Some aspects of system identification : order selection rules, a multivariable parameter estimation technique

van Beek, W.J.

Award date:
1985

5127 1

DEPARTMENT OF ELECTRICAL ENGINEERING
EINDHOVEN UNIVERSITY OF TECHNOLOGY
Group Measurement and Control

SOME ASPECTS OF SYSTEM IDENTIFICATION:

- ORDER SELECTION RULES
- A MULTIVARIABLE PARAMETER ESTIMATION TECHNIQUE

by W.J. van Beek

This report is submitted in partial fulfillment of the requirements for the degree of electrical engineer (M.Sc.) at the Eindhoven University of Technology.
The work was carried out from sept. 1984 until aug. 1985
in charge of Prof.Dr.Ir. P. Eykhoff
under supervision of Prof.Dr.Ir. P. Eykhoff
and Ir. A.A. van Rede

ABSTRACT

In this report the structural and parametrical identification is treated of stable, linear, time-invariant and time-discrete systems. The contents of this report can be divided into two different subjects:

The first part of this report treats the comparison of several order selection criteria for univariable systems.
A new type of criterion based on the concept of cross-validation has been compared with the existing criteria proposed by Akaike.
Making use of available software from the SATER program package of the group Measurement and Control the order selection rules have been implemented on a PDP 11/60 minicomputer.
For a number of simulated systems the performances of the different criteria have been compared.

The second part of this report treats the parametrical identification of multivariable systems.
A different concept for parameter estimation of multiple-input single-output systems has been developed.
This concept consists of an alternating iterative type of estimation making use of an existing univariable parameter estimation algorithm, minimizing the output error.
According to this concept a program has been developed on a VAX 11/750 computer.
The behaviour of this method has been tested extensively with simulated data and eventually the method has been applied to a set of practical data recorded in the field.

SAMENVATTING

Dit verslag behandelt de structuur- en parameterschatting van stabiele, lineaire, tijd-invariante en tijd-discrete systemen. De inhoud van dit verslag is te verdelen in twee verschillende onderwerpen:

Het eerste deel van dit verslag gaat over de vergelijking van enkele orde selectie criteria voor univariable systemen.
Een nieuw criterium dat gebaseerd is op het principe van de kruis-validatie is vergeleken met de bestaande orde selectie criteria van Akaike.
Met behulp van bestaande programmatuur afkomstig uit het programmapakket SATER, ontwikkeld in de vakgroep Meten en Regelen, zijn de orde selectie criteria geïmplementeerd op een PDP 11/60 minicomputer.
Voor een aantal verschillende gesimuleerde systemen zijn de resultaten van de orde selectie criteria met elkaar vergeleken.

Het tweede deel van dit verslag gaat over de parameter identificatie van multivariable systemen.
Een nieuw, afwijkend concept is ontwikkeld voor de parameterschatting van systemen met meerdere ingangen en één uitgang.
Dit concept behelst een alternerende iteratieve schatting met behulp van een bestaand algorithme voor de parameterschatting van univariabele systemen, dat de 'output error' minimaliseert.
Volgens dit concept is een programma ontwikkeld op een VAX 11/750 computer.
Het gedrag van deze methode is uitvoerig getest met behulp van gesimuleerde data en uiteindelijk is het programma toegepast op praktische data.

TABLE OF CONTENTS

1. Introduction

During the last two decades an increasing interest has arisen in the field of identification of systems.
This interest originates from various kinds of problems, not only directly related to the purely technical issue of controlling a system, but also related to the problem of measuring certain quantities in a system.
This problem reaches beyond the physical aspects of obtaining measurement data of a system; it concerns the extraction of knowledge about certain quantities that are hidden in a sequence of the acquired measurement data.
Thus it is not surprising that the interest also comes from other than purely technical scientific disciplines, such as economy, biology and medicine.
The development of computer technology has contributed considerably to the growing maturity of identification techniques, since the numerical problems, that identification of systems brings with it, could never be solved without the support of modern computers.
The aim of system identification is to build a model of a particular process, that satisfies certain selected requirements concerning the resemblance of that model with the system. These requirements normally comprise the minimization of certain quantities, dictated by the specific identification method that is being used.
The process of identification can be described in several steps:

- -choice of an appropriate model set
- -structural identification
- -parametrical identification
- -assessment of the obtained model

Fig. 1.1 illustrates the identification of a process in different steps, given the input and output data of the process.
Clearly the different steps of the identification process are closely related. The rather subjective choice of a model set will be decisive for the structural identification of a system. Similarly the structure that is chosen is essential for the objective parameter estimation.

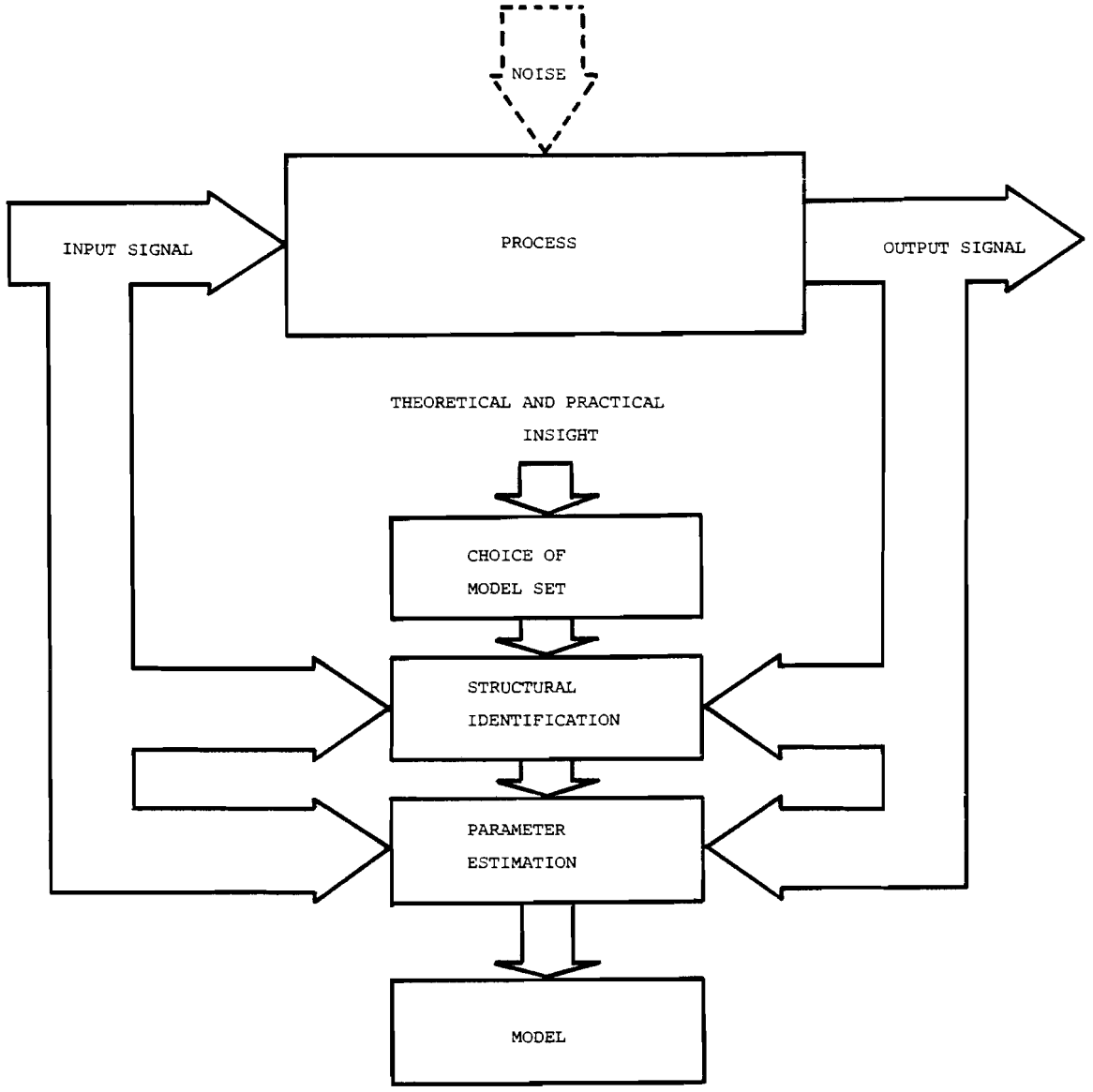


Fig. 1.1 The identification of a process divided in different steps

On purpose the structural identification is not mentioned using the adjective 'subjective' or 'objective', since this depends on the specific identification technique that is being used.

It is clear that an objective technique is preferable to a subjective one in any application. Note, however, that also the choice itself of the identification method, both structural and parametrical, still contains a subjective aspect.

Stoica (1984) gives a comprehensive review of literature, including many existing model structure selection methods in a relatively small number

of classes. Among these classes a division can be made between subjective and objective methods. Söderström (1977) analytically compares some of these methods.
Part One of this report concerns the structural identification of single-input single-output (SISO) systems.
In Stoica (1984) a new objective model structure selection rule is introduced for univariable systems, based on the concept of cross-validation.
During the first part of my project I have implemented this cross-validatory method and I have compared its performance with the methods derived by Akaike (1974,1977) and Hannan and Quinn (1979).
Chapter 2 treats the cross-validatory principle, the structure test criteria and their comparison. In chapter 3 some programming aspects are discussed of the interactive program SATER, the basis of which is used for the implementation and comparison of the criteria.

In system identification a division can be made between univariable and multivariable systems (more frequently multivariable is called multiple-input multiple-output or MIMO). The identification methods for the SISO case have reached a certain degree of maturity in the last two decades, whereas the MIMO case only has received attention more recently. El-Sherief (1982) gives a clear introduction to multivariable identification methods.
During the second part of my project I have been working on a different approach to the MIMO parametrical identification. This approach has lead to an estimation method based on the alternating iterative estimation of two SISO subsystems of a MIMO system.
Part Two of this report concerns the parametrical identification of multiple-input single-output systems. Chapter 4 describes the development of the estimation method. In chapter 5 the simulation results will be discussed to evaluate the performance of the method. Chapter 6 will treat the application of the method to data originating from Philips Glass.

PART ONE

2. Comparison of SISO order test criteria

2.1 Introduction

The field of order testing is still quite an underexplored area; most order determination procedures still need a human interpretation of the result to lead to an order choice. The need for independently operating (objective) decision rules for order determination therefore can not be denied.
Several attempts have been made to arrive at such a criterion.
An early attempt was made by applying the F-test. This order test makes use of the statistical properties (a χ^2-distribution) of the loss-functions, leading to an F-distribution of the criterion value. By imposing a significance level to the criterion, an order can be indicated (cf. Åström and Eykhoff, 1971). The principle of regarding the determination of the order as a decision problem is quite common in regression analysis.
Akaike (1974) has proposed a criterion based on information theory for selecting the best order (AIC). Further methods based on singularity of the information matrix and methods based on pole-zero cancellation have been designed (cf. Söderström, 1975).
It is shown that the F-test and AIC are equivalent (cf. Söderström, 1977). A clear review on order testing methods is given by Van den Boom (1982).
In the following, Akaike's Information Criterion (AIC) and the Generalized Akaike's Information Criterion (GAIC) will be compared with a new criterion, introduced by Stoica (1984), based on the principle of cross-validation.

2.2 A cross-validation order test criterion

2.2.1 The cross-validatory principle

The concept of cross-validation is an old one; Stone (1974) gives an overview in literature of the principle that goes back as far as 1931,

but surely it has even been applied before that time.
A quotation from Stone's article is probably the best way to describe the fundamentals of the concept:

> "In its most primitive but nevertheless useful form, it [cross-validation] consists in the controlled or uncontrolled division of the data sample into two subsamples, the choice of a statistical predictor, including any necessary estimation, on one subsample and then the assessment of its performance by measuring its predictions against the other subsample."

Further Stone remarks that, in this context, the use of the term 'validation' has a connotation of excessive confidence about it and therefore he prefers to call it 'assessment'.
More specifically, the cross-validation concept in its simplest form comprises the following.
Let S be a data sequence of N measurements (u,y) of a univariate process P. Thus we may write

$$S = \{ (u_i, y_i) \mid i=1,\ldots,N \} \tag{2.1}$$

Then split the sequence S into (equal) parts S_1 and S_2; this gives

$$S_1 = \{ (u_i, y_i) \mid i=1,\ldots,j \} \tag{2.2-a}$$

and

$$S_2 = \{ (u_i, y_i) \mid i=j+1,\ldots,N \} \tag{2.2-b}$$

with $j < N$.
Consider the issue for which only the u-value is available and the value of y has to be estimated by means of this data sequence. Assuming a model structure $M_m(\underline{\theta})$ of $\{ M_k(\underline{\theta}) \mid \dim \underline{\theta} = k, k \in N \}$ for the process P, leads to a prescription of the estimates $\hat{y}_i$ of y_i:

$$\{ \hat{y}_i(u;\underline{\theta}) \mid \underline{\theta} \in M_m(\underline{\theta}) \} \tag{2.3}$$

Now we can define a loss-function to describe the performance of the estimate $\hat{y}$:

$$V(\underline{\theta}) \equiv \frac{1}{m} \sum_{i=1}^{m} f[y_i, \hat{y}_i(u;\underline{\theta})] \tag{2.4}$$

where $f[y,\hat{y}]$ is some selected loss-function of $\hat{y}$ as an estimate of the actual value y.
Organizing the estimation over subsequence S_1 in such a way that the loss-function

$$V_1(\underline{\theta}) = \frac{1}{j} \sum_{i=1}^{j} f[y_i, \hat{y}_i(u;\underline{\theta})] \tag{2.5}$$

is minimized, leads to the parameter estimate $\hat{\underline{\theta}}_1$, with $\hat{\underline{\theta}}_1 \varepsilon M_m(\underline{\theta})$.
Assessment of this choice of the estimation by measuring the loss-function

$$V_2(\hat{\underline{\theta}}_1) = \frac{1}{N-j} \sum_{i=j+1}^{N} f[y_i, y_i(u, \hat{\underline{\theta}}_1)] \tag{2.6}$$

is called cross-validation.
Taking different model structures by varying m, the dimension of the parameter vector, (read: varying the model order) leads to a specific curve of $V_2(\hat{\underline{\theta}}_1)$ as a function of m; the minimum value of V_2 will indicate the true model dimension.
This can be explained by the following.
In case of overparametrization (m too large), the dynamics of the noise present in the process will be partly modelled as part of the process. Since the estimation is performed over S_1, minimizing V_1, the oversized parameter estimate $\hat{\underline{\theta}}_1$ will be oriented on the specific dynamics of the noise sequence during S_1. Due to the (to a certain extent) unpredictable behaviour of the noise, whether it be 'white' or 'coloured', the dynamics of the noise sequence during S_2 will differ. Therefore the value of $V_2(\hat{\underline{\theta}}_1)$ will not be a minimum realization of $V_2(\underline{\theta})$.

Increasing the dimension of $\underline{\theta}$, will lead to an increasingly better capture of the dynamics of the noise sequence during S_1, thus leading to a worse 'capture' of the specific noise dynamics during S_2.
Fig. 2.1 illustrates this behaviour.

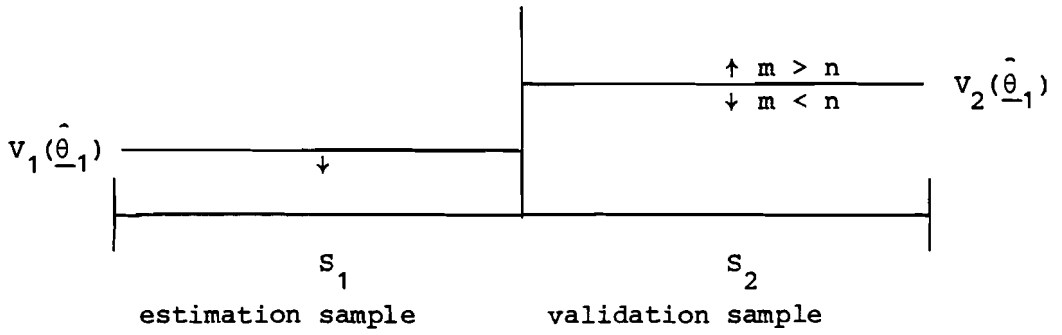


Fig. 2.1 The loss-function of the estimation sample and the validation sample; the arrows indicate the direction of the evolution of $V_1(\hat{\underline{\theta}}_1)$ and $V_2(\hat{\underline{\theta}}_1)$ for increasing model dimension, n is the 'true' dimension of the process.

To make better use of the available data, this principle can also be applied mutually to the data subsequences. This gives rise to the criterion

$$C \equiv \frac{N-j}{N} \cdot V_2(\hat{\underline{\theta}}_1) + \frac{j}{N} \cdot V_1(\hat{\underline{\theta}}_2) \tag{2.7}$$

with $\hat{\underline{\theta}}_2$ being the result of minimizing

$$V_2(\underline{\theta}) = \frac{1}{N-j} \sum_{i=j+1}^{N} f[y_i, \hat{y}_i(u;\underline{\theta})] \tag{2.8}$$

This is the basis of the criteria described in subsection 2.2.3, where division of S can occur also for any other value than two subsequences (limited of course by the number N of data samples).

2.2.2 The estimation method

For the parameter estimation the least squares method is implemented, based on an autoregressive model of the process; cf. fig. 2.2.

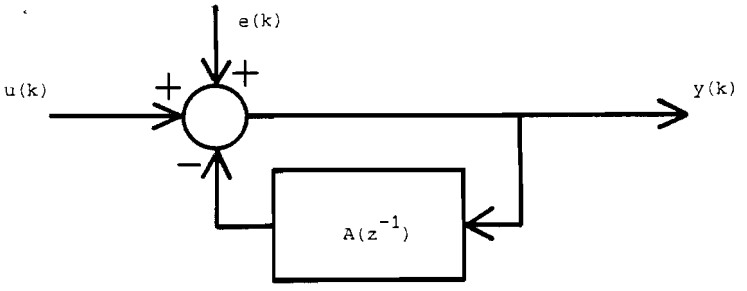


Fig. 2.2 Autoregressive model of a process with input u(k) and disturbance e(k)

Then the model description can be denoted as:

$$y(k) + a_1 y(k-1) + \ldots + a_n y(k-n) = u(k) + e(k) \qquad (2.9\text{-a})$$

or in vector notation:

$$y(k) = \begin{bmatrix} u(k) & y(k-1) & \ldots & y(k-n) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -a_1 \\ \vdots \\ -a_n \end{bmatrix} + e(k) \qquad (2.9\text{-b})$$

With

$$\underline{\phi}^T(k) = \begin{bmatrix} u(k) & y(k-1) & \ldots & y(k-n) \end{bmatrix} \qquad (2.9\text{-c})$$

and

$$\underline{\theta}^T = \begin{bmatrix} 1 & -a_1 & \ldots & -a_n \end{bmatrix} \qquad (2.9\text{-d})$$

this becomes:

$$y(k) = \underline{\phi}^T(k) \cdot \underline{\theta} + e(k) \qquad (2.9\text{-e})$$

An estimate of the parameter vector $\underline{\theta}$ is $\hat{\underline{\theta}}$, and the estimated output can

be denoted as:

$$\hat{y}(k) = \underline{\phi}^T(k).\hat{\underline{\theta}} \tag{2.10}$$

The equation error can be taken as:

$$\hat{e}(k) = y(k)-\hat{y}(k) = y(k)-\underline{\phi}^T(k).\hat{\underline{\theta}} \tag{2.11}$$

If we define the quadratic loss-function as:

$$V(\underline{\theta}) = \frac{1}{N-n} \sum_{k=n+1}^{N} \hat{e}(k)^2 \tag{2.12}$$

with N the number of samples and n the supposed order of the process, an estimate of $\underline{\theta}$ can be determined by minimizing V as a function of $\underline{\theta}$. With $\frac{\partial V(\theta)}{\partial \underline{\theta}} \Big|_{\hat{\underline{\theta}}} = \underline{0}$ and equation (2.11) we obtain:

$$\frac{-2}{N-n} \sum_{k=n+1}^{N} \underline{\phi}(k)\left[y(k)-\underline{\phi}^T(k)\underline{\theta}\right] \Big|_{\hat{\underline{\theta}}} = \underline{0} \tag{2.13}$$

From equation (2.13) the optimal parameter estimate $\hat{\underline{\theta}}$ (optimal in the sense of a minimal quadratic equation error) follows and it can be denoted as:

$$\hat{\underline{\theta}} = \left[\frac{1}{N-n} \sum_{k=n+1}^{N} \underline{\phi}(k)\underline{\phi}^T(k) \right]^{-1} .\left[\frac{1}{N-n} \sum_{k=n+1}^{N} \underline{\phi}(k)y(k) \right] \tag{2.14}$$

The first term on the right hand side of equation (2.14) can be written as follows:

$$\frac{1}{N-n} \sum_{k=n+1}^{N} \underline{\phi}(k)\underline{\phi}^T(k) =$$

$$= \frac{1}{N-n} \sum_{k=n+1}^{N} \begin{bmatrix} u^2(k) & u(k)y(k-1) & \cdots & u(k)y(k-n) \\ u(k)y(k-1) & y^2(k-1) & & \vdots \\ \vdots & & \ddots & \vdots \\ u(k)y(k-n) & \cdots & \cdots & y^2(k-n) \end{bmatrix} =$$

$$= \begin{bmatrix} \hat{\phi}_{uu}(0) & \cdots & \hat{\phi}_{uy}(n) \\ \vdots & \hat{\phi}_{yy}(0)' & \vdots \\ \hat{\phi}_{uy}(n) & \hat{\phi}_{yy}(n-1) \cdots & \hat{\phi}_{yy}(0)'' \end{bmatrix} \qquad (2.15)$$

Note that the diagonal terms $\hat{\phi}_{yy}$ all differ, since they are all derived with slightly different data.

The second term on the right hand side of equation (2.14) can be denoted as:

$$\frac{1}{N-n} \sum_{k=n+1}^{N} \underline{\phi}(k)y(k) = \frac{1}{N-n} \sum_{k=n+1}^{N} \begin{bmatrix} u(k) \\ y(k-1)y(k) \\ \vdots \\ y(k-n)y(k) \end{bmatrix} = \begin{bmatrix} \hat{\phi}_{uy}(0) \\ \hat{\phi}_{yy}(1) \\ \vdots \\ \hat{\phi}_{yy}(n) \end{bmatrix} \qquad (2.16)$$

From equations (2.14), (2.15) and (2.16) we obtain the expression:

$$\hat{\underline{\theta}} = \begin{bmatrix} \hat{\phi}_{uu}(0) & \cdots & \hat{\phi}_{uy}(n) \\ \vdots & \hat{\phi}_{yy}(0)' & \vdots \\ \hat{\phi}_{uy}(n) & \hat{\phi}_{yy}(n-1) \cdots & \hat{\phi}_{yy}(0)'' \end{bmatrix} \cdot \begin{bmatrix} \hat{\phi}_{uy}(0) \\ \hat{\phi}_{yy}(1) \\ \vdots \\ \hat{\phi}_{yy}(n) \end{bmatrix} \qquad (2.17)$$

This equation shows that the estimation method can be used both for the input-output situation and for the time series case.

The use for the input-output case is straightforward (cf. fig. 2.3). For the time series case the input u has to be discarded in equation (2.17) or can be taken as zero. In that case the matrix from equation (2.17) becomes singular because of its zero first row and column, and can not be

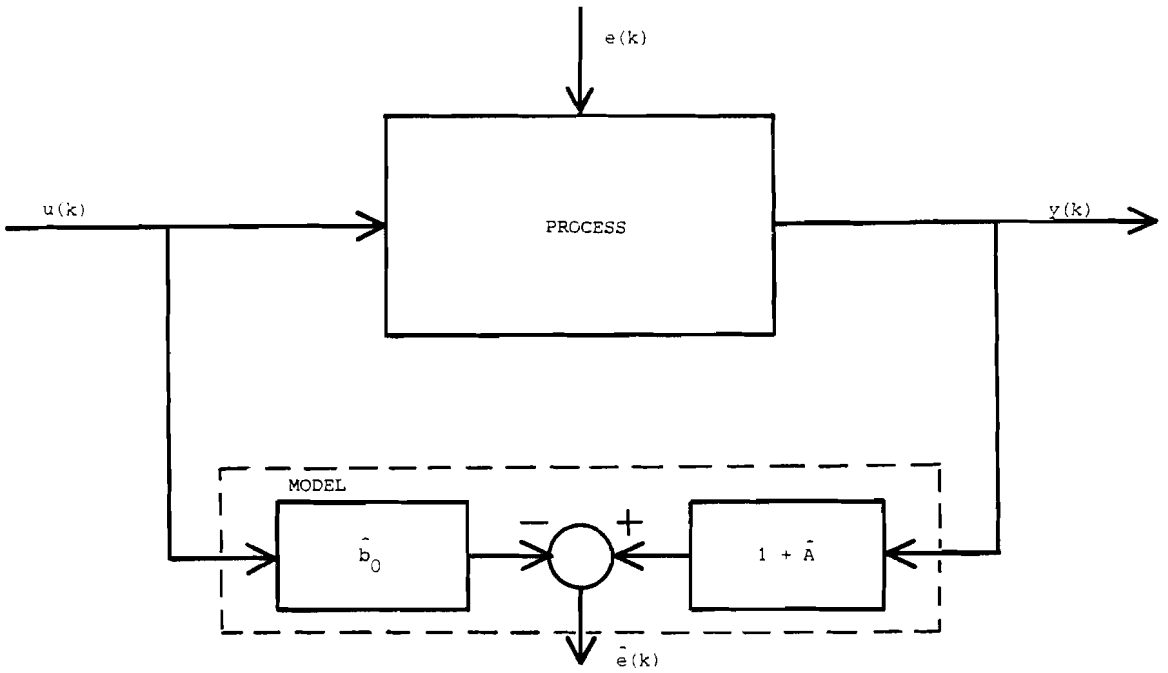


Fig. 2.3 Process with u(k) as input, e(k) as disturbance and an autoregressive model for estimation.

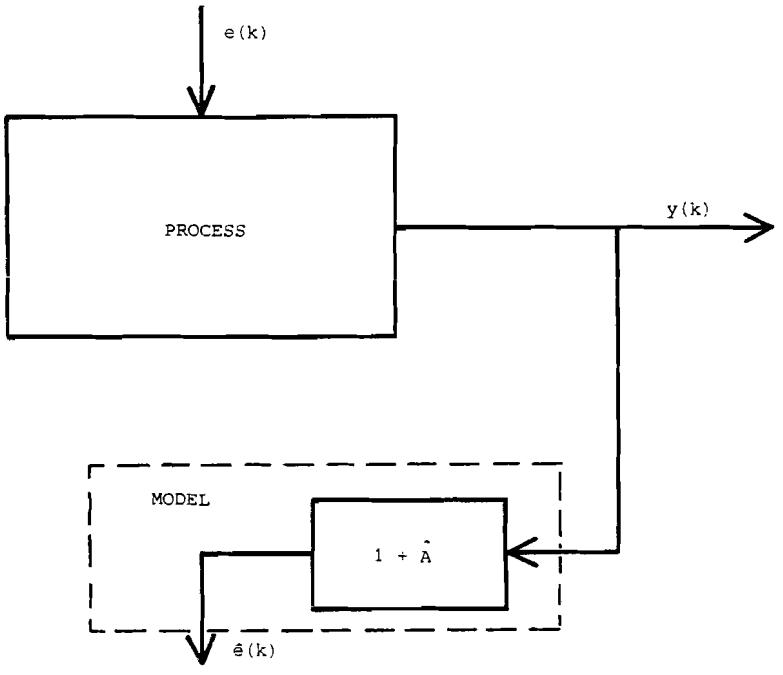


Fig. 2.4 Time series case with an autoregressive model.

inverted. To avoid this we leave out the first row and column.
Thus we obtain near-Yule-Walker equations (in matrix form) for time series estimation with an autoregressive model. The matrix only deviates in the sense that the diagonal terms slightly differ from each other.
We obtain the following explicit expression for the parameter estimate $\hat{\underline{\theta}}$ (except its first element, which has no meaning in the time series case), denoted as:

$$\hat{\underline{\theta}}' = \begin{bmatrix} \hat{\phi}_{yy}(0)' & \cdots & \hat{\phi}_{yy}(n-1) \\ \vdots & & \vdots \\ \hat{\phi}_{yy}(n-1) & \cdots & \hat{\phi}_{yy}(0)'' \end{bmatrix}^{-1} \cdot \begin{bmatrix} \hat{\phi}_{yy}(1) \\ \vdots \\ \hat{\phi}_{yy}(n) \end{bmatrix} \qquad (2.18)$$

Now $\hat{\underline{\theta}}'$ is the vector $\hat{\underline{\theta}}$ without its first element. In the algorithm this step is simplified by only inverting the submatrix (obtained by removing the first row and column) in equation (2.17). Fig. 2.4 illustrates the estimation in the time series case.

2.2.3 The order test method

Stoica (1984) proposes in his paper two criteria to select the model structure of a process. These criteria are based on the principle of cross-validatory assessment, as discussed previously in subsection 2.2.1.
The first criterion comprises the following:

$$C_I \equiv \sum_{p=1}^{k} \sum_{t \varepsilon I_p} \hat{e}^2(t, \hat{\underline{\theta}}_p) \qquad (2.19)$$

where

$$\hat{\underline{\theta}}_p \equiv \arg \min_{\theta \varepsilon \Theta} \sum_{t \varepsilon I - I_p} e^2(t, \underline{\theta}) \qquad (2.20)$$

with k the number of intervals in which the data sequence is divided and

I_p is the p^{th} interval, consisting of m samples.
Stoica shows, that the criterion C_I can be approximated by:

$$C_{1m} = V(\hat{\underline{\theta}}) + \frac{4}{(N-n)^2} \sum_{p=1}^{k} \underline{w}_p^T(\hat{\underline{\theta}})\, V_{\theta\theta}^{-1}(\hat{\underline{\theta}})\, \underline{w}_p(\hat{\underline{\theta}}) =$$
$$= V(\hat{\underline{\theta}}) + \frac{4}{(N-n)^2} \operatorname{tr} V_{\theta\theta}^{-1}(\hat{\underline{\theta}})\; W(\hat{\underline{\theta}}) \qquad (2.21)$$

with

$$\underline{w}_p(\hat{\underline{\theta}}) = \sum_{t \in I_p} \hat{e}(t,\underline{\theta})\, \hat{\underline{e}}_{\underline{\theta}}(t,\underline{\theta}) \qquad \text{for } p=1,\ldots,k \qquad (2.22)$$

and

$$W(\hat{\underline{\theta}}) = \sum_{p=1}^{k} \underline{w}_p(\hat{\underline{\theta}})\, \underline{w}_p^T(\hat{\underline{\theta}}) \qquad (2.23)$$

The <u>second</u> criterion suggested by Stoica, comprises:

$$C_{II} \equiv \sum_{p=1}^{k} \sum_{t \in I - I_p} \hat{e}^2(t,\hat{\underline{\theta}}_p) \qquad (2.24)$$

where

$$\hat{\underline{\theta}}_p \equiv \arg \min_{\theta \in \Theta} \sum_{t \in I_p} \hat{e}^2(t,\underline{\theta}) \qquad (2.25)$$

Stoica approximates C_{II} with:

$$C_{2k} = V(\hat{\underline{\theta}}) + \frac{2k}{(N-n)^2} \sum_{p=1}^{k} \underline{w}_p^T(\hat{\underline{\theta}})\, V_{\theta\theta}^{-1}(\hat{\underline{\theta}})\, \underline{w}_p(\hat{\underline{\theta}})$$
$$= V(\hat{\underline{\theta}}) + \frac{2k}{(N-n)^2} \operatorname{tr} V_{\theta\theta}^{-1}(\hat{\underline{\theta}})\, W(\hat{\underline{\theta}}) \qquad (2.26)$$

$V_{\theta\theta}^{-1}(\hat{\underline{\theta}})$, the inverse Hessian matrix, is found as follows (cf. equations (2.11) and (2.12)):

$$V_{\theta\theta}^{-1}(\hat{\underline{\theta}}) = \left[\frac{2}{N-n} \sum_{k=n+1}^{N} \underline{\phi}(k)\underline{\phi}^T(k) \right]^{-1} \qquad (2.27)$$

This term has already been used in the estimation method, cf. equation (2.14), apart from a constant factor.
The vectors $\underline{w}_p(\hat{\underline{\theta}})$ are also easily determined. With

$$\left.\frac{\partial \hat{e}}{\partial \underline{\theta}}\right|_{k,\hat{\underline{\theta}}} = -\underline{\phi}(k) \qquad (2.28)$$

a result of the least squares estimation, we can denote:

$$\underline{w}_p(\hat{\underline{\theta}}) = - \sum_{t \varepsilon I_p} \hat{e}(t,\hat{\underline{\theta}})\underline{\phi}(t) \qquad (2.29)$$

The two criteria and their approximated versions, as derived by Stoica, are now briefly described. Both C_1 and C_2, the approximations, consist of the loss-function $V(\underline{\theta})$ and an additional term. Since the loss-function is supposed to be a monotonously decreasing function of the model order, these additional terms must have an interpretation of correction; the resulting sum of $V(\underline{\theta})$ and the term has a unique minimum. Since the correction terms of both criteria are of the same structure and only differ in some additional factors, an interpretation to this penalizing term will be given in the following.

Define $V_p(\underline{\theta})$ the loss-function of the interval I_p:

$$V_p(\underline{\theta}) = \frac{1}{m} \sum_{t \varepsilon I_p} \hat{e}^2(t,\underline{\theta}) \qquad (2.30)$$

with $m = N/k$.
A good measure for the deviation of $\hat{\underline{\theta}}$ from the local optimum $\hat{\underline{\theta}}_p$ for I_p is:

$$\left.\frac{\partial V_p}{\partial \underline{\theta}}\right|_{\hat{\underline{\theta}}} = \frac{2}{m} \underline{w}_p \qquad (2.31)$$

(also cf. equation (2.22)).

A Taylor expansion of (2.31) can then be developped:

$$\underline{0} = \frac{\partial V_p(\underline{\theta})}{\partial \underline{\theta}} \Big|_{\hat{\underline{\theta}}_p} = \frac{\partial V_p(\underline{\theta})}{\partial \underline{\theta}} \Big|_{\hat{\underline{\theta}}} + \frac{\partial^2 V_p(\underline{\theta})}{\partial \underline{\theta}^2} \Big|_{\hat{\underline{\theta}}} (\hat{\underline{\theta}}_p - \hat{\underline{\theta}}) + O(|\hat{\underline{\theta}}_p - \hat{\underline{\theta}}|^2) \tag{2.32}$$

Hence we can write:

$$(\hat{\underline{\theta}} - \hat{\underline{\theta}}_p) \approx \left[\frac{\partial^2 V_p(\underline{\theta})}{\partial \underline{\theta}^2} \Big|_{\hat{\underline{\theta}}} \right]^{-1} \cdot \frac{\partial V_p(\underline{\theta})}{\partial \underline{\theta}} \Big|_{\hat{\underline{\theta}}} \tag{2.33}$$

Now define a norm for the deviation of the parameter vector as follows:

$$\| \hat{\underline{\theta}} - \hat{\underline{\theta}}_p \|^2 = (\hat{\underline{\theta}} - \hat{\underline{\theta}}_p)^T W (\hat{\underline{\theta}} - \hat{\underline{\theta}}_p) \tag{2.34}$$

with W a weighting matrix, since the Euclidean norm does not take into account the differing sensitivities of the different parameters. The choice of the covariance matrix $E[(\hat{\underline{\theta}} - \underline{\theta}) (\hat{\underline{\theta}} - \underline{\theta})^T] = V_{\theta\theta}^{-1}(\underline{\theta})$ places the norm in a statistical context (cf. Bard, 1974).

Thus we yield:

$$\| \hat{\underline{\theta}} - \hat{\underline{\theta}}_p \| = (\hat{\underline{\theta}} - \hat{\underline{\theta}}_p)^T V_{\theta\theta}(\hat{\underline{\theta}}) (\hat{\underline{\theta}} - \hat{\underline{\theta}}_p) =$$

$$\approx \frac{\partial V_p(\underline{\theta})}{\partial \underline{\theta}} \Big|_{\hat{\underline{\theta}}}^T \cdot \frac{\partial^2 V_p(\underline{\theta})}{\partial \underline{\theta}^2} \Big|_{\hat{\underline{\theta}}}^{-1} \cdot \frac{\partial^2 V(\underline{\theta})}{\partial \underline{\theta}^2} \Big|_{\hat{\underline{\theta}}} \cdot \frac{\partial^2 V_p(\underline{\theta})}{\partial \underline{\theta}^2} \Big|_{\hat{\underline{\theta}}}^{-1} \cdot \frac{\partial V_p(\underline{\theta})}{\partial \underline{\theta}} \Big|_{\hat{\underline{\theta}}} \tag{2.35}$$

Taking the average value of this quadratic norm over the whole sequence yields:

$$\frac{1}{k}\sum_{p=1}^{k} \|\hat{\underline{\theta}}-\hat{\underline{\theta}}_p\|^2 =$$

$$= \frac{1}{k}\sum_{p=1}^{k} \frac{\partial V_p(\underline{\theta})}{\partial \underline{\theta}} \Bigg|_{\hat{\underline{\theta}}}^{T} \cdot \frac{\partial^2 V_p(\underline{\theta})}{\partial \underline{\theta}^2} \Bigg|_{\hat{\underline{\theta}}}^{-1} \cdot \frac{\partial^2 V(\underline{\theta})}{\partial \underline{\theta}^2} \Bigg|_{\hat{\underline{\theta}}} \cdot \frac{\partial^2 V_p(\underline{\theta})}{\partial \underline{\theta}^2} \Bigg|_{\hat{\underline{\theta}}}^{-1} \cdot \frac{\partial V_p(\underline{\theta})}{\partial \underline{\theta}} \Bigg|_{\hat{\underline{\theta}}} \tag{2.36}$$

Since the Hessian matrix is not supposed to differ much from the value over the whole sequence, we can write:

$$\frac{1}{k}\sum_{p=1}^{k} \|\hat{\underline{\theta}}-\hat{\underline{\theta}}_p\|^2 = \frac{1}{k}\sum \frac{\partial V_p(\underline{\theta})}{\partial \underline{\theta}} \Bigg|_{\hat{\underline{\theta}}}^{T} \cdot \left[\frac{\partial^2 V(\underline{\theta})}{\partial \underline{\theta}^2} \Bigg|_{\hat{\underline{\theta}}} \right]^{-1} \cdot \frac{\partial V_p(\underline{\theta})}{\partial \underline{\theta}} \Bigg|_{\hat{\underline{\theta}}} \tag{2.37}$$

Invoking equation (2.31) then yields:

$$\sum_{p=1}^{k} \underline{w}_p^T(\hat{\underline{\theta}})\, V_{\theta\theta}^{-1}(\underline{\theta})\, \underline{w}_p(\hat{\underline{\theta}}) = \frac{km^2}{4} \cdot \overline{\|\underline{\theta}-\underline{\theta}_p\|^2} \tag{2.38}$$

The interpretation of this result as a penalizing term is clear. The variance of $\hat{\underline{\theta}}_p$, described by the mean square value of $\|\hat{\underline{\theta}}-\hat{\underline{\theta}}_p\|$ is expected to increase for growing order, for each $\hat{\underline{\theta}}_p$ is optimized for a different p^{th} noise subsequence. Clearly the variance of $\hat{\underline{\theta}}_p$ will be less if the intervals are chosen larger, so the size of the intervals also influences the variance.

2.3 Design of the program

For validation purposes of the order test criteria a program has been designed, partially making use of existing subprograms, originating from the SATER program package.
In this program the structure of the SATER package has been maintained, both for practical use of the existing subprograms and versatility.
All subroutines fit in the main program called STMAIN, derived directly from SATER's main program SMAIN. In the same way use is made of the program SINIT (after SATER) to initialize the necessary datafile and variable values.

Because of the limited amount of memory of the available PDP 11/60 minicomputer the processing of data has to take place in such a way that the sequence of data does not occupy the whole core memory. Therefore transfer of data occurs in datafiles, which do not take up memory space of the computer core, but stay in peripheral memory. The process of accessing and searching datafiles slows down the performance in time of the program, but it is inevitable when using a small computer system. An advantage is that intermediate results are not temporary but can be recalled after execution of the program. Fig. 2.5 illustrates the operation of the program. For the transfer of data in the program cf. fig. 2.6.

The subprogram LSQEST computes the least squares estimate of the parameters of the current order. The subprogram ORDCRT computes the order test criteria, given the result of LSQEST. Hereafter the subprogram MINIM compares these values with the former corresponding values of one order less. The minimum is determined when the previous value is less than the current corresponding value. Note that this procedure can cause erroneous results, in case of local minima of the loss-function (as a function of the order), but this is considered of minor importance for loss-functions that have a sufficiently smooth shape.

When all the different selection rules have detected a minimum, the selected orders are stored and a next simulation can be started for comparison.

