NINTH CONFERENCE
ON THE MATHEMATICS OF
OPERATIONS RESEARCH & SYSTEM THEORY

MB ST

BENELUX MEETING
ON SYSTEMS & CONTROL 1984

MINI COURSE:
TIME SERIES ANALYSIS AND SYSTEM IDENTIFICATION
NOTES

January 11-13, 1984

CONFERENCE CENTER "DE BLIJE WERELT"
LUNTEREN, THE NETHERLANDS

Centre for Mathematics and Computer Science
Amsterdam, The Netherlands

NINTH CONFERENCE
ON THE MATHEMATICS OF
OPERATIONS RESEARCH & SYSTEM THEORY

BENELUX MEETING
ON SYSTEMS & CONTROL 1984

MINI COURSE:
TIME SERIES ANALYSIS AND SYSTEM IDENTIFICATION
NOTES

January 11-13, 1984

CONFERENCE CENTER "DE BLIJE WERELT"
LUNTEREN, THE NETHERLANDS

Centre for Mathematics and Computer Science
Amsterdam, The Netherlands

MINICOURSE
TIME SERIES ANALYSIS AND SYSTEM IDENTIFICATION

PROGRAM

1. Time series analysis,
   by B.B. van der Genugten (Tilburg)
2. System identification,
   by P. Eykhoff and A.J.W. van den Boom (Eindhoven)

CONTENTS OF NOTES

TIME SERIES ANALYSIS
IDENTIFICATION IN STATISTICAL INFERENCE

B.B. van der Genugten

Subfaculteit Econometrie
Katholieke Hogeschool Tilburg
Postbus 90153
5000 LE Tilburg

SUMMARY

The problem of identification arises in various fields of inferential statistics. In linear regression it has led to the theory of estimable or identifiable functions. In simultaneous equation systems it culminates in the question of the identification of structural equations. Other fields that should be mentioned are the analyses of time-series, factor analyses, statistical decision theory and system theory.

The formulation and treatment of the identification problem varies with the field of application. Sometimes identification is treated as a basic concept that does not need a formal definition at all. Of course this makes theorems about identification rather questionable. Some examples of that will be given.

In this lecture I will

- carefully exploit a general concept of identification for inferential statistics;
- illustrate the definitions and the derived theorems with simple examples;
- apply the general theory developed in this way to dynamic models appearing in the field of time series and system theory.

(

SYSTEM IDENTIFICATION - mini course

P.Eykhoff
A.J.W. van den Boom

Eindhoven University
of Technology
EE Dept.
NL-5600 MB Eindhoven

I

System identification and parameter estimation still offer a rather chaotic picture. Among the reasons for this confusion one may recognize:

- the variety of schemes proposed by many authors without much regards to the coherence of the field,
- the wide spectrum of actual and potential applications in which, of course, the models have to be chosen in accordance with the applications in mind.

One of the ways to attain some order in the outlook on the field is by recognizing, that the majority (or even all ?) methods imply a projection of the process input- and output signals onto function spaces.
To emphasize this notion and its implications we will discuss the choice of 'template functions/-series' that span such function spaces. These templates range from ordinary sine/cosine components (the choice of which does not imply any specific relation to the observed signals and to the model chosen) until the cases where templates are derived from input and output signals (the choice of which templates may be closely related to the model structure chosen).

Based on those template functions/-series choices the variety of estimation methods, as well as their properties, can be discussed in an orderly way through the recognition of three classes:

	template functions	observations
class I	deterministic	deterministic
class II	deterministic	stochastic
class III	stochastic	stochastic

The 'information generating' capabilities of the estimators in these three classes are discussed. Depending on the class to which they belong, there are essential differences in the properties of the estimators, ranging from complete statistical information to only limit-measures for the estimation accuracy attainable. Here the Cramér-Rao relation as well as the information matrix play an essential role.
Related to such fundamental aspects are also the notions of linearity-, pseudo-linearity- and nonlinearity-in-the-parameters of the models used.

II

The second part deals with the relationships between the variety of existing (recursive) estimation schemes for single-input single-output systems. The relationships are established by presenting a general diagram of which various well-known estimators are special cases; such estimators imply the least squares -, generalized least squares -, extended matrix -, instrumental variable -, approximated maximum likelihood -, implicit quasi-linearization -, filtered instrumental variable estimation -, (sub)optimal instrumental variable - and the instrumental variable variant of the extended matrix method.

In the generalized diagram, three important operations can be distinguished:

- noise filtering,
- model extension and
- instrumental variable generation.

It will be shown that by proper combination of these three main operations, all the estimators mentioned above can be constructed.

Åström, K.J. (1981).
Maximum likelihood and prediction error methods. In: P.Eykhoff (editor), Trends and Progress in System Identification. Pergamon, Oxford. p.145-168.

Boom, A.J.W. van den (1982).
System Identification; on the Variety and Coherence in Parameter- and Order Estimation Methods. Dr. Thesis, EE Dept., Eindhoven University of Technology. 239 pp.

Eykhoff, P. (1974,1977).
System Identification; Parameter and State Estimation. Wiley, London. 555 pp.

Ljung, L., A.K. Hajdasinski et al. (1982).
Workshop on the state of the art of system identification. In: G.A. Bekey and G.N. Saridis (editors), Proc. IFAC Symp. on Identification and System Parameter Estimation, Washington D.C. Pergamon Press. p.45-84.

Söderström, T. and P.G. Stoica (1983).
Instrumental Variable Methods for System Identification. Springer, Berlin. 243 pp.

I ON THE COHERENCE AMONG THE MULTITUDE OF SYSTEM IDENTIFICATION METHODS

1 INTRODUCTION - A BAG OF TRICKS

During the development in time of a field of natural science, one expects to recognize the growth of a coherent picture, embracing the previous, partial contributions into a framework that provides order and that clarifies mutual relationships. Since Descartes and Kant, the aspects induction (derivation of the general principles from particular cases), and deduction (derivation of particular cases from general principles), have been recognized as essential entities in the development of science.
In the field of system identification it has been the hope that the maturity of that area would be recognizable in terms of:

- such a coherent picture, generated by induction, and
- availability for specific classes of applications by deduction.

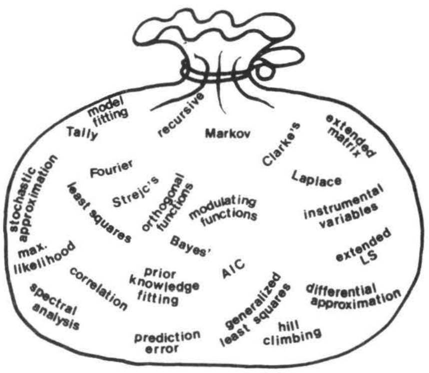


Fig. 1 "A bag of tricks"

In spite of this hope, one has to recognize that the field can still be well represented by fig. 1, where some of the multitude of methods advocated by numerous authors are held together by the peculiar notion .... "a bag of tricks".
Serious attempts and claims have been made by various authors to contribute to such a coherent picture (e.g. Akaike, 1981; Peterka, 1981), but in spite of such efforts still much is left to be desired.

2 THE USE OF A PRIORI KNOWLEDGE

2.1 Statistical knowledge

A first step towards order has to be sought in recognition of a priori knowledge, available in particular system identification/parameter estimation situations. This is indicated in Table 1, where

- N is the covariance matrix of the disturbances;
- $p(n)$ is the shape of the probability density function of the disturbances;
- $q(b)$ indicates the a priori probability function of the parameters to be estimated;
- $C(\hat{\beta},b)$ represents the cost function related to the estimated and the true parameter values $\hat{\beta}$ viz. b.

TABLE 1 A priori knowledge versus estimation schemes applicable

estimation techniques	a priori knowledge p(n)			
	N	shape	q(b)	$C(\hat{\beta},b)$
least squares (LS)	-	-	-	-
Markov (generalized LS)	+	-	-	-
maximum likelihood (ML)	+	+	-	-
Bayes	+	+	+	-
minimum risk/cost	+	+	+	+

Based on the a priori knowledge available and on the preparedness to actually use that knowledge, one has to choose from the methods indicated. Note that this choice implies an associated computational effort, which increases from top to bottom in Table 1.

Another type of potential a priori knowledge is the following. Suppose the probability density of the disturbances is not known, but some other information is, e.g. the range of the disturbing variable (its domain or support) and some constraint.
Given this a priori knowledge one may look for that particular probability model that is certainly not too optimistic, if the estimation results are based on its assumption (maximum information/uncertainty or maximum entropy approach); cf. Jaines (1968), Ponomarenko et al. (1982). Some simple cases are the following (Boekee, 1976):

a priori knowledge: domain	a priori knowledge: constraint	p(n) for maximum entropy choice
(a,b)	-	uniform distribution
$(0,\infty)$	$\varepsilon[n] = c_1$	negative exponential -
$(-\infty,+\infty)$	$\varepsilon[n^2] = c_2$	Gaussian -
$(-\infty,+\infty)$	$\varepsilon[\lvert n\rvert] = c_3$	Laplace -

2.2 Model knowledge

From the theoretical/methodological point of view it is appealing to build an identification approach on the assumption that no (subjective?) knowledge on the process is available.
From the engineering side, such an assumption is, of course, neither realistic nor sensible; any piece of knowledge that can be incorporated into the identification scheme might contribute to an enhanced feasibility of the estimation procedure to be implemented. Yet the following crude distinction with respect to the identification-goals is worth mentioning:

goals	type of model
- diagnostic applications - preventive maintenance	structural model (generic model) (explanatory model)
- interpretation of past behaviour - prediction of future behaviour - optimizing control - adaptive control	input/output model (non-generic model) (representation model)

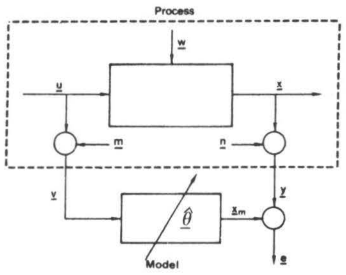


Fig.2 Structural model

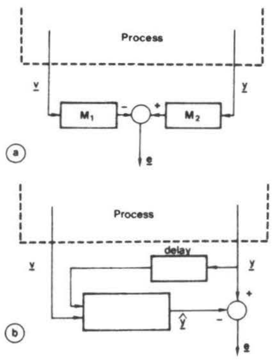


Fig.3 Input/output model

The structural model, cf. fig. 2, implies use of the "topology" of the process, the inclusion of known parameter values and, in many cases, the need to try to determine quantities which have a direct physical meaning. In this situation the attainment of good parameter estimates is the primary goal itself.

The input/output model, cf. fig. 3, implies the restriction of the attention to the overall cause-effect aspects of the process. The estimation quality is only a secondary, derived or intermediate goal.

The first type of model will generally be directed to the use of the output error; the second type of model provides more flexibility of model structures, viz. output error (fig. 3a, $M_2 = 1$), equation error (fig. 3a, $M_2 \neq 1$, generalized model) and prediction error (fig. 3b).

