Automatica, Vol. 7, pp. 123–162. Pergamon Press, 1971. Printed in Great Britain.

System Identification—A Survey*

Une revue de l'identification des systèmes

Systemidentifikation, eine Übersicht

Обзор опознавания систем

K. J. ÅSTRÖM† and P. EYKHOFF‡

Summary—The field of identification and process-parameter estimation has developed rapidly during the past decade. In this survey paper the state-of-the-art/science is presented in a systematic way. Attention is paid to general properties and to classification of identification problems. Model structures are discussed; their choice hinges on the purpose of the identification and on the available *a priori* knowledge.

For the identification of models that are linear in the parameters, the survey explains the least squares method and several of its variants which may solve the problem of correlated residuals, viz. repeated and generalized least squares, maximum likelihood method, instrumental variable method, tally principle.

Recently the non-linear situation, the on-line and the real-time identification have witnessed extensive developments. These are also reported. There are 230 references listed, mostly to recent contributions. In appendices a resumé is given of parameter estimation principles and a more detailed exposition of an example of least squares estimation.

Organization of the paper

The contents of the paper are arranged as follows:

* Received 9 February 1970; revised 24 August 1970. The original version of this paper was an Invited Survey Paper which was presented at the 2nd IFAC Symposium on Identification and Process Parameter Estimation, held in Prague, Czechoslovakia, during June 1970. It was recommended for publication in revised form by the editorial staff.

† Division of Automatic Control, Lund Institute of Technology, Lund, Sweden.

‡ Department of Electrical Engineering, Technical University, Eindhoven, Netherlands.

1. INTRODUCTION

IN RECENT years aspects of system identification have been discussed in a multitude of papers, at many conferences and in an appreciable number of university courses. Apparently the interest in this subject has different roots, e.g.

Definite needs by engineers in process industries to obtain a better knowledge about their plants for improved control. This holds not only for the chemical but also for the mechanical and other production industries.

The task of studying high performance aero and space vehicles, as well as the dynamics of more down-to-earth objects like railway carriages and hydrofoils.

Study of the human being in tracking action and in other types of control.

Research of biological functions, e.g. of neuro-muscular systems like the eye pupil response, arm or leg control, heart rate control, etc.

Not only the needs for, but also the possibilities of estimation have dramatically changed with the development of computer hardware and software. More or less apart from the "engineering" and "biological" approach the econometricians and statisticians have been working on dynamical economic models, leading to incidental cross-fertilization with engineering [1].

At many universities the field has been recognized as a legitimate subject for faculty and Ph.D. research. The net result of this development is a large number of publications, either accentuating a particular type of approach or describing a certain case study. In this survey paper the "motivation" of the identification is derived from control engineering applications.

Throughout the history of automatic control it has been known that the knowledge about a system and its environment, which is required to design a control system, is seldom available *a priori*. Even if the equations governing a system are known in principle it often happens that knowledge of particular parameters is missing. It is not uncommon that the models which are available are much too complex etc. Such situations naturally occur in many other fields. There are, however, two facts which are unique for the identification problems occurring in automatic control, i.e.

It is often possible to perform experiments on the system in order to obtain the lacking knowledge.

The purpose of the identification is to design a control strategy.

One of the factors which undoubtedly contributed very much to the great success of frequency response techniques in "classical" control theory was the fact that the design methods were accompanied by a very powerful technique for systems identification, i.e. frequency analysis. This technique made it possible to determine the transfer functions accurately, which is precisely what is needed to apply the synthesis methods based on logarithmic diagrams. The models used in "modern" control theory are with a few exceptions *parametric* models in terms of state equations. The desire to determine such models from experimental data has naturally renewed the interests of control engineers in parameter estimation and related techniques.

Status of the field

Although it is very difficult to get an overview of a field in rapid development we will try to point out a few facts which have struck us as being relevant when we prepared this survey.

The field of identification is at the moment rather bewildering, even for so-called experts. Many different methods and techniques are being analysed and treated. "New methods" are suggested en masse, and, on the surface, the field appears to look more like a bag of tricks than a unified subject. On the other hand many of the so-called different methods are in fact quite similar. It seems to be highly desirable to achieve some unification of the field. This means that an abstract frame-work to treat identification problems is needed. In this context it appears to us that the definition of an identification problem given by ZADEH [2] can be used as a starting point, i.e. an identification problem is characterized by three quantities: a class of models, a class of input signals, and a criterion. We have tried to emphasize this point of view throughout the paper.

For a survey paper like this it is out of question to strive for completeness. Limitations are given by: the number of relevant publications; the balance between the "educational" and the "expert" slant of this representation; the (in)coherence of the field and the wide spectrum of related topics.

Also it is desirable to keep in mind that until now a number of survey papers have been written, based on many references. For an indication where this new paper [0] stands with respect to the older ones the reader is presented with an enumeration of topics dealt with in the IFAC survey papers: EYKHOFF, VAN DER GRINTEN, KWAKERNAAK and VELTMAN [208], CUENOD and SAGE [206], EYKHOFF [22], and BALAKRISHNAN and PETERKA [35].

IDENTIFICATION IN AUTOMATIC CONTROL SYSTEMS

General aspects

The purpose of identification/estimation procedures

Identification	
—definition and formulation	[0]
—and control	[0], [35]
Model representation	
—*a priori* knowledge	[22]
—linear	[0], [22], [35], [206], [208]
—linear in the parameters	[0], [22]
—non-linear, general	[206]
—non-linear, Wiener	[206]
—non-linear, Volterra	[35]
—linear/non-linear-in-parameters	[0]
—multivariable	[0]
Industrial models	
—use	[208]
—examples dynamic/static	[208]

Formulation of the estimation problem

Classes of instrumentation	[0], [22], [35], [208]
—models	[0]
—input signal	[0]
—criteria	[0]

Other surveys and monographs of general interest are ALEKSANDROVSKII and DEICH [3], BEKEY [4], LEE [209], RAIBMAN [221] and STROBEL [6], [42].

2. GENERAL PROPERTIES OF IDENTIFICATION PROBLEMS

Purpose of identification

When formulating and solving an identification problem it is important to have the *purpose* of the identification in mind. In control problems the final goal is often to design control strategies for a particular system. There are, however, also situations where the primary interest is to analyse the properties of a system. Determination of rate coefficients in chemical reactions, heat transfer coefficients of industrial processes and reactivity coefficients in nuclear reactors are typical examples of such a "diagnostic" situation. In such a case determination of specific parameter values might be the final goal of the identification. Many problems of this type are also found in biology, economy, and medicine.

Even if the purpose of the identification is to design a control system the character of the problem might vary widely depending on the nature of the control problem. A few examples are given below:

Design a stable regulator.

Design a control program for optimal transition from one state to another.

Design a regulator which minimizes the variations in process variables due to disturbances.

In the first case it might be sufficient to have a fairly crude model of the system dynamics. The second control problem might require a fairly accurate model of the system dynamics. In the third problem it is also necessary to have a model of the environment of the system. Assuming that the ultimate aim of the identification is to design a control strategy for a system, what would constitute a satisfactory solution from a practical point of view?

In most practical problems there is seldom sufficient *a priori* information about a system and its environment to design a control system from *a priori* data only. It will often be necessary to make some kind of experiment, observe the process while using perturbations as input signals and observe the corresponding changes in process variables. In practice there are, however, often severe limitations on the experiments that can be performed. In order to get realistic models it is often necessary to carry out the experiments *during normal operation*. This means that if the system is perturbed, the perturbations must be small so that the production is hardly disturbed. It might be necessary to have several regulators in operation during the experiment in order to keep the process fluctuations within acceptable limits. This may have an important influence on the estimation results.

When carrying out identification experiments of this type there are many questions which arise naturally:

How should the experiment be planned? Should a sequential design be used, i.e. plan an experiment using the available *a priori* information, perform that experiment, plan a new experiment based on the results obtained, etc. When should the experimentation stop?

What kind of analysis should be applied to the results of the experiment in order to arrive at control strategies with desired properties? What confidence can be given to the results?

What type of perturbation signal should be used to get as good results as possible within the limits given by the experimental conditions?

If a digital computer is used what is a suitable choice of the sampling interval?

In spite of the large amount of work that has been carried out in the area of system identification we have at present practically no general answers to the problems raised above. In practice most of these general problems are therefore answered in an *ad hoc* manner, leaving the analysis to more specified problems. In a recent paper JACOB and ZADEH [7] discuss some of the questions in connection with the problem of identifying a finite state machine; c.f. also ANGEL and BEKEY [8] Some aspects of the choice of sampling intervals are given in FANTAUZZI [9] and SANO and TERAO [11]. Since the general problems discussed above are very difficult to formalize one may wonder if there will ever be rational answers to them. Nevertheless, it is worthwhile to recognize the fact that the final purpose of identification is often the design of a control system, since this simple observation may resolve many of the ambiguities of an identification problem. A typical example is the discussion whether the accuracy of an identification should be judged on the basis of deviations in the model *parameters* or in the *time response*. If the ultimate purpose is to design control systems then it seems logical that the accuracy of an identification should be judged on the basis of the *performance* of the control system designed from the results of the identification.

Formulation of identification problems

The following formulation of the identification problem was given by ZADEH [2]:

"Identification is the determination, on the basis of input and output, of a system within a specified class of systems, to which the system under test is equivalent."

Using Zadeh's formulation it is necessary to specify a class of systems, $\mathscr{S} = \{S\}$, a class of input signals, $\mathscr{U}$, and the meaning of "equivalent". In the following we will call "the system under test" simply the *process* and the elements of $\mathscr{S}$ will be called *models*. Equivalence is often defined in terms of a criterion or a *loss function* which is a functional of the process output y and the model output y_m, i.e.

$$V = V(y, y_m). \tag{1}$$

Two models m_1 and m_2 are then said to be *equivalent* if the value of the loss function is the same for both models, i.e.

$$V(y, y_{m_1}) = V(y, y_{m_2}).$$

There is a large freedom in the problem formulation which is reflected in the literature on identification problems. The selection of the class of models, $\mathscr{S}$, the class of input signals, $\mathscr{U}$, and the criterion is largely influenced by the *a priori* knowledge of the process as well as by the purpose of the identification.

When equivalence is defined by means of a loss function the identification problem is simply an *optimization problem:* find a model $S_0 \in \mathscr{S}$ such that the loss function is as small as possible. In such a case it is natural to ask several questions:

Is the minimum achieved?

Is there a unique solution?

Is the uniqueness of the solution influenced by the choice of input signals?

If the solution is not unique, what is the character of the models which give the same value of the loss function and how should $\mathscr{S}$ be restricted in order to ensure uniqueness?

Answers to some of these problems have been given for a simple class of linear systems arising in biomedical applications by BELLMAN and ÅSTRÖM [12]. The class of models $\mathscr{S}$ has been called *identifiable* if the optimization problem has a unique solution. Examples of identifiable and non-identifiable classes are also given.

The formulation of an identification problem as an optimization problem also makes it clear that there are connections between identification theory and approximation theory. Many examples are found in the literature, e.g. LAMPARD [13], KITAMORI [14], BARKER and HAWLEY [15], ROBERTS [16], [17], and others, where covariance functions are identified as coefficients in orthogonal series expansions. Recent examples are SCHWARZE [18], GORECKI and TUROWICZ [19], and DONATI and MILANESE [207].

Another type of identification problem is obtained by imbedding in a probabilistic framework. If $\mathscr{S}$ is defined as a parametric class, $\mathscr{S} = \{S_\beta\}$, where β is a parameter, the identification problem then reduces to a parameter estimation problem. Such a formulation makes it possible to exploit the tools of estimation and decision theory. In particular it is possible to use special estimation methods, e.g. the maximum likelihood method, Bayes' method, or the min–max method. It is possible to assign accuracies to the parameter estimates and to test various hypotheses.

Also in many probabilistic situations it turns out that the estimation problem can be reduced to an optimization problem. In such a case the loss function (1) is, however, given by the probabilistic assumptions. Conversely to a given loss function it is often possible to find a probabilistic interpretation. The problem can usually be handled analytically under gaussian assumptions. An approximation technique in the non-gaussian case is discussed by RAIBMAN and collaborators [215], [216].

There are several good books on estimation theory available, e.g. DEUTSCH [20] and NAHI [21]. A summary of the important concepts and their application to process identification is given by EYKHOFF [22]. An exposé of the elements of estimation theory is also given in Appendix A.

Also in the probabilistic case it is possible to define a concept of *identifiability* using the framework of estimation theory. In ÅSTRÖM and BOHLIN [23] a system is called identifiable if the estimate is consistent. A necessary condition is that the information matrix, associated with the estimation problem, is positive definite. Another concept of identifiability has been given by BALAKRISHNAN [24]. The question of identifiability has also been pursued by STALEY and YUE [25]. The concept of determinable classes introduced by ROOT [228], [229], [230] is also closely related to identifiability.

Relations between identification and control. The separation hypothesis.

Whenever the design of a control system for a partially known process is approached via identification it is an *a priori* assumption that the design can be divided into two steps: identification and control. In analogy with the theory of stochastic control we refer to this assumption as the *separation hypothesis*. The approach is very natural, in particular if we consider the multitude of techniques which have been developed for the design of systems with known process dynamics and known environments. However, it is seldom true that optimum solutions are obtained if a process is identified and the results of the identification are used in a design procedure, developed under the assumption that the process and its environment are known precisely. It can be necessary to modify the control strategy to take into account the fact that the identification is not precise. Conceptually, it is known how these problems should be handled. In the extreme case when identification and control are done simultaneously for a system with time-varying parameters the *dual control* concept of FEL'DBAUM [26] can be applied. This approach will, however, lead to exorbitant computational problems even for simple cases. C.f. MENDES [27] and the discussion in section 7 of this paper.

It can also be argued that the problem of controlling a process with unknown parameters can be approached without making reference to identification at all. As a typical example we mention online tuning of PID regulators. In any case it seems to be a worthwhile problem to investigate rigorously under what conditions the separation hypothesis is valid. Initial attempts in this direction have been made by SCHWARTZ and STEIGLITZ [28], ÅSTRÖM and WITTENMARK [29]. Apart from the obvious fact that it is desirable to choose a class of model $\mathscr{S}$ for which there is a control theory available, there are also many other interesting questions in the area of identification and control, e.g.

> Is it possible to obtain rational choices of model structures and criteria for the identification if we know that the results of identification will be used to design control strategies?
>
> What "accuracy" is required of the solution of an identification problem if the separation hypothesis should be valid at least with a specified error?

Partial answers to these questions are given by ÅSTRÖM and WITTENMARK [29] for a restricted class of problems.

Accuracy of identification

The problem of assigning accuracy to the result of an identification is an important problem and also a problem which always seems to give rise to discussions; e.g. QVARNSTRÖM [30], HERLES [31]. The reason is that it is possible to define accuracy in many different ways and that an identification which is accurate in one sense may be very inaccurate in another sense.

For example in the special case of linear systems it is possible to define accuracy in terms of deviations in the transfer function, in the impulse response, or in the parameters of a parametric model. Since the Fourier transform is an unbounded operator, small errors in the impulse response can very well give rise to large errors in the transfer function and *vice versa*. A discussion of this is given by UNBEHAUEN and SCHLEGEL [32] and by STROBEL [5]. It is also possible to construct

examples where there are large variations in a parametric model in spite of the fact that the corresponding impulse response does not change much. See e.g. STROBEL [42].

Many controversies can be resolved if we take the ultimate goal of the identification into account. This approach has been taken by ŠTĚPÁN [33] who considers the variation of the amplitude margin with the system dynamics. Such an analysis is ideally suited to analyse the case when the purpose of the identification is to design a stable proportional regulator. The following example illustrates the difficulties in a universal concept of accuracy.

Example. Consider the process S_T given by

$$\frac{dx}{dt}=u(t-T). \tag{2}$$

The transfer function is

$$H_T(s)=\frac{1}{s}\cdot e^{-sT} \tag{3}$$

and the unit step response is

$$h_T(t)=\begin{cases}0 & 0\leqslant t\leqslant T\\ t-T & t>T.\end{cases} \tag{4}$$

Assume that the process S_T is identified as S_0. Is it possible to give a sensible meaning to the accuracy of the identification? It is immediately clear that the differences

$$\max_t |h_T(t)-h_0(t)| \tag{5}$$

$$\max_\omega |H_T(j\omega)-H_0(j\omega)| \tag{6}$$

can be made arbitrarily small if T is chosen small enough. On this basis it thus seems reasonable to say that S_0 is an accurate representation of S_T if T is small. On the other hand the difference

$$|\log H_T(j\omega)-\log H_0(j\omega)|=\omega T \tag{7}$$

i.e. the difference in phase shift, can be made arbitrarily large, no matter how small we choose T. Finally, assume that it is desired to control the system (2) with the initial condition

$$x(0)=1 \tag{8}$$

in such a way that the criterion

$$V=\int_0^\infty \{\alpha^2x^2(t)+u^2(t)\}dt \tag{9}$$

is minimal. Suppose, that an identification has resulted in the model S_0 while the process is actually S_T. How large a deviation of the loss function is obtained? For S_0 the control strategy which minimizes (9) is given by

$$u(t)=-\alpha x(t). \tag{10}$$

The minimal value of the loss is

$$\min V=\alpha.$$

If $\alpha=1$ it can be shown that a very slight increase of the loss function is obtained if say $T=0{\cdot}001$. However, if $\alpha=2000$ ($>\pi/2T$) the criterion (9) will be infinite for the strategy (10) because the system is unstable. We thus find that the same model error is either negligible or disastrous depending on the properties of the loss function.

Limitations in the possibilities of identification are discussed by ŠTĚPÁN [34].

3. CLASSIFICATION OF IDENTIFICATION METHODS

The different identification schemes that are available can be classified according to the basic elements of the problem, i.e. the class of systems $\mathscr{S}$, the input signals $\mathscr{U}$, and the criterion. Apart from this it might also be of interest to classify them with respect to implementation and data processing requirements. For example: in many cases it might be sufficient to do all computations *off line*, while other problems might require that the results are obtained *on line*, i.e. at the same time the measurements are done. Classifications have been done extensively in EYKHOFF [22], BALAKRISHNAN and PETERKA [35].

The class of models $\mathscr{S}$

The models can be characterized in many different ways, by *nonparametric* representations such as impulse responses, transfer functions, covariance functions, spectral densities, Volterra series and by *parametric* models such as state models

$$\frac{dx}{dt}=f(x, u, \beta)$$

$$y=g(x, u, \beta) \tag{11}$$

where x is the state vector, u the input vector, y the output vector and β a parameter (vector). It is known that the parametric models can give results with large errors if the order of the model does not agree with the order of the process. An illustration of this is given in an example of section 5. A more detailed discussion of parametric model structures is given in section 4. The nonparametric representations have the advantage that it is not necessary to specify the order of the process explicitly. These representations are, however,

intrinsically infinite dimensional, which means that it is frequently possible to obtain a model such that its output agrees exactly with the process output. A typical example is given below.

Example. Suppose, that the class of models is taken as the class of linear time-invariant systems with a given transfer function. A reasonable estimate of the transfer function is then given by

$$\hat{H}_1(s)=\frac{\int_0^T y(t)e^{-st}dt}{\int_0^T u(t)e^{-st}dt}$$

where u is the input to the process and y is the output. To "eliminate disturbances" we might instead first compute the input covariance function

$$R_u(\tau)=\frac{1}{T}\int_0^{T-|\tau|} u(t)u(t+\tau)dt$$

and the input–output covariance

$$R_{uy}(\tau)=\begin{cases}\frac{1}{T}\int_0^{T-\tau} u(t)y(t+\tau)dt & \tau>0\\ \frac{1}{T}\int_\tau^T u(t)y(t+\tau)dt & \tau\leqslant 0\end{cases}$$

and then estimate the transfer function by

$$\hat{H}_2(s)=\frac{\int_{-T}^T R_{uy}(\tau)e^{-s\tau}d\tau}{\int_{-T}^T R_u(\tau)e^{-s\tau}d\tau}.$$

It is easy to show that $\hat{H}_1=\hat{H}_2$. The reason is simply that the chosen transfer function will make the model output exactly equal to the process output, at least if the process is initially at rest.

Interesting aspects of parametric versus nonparametric models are found in the literature on time series analysis. See for example MANN and WALD [36], WHITTLE [37], GRENANDER and ROSENBLATT [38], JENKINS and WATTS [39]. Needless to say the models must, of course, finally be judged with respect to the ultimate aim.

The class of input signals

It is well known that significant simplifications in the computations can be achieved by choosing input signals of a special type, e.g. impulse functions, step functions, "colored" or white noise, sinusoidal signals, pseudo-random binary noise (PRBS), etc. A bibliography on PRBS is given in NIKIFORUK and GUPTA [40]. See also GODFREY [41]. For the use of deterministic signals c.f. STROBEL [42], GITT [43], WILFERT [44], VAN DEN BOS [45], WELFONDER and HASENKOPF [46], CUMMING [47].

From the point of view of applications it seems highly desirable to use techniques which do not make strict limitations on the inputs. On the other hand if the input signals can be chosen, how should this be done? It has been shown by ÅSTRÖM and BOHLIN [23], ÅSTRÖM [48], AOKI and STALEY [49] that the condition of *persistent excitation* (of order n), i.e. that the limits

$$\bar{u}=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N u(k)$$

and

$$R_u(i)=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^N \{u(k)-\bar{u}\}\{u(k+i)-\bar{u}\} \tag{12}$$

exist and the matrix A_n defined by

$$A_n=\{a_{ij}=R_u(i-j)\}\quad i,j=1,\ldots,n \tag{13}$$

is positive definite, is sufficient to get consistent estimates for least squares, maximum likelihood and maximum likelihood in the special case of white measurement errors.

