{\partial b} = b^{-1}(a^2+b^2)^{-1}\{de^{-2ah} - h[c(a^2+b^2)-ad]e^{-ah}\cos(bh) -$$

$$- 2bce^{-ah}\sin(bh) - de^{-ah}\cos(bh) + hbde^{-ah}\sin(bh)\} -$$

$$- (a^2+3b^2)b^{-2}(a^2+b^2)^{-2}\{bde^{-2ah} - [c(a^2+b^2)-ad] \cdot$$

$$\cdot e^{-ah}\sin(bh) - bde^{-ah}\cos(bh)\}$$

$$\frac{\partial b_2}{\partial c} = - b^{-1}e^{-ah}\sin(bh)$$

$$\frac{\partial b_2}{\partial d} = b^{-1}(a^2+b^2)^{-1}[be^{-2ah} + ae^{-ah}\sin(bh) -$$
$$- be^{-ah}\cos(bh)]$$

APPENDIX C

Let

$$y(t) + ay(t-1) = bu(t-1) + \lambda e(t)$$

We get

$$\varepsilon(t) = (1 - aq^{-1})y(t) - bq^{-1}u(t)$$

i.e.

$$\begin{cases} \dfrac{\partial \varepsilon(t)}{\partial a} = -q^{-1}y(t) = -\dfrac{bq^{-2}}{1+aq^{-1}}u(t) - \dfrac{\lambda q^{-1}}{1+aq^{-1}}e(t) \\ \\ \dfrac{\partial \varepsilon(t)}{\partial b} = -q^{-1}u(t) \end{cases}$$

If u(t) white noise and independent of e(t) we get

$$\mathrm{Cov}(\hat{a},\hat{b}) \geq \frac{\lambda^2}{N}\begin{bmatrix} \dfrac{b^2\sigma_u^2 + \lambda^2}{1-a^2} & 0 \\ 0 & \sigma_u^2 \end{bmatrix}^{-1} =$$

$$= \frac{\lambda^2}{N\sigma_u^2(b^2\sigma_u^2+\lambda^2)}\begin{bmatrix} (1-a^2)\sigma_u^2 & 0 \\ 0 & b^2\sigma_u^2+\lambda^2 \end{bmatrix}$$

where σ_u^2 is the variance of u(t).

