Some identification schemes for non-linear noisy processes

SOME IDENTIFICATION SCHEMES FOR NON-LINEAR NOISY PROCESSES

by

J.Z. Westenberg

SOME IDENTIFICATION SCHEMES FOR NON-LINEAR NOISY PROCESSES

by

J.Z. Westenberg

TH-Report 69-E-09

Submitted in partial fulfillment of the requirements
for the degree of Ir. (M.Sc.) in the Eindhoven
University of Technology.

SOME IDENTIFICATION SCHEMES FOR NON-LINEAR NOISY PROCESSES

J.Z. Westenberg
Eindhoven University of Technology
Department of Electrical Engineering
Eindhoven, Netherlands

Abstract

A general mathematical description for non-linear systems in the time-domain is elaborated in this paper. Four algorithms for estimating the parameters of time-discrete processes corrupted by additive noise are studied and instrumented on a digital computer.
All identification schemes may be employed in on-line applications.
The properties of each algorithm are given; numerical results and amount of complexity are compared.

Contents

1. Introduction

In optimization problems of physical processes a certain amount of knowledge of system properties is required for effective control. The analysis of a process with unknown properties is therefore essential.

In most realistic situations the experimental determination is a matter of estimation; the process may be disturbed by noisy influences, and measurements are affected by measurement errors.

Process parameter estimation has become increasingly important in control theory and many investigators have contributed to the solution of the identification problem.

A survey paper of parameter identification is given by Eykhoff (ref.1) and recently by Balakrishnan and Peterka (ref.2).

In this paper we shall consider single-input and single-output processes represented in fig. 1.

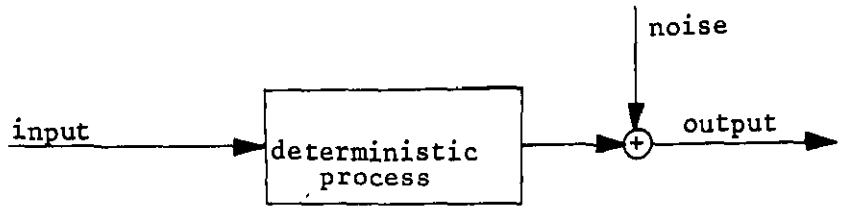

fig. 1

The process can be decomposed in a deterministic part acting on the input signal and an additive noisy disturbance corrupting the ideal output.

We consider discrete-time processes, since, apart from mathematical simplicities the use of digital techniques in control problems has many advantages. Observation of input-output data may give us an estimate of the process parameters in a situation where no a priori knowledge about the process is available. In realistic situations, as we shall see, this black box approach has to be limited in order to obtain practical results.

The process may be linear as well as non-linear; the only assumption

will be time-invariance of the parameters. The input signal used for identification can be chosen arbitrarily; any type of input signal being stationary and noise free measurable is permitted.

The common way for solving identification problems is the approximation by means of a physical or mathematical model illustrated in fig. 2.

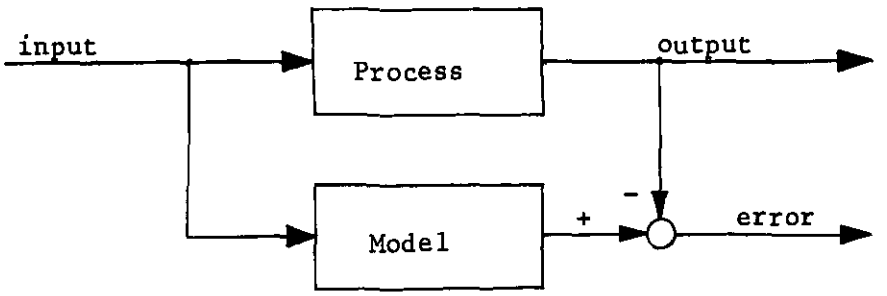

fig. 2

Some functional of the difference between model- and process-output has to be minimized. The choice of this functional depends on the available knowledge of the statistics of process parameters and disturbing noise signal, which knowledge is not presupposed within the scope of this paper. Accordingly, we deal with a least squares error criterion.

Subsequent to a general mathematical description of non-linear systems four identification methods are considered. Updating schemes are devised as new input-output observations become available. So, in practice, all processed data need not be memorized.

The presented identification schemes can be applied in a wide variety of control situations. We may think of the determination of non-linear dynamics of a human operator, characteristics of airplanes during flight, optimal control of chemical plants, etc.