One might therefore perhaps dare to conjecture that a condition of this nature will be required in general.

Notice, that if the "mean square fourier transform" of the input, i.e.

$$f_u(\omega)=\lim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{k=1}^N [u(k)-\bar{u}]e^{i\omega k} \tag{14}$$

exists, then

$$R_u(\tau)=\int |f_u(\omega)|^2 e^{i\omega\tau}d\omega \tag{15}$$

and the matrix A_n given by (12) is automatically nonnegative definite for arbitrary n. Moreover, if

$$f_u(\omega)>0 \tag{16}$$

the matrix A_n is also positive definite for arbitrary n. The condition (16) is thus sufficient to guarantee that an input signal is persistently exciting of arbitrary order. Also notice that if the input signal is an ergodic stationary stochastic process then its spectral density is given by

$$\phi_u(\omega)=|f_u(\omega)|^2. \tag{17}$$

The condition (16) then implies that the spectral density of the input signal does not vanish for any ω, a condition which is well known for correlation analysis.

Apart from persistent excitation many applications will require that the output is kept within specified limits during the experiment. The problem of designing input signals, energy and time constrained, which are optimal e.g. in the sense that they minimize the variances of the estimates,

has been discussed by LEVADI [50], AOKI and STALEY [49], [198], and NAHI [213]. The same problem is also discussed in RAULT *et al.* [51]. It is closely related to the problem of optimal signal design in communication theory; see e.g. MIDDLETON [52].

Many identification procedures require that the input signal is independent of the disturbances acting on the process. If this condition is not fulfilled it might still be possible to identify the parameters. Each specific case must, however, be analyzed in detail. Typical cases where the input may depend on the disturbances are when normal operating records are used, and the system is under closed loop control as in adaptive systems.

The danger of identifying systems under closed loop control deserves to be emphasized. Consider the classical example of Fig. 1.

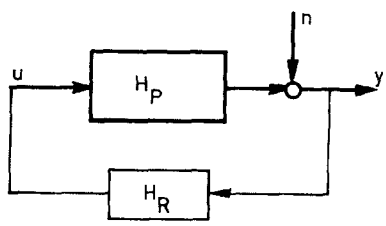

FIG. 1. Identification of a process in a closed loop.

An attempt to identify H_P from measurements of u and y will give

$$\hat{H}_P = \frac{1}{H_R}$$

i.e. the inverse of the transfer function of the feedback. In industrial applications the feedback can enter in very subtle ways e.g. through the action of an operator who makes occasional adjustment. FISHER [53] has shown the interesting result that the process may be identified if the feedback is made non-linear.

The criterion

It was mentioned in section 2 that the criterion is often a minimization of a scalar loss function. The loss function is chosen *ad hoc* when the identification problem is formulated as an optimization problem and it is a consequence of other assumptions when the problem is formulated as an estimation problem.

Mostly the criterion is expressed as a functional of an error e.g.

$$V(y, y_m) = \int_0^T e^2(t)dt \qquad (18)$$

where y is the process output, y_m the model output, and e the error; y, y_m, and e are considered as functions defined on $(0, T)$. Notice, that the criterion (18) can be interpreted as a least squares criterion for the error e. In the case

$$e = y - y_m = y - M(u) \qquad (19)$$

where $M(u)$ denotes the output of the model when the input is u, e is called the *output error*. It is the natural definition when the only disturbances are white noise errors in the measurement of the output.

In the case

$$e = u - u_m = u - M^{-1}(y) \qquad (20)$$

where $u_m = M^{-1}(y)$ denotes the input of the model which produces the output y, e is called the *input error*. The notation M^{-1} implies the assumption that the model is *invertible*, roughly speaking that it is always possible to find a unique input which produces a given output. Rigorous definitions of the concept of invertibility are discussed by BROCKETT and MESAROVIC [54], SILVERMAN [55], SAIN and MASSEY [56]. From the point of view of estimation theory the criterion (18) with the error defined as the input error (20) would be the natural criterion if the disturbances are white noise entering at the system input.

Since the output of a system depends not only on the input but also on the state, the error as defined by (19) and (20) will also depend on the initial state of the model M. This creates little difficulty in the case of finite dimensional state models since the initial state can always be included in the parameters. For nonparametric problems there might, however, be some difficulties. The problem has been discussed by SILVERMAN [57] in connection with correlation methods.

In a more general case the error can be defined as

$$e = M_2^{-1}(y) - M_1(u) \qquad (21)$$

where M_2 represents an invertible model. This type of model and error (21) are referred to as *generalized model* and *generalized error* EYKHOFF [113]. A special case of the generalized error is the "equation error" introduced by POTTS, ORNSTEIN and CLYMER [58]. Figure 2 gives an interpretation of the different error concepts in terms of a block diagram.

Computational aspects

All solutions to parametric identification problems consist of finding the extremum of the loss function V considered as a function of the parameters β. The minimization can be done in many different ways, e.g.

as a "*one-shot*" approach, i.e. solving the relations that have to be satisfied for the extremum of the function or functional V or:

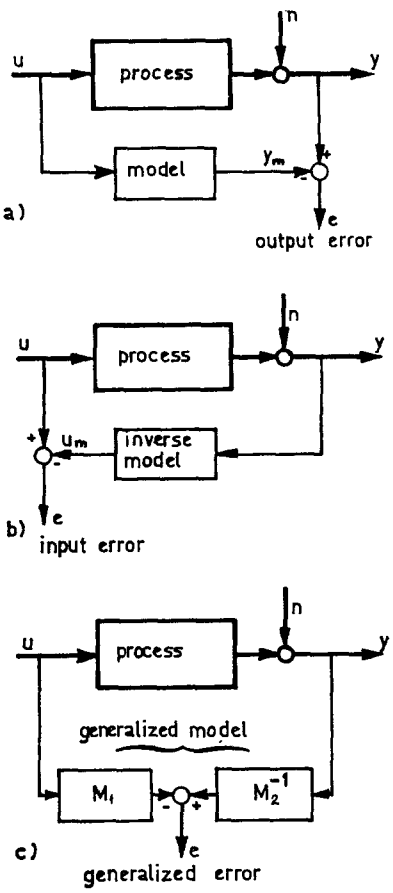

FIG. 2. Some error concepts.

as an *iterative* approach, i.e. by some type of hillclimbing. In this case numerous techniques are available, e.g.

(a) cyclic adjustment of the parameters one-by-one, a.o. Southwell relaxation method

(b) gradient method:

$$\beta(i+1)=\beta(i)-\Gamma V_\beta[\beta(i)] \qquad \Gamma>0^*$$

(c) steepest descent method:

$$\beta(i+1)=\beta(i)-\Gamma(i)V_\beta[\beta(i)] \qquad \Gamma(i)>0$$

$\Gamma(i)$ chosen such that $V(\beta)$ is minimum in the direction of the gradient.

(d) Newton's method:

$$\beta(i+1)=\beta(i)-\Gamma(i)V_\beta[\beta(i)]$$
$$\Gamma(i)=[V_{\beta\beta}(\beta(i))]^{-1}$$

(e) conjugate gradient method:

$$\beta(i+1)=\beta(i)-\Gamma(i)s(i)$$
$$s(i)=V_\beta(\beta(i))-\frac{\|V_\beta(\beta(i))\|^2}{\|V_\beta(\beta(i-1))\|^2}s(i-1)$$

$\Gamma(i)>0$ chosen to minimize $V(\beta(i)-\Gamma(i)s(i))$.

This method, applied to a positive definite quadratic function of n variables, can reach the minimum in at most n steps.

In these methods it has not been taken into account that in the practice of estimation the determination of the gradient is degraded through the stochastic aspects of the problem. A method which considers this uncertainty in the gradient-deetrmination is the:

(f) stochastic approximation method:

$$\beta(i+1)=\beta(i)-\Gamma(i)V_\beta(\beta(i))$$

where $\Gamma(i)$ has to fulfil the conditions:

$$\Gamma(i)\geqslant 0$$
$$\sum_{i=1}^{\infty}\Gamma^2(i)<\infty$$

and

$$\sum_{i=1}^{n}\Gamma(i)\to\infty$$

as $n\to\infty$.

In the scalar case $\Gamma(i)=1/i$ satisfies the above conditions.

A good survey of optimization techniques is found in the book by WILDE [59]. See also BEKEY and MCGEE [60].

4. CHOICE OF MODEL STRUCTURE

The choice of model structure is one of the basic ingredients in the formulation of the identification problem. The choice will greatly influence the character of the identification problem, such as: the way in which the results of the identification can be used in subsequent operations, the computational effort, the possibility to get unique solutions, etc. There are very few general results available with regard to the choice of structures.

In this section we will first discuss the concept of linearity in the parameters and we will then discuss the structure of linear systems.

The concept of linearity in the parameters

In control theory the distinction between linear and non-linear is usually based on the dynamic behaviour, i.e. the relation between *the dependent and the independent time variables.* For parameter estimation another distinction between linearity and nonlinearity is of as much importance, viz. with respect to the relation between *the dependent variables and the parameters.* To be specific a system is said to be linear in the parameters if the generalized error is linear in the parameters. Apparently, these two notions of linearity have no immediate relation as can be seen from the following examples.

* V_β is used as a shorthand notation for the gradient of V with respect to β:

$$V_\beta \overset{\text{def}}{=} \nabla_\beta V=\left[\frac{\partial V}{\partial\beta_1},\ldots,\frac{\partial V}{\partial\beta_m}\right],$$

We assume a process with input signal u and output signal y. Then the "model" may be chosen to form an "error" e between process and model output in the following way.

Such non-linear expressions, that can be made linear in the parameters through transformation, are called *intrinsically linear*. If such a linearization is not possible then *intrinsically non-linear* is used.

Process		Model: linear in the parameters	non-linear in the parameters
$\dot{y}+ay=u$	dynamic behaviour: linear	$e=\dot{y}+\alpha y-u$	$e=y-w=y-\dfrac{u}{D+\alpha}$ $\dot{w}+\alpha w=u$
$\dot{y}+ay^3=u$	dynamic behaviour: non-linear	$e=\dot{y}+\alpha y^3-u$	$e=y-w=y-\sigma[u,\ \alpha]$ $\dot{w}+\alpha w^3=u$

Henceforth we will use the term "linear" for the dynamic behaviour and use "linear in the parameters" for the other type.*

In connection with estimation schemes the great importance of linearity in the parameters will become clear. Therefore it pays to try to find transformations of the variables to obtain such a linearity if possible. Some simple examples may illustrate this.

Example 1. The transformation

$$\frac{1}{x_1}\to u_1,\quad \frac{1}{x_2}\to u_2,\quad \frac{\alpha_2}{\alpha_1}\to\beta_1,\quad \frac{1}{\alpha_1}\to\beta_2$$

brings the non-linear relation

$$z=\frac{\alpha_2 x_2+x_1}{\alpha_1 x_1 x_2}$$

to the linear form

$$z=\beta_1 u_1+\beta_2 u_2.$$

Example 2. The transformation

$$\log z\to y,\quad \log x_1\to u_1,\quad \log x_2\to u_2$$
$$\log c\to\beta_0,\quad \alpha_1\to\beta_1,\quad \alpha_2\to\beta_2$$

brings

$$z=cx_1^{\alpha 1}x_2^{\alpha 2}$$

to the linear form

$$y=\beta_0+\beta_1 u_1+\beta_2 u_2.$$

It may pay to make transformations even if the system is intrinsically non-linear; see e.g. DISKIND [61]. A typical example is the identification of discrete-time linear system when the output is measured with white measurement noise. The representation of the system by the coefficients of the pulse transfer function leads to a non-linear regression problem while the representation of the model by coefficients of a generalized model or by the ordinates of the weighting function leads to an estimation problem which is linear in the parameters.

Representation of linear systems

Linear time-invariant systems can be represented in many different ways by input–output descriptions, such as impulse response or transfer function H or by the state model $S(A, B, C, D)$ defined by

$$\frac{dx}{dt}=Ax+Bu$$
$$y=Cx+Du \qquad (22)$$

where x is an n-vector, the input u is a p-vector, and the output y is an r-vector. It is well-known that the systems $S(A, B, C, D)$ and $S(TAT^{-1}, TB, CT^{-1}, D)$ where T is a nonsingular matrix are equivalent in the sense that they have the same input–output relation. It is also easy to verify that the systems $S(A, B, C, D)$ and $S(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$ are equivalent in the sense that they have the same input–output relation if

$$D=\tilde{D}$$
$$CA^kB=\tilde{C}\tilde{A}^k\tilde{B} \qquad k=0, 1, \ldots, n. \qquad (23)$$

The relations between the different representations were clarified by Kalman's work; see e.g. KALMAN

* Note that also the term "order" may cause confusion. In regression analysis this term refers to the highest degree of the independent variable:
$y=\beta_0+\beta_1 u_1+n$ models of the first order
$y=\beta_0+\beta_1 u_1+\beta_2 u_2+\ \ldots\ +\beta_m u_m+n$ order
$y=\beta_0+\beta_1 u_1+\beta_2 u_1^2$ model of the second order

[62]. The impulse response and the transfer function only represent the part of the system S which is completely controllable and completely observable. It is thus clear that only the completely controllable and completely observable part of a state model $S(A, B, C, D)$ can be determined from input–output measurements. The impulse response and the transfer function are easily obtained from the state description. The problem of determining a state model from the impulse response is more subtle, even if we disregard the fact that only the controllable and observable subsystem can be determined from the impulse response. The problem of assigning a state model of the lowest possible order which has a given impulse response has been solved by HO and KALMAN [63]. See also KALMAN, FALB and ARBIB [64], and BRUNI, ISIDORI and RUBERTI [201]. Again the solution is not unique. The model $S(A, B, C, D)$ contains

$$N_1 = n^2 + np + nr + pr \tag{24}$$

parameters. The fact that the input–output relation is invariant under a linear transformation of the state variables implies that all N_1 parameters cannot be determined from input–output measurements. To obtain unique solutions as well as to be able to construct efficient algorithms it is therefore of great interest to find representations of the system which contain the smallest number of parameters, i.e. *canonical representations.*

Canonical forms for linear deterministic systems

Canonical forms for linear systems are discussed e.g. by KALMAN [62]. When the matrix A has distinct eigenvalues canonical forms can be obtained as follows. By a suitable choice of coordinates the matrix A can be brought to diagonal form.

$$\frac{dx}{dt} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} x + \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2p} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \beta_{n1} & \beta_{n2} & \dots & \beta_{np} \end{bmatrix} u$$

$$y = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \gamma_{r1} & \gamma_{r2} & \dots & \gamma_{rn} \end{bmatrix} x + \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1p} \\ d_{21} & d_{22} & \dots & d_{2p} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ d_{r1} & d_{r2} & \dots & d_{rp} \end{bmatrix} u \,. \tag{25}$$

This representation contains $n+np+nr+pr$ parameters. n of these are redundant since all state variables can be scaled without affecting the input–output relations. The input–output relation can thus be characterized by

$$N_2 = n(p+r) + pr \tag{26}$$

parameters. Since the system is completely controllable and observable there is at least one non zero element in each row of the B matrix and of each column of the C matrix. The redundancy in (25) can thus be reduced by imposing conditions like

$$\max_j \beta_{ij} = 1 \qquad i = 1, 2, \dots, n \tag{27}$$

$$\sum_j |\beta_{ij}| = 1 \qquad i = 1, 2, \dots, n \tag{28}$$

or similar conditions on the C matrix.

When the matrix A has multiple eigenvalues the problem of finding a minimal parameter representation is much more complex. If A is *cyclic*, i.e. there exists a vector b such that the vectors b, Ab, $A^2b, \dots, A^{n-1}b$ span the n-dimensional space, the matrix can be transformed to companion form and a minimal parameter representation is then given by

$$\frac{dx}{dt} = \begin{bmatrix} -a_1 & 1 & 0 & \dots & 0 \\ -a_2 & 0 & 1 & \dots & 0 \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ -a_{n-1} & 0 & 0 & \dots & 1 \\ -a_n & 0 & 0 & \dots & 0 \end{bmatrix} x$$

$$+\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ b_{n-1,1} & b_{n-1,2} & \dots & b_{n-1,p} \\ b_{n1} & b_{n2} & \dots & b_{np} \end{bmatrix} u$$

$$y=\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ c_{r1} & c_{r2} & \dots & c_{rn} \end{bmatrix} x$$

$$+\begin{bmatrix} d_{11} & d_{12} & \dots & d_{1p} \\ d_{21} & d_{22} & \dots & d_{2p} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ d_{r1} & d_{r2} & \dots & d_{rp} \end{bmatrix} u \qquad (29)$$

where n additional conditions, e.g. of the form (27) or (28) are imposed on the elements of the matrices B and C.

In the case of processes with one output the additional conditions are conveniently introduced by specifying all elements of the vector C, e.g. $C'=[1\ 0\ \dots\ 0]$. The canonical form (29) then becomes

$$Y(s)=\left[d_{11}+\frac{b_{11}s^{n-1}+b_{21}s^{n-2}+\dots+b_{n1}}{s^n+a_1s^{n-1}+\dots+a_n}\right]U_1(s)$$
$$+\dots+\left[d_{1p}+\frac{b_{1p}s^{n-1}+b_{2p}s^{n-2}+\dots+b_{np}}{s^n+a_1s^{n-1}+\dots+a_n}\right]U_p(s) \qquad (30)$$

where Y and U_i denote the Laplace transforms of y and u_i. A canonical representation of a process of the nth order with p inputs and one output can thus be written as

$$\frac{d^ny}{dt^n}+a_1\frac{d^{n-1}y}{dt^{n-1}}+\dots+a_ny=\left[b'_{o1}\frac{d^nu_1}{dt^n}+\dots\right.$$
$$\left.+b'_{n1}u_1\right]+\dots+\left[b'_{op}\frac{d^nu_p}{dt^n}+\dots+b'_{np}u_p\right] \qquad (31)$$

An analogous form for systems with several outputs is

$$\frac{d^ny}{dt^n}+A_1\frac{d^{n-1}y}{dt^{n-1}}+\dots+A_ny=\left[B_{o1}\frac{d^nu_1}{dt^n}+\dots\right.$$
$$\left.+B_{n1}u_1\right]+\dots+\left[B_{op}\frac{d^nu_p}{dt^n}+\dots+B_{np}u_p\right] \qquad (32)$$

where y is a r-vector and A_i $r\times r$ matrices.

This form was introduced by KOEPCKE [65]. It has been used among others by WONG *et al.* [66] and ROWE [67]. The determination of the order of the process (32), which in general is different from n, as well as the reduction of (32) for state form has been done by TUEL [68]. Canonical forms for linear multivariable systems have also been studied by LUENBERGER [69] and BUCY [205]. Analogous results hold for discrete time systems. When the matrix A has multiple eigenvalues and is not cyclic it is not clear what a "minimal parameter representation" means. The matrix A can, of course, always be transformed to Jordan canonical form. Since the eigenvalues of A are not distinct, the matrix A can, strictly speaking, be characterized by fewer than n parameters. The one's in the superdiagonal of the Jordan form can, however, be arranged in many different ways depending on the internal couplings. This leads to many different structures.

Canonical forms for linear stochastic systems

We will now discuss canonical forms for stochastic systems. To avoid the technical difficulties associated with continuous-time white noise we will present the results for discrete-time systems. The analogous results are, however, true also for continuous-time systems. Consider the system

$$x(k+1)=\Phi x(k)+\Gamma u(k)+v(k)$$
$$y(k)=\theta x(k)+Du(k)+e(k) \qquad (33)$$

where k takes integer values. The state vector x, the input u and, the output y have dimensions n,p, and r; $\{v(k)\}$ and $\{e(k)\}$ are sequences of independent equally-distributed random vectors with zero mean values and covariances R_1 and R_2. Since the covariance matrices are symmetric the model (33) contains

$$N_3=n^2+np+nr+pr+\tfrac{1}{2}n(n+1)+\tfrac{1}{2}r(r+1)$$
$$=n(\tfrac{3}{2}n+\tfrac{1}{2}+p+r)+r\left(p+\frac{r}{2}+\tfrac{1}{2}\right) \qquad (34)$$

parameters. Two models of the type (33) are said to be *equivalent* if: (i) their input–output relations are the same when $e=0$ and $v=0$, and (ii) the stochastic properties of the outputs are the same when

$u=0$. The parameters of Φ, Γ, and θ can be reduced by the techniques applied previously to deterministic systems.

It still remains to reduce the parameters representing the disturbances. This is accomplished e.g. by the Kalman filtering theorem. It follows from this that the output can be represented as

$$\begin{aligned} \hat{x}(k+1)&=\Phi\hat{x}(k)+\Gamma u(k)+K\varepsilon(k)\\ y(k)&=\theta\hat{x}(k)+Du(k)+\varepsilon(k) \end{aligned} \tag{35}$$

where $\hat{x}(k)$ denotes the conditional mean of $x(k)$ given $y(k-1), y(k-2), \ldots$, and $\{\varepsilon(k)\}$ is a sequence of independent equally distributed random variables with zero mean values and covariance R.