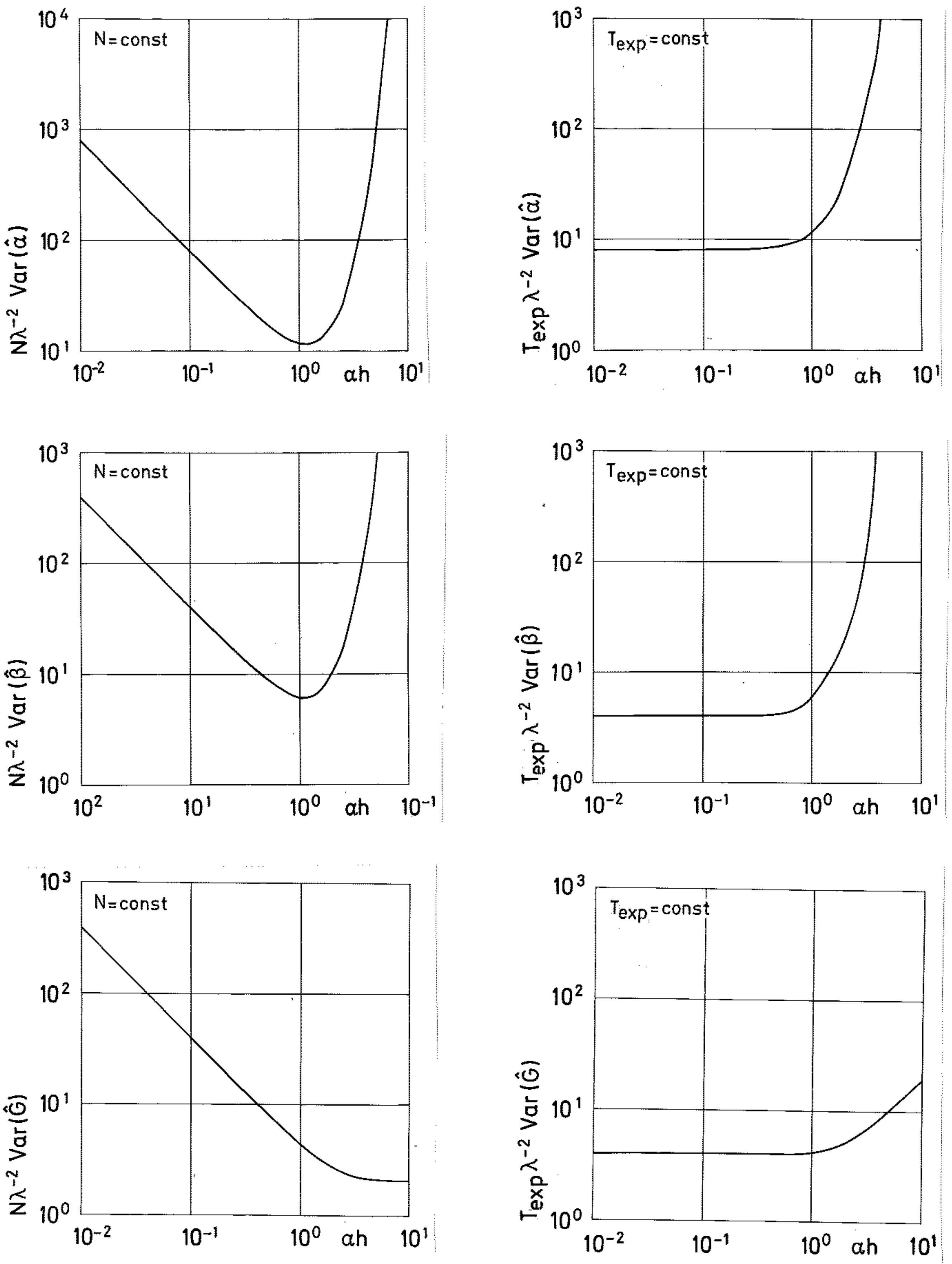

Fig.1 The achievable accuracy for estimates of parameters of a first order process $\beta/(s + \alpha)$ with white output noise shown as functions of αh. T_{exp} = Nh. Apart from a scaling factor the variance of $\hat{T}$ is equal to the variance of $\hat{\alpha}$.

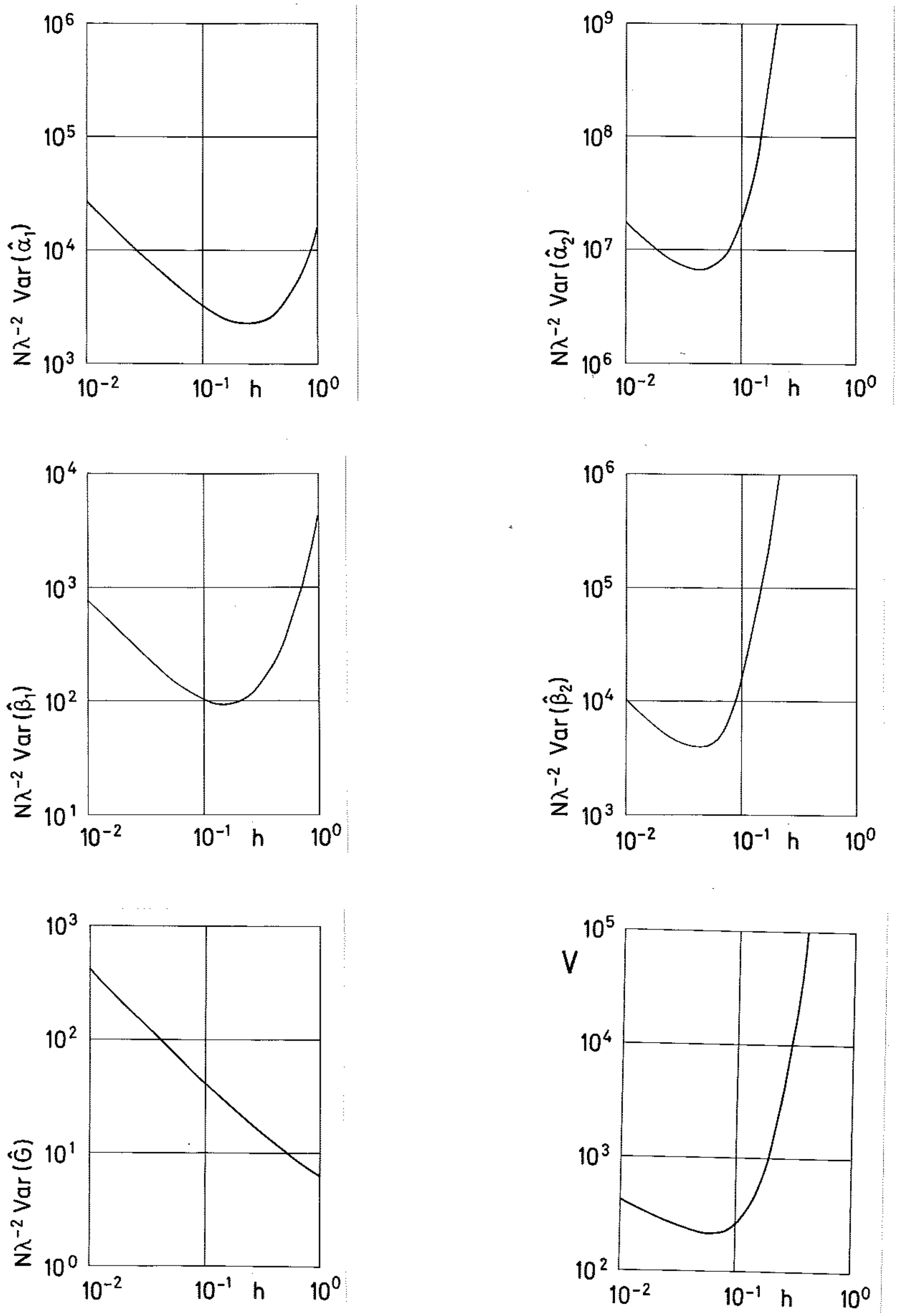

Fig.2 The achievable accuracy for estimates of parameters of a second order process $\alpha_1/(s + \beta_1) + \alpha_2/(s + \beta_2)$, where $\alpha_1=1$, $\alpha_2=25$, $\beta_1=0.2$ and $\beta_2=1$ with white output noise shown as function of h. In the lower right diagram the sum of relative errors for one parameter combination is shown. $V=\lambda^{-1}\sqrt{N}\sum_i\{\sqrt{Var(\hat{\theta}_i)}/|\theta_i|\}$, $\theta_i=T_1,T_2,G$. N=const. for all diagrams.