This flexibility can be used to create models that are linear-in-the-parameters and, due to this property, that lead to some interesting simplifications in estimation algorithms. The: AR (autoregressive; all pole), MA (moving average; all zero), ARMA (autoregressive-moving average; pole-zero), ARMAX (ARMA, with characterization of the noise) representations are examples of such flexibility.

2.3 Errors

Closely related to the intended use of the model is the aspect of errors that may occur in the modelling, measurement and estimation procedures. Generally speaking, this carries a particular weight in the case of structural models, where parameters with a direct physical meaning have to be determined. In an interesting paper, Grove, Bekey and Haywood (1980) recognize and discuss the following error sources:

- measurement errors: input; output; non-estimated (a priori "known") parameters
- modelling errors: structural; numerical
- optimization error.

In the light of these error sources they discuss the variability of the estimated parameters.

3 MODELS, LINEAR-IN-THE-PARAMETERS

3.1 Model structures

Definition of the notation to be used:

- process viz. process-with-measurement-noises, cf. fig. 4a and the equations (1) viz. (2)

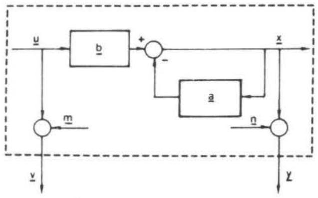


Fig.4a Linearity-in-parameters; observation noise

$$x(k) = -\sum_{i=1}^{p} a_i x_i(k) + \sum_{i=0}^{q} b_i u_i(k)$$

with

$$\underline{x} = [x(1), \ldots , x(k)]^T$$

$$\Omega(u,x) = [U \mid X] =$$

$$= \left[\begin{array}{ccc|ccc} u_o(1) & \ldots & u_q(1) & x_1(1) & \ldots & x_p(1) \\ \vdots & & \vdots & \vdots & & \vdots \\ u_o(k) & \ldots & u_q(k) & x_1(k) & \ldots & x_p(k) \end{array}\right]$$

$$\underline{\theta} = [\underline{b}^T \mid -\underline{a}^T]^T = [b_o, \ldots , b_q \mid -a_1, \ldots , -a_p]^T$$

this can be noted as:

$$\underline{x} = \Omega(u,x)\underline{\theta} \tag{1}$$

with noises $\underline{v} = \underline{u} + \underline{m}$ and $\underline{y} = \underline{x} + \underline{n}$:

$$\underline{y} = \Omega(v,y)\underline{\theta} + \underline{r} \tag{2}$$

- process with system noise, characterized through an ARMA structure operating on white noise w, cf. fig. 4b and eq. (3)

$$y(k) = -\sum_{i=1}^{p} a_i y_i(k) + \sum_{i=0}^{q} b_i u_i(k) +$$

$$+ \sum_{i=1}^{r} c_i w_i(k) - \sum_{i=1}^{s} d_i e_i(k) + w(k)$$

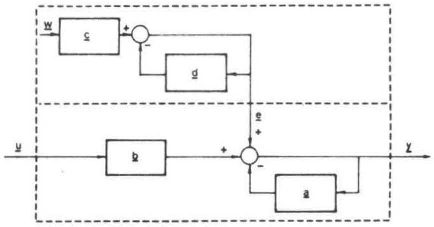


Fig.4b Linearity-in-parameters; process noise

The dynamics is represented by the forward process parameters $\underline{b}$ (moving-average-; MA model) and the backward process parameters $\underline{a}$ (autoregressive-; AR model). The samples w(k) represent white discrete noise; this is coloured by the forward (MA) noise parameters $\underline{c}$ and/or the backward (AR) noise parameters $\underline{d}$. With the appropriate vector definitions, this can be written:

$$y(k) = [\underline{u}^T \mid \underline{y}^T \mid \underline{w}^T \mid \underline{e}^T] \begin{bmatrix} \underline{b} \\ -\underline{a} \\ \underline{c} \\ -\underline{d} \end{bmatrix} + w(k)$$

or

$$\underline{y} = [U \mid Y \mid W \mid E]\, \underline{\theta} + \underline{w}$$

or

$$\underline{y} = \Omega(u,y,\dot{w},e)\underline{\theta} + \underline{w} \tag{3}$$

Ω will be interpreted as being a matrix composed from values of the variables used as arguments.

3.2 Template functions

The unifying approach to estimation to be considered here is based on the notion of "template functions" (Eykhoff, 1980). This notion is based on the recognition

- that most parameters of interest, governing the process dynamics, manifest themselves through the development of the (input and)output signal(s) in the course of time;
- that such signal(s) may be contaminated by process disturbances or unobserved inputs;
- that the measured process input and output data may be disturbed by measurement errors/noise.

These aspects require observation and data processing over an interval of time, say $0 \leqslant t \leqslant T$. This can be described through the notion of (integral) operators working over such a time interval on, say, signal y(t):

$$\langle f_i, y \rangle = \int_0^T f_i(t)y(t)dt \tag{4}$$

The time functions $f_i(t)$ are called "template functions" and they distinguish the operators. These operators may be:

(i) continuous in time ("analog" case), cf. eq. (4);

(ii) discrete in time ("sampled" case), cf. eq. (5).
The second type can be derived from the first one by introduction of Dirac "functions" $\delta(t-jT)$, viz.

$$<f_i,y> = \int_0^{kT_s} f_i(t)\delta(t-jT_s)y(t)dt = \sum_{j=0}^{k} f_i(jT_s)y(jT_s) = \underline{f}_i^T \underline{y} \qquad (5)$$

The notation < > refers to the notion of scalar product in function (abstract) spaces (e.g. E_k, E_∞, L_2), which may imply both discrete (finite and infinitely dimensional, E_k viz. E_∞) and continuous (L_2) cases. It also implies the utility of the associated notions and theory like manifolds, (sub)spaces, basis, orthogonalization, projection, adjoint operator, inverse operator, quadratic form, etc. (cf. e.g. Friedman, 1956).
Of particular importance is the projection theorem which states:
"If y is any vector not in a subspace M, there exists in M a vector w, called the projection of y on M, such that y-w is orthogonal to M, that is

$$<f,y-w> = 0 \qquad (6)$$

for every f in M," (ibid, p. 51 for statement and proof; cf. also Luenberger, 1969).

Another quite relevant notion is that of linear operator, say f = Ax; in the general framework of abstract spaces indicated this may represent (ibid, p. 23):

- a k x k matrix in E_k such that f = Ax represents k simultaneous linear equations;
- an integral operator in L_2;
- a translation operator in L_2;
- a differential operator in L_2.

The projection theorem allows a direct application to the estimation procedures; one example is by equating y to the process output, w to the model output and f to one or more of the template functions that belong to the subspace M.

Example: One process parameter θ, estimated by $\hat{\theta}$:

$$\begin{aligned} &\text{process} \quad y(t) = u(t)\theta + n(t) \\ &\text{model} \quad y(t) - u(t)\hat{\theta} = e(t) \end{aligned} \qquad (7)$$

projection on f(t):

$$<f,y-u\hat{\theta}> = <f,e> = 0 \qquad (8)$$

or, with $\hat{\theta}$ constant:

$$\hat{\theta} = \frac{\langle f,y\rangle}{\langle f,u\rangle} \tag{9}$$

On this general principle several remarks can be made:

(i) Note that one way to derive template functions is through the definition of optimality, i.e. the type of error criterion that has to be minimized in the estimation procedure. Consider the case described by eq. (7). Then among the error criteria we have the following:

- integral squared error: $G_1 = \int_0^T e^2 dt \stackrel{\text{def}}{=} \langle e,e\rangle$

- integral absolute error: $G_2 = \int_0^T |e| dt \stackrel{\text{def}}{=} \langle |e|,1\rangle$

Differentiation with respect to $\hat{\theta}$ leads to

$$\frac{\partial G_1}{\partial \theta} = -2\ \langle u,e\rangle$$ template function: u

$$\frac{\partial G_2}{\partial \theta} = \langle u, \frac{|e|}{e}\rangle = \langle u, \mathrm{sgn}\,[e]\rangle$$ template function: u

signum operation on e.

More general error criteria can imply weighting functions, e.g. exponential forgetting.

(ii) The template function may be derived from e.g. u(t) through a <u>linear operator</u> L; in that case:

$$\hat{\theta} = \frac{\langle Lu,y\rangle}{\langle Lu,u\rangle} \tag{10}$$

Note that this is equivalent to the use of adjoint operator L* operating on the second terms of the scalar products:

$$\hat{\theta} = \frac{\langle u,L^*y\rangle}{\langle u,L^*u\rangle} \tag{11}$$

The cases indicated by eqs. (10) and (11) illustrate the fundamental distinction between constructing a special template function, viz. "filtering" the observation signals; cf. Van den Boom (1982).

(iii) The use of linear operators can be made even more extensive, e.g.:

$$\langle u,L_1^*L_2y\rangle = \langle L_1u,L_2y\rangle = \langle L_2^*L_1u,y\rangle$$

(iv) It is noteworthy that the template functions may also be derived by non-linear operations; e.g. quantizing Q[ ] and clipping sgn[ ] leading to expressions such as:

$$\langle Q[u],e\rangle \quad \text{and} \quad \langle \mathrm{sgn}[u],e\rangle \tag{12}$$

This provides a variety of simplified correlation schemes.

(v) The template functions to be used may be chosen according to the following purposes:

- to provide a number of equations that is sufficient to estimate the number of parameters;
- to weight the data in time and/or in frequency to enhance or emphasize the particular aspects;
- to add more weight to new data than to old (in the case of quasi-stationary behaviour of the parameters);
- to implement particular error criteria, taking into account the properties of the disturbances.

(vi) Although much theory is available for scalar products and for operators in infinite dimensional discrete (E_∞) and continuous (L_2) spaces, in the sequel the emphasis will lie on finite dimensional space (E_k), i.e. the discrete case, finite data lengths, with representations by vectors/matrices

Along the lines given before we consider a premultiplication of the process-model equation (2) with a matrix F^T, where the dimensions of F are equal to the dimensions of Ω. F is called "template function matrix". Then it is recognized that $\underline{r}$ is unobservable and its projection on the subspace spanned by F,

$$\frac{1}{k} F^T \underline{r},$$

can be chosen close to $\underline{0}$. If we assume this equal to zero in some sense (expectation, plim) then:

$$\frac{1}{k}[F^T\underline{y} - F^T\Omega\underline{\hat{\theta}}] = \frac{1}{k} F^T\underline{r} = \underline{0} \tag{13}$$

where $\underline{\hat{\theta}}$ is the estimator of the unknown parameters $\underline{\theta}$. Consequently this estimator can be written as:

$$\underline{\hat{\theta}} = [F^T\Omega]^{-1} F^T\underline{y} \tag{14}$$

provided of course that the inverse matrix exists. Substitution of the expression (2) for the process output into eq. (14) leads to:

$$\underline{\hat{\theta}} = \underline{\theta} + [F^T\Omega]^{-1} F^T\underline{r} \qquad (15)$$

from which statistical properties like (asymptotic) bias and (asymptotic) covariance can be found. In the sequel we will assume $\varepsilon[\underline{r}] = \underline{0}$ and we will need $\mathrm{cov}[\underline{r}] = \varepsilon[\underline{r}\,\underline{r}^T] = R$, the covariance matrix of the disturbances.