The single output version of the model (35) was used in ÅSTRÖM [199]. KAILATH [71] calls (35) an *innovations representation* of the process. A detailed discussion is given in ÅSTRÖM [72]. The model (35) is also used by MEHRA [73].

Notice, that if the model (35) is known, the steady state filtering and estimation problems are very easy to solve. Since K is the filter gain it is not necessary to solve any Riccati equation. Also notice that the state of the model (35) has physical interpretation as the conditional mean of the state of (33). If Φ is chosen to be in diagonal form and if conditions such as (27) are introduced on Γ and θ the model (35) is a canonical representation which contains

$$N_4=n(p+2r)+r\left(p+\frac{r}{2}+\tfrac{1}{2}\right) \tag{36}$$

parameters.

For systems with one output, where the additional conditions are as $\theta'=[1\,0\ldots0]$, the equation (35) then reduces to

$$\begin{aligned} &y(k)+a_1y(k-1)+\ldots+a_ny(k-n)\\ &\quad=[b'_{01}u_1(k)+\ldots+b'_{n1}u_1(k-n)]+\ldots\\ &\qquad+[b'_{op}u_p(k)+\ldots+b'_{np}u_p(k-n)]+\varepsilon(k)\\ &\qquad+c_1\varepsilon(k-1)+\ldots+c_n\varepsilon(k-n). \end{aligned} \tag{37}$$

By introducing the shift operator q defined by

$$qy(k)=y(k+1) \tag{38}$$

the polynomials

$$\begin{aligned} A(q)&=q^n+a_1q^{n-1}+\ldots+a_n\\ B_i(q)&=b'_{0i}q^n+b'_{1i}q^{n-1}+\ldots+b'_{ni}\\ &\quad i=1,2,\ldots,p\\ C(q)&=q^n+c_1q^{n-1}+\ldots+c_n \end{aligned} \tag{39}$$

and the corresponding reciprocal polynomials

$$\begin{aligned} A^*(q)&=q^nA(q^{-1})\\ B_i^*(q)&=q^nB_i(q^{-1})\\ C^*(q)&=q^nC(q^{-1}) \end{aligned} \tag{40}$$

the equation (37) can be written as

$$A^*(q^{-1})y(k)=\sum_{i=1}^{p}B_i^*(q^{-1})u_i(k)+C^*(q^{-1})\varepsilon(k) \tag{37'}$$

or

$$A(q)y(k)=\sum_{i=1}^{p}B_i(q)u_i(k)+C(q)\varepsilon(k). \tag{37''}$$

This canonical form of an nth order system was introduced by ÅSTRÖM, BOHLIN and WENSMARK [74] and has since then been used extensively. The corresponding form for multivariable systems is obtained by interpreting y and u_i as vectors and A, B_i and C as polynomials whose coefficients are matrices. Such models have been discussed by EATON [75], KASHYAP [76], ROWE [67] and VALIŠ [219].

The following canonical form

$$\begin{aligned} y(k)=&\frac{B_1(q)}{A_1(q)}u_1(k)+\frac{B_2(q)}{A_2(q)}u_2(k)+\ldots\\ &+\frac{B_p(q)}{A_p(q)}u_p(k)+\frac{C(q)}{A(q)}\varepsilon(t) \end{aligned} \tag{41}$$

has been used by BOHLIN [77], and STEIGLITZ and MCBRIDE [94] as an alternative to (37).

The choice of model structure can greatly influence the amount of work required to solve a particular problem. We illustrate this by an example.

A filtering example. Assume, that the final goal of the identification is to design a predictor using Kalman filtering. If the process is modeled by

$$\begin{aligned} x(k+1)&=\Phi x(k)+v(k)\\ y(k)&=\theta x(k)+e(k) \end{aligned} \tag{42}$$

where $\{e(k)\}$ and $\{v(k)\}$ are discrete-time white noise with covariances R_1 and R_2, the likelihood function for the estimation problem can be written as

$$\begin{aligned} -\log L=&\tfrac{1}{2}\sum_{k=1}^{n}[v'(k)R_1^{-1}v(k)+e'(k)R_2^{-1}e(k)]\\ &+\frac{n}{2}\log(\det R_1\cdot\det R_2)+\text{const}. \end{aligned} \tag{43}$$

where the system equations are considered as constraints. The evaluation of gradients of the loss function leads to a two point boundary value problem. Also when the identification is done, the solution of the Kalman filtering problem requires the solution of a Riccati equation.

Assume instead that the process is identified using the structure

$$z(k+1)=\Phi z(k)+K\varepsilon(k)$$
$$y(k)=\theta z(k)+\varepsilon(k) \tag{44}$$

the likelihood function then becomes

$$-\log L=\tfrac{1}{2}\sum_{k=1}^{n}\varepsilon'(k)R^{-1}\varepsilon(k)+\frac{n}{2}\log\det R\,. \tag{45}$$

The evaluation of gradients of the loss function in this case is done simply as an initial value problem. When the identification is done the steady state Kalman filter is simply given by

$$\hat{x}(k+1)=\Phi\hat{x}(k)+K[y(k)-\theta\hat{x}(k)]\,. \tag{46}$$

Hence if the model with the structure (44) is known there is no need to solve a Riccati equation in order to obtain the steady state Kalman filter.

5. IDENTIFICATION OF LINEAR SYSTEMS

Linear systems naturally represent the most extensively developed area in the field of systems identification. In this section we will consider linear systems as well as "linear environments", i.e. environments that can be characterized by linear stochastic models. In most control problems the properties of the environment will be just as important as the system dynamics, because it is the presence of disturbances that creates a control problem in the first place.

To formulate the identification problem using the framework of section 2 the class of models $\mathscr{S}$, the inputs $\mathscr{U}$, and the criterion must be defined. These problems were discussed in sections 3 and 4. If classical design techniques are to be used, the model can be characterized by a transfer function or by an impulse response. Many recently developed design methods will, however, require a state model, i.e. a parametric model.

Several problems naturally arise:

Suppose the impulse response is desired. Should this be identified directly or is it "better" to identify a parametric model and then compute the impulse response?

Assume, that a parametric model is desired. Should this be fitted directly or is it "better" to first determine the impulse response and then fit a parametric model to that?

Since a parametric model contains the order of the system explicitly what happens if the wrong order is assumed in the problem formulation?

A linear stochastic system can be decomposed into a deterministic dynamic system with additive noise corrupting the ideal output. Is it "better" to determine the deterministic model separately and then (if required) the statistical characteristics of the additive noise, or should the overall model be determined first and then decomposed, if necessary?

There are not yet any general answers to these problems. Special cases have been investigated by GUSTAVSSON [78] in connection with identification of nuclear reactor and distillation tower dynamics as well as on simulated data. Since correlation techniques, their properties and applications by now are very well known we will not discuss these here. Let it suffice to mention the recent papers by RAKE [79], WELFONDER [80], BUCHTA [81], HAYASHI [82], REID [83], [84], STASSEN [85], GERDIN [86], BROWN [202], KOSZELNIK *et al.* [203], and RAIBMAN *et al.* [204]. Instead we will concentrate on the more recent results on the identification of parametric models.

Least squares identification of a parametric model

Consider a linear, time invariant, discrete-time model with one input and one output. A canonical form for the model is

$$y_m(k)+a_1y_m(k-1)+\ \ldots\ +a_ny_m(k-n)$$
$$=b_1u(k-1)+\ \ldots\ +b_nu(k-n) \tag{47}$$

where u is the input and y_m the output of the model. Using the notation introduced in section 4 the model (47) can be written as

$$A(q)y_m(k)=B(q)u(k) \tag{47'}$$

or

$$A^*(q^{-1})y_m(k)=B^*(q^{-1})u(k)\,. \tag{47''}$$

Let the criterion be chosen as to minimize the loss function (1), i.e.

$$V=V(y,\,y_m)=\sum_{k=n}^{N+n}e^2(k) \tag{48}$$

where e is the generalized error defined by

$$e(k)=A^*(q^{-1})[y(k)-y_m(k)]$$

or

$$e(k)=A^*(q^{-1})y(k)-B^*(q^{-1})u(k) \tag{49}$$

and the last equality follows from (47''). The main reason for choosing this particular criterion is that

the error e is linear in the parameters a_i and b_i. The function V is consequently quadratic and it is easy to find its minimum analytically. Notice that (49) implies

$$y(k)+a_1y(k-1)+\ldots+a_ny(k-n) = b_1u(k-1)+\ldots+b_nu(k-n)+e(k). \quad (50)$$

The quantities $e(k)$ are also called *residuals* or equation errors. The criterion (48) is called minimization of "equation error". In Fig. 3 we give a block diagram which illustrates how the generalized error can be obtained from the process inputs and outputs and the model parameters a_i and b_i in the least squares method.

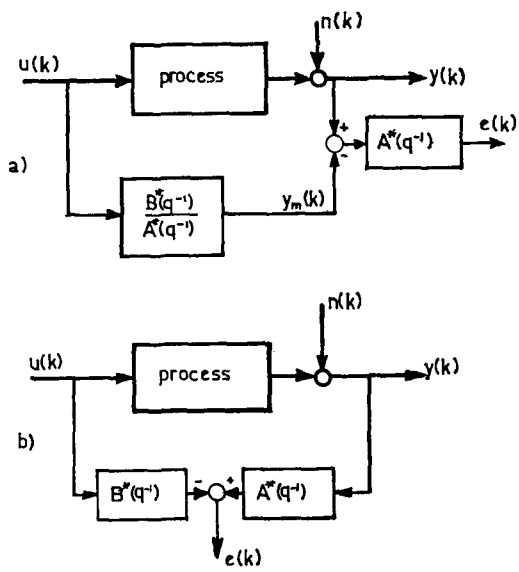

FIG. 3. Definition of the "equation error".

To find the minimum of the loss function V we introduce

$$y=\begin{bmatrix} y(n+1) \\ y(n+2) \\ \vdots \\ y(n+N) \end{bmatrix}$$

$$\Phi=\left[\begin{array}{cccc:cccc} -y(n) & -y(n-1) & \ldots & -y(1) & u(n) & u(n-1) & \ldots & u(1) \\ -y(n+1) & -y(n) & \ldots & -y(2) & u(n+1) & u(n) & \ldots & u(2) \\ \vdots & & & \vdots & \vdots & & & \vdots \\ -y(N+n-1) & -y(N+n-2) & \ldots & -y(N) & u(N+n-1) & \ldots & \ldots & u(N) \end{array}\right]$$

$$\beta'=[a_1a_2\ \ldots\ a_nb_1b_2\ \ldots\ b_u]. \quad (51)$$

The equation defining the error (49) then becomes

$$e=y-\Phi\beta. \quad (52)$$

The minimum of the loss function is found through $\nabla V=0$. If $\Phi'\Phi$ is not singular then this minimum is obtained for $\beta=\hat{\beta}$:

$$\hat{\beta}=[\Phi'\Phi]^{-1}\Phi'y. \quad (53)$$

It is thus a simple matter to determine the least squares estimate. The matrices $\Phi'y$ and $\Phi'\Phi$ are given in (54) and (55). Notice that $\Phi'\Phi$ is symmetric.

$$\Phi'y=\begin{bmatrix} -\sum_{k=n+1}^{N+n} y(k)y(k-1) \\ -\sum_{k=n+1}^{N+n} y(k)y(k-2) \\ \vdots \\ -\sum_{k=n+1}^{N+n} y(k)y(k-n) \\ \hline \sum_{k=n+1}^{N+n} y(k)u(k-1) \\ \sum_{k=n+1}^{N+n} y(k)u(k-2) \\ \vdots \\ \sum_{k=n+1}^{N+n} y(k)u(k-n) \end{bmatrix} \quad (54)$$

$$\Phi'\Phi = \left[\begin{array}{cccc|cccc}
\sum_{k=n}^{N+n-1} y^2(k) & \sum_{k=n}^{N+n-1} y(k)y(k-1) & \cdots & \sum_{k=n}^{N+n-1} y(k)y(k-n+1) & -\sum_{k=n}^{N+n-1} y(k)u(k) & -\sum_{k=n}^{N+n-1} y(k)u(k-1) & \cdots & -\sum_{k=n}^{N+n-1} y(k)u(k-n+1) \\
 & \sum_{k=n-1}^{N+n-2} y^2(k) & \cdots & \sum_{k=n-1}^{N+n-2} y(k)y(k-n+2) & -\sum_{k=n-1}^{N+n-2} y(k)u(k+1) & -\sum_{k=n-1}^{N+n-2} y(k)u(k) & \cdots & -\sum_{k=n-1}^{N+n-2} y(k)u(k-n+2) \\
 & & & \vdots & \vdots & \vdots & & \vdots \\
 & & & \sum_{k=1}^{N} y^2(k) & -\sum_{k=1}^{N} y(k)u(k+n-1) & -\sum_{k=1}^{N} y(k)u(k+n-2) & \cdots & -\sum_{k=1}^{N} y(k)u(k) \\
\hline
 & & & & \sum_{k=n}^{N+n-1} u^2(k) & \sum_{k=n}^{N+n-1} u(k)u(k-1) & \cdots & \sum_{k=n}^{N+n-1} u(k)u(k-n+1) \\
 & & & & & \sum_{k=n-1}^{N+n-2} u^2(k) & \cdots & \sum_{k=n-1}^{N+n-2} u(k)u(k-n+2) \\
 & & & & & & & \vdots \\
 & & & & & & & \sum_{k=1}^{N} u^2(k)
\end{array}\right] \tag{55}$$

For literature on matrix inversion the reader is referred to WESTLAKE [87].

Notice, that the technique of this section can immediately be applied to the identification of non-linear processes which are linear in the parameters, e.g.

$$y(k)+ay(k-1)=b_1u(k-1)+b_2u^2(k-1). \quad (56)$$

A probabilistic interpretation

Consider the least squares identification problem which has just been discussed. Assume, that it is of interest to assign accuracies to the parameter estimates as well as to find methods to determine the order of the system if it is not known. Such questions can be answered by imbedding the problem in a probabilistic framework by making suitable assumptions on the residuals. For example if it is assumed that

The input–output data is generated by the model where the residuals $\{e(k)\}$ are independent, equally distributed have zero mean values and finite fourth order moments;

The input u is independent of e and persistently exciting of order n;

All the roots of the equation

$$z^n+a_1z^{n-1}+\ldots+a_n=0 \quad (57)$$

have magnitudes less than one.

Then it is possible to show that the least squares estimate $\hat{\beta}$ given by (53) converges to the true parameter value as $N\to\infty$.

The special case of this theorem when $b_i=0$ for all i, which correspond to the identification of the parameters in an autoregression, was proven by MANN and WALD [36]. The extension to the case with $b_i\neq 0$ is given in ÅSTRÖM [48].

For large N the covariance of $\hat{\beta}$ is given by

$$\operatorname{cov}[\hat{\beta}]=\sigma^2[\Phi'\Phi]^{-1}$$

where σ^2 is the variance of $e(t)$.

Estimates of the variances of the parameter estimates are obtained from the diagonal elements of this matrix.

If it is also assumed that the residuals are gaussian we find that the least squares estimate can be interpreted as the maximum likelihood estimate, i.e. we obtain the loss function (48) in a natural way.

It has been shown that the estimate β is asymptotically normal with mean β and covariance $\sigma^2[\Phi'\Phi]^{-1}$. Notice, that this does *not* follow from the general properties of the maximum likelihood estimate since they are derived under the assumption of independent experiments.

In practice the order of the system is seldom known. It can also be demonstrated that serious errors can be obtained if a model of the wrong order is used. It is therefore important to have some methods available to determine the order of the model, i.e. we consider $\mathscr{S}$ as the class of linear models with arbitrary order.

To determine the order of the system we can fit least squares models of different orders and analyse the reduction of the loss function.

The loss function will of course always decrease when the model parameters are increased. To test if the reduction of the loss function is significant when the number of parameters are increased from n_1 to n_2 we can use the following test quantity.

$$t=\frac{V_1-V_2}{V_2}\,\frac{N-n_2}{n_2-n_1} \quad (58)$$

where V_i is the minimum value of the loss function for a model with n_i parameters ($i=1, 2$) and N is the number of input–output pairs. It can be shown that the random variable t for large N is asymptotically $F(n_2-n_1, N-n_2)$ distributed.

If follows from the properties of the F-distribution that $(n_2-n_1)\ t$ is also asymptotically χ^2 distributed.

When applying the least squares method to a single-input single-output system and the order of the model is increased from n to $n+1$ the number of parameters is increased by two. We have

$$F(2,\ 100)=3{\cdot}09\Rightarrow P\{t>3{\cdot}09\}=0{\cdot}05$$

$$F(2,\ \infty)=3{\cdot}00\Rightarrow P\{t>3{\cdot}00\}=0{\cdot}05.$$

Hence at a risk level of 5 per cent and $N>100$ the quantity t should be at least 3 for the corresponding reduction in loss function to be significant.

The idea to view the test of order as a decision problem has been discussed by ANDERSON [89]. It is also a standard tool in regression analysis. Other techniques for testing order are discussed in WOODSIDE [88].

Notice, that the least squares method also includes parametric time series analysis in the sense of fitting an autoregression. This has been discussed by WOLD [90] and WHITTLE [37]. Recent applications to EEG analysis have been given by GERSCH [91]. Using the probabilistic framework we can also give another interpretation of the least squares method in terms of the general definition of an identification problem given in section 3.

First observe that in the generalized error defined by (49) another y_m can be used:

$$e(k)=A^*(q^{-1})y(k)-B^*(q^{-1})u(k)$$
$$=y(k)-y_m(k). \tag{59}$$

Consequently:

$$y_m(k)=\hat{y}(k|k-1)=[1-A^*(q^{-1})]y(k)$$
$$+B^*(q^{-1})u(k)$$
$$=-a_1y(k-1)-\ldots-a_ny(k-n)$$
$$+b_1u(k-1)+\ldots+b_nu(k-n). \tag{60}$$

Notice, that $y_m(k)=\hat{y}(k|k-1)$ has a physical interpretation as the best linear mean squares predictor of $y(k)$ based on $y(k-1)$, $y(k-2)$, ... for the system (50). The generalized error (49) can thus be interpreted as the difference between the actual output at time k and its prediction using the model (60).

The least squares procedure can thus be interpreted as the problem of finding the parameters for the (prediction) model (60) in such a way that the criterion

$$V(y, y_m)=\sum_{k=1}^{N}[y(k)-y_m(k)]^2 \tag{61}$$

is as small as possible. Compare with the block diagram of Fig. 4. This interpretation is useful, because it can be extended to much more general cases. The interpretation can also be used in situations where there are no inputs, e.g. in time series analysis.

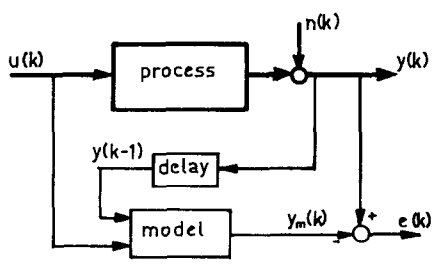

FIG. 4. Interpretation as predictor model.

Comparison with correlation methods

Now we will compare the least squares method with the correlation technique for determining the impulse response. When determining process dynamics for a single-input single-output system using correlation methods the following quantities are computed:

$$\left.\begin{aligned}
R_u(i)&=\frac{1}{N-i}\sum_{k=1}^{N-i}u(k)u(k+i)\\
R_y(i)&=\frac{1}{N-i}\sum_{k=1}^{N-i}y(k)y(k+i)\\
R_{yu}(i)&=\frac{1}{N-i}\sum_{k=1}^{N-i}y(k)u(k+i)\\
R_{uy}(i)&=\frac{1}{N-i}\sum_{k=1}^{N-i}u(k)y(k+i)
\end{aligned}\right\}. \tag{62}$$

Comparing with the least squares identification of process dynamics we find that the elements of the matrices $\Phi'\Phi$ and $\Phi'y$ of the least squares procedure given by (54) and (55) are essentially correlations or cross-correlations. Neglecting terms in the beginning and end of the series we find

$$\Phi'\Phi=N\begin{bmatrix}
R_y(0) & R_y(1) & \ldots & R_y(n-1) & -R_{yu}(0) & -R_{uy}(1) & \ldots & -R_{uy}(n-1)\\
 & R_y(0) & \ldots & R_y(n-2) & -R_{yu}(1) & -R_{yu}(0) & & -R_{uy}(n-2)\\
 & & & \vdots & \vdots & \vdots & & \vdots\\
 & & & R_y(0) & -R_{yu}(n-1) & -R_{yu}(n-2) & \ldots & -R_{yu}(0)\\
\hline
 & & & & R_u(0) & R_u(1) & \ldots & R_u(n-1)\\
 & & & & & R_u(0) & \ldots & R_u(n-2)\\
 & & & & & & & \vdots\\
 & & & & & & & R_u(0)
\end{bmatrix} \tag{63}$$

$$\Phi'y=N\begin{bmatrix}
-R_y(1)\\
-R_y(2)\\
\vdots\\
-R_y(n)\\
\hline
R_{uy}(1)\\
R_{uy}(2)\\
\vdots\\
R_{uy}(n)
\end{bmatrix}. \tag{64}$$

Hence if a correlation analysis is performed, it is a simple matter to calculate the least squares estimate by forming the matrices $\Phi'\Phi$ and $\Phi'y$ from the values of the sample covariance functions and solving the least squares equation. Since the order of the system is seldom known *a priori* it is often convenient to compute the least squares estimate recursively using the test of order we have described previously.