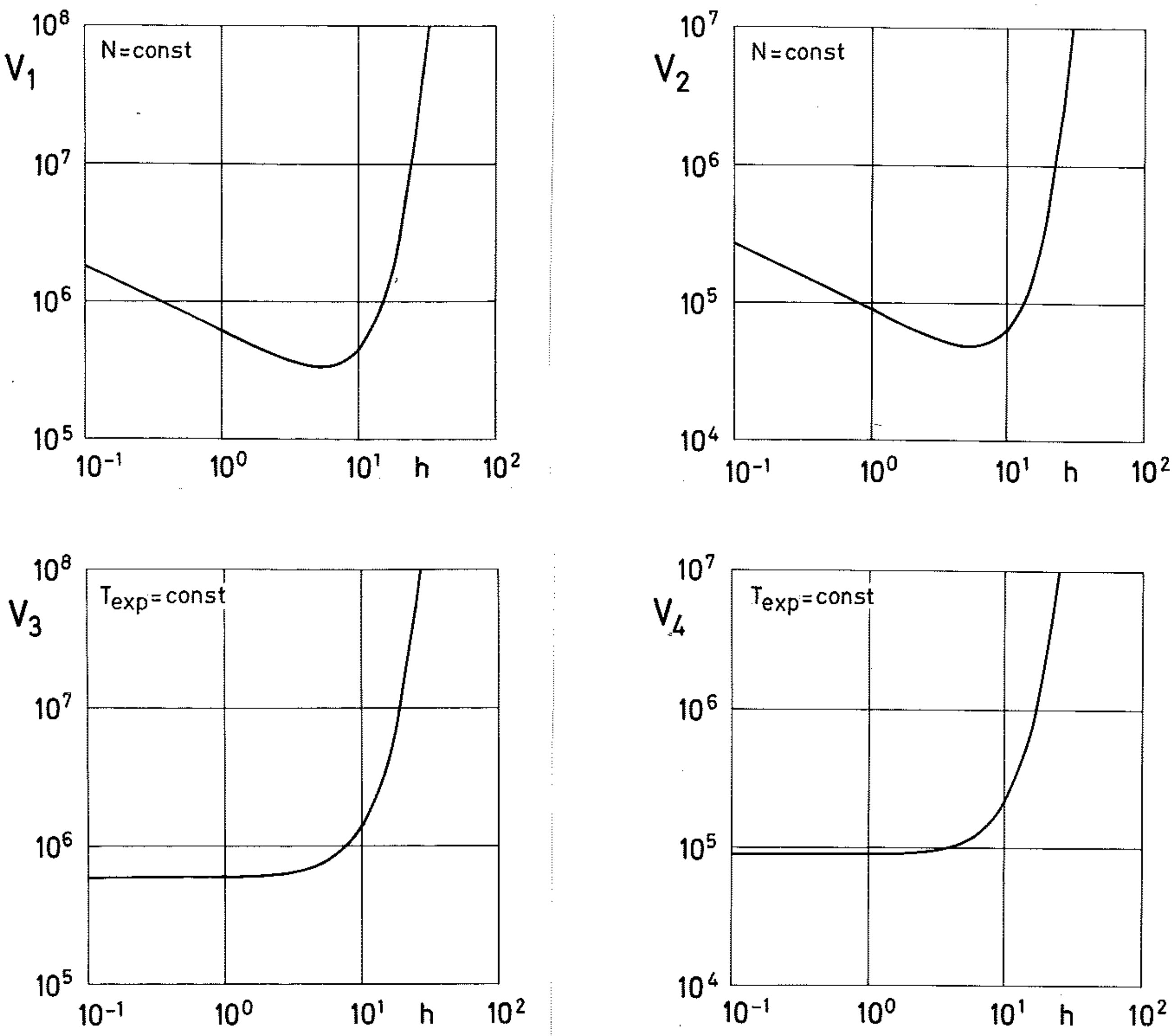

Fig.3 The sum of relative errors for different parameter combinations of a fourth order approximation of a one dimensional heat diffusion process as function of h. $T_{exp} = Nh$.

$$V_1 = \lambda^{-1}\sqrt{N}\sum_i\{\sqrt{Var(\hat{\theta}_i)}/|\theta_i|\} \quad \theta_i = \alpha_1,\ldots,\alpha_4,\beta_1,\ldots,\beta_4$$

$$V_2 = \lambda^{-1}\sqrt{N}\sum_i\{\sqrt{Var(\hat{\theta}_i)}/|\theta_i|\} \quad \theta_i = T_1,\ldots,T_4,G$$

$$V_3 = \lambda^{-1}\sqrt{T_{exp}}\sum_i\{\sqrt{Var(\hat{\theta}_i)}/|\theta_i|\} \quad \theta_i = \alpha_1,\ldots,\alpha_4,\beta_1,\ldots,\beta_4$$

$$V_4 = \lambda^{-1}\sqrt{T_{exp}}\sum_i\{\sqrt{Var(\hat{\theta}_i)}/|\theta_i|\} \quad \theta_i = T_1,\ldots,T_4,G$$

