INTRODUCTION: AN ANNOTATED NAVIGATION CHART AND SKETCH OF THE MAIN THEMES IN THIS VOLUME AND THEIR INTERRELATIONS.

Michiel Hazewinkel
Dept. Math., Erasmus Univ.
Rotterdam,
P.O. Box 1738,
3000 DR Rotterdam
The Netherlands

Jan C. Willems
Dept. Math., Univ. of Groningen
P.O. Box 800,
9700 AV Groningen
The Netherlands

This volume is devoted to modeling and analysis of uncertain dynamical systems in an uncertain environment and the synthesis of filters, identifiers and adaptive controllers in such a setting. All this with particular emphasis on recursive (and/or on-line) techniques.

This is a large and varied field of inquiry. It was the intention of the conference, of which this volume constitutes the proceedings, to review the most important themes and new developments in a coherent manner without making too many demands on the audience in the matter of prerequisites.

As a result this volume contains tutorial material, reviews and surveys, as well as research papers on the topics of modeling, adaptive control, identification and filtering and applications. The present introduction is intended to provide an informal outline of the main themes of the volume: identification and filtering and recursiveness, and to indicate how the various contributions fit together. That is, it is essentially an (annotated) navigation chart. We have concentrated mostly on the

M. Hazewinkel and J. C. Willems (eds.), Stochastic Systems: The Mathematics of Filtering and Identification and Applications, 3–26.

tutorial and the invited survey-and-state-of-the-art papers (marked with a *) or **) in the table of contents).

1. THE SETTING AND THE BASIC THEMES.

An uncertain dynamical system may be defined as a map F from an input space $\mathcal{U}$ (which is a family of maps from the time axis $T \subset \mathbb{R}$ to the space of input values U) and an uncertainty space N to an output space $\mathcal{Y}$ (which is a family of maps from T to the space of output values Y) which is nonanticipating, that is to say that for all values of the uncertainty parameter n the output y is independent of future values of the input u. The uncertain system under consideration is often called the plant and is depicted by the following signal flow diagram (fig. 1):

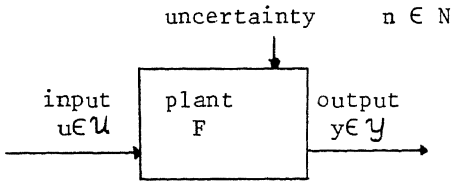


figure 1

We think of the inputs as variables which can be manipulated (controls) or, more generally, through which the environment can influence the system; we think of the outputs as variables which can be measured (observations) or, more generally, through which the system can influence the environment. The uncertainty reflects the fact that the dynamic behaviour is unknown (for example because the numerical value of a parameter is not known) or that it may depend on a stochastic phenomenon. We think of the uncertainty as a parameter n being choosen by 'nature'.

For the purpose of the contributions in this volume it is in fact insightful to assume that the uncertainty space N is a product space, $N = P \times R$ with P a set of unknown parameters and R the outcome space of a general random variable. Formally, there

is a probability space $\{\Omega,A,P\}$ and a map $\tau : \Omega \to R$ which selects the value in R in a random fashion. Finding out from measurements the actual value (or 'best' approximation) of the unknown parameter $p \in P$ is the basic problem in system identification while finding the actual value (or 'best' estimate) of the random parameter (vector) $r \in R$ is (indirectly) the basic problem in filtering. Preferably one wants to do this in a recursive manner that is, roughly, by a technique which updates a 'state-type' parameter vector η_t by means of the new information gathered at time t while the desired unknown parameters $p \in P$ or $r \in R$ are calculated as (known) functions of η_t. The vector η_t, so to speak, embodies or codifies all the useful information gathered up to time t.

2. MODELING ISSUES.

The study of stochastic dynamic systems brings with it the problem of modeling, particularly if one wants to use differential equation models. The reason why one wants to use such models is, as in the deterministic case, connected with the fact that one much prefers, for good (computational) reasons and also from a basic mathematical point of view to use recursive models, that is, models which display the state of the system explicitely. In a stochastic framework the idea of state leads to modeling in terms of a Markov process (since in general there is also an input we should really think of a controlled Markov process). Writing down the evolution of a Markov process leads to differential equations with a white noise term on the right hand side and the rigorous interpretation of such equations leads to Itô calculus.