Correlated residuals

Many of the nice properties of the least squares method depend critically upon the assumption that the residuals $\{e(k)\}$ are uncorrelated. It is easy to find real-life examples where this assumption does not hold.

Example. Consider a noise-free first order system

$$x(k+1)+ax(k)=bu(k).$$

Assume that x is observed with independent measurement errors (additive noise), i.e.

$$y(k)=x(k)+n(k)$$

th n

$$y(k+1)+ay(k)=bu(k)+n(k+1)+an(k).$$

We thus get a system similar to (50) but with correlated residuals.

When the residuals are correlated the least squares estimate will be biased. Asymptotically the bias is given by

$$E(\hat{\beta}-b)=[E(\Phi'\Phi)]^{-1}E(\Phi'e) \qquad (65)$$

where $\hat{\beta}$ is the estimate and b is the true value of the parameter. The reason for this bias can be indicated as follows; c.f. Fig. 5.

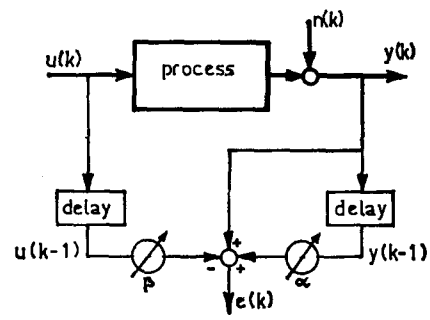

FIG. 5. Parameter estimation using a generalized model.

The estimate results from a minimization of the loss function

$$V=\sum_{k=1}^{N} e^2(k).$$

Necessary conditions are:

$$\frac{\partial V}{\partial \beta}=\sum_{k=1}^{N} e(k)u(k-1)=0 \qquad (66)$$

$$\frac{\partial V}{\partial \alpha}=\sum_{k=1}^{N} e(k)y(k-1)=0. \qquad (67)$$

In (67) the additive noise is present both in e and y; this leads to a term

$$\sum_{k=1}^{N} n^2(k-1)$$

which causes a bias.

To see how this works out consider an example.

Example. Assume, that the process is actually described by the model

$$y(k+1)-0{\cdot}5y(k)=1{\cdot}0u(k)+n(k+1)+0{\cdot}1n(k)$$

where $\{n(k)\}$ is a sequence of independent normal (0, 1) random variables, but that the system is identified using the least squares method under the assumption that the residuals are uncorrelated. Below we give a typical result obtained from 500 pairs of inputs and outputs

Process parameters	Estimates
$a=-0{\cdot}5$	$\hat{\alpha}=-0{\cdot}643\pm0{\cdot}029$
$b=\ 1{\cdot}0$	$\hat{\beta}=\ 1{\cdot}018\pm0{\cdot}062$

We are apparently in a very bad situation; not only is the estimate $\hat{\alpha}$ wrong but we have also a great deal of confidence in the wrong result. Note, that the true value 0·500 deviates from the estimate with about 5σ.

The correlation of the residuals can thus easily lead to wrong conclusions. Several techniques have been suggested to deal with correlated residuals, viz:

(a) repeated least squares
(b) generalized least squares
(c) the maximum likelihood method
(d) instrumental variables
(e) Levin's method
(f) tally principle.

(a) *Repeated least squares.* Suppose that we did not know the order of the system discussed in the previous example. It would then be natural to continue the least squares procedure and to test the order of the system. The results for the particular example are shown in Table 1. Assuring a risk level of 5 per cent and using the test quantity of

(58) we find from Table 1 that the increase of the

TABLE 1

n	$\alpha \pm \sigma(\alpha)$	$\beta \pm \sigma(\beta)$	V	t
1	$-0{\cdot}643 \pm 0{\cdot}029$	$1{\cdot}018 \pm 0{\cdot}062$	592·65	
2	$-1{\cdot}015 \pm 0{\cdot}045$ $0{\cdot}377 \pm 0{\cdot}039$	$1{\cdot}086 \pm 0{\cdot}056$ $-0{\cdot}520 \pm 0{\cdot}072$	469·64	50·94
3	$-1{\cdot}118 \pm 0{\cdot}050$ $0{\cdot}624 \pm 0{\cdot}068$ $-0{\cdot}178 \pm 0{\cdot}043$	$1{\cdot}115 \pm 0{\cdot}055$ $-0{\cdot}660 \pm 0{\cdot}078$ $0{\cdot}263 \pm 0{\cdot}076$	447·25	9·67
4	$-1{\cdot}157 \pm 0{\cdot}050$ $0{\cdot}756 \pm 0{\cdot}074$ $-0{\cdot}412 \pm 0{\cdot}074$ $0{\cdot}187 \pm 0{\cdot}044$	$1{\cdot}085 \pm 0{\cdot}055$ $-0{\cdot}733 \pm 0{\cdot}078$ $0{\cdot}409 \pm 0{\cdot}083$ $-0{\cdot}146 \pm 0{\cdot}076$	426·40	9·43
5	$-1{\cdot}185 \pm 0{\cdot}051$ $0{\cdot}814 \pm 0{\cdot}077$ $-0{\cdot}518 \pm 0{\cdot}083$ $0{\cdot}349 \pm 0{\cdot}076$ $-0{\cdot}117 \pm 0{\cdot}044$	$1{\cdot}080 \pm 0{\cdot}054$ $-0{\cdot}745 \pm 0{\cdot}078$ $0{\cdot}475 \pm 0{\cdot}086$ $-0{\cdot}252 \pm 0{\cdot}086$ $0{\cdot}123 \pm 0{\cdot}076$	418·72	3·51
6	$-1{\cdot}195 \pm 0{\cdot}051$ $0{\cdot}339 \pm 0{\cdot}079$ $-0{\cdot}555 \pm 0{\cdot}088$ $0{\cdot}410 \pm 0{\cdot}088$ $-0{\cdot}208 \pm 0{\cdot}079$ $0{\cdot}061 \pm 0{\cdot}045$	$1{\cdot}079 \pm 0{\cdot}055$ $-0{\cdot}751 \pm 0{\cdot}078$ $0{\cdot}487 \pm 0{\cdot}087$ $-0{\cdot}290 \pm 0{\cdot}090$ $0{\cdot}183 \pm 0{\cdot}087$ $-0{\cdot}080 \pm 0{\cdot}076$	416·56	0·99
7			414·62	0·89

model order from 5 to 4 gives a significant reduction in the loss function ($t=3{\cdot}51$) but that the reduction in the loss function when the order is increased from 5 to 6 is not significant ($t=0{\cdot}99$). The order test will thus indicate that a fifth order model is appropriate.

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

where

$$A^*(q^{-1})=1-1{\cdot}19q^{-1}+0{\cdot}81q^{-2}-0{\cdot}52q^{-3}+0{\cdot}35q^{-4}-0{\cdot}12q^{-5}$$

$$B^*(q^{-1})=1{\cdot}08q^{-1}-0{\cdot}75q^{-2}+0{\cdot}48q^{-3}-0{\cdot}25q^{-4}+0{\cdot}12q^{-5}$$

Dividing A by B we find

$$\frac{1{\cdot}08A(z)}{B(z)}=z-0{\cdot}495+\frac{0{\cdot}03z^3-0{\cdot}072^2+0{\cdot}12z-0{\cdot}06}{z^4-0{\cdot}69z^3+0{\cdot}44z^2-0{\cdot}23z+0{\cdot}11}.$$

Taking the uncertainties of the coefficients α and β into account it does not seem unreasonable to neglect the rest in the above expression. We find that the system under test can be represented by

$$y(k)=\frac{1{\cdot}08}{q-0{\cdot}5}u(k)+\frac{\lambda}{A^*(q^{-1})}e(k).$$

We can thus conclude that if we choose $\mathscr{S}$ not as the class of linear first order systems but as the class of linear systems of arbitrary order it is at least possible to overcome the difficulty of correlated residuals in the specific example. This idea was mentioned briefly in ÅSTRÖM [92]. It has, however, not been pursued in general.

(b) *Generalized least squares.* Another way to overcome the difficulty with correlated residuals is to use the method of generalized least squares. See e.g. CLARKE [93].

The basic idea is as follows. Let the process be governed by

$$A^*(q^{-1})y(k)=B^*(q^{-1})u(k)+v(k) \tag{68}$$

where A^* and B^* are polynomials and $\{v(k)\}$ a sequence of correlated random variables. Suppose that the correlations of the residuals are known. Say that $\{v(k)\}$ can be represented as

$$v(k)=G^*(q^{-1})e(k) \tag{69}$$

where $\{e(k)\}$ is a sequence of uncorrelated random variables and G a pulse transfer function. The equation describing the process can then be written as

$$A^*(q^{-1})y(k)=B^*(q^{-1})u(k)+G^*(q^{-1})e(k) \tag{70}$$

or

$$A^*(q^{-1})\tilde{y}(k)=B^*(q^{-1})\tilde{u}(k)+e(k) \tag{71}$$

where

$$\tilde{y}(k)=\frac{1}{G^*(q^{-1})}y(k) \tag{72}$$

$$\tilde{u}(k)=\frac{1}{G^*(q^{-1})}u(k). \tag{73}$$

Hence if the signals $\tilde{u}$ and $\tilde{y}$ are considered as the inputs and outputs we have an ordinary least squares problem. Compare with (49). We thus find that the generalized least squares can be interpreted as a least squares identification problem where the criterion is chosen as (48) with the generalized error defined as

$$e(k)=\frac{A^*(q^{-1})}{G^*(q^{-1})}y(k)-\frac{B^*(q^{-1})}{G^*(q^{-1})}u(k)$$

$$=A^*(q^{-1})\left[\frac{1}{G^*(q^{-1})}y(k)\right]-B^*(q^{-1})\left[\frac{1}{G^*(q^{-1})}u(k)\right]. \tag{74}$$

Compare with the block diagram of Fig. 6. This shows how the generalized error can be obtained from the process inputs and outputs and the model parameters α and β in the generalized least squares method.

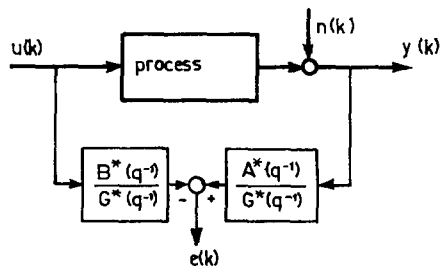

FIG. 6. Parameter estimation using "whitening filters" G^*.

The correlation of the residuals and the pulse transfer function G are seldom known in practice. CLARKE [93] has proposed an iterative procedure to determine G which has been tested on simulated data as well as on practical measurements (distillation column identification).

The procedure consists of the following steps:

(1) Make an ordinary least squares fit of the model

$$A_j^*(q^{-1})y(k)=B_j^*(q^{-1})u(k)+v(k). \quad (75)$$

(2) Analyze the residuals v and fit an autoregression i.e.

$$D_j^*(q^{-1})v(k)=e(k) \quad (76)$$

where $\{e(k)\}$ is discrete-time white noise.

(3) Filter the process inputs and outputs through

$$\tilde{y}(k)=D_j^*(q^{-1})y(k);$$
$$\tilde{u}(k)=D_j^*(q^{-1})u(k). \quad (77)$$

(4) Make a new least squares fit to the filtered inputs and outputs and repeat from 2.

Another approach along these lines is the algorithm of STEIGLITZ and MCBRIDE [94]. They use at the jth iteration:

$$D^*(q^{-1})=A_{j-1}^*(q^{-1})$$

thus making a least squares fit of the model

$$e(k)=\frac{A^*(q^{-1})}{A_{j-1}^*(q^{-1})}y(k)-\frac{B_j^*(q^{-1})}{A_{j-1}^*(q^{-1})}u(k)$$
$$=\left[y(k)-\frac{B_j^*(q^{-1})}{A_j^*(q^{-1})}u(k)\right]\frac{A_j^*(q^{-1})}{A_{j-1}^*(q^{-1})}. \quad (78)$$

The non-linearity in the parameters can also be handled by means of quasilinearization, c.f. SCHULZ [218]. The techniques proposed in [111] and [217] can also be interpreted as generalized least squares methods.

These procedures have the drawbacks that there are no systematic rules for the choice of order of the model (75) and of the autoregression (76). Neither are any convergence proofs yet available. It has, however, been shown to work very well with reasonable *ad hoc* choices of order in specific examples.

The following observation might also be worthwhile. Assume, that the generalized least squares procedure will converge. Say $A_j\to A$, $B_j\to B$, and $D_j\to D$. We will then obtain

$$A^*(q^{-1})D^*(q^{-1})y(k)$$
$$=B^*(q^{-1})D^*(q^{-1})u(k)+e(k) \quad (79)$$

i.e. a description of the process with uncorrelated residuals. It thus appears that the differences between the repeated least squares and the generalized least squares are small.

(c) *The maximum likelihood method.* Another way to deal with the problem of correlated residuals is to postulate a system with correlated residuals, e.g. a canonical representation of an n:th order system with one input and one output

$$A^*(q^{-1})y(k)=B^*(q^{-1})u(k)+\lambda C^*(q^{-1})e(k) \quad (80)$$

where u is the input, y the output and $\{e(k-\}$ a sequence of independent normal (0, 1) random variables. Compare with section 4. The parameters of (80) can be determined using the method of maximum likelihood. This has been discussed in [74], [23], [97], [78], [199].

The likelihood function L is given by

$$-\log L(\theta,\lambda)=\frac{1}{2\lambda^2}\sum_{k=1}^{N}\varepsilon^2(k)$$
$$+\frac{N}{2}\log\lambda+\frac{N}{2}\log 2\pi \quad (81)$$

where

$$C^*(q^{-1})\varepsilon(k)=A^*(q^{-1})y(k)-B^*(q^{-1})u(k) \quad (82)$$

and $\{u(k),\ k=1, 2, \ldots, N\}$ is the applied input signal and $\{y(k),\ k=1, 2, \ldots, N\}$ is the observed output signal. The likelihood function is considered as a function of θ and λ, where θ is a vector whose components are the parameters $a_1, a_2, \ldots, a_n$, $b_1, b_2, \ldots, b_n$, $c_1, c_2, \ldots, c_n$ and the n initial conditions of (82). Notice, that the logarithm of the likelihood function is linear in the parameters a_i and b_i but strongly non-linear in c_i.

Also notice, that the optimization of L with respect to θ and λ can be performed separately in the following way: First determine θ such that the loss function

$$V(\theta)=\sum_{k=1}^{N}\varepsilon^2(k) \tag{83}$$

is minimal with respect to θ. The optimization with respect to λ can then be performed analytically. We get

$$\hat{\lambda}^2=\frac{1}{N}\min_{\theta}V(\theta). \tag{84}$$

The maximum likelihood estimate can be shown to have nice asymptotic properties. Estimates of the accuracy of the parameters can also be provided. See ÅSTRÖM, BOHLIN and WENSMARK [74], ÅSTRÖM and BOHLIN [23]. A test for possible violation of prerequisites of ML estimation is given by BOHLIN [99].

The maximum likelihood procedure can also be interpreted as finding the coefficients of the prediction model

$$y_m(k)=\hat{y}(k|k-1)=\frac{B^*(q^{-1})}{C^*(q^{-1})}u(k) - \frac{A^*(q^{-1})-C^*(q^{-1})}{C^*(q^{-1})}y(k) \tag{85}$$

in such a way that the criterion

$$V=V(y,\,y_m)=\sum_{k=1}^{N}[y(k) - y_m(k)]^2=\sum_{k=1}^{N}\varepsilon^2(k) \tag{86}$$

is as small as possible.

Notice, that (80) can also be written as

$$A^*(q^{-1})\left[\frac{1}{C^*(q^{-1})}y(k)\right] =B^*(q^{-1})\left[\frac{1}{C^*(q^{-1})}u(k)\right]+\lambda e(k). \tag{87}$$

This means that the maximum likelihood method can also be interpreted as a generalized least squares method where the filter function $G=1/C$ is determined automatically.

It has been shown that predictors and minimal variance control algorithms are easily determined from the model (80); c.f. ÅSTRÖM [72], [92]. The maximum likelihood method has been applied extensively to industrial measurements. See e.g. ÅSTRÖM [92], [199], and GUSTAVSSON [100] in this paper comparisons with other techniques such as correlation methods and generalized least squares are also given. The maximum likelihood method has also been applied to time series analysis (put $B=0$). The maximum likelihood estimate is a strongly non-linear function of the parameters. Since time series analysis is mostly concerned with quadratic functions, such as covariances and spectral densities, one might expect that the estimates can be expressed as non-linear functions of the sample covariances. Estimates of this nature which are asymptotically equivalent to the maximum likelihood estimates for parametric time series analysis have been given by ZETTERBERG [101]. The generalization of the maximum likelihood method to the multivariable case has been done by WOO [97] and CAINES [222].

(d) *Instrumental variables.* Repeated least squares, generalized least squares and the maximum likelihood method all give a model of the environment in terms of a model for the disturbances as a filter driven by white noise. If we are only interested in the system dynamics there are other methods to avoid the difficulties with correlated residuals, e.g. the instrumental variable method.

The equation for the least squares estimate can be obtained from the equation

$$y=\Phi\beta+e \tag{88}$$

by premultiplying with Φ', neglecting the term $\Phi'e$ and solving the equation

$$\Phi'y=\Phi'\Phi\hat{\beta}. \tag{89}$$

The estimate $\hat{\beta}$ will be unbiased if the term $\Phi'e$ has zero mean. When the residuals are correlated it is easy to show that $E\Phi'e\neq 0$.

In the instrumental variable method, [102], [103], the equation (88) is multiplied with W' where W, called the instrumental matrix, is a matrix whose elements are functions of the data with the properties

$$EW'\Phi \text{ nonsingular} \tag{90}$$

$$EW'e=0. \tag{91}$$

The parameter estimate obtained from

$$W'y=W'\Phi\hat{\beta} \tag{92}$$

will then be unbiased. It is also possible to find instrumental variables such that the estimate has optimal properties. The schemes proposed by PETERKA and ŠMUK [104], HSIA and LANDGREBE [105] are closely related to the instrumental variable

technique. VUŠKOVIĆ, BINGULAC and DJOROVIĆ [106] indicate the use of pseudo sensitivity functions as instrumental variables.

(e) *Levin's method.* A particular method for estimating the bias due to correlated residuals has been proposed by LEVIN [107], [210], [211] for the case of a deterministic system with independent observation errors. Levin's results are based on a technique due to Koopmans and it gives the estimates in terms of an eigenvalue problem for the matrix $\Phi'\Phi$. A careful analysis of Levin's method which includes convergence proofs and error estimates has been done by AOKI and YUE [108]. The method has also been used by SMITH [109].

(f) *Tally principle.* This estimation procedure as described by PETERKA and HALOUSKOVÁ [110] starts from the canonical fo m of an nth order process given by equation (37). Now the terms with ε are combined to a random variable $\delta(k)$:

$$\delta(k)=\varepsilon(k)+\sum_{i=1}^{n} c_i\varepsilon(k-i).$$

Based on the knowledge that $\{\varepsilon(k)\}$ is a sequence of uncorrelated random variables it is found that

$$E[\delta(k)y(k-i)]=0 \text{ for } i>n$$
$$E[\delta(k)u(k-i)]=0 \text{ for } i>0\,. \qquad (93)$$

For any estimate of the unknown parameters one finds a sequence $\delta^*(k)$. Now

$$\frac{1}{N}\sum_{k=1}^{N}\delta^*(k)u(k-i) \qquad (94)$$

and

$$\frac{1}{N}\sum_{k=1}^{N}\delta^*(k)y(k-i) \qquad (94')$$

have to "tally" (duplicate) the corresponding expectations (93) as well as possible in a least square sense.

In the special case of time series analysis the "tally" principle reduces to the well-known technique of fitting the parameters of the Yule–Walker equation to the sample covariances. See e.g. Ref. [224].

Multivariable systems

The essential difficulty in the identification of multivariable systems is to find a suitable representation of the system. Once a particular representation is chosen it is a fairly straight-forward procedure to construct identification methods analogous to those given for the case of systems with one input and one output. We refer to section 4 for a discussion of structures for multivariable systems. Also c.f. GRAUPE, SWANICK and CASSIR [112]. For the structure (35) the likelihood function is given by

$$-\log L(\theta, R)=\frac{N}{2}\log\det R + \frac{1}{2}\sum_{k=1}^{N}\varepsilon'(k)R^{-1}\varepsilon(k)+\frac{nN}{2}\log 2\pi\,. \qquad (95)$$

Even for multivariable systems the maximization of $L(\theta, R)$ can be performed separately with respect to θ and R. It was shown by EATON [75] that the maximum of $L(\theta, R)$ is obtained by finding θ which minimizes

$$V(\theta)=\det\left[\sum_{k=1}^{N}\varepsilon(k)\varepsilon'(k)\right]. \qquad (96)$$

The maximization with respect to R can then be done analytically to yield

$$\hat{R}=\frac{1}{N}\sum_{k=1}^{N}\varepsilon(k)\varepsilon'(k)\,. \qquad (97)$$

This fact is also mentioned in ROWE [67]. The identification problem for linear multivariable systems is also closely related to the problem of simplifying a large linear system. See e.g. DAVISON [70].