2. System Representation

An essential choice to be made is the form of the mathematical relation between input and output of the process.
Since we consider non-linear processes, we are looking after a non-linear functional F mapping the input x(t) into the output y(t) explicitly.

$$y(t) = F\{x(t)\} \tag{1}$$

Frechet (ref.3) has shown that any continuous functional defined on a set of continuous functions over a finite interval can be represented by a functional power series. An attractive form of this functional for practical applications is the Volterra expansion

$$y(t) = \sum_{n=1}^{\infty} \left\{ \int_0^{\infty} \cdots \int_0^{\infty} b^{(n)}(\tau_1,\ldots,\tau_n).x(t-\tau_1)\ldots x(t-\tau_n)\, d\tau_1 \ldots d\tau_n \right\} \tag{2}$$

$b^{(n)}(\tau_1,\ldots,\tau_n)$ is called n-th-order kernel of the Volterra series. For linear systems only the first-order kernel $b^{(1)}(\tau)$ is considered.

The time-discrete version of this functional with truncated power series and finite memorylength of each kernel represents a non-linear system by

$$y(k) = \sum_{n=1}^{N} \left\{ \sum_{m_1=0}^{M_n-1} \cdots \sum_{m_n=0}^{M_n-1} b^{(n)}(m_1,\ldots,m_n).x(k-m_1)\ldots x(k-m_n) \right\} \tag{3}$$

N is the number of kernels, M_n the memorylength of the n-th-order kernel.

This type of description is very general. An important advantage is its linearity in the parameters, which makes it well-suited for the identification schemes studied here.

The number of parameters in expression (3) can be reduced if similar terms

are considered as one term, e.g.

$$b^{(2)}(i,j).x(k-i).x(k-j)+b^{(2)}(j,i).x(k-j).x(k-i) =$$

$$= b^{(2)^*}(i,j)\ x(k-i).x(k-j)$$

Applying this reduction of dimensionality, we obtain

$$y(k) = \sum_{m_1=0}^{M_1-1} b^{(1)}(m_1).x(k-m_1)+ \sum_{m_1=0}^{M_2-1}\sum_{m_2=0}^{m_1} b^{(2)}(m_1,m_2)x(k-m_1).x(k-m_2)+$$

$$+ \sum_{m_1=0}^{M_3-1}\sum_{m_2=0}^{m_1}\sum_{m_3=0}^{m_2} b^{(3)}(m_1,m_2,m_3).x(k-m_1).x(k-m_2).x(k-m_3)+ \ldots \qquad (4)$$

Expression (4) can be represented as an inner product of two vectors

$$y(k) = \underline{u}_k^T\, \underline{b} \qquad (5)$$

where

$$\underline{u}_k = \begin{bmatrix} x(k) \\ x(k-1) \\ \vdots \\ x(k-M_1+1) \\ x^2(k) \\ \vdots \\ x^2(k-M_2+1) \\ x^3(k) \\ \vdots \\ x^N(k-M_N+1) \end{bmatrix} \quad \text{and} \quad \underline{b} = \begin{bmatrix} b^{(1)}(0) \\ b^{(1)}(1) \\ \vdots \\ b^{(1)}(M_1-1) \\ b^{(2)}(0,0) \\ \vdots \\ b^{(2)}(M_2-1,M_2-1) \\ b^{(3)}(0,0,0) \\ \vdots \\ b^{(N)}(M_N-1,\ldots,M_N-1) \end{bmatrix} \qquad (6,7)$$

$\underline{u}_k$ is the input vector and $\underline{b}$ the process parameter vector, both with p elements, where p is the total number of parameters.
According to this notation the process output vector

$$\underline{y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(k) \end{bmatrix} \qquad (8)$$

can be written as

$$\underline{y} = U\,\underline{b} \tag{9}$$

with

$$U = \begin{bmatrix} \underline{u}_1^T \\ \underline{u}_2^T \\ \vdots \\ \underline{u}_k^T \end{bmatrix} = \begin{bmatrix} u_1(1) & u_1(2) & \dots & u_1(p) \\ u_2(1) & u_2(2) & \dots & u_2(p) \\ \vdots & \vdots & & \vdots \\ u_k(1) & u_k(2) & \dots & u_k(p) \end{bmatrix} \tag{10}$$

If the process output is corrupted by additive noise n we find for the measurable output vector

$$\underline{z} = \underline{y} + \underline{n} \tag{11}$$

In this way any non-linear continuous process can be represented by a linear relation between the output and the parameters.