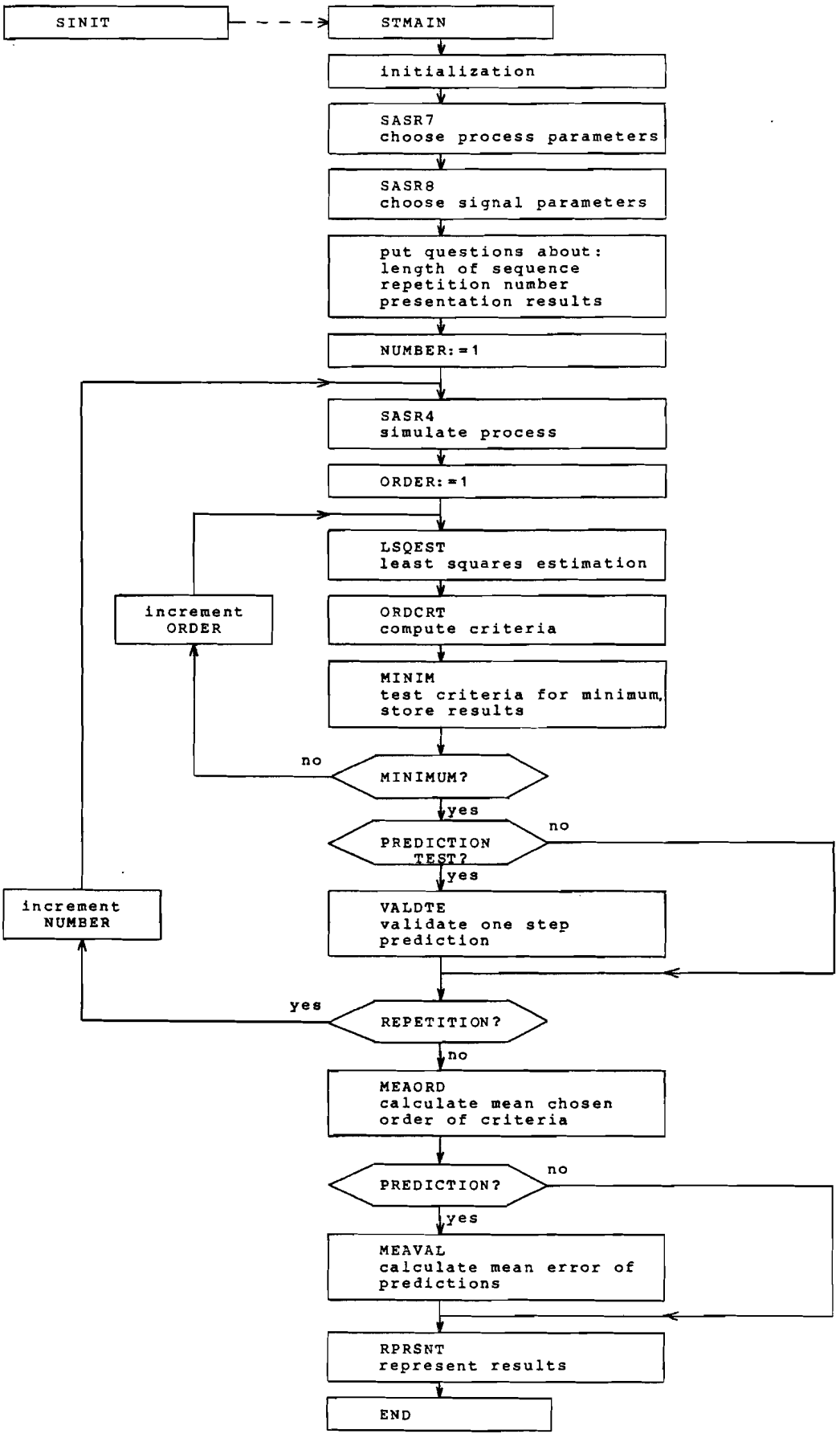
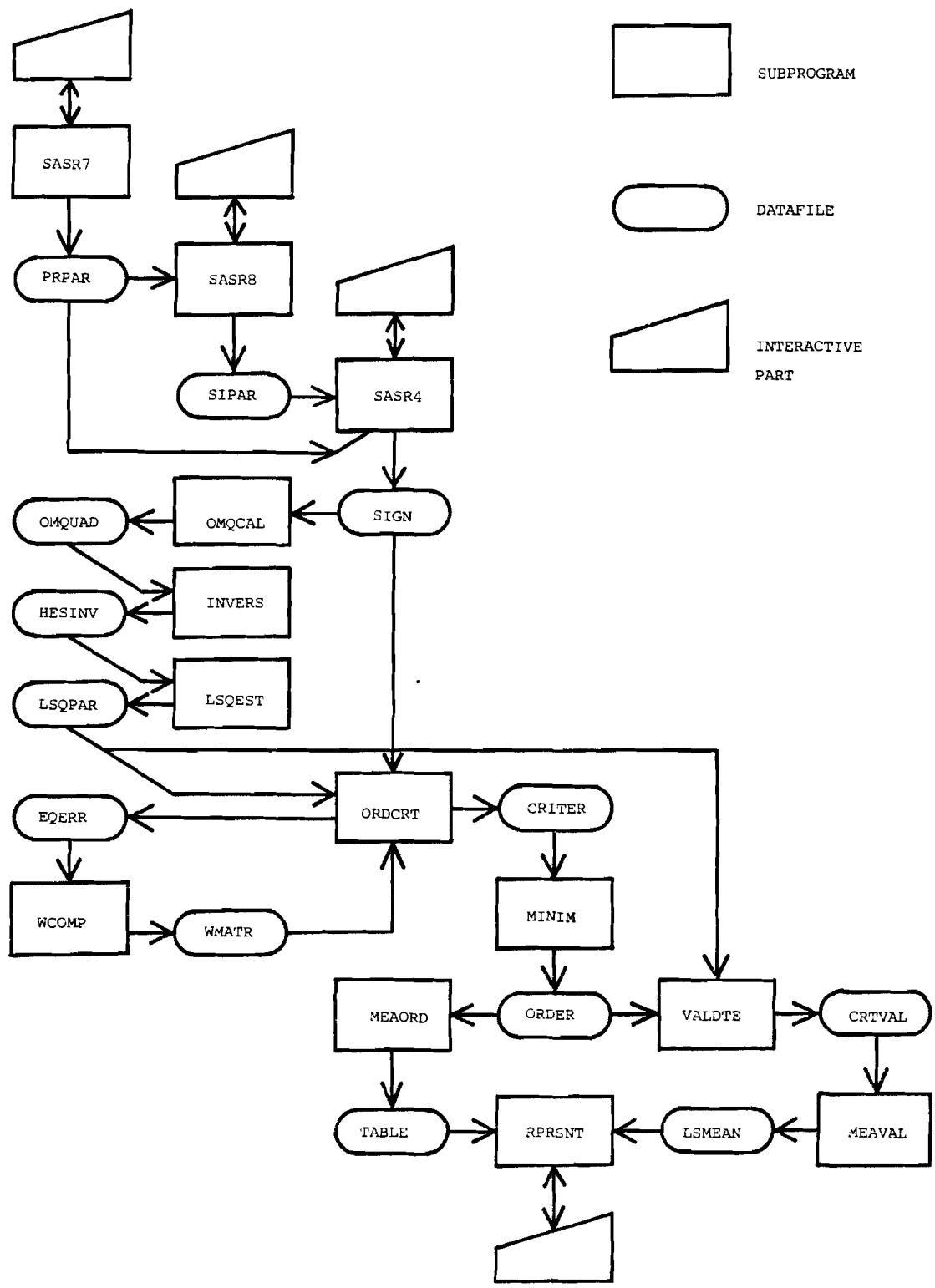


Fig. 2.5 Flow diagram of the program STMAIN

Fig. 2.6 Data transfer in the program STMAIN



2.4 A comparison of several criteria

Akaike (1974) introduced a model selection rule, as an extension of the maximum likelihood principle. Maximizing the log-likelihood function would lead to models with parameter vectors of high dimensionality, for which Akaike introduced an appropriate correction term.
In the case of a least squares estimation, the Akaike Information Criterion (AIC) comprises:

$$AIC = \log V(\underline{\theta}) + \frac{2}{N} \dim \underline{\theta} \qquad (2.39)$$

Minimization of AIC is supposed to lead to the 'true' order. Shibata (1976) and Kashyap (1980) have shown that the AIC gives an asymptotic inconsistent estimate of the order for the autoregressive case; AIC is shown to be not a weakly consistent estimator.
Akaike (1977) and Rissanen (1978) both introduced equivalent criteria that can be described as:

$$BIC = \log V(\underline{\theta}) + \frac{1}{N} \log N \cdot \dim \underline{\theta} \qquad (2.40)$$

Hannan and Quinn (1979) derived a similar procedure to determine the order of an autoregression, which for the case of a least squares estimation procedure can be described as:

$$\Phi = \log V(\underline{\theta}) + \frac{1}{N} c \log \log N \cdot \dim \underline{\theta} \qquad (2.41)$$

Both BIC and Φ are generalized in the Generalized Akaike's Information Criterion:

$$GAIC = \log V(\underline{\theta}) + \frac{k_N}{N} \dim \underline{\theta} \qquad (2.42)$$

Hannan (1980) gives a review of the three tests of order and shows that both BIC and Φ are strongly consistent for the case of a time series ARMA estimation with independent noise.
Stone (1977) has shown the asymptotic equivalence of the choice of a model by cross-validation and Akaike's criterion. Stoica (1984) proves this under more general conditions, namely:

$$\log(\frac{1}{N} C_{I1}) = AIC + O(1/N^{1\frac{1}{2}}) \qquad (2.43)$$

under the assumption $E[\hat{e}.\underline{\hat{e}}_{\theta\theta}] = \underline{0}$.
Further he proves that C_{IIk}, cf. equations (2.24) and (2.26), is asymptotically equivalent with the Generalized Akaike's Information Criterion:

$$\log(\frac{1}{(k-1)N} C_{II}) = GAIC + O(\frac{1}{m.\min(k^{\frac{1}{2}}, m^{\frac{1}{2}})}) \qquad (2.44)$$

under the assumption that $\hat{e}$ is white noise.
In the program STMAIN these order test criteria are implemented in subprogram ORDCRT, cf. figures 2.5 and 2.6.
In order to test the usefulness of our practical selection rules based on C_1 and C_2 Stoica proposes to compare AIC with $C_1(m=1)$ and $C_1(m=5)$, GAIC(k=4) with $C_2(k=4)$ and finally GAIC(k=6) with $C_2(k=6)$. The choices of k_N are based on BIC and Φ, with $c > 2$, and a data sequence of $N = 500$ data samples.
The first comparison is the following.

2.4.1 Order estimation performance

The aim of this example is to test the ability of the selection rules to choose the true structure of the system. This implies that the system that generated the data must belong to the model set that is being handled by the estimation procedure. In the program the process is simulated with a fourth order autoregressive model. The estimation model of STMAIN is also based on an autoregressive model.
The simulations are done with zero input; u(k)=0. To attain a good impression of the performance, the simulation runs are done 100 times with random sequences for the disturbance e(k). These sequences are Gaussian, white and are generated sequentially.
Each starting number of a noise sequence (of 500 samples) is generated by the preceding sequence of random numbers. This is done in order to avoid using an identical starting number twice, which would lead to identical

or partially identical noise sequences.
The poles of the system are chosen not too close to the unit circle in the z-plane, to avoid numerical instability in the estimation process. Several pole configurations have been submitted to the performance comparison procedure.
The results are presented in the Tables 2.1-a to 2.1-c (see Tables). Table 2.1-a represents the most typical pole configuration of a properly sampled process. Table 2.1-b is an example of practical sampling, the sample frequency is relatively high. Table 2.1-c is an example of a fourth order process that, at a four times larger sample period, becomes a first order process. Fig. 2.7-a to 2.7-c represent the corresponding pole-zero configurations in the z-plane. For a discussion of these results, see Section 2.5.

2.4.2 Comparison of predictive performance

Another way to compare the performances of the various criteria is to test whether a criterion selects a good one step ahead predictor structure. This is clearly necessary where the process generating the data does not have to belong to the model set at hand.
Therefore simulations are done of ARMA(2,2) systems, in which we distinguish two cases, namely systems with poles of the MA-part close to zero and systems with the MA poles close to the unit circle. The latter case will result in selection of a higher order of the approximating AR model. The evaluation of the performance of a selected model can be done with the following criterion:

$$L_S = \frac{1}{S} \sum_{t=1}^{S} \left[y(500+t) - \hat{y}(500+t) \right]^2 \qquad (2.45)$$

where $\hat{y}$ is the one step ahead prediction calculated with the selected model and S the number of samples for which performance of prediction is examined.
Stoica suggests as values of S:
S = 1, S = 5, S = 1500 and S = 2500.
These values are based on the assumption that S/N = m/(N-m) is the most

appropriate ratio for the first criterion and S/N = (N-m)/m for the second criterion (cf. Stoica (1984) pp. 12,23).
Also the simulation has been performed 100 times for each of a number of pole configurations, with different random sequences for the input disturbances. Each noise sequence (of 3000 samples; 500 for estimation and 2500 for validation) is again initiated with the start number generated as the last number of the preceding sequence.
The interpretation of the criterion of equation (2.45) is that the lower L_S is, the better the performance of the selected model is for S samples ahead. The simulations of each configuration of the system have been carried out for both zero input (u(k)=0; time series case) and stochastic-like input u(k), independent of the disturbing noise e(k). The values of L_S are averaged for each selection rule and for each value of S. The results are presented in the Tables 2.2-a to 2.2-d (see Tables). Tables 2.2-a present the examples of the different processes with zero input. Tables 2.2-e to 2.2-h are the same processes but with a stochastic-like input signal. Fig. 2.8-a to 2.8-d represent the corresponding pole-zero configurations in the z-plane. For a discussion of the results, cf. Section 2.5.

2.5 Conclusion

The rule for dividing the data sequence into different intervals is an unsolved problem. Division of the sequence has to be done with caution, since interrupting the estimation in the sequence can ignore the 'tail'effects of the process. However, this can be overcome by using data from the validation interval in the following estimation sample in such a way that the minimization only concerns the prediction of the next sample that proceeded from data of the validation sample.
The original cross-validatory selection rules C_I and C_{II} contain this drawback of dividing the data sequence, whereas the suggested approximations for both rules do not have this practical disadvantage and therefore can be useful for application.

The first experiment, intended to confirm the asymptotical equivalence between C_1 (m=1) and AIC, proves that the new criterion has the same

ability of selecting the right order as Akaike's Information Criterion. As for C_1 (m=5), a difference in performance is barely noticeable. The supposed equivalences between GAIC (k=4), (k=6) and C_2 (k=4), (k=6) respectively are less obvious; the C_2 criterion tends to be less effective than its opponent, the Generalized Akaike's Information Criterion. Table 2.3 gives us an impression of the computational labor at stake for both AIC and C_1. GAIC and C_2 have identical figures.

Table 2.3
The number of multiplications and divisions compared for the AIC and C_1 for both Newton_Raphson and least squares estimation (time series case, autoregression).

	AIC	C_1
Newton-Raphson	N + l + 4	$N + (\frac{1}{2}k+1)n^2 + \frac{1}{2}kn + 4$
least squares	N(n+1) + l + 4	$N(2n+1) + \frac{1}{2}kn^2 + \frac{1}{2}kn + 4$

With N is the number of samples, k the number of intervals, n the order of the model and l the number of multiplications equivalent to a logarithmic operation. Note that for m=1, then k=N.

It appears that C_1 and C_2 bear a substantially greater computational burden. Together with the foregoing results this disfavours the use of C_1 and C_2 as a model selection rule.

The second experiment, which was intended to be a confirmation of a better predictive performance, shows that this does not hold. Simply stated, the assumption was that a model, selected by a rule based on a certain ratio between 'check' and 'estimation' sample lengths, would reflect this ratio in its ability to predict; cf. Stoica (1984) pp.12 and 23.
The results of this experiment (cf. Tables 2.2-a to 2.2-h) strongly contradict this assumption, since L_S varies barely for the different criteria.
An explanation of this behaviour is the following. The idea that higher

order models perform better than lower order models in 'short term' prediction and, vice versa, that lower order models have better performance in 'long term' prediction originates from polynomial estimation. Higher order polynomial models are more accurate near the estimation sample, but become unstable in the end, while lower order models remain stable but are less accurate.
However, this does not always hold for stochastic models. In the case where the system is not included in the model set, the highest order model has the best performance, since a model of infinite order will describe the system completely.
The fact that the accuracy of the estimate decreases with the number of parameters to be estimated restricts this statement.
Ignoring this loss of accuracy, leads to the following.
The process is considered to be an ARMA type,while the model is taken to be an AR type. The system can then be described as:

$$y(k) = b_0 e(k) + \ldots + b_n e(k-n) - a_1 y(k-1) - \ldots - a_n y(k-n) \quad (2.46)$$

with n the order of the process.
Equation (2.46) can also be denoted as a function of all disturbances after the estimation sample and the output during the estimation:

$$y(k) = \sum_{j=0}^{k-N-1} h_j e(k-j) + \sum_{i=0}^{N-1} \gamma_{ki} y(N-i) \quad (2.47)$$

The second term on the right hand side of this equation accounts for the tail effect of the disturbances during the estimation interval. The model can be described as:

$$\hat{y}(k) = -\hat{a}_1 y(k-1) - \ldots - \hat{a}_{\hat{n}} y(k-\hat{n}) \quad (2.48)$$

with $\hat{n}$ the selected order and $\hat{a}_1, \ldots, \hat{a}_n$ the corresponding estimated parameters. With $y(k) - \hat{y}(k) = e(k)$ we can denote:

$$y(k) = \hat{e}(k) - \hat{a}_1 y(k-1) - \ldots - \hat{a}_{\hat{n}} y(k-\hat{n}) \quad (2.49)$$

Also this equation can be written as a function of all errors $\hat{e}$ after the estimation interval and the output of the last $\hat{n}$ data samples during the estimation:

$$\hat{y}(k) = \sum_{j=1}^{k-N-1} h'_j e(k-j) + \sum_{i=0}^{\hat{n}-1} \gamma'_{ki} y(N-i) \qquad (2.50)$$

Comparing the equivalence of equations (2.47) and (2.50) we can conclude that for the second term of (2.50), the highest model order $\hat{n}$ results in the best description of the tail effect of the estimation part, since the information hidden in the output can be exploited more with a larger number of data.

This confirms the assumption that higher order models have a better performance just after the estimation part.

As for long term prediction, a higher order model still uses better information of the preceding samples and therefore is certainly not worse than a lower order model, within the restrictions of loss of accuracy of the estimate. To clarify this, think of an estimated model of order n. If we consider a model of the same set of models of order n+1, this model includes the former model.

If the n^{th} order model would be better, an optimal estimator would choose the $n+1^{st}$ parameter zero. If this is not the case, the $n+1^{st}$ order model is better, as long as $n+1 << N$, with N the size of the estimation interval. Note that this reasoning concerns the case of a process outside of the model set.

3. Programming aspects of the use of the SATER package

3.1 Introduction

The SATER program package (System Analysis Techniques, developed by the group ER) has been developed by various people to make a number of well-known estimation and order test algorithms for SISO systems practically available to a large group of users (cf. v.d. Boom and Bollen, 1983). Therefore the package is designed as an interactive program, which has several advantages.
The SATER program package is written for application on a PDP 11/60 minicomputer system. Minor alterations will suffice to obtain an executable version on different computer types.
However the possibilities and restraints of the PDP system have shackled the development of the program. Clearly this is a general symptom of any arbitrary program, that makes use of possibilities that reach beyond the level of the programming language.
Intensive confrontation with the structure of the SATER program has lead to a more clear insight into matters concerning the interactivity and general applicability of programs.
In the following comment will be given on several aspects of the design of interactive programs in the field of control research and development, to contribute to the success of future projects.

3.2 Interactive aspects

A crucial point of general purpose programs, intended for application by a large group of users, is the degree of interactivity that is implemented in the program.
Clearly the ease of using a program influences the frequency of its application; furthermore a possible discomfort in the application can make potential users decide to abandon its use and to write a proper implementation for a special application.
The SATER package contains a structure to handle questions in a user-friendly way. Fig. 3.1 illustrates this principle in a diagram.

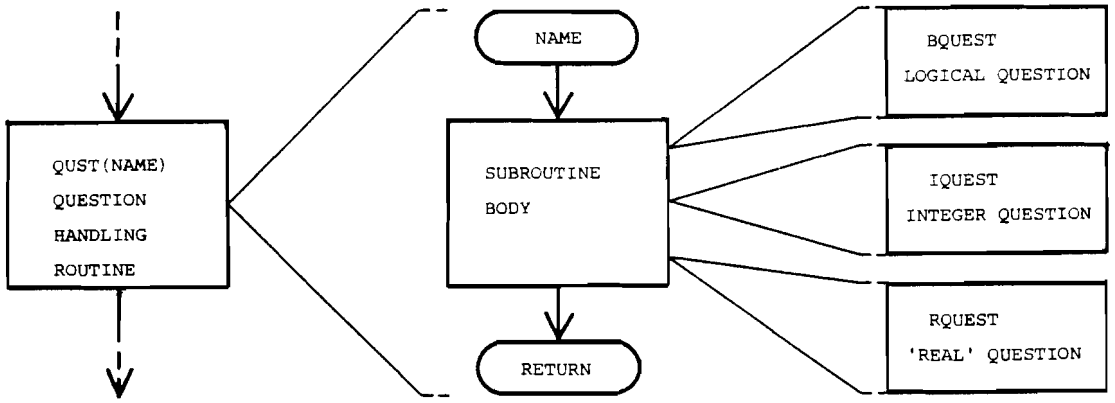


Fig. 3.1 Application of the question subroutine QUST; the users application NAME contains the SATER question routines.

The subroutine QUST calls a subroutine specified in the argument list. This subroutine is meant to pose questions about a certain application, by means of question routines such as BQUEST, IQUEST and RQUEST for logical, integer and real value questions respectively.
Thus within the block NAME any answer can be given a default value and any answer can be altered while the user has not left the block.
Subroutine QUST also offers the possibility of terminating the program or restarting the specific application. Also a limited 'help' facility is available to inform the user about the questions present in block NAME.
The 'help' facility could be more general in the sense that it might also inform the user about the functions of the keys and the program package in general. Furthermore the absence of a facility to return to the former question block in the application is quite a disadvantage.
This could be implemented by organizing the questions in several levels, and at each level enabling the user to return to the previous level.
Doing so, a lot of annoyance can be avoided, since the necessity of restarting the application by some mistake, which can be quite time consuming and temper loosing, will occur less frequently.

The principle of using a large datafile (called MESG.DAT) containing all the messages, questions and answers of the whole program package appears quite practical. However, alterations on the messages as well as extensions to new applications can only be introduced in an impractical

way; the datafile is written in a special format. This special format needs a conversion program; first the sourcefile has to be updated and then the conversion can take place to produce the specially formatted file.
A better solution might be to have for each application a different message file, in such a format that it can be changed directly, and a system message file containing all the basic messages, not to be altered for new applications.

Another feature that directly concerns the interactive ability of a program package is the graphics facility. The SATER package contains a number of system routines that are to be used in the various application blocks. Thus each application needs its own configuration of these system routines to produce a nice graphical representation of the achieved results.
This might also be organized in a different manner; a graphical block at the same level of the various application blocks can be introduced. Each application block can thus deliver data that can be represented in a uniform way by the graphical block.
This way the user can decide better how the results should be represented and applications do not need to be repeated to give different graphical representation forms.
A last remark on interactive aspects is that the lay-out on the screen is also quite an important aspect; a less carefully arranged lay-out on the screen does not improve the credibility of the quality of the program package. In this respect the SATER program performs quite well.

3.3 Data handling

The transport of data is organized via a certain file structure in SATER. The different application blocks can transfer the produced data via these files. There exists a considerable number of ways to organize a datafile and unfortunately each program package, so it seems, uses its own structure. The structure of SATER is a practical one, but before application external files will have to be converted to the right structure. The PDP computersystem prohibits the use of large data arrays, due to the limited memory space in the core memory. This is why

applications that need the data a number of times, i.e. iterative algorithms, are slowed down considerably by the necessary I/O actions, that take place with peripheral memory.
Larger computersystems need a different programming strategy, for their core memory will be sufficient to contain the whole contents of a datafile.

A universal program package will be complicated to realize, but an approach that enables both strategies can be considered. Thus the computational performance of the package can be adapted to the features of the host computer system.
A limitation of the programming language can also play an important part. Fortran does not have the possibility of declaring variable arrays in the main program. Therefore the operating system can offer facilities, which make the whole package quite machine-dependent. Other languages, such as PASCAL, do not have this drawback but are, at this moment, less frequently used.
The declaration of variable array length can be advantageous, this way the program only occupies the memory space that is really necessary, thus facilitating a better use of the host computer.

3.4 Universality

The SATER program package contains numerous subroutines, that perform interesting algorithms. Therefore it seems reasonable that the tools, gathered in this package, are also available for application outside the SATER shackle. This requires an organization of the existing subroutines in a library structure. In doing so the potential user will not have to be concerned about various subroutines that are called by the subroutine he wishes to apply, since all are available in the library.
Clearly this requires a disciplined programming structure for all subroutines, in which there can be no room for the use of common blocks. All variables must be transferred as argument values of the subroutines, thus avoiding different levels of parameter transfer.
It is advisable to use a standard implementation of the programming language, so that the number of machine dependent routines is reduced to a minimum. Thus the amount of work to alter these routines for other host

machines will be calculable. Furthermore it is advisable to standardize input and output actions completely, so that the applications that make use of the library can also be executed in a non-interactive mode (batch mode).

PART TWO

4. Estimation of MISO systems via a SISO model approach

4.1 Introduction

In the case of a two-input-two-output system it seems feasible to model this system in four different SISO blocks (cf. fig. 4.1).

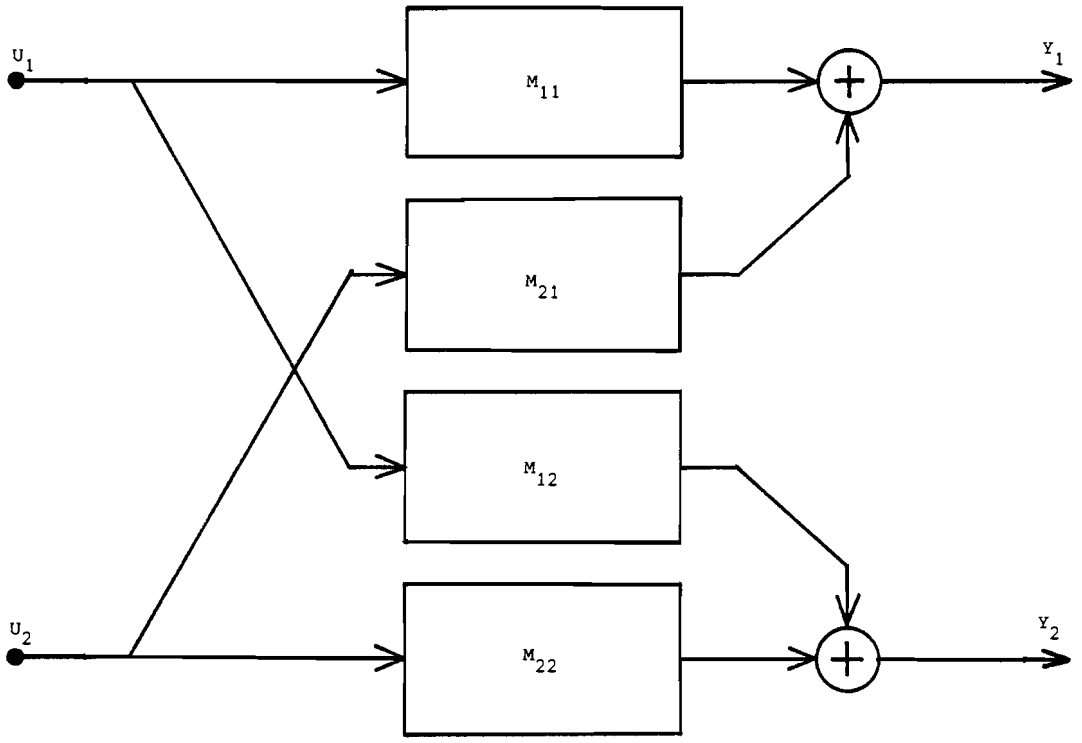


Fig. 4.1 A two-input two-output system.

For estimation purposes this model can be split in two separate multiple-input single-output (MISO) system models, which can be identified separately.
Clearly this division is only allowed, assuming that the outputs do not influence each other mutually.
On this assumption, we can state the following. Assuming a matrix fraction description

$$P(z^{-1}).\underline{y}(k) = Q(z^{-1}).\underline{u}(k)$$

of a MIMO transfer system, with $\underline{u}$ the q-dimensional input signal and $\underline{y}$

the p-dimensional output signal and P and Q the polynomial matrices of z^{-1}, the unitary time shift operator, we can write

$$\underline{y}(k) = P^{-1}(z^{-1}).Q(z^{-1}).\underline{u}(k)$$

if P is non-singular.
Thus the i^{th} output signal can be denoted as

$$y_i(k) = \underline{r}_i(z^{-1}).\underline{u}(k) \qquad i = 1,\ldots,p$$

with $\underline{r}_i$ the i^{th} row vector of $P^{-1}.Q$.
This implies that each output can be expressed as a function of the inputs only, so the division into MISO systems is allowed for this method of description.
The subject of this chapter mainly concerns the estimation of those MISO systems by means of a SISO model structure. Some examples of existing estimation methods based on this concept will be given below.

Bollen (1980) uses standard SISO estimation algorithms for the parameter estimation of MIMO systems in a Guidorzi form. This direct application of SISO methods seems quite hazardous, but a previous structural identification of the system is performed, which apparently allows the SISO approach. Nevertheless, the SISO estimation will require excessive modelling of the additive 'disturbances', which leads to 'double' modelling. First a transfer is considered a noise model, then, performing the estimation of the specific transfer, it will be considered a transfer model. Doing so, each SISO subsystem of the MIMO will be estimated twice, namely as a noise model and as a transfer model.

Barrett-Lennard and Blair (1981) described a method to estimate multiple-input single-output parameter coefficients in a repetitive way (iteratively) with a recursive refined instrumental variable approximate maximum likelihood (IVAML) method for the different SISO transfer systems. This method was designed for both the similar and the dissimilar denominator structure of a MISO system. Theoretical justification remained a problem, but the method has been shown to work.

In the following, a dissimilar denominator structure of the system is assumed, in other words, the subsystems are supposed to have different autoregressive parts. Fig. 4.5 illustrates this model structure. Barrett-Lennard and Blair suggest that the use of such a model structure may be considerably more fruitful than the common denominator structure, especially in the case of two or three inputs. This may arise from the fact that a smaller number of parameter coefficients need to be estimated.

4.2 Introductory comparison of various methods

Three methods for estimating a MISO model have been compared. First a brief description will be given of each of these methods.

4.2.1 'One shot' method

This method is based on the least squares method and minimizes the equation error (cf. fig. 4.2).

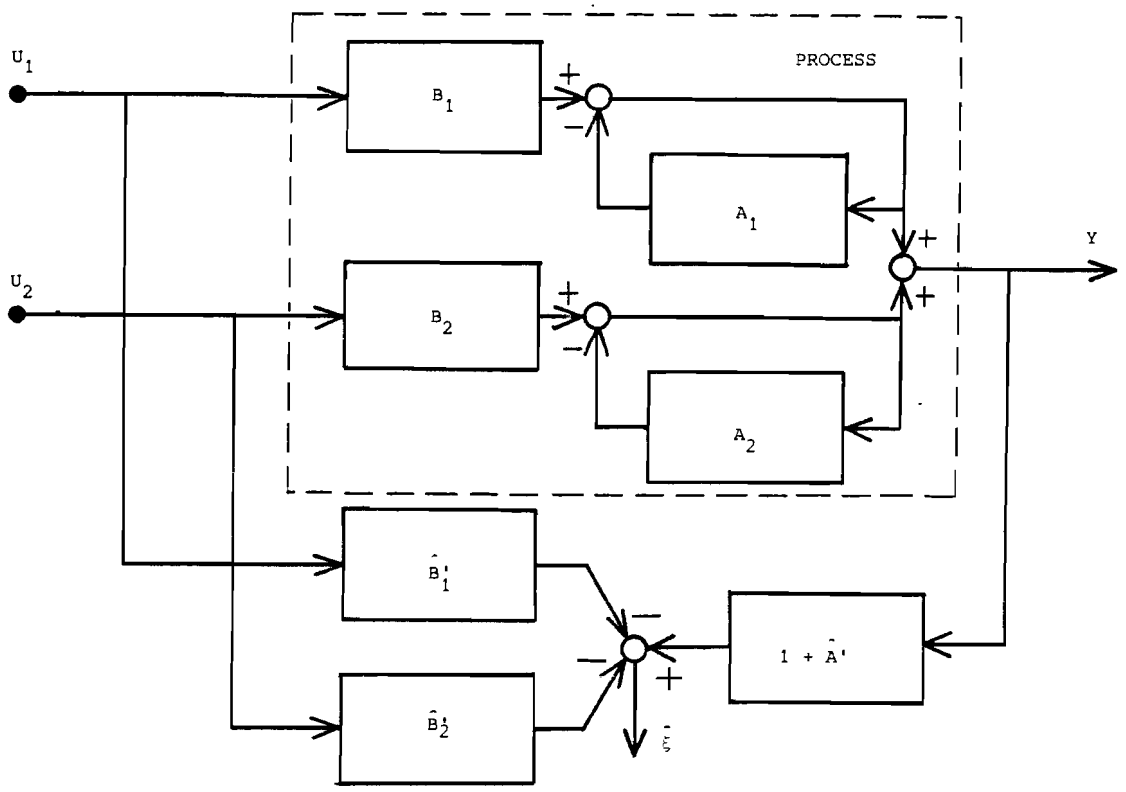


Fig. 4.2 The 'One shot' method.

Clearly this is an equation error method, but it is not obvious at what point the supposedly white noise, corresponding with this error, is assumed to be operating in the system, since the model contains only one autoregressive part and the system has two supposedly different autoregessive parts.

Furthermore the estimated model has to be checked for partial pole-zero cancellation between the two moving average parts and the autoregressive part. This is caused by the fact that $\hat{B}_1$ and $\hat{B}_2$ are

estimates of $B_1(1+A_2)$ and $B_2(1+A_1)$ respectively and $1+\hat{A}'$ is an estimate of the product $(1+A_1)(1+A_2)$.

4.2.2 Iterative equation error method

This is an alternating method; the two SISO models are estimated in an alternating fashion, first the second model is considered to be producing equation noise in the first SISO model, then the second model is estimated with the first model producing equation noise (cf. fig. 4.3).

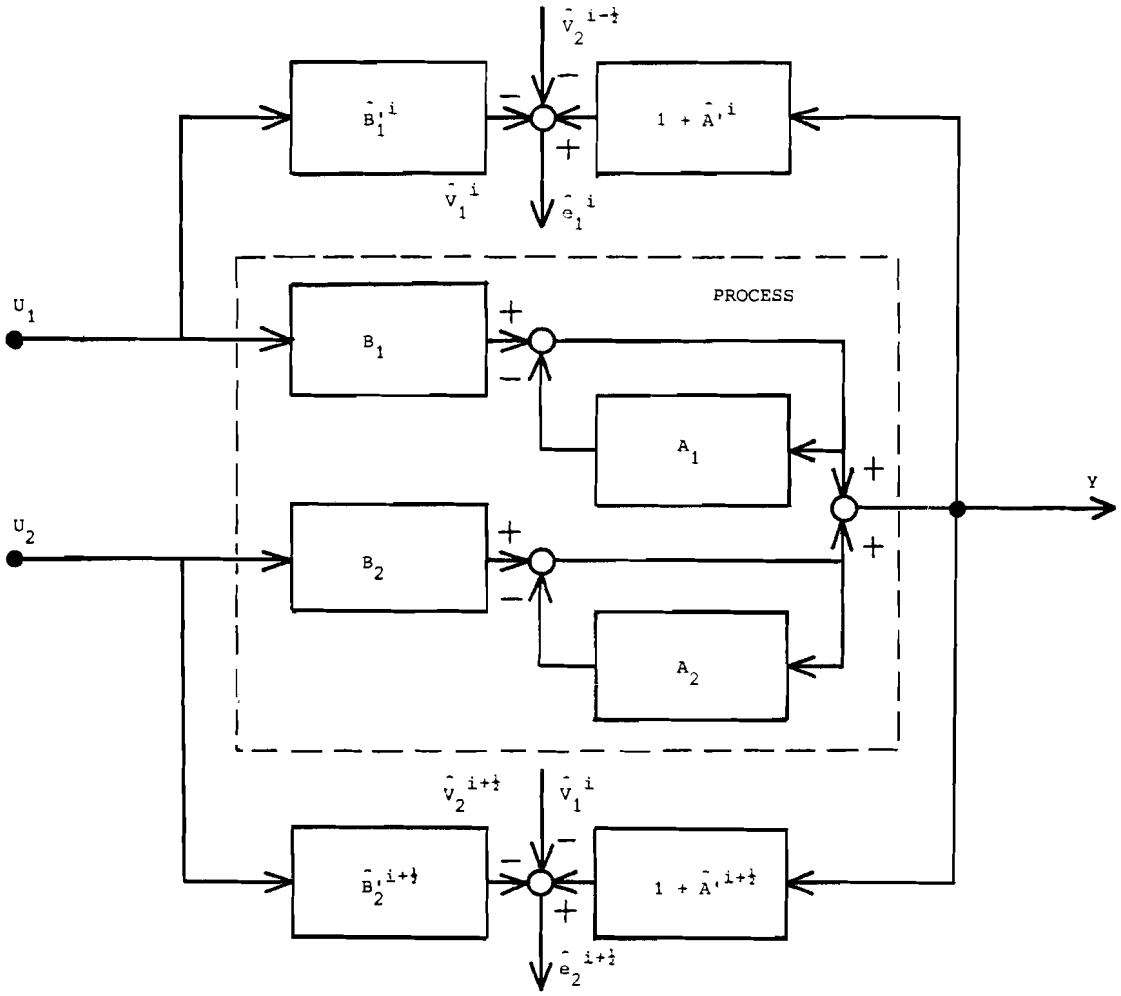


Fig. 4.3 The iterative equation error method.

This method is afflicted with the same drawbacks as the first method; consequently extensive analysis has to be done to check for partial pole-zero cancellation and the equation noise is vaguely defined.

4.2.3 Output error method

This is also an alternating method, the two SISO models being estimated in an alternating way.
First the second model is considered a compensatory part of the output noise of the first model, then the second model is estimated with the first model as additive output noise (cf. fig. 4.4).

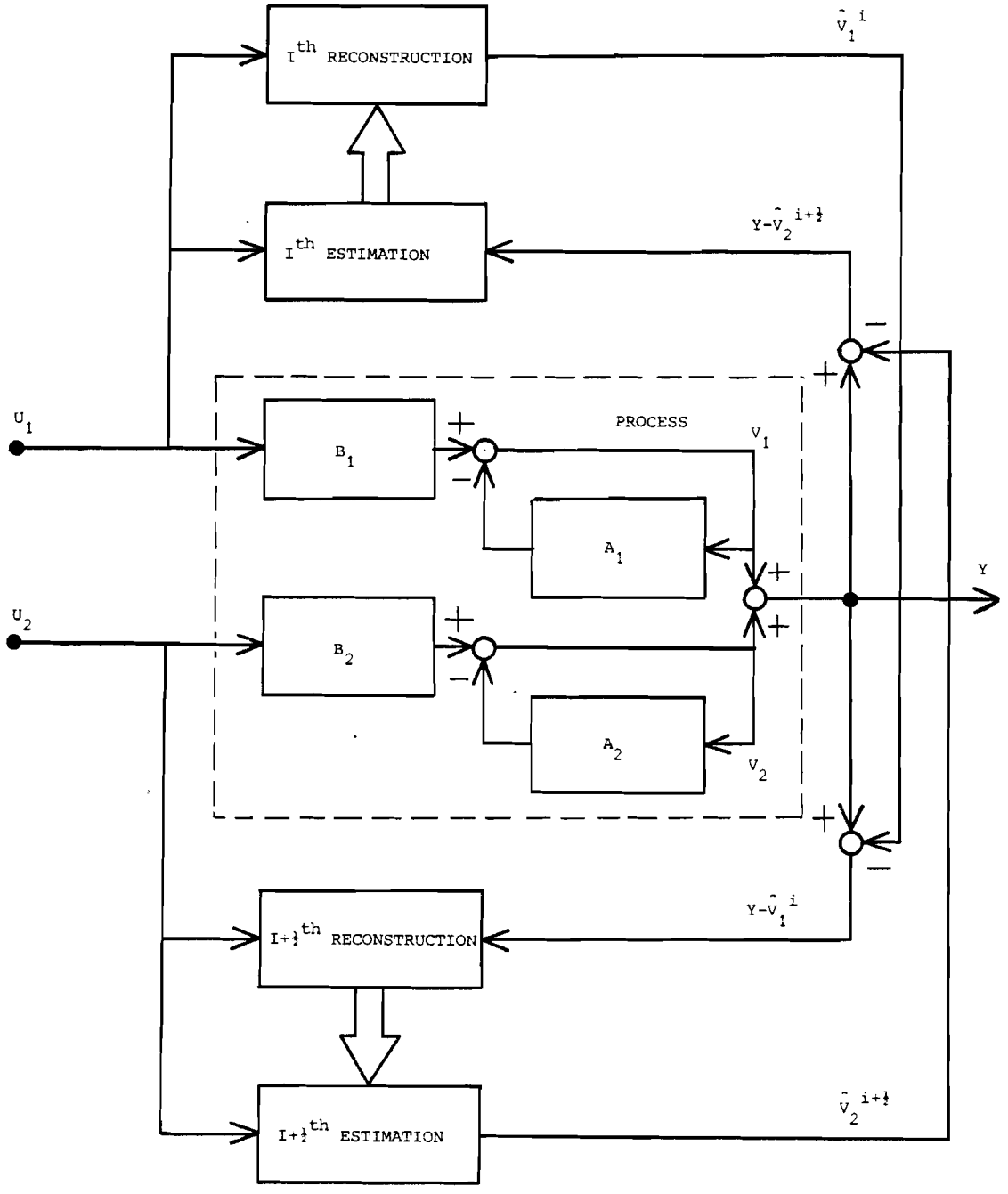


Fig. 4.4 The alternating iterative output error method.

The advantage of this method is that it inherently is an output error method, for which the pole-zero cancellation of the former methods does not occur.

4.2.4 Recapitulation

A summary of these three methods is given in Table 4.1.

