State and Parameter Estimation in Non-Linear Systems

By B. D. O. Anderson, B.Sc., B.E., Ph.D., M.I.E.Aust.*

Summary.—The paper examines the estimation of state variables and parameters for non-linear, noisy, discrete-time systems. The estimation problem is formulated as one requiring the calculation of probability density functions of the states and parameters, with these density functions being "conditioned" on the plant measurements in the sense that they reflect all knowledge about the plant. Formulas are developed for updating the probability densities as further measurements become available.

Using the probability densities, certain parameters, e.g., mean and variance, of a plant state or trajectory could be estimated. However, computation of the probability densities can present practical difficulties, and a dynamic programming procedure avoiding their computation is presented which derives the modal trajectory, that is, the most likely trajectory, which maximizes the associated probability density.

1.—INTRODUCTION

The control engineer is often faced with the problem of wishing to measure certain variables in a plant, but he is prevented by various practical considerations from doing so. Thus knowledge of the state-variables of a plant, so often apparently required for implementation of a controller, may be missing, while variables are available representing transformations of these state-variables. Often any plant measurements are noisy, i.e., the measured variable is some transformation of the state-variables together with a random variable. The natural question then arises as to how the state variables of the plant may be estimated from the measured variables.

Beginning with the noiseless case and a linear plant, theories have become available outlining techniques for the recovery of the state-variable. The work of Kalman on linear system observability is well known (Ref. 1), and hardly less well known is the work of Kalman and Bucy (Ref. 2), dealing with state estimation in a noisy environment.

Extensions to non-linear systems are clearly difficult. Among such we note the work of Albrekht and Krasovskii (Ref. 3) for deterministic systems and Ref. 4, which discusses the application of the Kalman-Bucy filter to noisy non-linear systems.

Whilst it is certainly a sensible policy to seek the state estimates of a linear plant subjected to Gaussian noise, the case where the plant is non-linear or the noise non-Gaussian should cause a re-thinking of the aim of estimation. As described in Ref. 5 and applied in Ref. 6, the problem of optimal control given noisy measurements is best solved using, not a state estimate, but the *information state of the plant, i.e., a probability density function of the plant state, where this function is conditioned (in the technical sense of probability theory) on a knowledge of all available measurements of the plant.* The reasons for this are that (i) the information state truly sums up all available knowledge about the state of the plant, and (ii) the information state can normally be calculated, at least in principle.

Unfortunately, all too often the calculation of the information state is prohibitive in its consumption of computer time. In a few situations, e.g., linear plant and Gaussian noise, this is not so, but in general hardly any continuous plants can be considered; nor is it true that information states can always be calculated in practice for discrete plants. Interest therefore may be centred around using certain quantities associated with the information state which (i) can be more easily computed, and (ii) can be used to generate a control strategy.

*Paper No. 2528, presented at the IFAC Symposium held in Sydney from 26th to 30th August, 1968.

The author is Professor of Electrical Engineering, University of Newcastle.

The idea of using such quantities arises also in optimal filtering problems, as distinct from optimal control problems; knowledge of the information state may be excessive, in the sense that knowledge of, e.g., only the state mean and variance may be desired.

Another class of problems centres around developing an estimate of the *trajectory* (i.e., sequence of states over a time interval) of a plant, rather than the state of the plant at some particular time. Of course, the estimation is based on knowledge of certain (noisy) measurements. Just as a conditional probability density can be defined for the state of a plant, so can a conditional probability density for its trajectory be found, and we may regard estimation problems as including the problem of determining this conditional probability density, and of determining certain associated relevant quantities which can more easily be computed.

In this paper, we consider discrete-time plants; there are noisy inputs and noisy outputs, i.e., in addition to a deterministic or known input there is a stochastic or unknown input, and the measured outputs are not merely functions of the plant state, but also of some random variable. The state at time $k + 1$ will be defined as a non-linear function of the state and input (deterministic and random) at time k, and the measured output at time k as a non-linear function of the state at time k and the random output noise.

We shall derive equations describing the evolution of the probability density of the system trajectory with time; thus these deterministic equations may be used to describe the system rather than the original stochastic equation describing the evolution of the state with time.

Section 2 presents the precise plant equations, and explicitly defines the estimation problems to be considered, also indicating the currently available results. In Section 3, calculations are presented for the conditional probability density functions of the various estimation problems; of particular interest is the connection between trajectory (sequence of states) estimation and (single) state estimation.

