A comparison of some process parameter estimating schemes

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

A COMPARISON OF SOME PROCESS PARAMETER
ESTIMATING SCHEMES

by

A.J.W. van den Boom and J.H.A.M. Melis

A COMPARISON OF SOME PROCESS PARAMETER ESTIMATING SCHEMES

by

A.J.W. van den Boom and J.H.A.M. Melis

TH-Report 68-E-03

A COMPARISON OF SOME PROCESS PARAMETER ESTIMATING SCHEMES

A.J.W. van den Boom and J.H.A.M. Melis
Department of Electrical Engineering
Technological University
Eindhoven, Netherlands

Summary

In this paper some principal problems in the field of process parameter
estimation are discussed, especially with respect to the uncertainty in
the estimation, caused by additive noise and approximated model structures.
Two basic approaches, i.e. the instrumentation of explicit mathematical
relations and the model matching technique, are compared theoretically with
respect to the accuracy as a function of the observation interval.
In some situations both methods, each requiring matrix inversion, can
sufficiently be approximated by the use of a non-orthogonal estimating
scheme without matrix inversion.
A summarizing discussion is given on additional errors due to the
approximation of the ideal estimating procedure.

Contents

1. Introduction

In many modern applications of control systems, it is a matter of importance to obtain the required information of the process to be controlled, in order to aim at an optimal action or correct adaptation of the system. In optimal, self-optimizing and adaptive systems the quality of the control greatly depends on the amount of knowledge supplied to the controller. Besides, computers are increasingly used as control elements for complicated processes.

The growing demands concerning speed and quality of the control and consequently the insight into relevant process parameters justify an extension of the computer's task to process input and output data in order to obtain a better knowledge of the proces.

In an effort to determine the characteristics of a process on the basis of input and output data, we are searching for a dynamic operator which, acting on the input signal, results in the "best" estimate of the output signal.

The numerical values of the estimates of the process parameters can be obtained from a model. The difference $e(t)$ between process and model output can be used as a measure for the error in the estimate, cf. fig. 1. Minimization of some function or functional of the error yields as model parameters $\tilde{\beta}$ estimates which are optimal with respect to the criterion chosen. Minimization of a quadratic type of criterion yields the least squares estimate.

Actually fig. 1 does not represent the most general situation. Bayes and maximum likelihood estimation may take into account more a priori knowledge that may be available [1]. Fig. 1, however, represent by far the most important estimation situation. Moreover it is easily shown that the maximum likelihood estimation method reduces to the (generalized) least squares method if the additive noise is Gaussian. As these least squares methods represent a great majority of parameter estimation cases, this paper is devoted to a closer investigation of their properties.

With respect to the instrumentation we distinct two classes [2] :

1) the explicit method

the parameters are determined by instrumentation of the
mathematical relations resulting from the minimization of the
criterion;

2) the implicit method

the parameters are determined by adjustment of a model.

Many publications deal with different approaches to the problem of parameter
estimation, some of them discussing a specific process. A comparison of
properties of the different types of instrumentation hardly ever occurs.
Therefore, it is desirable to compare explicit and implicit methods for
the case of models, which are linear in the parameters.

In principle the estimation of process parameters can be done both on analog
and digital computers. Especially with respect to the simulation the digital
machine has the following important advantages:

1) the memory of the digital machine is well suited for the simulation
   of processes with long response times or processes with delay times;

2) the errorless repeating generation of identical test and disturbing
   signals is apart from being throughout possible also of great im-
   portance where the influence of relatively short measuring intervals
   is studied in relation to noise effects.

With respect to the instrumentation of the estimation mechanism, the digital
computer has the following practical advantages:

1) an intermittent measuring and model adjusting procedure can easily
   be instrumented;

2) fundamental mathematical manipulations such as matrix calculations
   and statistical computations over a certain amount of estimates
   can proceed accurately and be easily programmed.

Our investigations mainly concentrated toward an instrumentation using digital
techniques, leading to a way of description of the estimation procedure as given
below.

If the data are derived from analog processes, one has to count with
additional errors caused by imperfect analog-digital conversion, e.g. due
to quantization errors. In order to avoid this type of errors we only
consider discrete processes.

As the example to be discussed we take a linear system, the impulse response of which is represented by the vector $\underline{h}^T = (h(1),h(2),...,h(p))$ cf. [3] . This can be simulated by means of a delay line, cf. fig. 2.

For the description of the explicit method we define:

$$\underline{x}_1^T = (x(1),x(2),.....,x(\ell),0...,0) \quad \text{vector of inputs}$$

$$\underline{x}_2^T = (0,x(1),.............,\ x(\ell),0,....,0)$$

$$\underline{y}^T = (y(1),y(2),.....,y(\ell+p-1)) \qquad \text{vector of outputs}$$

$$\underline{n}^T = (n(1),n(2),.....,n(\ell+p-1)) \qquad \text{vector of additive noise}$$

$$\underline{w}^T = (w(1),w(2),.....,w(\ell+p-1)) \qquad \text{vector of model outputs}$$

$$\underline{\beta}^T = (\beta(1),\beta(2),.....,\beta(p)) \qquad \text{parameter vector of the model}$$

$$\underline{h}^T = (h(1),h(2),.....,h(p)) \qquad \text{parameter vector of the process}$$

$$X = \begin{bmatrix} x(1) & 0 & .... & 0 \\ x(2) & x(1) & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & x(1) \\ x(\ell) & \vdots & & \vdots \\ 0 & x(\ell) & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ....& x(\ell) \end{bmatrix} \qquad = (\underline{x}_1,\ \underline{x}_2,...,\ \underline{x}_p) \qquad (1)$$

This yields:

$$\underline{z} = \underline{y} + \underline{n} = X\ \underline{h} + \underline{n}$$

$$\underline{w} = X\ \underline{\beta} \qquad\qquad\qquad\qquad (2)$$

$$\underline{e} = \underline{z} - \underline{w}$$

The length of the observation sequence is ℓ.