6. IDENTIFICATION OF NON-LINEAR SYSTEMS

Representation of non-linear systems

Note again, that non-linearity does not necessarily imply a non-linearity in the parameters too (c.f. section 4).

For linear systems the impulse response offers a non-parametric system description that does not depend on specific *a priori* assumptions. For a wide class of non-linear systems a Volterra series expansion offers analogous possibilities, using impulse responses of increasing dimensionality. Approximation of these functions by a finite number of points leads to a model that is linear in the parameters; c.f. EYKHOFF [113], ALPER [114], ROY and SHERMAN [115]. For many practical cases the number of parameters needed for this description is too large.

Another approach for non-linear model building, called group method of data handling, is given by IVAKHNENKO [116], [117].

When considering non-linear systems as well as linear systems with multiple inputs and multiple outputs it therefore is necessary to make specific assumptions concerning the model structure. It is usually assumed that the system equations are known except for a number of parameters b.

In a typical case the identification problem can then be formulated as follows:
Let the class of models be all systems described by the state equation

$$\frac{dx}{dt}=f(x, u, \beta, t)$$
$$y_m=g(x, u, \beta, t) \quad (98)$$

where the parameter vector β belongs to a given set. Let the criterion be given by the loss function

$$V(y, y_m)=V(\beta)=\int_0^T [y(k)-y_m(k, \beta)]^2 dt \quad (99)$$

where y is the process output and y_m the model output.

Estimation for a parametric model

The estimation problem for a non-linear parametric model thus reduces to a non-linear optimization problem. As was discussed in section 3—Computational aspects, there are many techniques available for the solution of the non-linear estimation problem.

Techniques which exploit the particular structure of the problem is discussed by MARQUARDT [120].

The particular case when the non-linear system can be separated into linear dynamics and non-linearities without memory are discussed by BUTLER and BOHN [118] and NARENDRA and GALLMAN [119]. BELLMAN and KALABA [121] have used quasilinearization to solve the non-linear optimization problem.

Interesting applications of this technique are found in BUEL, KAGIWADA and KALABA [122], BUEL and KALABA [123]. A fairly general computer program to solve the problem has been written by BUEL. Also c.f. VANĚČEK and FESSL [124].

Another method to solve the non-linear optimization problem has been given by TAYLOR, ILIFF and POWERS [125] in connection with application to inflight determination of stability derivatives.

Again the criterion (99) can be given a probabilistic interpretation if it is assumed that the only disturbances are white noise measurement errors. A technique which admits the measurement errors to be a stationary process with unknown rational spectral density has been proposed by ÅSTRÖM, BOHLIN and WENSMARK [74].

Due to specific assumptions that are made concerning the structure of (98) one might expect that serious mistakes can be made if these assumptions are not true. Results which prove or disprove this are not known.

Rather few publications have appeared on the use of Bayes' method (c.f. Appendix A) in identification techniques; MASLOV [126], GALTIERI [127]. This is probably due to the computational problems when evaluating the conditional expectations. MCGEE and WALFORD [128] propose a solution by using a Monte Carlo approach.

7. ONLINE AND REAL-TIME IDENTIFICATION

In many applications it is highly desirable to obtain the result of the identification recursively as the process develops. For example it might be of interest to proceed with the experiment until a specified parameter accuracy is achieved. The problem is then as follows. Assume, that an estimate β_N is obtained based on N pairs of input–output samples. Is it necessary to repeat the whole identification procedure from the beginning, using the whole string of input–output data in order to obtain β_{N+1} or is it possible to arrange the computations recursively? An identification scheme which is recursive and which does not require that the whole string of input–output data is brought in at each step is called an *on-line* method.

On-line identification can thus be looked upon as a convenient way of arranging the computations. Apart from being of practical interest this point of view on identification problems will also make it possible to establish connections with other fields, e.g. non-linear filtering, stochastic approximation, learning and adaption.

If the parameters of the process are truly time varying it is, of course, meaningless to do anything else but track the parameters in real-time. This is called *real-time identification*. One may recognize two computational procedures: an *accumulative* solution, open loop with respect to the parameter estimate, and a *recursive* solution, closed loop with respect to the parameter estimate; c.f. GENIN [129].

There are many different ways to obtain algorithms for real-time identification. Practically all methods, however, will yield algorithms with the structure

$$\beta(N+1)=\beta(N)+\Gamma(N)e(N) \quad (100)$$

where $e(N)$ is the generalized error discussed earlier and $\Gamma(N)$ is a gain factor which can be of varying complexity.

Model reference techniques

The on-line identification problem is sometimes formulated as a model tracking problem. A simple case is illustrated in Fig. 7. Note, that this is an example of a recursive, closed loop, approach.

The input is simultaneously fed to the process and to a model with adjustable parameters. The adjustable parameters are changed by an adjustment mechanism which receives the process output y and the model output y_m as inputs. This formulation of the on-line identification problem was first

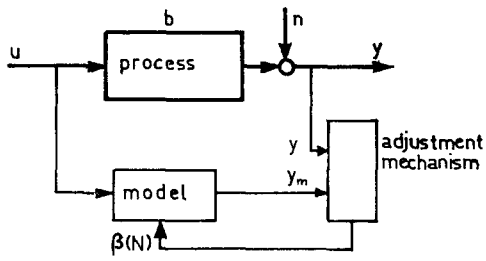

FIG. 7. Model adjustment technique.

considered by WHITAKER [130]. The essential problem is to determine the adjustment mechanism such that the model parameters in some sense will be close to process parameters. There are many ways to obtain suitable adjustment mechanisms. One way is to introduce a criterion in terms of a loss function as was done in section 2 and to change parameters in such a way that the loss function decreases. Since there are so many different ways to define both the generalized error and the loss function we will not enumerate all cases but just illustrate the principle. For example, if the loss function is based on $e(N)=y(N)-y_m(N)$ and a least squares structure is used, it follows that the optimal choice of gain $\Gamma(N)$ of the adjustment mechanism (100) is proportional to

$$u(N)\Big/\sum_{k=1}^{N}u^2(k).$$

This was e.g. shown by KACZMARZ [131] and has later been exploited by several authors, e.g. NAGUMO and NODA [132], BÉLANGER [133], RICHALET [134].

The effect of working with a quantized signal sign u is discussed in CRUM and WU [135]. In simple cases the adjustment mechanism makes the rate of change of the parameter adjustment proportional to the sensitivity derivatives. See e.g. MEISSINGER and BEKEY [136]. A recent application of this idea is given by ROSE and LANCE [137]. The requirement that the closed loop is stable is a necessary design criterion. Since the system consisting of the adjustable model, the process and the adjustment mechanism is highly non-linear the stability problem is not trivial. Using Liapunov methods, LION [138], SHACKCLOTH and BUTCHART [139], PARKS [140], PAZDERA and POTTINGER [142] have designed stable systems. Lion's results have recently been generalized to stochastic systems by KUSHNER [143]. The powerful stability tests developed by POPOV [144] and ZAMES [145] have given new tools to design adjustment mechanisms which will result in stable systems. Initial efforts in this direction have been done by LANDAU [146] who has proposed stable model reference systems using the Popov criterion.

On-line least squares

The conversion of any identification method to an on-line technique consists of showing that the estimate satisfies a recursive equation. This is easily done for the least squares method. Consider the least squares model of section 5, i.e.

$$y(k)+a_1y(k-1)+\ldots+a_ny(k-n)=b_1u(k-1)+\ldots+b_nu(k-n). \quad (101)$$

Define

$$\beta'=[a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n] \quad (102)$$

and

$$\varphi(N+1)=[-y(N), -y(N-1), \ldots, -y(N-n+1), u(N), u(N-1), \ldots, u(N-n+1)]. \quad (103)$$

The least squares estimate is then given by (53). It can be shown by simple algebraical manipulations that the least squares estimate satisfies the recursive equation

$$\beta(N+1)=\beta(N)+\Gamma(N)[y(N+1)-\varphi(N+1)\beta(N)] \quad (104)$$

where $\beta(N)$ denotes the least squares estimate based on N pairs of input–output data and

$$\Gamma(N)=P(N)\varphi'(N+1)[\alpha+\varphi(N+1)P(N)\varphi'(N+1)]^{-1} \quad (105)$$

$$P(N+1)=P(N)-P(N)\varphi'(N+1)[\alpha+\varphi(N+1)P(N)\varphi(N+1)]^{-1}\cdot\varphi(N+1)P(N)=P(N)-\Gamma(N)\varphi(N+1)P(N)=[I-\Gamma(N)\varphi(N+1)]P(N) \quad (106)$$

$$P(N_0)=\alpha[\Phi'(N_0)\Phi(N_0)]^{-1} \quad (107)$$

where N_0 is a number such that $\Phi'(N_0)\Phi(N_0)$ is positive definite.

The recursive equation (104) has a strong intuitive appeal. The next estimate $\beta(N+1)$ is formed by adding a correction to the previous estimate. The correction is proportional to $y(N+1)-\varphi(N+1)\beta(N)$. The term $\varphi\beta$ would be the value of y at time $N+1$ if the model were perfect and there were no disturbances. The correction term is thus proportional to the difference between the measured value of $y(N+1)$ and the prediction of $y(N+1)$ based on the previous model parameters. The components of the vector $\Gamma(N)$ are weighting factors which tell how the corrections and the previous estimate should be weighted.

Notice, that in order to obtain the recursive equations it is necessary to introduce the auxiliary quantity P. The state of the systems (104), (105) and (106) governing the on-line estimator is thus the vector β, which is the current estimate, and the symmetric matrix P. The pair (β, P) thus represents the smallest number of variables, characterizing the input–output data, which are necessary to carry along in the computations.

If the model of the system is actually given by (50) where $\{e(k)\}$ are independent residuals with variance σ^2, the matrix P can be interpreted as the covariance matrix of the estimate if α is chosen as σ^2. A special discussion is needed about the start of the iterative procedure. That can be done by:

using possible *a priori* knowledge about β and P;

by using a "one-shot" least squares estimation procedure using the first series of observations;

by starting with $P(0)=cI$ c.f. KLINGER [147].

Notice, that since $\beta(N)$ given by (104) is the least squares estimate the convergence of the equations (104)–(106) follows directly from the consistency proofs of the least squares estimate.

Recursive versions of the generalized least squares procedure have been derived by YOUNG [148]. Recursive versions of an instrumental variable method has been derived by PETERKA and ŠMUK [104]. An approximative on-line version of the maximum likelihood method has been proposed by PANUSKA [96].

A discussion of on-line methods is also given in LEATHRUM [149], STANKOVIĆ, VELAŠEVIĆ and ČARAPIĆ [150].

Contraction mappings

A technique of constructing recursive algorithms have been suggested by OZA and JURY [151], [152]. We will explain the technique in connection with the least squares problem. Instead of solving the equation

$$\Phi'(N)\Phi(N)\beta(N)=\Phi'(N)y \tag{108}$$

for each N and showing that $\beta(N)$ satisfies a recursive equation, Oza and Jury introduce the mapping

$$T_N(\beta)=\beta-\gamma[\Phi'(N)\Phi(N)\beta-\Phi'(N)y] \tag{109}$$

where γ is a scalar. It is then shown that the sequence

$$\beta(N+1)=T_N(\beta(N)) \tag{110}$$

under suitable conditions converges to the true parameters as $N\to\infty$. When applied to the ordinary least squares problems the algorithm (109) is not efficient in contrast with the recursive least squares method. To obtain an efficient algorithm it is necessary to make γ a matrix.

With the choice

$$\gamma=[\Phi'(N)\Phi(N)]^{-1} \tag{111}$$

the algorithm becomes equivalent to the least squares.

The method of Oza and Jury can be applied to more general cases than the least squares. It was actually proven for the case when there are errors in the measurements of both inputs and outputs, provided the covariance function of the measurement errors are known. The assumption of known covariances of the measurement errors severely limits the practical applicability of the method.

Stochastic approximations

The formula for the recursive least squares is

$$\beta(k+1)=\beta(k)+\Gamma(k)[y(k+1)-\varphi(k+1)\beta(k)] \tag{112}$$

where Γ was chosen by the specific formula (105). It can be shown that there are many other choices of Γ for which the estimate β will converge to the true parameter value b. Using the theory of stochastic approximations it can be shown that the choice

$$\Gamma(k)=\frac{1}{k}A\varphi'(k+1) \tag{113}$$

will ensure convergence if A is positive definite. See e.g. ALBERT and GARDNER [153]. A particularly simple choice is e.g. $A=I$. The algorithms obtained by such choices of Γ will in general give estimates with variances that are larger than the variance of the least squares estimate. The algorithms are, however, of interest because they make it possible to reduce computations, at the price of a larger variance. Using stochastic approximations it is also possible to obtain recursive algorithms in cases where the exact on-line estimate is either very complicated or very difficult to derive. There are excellent surveys available on stochastic approximations. See e.g. ALBERT and GARDNER [153] and TSYPKIN [154]. Recent applications are given by SAKRISON [155], SARIDIS and STEIN [156], [157], HOLMES [158], ELLIOTT and SWORDER [159], NEAL and BEKEY [160].

Real time-identification

The recursive version of the least squares method is closely related to the Kalman filtering theory. Kalman considers a dynamical system

$$\begin{aligned} x(k+1)&=\Phi x(k)+e(k)\\ y(k)&=Cx(k)+v(k) \end{aligned} \tag{114}$$

where $\{e(k), k=1, 2, \ldots\}$ and $\{v(k), k=1, 2, \ldots\}$ are sequences of independent equally distributed random vectors with zero mean values and covariance matrices R_1 and R_2 respectively. Kalman has proven the following theorem.

Theorem (Kalman). Let the initial condition of (114) be a normal random variable (m, R_0). The best estimate of $x(k)$, in the sense of least squares, given the observed outputs $y(1), y(2), \ldots, y(k)$ is given by the recursive equations

$$\hat{x}(k)=\Phi\hat{x}(k-1)+\Gamma(k)[y(k)-C\Phi\hat{x}(k-1)]$$
$$\hat{x}(0)=m \tag{115}$$

where

$$\Gamma(k)=S(k)C'[CS(k)C'+R_2]^{-1}$$
$$S(k)=\Phi P(k-1)\Phi'+R_1$$
$$P(k)=S(k)-\Gamma(k)CS(k)$$
$$S(0)=R_0. \tag{116}$$

The matrix $S(k)$ has a physical interpretation as the covariance matrix of the *a priori* estimate of $x(k)$ given $y(1), \ldots, y(k-1)$ and the matrix $P(k)$ as the covariance of the posterior estimate of $x(k)$ given $y(1), \ldots, y(k)$.

Now consider the least squares identification of the system

$$y(k)+a_1y(k-1)+\ldots+a_ny(k-n)$$
$$=b_1u(k-1)+\ldots+b_nu(k-n)+e(k) \tag{117}$$

where $\{e(k)\}$ is a sequence of normal $(0, \lambda)$ random variables.

Introduce the coefficients of the model as state variables

$$x_1(k)=a_1$$
$$x_2(k)=a_2$$
$$\vdots$$
$$x_n(k)=a_n$$
$$x_{n+1}(k)=b_1$$
$$x_{n+2}(k)=b_2$$
$$\vdots$$
$$x_{2n}(k)=b_n \tag{118}$$

and define the following vector

$$C(k)=[-y(k-1), \ldots, -y(k-n), u(k-1), \ldots, u(k-n)]. \tag{119}$$

Since the coefficients are constant we have

$$x(k+1)=x(k). \tag{120}$$

The equation (117) can now be written as

$$y(k)=C(k)x(k)+e(k) \tag{121}$$

and the least squares identification problem can be stated as a Kalman filtering problem with $\Phi=I$, $R_1=0$, $R_2=\lambda^2$.

The recursive equations of the least squares estimate can thus be obtained directly from Kalman's theorem. This has an interesting consequence because it turns out that if the parameters a_i are not constants but gauss-markov processes, i.e.

$$a_i(k+1)=\alpha a_i(k)+v_i(k) \tag{122}$$

the Kalman theorem can still be applied. (This requires a slight generalization of Kalman's proof since the parameters of the C vector are stochastic processes.)

BOHLIN [161] has extended the argument to processes of the structure (117) with coefficients which are processes with rational spectral densities.

It thus is possible to obtain parameter estimators for linear models with time varying parameters. It is in fact not necessary to assume a first order process but the parameters a_i can be chosen to be stationary processes with rational spectral densities.

In this way it is possible to obtain identifiers for processes with rapidly varying parameters. This has been discussed by BOHLIN [77] and WIESLANDER [163]. The method proposed by SEGERSTÅHL [164] can be considered as the special case $v_i=0$. Notice, that with this approach it is necessary to know the covariances of the processes $\{v_i\}$ characterizing the variations in the model parameters. Such assumptions will, of course, limit the practical applicability of the methods. One way to overcome this difficulty is to use the approximative techniques for estimating the covariances in a Kalman filter proposed by MEHRA [73] and SAGE and HUSA [165]. Techniques for validating the assumptions of the Kalman filtering problem have been proposed by BERKOVEC [166]. Recursive estimation of the transition matrix is discussed by PEARSON [167]. An analog implementation is given by HSIA and VIMOLVANICH [168].

Non-linear filtering

The relationship between the recursive least squares and the Kalman filtering theory was obtained by introducing the parameters of the identification problem as state variables. We thus find that there are in principle no differences between parameter estimation and state estimation. A

parameter estimation problem can be extended to a state estimation problem by introducing the parameters as auxiliary state variables. A constant parameter b corresponds to the state equation

$$\frac{db}{dt}=0 \tag{123}$$

for continuous time systems and

$$b(k+1)=b(k) \tag{124}$$

for discrete time systems. The (state) estimation problem obtained in this way will, however, in general be a non-linear problem since the parameters frequently occur in terms like $bx(t)$ in the original problem. In the general continuous time case we are thus faced with a filtering problem for the model

$$\begin{aligned} dx &= f(x, t)dt + \sigma(t, x)dv \\ dy &= g(x, t)dt + \mu(t, x)de \end{aligned} \tag{125}$$

where $\{v(t)\}$ and $\{e(t)\}$ are Wiener processes with incremental covariances $I_v dt$ and $I_e dt$. Some of the components of x are state variables and others are parameters of the identification problem. The non-linear filtering problem is completely solved if the conditional probability of $x(t)$ given $\{y(s), t_0 \leqslant s \leqslant t\}$ can be computed. The maximum likelihood estimate is e.g. obtained by finding the value of x for which the conditional density has its maximum. The least squares estimate is given by the conditional mean etc. Compare the resumé of estimation theory in Appendix A.

It is in principle easy to obtain a recursive formula for the conditional probability distribution simply by applying Bayes' rule. For the model (125) the problem is, however, technically difficult and great care must be given to the appropriate interpretation of the equation (125). It has been shown by KUSHNER [172] and STRATONOVICH [223] that under suitable regularity conditions the conditional probability density of $x(t)$ given $\{y(s), t_0 \leqslant s \leqslant t\}$ satisfies the following functional equation.

$$\begin{aligned} d_t p(t, x) = &-\left[\sum_{i=1}^{n} \frac{\partial}{\partial x}(f_i p) + \tfrac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}(\sigma_i \sigma_j p)\right] dt \\ &+ [dy - \int g(x, t)p(x, t)dx]'[\mu\mu']^{-1} \\ &\times [dy - \int g(x, t)p(x, t)dx] \end{aligned} \tag{126}$$

where the differential $d_t p$ is interpreted in the Ito sense.

In the special case of linear systems with gauss-markov parameters, discussed before, the functional equation has a solution which is a gaussian distribution. Apart from this special case the solution of the functional equation is an extremely difficult numerical problem even in simple cases. For a system of second order with two parameters the vector x will have four components. If we approximate crudely, e.g. by quantizing each state variable in 100 levels, the storage of the function $p(x, t)$ for a fixed t will require $100^4 = 10^8$ cells.

Contributions to the theory of non-linear filtering are also given by BUCY [170], SHIRYAEV [171], FISHER [226], MORTENSEN [227], WONHAM [174], [220]. A survey is given in the book [225].

Approximations

From the functional equation for the conditional distribution it is possible to derive equations for the maximum likelihood estimate (the mode), the minimum variance estimate (the conditional mean) etc. It turns out that these equations are not well suited for numerical computations. If we want to compute the conditional mean we find that the differential equation for the mean will contain not only covariances but also higher moments. It is therefore of great interest to find approximative methods to solve the nonlinear filtering problem. Approximative schemes have been suggested by BASS and SCHWARTZ [175], NISHIMURA *et al.* [176] SUNAHARA [177], KUSHNER [178], JASZWINSKI [179], [180], LARMINAT and TALLEC [181]. Kushner's article contains a good survey of many of the difficulties associated with the approximative techniques.

The following type of approximation has been suggested by several authors. Estimate x of (125) by $\hat{x}$ given by

$$d\hat{x} = f(\hat{x}, t)dt + \Gamma(t)[dy - g(\hat{x}, t)dt]. \tag{127}$$

The estimate $\hat{x}$ is referred to as the *extended* Kalman filter. The gain matrix Γ of (127) can be chosen in many different ways, e.g. by evaluating the optimal gain for the linearized problem or simply by choosing

$$\Gamma(t) = \frac{1}{t} g'(x, t) \tag{128}$$

in analogy with stochastic approximations. See e.g. ATHANS *et al.* [200], COX [182] and JASZWINSKI [225]. The essential difficulty with all the approximative techniques is to establish convergence. In practice it is often found that the algorithms converge very well if good initial conditions are available, i.e. if there are good initial estimates of the parameters, and if suitable computational "tricks" are used. A computational comparison of several non-linear filters is given by SCHWARTZ and STEAR

[183]. A technique allowing second order non-linearity in system measurements is discussed in NEAL [184].