Section 4 is concerned with showing how a predicted modal (i.e., most likely) trajectory can be found by dynamic programming methods, while in Section 5 there is a brief discussion of the results, in terms of the computational problems involved in their application.

The significance of Section 4 is that it suggests a possible computational saving. As remarked earlier, actual computation of the conditional probability density functions may prove impossible in practice; Section 4 bypasses this calculation, giving an alternative route to calculating the state or trajectory which maximizes the associated probability density functions. The modal trajectory is thus one "quantity" associated with the probability density function whose computation is not as difficult as that of the probability density function.

2.—SYSTEM AND PROBLEM DESCRIPTIONS

We consider discrete-time dynamical systems with noisy inputs and noisy outputs. Since the systems are not restricted to being time-invariant, deterministic inputs are readily incorporated by appropriate time-variation of the system equations.

The evolution of the state vector is described by

$$x_{k+1} = f(x_k, w_k, k) \qquad \text{.........................(1)}$$

where x_k denotes the state vector at time k, and w_k is the noisy, i.e., random, input at time k.

The initial value of the state, x_0, is assumed to be a random variable with known probability density function $p(x_0)$.

The statistics of w, that is, the probability density functions $p(w_i)$, $i = 0, 1, 2, \ldots$ are also assumed known, and the random variables w_i and w_j for $i \neq j$ are assumed independent.

The formulation of Eq. (1) has the advantage of possibly including in the vector x_{k+1} entries more closely representing plant parameters than true plant state variables. Thus, consider a plant with state vector $\hat{x}$ and a set of unknown parameters represented by a vector $\hat{p}$. The plant equations are then of the form

$$\hat{x}_{k+1} = \hat{f}(x_k, \hat{p}, w_k, k) \quad \ldots\ldots\ldots\ldots(2)$$

to which may be adjoined

$$\hat{p}_{k+1} = \hat{p}_k \quad \ldots\ldots\ldots\ldots(3)$$

By defining x through $\hat{x}' = [\, x' \,|\, \hat{p}' \,]$, and with a suitable definition of f, Eq. (1) is recovered, and estimation procedures developed for Eq. (1) yield state *and parameter* estimation procedures for Eq. (2).

The measurable output of the system at time k is

$$z_k = h(x_k, v_k, k) \quad \ldots\ldots\ldots\ldots(4)$$

where the dimension of z_k need not be the same as that of x_k. Measurement noise is represented by the random variable v_k. We assume as for w_i that each $p(v_i)$, the probability density function of v_i, is known and that the random variables v_i and v_j for $i \neq j$ are independent. Finally, we assume the random variables v_i and w_j are independent for all i and j.

Problems such as the following then arise:

(1) Given the set of measurements $z_1, z_2, \ldots, z_k$ determine for some i the conditional probability density $p(x_i|z_1, z_2, \ldots, z_k)$; here i may be less than k, (state smoothing problem), equal to k, (state filtering problem), or greater than k (state prediction problem).

(2) Given the set of measurements $z_1, z_2, \ldots, z_k$, determine for some i the conditional probability density $p(x_0, x_1 \ldots x_i|z_1, z_2 \ldots z_k)$; for i less than k, we have the trajectory smoothing problem, for i equal to k the trajectory filtering problem, and for i greater than k the trajectory prediction problem.

(3) Given the set of measurements $z_1, z_2, \ldots, z_k$, determine for some i the conditional probability density $p(x_k, x_{k+1}, \ldots, x_{k+i}| z_1, z_2, \ldots, z_k)$. (Problem 3 differs from the third part of problem 2 by requiring only prediction of the trajectory, rather than simultaneous prediction and smoothing.)

In all problems, a technique is desired for incorporating the knowledge that one new measurement gives. Thus in problem 2, some hopefully simple procedures should be available for computing $p(x_1, x_2, \ldots, x_i|z_1, z_2, \ldots, z_{k+1})$ or $p(x_1, x_2, \ldots, x_{i+1}|z_1, z_2, \ldots, z_{k+1})$ when $p(x_1, x_2, \ldots, x_i|z_1, z_2, \ldots, z_k)$ is known.

Associated with each of the above problems are *modal estimation* problems, where the aim is to find that estimate of the state or the trajectory which will maximize the associated conditional probability density.