For the description of the implicit method we will consider the jth observation interval; each interval having a length ℓ^*.
During this observation time the model vector is kept constant $(= \underline{\beta}_{j-1})$ prior to the adjustment from $\underline{\beta}_{j-1}$ to $\underline{\beta}_j$ (intermittent adjustment procedure). For the input and output quantities of this interval we write:

$$\underline{z}_j^T = (z\ \{(j-1)\ell^*+1\ \},......,\ z\{\ j\ell^*\ \})$$

$$\underline{e}_j^T = (e\ \{(j-1)\ell^*+1\ \},......,\ e\{\ j\ell^*\ \}) \qquad (3)$$

$$\underline{\beta}_{j-1}^T = (\beta_{j-1}(1),......,\ \beta_{j-1}(p))$$

$$
X_j = \begin{bmatrix}
x\{(j-1)\, \ell^{*}+1\} & x\{(j-1)\, \ell^{*}\} & \ldots\ldots\ldots & x\{(j-1)\, \ell^{*}-p+2\} \\
x\{(j-1)\, \ell^{*}+2\} & & & \vdots \\
\vdots & & & \vdots \\
\vdots & & & \vdots \\
\vdots & & & \vdots \\
x\{j\, \ell^{*}\} & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & & x\{j\, \ell^{*}-p+1\}
\end{bmatrix}
$$

This yields:

$$
\begin{aligned}
\underline{z}_j &= X_j\, \underline{h} + \underline{n}_j \\
\underline{w}_j &= X_j\, \underline{\beta}_{j-1} \\
\underline{e}_j &= \underline{z}_j - \underline{w}_j
\end{aligned}
\tag{4}
$$

In all cases we will suppose that the output noise has zero mean and is independent of the input signal, i.e.

$$
\begin{aligned}
E\left[\underline{n}\right] &= \underline{0} \\
E\left[X^T\underline{n}\right] &= \underline{0}
\end{aligned}
\tag{5}
$$

The criterion R to be minimized is

$$
R = \underline{e}^T\, \Phi\, \underline{e}
\tag{6}
$$

The choice of Φ is guided by the available a priori knowledge.

The __explicit__ method yields an estimate $\tilde{\underline{\beta}}$ of $\underline{h}$ according to

$$
\tilde{\underline{\beta}} = (X^T\Phi X)^{-1}\, X^T\Phi\underline{z}
\tag{7}
$$

under the following conditions:

   a) the numerical value of the impulse response h(i) must be constant after a given i;

   b) at the beginning of the observation the model must have reached a state identical to that of the process, if no disturbances affect the process; i.e. either $x(i) = 0 \quad i<0$
or, in the case the test signal constituting a part of a continuous sequence, only those measured outputs $\underline{z}$ appearing after i=p are taken into account;

   c) the matrix $X^T\Phi X$ must not be singular and must have an inverse.

Under the same conditions, the __implicit__ method yields an estimate $\underline{\beta}_j$ after j adjustments according to

$$\underline{\beta}_j = \underline{\beta}_{j-1} - \frac{g_j}{2} \; (X_j^T \Phi X_j)^{-1} \; \nabla_{\underline{\beta}_{j-1}} \; R_j \tag{8}$$

where g_j is a factor governing the adjusting speed and $\nabla_{\underline{\beta}_{j-1}}$ denotes the gradient with respect to $\underline{\beta}_{j-1}$.

The least squares estimate is that estimate where no a priori knowledge is available, i.e. $\Phi = I$, I being the identity matrix. The input signal X is measurable. Its bandwidth is assumed large compared to the bandwidth of the process ("white" input signal).

Both explicit and implicit techniques request matrix inversion, which of course is a very expensive operation, but which has the advantage that arbitrary stochastic testsignals can be used.

