

ALGORITHMS FOR THE INCORPORATION  
OF PREDICTIVE INFORMATION IN SURVEILLANCE THEORY\*

David A. Castanon<sup>†</sup>  
Bernard C. Levy\*\*  
Alan S. Willsky\*\*

ABSTRACT

This paper discusses the problem of tracking a platform given probabilistic information concerning its possible destinations. Using a Bayesian approach, we derive an algorithm which optimally incorporates the real-time measurements of the platform position with the predictive information about its destination. This algorithm can be efficiently implemented using some results from optimal smoothing. The results are illustrated by a simple example.

---

\* This research was performed at the Laboratory for Information and Decision Systems, M.I.T., and was supported by the Office of Naval Research under contract ONR/N00014-76-C-0346 and by contract ONR/N00014-77-C-0532.

\*\* Laboratory for Information and Decision Systems, Room 35-316, Massachusetts Institute of Technology, Cambridge, MA. 02139.

<sup>†</sup> Alphatech, Inc., 3 New England Executive Park, Burlington, MA.

1. Introduction

An important problem in the design of current U.S. Navy systems is the designs of an overall surveillance systems. This surveillance system has two primary objectives: The first one is to keep track of all the ships located in a given area and the second one is to maintain a record of all the information related to these ships (status, nationality, possible destination...). The information which is available to such a surveillance system consists of a combination of the following data:

- (i) the identification of the ship's class and its dynamics,
- (ii) some information about the ship's origin, its destination and its refueling stops,
- (iii) sightings by other ships (submarines or surface ships), by airplanes or by satellites,
- (iv) some sonar information obtained from tracking sonar arrays.

A close look at the available data indicates that it can be divided into two categories. Sightings by ships, airplanes, sonar or radar, and the ship's origins provides information about the past of the ships trajectory, whereas information about the ship's destination and refueling stops provide information about the future of the ship trajectory. In this paper, we derive an algorithm which can be used to incorporate both categories of information, in order to obtain an overall best estimate of the ship's position.

Our basic approach is to assume that the ship's dynamics are governed by a state-space model of a stochastic dynamical system, of the form

$$dx(t) = A(t) \ x(t)dt + dU(t) \tag{1}$$

where  $U(t)$  is a Wiener process with covariance

$$E\{U(t) U'(s)\} = \int_0^{\min(t,s)} Q(\tau) d\tau \quad (2)$$

Models of this form have been proposed by a number of authors (Singer [4], Singer and Behnke [5], Morgan [6]) for the modeling of ship trajectories with and without autopilots, or in maneuvering situations. The detailed modeling of ship dynamics is not important to the subsequent results; the reader is referred to [4] - [6] for further discussions of these issues.

The model in (1) describes the ship dynamics before any additional information its destination is known. In the next sections, we describe an algorithm which modifies (1) to obtain a new Markovian model which has incorporated the predictive information concerning a single destination. This algorithm uses the basic approach of obtaining backwards Markovian models, as in Verghese and Kailath [1]. Based on this new model, the optimal estimate of ship position using past measurements can be obtained from standard estimation algorithms such as Kalman filtering [15].

In Section 3, we extend the results of Section 2 to incorporate probabilistic predictive information concerning multiple possible destinations. The optimal estimate of ship position, based on past and predictive information, is computed using a multiple model approach [15]. Furthermore, a likelihood ratio method to identify the validity of the intelligence information is presented, leading to an efficient implementation of the tracking algorithm using fixed-interval smoothing formulae [12], [13], [18]. In Section 4, we discuss how this method can be adapted for maneuver detection using the results of Willsky and Jones [3]. Section 5 contains an example which illustrates the basic algorithm.

## 2. Incorporation of Predictive Information

Assume that a state-space model of a ship trajectory is given by equations (1) - (2) and the initial condition

$$x(0) \sim N(m_0, \Pi_0) \quad (3)$$

where  $x(0)$  is independent of  $U(t)$  for all  $t$  in  $[0, T]$ . Assume that we receive the additional predictive information that, at time  $T$ , the ship's position will be observed probabilistically, in terms of a mean value and an error, as

$$x_T \sim N(m_T, \Pi_T) \quad (4)$$

Our goal is to obtain a new model which already incorporates this predictive information. Mathematically, we are interested in obtaining a new description of the trajectories of the process, of the form

$$dx(t) = \bar{A}(t) x(t)dt + d\bar{U}(t) \quad (5)$$

$$x(0) = N(\bar{x}_0, \bar{\Sigma}_0) \quad (6)$$

such that  $x(0)$  is independent of  $\bar{U}(t)$  for all  $t$  in  $[0, T]$ , and such that the probability distribution induced by (5), (6) on the space of sample paths  $C\{[0, T]; \mathbb{R}^n\}$  is equal to the probability induced by (1), (2), (3), conditioned on knowledge of the predictive information (4). The first step in this process is to describe the predictive information as an observation. Let

$$m_T = x_T + v_T \quad (7)$$

where  $v_T \sim N(0, \Pi_T)$  is independent of  $U(t)$  for  $t$  in  $[0, T]$ . Let  $(\Omega, \mathcal{F}, P)$  be our basic probability space, where  $x_0, v_T$  and  $U(s), 0 \leq s \leq T$  are defined. Define the two increasing sequences of  $\sigma$ -fields

$$G_t = U_t \vee X_0$$

$$0 \leq t \leq T$$

$$F_t = U_t \vee X_0 \vee M_T$$

where  $U_t = \sigma\{U(s), 0 \leq s \leq t\}$  is the sequence of  $\sigma$ -fields generated by the Wiener process  $U(\cdot)$ , and where  $X_0$  and  $M_T$  are the  $\sigma$ -fields generated by  $x(0)$  and  $m_T$ . Since the noise  $U(\cdot)$  is independent of  $x(0)$ , the process  $(U(t), G_t)$  is a martingale; that is,

$$E\{U(s) - U(t) | G_t\} = 0$$

for  $0 \leq t \leq s \leq T$ . Hence, we can write (1), in integrated form, as

$$x(t) = x(0) + \int_0^t Ax(s)ds + \int_0^t dU(s) \quad (8)$$

where  $x(t)$  is a Markov process, the middle term is  $G_t$  predictable, and the last term is a  $G_t$  martingale. A model satisfying this property is said to be a Markovian model of  $x(\cdot)$ . Most filtering or estimation results have been obtained for processes described by Markovian models.