Results in the non-Gaussian, non-linear case defined by Eqs. (1) and (4) are not common. Among the principal references we note the work of Lee (Ref. 7) and Ho and Lee (Ref. 8), and that of Larson and Peschon (Ref. 9). The former two references restrict consideration to the state filtering and smoothing problems. The approach is to obtain equations for the relevant probability densities which are recursive in i, the index of the appropriate state, or in k, the index of the final measurement. The last reference is concerned with developing a recursive equation for the trajectory filtering problem, and then applying a dynamic programming technique to this equation to deduce the modal trajectory by a sequence of minimizations.

Cox (Ref. 10) formulates trajectory estimation problems for non-linear systems with *additive Gaussian* noise at the input and output; his equations are thus modified versions of Eqs. (1) and (4). He too uses dynamic programming techniques to obtain modal trajectories.

This paper extends an earlier report (Ref. 11).

3.—TRAJECTORY ESTIMATION—FILTERING, SMOOTHING AND PREDICTION

We shall adopt the notation

$$X_i = \{x_1, x_2, \ldots, x_i\} \quad \ldots\ldots\ldots\ldots$$

and

$$Z_i = \{z_1, z_2, \ldots, z_i\} \quad \ldots\ldots\ldots\ldots$$

3.1 Filtering:

Ref. 9 establishes the following equation.—

$$p(X_{k+1}|Z_{k+1}) = \frac{p(z_{k+1}|x_{k+1})\, p(x_{k+1}|x_k)}{p(z_{k+1}|Z_k)}\, p(X_k|Z_k) \quad \ldots\ldots\ldots$$

The first point to note is that Eq. (7) is a recursive formu in the sense that the filtering problem when given measuremen Z_{k+1} is solved using the solution of the filtering problem wh given Z_k, always assuming that the three densities $p(z_{k+1}|x_{k+}$ $p(x_{k+1}|x_k)$ and $p(z_{k+1}|Z_k)$ are known. Of course, the starti point for the iteration is

$$p(X_1|Z_1) = \frac{p(z_1|x_1)\, p(x_1|x_0)}{p(z_1)}\, p(x_0) \quad \ldots\ldots\ldots\ldots$$

The densities $p(z_{k+1}|x_{k+1})$ and $p(x_{k+1}|x_k)$ required to u Eq. (7) are in theory derivable from Eqs. (1) and (4) and knowled of the densities $p(v_k)$ and $p(w_k)$. In practice the calculation m be difficult, though there are easy cases, for example, linear plan with Gaussian noise.

The derivation of $p(z_{k+1}|Z_k)$ in Eq. (7) is naturally more dif cult. This density can however be related to the state probabil density function $p(x_k|Z_k)$ or the trajectory probability density fun tion $p(X_k|Z_k)$. Certainly the latter will be formed in the iterati process, and thus all terms in Eq. (7) may be considered known.

The precise relations may be derived as follows:

Observe that,

$$p(x_{k+1}, z_{k+1}) = p(z_{k+1}|x_{k+1})\, p(x_{k+1})$$

$$= \int p(z_{k+1}|x_{k+1})\, p(x_{k+1}|x_k)\, p(x_k)\, dx_k$$

This equation remains valid when each probability dens function is conditioned on Z_k. Noting that from Eq. (2) knowled of x_k and z_k is no more use than knowledge of merely x_k in esti ating x_{k+1}, we have

$$p(x_{k+1}|x_k, Z_k) = p(x_{k+1}|x_k).$$

Likewise, from Eq. (4), knowledge of x_{k+1} and Z_k is of no mo use than knowledge of merely x_{k+1} in estimating z_{k+1}, and thus

$$p(z_{k+1}|x_{k+1}, Z_k) = p(z_{k+1}|x_{k+1}).$$

Consequently,

$$p(x_{k+1}, z_{k+1}|Z_k) = \int p(z_{k+1}|x_{k+1})\, p(x_{k+1}|x_k)\, p(x_k|Z_k)\, dx_k$$

Now

$$p(z_{k+1}|Z_k) = \int p(x_{k+1}, z_{k+1}|Z_k)\, dx_{k+1}$$

and thus we deduce

$$p(z_{k+1}|Z_k) = \iint p(z_{k+1}|x_{k+1})\, p(x_{k+1}|x_k)\, p(x_k|Z_k)\, x_k\, dx_{k+1} \quad \ldots\ldots$$