Table 4.1 Summary of the three methods.

method	1	2	3
estimation	'one shot'	iterative	iterative
solution	analytical	analytical	iterative
minimizing *)	equation error	equation error	output error
computation time **)	$18n^2N$	$16n^2N + 8knN$	$2k(2+\frac{1}{2}l)n^2N$
pole-zero cancellation	yes	yes	no

*) $n_a = n_b = n$, N = number of samples with $N >> n$

**) method 1 can also be changed to an output error method at the cost of more computation time.

Out of the three methods mentioned above the third method is chosen for further investigation.
This alternating method seems to have the most appropriate characteristics, such as an output error approach of the estimation and the absence of pole-zero cancellation. It is not clear on forehand however, whether this method has acceptable converging properties.

4.3 Investigation into convergence properties

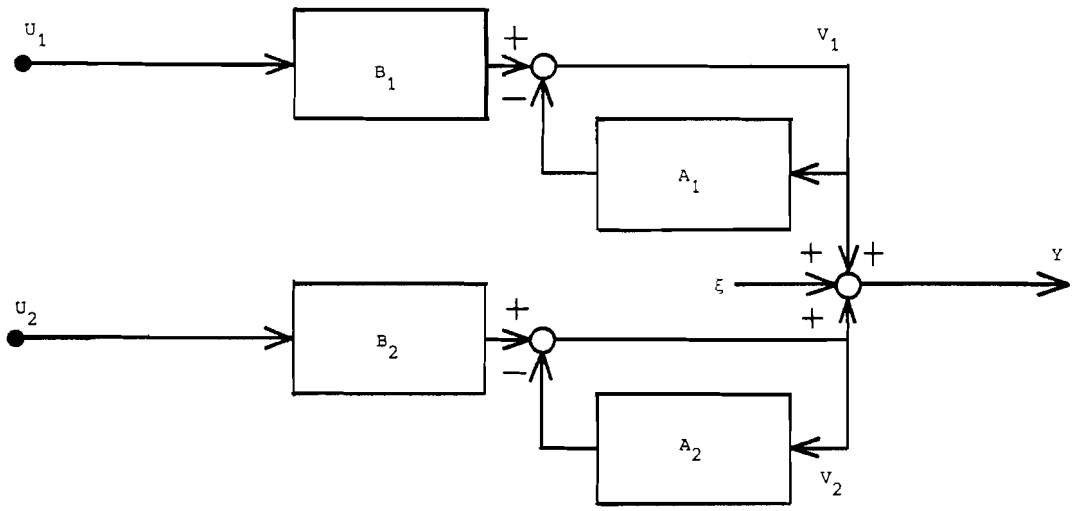


Fig. 4.5 An ARMA model of a MISO system.

The two SISO subsystems can both be described as follows:

$$v(k) = -a_1 v(k-1) - \ldots - a_{n_a} v(k-n_a) + b_0 u(k) + \ldots b_{n_b} u(k-n_b) \quad (4.1)$$

with a_i the autoregressive parameters, n_a the order of the autoregression, b_i the moving average parameters and n_b the moving average order.

The output noise ξ is not included in the description of the two SISO subsystem outputs, for it is operating on the summation of both signals (cf. fig. 4.5 and eq. (4.12)).

A model can be denoted as follows:

$$\hat{v}(k) = -\hat{a}_1 \hat{v}(k-1) - \ldots - \hat{a}_{\hat{n}_a} v(k-\hat{n}_a) + \hat{b}_0 u(k) + \ldots + \hat{b}_{\hat{n}_b} u(k-\hat{n}_b) \quad (4.2)$$

with the superscript 'hat' indicating that it concerns an estimate.

In the following the input signals u_1 and u_2 and the output noise signal ξ are assumed to have white noise properties and to be mutually independent. This can be denoted as:

$$E[u_i(k)u_i(k+j)] = \delta(j) \qquad i=1,2 \qquad (4.3\text{-a})$$

$$E[u_1(k)u_2(k+j)] = 0 \qquad (4.3\text{-b})$$

$$E[u_i(k)\xi(k+j)] = 0 \qquad i=1,2 \qquad (4.3\text{-c})$$

Suppose that the model orders $\hat{n}_a$ and $\hat{n}_b$ are chosen correctly, then we can say that the parameter estimate differs as follows from the true value:

$$\hat{a}_i - a_i = \Delta a_i \text{ and } \hat{b}_i - b_i = \Delta b_i \qquad (4.4)$$

Assume also that the reconstructed SISO output values $\hat{v}(k-j)$, preceding the reconstruction $\hat{v}(k)$, differ as:

$$\hat{v}(k-j) - v(k-j) = d(k-j) \qquad (4.5)$$

Thus the difference signal between the true SISO output and the reconstructed SISO output signal can be denoted as:

$$d(k) = \hat{v}(k) - v(k) = -\sum_{j=1}^{n_a} a_j d(k-j) - \sum_{j=1}^{n_a} \Delta a_j v(k-j) + \sum_{j=0}^{n_b} \Delta b_j u(k-j) \qquad (4.6)$$

Fig. 4.6 shows this difference signal in a block diagram.

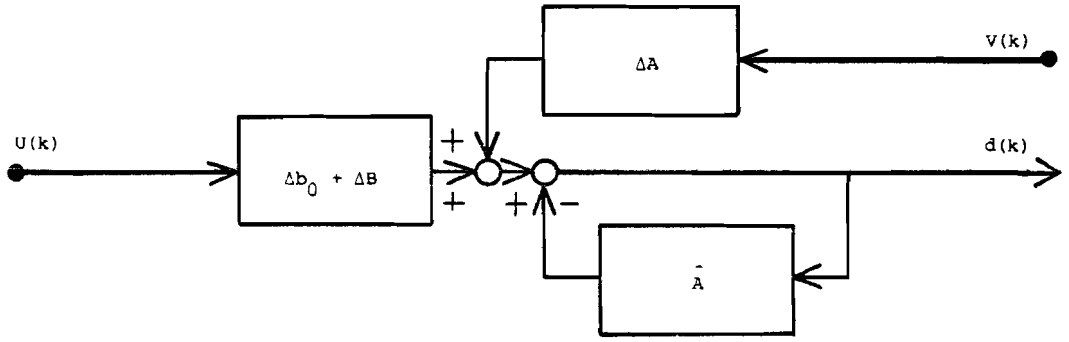


Fig. 4.6 Construction of the difference signal $d_i(k)$

Note that the difference signal $d_i(k)$ only depends on $u_i(k)$, with i=1 or 2, and that the noise signal $\xi(k)$ independently operates on the summation of both SISO output signals $v_1(k)$ and $v_2(k)$ (also cf. fig. 4.5 and eq.

(4.12)).

The magnitude of d(k) is mainly dependent on Δa_j (j=1,...,n_a) and Δb_j (j=1,...,n_b), with given input and output signal.

If $\hat{A}_i$, the i^{th} estimate of A, is close enough to A, we can write:

$$E[d_i^2] = \alpha\{ \sum_{j=1}^{n_a} \Delta a_j^{2_i} E[v^2] + \sum_{j=0}^{n_b} \Delta b_j^{2_i} E[u^2] + E[\text{cross terms(i)}]\} \tag{4.7}$$

With $\hat{A}_i \approx A$, α can be regarded as a factor relatively independent of the iteration i.

The cross terms, on the other hand, can change per iteration, but are assumed to be small enough. Eventually the expectation of the crossterms can be denoted as:

$$E[\text{cross terms(i)}] = \gamma_i E[u^2] \tag{4.8}$$

with γ_i accounting for the iterative dependence.

The expectation of v^2 can be written as:

$$E[v^2] = \beta E[u^2] \tag{4.9}$$

with β invariant for successive iterations.

With equations (4.7), (4.8) and (4.9) we can denote for the i^{th} iteration:

$$E[d_i^2] = \{\alpha(\beta \sum_{j=1}^{n_a} \Delta a_j^{2_i} + \sum_{j=0}^{n_b} \Delta b_j^{2_i} + \gamma_i)\} E[u^2] \tag{4.10}$$

Thus the subjoined requirement:

$$E[d_{i+1}^2] < E[d_i^2] \tag{4.11}$$

will be sufficient for convergence, since it implies a decrease of the magnitude of Δa_j and Δb_j, assuming that γ_i remains small.

The MISO system can be described as follows:

$$y(k) = v_1(k) + v_2(k) + \xi(k) \tag{4.12}$$

In eq. (4.12) $v_1(k)$ and $v_2(k)$ are the two SISO output signals according to eq. (4.1) and $\xi(k)$ is the noise component at the output (also cf. fig. 4.5).
Assume that the j^{th} estimation results in the reconstructed SISO signal $\hat{v}_1^j$:

$$\hat{v}_1^j = v_1(k) + d_1^j(k) \tag{4.13}$$

Note that the difference signal d_1^i now includes a noise component of ξ, for the estimation is performed with a limited number of samples. The second SISO subsystem is then estimated from the input signal $u_2(k)$ and the reconstructed output signal :

$$v_2(k) + \xi(k) - d_1^j(k) \tag{4.14}$$

this results in the estimated signal :

$$\hat{v}_2^{j+\frac{1}{2}}(k) = v_2(k) + d_2^{j+\frac{1}{2}}(k) \tag{4.15}$$

Fig. 4.7 illustrates this.
Because of the finite number of samples and in order to minimize the loss function, the estimation algorithm will succeed in modelling part of the output noise.
Since the model is supplied with $F = n_a+n_b+1$ degrees of freedom, the estimation procedure is expected to model a part of the noise contents having N degrees of freedom, with N the number of samples. For F degrees of freedom enable to incorporate F of the N degrees of freedom of the noise, we assume that over N samples a fraction r(F,N) of the noise power will be modeled. The reconstructed output signal used for the estimation, according to eq. (4.14) can be written in terms of the estimated signal and the corresponding error as :

$$v_2(k) + \xi(k) - d_1^j(k) = \hat{v}_2^{j+\frac{1}{2}}(k) + \hat{\xi}^{j+\frac{1}{2}}(k) \tag{4.16}$$

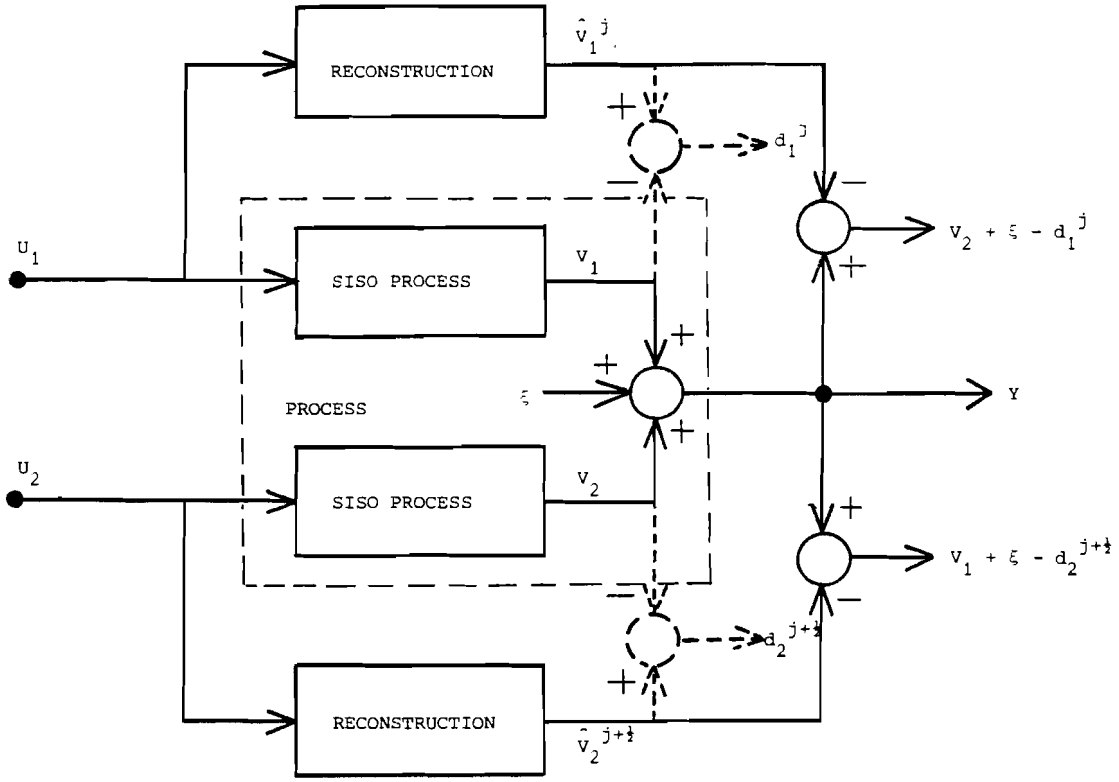


Fig. 4.7 Explanation of the estimation procedure, d_1 and d_2 are imaginary signals, but $v_2+\xi-d_1^j$ and $v_1+\xi-d_2^{j+\frac{1}{2}}$ can be obtained by subtracting the reconstructed SISO signal from the output.
The estimate of the parameter vector will differ from the true parameter vector, the differences Δa_i, Δb_i being dependent on the noise signal $\xi(k) - d_1^j(k)$.

with $\hat{v}_2^{j+\frac{1}{2}}$ from eq. (4.15) and $\hat{\xi}^{j+\frac{1}{2}}$ the output error of the $j+\frac{1}{2}^{th}$ iteration. Invoking eq. (4.15) in eq. (4.16) yields :

$$\hat{\xi}^{j+\frac{1}{2}}(k) = \xi(k) - d_1^j(k) - d_2^{j+\frac{1}{2}}(k) \qquad (4.17)$$

By raising both left hand and right hand terms of eq. (4.17) to a square we obtain :

$$\hat{\xi}^{2}{}^{j+\frac{1}{2}}(k) = \xi^2(k) + d_1^{2}{}^j(k) + d_2^{2}{}^{j+\frac{1}{2}}(k) - \\ -2\left[\xi(k)d_1^j(k) + \xi(k)d_2^{j+\frac{1}{2}}(k) - d_1^j(k)d_2^{j+\frac{1}{2}}(k)\right] \qquad (4.18)$$

The loss function corresponding with the $(j+\frac{1}{2})^{th}$ estimation can be represented as

$$V^{j+\frac{1}{2}} = \sum_{k=1}^{N} \hat{\xi}^{2}{}^{j+\frac{1}{2}}(k) \qquad (4.19)$$

The noise component in the reconstructed output signal of eq. (4.14) is:

$$\xi(k) - d_1^j(k) \qquad (4.20)$$

Hence an exact estimation would lead to the following :

$$\hat{\xi}_e{}^{j+\frac{1}{2}}(k) = \xi(k) - d_1{}^j(k) \qquad (4.21\text{-a})$$

and

$$V_e{}^{j+\frac{1}{2}} = \sum_{k=1}^{N} \{\xi(k) - d_1{}^j(k)\}^2 \qquad (4.21\text{-b})$$

Subscript e indicates that it concerns an exact estimate.

Modelling of the noise by the algorithm will result in a reduction of the expected loss-function with a factor $[1-r(F,N)]$:

$$E[V^{j+\frac{1}{2}}] = [1-r(F,N)]E[\sum_{k=1}^{N} \{\xi(k) - d_1{}^j(k)\}^2] \qquad (4.22)$$

This leads to:

$$E[\sum_{k=1}^{N} \hat{\xi}^{2}{}^{j+\frac{1}{2}}(k)] = [1-r(F,N)]E[\sum_{k=1}^{N} \{\xi(k) - d_1{}^j(k)\}^2] \qquad (4.23)$$

With eq. (4.18) this results in:

$$E[\sum_{k=1}^{N} \{\xi^2(k) + d_1^{2}{}^j(k) + d_2^{2}{}^{j+\frac{1}{2}}(k) - 2[\xi(k)d_1{}^j(k) + \xi(k)d_2{}^{j+\frac{1}{2}}(k) - d_1{}^j(k)d_2{}^{j+\frac{1}{2}}(k)]\}] =$$

$$= [1-r(F,N)]E[\sum_{k=1}^{N} \{\xi^2(k) - 2\xi(k)d_1{}^j(k) + d_1^{2}{}^j(k)\}] \qquad (4.24)$$

Since $d_2^{j+\frac{1}{2}}(k)$ contains the modelling by the algorithm of the noise $\xi(k)-d_1^{j}(k)$, the last two terms of the left hand expression can be rewritten as:

$$2E\Big[\sum_{k=1}^{N} d_2^{j+\frac{1}{2}}(k)\{\xi(k)-d_1^{j}(k)\}\Big] = 2r(F,N).E\Big[\sum_{k=1}^{N}\{\xi(k)-d_1^{j}(k)\}^2\Big] \tag{4.25}$$

Invoking eq. (4.25) in eq. (4.24) and some rearranging yields:

$$E\Big[\sum_{k=1}^{N} d_2^{2\,j+\frac{1}{2}}(k)\Big] = r(F,N)\Big\{E\Big[\sum_{k=1}^{N}\xi^2(k)\Big] + E\Big[\sum_{k=1}^{N} d_1^{2\,j}(k)\Big] - 6E\Big[\sum_{k=1}^{N}\xi(k)d_1^{j}(k)\Big]\Big\} \tag{4.26}$$

The last term on the right hand side of eq. (4.26) can be worked out as follows:

$$E\Big[\sum_{k=1}^{N}\xi(k)d_1^{j}(k)\Big] = E\Big[\sum_{k=1}^{N} d_1^{j}(k)\{\xi(k)-d_2^{j+\frac{1}{2}}(k)\}\Big] + E\Big[\sum_{k=1}^{N} d_1^{j}(k)d_2^{j+\frac{1}{2}}(k)\Big] \tag{4.27}$$

Invoking a similar equation as eq. (4.25) yields:

$$E\Big[\sum_{k=1}^{N}\xi(k)d_1^{j}(k)\Big] = r(F,N).E\Big[\{\xi(k)-d_2^{j-\frac{1}{2}}(k)\}^2\Big] + E\Big[\sum_{k=1}^{N} d_1^{j}(k)d_2^{j-\frac{1}{2}}(k)\Big] \tag{4.28}$$

This explains for the fact that the last term of eq. (4.26) can be neglected, since both terms of eq. (4.28) are very small ($\approx 1/N$).
We thus obtain:

$$E\Big[\sum d_2^{2\,j+\frac{1}{2}}(k)\Big] = r(F,N)\Big\{E\Big[\sum\xi^2(k)\Big] + E\Big[\sum d_1^{2\,j}(k)\Big]\Big\} \tag{4.29}$$

Since the transition of $d_2^{j+\frac{1}{2}}$ to d_1^{j+1} is similar to the above, it holds

that:

$$E\Big[\sum_{k=1}^{N} d_1^{2j+1}(k)\Big] = r(F,N)\Big\{(1+r(F,N))E\Big[\sum_{k=1}^{N}\xi^2(k)\Big] + r(F,N).E\Big[\sum_{k=1}^{N} d_1^{2j}(k)\Big]\Big\} \qquad (4.30)$$

Assuming ergodicity and stationarity and applying the requirement for convergence (4.11) we see that:

$$E[\xi^2] < [1-r(F,N)]/r(F,N).E[d_1^{2j}] \qquad (4.31)$$

seems a sufficient condition for convergence of the estimation algorithm.

Remark:

Note that equation (4.31) also gives us an indication of the accuracy we may expect; with given variance of the noise, order of the system and available number of samples we obtain the following:

$$E[d_1^2] > \frac{r(F,N)}{1-r(F,N)}\,E[\xi^2] \qquad (4.32)$$

Clearly eq. (4.32) neglects the numerical aspects of the method; this result solely accounts for the accuracy of the applied principle.

4.4 The output error method

Using an output error method, the noise operating in a system is expected to be concentrated at the output of the system; see the process of fig. 4.8.

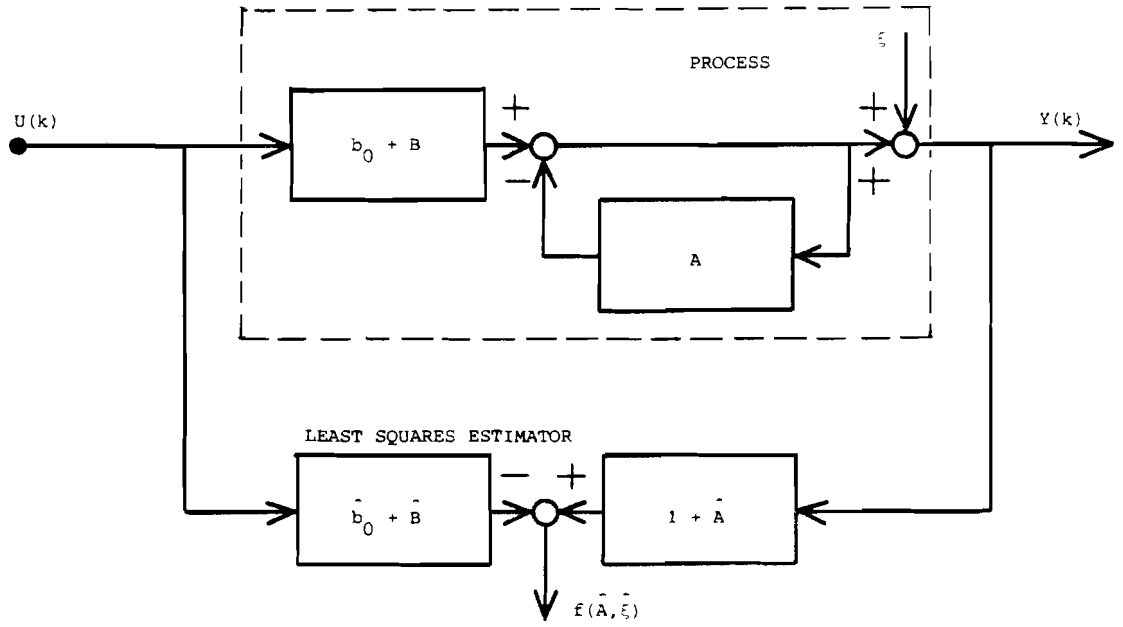


Fig. 4.8 An output error model of the process, with a least squares estimator.

For an ARMA model this can be described as:

$$y(k)= \frac{b_0+B(q^{-1})}{1+A(q^{-1})}u(k)+\xi(k) \qquad (4.33)$$

An output error method minimizes the quadratic loss-function constructed of the output error, with respect to the parameters of the model:
with $\hat{\xi} = \hat{y}-y$

$$\min V = \min \Sigma \hat{\xi}^2 \qquad (4.34)$$

Since the loss-function is not linear-in-the-parameters, the minimization has to be performed in an iterative way.
Introduce the function $f(A,\xi)$

$$f(\hat{A},\hat{\xi}) = \left[1+\hat{A}\right]\hat{\xi} \tag{4.35}$$

as is shown in fig. 4.8. Applying a simple least squares estimation will not deliver a consistent estimate, but since we know the structure of the function f, we can find a strategy to obtain an estimate $\hat{\xi}$ of ξ and minimize its square sum.

Thus we can derive a Taylor expansion developed around f^i, where we have the estimated parameters $\left[\hat{A}^i\right]$, $\left[\hat{b}_0{}^i+\hat{B}^i\right]$ and residual $\hat{\xi}^i$:

$$f^{i+1} = f^i + \left.\frac{\partial f}{\partial \hat{A}}\right|_i \left[A^{i+1}-A^i\right] + \left.\frac{\partial f}{\partial \hat{\xi}}\right|_i (\hat{\xi}^{i+1}-\hat{\xi}^i) \tag{4.36}$$

where f^{i+1} represents the function f with the estimated values in the $i+1^{st}$ iteration and f^i for the i^{th} iteration.

With

$$\left.\frac{\partial f}{\partial \hat{A}}\right|_i = \hat{\xi}^i(k) \tag{4.37-a}$$

$$\left.\frac{\partial f}{\partial \hat{\xi}}\right|_i = \left[1+\hat{A}^i\right] \tag{4.37-b}$$

$$f^{i+1}(k) = \left[1+A^{i+1}\right]\, y(k) - \left[\hat{b}_0{}^{i+1}+\hat{B}^{i+1}\right]\, u(k) \tag{4.37-c}$$

$$f^i(k) = \left[1+\hat{A}^i\right]\, \hat{\xi}^i \tag{4.37-d}$$

and some rearranging we obtain:

$$\hat{\xi}^{i+1}(k) = \frac{\left[1+\hat{A}^{i+1}\right]}{\left[1+\hat{A}^i\right]}\, y(k) - \frac{\left[\hat{b}_0{}^{i+1}+\hat{B}^{i+1}\right]}{\left[1+\hat{A}^i\right]}\, u(k) - \\ - \frac{\left[1+\hat{A}^{i+1}\right]}{\left[1+\hat{A}^i\right]}\, \hat{\xi}^i(k) + \hat{\xi}^i(k) \tag{4.38}$$

Now two possibilities arise to linearize this expression, in order to yield a least squares minimization procedure:

4.4.1 Quasi-linear method

First we rearrange eq. (4.38) to the subjoined expression:

$$\hat{\xi}^{i+1}(k) + [\hat{A}^{i}](\hat{\xi}^{i+1}(k) - \hat{\xi}^{i}(k)) =$$
$$[1+\hat{A}^{i+1}]y(k) - [\hat{b}_0{}^{i+1}+\hat{B}^{i+1}]u(k) - [\hat{A}^{i+1}]\hat{\xi}^{i}(k) \qquad (4.39)$$

If $\hat{\xi}^{i+1}(k) \approx \hat{\xi}^{i}(k)$ then the second term on the left hand side of eq. (4.39) can be discarded. We thus yield the following expression:

$$\hat{\xi}^{i+1}(k) = [1+\hat{A}^{i+1}]y(k) - [\hat{b}_0{}^{i+1}+\hat{B}^{i+1}]u(k) - [\hat{A}^{i+1}]\hat{\xi}^{i}(k) \qquad (4.40)$$

The loss-function $\Sigma\, \hat{\xi}^{2}{}^{i+1}(k)$ now seems quadratic-in-the-parameters, disregarding the fact that $\hat{\xi}^{i}(k)$ is also a function of the parameters. The non-linearity of the error ξ in the model is handled by means of a quasi-linearization technique. An iterative algorithm can be developed assuming that the estimated errors $\hat{\xi}^{i}$ from the i^{th} iteration are available. Initiating the algorithm with a least squares estimation will deliver the first estimated error $\hat{\xi}^{0}$. Minimization of the loss-function now with a least squares method will yield an explicit solution of the parameters and herewith $\hat{\xi}^{i+1}(k)$ can be updated iteratively. This is a quasi-Newton iterative method.

4.4.2 Steiglitz-McBride method

If $[\hat{A}^{i+1}] \approx [\hat{A}^{i}]$, then the last two terms on the right hand side of eq. (4.38) can be neglected.
This yields:

$$\hat{\xi}^{i+1}(k) = \frac{[1+\hat{A}^{i+1}]}{[1+\hat{A}^{i}]}\, y(k) - \frac{[\hat{b}_0{}^{i+1}+\hat{B}^{i+1}]}{[1+\hat{A}^{i}]}\, u(k) \qquad (4.41)$$

This can be interpreted as 'prefiltering' of the input u(k) and the output y(k) with the previous estimate of the autoregressive part:

$$\tilde{y}^i(k) = \frac{1}{\left[1+\hat{A}^i\right]} y(k) \qquad (4.42\text{-a})$$

and

$$\tilde{u}^i(k) = \frac{1}{\left[1+\hat{A}^i\right]} u(k) \qquad (4.42\text{-b})$$

Fig. 4.9 illustrates the prefiltering of both signals.

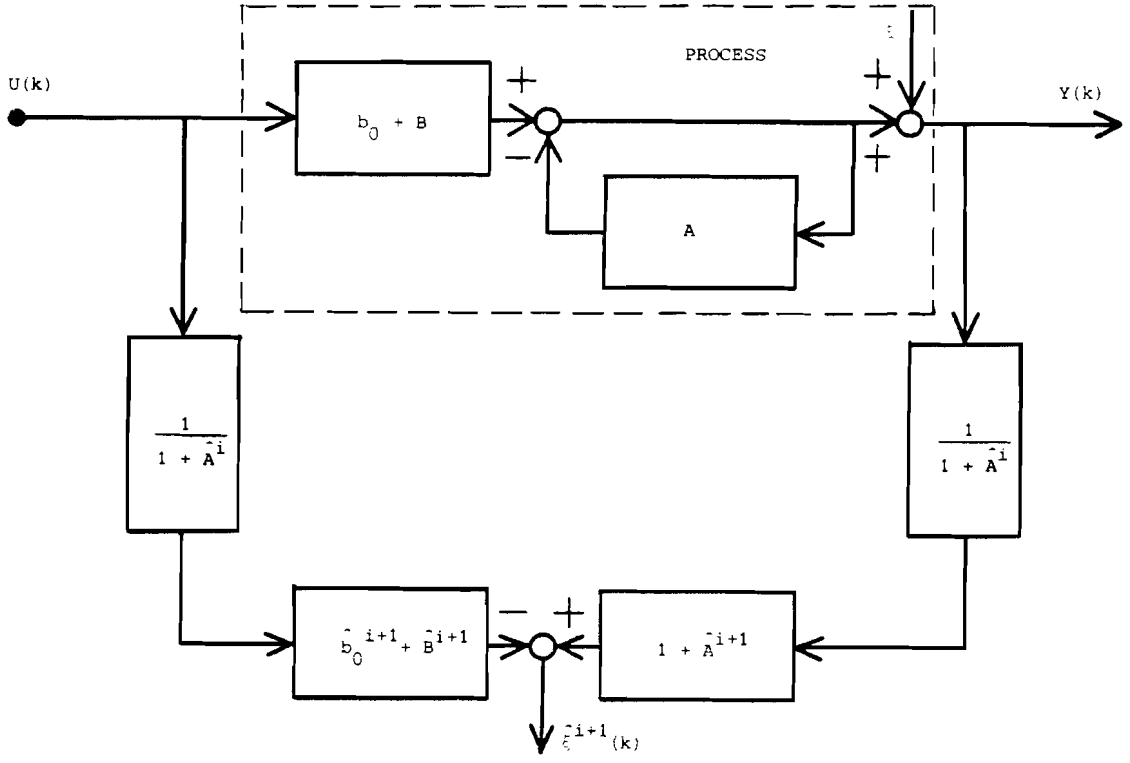


Fig. 4.9 The prefiltering of input signal and output signal as used in the Steiglitz-McBride Identification Algorithm.

After prefiltering the error $\tilde{\xi}^{i+1}(k)$ is linear-in-the-parameters and a least squares minimization of the loss-function $\Sigma\, \tilde{\xi}^{i+1}(k)$ yields an explicit solution of the parameters. With this solution the prefiltering can be updated and the estimation process can be repeated.
This iterative algorithm is known in literature as the Steiglitz-McBride Identification Algorithm; cf. Steiglitz and McBride (1965).

4.4.3 Recapitulation

The first method seems to be advantageous, since the elaborate prefiltering of the second method is not required.
However, simulation results have lead to the conclusion that the first method appears to have poor convergence properties. Apparently the supposition that $\hat{\xi}^{i+1}(k) \approx \hat{\xi}^{i}(k)$ does not always hold, especially the initial values will differ substantially.
The second method is proven to be convergent to the true parameter vector if the additive output noise is white; cf. Stoica and Söderström (1981). Since the alternating iterative estimation method (cf. subsection 4.2.3) converges and additional white noise is assumed at the output, the Steiglitz-McBride method seems an appropriate choice. The more the iterative process proceeds, the 'whiter' the additional noise for the algorithm gets, for the 'coloured' component of the other SISO subsystem misfit decreases.

4.5 A termination criterion

The foregoing merely concerns the estimating part of the method, but since the estimation is performed in an iterative way, the importance of detecting the progress of the algorithm can not be denied.
The termination criterion detects the progress, or at least a measure for it, and when a certain previously adjusted threshold level is exceeded, the iterative loop in the algorithm is terminated.
For the method under study, a distinction can be made between the iterative algorithm of the SISO output error method, and the alternating iterative MISO algorithm, in which the former algorithm will be implemented as a subroutine.

For the output error algorithm, the loss-function as defined in eq. (4.31) well reflects the progress of the iterative process. Therefore the termination criterion here is based on the relative decrease of the loss-function; if the decrease is small enough, the method is assumed to be close enough to the minimum value attainable and the iteration is stopped.
Practical application of the Steiglitz-McBride algorithm demonstrates that to a certain extent divergence of the iteration can occur. Apparently the estimation process acts less sensitive close to the presumed minimum. Therefore in case of an increase of the loss-function the criterion algorithm acts accordingly by restoring the previous situation and terminating the iteration.

For the alternating MISO algorithm, three ways have been considered to take a measure of the presumed progress:

- The number of iterations; this merely reflects the computational effort performed; in general no indication of attained accuracy can be derived.
  This criterion is used to restrict the maximum computation time, which can be substantial in case of slow convergence.
- The loss-function of the two SISO estimations; when both values, within certain limits, draw level with each other, then no appreciable progress will be made in succeeding iterations. The criterion can be described with:

$$C_2^{i+\frac{1}{2}} = \sum_{k=1}^{N} \hat{\xi}_1^{2\,i}(k) - \sum_{k=1}^{N} \hat{\xi}_2^{2\,i+\frac{1}{2}}(k) \qquad (4.43)$$

- Define the loss-function:

$$V^{i+\frac{1}{2}} = \sum_{k=1}^{N} (\hat{\xi}_1^{\,i}(k) - \hat{\xi}_2^{\,i+\frac{1}{2}})^2 \qquad (4.44)$$

With $\hat{\xi}_1^{\,i}(k)$ the resulting reconstruction error of the i^{th} estimate of the first SISO subsystem and $\hat{\xi}_2^{\,i+\frac{1}{2}}(k)$ the reconstruction error corresponding with the $i+\frac{1}{2}^{th}$ estimation of the second SISO subsystem. Since both errors converge asymptotically to the same value $\xi(k)$ this function will be decreasing to zero in case of convergence. When this function is small enough, then no considerable progress will be made by continuing the iterative process.

For practical purposes, the two last mentioned criteria are always applied in combination with the first criterion, in order to deal with the imaginary case of utterly slow convergence.
In the following the properties of the last two criteria will be treated more thoroughly.

Properties of the criteria

In the previous section I have shown, that a reconstructed signal of the $i-\frac{1}{2}^{th}$ estimation of the second SISO subsystem can be denoted as:

$$\hat{v}_2^{\,i-\frac{1}{2}}(k) = v_2(k) + d_2^{\,i-\frac{1}{2}}(k) \qquad (4.45)$$

With $d_2^{\,i-\frac{1}{2}}(k)$ only dependent on the input signal $u_2(k)$.
The following estimation of the first SISO subsystem is then performed with the input and output data $u_1(k)$ and $y(k) - \hat{v}_2^{\,i-\frac{1}{2}}(k)$.
This results in a reconstructed signal:

$$\hat{v}_1^{\,i}(k) = v_1(k) + d_1^{\,i}(k) \qquad (4.46)$$

With $d_1^{\,i}(k)$ only dependent on the input signal u_1.
The resulting reconstruction error can thus be denoted as follows:

$$\hat{\xi}_1^{\,i}(k) = \{y(k)-\hat{v}_2^{\,i-\frac{1}{2}}(k)\} - \hat{v}_1^{\,i}(k) \tag{4.47}$$

With equations (4.20), (4.45) and (4.46) we obtain:

$$\hat{\xi}_1^{\,i}(k) = \xi(k) + d_1^{\,i}(k) + d_2^{\,i-\frac{1}{2}}(k) \tag{4.48}$$

Similarly the reconstruction error corresponding with the next second SISO subsystem estimation can be denoted as:

$$\hat{\xi}_2^{\,i+\frac{1}{2}}(k) = \xi(k) + d_1^{\,i}(k) + d_2^{\,i+\frac{1}{2}}(k) \tag{4.49}$$

Fig. 4.10 illustrates this.

The second termination criterion comprises the following:
Eq. (4.43) describes the criterion. With N, the number of samples, large enough, there are no objections against approaching this with statistical expectations:

$$C_2^{\,i+\frac{1}{2}} \approx N\{E[\hat{\xi}_1^{\,2i}(k)] - E[\hat{\xi}_2^{\,2i+\frac{1}{2}}(k)]\} \tag{4.50}$$

Invoking equations (4.48) and (4.49) and the fact that ξ, d_1 and d_2 are statistically independent, this yields:

$$C_2^{\,i+\frac{1}{2}} \approx N\{E[d_2^{\,2i-\frac{1}{2}}(k)] - E[d_2^{\,2i+\frac{1}{2}}(k)]\} \tag{4.51}$$

Hence when C_2 gets small, this implies that the variance of d_2, the difference signal of the second SISO subsystem estimation (cf. eq. (4.45)), has an insignificant decrease in value in successive iterations.
The magnitude of C_2 informs about the improvement of the estimate of the second SISO subsystem that we might expect in future iterations. Clearly C_2 merely indicates the progress of the second SISO subsystem estimate. Therefore we can also define:

$$C_1^{\,i} = \sum_{k=1}^{N} \hat{\xi}_2^{\,2\,i-\frac{1}{2}}(k) - \sum_{k=1}^{N} \hat{\xi}_1^{\,2\,i}(k) \qquad (4.52$$

$C_1^{\,i}$ similarly is a measure for the improvement of the first SISO system estimate.

Invoking equations (4.48) and (4.49) in the third termination criterion (cf. eq. (4.44)), we obtain the following:

$$V_2^{\,i+\frac{1}{2}} = \sum_{k=1}^{N} \left\{ d_2^{\,i-\frac{1}{2}}(k) - d_2^{\,i+\frac{1}{2}}(k) \right\}^2 \qquad (4.53)$$

With N large enough, we can introduce the statistical expectation:

$$V_2^{\,i+\frac{1}{2}} \approx N.E\left[\left\{ d_2^{\,i-\frac{1}{2}}(k) - d_2^{\,i+\frac{1}{2}}(k) \right\}^2 \right] \qquad (4.54)$$

This indicates that V_2 represents the variance of the difference between successive difference signals of the second SISO subsystem estimations. Similar to the second criterion, the magnitude of V_2 informs about the gain that we might expect for the second SISO subsystem estimate in future iterations.
For V_2 merely concerns the second SISO subsystem estimation, we also require:

$$V_1^{\,i} = \sum_{k=1}^{N} \left\{ \hat{\xi}_2^{\,i-\frac{1}{2}}(k) - \hat{\xi}_1^{\,i}(k) \right\}^2 \qquad (4.55)$$

Consequently V_1 being a measure for the amelioration of the first SISO subsystem estimate.

Comparing the second and third criterion we conclude that C, the second criterion, requires less computational effort, since both loss-functions are already available, while V requires additional calculation. On the other hand, V gives a more precise indication of the improvement, for it calculates the variance of the difference, while C calculates the difference of the successive variances.

4.6 Recapitulation

The foregoing sections treated the development of the proposed MISO estimation method.

The method comprises an alternating estimation of the MISO process, by applying a Steiglitz-McBride output error minimization method alternately to the first and the second SISO subsystem. This is done in such a way that each successive output error minimization is applied to the previously rearranged data. This previous rearrangement is performed by subtracting the reconstructed output signal of the other subsystem from the given output data. Fig. 4.10 illustrates the procedure described above.

Section 4.5 describes two ways of defining a justifiable criterion to terminate the iterative estimation process. These methods are both based on the behaviour of different quadratic loss-functions.

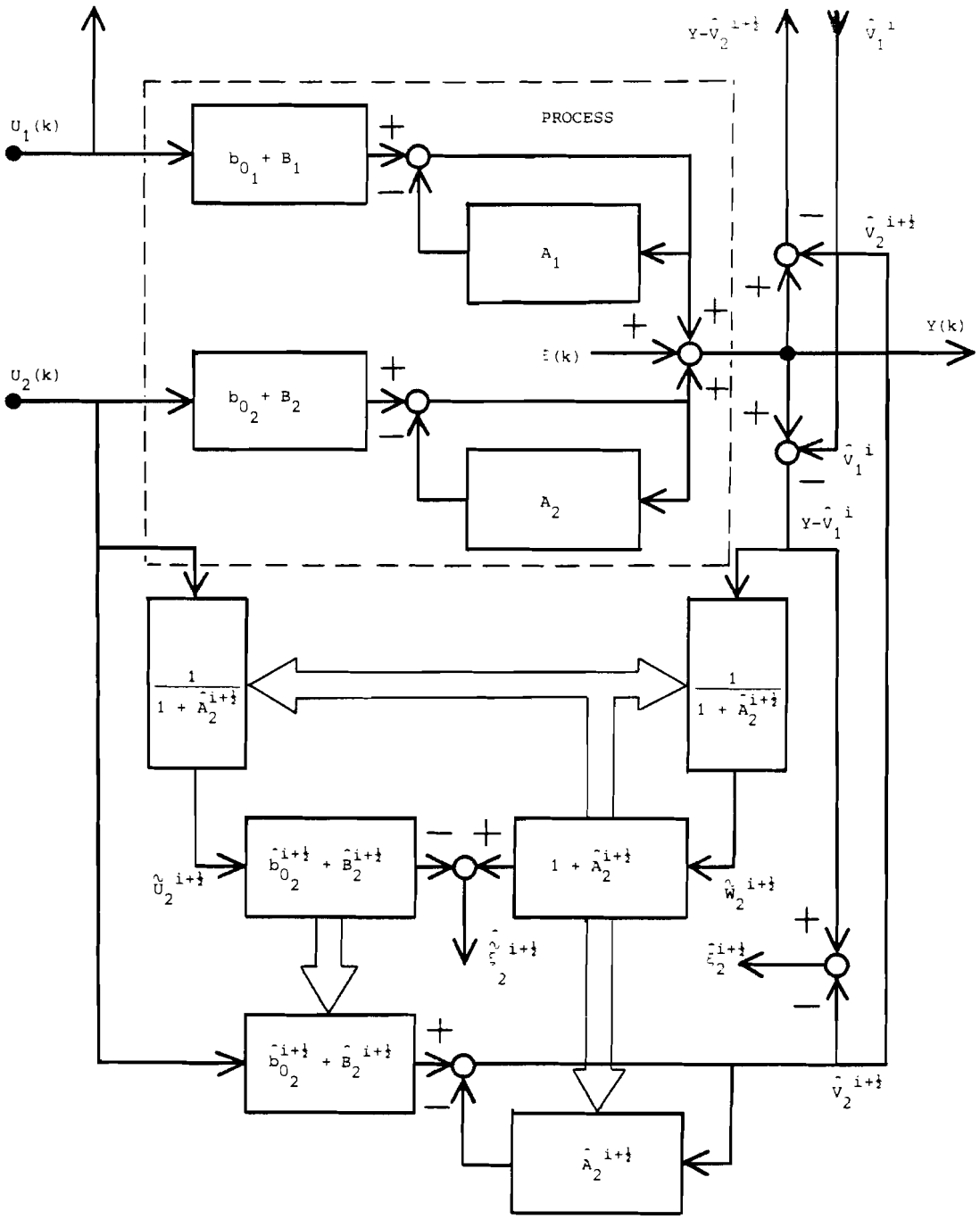


Fig. 4.10 Estimation scheme of the second SISO subsystem, the upper half of the scheme has been omitted for reasons of clarity.