Note that the process  $U(t)$  is not a martingale with respect to  $F_t$  (indeed,  $U(t)$  and  $m_T$  are correlated). Therefore, the problem of incorporating the information (4) in the model (1), (2) is essentially the one of constructing a Markovian model of  $x(\cdot)$  with respect to  $F_t$ ; i.e., with respect to all the a priori information available, including predictive information.

To do so, we shall follow a method similar to the one used by Verghese and Kailath [1] for the derivation of backwards Markovian models. The first step is to construct a Doob-Meyer decomposition of  $U(t)$  with respect to  $F_t$ , i.e.,

$$U(t) = \int_0^t E\{dU(\tau) | F_\tau\} + \tilde{U}(t) \quad (9)$$

where

$$E\{\tilde{U}(s) - \tilde{U}(t) | F_t\} = 0 \text{ for } 0 \leq t \leq s \leq T .$$

The reader is referred to Meyer [7] and Wong [8] for a description of the properties of such a decomposition. From (9) it is clear that to obtain the martingale  $\tilde{U}(t)$ , one needs only to compute the predictable projection  $E\{dU(t) | F_t\}$ , where

$$dU(t) \triangleq U(t + dt) - U(t).$$

However, before doing so, it will be useful to decompose  $F_t$  as

$$F_t = U_t \vee X_0 \vee \tilde{X}(t, T) \quad (10)$$

where  $\phi(\cdot, \cdot)$  denotes the transition matrix of  $A(\cdot)$ , and where

$$\tilde{X}(t, T) = x(t) - \phi(t, T)m_T = - \int_t^T \phi(t, s) dU(s) - \phi(t, T)v_T . \quad (11)$$

Let  $\tilde{X}(t, T)$  denote the  $\sigma$ -field generated by  $\tilde{X}(t, T)$ . This  $\sigma$ -field, by construction, is independent from  $G_t = U_t \vee X_0$ . To justify the decomposition (10) note that

$$x(t) \in U_t \vee X_0 \quad (12)$$

and also observe that

$$x(t) \vee \tilde{X}(t,T) = x(t) \vee M_T \quad (13)$$

(i.e., the knowledge of  $(x(t), \tilde{X}(t,T))$  is equivalent to the knowledge of  $(x(t), m_T)$ ).

We are now in a position to prove our main Theorem.

Theorem 1: Modified Model

Let  $\Pi(t,T)$  denote the solution of the equation

$$\begin{aligned} -\frac{d}{dt} \Pi(t,T) &= -A\Pi - \Pi A^* + Q \\ \Pi(T,T) &= \Pi_T \end{aligned} \quad (14)$$

Then, the model

$$\begin{aligned} x(t) &= x(0) + \int_0^t Ax(s) ds \\ &+ \int_0^t Q(s) \Pi^{-1}(s,T) (\phi(s,T)m_T - x(s)) ds \\ &+ \tilde{U}(t) \end{aligned} \quad (15)$$

is a Markovian model of  $x(\cdot)$  with respect to  $F_t$ , where  $(\tilde{U}(t), F_t)$  is a Brownian motion with intensity  $\int_0^t Q(s) ds$ .

Proof: The independence of  $G_t$  and  $\tilde{X}(t,T)$  gives

$$E\{dU(t) | F_t\} = E\{dU(t) | G_t\} + E\{dU(t) | \tilde{X}(t,T)\} = E\{dU(t) | \tilde{X}(t,T)\}$$

because  $U_t$  is a  $G_t$  martingale and  $\tilde{X}(t,T)$  is generated by  $\tilde{x}(t,T)$ . Now,

$dU(t)$  and  $\tilde{x}(t,T)$  are Gaussian random variables, so that

$$E\{dU(t) | \tilde{x}(t,T)\} = E\{dU(t)\tilde{x}'(t,T)\} E\{\tilde{x}(t,T)\tilde{x}'(t,T)\}^{-1} \tilde{x}(t,T) . \quad (16)$$

From (11), we get

$$E\{dU(t) \tilde{x}'(t,T)\} = - Q(t)dt$$

$$E\{\tilde{x}(t,T) \tilde{x}'(t,T)\} = \int_t^T \phi(t,s)Q(s)\phi'(t,s)ds + \phi(t,T)\Pi_T\phi'(t,T) . \quad (17)$$

Note that (17) is the solution of (14), so that

$$E\{\tilde{x}(t,T) \tilde{x}'(t,T)\} = \Pi(t,T) .$$

Hence

$$E\{dU(t) | F_t\} = Q(t) \Pi^{-1}(t,T) (\phi(t,T)m_T - x(t))dt . \quad (18)$$

Equation (18) represents the predictable part of the Doob-Meyer decomposition of  $U(t)$  with respect to the  $\sigma$ -fields  $F_t$ . Since  $U(t)$  has continuous sample paths, the quadratic variation of  $\tilde{U}(t)$  is the same as the quadratic variation of  $U(t)$ , so it is an  $F_t$  Wiener process with covariance

$$E\{\tilde{U}(t) \tilde{U}(s)\} = \int_0^{\min(t,s)} Q(\tau)d\tau .$$

Now, substituting (18), into the integrated form of (1) yields the results of the theorem.

Remark:

The model (15) is the same as the backwards Markovian model obtained by Verghese and Kailath [1] (also [9] , [10]). However, the model (15) is a forwards model. The martingale method that we have used here to construct (15) is very general and applies also to a large class of problems in the study of random fields (Willsky [11]).

The results of Theorem 1 provide us with a model which describes the evolution of the sample paths of  $x(t)$  from an initial condition  $x(0)$ . In order to properly specify the distribution of sample paths, we must provide a description of the initial distribution for  $x(0)$ . This initial distribution will be the conditional probability of  $x(0)$ , given the prior statistics

$$x(0) \sim N(m_0, \Pi_0)$$

and the posterior observation

$$m_T = x_T + v_T .$$

From (11), we can transform this observation to

$$\phi(0,T)m_T = x(0) - \tilde{x}(0,T) \tag{19}$$

where  $\tilde{x}(0,T)$  is independent of  $x(0)$ , with distribution

$$\tilde{x}(0,T) \sim N(0, \Pi(0,T)) .$$

Hence, the posteriori distribution of  $x(0)$ , given  $M_T$ , is given by

$$x(0) \sim N(\hat{x}_s(0), \Pi_s(0)), \tag{20}$$

where

$$\Pi_S(0) = (\Pi_0^{-1} + \Pi^{-1}(0,T))^{-1} \quad (21)$$

$$\hat{x}_S(0) = \Pi_S(0) (\Pi_0^{-1} m_0 + \Pi^{-1}(0,T) \phi(0,T) m_T) \quad (22)$$

This last expression corresponds to the two-filter Mayne-Fraser smoothing formula [12], [13], obtained from the observations (3) and (4).