An Itô equation is a differential equation of the form

$$(1) \qquad dx = f(x)dt + g(x)dw \qquad x(t_o) = x_o$$

with $x \in \mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}^n$, $g: \mathbb{R}^n \to \mathbb{R}^{nxm}$ (the nxm-matrices), w an $\mathbb{R}^m$-valued Wiener stochastic process, and $x_o \in \mathbb{R}^n$ a random vector. Assume that w and x_o are defined on the probability space (Ω,A,P). The above equation can be thought of intuitively as the equation

$$\dot{x} = f(x) + g(x)\frac{dw}{dt} \qquad x(t_o) = x_o$$

(at least as long as w_1 is scalar valued) with $\frac{dw}{dt}$ 'white noise'. This, however, is not a process defined in the convential way. The rigorous interpretation of equation (1) is made in terms Itô calculus and is the subject of CURTAIN's tutorial [section 2.2 in this volume]. Under suitable assumptions, explained in the tutorial, (1) yields a well-defined Markov process x. We may add inputs and outputs to (1) which leads to the usual form of a stochastic differential system given by:

$$\begin{aligned} dx &= f(x,u(t))dt + g(x,u(t))dw \qquad && x(t_o) = x_o \\ dy &= g(x)dt + dv && y(t_o) = 0 \end{aligned} \tag{2}$$

where v is a stochastic Wiener process assumed to be independent of w. The noises w and v are respectively called the system noise and the output noise. (Problems where the system noise w and the output noise v are dependent are of interest but are usually not given much attention ; cf., however, e.g. section 7.3 in this volume; this introduces fundamental extra difficulties). Model (2) leads then to an uncertain dynamical system of the type informally discussed in section 1, with uncertainty random variables.

Two 'case studies' of modeling of stochastic systems are contained in part 3 of this volume, both taken from areas where

there is a great deal of recent activity in applied mathematics The first one of these contributions is by BOEL [section 3.3] and describes how one may set up stochastic models of computer networks. The models proposed are in terms of queues and contrary to (1) involve mainly Poisson noise. An interesting feature in the analysis of these models is the important role played by 'quasi-(time) reversibility'.
The second paper about modeling is by ARNOLD [section 3.1] and treats chemical reactions. Such reactions show irregularities in space and one can consider the local behaviour versus the global behaviour where one expects to be able to derive some type of space average behaviour. Chemical reactions also have a stochastic feature due to the fact that particles react when they 'meet' which is modeled as a random phenomenon. The purpose of ARNOLD's paper is to demonstrate how global deterministic models may be viewed as suitable limits of global stochastic models or of local deterministic models both of which may in turn be viewed as a limit of a local stochastic model.

One of the important issues in mathematical control theory is the realization theory problem. This means essentially the realization (or modeling) of a given input/output operator by means of a 'machine' of type (2). It also means the construction of a stochastic process of a certain type with a pregiven covariance function.

As we have already mentioned many applications (in fact most of those discussed in this volume as Kalman filtering and nonlinear filtering) need, in order to carry out the required calculations, a model in state space form. Often, one starts with a model in input/output form - some model of the type introduced in section 1 - and the question then arises how to construct an equivalent state space model. In the context of random processes, this problem becomes the following: Let $y(t)$, $t \in T \subset \mathbb{R}$ be a given stochastic process with outcome

space Y. The problem then is to construct a space X, a Markov process $x(t)$, $t \in T \subset \mathbb{R}$ with outcome space X, and a map $h: X \to Y$ such that $h(x(t))$ is in some sense equivalent to the original process $y(t)$. In the paper by LINDQUIST and PICCI the realization theory for multivariate stationary gaussian stochastic processes is presented.

In addition part 3 of this volume contains two papers on the more qualitative properties of stochastic differential equations

$$\dot{x}(t) = f(x(t),\xi(t)),\ x(0) = x_o$$

$$\dot{x}(t) = f(x(t),\ \xi(t)),\ x(0) = x_o$$

with random initial condition x_o and $\xi(t)$ a random process. Here solutions are to be interpreted pathwise, i.e. this equation is really a collection of equations, one for each noise trajectory (and initial condition).

The paper by Arnold [section 3.2] is a survey in extended abstract form of problems, concepts and results of the qualitative theory of such equations. Qualitative concepts include such things as stationary solutions, attractors, stability and ergodicity. This last topic is the subject of the paper by Wihstutz. Obviously something like ergodicity for instance is of relevance when discussing the compatibility between local (micro) stochastic models and global (average) deterministic models. Think of statistical mechanics.

3. NONLINEAR FILTERING.

The filtering problem takes up by far the largest part of this volume. In abstract terms the filtering problem is a stochastic version of an obtimal observer design problem. Take an uncertain plant as introduced in section 1, and make, to simplify the discussion, the (inessential) restriction that there are no inputs. Assume furthermore that there are two types

of outputs: one output, which we will denote by y, which is a signal which can be measured - the observations - and another output, which we will denote by z, which is a signal which we would like to know - the to-be-estimated output. These outputs take on their values in a space Z; often z = x, the state of the plant processor, which accepts as inputs the observations y and produces as outputs estimates $\hat{z}$ of z. Formally we have a plant (F_y,F_z): N → Y x Z and we wish to construct a nonanticipating map K: Y → Z such that, in some sense, $\hat{z} = Ky = K_y^F(n)$ is close to $z = F_z(n)$ (see Figure 2). Expressing