As before, the probabilities $p(z_{k+1}|x_{k+1})$ and $p(x_{k+1}|$ follow from Eqs. (1) and (4). We have as yet not indicated h $p(x_k|Z_k)$ may be found, but this is related to $p(X_k|Z_k)$ simply by

$$p(x_k|Z_k) = \iint \ldots \int p(X_k|Z_k)\, dx_0\, dx_1 \ldots dx_{k-1} \quad \ldots\ldots\ldots($$

so that

$$p(z_{k+1}|Z_k) = \iint \ldots \int p(z_{k+1}|x_{k+1})\, p(x_{k+1}|x_k)\, p(X_k|Z_k) \times$$
$$\times\, dx_0\, dx_1 \ldots dx_{k+1} \quad \ldots\ldots\ldots($$

Actually, this equation also follows directly from Eq. (7) observing that the integral of the left hand side with respect $(x_0, x_1, \ldots, x_{k+1})$ must be unity, by the normalization prope of probability densities.

In summary, we have found the following:

(*a*) Eq. (7) allows recursive calculation of $p(X_{k+1}|Z_{k+1})$ as k increases.

(*b*) To carry out the recursive calculation two of the probability density functions are computable from the system Eqs. (1) and (4). The third is given by Eq. (11) and its computation requires $p(X_k|Z_k)$, as well as densities derived from Eqs. (1) and (4).

(*c*) Eq. (11) really reflects the normalization property of $p\ (X_{k+1}|Z_{k+1})$, and, as such, is not of fundamental significance. Thus there may be situations where the calculation of Eq. (11) need not be carried out. For example, as $p(z_{k+1}|Z_k)$ is a function of $z_1, z_2, \ldots, z_{k+1}$ but not explicitly of $x_0, x_1, \ldots$ or x_k, then the modal trajectory X_k, i.e., arg min $\{p(X_k|Z_k)\}$, is also from Eq. (7) arg min $\{p(z_k|x_k)\, p(x_k|x_{k-1})\, p(X_{k-1}|Z_{k-1})\}$.

(*d*) Given the probability density function $p(X_k|Z_k)$ applicable to trajectory filtering, the probability density function $p(x_k|Z_k)$ applicable to state filtering follows simply, using Eq. (10).

(*e*) One interpretation of Eq. (7) is that it is a deterministic equation replacing the stochastic Eq. (1). The solutions of Eq. (1), i.e., the members of the sequence $x_1, x_2, \ldots$, are random variables; the solutions of Eq. (7), i.e., the members of the sequence $p(X_1|Z_1)$, $p(X_2|Z_2), \ldots$, are definite functions. Unfortunately, the number of arguments in $p(X_k|Z_k)$, viz. $2k + 1$, increases with k in general.

Just as there is a recursive formula for $p(X_k|Z_k)$, so there is one for $p(x_k|Z_k)$. The derivation may be found in Ref. 8, for completeness we include the result.—

$$p(x_{k+1}|Z_{k+1}) = \frac{\int p(z_{k+1}|x_{k+1})\, p(x_{k+1}|x_k)\, p(x_k|Z_k)\, dx_k}{p(z_{k+1}|Z_k)} \qquad \ldots\ldots(12)$$

Note that the evaluation of $p(z_{k+1}|Z_k)$ in terms of $p(x_k|Z_k)$ has already been discussed. Note also that the numerator in Eq. (12) is very similar to that in Eq. (7), the corresponding trajectory result, but that an integration is required (and thus further calculation) for Eq. (7).

3.2 Trajectory Smoothing:

In the trajectory smoothing problem interest centres around $p(X_i|Z_k)$, for $i < k$. Here one is trying to estimate the behaviour of a plant up till some critical time in the past. Behaviour beyond this time is unimportant, but some measurements are available. Clearly these measurements contain information concerning earlier behaviour, and thus $p(X_i|Z_k)$ should be more useful than say $p(X_i|Z_i)$.