Based on these initial conditions, the correspondence between the model in (1), (2) with initial condition (3) and predictive observation (4) and the model in Theorem 1 can be precisely stated.

Theorem 2: Consider the Markovian model of Theorem 1, with initial distribution (20). The probability measure defined on the sample paths  $x(\cdot)$  by this model is equal to the probability measure defined on the sample paths  $x(\cdot)$  by the model (1), (2) with initial condition (3), conditioned on the  $\sigma$ -field  $M_T$ , almost surely.

Proof;

Let  $P_1$  be the distribution induced by (1), and  $P_2$  the distribution induced by (15). By construction, both models yield continuous sample paths with Gaussian statistics. Hence, it is sufficient to show that the mean and covariance of  $x(\cdot)$  under both models is identical. By (20), we know

$$E_{P_1} \{x(0) | M_T\} = E_{P_2} \{x(0)\} = \hat{x}_S(0) \quad \text{a.s.}$$

$$E_{P_1} \{ (x(0) - \hat{x}_s(0)) (x(0) - \hat{x}_s(0))' | M_T \} =$$

$$E_{P_2} \{ (x(0) - \hat{x}_s(0)) (x(0) - \hat{x}_s(0))' \} = \Pi_s(0) .$$

Furthermore,

$$E_{P_1} \{ x(t) | M_T \} = E_{P_1} \{ x(0) | M_T \} + E_{P_1} \left\{ \int_0^t Ax(s) ds | M_T \right\}$$

$$+ E_{P_1} \{ U(s) | M_T \}$$

$$= E_{P_1} \{ x(0) | M_T \} + \int_0^t A E_{P_1} \{ x(s) | M_T \} ds$$

$$+ \int_0^t Q(s) \Pi^{-1}(s, T) (\phi(s, T) M_T - E_{P_1} \{ x(s) | M_T \}) ds$$

because

$$E_{P_1} \{ \tilde{U}(t) | M_T \} = E_{P_1} \{ E_{P_1} \{ \tilde{U}(t) | X_0, M_T \} | M_T \}$$

$$= 0$$

which implies that, for all  $t$ ,

$$E_{P_1} \{ x(t) | M_T \} = E_{P_2} \{ x(t) \} \quad \text{a.s.}$$

An identical argument using Ito's formula [14] establishes

$$E_{P_1} \{ f(x(t)) | M_T \} = E_{P_2} \{ f(x(t)) \}$$

and

$$E_{P_1} \{ f(x(s)) g(x(t)) | M_T \} = E_{P_2} \{ f(x(s)) g(x(t)) \}$$

for  $0 \leq s, t \leq T$ , and for twice differentiable functions  $f, g$ . Hence, the finite dimensional distributions of  $P_1$  conditioned on  $M_T$  and  $P_2$  agree, proving the theorem.

Example: The Pinned Case

Let  $\Pi_T = 0$ , and  $x(T) = m$ . In this case, the value of  $x(t)$  is fixed and the model (15) has the property that  $\Pi(T,T) = 0$ . Thus by denoting  $\delta(t) = x(t) - \phi(t,T)m$  one gets

$$d\delta(t) = A(t) - Q(t)\Pi^{-1}(t,T)\delta(t)dt + d\tilde{U}(t)$$

where  $\Pi^{-1}(t,T) \rightarrow \infty$  as  $t \rightarrow T$ . This means that as  $t \rightarrow T$  the state  $x(t)$  is steered stronger and stronger towards  $m$  and

$$\delta(t) \rightarrow 0 \text{ almost surely as } t \rightarrow T.$$

This property is consistent with previous studies of the pinned Brownian motion (cf. Breiman [14]).

The main feature of the procedure that has been used here to obtain the model (15) with initial conditions (20) - (22) is that it is Bayesian. This means that the model (1)-(2) has been assumed a priori for  $x(\cdot)$ . The advantage of this procedure is that it can be used to incorporate sequentially all the information available on the ship trajectory.

For example, suppose that additional predictive information concerning possible refueling stops is available. That is, we are given

$$x(t_i) \sim N\{m_i, \Pi_i\} \tag{23}$$

with  $0 \leq t_i \leq T$ , for  $i = 1, \dots, N$ .

By an analogous procedure to the construction in Theorem 1, we can construct a Markovian model which incorporates the above predictive information.

Now, assume that real time measurements

$$dy(t) = Cx(t)dt + dV(t) \quad (24)$$

are available, where the noises  $dU(t)$  and  $dV(t)$  are independent. Then, based on the Markovian model of Theorem 1, the optimal estimate of  $x(t)$ , given the predictive information (23) and the real time observations (24), can be obtained from a Kalman filter based on the Markovian model (15) with observation (24).

An alternative approach to obtaining a Markovian model of ship trajectories which included destination information was derived in [2], in the context of simple Brownian motion. In [2], a nonBayesian approach was used to incorporate the predictive information. In the case that we are considering here, a simple generalization of [2] would yield the model

$$d \begin{pmatrix} x(t) \\ x(T) \end{pmatrix} = \begin{pmatrix} A(t) - Q(t)\Pi^{-1}(t,T) & Q(t)\Pi^{-1}(t,T)\phi(t,T) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ x(T) \end{pmatrix} dt + \begin{pmatrix} d\tilde{U}(t) \\ 0 \end{pmatrix} \quad (25)$$

where the initial conditions are given by

$$\begin{pmatrix} x(0) \\ x(T) \end{pmatrix} \sim N \left( \begin{pmatrix} m_0 \\ m_T \end{pmatrix}, \begin{pmatrix} \Pi_0 & 0 \\ 0 & \Pi_T \end{pmatrix} \right)$$

and where  $\tilde{U}(\cdot)$  is a Wiener process with covariance (2). The main aspect of this model is that it is obtained by replacing  $m_T$  by  $x(T)$  in (15) and by setting  $\Pi(T,T) = 0$  in the differential equation for  $\Pi(\cdot,T)$ . To see how this model differs from (15) we need only to note that its transition

matrix is given by

$$\psi(t,s) = \begin{pmatrix} \psi_1(t,s) & \psi_2(t,s) \\ 0 & I \end{pmatrix} \quad (26)$$

where  $\psi_1(t,s) = \Pi(t,T)\phi'(s,t)\Pi^{-1}(s,T)$

$$\psi_2(t,s) = \phi(t,T) - \psi_1(t,s)\phi(s,T) .$$

Then if  $m(t)$  and  $\Pi(t)$  denote respectively the mean and variance of  $x(t)$  one has

$$m(t) = \psi_1(t,0)m_0 + \psi_2(t,0)m_T$$

$$\begin{aligned} \Pi(t) &= \psi_1(t,0)\Pi_0\psi_1^*(t,0) + \psi_2(t,0)\Pi_T\psi_2^*(t,0) \\ &\quad + \int_0^t \psi_1(t,s)Q(s)\psi_1^*(t,s)ds . \end{aligned}$$

