Distributed Estimation Algorithms for Nonlinear Systems

DAVID A. CASTANON AND DEMOSTHENIS TENEKETZIS

Abstract—**In this paper, we consider the problem of combining the local conditional distributions of a random variable which have been generated by local observers having access to their private information. Sufficient statistics for the local distributions are communicated to a coordinator, who attempts to reconstruct the global centralized distribution using only the communicated statistics. We obtain a distributed processing algorithm which recovers exactly the centralized conditional distribution. The results can be applied in designing distributed hypothesis-testing algorithms for event-driven systems.**

I. Introduction

CONSIDER the following estimation problem. The state trajectory of a random process is observed by K distinct observers, using noise-corrupted observations. Each observer processes his own observation history to obtain the local conditional distribution of the state as a function of time. Assume that sufficient statistics representing each local conditional distribution are communicated to a coordinator at a central location at each point in time. The coordinator's estimation problem consists of constructing the overall conditional distribution of the state, conditioned on knowing all of the observations, while using only the sufficient statistics communicated to him. The above estimation structure is illustrated in Fig. 1.

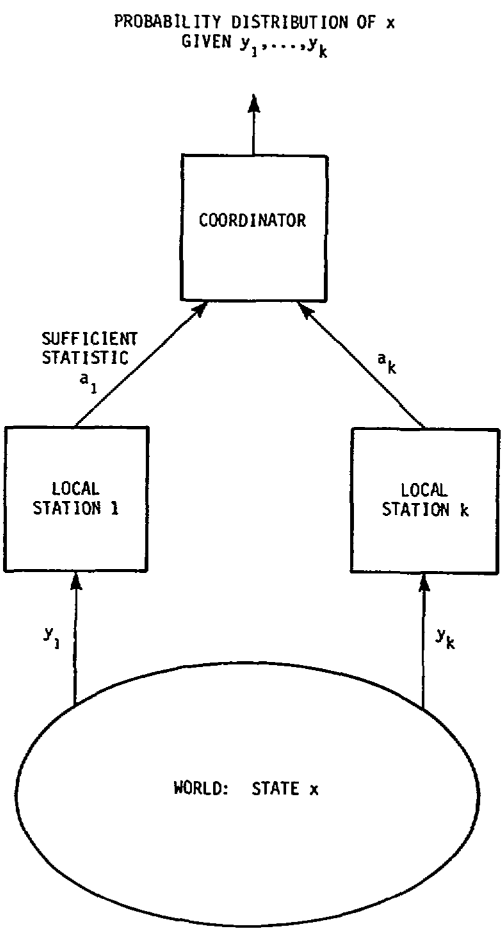


Fig. 1.

When the state process is a Gauss–Markov process, and the local observations are linear measurements of the state corrupted by the noise, the solution to the coordinator's problem has been obtained by many authors, notably Speyer [1], Chong [2], and Willsky *et al.* [3]. In this case, the sufficient statistic is provided by the local conditional mean and covariance. The results of [1] and [2] show that the centralized conditional mean and covariance can be obtained using linear operations on the local estimates and covariances. The results of [3] extend these results to consider problems in optimal smoothing, as well as problems where the local models used in producing local estimates differ from the true global model available to the coordinator.

In this paper, we extend the results of [1]–[3] to include general Markov stochastic processes. A major assumption in our results is that all agents have common knowledge of the *a priori* statistics of the uncertainties in the system. We deal both with discrete-time and continuous-time Markov processes. As an introduction, we derive the solution of the coordinator's problem for discrete-time finite-state Markov processes using Bayes' rule. The structure of the coordinator's solution is a generalization of the results of [2] to finite-state Markov processes. We then extend these results for the problem of estimating a continuous-time Markov process observed through additive white noise. We use recent results in nonlinear filtering [4], [5] to characterize the evolution of the local and the centralized conditional probability densities of the state. Based on the solution structure for the discrete-time case, we develop the equations for the optimal coordination algorithm.

The results for the continuous-time coordination problem are illustrated for the problem of designing optimal hierarchical estimation and parameter estimation algorithms for a class of event-driven systems. These algorithms provide the basis for designing distributed hypothesis testing algorithms which perform as well as centralized algorithms.

The rest of the paper is organized as follows. In Section II, we introduce and solve the discrete-time formulation of the coordinator's problem; Theorem 1 in that section presents the solution to the coordinator's problem. In Section III, we introduce the continuous time version of the coordinator's problem, and we solve it in Theorem 2. Section IV illustrates the applications of the results of Section III to the problem of parameter estimation. Section V discusses the possibility of periodic or asynchronous implementation of the coordinator's algorithms. Section VI summarizes the results and indicates directions of future research.