If $p(X_k|Z_k)$ is available it is immediate that

$$p(X_i|Z_k) = \int\int \ldots \int p(X_k|Z_k)\, dx_{i+1}\, dx_{i+2} \ldots dx_k \qquad \ldots\ldots\ldots(13)$$

However, assuming this is not the case, we are led to considering the derivation of a recursive formula for $p(X_i|Z_k)$. First, observe that

$$p(X_i|z_k)\, p(z_k) = p(z_k|X_i)\, p(X_i)$$

and so, conditioning each probability on a knowledge of Z_{k-1},

$$p(X_i|Z_k)\, p(z_k|Z_{k-1}) = p(z_k|Z_{k-1}, X_i)\, p(X_i|Z_{k-1})$$

Immediately,

$$p(X_i|Z_k) = \frac{p(z_k|Z_{k-1}, X_i)}{p(z_k|Z_{k-1})}\, p(X_i|Z_{k-1}) \qquad \ldots\ldots\ldots\ldots\ldots(14)$$

which should be compared with the corresponding result for state smoothing, see Ref. 7:

$$p(x_i|Z_k) = \frac{p(z_k|Z_{k-1}, x_i)}{p(z_k|Z_{k-1})}\, p(x_i|Z_{k-1}) \qquad \ldots\ldots\ldots\ldots\ldots(15)$$

The evaluation of the probability density function $p(z_k|Z_{k-1}, X_i)$ in terms of simpler density functions appears very complex, and will not be discussed here. (Even in the linear plant, Gaussian noise case when the density functions can be characterized by a mean and covariance, the extension of the filtering and prediction theory of Ref. 2 to smoothing theory, see Refs. 7 and 12, occurred several years after Ref. 2. The density function $p(z_k|Z_{k-1})$ in the denominators of Eqs. (14) and (15) has already been discussed.)

Eq. (14) is suited to dealing with the situation where the measurement data are increasing. Alternatively, of course, it may simply be used in the situation where the measurement data are fixed. Then iteration of Eq. (14) yields

$$p(X_i|Z_k) = \frac{p(z_k|Z_{k-1}, X_i) \ldots p(z_{i+1}|Z_i, X_i)}{p(z_k|Z_{k-1}) \ldots p(z_{i+1}|Z_i)}\, p(X_i|Z_i) \qquad \ldots\ldots\ldots(1$$

which is also very similar to the corresponding state estimatio formula (Ref. 7)

$$p(x_i|Z_k) = \frac{p(z_k|Z_{k-1}, x_i) \ldots p(z_{i+1}|Z_i, x_i)}{p(z_k|Z_{k-1}) \ldots p(z_{i+1}|Z_i)}\, p(x_i|Z_i) \qquad \ldots\ldots\ldots(1$$

A recursive formula for generating $p(X_i|Z_k)$ from $p(X_{i-1}|Z_{k-}$ is also available. To obtain this, observe first that

$$\begin{aligned} p(X_i|Z_k) &= p(x_i, X_{i-1}|Z_k) \\ &= p(x_i|X_{i-1}, Z_k)\, p(X_{i-1}|Z_k) \\ &= p(x_i|x_{i-1}, Z_k)\, p(X_{i-1}|Z_k) \end{aligned}$$

and thus, using Eq. (14)

$$p(X_i|Z_k) = \frac{p(x_i|x_{i-1}, Z_k)\, p(z_k|x_{i-1}, Z_{k-1})}{p(z_k|Z_{k-1})}\, p(X_{i-1}|Z_{k-1}) \qquad \ldots\ldots(18$$

The point of this result is that it allows estimation of the tra jectory up to a point in the past for which the time interval betwee this point and the present is constant. When more measuremen data become available, the trajectory is estimated further forwar in time.

3.3 State Prediction, Trajectory Prediction with Smoothin and Pure Trajectory Prediction:

The prediction of the future behaviour of a system on the bas of certain available measurements is clearly of great interest. W distinguish between three cases:

(i) Estimating the system state at some future time;

(ii) Estimating the system trajectory, from some starting time in th past through to some future time;

(iii) Estimating the system trajectory in future time.

It turns out that the calculation of $p(x_i|Z_k)$, $p(X_i|Z_k)$ an $p(x_k, \ldots x_i|Z_k)$ for $i > k$ can be easily described in iterative term

Observe that

$$\begin{aligned} p(x_i|Z_k) &= \int p(x_i|x_{i-1}, Z_k)\, p(x_{i-1}|Z_k)\, dx_{i-1} \\ &= \int p(x_i|x_{i-1})\, p(x_{i-1}|Z_k)\, dx_{i-1} \qquad \ldots\ldots\ldots\ldots\ldots(1 \end{aligned}$$

which is an immediate formula for predicting the state recursivel As before, $p(x_i|x_{i-1})$ follows from the fundamental system Eq. (1 and knowledge of $p(w_{i-1})$. The corresponding formula for tra jectory prediction is even simpler:

$$\begin{aligned} p(X_i|Z_k) &= p(x_i|X_{i-1}, Z_k)\, p(X_{i-1}|Z_k) \\ &= p(x_i|x_{i-1})\, p(X_{i-1}|Z_k) \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2 \end{aligned}$$