These estimators require all data to be available before the algorithm can be executed; these estimation methods carry a variety of names; e.g.

explicit; one-shot; batch; off-line.

Note that, due to the linearity-in-the-parameters, these estimators can easily be brought into a recursive form, having the same statistical properties as the "one-shot" estimators indicated by eq. (14). Such estimation methods are named:

implicit; sequential; recursive; on-line.

3.3 Three classes of instrumentation

In discussing the statistical properties, one notices that we may recognize three classes of situations:

class	template matrix F	observations matrix Ω
I	deterministic	deterministic
II	deterministic	stochastic
III	stochastic	stochastic

Here "deterministic" refers to the situation where the signals concerned can be measured without uncertainty (cf. also Van den Boom, 1976). For a summary of these classes, their methods and some examples, cf. Table 2. Note that many of the celebrated system estimation schemes are found in the examples.

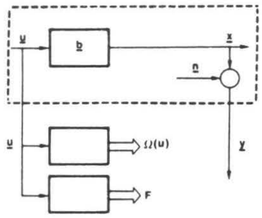
TABLE 2 Three classes of template functions, and some examples belonging to those classes

CLASS I:
- template function matrix F deterministic
- observation matrix Ω deterministic

The template function matrix F can be chosen appropriately i.e. as samples of functions that are independent of the process noise.

I-a cf. fig.(right)

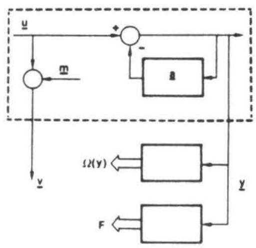
F is derived from the process input signal, measurable without disturbances/measurement errors;
Ω is also derived from the same signal; only forward parameters are being estimated.



Some relevant methods	F chosen as (consisting of):
- correlation method - matrix method with retarded observations	$\Omega\big(u(k-i)\big)$
- Volterra method - Wiener method	$\Omega\big(u(k-i),\ u^2(k-i),\ u^3(k-i),\ \ldots,\ \ldots,\ u(k-i)\ u(k-j),\ \ldots\ldots\ldots\big)$
- least squares method	$\Omega(u) = U$
- weighted l.s.method	$W\Omega(u) = WU$
- Markov method	$R^{-1}\Omega(u) = R^{-1}U$

I-b cf. fig.(right)

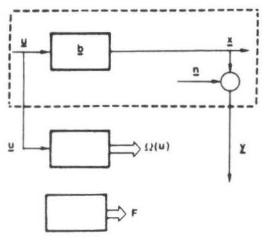
F is derived from the process output signal, measurable without disturbances/measurement errors;
Ω is also derived from the same signal; only backward parameters are being estimated.



Some relevant methods	F chosen as (consisting of):
- comparable to I-a	e.g. $\Omega\big(y(k-i)\big)$

I-c cf. fig.(right)

F consists of a set of deterministic functions, chosen independently of the process signals, e.g. sine/cosine-, orthogonal-, Walsh functions.
Ω is chosen as before.



Some relevant methods	F chosen as (consisting of):
This will be discussed as subclass II-C	

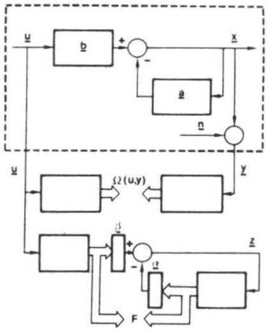
CLASS II: - template function matrix F deterministic
- observation matrix Ω stochastic

Now there are additive disturbances at the process output (and possibly also at the input). Forward as well as backward parameters are being determined:

II-a cf. fig. (right)

F is derived from the process input signal, measurable without disturbances/measurement errors;
Ω is derived from the input and output signals;

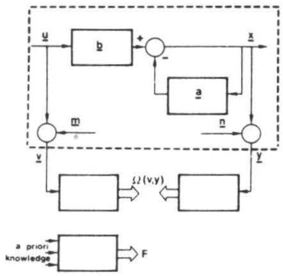
Note that to this class belong the instrumental variable techniques.



Some relevant methods	F chosen as (consisting of):
- instrumental variable method	$\Omega(u,z)$
- (optimal covariance) instrum.var.method	$R^{-1}\Omega(u,x)$
- delayed inputs instrum.var.method	$\Omega\{u,u(k-i)\}$

II-b cf. fig. (right)

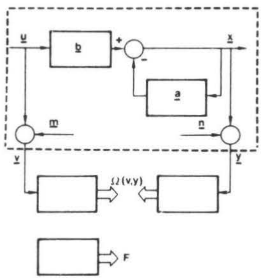
F is chosen independently of noisy measurements, e.g. based on rough knowledge that may be available a priori.
Such situations may occur in estimation on batch processes or on biomedical systems.
Ω is chosen as before.



Some relevant methods	F chosen as (consisting of):
- experience-derived template function method	F based on rough a priori e.g., in the case of batch processes, averaged over an "ensemble" of previous batches.

II-c cf. fig. (right)

F consists of a set of deterministic functions, chosen independently of the process signals, e.g. sine/cosine-, orthogonal-, Walsh functions
Ω is chosen as before.



Some relevant methods	F chosen as (consisting of):
- Fourier method	$\sin\omega_i t$; $\cos\omega_i t$
- Laplace method	$\exp[\alpha_i t]\sin\omega_i t$, etc.
- orthogonal development	orthogonal functions
- Shinbrot's method	method functions
- Strejc's method	reducing functions

CLASS III: - template function matrix F stochastic
- observation matrix Ω stochastic

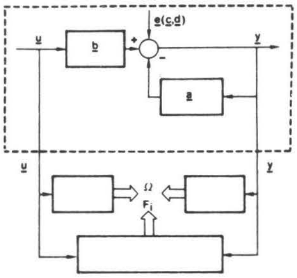
The template matrix is derived using disturbed process signals. One such a method is the differential approximation; as it leads to highly biased result we will leave this out of discussion. Other methods include:

III-a cf. fig. (right)

F consists of observations of measurable signals and estimations of unobservable quantities, e.g. in the representation of eq.(3):

$$\Omega(u,y,\hat{w},\hat{e})$$

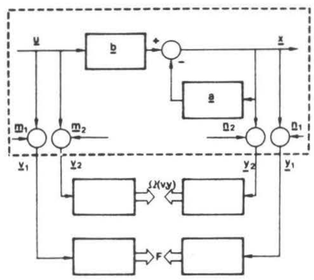
where $\hat{w}$ and $\hat{e}$ can only be estimated after a rough estimate of the parameters $\underline{a}$ and $\underline{b}$ from the input and output signals. Consequently the template function matrix develops in an iterative way.
The same holds for the observation matrix Ω.



Some relevant methods	F chosen as (consisting of):
- generalized l.squares - Clarke's method - Hastings-James' method	$D\Omega(u,y)$
- Young's method	$\Omega(u,y,\hat{w})$
- extendend matrix method (Talmon and Van den Boom, 1973) - extended l.squares - Panuska's method	$\Omega(u,y,\hat{w},\hat{e})$
- approx. ML method	$D\Omega(u,y,\hat{e})$

III-b cf. fig. (right)

F consists of noisy observations of process signals. Here it is assumed that two (or more) sequences of the same variable can be measured with errors that are mutually independent between these sequences. Such situations may occur if a quantity can be measured using different physical principles, e.g. measuring thickness by capacitive-, radiation-, or ultrasonic means;



Some relevant methods	F chosen as (consisting of):
- least-squares-like method (Van den Dungen, 1978,1981)	$\Omega(v_1,y_1)$

Some general remarks are in order here.
(i) The computational aspect of handling matrices leads to a variety of techniques, e.g. lattice algorithms, square root algorithms (Cholesky, Householder, Gram Schmidt); cf. Graupe et al. (1980), Strejc (1981).
(ii) Note that the time series identification/estimation can be considered as a special case of system identification, viz. the situation where the input signal u is not available/measurable. This implies that from Table 2 only a limited part is relevant for such time series characterization problems, i.e. I-c/II-b/II-c; III-a; III-b.
Again batch- and sequential algorithms can be distinguished; cf. Graupe, op.cit., who discusses: direct LS-, sample covariance-, autocorrelation- and partial correlation (PARCOR) methods, even though it does not consider additive (measurement) disturbances.
(iii) Of course the "estimation efficiency" and the "information-density-of-the-parameters-used" strongly depends on the choice of the set of template functions. If realizable, a close relation between the parameter set and the template function set is, of course, "optimal".
(iv) If test signals can be applied, then the template matrix will be derived from this signal and hence be deterministic, i.e. instrumentation class I or II. By the choice of the test functions, the information-efficiency properties can be optimized.
(v) Note that these considerations are not limited to SISO (single input - single output) situations, but can also be used for MIMO (multi input - multi output) processes that can be represented by models linear-in-the-parameters, e.g. the Hankel model.

On several subclasses and examples some notes can also be made, e.g.:

- ad I-a and I-b: The simplest cases may provide some interesting additional insight. Consider the matrix

$$P = U[U^TU]^{-1}\, U^T \tag{16}$$

which is called an orthogonal projection matrix, having the following properties:

$P^2 = P$ idempotent; same for (I-P)

$P = P^T = P^*$ symmetric, self adjoint

If $\underline{y} = \underline{x} + \underline{n} = U\underline{\theta} + \underline{n}$, then $P\underline{y} = \hat{\underline{x}}$ with the property that the error $\underline{y} - \hat{\underline{x}}$ is orthogonal to $\underline{y}$:

$$\langle\hat{x}, y - \hat{x}\rangle = \langle Py, (I-P)y\rangle = \langle y, (P^*-P^*P)y\rangle = 0$$

i.e. $\hat{\underline{x}}$ is the orthogonal projection of $\underline{y}$. Using the LS estimator:

$$\hat{\underline{\theta}} = [U^TU]^{-1}U^T\underline{y} \tag{17}$$

one finds indeed

$$\hat{\underline{x}} = U\hat{\underline{\theta}} = U\left[U^TU\right]^{-1} U^T\underline{y} = P\underline{y} \tag{18}$$

- ad I-c and II-c: For estimating m parameters it could be proposed to project $\underline{e}$, viz. $\underline{y}$ and $\Omega\hat{\underline{\theta}}$ on a template matrix consisting of m (sampled) sine and cosine functions. Generally speaking, this might well result in little information per parameter. One may use $n > m$ such functions in F, but then $F^T\Omega$ is not square and not invertible. A solution is to premultiply with matrix M with proper dimensions

$$MF^T\Omega\hat{\underline{\theta}} = MF^T\underline{y} \tag{19}$$

or

$$\hat{\underline{\theta}} = \left[MF^T\Omega\right]^{-1} MF^T\underline{y}. \tag{20}$$

This can be interpreted as creating a new template matrix $F_1^T = MF^T$ consisting of linear combinations of those sine/cosine functions in such a way that they provide a better "information efficiency".

- ad II-a: Note that the instrumental variable method with delayed, noise corrupted outputs belongs to class III.