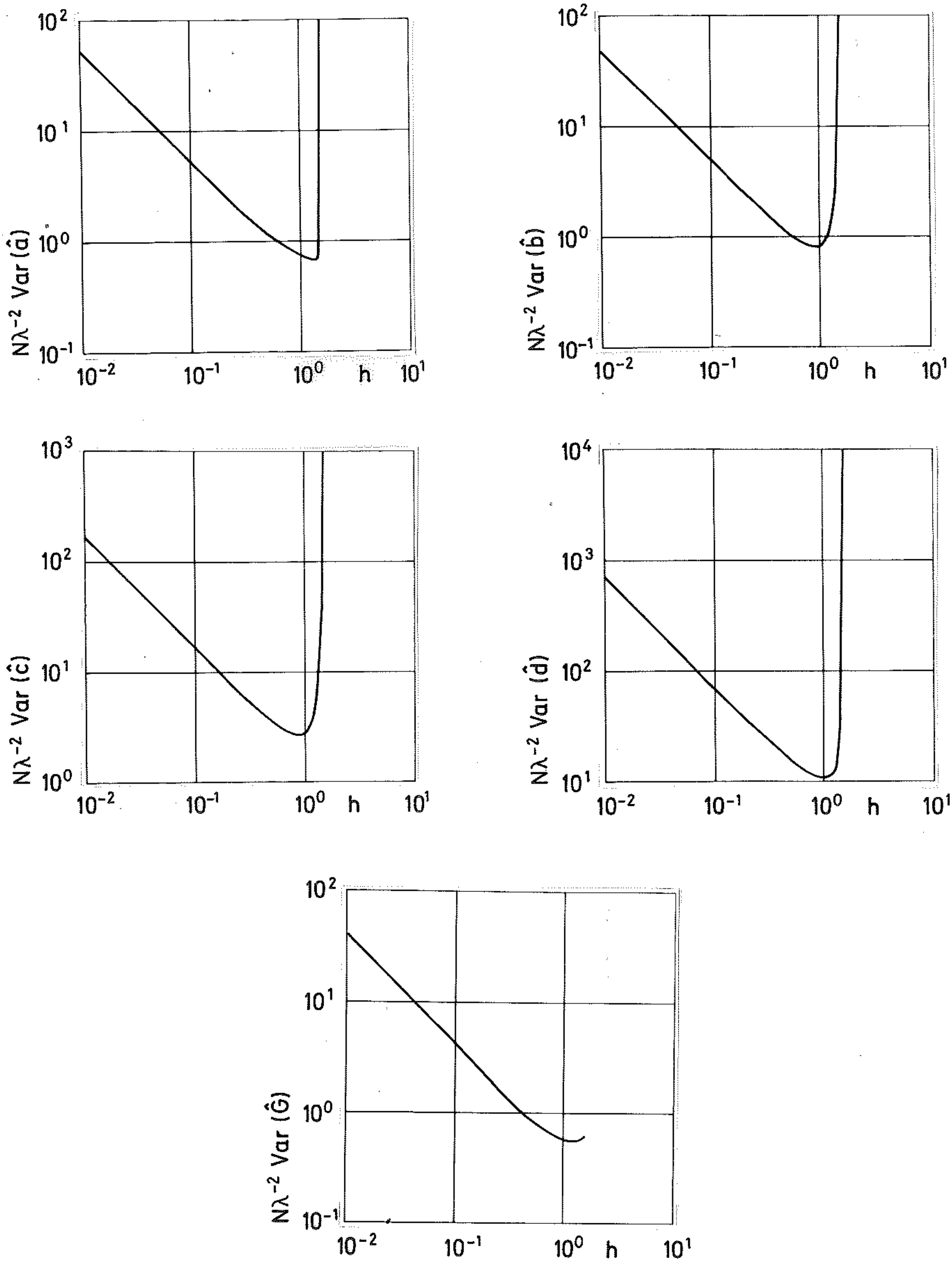

Fig.4 The achievable accuracy for estimates of parameters of a second order process $(cs + d)/((s + a)^2 + b^2)$ with $a=0.2$, $b=\sqrt{3.96}$, $c=0$, $d=1$ and with white output noise shown as function of h. N=const.