5. Realization of the method

5.1 Implementation

The preceding chapter deals with the basic ideas concerning the development of method based on the alternating separate estimation of the two SISO subsystems. This section describes briefly how this algorithm is implemented in Fortran-77 on a VAX-11 mini-computer system.
The alternating MISO estimation is organized in the subroutine MSLSQE. This subroutine delivers the estimates of the two SISO subsystems, given a certain datafile containing the input and output samples (u_1,u_2,y). Further the routine requires a number of parameters to perform the estimation in a proper way. Fig. 5.1 represents the flow diagram of this routine.
The logical function BRKOFF realizes the termination criteria as mentioned in the foregoing section, by means of a 'mode' argument variable a criterion can be chosen.
The subroutine OELSQE performs the Steiglitz-McBride output error minimization, as pointed out in Sect. 4.4, and delivers the parameter vector of each SISO subsystem. Fig. 5.2 represents the flow diagram of this subroutine.
Since the MISO estimation is implemented in a subroutine form, application is relatively simple in a main program.
For test purposes the interactive program MISO has been developed. This program enables the user to enter a MISO datafile in THE-structure, select the proper features for the estimation and obtain graphical files for representation of the results.

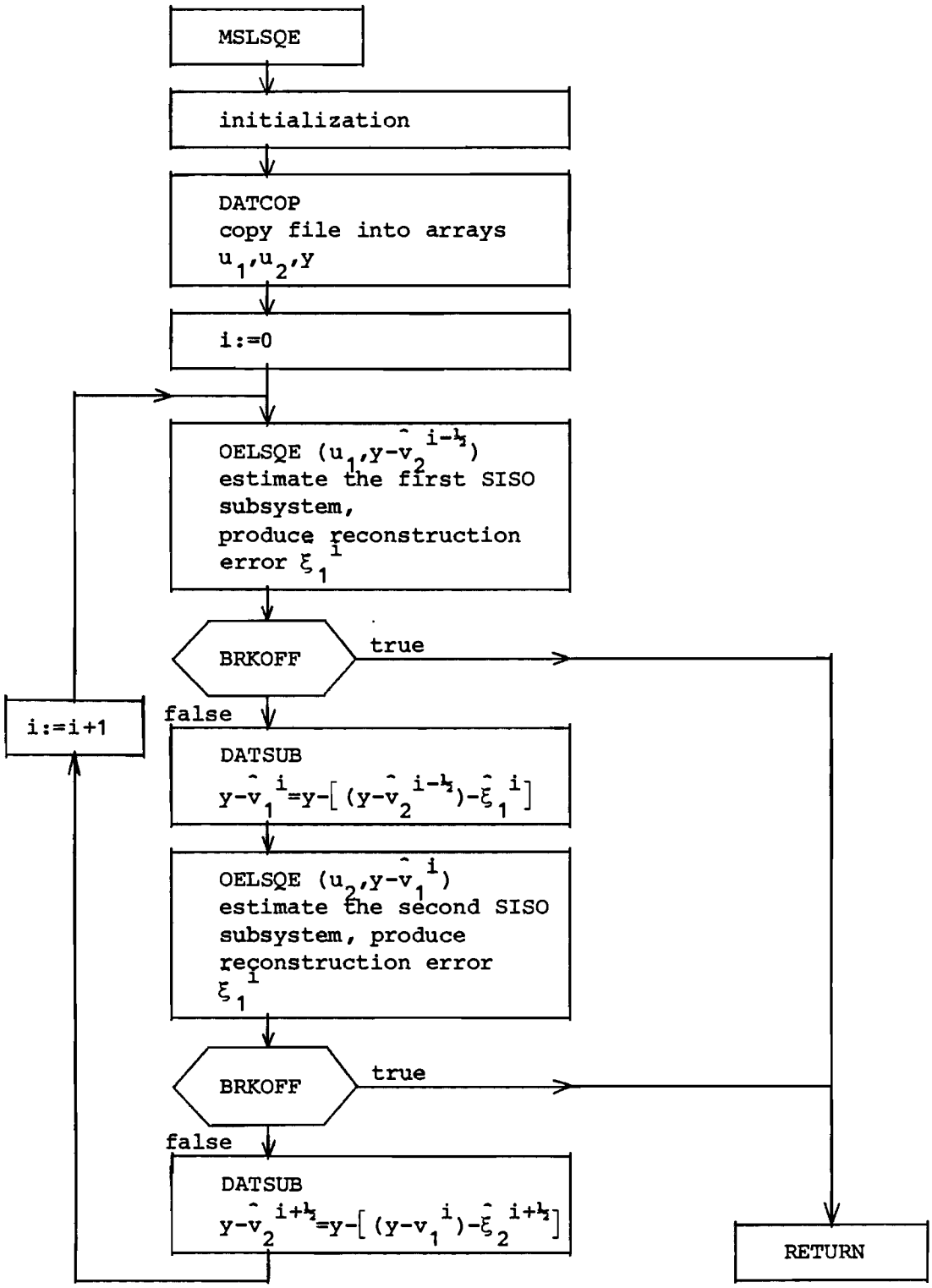


Fig. 5.1 Flow diagram of the MISO estimation routine

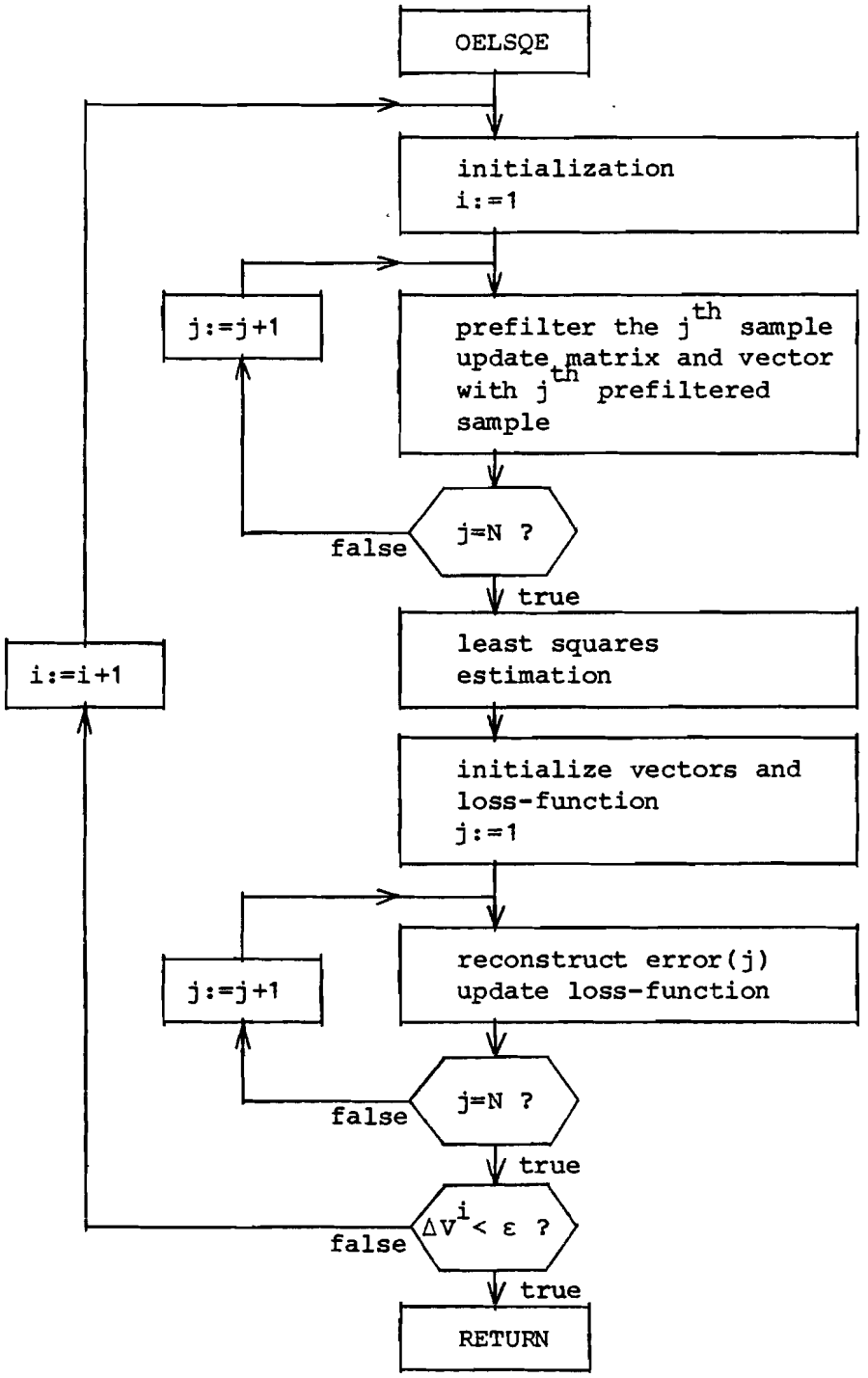


Fig. 5.2 Flow diagram of the SISO output error minimization routine

5.2 Simulation of systems

The interactive program SIMUL, derived from SYSIMUL, facilitates the simulation of linear processes. By means of the program package MATLAB a file must be created, containing a state-space description of the desired process. This file is used in SIMUL to describe the process. A state-space description of an ARMA representation can be given as undermentioned.
Given the ARMA process:

$$v(k) = b_0u(k)+\ldots+b_{n_b}u(k-n_b)-a_1v(k-1)-\ldots-a_{n_a}v(k-n_a) \quad (5.1)$$

the state-space description can be denoted as:

$$\underline{x}(k+1)= A\,\underline{x}(k) +\underline{b}\,u(k) \quad (5.2\text{-a})$$
$$v(k) = \underline{c}^T\underline{x}(k) +d\,u(k) \quad (5.2\text{-b})$$

With $n \equiv \max(n_a,n_b)$ we find:

$$A = \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdot & \cdot & & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_n & -a_{n-1} & \cdots & -a_2 & -a_1 \end{bmatrix} \quad (5.3\text{-a})$$

$$\underline{b}^T= \begin{bmatrix} h_0 & \cdots & h_n \end{bmatrix} \quad (5.3\text{-b})$$

h_i, i=0,...,n is defined as:

$$\begin{aligned} h_0&=b_0 \\ h_1&=b_1-a_1h_0 \\ &\vdots \\ h_n&=b_n-a_1h_{n-1}-\ldots-a_{n-1}h_1-a_nh_0 \end{aligned} \quad (5.3\text{-c})$$

$$\underline{c}^T = [\ 1\ 0\ .\ .\ .\ .\ .\ 0\ ] \qquad (5.3\text{-d})$$

and

$$d = [h_0] \qquad (5.3\text{-e})$$

(also cf. Ogata, 1970)

A MISO system can thus be described as:

$$\underline{x}(k+1) = \begin{bmatrix} A_1 & \emptyset \\ \emptyset & A_2 \end{bmatrix} \underline{x}(k) + \begin{bmatrix} \underline{b}_1 & \underline{0} \\ \underline{0} & \underline{b}_2 \end{bmatrix} \underline{u}(k) \qquad (5.4\text{-a})$$

$$y(k) = [\ \underline{c}_1^T \underline{c}_2^T\ ]\ \underline{x}(k) + [\ d_1\ d_2\ ]\ \underline{u}(k) + \xi(k) \qquad (5.4\text{-b})$$

with $\underline{u}^T(k) = [u_1(k)\ u_2(k)]$, $\xi(k)$ the output disturbance and the subscripts 1,2 corresponding with the respective SISO subsystems.

5.3 Simulation results

In order to know more about the developed method, a number of tests have been performed with the program.

The test system was chosen as follows:

Table 5.1 configuration of test system

	first subsystem	second subsystem
b_0	1.0	2.5
b_1	0.7	1.25
b_2	0.3	0.1
a_1	-1.5	-1.0
a_2	0.7	0.5

Both SISO subsystems of ARMA(2,2) structure have poles not too close to the unit circle, also cf. fig. 5.3-a and -b.

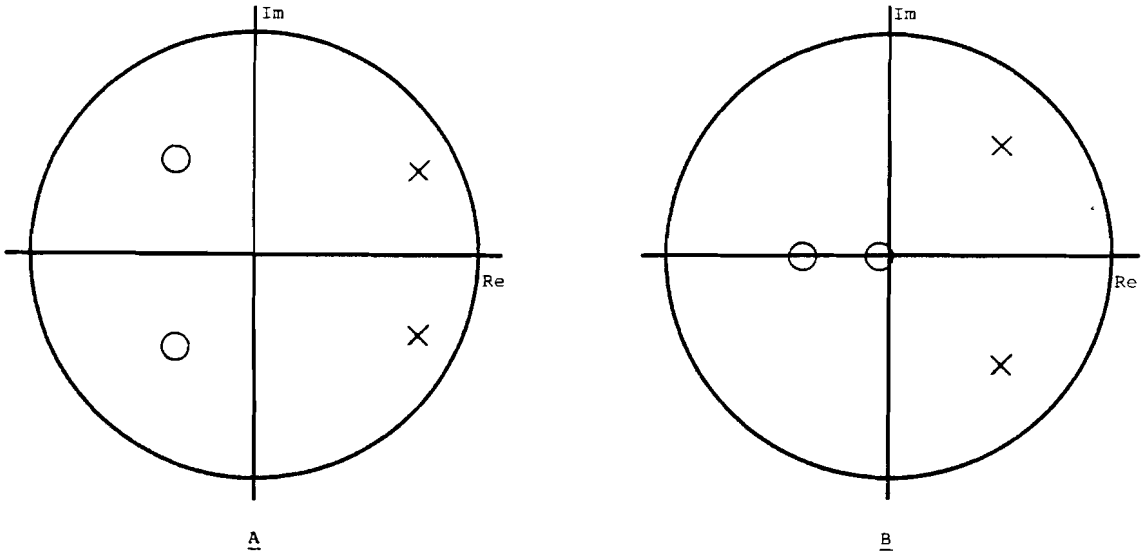


Fig. 5.3 Pole-zero plots of both SISO subsystems;
a first subsystem
b second subsystem

The SISO subsystems are chosen in such a way, that the output energy contents of both systems will be equal in case of normal white noise excitation of equal power at both inputs.
The experiments that are performed relate to the following subjects:

- general convergence of the method
- convergence for unequal energy contents
- behaviour for a non-white disturbance
- behaviour for a non-white input
- behaviour for dependent input signals
- behaviour for a wrong model order

5.3.1 General convergence

The performance of the method has been examined for the case described in Sect. 4.3. Here convergence can be shown under the assumption that the input signals are independent and 'white', the additive output noise is also 'white' and independent of the two input signals.
The estimation has been performed over 500 data samples for several signal-to-noise-ratios and the results are presented in Table 5.2.
The signal-to-noise-ratio is defined as:

$$SNR \equiv 20 \log \left[(\overline{v_1+v_2}) / \bar{\xi} \right] \tag{5.5}$$

with $(\overline{v_1+v_2})$ the RMS value of the non-disturbed output and $\bar{\xi}$ the RMS value of the additive output disturbance.

The results of this experiment show that in the general case the method performs as expected. The first column of Table 5.2 indicates that the method results in an unbiased estimate in the absence of additive disturbances.
As can be seen in the next three columns, the resulting estimate grows worse as the signal-to-noise-ratio grows worse. Clearly the estimation at low signal-to-noise-ratios can be improved by estimating over a larger number of samples, but this has not been done for comparative reasons. Note that this statement is relative, for an increase of the number of samples will result in an increase of the numerical inaccuracy. (also cf. Zhu, 1984).
Table 5.2 also shows the variance of the reconstruction error (the loss-function). This quantity shows an increase with a factor 10 for each decrease of the signal-to-noise-ratio with 10 dB.
Finally Table 5.2 shows that the total number of output error iterations performed by the algorithm does not differ substantially for different signal-to-noise-ratios. Note however that this number depends on the imposed accuracy. In this case the imposed threshold values for the termination criteria (cf. Sect. 4.5) were 0.0001 and 0.0005 respectively for the Steiglitz-McBride output error method and the MISO estimation method (cf. eq. (4.43)).

5.3.2 Convergence for unequal energy contents

Due to unequally distributed energy over the outputs of both SISO subsystems the estimation is expected to become less reliable for unfavourable signal-to-noise-ratios. Here the question arises whether the choice of the first subsystem (first in the sense of the estimation sequence) influences the estimation result. To investigate this, two experiments have been performed.
The first experiment comprises the case in which the SISO subsystem that is estimated at first in the algorithm has the highest energy output. The moving-average parameters b_0, b_1 and b_2 of the first SISO subsystem have been tripled for the simulation, in order to get an output signal three times as strong as the second output signal. Thus the $signal_1$-to-$signal_2$-ratio is almost + 10 dB. With this simulated system estimations have been performed under the same conditions as mentioned in subsection 5.3.1. The results of this experiment are presented in Table 5.3.

The second experiment comprises the case where the second SISO subsystem has the highest energy output. To establish this the sequence of estimation of the system that is used in the first experiment is altered. Thus the $signal_1$-to-$signal_2$-ratio is almost -10 dB. With this simulated system estimations have been performed under the conditions of subsection 5.3.1. The results of this experiment are presented in Table 5.4.

A comparison of the results of both experiments shows that in the case of a $signal_1$-to-$signal_2$-ratio of -10 dB, the estimates of both subsystems do not differ substantially in accuracy from the results of the general case (see subsection 5.3.1). However, in the case of a $signal_1$-to-$signal_2$-ratio of +10 dB, the estimates of the second subsystem differ more in accuracy. The estimates of the first subsystem are of the same accuracy as the results of the general case (see subsection 5.3.1).
Tables 5.3 and 5.4 also show the variance of the reconstruction error (the loss-function) and the total number of output error iterations performed.
Note that the number of iterations is quite different for several different signal-to-noise-ratios.
A comparison of the loss-functions (i.e. the variance of the

reconstruction error) of Tables 5.3 and 5.4 shows that the asymmetry in estimation performance only shows up for the worst signal-to-noise ratios; for 10 dB and 0 dB the loss-function is smaller in the second experiment.
From these two experiments we can conclude that the sequence of estimation can influence the result of the estimation considerably. This phenomenon occurs in the case of less favourable signal-to-noise-ratio conditions, when the SISO subsystem that is to be estimated at first has a substantially higher output signal energy than the second SISO subsystem. If this occurs, the remedy is to alter the sequence of estimation, but in the case of practical data the phenomenon can be difficult to recognize.

5.3.3 Behaviour for non-white disturbance

As is mentioned in subsection 4.4.3, the applied output error method can converge to the true parameter vector only when the additive output noise is white. In that case the method is proved to be locally convergent to the true parameters (cf. Stoica and Söderström, 1981).
The previous two sections dealt with the case of additive white output noise, but since the case of a non-white additive disturbance seems of practical interest, the estimation performance has been examined. For this purpose a first order ARMA(1,1) filter has been chosen to produce the non-white noise. The filter has the following parameter values:

moving average parameters $c_0 = 1.0$ $c_1 = 0.3$
autoregressive parameter $d_1 = -0.5$

Fig. 5.4-a shows the pole-zero plot of this filter and fig. 5.4-b shows the autocorrelation of the output with white noise excitation.

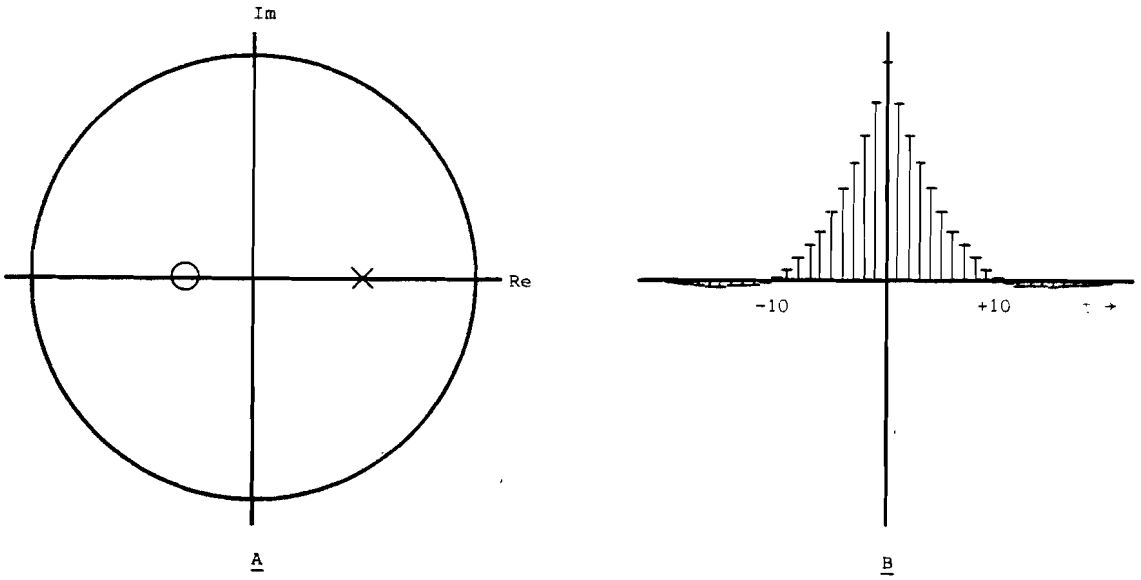


Fig. 5.4 a Pole-zero plot of the noise filter
b Autocorrelation of the disturbance

The estimation has been performed over 500 data samples for several signal-to-noise-ratios. The imposed accuracy is taken the same as in the previous experiments, but it should be noted that in cases of malconvergence of the output error method the algorithm is forced to restore the preceding estimate. This can effect a premature termination of the algorithm, in case of malconvergence of the estimation for both subsystems.

The results of the experiment are presented in Table 5.5.

A comparison with the results in Table 5.2 for additive white noise shows that the estimates are more biased, due to the non-whiteness of the disturbance. Furthermore the decreasing number of performed output iterations manifests that in the case of lower signal-to-noise-ratios malconvergence of the output error algorithm occurs. This contrasts with the case of a white disturbance, where convergence maintains.

The variance of the reconstruction error shows that the algorithm is still capable of modelling a sufficient part of the information contained by the data sequence.

Note that this experiment is just an indication for the behaviour in case of non-white disturbances. The non-whiteness of the applied disturbance is assumed to have general characteristics, but the experiment is not claimed to be general for a variety of non-white disturbances.

5.3.4 Behaviour for non-white input signals

For identification purposes it is advantageous to have input data with white noise properties. This way all frequency modes of the system are being excited equally. With two independent input signals the estimation performs well. In practice, however, the input normally will not have such attractive (from estimation point of view) properties; the input signals will not be 'white'.
To examine the method under study for such conditions an experiment has been done with a filtered input signal. Both inputs are excited with signals with independent white noise properties, filtered by the ARMA(1,1) filter mentioned in subsection 5.3.3.
Again the estimation has been performed over 500 samples for several signal-to-noise-ratios. The results are presented in Table 5.6.

The results of this experiment show that independent non-white input signals effect a deterioration of the estimates. Compared with the results of the estimation for white input excitation (cf. Table 5.2) we see that the estimates are considerably more biased. The experiment demonstrates that input signals with non-white properties severely worsen the estimation results. This is not considered to be a specific disability of the applied method; in general the non-whiteness of the input signals results in less accurate estimates.
This effect is known in literature and can be explained with the Cramér-Rao lower bound; non-white signals result in a covariance matrix (or Fisher information matrix) with non-zero off-diagonal terms. Since the covariance matrix is used to define a certain confidence region for the parameter estimate (cf. Bard, 1974), the achievable accuracy of the estimate will decrease. The so-called concentration ellipsoid, describing the confidence region, will cover a larger area than the optimal sphere for a 'white' input signal (also cf. Eykhoff, 1974).

5.3.5 Behaviour for dependent input signals

One of the assumptions made in Section 4.3 to show convergence was independence of both input signals. Clearly this is a very important restriction, discarding of which can lead to erroneous results.
To examine this issue, an experiment has been done. The dependence of the input signals has been realized by filtering input signal 2, which has white noise characteristics, and using this result as input signal 1. Care has been taken to assure that the output signals of both subsystems have equal energy contents.
The experiment has been performed with two different filters. The first filter that has been used is the ARMA(1,1) filter mentioned in subsection 5.3.3. Fig. 5.5-a shows the cross-correlation of the inputs that is achieved with this type of filtering. The estimation results corresponding with this filter are presented in Table 5.7.
The second filter that has been used is a time-shift of 5 sample times. This way both inputs have white noise properties, but are not independent. Fig. 5.5-b shows the corresponding cross-correlation of the two input signals.

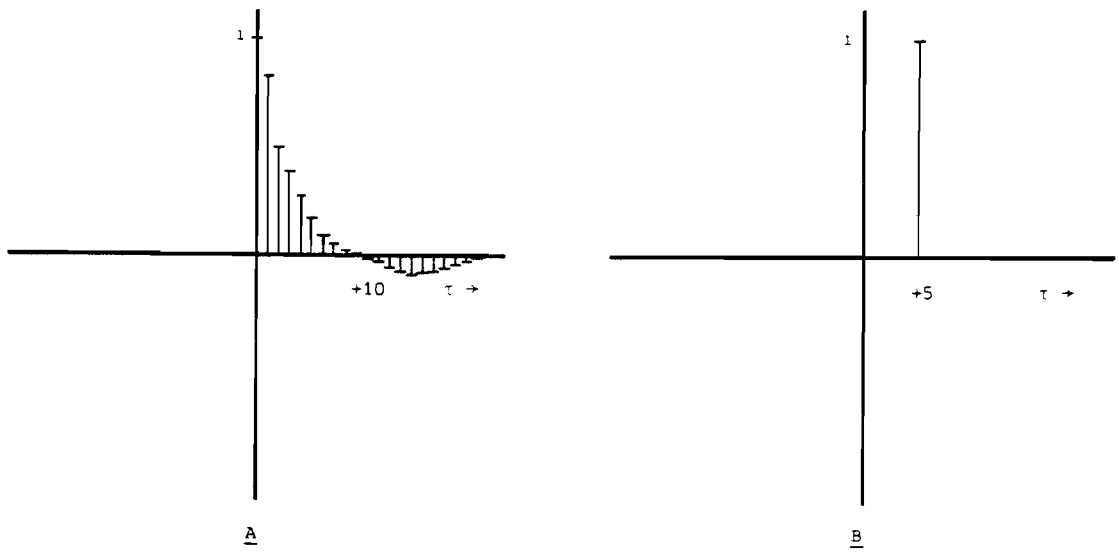


Fig. 5.5 Cross-correlations of the two input signals
a ARMA filtered
b time shifted

The estimation results for this filter are presented in Table 5.8.

Both the results of Table 5.7 and the results of Table 5.8 show that independence of the input signals is quite an important restriction for application of the method.
The estimates demonstrate that in this case the algorithm converges to a biased parameter vector. Even the noise-free case is afflicted with a considerable bias for both types of filtering of the input signals. The estimate with the time shifted 'white' input signals tends to a better result (Table 5.8); perhaps this can be explained by the non-whiteness of the ARMA filtered input signal, which also deteriorates the estimate. A peculiar effect is that the variance of the reconstruction error does not decrease significantly when the signal-to-noise-ratio increases above 30 dB. Furthermore the number of performed output error iterations is excessive for several cases (namely the 'better' signal-to-noise-ratios). This has lead to exceeding of the imposed maximum number of alternating iterations (500 iterations). Thus for these cases the imposed accuracy is not achieved (see Table 5.7 20 dB, 30 dB, 40 dB and the noisefree case). A contrast form the results for 0 dB and 10 dB, where the malconvergence leads to premature termination and very small numbers of iterations.
The above results show that the independence of the input signals should always be checked in practical cases, since dependence will lead to erroneous estimation results.

5.3.6 Wrong model order

Convergence of the algorithm has been shown for the correct order of the estimation model. Since in the case of practical data the order of the system is not known, but also can be a potential subject for identification, the behaviour of the method for a different model order is investigated.
The 'true' order is still 2, since the method is applied to the same ARMA(2,2) simulated data.
The first experiment comprises the estimation with too low an order, namely model order = 1. The results of this experiment are presented in Table 5.9. Note that the small number of performed output iterations indicate malconvergence of the algorithm and a premature termination of the program (also cf. subsection 5.3.3).
Clearly this model will not be capable of describing the whole system, so

the loss-function remains quite large.

The second experiment concerns the estimation of the system with too high an order, namely model order = 3. Table 5.10 represents the results of this experiment.
These numerical results can not be compared with the estimations for the 'true' order, due to the freedom of the algorithm to put the pole-zero cancellation anywhere in the z-plane. However, pole-zero representations of the estimates show that the estimates of order 3 describe the system quite well. Fig. 5.6-a to 5.6-l show these pole-zero plots of the estimates of the third order.
The superfluity of the third order is dealt by one pole-zero cancellation in each model of the subsystems. Note that the pole-zero cancellation takes place almost exactly until 10 dB, thus being a quite reliable method for selection of the order.

The third experiment concerns the estimation of the system with the fourth order. Table 5.11 represents the results of this experiment. Fig. 5.7-a to 5.7-l show the corresponding pole-zero representations of the estimates. Note that the pole-zero cancellation has a tendency to occur on or close to the unit circle (this applies only for higher signal-to-noise-ratios).
Estimation results of higher order modelling confirm the tendency of the algorithm to put the pole-zero pair near the unit circle. This implies that too high order modelling possibly leads to unstable models, since the pole-zero pair will not cancel exactly under worse noise conditions.

5.3.7 Recapitulation

Summarizing we can say that in general the estimation method under study performs well. Simulation results have demonstrated that the method can be applied in less restrictive cases than was assumed beforehand.
In case of non-white additive output disturbances or non-white input signals the estimation converges to slightly biased results. In case of energy unequally distributed over both subsystems the estimation can be improved by changing the sequence of the estimation, i.e. the subsystem with the weakest output signal should be estimated first.

If the input signals are dependent, care must be taken since the estimation results will be erroneous, therefore application of the method is dissuaded in such cases.
A crucial point is also the choice of the correct model order. Simulation results show that the occurance of unstable models is conceivable for too high orders.

Remark:

Note that the separate estimations of the two subsystems deal with a worse signal-to-noise-ratio than is expressed with the signal-to-noise-ratio of the whole MISO system.
In the case of energy equally distributed over both subsystems this results in a 3 dB drop of the signal-to-noise-ratio for the separate subsystems, since each subsystem deals with the total amount of additive noise, while the output energy of each subsystem is smaller than the sum of both output signal energies.

6. Application of the method

6.1 Introduction

The method of alternating estimation of SISO subsystems, as described in Chapter 4, has been tested extensively on simulated data in the previous chapter.
However, one must not forget that the eventual object of the method is to identify the parameters of a real process with real data of input and output signals.

Glass-tube manufacturing process

Practical data was available, recorded at the Philips glass production plant of Winschoten. Here manufacturing takes place of quartz tubes, that are being used in halogene lamps.
Fig 6.1 illustrates this production process.

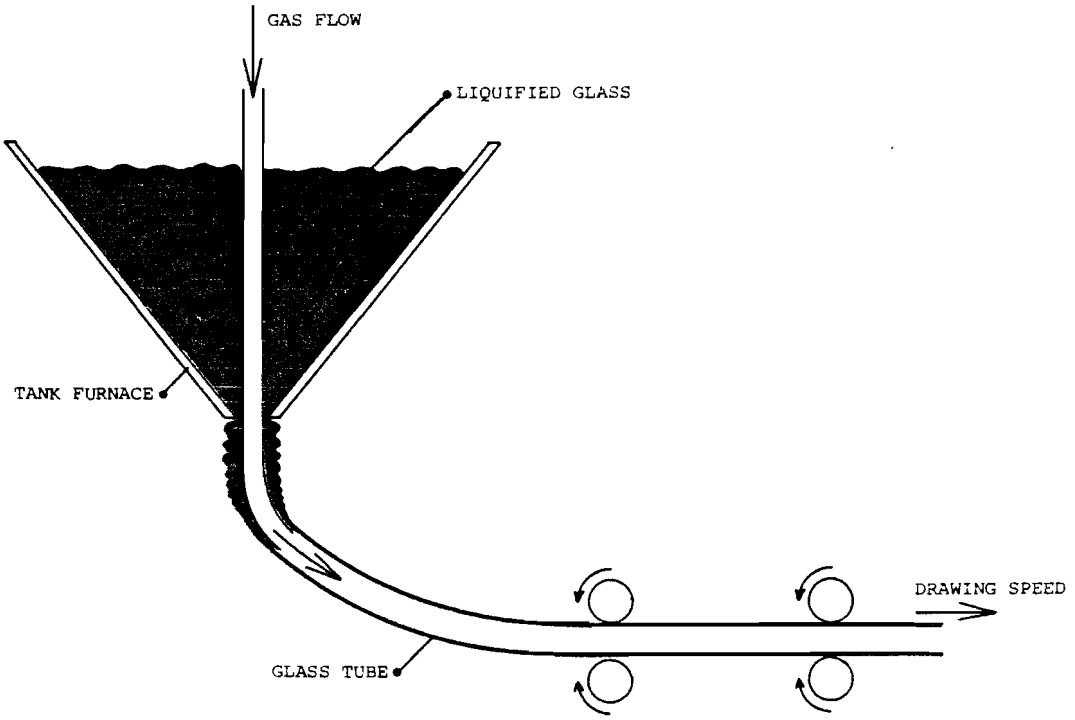


Fig. 6.1 Glass-tube manufacturing process

The basic materials for the glass are being melted in a tank furnace. Through an orifice in the bottom the liquified glass leaves the tank,

forced by gravity. Then through a small tube, located in the middle of the orifice, a special gas is blown, in order to provide the outcoming solidifying glass a cylindrical form. The solid glass tube is pulled, so that its drawing speed and therefore its physical proportions can be controlled.
This way the process can be regarded as a two-input-two-output system, with:

input 1 : gas flow
input 2 : drawing speed
output 1 : wall thickness of quartz tube
output 2 : diameter of quartz tube

Other quantities that might influence the process, such as temperature of glass and ambience, humidity etc., are considered to be disturbances (cf. Carrière, 1984).
Note that the temperature of the glass and the ambience clearly influences the process, since they are decisive for the solidity (or viscosity) of the glass and therefore influence the sensitivity of the process towards the control inputs as gas flow and drawing speed.
Data of the process was available from file SEL003.DAT (in PICOS format). The sample frequency was 2.5 Hz and the 10500 samples were corrected for trends.

6.2 Extension to MIMO systems

The MISO estimation algorithm as described in the preceding chapters can be applied to a MIMO system under certain assumptions.
Assume that the system can be divided into separate MISO systems without loss of generality. In other words, the output signals are not supposed to mutually influence each other; the structure is assumed to be non-nested.
Application of the method has been realized by implementing the subroutine MSLSQE, described briefly in Chapter 5, in program AMESA. the options that are implemented in this program, such as a determination of the time delay in the process and a determination of the order of the process, are discussed in subsections 6.4 and 6.5 respectively.
Fig. 6.2 illustrates the operation of the program AMESA.

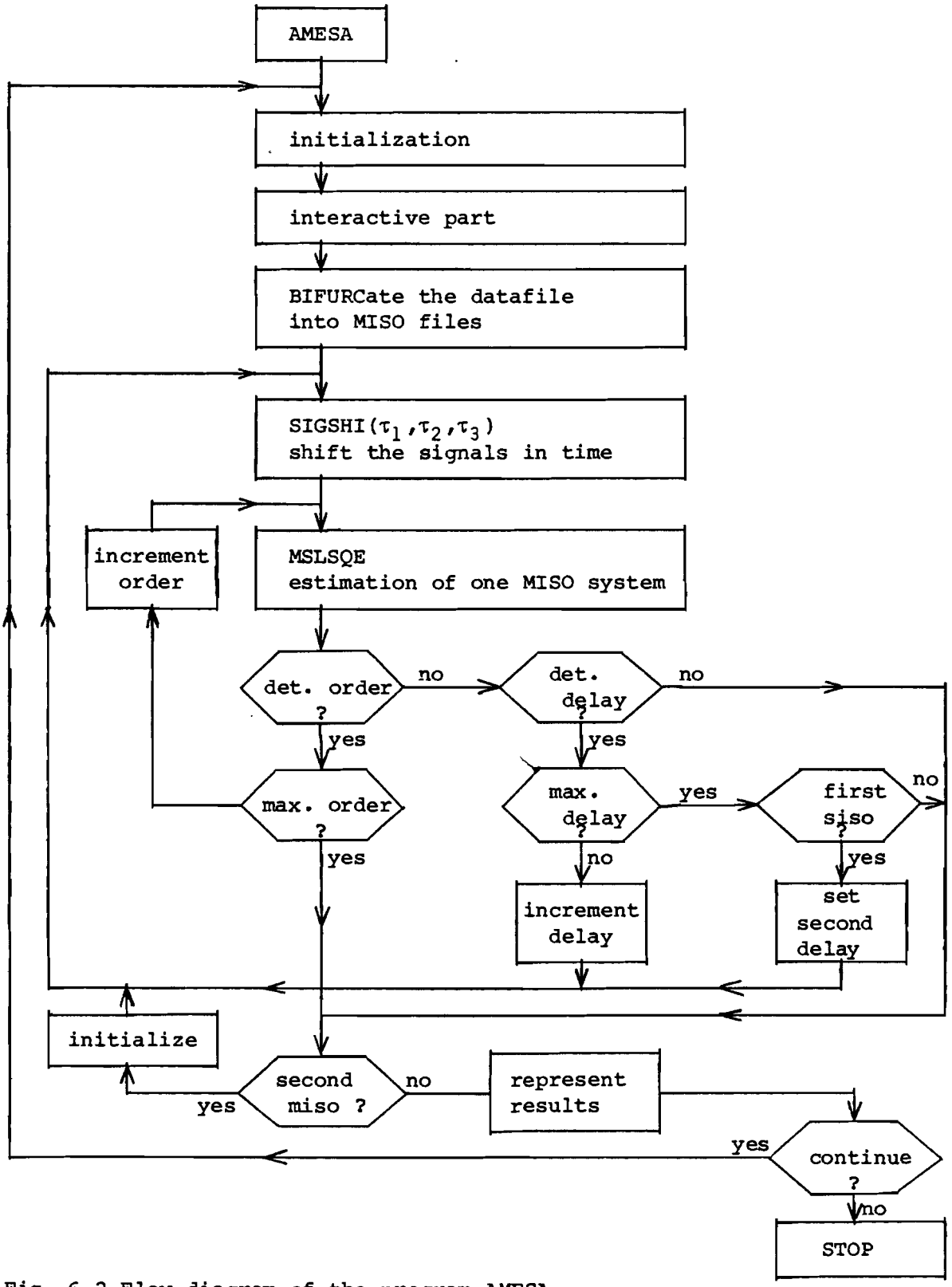


Fig. 6.2 Flow diagram of the program AMESA

6.3 Normalization

An unavoidable property of practical data is that the magnitude of the different signals can differ substantially. Furthermore the signals tend to have a mean value that is not equal to zero. These effects can influence the estimation and therefore a normalization of the data is recommended. Also a normalization improves the ease of interpreting the results, for example the loss-function then directly informs us about the achieved model accuracy.
This section treats the effects of such a normalization.
Assume a MISO process as illustrated in Fig. 6.3.

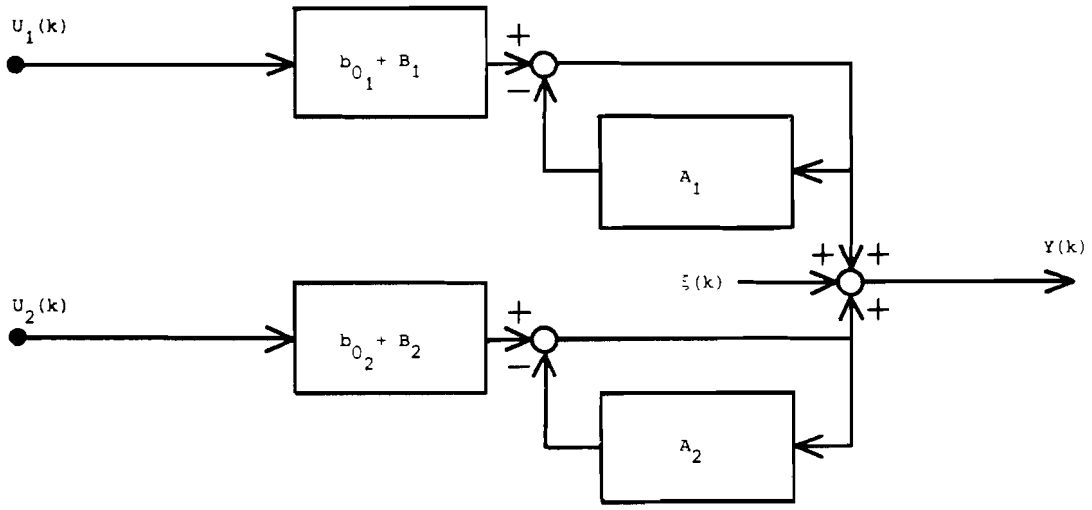


Fig. 6.3 Model of a MISO process.