Manuscript received June 23, 1983; revised June 5, 1984. Paper recommended by Past Associate Editor, P. R. Kumar. This work was supported by the Department of Energy under Contract DOE AC01-80-RA50418.

D. A. Castanon is with ALPHATECH, Inc., Burlington, MA 01803.
D. Teneketzis was with ALPHATECH, Inc., Burlington, MA 01803. He is now with the Department of Electrical Engineeing and Computer Science, University of Michigan, Ann Arbor, MI.

0018-9286/85/0500-0418$01.00 © 1985 IEEE

II. A Motivating Example: Finite-State Discrete-Time Problems

Consider the diagram of Fig. 1. Assume that the state process is the finite-state hybrid process z_t, evolving in discrete time, where z_t is a member of a finite space $S = (1, \cdots, N)$. Denote the transition probability distribution of z_t as

$$P_t(z_{t+1};\ z_t) \triangleq P(z_{t+1} \mid z_t). \tag{2-1}$$

There are K local agents taking observations of the state process. The observations of agent i at stage t are measurements y_t^i of z_t through memoryless channels with values in a finite space S^i, described by the transition probability

$$Q_t^i(c \mid z_t) \triangleq P(y_t^i = c \mid z_t). \tag{2-2}$$

In order to properly define the estimation problem, we must establish the relationship between the different observation channels. Let (Ω, F, P) be the underlying probability space for the random sequences $z_t, y_t^i, i = 1, \cdots, K, t = 0, \cdots, T$. We make the following assumption.

1) Under P, the memoryless channels of (2-2) are mutually independent.

Now, consider the estimation problem of each local agent. The purpose of each local agent is to produce the conditional distribution of z_t, given all of the past observations $y_0^i, \cdots, y_t^i$. Define the notation Y_t^i as

$$Y_t^i = \{y_0^i, \cdots, y_t^i\}$$
$$Y_t = \{Y_t^i,\ i = 1, \cdots, K\}.$$

Assumption 1) implies that the random variables $y_t^1, \cdots, y_t^K$ are conditionally independent given z_t. That is,

$$P(y_t^1, \cdots, y_t^K \mid z_t) = \prod_{i=1}^{K} Q_t^i(y_t^i \mid z_t). \tag{2-3}$$

Furthermore, the memoryless nature of the channels implies that

$$P(y_t^i \mid z_t, y_{t-1}^i, \cdots, y_0^i) = P(y_t^i \mid z_t). \tag{2-4}$$

Then, agent i's problem is solved recursively using Bayes' rule as [7]

$$P(z_t \mid Y_t^i) = \frac{P(y_t^i \mid z_t) P(z_t \mid Y_{t-1}^i)}{\sum_{z_t \in S} P(y_t^i \mid z_t) P(z_t \mid Y_{t-1}^i)} \tag{2-5}$$

where $P(z_t \mid Y_t^i)$ denotes the distribution $P(z_t = z \mid Y_t^i)$ throughout this paper. The derivation of these equations is standard, and can be found in [7] or [8].

The coordinator's problem can be described as follows. At each t, each agent i communicates the distribution $P(z_t \mid Y_t^i)$ for all z to the coordinator. The coordinator must use the information in these communications to estimate the state z_t.

Let $u_t^i(z_t)$ be the communications from agent i at time t, and

$$U_t^i \triangleq \{u_0^i, \cdots, u_t^i\};\ U_t \triangleq \{U_t^1, \cdots, U_t^K\}.$$

The coordinator's problem is to develop an algorithm for obtaining $P(z_t \mid U_t)$. Rather than deriving $P(z_t \mid U_t)$, we will establish that $P(z_t \mid Y_t)$, the conditional probability distribution assuming that all measurements were available to the coordinator, can be obtained as a function of U_t. Since U_t is a function of Y_t by definition, this establishes that $P(z_t \mid Y_t) = P(z_t \mid U_t)$. We call the map from U_t to $P(z_t \mid Y_t)$ the coordinator's algorithm. This algorithm is described in the following theorem.

Theorem 1: Under assumption 1), the coordinator can reconstruct the centralized conditional probability distribution using the recursive algorithm

$$P(z_t \mid Y_t) = \frac{\prod_{i=1}^{K} u_t^i(z_t)}{c_t(z_t)} \cdot \frac{1}{\sum_{z \in S} \frac{\prod_{i=1}^{K} u_t^i(z)}{c_t(z)}} \tag{2-6}$$

where

$$c_t(z_t) = \frac{\prod_{i=1}^{K} \sum_{z_{t-1} \in S} P_t(z_t;\ z_{t-1}) u_{t-1}^i(z_{t-1})}{\sum_{z_{t-1} \in S} P_t(z_t;\ z_{t-1}) P(z_{t-1} \mid Y_{t-1})}. \tag{2-7}$$