4 MODELS, NON-LINEAR-IN-THE-PARAMETERS

Now we consider the scheme indicated by fig. 2. It has been mentioned before that the use of structural knowledge of the process may easily lead to an error signal that is non-linear-in-the-parameters. Assumed are additive disturbances $\underline{n}$ at the output and not at the input ($\underline{m} = \underline{0}$).
The error is:

$$\underline{e} = \underline{y} - \underline{x}_m(\hat{\underline{\theta}}) = \underline{x} - \underline{x}_m(\hat{\underline{\theta}}) + \underline{n} \tag{21}$$

and the criterion of optimality can be simply the minimization of some (weighted) squared error, e.g.:

$$G = \underline{e}^T\underline{e} \tag{22}$$

The wide variety of minimization/hill descending algorithms available, e.g. steepest descent (without or with minimization along a line), Newton-Raphson, Gauss-Newton, Marquart, conjugate gradient/Fletcher-Powell/
Davidson, Jacobson and Oksman, stochastic approximation, all require at least (an approximation of) the gradient

$$\frac{\partial G}{\partial \hat{\underline{\theta}}} = 2\,\frac{\partial \underline{e}^T}{\partial \hat{\underline{\theta}}}\,\underline{e} = -2\,\frac{\partial \underline{x}_m^T(\hat{\underline{\theta}})}{\partial \hat{\underline{\theta}}}\,\underline{e} \overset{\text{def}}{=} F^T_{x,\hat{\theta}}\,\underline{e} \tag{23}$$

which implies, as in the previous cases, a projection of the error. Now, however, the dependence of F on $\underline{\hat{\theta}}$ makes the method more cumbersome, as in principle for each iteration the F matrix has to be determined through a sensitivity model or by means of a finite difference approximation. Consequently, bias, convergence rate, local and global minima, influence of input/test signals and a priori asssumptions are barely discussable in general terms.

One may try to perceive a relation with the GMDH (Group Method of Data Handling) method, suggested and explored by Ivakhnenko (1970). Here the structure of the model is not known or chosen beforehand, but derived in an iterative way, using pre-chosen (standard, non-linear) components as building blocks for the model.

5 INFORMATION ASPECTS

The goal of the techniques being discussed is to derive information on unknown (structure and) parameters. Consequently it is necessary to briefly indicate what is the situation with respect to the "information generating" capabilities of these respective classes of instrumentation.

TABLE 3 Information limits due to additive noise

class	condition	statistical properties	remarks
linear-in-the-parameters			
I F determ. Ω determ.	$\varepsilon[\underline{r}] = \underline{0}$ $\mathrm{cov}[\underline{r}] = R$ non-singular	$\varepsilon[\underline{\hat{\theta}}] = \underline{\theta}$ $\mathrm{cov}[\underline{\hat{\theta}}] = [F^T\Omega]^{-1}F^TRF[\Omega^TF]^{-1}$	$\underline{\hat{\theta}}$ Gaussian if $\underline{r}$ Gaussian for observation sequence of all lengths
II F determ. Ω stoch.	do. $\underset{k\to\infty}{\mathrm{plim}} \frac{1}{k} F^T\underline{r} = \underline{0}$ $\underset{k\to\infty}{\mathrm{plim}} \frac{1}{k} F^T\Omega$ exists, nonsingular	$\underset{k\to\infty}{\mathrm{plim}}\, \underline{\hat{\theta}} = \underline{\theta}$ asymptotically unbiased asymptotic covariance = $= [F^T\Omega]^{-1}F^TRF[\Omega^TF]^{-1}$	asymptotical properties only
III F stoch. Ω stoch.	EMM $\varepsilon[\underline{w}] = \underline{0}$ $\mathrm{cov}[\underline{w}] = \sigma_w^2 I$	$\mathrm{cov}[\underline{\hat{\theta}}] \geq [\Omega^T\Omega]^{-1}\sigma_w^2$	$F = \Omega$ is an unrealizable choice of template functions; yet it provides a limit measure of the accuracy
nonlinear-in-the-parameters			
IV	$\left.\begin{array}{l}\varepsilon[\underline{n}] = \underline{0}\\ \mathrm{cov}[\underline{n}] = N\end{array}\right\}$ coloured $\left.\begin{array}{l}\varepsilon[\underline{n}] = \underline{0}\\ \mathrm{cov}[\underline{n}] = \sigma_n^2 I\end{array}\right\}$ white	$\mathrm{cov}[\underline{\hat{\theta}}] = \begin{cases}[S^TS]^{-1}S^TNS[S^TS]^{-1}\\ [S^TS]^{-1}\sigma_n^2\end{cases}$	near the optimal model adjustment only

Class I F deterministic; Ω deterministic
Taking eq. (14) as being representative for this class, one notices that the $\underline{y}$ term is the only stochastic one. From eq. (15) one can derive the statistical properties of $\hat{\underline{\theta}}$ as given in Table 3.

At this point it may be interesting to consider the optimal unbiased (Markov) estimator:

$$\hat{\underline{\theta}} = [U^T R^{-1} U]^{-1} U^T R^{-1} \underline{y}$$

where R is the covariance matrix of the additive noise. This estimator is known to have the minimum variance of all linear unbiased estimators. It can be rewritten in two ways by recognizing:

$$R^{-1} = D^T D, \text{ thus}$$

$$\hat{\underline{\theta}} = [\{DU\}^T \{DU\}]^{-1} \{DU\}^T D\underline{y}$$

viz.

$$U^T R^{-1} = F^T \text{ , thus}$$

$$\hat{\underline{\theta}} = [F^T U]^{-1} F^T \underline{y}$$

The first represents an interpretation using a noise-whitening filter (Eykhoff, 1974, pag. 193; Van den Boom, 1982), and the second can be recognized as a template-function-interpretation; cf. also eq. (11) and (10). Now being interested in the latter representation it is worth noting that the optimal template function matrix apparently consists of the independent variables weighted by the noise covariance matrix:

$$F_{opt} = R^{-1} U$$

If $\underline{r}$ is Gaussian, then eq. (15) represents a linear operation on a Gaussian distribution, resulting in a Gaussian estimate. As this type of probability density function is completely determined by $\varepsilon[\hat{\underline{\theta}}]$ and $\text{cov}[\hat{\underline{\theta}}]$, we have complete statistical knowledge on the estimation convergence with an increasing number of measurements, even for short observation sequences.
Summarizing for class I, we observe:

- the information development during the estimation procedure is known even for short observation sequences;
- for Gaussian noise we obtain complete statistical information;
- the effect of the choice of a non-optimal template function matrix and its associated information loss can be studied.

Class II F deterministic; Ω stochastic
Starting again from eqs. (14) and (15) on notices that

$$\varepsilon\left[[F^T \Omega]^{-1} F^T \underline{r}\right] \qquad (24)$$

cannot be written as the product of the expectations of the two product terms, as Ω and $\underline{r}$ are statistically related. Therefore one has to be satisfied with the asymptotic development of the statistical properties of the estimator. The asymptotic properties are known to be as given in Table 3; cf. Ward (1977); Söderström, Ljung and Gustavsson (1978); Stoica and Söderström (1982).

Summarizing for class II we notice:

- the information development during the estimation procedure can be studied conveniently only for the asymptotic case, i.e. for sizable observation sequences;
- the effect of choosing various kinds of template functions can be studied. Such choices include the use of rough a priori knowledge and the use of sets of deterministic functions.

Class III F stochastic; Ω stochastic

For various schemes that belong to the class of 'pseudo-linear regression algorithms' (ELS; EMM; approximate ML) the asymptotic accuracy is discussed by Stoica et al. (1983). A simple analysis proceeds along the following lines. In Eykhoff et al. (1981) it is indicated that for eq. (3) with $\underline{w}$ = white noise, $\varepsilon[\underline{w}] = 0$, $\mathrm{cov}[\underline{w}] = \sigma_w^2 I$ the choice $F = \Omega$ leads to

$$\underline{\hat{\theta}} = [\Omega^T\Omega]^{-1}\Omega^T\underline{y} \tag{25}$$

and

$$\mathrm{cov}[\underline{\hat{\theta}}] = [\Omega^T\Omega]^{-1}\ \sigma_w^2 \tag{26}$$

As $\underline{w}$ and $\underline{e}$ are not measurable, this is an unrealizable choice of F. Yet it provides a simple limit to the maximum information attainable; cf. Table 3.

Example: Using the process and noise characteristics to be described by:

$$y(k) = -a_1 y(k-1) + b_o u(k) + c_1 w(k-1) + w(k) \tag{27}$$

with the assumptions

$$\varepsilon[u^2] = \sigma_u^2 \qquad \varepsilon[w^2] = \sigma_w^2 \qquad \text{u white; w white; u,w independent}$$

and the numerical values

$$a_1 = 0.7 \qquad b_o = 1 \qquad c_1 = 0.3 \qquad k = 250 \qquad \sigma_u = 1 \qquad \sigma_w = 1$$

Then eq. (26) is found to be

$$\mathrm{cov}\begin{bmatrix} a_1 \\ b_o \\ c_1 \end{bmatrix} \approx \frac{\sigma_w^2}{k}\begin{bmatrix} \dfrac{b_o^2\sigma_u^2 + (1+c_1^2 - 2a_1c_1)\sigma_w^2}{1-a_1^2} & 0 & \sigma_w^2 \\ 0 & \sigma_u^2 & 0 \\ \sigma_w^2 & 0 & \sigma_w^2 \end{bmatrix}^{-1}$$

Using the numerical values given, the covariance is:

$$10^{-2}\begin{bmatrix} 0.176 & 0 & -0.176 \\ 0 & 0.4 & 0 \\ -0.176 & 0 & 0.576 \end{bmatrix}$$

or

$$\sigma^2_{a_1} \approx .042 \qquad \sigma_{b_o} \approx .063 \qquad \sigma_{c_1} \approx .076$$

Once again we have to emphasize that this simple guess of the available information must be too optimistic due to the assumption that the "template functions matrix" can be determined without uncertainty, which certainly is not the case.

Another such measure can be obtained by using the Cramér-Rao theorem. For the sake of notational convenience, we change the symbols from those used in fig. 4 to the ones indicated in fig. 5, i.e. process transfer operators for linear dynamics, where:

$$\begin{aligned} A &= 1 + a_1 z^{-1} + \ldots + a_p z^{-p} \\ B &= b_o + b_1 z^{-1} + \ldots + b_q z^{-q} \\ C &= 1 + c_1 z^{-1} + \ldots + c_r z^{-r} \\ D &= 1 + d_1 z^{-1} + \ldots + d_s z^{-s} \end{aligned} \tag{28}$$

and $\hat{A}$, $\hat{B}$, $\hat{C}$ and $\hat{D}$ are the corresponding model operators. Using the Cramér-Rao theorem, assuming Gaussian noises along the lines indicated by Åström (1967), it can be shown (Costongs, 1979) that the information matrix is

$$J = \frac{k}{\sigma^2_w} \varepsilon[\underline{f}\,\underline{f}^T] \tag{29}$$

with

$$\underline{f}^T = [\frac{BD}{AC} u + \frac{1}{A} w, \ -\frac{D}{C} u, \ \frac{1}{D} w, \ -\frac{1}{C} w] \tag{30}$$

The physical interpretation of the components of this vector can be recognized from figures 5 and 6. This corresponds to the intuitive notion that the best estimate $\hat{A}$ of A is obtained if $\hat{B} = B$, $\hat{C} = C$, and $\hat{D} = D$, and that the same holds for the other model parameters. Consequently, the components of this vector are recognized as the signals that function as information carriers for the respective parameters.