The transfer of the SISO subsystems can be described as follows:

$$v_1(k) = b_0u_1(k)+\ldots b_nu_1(k-n)-a_1v_1(k-1)-\ldots a_nv_1(k-n) \qquad (6.1)$$

Input signal $u_1(k)$ can be written as:

$$u_1(k) = K_{i1}\tilde{u}_1(k)+\overline{u_1} \qquad (6.2)$$

with $\overline{u_1} \equiv E[u_1(k)]$

and $K_1^2 \equiv E[(u_1(k)-\overline{u_1})^2]$

so that $E[\tilde{u}_1^2(k)] = 1$

Invoking eq. (6.2) in eq. (6.1) we obtain:

$$v_1(k) = K_{i1}b_0\tilde{u}_1(k)+\ldots+K_{i1}b_n\tilde{u}_1(k-n) + \overline{u_1}\sum_{i=0}^{n} b_i - \\ -a_1v_1(k-1)-\ldots-a_nv_1(k-n) \qquad (6.3)$$

With $b_i' \equiv K_{i1}b_i$ and $\overline{c_1} \equiv \overline{u_1}\,\Sigma\, b_i$ this becomes:

$$v_1(k) = b_0'\tilde{u}_1(k)+\ldots+b_n'\tilde{u}_1(k-n) - a_1v_1(k-1)-\ldots a_nv_1(k-n) + \overline{c_1} \qquad (6.4)$$

Fig. 6.4 illustrates how the influence of $\overline{c_1}$ on the output signal $v_1(k)$ can be represented.

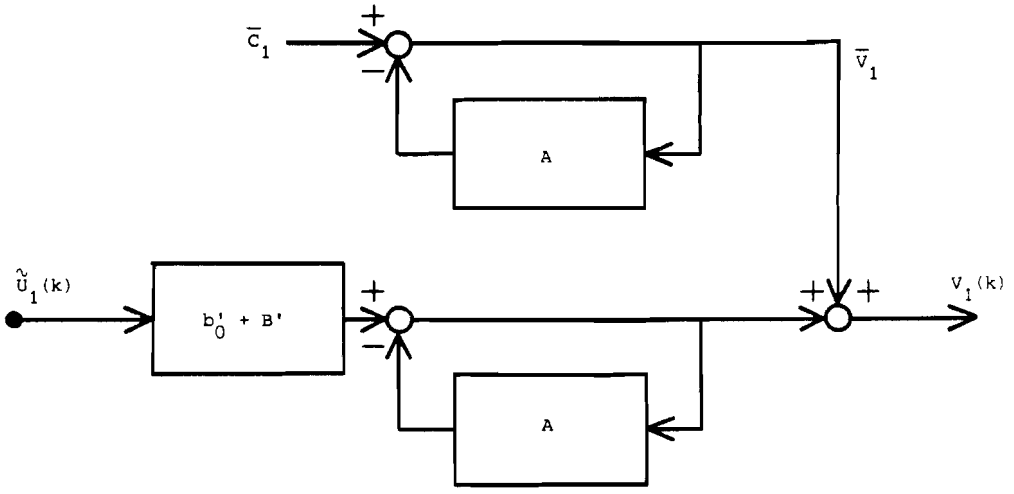


Fig. 6.4 Offset on the output signal $v_1(k)$.

The result is a bias $\overline{v_1}$ of $v_1(k)$ that can be denoted as

$$\overline{v_1} = \frac{1}{1 + \sum_{1} a_i} = \frac{\overline{u_1}\sum_{0} b_i}{1 + \sum_{1} a_i} \qquad (6.5)$$

The output of the SISO subsystem can be written as follows:

$$v_1(k) = v_1'(k) + \overline{v_1} \tag{6.6}$$

and so the output of the MISO system as:

$$y(k) = v_1'(k) + \overline{v_1} + v_2'(k) + \overline{v_2} + \xi(k) \tag{6.7}$$

Normalization of y(k) leads to:

$$y(k) = K_o \tilde{y}(k) + \overline{y} \tag{6.8}$$

with $K_o^2 \equiv E\left[(y(k)-\overline{y})^2\right]$.
Eq. (6.7) and eq. (6.8) give rise to:

$$\tilde{y}(k) = \frac{1}{K_o}\left[v_1'(k) + v_2'(k) + \xi(k) \right] \tag{6.9}$$

Thus we can write:

$$\begin{aligned}
\tilde{\xi}(k) &= \frac{1}{K_o}\xi(k) \\
\tilde{v}_1(k) &= \frac{1}{K_o} v_1(k) \\
\tilde{v}_2(k) &= \frac{1}{K_o} v_2(k) \\
\tilde{b}_{j_1}(k) &= \frac{K_{i1}}{K_o} b_{j_1} \\
\tilde{b}_{j_2}(k) &= \frac{K_{i2}}{K_o} b_{j_2}
\end{aligned} \tag{6.10}$$

Normalization only affects the values of the moving average parameters, the autoregressive parameters do not differ.
Fig. 6.5 illustrates the model after normalization.

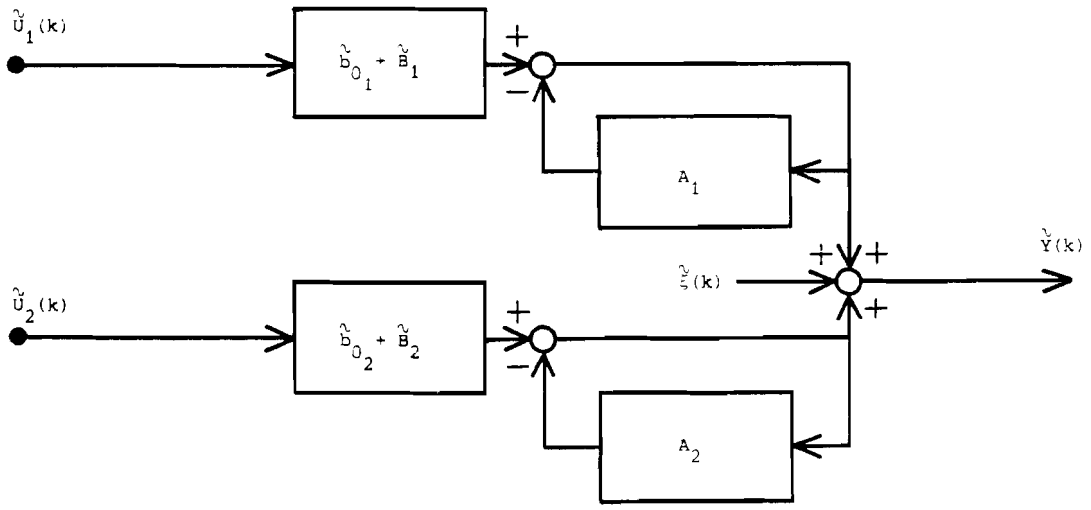


Fig. 6.5 The normalized model.

Estimation delivers the normalized parameters $\tilde{a}_j$ and $\tilde{b}_j$. The original scale of the parameters can be obtained as follows:

$$\begin{aligned} a_j &= \tilde{a}_j && j=1,\ldots,n \\ b_{k_j} &= \frac{K_o}{K_{ik}} \cdot \tilde{b}_{k_j} && k=1,2 \\ & && j=0,\ldots,n \end{aligned} \qquad (6.11)$$

6.4 Determination of time delays

A significant problem of real life data is that it is often afflicted with several different time delays.
This inevitably also occurs in the case of the glass tube manufacturing process, for the relevant quantities are being measured at different locations in the process, thus creating time delays. Furthermore the process inherently contains a certain amount of inertness in its response.
Fig. 6.6 illustrates the delay times in a two-input-two-output system.

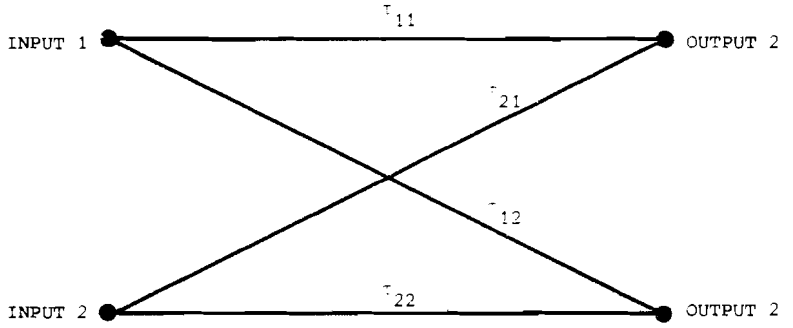


Fig. 6.6 Delay times of a two-input two-output system.

In the case of a system with p inputs and q outputs, pxq time delays occur. By shifting of the data p+q-1 time delays can be eliminated. This leaves pxq - (p+q-1) time delays which can not be eliminated.
This is a serious problem, since a time delay in general requires an increase of the model order, at the size of the remaining time delay. Splitting the model in MISO systems seems advantageous, since all the time delays can then be eliminated.
Another problem is to determine the time delay of a transfer. This can be done by examining the corresponding cross-correlation function. Another way is to perform several estimations with successive time delays adjusted in the algorithm. The corresponding curve of the loss-function is then supposed to indicate a minimum at the correctly adjusted time delay.
Practizing this method for simulated data shows that the estimations do not need to be performed for the correct order of the model, since a lower order will also result in the correct time delay. Clearly this is an advantage, since the order can also be unknown.
This method is implemented in the program AMESA, by introducing an extra loop in the program. Fig. 6.2 illustrates the operation of this program.

6.5 Determination of order

Insufficient knowledge of the process' physics or complications that go beyond the existing physical laws disable man to give a correct description of certain processes. In such a case, the model structure that should be used for identification of the process, is unknown.
It is assumed that the model that we apply is suitable for identification of the glass tube manufacturing process, but the orders of the different

SISO transfers are still unknown. The program AMESA contains a simple extension to enable an order determination; the estimation of a MISO transfer is simply performed for successive orders (cf. fig. 6.2). The curve of the loss-function then is supposed to lead to the proper order; at the right order the loss-function will show a considerable drop in value.
Another way to get some indication of the right order is to examine the pole-zero plots of estimates for different orders. If pole-zero cancellation occurs, then the model is overparametrized; the highest order for which no pole-zero cancellation takes place will be the best order for the process.

6.6 Results

First of all the signals have been normalized, to facilitate the interpretation of the achieved results. The data is checked for suitability. This is done by examining the correlation between the input signals. Fig. 6.7 shows the cross-correlation between the two inputs. Imposing a significance level of 99 % we get a reliability interval of $[-3\sigma,+3\sigma]$, where σ is the standard deviation of the cross-correlation:

$$\sigma = \frac{\sqrt{[\phi_{11}(0).\phi_{22}(0)]}}{\sqrt{N}} \qquad (6.12)$$

with $\phi_{11}(0)$ and $\phi_{22}(0)$ the autocorrelations of the respective input signals (τ=0), N the number of samples applied for the cross-correlation. Since we normalized the signals, we have $\sigma = 1/\sqrt{10500} = 0.01$, so the reliability interval is $[-0.03,+0.03]$. This is represented in fig. 6.7 by the dotted lines. Since almost all fluctuations of the cross-correlation are inside this interval, the hypothesis that both input signals are uncorrelated is accepted.
A check for the whiteness of the input signals leads to fig. 6.8-a and 6.8-b, representing the autocorrelation of signal 1 and 2 respectively. Interpretation of this, together with the cross-correlation curves of fig. 6.9-a to 6.9-d, gives rise to a reduction of the sample rate of 5 to 1. Thus the autocorrelations represent signals with almost white properties. The cross-correlations of fig. 6.9-a to 6.9-d give an indication of the time delay that is involved for the different

transfers, but do not give decisive answers.

Remark:

The cross-correlations give rise to the supposition that the process has not been measured in a complete open loop situation, since the variation of the cross-correlations before the supposed delay time indicate a non-causal behaviour of the system.
Possibly the variations of the cross-correlations can also be inherent to the process; variations of gas flow and drawing speed will not operate on one specific location in the process, but will influence the production of the tube over a longer range where the tube's glass viscosity still allows a certain deformation.

6.6.1 The first MISO system

First of all the time delays are verified with the program AMESA. Since the estimation is performed for a range of successive time delays, the verification is quite time consuming. Fig. 6.10-a and -b present the results of this part. The time delay of the transfer from input 1 to output 1 is estimated 8 sample units and the time delay of input 2 to output 1 is estimated 6 sample units (measured in the reduced sample rate).
The second step was to determine the order of the system. It is assumed that both SISO subsystems and MA and AR parts have equal orders. With the program AMESA the loss-function is computed for a range of successive orders. For orders higher than 8 the program is terminated because of the estimation of unstable models. Fig. 6.11 represents the curve of the loss-function as a function of the model order. No clear drop of the loss-function is detectable. However the estimation of models with order higher than eight leads to sometimes unstable models, so an eighth order seems suitable for this process.
A pole-zero plot of an eighth order model is presented in fig. 6.12-a and fig. 6.12-b.
Fig. 6.13 represents a reconstruction of the signal for the eighth order. The loss-function amounts to 0.16, which implies that the residual still is a considerable part of the signal.

6.6.2 The second MISO system

Verification of the time delays estimated with the cross-correlation curves leads to quite a disappointment; with the program AMESA no corresponding meaningful time delay can be indicated. Use of AMESA leads to quite different time delays. The time delays suggested by the program are 13 sample units for input 1 and 12 for input 2, see fig. 6.14-a and 6.14-b. Estimation for different orders leads to a seventh order model. Fig. 6.15-a and 6.15-b represent the pole-zero configurations that correspond with this model. The loss-function amounts to 0.56, which implies a quite bad 'fit' of the model.
A reconstruction of the output signal with this model is given in fig. 6.16.

7. Conclusion

As for the first part of this report, the conclusions already have been mentioned in Chapter 2; the cross-validatory selection rule C_1 proposed by Stoica demonstrates a remarkable equivalence in performance with Akaike's AIC, as shown in Stoica (1984). The second proposed criterion C_2 shows a less obvious equivalence with the Generalized AIC.
Thus the criteria function quite well as objective model structure selection rules, but the considerably more extensive computational effort compared to Akaike's criteria is detrimental.
As for the predictive performance, no causal relation could be discovered between the division rule of the data sequence and the best range of prediction of the criteria.
A subject for further research is the supposed superiority of the cross-validatory criteria in specific applications where the process is 'farther' from the model set, i.e. bilinear and non-linear processes.

The parameter estimation algorithm described in the second part of this report is a supplementary method to the group of existing MIMO estimation methods.
Simulation results show that the method performs quite well under several differing conditions. A specific advantage is that the method is quite fast, compared to other MIMO output error estimation methods, see subsection 4.2.4. The method is especially designed for the case of a two-input process. Extension to more inputs would be possible, but does not seem useful, since the alternating estimation would lead to a less attractive computational performance. The alternating nature of the method also imposes restraints on the attainable accuracy, see Section 4.3. Simulation results demonstrate that the method has a slight tendency to asymmetrical behaviour, in case the output signal of the process is mainly composed by one of the two SISO subsystems. Perhaps this can be resolved by altering the initial procedure of the method; i.e. the subtraction of the reconstructed signals can be introduced after the first estimation cycle.
The output error minimization algorithm that is applied in this method, the Steiglitz-McBride algorithm, can be replaced by any other output error algorithm. Application of a different algorithm might improve the performance of the method, since the convergence of the Steiglitz-McBride

method is not always guaranteed.

A theoretical justification of the method is initiated in Section 4.3, but the strong non-linearity of the alternating output error algorithm is quite an obstacle to more extensive proof of consistency of the estimator.

Application of the method to practical data has lead to interesting results. One MISO system could not be modeled properly, but this is perhaps inherent to the specific process.

LIST OF SYMBOLS

$\hat{\ }$	added for estimates
$A(z^{-1})$	autoregressive part of ARMA model
A	system matrix in state-space description
a_i	autoregression parameter
α	some coefficient
$b_0 + B(z^{-1})$	moving average part of ARMA model
B	distribution matrix in state-space description
b_i	moving average parameter
β	some coefficient
$C_{I,1m,II,2k}$	cross-validation criteria
AIC	Akaike's Information Criterion
GAIC	Generalized AIC
C	output matrix in state-space description
C_j^i	termination criterion
V_j^i	termination criterion
γ	some coefficient
D	I/O matrix in state-space description

$d_i(k)$	difference signal of i^{th} SISO subsystem
$e(k)$	disturbance signal
$\xi(k)$	output disturbance signal
F	number of degrees of freedom of a model
$f(A,\xi)$	filtered output disturbance ξ
$f[\hat{y},y]$	some selected loss-function
$\underline{\phi}^T(k)$	$[u(k)\ y(k-1),\ldots,y(k-n)]$
ϕ_{uu}	autocorrelation
ϕ_{uy}	cross-correlation
h_j	impulse response j^{th} step
I_p	p^{th} interval of the data sequence
i	index
K_{ij}	normalization coefficient
k	index
L_s	prediction performance criterion
$M_m(\underline{\theta})$	model structure
m	index
N	number of data samples
n	order of the process

n_a	order of autoregression
n_b	order of moving average part
P	linear, time-invariant, stable, stationary process
$P(z^{-1})$	left matrix fraction
p	index
$Q(z^{-1})$	right matrix fraction
$\underline{r}_i$	i^{th} row vector of $P^{-1}Q$
r(F,N)	fraction of signal contents
S	data sequence
SNR	signal-to-noise-ratio
σ	standard deviation
t	index
τ	time delay
$u_i(k)$	i^{th} input signal
$\tilde{u}_i(k)$	i^{th} prefiltered input signal
$V(\underline{\theta})$	loss-function
$v_i(k)$	i^{th} SISO subsystem output signal
W	some weighting matrix
$W(\hat{\underline{\theta}})$	$\Sigma\ w_p w_p^T$

$\underline{w}_p(\hat{\underline{\theta}})$	$\Sigma\ e.\underline{e}_\theta$
$x(k)$	state vector
$y_i(k)$	i^{th} output signal
$\tilde{y}_i(k)$	i^{th} prefiltered output signal
z^{-1}	unitary time delay operator

TABLES

ORDER SELECTION RULE	1	2	3	4	5	6	7	8	9	10	NOT FOUND	SELECTED AVERAGE ORDER	STANDARD DEVIATION OF ORDER
AIC	0	0	0	79	19	2	0	0	0	0	0	4.23	0.468
C_1 (M=1)	0	0	0	80	18	2	2	0	0	0	0	4.22	0.463
C_1 (M=5)	0	0	0	79	17	4	0	0	0	0	0	4.25	0.520
GAIC (K=4)	0	0	0	94	6	0	0	0	0	0	0	4.06	0.239
C_2 (K=4)	0	0	0	75	18	6	1	0	0	0	0	4.33	0.636
GAIC (K=6)	0	0	0	96	4	0	0	0	0	0	0	4.04	0.197
C_2 (K=6)	0	1	0	82	16	1	0	0	0	0	0	4.16	0.466

Table 2.1-a

Order estimation performance of several selection rules

The pole-zero plot of the process is represented in fig. 2.7-a.

ORDER SELECTION RULE	1	2	3	4	5	6	7	8	9	10	NOT FOUND	SELECTED AVERAGE ORDER	STANDARD DEVIATION OF ORDER
AIC	0	0	7	78	12	1	2	0	0	0	0	4.13	0.630
C_1 (M=1)	0	0	8	79	9	2	2	0	0	0	0	4.11	0.650
C_1 (M=5)	0	0	7	78	9	4	2	0	0	0	0	4.16	0.692
GAIC (K=4)	0	0	22	76	2	0	0	0	0	0	0	3.80	0.449
C_2 (K=4)	0	0	15	73	8	2	1	1	0	0	0	4.04	0.751
GAIC (K=6)	0	0	39	60	1	0	0	0	0	0	0	3.62	0.508
C_2 (K=6)	0	0	33	61	6	0	0	0	0	0	0	3.73	0.566

Table 2.1-b

Order estimation performance of several selection rules.

The pole-zero plot of the process is represented in fig. 2.7-b.

ORDER SELECTION RULE	1	2	3	4	5	6	7	8	9	10	NOT FOUND	SELECTED AVERAGE ORDER	STANDARD DEVIATION OF ORDER
AIC	83	13	2	1	1	0	0	0	0	0	0	1.24	0.637
C_1 (M=1)	83	14	1	1	1	0	0	0	0	0	0	1.23	0.617
C_1 (M=5)	80	17	1	1	1	0	0	0	0	0	0	1.26	0.629
GAIC (K=4)	95	4	0	1	0	0	0	0	0	0	0	1.07	0.355
C_2 (K=4)	74	20	4	2	0	0	0	0	0	0	0	1.34	0.655
GAIC (K=6)	99	1	0	0	0	0	0	0	0	0	0	1.01	0.100
C_2 (K=6)	84	13	2	1	0	0	0	0	0	0	0	1.20	0.513

Table 2.1-c

Order estimation performance of several selection rules.

The pole-zero plot of the process is represented in fig. 2.7-c.

Order estimation

ORDER SELECTION RULE	1	2	3	4	5	6	7	8	9	10	NOT FOUND	SELECTED AVERAGE ORDER	STANDARD DEVIATION OF ORDER
AIC	0	6	70	19	5	0	0	0	0	0	0	3.23	0.633
C_1 (M=1)	0	5	69	21	5	0	0	0	0	0	0	3.26	0.629
C_1 (M=5)	0	4	70	20	5	1	0	0	0	0	0	3.29	0.671
GAIC (K=4)	0	17	71	11	1	0	0	0	0	0	0	2.96	0.567
C_2 (K=4)	0	15	63	19	3	0	0	0	0	0	0	3.10	0.675
GAIC (K=6)	0	27	69	4	0	0	0	0	0	0	0	2.77	0.510
C_2 (K=6)	0	21	66	13	0	0	0	0	0	0	0	2.92	0.581

Predictive performance

SELECTION RULE	AIC	C_1 (M=1)	C_1 (M=5)	GAIC (K=4)	C_2 (K=4)	GAIC (K=6)	C_2 (K=6)
SELECTED AV. ORDER	3.23	3.26	3.29	2.96	3.10	2.77	2.92
L_1	0.838E+0	0.840E+0	0.839E+0	0.849E+0	0.851E+0	0.844E+0	0.843E+0
L_5	0.990E+0	0.989E+0	0.989E+0	0.990E+0	0.990E+0	0.984E+0	0.989E+0
L_{1500}	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1
L_{2500}	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1

Table 2.2-a

Predictive performance and order selection of several criteria.

The simulated process has the zeros of the MA part close to zero.

Fig. 2.8-a represents the corresponding pole-zero plot of the process.

The simulation is performed with zero input ($u_k = 0$) and a white Gaussian disturbance on the input ($\sigma = 1.0$).

Order estimation

ORDER SELECTION RULE	1	2	3	4	5	6	7	8	9	10	NOT FOUND	SELECTED AVERAGE ORDER	STANDARD DEVIATION OF ORDER
AIC	0	0	52	39	7	1	1	0	0	0	0	3.60	0.752
C_1 (M=1)	0	0	50	39	8	3	0	0	0	0	0	3.64	0.794
C_1 (M=5)	0	0	48	40	9	3	0	0	0	0	0	3.67	0.802
GAIC (K=4)	0	0	79	19	2	0	0	0	0	0	0	3.23	0.500
C_2 (K=4)	0	0	58	35	2	4	1	0	0	0	0	3.55	0.838
GAIC (K=6)	0	0	88	11	1	0	0	0	0	0	0	3.13	0.400
C_2 (K=6)	0	0	71	25	3	1	0	0	0	0	0	3.34	0.620

Predictive performance

SELECTION RULE	AIC	C_1 (M=1)	C_1 (M=5)	GAIC (K=4)	C_2 (K=4)	GAIC (K=6)	C_2 (K=6)
SELECTED AV. ORDER	3.60	3.64	3.67	3.23	3.55	3.13	3.34
L_1	0.858E+0	0.857E+0	0.857E+0	0.868E+0	0.855E+0	0.875E+0	0.877E+0
L_5	0.982E+0	0.981E+0	0.982E+0	0.985E+0	0.981E+0	0.986E+0	0.985E+0
L_{1500}	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1
L_{2500}	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1	0.101E+1

Table 2.2-b

Predictive performance and order selection of several criteria.

The simulated process has the zeros of the MA part close to zero.

Fig. 2.8-b represents the corresponding pole-zero plot of the process.

The simulation is performed with zero input ($u_k = 0$) and a white Gaussian disturbance on the input ($\sigma = 1.0$).

Order estimation

ORDER SELECTION RULE	1	2	3	4	5	6	7	8	9	10	11	12	13	NOT FOUND	SELECTED AVERAGE ORDER	STANDARD DEVIATION OF ORDER
AIC	0	21	0	50	0	8	2	5	5	2	5	2	0	0	4.88	2.34
C_1 (M=1)	0	19	0	52	0	8	3	7	3	1	3	3	1	0	4.90	2.65
C_1 (M=5)	0	30	0	47	0	5	2	3	6	2	3	2	0	0	4.47	2.62
GAIC (K=4)	0	35	0	54	0	3	4	3	1	0	0	0	0	0	3.65	1.59
C_2 (K=4)	0	37	0	46	0	6	2	4	4	0	0	0	0	0	3.89	2.09
GAIC (K=6)	0	46	0	46	0	5	0	3	0	0	0	0	0	0	3.30	1.43
C_2 (K=6)	0	54	0	40	0	1	0	3	2	0	0	0	0	0	3.16	1.58

Predictive performance

SELECTION RULE	AIC	C_1 (M=1)	C_1 (M=5)	GAIC (K=4)	C_2 (K=4)	GAIC (K=6)	C_2 (K=6)
SELECTED AV. ORDER	4.88	4.90	4.47	3.65	3.89	3.30	3.16
L_1	0.102E+1	0.102E+1	0.102E+0	0.101E+1	0.103E+1	0.103E+1	0.101E+1
L_5	0.116E+1	0.116E+1	0.118E+1	0.119E+1	0.118E+1	0.119E+1	0.117E+1
L_{1500}	0.114E+1	0.113E+1	0.116E+1	0.118E+1	0.118E+1	0.120E+1	0.122E+1
L_{2500}	0.114E+1	0.113E+1	0.116E+1	0.118E+1	0.128E+1	0.120E+1	0.122E+1

Table 2.2-c

Predictive performance and order selection of several criteria. The simulated process has the zeros of the MA part close to the unit circle. Fig. 2.8-c represents the corresponding pole-zero plot of the process. The simulation is performed with zero input ($u_k = 0$) and a white Gaussian disturbance on the input ($\sigma = 1.0$).

Order estimation

ORDER SELECTION RULE	1	2	3	4	5	6	7	8	9	10	11	12	NOT FOUND	SELECTED AVERAGE ORDER	STANDARD DEVIATION OF ORDER
AIC	0	28	0	2	0	0	32	0	35	0	2	1	0	6.37	2.97
C_1 (M=1)	0	28	0	2	0	0	32	0	35	0	2	1	0	6.37	2.97
C_1 (M=5)	0	35	0	2	0	0	24	0	36	0	1	2	0	6.05	3.21
GAIC (K=4)	0	44	0	2	0	4	28	1	21	0	0	0	0	5.13	2.94
C_2 (K=4)	0	47	0	1	0	6	21	3	17	1	4	0	0	5.12	3.15
GAIC (K=6)	0	52	0	4	0	7	30	2	5	0	0	0	0	4.33	2.57
C_2 (K=6)	0	62	0	5	2	3	17	2	9	0	0	0	0	3.88	2.60

Predictive performance

SELECTION RULE	AIC	C_1 (M=1)	C_1 (M=5)	GAIC (K=4)	C_2 (K=4)	GAIC (K=6)	C_2 (K=6)
SELECTED AV. ORDER	6.37	6.37	6.05	5.13	5.12	4.33	3.88
L_1	0.105E+1	0.105E+1	0.109E+1	0.107E+1	0.108E+1	0.106E+1	0.108E+1
L_5	0.115E+1	0.115E+1	0.118E+1	0.119E+1	0.119E+1	0.120E+1	0.128E+1
L_{1500}	0.115E+1	0.115E+1	0.117E+1	0.121E+1	0.122E+1	0.124E+1	0.127E+1
L_{2500}	0.115E+1	0.115E+1	0.117E+1	0.121E+1	0.122E+1	0.124E+1	0.127E+1

Table 2.2-d

Predictive performance and order selection of several criteria. The simulated process has the zeros of the MA part close to the unit circle. Fig. 2.8-d represents the corresponding pole-zero plot of the process. The simulation is performed with zero input ($u_k = 0$) and a white Gaussian disturbance on the input ($\sigma = 1.0$).

Order estimation

ORDER SELECTION RULE	1	2	3	4	5	6	7	8	9	10	NOT FOUND	SELECTED AVERAGE ORDER	STANDARD DEVIATION OF ORDER
AIC	0	0	66	30	1	1	1	1	0	0	0	3.44	0.808
C_1 (M=1)	0	0	66	30	1	1	1	1	0	0	0	3.44	0.808
C_1 (M=5)	0	0	65	31	1	0	1	2	0	0	0	3.47	0.893
GAIC (K=4)	0	0	83	15	1	1	0	0	0	0	0	3.20	0.492
C_2 (K=4)	0	0	69	28	2	1	0	0	0	0	0	3.35	0.575
GAIC (K=6)	0	0	89	11	0	0	0	0	0	0	0	3.11	0.314
C_2 (K=6)	0	0	81	17	2	0	0	0	0	0	0	3.21	0.456

Predictive performance

SELECTION RULE	AIC	C_1 (M=1)	C_1 (M=5)	GAIC (K=4)	C_2 (K=4)	GAIC (K=6)	C_2 (K=6)
SELECTED AV. ORDER	3.44	3.44	3.47	3.20	3.35	3.11	3.21
L_1	0.762E-1	0.762E-1	0.763E-1	0.761E-1	0.767E-1	0.761E-1	0.774E-1
L_5	0.890E-1	0.890E-1	0.890E-1	0.888E-1	0.889E-1	0.890E-1	0.894E-1
L_{1500}	0.911E-1	0.911E-1	0.911E-1	0.910E-1	0.910E-1	0.910E-1	0.910E+1
L_{2500}	0.911E-1	0.911E-1	0.911E-1	0.910E-1	0.910E-1	0.910E-1	0.910E-1

Table 2.2-e

Predictive performance and order selection of several criteria. The simulated process has the zeros of the MA part close to zero. Fig. 2.8-a represents the corresponding pole-zero plot of the process. The simulation is performed with noiselike input (σ = 1.0) and an independent white Gaussian disturbance on the input (σ = 0.3).

Order estimation

ORDER SELECTION RULE	1	2	3	4	5	6	7	8	9	10	NOT FOUND	SELECTED AVERAGE ORDER	STANDARD DEVIATION OF ORDER
AIC	0	0	0	62	29	5	4	0	0	0	0	4.51	0.772
C_1 (M=1)	0	0	0	63	28	5	4	0	0	0	0	4.50	0.772
C_1 (M=5)	0	0	0	62	26	8	4	0	0	0	0	4.54	0.809
GAIC (K=4)	0	0	0	77	20	2	1	0	0	0	0	4.27	0.548
C_2 (K=4)	0	0	0	59	30	10	1	0	0	0	0	4.53	0.717
GAIC (K=6)	0	0	0	86	13	1	0	0	0	0	0	4.15	0.386
C_2 (K=6)	0	0	0	77	18	3	2	0	0	0	0	4.30	0.628

Predictive performance

SELECTION RULE	AIC	C_1 (M=1)	C_1 (M=5)	GAIC (K=4)	C_2 (K=4)	GAIC (K=6)	C_2 (K=6)
SELECTED AV. ORDER	4.51	4.50	4.54	4.27	4.53	4.15	4.30
L_1	0.765E-1	0.765E-1	0.761E-1	0.758E-1	0.770E-1	0.767E-1	0.761E-1
L_5	0.887E-1	0.889E-1	0.891E-1	0.889E-1	0.891E-1	0.890E-1	0.889E-1
L_{1500}	0.913E-1	0.913E-1	0.913E-1	0.913E-1	0.913E-1	0.913E-1	0.913E-1
L_{2500}	0.913E-1	0.913E-1	0.913E-1	0.913E-1	0.913E-1	0.913E-1	0.913E-1

Table 2.2-f

Predictive performance and order selection of several criteria. The simulated process has the zeros of the MA part close to zero. Fig. 2.8-b represents the corresponding pole-zero plot of the process. The simulation is performed with noiselike input (σ = 1.0) and an independent white Gaussian disturbance on the input (σ = 0.3).

Order estimation

ORDER SELECTION RULE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	NOT FOUND	SELECTED AVERAGE ORDER	STANDARD DEVIATION OF ORDER
AIC	0	0	0	27	0	0	0	0	5	0	49	0	18	0	1	0	9.41	3.45
C_1 (M=1)	0	0	0	29	0	0	0	0	5	0	48	0	17	0	1	0	9.25	3.51
C_1 (M=5)	0	0	0	30	0	0	0	0	3	0	46	0	20	0	1	0	9.28	3.60
GAIC (K=4)	0	1	0	46	0	0	0	0	5	0	38	0	10	0	0	0	7.79	3.71
C_2 (K=4)	0	1	0	46	0	0	0	0	4	0	36	0	11	1	1	0	7.90	3.84
GAIC (K=6)	0	2	0	60	0	0	0	0	6	0	28	0	4	0	0	0	6.58	3.46
C_2 (K=6)	0	6	0	57	0	0	0	0	6	0	28	0	3	0	0	0	6.41	3.49

Predictive performance

SELECTION RULE	AIC	C_1 (M=1)	C_1 (M=5)	GAIC (K=4)	C_2 (K=4)	GAIC (K=6)	C_2 (K=6)
SELECTED AV. ORDER	9.41	9.25	9.28	7.79	7.90	6.58	6.41
L_1	0.982E-1	0.990E-1	0.105E+0	0.125E+0	0.125E+0	0.139E+0	0.144E+0
L_5	0.117E+0	0.119E+0	0.122E+0	0.138E+0	0.137E+0	0.152E+0	0.162E+0
L_{1500}	0.129E+0	0.131E+0	0.132E+0	0.153E+0	0.153E+0	0.172E+0	0.181E+0
L_{2500}	0.129E+0	0.131E+0	0.132E+0	0.153E+0	0.153E+0	0.172E+0	0.181E+0

Table 2.2-g

Predictive performance and order selection of several criteria. The simulated process has the zeros of the MA part close to the unit circle. Fig. 2.8-c represents the corresponding pole-zero plot of the process. The simulation is performed with noiselike input (σ = 1.0) and an independent white Gaussian disturbance on the input (σ = 0.3).

Order estimation

ORDER SELECTION RULE	1	2		9	10	11	12	13	14	15	16	17	18	19	20	NOT FOUND	SELECTED AVERAGE ORDER	STANDAR DEVIATI OF ORDE
AIC	0	1	0	67	0	12	0	4	2	5	5	0	3	1	0	0	10.45	2.78
C_1 (M=1)	0	0	0	61	0	14	0	5	2	6	4	1	5	2	0	0	10.95	2.99
C_1 (M=5)	0	2	0	64	0	12	0	4	1	5	5	1	1	0	0	0	10.60	3.17
GAIC (K=4)	0	4	0	79	0	12	0	3	0	1	0	0	1	0	0	1	9.687	3.47
C_2 (K=4)	0	10	0	55	0	16	0	6	2	2	5	2	0	0	0	0	8.18	2.56
GAIC (K=6)	0	14	0	78	0	8	0	0	0	0	0	0	0	1	0	0	7.74	3.77
C_2 (K=6)	0	26	0	60	0	6	0	3	2	2	0	0	0	0	0	0	2.92	0.581

Predictive performance

SELECTION RULE	AIC	C_1 (M=1)	C_1 (M=5)	GAIC (K=4)	C_2 (K=4)	GAIC (K=6)	C_2 (K=6)
SELECTED AV. ORDER	10.45	10.95	10.60	9.23	9.69	8.18	7.74
L_1	0.101E+0	0.971E-1	0.107E+0	0.115E+0	0.118E+0	0.164E+0	0.195E+0
L_5	0.109E+0	0.104E+0	0.111E+0	0.119E+0	0.142E+0	0.157E+0	0.190E+0
L_{1500}	0.112E+0	0.107E+0	0.116E+0	0.127E+0	0.149E+0	0.169E+0	0.217E+0
L_{2500}	0.112E+0	0.107E+0	0.116E+0	0.127E+0	0.149E+0	0.169E+0	0.216E+0

Table 2.2-h

Predictive performance and order selection of several criteria. The simulated process has the zeros of the MA part close to the unit circle. Fig. 2.8-d represents the corresponding pole-zero plot of the process. The simulation is performed with noiselike input (σ = 1.0) and an independent white Gaussian disturbance on the input (σ = 0.3).

SIGNAL-TO-NOISE-RATIO IN dB		TRUE VALUE	INF	40	30	20	10	0
FIRST SUB-SYSTEM	B_0	1.0	0.100E+1	0.100E+1	0.102E+1	0.994E+0	0.116E+1	0.151E+1
	B_1	0.7	0.700E+0	0.692E+0	0.676E+0	0.650E+0	0.442E+0	-0.572E-1
	B_2	0.3	0.300E+0	0.308E+0	0.287E+0	0.387E+0	0.553E+0	0.113E+1
	A_1	-1.5	-0.150E+1	-0.150E+1	-0.150E+1	-0.150E+1	-0.147E+1	-0.138E+1
	A_2	0.7	0.700E+0	0.699E+0	0.702E+0	0.700E+0	0.676E+0	0.604E+0
SECOND SUB-SYSTEM	B_0	2.5	0.250E+1	0.250E+1	0.249E+1	0.243E+1	0.255E+1	0.266E+1
	B_1	1.25	0.125E+1	0.126E+1	0.129E+1	0.126E+1	0.146E+1	0.210E+1
	B_2	0.1	0.100E+0	0.110E+0	0.126E+0	0.127E+0	0.349E+0	0.910E+0
	A_1	-1.0	-0.100E+1	-0.997E+0	-0.991E+0	-0.998E+0	-0.932E+0	-0.747E+0
	A_2	0.5	0.500E+0	0.499E+0	0.497E+0	0.506E+0	0.472E+0	0.370E+0
	$\overline{Y^2}$		0.734E+2	0.735E+2	0.623E+2	0.501E+2	0.805E+2	0.141E+3
	$\overline{(Y-\tilde{Y})^2}$		0.961E-9	0.648E-2	0.622E-1	0.547E+0	0.648E+1	0.647E+2
NUMBER OF ITERATIONS			22	20	21	16	24	67

Table 5.2 Estimation results for the case of white noise disturbances and independent white input signals. The subsystem output signals are equally strong.

SIGNAL-TO-NOISE-RATIO IN dB		TRUE VALUE	INF	40	30	20	10	0
SECOND SUB-SYSTEM	B_0	1.0	0.100E+1	0.100E+1	0.104E+1	0.955E+0	0.116E+1	0.238E+1
	B_1	0.7	0.700E+0	0.690E+0	0.658E+0	0.761E+0	0.562E+0	-0.985E+0
	B_2	0.3	0.300E+0	0.322E+0	0.324E+0	0.317E+0	0.420E+0	0.198E+1
	A_1	-1.5	-0.150E+1	-0.150E+1	-0.150E+1	-0.150E+1	-0.143E+1	-0.496E+0
	A_2	0.7	0.700E+0	0.698E+0	0.698E+0	0.702E+0	0.672E+0	0.186E+0
FIRST SUB-SYSTEM	B_0	7.5	0.750E+1	0.751E+1	0.749E+1	0.754E+1	0.743E+1	0.815E+1
	B_1	3.75	0.375E+1	0.372E+1	0.372E+1	0.373E+1	0.377E+1	0.524E+1
	B_2	0.3	0.100E+0	0.316E+0	0.263E+0	0.466E+0	0.892E+0	0.175E+1
	A_1	-1.0	-0.100E+1	-0.100E+1	-0.100E+1	-0.995E+0	-0.988E+0	-0.840E+0
	A_2	0.5	0.500E+0	0.500E+0	0.500E+0	0.502E+0	0.492E+0	0.415E+0
	$\overline{Y^2}$		0.231E+3	0.273E+3	0.300E+3	0.305E+3	0.363E+3	0.692E+3
	$\overline{(Y-\hat{Y})^2}$		0.413E-10	0.256E-1	0.294E+0	0.300E+1	0.335E+2	0.330E+3
NUMBER OF ITERATIONS			22	20	18	17	35	111

Table 5.3 Estimation results for the case of white noise disturbances and independent white input signals. The first subsystem output signal is stronger than the second subsystem output signal; the signal_1-to-signal_2-ratio is +10 dB.

SIGNAL-TO-NOISE-RATIO IN dB		TRUE VALUE	INF	40	30	20	10	0
FIRST SUB-SYSTEM	B_0	1.0	0.100E+1	0.100E+1	0.104E+1	0.955E+0	0.117E+1	0.213E+1
	B_1	0.7	0.700E+0	0.690E+0	0.658E+0	0.761E+0	0.567E+0	-0.839E+0
	B_2	0.3	0.300E+0	0.322E+0	0.324E+0	0.317E+0	0.440E+0	0.176E+1
	A_1	-1.5	-0.150E+1	-0.150E+1	-0.150E+1	-0.150E+1	-0.145E+1	-0.128E+1
	A_2	0.7	0.700E+0	0.698E+0	0.698E+0	0.702E+0	0.669E+0	0.511E+0
SECOND SUB-SYSTEM	B_0	7.5	0.750E+1	0.751E+1	0.749E+1	0.754E+1	0.738E+1	0.784E+1
	B_1	3.75	0.375E+1	0.372E+1	0.372E+1	0.373E+1	0.372E+1	0.565E+1
	B_2	0.3	0.100E+0	0.316E+0	0.263E+0	0.466E+0	0.896E+0	0.236E+1
	A_1	-1.0	-0.100E+1	-0.100E+1	-0.100E+1	-0.995E+0	-0.982E+0	-0.797E+0
	A_2	0.5	0.500E+0	0.500E+0	0.500E+0	0.502E+0	0.499E+0	0.406E+0
	$\overline{Y^2}$		0.231E+3	0.273E+3	0.300E+3	0.305E+3	0.363E+3	0.692E+3
	$\overline{(Y-\hat{Y})^2}$		0.413E-10	0.256E-1	0.294E+0	0.300E+1	0.330E+2	0.326E+3
NUMBER OF ITERATIONS			22	20	18	17	34	119

Table 5.4 Estimation results for the case of white noise disturbances and independent white input signals. The second subsystem output signal is stronger than the first subsystem output signal; the signal_1-to-signal_2-ratio is -10 dB.