Proof:

$$P(z_t \mid Y_t) = \frac{P(y_t^1, \cdots, y_t^K \mid z_t) P(z_t \mid Y_{t-1})}{\sum_{z_t \in S} (P(y_t^1, \cdots, y_t^K \mid z_t) P(z_t \mid Y_{t-1})} \tag{2-8}$$

$$= k_t(Y_t) \left(\prod_{i=1}^{K} Q_t^i(y_t^i \mid z_t) \right) (P(z_t \mid Y_{t-1}) \tag{2-9}$$

where $k_t(Y_t)$ is a normalization constant, using assumption 1). From the expression for the optimal local estimates in (2-5), we obtain

$$Q_t^i(y_t^i \mid z_t) = k_t^i(Y_t^i) \cdot \frac{P(z_t \mid Y_t^i)}{P(z_t \mid Y_{t-1}^i)}. \tag{2-10}$$

Combining (2-9) and (2-10) establishes the theorem, where the normalization constants have been explicity evaluated. Note that (2-7) is a function of U_{t-1}; thus, (2-6) expresses $P(z_t \mid Y_t)$ as a function of U_t.

The structure of the coordinator's solution in Theorem 1 is interesting. Basically, the coordinator accounts for the presence of correlations between the local estimators, due to the fact that they all observe the same state process, by compensating the product of the local conditional densities with the ratio of the centralized conditional predicted density and the product of the local conditional predicted densities. Formally,

$$\frac{p(z_t \mid Y_t)}{\prod_{i=1}^{K} p(z_t \mid Y_t^i)} = K_t \frac{p(z_t \mid Y_{t-1})}{\prod_{i=1}^{K} p(z_t \mid Y_{t-1}^i)} \tag{2-11}$$

where K_t is a proportionality constant to account for the normalization of the probability densities.

The results of Theorem 1 can be generalized to arbitrary discrete-time Markov processes as long as an independence assumption equivalent to assumption 1) is included, and enough regularity assumptions are made to guarantee the existence of the conditional distributions. Rather than developing these extensions here, we will focus the rest of this paper on distributed estimation algorithms for continuous-time Markov processes.

III. The Continuous-Time Coordinator's Problem

Assume that the state process can be described by the stochastic differential equation

$$dx_t = f(t, x_t, \rho_t)dt + \sigma(t, x_t, \rho_t)dw_t \tag{3-1}$$

where w_t is a standard Brownian motion with values in R^n, and where ρ_t is an element of a finite set $S = \{1, \cdots, N\}$ whose transitions are described in terms of the infinitesimal rates

$$P\{\rho_{t+\Delta}=i \mid \rho_t=j,\ x_t=x\}=\lambda_{ji}(x)\Delta+0(\Delta). \tag{3-2}$$

Under appropriate assumptions, (3-1) and (3-2) define the evolution of a strong Markov process $(x_t, \rho_t) = z_t$ with values in $R^n \times S$. For our purposes, we assume the following.

2) $\sigma(t, x, \rho)$, $f(t, x, \rho)$, and $\lambda_{ji}(x)$ are sufficiently regular to guarantee the existence of a strong Markov process $z_t = (x_t, \rho_t)$ which is Feller continuous for any x_0, ρ_0 [9].

3) The initial distribution of z_0 is known and independent of w.

4) $\lambda_{ij}(x) \geqslant \epsilon > 0$ for all i, j.

Suppose that each local station $i = 1, \cdots, K$ has measurements or the state process z_t described by the stochastic differential equation

$$dy_t^i = h^i(t, x_t, \rho_t)dt + dv_t^i. \tag{3-3}$$

We assume additionally the following.

5) The processes v^i are mutually independent standard Brownian motions which are also independent of w and z_0.

6) The functions f, σ, and h^i are smooth enough so that conditional probability densities of z_t given the local information y_s^i, $0 \leqslant s \leqslant t$ exist for each station i.

Conditions which guarentee the existence and smoothness of the conditional probability density function can be found in [4] and [5]. Typically, the functions h^i will be assumed to be uniformly bounded with bounded x derivatives.

Under these assumptions, the solution to each agent's problem can be obtained using the equations for optimal nonlinear filtering [4], [5]. As before, denote the information available at time t as

$$Y_t^i = \{y_s^i,\ 0 \le s \le t\}$$
$$Y_t = \{Y_t^1, \cdots, Y_t^K\}.$$

Define A_t as the differential operator on functions on $R^n \times S$ into R as

$$A_t v(x, \rho) = \sum_{i=1}^{n} f_i(t, x, \rho) \frac{\partial}{\partial x_i} v(x, \rho) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(t, x, \rho) \frac{\partial^2}{\partial x_i \partial x_j} v(x, \rho) + \sum_{i=1}^{K} \lambda_{\rho i}(x)(v(x, i) - v(x, \rho)) \tag{3-4}$$

where $a(t, x, \rho) = \sigma(t, x, \rho)\sigma^T(t, x, \rho)$.