A variant of the above technique is frequently used and can be instrumented in a simple way, because the matrix inversion can be avoided. A restriction inherent to this method is the fact that only a "white" input signal should be used as test signal.
For the <u>explicit</u> method the estimate without matrix inversion follows from

$$\tilde{\beta}(j) = \frac{\tilde{\psi}_{xy}(j,\ell)}{\tilde{\psi}_{xx}(0,\ell)} \quad \text{with} \quad \tilde{\psi}_{xy}(j,\ell) = \frac{1}{\ell} \sum_{i=1}^{\ell} x(i)y(j+i) \tag{9}$$

A general insight into the estimates (7) and (9) can be obtained by considering that h(i) can be written explicitly in the two following ways if a "white" test signal is used.

a) $\quad h(i) = \dfrac{\psi_{xy}(i)}{\psi_{xx}(0)}$

$$b) \quad h(i) = \frac{\tilde{\psi}_{xy}(i,\ell)}{\tilde{\psi}_{xx}(0,\ell)} - \frac{1}{\tilde{\psi}_{xx}(0,\ell)} \left[\{h(i+1)+h(i-1)\} \; \tilde{\psi}_{xx}(1,\ell-1)+\ldots\ldots\ldots \right] \tag{10}$$

compensating terms

The procedure without matrix inversion neglects the compensating terms.
For the <u>implicit</u> method the estimate without matrix inversion follows from

$$\underline{\beta}_j = \underline{\beta}_{j-1} + g_j \{ \ell^* \; \psi_{xx}(0) \}^{-1} \; X_j^T \; \underline{e}_j \tag{11}$$

- 6 -

The neglection of the compensating terms causes an additional error. This error is called the "truncation error", which means in terms of model structure that the model has a shorter impulse response than that of the process, which means that condition a) is not fulfilled. If the conditions a), b) and c) are fulfilled, the uncertainty in the estimate is only caused by the disturbing noise.

2. Estimation schemes

a) the explicit method with matrix inversion

For the explicit method we found

$$\tilde{\underline{\beta}} = (X^T X)^{-1} X^T \underline{z} \tag{12}$$

Taking the expectation of $\tilde{\underline{\beta}}$ yields

$$E\left[\tilde{\underline{\beta}}\right] = E\left[(X^T X)^{-1} X^T \underline{z}\right] = \underline{h} \tag{13}$$

The estimate $\tilde{\underline{\beta}}$ is unbiased.

The noise error is written as follows

$$\Delta\tilde{\underline{\beta}} = \tilde{\underline{\beta}} - \underline{h} = (X^T X)^{-1} X^T (\underline{y} + \underline{n}) - \underline{h} =$$

$$= (X^T X)^{-1} X^T (X\underline{h} + \underline{n}) - \underline{h} = (X^T X)^{-1} X^T \underline{n} \tag{14}$$

From a statistical point of view the standard deviation of $\Delta\tilde{\underline{\beta}}$ is a useful measure for the error in the estimate. This can be found by calculating the covariance matrix of $\tilde{\underline{\beta}}$

$$\text{cov } \tilde{\underline{\beta}} = E\left[\Delta\tilde{\underline{\beta}} \cdot \Delta\tilde{\underline{\beta}}^T\right] = (X^T X)^{-1} X^T E\left[\underline{n} \cdot \underline{n}^T\right] X (X^T X)^{-1} \tag{15}$$

For "white" additive noise $\underline{n}$

$$E\left[\underline{n} \cdot \underline{n}^T\right] = \psi_{nn}(0) I \tag{16}$$

$\psi_{nn}(i)$ being the autocorrelation function of $\underline{n}$. This yields for the covariance

$$\text{cov } \tilde{\underline{\beta}} = \psi_{nn}(0) (X^T X)^{-1} \tag{17}$$

For large ℓ and "white" input noise:

$$\text{cov } \tilde{\underline{\beta}} \approx \frac{\psi_{nn}(0)}{\ell \psi_{xx}(0)} I \tag{18}$$

For the standard deviation σ in each parameter we find:

$$\sigma = \sqrt{\frac{\psi_{nn}(0)}{\ell \psi_{xx}(0)}} \tag{19}$$

Listing the properties of the explicit method with matrix inversion, we come to the following specifications:

a) the estimate is linear: $\tilde{\underline{\beta}} = Q \underline{z}$;

b) the estimate is unbiased: $E\left[\tilde{\underline{\beta}}\right] = \underline{h}$;

c) $\lim\limits_{\ell \to \infty} \text{cov } \tilde{\underline{\beta}} = 0 \ I$;

d) the method is suited for all types of input signals, provided all process modes are excited;

e) the method is suited for all types of additive noise;

f) the instrumentation requires matrix inversion, matrix-matrix multiplication and two matrix-vector multiplications.

b) the implicit method with matrix inversion

For the implicit method we found:

$$\underline{\beta}_j = \underline{\beta}_{j-1} - \frac{g_j}{2} (X_j^T X_j)^{-1} \nabla_{\underline{\beta}_{j-1}} R_j \tag{20}$$

or

$$\underline{\beta}_j = \underline{\beta}_{j-1} + g_j (X_j^T X_j)^{-1} X_j^T \underline{e}_j$$

Taking the expectation of $\underline{\beta}_j$ we find:

$$E\left[\underline{\beta}_j\right] = g_j \ \underline{h} + (1-g_j) \ E\left[\underline{\beta}_{j-1}\right] \tag{21}$$

leading to

$$E\left[\underline{\beta}_j\right] = \underline{h} \sum_{i=1}^{i} g_i \prod_{k=i+1}^{j} (1-g_k) + \underline{\beta}_0 \prod_{i=1}^{j} (1-g_i) \tag{22}$$

where $\underline{\beta}_0$ is the initial model guess (a priori knowledge). Where a priori information is lacking, $\underline{\beta}_0$ can be chosen arbitrary e.g. $\underline{\beta}_0 = \underline{0}$.
Considering first

$$\underline{\beta}_j - E\left[\underline{\beta}_j\right] = (1-g_j) \left(\underline{\beta}_{j-1} - E\left[\underline{\beta}_{j-1}\right]\right) + g_j (X_j^T X_j)^{-1} X_j^T \underline{n}_j \tag{23}$$

we derive

$$\text{cov } \underline{\beta}_j = (1-g_j)^2 \text{ cov } \underline{\beta}_{j-1} + g_j^2 (X_j^T X_j)^{-1} X_j^T E\left[\underline{n} \ \underline{n}^T\right] X_j (X_j^T X_j)^{-1} \tag{24}$$