SIGNAL-TO-NOISE-RATIO IN dB		TRUE VALUE	INF	40	30	20	10	0
FIRST SUB-SYSTEM	B_0	1.0	0.100E+1	0.100E+1	0.986E+0	0.989E+0	0.112E+1	0.114E+1
	B_1	0.7	0.700E+0	0.696E+0	0.706E+0	0.697E+0	0.757E+0	0.672E-1
	B_2	0.3	0.300E+0	0.305E+0	0.282E+0	0.165E+0	-0.470E+0	-0.134E+1
	A_1	-1.5	-0.150E+1	-0.150E+1	-0.150E+1	-0.152E+1	-0.161E+1	-0.177E+1
	A_2	0.7	0.700E+0	0.699E+0	0.703E+0	0.717E+0	0.773E+0	0.836E+0
SECOND SUB-SYSTEM	B_0	2.5	0.250E+1	0.249E+1	0.251E+1	0.252E+1	0.248E+1	0.240E+1
	B_1	1.25	0.125E+1	0.125E+1	0.124E+1	0.109E+1	0.863E+0	0.100E+1
	B_2	0.1	0.100E+0	0.995E-1	0.569E-1	-0.367E+0	-0.462E+0	0.699E+0
	A_1	-1.0	-0.100E+1	-0.100E+1	-0.101E+1	-0.109E+1	-0.113E+1	-0.105E+1
	A_2	0.5	0.500E+0	0.500E+0	0.502E+0	0.523E+0	0.478E+0	0.241E+0
	$\overline{Y^2}$		0.734E+2	0.734E+2	0.734E+2	0.496E+2	0.575E+2	0.130E+3
	$\overline{(Y-\hat{Y})^2}$		0.961E-9	0.785E-2	0.520E-1	0.596E+0	0.109E+2	0.708E+2
NUMBER OF ITERATIONS			22	21	24	23	9	7

Table 5.5 Estimation results for the case of non-white noise disturbances and independent white input signals. The disturbances are ARMA(1,1) filtered.

SIGNAL-TO-NOISE-RATIO IN dB		TRUE VALUE	INF	40	30	20	10	0
FIRST SUB-SYSTEM	B_0	1.0	0.100E+1	0.103E+1	0.976E+0	0.931E+0	0.836E+1	0.207E+1
	B_1	0.7	0.700E+0	0.640E+0	0.742E+0	0.798E+0	0.104E+0	-0.149E+1
	B_2	0.3	0.300E+0	0.336E+0	0.283E+0	0.292E+0	0.274E+0	0.209E+1
	A_1	-1.5	-0.150E+1	-0.150E+1	-0.150E+1	-0.149E+1	-0.148E+1	-0.137E+1
	A_2	0.7	0.700E+0	0.701E+0	0.699E+0	0.698E+0	0.692E+0	0.612E+0
SECOND SUB-SYSTEM	B_0	2.5	0.250E+1	0.249E+1	0.248E+1	0.229E+1	0.280E+1	0.322E+1
	B_1	1.25	0.125E+1	0.129E+1	0.133E+1	0.154E+1	0.840E+0	0.456E+1
	B_2	0.1	0.100E+0	0.661E+0	0.777E-1	0.147E+0	0.109E+1	0.401E+1
	A_1	-1.0	-0.100E+1	-0.998E+0	-0.993E+0	-0.967E+0	-0.931E+0	0.148E+0
	A_2	0.5	0.500E+0	0.497E+0	0.497E+0	0.489E+0	0.533E+0	0.486E-1
	$\overline{Y^2}$		0.630E+3	0.499E+3	0.451E+3	0.453E+3	0.557E+3	0.123E+4
	$\overline{(Y-\hat{Y})^2}$		0.500E-9	0.601E-1	0.492E+0	0.519E+1	0.570E+2	0.520E+3
NUMBER OF ITERATIONS			28	38	13	13	26	156

Table 5.6 Estimation results for the case of white noise disturbances and independent non-white (ARMA(1,1) filtered) input signals.

SIGNAL-TO-NOISE-RATIO IN dB		TRUE VALUE	INF	40	30	20	10	0
FIRST SUB-SYSTEM	B_0	1.0	-0.330E+1	0.310E+1	0.278E+1	0.622E+0	0.767E+1	0.842E+1
	B_1	0.7	0.138E+2	0.129E+2	0.105E+2	0.125E+2	0.606E+1	0.578E+1
	B_2	0.3	-0.945E+1	-0.908E+1	-0.874E+1	-0.918E+1	0.122E+1	0.451E+1
	A_1	-1.5	-0.126E+1	-0.126E+1	-0.123E+1	-0.125E+1	-0.115E+0	-0.119E+0
	A_2	0.7	0.489E+0	0.490E+0	0.462E+0	0.476E+0	-0.907E-2	0.287E+0
SECOND SUB-SYSTEM	B_0	2.5	0.299E+1	0.275E+1	0.186E+1	0.264E+1	-0.945E-1	0.212E-1
	B_1	1.25	-0.390E+1	-0.380E+1	-0.360E+1	-0.379E+1	-0.196E-1	-0.922E-1
	B_2	0.1	-0.223E+0	-0.233E+0	-0.286E-2	-0.784E-1	0.185E+0	0.895E-1
	A_1	-1.0	-0.131E+1	-0.132E+1	-0.134E+1	-0.132E+1	-0.246E+0	-0.460E-1
	A_2	0.5	0.661E+0	0.669E+0	0.674E+0	0.663E+0	-0.162E+0	0.631E-1
	$\overline{Y^2}$		0.136E+3	0.124E+3	0.136E+3	0.124E+3	0.149E+3	0.250E+3
	$\overline{(Y-\hat{Y})^2}$		0.643E-1	0.648E-1	0.192E+0	0.133E+1	0.165E+2	0.128E+3
	NUMBER OF ITERATIONS		2012	2012	1882	1296	7	11

Table 5.7 Estimation results for the case of white noise disturbances and dependent input signals.

The applied filter is an ARMA(1,1) type.

SIGNAL-TO-NOISE-RATIO IN dB		TRUE VALUE	INF	40	30	20	10	0
FIRST SUB-SYSTEM	B_0	1.0	0.173E+1	0.168E+1	0.175E+1	0.130E+1	0.125E+1	0.137E+1
	B_1	0.7	0.106E+1	0.110E+1	0.103E+1	0.107E+1	0.101E+1	0.140E+1
	B_2	0.3	0.545E+0	0.531E+0	0.577E+0	0.331E+0	-0.602E-1	-0.698E+0
	A_1	-1.5	-0.147E+1	-0.147E+1	-0.147E+1	-0.145E+1	-0.148E+1	-0.149E+1
	A_2	0.7	0.715E+0	0.714E+0	0.714E+0	0.670E+0	0.676E+0	0.668E+0
SECOND SUB-SYSTEM	B_0	2.5	0.250E+1	0.251E+1	0.251E+1	0.251E+1	0.270E+1	0.313E+1
	B_1	1.25	-0.601E-1	-0.861E-1	-0.611E-1	0.293E+0	0.348E+1	0.662E+0
	B_2	0.1	-0.141E+1	-0.137E+1	-0.140E+1	-0.107E+1	-0.301E+0	0.446E+0
	A_1	-1.0	-0.154E+1	-0.154E+1	-0.154E+1	-0.138E+1	-0.124E+1	-0.131E+1
	A_2	0.5	0.786E+0	0.784E+0	0.786E+0	0.672E+0	0.634E+0	0.730E+0
	$\overline{Y^2}$		0.566E+2	0.543E+2	0.575E+2	0.542E+2	0.633E+2	0.118E+3
	$\overline{(Y-\hat{Y})^2}$		0.437E-1	0.558E-1	0.104E+0	0.684E+0	0.578E+1	0.580E+2
	NUMBER OF ITERATIONS		600	706	626	80	110	116

Table 5.8 Estimation results for the case of white noise disturbances and dependent input signals.

The applied filter is a time shift of 5 sample steps.

SIGNAL-TO-NOISE-RATIO IN dB		TRUE VALUE	INF	40	30	20	10	0
FIRST SUB-SYSTEM	B_0	1.0	0.123E+1	0.124E+1	0.125E+1	0.125E+1	0.137E+1	0.171E+1
	B_1	0.7	0.216E+1	0.216E+1	0.220E+1	0.187E+1	0.221E+1	0.204E+1
	B_2	0.3						
	A_1	-1.5	-0.787E+0	-0.787E+0	-0.780E+0	-0.785E+0	-0.767E+0	-0.756E+0
	A_2	0.7						
SECOND SUB-SYSTEM	B_0	2.5	0.301E+1	0.301E+1	0.238E+1	0.247E+1	0.266E+1	0.292E+1
	B_1	1.25	0.156E+1	0.157E+1	0.198E+1	0.222E+1	0.252E+1	0.311E+1
	B_2	0.1						
	A_1	-1.0	-0.619E+0	-0.618E+0	-0.628E+0	-0.639E+0	-0.521E+0	-0.445E+0
	A_2	0.5						
	$\overline{Y^2}$		0.734E+2	0.735E+2	0.623E+2	0.501E+2	0.805E+2	0.141E+3
	$\overline{(Y-\hat{Y})^2}$		0.157E+2	0.157E+2	0.179E+2	0.148E+2	0.187E+2	0.748E+2
NUMBER OF ITERATIONS			31	31	8	9	9	12

Table 5.9 Estimation results for the case of white noise disturbances and independent white input signals, for a wrong model order; the model order is taken as 1.

SIGNAL-TO-NOISE-RATIO IN dB		TRUE VALUE	INF	40	30	20	10	0
FIRST SUB-SYSTEM	B_0	1.0	0.100E+1	0.100E+1	0.102E+1	0.994E+0	0.116E+1	0.151E+1
	B_1	0.7	0.403E-1	-0.220E+0	0.381E+0	0.338E+0	0.793E-1	0.241E+0
	B_2	0.3	-0.162E+0	-0.322E+0	0.766E-1	0.189E+0	0.488E+0	0.109E+1
	B_3		-0.198E+0	-0.278E+0	-0.690E-1	-0.126E+0	-0.225E+0	0.327E+0
	A_1	-1.5	-0.216E+1	-0.241E+1	-0.180E+1	-0.181E+1	-0.176E+1	-0.120E+1
	A_2	0.7	0.169E+1	0.206E+1	0.114E+1	0.117E+1	0.111E+1	0.384E+0
	A_3		-0.462E+0	-0.635E+0	-0.205E+0	-0.218E+0	-0.198E+0	0.901E-1
SECOND SUB-SYSTEM	B_0	2.5	0.250E+1	0.250E+1	0.249E+1	0.243E+1	0.255E+1	0.267E+1
	B_1	1.25	0.315E+1	-0.117E+1	0.312E+0	0.172E+1	0.291E+1	0.339E+1
	B_2	0.1	0.105E+1	-0.111E+1	-0.381E+0	0.362E+0	0.123E+1	0.199E+1
	B_3		0.761E-1	-0.103E+0	-0.365E-1	0.420E-1	0.239E+0	0.317E+0
	A_1	-1.0	-0.240E+0	-0.197E+1	-0.138E+1	-0.812E+0	-0.363E+0	-0.251E+0
	A_2	0.5	-0.260E+0	0.147E+1	0.887E+0	0.323E+0	-0.394E-1	-0.130E-1
	A_3		0.380E+0	-0.486E+0	-0.195E+0	0.934E-1	0.256E+0	0.185E+0
	$\overline{Y^2}$		0.734E+2	0.735E+2	0.623E+2	0.501E+2	0.805E+2	0.141E+3
	$\overline{(Y-\hat{Y})^2}$		0.434E-10	0.636E-2	0.621E-1	0.547E+0	0.647E+1	0.647E+2
	NUMBER OF ITERATIONS		30	32	24	18	19	22

Table 5.10 Estimation results for the case of white noise disturbances and independent white input signals, for a wrong model order; the model order is taken as 3.

SIGNAL-TO-NOISE-RATIO IN dB		TRUE VALUE	INF	40	30	20	10	0
FIRST SUB-SYSTEM	B_0	1.0	0.100E+1	0.101E+1	0.103E+1	0.993E+0	0.116E+1	0.152E+1
	B_1	0.7	-0.666E+0	-0.649E+0	-0.630E+0	-0.371E+0	-0.128E+1	0.641E+0
	B_2	0.3	0.340E+0	0.343E+0	0.516E-1	0.508E+0	0.114E+1	0.110E+1
	B_3		0.287E+0	0.246E+0	0.674E-1	0.139E+0	-0.685E+0	0.836E+0
	B_4		0.299E+0	0.291E+0	0.173E+0	0.297E+0	0.587E+0	-0.202E+0
	A_1	-1.5	-0.287E+1	-0.283E+1	-0.278E+1	-0.253E+1	-0.290E+1	-0.929E+0
	A_2	0.7	0.374E+1	0.365E+1	0.323E+1	0.304E+1	0.365E+1	0.816E-1
	A_3		-0.245E+1	-0.236E+1	-0.183E+1	-0.192E+1	-0.225E+1	0.128E+0
	A_4		0.697E+0	0.665E+0	0.435E+0	0.563E+0	0.593E+0	0.605E-1
SECOND SUB-SYSTEM	B_0	2.5	0.250E+1	0.250E+1	0.250E+1	0.243E+1	0.253E+1	0.258E+1
	B_1	1.25	-0.267E+1	-0.173E+1	-0.137E+1	0.155E+1	0.500E+1	0.378E+1
	B_2	0.1	0.598E+0	-0.313E+0	0.710E+0	-0.185E+0	0.493E+1	0.275E+1
	B_3		0.107E+1	0.416E+0	0.914E+0	-0.217E+0	0.209E+1	0.617E+0
	B_4		0.983E-1	0.519E-1	0.848E-1	-0.220E-1	0.448E+0	-0.405E+0
	A_1	-1.0	-0.257E+1	-0.219E+1	-0.206E+1	-0.880E+0	0.464E+0	-0.636E-1
	A_2	0.5	0.305E+1	0.212E+1	0.235E+1	0.199E+0	0.174E+0	-0.626E-2
	A_3		-0.177E+1	-0.103E+1	-0.131E+1	0.250E+0	-0.230E+0	0.208E-1
	A_4		0.492E+0	0.215E+0	0.391E+0	-0.965E-1	0.447E+0	0.111E+0
	$\overline{Y^2}$		0.734E+2	0.735E+2	0.623E+2	0.501E+2	0.805E+2	0.141E+3
	$\overline{(Y-\hat{Y})^2}$		0.246E-8	0.632E-2	0.615E-1	0.547E+0	0.638E+1	0.648E+2
NUMBER OF ITERATIONS			31	24	32	16	82	21

Table 5.11 Estimation results for the case of white noise disturbances and independent white input signals, for a wrong model order; the model order is taken as 4.

FIGURES

Fig. 2.7-a

Pole-zero plot in the z-plane of the process parameters.

The process is a fourth order autoregression with the following parameters:

$a_0 = 1.0$

$a_1 = 0.8$

$a_2 = 0.7$

$a_3 = 0.6$

$a_4 = 0.5$

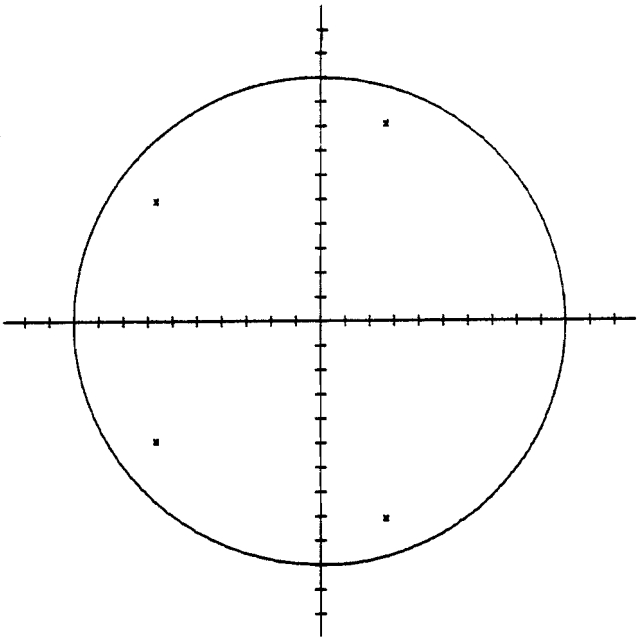


Fig. 2.7-b

Pole-zero plot in the z-plane of the process parameters.

The process is a fourth order autoregression with the following parameters:

$a_0 = 1.0$

$a_1 = -2.0$

$a_2 = 1.7525$

$a_3 = -0.7525$

$a_4 = 0.12625$

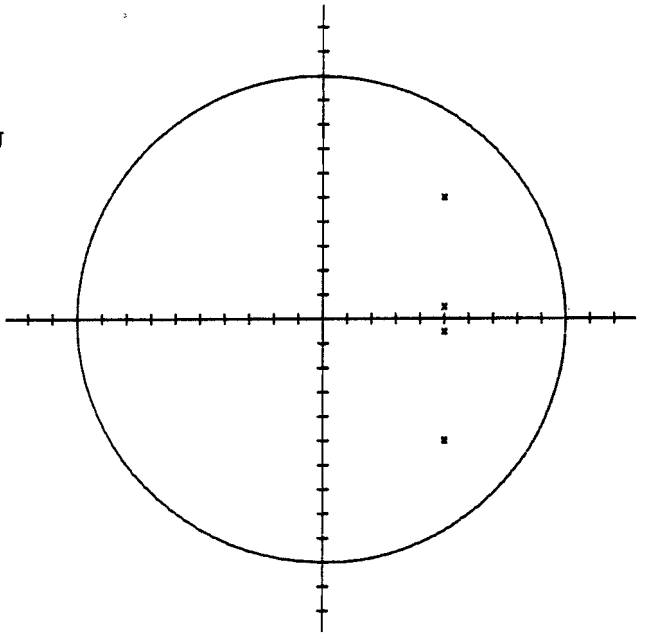


Fig. 2.7-c

Pole-zero plot in the z-plane of the process parameters.

The process is a fourth order autoregression with the following parameters:

$a_0 = 1.0$

$a_1 = 0.0$

$a_2 = 0.0$

$a_3 = 0.0$

$a_4 = 0.0625$

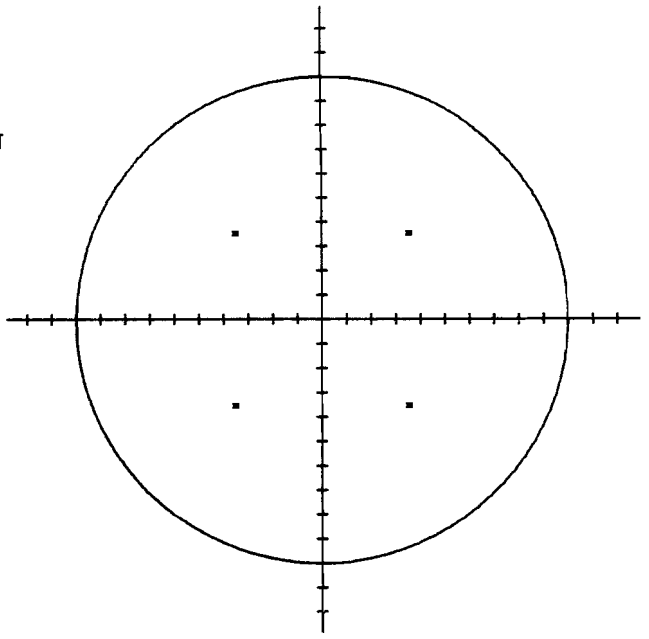


Fig. 2.8-a
Pole-zero plot in the z-plane of the process parameters.
The process has an ARMA(2,2) configuration with the zeros of the MA part close to zero. The parameters are:

$a_0 = 1.0$
$a_1 = 0.3$
$a_2 = 0.7$
$b_0 = 1.0$
$b_1 = -0.2$
$b_2 = 0.02$

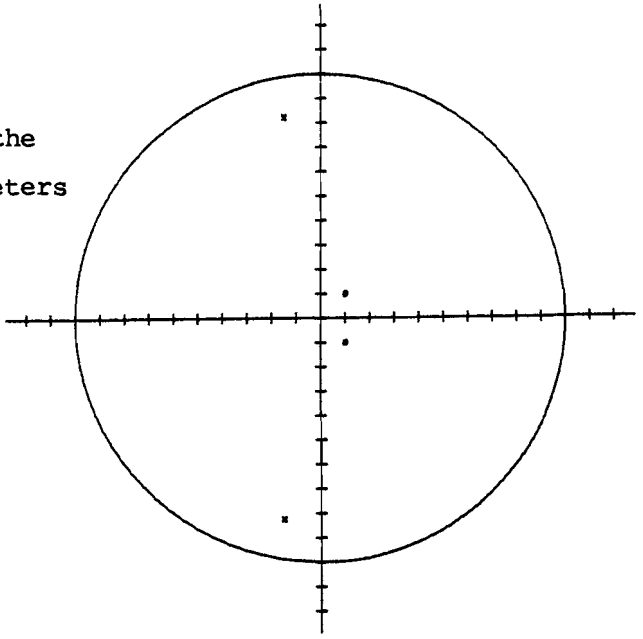


Fig. 2.8-b
Pole-zero plot in the z-plane of the process parameters.
The process has an ARMA(2,2) configuration with the zeros of the MA part close to zero. The parameters are:

$a_0 = 1.0$
$a_1 = -1.0$
$a_2 = 0.5$
$b_0 = 1.0$
$b_1 = 0.4$
$b_2 = 0.05$

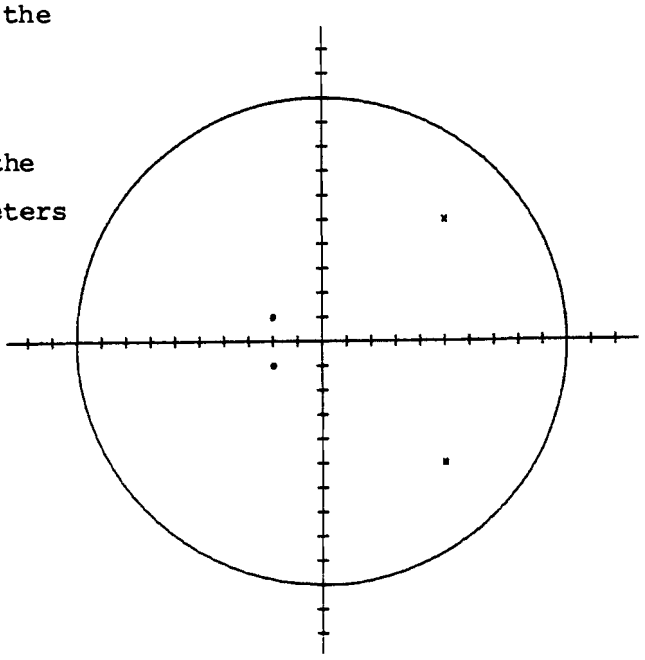


Fig. 2.8-c
Pole-zero plot in the z-plane of the process parameters.
The process has an ARMA(2,2) configuration with the zeros of the MA part close to the unit circle.
The parameters are:

$a_0 = 1.0$
$a_1 = -1.0$
$a_2 = 0.5$
$b_0 = 1.0$
$b_1 = 0.3$
$b_2 = 0.7$

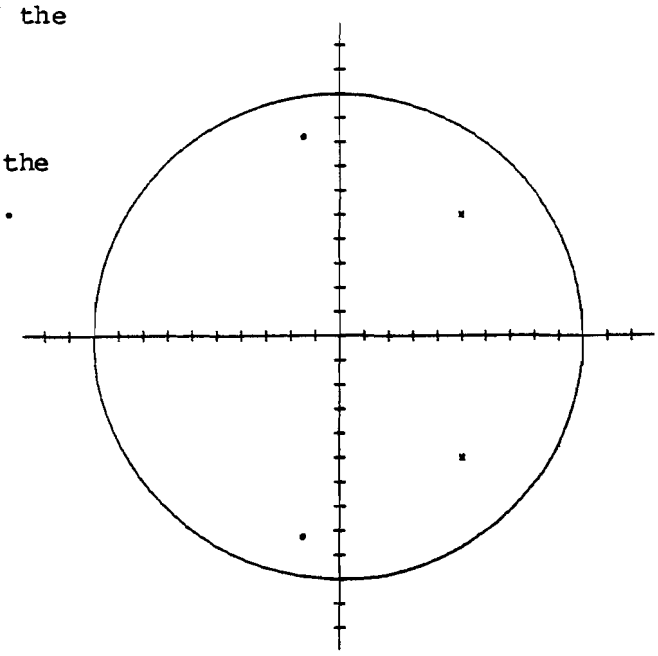


Fig. 2.8-d
Pole-zero plot in the z-plane of the process parameters
The process has an ARMA(2,2) configuration with the zeros of the MA part close to the unit circle.
The parameters are:

$a_0 = 1.0$
$a_1 = -1.5$
$a_2 = 0.7$
$b_0 = 1.0$
$b_1 = 0.3$
$b_2 = 0.7$

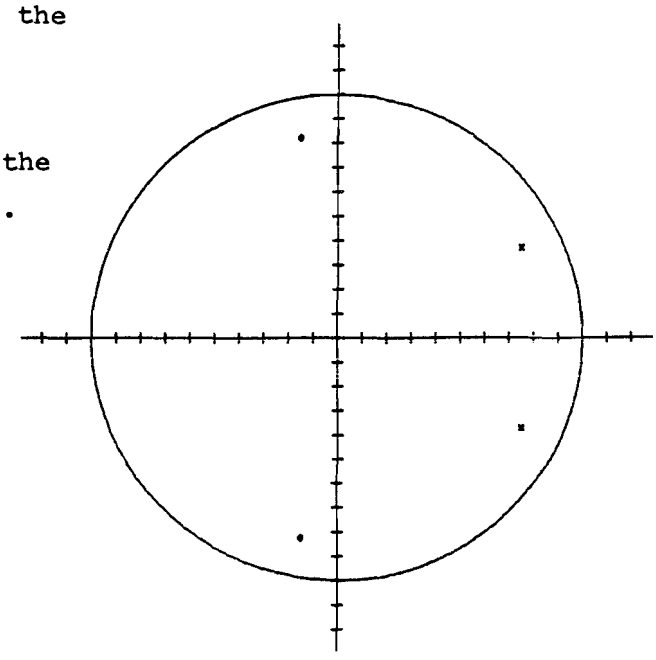
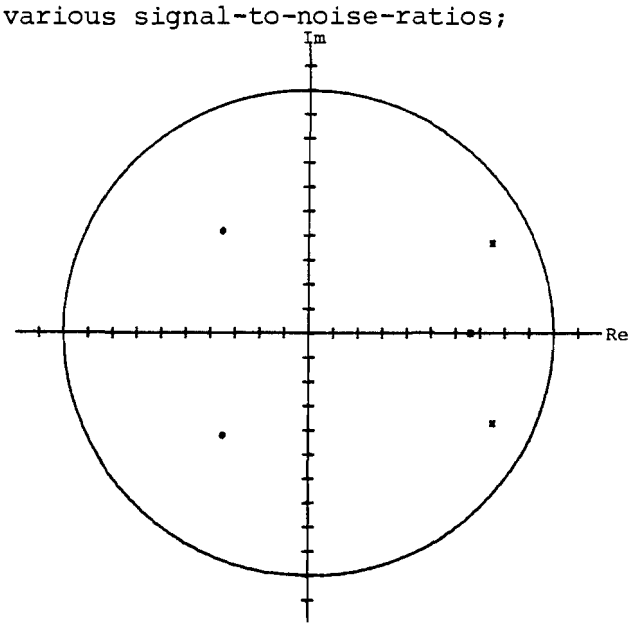
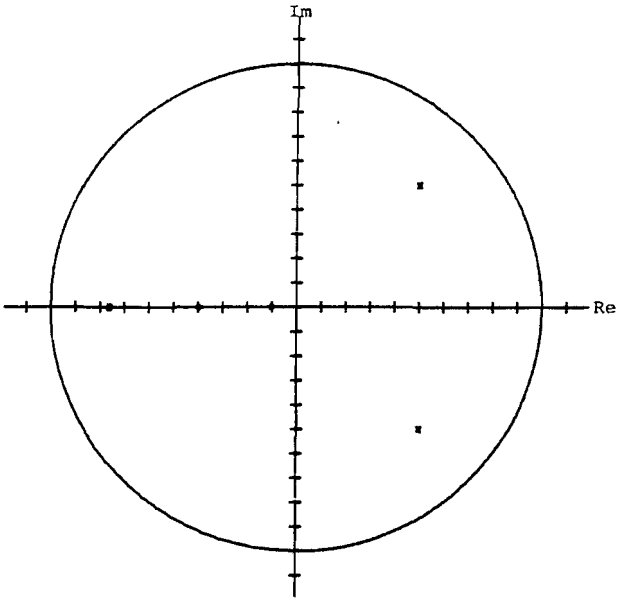


Fig. 5.6-a to 5.6-l

Pole-zero plots of the estimations with a third order model for various signal-to-noise-ratios;

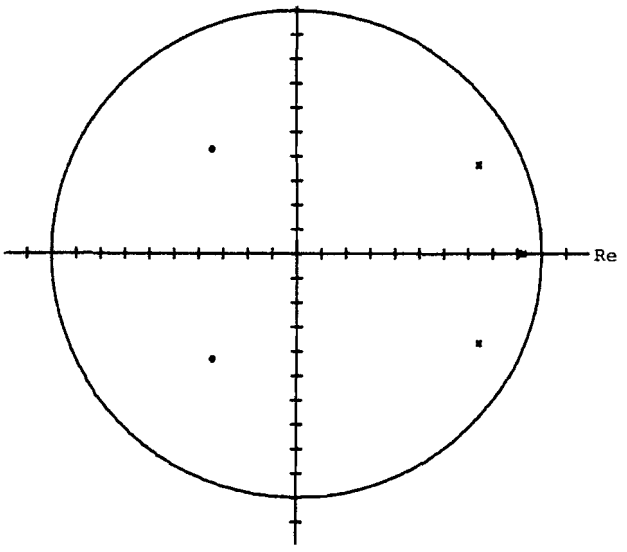


A S/N = INF

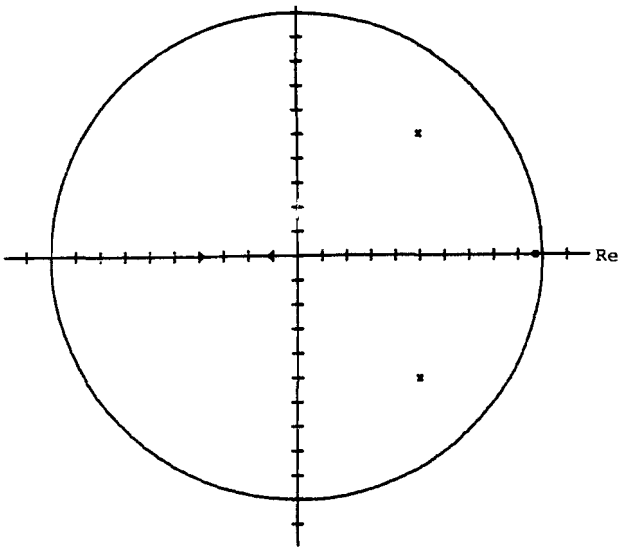


G S/N = INF

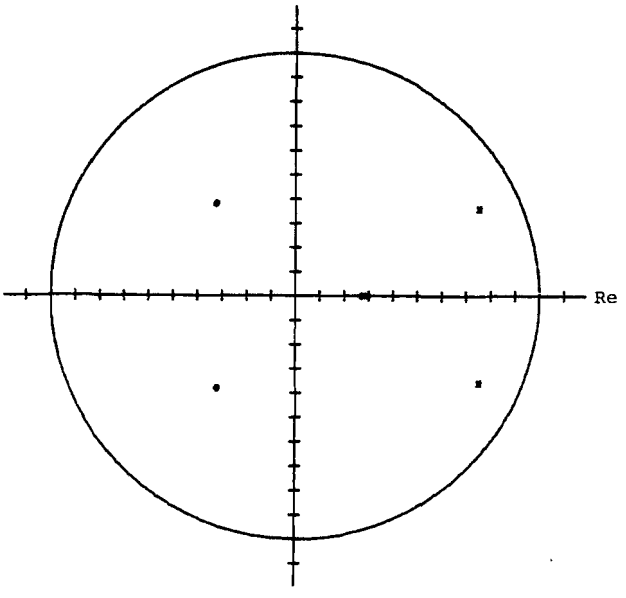
a to f concern the first SISO subsystem g to l concern the second SISO subsyste



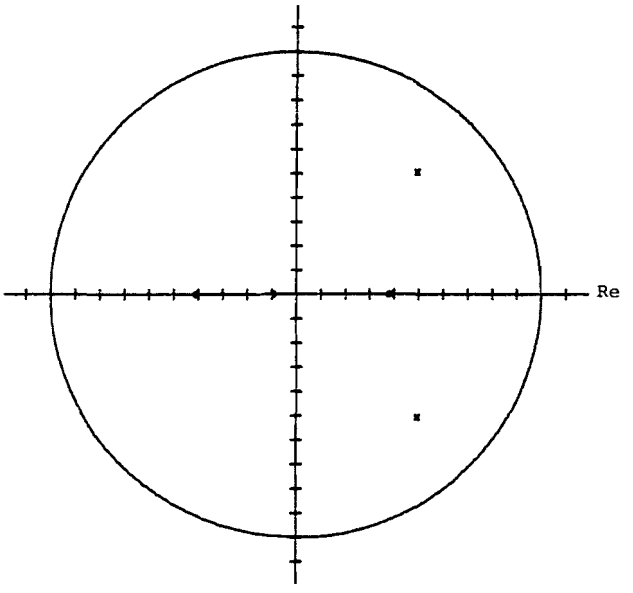
B S/N = 40 dB



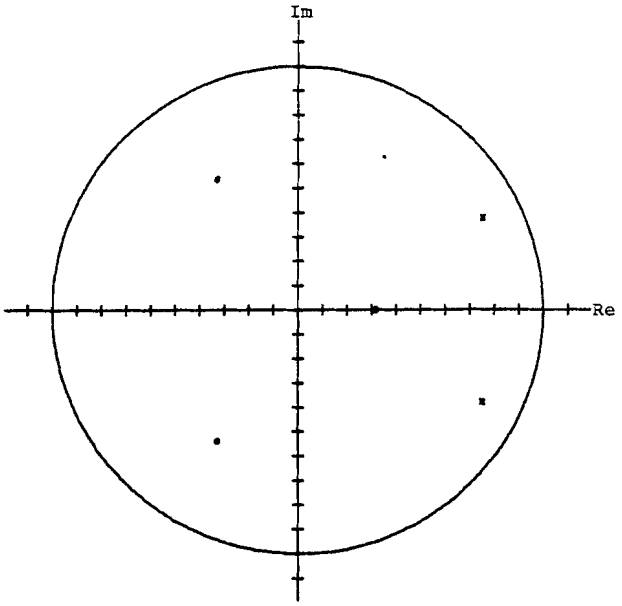
H S/N = 40 dB



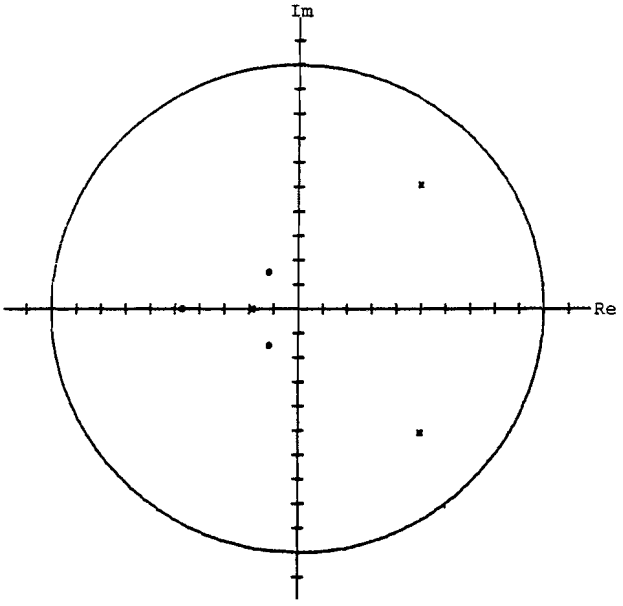
C S/N = 30 dB



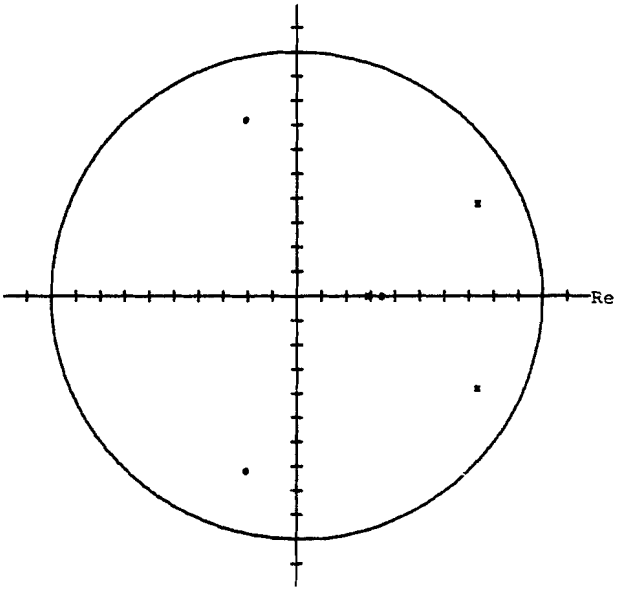
I S/N = 30 dB



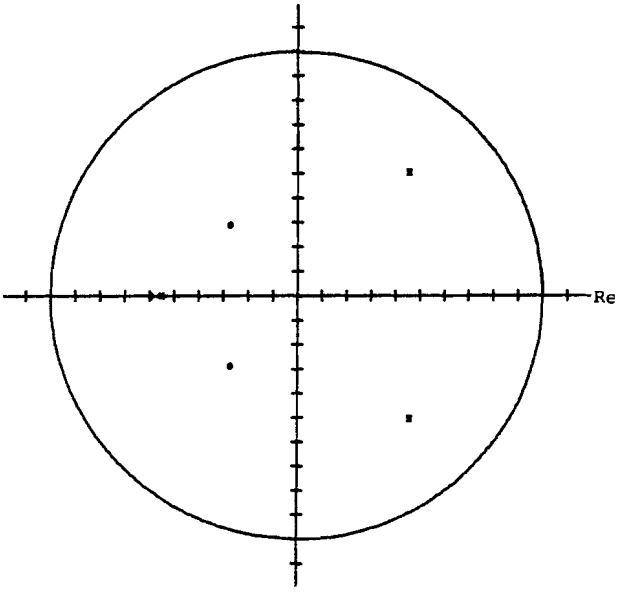
D S/N = 20 dB



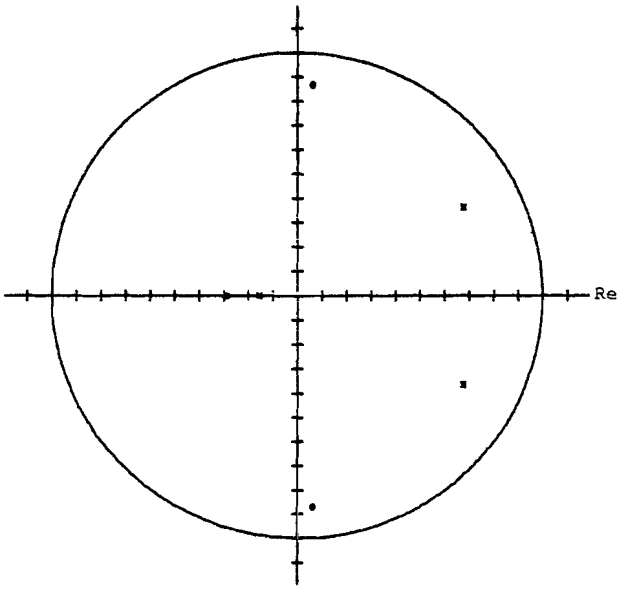
J S/N = 20 dB



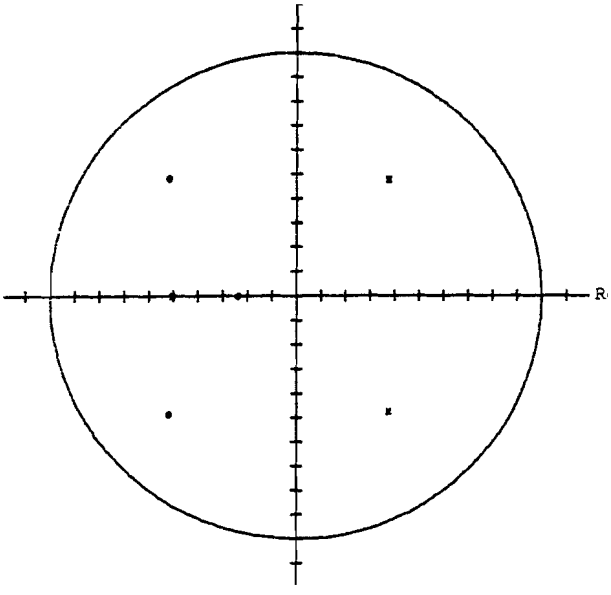
E S/N = 10 dB



K S/N = 10 dB

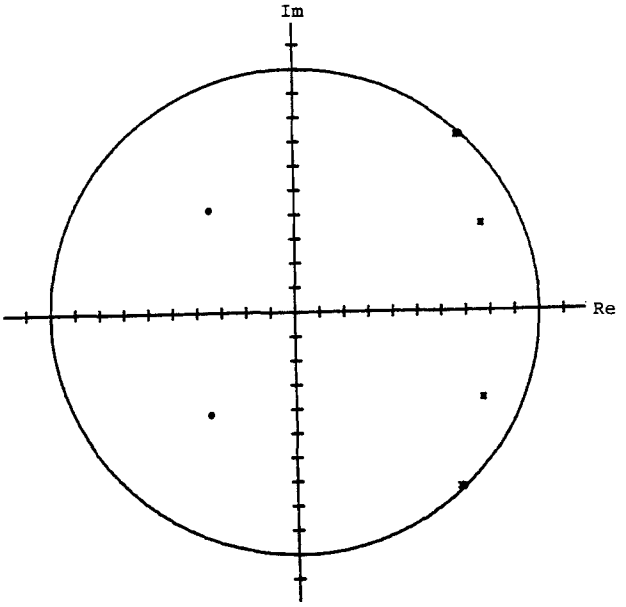


F S/N = 0 dB

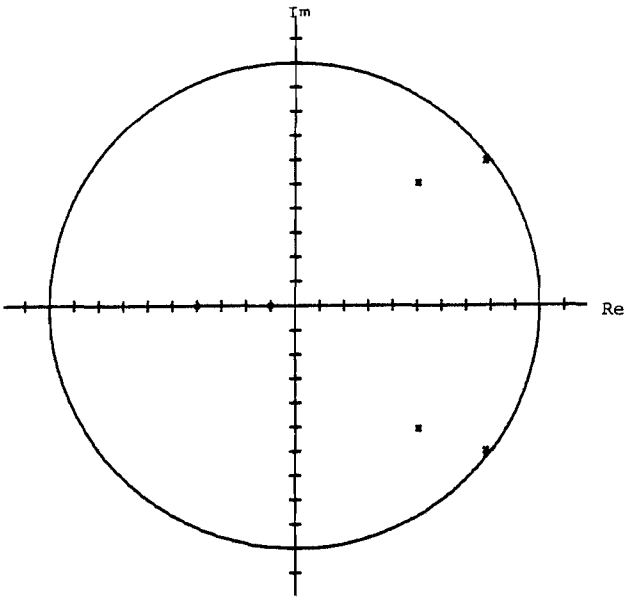


L S/N = 0 dB

Pole-zero plots of the estimations with a fourth order model for various signal-to-noise-ratios;