The application of (127) to system identification was first proposed by KOPP and ORFORD [185]. It has recently been applied to several industrial identification problems, e.g. nuclear reactors in HABEGGER and BAILEY [186], stirred tank reactors in WELLS [187], and head box dynamics in SASTRY and VETTER [188]. The use of the extended Kalman filter in connection with reduced models is discussed by BAGGERUD and BALCHEN [189].

8. CONCLUSIONS

In the previous sections applicable identification techniques as well as (yet) unsolved problems have been mentioned. Particularly from the point of view of the practising engineers (being university professors we are really sticking our necks out now) there are many important questions that remain to be answered. For example how should the sampling interval be chosen? What is a reasonable model structure? How should an identification experiment be planned? How can it be ensured that the *a priori* assumptions required to use a particular method are not violated? If we leave the area of general problems and study specific methods the situation is better.

Linear time-invariant systems. There are good techniques available for identifying linear stationary systems as well as linear environments, stationary stochastic processes. The relations between different techniques are reasonably well understood and the choice of methods can be done largely on the basis of the final purpose of the identification. There are, however, unresolved problems also in this area, for example convergence proofs for the generalized least squares method.

Multivariable systems. The essential difficulty with multivariable systems is to choose a suitable canonical form. A few such forms are available but the structural problems are not yet neatly resolved. For example there are few good techniques to incorporate *a priori* knowledge of the nature that there is no coupling between two variables or that there is a strong coupling between two other variables. If a multivariable system is identified as a conglomerate of single-input single-output systems, how should the different single-input single-output systems be combined into one model. How do we decide if a particular mode is common to several loops taking uncertainties into account.

Once a particular structure is chosen the solution of the multivariable identification problem is straightforward.

Non-linear systems. The techniques currently used simply convert the identification problem to an approximation problem by postulating a structure. The few non-parametric techniques available are computationally extremely time consuming.

On-line and real-time identification. This is a fiddler's paradise. Much work remains to be done to prove convergence as well as to devise approximation techniques.

Empirical knowledge available. The extensive applications of identification methods which are now available provide a source of empirical information which might be worth a closer analysis.

One of the most striking facts is that most methods yield very simple models even for complex systems. It seldom happens that models of a single-input single-output system are of an order higher than 5. This fact, which is extremely encouraging from the point of view of complexity of the regulator, is not at all well-understood.

Most methods seem to work extremely well on simulated data, but not always that well on actual industrial data. This indicates that some methods might be very sensitive to the *a priori* assumptions. It therefore seems highly desirable to develop tests which insure that the *a priori* assumptions are not contradicted by the experimental data. It also means that it is highly desirable to have techniques which are flexible with respect to *a priori* assumptions. A typical example is the assumption that the measurement errors or the residuals have a known covariance function. It is, of course, highly unnatural from a practical point of view to assume that we have an unknown model but that the residuals of this unknown model have known statistics.

Comparison of different techniques

In spite of the large literature on identification there are few papers which compare different techniques. The exceptions are VAN DEN BOOM and MELIS [190], CHERUY and MENENDEZ [191], GUSTAVSSON [78], [100], ŠMUK [192]. Of course, it is more fun to dream up new methods than to work with somebody else's scheme. Nevertheless, for a person engaged in applications it would be highly desirable to have comparisons available. It would also be nice to have a selection of data to which several known techniques are tried which can be used to evaluate new methods.

Where is the field moving?

It is our hope and expectation that the field is moving towards more unification and that there will be more comparisons of different techniques. The textbooks which up to now have been lacking will definitely contribute to that; forthcoming are EYKHOFF [193], SAGE and MELSA [194]. The area of identification will certainly also in the future be influenced by vigorous development of other fields

of control systems theory. One may guess that the presently active work on multivariable systems will result in a deeper understanding of such systems and consequently also of the structural problems. The recent results in stability theory might influence the real time identification algorithms. Pattern recognition, learning theory and related theories will contribute also to the field of identification; e.g. TSYPKIN [195–197]. Also after the IFAC, Prague, 1970 symposium much work remains to be done.

Acknowledgement—The authors like to mention gratefully discussions with many colleagues and material provided by several institutes and individuals. When this paper was written Åström was visiting professor at the Division of Applied Mathematics, Brown University, Providence, Rhode Island, U.S.A. His work on this paper has been partially supported by National Aeronautics and Space Administration under grant No. NGL 40-002-015 and by the National Science Foundation under grant No. GP 7347.

REFERENCES

[1] F. M. FISHER: *The Identification Problem in Econometrics*. McGraw-Hill, New York (1966).
[2] L. A. ZADEH: From circuit theory to system theory. *Proc. IRE* **50,** 856–865 (1962).
[3] N. M. ALEKSANDROVSKII and A. M. DEICH: Determination of dynamic characteristics of non-linear objects (review). 134 ref. *Automn Remote Control* **29,** 142–160 (1968).
[4] G. A. BEKEY: Identification of dynamic systems by digital computers. Joint automatic control conf. (Computer control workshop), 7-18 (1969).
[5] H. STROBEL: On the limits imposed by random noise and measurement errors upon system identification in the time domain. IFAC symp. Identification in autom. control systems, Prague, paper 4.4 (1967).
[6] H. STROBEL: The approximation problem of the experimental system analysis (in German). 161 references. *Zeitschr. Messen, Steuern Regeln* **10,** 460–464 (1967); **11,** 29–34; 73–77 (1968).
[7] J. P. JACOB and L. A. ZADEH: A computer-oriented identification technique. Preprint JACC, 979–981 (1969).
[8] E. S. ANGEL and G. A. BEKEY: Adaptive finite-state models of manual control systems. *IEEE Trans. Man-Machine Syst.* **MMS-9,** 15–20 (1968).
[9] G. FANTAUZZI: The relation between the sampling time and stochastic error for the impulsive response of linear time-independent systems (short paper). *IEEE Trans. Aut. Control* **AC-13,** 426–428 (1968).
[10] K. J. ÅSTRÖM: Computational aspects of a class of identification problems. Report, Dept. of EE, University of Southern California, Los Angeles (1969).
[11] A. SANO and M. TERAO: Measurement optimization in batch process optimal control. Proc. fourth IFAC congress, Warsaw, paper 12.4. Also *Automatica* **6** (1969).
[12] R. BELLMAN and K. J. ÅSTRÖM: On structural identifiability. *Math. Biosci.* **1,** 329–339 (1969).
[13] D. G. LAMPARD: A new method of determining correlation functions of stationary time series. *Proc. IEE* **102C,** 35–41 (1955).
[14] T. KITAMORI: Application of orthogonal functions to the determination of process dynamic characteristics and to the construction of self-optimizing control systems. Proc. first IFAC congress, Moscow, 613–618 (1960).
[15] H. A. BARKER and D. A. HAWLEY: Orthogonalising and optimization. Proc. third IFAC congress, London, paper 36A (1966).
[16] P. D. ROBERTS: The application of orthogonal functions to self-optimizing control systems containing a digital compensator. Proc. third IFAC congress, London, paper 36C (1966).
[17] P. D. ROBERTS: Orthogonal transformations applied to control system identification and optimization. IFAC symp. Identification in autom. control systems, Prague, paper 5.12 (1967).
[18] G. SCHWARZE: Determination of models for identification with quality characteristics in the time domain. Proc. fourth IFAC congress, Warsaw, paper 19.3 (1969).
[19] H. GORECKI and A. TUROWICZ: The approximation method of identification. Proc. fourth IFAC congress, Warsaw, paper 5.5 (1969).
[20] R. DEUTSCH: *Estimation Theory*. Prentice-Hall, Englewood Cliffs, N.J. (1965).
[21] N. E. NAHI: *Estimation Theory and Applications*. John Wiley, New York (1969).
[22] P. EYKHOFF: Process parameter and state estimation. IFAC symp. Identification in autom. control systems, Prague, survey paper; also in *Automatica* **4,** 205–233 (1968).
[23] K. J. ÅSTRÖM and T. BOHLIN: Numerical identification of linear dynamic systems from normal operating records. Paper IFAC symp. Theory of self-adaptive control systems, Teddington, Engl. In *Theory of Self-Adaptive Control Systems* (Ed. P. H. HAMMOND). Plenum Press, New York (1966).
[24] A. V. BALAKRISHNAN: Techniques of system identification. Lecture notes, Institutio Electrotecnio, University of Rome (1969).
[25] R. M. STALEY and P. C. YUE: On system parameter identifiability. Inform. Sci. **2,** 127–138 (1970).
[26] A. A. FEL'DBAUM: Dual-control theory. *Automn. Remote Control* **21,** 874–880, 1033–1039 (1960); **22,** 1–12, 109–121 (1961).
[27] M. MENDES: An on line adaptive control method. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 6.5; see also *Automatica* **7,** No. 3, 1971, and Dual adaptive control of time-varying systems (in German). *Regelungstechnik*; **18,** 157–164 (1970).
[28] S. C. SCHWARTZ and K. STEIGLITZ: The identification and control of unknown linear discrete systems. Report 24, Information Sciences and Systems Laboratory, dept. EE., Princeton University (1968).
[29] K. J. ÅSTRÖM and B. WITTENMARK: Problems of identification and control. Report, Electronic Sciences Laboratory, University of South. Calif., Los Angeles, Calif. (1969). *J. Math. Anal. Appl.* to appear.
[30] B. QVARNSTRÖM: An uncertainty relation for linear mathematical models, Proc. second IFAC congress, Basle, 634 (1963).
[31] V. HERLES: Limits of precision in identification procedures. Second IFAC symp. Identification and process parameter estimation, Prague, paper 9.8 (1970).
[32] H. UNBEHAUEN and G. SCHLEGEL: Estimation of the accuracy in the identification of control systems using deterministic test signals. IFAC symp. Identification autom. control systems, Prague, paper 1.12 (1967).
[33] J. ŠTĚPÁN: The identification of controlled systems and the limit of stability. IFAC symp. Identification in autom. control systems, Prague, paper 1.11 (1967).
[34] J. ŠTĚPÁN: The uncertainty problem in system identification. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 1.8 (1970).
[35] A. V. BALAKRISHNAN and V. PETERKA: Identification in automatic control systems. Proc. fourth IFAC congress, Warsaw, survey paper. Also *Automatica* **5,** 817–829 (1969).
[36] H. B. MANN and A. WALD: On the statistical treatment of linear stochastic difference equations. *Econometrica* **11,** No. 3+4 (1943).

[37] P. WHITTLE: *Prediction and Regulation.* Van Nostrand, Princeton, N.J. (1963).

[38] U. GRENANDER and M. ROSENBLATT: *Statistical Analysis of Stationary Time Series.* John Wiley, New York (1957).

[39] G. JENKINS and D. WATTS: *Spectral Analysis and Its Applications.* Holden Day, San Francisco (1968).

[40] P. N. NIKIFORUK and M. M. GUPTA: A bibliography of the properties, generation and control systems applications of shift-register sequences. *Int. J. Control* **9,** 217–234 (1969).

[41] K. R. GODFREY: The application of pseudo random sequences to industrial processes and nuclear power plant. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 7.1 (1970).

[42] H. STROBEL: *System Analysis Using Deterministic Test Signals* (in German). V.E.B. Verlag Technik, Berlin (1968).

[43] W. GITT: Systematic errors in the determination of frequency response by means of impulse-testing (in German). *Industrie-Anzeiger,* No. 52. Ausgabe "Antrieb-Getriebe-Steuerung" (1969).

[44] H. H. WILFERT: *Signal- and Frequency Response Analysis on Strongly-Disturbed Systems* (in German). V.E.B. Verlag Technik, Berlin (1969).

[45] A. VAN DEN BOS: Estimation of linear system coefficients from noisy responses to binary multifrequency test signals. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 7.2 (1970).

[46] E. WELFONDER and O. HASENKOPF: Comparison of deterministic and statistical methods for system identification by test-measurements at a disturbed industrial control system. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 7.5 (1970).

[47] I. G. CUMMING: Frequency of input signal in identification. 2nd IFAC symposium Identification and process parameter estimation, Prague, paper 7.8. (1970).

[48] K. J. ÅSTRÖM: Lectures on the identification problem. The least squares method. Report, Division of Automatic Control, Lund Institute of Technology, Lund, Sweden (1968).

[49] M. AOKI and R. M. STALEY: On input signal synthesis in parameter identification. Proc. fourth IFAC congress, Warsaw, paper 26.2. Also *Automatica* **6,** 431–440 (1969).

[50] V. S. LEVADI: Design of input signals for parameter estimation. *IEEE Trans. Aut. Control* **AC-11,** 205–211 (1966).

[51] A. RAULT, R. POULIQUEN and J. RICHALET: Sensibilizing input and identification. Proc. fourth IFAC congress, Warsaw, paper 5.2 (1969).

[52] D. MIDDLETON: *An Introduction to Statistical Communication Theory.* McGraw-Hill, New York (1960).

[53] E. E. FISHER: Identification of linear systems. Paper JACC, 473–475 (1965).

[54] R. W. BROCKETT and M. D. MESAROVIC: The reproducibility of multivariable control systems. *JMAA* **11,** 548–563 (1965).

[55] L. M. SILVERMAN: Inversion of multivariable linear systems. *IEEE Trans. Aut. Control* **AC-14,** 270–276 (1969).

[56] M. K. SAIN and J. L. MASSEY: Invertibility of linear, time-invariant dynamical systems. *IEEE Trans. Aut. Control* **AC-14,** 141–149 (1969).

[57] H. SILVERMAN: Identification of linear systems using fast fourier transform technique. Ph.D. thesis, Brown University (1970).

[58] T. F. POTTS, G. N. ORNSTEIN and A. B. CLYMER: The automatic determination of human and other system parameters. Proc. West. Joint Computer Conf., Los Angeles, **19,** 645–660 (1961).

[59] D. J. WILDE: *Optimum Seeking Methods.* Prentice-Hall, Englewood Cliffs, N.J. (1964).

[60] G. A. BEKEY and R. B. MCGEE: Gradient methods for the optimization of dynamic system parameters by hybrid computation. In *Computing methods in Optimization Problems* (Eds. BALAKRISHNAN and NEUSTADT). Academic Press, New York (1964).

[61] T. DISKIND: Parameter evaluation by complex domain analysis. Preprint JACC, 770–781 (1969).

[62] R. E. KALMAN: Mathematical description of linear dynamical systems. *SIAM J. Control* **1,** 152–192 (1963).

[63] B. L. HO and R. E. KALMAN: Effective construction of linear state-variable models from input/output functions. *Regelungstechnik* **14,** 545–548 (1966).

[64] R. E. KALMAN, P. L. FALB and M. A. ARBIB: *Topics in Mathematical System Theory.* McGraw-Hill, New York (1969).

[65] R. W. KOEPCKE: Private communication (1963).

[66] K. Y. WONG, K. M. WIIG and E. J. ALLBRITTON: Computer control of the Clarksville cement plant by state space design method. *IEEE Cement Industry Conf.,* St. Louis, Mo. (1968).

[67] I. H. ROWE: A statistical model for the identification of multivariable stochastic systems. Proc. IFAC symp. on multivariable systems, Düsseldorf (1968).

[68] W. G. TUEL, JR.: Canonical forms for linear systems —I and II. IBM Research report. R.J. 375 and 390 (1966).

[69] D. G. LUENBERGER: Canonical forms for linear multivariable systems (short paper). *IEEE Trans. Aut. Control* **AC-12,** 290–293 (1967).

[70] E. J. DAVISON: A new method for simplifying large linear dynamic systems (correspondence). *IEEE Trans. Aut. Control* **AC-13,** 214–215 (1968).

[71] T. KAILATH: An innovations approach to least-squares estimation. Part I: Linear filtering in additive white noise. *IEEE Trans. Aut. Control* **AC-13,** 646–655 (1968).

[72] K. J. ÅSTRÖM: *Introduction to Stochastic Control Theory.* Academic Press, New York (1970).

[73] R. K. MEHRA: Identification of stochastic linear dynamic systems. Proc. 1969 IEEE symposium on adaptive processes. Pennsylvania State University (1969).

[74] K. J. ÅSTRÖM, T. BOHLIN and S. WENSMARK: Automatic construction of linear stochastic dynamic models for industrial processes with random disturbances using operating records. Report, IBM Nordic Laboratory, Stockholm (1965).

[75] J. EATON: Identification for control purposes. Paper, IEEE Winter meeting, N.Y. (1967).

[76] R. L. KASHYAP: Maximum likelihood identification of stochastic linear systems. *IEEE Trans. Aut. Control* **AC-15,** 25–34 (1970).

[77] T. BOHLIN: Real-time estimation of time-variable processes characteristics. Report, IBM Nordic Lab., Lidingö, Sweden (1968).

[78] I. GUSTAVSSON: Maximum likelihood identification of dynamics of the Ågesta reactor and comparison with results of spectral analysis. Report 6903. Division of Automatic Control, Lund Institute of Technology, Lund, Sweden (1969).

[79] H. RAKE: The determination of frequency responses from stochastic signals by means of an electronic analog computer (in German). *Regelungstechnik* **16,** 258–260 (1968).

[80] E. WELFONDER: Stochastic signals; statistical averaging under real conditions (in German). *Fortsch. Ber. VDI-Z* **8,** 165 (1969).

[81] H. BUCHTA: Method of estimation of random errors in the determination of correlation functions of random infralow frequency signals in control systems. Proc. fourth IFAC congress, Warsaw, paper 33.4 (1969).

[82] N. HAYASHI: On a method of correlation analysis for multivariate systems. Proc. fourth IFAC congress, Warsaw, paper 33.1 (1969).

[83] K. N. REID: Lateral dynamics of an idealized moving web. Preprint JACC, 955–963 (1969).
[84] K. N. REID: Lateral dynamics of a real moving web. Preprint JACC, 964–976 (1969).
[85] H. G. STASSEN: The polarity coincidence correlation technique—a useful tool in the analysis of human operator dynamics. *IEEE Trans. Man–Machine Syst.* **MMS-10,** 34–39 (1969).
[86] K. GERDIN: Markov modelling via correlation analysis applied to identification of a paper machine in the absence of test signal. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 2.8 (1970).
[87] J. R. WESTLAKE: *A Handbook of Numerical Matrix Inversion and Solution of Linear Equations.* John Wiley, New York (1968).
[88] C. M. WOODSIDE: Estimation of the order of linear systems. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 1.7 (1970).
[89] T. W. ANDERSON: The choice of the degree of a polynomial regression as a multiple decision problem. *Ann. math. stat.* **33,** 255–265 (1962).
[90] H. WOLD: A study in the analysis of stationary time series. Uppsala, Almqvist & Wiksell (1938).
[91] W. GERSCH: Spectral analysis of e.e.g.'s by autoregressive decomposition of time series. Report, Center of Applied Stochastics, School of Aeronautics, Purdue University; *Math. Biosci.* **7,** 205–222 (1970).
[92] K. J. ÅSTRÖM: Computer control of a paper machine —an application of linear stochastic control theory. *IBM J. Res. Dev.* **11,** 389–405 (1967).
[93] D. W. CLARKE: Generalized-least-squares estimation of the parameters of a dynamic model. IFAC symp. Identification in autom. control system, Prague, paper paper 3.17 (1967).
[94] K. STEIGLITZ and L. E. MCBRIDE: A technique for the identification of linear systems (short paper). *IEEE Trans. Aut. Control* **AC-10,** 461–464 (1965).
[95] A. E. ROGERS and K. STEIGLITZ: Maximum likelihood estimation of rational transfer function parameters (short paper). *IEEE Trans. Aut. Control* **AC-12,** (594–597 (1967).
[96] V. PANUSKA: On the maximum likelihood estimation of rational pulse transfer function parameters (correspondence). *IEEE Trans. Aut. Control* **AC-13,** 304 (1968).
[97] K. T. WOO: Maximum likelihood identification of noisy systems. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 3.1 (1970).
[98] K. J. ÅSTRÖM: On the achievable accuracy in identification problems. Proc. IFAC symp. Identification in autom. control systems, Prague, paper 1.8 (1967).
[99] T. BOHLIN: On the problem of ambiguities in maximum likelihood identification. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 3.2 (1970). Also see this issue.
[100] I. GUSTAVSSON: Comparison of different methods for identification of linear models for industrial processes. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 11.4 (1970).
[101] L. H. ZETTERBERG: Estimation of parameters for a linear difference equation with application to EEG analysis. *Math. Biosci.* **5,** 227–275 (1969).
[102] M. G. KENDALL and A. STUART: *The Advanced Theory of Statistics,* Vol. II. Griffin, London (1961).
[103] K. Y. WONG and E. POLAK: Identification of linear discrete time systems using the instrumental variable method. *IEEE Trans. Aut. Control* **AC-12,** 707–718 (1967).
[104] V. PETERKA and K. ŠMUK: On-line estimation of dynamic model parameters from input–output data. Proc. fourth IFAC congress, Warsaw, paper 26.1 (1969).
[105] T. C. HSIA and D. A. LANDGREBE: On a method for estimating power spectra. *IEEE Trans. Instrum. Meas.* **IM-16,** 255–257 (1967).
[106] M. I. VUŠKOVIĆ, S. P. BINGULAC and M. DJOROVIĆ: Applications of the pseudo sensitivity functions in linear dynamic systems identification. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 2.5 (1970).
[107] M. J. LEVIN: Estimation of the characteristics of linear systems in the presence of noise. Eng. Sc.D. dissertation, Dept. EE., Columbia University, New York (1959).
[108] M. AOKI and P. C. YUE: On certain convergence questions in system identification. *SIAM J. Control* **8,** 239–256 (1970).
[109] F. W. SMITH: System Laplace-transform estimation from sampled data. *IEEE Trans. Aut. Control* **AC-13,** 37–44 (1968).
[110] V. PETERKA and A. HALOUSKOVÁ: Tally estimate of Åström model for stochastic systems. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 2.3 (1970).
[111] W. OCHEM, A. RAULT and H. WEINGARTEN: A new identification algorithm for linear discrete time systems. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 2.9 (1970).
[112] D. GRAUPE, B. H. SWANICK and G. R. CASSIR: Reduction and identification of multivariable processes using regression analysis (short paper). *IEEE Trans. Aut. Control* **AC-13,** 564–567 (1968).
[113] P. EYKHOFF: Some fundamental aspects of process-parameter estimation. *IEEE Trans. Aut. Control* **AC-8,** 347–357 (1963).
[114] P. ALPER: A consideration of the discrete Volterra series. *IEEE Trans. Aut. Control* **AC-10,** 322–327 (1965).
[115] R. J. ROY and J. SHERMAN: A learning technique for Volterra series representation (short paper). *IEEE Trans. Aut. Control* **AC-12,** 761–764 (1967).
[116] A. G. IVAKHNENKO: Heuristic self-organization in problems of engineering cybernetics. *Automatica* **6,** 207–219 (1970).
[117] A. G. IVAKHNENKO and JU.V. KOPPA: The group method of data handling for the solution of the various interpolation problems of cybernetics. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 2.1 (1970).
[118] R. E. BUTLER and E. V. BOHN: An automatic identification technique for a class of non-linear systems (short paper). *IEEE Trans. Aut. Control* **AC-11,** 292–296 (1966).
[119] K. S. NARENDRA and P. G. GALLMAN: An iterative method for the identification of non-linear systems using a Hammerstein model (short paper). *IEEE Trans. Aut. Control* **AC-11,** 546–550 (1966).
[120] D. W. MARQUARDT: An algorithm for least-squares estimation of non-linear parameters. *J. Soc. ind. appl. math.* **11,** 431–441 (1963).
[121] R. E. BELLMAN and R. E. KALABA: *Quasi-linearization and non-linear boundary-valve problems.* Elsevier, New York (1965).
[122] J. D. BUEL, H. H. KAGIWADA and R. E. KALABA: A proposed computational method for estimation of orbital elements drag coefficients and potential fields parameters from satellite measurements. *Ann. Géophys.* **23,** 35–39 (1967).
[123] J. D. BUEL and R. E. KALABA: Quasilinearization and the fitting of non-linear models of drug metabolism to experimental kinetic data. *Math. Biosci.* **5,** 121–132 (1969).
[124] A. VANĚČEK and J. FESSL: A contribution to parameters and state estimation and synthesis via quasi-linearization. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 2.6 (1970).
[125] L. W. TAYLOR, K. W. ILIFF and B. G. POWERS: A comparison of Newton–Raphson and other methods for determining stability derivatives from flight data. Paper, AIAA third Flight Test Simulation and Support Conf., Houston, Texas (1969).
[126] E. P. MASLOV: Application of the theory of statistical decisions to the estimation of object parameters. *Automn Remote Control* **24,** 1214–1226 (1963).