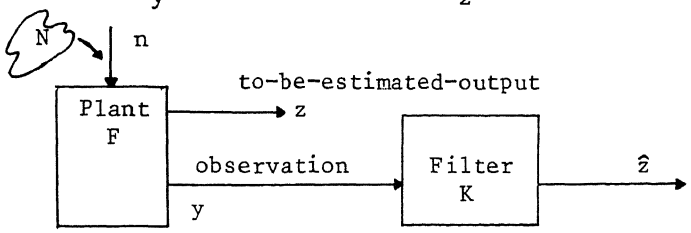
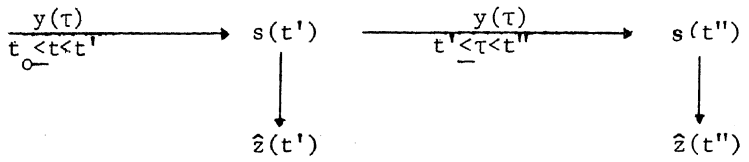


figure 2

'being close to' in terms of a loss functional and assuming the uncertainty to be a random variable it is natural to express this problem in terms of the minimization of the average loss $E\{d(z,\hat{z})\}$. It is furthermore clear that one can formulate this minimization for all times t ∈ T which leads to the problem of finding, for all t, a K_t: Y → Z which minimizes $E\{d(z(t))\}$, where d is an appropriate distance function. Now, since one wants to obtain this estimate $\hat{z}(t)$ for all t, it is very natural and advantageous to attempt to do this computation recursively. This is done by trying to find a 'state' of the observer s such that the computation of $\hat{z}$ may be carried out according to the diagram:



Storing $s(t')$ instead of $y(\tau)$ for $t_o \leq \tau < t'$ will hopefully also lead to an automatic data reduction which could be very advantageous from the point of view of computational complexity and memory storage requirements.

Let us be a little more specific. Assume that in continuous time we have the Itô equation

$$\begin{aligned} dx &= f(x)dt + g(x)dw & x(t_o) &= x_o \\ dy &= h(x)dt + dv & y(t_o) &= 0 \\ z &= k(x) \end{aligned} \tag{3}$$

with $x \in X := \mathbb{R}^n$, $y \in \mathbb{R}^p$, $z \in \mathbb{R}^q$, and v and w mutually independent Wiener processes and independent of the initial randomness $x_o \in \mathbb{R}^n$. Assume that we want to obtain the best estimate in the quadratic sense of $z(t)$ based on observations $y(\tau)$ for $t_o \leq \tau \leq t$. This is the filtering problem. The prediction problem asks for the best estimate of $z(s)$ given $y_\tau, t_o \leq \tau \leq t$, $t < s$ and the smoothing problem asks for the best estimate of $z(s)$ given $t_o \leq \tau \leq t$, $s < t$, i.e. given also future observations. It is wellknown that the conditional expectation $z^*(t) := E\{z(t) \,|\, y(\tau),\ t_o \leq \tau \leq t\}$ is the best estimate in the leastsquares sense, i.e. it minimizes every quadratic loss of the form $E\{||z(t)-\hat{z}(t)||^2\}$. The filtering problem is then to give a (recursive) algorithm for computing this conditional expectation.

Because of the special structure of the system (3), in particular, because of the Markov property of x, it follows that

the conditional distribution $\pi_t := p(x(t)|y(\tau),\ t_o \leq \tau \leq t)$ can act as a state for the filter. That is to say that there exists an update equation of the type

$$
\begin{aligned}
d\pi &= A_1(\pi)dt + B_1(\pi)dy(t) \\
z^*(t) &= \int_X k(x)\pi(t)dx
\end{aligned}
\tag{4}
$$

with $\pi(t_o)$ = the distribution of x_o. Since $x \in \mathbb{R}^n$, π is a function on $\mathbb{R}^n$ and hence one may expect that (4) will be a type of partial differential equation. In fact A_1 and B_1 are integro-differential operators on X.

In the tutorial article of DAVIS & MARCUS [section 2.3] this equation and the rigorous derivation of it is discussed together with the role of the so-called Duncan-Mortensen-Zakai-equation which is an unnormalized version of (4). That is, instead of having an update equation for $\pi(t)$, the D-M-Z equation computes a function $\rho(t)$ with the property that $\pi(t)$ is related to it by a simple formula of the type

$$\pi(t) = \frac{\rho(t)}{\int_X \rho(t)dx}$$

Working with $\rho(t)$ has certain advantages: ρ satisfies a much simpler looking equation than π. The equation of ρ is a stochastic partial differential equation:

$$d\rho = A_2\rho\ dt + h\rho\ dy(t) \tag{5}$$

from which $z^*(t)$ is calculated by means of the (output) map

$$z^*(t) = \left(\int_X \rho(t)dx\right)^{-1} \int_X k(x)\rho(t)dx \tag{6}$$

Here A_2 is a suitable linear differential operator defined in DAVIS-MARCUS [section 2.3]. This is a bilinear equation in the sense that ρ satisfies a linear equation in which the driving term is a linear function of the 'input' y.