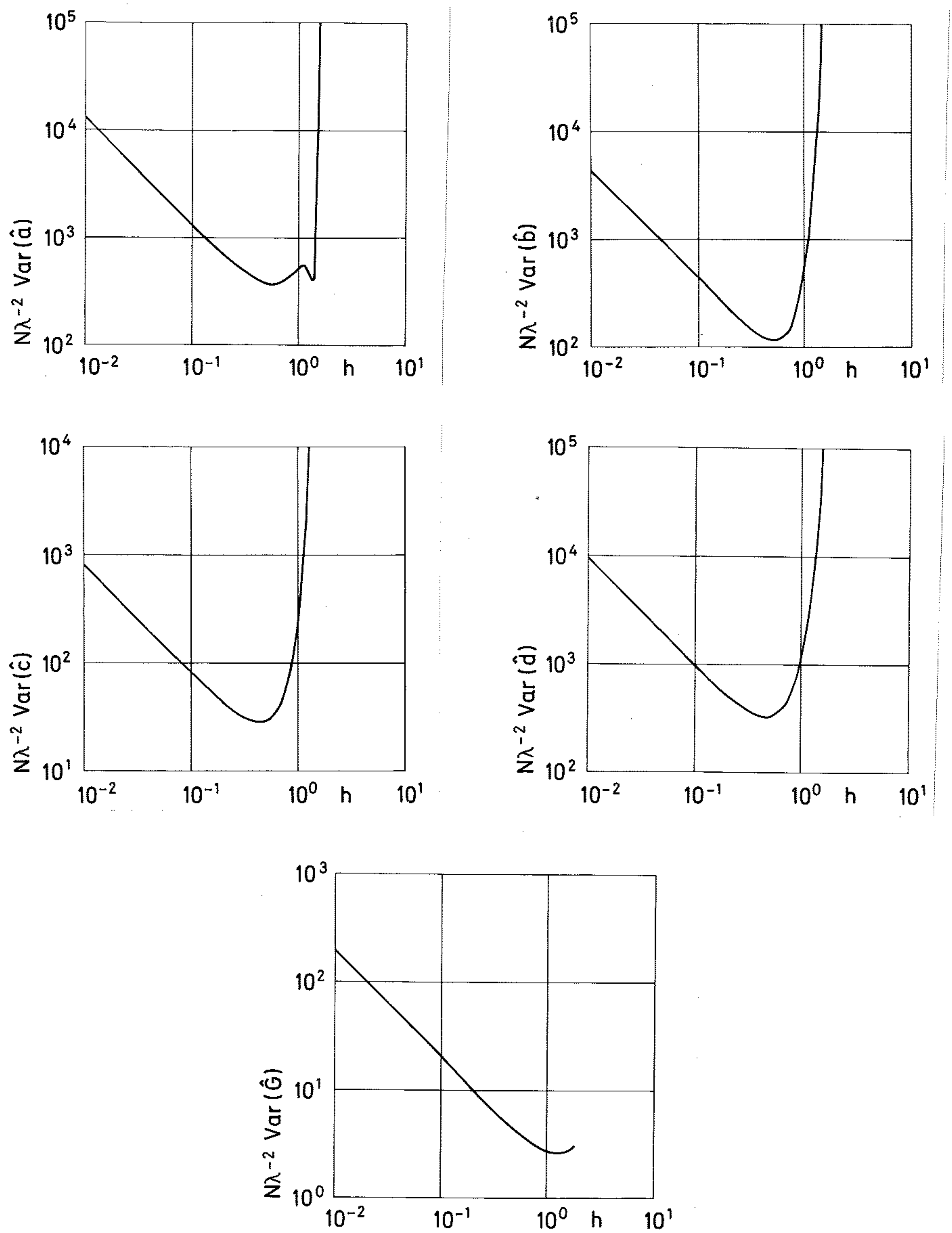

Fig.5 The achievable accuracy for estimates of parameters of a second order process $(cs + d)/((s + a)^2 + b^2)$ with $a=1$, $b=\sqrt{3}$, $c=0$, $d=1$ and with white output noise shown as function of h. N=const.

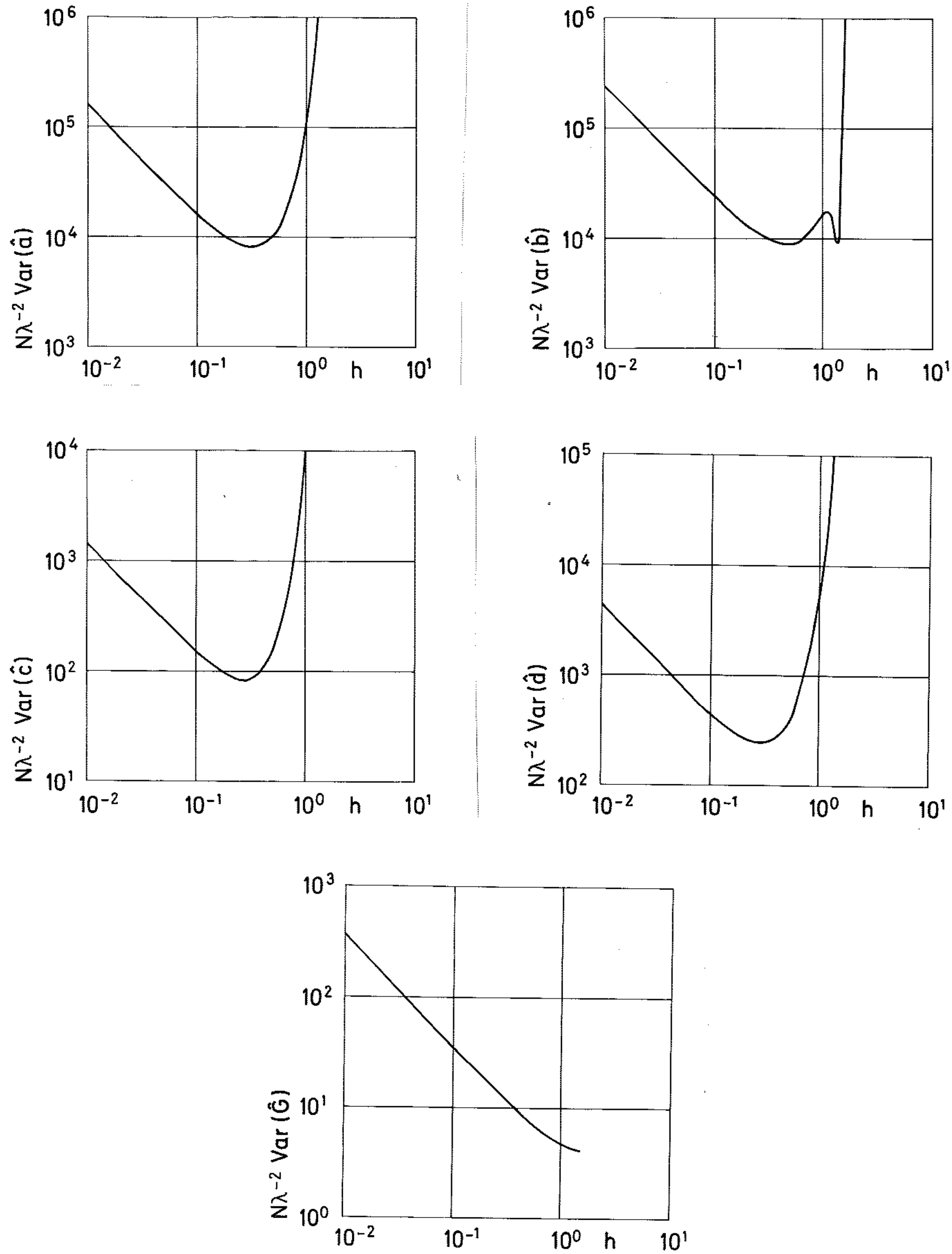

Fig.6 The achievable accuracy for estimates of parameters of a second order process $(cs + d)/((s + a)^2 + b^2)$ with $a=1.8$, $b=\sqrt{0.76}$, $c=0$, $d=1$ and with white output noise shown as function of h. N=const.