3. Identification Schemes

In accordance with expression (7) a non-linear model of the unknown process is characterized by a parameter vector $\underline{\beta}$ with also p elements. Similarly to eq. (9) we obtain for the model-output

$$\underline{w} = U\,\underline{\beta} \tag{12}$$

The error between model-output and process-output is represented by a vector with k elements

$$\underline{e} = \underline{w} - \underline{z} = U\,\underline{\beta} - \underline{z} \tag{13}$$

The sum of the squares of the elements of $\underline{e}$ is

$$E = \sum_{i=1}^{k} e^2(i) = \underline{e}^T\underline{e} \tag{14}$$

The least squares error criterion yields minimization of E by choosing an optimal estimate $\underline{\beta}$.

3.1. Explicit Estimation

Assuming a finite sequence of K input-output observations, we can directly obtain a least squares estimate in the following way.
We try to minimize

$$E = \underline{e}^T \underline{e} = \left[U \underline{\beta} - \underline{z} \right]^T \left[U \underline{\beta} - \underline{z} \right] \tag{15}$$

Differentiating (14) with respect to $\underline{\beta}$ and equating the result to zero, we obtain the optimal solution

$$\frac{\partial E}{\partial \underline{\beta}} = 2\, U^T U\, \underline{\beta} - 2\, U^T \underline{z} = \underline{0}$$

So

$$\underline{\beta} = \left[U^T U \right]^{-1} U^T \underline{z} \tag{16}$$

Actually, it is not necessary to store all K elements of the input-output sequences. An updating scheme can be developed by indexing matrix U and vector $\underline{z}$.
After k-1 observations we define

$$R_{k-1} = U_{k-1}^T\, U_{k-1} \tag{17}$$

$$\underline{q}_{k-1} = U_{k-1}^T\, \underline{z}_{k-1} \tag{18}$$

A new k-th input-output observation leads to

$$U_k = \begin{bmatrix} U_{k-1} \\ \underline{u}_k^T \end{bmatrix} \qquad \text{and} \qquad \underline{z}_k = \begin{bmatrix} \underline{z}_{k-1} \\ z(k) \end{bmatrix}$$

We then find

$$R_k = U_k^T U_k = U_{k-1}^T U_{k-1} + \underline{u}_k\, \underline{u}_k^T = R_{k-1} + \underline{u}_k\, \underline{u}_k^T \tag{19}$$

$$\underline{q}_k = U_k^T \underline{z}_k = U_{k-1}^T\, \underline{z}_{k-1} + \underline{u}_k\, z(k) = \underline{q}_{k-1} + \underline{u}_k\, z(k) \tag{20}$$

Starting with $R_0 = 0\ I$ and $\underline{q}_0 = \underline{0}$, R_k and $\underline{q}_k$ are updated for k = 1,2,..., K.
The estimate of the process parameters then is

$$\underline{\beta}_K = R_K^{-1}\, \underline{q}_K \tag{21}$$

The pxp matrix R_K is symmetric around its main diagonal, and positive definite; for a numerical solution Cholesky's method for matrix inversion may be used (ref. 5).

The quality of the estimate is determined by its statistical expectation and covariance. If the additive noise is statistically independent of the input signal and has zero mean, it follows for the expectation

$$\mathcal{E}\{\underline{\beta}\} = \mathcal{E}\left\{\left[U^T U\right]^{-1} U^T \underline{z}\right\} = \underline{b} \tag{22}$$

The estimate is unbiased. A measure for the accuracy of $\underline{\beta}$ is the covariance matrix

$$\text{cov}\ \underline{\beta} = \mathcal{E}\left\{\left[\underline{\beta}-\underline{b}\right]\left[\underline{\beta}-\underline{b}\right]^T\right\} = \left[U^T U\right]^{-1} U^T\, \mathcal{E}\left\{\underline{n}\underline{n}^T\right\}\, U \left[U^T U\right]^{-1}$$

If the additive noise $\underline{n}$ is white then we find

$$\mathcal{E}\left\{\underline{n}\,\underline{n}^T\right\} = \sigma_n^2\, I \tag{23}$$

Consequently

$$\text{cov}\ \underline{\beta} = \sigma_n^2 \left[U^T U\right]^{-1} \tag{24}$$

The elements of $\left[U^T U\right]^{-1}$ and hence cov $\underline{\beta}$ will converge to zero for growing data.