This bilinear structure of the Zakai-equation is very much exploited by BROCKETT [section 7.1] in his expository article in which he explains the geometric structure of the Zakai-equation, with an eye towards finding conditions for the existence of finite-dimensional filters.

The issue of the finite-dimensionality of the filter receives a great deal of attention in this volume. Let us explain in an informal way what this fuss is all about. Consider equation (4) or (5). This defines (the filtering problem was precisely set-up this way) a non-anticipating map from the observation y which acts as inputs to the filter to produce estimates z^* which are the outputs of the filter. Now (4) and (5) are realizations of this map, but they are infinite dimensional realizations because the state $\pi(t)$ or $\rho(t)$ is a map from $X = \mathbb{R}^n$ to $\mathbb{R}$, i.e. it is an infinite dimensional object (a function space). Now, it may be the case that this filter (input/output map) admits a finite dimensional realization.

This means that there would be a finite dimensional manifold M and a differential equation with output map

$$\dot{m} = v(m,\dot{y},t), \quad z^* = w(m) \tag{7}$$

on it such that (7) defines the same input/output map as (5) and (6). Obviously finite dimensionality of a filter is a very desirable (if not necessary) feature if one actually wants to implement it.

Thus assuming that a finite dimensional machine for calculating z^* (a filter) exists we would have two equivalent ways for processing the data y_s, $0 \leq s \leq t$ to produce $z^*(t)$).

The finite dimensional machine can be assumed to be of minimal dimension and assuming this one expects that there exists a map from (the from ρ_o or π_o accessible part of the function space) to the manifold M which takes the evolution equation for ρ_t (or π_t) to the equation for m. (This infinite dimensional extension of a result of Sussmann still has to be proved; it seems now very likely to be true in one sense or another). In the case of ρ_t there would result a filter of the form

$$\dot{m} = \alpha(m) + \beta(m)\dot{y}_t, \quad z^* = \gamma(m) \tag{8}$$

where $\alpha(m)$ and $\beta(m)$ are vectorfields on the manifold M.

It is also definitely not unreasonable to look for a filter of the form (8) because (for linear systems) the Kalman-Bucy filter of considerable fame and enormous applicability is precisely a machine of the form (8). And so is for that matter the extended Kalman filter.

A main tool in this analysis is the Lie algebra of operators generated by the two operators A_2 and 'multiplication by h' which occur in the equation (5). This Lie algebra is called the estimation Lie algebra. The necessary differential topology and Lie-algebra background material for all this can be found in the tutorials of Hazewinkel [sections 2.4 and 2.5].

One particular most interesting feature of the estimation Lie algebra of a system (5) is that it is intrinsic. That is, it is (up to isomorphism) invariant under (nonlinear) changes of coordinates (cf. Brockett's lectures [section 7.1]). As such it could help e.g. in recognizing that a certain highly non-linear looking system is in fact a linear system to which a nonlinear change of coordinates has been applied. This Lie algebraic criterion will not be a sufficient, though, e.g. because the estimation Lie algebra is also invariant under socalled Gauge transformations, which do not correspond to coordinate changes.

One consequence of the existence of a map as discussed just above equation (8) above is the existence of a homomorphism of Lie algebras from the estimation Lie algebra to the Lie algebra of vectorfields on M generated by the vectorfields α and β in (8). In the particular case of linear systems and the Kalman-Bucy filter this can be checked by hand (Brockett [section 7.1]). Thus finite dimensional exact filters give rise to certain homomorphisms of Lie algebras and as a matter of fact there is evidence for a reverse statement as well. One collection of results which we shall need for this are uniqueness existence and regularity results for stochastic partial differential equations of the type (5), which is the subject of the contributions by Michel (section. 7.12) and Sussmann (section 7.14) cf. also Pardoux (section 7.4) (Michel uses the socalled Malliavin Stochastic variational calculus (currently a hot topic which was the subject of a conference in Durham later in 1980); additional or similar results on existence, regularly, uniqueness will probably result from the variational path integral formulation of Fleming and Mitter discussed in [section 7.2]). Given these one can exploit certain theorems concerning Lie algebras discussed in Hazewinkel - Marcus [section 7.9] to conclude e.g. that there exist no finite dimensional exact filters for any nonconstant statistic of the socalled cubic sensor. Though some of the things mentioned above are still conjectural this is now a firm theorem. Indeed it seems likely that we shall be able to prove that as a rule finite dimensional exact filters will not exist, which brings us to approximate calculation devices, a topic to which we shall return below.