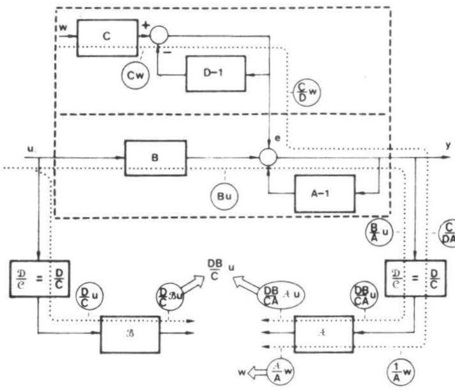


Fig.5 Information carriers; process parameters

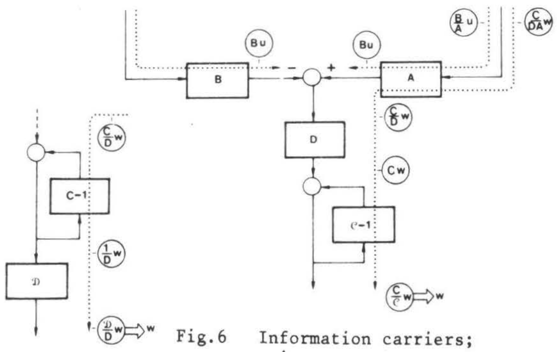


Fig.6 Information carriers; noise parameters

Example: Using the same process and noise description as before, eq. (27), we find the parameter covariance matrix

$$\operatorname{cov}\begin{bmatrix} a_1 \\ b_o \\ c_1 \end{bmatrix} \geqslant \frac{\sigma_w^2}{k}\begin{bmatrix} .11 & .04 & .12 \\ .04 & .38 & .05 \\ .12 & .05 & .51 \end{bmatrix} \tag{31}$$

or

$$\sigma_{a_1} \geqslant .03 \qquad \sigma_{b_o} \geqslant .06 \qquad \sigma_{c_1} \geqslant .07$$

The estimation situation given in Table 2, ad III-b, where two or more sequences of the same variable can be measured with errors that are mutually independent between these sequences, remains to be discussed. This type of stochastic template function matrix needs more extensive treatment and is discussed separately (Van den Dungen, 1978; Van den Dungen and Eykhoff, 1981).

Summarizing for class III we notice:

- a rough and simple estimation of the information available can be obtained quite simply;
- the information development during the estimation procedure can be discussed using the Cramér-Rao bound, indicating the maximum amount of information that can be derived on the estimated parameters. Simulation for particular cases of the extended matrix method indicates a rather close approach to this bound, whereas the amount of computation time needed compares favourably with the maximum likelihood method.

Class IV non-linearity-in-the-parameters

A general derivation of the information development still seems out of reach. An indication of the ultimately attainable information, as limited by the additive noise, can be discussed by considering a well-adjusted model. Near the optimal adjustment of the model, the error can be expressed as follows

$$\underline{e} = \underline{x} + \underline{n} - \underline{x}_m - \frac{\partial \underline{m}}{\partial \underline{\theta}^T}\Delta\underline{\theta} \approx \underline{n} - S\Delta\underline{\theta} \tag{32}$$

Minimization of $G = \underline{e}^T\underline{e}$ leads to

$$\underline{0} = \frac{\partial G}{\partial \underline{\theta}} = -\ 2S^T\underline{e} \tag{33}$$

or

$$\Delta\underline{\theta} = [S^TS]^{-1}\ S^T\underline{n} \tag{34}$$

cf. Table 3.
In this way one may obtain an indication of the accuracy that is maximally attainable in such a situation, provided the algorithm used converges to the global minimum.

6 CONCLUSIONS

In response to the chaotic picture of the identification/parameter estimation methods as suggested by fig. 1, some unifying concepts are recognized, viz:

- the elements of a priori knowledge available and being used;
- the notion of template function(s) (matrices).

The latter notion leads to a coherent picture that includes the majority of parameter estimation schemes proposed until now. Also it suggests some new approaches to identification, e.g. the use of template functions based on rough a priori knowledge (cf. Table 2, case II-b).

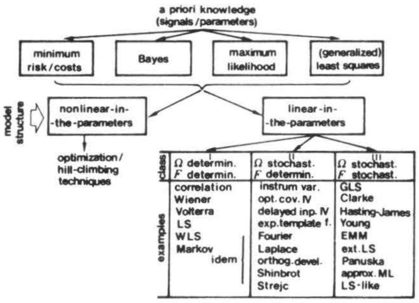


Fig.7 The bag of tricks ordered

Based on essential properties (deterministic versus stochastic) of the template matrix and the observation matrix, three classes of instrumentation are recognized. The estimation methods belonging to each of these respective classes share properties related to the information-development during the identification procedure. For linearity-in-the-paramters this holds for the 'one-shot' algorithm as well as for the corresponding recursive version.

Consequently, one may recognize a crucial step towards a coherent picture, cf. also fig. 7, and a basis for further comparison of parameter estimation methods.

References

- Akaike, H. (1981). Modern development of statistical methods. In: P. Eykhoff (edit.), Trends and Progress in System Identification. Pergamon, Oxford, p. 169-184.
- Åström, K.J. (1967). On the achievable accuracy in identification problems. Preprints 1st IFAC Symp. on Identification, Academia, Prague. Paper 1.8.
- Boekee, D.E. (1976). Maximum information in continuous systems with constraints. Proc. 8th Int. Congress on Cybernetics, Namur, Sept. 1976, p. 243-259.
- Boom, A.J.W. van den, cf. Van den Boom.
- Costongs, J.W. (1979). Accuracy of the parameter estimation techniques in the program package SATER (in Dutch). M.Sc. Thesis, EE Dept., Eindhoven University of Technology.
- Dungen, W.T.H.M. van den, cf. Van den Dungen.
- Eykhoff, P. (1974). System Identification. Wiley, London, 555 pp.
- Eykhoff, P. (1980). System identification: approach to a coherent picture through template functions. Electronics Letters, 16, p. 502-503.
- Eykhoff, P., A.J.W. Van den Boom and A.A. Van Rede (1981). System identification methods - unification and information development using template functions. Preprints/Proceed. 8th IFAC World Congress, Kyoto, Aug. 1981.
- Friedman, B. (1956). Principles and Techniques of Applied Mathematics. Wiley/Chapman and Hall, New York/London, 315 pp.
- Graupe, D., V.K. Jain and J. Salahi (1980). A comparative analysis of various least-squares identification algorithms. Automatica, 16, p. 663-681.
- Grove, T.M., G.A. Bekey and L.J. Haywood (1980). Analysis of errors in parameter estimation with application to physiological systems. Amer. J. of Physiology, 239, p. R390-R400.
- Ivakhnenko, A.G. (1970). Heuristic self-organization in problems of engineering-cybernetics. Automatica, 6, p. 207-219.
- Jaines, E.T. (1968). Prior probabilities. IEEE Trans. on Systems Science and Cybernetics, SSC-1, p. 227-241.
- Luenberger, D.G. (1969). Optimization by Vector Space Methods, Wiley, London, 326 pp.
- Peterka, V. (1981). Bayesian approach to system identification. In: P. Eykhoff (edit.), Trends and Progress in System Identification. Pergamon, Oxford, p. 239-304.
- Ponomarenko, M.F., P. Eykhoff and A.A. van Rede (1982). Information theory and identification; an account of an effective interaction. Preprints/Proc. 6th IFAC Symp. on Identification and Syst. Param. Estimation, Washington D.C.
- Söderström, T., L. Ljung and I. Gustavsson (1978). A theoretical analysis of recursive identification methods. Automatica, 14, p. 231-244.
- Söderström, T. and P. Stoica (1981). Comparison of some instrumental variable methods - consistency and accuracy aspects. Automatica, 17, p. 101-115.
- Stoica, P. and T. Söderström (1982). Identification of multivariable systems using instrumental variable methods. Proc. 6th IFAC Symp. on Identification and System Param. Estimation, Washington. Pergamon, Oxford, 1982.

- Stoica, P., T. Söderström, A. Ahlen and G. Solbrand (1983). On asymptotic accuracy of pseudo-linear regression algorithms. Preprints 5th Int. Conf. on Systems and Computer Science, Bucharest, June 1983.
- Strejc, V. (1981). Least squares and regression methods. In: P. Eykhoff (edit.), Trends and Progress in System Identification. Pergamon, Oxford, p. 103-143.
- Talmon, J.L. and A.J.W. Van den Boom (1973). On the estimation of the transfer function parameters of process- and noise dynamics using a single-stage estimator. Proc. 3rd IFAC Symp. on Identification and Syst. Param. Estimation. The Hague/Delft. North-Holland, Amsterdam, 1973.
- Van den Boom, A.J.W. (1976). On the relation between weighted least-squares estimators and instrumental variable estimators. Proc. 4th IFAC Symp. on Identification and System Param. Estimation, Tbilisi. North-Holland, Amsterdam, 1978, p. 1261-1271.
- Van den Boom, A.J.W. (1982). System Identification; on the Variety and Coherence in Parameter and Order Estimation Methods. Dr.-Dissertation, Eindhoven University of Technology, Sept. 1982.
- Van den Dungen, W.T.H.M. (1978). A generalization of the least squares estimator (in Dutch). M.Sc. Thesis, EE Dept., Eindhoven University of Technology.
- Van den Dungen, W.T.H.M. and P. Eykhoff (1981). A generalization of least squares estimators. Preprints/Proceed. 8th IFAC World Congress, Kyoto, Aug. 1981.
- Ward, R.K. (1977). Notes on the instrumental variable method. IEEE Trans. Autom. Control, AC-22, p. 482-484.

The author gratefully acknowledges helpful and stimulating contributions by dr. P. Stoica, Bucharest and by members of the group Measurement and Control, Eindhoven Univ. of Technology.
Ideas developed in this paper have previously been presented at the 6th IFAC Symp. on Identification and System Param. Estimation, Washington, 1982.

II A GENERALIZED DIAGRAM FOR VARIOUS PARAMETER ESTIMATION METHODS; THE COHERENCE IN PSEUDO-LINEAR REGRESSION SCHEMES

1. INTRODUCTION

For the estimation of the parameters of dynamical systems, several methods have been proposed in the last decade, often on a rather ad hoc basis. A characterization of the field by a "bag of tricks" seems to be appropriate, cf. Eykhoff e.a. (1981). Such a status of the field automatically asks for a thorough inventarization and schematization, resulting in a coherent picture. The advantage of a coherent presentation is threefold: - it gives a better understanding of the interrelations of the available estimators, - it clarifies their specific properties, - computer programs for such estimators can be designed in such a way that one program can represent each of the estimators considered by suitable "switches" in the program to be set by the user. The concepts which will be developed in the next sections and which will be the basis for the schematization of the field can then also be used fruitfully to develop new estimation schemes, e.g. for the situation where both input- and output signals are noise contaminated; for details cf. Van den Boom (1982, 1983).