3.2. Iterative Estimation

In many situations an iterative identification scheme or model-adjusting method is desirable. With respect to the least squares error criterion an iterative procedure can be derived from our explicit estimating scheme (ref. 6). Indexing by k, i.e. the observation number, we have explicitly

$$\underline{\beta}_k = \left[U_k^T U_k\right]^{-1} U_k^T\, \underline{z}_k \tag{25}$$

Define

$$P_k^{-1} = U_k^T U_k \tag{26}$$

Similarly to eq. (19) we obtain

$$P_k^{-1} = P_{k-1}^{-1} + \underline{u}_k \underline{u}_k^T \tag{27}$$

As P_k^{-1} is a symmetrical matrix, inversion leads to

$$P_k = P_{k-1} - P_{k-1} \underline{u}_k \left[\underline{u}_k^T P_{k-1} \underline{u}_k + 1 \right]^{-1} \underline{u}_k^T P_{k-1} \tag{28}$$

The quantity $\left[\underline{u}_k^T P_{k-1} \underline{u}_k + 1 \right]$ is a scalar; the inversion of a pxp matrix is reduced to simply computing a numerical reciprocal and updating after each sample. The estimate then follows from

$$\underline{\beta}_k = P_k U_k^T \underline{z}_k = P_k \left[U_{k-1}^T \underline{z}_{k-1} + \underline{u}_k z(k) \right] \tag{29}$$

Substituting of (28) into (29) gives

$$\underline{\beta}_k = \underline{\beta}_{k-1} - P_{k-1} \underline{u}_k \left[\underline{u}_k^T P_{k-1} \underline{u}_k + 1 \right]^{-1} \left[\underline{u}_k^T \underline{\beta}_{k-1} - z(k) \right] \tag{30}$$

or

$$\underline{\beta}_k = \underline{\beta}_{k-1} - P_k \underline{u}_k \left[\underline{u}_k^T \underline{\beta}_{k-1} - z(k) \right] \tag{31}$$

Now, we have found recursive relationships for P_k and $\underline{\beta}_k$. After each input-output observation matrix P_k and parameter vector $\underline{\beta}_k$ are updated, direct matrix inversion is discarded. P_k is a symmetrical matrix, hence only one triangular part needs to be updated.
In expression (30) vector $P_{k-1} \underline{u}_k \left[\underline{u}_k^T P_{k-1} \underline{u}_k + 1 \right]^{-1}$ represents the weighting factors for every element of the parameter vector with respect to the difference between model- and process-output $\left[\underline{u}_k^T \underline{\beta}_{k-1} - z(k) \right]$.

Since this iterative scheme has been derived from the explicit method, the statistical properties are equal to those obtained by explicit estimating.

The recursive relations (28) and (30) can only be used if a starting matrix P_k and estimate $\underline{\beta}_k$ are initiated. In order to obtain an optimal estimate in least squares sense the initial values must be chosen accurately. For this purpose two ways can be followed.

Firstly, initial estimates are derived from the explicit method. In performing matrix inversion, a starting matrix follows from a minimal set of p observations

$$P_p = \left[U_p^T U_p \right]^{-1} \tag{32}$$

The initial estimate of the parameter vector follows from

$$\underline{\beta}_p = P_p U_p^T \underline{z}_p \tag{33}$$

Next, the iterative scheme is used for subsequent data.

Secondly, a least squares optimal solution without any matrix inversion can be achieved with the iterative scheme, assuming a special form for P_o, i.e.

$$P_o = a\, I \tag{34}$$

where a is a very large scalar; $\underline{\beta}_o$ can be chosen arbitrarily. After one observation we obtain

$$P_1 = \left[P_o^{-1} + \underline{u}_1 \underline{u}_1^T \right]^{-1}$$

and with respect to eq. (31)

$$\underline{\beta}_1 = \underline{\beta}_o - P_1 \underline{u}_1 \left[\underline{u}_1^T \underline{\beta}_o - z(1) \right] = P_1 \left[P_o^{-1} \underline{\beta}_o + \underline{u}_1 z(1) \right]$$

Continuing in this manner up to the p-th observation, we can derive

$$P_p = \left[P_o^{-1} + U_p^T U_p \right]^{-1} \tag{35}$$

$$\underline{\beta}_p = P_p \left[P_o^{-1} \underline{\beta}_o + U_p^T \underline{z}_p \right] \tag{36}$$

As $P_o^{-1} = \frac{1}{a} I$ and taking the limit of (35) and (36) for $a \to \infty$, it follows

$$P_p \to \left[U_p^T U_p \right]^{-1}$$

$$\underline{\beta}_p \to P_p U_p^T \underline{z}_p$$

Hence, in least squares sense the optimal solution of the estimating problem is obtained after p and subsequent observations. The latter approach permits the use of the iterative algorithm for all input-output data, k = 1,2,...,K.