The solution to each local agent's problem is described by Zakai's equation [10] for the unnormalized conditional density of the state z_t given the observations Y_t^i. Let $q_t^i(z)$ denote this density for the local station i. Then

$$dq_t^i(z) = A_t^*(z)q_t^i(z)dt + h_t^i(z)^T q_t^i(z)dy_t^i \tag{3-5}$$

where A_t^* is the formal adjoint of the operator A_t defined in (3-4). The differentials used in (3-5) are the Ito differentials; in Stratonovich form [11], using symmetric differentials, (3-5) becomes

$$\bar{d}q_t^i(z) = (A_t^*(z) - 1/2 h_t^{iT}(z)h_t^i(z))q_t^i(z)dt + h_t^{iT}(z)q_t^i(z)\bar{d}y_t^i \tag{3-6}$$

where we have adopted the notation $\bar{d}$ of [11] to indicate the Stratonovich symmetric differential.

We formulate the coordinator's problem as follows. At each time t, the coordinator receives, from each location i, a multiple of $q_t^i(z)$ or a sufficient statistic which enables him to reconstruct $q_t^i(z)$ exactly. As in the previous section, our objective is to develop an algorithm whereby the coordinator can reconstruct $P(z_t \in A \mid Y_t)$ from the histories $q_s^i(z)$, $i = 1, \cdots, K$, $s \leqslant t$. Denote the coordinator's received message as

$$a_t^i(z) = k_t^i q_t^i(z)$$

where k_t^i is the scale constant at time t.

The solution of the coordinator's problem is summarized in Theorem 2.

Theorem 2: Under assumptions 2)–6), the coordinator can reconstruct the centralized conditional probability distribution recursively as

$$P(z_t \in A \mid Y_t) = \frac{\int_A q_t(z)\, dz}{\int_{R^n \times S} q_t(z)\, dz}$$

where

$$q_t(z) = \frac{\prod_{i=1}^{K} a_t^i(z)}{C_t(z)}$$

and C_t satisfies

$$\frac{1}{C_t} \frac{dC_t(z)}{dt} = \sum_{i=1}^{K} \frac{A_t^* a_t^i(z)}{a_t^i(z)} - \frac{C_t(z)}{\prod_{i=1}^{K} a_t^i(z)} A_t^* \frac{\prod_{i=1}^{K} a_t^i(z)}{C_t(z)} \tag{3-7}$$

$$C_0(z) = P_0(z)^{K-1}$$

where $P_0(z)$ is the initial density of z_0.

Proof: Since the Brownian motions v_t^i are mutually independent, by assumption 5), the Zakai equation for the unnormalized conditional density $q_t(z)$ is given by

$$dq_t(z) = A_t^*(z)q_t(z)dt + \sum_{i=1}^{K} h_t^i(z)^T dy_t^i q_t(z) \tag{3-8}$$

which, when converted to Stratonovich form, reads

$$\bar{d}q_t = \left(A_t^*(z) - \frac{1}{2} \sum_{i=1}^{K} h_t^i(z)^T h_t^i(z)\right) q_t(z)dt + \sum_{i=1}^{K} h_t^i(z)^T \bar{d}y_t^i q_t(z). \tag{3-9}$$

Now define

$$C_t(z) = \frac{\prod_{i=1}^{K} q_t^i(z)}{q_t(z)}. \tag{3-10}$$

Then, using the calculus of Stratonovich differentials,

$$\bar{d}C_t(z) = \sum_{i=1}^{K} C_t(z) \frac{\bar{d}q_t^i(z)}{q_t^i(z)} - C_t(z) \frac{\bar{d}q_t(z)}{q_t(z)}. \tag{3-11}$$

Hence, using (3-6) and (3-9),

$$\frac{dC_t(z)}{C_t(z)}=\sum_{i=1}^{K}\frac{A_t^*(z)q_t^i(z)}{q_t^i(z)}\,dt-\frac{A_t^*(z)q_t(z)}{q_t(z)}\,dt. \tag{3-12}$$

Equation (3-12) defines $C_t(z)$ as a deterministic equation, driven by $q_t^i(z)$ and $q_t(z)$. Substituting (3-10) with (3-12) and multiplying the numerator and denominator of the first sum in (3-12) by k_t^i results in (3-7).