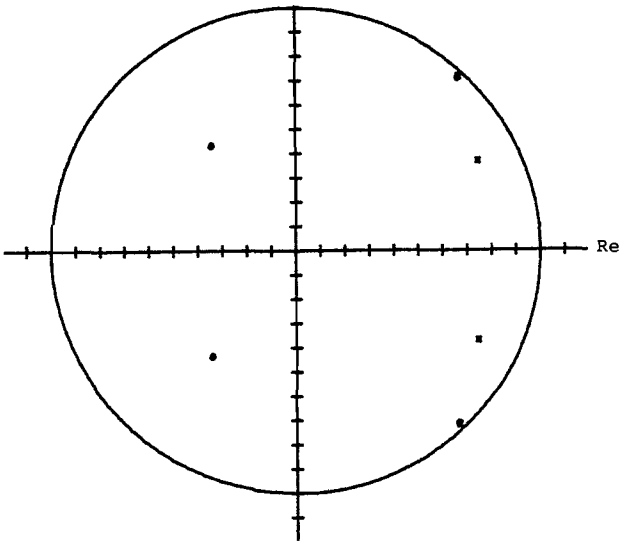


A S/N = INF

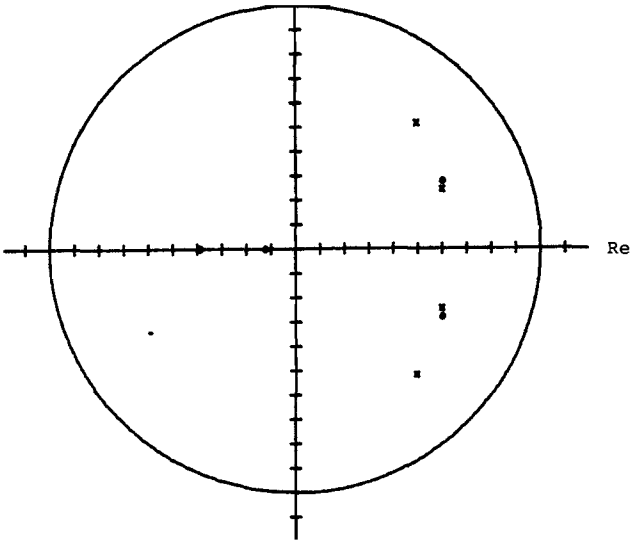


G S/N = INF

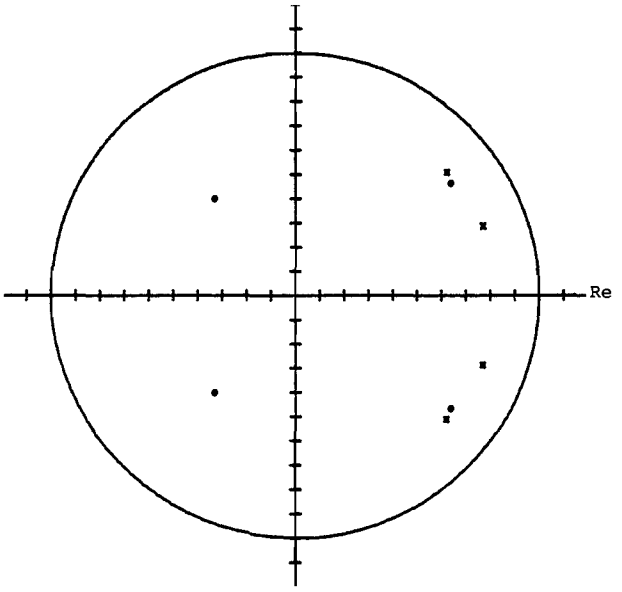
a to f concern the first SISO subsystem, g to l concern the second SISO subsystem



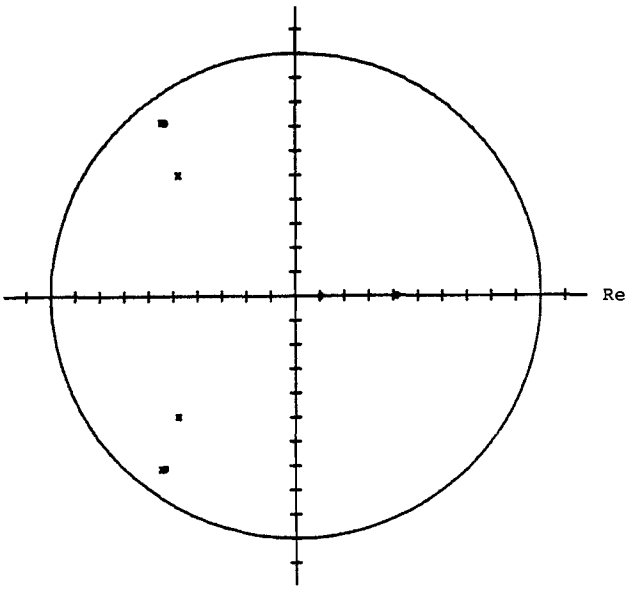
B S/N = 40 dB



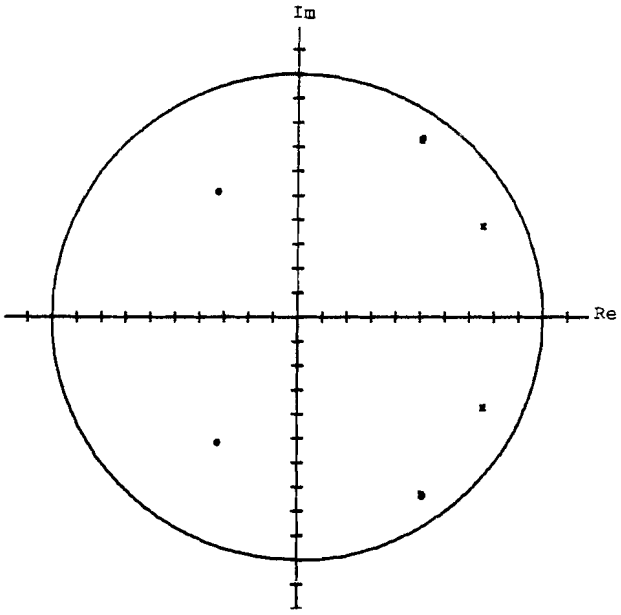
H S/N = 40 dB



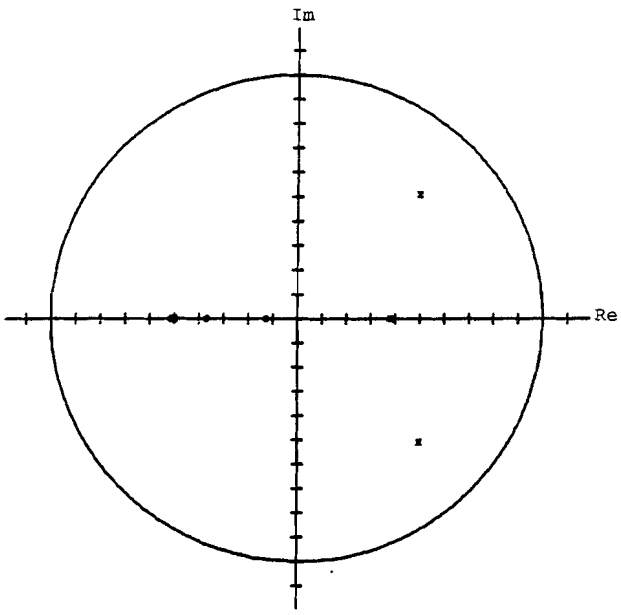
C S/N = 30 dB



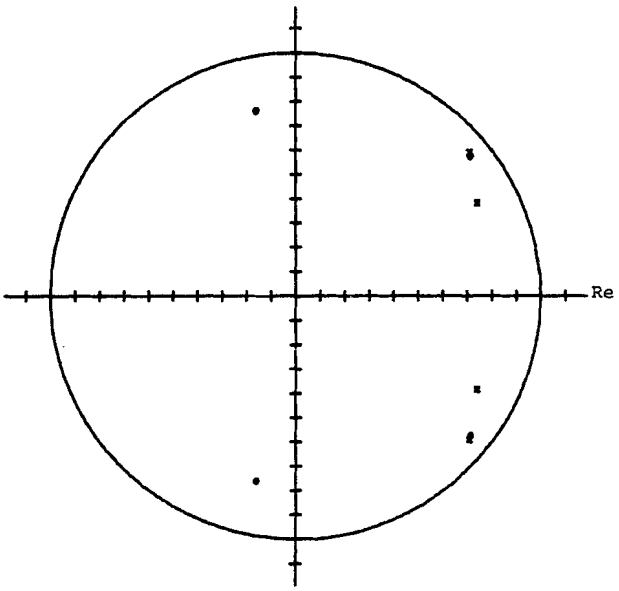
I S/N = 30 dB



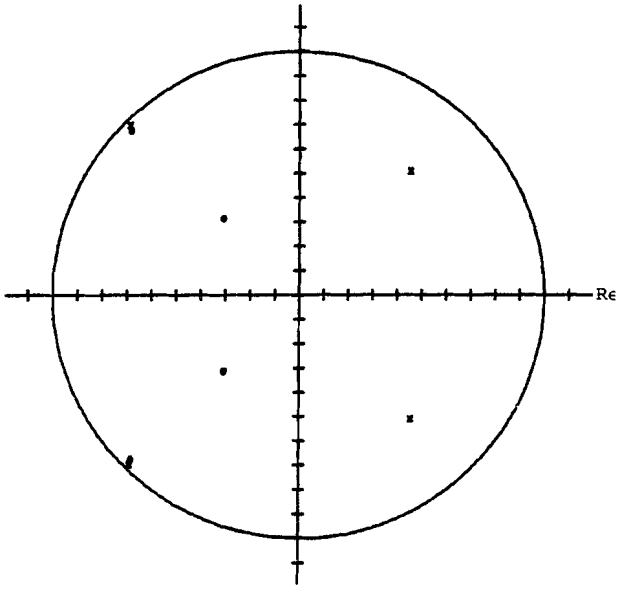
D S/N = 20 dB



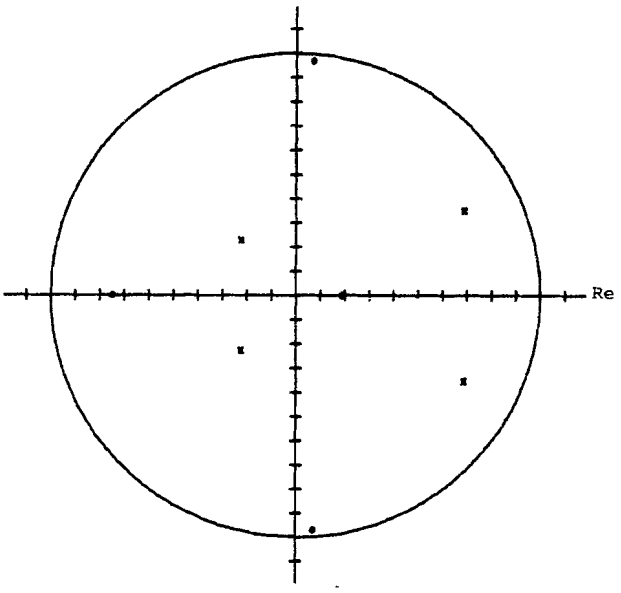
J S/N = 20 dB



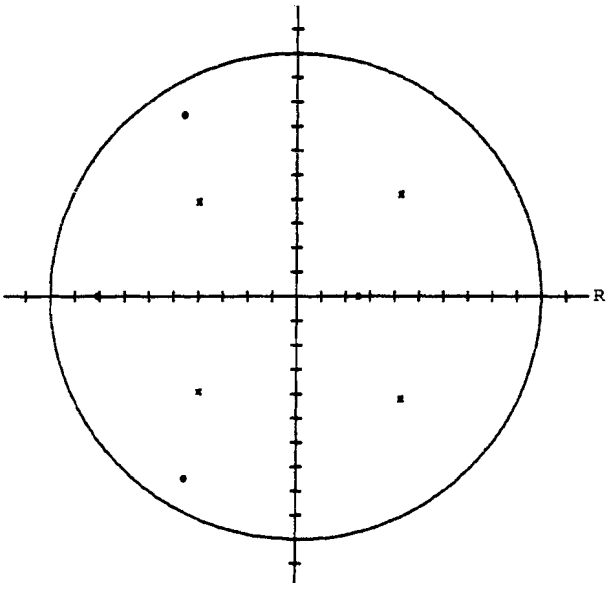
E S/N = 10 dB



K S/N = 10 dB



F S/N = 0 dB



L S/N = 0 dB

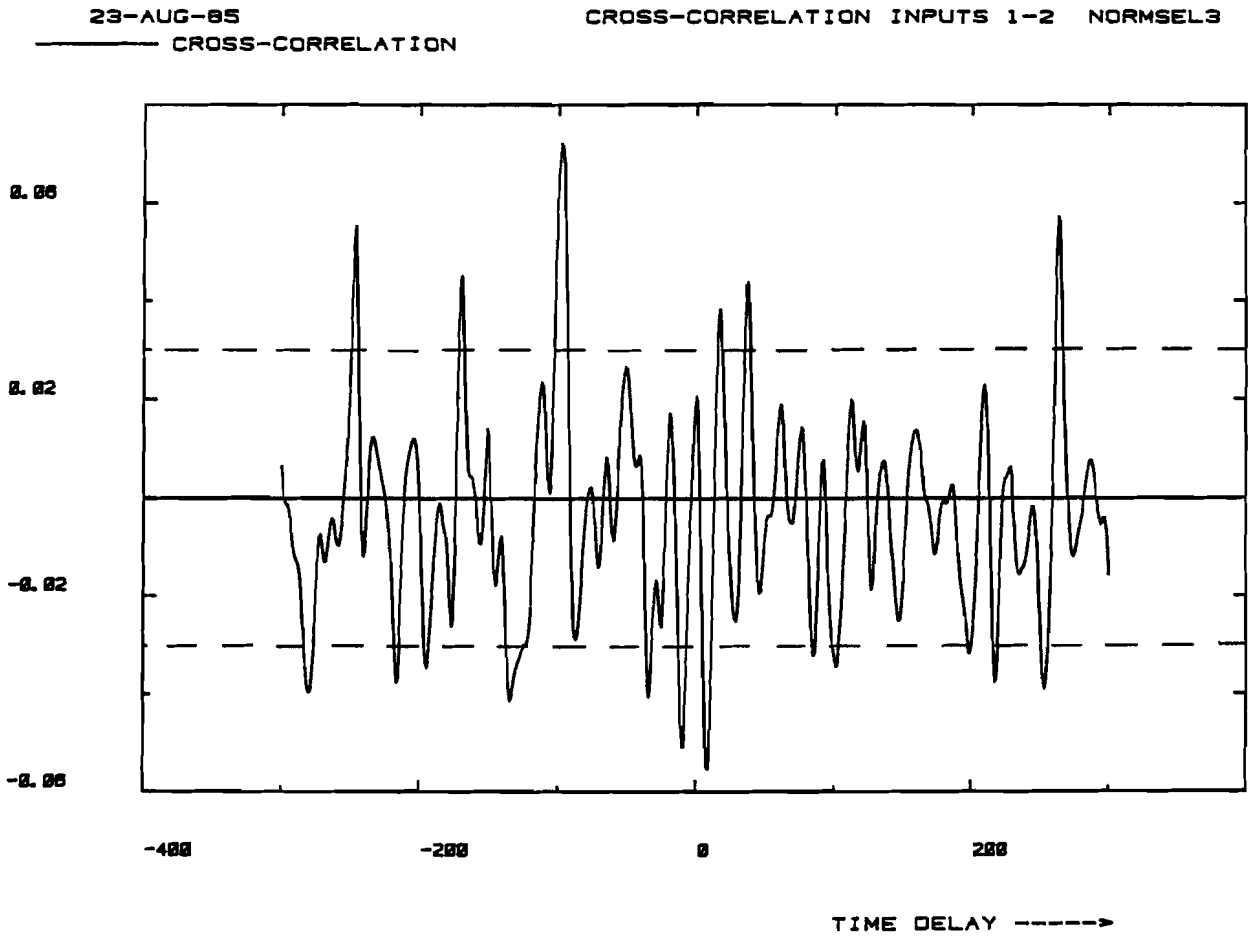


Fig. 6.7 Cross-correlation of the two input signals.

Fig. 6.8-a Autocorrelation of the first input signal; the gasflow (sample frequency 0.5 Hz).

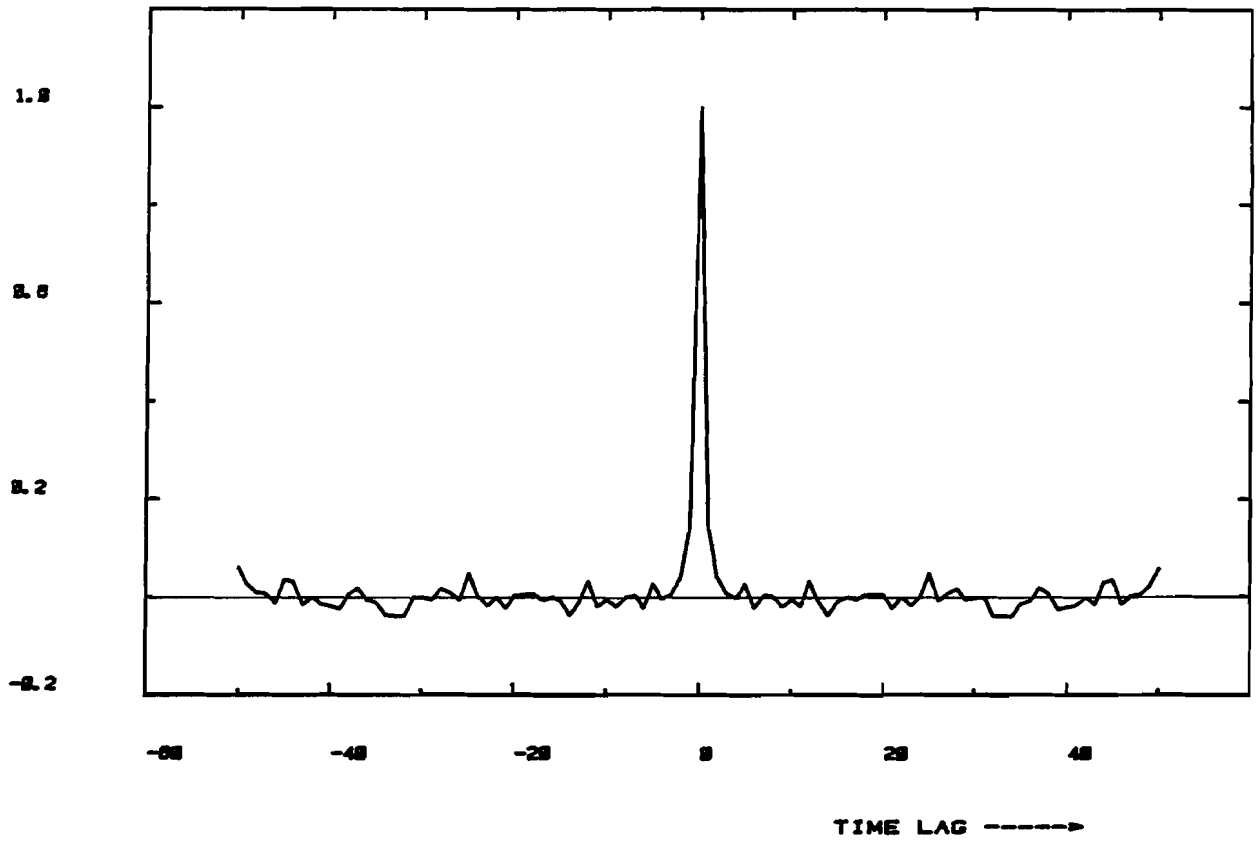


Fig. 6.8-b Autocorrelation of the second input signal; the drawing speed (sample frequency 0.5 Hz).

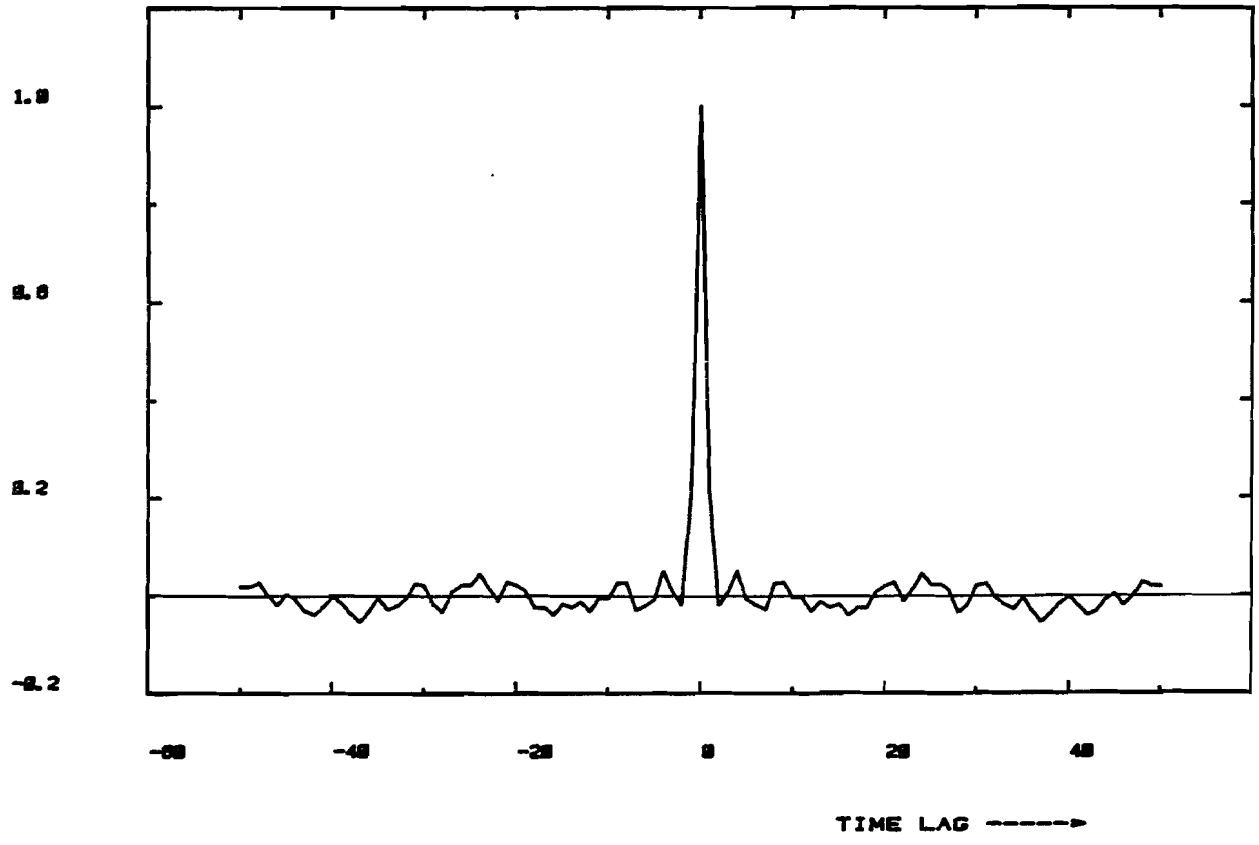


Fig. 6.9-a Cross-correlation of the first input signal (the gas flow) and the first output signal (the wall thickness) (sample frequency 0.5 Hz).

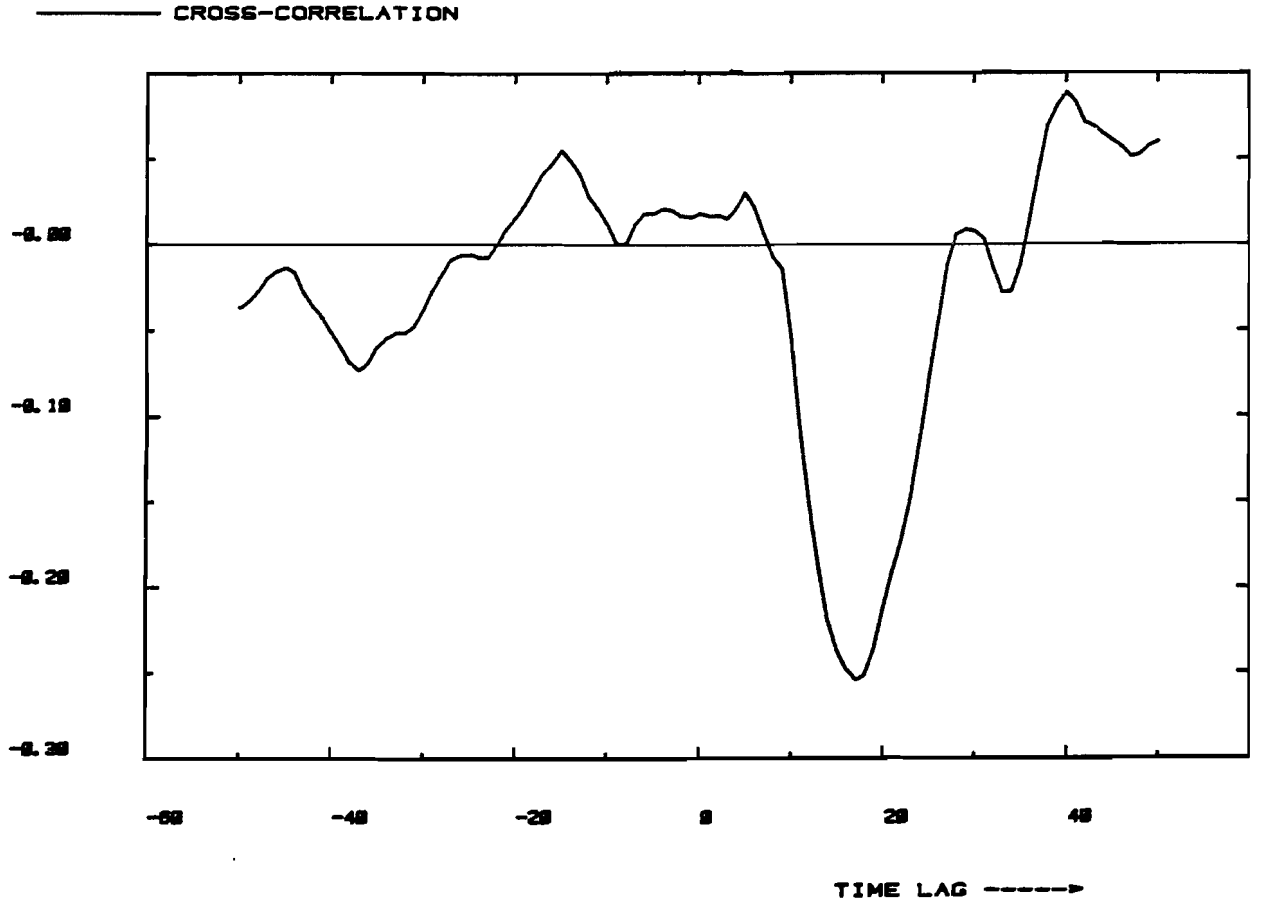


Fig. 6.9-b Cross-correlation of the second input signal (the drawing speed) and the first output signal (the wall thickness) (sample frequency 0.5 Hz).

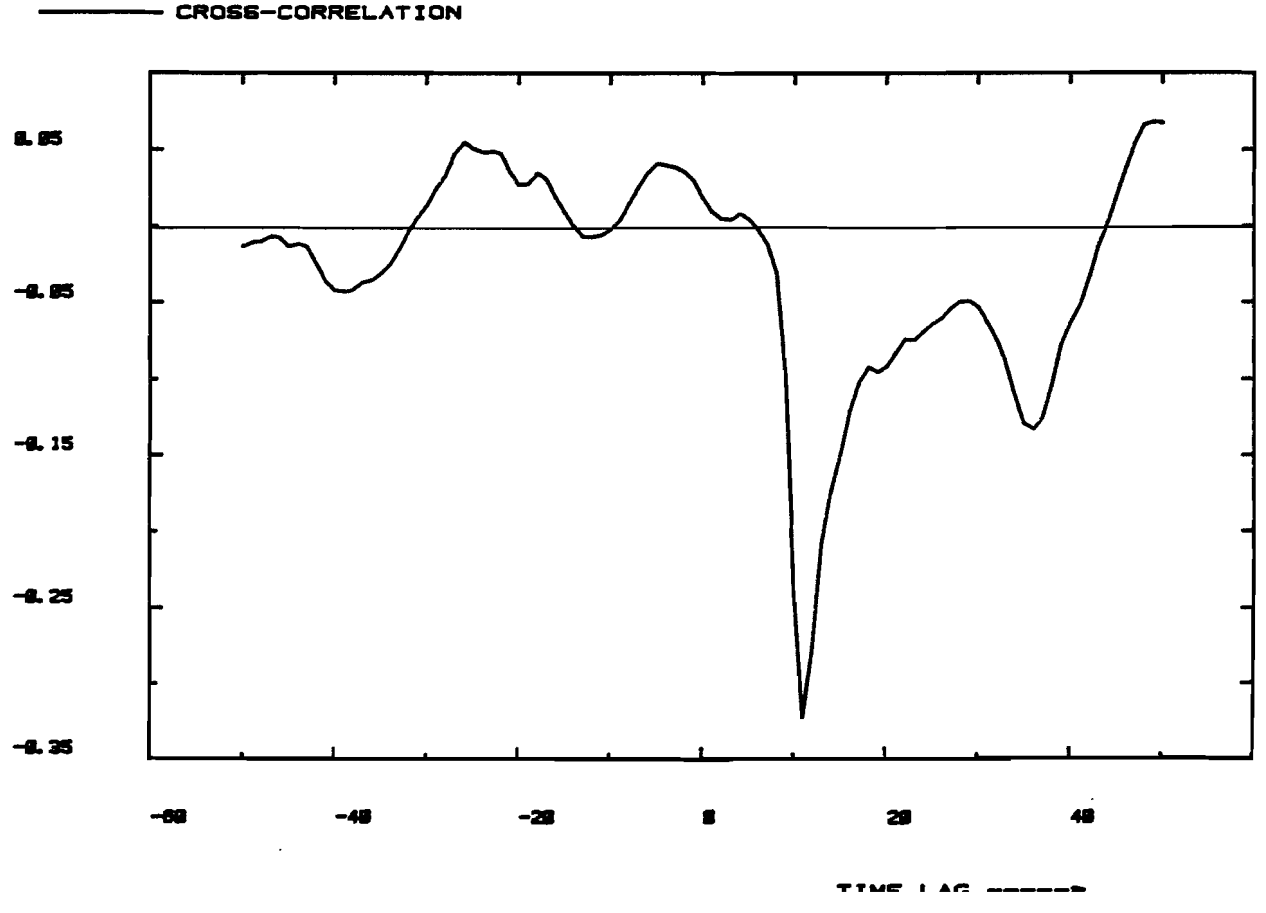


Fig. 6.9-c Cross-correlation of the first input signal (the gas flow) and the second output signal (the diameter of the tube)

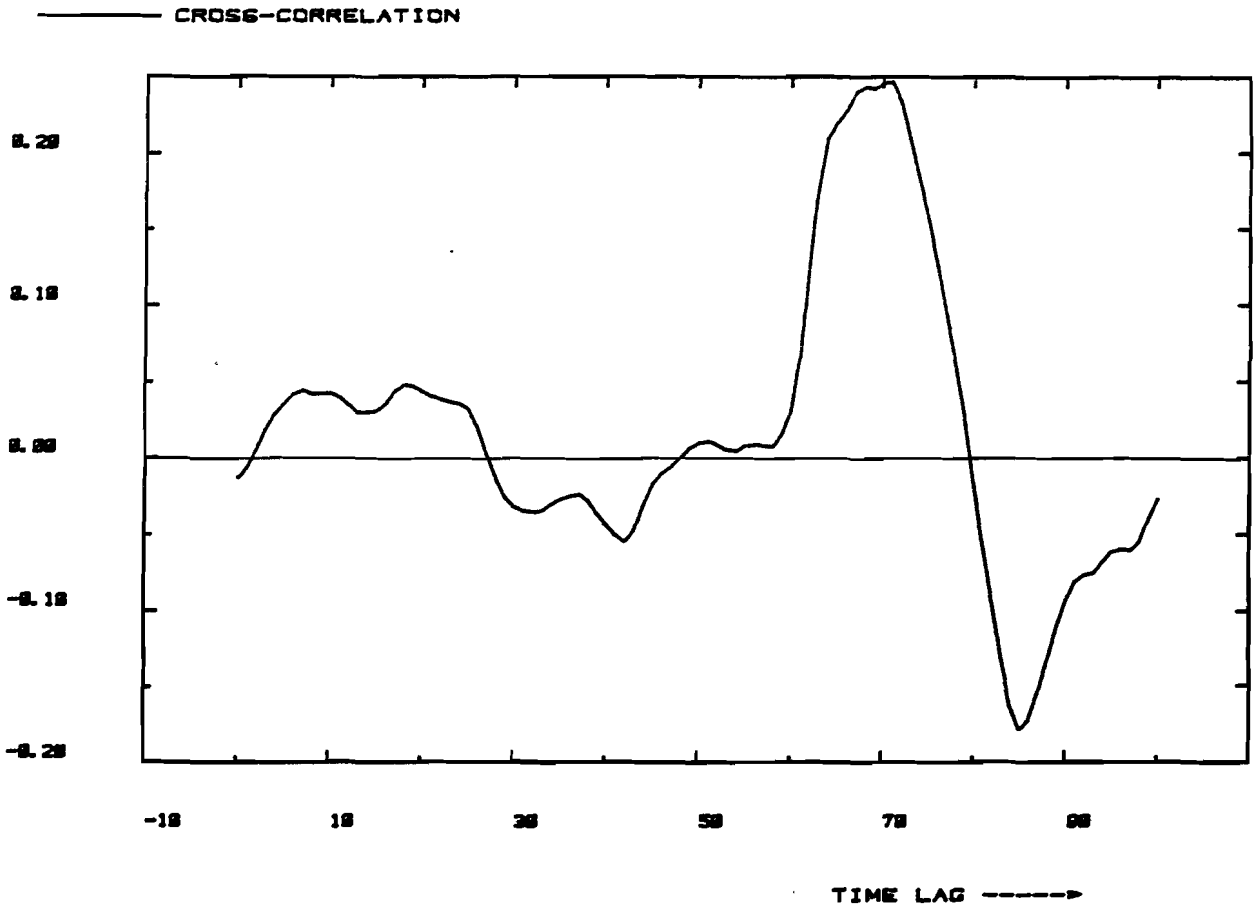


Fig. 6.9-d Cross-correlation of the second input signal (the drawing speed) and the second output signal (the diameter of the tube) (sample frequency 0.5 Hz).