3.3. Stochastic Approximation

An often applied procedure for minimizing the mean-squared error between the output of model and process is the generally known gradient method. This is an iterative estimating scheme and its general form is

$$\underline{\beta}_k = \underline{\beta}_{k-1} - a_k \nabla E(\underline{\beta}_{k-1}) \tag{37}$$

a_k is a relaxation factor and has to satisfy some conditions for the sake of convergence of relation (37).

In the discrete-time case a stochastic approximation method is given by Holmes (ref. 7).

The parameter vector is adjusted after each new observation; input and additive noise sequences are assumed to be stationary and independent.

In the k-th instant the error can be written as

$$e(k) = \underline{u}_k^T \underline{\beta}_{k-1} - z(k) \tag{38}$$

For the instantaneous squared error we denote

$$E_k = e^2(k) \tag{39}$$

As e(k) is a stochastic variable we have to minimize the expected value of the squared error

$$L_k = \mathcal{E}\left\{E_k \,|\, \underline{\beta}_{k-1}\right\}$$

Actually, only the variable E_k is available and hence the i-th component of the gradient in (37) is approximated by

$$\frac{\partial L_k}{\partial \beta_{k-1}(i)} \simeq \frac{\partial E_k}{\partial \beta_{k-1}(i)} \simeq \frac{E_k(\underline{\beta}_{k-1} + c_k \underline{e}_i) - E_k(\underline{\beta}_{k-1} - c_k \underline{e}_i)}{2\, c_k} \tag{40}$$

$\underline{e}_i$ is an orthogonal unit vector in the p-dimensional space, c_k is a small scalar.

Substituting expressions (38) and (39) into (40) and involving (37) we obtain for the i-th component of the parameter vector

$$\beta_k(i) = \beta_{k-1}(i) - 2\, a_k\, u_k(i)\, e(k) \tag{41}$$

where $i = 1,2,\ldots,p$.
Note that (41) does not depend on c_k.
The adjusting algorithm in vector notation follows from (38) and (41)

$$\underline{\beta}_k = \underline{\beta}_{k-1} - 2\, a_k\, \underline{u}_k \left[\underline{u}_k^T\, \underline{\beta}_{k-1} - z(k) \right] \tag{42}$$

It can be shown (ref. 7) that this stochastic approximation estimate converges in probability to the actual parameters $\underline{b}$ of the process if a_k is assumed to satisfy

$$\sum_{k=1}^{\infty} a_k \to \infty; \quad \sum_{k=1}^{\infty} a_k^2 < \infty \quad \text{and} \quad a_k \geq 0 \tag{43}$$

Here, the estimate is squared-error consistent.

A possible choice for a_k is

$$a_k = \frac{A}{k^{\alpha}} \tag{44}$$

A being a positive constant factor and

$$\frac{1}{2} < \alpha \leq 1 \tag{45}$$

An initial vector $\underline{\beta}_o$ can be chosen arbitrarily, e.g. $\underline{\beta}_o = \underline{0}$. Comparing the iterative schemes (31) and (42), we note the similarity of both algorithms. In expression (42) a matrix P_k linked with the covariance matrix of the k-th estimate (eq. (24)) is replaced by a diagonal matrix $2\, a_k\, I$. Due to this simplification the stochastic approximation method does not lead to the optimal estimate in least squares sense, but yields an approximate solution.

3.4. Orthogonal Transformation

The identification of noisy processes by minimization of the mean squared error deals with the solution of a set of linear equations. A well-suited method for our problem, employing an orthogonal transformation, has recently been given by Peterka and Šmuk (ref. 8).

The set of k linear equations representing the error

$$\underline{U\beta} - \underline{z} = \underline{e}$$

where a function $E = \underline{e}^T\underline{e}$ is to be minimized, may be premultiplied by a kxk square matrix T.

We denote

$$\begin{aligned} TU &= \tilde{U} \quad \text{(dimensions kxp)} \\ T\underline{z} &= \underline{\tilde{z}} \quad \text{(k elements)} \\ T\underline{e} &= \underline{\tilde{e}} \quad \text{(k elements)} \end{aligned} \tag{46}$$

Now, the transformed set of equations is

$$\tilde{U}\underline{\beta} - \underline{\tilde{z}} = \underline{\tilde{e}} \tag{47}$$

The sum of the squares of the elements of $\underline{\tilde{e}}$ is

$$\tilde{E} = \underline{\tilde{e}}^T\underline{e} = \underline{e}^T T^T T\,\underline{e} \tag{48}$$