If $\underline{n}$ is "white" noise (24) can be rewritten as a vari-linear difference equation:

$$(1-g_j)^2 \ \Delta \text{cov } \underline{\beta}_j - g_j (g_j-2) \text{ cov } \underline{\beta}_j = \frac{g_j^2}{\ell^*} \ \psi_{nn}(0) \ \tilde{\psi}_{x_j}^{-1} \tag{25}$$

- 8 -

where

$$\Delta \text{cov}\, \underline{\beta}_j = \text{cov}\, \underline{\beta}_j - \text{cov}\, \underline{\beta}_{j-1} \tag{26}$$

and

$$\tilde{\psi}_{X_j} = \frac{1}{\ell^*}\, X_j^T X_j \tag{27}$$

Now we can distinguish the following cases:

1) $g_j = 1/j$ "stochastic approximation", cf. [4].

2) $g_j = \text{constant}$ "parameter tracking"

<u>ad 1)</u> $g_j = 1/j$

The algorithm is:

$$\underline{\beta}_j = \underline{\beta}_{j-1} + 1/j\, (X_j^T X_j)^{-1}\, X_j^T \underline{e}_j \tag{28}$$

Considering that in this particular case

$$\sum_{i=1}^{j} g_i \prod_{k=i+1}^{j} (1-g_k) = 1$$

$$\prod_{i=1}^{j} (1-g_i) = 0 \tag{29}$$

$$\lim_{j \to \infty} \prod_{i=a}^{j} (1-g_i) = 0 \qquad a \geq 1$$

it follows from equation (22) that

$$E\left[\underline{\beta}_j\right] = \underline{h} \tag{30}$$

This method of model matching yields an unbiased estimate for all j, irrespective of the initial model guess!

Equation (25) becomes

$$(j-1)^2 \Delta \text{cov}\, \underline{\beta}_j + (2j-1)\, \text{cov}\, \underline{\beta}_j = \frac{\psi_{nn}(0)}{\ell^*}\, \tilde{\psi}_{X_j}^{-1} \tag{31}$$

If the observation interval ℓ^* is large enough and if X_j is "white" noise the following approximation can be used for $\tilde{\psi}_{X_j}^{-1}$

$$\tilde{\psi}_{X_j}^{-1} \approx \frac{I}{\psi_{xx}(0)} \tag{32}$$

Equation (26) then becomes

$$(j-1)^2 \Delta \text{cov}\, \underline{\beta}_j + (2j-1)\, \text{cov}\, \underline{\beta}_j = \frac{\psi_{nn}(0)}{\ell^* \psi_{xx}(0)}\, I \tag{33}$$

with the solution

$$\text{cov } \underline{\beta}_j = \frac{\psi_{nn}(0)}{j\ell^* \psi_{xx}(0)} \ I = \frac{\psi_{nn}(0)}{\ell \psi_{xx}(0)} \ I \tag{34}$$

where ℓ is the equivalent observation length used in the explicit method. In fig. 3 the diagonal elements of cov $\underline{\beta}_j$ are plotted as a function of $j\ell^*$.

For relatively small ℓ^* the approximation of eq. (32) is not longer valid: experimental results show a shift (dotted lines) towards greater variances.

Recapulating the properties of the adjustment procedure with matrix inversion and $g_j = 1/j$, we come to the following specification:

   a) the adjustment is unbiased for all j;

   b) $\lim\limits_{j \to \infty} \text{cov } \underline{\beta}_j = 0 \ I$ for ℓ^* large enough

   c) the method is suited for all types of input noise;

   d) the method is suited for all types of additive noise;

   e) the method requires for every interval a matrix inversion, a matrix-matrix multiplication and two matrix-vector multiplications;

   f) the method yields intermediate results.

   g) the method is not suited for parameter tracking.