Meanwhile there is obvious interest in analysing the estimation algebra in various cases. Finite dimensionality of this algebra would be nice to have and this is the topic of Ocone [section 7.13], though of course a Lie algebra of

vectorfields on a finite dimensional manifold need not be finite dimensional. Low dimensionality of the estimation algebra and ease of computation ought to be related, cf. Baillieul [section 7.5] and the question whether similar estimation algebras correspond to filtering problems of equal computational complexity is addressed by Baras [section 7.6]. It is perhaps too early in the game to say just how useful the estimation Lie algebra and its concomitant geometrical considerations will be in the actual construction of (approximate) filters. (Its intrinsic nature exerts of course a powerful appeal and the writers of the present words are quite optimistic in this regard). Meanwhile, however, these geometrical ideas have certainly helped our theoretical understanding and have also helped in the actual construction of unexpectedly low dimensional filters (for finite state Markov chains, cf. Brockett [section 7.1]).

In our informal exposition of the nonlinear filtering problem we have up to now skipped over an important point or rather several much related points. Equations (4) and (5) are stochastic differential equations. This implies that abstractly they define a map from the probability space Ω to the observations y and then via the non-anticipating filter map to the optimal estimates z^*. However, from the construction of stochastic integrals it follows that in principle these maps depend on the probability measure on Ω. This is, of course, an unpleasant situation since it says that we cannot just consider the filter map as simply acting on realizations of the observation process, in other words the filter map does not act (necessarily) 'sample pathwise'. In DAVIS' contribution [section 7.3] it is shown that in a large class of filtering problems one can in fact prove that the filter acts indeed sample-pathwise.

There is a second point, much related, as it turns out, to the first. The conditional expectation

$z^* = E[k(x)|y^t] = \int k(x)\pi_t(x)dx$ is a functional of y_t. I.e. given by some function ϕ which is only determined up to sets of measure zero (with respect to the measure on the function space $C([0,T])$ induced by the y and this measure has the same sets of measure zero as Wiener measure. Since the set of functions of bounded variations has measure zero ϕ is so to speak undefined on these. However, physical observation paths will be of bounded variation and so this approach to filtering would seem to be inapplicable unless there exists a version which is e.g. continuous w.r.t. the supremum norm on $C([0,T])$, giving us a 'robust' form of the filter (Robustness is, roughly, the property of a statistical procedure, or observer, or model, or ... to perform well even when the assumptions underlying its construction are not fully met). This fortunately turns out to be the case if the observation noises are independent of the system noises and also more generally provided the output y_t is scalar. The issue is much related to the pathwise issue discussed above; cf. Davis [section 7.3].

This robust-pathwise approach goes via a Feynman-Kac formula and thus suggests links with the path-integral approach to Quantum mechanics (à la Nelson). Another interesting and stimulating observation in this respect is that the estimation Lie algebra of the simplest (nonzero) linear system $dx = dw_t$, $dy_t = x_t dt + dv_t$ is the four dimensional oscillator Lie algebra (of some fame), whose derived Lie algebra is the even more famous Heisenberg Lie algebra of the canonical quantum mechanical commutation relations. And indeed the Kalman-filter for this system turns out to be gauge equivalent to a forced (euclidean)harmonic oscillator. The deep and fundamental relations of (nonlinear) filtering with quantum theory of which the two observations above are symptomatic are the subject of Mitter [section 7.2].

As the quantum constant h goes to zero quantum mechanics

goes to deterministic mechanics and one may ask to what deterministic limit nonlinear filtering converges if the noise intensity goes to zero. This matter is discussed in Hijab [section 7.10].

Both the estimation algebra approach and the robust-pathwise approach offer approximation possibilities. For the former some speculations are offered in Hazewinkel-Marcus [section 7.9]. Approximation by continuous time Markov chains is the subject of Di Masi-Runggaldier [section 7.8]. As soon as one starts approximating the question of a priori lower and upper bounds on the errors arises and whether these bounds are perhaps attained asymptotically. This is discussed by Bobrovsky-Zakai [section 7.7]. Finally Le-Gland uses the nonlinear filtering equations (and robustness) as an approach to maximum likelihood estimation for an astronomical observation problem.

So far, in this section we have concentrated on the filtering problem, neglecting the closely related and equally interesting prediction and smoothing problems. Pardoux [section 7.4] discusses the matter of finding DE's driven by the observations for optimal smoothers and predictors by means of a novel method involving both backward and forward equations. (The latter is the Duncan-Mortensen-Zakai-equation). This also yields new results for the smoothing problem extending the known results for finite state Markov processes.