Thus, by noting that

$$\psi_1(0,0) = I , \quad \psi_1(T,0) = 0$$

$$\psi_2(0,0) = 0, \quad \psi_2(T,0) = I$$

we see that  $x(0) \sim N(m_0, \Pi_0)$  and  $x(T) \sim N(m_T, \Pi_T)$  exactly. By comparison the initial conditions that were used for the Bayesian model (15) were

$$x(0) \sim N(\hat{x}_S(0), \Pi_S(0))$$

where  $\hat{x}_S(0)$  is a linear combination of  $m_0$  and  $\phi(0,T)m_T$ . This means that the information on the time of arrival of the ship was used to provide

some information on its time of departure. Such a procedure makes sense: if a ship crosses the Atlantic and arrives on a certain date, it should leave at least one week before.

The previous comments illustrate the main advantages of Bayesian vs. nonBayesian methods for the fusion of information

(i) Bayesian methods are sequential (any additional information can be incorporated in the model)

(ii) no piece of information is overweighted with respect to another.

### 3. Predictive Information about Multiple Destinations

The results of Theorems 1 and 2 describe an approach for deriving Markovian models which incorporate predictive information about a single possible destination. However, it is common that predictive information consists of multiple hypotheses concerning the possible destinations of the object of interest. A typical intelligence report might read: "The ship x will be heading for point A with 75% certainty; otherwise, it will head for point B". In this section, we discuss a Bayesian estimation scheme which will incorporate this information, together with some real-time measurements, to obtain the optimal state estimate.

Figure 1 illustrates the case when there exist a number of possible destinations for the ship characterized by

$$H_i: x(t_i) \sim N(m_i, \Pi_i) \quad 1 \leq i \leq N.$$

For each of the possible hypotheses  $H_i$ , we can use the results of Theorems 1 and 2 to obtain the Markovian models

$$dx(t) = (A_i(t)x(t) + s_i(t))dt + \tilde{U}_i(t) \quad (27)$$

where

$$A_i(t) = A - Q(t) \Pi_i^{-1}(t, t_i) \quad (28)$$

$$s_i(t) = Q(t) \Pi_i^{-1}(t, t_i) \phi(t, t_i) m_i \quad (29)$$

where  $\Pi(t, t_i)$  is the solution of (14) with  $\Pi(t_i, t_i) = \Pi_i$ .

The model described by (27), (28), (29) is a Markovian model of  $x(\cdot)$ , conditioned on the information  $m_i$  (that is,  $H_i$  is true), represented as

$$m_i = x(t_i) + v_i$$

$$v_i \sim N(0, \Pi_i) .$$

Suppose now that real time observations are available, to discriminate the hypotheses  $H_i$ . Assume that these observations are given by

$$dy(t) = C_i(t)x(t)dt + R^{1/2}(t)dV(t) \quad (30)$$

where  $V(t)$  is a standard Wiener process independent of  $U(t)$ . The likelihood ratio based on the sequence of measurements  $Y_t = \{y(s), 0 \leq s \leq t\}$  is defined for each pair of hypotheses as

$$L_{ij}(t) = \frac{P(H_i | Y_t)}{P(H_i)} \frac{P(H_j)}{P(H_j | Y_t)} \quad (31)$$

where  $P(H_i)$  denotes the a priori probability of  $H_i$ .

Thus, evaluation of the likelihood ratio between models can be obtained from a Bayesian approach by evaluating the a posteriori probabilities  $P_i(t) = P(H_i | Y_t)$ . These probabilities can be evaluated recursively using Kalman filters (see [16], [17] for a derivation of these equations) as

$$d\hat{x}_i(t) = (A_i(t)\hat{x}_i(t) + s_i(t))dt + \Sigma_i(t)C_i'(t)R^{-1}(t) \cdot (dy(t) - C_i(t)\hat{x}_i(t)dt) \quad (32)$$

$$dP_i(t) = P_i(t)(C_i(t)\hat{x}_i(t) - \sum_{j=1}^N P_j(t)C_j(t)\hat{x}_j(t))'R^{-1}(t) \cdot (dy(t) - \sum_{j=1}^N P_j(t)C_j(t)\hat{x}_j(t)dt) \quad (33)$$

$$\begin{aligned} \frac{d}{dt} \Sigma_i(t) &= A_i(t)\Sigma_i(t) + \Sigma_i(t)A_i'(t) + Q(t) \\ &\quad - \Sigma_i(t)C_i'(t)R^{-1}(t)C_i(t)\Sigma_i(t) \end{aligned} \quad (34)$$

where  $\hat{x}_i(t)$  and  $\Sigma_i(t)$  are the conditional estimate and covariance given that  $H_i$  is true. The initial conditions are

$$\hat{x}_i(0) = \hat{x}_{Si}(0)$$

$$P_i(0) = P(H_i)$$

$$\Sigma_i(0) = \Pi_{Si}(0)$$

where  $\hat{x}_{Si}(0)$  and  $\Pi_{Si}(0)$  are given by (21) and (22). We can now state the main result of this section:

**Theorem 3:** Denote the  $\sigma$ -field generated by the predictive information as  $K$ . Then, the conditional density of  $x(t)$  given the information  $K$ , and the observations  $y(s)$  over  $0 \leq s \leq t$ , is given by

$$p(x,t) = \sum_{i=1}^N P_i(t) p_i(x,t)$$

where

$p_i(x, t)$  is a Gaussian density with mean  $\hat{x}_i(t)$  and variance  $\Sigma_i(t)$ .

Proof: The proof follows directly from Bayes' formula, as, for Borel set  $A$  in  $\mathbb{R}^n$ ,

$$\text{Prob}\{x(t) \in A \mid K, y(s), 0 \leq s \leq t\} =$$

$$\sum_{i=1}^N \text{Prob}\{x(t) \in A \mid H_i, y(s), 0 \leq s \leq t\} \cdot$$

$$\text{Prob}\{H_i \text{ is true} \mid K, y(s), 0 \leq s \leq t\}$$

which establishes the theorem, using Theorems 1 and 2.