<u>ad 2)</u> $g_j = $ constant

The algorithm now is

$$\underline{\beta}_j = \underline{\beta}_{j-1} + c \ (X_j^T X_j)^{-1} \ X_j^T \ \underline{e}_j \tag{35}$$

Substitution of

$$\sum_{i=1}^{j} g_i \prod_{k=i+1}^{j} (1-g_k) = 1 - (1-c)^j \tag{36}$$

in eq. (22) and considering

$$\left. \begin{array}{l} \lim\limits_{j \to \infty} \ 1 - (1-c)^j = 1 \\[2em] \lim\limits_{j \to \infty} \ (1-c)^j \ \ \ = 0 \end{array} \right\} \ \ 0 < c < 2 \tag{37}$$

yields

$$E\left[\underline{\beta}_j\right] = \{ \ 1-(1-c)^j \ \} \ \underline{h} + (1-c)^j \ \underline{\beta}_0 \tag{38}$$

For the covariance we find

$$\text{cov } \underline{\beta}_j = (1-c)^2 \text{ cov } \underline{\beta}_{j-1} + c^2 \frac{\psi_{nn}(0)}{\ell^*} \tilde{\Psi}_{X_j}^{-1} \tag{39}$$

or written as a difference equation

$$(1-c)^2 \Delta\text{cov } \underline{\beta}_j - c(c-2) \text{ cov } \underline{\beta}_j = c^2 \frac{\psi_{nn}(0)}{\ell^*} \tilde{\Psi}_{X_j}^{-1} \tag{40}$$

If ℓ^* is large enough and if a "white" input signal is used:

$$(1-c)^2 \Delta\text{cov } \underline{\beta}_j - c(c-2) \text{ cov } \underline{\beta}_j = c^2 \frac{\psi_{nn}(0)}{\ell^* \psi_{xx}(0)} I \tag{41}$$

The asymptotic solution is:

$$\lim_{j \to \infty} \text{cov } \underline{\beta}_j = \frac{c}{2-c} \frac{\psi_{nn}(0)}{\ell^* \psi_{xx}(0)} I \tag{42}$$

This method of model matching yields even after an infinite observation interval an estimate $\underline{\beta}_\infty$ with a variance greater than zero. This version of the adjustment algorithm is important for the identification of processes with slowly varying parameters (parameter tracking).

c) the explicit method without matrix inversion

As already pointed out above, the algorithm for the explicit estimation without matrix inversion is

$$\tilde{\underline{\beta}} = \frac{1}{\ell \tilde{\psi}_{xx}(0,\ell)} X^T \underline{z} \tag{43}$$

where X is a "white" input signal.
In eq. (10) it can be seen that in this situation a truncation occurs:

$$\tilde{\underline{\beta}} = \underline{h} + \Delta\underline{\beta}_{trunc} + \Delta\underline{\beta}_{noise} \tag{44}$$

$$\Delta\underline{\beta}_{noise} = \frac{1}{\ell \tilde{\psi}_{xx}(0,\ell)} X^T \underline{n} = \Delta\underline{\beta}_n \tag{45}$$

Considering an arbitrary "white" input matrix X, we obtain:

$$E\left[\tilde{\underline{\beta}}\right] = \underline{h} + \Delta\underline{\beta}_{trunc} \tag{46}$$

where $\Delta\underline{\beta}_{trunc}$ depends on the matrix X.
The algorithm (43) yields a biased estimate.

Taking

$$\underline{\Delta\beta}_n \cdot \underline{\Delta\beta}_n^T = \frac{1}{\ell^2 \tilde{\psi}_{xx}^2(0,\ell)} X^T \underline{n}\, \underline{n}^T X \tag{47}$$

we get in the case of "white" additive noise

$$\text{cov } \underline{\beta}_n = \frac{\psi_{nn}(0)}{\ell^2 \tilde{\psi}_{xx}^2(0,\ell)} X^T X \tag{48}$$

For large ℓ we write

$$\text{cov } \underline{\beta}_n \approx \frac{\psi_{nn}(0)}{\ell \psi_{xx}(0)} \; I \tag{49}$$

In terms of standard deviation for each parameter

$$\sigma = \sqrt{\frac{\psi_{nn}(0)}{\ell \psi_{xx}(0)}} \tag{50}$$

Recapitulating the properties of the explicit method without matrix inversion we find:

    a) the estimate is linear $\tilde{\underline{\beta}} = Q \underline{z}$;

    b) the estimate is biased because $\frac{1}{\ell} X^T X = \tilde{\Psi}_X \neq \Psi_X$

        however the following holds

$$\lim_{\ell \to \infty} E\left[\tilde{\underline{\beta}}\right] = \underline{h}$$

$$E_X \; E\left[\tilde{\underline{\beta}}\right] = \underline{h}, \text{ where } E_X \text{ denotes the expectation with respect to X.}$$

    c) the method is only suited for "white" input noise;

    d) the method is suited for all types of output noise;

    e) the instrumentation only requires a multiplication of a

       $(\ell+p-1) \times p$ matrix by a $\ell+p-1$ vector.

c) the implicit method without matrix inversion

In equation (20) we can approximate the matrix inversion in the case of a "white" input signal:

$$(X_j^T X_j)^{-1} \approx \frac{1}{\ell^* \psi_{xx}(0)} \; I \tag{51}$$

This yields for the algorithm:

$$\underline{\beta}_j = \underline{\beta}_{j-1} + g_j (\ell^* \psi_{xx}(0))^{-1} X_j^T \underline{e}_j \tag{52}$$

Equation (52) is an interesting result as the total of four operations per iteration (matrix inversion, matrix-matrix multiplication and two matrix-vector multiplications) is reduced to one single operation (matrix-vector product) only.