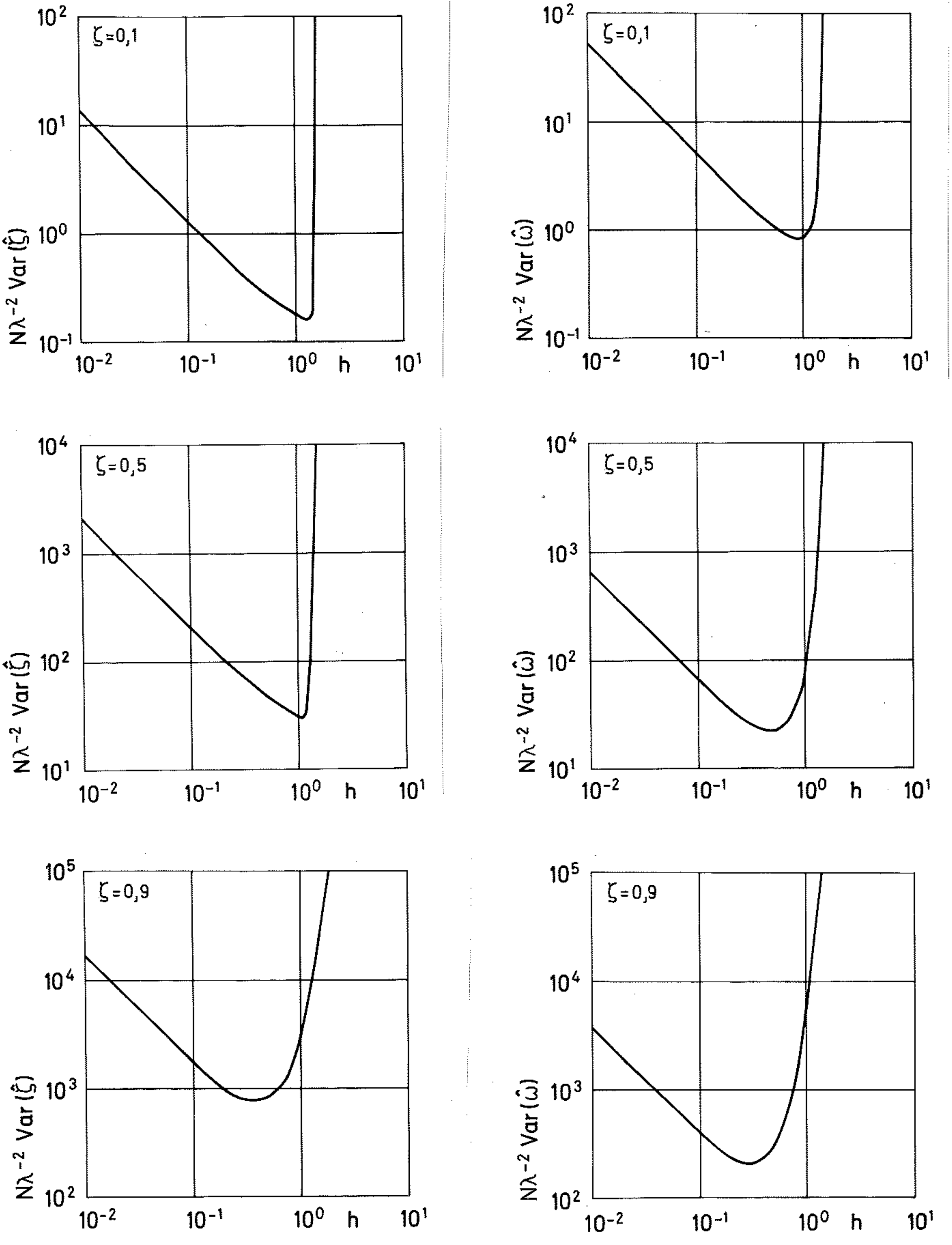

Fig.7 The achievable accuracy for estimates of parameters of second order processes $(cs + d)/(s^2 + 2\zeta\omega + \omega^2)$ with $\omega=2, c=0, d=1$ and varying damping factor, ζ, with white output noise shown as function of h. N=const.