The equations (32) - (34) require the parallel processing of the observations  $dy(t)$  by  $N$  Kalman filters. By taking advantage of the structure of the models in Section 2, one can obtain an algorithm where the observations are processed by a single Kalman filter. To do so, we denote the model of the state process before incorporating predictive information as

$$dx(t) = A(t)x(t)dt + dU(t). \quad (35)$$

The hypothesis  $H_i$  corresponds to an added observation of the form

$$m_i = x(t_i) + v_i, \quad v_i \sim N(0, \Pi_i). \quad (36)$$

Hence the estimate  $\hat{x}_i(t)$  is given by

$$\hat{x}_i(t) = E\{x(t) \mid Y_t, m_i\}.$$

Consequently,  $\hat{x}_i(t)$  can be viewed as a smoothed estimate that can be obtained by using the two filter procedure described by Mayne [12], Fraser [13] and Wall, et al. [18] among others. The forward filter processes the information  $Y_t$  as

$$d\hat{x}(t) = A(t)\hat{x}(t)dt + \Sigma(t)C'(t)R^{-1}(t)(dy(t) - C(t)\hat{x}(t)dt) \quad (37)$$

$$\begin{aligned} \frac{d}{dt} \Sigma(t) &= A(t)\Sigma(t) + \Sigma(t)A'(t) + Q(t) \\ &\quad - \Sigma(t)C'(t)R^{-1}(t)C(t)\Sigma(t) \end{aligned} \quad (38)$$

$$\hat{x}(0) = m_0, \quad \Sigma(0) = \Pi_0$$

where the matrices  $C_i(t)$  are assumed independent of  $i$ . Note that the forward filter is the same for all hypotheses. The backward filter for  $H_i$  is given by

$$\frac{d}{dt} \bar{x}_i(t) = A(t)\bar{x}_i(t) \quad (39)$$

$$\frac{d}{dt} \bar{\Sigma}_i(t) = A(t)\bar{\Sigma}_i(t) + \bar{\Sigma}_i(t)A'(t) - Q(t) \quad (40)$$

with the terminal conditions

$$\bar{x}_i(t_i) = m_i, \quad \bar{\Sigma}_i(t_i) = \Pi_i \quad (41)$$

so that we can identify

$$\bar{x}_i(t) = \phi(t, t_i)m_i \quad \text{and} \quad \bar{\Sigma}_i(t) = \Pi(t, t_i) \quad .$$

$\bar{x}_i(t)$  and  $\bar{\Sigma}_i(t)$  can be precomputed off line, since they require no processing of measurements. The forward and backward estimates are combined to yield

$$\hat{x}_i(t) = (\Sigma^{-1}(t) + \bar{\Sigma}_i^{-1}(t))^{-1}(\Sigma^{-1}(t)\hat{x}(t) + \bar{\Sigma}_i^{-1}(t)\bar{x}_i(t)) \quad (42)$$

where  $v_i$  is assumed independent of the Wiener processes  $U(t)$  and  $V(t)$ .

The equation (42) can be substituted in (33) to obtain a recursive expression for  $P_i(t)$  driven by the observations  $dy(t)$  and the forward filter estimates  $\hat{x}(t)$ . Hence, the likelihood ratio given by the equation (31) can be obtained by processing the observations with a single Kalman filter. This implementation has the advantage that the tracking filter produces the same estimate  $\hat{x}(t)$  whether there are predictive hypotheses, or not, until a specific hypothesis is accepted, whereupon a new single model can be produced for the tracking algorithm.

#### 4. Detection of Branching Times

In Sections 2 and 3 we have studied the problem of selecting a tracking model for a trajectory, given some predictive information on the possible destination. However these results were assuming a known origin, i.e., a known branching time. In practice, this branching time is seldom known, as ships are likely to change their mission or their destination while en route. In this section, we develop similar hypothesis testing procedures for processes where the branching times are not known. The basic approach is motivated by previous work by Willsky and Jones [3] for detection of jumps in linear systems using a generalized maximum likelihood technique.

Assume that the original trajectory is modeled by

$$H_0: dx(t) = (A_0(t)x(t) + s_0(t))dt + B(t)dU(t) \quad (43)$$

and that at an unknown time  $\theta$ , the ship changes destination to one of several locations described by Figure 2, yielding N possible models described by equations (27), (28), (29).

We assume that the observations  $dy(t)$  have been stored over a time interval  $[0,t]$ , as well as the state estimate  $\hat{x}_0(t)$ . The problem consists of identifying two parameters: the destination  $i$  and the switch time  $\theta$ . We define the likelihood ratios for a given  $\theta$  as

$$L_{ij}(t; \theta) = \frac{dP(Y_t | \theta, H_i)}{dP(Y_t | \theta, H_j)} \quad (44)$$

For a given  $\theta$ , (44) can be expressed as

$$L_{ij}(t; \theta) = \frac{P(H_i | \theta, Y_t)}{P(H_i | \theta)} \frac{P(H_j | \theta)}{P(H_j | \theta, Y_t)}$$

where the conditional probabilities  $P(H_i | \theta)$  are specified a priori and

$$P(H_i | \theta, Y_t) \triangleq P_i(t; \theta) \quad (45)$$

Separate  $Y_t$  into  $Y_\theta$  and  $Y[\theta, t]$  where

$$Y[\theta, t] = \sigma\{y(s), \theta \leq s \leq t\}$$

Then, one can process  $Y_\theta$  to get an initial distribution for  $x(\theta)$  as

$$P(x(\theta) | Y_\theta) = N(\hat{x}_0(\theta), \Sigma_0(\theta)). \quad (46)$$

The statistics (46) can now be used as initial conditions in (32)-(34) with the initial conditions

$$P_i(\theta; \theta) = P(H_i | \theta)$$

and the information  $Y[\theta, t]$  to obtain  $P_i(t; \theta)$   $0 \leq i \leq N$ . The likelihood ratios can be obtained as in (31), so that

$$L_{i0}(t; \theta) = \exp \frac{1}{2} \int_{\theta}^t (C(s) \hat{x}_i(s) - C(s) \hat{x}_0(s)) R^{-1}(s) (2dy(s) - C(s) (\hat{x}_i(s) + \hat{x}_0(s)) ds) \quad (47)$$

which can be implemented by processing the information  $Y[\theta, t]$  using only one Kalman filter, as indicated in equations (37)-(42). Furthermore, this filter is identical for all values of  $\theta$ . This implies that the estimates  $\hat{x}_i(t)$  are independent of the branching time  $\theta$ .