Taking the expectation of $\underline{\beta}_j$ yields

$$E\left[\underline{\beta}_j\right] = g_j \; (\ell^* \psi_{xx}(0))^{-1} \; X_j^T X_j \underline{h} + \{I - g_j \; (\ell^* \psi_{xx}(0))^{-1} \; X_j^T X_j\} \; E\left[\underline{\beta}_{j-1}\right] \tag{53}$$

The input signal in the jth iteration influences $E\left[\underline{\beta}_j\right]$: the adjustment is biased. For sufficiently large ℓ^* equation (53) leads to

$$E\left[\underline{\beta}_j\right] = g_j \; \underline{h} + (1 - g_j) \; E\left[\underline{\beta}_{j-1}\right] \tag{54}$$

We can take the expectation of eq. (53) with respect to the input signals in all preceding intervals

$$E_{X_1} E_{X_2} \ldots E_{X_j} \; E\left[\underline{\beta}_j\right] = \underline{h} \sum_{i=1}^{j} g_i \prod_{k=i+1}^{j} (1 - g_k) + \underline{\beta}_0 \prod_{i=1}^{j} (1 - g_i) \tag{55}$$

The expression for the covariance is given by

$$\text{cov} \; \underline{\beta}_j = E\left[(\underline{\beta}_j - E\left[\underline{\beta}_j\right]) \; (\underline{\beta}_j - E\left[\underline{\beta}_j\right])^T\right] =$$

$$= \{I - g_j \; (\ell^* \psi_{xx}(0))^{-1} \; X_j^T X_j\} \; \text{cov} \; \underline{\beta}_{j-1} \{I - g_j \; (\ell^* \psi_{xx}(0))^{-1} \; X_j^T X_j\}$$

$$+ g_j^2 \; (\ell^* \psi_{xx}(0))^{-2} \; \psi_{nn}(0) \; X_j^T X_j \tag{56}$$

In order to get some insight, this expression is instrumented. The results are briefly summarized as follows:

1) $g_j = 1/j$

In fig. 4 $\text{cov} \; \underline{\beta}_j$ is plotted against $j\ell^*$.

For large $j\ell^*$ this diagram shows that $\text{cov} \; \underline{\beta}_j$ has reached approximately the same value as in the adjustment with matrix inversion.

Only for small $j\ell^*$ a deviation from the line

$$\text{cov} \; \underline{\beta}_j = \frac{1}{j\ell^*} \; \frac{\psi_{nn}(0)}{\psi_{xx}(0)} \; I$$

can be observed.

2) $g_j = \text{constant}$

For relatively large ℓ^* (56) can be approximated by

$$\text{cov} \; \underline{\beta}_j \approx (1-c)^2 \; \text{cov} \; \underline{\beta}_{j-1} + c \; \frac{\psi_{nn}(0)}{\ell^* \psi_{xx}(0)} \; I \tag{57}$$

- 13 -

yielding as asymptotic solution the same expression as in the case
of adjustment with matrix inversion:

$$\text{cov } \underline{\beta}_\infty = \frac{c}{2-c} \frac{\psi_{nn}(0)}{\ell^* \psi_{xx}(0)} \tag{58}$$

3. Some remarks on other types of errors in a practical situation.

Considering the noise error we assumed that certain conditions were ful-
filled. This, however, will mostly not be the case. It is therefore useful
to indicate some additional errors that occur when these conditions are
not satisfied.

Truncation error.

The delay line in our mathematical model will often be too small, which
then causes the phenomenon known as truncation error. The existence of
this kind of error may be considered in an analogous way as was done with
the noise error, namely as if it was caused by additive correlated noise
added to the output of the process, cf. fig. 5.
The explicit expression of the truncation error vector $\Delta\beta$ is

$$\underline{\Delta\beta} = \underline{\tilde{\beta}} - \underline{h}^* = (U^T U)^{-1} U^T (\underline{y} - \underline{\gamma}) = (U^T U)^{-1} U^T \underline{r} \tag{59}$$

U is a partial matrix of X satisfying

$$\begin{aligned} X\underline{h} &= \underline{y} \\ X\underline{h}^* &= \underline{\gamma} \end{aligned} \tag{60}$$

The truncation error does not affect the unbiasedness of the estimation
when the input signal is "white" noise, as appears from:

$$(U^T U)\underline{\tilde{\beta}} = U^T \underline{y} = U^T \underline{\gamma} + U^T \underline{r} \tag{61}$$

$$E\left[U^T U\right] E\left[\underline{\tilde{\beta}}\right] = E\left[U^T \underline{\gamma}\right] + E\left[U^T \underline{r}\right] =$$

$$= E\left[U^T U\right]\underline{h}^* E\left[U^T \underline{r}\right] \tag{62}$$

$U^T \underline{r}$ appears to be:

$$U^T \underline{r} = \begin{bmatrix} \displaystyle\sum_{v=1}^{k-p} \left\{ h_{p+v} \sum_{i=1}^{\ell-p-v+1} x_i x_{i+p+v-1} \right\} \\[2ex] \displaystyle\sum_{v=1}^{k-p} \left\{ h_{p+v} \sum_{i=1}^{\ell-p-v+2} x_i x_{i+p+v-2} \right\} \\[2ex] \vdots \\[2ex] \displaystyle\sum_{v=1}^{k-p} \left\{ h_{p+v} \sum_{i=1}^{\ell-v} x_i x_{i+v} \right\} \end{bmatrix} \tag{63}$$

If the input is "white" noise it is easy to see that

$$E\left[U^T \underline{r} \right] = \underline{0}$$

so $\quad E\left[\underline{\tilde{\beta}} \right] = \underline{h}^*$

Hence, also $\underline{\Delta\beta} \to 0$ for $\ell \to \infty$.