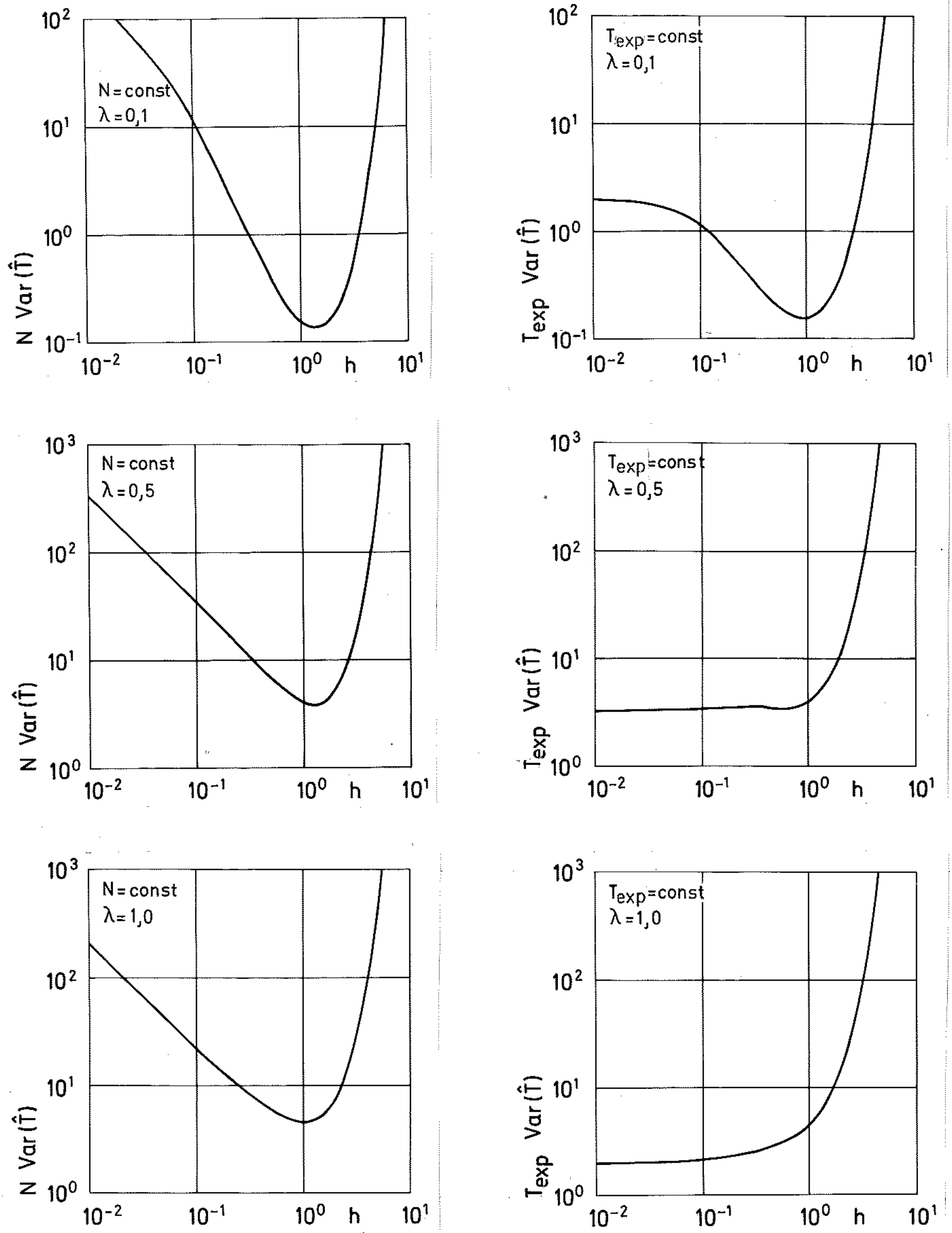

Fig.8 The achievable accuracy of the time constant, $T=1/\alpha$, of the system $\beta/(s+\alpha)$ with $\alpha=\beta=1$, using the least squares structure for the modelling shown as function of h. The variance of the noise in the discrete time model, λ, has been varied. $T_{exp}=Nh$.