However, our evolution equations for the identification probabilities  $P_i(t; \theta)$ , as well as the likelihood ratio  $L_{i0}(t; \theta)$ , depend on the value of the parameter  $\theta$ . This suggests a generalized maximum likelihood identification method, where

$$L_0^*(t; \theta) = \text{Max}_{0 \leq i \leq N} L_{i0}(t; \theta) \quad (48)$$

and

$$L_0^*(t) = \text{Max}_{0 \leq \theta \leq t} L_0^*(t; \theta). \quad (49)$$

Other detection rules can be based on observation of the residuals. We have exhibited a simple detection procedure which can be implemented in an efficient manner. Note that the only reprocessing of information is done when evaluating the likelihood ratios. Once a destination has been identified, the new estimate of the state can be obtained from (42) without reprocessing the observations. Furthermore, if the optimal selection of models is independent of  $\theta$  over the range of interest, then the state

estimate  $\hat{x}(t)$  can be obtained independently of equation (49). Thus, our implementation has the advantage that all the processing of information is done by a single Kalman filter, and while the identification procedure is taking place, this filter continues to produce state estimates which can be used for decision purposes.

### 5. A Simple Example

Consider the following two-dimensional Markov process:

$$x_1(t) = w_1(t)$$

$$x_2(t) = w_2(t)$$

where  $w_1, w_2$  are standard independent Brownian motions. Consider the following predictive information:

At time  $t = 14$ , the state  $x(t)$  will be distributed as

$$N\left(\begin{pmatrix} +10 \\ -10 \\ +10 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

with equal probability.

The situation is shown in Figure 3.

Assume that, real-time measurements are taken, according to the equation

$$dz_1(t) = x_1(t)dt + dv_1(t)$$

$$dz_2(t) = x_2(t)dt + dv_2(t)$$

where  $v_1, v_2, w_1, w_2$  are all independent Brownian motions. In this example, the equations (39), (40) decouple in each coordinate, yielding

$$(\bar{x}_i(t), \bar{y}_i(t)) = (\bar{x}_i(14), \bar{y}_i(14))$$

$$\bar{\Sigma}_i(t) = \begin{pmatrix} 15-t & 0 \\ 0 & 15-t \end{pmatrix} \quad \text{for } i = 1, 2, 3, 4 .$$

Furthermore, the Riccati equation (38) decouples into two identical equations of the form

$$\frac{d}{dt} \Sigma(t) = -\Sigma^2 + 1$$

$$\Sigma(0) = 0 .$$

Table 1 shows the evolution of the probabilities when the true destination is (10,10), and the trajectory from the origin to (10,10), is a straight line of constant speed, using a sample of random numbers. The column  $\hat{x}$  indicates the estimated position using no predictive information. The column  $\hat{x}_s$  is the optimal mean, using the predictive information that all four estimations are equally likely. The actual random trajectory is indicated under the column x.

Note the relatively quick identification of the destination, even though the local properties of Brownian motion make the destination difficult to identify. In addition, note the improvement in performance of the filter with predictive information. Even though the average of the apriori predictive information corresponded to no information at all, the use of a recursive hypothesis estimation logic resulted in correct a-posteriori identification which improved the estimator.

The results of Table 1 indicate how the predictive information is useful when the actual mean trajectory is a straight line to the destination.

Assume now that the actual model for the trajectory is

$$\begin{aligned}x_1(t) &= \frac{\sqrt{2}}{2} \left( t - \sin \frac{t\pi}{7} \right) + w_1(t) \\x_2(t) &= \frac{\sqrt{2}}{2} \left( t + \sin \frac{t\pi}{7} \right) + w_2(t)\end{aligned}\tag{50}$$

This trajectory is depicted in Figure 4.

Table 2 shows the resulting evolutions of the probabilities, the sample path of the process, and the filtered estimates with and without predictive information.

Although the identification of the correct destination occurs promptly, the results in Table 2 suggest that incorporating the correct predictive destination does not improve the performance of the estimator as much as it did in Table 1. This is the result of the mismatch between the linear models, incorporating predictive information, and the nonlinear trajectories generated by (50).

## 6. Conclusions

This paper has shown how some information about the possible destinations of a ship can be incorporated in the modeling of its trajectories. These results are potentially useful for the tracking of multiple objects (cf. [19], [20]). One possible approach is to generate multiple models based on the likely associations of data and use a sequential evaluation of these models as described in Section 3. In addition, when a region is congested with traffic, the additional knowledge of the destination of the various ships in the area is very valuable for the track reconstruction problem (one needs only to match the outgoing tracks with the ships' destinations).

Another appealing feature of the algorithms presented here is their capacity to evaluate the validity of predictive information. In particular, Sections 3 and 4 discuss algorithms where the predictive information is not used in the tracking filter until a likelihood ratio test confirms its validity.

Acknowledgements. The authors are very grateful to Prof. N. Sandell, Jr. for suggesting the problem studied in this paper, to Prof. F. Schweppe for some valuable comments, and to Mr. Lawrence Ng, for developing the simulation results in Section 5.

#### References

- [1] G. Verghese and T. Kailath, "A further note of backwards Markovian models," IEEE Trans. Inform. Theory, Vol. IT-25, No. 1, pp. 121-124, January 1979.
- [2] J.R. Weisinger, "A transistor model: theory," Daniel H. Wagner, Associates, Interim memorandum to Naval Electronic Systems Command, November, 1978.
- [3] A.S. Willsky and H.L. Jones, "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," IEEE Trans. Automat. Contr., Vol. AC-21, no. 1, pp. 108-112, February 1976.
- [4] R.A. Singer, "Estimating optimal tracking filter performance for manned maneuvering targets," IEEE Trans. Aerosp. Electron. Syst., Vol. AES-6, no. 4, pp. 473-483, July 1970.
- [5] R.A. Singer and K.W. Behnke, "Real-time tracking filtering evaluation and selection for tactical applications," IEEE Trans. Aerosp. Electron. Syst., Vol. AES-7, no. 1, pp. 100-110, January 1971.
- [6] D.R. Morgan, "A target trajectory noise model for Kalman trackers," IEEE Trans. Aerosp. Electron. Syst., Vol. AES-12, no. 3, pp. 405-408, May 1976.