A direct analogy between Theorems 1 and 2 can be drawn as follows:

$$\frac{A_t^*q_t^i(z)}{q_t^i(z)}$$

correspond to predictors in the discrete-time case, as in the heuristic formula

$$\ln\,(P_{t+dt}(z\mid Y_t^i))=\ln\,P_t(z\mid Y_t^i)+\frac{A_t^*q_t^i(z)}{q_t^i(z)}\,dt. \tag{3-13}$$

Furthermore,

$$\frac{-A_t^*q_t(z)}{q_t(z)}\,dt=\ln\,P_t(z\mid Y_t)-\ln\,(P_{t+dt}(z\mid Y_t)). \tag{3-14}$$

Substituting into (3-7) gives (formally)

$$\ln\,C_{t+dt}(z)-\ln\,C_t(z)$$

$$=\ln\frac{\prod_{i=1}^{K}P_{t+dt}(z\mid Y_t^i)}{P_{t+dt}(z\mid Y_t)}-\ln\,C_t(z)$$

$$-\ln\,K_t \tag{3-15}$$

where K_t is a constant factor due to the scaling in the definition of $C_t(z)$. Hence,

$$C_{t+dt}(z)=K_t\,\frac{\prod_{i=1}^{K}P_{t+dt}(z\mid Y_t^i)}{P_{t+dt}(z\mid Y_t)} \tag{3-16}$$

which is the ratio of the product of the decentralized predictors to the centralized predictor. Thus, C_t is the same compensator as was used in Theorem 1.

IV. Application to Distributed Parameter Estimation

In this section, we apply the results of Section III to the problem of parameter estimation of continuous-time hybrid systems.

Assume that at time $t=0$ an event H occurs where

$$H\in\{1,2,\cdots,N\} \tag{4.1}$$

and $P(H=i)(i=1,2,\cdots,N)$ is known *a priori*. The event H determines the parameters of evolution of a Gauss–Markov continuous-time R-valued[1] process described by

$$dx_t=A(H)x_tdt+b(H)dw_t. \tag{4-2}$$

This process is being observed by K independent stations whose measurements are described by

$$dy_t^i=C_t^i(H)x_tdt+dv_t^i,\qquad i=1,2,\cdots,K. \tag{4-3}$$

[1] For clarity of the presentation of results, we restrict attention to scalar systems. The results can be easily generalized to R^n-valued Gauss–Markov processes.

The assumptions of Section III are assumed to hold. In addition, we assume the following.

7) x_0 is a Gaussian random variable conditioned on H, with mean $x_0(H)$ and covariance $\sigma_0(H)^2$.

8) The initial probability distribution of H is given by $P_0(H)$, and is positive for all H.

The state of the hybrid system at any time t is

$$z_t\triangleq(H,x_t). \tag{4-4}$$

The local conditional density of z_t given the past measurements Y_t^i can be summarized by the finite dimensional statistic

$$S_t^i=\begin{bmatrix}P(H=1\mid Y_t^i)\\P(H=2\mid Y_t^i)\\ \cdot\\ \cdot\\ \cdot\\ P(H=N\mid Y_t^i)\\ x_t^i(1)\\ \cdot\\ \cdot\\ \cdot\\ x_t^i(N)\end{bmatrix} \tag{4-5}$$

where

$$P_t^i(k)\triangleq P(H=k\mid Y_t^i) \tag{4-6}$$

is the conditional probability that $H=k(k=1,2,\cdots,N)$ given Y_t^i and

$$\hat{x}_t^i(k)=E\{x_t\mid H=k,\ Y_t^i\}. \tag{4-7}$$

The local conditional density of z_t given Y_t^i is given by

$$P(z_t\mid Y_t^i)=P_t^i(H)\frac{1}{\sqrt{2\pi}\sigma_t^i(H)}\exp\left\{-\frac{(x_t-\hat{x}_t^i(H))^2}{2\sigma_t^i(H)^2}\right\} \tag{4-8}$$

where

$$\sigma_t^i(H)^2=E\{(x_t-\hat{x}_t^i(H))^2\mid Y_t^i,\ H\} \tag{4-9}$$

can be computed *a priori* from the parameters of the problem, and

$$P_t^i(H)=\frac{L_t^i(H)P_0(H)}{\sum_k L_t^i(k)P_0(k)} \tag{4-10}$$

where $L_t^i(H)$ is the likelihood of event H satisfying

$$dL_t^i(H)=L_t^i(H)C_t^i(H)\hat{x}_t^i(H)dy_t^i. \tag{4-11}$$

A simple derivation of these equations is provided in [12].

The conditional density of the hybrid state z_t given Y_t (i.e., the centralized conditional density) has a sufficient statistic defined by

$$S_t=\begin{bmatrix}P_t(1)\\ \cdot\\ \cdot\\ \cdot\\ P_t(N)\\ \hat{x}_t(1)\\ \cdot\\ \cdot\\ \cdot\\ x_t(N)\end{bmatrix} \tag{4-12}$$

where

$$P_t(l)=P(H=l \mid Y_t) \tag{4-13}$$

$$\hat{x}_t(l)=E\{x_t \mid Y_t,\ h=l\}. \tag{4-14}$$

The coordinator's task is to attempt to reconstruct S_t based on the past reports $S_s^i (i = 1, 2, \cdots, K)$, $s \le t$, that he receives from the local observation stations.