Assuming an orthogonal matrix T the sum of the squares will not change, so $\tilde{E} = E$ if

$$T^T T = I \tag{49}$$

The orthogonal matrix T will be chosen to have a special form, i.e.

$$T_{ij} = \begin{bmatrix} 1 & & & & & & & & \\ & 1 & & & & & & & \\ & & \ddots & & & & & & \\ & & & 1 & & & & & \\ & & & & r & & s & & i \\ & & & & & 1 & & & \\ & & & & & & \ddots & & \\ & & & & & & & 1 & \\ & & & & s & & r & & j \\ & & & & & & & 1 & \\ & & & & & & & & \ddots \\ & & & & & & & & & 1 \end{bmatrix} \tag{50}$$

The non-trivial elements of T are zero. Matrix (50) is called elementary matrix of rotation and is orthogonal provided that

$$r^2 + s^2 = 1 \tag{51}$$

In the set of equations (47) only the i-th and j-th equation have been changed in multiplicating by matrix T_{ij}; with reference to (46) we have for $n \neq i$ and $n \neq j$

$$\begin{aligned} \tilde{u}(n,1) &= u(n,1), \qquad (1 = 1,2,\ldots,p) \\ \tilde{z}(n) &= z(n) \\ \tilde{e}(n) &= e(n) \end{aligned} \tag{52}$$

for $n = i$

$$\begin{aligned} \tilde{u}(i,1) &= r\,u(i,1) + s\,u(j,1), \qquad (1 = 1,2,\ldots,p) \\ \tilde{z}(i) &= r\,z(i) + s\,z(j) \\ \tilde{e}(i) &= r\,e(i) + s\,e(j) \end{aligned} \tag{53}$$

for $n = j$

$$\begin{aligned} \tilde{u}(j,1) &= -s\,u(i,1) + r\,u(j,1), \qquad (1 = 1,2,\ldots,p) \\ \tilde{z}(j) &= -s\,z(i) + r\,z(j) \\ \tilde{e}(j) &= -s\,e(i) + r\,e(j) \end{aligned} \tag{54}$$

Refering to relation (51) only one of the coefficients r and s can be chosen arbitrarily. We will make a selection so that an element of the j-th row of $\tilde{u}$ equals zero, i.e.

$$\tilde{u}(j,m) = -s\,u(i,m) + r\,u(j,m) = 0 \tag{55}$$

Then for r and s we obtain

$$r = \frac{u(i,m)}{\sqrt{u^2(i,m) + u^2(j,m)}}, \quad s = \frac{u(j,m)}{\sqrt{u^2(i,m) + u^2(j,m)}} \tag{56}$$

Employing a successive application of this transformation, any element of matrix u can be eliminated for any parameter vector ; the set of equations (47) can be arranged as follows

$$\begin{bmatrix} u^*(1,1) & u^*(1,2) & \dots & u^*(1,p) \\ & u^*(2,2) & \dots & u^*(2,p) \\ & & & \vdots \\ & & & u^*(p,p) \\ & 0 & & \end{bmatrix} \begin{bmatrix} \beta(1) \\ \beta(2) \\ \vdots \\ \beta(p) \end{bmatrix} - \begin{bmatrix} z^*(1) \\ z^*(2) \\ \vdots \\ z^*(p) \\ z^*(p+1) \\ \vdots \\ z^*(k) \end{bmatrix} = \begin{bmatrix} e^*(1) \\ e^*(2) \\ \vdots \\ e^*(p) \\ e^*(p+1) \\ \vdots \\ e^*(k) \end{bmatrix} \qquad (57)$$

After this transformation it still holds that

$$E = \sum_{i=1}^{k} e^2(i) = \sum_{i=1}^{k} e^{*2}(i) \qquad (58)$$

This function has to be minimized by a suitable selection of vector $\underline{\beta}$. As only the first p elements at the righthand side of (57) can be governed by $\underline{\beta}$, we provide these elements to be zero. Estimate $\underline{\beta}$ can then be obtained from a reduced set of p linear equations, i.e.

$$U^*\underline{\beta} - \underline{z}^* = 0 \qquad (59)$$

With

$$U^* = \begin{bmatrix} u^*(1,1) & u^*(1,2) & \dots & u^*(1,p) \\ & u^*(2,2) & \dots & u^*(2,p) \\ & & & \vdots \\ & & & u^*(p,p) \\ & 0 & & \end{bmatrix} \text{ and } \underline{z}^* = \begin{bmatrix} z^*(1) \\ z^*(2) \\ \vdots \\ z^*(p) \end{bmatrix} \qquad (60)$$

Now, the solution of (59) can simply be found since U^* is an upper triangular matrix

$$\beta(p) = \frac{z^*(p)}{u^*(p,p)}$$

$$\beta(p-n) = \frac{1}{u^*(p-n,p-n)} \left[z^*(p-n) - \sum_{i=0}^{n-1} \beta(p-i)u^*(p-n,p-i) \right] \qquad (61)$$

$$(n = 1,2, \dots,p-1)$$

With respect to the least squares criterion we have an optimal estimate of the process parameters after k input-output observations explicitly without the necessity of matrix inversion.