In the following, we will restrict the discussion of the coherence of estimation schemes to recursive estimators. A classification for one-shot and/or iterative estimators can be given analogously; cf. Van den Boom (1982).
We will consider linear SISO processes, where, at the k-th time instant the input signal u_k is measured noise-free, and the measured output signal y_k is disturbed. The disturbances e_k may be due to measurement noise, to system noise or they may represent a non-adequate process description. We will use polynomial operators $[1+A(z^{-1})]$ and $[b_o+B(z^{-1})]$ for the description of the dynamics of the system and polynomial operators $[1+C(z^{-1})]$ and/or $[1+D(z^{-1})]$ for the description of the noise dynamics. The quantity ξ_k will represent a white noise which is assumed to be uncorrelated with u_k.

The simplest ARMA model for the process dynamics is

$$[1+A(z^{-1})]y_k = [b_o+B(z^{-1})]u_k + e_k \tag{1.1}$$

The following model incorporates ARMA modelling for both process- and noise dynamics; cf. Talmon and Van den Boom (1973) and see fig. 1.

$$[1+A(z^{-1})]y_k = [b_o+B(z^{-1})]u_k + \frac{[1+C(z^{-1})]}{[1+D(z^{-1})]}\xi_k \tag{1.2}$$

$$[1+A(z^{-1})]y_k = [b_o+B(z^{-1})]u_k - [D(z^{-1})]e_k + [1+C(z^{-1})]\xi_k \tag{1.3}$$

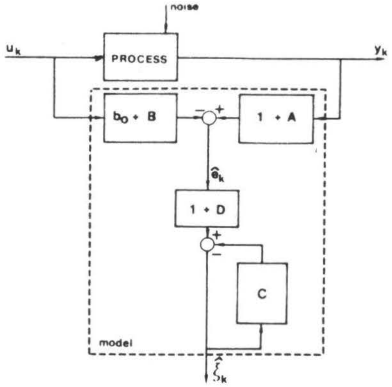


Fig. 1   The model of Talmon and Van den Boom

This model is a generalization of two models which have been previously proposed in the literature, cf. Åström and Bohlin (1965) and Clarke (1967) where the noise is modelled as a pure MA- or as a pure AR signal respectively. The model of (1.2) incorporates the advantages of both models, i.e. by properly choosing the degrees of the autoregressive AR and moving average MA parts, one can model the MA part of the noise by the MA part of the noise model and the AR part of the noise by the AR part of the noise model. There is no need for MA noises to be modelled by pure AR models as with Clarke's model. The Talmon and Van den Boom model gives therefore more flexibility to arrive at a minimal parameter set.

An important observation is that the above models are linear-in-the-parameters, provided that the signals e_k and ξ_k are available.
It is obvious that this will not be the case in practice, as only input and output samples u_k and y_k are available. The models which will be used then will need an estimate of these signals e_k and ξ_k.
These estimates can be obtained by making use of previous estimates of process- and noise parameters. Therefore, the above models are linear-in-the-process-parameters A and B but are non-linear-in-the-noise-parameters C and D. This will cause the appropriate estimation methods to need iterations or recursions to handle these non-linearities leading to various pseudo-linear regression algorithms.

2. RECURSIVE ESTIMATORS

In this section we will review briefly, mainly for establishing the notation and terminology, the recursive estimators. Define:

$$\left.\begin{aligned} &\underline{y}^T = (y_{q+1},\ldots,y_N) \\ &\Omega_N(u,y) = [U_N\ Y_N] = \left[\begin{array}{ccc|ccc} u_{q+1} & \cdots & u_{q+1-p} & y_q & \cdots & y_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ u_N & \cdots & u_{N-p-1} & y_{N-1} & \cdots & y_{N-q-1} \end{array}\right] \quad (2.1) \\ &\underline{\theta}^T = (b_o,\ldots,b_p, -a_1,\ldots,-a_q) \end{aligned}\right]$$

where p,q denote the degrees of the polynomials $[b_o+B(z^{-1})]$ and $[1+A(z^{-1})]$ resp., and the parameter vector $\underline{\theta}$ contains the elements of these polynomials. We know that a one-shot least squares estimator $\hat{\underline{\theta}}$ is given by, cf. Eykhoff (1974):

$$\hat{\underline{\theta}} = [\Omega_N^T(u,y)\Omega_N(u,y)]^{-1}\ \Omega_N^T(u,y)\underline{y} \qquad (2.2)$$

This estimate is based on N samples, and is calculated after these N samples have been acquired. When collecting these samples, such a type of explicit estimator gives no intermediate estimation results. We can construct a recursive estimator which, after the acquisition of the new measurements $(u_k,\ y_k)$, updates the previous estimate $\hat{\underline{\theta}}_{k-1}$ with $k \leq N$. This implicit or recursive estimator yields results identical to the explicit form (2.2), if initiated appropriately; cf. Eykhoff (1974).

$$\left.\begin{aligned} \hat{\underline{\theta}}_k &= \hat{\underline{\theta}}_{k-1} + (1+P_{k-1}\underline{\omega}_k^T P_{k-1}\underline{\omega}_k)^{-1}(y_k-\underline{\omega}_k^T\hat{\underline{\theta}}_{k-1}) \\ P_k &= P_{k-1} - P_{k-1}\underline{\omega}_k(1+\underline{\omega}_k^T P_{k-1}\underline{\omega}_k)^{-1}\underline{\omega}_k^T P_{k-1} \end{aligned}\right] \qquad (2.3)$$

where $\underline{\theta}_k$ is the estimate after k samples and

$$P_{k-1} = [\Omega_{k-1}^T(u,y)\Omega_{k-1}(u,y)]^{-1} \qquad (2.4)$$

$$\underline{\omega}_k^T = (u_k,\ldots,u_{k-p},y_{k-1},\ldots,y_{k-q}) \qquad (2.5)$$

so that $\underline{\omega}_k^T$ is the last row of $\Omega_k(u,y)$.

If we consider (2.3) in more detail we notice the term $y_k-\underline{\omega}_k^T\hat{\underline{\theta}}_{k-1}$ which is usually referred to as the <u>prediction error</u>. This prediction error is the difference between the next measurement y_k and its one-step-ahead prediction based on the model resulting from the previous parameter estimates. The prediction error is used to evaluate the gradient direction for obtaining the next model setting. Depending on the extent of the model used, we have the choices $\hat{e}_k$ and $\hat{\xi}_k$.

$$\left.\begin{aligned} \hat{e}_k &= [1+\hat{A}^{k-1}]y_k - [\hat{b}_o^{k-1}+\hat{B}^{k-1}]u_k \\ \hat{\xi}_k &= [1+\hat{D}^{k-1}]\hat{e}_k - [\hat{C}^{k-1}]\hat{\xi}_k \end{aligned}\right] \tag{2.6}$$

where $[1+\hat{A}^{k-1}]$, $[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$, $[1+\hat{C}^{k-1}]$ and $[1+\hat{D}^{k-1}]$ are the polynomials containing the estimated parameters after the k-1[st] recursion step.

Also <u>residuals</u> $\hat{\hat{e}}_k$ and $\hat{\hat{\xi}}_k$ can be defined, which can be computed in the k-th recursion step when the k-th estimate is available by analogous expressions such as (2.6), but using the estimated parameters obtained in the k-th step.
Taking $\hat{e}_k$ as prediction error, i.e. modelling only the ARMA (process) part of the model, yields the result of eq. (2.3). Extension of this model with ARMA noise parameters yields the following:

$$\Omega_{k-1}(u,y,\hat{\xi},\hat{e}) = [U_{k-1}|Y_{k-1}|\hat{\Xi}_{k-1}|\hat{E}_{k-1}] = \tag{2.7}$$

$$= \left[\begin{array}{ccc|ccc|ccc|ccc} u_{q+1} & \cdots & u_{q+1-p} & y_q & \cdots & y_1 & \hat{\xi}_q & \cdots & \hat{\xi}_{q-s+1} & \hat{e}_q & \cdots & \hat{e}_{q-r+1} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ u_{k-1} & \cdots & u_{k-p-1} & y_{k-2} & \cdots & y_{k-q-1} & \hat{\xi}_{k-2} & \cdots & \hat{\xi}_{k-s-1} & \hat{e}_{k-2} & \cdots & \hat{e}_{k-r-1} \end{array}\right]$$

$$\underline{\omega}_k^T = (u_k,..,u_{k-p},y_{k-1},..,y_{k-q},\ \hat{\xi}_{k-1},..,\hat{\xi}_{k-s},\ \hat{e}_{k-1},..,\hat{e}_{k-r}) \tag{2.8}$$

Applying a minimalization procedure to minimize in the k-th recursion, based on the k-1[st] model setting, the squared residual

$$V_k = \tfrac{1}{2}\hat{\hat{\xi}}_k^2 \tag{2.9}$$

using

$$\underline{\hat{\theta}}_k = \underline{\hat{\theta}}_{k-1} - Q_k \left.\frac{\partial V}{\partial \underline{\theta}}\right|_{\underline{\hat{\theta}}_{k-1}} \tag{2.10}$$

we find

$$\underline{\hat{\theta}}_k = \underline{\hat{\theta}}_{k-1} + Q_k \underline{\tilde{\omega}}_k \hat{\xi}_k \tag{2.11}$$

where the gradient direction is:

$$\left.\frac{\partial V}{\partial \underline{\theta}}\right|_{\underline{\hat{\theta}}_{k-1}} = -\underline{\tilde{\omega}}_k \; \hat{\xi}_k \tag{2.12}$$

with

$$\underline{\tilde{\omega}}_k = \frac{\partial \hat{\xi}_k}{\partial \underline{\theta}^T} = (\tilde{u}_k, \ldots, \tilde{u}_{k-p}, \tilde{y}_{k-1}, \ldots, \tilde{y}_{k-q}, \tilde{\hat{\xi}}_{k-1}, \ldots, \tilde{\hat{\xi}}_{k-s}, \tilde{\hat{e}}_{k-1}, \ldots, \tilde{\hat{e}}_{k-r}) \tag{2.13}$$

where we have the filtering:

$$\left.\begin{array}{ll} \tilde{u}_k = \dfrac{[1+\hat{D}]}{[1+\hat{C}]} u_k \; ; & \tilde{y}_k = \dfrac{[1+\hat{D}]}{[1+\hat{C}]} y_k \\ \tilde{\hat{e}}_k = \dfrac{1}{[1+\hat{D}]} \hat{e}_k \; ; & \tilde{\hat{\xi}}_k = \dfrac{1}{[1+\hat{C}]} \hat{\xi}_k \end{array}\right] \tag{2.14}$$

For the choice of Q_k, various possibilities have been proposed, ranging from a scalar quantity, as in the stochastic approximation schemes, to a matrix quantity meant for orthogonalization of the scheme such as the Newton-Raphson variants where Q_k can be interpreted as the (matrix) second derivative of the criterion function.