- [7] P.A. Meyer, Probability and Potentials, Waltham, Massachusetts: Blaisdell, 1966.
- [8] E. Wong, "Recent progress in stochastic processes," IEEE Trans. Inform. Theory, Vol. IT-19, no. 3, pp. 262-275, May 1973.
- [9] L. Ljung and T. Kailath, "Backwards Markovian models for second-order stochastic processes," IEEE Trans. Inform. Theory, Vol. IT-22 no. 4, pp. 488-491, July 1976.
- [10] D.G. Lainiotis, "General backwards Markov models," IEEE Trans. Automat. Contr. Vol. AC-21, no. 4, pp. 595-598, August 1976.
- [11] A.S. Willsky, "Covariance analysis and estimation algorithms for spatially-distributed random data," TASC internal memorandum, April 1979.
- [12] D.Q. Mayne, "A solution of the smoothing problem for linear dynamic systems," Automatica, Vol. 4, no. 2, pp. 73-92, November 1966.
- [13] D.C. Fraser, "A new technique for the optimal smoothing of data," Sc.D. Dissertation, Mass. Inst. of Technology, Cambridge, MA, January 1967.
- [14] L. Breiman, Probability, Reading Mass.: Addison-Wesley, 1968.
- [15] H. Van Trees, Detection, estimation and modulation theory, New York: Wiley, 1968.
- [16] K.P. Dunn, "Measure transformation, estimation, detection and stochastic control," Ph.D. Thesis, Washington University, St. Louis, MO, 1974.
- [17] M. Athans, D. Castanon et. al., "The stochastic control of the F-8C aircraft using a multiple model adaptive control method," IEEE Trans. Automat. Contr., Vol. AC-22, no. 5, pp. 768-780, October 1977.
- [18] J.E. Wall, Jr., A.S. Willsky and N.R. Sandell, Jr., "On the fixed-interval smoothing problem," Stochastics, Vol. 5, nos. 1 and 2, pp. 1-42, 1981.
- [19] Y. Bar-Shalom, "Tracking methods in a multi-target environment," IEEE Trans. Automat. Contr., Vol. AC-23, no. 4, pp. 618-626, August 1978.
- [20] C.L. Morefield, "Application of 0-1 integer programming to multi-target problems," IEEE Trans. Automat. Contr., Vol. AC-22, no. 3, pp. 302-312, June 1977.

Table 1

Time	$P_1$	$P_2$	$P_3$	$P_4$	$x_1$	$x_2$	$\hat{x}_1$	$\hat{x}_2$	$\hat{x}_{1s}$	$\hat{y}_{2s}$
.6	.03	.43	.51	.03	-.37	.43	-.02	.38	-.04	.92
1.2	0	.75	.25	0	.93	.15	.14	.76	.48	1.38
1.8	0.05	.95	0	0	1.0	.14	.74	.37	1.4	1.0
2.4	.08	.92	0	0	2.7	.23	1.5	.29	2.2	.96
3.0	.01	.99	0	0	2.5	1.8	2.5	.56	3.0	1.26
3.6	0	1	0	0	2.3	2.6	2.5	.97	3.1	1.7
5.0	0	1	0	0	2.4	4.93	1.9	4	2.6	4.6

Table 2

Time	$\frac{P_1}{P_1}$	$\frac{P_2}{P_2}$	$\frac{P_3}{P_3}$	$\frac{P_4}{P_4}$	$\frac{x_1}{x_1}$	$\frac{x_2}{x_2}$	$\frac{\hat{x}_1}{\hat{x}_1}$	$\frac{\hat{x}_2}{\hat{x}_2}$	$\frac{\hat{x}_{1s}}{\hat{x}_{1s}}$	$\frac{\hat{x}_{2s}}{\hat{x}_{2s}}$
.6	.03	.19	.66	.12	-.38	-1.6	-.17	.24	-.45	.83
1.2	0	.88	.12	0	1.17	1.13	.28	.73	1.09	1.35
1.8	0	.95	.05	0	+1.56	.65	.38	.86	1.06	1.5
2.4	0	1	0	0	2.1	3.7	1.1	1.4	1.7	2.0
3.0	0	1	0	0	.5	3.3	1.3	2.1	2.0	2.78
5.0	0	1	0	0	3.3	5.4	2.1	4.1	2.8	4.6
7.0	0	1	0	0	7.0	4.2	4.7	4.6	5.3	5.2
8.0	0	1	0	0	5.3	5.8	5.5	4.4	6.1	5.1

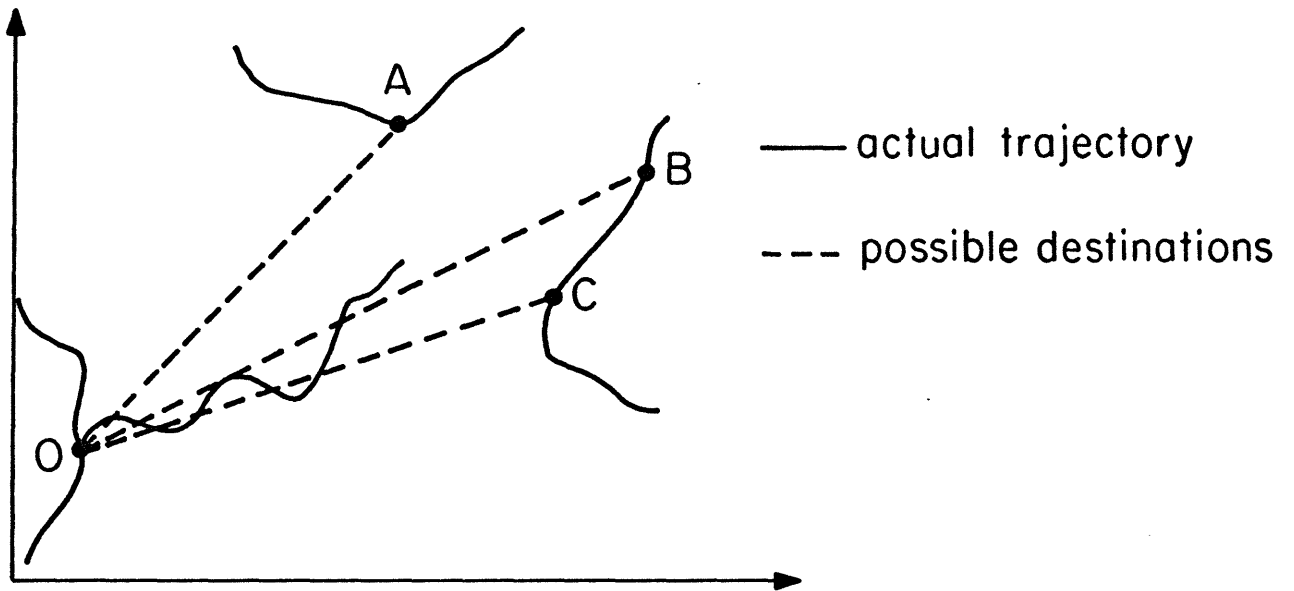


Figure 1. Determination of the ship's destination from multiple hypotheses.