4. LINEAR FILTERING.

Of course, there is one case in which the filtering problem may be solved explicitely, namely where the maps f, g, h, and k of (3) are linear, which leads to the model

$$dx = F(t)x\,dt + G(t)dw \qquad x(t_o) = x_o \text{ (gaussian)}$$
$$dy = H(t)x\,dt + I(t)dw \qquad y(t_o) = 0$$
$$z = K(t)x$$

with F,G,H and I matrices of suitable dimension. The solution of the filtering problem in this case is given by the celebrated Kalman-Bucy filter. These filtering equations are very wellknown and play an important role in some of the other papers of these volume. The tutorial by WILLEMS [section 2.1] gives a brief introduction to the Kalman filter in the context of the general LQG (linear-quadratic-gaussian) stochastic control problem.

If one drops the assumption that there is a state-space model the filtering, smoothing and prediction problems take the following form. We have the following model for the observed process y_t

$$y_t = z_t + v_t$$

where z_t is the (stationary) signal and v_t is white random noise. The smoothing problem, filtering problem and prediction problem now take the form: find the best estimate $\hat{y}_{t|\tau}$ given observations up to and including time τ where respectively $\tau > t$ (smoothing), $\tau = t$ (filtering), $\tau < t$ (prediction). This is the problem studied and solved by Wiener and Kolmogorov in the early forties. The techniques involved in this solution, their extension to the case of finite time interval observations and associated problems of (efficient) computation are discussed in Kailath [section 5.1] (Wiener-Hopf technique, Ambartzumian-Chandrasekhar equations, Krein-Levinson equations). Kailath then goes on to discuss an extension to nonstationary models and a scattering theory framework for linear estimation.

Now scattering theory compares the asymptotic behaviour of an evolving system as $t \to -\infty$ with its asymptotic behaviour as $t \to \infty$. It is especially relevant when comparing the behaviour of a reference system (no scattering object) with that of a perturbed system (a scattering object is present) when the perturbations are negligible for large $|t|$. Think e.g. of a

wave packet traveling from left to right being scattered by some object at the origin. Let U^t and U_o^t denote the evolution operators giving the state of the system at time t in terms of the state at time 0 for the perturbed and unperturbed system respectively. Then there are two states x_+ and x_- of the unperturbed system such that $U^t x$ behaves as $U_o^t x_-$ for $t \to -\infty$ and $U_o^t x_+$ for $t \to +\infty$. The scattering operator is the mapping $S: x_- \to x_+$ and the inverse scattering problem is the reconstruction of the scatterer from the scattering operator.

The relation of inverse scattering with linear prediction is the main theme of DeWilde, Fokkema en Widya [section 5.2]. Here, as in Kailath, the 'scatterer' is a transmission line with incident and reflected (light) waves from both sides. DeWilde e.a. first discuss (Redheffer) scattering, then the main theoretical result which says that the predictor filter may be obtained by solving a (very special) inverse scattering problem and then proceed how this fact can be used to produce concrete algorithms.

As was mentioned above (in the section on nonlinear filtering) there are links between quantum theory and the Duncan-Mortensen-Zakai-equation-approach to state-space-model filtering. This is not the first time that links between filtering problems and quantum theory have appeared. In fact, in a Seminaire Bourbaki exposé in 1961 Cartier discusses how a certain number of results of the spectral theory of Wiener and Kolmogorov filtering can be grouped around the ideas related to the Stone-von Neumann uniqueness theorem on representations of the Heisenberg Lie algebra (canonical commutation relations), and how the Wiener-Kolmogorov theory can be deduced from this point of view. This was the subject of the lectures by Hazewinkel [section 5.3].

In this connection it is interesting to observe that Wiener-Kolmogorov filtering can be viewed as a limit of Kalman-Bucy filtering and that on the other hand a main result of scattering

theory (the translation representation theorem) is in fact equivalent to (the Weyl form of) the Stone-von Neumann theorem. There seems to be room for future work here.

5. IDENTIFICATION.

In the context of Section 1, the identification problem typically arises in a context where an uncertain system has, in addition to a stochastic component, also an uncertain non-stochastic 'parameter'. The basic problem is then to find out from measurements of the input and the output variables what the value of this unknown parameter is. There are, of course, more general situations where one may use identification ideas. For example one could try to fit a linear model to a nonlinear plant or one could try to fit a low dimensional linear model to a (very) high dimensional linear plant. In these cases it is not really fair to say that one tries to determine the unknown parameters of the plant. However, for the purposes of the present discussion, it suffices to think of the identification problem in this simple minded context.

Let us denote the unknown parameter(s) by Θ
If the input used is u then we will observe $y = F(u,\Theta,\omega)$ which, of course, will in general also depend on the parameters Θ and the random element $\omega \in \Omega$. In a dynamic situation it is natural to introduce also the time $t \in T$. At each instant one will then have available the past of u and y and an identification scheme will give us an estimate $\hat{\Theta}(t)$ of Θ(see figure 3).