[127] C. A. GALTIERI: The problem of identification of discrete time processes. Report, Electronics Research Lab., University of California, Berkeley (1963).
[128] R. B. MCGHEE and R. B. WALFORD: A Monte Carlo approach to the evaluation of conditional expectation parameter estimates for non-linear dynamic systems. *IEEE Trans. Aut. Control* **AC-13**, 29–37 (1968).
[129] Y. GENIN: A note on linear minimum variance estimation problems (correspondence). *IEEE Trans. Aut. Control* **AC-13**, 103 (1968).
[130] H. P. WHITAKER *et al.*: Design of a model reference adaptive control system for aircraft. Report R-164, Instrumentation Lab., MIT, Boston (1958).
[131] S. KACZMARZ: Approximate solution of systems of linear equations (in German). *Bull. Int. Acad. polon. Sci., Cl. Sci. Math. Nat. Ser. A.* 355–357 (1937).
[132] J. NAGUMO and A. NODA: A learning method for system identification. *IEEE Trans. Aut. Control* **AC-12**, 282–287 (1967).
[133] P. R. BÉLANGER: Comments on "A learning method for system identification" (correspondence). *IEEE Trans. Aut. Control* **AC-13**, 207–208 (1968).
[134] J. RICHALET, F. LECAMUS and P. HUMMEL: New trends in identification: structural distance and weak topology. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 1.2 (1970).
[135] L. A. CRUM and S. H. WU: Convergence of the quantizing learning method for system identification (correspondence). *IEEE Trans. Aut. Control* **AC-13**, 297–298 (1968).
[136] H. F. MEISSINGER and G. A. BEKEY: An analysis of continuous parameter identification method. *Simulation* **6**, 95–102 (1966).
[137] R. E. ROSE and G. M. LANCE: An approximate steepest descent method for parameter identification. *Proc. JACC*, 483–487 (1969).
[138] P. M. LION: Rapid identification of linear and non-linear systems. Preprint JACC, Washington (1966).
[139] B. SHACKCLOTH and R. L. BUTCHART: Synthesis of model reference adaptive systems by Liapunov's second method. IFAC symp. theory of selfadaptive control systems, Teddington, 145–152 (1965).
[140] P. C. PARKS: Liapunov redesign of model reference adaptive control systems. *IEEE Trans. Aut. Control* **AC-11**, 362–367 (1966).
[141] R. P. WISHNER, J. A. TABACZYNSKI and M. ATHANS: A comparison of three non-linear filters. *Automatica* **5**, 487–496 (1969).
[142] J. S. PAZDERA and H. J. POTTINGER: Linear system identification via Liapunov design techniques. Preprints JACC, 795–801 (1969).
[143] H. J. KUSHNER: On the convergence of Lion's identification method with random inputs. Report, Division of Applied Mathem., Brown University, R.I. (1969).
[144] V. M. POPOV: Absolute stability of non-linear systems of automatic control. *Automn. Remote Control* **22**, 857–875 (1962).
[145] G. ZAMES: On the input–output stability of time-varying non-linear feedback systems. *IEEE Trans. Aut. Control* **AC-11**, 228–238 (1966).
[146] I. D. LANDAU: An approach to synthesis of model reference adaptive control systems. Report, Alsthon-Recherches; to appear in *IEEE Trans. Aut. Control* (1970).
[147] A. KLINGER: Prior information and bias in sequential estimation (correspondence). *IEEE Trans. Aut. Control* **AC-13**, 102–103 (1968).
[148] P. C. YOUNG: An instrumental variable method for real-time identification of a noisy process. Proc. fourth IFAC congress, Warsaw, paper 26.6. Also *Automatica* **6**, 271–287 (1970).
[149] J. F. LEATHRUM: On-line recursive estimation. *Proc. JACC*, 173–178 (1969).
[150] S. STANKOVIĆ, D. VELAŠEVIĆ and M. ČARAPIĆ: An approach to on-line discrete-time system identification. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 4.2 (1970).
[151] K. G. OZA and E. I. JURY: System identification and the principle of contraction mapping. *SIAM J. Control* **6**, 244–257 (1968).
[152] K. G. OZA and E. I. JURY: Adaptive algorithms for identification problem. Proc. fourth IFAC congress, Warsaw, paper 26.4. Also *Automatica* **6**, 795–799 (1970).
[153] A. E. ALBERT and L. A. GARDNER JR.: *Stochastic Approximation and Non-linear Regression.* MIT Press, Cambridge (Mass.) (1967).
[154] YA. Z. TSYPKIN: Adaption, learning and selflearning in control systems. Proc. third IFAC congress, London, survey paper; also: Adaptation, training and self-organization in automatic systems, *Automn. Remote Control* **27**, 16–51, 160 lit. ref. (1966).
[155] D. J. SAKRISON: The use of stochastic approximation to solve the system identification problem. *IEEE Trans. Aut. Control* **AC-12**, 563–567 (1967).
[156] G. N. SARIDIS and G. STEIN: Stochastic approximation algorithms for linear discrete-time system identification. *IEEE Trans. Aut. Control* **AC-13**, 515–523 (1968).
[157] G. N. SARIDIS and G. STEIN: A new algorithm for linear system identification (correspondence). *IEEE Trans. Aut. Control* **AC-13**, 592–594 (1968).
[158] J. K. HOLMES: System identification from noise corrupted measurements. *J. Optimizn Theory Appl.* **2**, 103–116 (1968).
[159] D. F. ELLIOTT and D. D. SWORDER: Applications of a simplified multidimensional stochastic approximation algorithm. *Proc. JACC*, 148–154 (1969).
[160] C. B. NEAL and G. A. BEKEY: Estimation of the parameters of sampled data systems by means of stochastic approximation. Preprints 1969 Joint autom. control conf. (sess. VI-E), 977–978 (1969).
c.f. also: C. B. NEAL: Estimation of the parameters of sampled data systems by stochastic approximation, Univ. of South Calif., Electronic Sciences, Lab., techn. report (1969).
[161] T. BOHLIN: Information pattern for linear discrete-time models with stochastic coefficients. *IEEE Trans. Aut. Control* **AC-15**, 104–106 (1970).
[162] T. BOHLIN: On the maximum likelihood method of identification. *IBM J. Res. Develop.* **14**, 41–51 (1970).
[163] J. WIESLANDER: Real time identification Parts I and II. Reports, Division of Automatic Control, Lund Institute of Technology, Lund, Sweden (1969).
[164] B. SEGERSTÅHL: A theory of parameter tracking applied to slowly varying non-linear systems. Proc. fourth IFAC congress, Warsaw, paper 5.4 (1969).
[165] A. P. SAGE and G. W. HUSA: Adaptive filtering with unknown prior statistics. Proc. *JACC*, 760 [1969).
[166] J. W. BERKOVEC: A method for the validation of Kalman filter models. *Proc. JACC*, 488–493 (1969).
[167] J. D. PEARSON: Deterministic aspects of the identification of linear discrete dynamic systems (short paper). *IEEE Trans. Aut. Control* **AC-12**, 764–766 (1967).
[168] T. C. HSIA and V. VIMOLVANICH: An on-line technique for system identification (short paper). *IEEE Trans. Aut. Control* **AC-14**, 92–96 (1969).
[169] J. B. FARISON: Parameter identification for a class of linear discrete systems (correspondence). *IEEE Trans. Aut. Control* **AC-12**, 109 (1967).
[170] R. S. BUCY: Non-linear filtering theory. *IEEE Trans. Aut. Control* **AC-10**, 198 (1965).
[171] A. N. SHIRYAEV: On stochastic equations in the theory of conditional Markov processes. *Theory of probability and its applications, U.S.S.R.* **11**, 179 (1966).
[172] H. J. KUSHNER: Non-linear filtering: the exact dynamical equations satisfied by the conditional mode. *IEEE Trans. Aut. Control* **AC-12**, 262–267 (1967).
[173] R. L. STRATONOVICH: Conditional Markov processes. *Theory Probab. Appl.* **5**, 156–178 (1967).
[174] W. M. WONHAM: Some applications of stochastic differential equations to non-linear filtering. *SIAM J. Control* **2**, 347–369 (1964).

[175] R. BASS and L. SCHWARTZ: Extensions to multichannel non-linear filtering. Hughes report SSD 604220R (1966).
[176] M. NISHIMURA, K. FUJII and Y. SUZUKI: On-line estimation and its application to an adaptive control system. Proc. fourth IFAC congress, Warsaw, paper 26.3 (1969).
[177] Y. SUNAHARA: An approximate method of state estimation for non-linear dynamical systems. *Proc. JACC*, 161–172 (1969).
[178] H. J. KUSHNER: Approximations to optimal non-linear filters. Preprints JACC, Philadelphia, Pa., 613–623 (1967).
[179] A. JASZWINSKI: Filtering for non-linear dynamical systems. *IEEE Trans. Aut. Control* **AC-11,** 765 (1966).
[180] A. H. JASZWINSKI: Adaptive filtering. *Automatica* **5,** 475–485 (1969).
[181] PH. DE. LARMINAT and M. H. TALLEC: On-line identifications of time varying systems. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 3.3 (1970).
[182] H. COX: On the estimation of state variables and parameters for noisy dynamic systems. *IEEE Trans. Autom. Control* **AC-9,** 5–12 (1964).
[183] L. SCHWARTZ and E. B. STEAR: A computational comparison of several non-linear filters (short paper). *IEEE Trans. Aut. Control* **AC-13,** 83–86 (1968).
[184] S. R. NEAL: Non-linear estimation techniques (short paper). *IEEE Trans. Aut. Control* **AC-13,** 705–708 (1968).
[185] R. E. KOPP and R. J. ORFORD: Linear regression applied to system identification for adaptive control systems. *AIAA Journal* 2300–2306 (1963).
[186] L. J. HABEGGER and R. E. BAILEY: Minimum variance estimation of parameters and states in nuclear power systems. Proc. fourth IFAC congress, Warsaw, paper 12.2 (1969).
[187] C. H. WELLS: Application of the extended Kalman estimator to the non-linear well stirred reactor problem. *Proc. JACC*, 473–482 (1969).
[188] V. A. SASTRY and W. J. VETTER: The application of identification methods to an industrial process. Preprints JACC, 787–794 (1969).
[189] A. BAGGERUD and J. G. BALCHEN: An adaptive state estimator. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 10.1 (1970).
[190] A. J. W. VAN DEN BOOM and J. H. A. M. MELIS: A comparison of some process parameter estimating schemes. Proc. fourth IFAC congress, Warsaw, paper 26.5 (1969).
[191] J. CHERUY and A. MENENDEZ: A comparison of three dynamic identification methods. Preprints JACC, 982–983 (1969).
[192] K. ŠMUK: Comparison of some algorithms for least squares updating. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 2.10 (1970).
[193] P. EYKHOFF: System parameter and state estimation. John Wiley, London; to appear (1971).
[194] A. P. SAGE and J. L. MELSA: *System Identification.* Academic Press, New York; to appear.
[195] YA. Z. TSYPKIN: Optimization, adaptation, and learning in automatic systems. *Computer Inform. Sci.* **2,** 15–32 (1967).
[196] YA. Z. TSYPKIN: Self-learning—What is it? *IEEE Trans. Aut. Control* **AC-13,** 608–612 (1968).
[197] YA. Z. TSYPKIN: Generalized algorithms of learning. *Automn Remote Control* **31,** 97–104 (1970).
[198] M. AOKI and R. M. STALEY: Some computational considerations in input signal synthesis problems. In *Computing Methods in Optimization Problems*—Vol. 2 (Eds. ZADEH, NEUSTADT, BALAKRISHNAN). Academic Press, New York (1969).
[199] K. J. ÅSTRÖM: Control problems in papermaking. Proc. IBM Scientific computing symposium on control theory and applications, Yorktown Heights, New York (1964).
[200] M. ATHANS, R. P. WISHNER and A. BERTOLINI: Suboptimal state estimation for continuous-time non-linear systems from discrete noisy measurements. *IEEE Trans. Aut. Control* **AC-13,** 504–514 (1968).
[201] C. BRUNI, A. ISIDORI and A. RUBERTI: Order and factorization of the impulse response matrix. Proc. fourth IFAC congress, Warsaw, paper 19.4 (1969).
[202] R. F. BROWN: Review and comparison of drift correction schemes for periodic cross correlation. *Electron. Letters* **5,** 179–181 (1969).
[203] M. KOSZELNIK, J. MALKIEWICZ and ST. TRYBULA: A method to determine the transfer functions of power systems. Proc. fourth IFAC congress, Warsaw, paper 33.2 (1969).
[204] N. S. RAIBMAN, S. A. ANISIMOV and V. M. CHADEYE: Linear plant type identification technique. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 1.6 (1970).
[205] R. S. BUCY: Canonical forms for multivariable systems (short paper). *IEEE Trans. Aut. Control* **AC-13,** 567–569 (1968).
[206] M. CUENOD and A. SAGE: Comparison of some methods used for process identification. IFAC symp. Identification in autom. control systems, Prague, survey paper; also in *Automatica* **4,** 235–269 (1968).
[207] F. DONATI and M. MILANESE: System identification with approximated models. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 1.9 (1970).
[208] P. EYKHOFF, P. M. VAN DER GRINTEN, H. KWAKERNAAK and B. P. VELTMAN: Systems modelling and identification. Proc. third IFAC congress, London, survey paper (1966).
[209] R. C. K. LEE: Optimal estimation, identification and control. MIT Press, Cambridge (Mass.) (1964).
[210] M. J. LEVIN: Optimum estimation of impulse response in the presence of noise. *IRE Trans. Circuit Theory* **CT-7,** 50–56 (1960).
[211] M. J. LEVIN: Estimation of system pulse transfer function in the presence of noise. Preprints JACC, Minneapolis, 452–458 (1963).
[212] D. V. LINDLEY: *Introduction to Probability and Statistics, from a Bayesian Viewpoint.* Cambridge Univ. Press, London (1965).
[213] N. E. NAHI: Optimal inputs for parameter estimation in dynamic systems with white observation noise. *Proc. JACC*, 506–513 (1969).
[214] V. PANUSKA: An adaptive recursive least-squares identification algorithm. Proc. 1969, IEEE symposium on adaptive processes. Pennsylvania State University (1969).
[215] N. S. RAIBMAN and O. F. HANS: Variance methods of identifying multidimensional non-linear controlled systems. *Automn Remote Control* **28,** 695–714 (1967).
[216] N. S. RAIBMAN, S. A. ANISIMOV and F. A. HOVSEPIAN: Some problems of control plants identification. Proc. fourth IFAC congress, Warsaw, paper 5.1 (1969).
[217] I. H. ROWE: A bootstrap method for the statistical estimation of model parameters. Preprints JACC, 989–987 (1969).
[218] E. R. SCHULZ: Estimation of pulse transfer function parameters by quasilinearization (short paper). *IEEE Trans. Aut. Control* **AC-13,** 424–426 (1968).
[219] J. VALIŠ: On-line identification of multivariable-linear systems of unknown structure from input output data. 2nd IFAC symp. Identification and process parameter estimation, Prague, paper 1.5 (1970).
[220] W. M. WONHAM: Random differential equations in control theory. In *Probabilistic Methods in Applied Mathematics* (Ed. A. T. BHARUCHA-REID). Academic Press, New York (1969).
[221] N. S. RAIBMAN: What is system identification? (in Russian). Moscow (1970).
[222] P. CAINES: Maximum likelihood identification of multivariable systems. Ph.D. thesis, Imperial College, London (1970).

[223] R. L. STRATONOVICH: *Conditional Markov Processes and Their Application to the Theory of Optimal Control.* American Elsevier, New York (1968).
[224] K. J. ÅSTRÖM and S. WENSMARK: Numerical identification of stationary time series (1964).
[225] A. H. JASZWINSKI: *Stochastic Processes and Filtering Theory.* Academic Press, New York (1970).
[226] J. FISHER: Conditional probability density functions and optimal non-linear estimation. Ph.D. thesis, UCLA, Los Angeles (1966).
[227] R. E. MORTENSEN: Optimal control of continuous-time stochastic systems. Ph.D. thesis (engineering), University of California, Berkeley (1966).
[228] W. L. ROOT: On system measurement and identification. Proc. Symposium on system theory, Polytechnic Institute of Brooklyn, April 1965.
[229] R. T. PROSSER and W. L. ROOT: Determinable classes of channels. *J. Math. Mech.* **16**, 365–398 (1966).
[230] W. L. ROOT: On the structure of a class of system identification problems. 2nd IFAC Symposium. Identification and process parameter estimation, Prague, paper 8.1 (1970).

APPENDIX A

A resumé of parameter estimation

As a tutorial resumé of the statistical methods the following example of utmost simplicity may suffice. Consider the situation of Fig. 8 where an

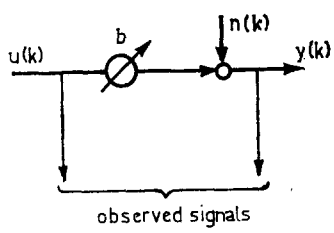

FIG. 8. The simplest example of parameter estimation.

estimate β has to be found for parameter b. This estimate has to be derived from a number of signal samples $u(k)$ and $y(k)$ where

$$y(k)=bu(k)+n(k) \qquad (A.1)$$

and where the average (expected) value of $n(k)$ is zero.

Using the *least squares* method the estimate is chosen in such a way that the loss function, defined as

$$V(\beta)=\sum_k [y(k)-\beta u(k)]^2=(y-\beta u)'(y-\beta u)$$

is a minimum.