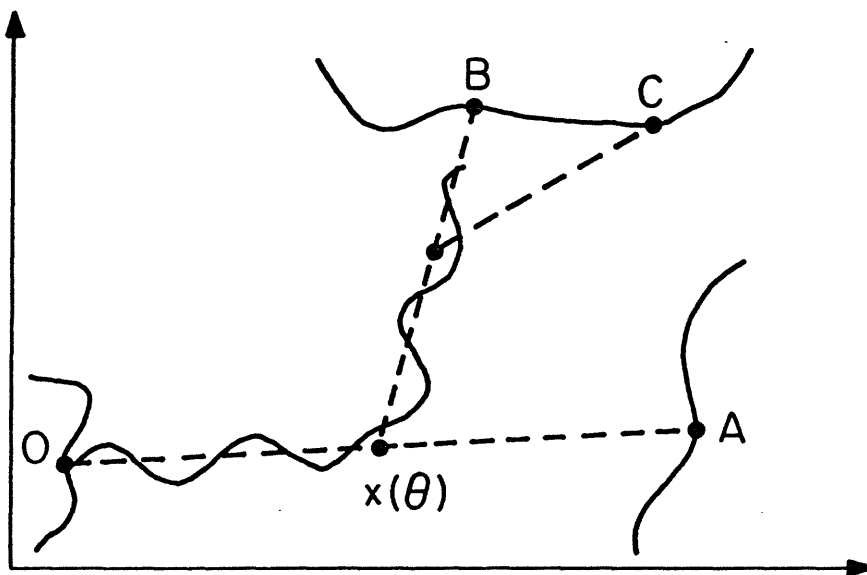


Figure 2. The detection of maneuvers.

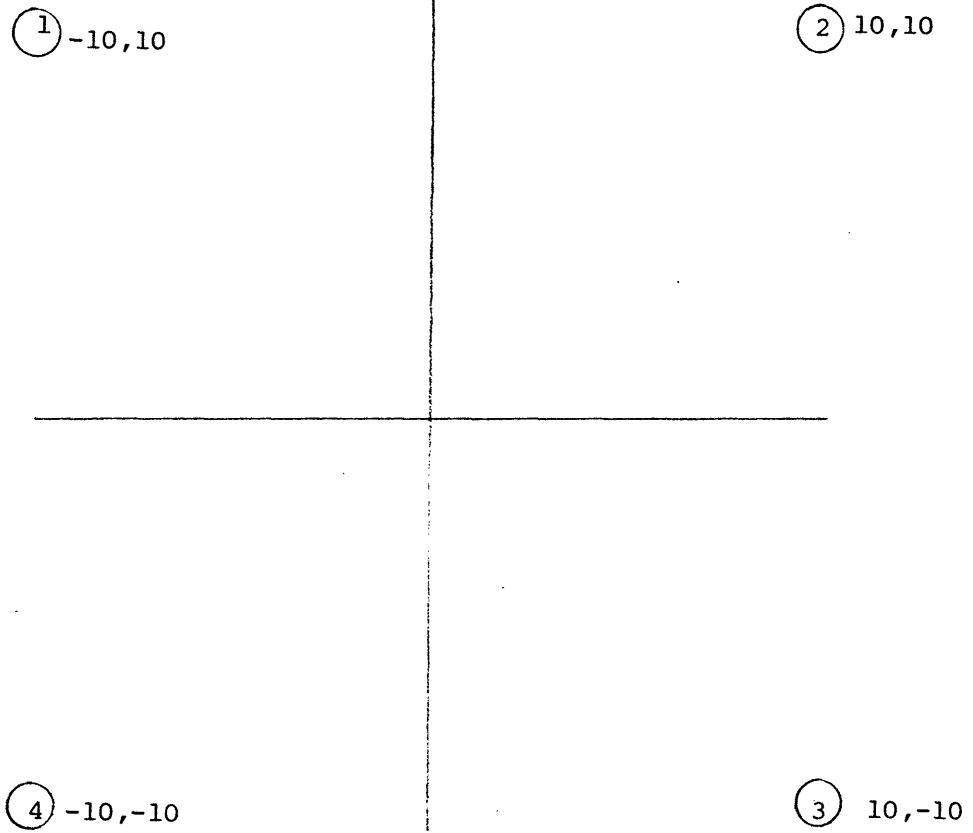


Figure 3. Possible Destinations for Example 1

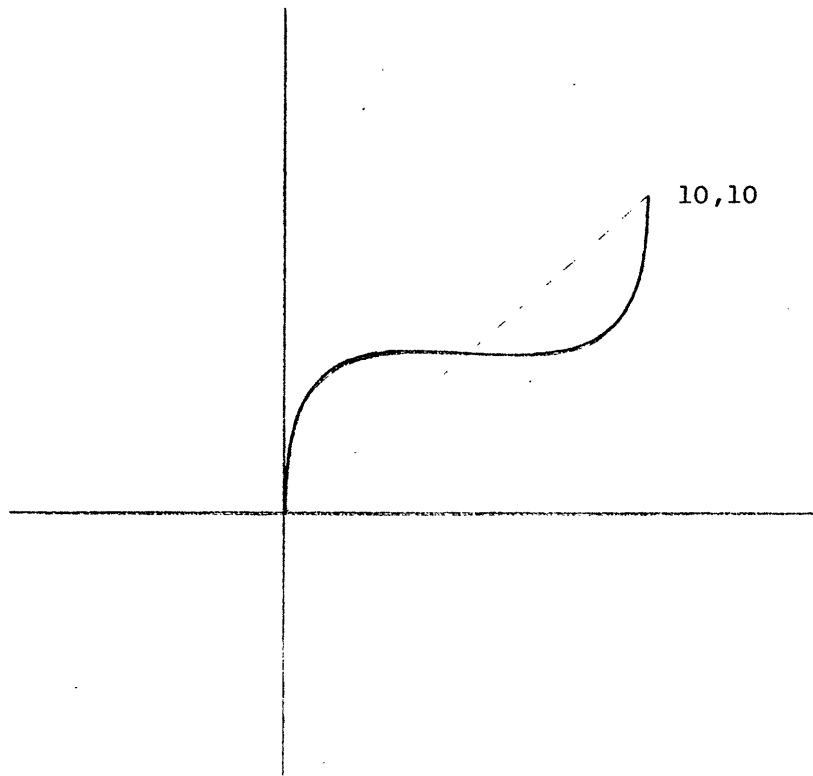


Figure 4. Mean Ship trajectory for Table 2.