The system of (4-2), (4-3) satisfies all the assumptions of Section III. Thus, according to Theorem 2, the coordinator can reconstruct the centralized conditional density of the state z_t based on the reports S_t^i that he receives from the local observation stations. The problem of parameter estimation consists of constructing the conditional distribution $P_t(H)$ over the set of possible parameters H. The coordinator's optimal algorithm is given below.

Theorem 3: The conditional probabilities $P_t(H)$ of the events are given by

$$P_t(H)=\frac{Q_t(H)}{\sum_l Q_t(l)} \tag{4-15}$$

where

$$Q_t(H)=\frac{\exp\left[-1/2\left\{\sum_{i=1}^{K}\frac{\hat{x}_t^i(H)^2}{\sigma_t^i(H)^2}-\frac{\hat{x}_t(H)^2}{\sigma_t(H)^2}\right\}\right]}{f_1(t,H)\exp\left[-\frac{f_2(t,H)^2}{2f_3(t,h)}\right]}$$

$$\cdot\frac{\sigma_t(H)}{\prod_{i=1}^{K}\sigma_t^i(H)}\prod_{i=1}^{K}P_t^i(H) \tag{4.16}$$

and $f_1(t, H)$, $f_2(t, H)$, $f_3(t, H)$ are determined by the following equations:

$$\frac{d}{dt}\cdot\frac{1}{f_3(t,H)}=-2A(H)\frac{1}{f_3(t,H)}-2b^2(H)\frac{1}{f_3(t,H)}\sum_{i=1}^{K}\frac{1}{\sigma_t^i(H)^2}$$

$$+b^2(H)\frac{1}{f_3(t,H)^2}+2b^2(H)\sum_{i=1}^{K}\sum_{j=1}^{i-1}\frac{1}{\sigma_t^i(H)^2\sigma_t^j(H)^2} \tag{4-17}$$

$$f_3(0,H)=\frac{\sigma_0(H)^2}{K-1} \tag{4-18}$$

$$\frac{d}{dt}f_2(t,H)=\left\{A(H)-2b^2(H)f_3(t,H)\sum_{i=1}^{K}\sum_{j=1}^{i-1}\frac{1}{\sigma_t^j(H)^2}\frac{1}{\sigma_t^i(H)^2}\right.$$

$$\left.+b^2(H)\sum_{i=1}^{K}\frac{1}{\sigma_t^i(H)^2}\right\}f_2(t,H)$$

$$-b^2(H)\sum_{i=1}^{K}\frac{\hat{x}_t^i(H)}{\sigma_t^i(H)^2}$$

$$+b^2(H)\sum_{i=1}^{K}\sum_{j=1}^{i-1}\frac{(\hat{x}_t^i(H)+\hat{x}_t^j(H))}{\sigma_t^i(H)^2\sigma_t^j(H)^2}f_3(t,H) \tag{4-19}$$

$$f_2(0,H)=\hat{x}_0(H) \tag{4-20}$$

$$f_1(t,H)=c_1\exp-(k-1)A(H)t\cdot\exp\left[-\frac{b^2(H)}{2}\int_0^t\frac{1}{f_3(s,H)}\,ds\right]$$

$$\cdot\exp\left[-\frac{b^2(H)}{2}\int_0^t\sum_{i=1}^{K}\sum_{j=1}^{i-1}(f_2(s,H)\right.$$

$$\left.-\hat{x}_s^i(H))(f_2(s,H)-\hat{x}_s^j(H))\,ds\right] \tag{4-21}$$

$$f_1(0,H)=\frac{1}{\sigma_0(H)^{K-1}}(P_0(H))^{K-1}=C_1. \tag{4-22}$$

The conditional means $x_t(H)$ are given by

$$x_t(H)=\sigma_t(H)^2\left\{\sum_{i=1}^{K}\frac{\hat{x}_t^i(H)}{\sigma_t^i(H)^2}-\frac{f_2(t,H)}{f_3(t,H)}\right\} \tag{4-23}$$

and

$$\sigma_t(H)^2=\left\{\sum_{i=1}^{K}\frac{1}{\sigma_t^i(H)^2}-\frac{1}{f_3(t,H)}\right\}^{-1}. \tag{4-24}$$

Proof: See the Appendix.