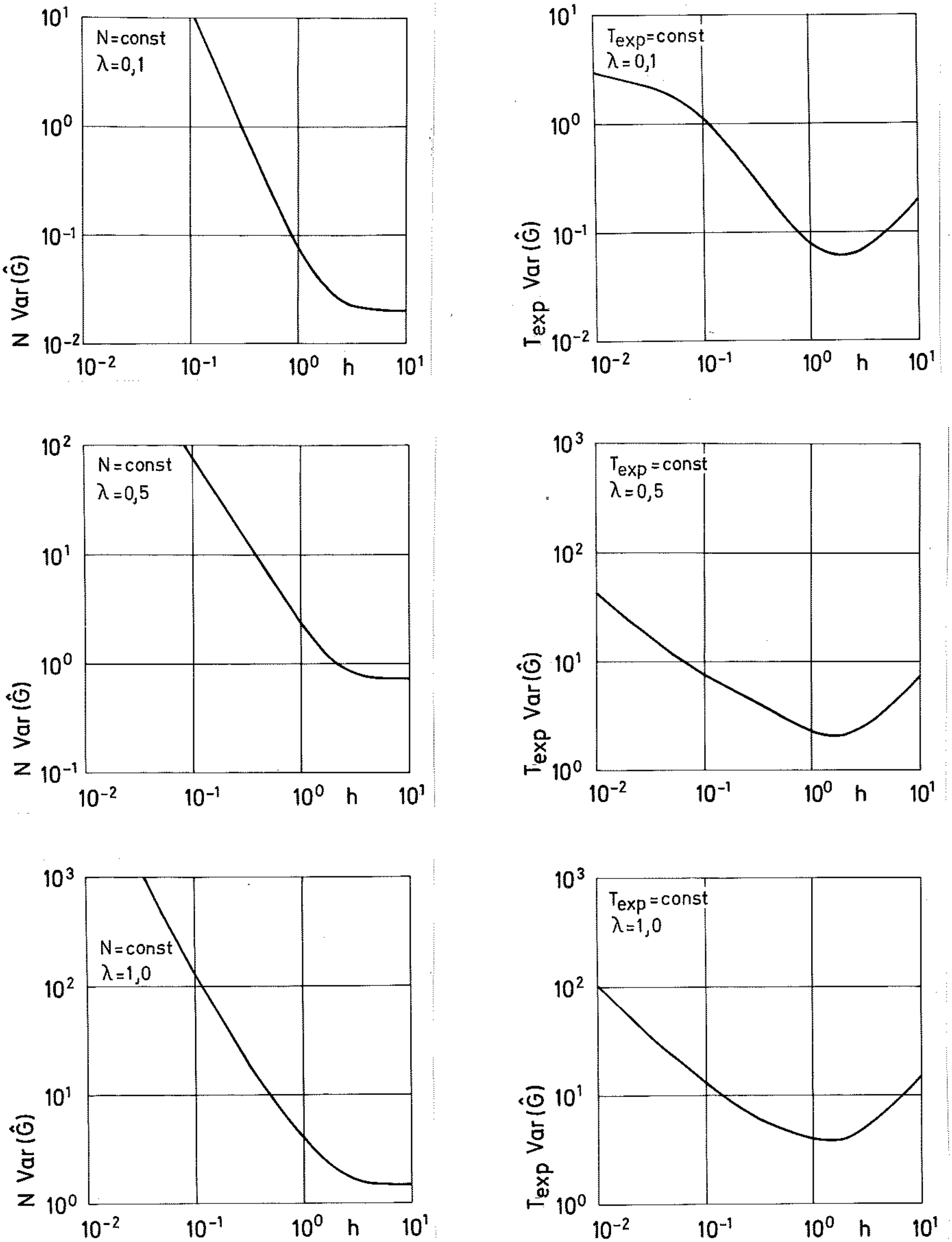

Fig.9 The achievable accuracy of the gain, $G=\beta/\alpha$, of the system $\beta/(s + \alpha)$ with $\alpha=\beta=1$, using the least squares structure for the modelling shown as function of h. The variance of the noise in the discrete time model, λ, has been varied. $T_{exp}=Nh$.

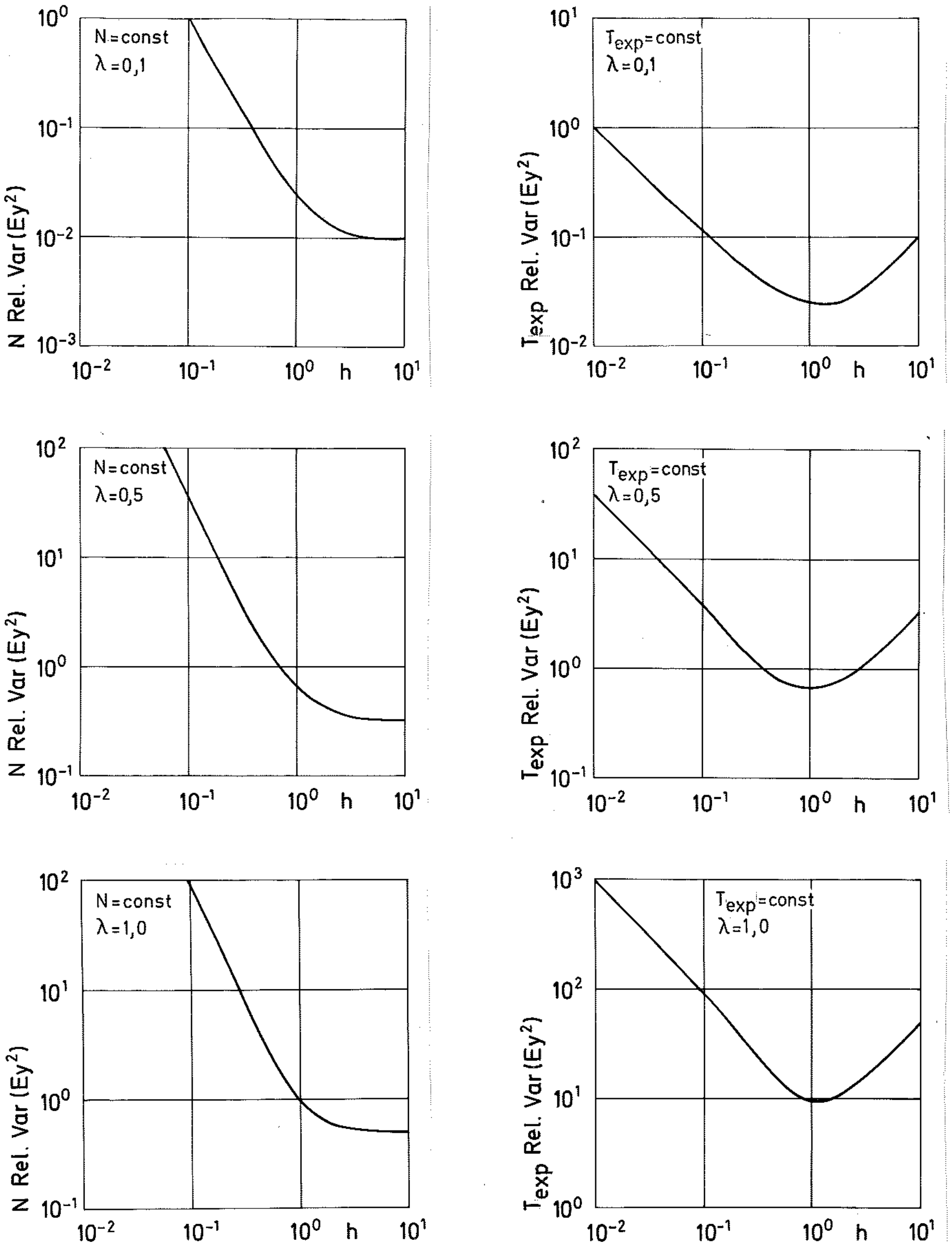

Fig.10 The relative variation of the expected output variance (Rel. var(Ey^2)) for the system $\beta/(s + \alpha)$ with $\alpha=\beta=1$, using the least squares structure for the modelling shown as function of h. The variance of the noise in the discrete time model, λ, has been varied. $T_{exp}=Nh$.