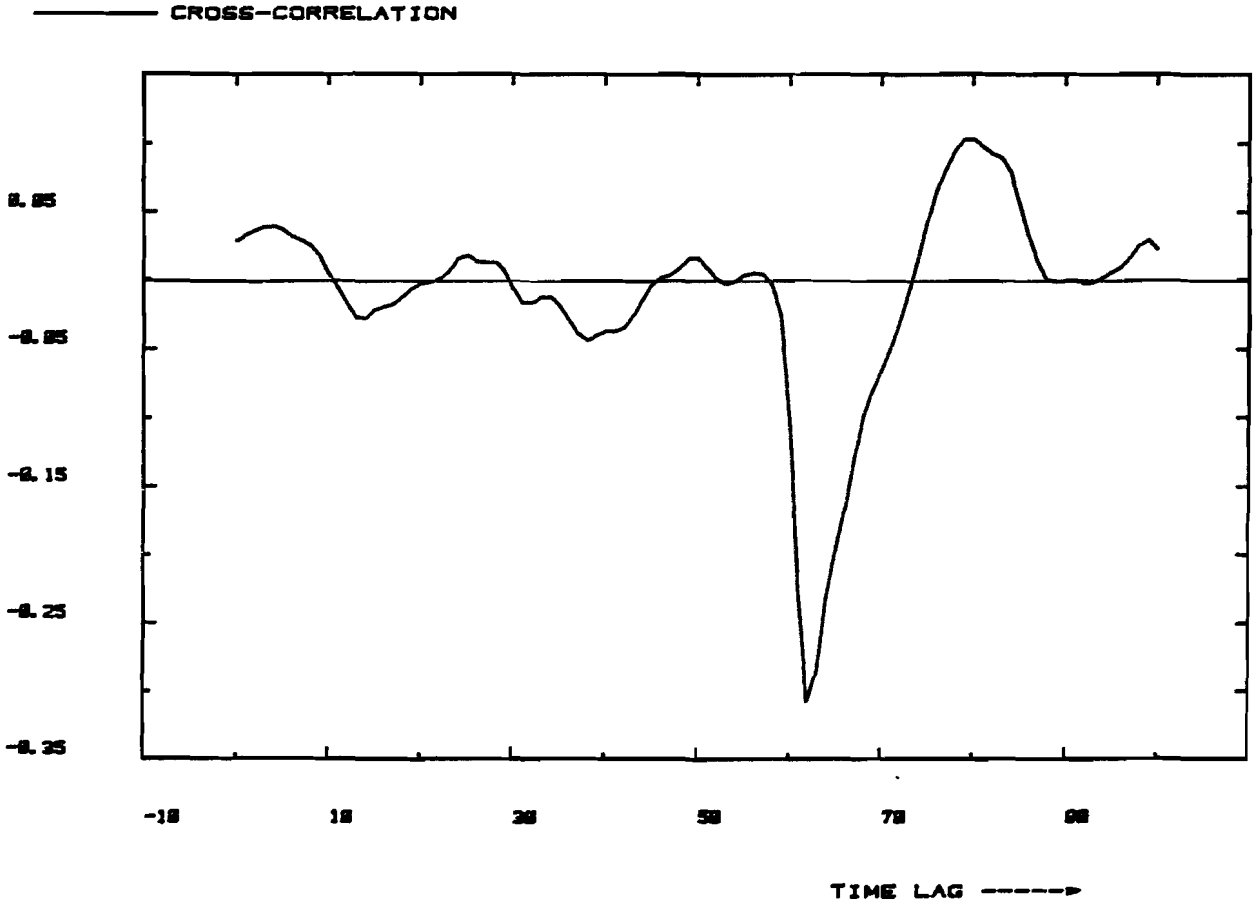


Fig. 6.10-a Determination of the time delay in the transfer of input 1 to output 1;

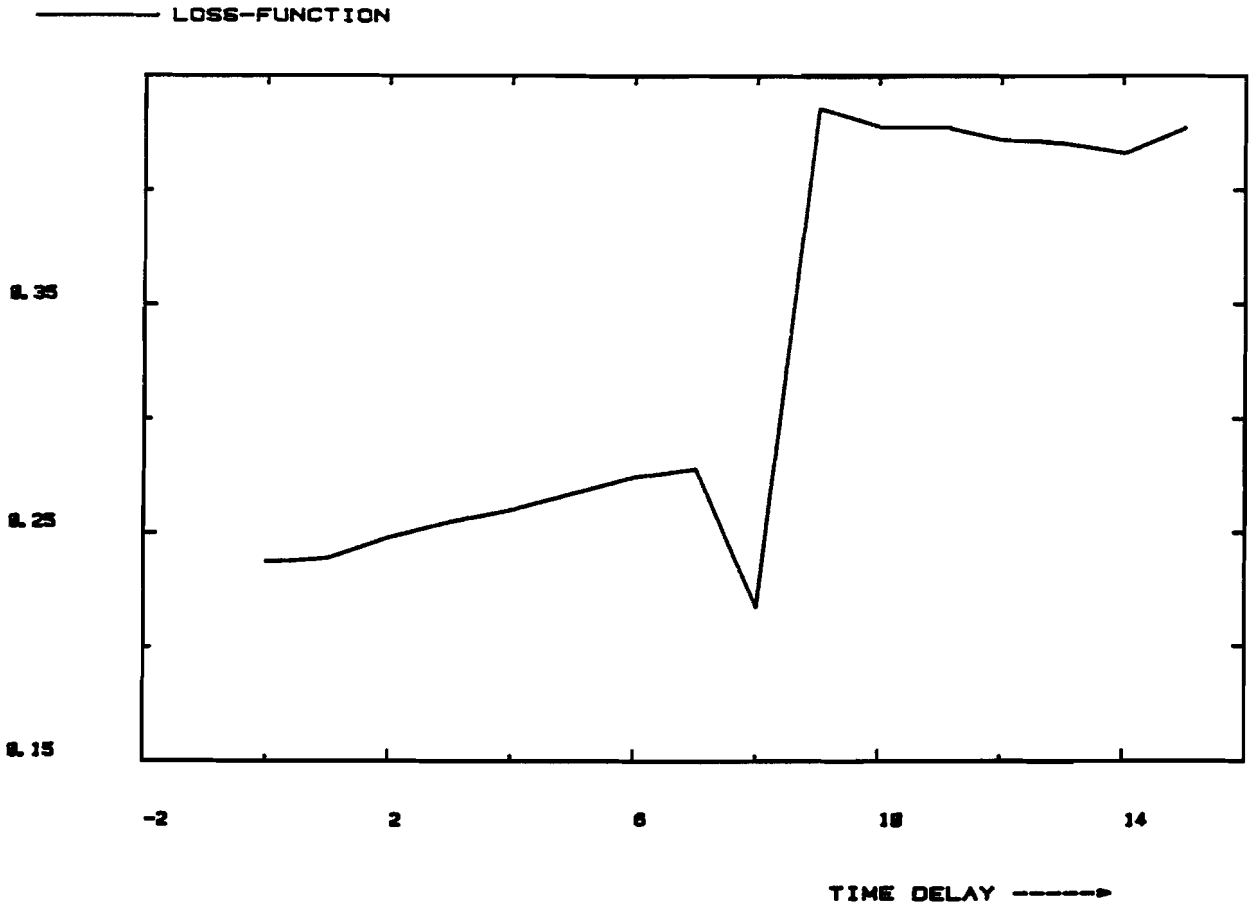
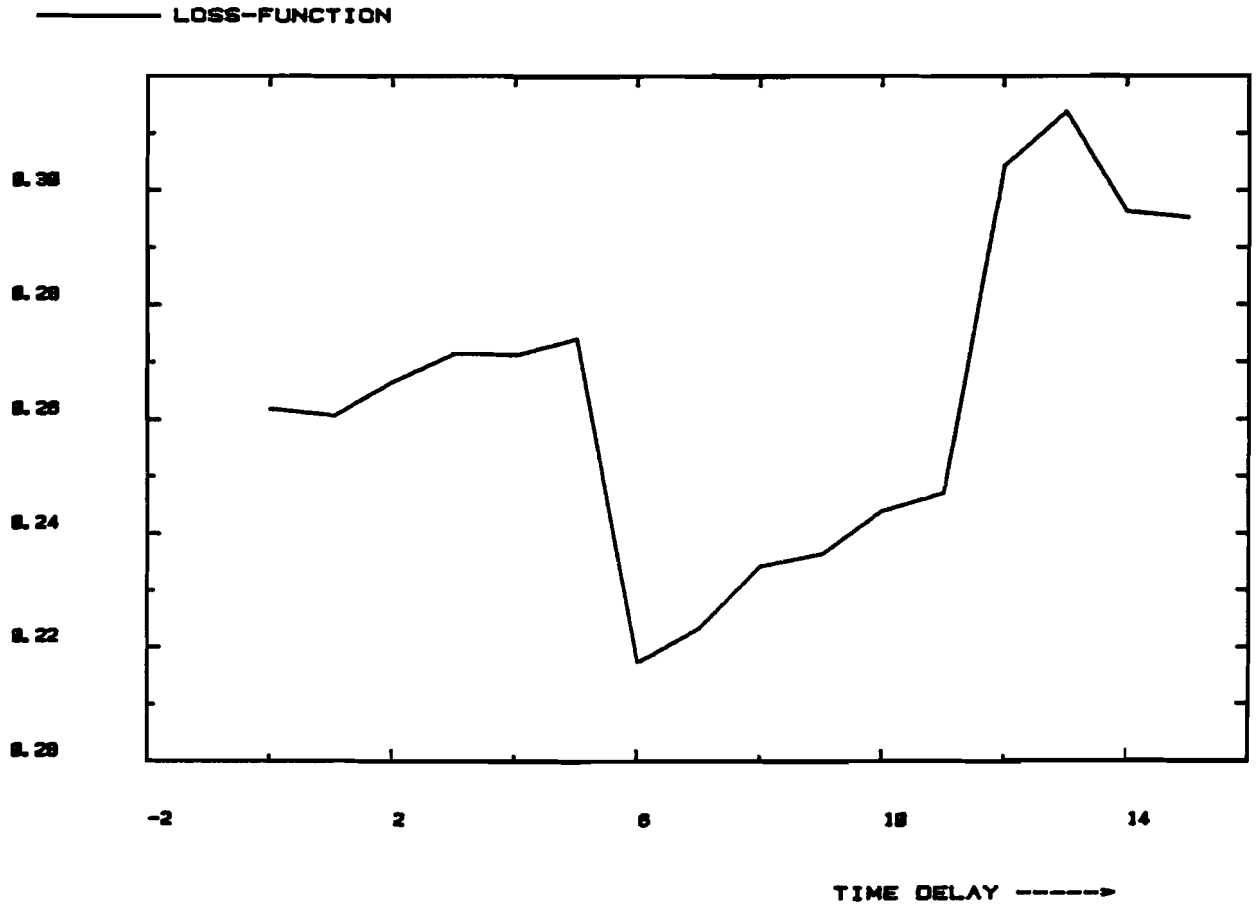


Fig. 6.10-b Determination of the time delay in the transfer of input 2 to output 1; the minimum value of the loss-function indicates the time delay.



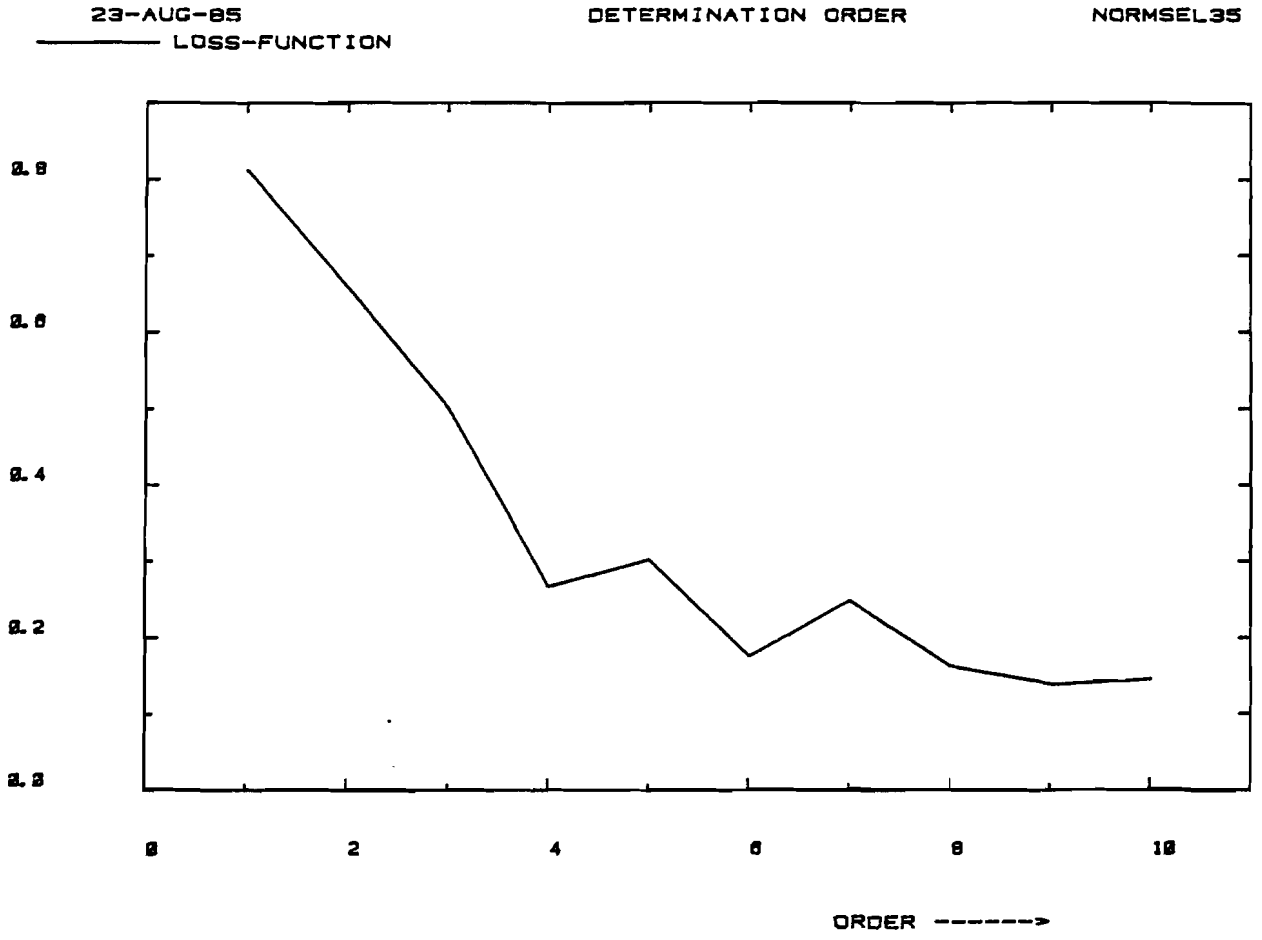


Fig. 6.11 Curve of the loss-function as a function of the model order.

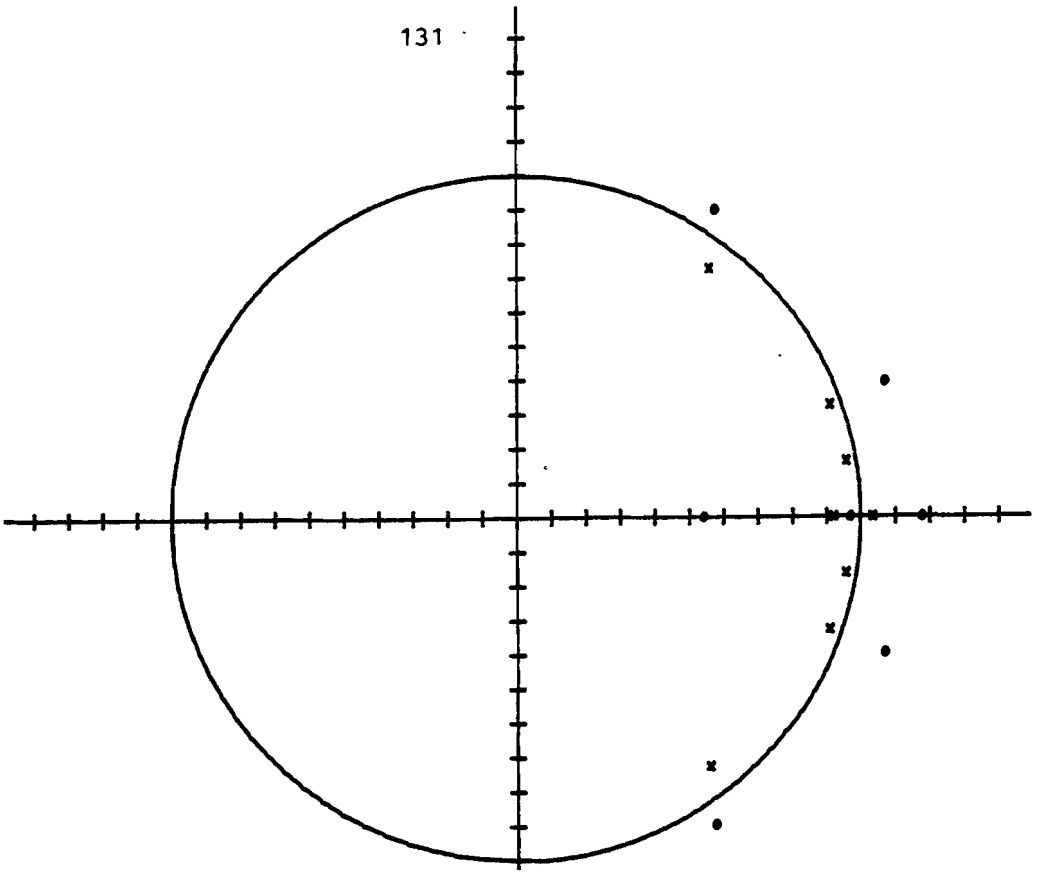


Fig. 6.12-a Pole-zero plot of the eighth order estimated model of the transfer input 1 to output 1. Due to numerical inaccuracy of the pole-zero plot program one pole is located just outside of the unit circle. One zero at (1.79,0) could not be plotted.

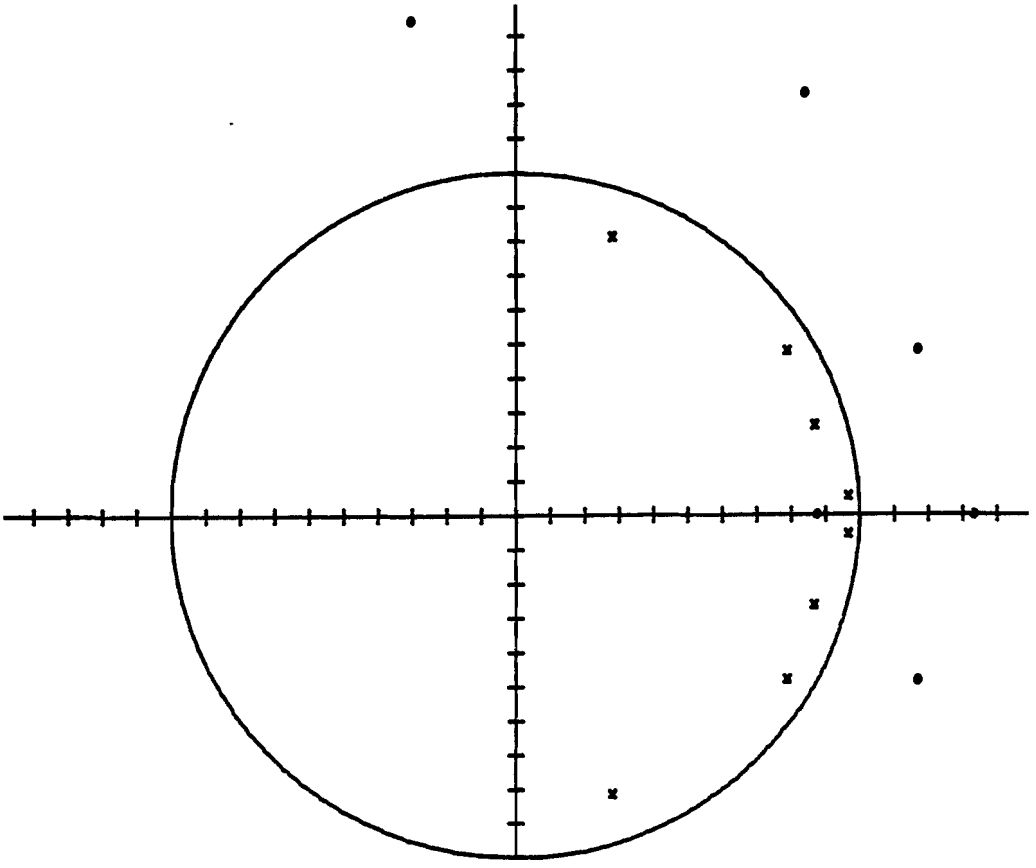


Fig. 6.12-b Pole-zero plot of the eighth order estimated model of the transfer input 2 to output 1.

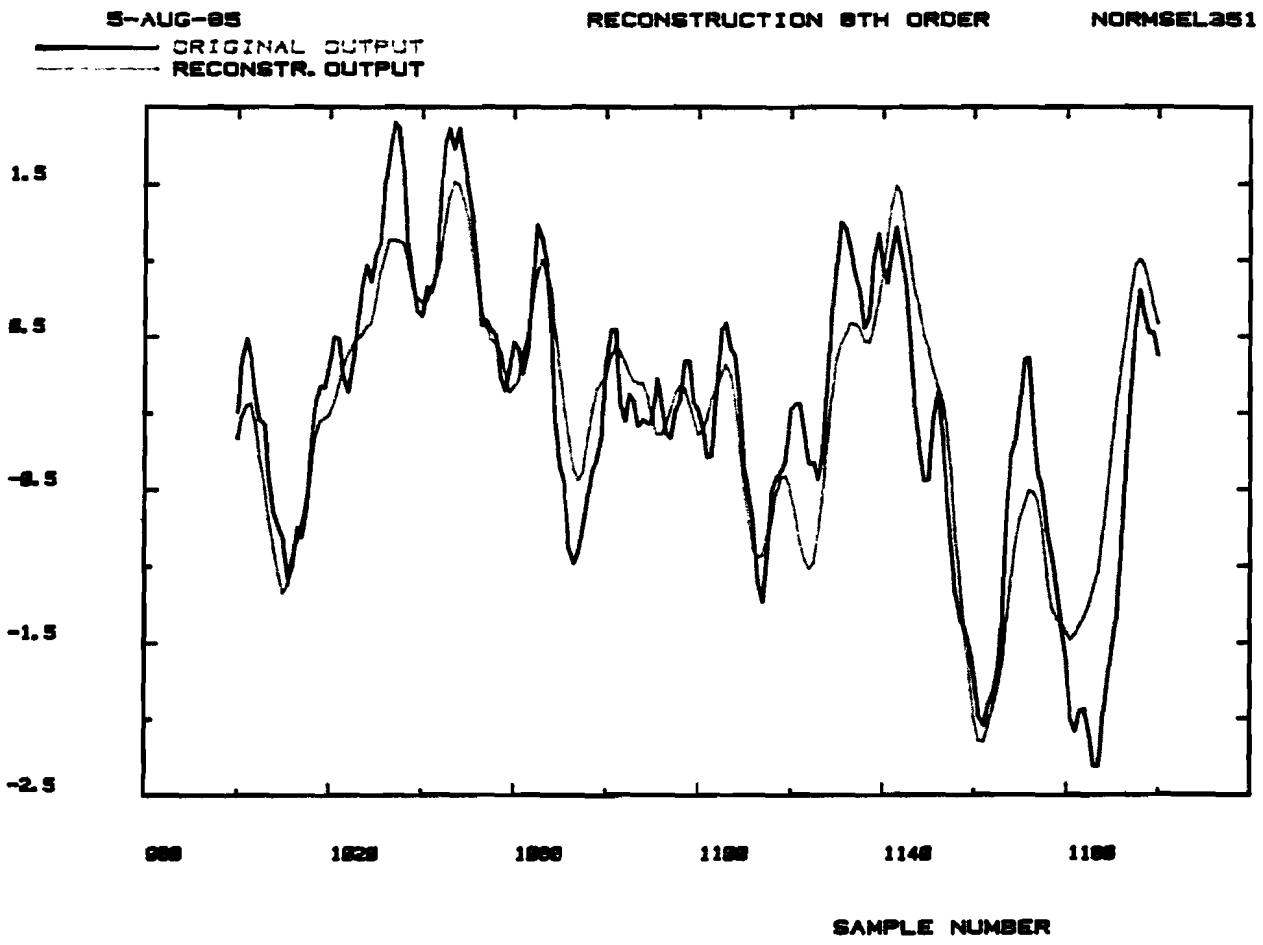


Fig. 6.13 Reconstruction of the first output signal (the wall thickness) with an eighth order MISO model.

Fig. 6.14-a Determination of the time delay in the transfer of input 1 to output 2;

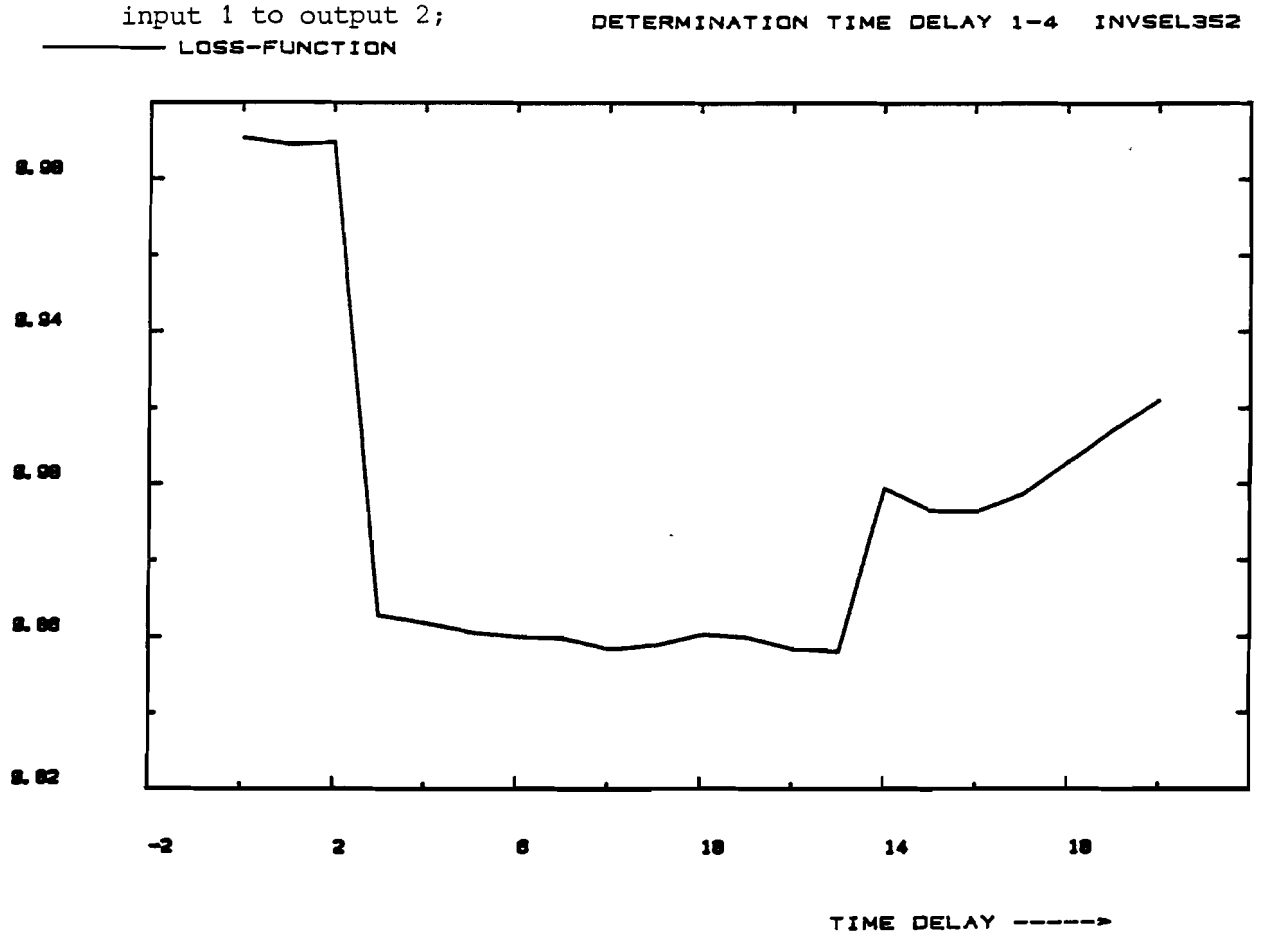
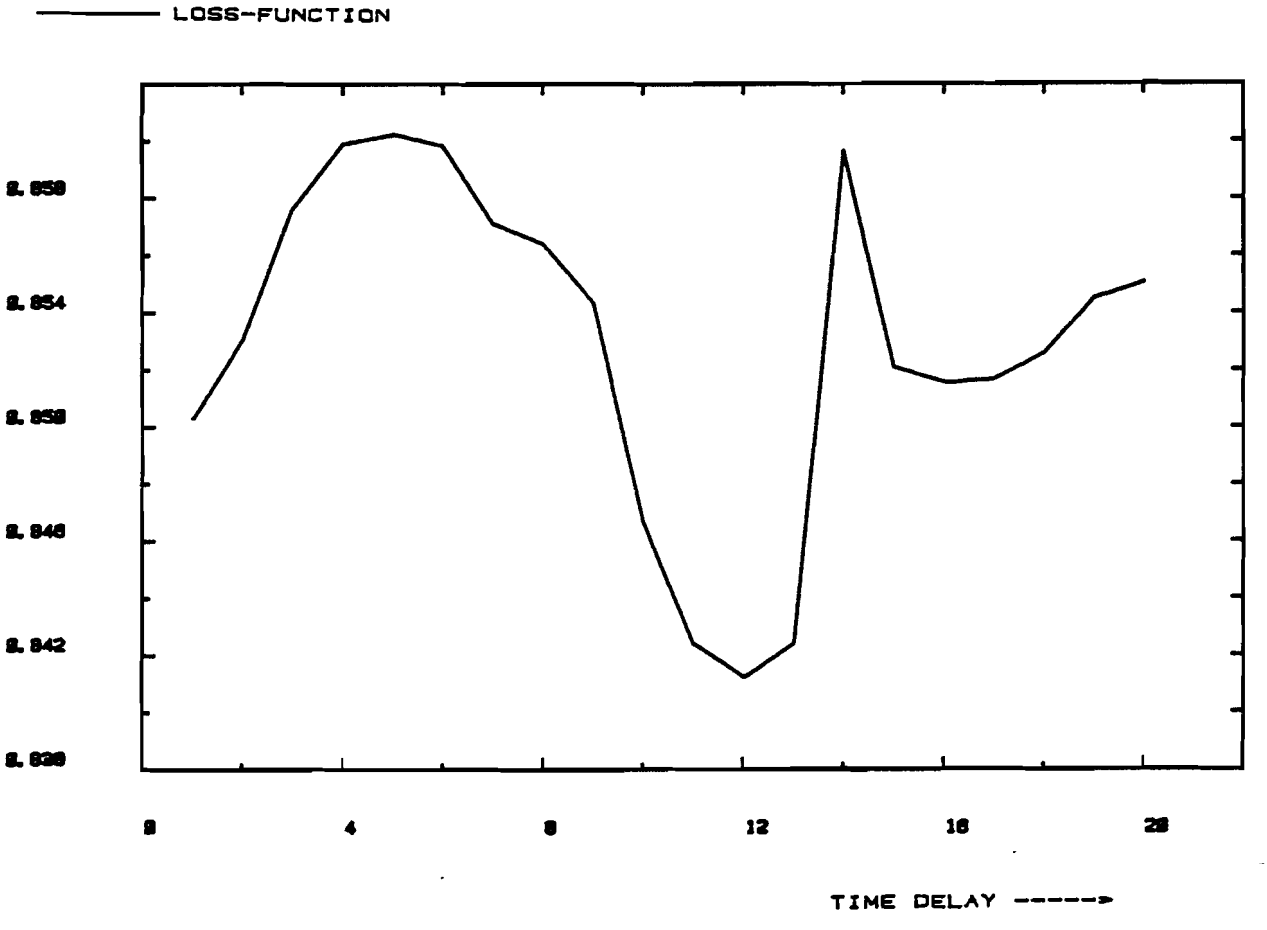


Fig. 6.14-b Determination of the time delay in the transfer of input 2 to output 2; the minimum value of the loss-function is supposed to indicate the correct time delay.



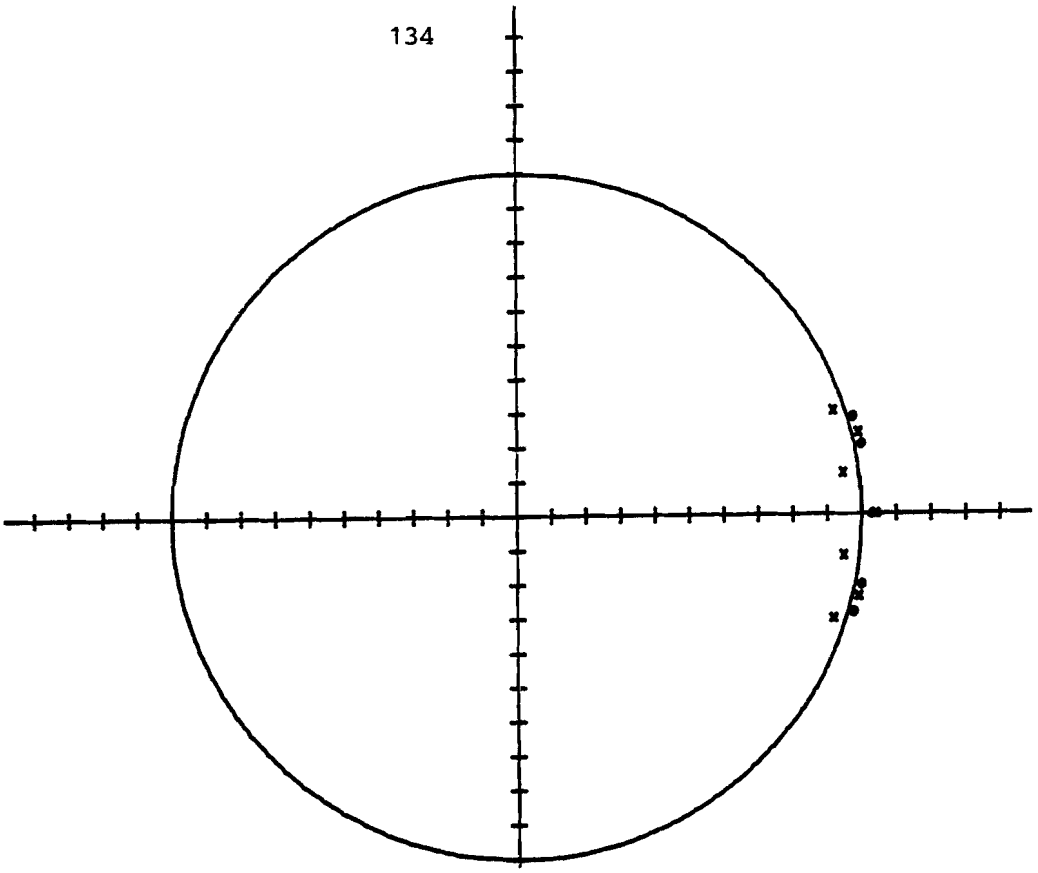


Fig. 6.15-a Pole-zero plot of the seventh order estimated model of the transfer input 1 to output 2. Due to numerical inaccuracy of the pole-zero plot program three poles are located just outside of the unit circle

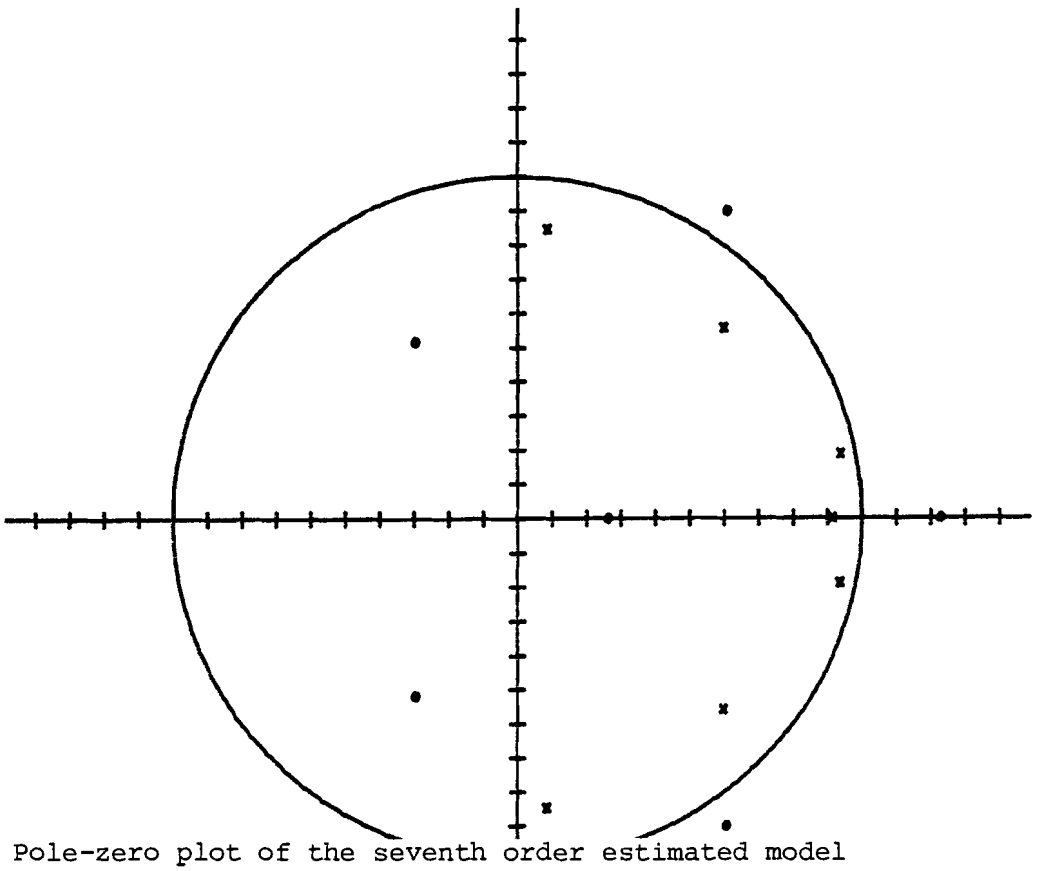


Fig. 6.15-b Pole-zero plot of the seventh order estimated model of the transfer input 2 to output 2. One zero at (2.14,0) could not be plotted.

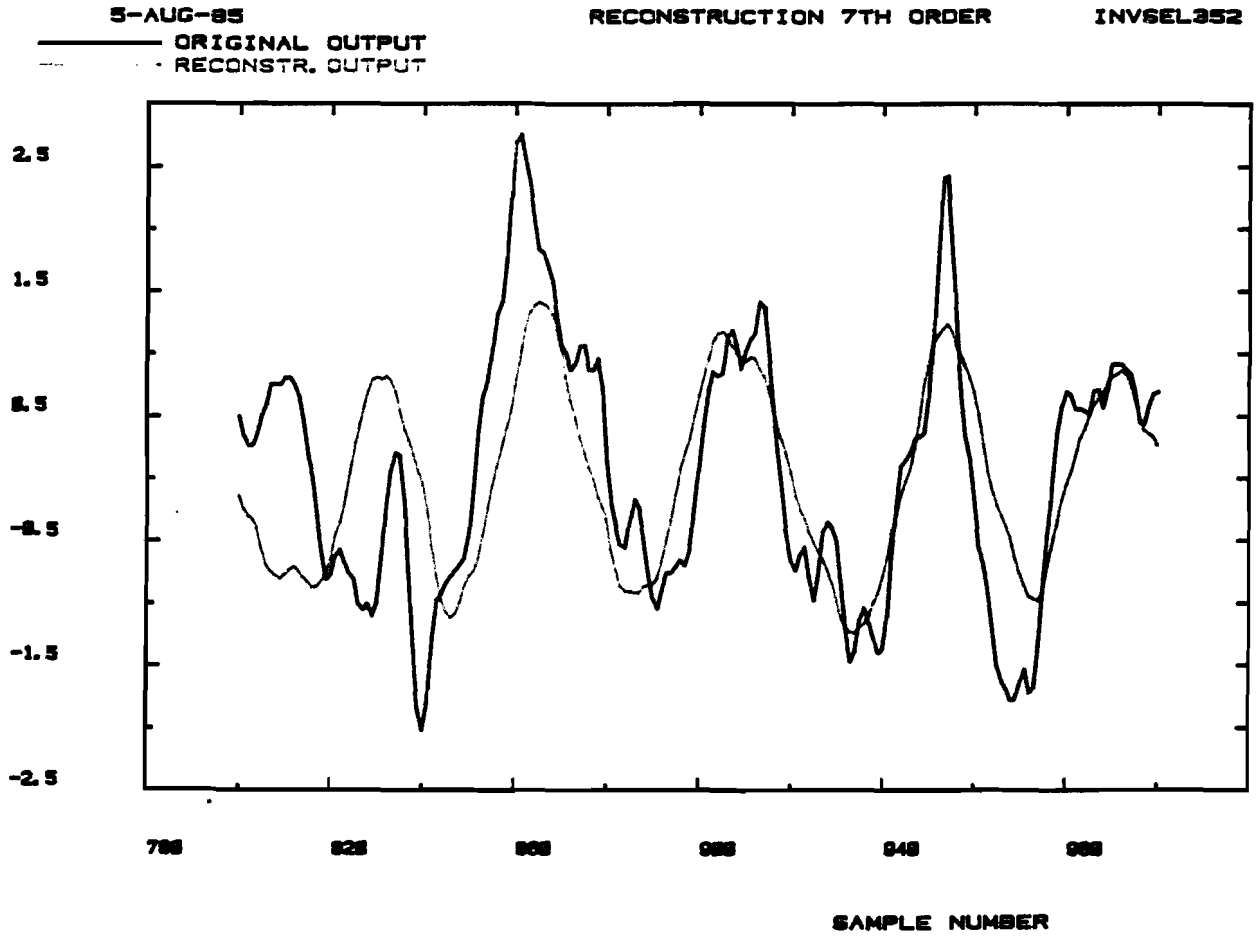


Fig. 6.16 Reconstruction of the second output signal (the diameter of the glass tube) with a seventh order MISO model.

REFERENCES

Akaike, H. (1974)
"A new look at the statistical model identification."
IEEE Transactions on Automatic Control, AC-19, pp. 716-723

Akaike, H. (1977)
"On entropy maximisation principle."
Applications of Statistics. pp. 27-41 (Krishnaiah, ed.), North-Holland, Amsterdam.

Åström, K.J. and P. Eykhoff (1971)
"System identification - a survey."
Automatica, vol. 7, pp. 123-162.

Bard, Y. (1974)
"Nonlinear parameter estimation."
Academic Press, New York, USA.

Barrett-Lennard, I. and J.R. Blair (1981)
"Estimation of coefficients for multiple input system models without employing common denominator structure."
Int. Journ. of Control, vol. 33, no. 1, pp. 123-136.

Bollen, R. (1980)
"Het interactieve programmapakket SATER met betrekking tot multivariabele systemen en de identificatie daarvan volgens Guidorzi."
M.Sc.Thesis, Group Systems and Control, Ph.E Dept., Eindhoven University of Technology, Netherlands (in Dutch).

Bollen, R. and A.J.W. van den Boom (1983)
"The interactive package SATER :
Part I : Description of the SATER system.
Part II : SATER programmer's manual.
Part III: SATER user's manual."
Internal Reports, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

Boom, A.J.W. van den (1982)
"System identification: on the variety and coherence in parameter and order estimation methods."
Dr. Dissertation, Eindhoven University of Technology, Netherlands.

Carrière, P.M. (1984)
"Identification of multivariable systems and representation in canonical and pseudo-canonical form."
M.Sc.Thesis, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.

El-Sherief, H. (1982)
"Recent advances in modeling and identification of stochastic multivariable systems and their applications."
IFAC Symp. on Identification and System Parameter Estimation 1982, Washington D.C., USA.

Eykhoff, P. (1974)
"System Identification, Parameter and State Estimation."
John Wiley, London.

Hannan, E.J. and B. Quinn (1979)
"The determination of the order of an autoregression."
Journ. of Roy. Statistic. Soc. B, 41, pp. 190-195.

Hannan, E.J. (1980)
"The estimation of the order of an ARMA process."
Annals of Statistics, vol. 8, no. 5, pp. 1071-1081.

Kashyap, R.L. (1980)
"Inconsistency of the AIC rule for estimating the order of autoregressive models."
IEEE Transactions on Automatic Control, AC-25, no. 5, pp. 996-998.

Ogata, K. (1970)
"Modern Control Engineering."
Prentice-Hall, New York, USA.

Rissanen, J. (1978)
"Modeling by shortest data description."
Automatica, vol. 14, pp. 465-471.

Shibata, R. (1976)
"Selection of the order of an autoregressive model by Akaike's Information Criterion."
Biometrika, vol. 63, no. 1, pp. 117-126.

Söderström, T. (1975)
"Test of pole-zero cancellation in estimated models."
Automatica, vol. 11, pp. 537-540.

Söderström, T. (1977)
"On model structure testing in system identification."
Int. Journ. of Control, vol. 26, pp. 1-18.

Steiglitz, K. and L.E. McBride (1965)
"A technique for the identification of linear systems."
IEEE Transactions on Automatic Control, AC-10, pp. 461-464.

Stoica, P. and T. Söderström (1981)
"The Steiglitz-McBride identification algorithm revisited - convergence analysis and accuracy aspects."
IEEE Transaction on Automatic Control, AC-26, pp. 712-717.

Stoica, P. (1984)
"Model structure selection by cross-validation."
preliminary paper (to be published).

Stone, M. (1974)
"Cross-validatory choice and assessment of statistical predictions."
Journ. of Roy. Statistic. Soc. B, 36, pp. 111-147.

Stone, M. (1977)
"An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion."
Journ. of Roy. Statistic. Soc. B, 39, pp. 44-47.

Stone, M. (1978)
"Cross-validation; a review."
Math. Operationsforsch. Statist., Ser. Statistics, vol. 9, no. 1, pp. 127-129.

Zhu, Y. (1984)
"Finite word length effects related to recursive identification."
Report 4th-year project, Group Measurement and Control, EE Dept., Eindhoven University of Technology, Netherlands.