Equations (4-15)–(4-24) accept as input the communications $P_t^i(H)$, $\hat{x}_t^i(H)$, using the *a priori*-determined parameters $\sigma_t^i(H)^2$, to obtain the compensator $C_t(H, x)$ defined in Theorem 2. A direct derivation of Theorem 3 is possible, without using the results of Theorem 2. Essentially, one defines

$$C_t(H,x)=\frac{\prod_{i=1}^{K}\frac{1}{\sqrt{2\pi}\sigma_t^i(H)}e^{-1/2(x_t-\hat{x}_t^i(H))^2/\sigma_t^i(H)^2}\cdot p_t^i(H)}{\frac{1}{\sqrt{2\pi}\sigma_t(H)}e^{-1/2(x_t-\hat{x}(H))^2/\sigma_t(H)^2}\cdot p_t(H)}. \tag{4-25}$$

A direct computation establishes

$$C_t(H,x)=cf_1(t,H)e^{-1/2(x_t-f_2(t,H))^2/f_3(t,H)} \tag{4-26}$$

where

$$\frac{1}{f_3(t,H)}=\sum_{i=1}^{n}\frac{1}{\sigma_t^i(H)^2}-\frac{1}{\sigma_t(H)^2} \tag{4-27}$$

$$f_2(t,H)=f_3(t,H)\cdot\left(\sum_{i=1}^{K}\frac{\hat{x}_t^i(H)}{\sigma_t^i(H)^2}-\frac{\hat{x}_t(H)}{\sigma_t(H)^2}\right) \tag{4-28}$$

and $f_1(t, H)$ defined appropriately. Then, differentiation of (4-27) and (4-28) yields (4-17) and (4-19). However, we have used the general result of Theorem 2 to obtain the specialized results of Theorem 3. Notice also that Theorem 3 reduces to the distributed estimation results in [1] and [3] when only one hypothesis is considered.

V. Communication Requirements

The algorithms described in Sections II–IV provide a coordination algorithm for obtaining the centralized conditional distribution of the state, given the past histories of the local conditional distributions of the state. In general, these algorithms require that sufficient statistics for the local conditional distributions of the state be transmitted at every possible time instant. These statistics may be infinite dimensional, thus requiring more communication than direct transmission of the finite-dimensional unprocessed observations.

There are classes of systems for which the communication requirements can be significantly reduced. For example, the results in [1] and [3] imply that, for linear Gaussian systems, the computation of the dynamic compensator required in the coordinator's algorithm can be divided into local dynamic compensators. In essence, each local station can compute a sufficient statistic for its contribution to the coordinator's dynamic compensator. Hence, the reconstruction of the centralized conditional distribution can be accomplished using only the communications at the most recent time, thereby reducing the required communications to the times when a centralized conditional distribution is required.

Unfortunately, the coordinator's algorithms in Sections II–IV cannot be separated into a memoryless combination of locally computed processes. This is illustrated clearly in Theorem 3 where the computation of $f_1(t, H)$ requires the computation of

$$\int_0^t \sum_{i=1}^{K} \sum_{j=1}^{i-1} (f_2(s, H) - \hat{x}_s^i(H))(f_2(s, H) - x_s^j(H))\, ds.$$

This term involves cross couplings between the information available to each local station at each point in time; hence, its computation cannot be divided into the product of integrals computed at the local stations. Note that the linearity of (4-19) would permit the computation of $f_2(t, H)$ as a linear combination of locally computed $f_2^i(t, H)$ as

$$f_2(t, H) = \sum_{i=1}^{K} f_2^i(t, H) \tag{5-1}$$

where

$$\frac{d}{dt} f_2^i(t, H) = \left(A(H) - 2b^2(H) f_3(t, H) \sum_{l=1}^{K} \sum_{j=1}^{l-1} \frac{1}{\sigma_t^l(H)\sigma_t^j(H)^2} + b^2(H) \sum_{l=1}^{K} \frac{1}{\sigma_t^l(H)^2} \right) f_2^i(t, H) - \frac{b^2(H)}{\sigma_t^i(H)^2} \hat{x}_t^i(H) + b^2(h) \frac{\hat{x}_t^i(H)}{\sigma_t^i(H)^2} \cdot f_3(t, H) \left(\sum_{j=1}^{K} \frac{1}{\sigma_t^j(H)^2} - \frac{1}{\sigma_t^i(H)^2} \right) \tag{5-2}$$

$$f_2^i(0, H) = \frac{1}{K} \bar{x}_0(H) \tag{5-3}$$

and that $f_3(t, H)$ would be computable at each local station, assuming that the global statistics of the processes were known at the stations. Hence, the only term which requires the continuous communication of information is $f_1(t, H)$.