In contrast with the noise error, the truncation error is independent of the power of the test signal.

Even in processes with "infinite memory" the truncation error need not cause predominantly bad estimates, because always in an actual estimation procedure only a finite number $(\ell+p-1)$ of values of the impulse response is calculated with. It is even possible to estimate without errors by joining an integrator to the end of the model's delay line, which is assumed to be sufficiently large.

In some publications, cf. [3], the integrator is placed in front of the delay line but this implies that the estimates will have to be provided with a correction factor afterwards.

In the case of estimating a process of "infinite memory" without the use of an integrator the estimates, through unbiased, will even for $\ell \to \infty$ remain uncertain with a finite variance. This is caused by the fact that the assumptions required for ergodicity do not match. The variance is equal in both cases of instrumentation with or without matrix inversion.

It is reasuring that in a practical situation the truncation errors in the several estimated parameters for large ℓ are as good as equal in value and polarity, since the components in the error vector $\underline{\Delta\beta}$ are strongly correlated. Hence we may conclude that in a single estimation we indeed get a rather good idea of the shape of the impulse response, although it is uncertain to what extent every estimated value deviates from the

corresponding value of the impulse response. So every estimated value
can be corrected, if a priori knowledge about at least one value of the
impulse response is available, e.g. when the amplification factor is
known.

Observation error.

Observation errors appear if the assumption b) turns out to be unjusti-
fied. We can avoid this kind of errors in our estimates by neglecting
the first and last p measured output values. However, when it is only
possible to observe for a very short time, all output values have to be
used in the calculations and the errors have to be weighted in a proper
way, cf. [5] .

4. Conclusions.

The described explicit and implicit methods appear to yield equal results
with respect to the uncertainty in the estimate caused by additive noise.
Likewise, this uncertainty (noise error) appears to be independent of
instrumentation with or without matrix inversion. This offers the experimenter
considerable freedom to select a method satisfying his needs and being as
simple as possible.
Fundamentally, the instrumentation without matrix inversion causes some
uncertainty due to truncation errors. In a practical situation, however,
for sufficiently large ℓ the truncation error does not predominantly
influence the total uncertainty in the estimate. Besides, such an in-
fluence can often be recognized easily.

Literature.

[1]. P. Eykhoff : "Process Parameter and State Estimation", invited survey
         paper for the IFAC Symposion, Prague, June 1967;
         Automatica, vol. 4 (1968), p.p. 205-233.

[2]. P. Eykhoff : "Some Fundamental Aspects of Process-Parameter Estimation",
         IEEE Trans. on Automatic Control, October 1963, p.p. 347-357.

[3]. M.J. Levin : "Estimation of the Characteristics of Linear Systems in the
         Presence of Noise", Ph.D. Thesis, Dpt. of El. Eng., Columbia
         University, April 1959.

[4]. Ya.Z. Tsypkin : "Adaptation, Training and Self-Organisation in Automatic
         Systems", Automation and Remote Control, vol. 27 (1966),
         no. 1, p.p. 16-51.

[5]. N.R. Draper and H. Smith: "Applied Regression Analysis", John Wiley and
         Sons Inc., New York, 1966.

[6]. R. Deutsch : "Estimation Theory", Prentice Hall Inc., Englewood Cliffs,
         N.J., 1965.

[7]. E. Blandhol : "On the Use of Adjustable Models for Determination of
         System Dynamics", Division of Automatic Control, Technical
         University of Norway, Trondheim, Norway, March 1962.

[8]. J.H.A.M. Melis : "The Explicit Estimation of Process Parameters Using
         a Digital Computer", M.Sc. Thesis, Technological University
         Eindhoven, Netherlands, June 1967 (in Dutch).