In Fig. 9 the differences between the observations y and the "predictions" βu are indicated. The minimization can be pursued analytically; a necessary condition for the minimum is

$$\frac{d}{d\beta}\sum_k [y(k)-\beta u(k)]^2\Big|_{\beta=\hat{\beta}}=0 \qquad (A.2)$$

or

$$\sum_k u(k)[y(k)-\hat{\beta}u(k)]=0 \qquad (A.3)$$

or

$$\hat{\beta}=\frac{\sum_k u(k)y(k)}{\sum_k u(k)u(k)}. \qquad (A.4)$$

$\hat{\beta}$ is the optimal estimate under the conditions given. Note from (A.3) that the terms $y(k)-\beta u(k)$ are weighted with respect to $u(k)$; quite naturally the larger the input signal, the more importance is assigned to the deviation between observation $y(k)$ and "prediction" $\beta u(k)$! Equation (A.4) refers to the correlation methods. For the extension to the more-parameter case and to the generalized least-squares method c.f. section 5.

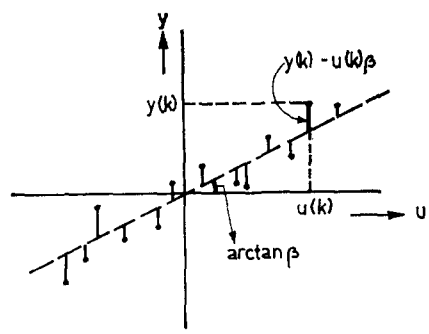

FIG. 9. Interpretation of least-squares estimation.

Using the *maximum-likelihood* method for the same case as before, we have to know p_n, the probability density function of $n(k)$. In that case the measurement of $u(k)$ provides us with the knowledge sketched in Fig. 10, the *a priori* probability

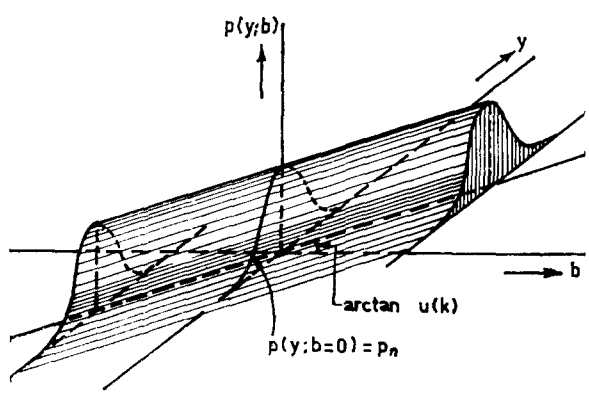

FIG. 10. Interpretation of maximum-likelihood estimation; *a priori* knowledge.

density of $y(k)$ with b as parameter. Now the measurement, or *a posteriori* knowledge, of $y(k)$ brings us to the situation indicated in Fig. 11. The function $L(y(k)\,;\beta)$ is called *likelihood function.*

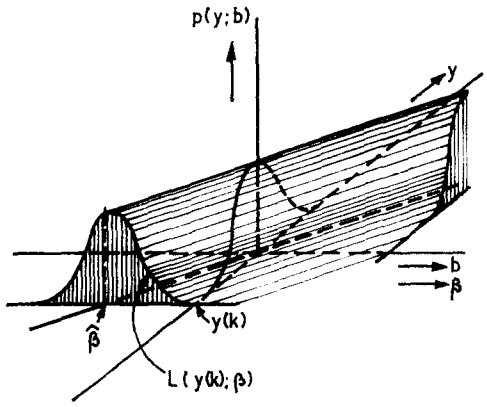

FIG. 11. Interpretation of maximum-likelihood estimation; *a posteriori* knowledge.

We have to assign an estimate $\hat{\beta}$ from this function L. A reasonable and popular choice is to take that value $\hat{\beta}$, for which $L(y(k); \beta)$ has its maximal value. Again this can be generalized to more-parameter cases.

Using the *Bayes'* method for the same case, one needs, as before, p_n, but also the *a priori* probability density function p_b of b. Note, that previously b was an unknown constant, and that now b is a random variable. From Bayes' rule

$$p(b|y)=\frac{p(y, b)}{p(y)}=\frac{p(y|b)p(b)}{p(y)}.$$

This can be interpreted as: the probability density function of the parameter b, given the result of the measurement on y. This can be rewritten as

$$p(y, b)=p(y-ub, b)\underbrace{\left|\frac{\partial(y-ub)}{\partial y}\right|}_{1}=p_n(y-ub)p_b(b)$$

since b and n are statistically independent. p_b and p_n give a probability density function as indicated in Fig. 12. Note, that this is the *a priori* knowledge,

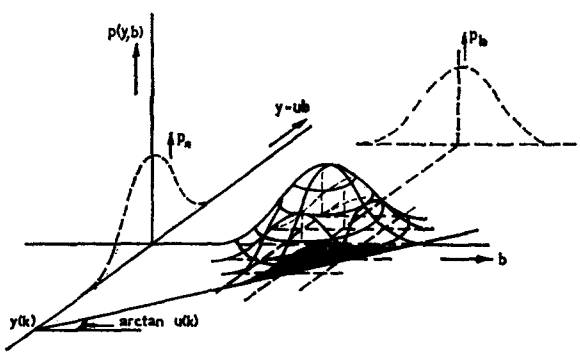

FIG. 12. Interpretation of Bayes' estimation.

available before the measurement of $u(k)$ and $y(k)$. These measurements provide us with a "cut" through the probability-density function, from which the *a posteriori* probability function for b follows. This new probability function now may be used as the *a priori* knowledge for the following measurement. In this way the development of p_b with increasing number of observations can be followed; c.f. Fig. 13. Note, that p_n, the additive noise being stationary, does not change.

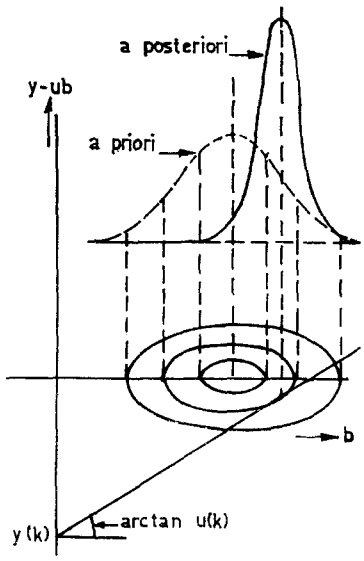

FIG. 13. Interpretation of Bayes' estimation.

The reader is invited to consider special cases like $u(k)=0$ and $u(k)\to\infty$. Again the method can be generalized to more parameters; its vizualization has severe limitations, however. Note that in this case the knowledge on b is given in terms of p_b, a function. In practice the reduction from a function to a single value (estimate) can be done by using a cost or loss function, providing a *minimum cost*— or *minimum loss* estimate.

The problem of input noise. The simple example of Fig. 8 also serves very well to illustrate the so-called "problem of input noise". Consider the system illustrated by the block diagram in Fig. 14 where neither the input nor the output can be observed exactly.

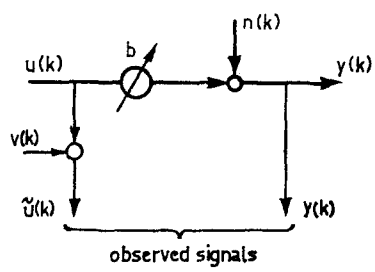

FIG. 14. The problem of input noise.

It is well-known in statistics (see e.g. LINDLEY [212]) that, unless specific assumptions are made concerning the variations in u, n, and v, it is not possible to estimate the dynamics of the process.

Consider e.g.

$$y(k)=bu(k)+n(k)$$
$$\tilde{u}(k)=u(k)+v(k).$$

Assume, that $v(k)$ and $n(k)$ are independent stochastic variables with zero mean values. If $v(k)=0$ we find that the estimate of b is given by (A.3)

$$\hat{\beta}=\frac{\sum_k \tilde{u}(k)y(k)}{\sum_k \tilde{u}(k)\tilde{u}(k)}.$$

However, if $n(k)=0$ we find by the same argument that the estimate of b is

$$\hat{\beta}=\frac{\Sigma\tilde{u}(k)y(k)}{\Sigma y(k)y(k)}.$$

This corresponds to choosing β such that the difference between the observations $u(k)$ and the predictions $y(k)/\beta$ are as small as possible in the least squares sense. See Fig. 15.

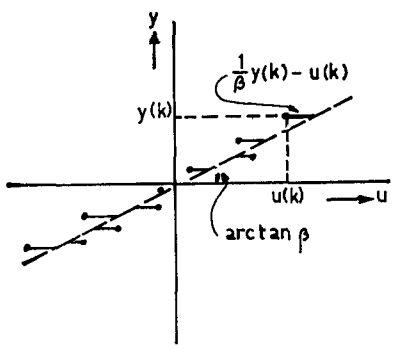

FIG. 15. The problem of input noise.

Without additional information it is, of course, impossible to tell which estimate to choose.

APPENDIX B

An example of least squares identification of a parametric model

To illustrate the least squares method, its applications and some numerical problems which might arise, we provide an example. The computations are carried out on input–output data from a known process.

In Fig. 16 we show input–output pairs which are generated by the equation

$$\underset{a_1}{y(k)} \underset{-1{\cdot}5y(k-1)}{\underset{a_2}{+0{\cdot}7y(k-2)}} = \underset{b_1}{1{\cdot}0u(k-1)} \underset{+0{\cdot}5u(k-2)+\lambda e(k)}{\underset{b_2}{}} \quad \text{(A.5)}$$

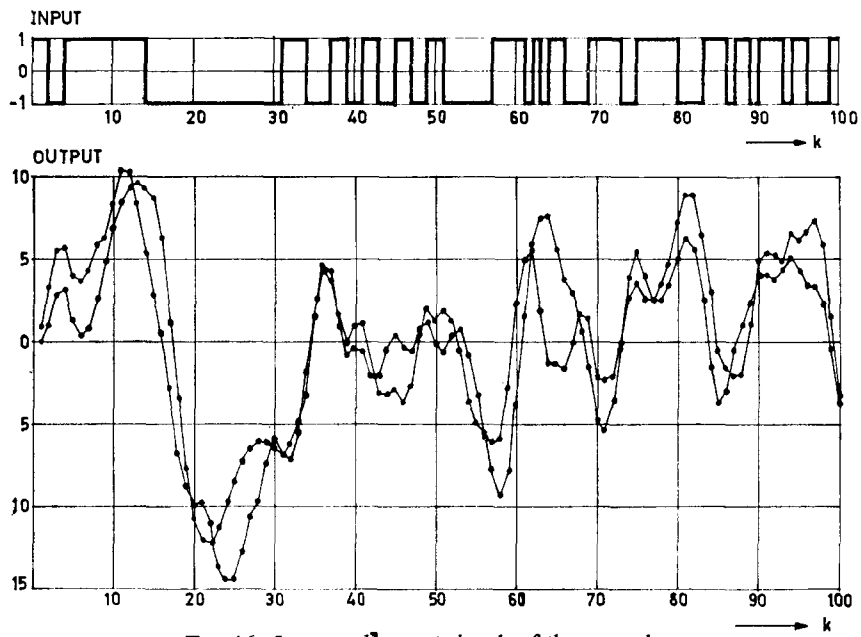

FIG. 16. Input and output signals of the example.

where $\{e(k)\}$ is a sequence of independent normal (0, 1) random numbers generated by a pseudo random generator. The following values of λ have been used: 0, 0·1, 0·5, 1·0 and 5·0.

In Table B.1 we show the results obtained when a model having the structure (B.1) is fitted to the generated input–output data using the least squares procedure.

The estimates are calculated recursively for models of increasing order to illustrate the very typical situation in practice when the order of the model is not known. In Table B.1 we have shown the least squares parameter estimates and their estimated accuracy, the loss function and a conditioning number of the matrix $\Phi'\Phi$. The conditioning number $\mu=2n\{\max(A)_{ij}\}\max\{A^{-1})_{ij}\}$ is chosen rather arbitrarily.

We will now analyse the result of Table B.1. Let us first consider the case of no disturbances $\lambda=0$. In this case we find that it is not possible to compute the third order model, we find that the matrix $\Phi'\Phi$ is singular as would be expected. The conditioning number is $1{\cdot}3\times10^6$. We also find that the estimated standard deviations of the second order model are zero.

To handle the numerical problems for a model of third order in a case like this, we must use numerical methods which do not require the inversion of $\Phi'\Phi$ e.g. the reduction of Φ to triangular form using a QR algorithm. A reduction based on a sequence of plane rotations has been pursued by PETERKA

TABLE B.1. LEAST SQUARES ESTIMATES OF THE PARAMETERS

	Model order	$-\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\lambda}$	V	μ
True parameter		1·50	0·70	0·0	1·0	0·5	0·0			
$\lambda=0{\cdot}0$	1	0·88 ± 0·03			1·23 ± 0·18			1·71	265·863	59
	2	1·50 ± 0·00	0·70 ± 0·00		1·00 ± 0·00	0·50 ± 0·00		0·00	0·000	203
	3	—	—	—	—	—	—	—	—	$1{\cdot}3\times10^6$
$\lambda=0{\cdot}1$	1	0·88 ± 0·03			1·19 ± 0·18			1·66	248·447	60
	2	1·50 ± 0·01	0·69 ± 0·01		0·99 ± 0·01	0·49 ± 0·02		0·11	0·987	205
	3	1·52 ± 0·11	0·73 ± 0·16	0·02 ± 0·08	0·99 ± 0·01	0·47 ± 0·11	−0·02 ± 0·08	0·11	0·983	35974
$\lambda=0{\cdot}5$	1	0·88 ± 0·03			1·10 ± 0·17			1·59	227·848	63
	2	1·48 ± 0·04	0·67 ± 0·03		0·96 ± 0·06	0·48 ± 0·07		0·53	24·558	206
	3	1·54 ± 0·11	0·76 ± 0·16	0·04 ± 0·08	0·96 ± 0·06	0·42 ± 0·12	−0·04 ± 0·09	0·53	24·451	1518
$\lambda=1{\cdot}0$	1	0·88 ± 0·03			1·00 ± 0·20			1·84	308·131	75
	2	1·47 ± 0·06	0·66 ± 0·06		0·93 ± 0·12	0·46 ± 0·14		1·06	99·863	212
	3	1·55 ± 0·11	0·82 ± 0·17	0·09 ± 0·09	0·93 ± 0·13	0·37 ± 0·16	−0·02 ± 0·15	1·07	98·698	476
$\lambda=5{\cdot}0$	1	0·86 ± 0·05			0·20 ± 0·83			7·55	5131·905	520
	2	1·48 ± 0·07	0·74 ± 0·08		0·98 ± 0·61	0·41 ± 0·61		5·29	2462·220	1174
	3	1·55 ± 0·11	0·87 ± 0·18	0·09 ± 0·09	0·97 ± 0·64	0·24 ± 0·66	0·13 ± 0·64	5·32	2440·245	2031

and ŠMUK [104]. They have shown that this calculation can also be done recursively in the order of the system.

Proceeding to the case of $\lambda=0{\cdot}1$, i.e. the standard deviation of the disturbances is one tenth of the magnitude of the input signals, we find that the matrix $\Phi'\Phi$ is still badly conditioned when a third order model is computed.

Analysing the details we find, however, that the Gauss Jordan method gives a reasonable accurate inverse of $\Phi'\Phi$. Pre- and postmultiplying the matrix with its computed inverse, we find that the largest off-diagonal element is 0·011 and the largest deviation of diagonal elements from 1·000 is 0·0045. We also find that the estimates $\hat{\alpha}_3$ and $\hat{\beta}_3$ do not differ significantly from zero.

We will also discuss some other ways to find the order of the system. We can e.g. consider the variances of the parameters. We find e.g. from Table B.1 that the coefficients $\hat{\alpha}_3$ and $\hat{\beta}_3$ do not differ significantly from zero in any case. In Table B.2 we also summarize the values of the loss function as well as the values of the test variable (58) when testing the reduction of the loss function for a model of order n_1 compared to a model of order n_2 as was discussed before. We have

$$F(2,\ 100)=3{\cdot}09\Rightarrow P\{t>3{\cdot}09\}=0{\cdot}05.$$

Hence at the 5 per cent risk level we find in all cases that the loss function is significantly reduced when the order of the system is increased from 1 to 2 but that the reductions in the loss function by a further increase of the loss function are not significant.

We thus find that by applying the F-test in this case we get as a result that the system is of second order for all samples. The actual parameter values of Table B.1 as well as the estimated accuracies give an indication of the accuracy that can be obtained in a case like this.

It should, however, be emphasized that when the same procedure is applied to practical data the results are very seldom as clearcut as in this simulated example. See e.g. GUSTAVSSON [100].

TABLE B.2. GIVES THE VALUES OF THE LOSS FUNCTION V, THE CONDITIONING NUMBER μ OF $\Phi'\Phi$ AND A TABLE OF t-VALUES WHEN IDENTIFYING MODELS OF DIFFERENT ORDER TO THE EXAMPLE OF FIG. 16

case 1 $\lambda=0$			n_2		
n	V	μ	n_1	1	2
0	2666·23		0	442·4	∞
1	285·86	59	1		∞
2	0·00	205			

case 2 $\lambda=0{\cdot}1$			n_2			
n	V	μ	n_1	1	2	3
0	2644·15		0	472	57000	42100
1	248·447	60	1		12000	5900
2	0·587	205	2			0·191
3	0·983	35974				

case 3 $\lambda=0{\cdot}5$			n_2				
n	V	μ	n_1	1	2	3	4
0	2701·35		0	532	2616	1715	1283
1	227·848	63	1		397	195	130
2	24·558	206	2			0·2	0·1
3	24·451	1518	3				0·8
4	24·006	2982					

case 4 $\lambda=1{\cdot}0$			n_2					
n	V	μ	n_1	1	2	3	4	5
0	3166·78		0	455	734	487	365	292
1	308·131	212	1		100	50	33	25
2	99·863	476	2			0·55	0·72	0·80
3	98·698	873	3				0·88	0·92
4	96·813	1351	4					0·96
5	94·800	1647						

case 5 $\lambda=5{\cdot}0$			n_2					
n	V	μ	n_1	1	2	3	4	5
0	21467·64		0	156	185	122	92	75
1	5131·905	520	1		52	26	18	14
2	2462·220	1174				0·21	0·94	1·2
3	2440·245	2031	3				1·2	1·5
4	2375·624	2910	4					1·6
5	2290·730	3847						

Résumé—Le domaine de l'identification et de l'estimation des paramètres de processus s'est rapidement developpé durant la décade passée. Dans cette revue, l'état de l'art de cette science est présenté d'une manière systématique. Une attention est accordée aux proprietés générales et à la classification des problèmes d'identification. Les structures des modèles sont discutées; leur choix depend du but de l'identification et des informations disponibles à priori. Pour l'identification des modèles qui sont linéaires dans les paramètres, la revue explique la méthode des moindres carrés et plusieurs de ses variantes qui peuvent resoudre le problème des résidus correlés, telles que les moindres carrés iterés et généralisés, la méthode de vraisemblance maximale, la méthode de la variable influente, le principe de pointage. Rebemment, la situation non-linéaire, l'identification dans la boucle et l'identification en temps réel ont donné lieu à des developpements considerables qui sont décrits d'une manière cohérente.

Plus de 230 références sont données, la plupart en ce qui concerne des publications récentes. Un appendix contient un résumé des principes d'estimation de paramètres et un autre donne un exposé plus détaillé d'un exemple d'une estimation au moyen des moindres carrés.

Zusammenfassung—Das Gebiet der Identifikation und Prozeßparameterschätzung entwickelte sich im letzten Jahrzehnt sehr schnell. In dieser Übersicht ist der Stand dieses Wissenschaftsweiges systematisch dargestellt. Aufmerksamkeit wurde allgemeinen Eigenschaften und der Klassifikation der Identifikationsprobleme geschenkt. Modellstrukturen werden diskutiert; ihre Wahl dreht sich um den Zweck der Identifikation und um die verfügbare a

priori-Kenntnis. Für die Identifikation von Modellen, die die Methode der kleinsten Quadrate und verschiedene ihrer Varianten, die das Problem der korrelierten Gleichungsfehler, nämlich wiederholter und verallgemeinerter Methoden der kleinsten Quadrate, Maximum likelihood Methode, "tally"—Prinzip charakterisieren. Neuerdings legte die nichtlineare Situation, die on-line und die real-time Identifikation ausgedehnte Entwicklungen nahe, über die zusammenhängend berichtet wird. Mehr als 230 Literaturstellen sind angegeben, die sich meist auf neue Arbeiten beziehen. In Anhängen wird eine Zusammenfassung der Prinzipien zur Parameterschätzung und ein genauer behandeltes Beispiel einer Schätzung nach der Methode der kleinsten Quadrate angeführt.

Резюме—Область опознавания и оценки параметров процессов быстро развилась в течении прошлого десятилетия. В настоящем обзоре систематически представлено настоящее состояние этой науки. Уделяется внимание общим свойствам и классификации проблем опознавания. Обсуждаются структуры моделей; их выбор зависит от цели опознавания и от заранее доступных сведений. Для опознавания моделей линейных в параметрах, обзор объясняет метод наименьших квадратов и некоторые из его вариантов могущих решить проблему коррелированных остатков, как на пример повторительные наименьшие квадраты, метод наибольшего правдопододия, метод влияющей переменной, принцип подсчёта. Недавно нелинейное состояние, опознавание в контуре и опознавание в действительном времени привели к значительным развитиям которые описываются систематическим образом.

Приведено свыше 230 референций относящихся главным образом к недавным публикациям. Приложения содержат сводку принципов оценки параметров и более детальное описание примера оценки при помощи наименьших квадратов.