From this review of well-known principles leading to developing a recursive estimator, already two main operations have been found: model extension and filtering. These two basic operations are used separately from each other or combined in different schemes, as we will show in the next section. A third basic operation is the generation of an extra instrumental signal which has to be uncorrelated with the noise as appearing in eq. (1.1). (1.2) and (1.3) but which has to be correlated with the undisturbed process signals i.e. the input- and undisturbed output signals. This leads to the following instrumental variable (IV) estimator (which can also be written in the recursive form):-

$$\underline{\hat{\theta}} = [\Omega_N^T(v,w)\Omega_N(u,y)]^{-1} \; \Omega_N^T(v,w)\underline{y} \tag{2.15}$$

Here the matrix Ω_N (v,w) is defined analogously as $\Omega_N(u,y)$, the signals v,w being the instrumental variables. There is some freedom in choosing the IV signals, as long as the above-mentioned requirements are fulfilled; cf. Van den Boom (1982).

3. GENERAL CLASSIFICATION OF RECURSIVE ESTIMATORS

In this section we will classify the various recursive estimation schemes as presented in the literature. We will use the already mentioned concept of three basic operations for this classification. We will present diagrams which have to be interpreted within the context of the recursive character of these estimators.

Central in these diagrams, a block representing the recursive estimator is used, cf. fig. 2. It makes a recursion step from k-1 to k, as given by eq. (2.3) for the simple least squares case. It uses the necessary (filtered) input-output signals as in eq. (2.14), the IV quantities, the signals gained from the model extension as in eq. (2.7) and the prediction error. The prediction error is denoted by a dot in all the forthcoming diagrams.

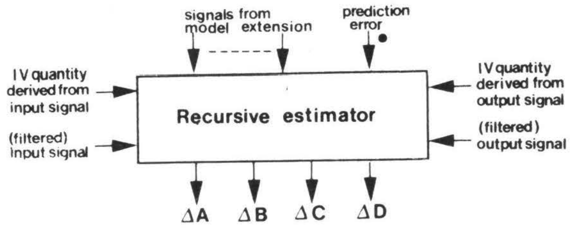


Fig. 2 Recursive estimator block

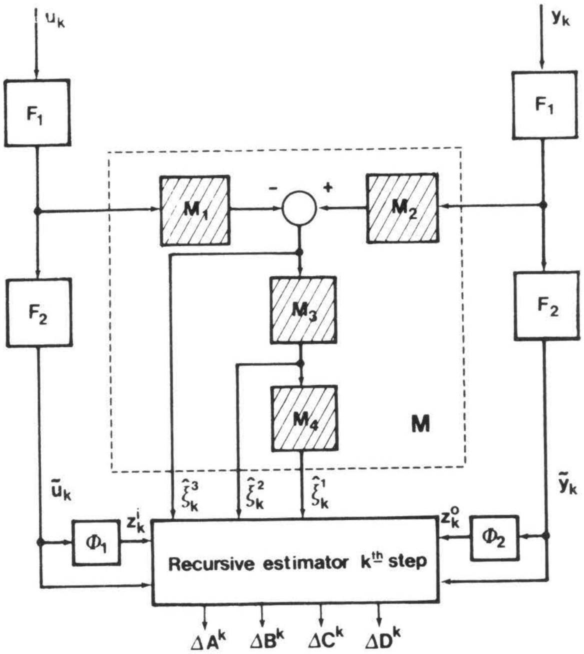


Fig. 3 The general recursive estimator

The diagrams of the various estimation schemes consist, apart from this recursive estimator block, of the following elements, cf. fig. 3.

- the model blocks M_1, M_2, M_2, M_4 having the values of the previously estimated parameters as model setting. The model has, in its most extended version, the form of the model of Talmon and Van den Boom (see section 1).
- the filters F_1 and F_2 for the additional noise whitening filtering, as already seen in section 2, eq. (2.14). The parameters of these filters may be tuned according to any possibly available information of the noise colouring, but usually they will follow from recently, i.e. in previous recursion step, estimated noise parameters, e.g. from the model extension part (as in EMM and AML) or from a separate estimator for the noise parameters (as in GLS).
- the IV filters Φ_1 and Φ_2. They generate the IV quantity from the input- or the output signal or both, or from extra signals, which have to be additionally available then. The latter gives the possibility of applying IV signals from extra measurements of signals which are related to the input- or output signals.

From fig. 3 it can also quite easily be seen how the noise is handled in the estimator. We have already seen that the correlation between the resulting error and the (shifted) input-output samples should be zero. This can be done in two ways:

a) operations which affect the noise colouring.
   Here we have two possibilities of making the resulting error white: a1) add extra filtering and/or
   a2) through the model extension resulting also in additional filtering. Depending on the choice of prediction error among ($\hat{\xi}_k^1$, $\hat{\xi}_k^2$, $\hat{\xi}_k^3$) the path for the equation error contained in y_k through F_1, M_2, M_3 and M_4 should be noise-whitening.

b) substitute the process signals by a related signal which contains no (or less) noise components. This is the IV approach and is realized by the filters Φ_1 and/or Φ_2.

The approaches listed under a) and b) may be combined, resulting in a family of estimators.

Fig. 4 shows the relationships between different recursive estimators. They will be discussed in more detail hereafter. It follows from this figure that estimators can be distinguished at four levels of complexity. The simplest estimator, LS, is very efficient with respect to computer time, but gives biased results, particularly for low S/N ratios. In cases where this is unwanted, the estimators of level 1 are of interest. They are more expensive with respect to computer time, but yield better results for low S/N ratios. Only in very specific cases, will these estimators diverge. The estimators of level 2 and 3 are even more complex and time consuming. With respect to noise sensitivity in relation to consistence, these estimators are superior to the estimators of level 0 and 1. Those estimators are not easy to apply as they have to be started by estimators of class 1 which, in their turn, have to be started by the LS estimator, except some variants of the IV estimator.

In the rest of this section we will show that all these estimators fit very well into the general scheme of figure 3. It can be observed that some estimation schemes have a bootstrap nature, as they use separate estimators for obtaining the necessary information for the filtering (such as GLS, IV-AML and suboptimal IV). These extra estimators are usually simple LS estimators for obtaining the AR or ARMA parameters of the model of the noise.

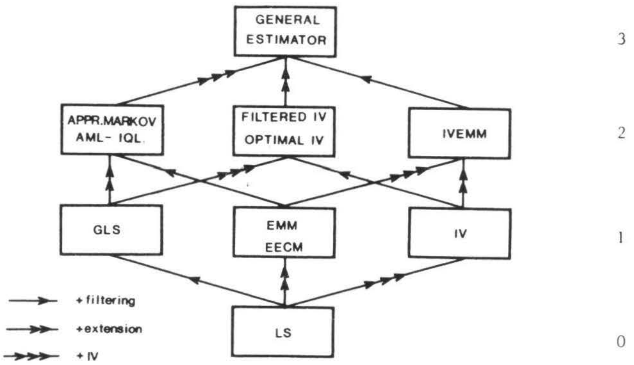


Fig. 4 Relation between recursive estimators

In table 1 a summary of the important aspects of the different recursive estimators is given. Here the particular choices for the model parts M_1, M_2, M_3 and M_4, the filters F_1 and F_2 and the IV filters Φ_1 and Φ_2 are shown.

The expression for the recursive estimator is given by

$$\left.\begin{aligned} \hat{\underline{\theta}}_k &= \hat{\underline{\theta}}_{k-1} + P_{k-1}\, z_k(1+\underline{\omega}_k^T\, P_{k-1}\underline{z}_k)^{-1}\, \hat{\xi}_k^i \\ P_k &= P_{k-1} - P_{k-1}\, \underline{z}_k(1+\underline{\omega}_k^T\, P_{k-1}\underline{z}_k)^{-1}\, \underline{\omega}_k^T\, P_{k-1} \end{aligned}\right\} \quad (3.1)$$

In this expression $\hat{\xi}_k^i$, $i = \{1,2,3\}$, is the prediction error, as listed in table 1 for the different methods and shown in fig. 3; in the figures 5,6,7,8,9,10,11,12 the prediction errors are also indicated for the different methods. The vector $\hat{\underline{\theta}}_k$ is the estimate. Its extent can be verified from table 1 under the heading "Model". Closely related to the extension of the parameter vector $\hat{\underline{\theta}}$ is the vector of measurables $\underline{\omega}_k$. This vector contains (filtered) measurements of the last inputs and outputs with an extension with the last estimates of the noise signals which are taken from the model. The contents of this vector $\underline{\omega}_k$ is shown in detail in the above mentioned figures for the different methods.

Comments with respect to the different estimators:

	Method	Model M_1	M_2	M_3	M_4	Signal filters F_1	F_2	Instr. Variable Filters Φ_1	Φ_2	prediction error	remarks
0	Least Squares	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-	-	-	-	-	$\hat{\xi}_k^3$	
1a	GLS	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-	$[1+\hat{D}^{k-1}]$	-	-	-	$\hat{\xi}_k^3$	separate est. for $[1+\hat{D}]$
1b	OLS	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-	-	-	-	-	$\hat{\xi}_k^3$	common factors in M_1,M_2
1b	EMM a)	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	-	-	-	-	-	$\hat{\xi}_k^2$	
	b)	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	$[1+\hat{C}^{k-1}]^{-1}$	-	-	-	-	$\hat{\xi}_k^1$	
	c)	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$	-	-	-	-	$\hat{\xi}_k^1$	
1c	IV	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-	-	-	a) fixed	-	$\hat{\xi}_k^3$	
								b) $\frac{[\hat{b}_o^{k-1}+\hat{B}^{k-1}]}{[1+\hat{A}^{k-1}]}$	-	$\hat{\xi}_k^3$	
								c) delay	-	$\hat{\xi}_k^3$	
								d) -	delay	$\hat{\xi}_k^3$	
								e) delay	delay	$\hat{\xi}_k^3$	
2ab	AML/IQL	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	$[1+\hat{C}^{k-1}]^{-1}$	$[1+\hat{C}^{k-1}]^{-1}$	-	-	-	$\hat{\xi}_k^3$	
2ab	Appr. Markov	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$	$[1+\hat{C}^{k-1}]^{-1}$	$[1+\hat{D}^{k-1}]$	-	-	$\hat{\xi}_k^2$	
2ac	Subopt. IV	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-	$[1+\hat{D}^{k-1}]$	-	$\frac{[\hat{b}_o^{k-1}+\hat{B}^{k-1}]}{[1+\hat{A}^{k-1}]}$	-	$\hat{\xi}_k^3$	separate est. for $[1+\hat{D}]$
2ac	IV-AML IV:	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	-	-	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$	$\frac{[\hat{b}_o^{k-1}+\hat{B}^{k-1}]}{[1+\hat{A}^{k-1}]}$	-	$\hat{\xi}_k^3$	
	AML:			$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$		$[1+\hat{C}^{k-1}]^{-1}$				
2bc	IVEMM	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$	-	-	see choices a-e under 1c		$\hat{\xi}_k^1$	
3abc	general	$[\hat{b}_o^{k-1}+\hat{B}^{k-1}]$	$[1+\hat{A}^{k-1}]$	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$	$[1+\hat{D}^{k-1}]$	$[1+\hat{C}^{k-1}]^{-1}$	see choices a-e under 1c		$\hat{\xi}_k^1$	