Similarly,

$$\begin{aligned} p(x_k, \ldots x_i|Z_k) &= p(x_i|x_k \ldots, x_{i-1}, Z_k)\, p(x_k \ldots, x_{i-1}|Z_k) \\ &= p(x_i|x_{i-1})\, p(x_k \ldots, x_{i-1}|Z_k) \qquad \ldots\ldots(2 \end{aligned}$$

Eq. (20) can be iterated to yield

$$p(X_i|Z_k) = p(x_i|x_{i-1}) \ldots p(x_{k+1}|x_k)\, p(X_k|Z_k) \qquad \ldots\ldots\ldots\ldots(2$$

and similarly for Eq. (21). Further iteration is best based on rela ing $p(X_k|Z_k)$ to $p(X_{k-1}|Z_{k-1})$ via Eq. (7).

Eq. (7) also permits the relating of $p(X_{i+1}|Z_{k+1})$ to $p(X_i|Z$ for $i > k$ or $p(x_{k+1}, \ldots x_{i+1}|Z_k)$ to $p(x_k, \ldots x_i|Z_k)$. Thus if inte est centres around predicting the trajectory, a fixed time interv into the future, and further measurements arrive, this relationshi is the appropriate one to use.

From Eq. (22),

$$\begin{aligned} p(X_{i+1}|Z_{k+1}) &= p(x_{i+1}|x_i) \ldots p(x_{k+2}|x_{k+1})\, p(X_{k+1}|Z_{k+1}) \\ &= p(x_{i+1}|x_i) \ldots p(x_{k+2}|x_{k+1})\, \frac{p(z_{k+1}|x_{k+1})\, p(x_{k+1}|x_k)}{p(z_{k+1}|Z_k)}\, p(X_k|Z \end{aligned}$$

from Eq. (7). Now use Eq. (22) *in reverse*, to get

$$p(X_{i+1}|Z_{k+1}) = p(x_{i+1}|x_i)\, \frac{p(z_{k+1}|x_{k+1})}{p(z_{k+1}|Z_k)}\, p(X_i|Z_k) \qquad \ldots\ldots\ldots\ldots(2$$

A similar result of course holds when pure prediction of the trajectory is being considered. Here we observe that from Eq. (21)

$$p(x_{k+1}, \ldots x_{i+1}|Z_{k+1}) = p(x_{i+1}|x_i) \ldots p(x_{k+2}|x_{k+1})\, p(x_{k+1}|Z_{k+1})$$

and using Eq. (12)

$$= \frac{p(z_{k+1}|x_{k+1})}{p(z_{k+1}|Z_k)} p(x_{i+1}|x_i) \times$$

$$\times \int p(x_i|x_{i-1}) \ldots p(x_{k+1}|x_k)\, p(x_k|Z_k)\, dx_k$$

$$= p(x_{i+1}|x_i) \frac{p(z_{k+1}|x_{k+1})}{p(z_{k+1}|Z_k)} \int p(x_k, \ldots x_i|Z_k)\, dx_k \qquad \ldots\ldots\ldots\ldots\ldots(24)$$

Despite the simplicity of some of the above relations, it is evident that the question " What is the most likely trajectory X_i, given a knowledge of Z_k?" is not easily answered. The probability density function $p(X_i|Z_k)$ is a function of the $(i+1)$ variables $x_0, x_1, \ldots x_i$ and k variables $z_1, z_2, \ldots, z_k$, and is thus a function of $i + k + 1$ variables in all*. To find the most likely trajectory via the methods of differential calculus, $(i+1)$ differential coefficients would be required, and the non-linear equations resulting from equating these differential coefficients to zero would need to be solved.

In the next section, an alternative approach using dynamic programming is presented.