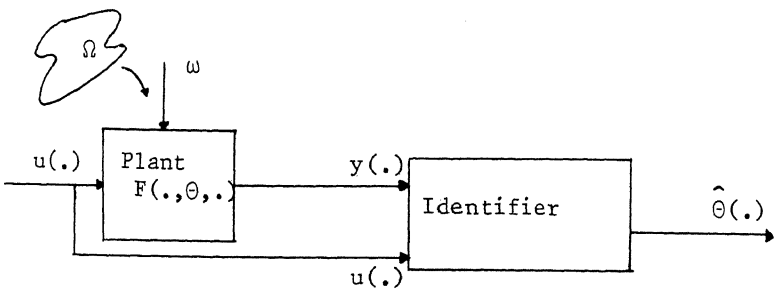


figure 3

There are two basic issues which are discussed in this volume regarding identification:

1. conditions for convergence of $\Theta^*(t)$ to the true parameter

2. recursive implementation of identification algorithms.

The article by HANNAN [section 4.1] contains a general convergence result for a class of identification problems. The model considered is a discrete time multivariable ARMAX (= autoregressive moving average with exogenous components) model which relates the input and output by

$$(10) \qquad \sum_{j=o}^{p} A(j)y(t-j) = \sum_{j=1}^{m} D(j)u(t-j) + \sum_{j=o}^{q} B(j)\varepsilon(t-j)$$

where one can think of ε as white random noise (the precise assumptions are given in the paper) and $A(0), \ldots, A(p)$, $D(1), \ldots, D(m)$, and $B(0), \ldots, B(q)$ as matrices with unknown coefficients. Let us denote this string of matrices by Θ. In this case 4 is thus a high dimensional Euclidean space. The identification principle used is that of maximum likelihood. The principle behind this idea is wellknown: at each instance of time there is a certain probability density $p(u(0), \ldots, u(t-1), y(0), \ldots, y(t); \Theta)$ which expenses the likelihood that the string $y(0), \ldots, y(t)$ would have been observed with the input $u(.)$ and the parameter value Θ. At each instant of time one then chooses the parameter $\hat{\Theta}(t)$ such that it maximizes this likelihood function over all possible Θ. The convergence question is whether or not $\lim_{t\to\infty} \hat{\Theta}(t) = \Theta^*$, where Θ^* equals the true value of the parameter matrices $A(0), \ldots, B(q)$ which generate the data y from the input u. HANNAN proves a nice and very general result in this direction.

Of course to state and prove such a result one needs a

topology on the space of all possible models and if one wants to go beyond this result and discuss also how fast the convergence is, one needs more, namely a metric or a Riemannian structure on the space of all possible systems of a given type. It is here that the geometry and topology of the space of linear systems enter the picture and the discrete invariants called Kronecker indices turn out to have an important role to play. As such the results presented in this paper are a primeur in giving hard evidence of the relevance of this geometric structure issue in system identification.

Basically the same questions as in Hannan's paper are addressed by Deistler [section 4.3] for the case that some initial structural information on the to be identified system is already given.

Statistical tests to decide whether ARMA models will be adequate are considered by Guegan in [section 4.4].

Both from a conceptual and from a practical point of view it is important to implement an identification scheme in a recursive algorithm. The idea behind this is basically the same as explained in the context of nonlinear filtering. However, since one in general does not like to treat the unknowns as random variables, the procedure for obtaining recursive algorithms goes differently. In addition there are many different ways of approaching an identification problem (contrary to the situation in nonlinear filtering where there are many reasons for considering in the first place the conditional mean of the to be estimated variables). LJUNG's contribution [section 4.2] provides a very readable account of various aspects of recursive system identification basically all in the context of scalar ARMAX models as (10). He describes a number of identification routines and discusses their convergence properties. He then gives some practical guidelines for the implementation of these algorithms and closes by giving some results on the application of

identification algorithms in adaptive control (see section 6 of this article).

It is possible, of course, to treat an identification problem from a so-called Bayesian point of view. In the context of the model introduced in the beginning of this section, one then puts a probability measure on Θ, the space of unknown parameters. By considering now the product measure on $\Theta \times \Omega$, the total uncertainty space, this problem becomes a purely stochastic one and it is possible, for instance, to use nonlinear filtering ideas in system identification. This approach applied to ARMAX models (written, however, in state space form) is the subject of the article by KRISHNAPRASAD & MARCUS [section 4.5]. The estimation Lie algebras of these problems have a particularly pleasing structure with interesting possibilities for the existence of explicit recursive (approximate) filters.

6. ADAPTIVE CONTROL.

The last main topic discussed in this book is that of adaptive control. This is really one of the very early motivations of control theory: the design of control algorithms which will automatically learn the value of the (changing) plant parameters and self-adjust their control strategy accordingly.