For matrix U^* and vector $\underline{z}^*$ an updating algorithm can be developed as the number of input-output data progresses. Starting from the transformed set of equations (57) the input-output data of a new sample add a new k-th equation and hence a new row to the matrix U. All elements of this new row can be eliminated sequentially by successive application of the transformation algorithm. The indices i and j of the elementary matrix of rotation T_{ij} must be selected. Since elements of the k-th row are to be eliminated, we choose $j = k$.

The indexnumber i runs from m to p, where m is the elementnumber of the row to be eleminated, successivily $m = 1,2,\ldots,p$.

According to this procedure all new information is added to the upper triangular matrix U^* and vector $\underline{z}^*$. The sum of the squares (58) will not change if matrix U and vector $\underline{z}$ are extended by a pxp zero matrix and a p-elements zero vector, i.e.

$$\begin{bmatrix} 0 \\ U \end{bmatrix} \underline{\beta} - \begin{bmatrix} \underline{0} \\ \underline{U} \end{bmatrix} = \begin{bmatrix} \underline{0} \\ \underline{e} \end{bmatrix} \tag{62}$$

Considering this zero matrix and zero vector as the initial state of U^* and $\underline{z}^*$, we can employ the updating scheme from the first input-output data. Only U^* and $\underline{z}^*$ need to be memorized; the optimal least squares estimate can be computed at any instant using eq.(61).

4. Experimental Results

The estimating schemes described in the preceding section are all programmed in Algol 60 and tested on a digital computer.*) A second-order non-linear process is simulated on the computer as well. The memorylength of the first-order kernel is 10 and of the second-order kernel 3, so that 16 parameters are to be determined.

Random input and noise sequences are generated with a subroutine. The input signal is either white with rectangular probability density or coloured by a discrete low-pass filter and has zero mean.

*) The procedures used for updating and/or estimating are available at the Eindhoven University of Technology, Department of Electrical Engineering, Group Measurement and Control, Eindhoven, Netherlands.

The additive noisy disturbance is white with zero mean and has rectangular probability density.
The parameters are computed and printed out after every interval of 50 samples in order to determine the rate of convergence. The total observation length is 1000 samples. A situation is considered where about 10% of the output power is caused by the noisy disturbance. The relative noise power and the relative mean squared error for the computed parameters are determined over an interval of 1000 samples.
Numerical data and results of the estimates for white input signal are given in table 1. Similar results are obtained by using a random input signal coloured by filter $G(m) = \exp(-m)$ with $m = 1,\ldots,4$.

The explicit estimating method with inversion of a 16x16 matrix gives the optimal solution in least squares sense.
The standard deviation

$$\sigma(i) = \sigma_n \sqrt{R^{-1}(i,i)} \qquad (63)$$

where $R = U^T U$, can easily be computed if σ_n is a known factor.
The same results are achieved when using the iterative scheme, where P_k and $\underline{\beta}_k$ are updated. Here, we may start with either a matrix P_p obtained by explicit estimation after a minimal data set or initial matrix $P_o = a\,I$, where a is chosen to be 10^6.
The stochastic approximation procedure yields a somewhat different solution due to the simplifications involved in the derivation of this algorithm. From experiments it appears that the best results are obtained when choosing relaxation factor

$$a_k = \frac{A}{k^{\alpha}}$$

with $A = 0.4$ and $\alpha = 0.6$

The procedure using orthogonal transformations also yields the optimal solution explicitly, but without any matrix inversion.
About 60% of the estimated parameters lie within the standard deviation to be expected according to formula (63). This applies to all four estimating schemes tested with either white random input or coloured input.
After each input-output observation the number of most time consuming operations in the execution of the estimating algorithms, being multiplications and subscripts, are given in table 2 for a number of p parameters.

5. Concluding Remarks

All estimating schemes studied in this paper provide for data reduction simultaneously with processing. The quantity of information to be memorized is independent of the observationlength, and only a rather small demand on storage capacity of the computer is required. The algorithms may therefore be useful in on-line applications.