4.—MODAL TRAJECTORY PREDICTION VIA DYNAMIC PROGRAMMING

The computational difficulties inherent in computing $p(X_i|Z_k)$ in Eq. (20) are worsened by the fact that in carrying through the iterative procedure, we are forced to terminate at $p(X_k|Z_k)$; at this point, for explicit computation of $p(X_i|Z_k)$ the procedure suggested by Eq. (7) must then be applied. Recognizing this difficulty in the filtering *as distinct from prediction* case, Ref. 9 exhibits a technique for computing the trajectory X_k which will maximize $p(X_k|Z_k)$, i.e., the modal trajectory; the technique has far less severe computational requirements than those associated with obtaining $p(X_k|Z_k)$ explicitly. Knowledge of the modal trajectory alone, rather than the probability density function of all trajectories, will often be sufficient for some applications.

In this section the result of Ref. 9 is improved to the extent of prescribing a procedure for predicting the modal trajectory, as well as estimating it up to the present time. In other words, we give a procedure for determining the trajectory X_i which maximizes the value of the function $p(X_i|Z_k)$ for $i > k$.

Full appreciation of the extension presented here depends upon a good knowledge of Ref. 9 and the associated dynamic programming principles.

For $j \geqslant k$, define

$$I(x_j, j) = \max_{X_{j-1}} p(X_j|Z_k) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(25)$$

and for $j \leqslant k$, define

$$I(x_j, j) = \max_{X_{j-1}} p(X_j|Z_j) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(26)$$

(Observe that Eqs. (25) and (26) agree when $j = k$.) The final (j-th) state on the modal trajectory will be the value of x_j which maximizes $I(x_j, j)$.

From Eqs. (20) and (25), it follows that for $j > k$

$$I(x_{j+1}, j+1) = \max_{X_j} \{p(x_{j+1}|x_j)\, p(X_j|Z_k)\}$$

$$= \max_{x_j} \{p(x_{j+1}|x_j) \max_{X_{j-1}} p(X_j|Z_k)\}$$

$$= \max_{x_j} \{p(x_{j+1}|x_j)\, I(x_j, j)\} \qquad \ldots\ldots\ldots(27)$$

For $j < k$, from Eq. (7),

$$I(x_{j+1}, j+1) = \max_{X_j} \left\{\frac{p(z_{j+1}|x_{j+1})\, p(x_{j+1}|x_j)}{p(z_{j+1}|Z_j)} p(X_j|Z_j)\right\}$$

$$= \max_{x_j} \left\{\frac{p(z_{j+1}|x_{j+1})\, p(x_{j+1}|x_j)}{p(z_{j+1}|Z_j)} I(x_j, j)\right\} \quad \ldots\ldots($$

Our aim is of course to find the X_j maximizing $p(X_j|Z$ rather than the value of $p(X_j|Z_k)$ for this maximizing X_j, equivalently, to find the X_j maximizing $I(x_j, j)$ and then all t earlier states of the trajectory. Define the functions

$$I_*(x_{j+1}, j+1) = \max_{x_j} \{p(x_{j+1}|x_j)\, I_*(x_j, j)\}, \quad j \geqslant k \quad \ldots\ldots\ldots\ldots($$

$$I_*(x_{j+1}, j+1) = \max_{x_j} \{p(z_{j+1}|x_{j+1})\, p(x_{j+1}|x_j)\, I_*(x_j, j)\}, \quad j < k \;($$

which differ from Eqs. (28) and (30) by the omission of the ter $p(z_{j+1}|Z_j)$. The functions I_* are proportional to I, so that the sam values of x_j maximize Eqs. (29) and (30) as maximize Eqs. (2 and (28). Note that although $p(z_{j+1}|Z_j)$ does not appear in Eq. (2 this equation must still be altered, because the starting point f iterating Eq. (27), namely $j = k$ requires the end part of the iter tions associated with Eq. (28). By changing this from $I(x_k, k)$ Eq. (28) to $I_*(x_k, k)$ in Eq. (30), we are forced to change Eq. (2 accordingly.

It is now clear that Eq. (29) is really a specialized version Eq. (30), corresponding to a situation where $p(z_{j+1}|x_{j+1})$ is constan in fact unity, though the value is immaterial, as a scaling has be introduced.