Most of the adaptive control strategies proposed in the literature work according to a separation principle of identification and control. This is easily explained in the context of the general set-up discussed in Section 1. Assume that we have given an uncertain plant F with observed output $y = F(x,\Theta,\omega)$, with control input $u \in U$, unknown parameter $\Theta \in \Theta$, and stochastic uncertainty $\omega \in \Omega$. The problem is to design a feedback compensator, i.e., a nonanticipating map $G: Y \to U$, such that the closed loop system has some desirable properties. This control design purpose may be expressible in terms of closed loop stability, an optimal stochastic control

criterion, or some of the design formulations of multivariable control as, for example, model matching, pole placement, disturbance decoupling, etc. The difficulty, however, is that the unknown parameter Θ is indeed unknown.

If one uses a recursive identification scheme as explained in section 5 one will have at each instant of time an estimate $\hat{\Theta}(t)$ of the unknown parameter. Assume now that if Θ were known one would use the feedback control law G which, since it will depend on Θ, we denote by G_Θ. If G_Θ is implemented recursively, this will lead to a set of update equations with coefficients depending on Θ. The idea of using separation is to use for these parameters the estimate $\hat{\Theta}(t)$ at time t. This is illustrated in figure 4

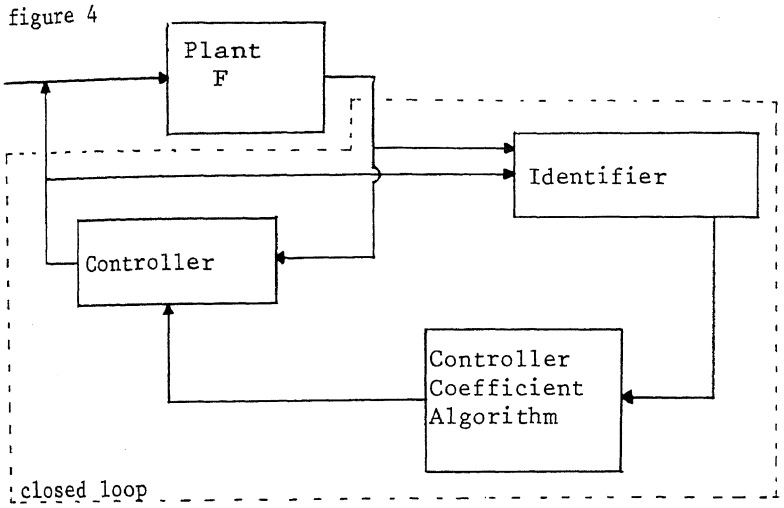


figure 4

The ensuing closed loop system will be very nonlinear and its properties are difficult to analyze. Moreover, one cannot simply conclude that a convergent identification routine will remain convergent when used in this closed loop framework. Indeed assumptions like u is bounded, deterministic, and if it is stochastic, independent of the stochastic disturbance of the plant, which one may have to make in order to prove the convergence

of the identification scheme, need not be satisfied.

The adaptive separation scheme induces a map $\tilde{G}: Y \to U$. The adaptive control scheme is said to be <u>self-tuning</u> if, as $t \to \infty$, the map $\tilde{G}$ converges to G_{Θ^*} in some sense. Here Θ^* denotes the true value of the parameter Θ and G_{Θ^*} denotes the controller which achieves the control objective (optimal performance, model matching, etc.) at the true value Θ^*. The surprising part of the results obtained so far is that self-tuning may occur even when inside the controller the convergence $\hat{\Theta}(t) \underset{t\to\infty}{\to} \Theta^*$ does not hold.

In this volume we have three papers on self-tuning control. The first one is by LANDAU [section 6.1] and treats self-tuning results for model reference adaptive control algorithms for scalar systems of the ARMAX type (10). The second paper is by CAINES & DORER [section 6.2]. It discusses a stabilization property for a class of (TV) ARMAX models, that is models of the type (10) but with time-varying coefficients. These coefficients are assumed to be stochastically time varying and the purpose of the paper is to prove asymptotic stability of the closed loop system.

The third paper in this chapter by FUCHS [section 6.3] discusses the stability of the overall system in terms of properties of the separate control subsystem and the identification subsystem.

7. CONCLUSIONS.

It is perhaps safer to leave the conclusions and statements of future prospects to the reader (after he has carefully read and digested the papers in this volume). For ourselves let us say that the future seems very bright, strong new impulses seem present everywhere in this field of filtering and identification and there seems to be a most promising gathering of forces in the sense that more and more new mathematical subjects are

brought in to bear upon the subject, which, when knowledgeably used, seem likely to enhance our understanding and improve our techniques.

To quote Joseph Louis Lagrange:

'As long as algebra and geometry proceeded along separate paths their advance was slow and their applications limited. But when these sciences joined company, they drew from each other fresh vitality and whence-forward marched on at a rapid pace toward perfection'.

It may well be (in our opinion) that in the field of enquiry of these proceedings we are witnessing today the beginnings of a similar joining of forces.