The fastest sampling frequency is allowed if the stochastic approximation algorithm is used, since this requires the fewest computations after each sample.

Volterra power series do represent non-linear systems in a very general way and due to this fact a generally large number of parameters need to be estimated. In realistic problems, where some a priori information of the process characteristics is available, a model with lower dimensionality is often applicable, e.g. frequency-domain model or difference-equation model.

Other criterion functions may be used for estimating if more knowledge of the statistical properties of process and noisy disturbances is available.

6. References

1. Eykhoff, P.,
   Process Parameter and State Estimation,
   IFAC Symposium on Identification in Automatic Control Systems,
   Prague, june 1967;
   Automatica, Vol. 4, 1968, p.p. 205-233.

2. Balakrishnan, A.V., V. Peterka,
   Identification in Automatic Control Systems,
   Fourth Congress IFAC, Warszawa, june 1969, Survey paper no. 9.

3. Frechet, M.,
   Sur les Functionelles Continues,
   Annales Scientifiques de l'Ecole Normal Supérieure,
   Vol. 27, Series 3, 1910.

4. Volterra, V.,
   Theory of Functionals and of Integro-Differential Equations,
   Dover Publications, Inc., New York, 1959.

5. Westlake, J.R.,
   A Handbook of Numerical Matrix Inversion and Solution of Linear Equations,
   John Wiley & Sons, Inc., New York, 1968.

6. Lee, R.C.K.,
   Optimal Estimation, Identification, and Control,
   M.I.T. Press, Cambridge, Massachusetts, 1964.

7. Holmes, J.K.,
   System Identification from Noise-Corrupted Measurements,
   Journal of Optimization Theory and Applications,
   Vol 2, No. 2, 1968, p.p. 103-116.

8. Peterka, V., K. Šmuk,
   On-Line Estimation of Dynamic Model Parameters from Input-Output Data,
   Identification, Methods for Parameter Estimation, Technical Session No. 26, Fourth Congress IFAC, Warszawa, june 1969, p.p. 3-26.

Table 1

Estimates of the parameters of a second-order non-linear process

Process: $M_1 = 10$, $M_2 = 3$
Input signal: rectangular-distributed between -1.0 and +1.0
Additive noise: rectangular-distributed between -0.7 and +0.7
Length of observation: K = 1000

parameter	actual value	explicit estimation	iterative estimation	stochastic approximation	orthogonal transformation	standard deviation
b(0)	1.000	1.044	1.044	1.017	1.044	0.022
b(1)	0.900	0.915	0.915	0.886	0.915	0.022
b(2)	0.800	0.769	0.769	0.808	0.769	0.022
b(3)	0.700	0.697	0.697	0.698	0.697	0.022
b(4)	0.600	0.591	0.591	0.574	0.591	0.022
b(5)	0.500	0.478	0.478	0.473	0.478	0.022
b(6)	0.400	0.420	0.420	0.420	0.420	0.022
b(7)	0.300	0.325	0.325	0.301	0.325	0.022
b(8)	0.200	0.256	0.256	0.258	0.256	0.022
b(9)	0.100	0.106	0.106	0.127	0.106	0.022
b(0,0)	0.300	0.280	0.280	0.254	0.280	0.039
b(1,0)	0.250	0.235	0.235	0.200	0.235	0.040
b(1,1)	0.200	0.226	0.226	0.213	0.226	0.039
b(2,0)	0.150	0.055	0.055	-0.035	0.055	0.040
b(2,1)	0.100	0.171	0.171	0.179	0.171	0.040
b(2,2)	0.050	0.067	0.067	0.087	0.067	0.039
relative mean squared error:		10.463%	10.463%	10.686%	10.463%	

relative noise power: 10.385%

Table 2

Number of multiplications and subscripts required for updating

Number of parameters: p

procedure:	multiplications:	1-dim.subscripts:	2-dim.subscripts:
explicit estimation	$\frac{1}{2} p^2 + \frac{3}{2} p$	$p^2 + 4 p$	$2 p^2 + 2 p$
iterative estimation	$\frac{3}{2} p^2 + \frac{9}{2} p$	$4 p^2 + 10 p$	$3 p^2 + 2 p$
stochastic approximation	$2 p$	$5 p$	0
orthogonal transformation	$2 p^2 + 17 p$	$\frac{3}{2} p^2 + \frac{15}{2} p$	$\frac{3}{2} p^2 + \frac{7}{2} p$