[9]. A.J.W. van den Boom : "The Implicit Estimation of Process Parameters
         Using a Digital Computer", M.Sc. Thesis, Technological
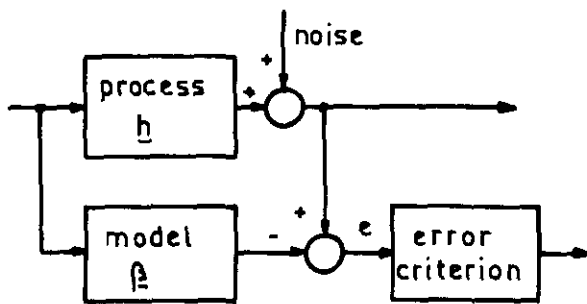   University Eindhoven, Netherlands, June 1967 (in Dutch).
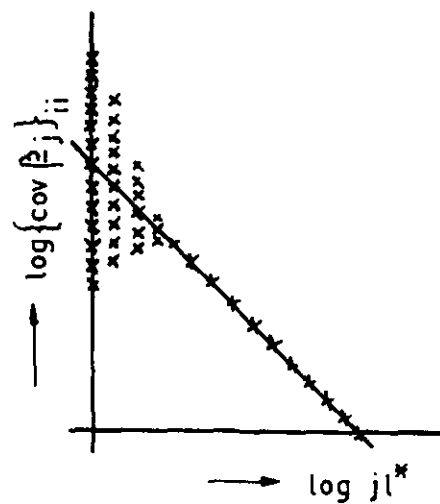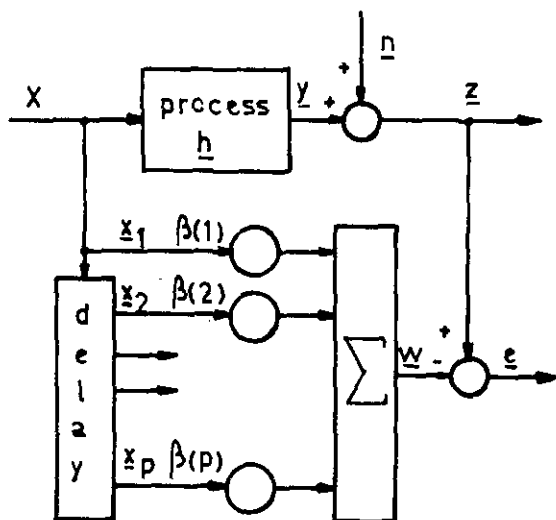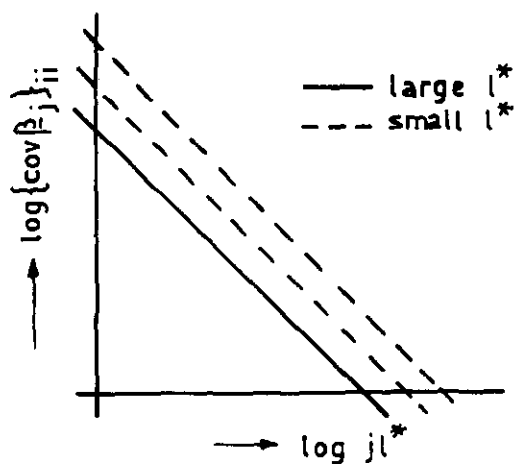
Fig. 1.

Fig. 4.

Fig. 2.

Fig. 3.
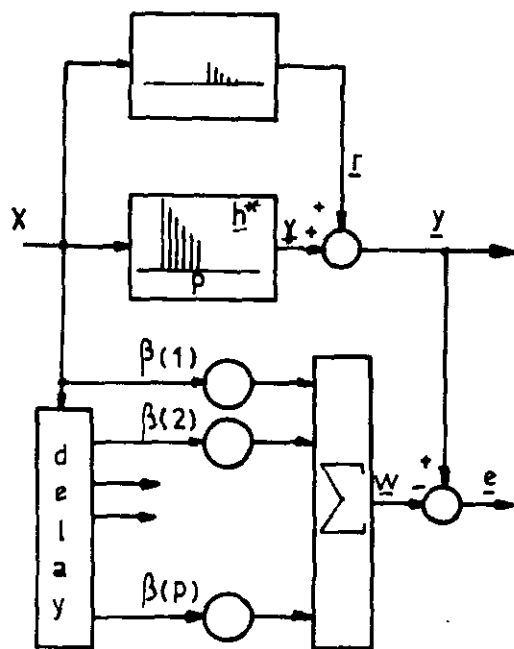
Fig. 5.

List of symbols.

E	expectation
I	identity matrix
Q	matrix
R	criterion matrix
R_j	criterion matrix in the jth iteration
U	partial matrix of X
X	input matrix
X_j	input matrix in the jth iteration
a	integer
c	constant
$\underline{e}$	error vector
$\underline{e}_j$	error vector in the jth iteration
g_j	gain in the jth iteration
$\underline{h}$	parameter vector of the process
$\underline{h}^*$	parameter vector of the truncated process
i	index
j	index
ℓ	total length of the observation sequence
ℓ^*	interval length
$\underline{n}$	additive noise
$\underline{n}_j$	additive noise in the jth iteration
p	dimension of process and model parameter vector
$\underline{r}$	output of the remnant process
v	index
$\underline{w}$	model output signal
$\underline{x}_i$	input signal at i^{th} tap of the delay line
$\underline{y}$	output signal
$\underline{y}_j$	output signal in the jth iteration
$\underline{z}$	by noise corrupted output signal
$\underline{z}_j$	corrupted output signal in the jth iteration
Φ	weighting matrix containing a priori knowledge
Ψ_{X_j}	autocorrelation matrix of the input signal in the jth iteration
$\tilde{\Psi}_{X_j}$	approximated autocorrelation matrix of the inputsignal in the jth iteration

$\underline{\beta}$	parameter vector of the model
$\underline{\beta}_j$	parameter vector of the model after the jth iteration
$\underline{\Delta\beta}$	noise error
$\underline{Y}$	output of the truncated process
σ	standard deviation
$\psi_{xx}(i)$	autocorrelation function
$\psi_{xy}(i)$	cross-correlation function
$\tilde{\psi}_{xx}(i,\ell)$	approximated autocorrelation function
$\tilde{\psi}_{xy}(i,\ell)$	approximated cross-correlation function
Δ	difference operator
∇	gradient operator
π	continued product
Σ	sum