Now the dynamic programming procedure of Ref. 3, not be discussed here, revolves around the use of the iterative equati

$$I_*(x_{j+1}, j+1) = \max_{x_j} \{p(z_{j+1}|x_{j+1})\, p(x_{j+1}|x_j)\, I_*(x_j, j)\} \quad \ldots\ldots($$

Consequently, *this procedure may be carried over directly our estimation problem, by equating $p(z_{j+1}|x_{j+1})$ to unity for $j \geqslant$*

It should be evident from the preceding and a study of Ref. that the part of the modal trajectory associated with instants to time k, i.e., the sequence of states $x_1, x_2, \ldots, x_k$ rather th $x_1, x_2, \ldots, x_i$, will not in general be the same as the modal tr jectory obtained in the smoothing problem, where $i = k$. Th means that there is no simple procedure for extrapolating the mod trajectory obtained in the smoothing problem to yield a mod trajectory for the prediction problem, though undoubtedly in ma situations such an extrapolation would be of significance.

5.—SIGNIFICANCE OF THE RESULTS

The definitive collection of formulas presented, even thou they suggest computational problems of enormous magnitud still have considerable potential for practical application. have some formulas is better than having none, and, as Section shows, the computational problems are indeed capable of reductio certainly too, as time passes such reductions will become less a less necessary as the available computers improve.

It is worth pointing out that there are other cases of inter besides the determination of the modal trajectory by dynam programming where computational simplifications are possib Such cases occur when the probability densities are simple types functions. Thus the Cox assumption (Ref. 10) of Gaussi additive noise *without plant linearity* guarantees that many densit of interest are exponentials and are thus easily computable; th are presumably a number of other densities with as pleasant prop ties as the exponential ones, though perhaps not so many with associated physical origin.

Another class of densities for which all formulas simplify those which are not continuous functions but rather a class generalized functions, namely delta functions. Such densit correspond to the situation where only discrete values of the st are permitted, certainly the case in some physical systems. such instances the methods of dynamic programming for mod trajectory determination prove very efficient, see e.g., Refs. 8 and

*Each variable may itself be a vector also.

References

1. Kalman, R. E.—On the General Theory of Control Systems. *P ceedings of the First International Congress of the International Fede tion of Automatic Control (IFAC), held in Moscow*, 1960, Vol. I, 481-492. London, Butterworths, 1961.

2. KALMAN, R. E. and BUCY, R. S.—New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering, Trans. A.S.M.E., Series D*, Vol. 83, No. 1, March, 1961, pp. 95-108, (Paper No. 60-JAC-12).

3. ALBREKHT, E. G. and KRASOVSKII, N. N.—Observability of a Nonlinear Controlled System in the Neighbourhood of Given Motion. *Automation and Remote Control.* Vol. 25, No. 7, July, 1964, pp. 934-944.

4. LARSON, R. E., DRESSLER, R. M. and RATNER, R. S.—Application of the Extended Kalman Filter to Ballistic Trajectory Estimation. *Stanford Research Institute, Menlo Park, California, Final Technical Report, Contract No.* DA-01-021-AMC-90006(Y).

5. MEIER, L.—Combined Optimal Control and Estimation. *Proceedings of the Third Allerton Conference on Circuit and System Theory, October,* 1965.

6. MEIER, L., PESCHON, J., HO, R., LARTSON, R. and DRESSLER, R.—Design Guidance and Control Systems for Optimum Utilization of Information. *Stanford Research Institute, Menlo Park, California, Final Technical Report, Contract No.* NAS2-3476.

7. LEE, R. C. K.—*Optimal Estimation, Identification and Control.* Cambridge, Massachusetts, MIT Press, 1964, 152 p.

8. HO, Y. C. and LEE, R. C. K.—A Bayesian Approach to Problems in Stochastic Estimation and Control. *I.E.E.E. Transactions on Automatic Control*, Vol. AC-9, No. 4, October, 1964, pp. 333-339.

9. LARSON, R. E. and PESCHON, J.—A Dynamic Programming Approach to Trajectory Estimation. *I.E.E.E. Transactions on Automatic Control*, No. AC-11, No. 3, July, 1966, pp. 537-540.

10. COX, H.—On the Estimation of State Variables and Parameters for Noisy Dynamic Systems. *I.E.E.E. Transactions on Automatic Control*, Vol. AC-9, No. 1, January, 1964, pp. 5-12.

11. ANDERSON, B. D. O.—Topics in Estimation Theory. *Stanford University, Electronics Laboratories, Stanford, California, Report No. SEL*-66-108, (*TR No.* 6560-9), December, 1966.

12. MEDITCH, J. S.—Orthogonal Projection and Discrete Optimal Linear Smoothing. *SIAM Journal on Control*, Vol. 5, No. 1, February, 1967, pp. 74-89.