Table 1 Choice of Model, Signal Filters and Instrumental Variable Filters for different Estimators; see also fig. 3

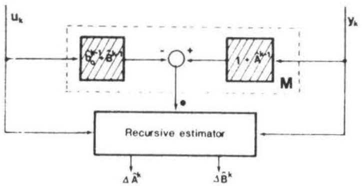


Fig. 5 The recursive least squares estimator

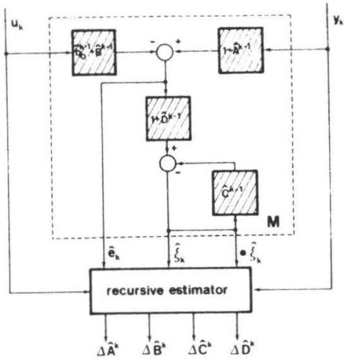


Fig. 7 The Extended Matrix Estimator (EMM)

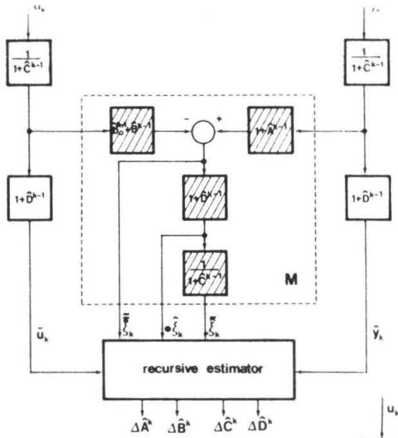


Fig. 10 The approximate Markov estimator

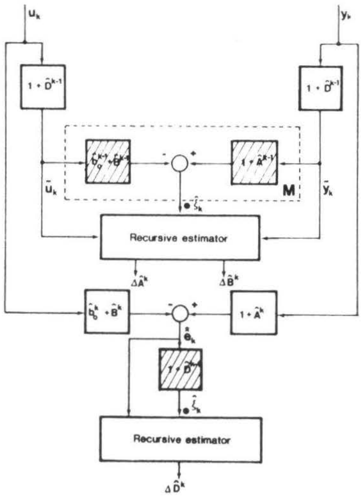


Fig. 6 The GLS estimator

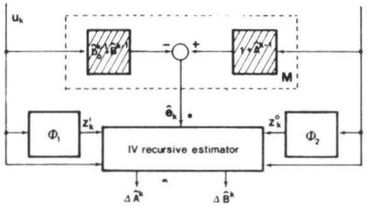


Fig. 8 The Instrumental Variable Estimator (IV)

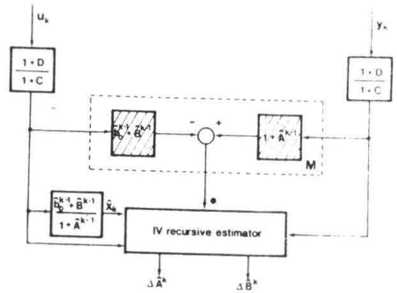


Fig. 11 The Suboptimal IV estimator

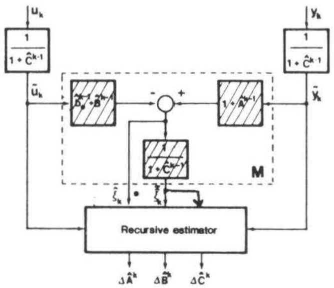


Fig. 9 The IQL/AML estimator (rearranged scheme)

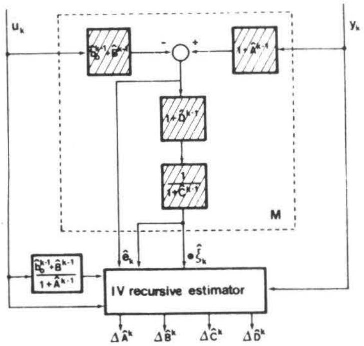


Fig. 12 The IV Extended Matrix (IVEMM) estimator

- Least Squares estimator (LS); cf. fig. 5.

Only ARMA process parameters are estimated. If the resulting prediction error is non-white, then the estimates will be biased, which can be undesired, depending on the intended use of the estimates.

- Generalized Least Squares estimator (GLS); cf. fig. 6.

A filtering operation is added in order to obtain white residuals. The filter parameters are obtained from a separate estimator of the noise parameters, which is also shown in fig. 6 for cases where the noise is modelled by an AR model, according to the formulation of Clarke (1967) and Hastings-James and Sage (1969). Stoica and Söderström (1977) proposed a variant where the noise is modelled by a MA model. An extension to ARMA modelling of the noise is straightforward in this respect, but has not yet been proposed for GLS.

- Over-parametrized Least Squares estimator (OLS)

This estimator has the same form as the LS estimator except for an extended ARMA model for the process parameters. If the extension is limited then pole-zero cancelling pairs in these ARMA parameters will occur which denote the AR noise model. This can be seen nicely if the filters $[1+\hat{D}^{k-1}]$ of the GLS estimator are shifted inwards the model block M. Then these AR noise parameters will be estimated as extended ARMA process parameters, which cause, at the same time, filtering as in GLS for obtaining white residuals. If a too large extension of the ARMA model of the process parameters is used then problems with respect to the uniqueness of the estimation results will occur.

- Extended Matrix Method estimator, (EMM); cf. fig. 7.

This method is a result of the extension of the parameter vector $\underline{\theta}$ with the noise parameters and consequently the extension of the vector of measurables with the noise signals, as appearing in eq. (1.3):

$$y_k = \underline{\omega}_k^T(u,y,\xi,e)\ \underline{\theta} + \xi_k \qquad (3.2)$$

so that a white equation error is left, cf. Talmon and Van den Boom (1973). As the signals ξ and e are not available, the residuals resulting from previous recursion steps are used. Usually this estimator produces good results except in special cases where the nature of the input signal and the noise colouring unfavourably coincide, cf. Van den Boom (1982).

- Instrumental Variable estimator (IV); cf. fig. 8.

This estimator relies on the availability of an IV signal z_k which originates from extra measurements, or which has to be generated from the available input-output signals. Several possibilities for the latter case are indicated in table 1.

- Approximate Maximum Likelihood estimator (AML)
  the Implicit Quasi Linearization estimator (IQL) and the
  Approximate Markov estimator, cf. fig. 9.

These estimators are according to the formulation in section 2, except for the differences in noise modelling.

In the formulation of the AML estimator, cf. Soderström (1973), only MA noise modelling is used, as is also shown in fig. 9. Also the IQL estimator, cf. Fuhrt (1973), uses MA noise modelling.

At first glance the formulation of the IQL estimator results in a more complex diagram than fig. 9 but by simple rearrangement of blocks, the diagram of fig. 9 can be obtained, cf. Van den Boom (1982).

The Approximate Markov estimator, cf. fig. 10, is closely related to the AML/IQL estimator; cf. Goedheer (1976) and fig. 10. The difference is in the noise modelling which is chosen here as an ARMA model and consequently the introduction of the filters F_2. The EMM estimator can be seen now as a simplified version of Approximate Markov estimator due to the lacking of the filtering operation.

- Suboptimal IV estimator (SIV); cf. fig. 11.

The IV estimator can be modified by introducing filtering of the input- and output signals and the IV signals. The resulting estimator is the SIV, which has been proposed by Wong and Polak (1967) for the case where the filtering is fixed and its parameters known. Usually these parameters are not known so that a separate estimator is necessary. Jakeman and Young (1983) proposed the use of the AML estimator for this purpose.

- IVEMM estimator, cf. fig. 12.

This is a modification of the EMM estimator as the IV operation has been added to the EMM scheme.

- General estimator

By comparing the diagrams of the approximate Markov estimator, the suboptimal IV estimator and the IV extended matrix estimator, we may propose a general diagram for these types of advanced estimators, leading to an even more versatile estimator: the general estimator, cf. fig. 3. This estimator is, in fact,

- the IV variant of the approximate Markov estimator or
- the extended model variant of the optimal IV estimator or
- the filtering variant of IVEMM.

4. CONCLUSION

In this paper we have recognized and presented a general set-up for simple and more complex least squares estimators. We showed that within this general framework the recursive estimators can be classified. This framework is based on the distinction of three basic operations related to estimation: noise filtering, model extension and instrumental variable generation. Based on these three main operations, a generalized estimator could be proposed, such that the existing recursive estimators are special cases of this general estimator. From this general concept, the relationship between the existing estimators can then be shown nicely and judicious choices can be made from the abundance of estimation methods advocated in the literature.

Literature

Åström, K.J. and T. Bohlin (1965)
"Numerical identification of linear dynamic systems from normal operating records". Proc. IFAC Symposium on Self Adaptive Control Systems, Teddington, U.K.

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

Boom, A.J.W. van den (1983)
"Estimators for linear dynamical SISO systems with disturbed input-output measurements".
Systems and Control Letters, vol. 3, pp. 133-144.

Clarke, D.W. (1967)
"Generalised least squares estimation of the parameter of a dynamic model". Prepr. IFAC Symposium on Identification in Automatic Control Systems, Prague, paper 3.17.

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

Eykhoff, P., A.J.W. van den Boom and A.A. van Rede (1981)
"System Identification methods - unification and information development using template functions". Prepr./Proc. 8th IFAC World Congress, Kyoto.

Fuhrt, B. (1973)
"Maximum likelihood identification of Åström model by quasilinearisation". Proc. Third IFAC Symposium on Identification and System Parameter Estimation. The Hague/Delft, the Netherlands, pp. 737-740.

Goedheer, L.D. (1976)
"Comparison of several identification methods and their mutual relations". M.Sc. Thesis, EE Dept., Eindhoven University of Technology, Netherlands.

Hastings-James, R. and M.W. Sage (1969)
"Recursive generalised least squares procedure for on line identification of process parameters". Proc. IEE, 116, pp. 2057-2062.

Jakeman, A.J. and P.C. Young (1983)
"Advanced methods of recursive time-series analysis". Int. J. of Control, vol. 37, pp. 1291-1310.

Söderström, T. (1973)
"An on-line algorithm for approximate maximum likelihood identification of linear dynamic systems". Report 7308, Lund Institute of Technology, Lund, Sweden.

Stoica P. and T. Söderström (1977)
"A method for the identification of linear systems using the generalized least squares principle", IEEE Trans. Aut. Control, vol. AC-22, pp. 631-634.

Talmon, J.L. and A.J.W. van den Boom (1973)
"On the estimation of the transfer function parameters of process- and noise dynamics using a single-stage estimator". Proc. Third IFAC Conference on Identification and System Parameter Estimation, the Hague/Delft, Netherlands, pp. 711-720.

Wong, K.Y. and E. Polak (1967)
"Identification of linear discrete time systems using the instrumental variable method". IEEE Trans. AC, AC-12, pp. 707-718.