Theorem 3 provides the basis for developing a coordination algorithm when communications are limited to periodic intervals with period Δt. Using (5-1)–(5-3), the values of $f_2(t, H)$ and $f_3(t, H)$ can be computed exactly at each instant of communication. Then, the computation of $f_1(t, H)$ can be approximated, for $t = M\Delta t$, as

$$f_1(t, H) \approx c_1 e^{-(k-1)A(H)t} \exp\left[-\frac{b^2(H)}{2} \int_0^t \frac{1}{f_3(s, H)} ds \right] \cdot \prod_{l=1}^{M} e^{-b^2(H)} \sum_{i=1}^{K} \sum_{j=1}^{i-1} (f_2(l\Delta t, H) - \hat{x}_{l\Delta t}^i(H))(f^2(l\Delta t, H) - \hat{x}_{l\Delta t}^j(H))\Delta t. \tag{5-4}$$

Although the algorithm does not reconstruct the centralized conditional distribution exactly, the error introduced by (6-4) will be of order $0(\Delta t^{1/2})$, the modulus of continuity of x_t^i.

One particular case where the centralized conditional probability can be computed exactly using only local communications at a single time instant consists of the class of systems in Section IV when the driving noise intensity $b(H)$ is zero for each event H. In this case,

$$f_1(t, H) = \frac{P_0(H)^{K-1}}{\sigma_0(H)^{K-1}} e^{-(K-1)A(H)t}. \tag{5-5}$$

Equation (5-5) does not require any dynamic integration of data; hence, the centralized conditional distribution can be computed using only a single time communication, consisting of the sufficient statistics S_t^i, because the equation for $f_2(t, H)$ is now

$$f_2(t, H) = \bar{x}_0 e^{A(H)t}. \tag{5-6}$$

Similar results can be obtained for the general jump-diffusion model of Section III when the driving noise intensity $\sigma(t, x, \rho) = 0$ and the jump rates $\lambda_{ij}(x) = 0$ for all $i \neq j$. In this case, the centralized conditional probability of z_t can be obtained using only the communications from the local stations at time t.

VI. Conclusions

The algorithms presented in Sections II and III solve the problem of data fusion for nonlinear systems under the conditions that, at each measurement time, the local stations compute their local conditional distribution and communicate a sufficient statistic to the coordination station. Under the conditions of Sections II and III, the coordination station can reconstruct exactly the centralized conditional distribution of the state process. The results in Section IV illustrate how, for a particular class of nonlinear Markov processes, finite-dimensional sufficient statistics can be found for the local conditional distributions, and how the results of Section III yield the centralized conditional distribution from these local sufficient statistics. These results represent a direct generalization of the results in [1]–[3] to non-Gaussian–Markov processes.

In most practical applications, the rate of communications between local stations and the coordinator is substantially lower than the measurement rate of the local stations. In Section V, we investigated the possibility of designing coordination algorithms which require less frequent communications. Under communication restrictions, the exact centralized conditional distribution can be obtained only under severe restrictions in the evolution of the process under observation. Nevertheless, accurate approximations to this conditional distribution can be obtained using periodic communications. These approximations can serve as the basis for implementable algorithms, provided that approximate finite-dimensional sufficient statistics can be determined.

Appendix
Proof of Theorem 3

From Theorem 2, we have

$$q_t(x, H) = \frac{\prod_{i=1}^{K} q_t^i(x, H)}{C_t(x, H)} \tag{A-1}$$

where

$$q_t^i(x, H) = \frac{P_t^i(H)}{\sqrt{2\pi}} \frac{1}{\sigma_t^i(H)} \exp\left[-\frac{(x - \hat{x}_t^i(H))^2}{2\sigma_t^i(H)^2} \right] \tag{A-2}$$

and $C_t(x, H)$ satisfies (3-7). For the system of (4-2), (3-4) becomes

$$A_t^* f(x, H) = \frac{1}{2} b(H)^2 \frac{d^2}{dx^2} f(x, H) - A(H)x \frac{d}{dx} f(x, H) - A(H) f(x, H). \quad \text{(A-3)}$$

Hence, (3-7) becomes

$$\frac{d}{dt} C_t(x, H) = -(K-1)A(H)C_t(x, H) - A(H)x \frac{d}{dx} C_t(x, H) + b^2(H) \frac{d}{dx} C_t(x, H) \cdot \sum_{i=1}^{K} \frac{\frac{d}{dx} q_t^i(x, H)}{q_t^i(x, H)} + \frac{b^2(H)}{2} \frac{d^2}{dx^2} C_t(x, H) - \frac{b^2(H)}{C_t(x, H)} \left(\frac{d}{dx} C_t(x, H) \right)^2 - b^2(H) \cdot C_t(x, H) \sum_{i=1}^{K} \sum_{j=1}^{i-1} \frac{\frac{d}{dx} q_t^i(x, h) \frac{d}{dx} q_t^j(x, H)}{q_t^i(x, H) q_t^j(x, H)}. \quad \text{(A-4)}$$

Substituting (A-2) into (A-4) and differentiating gives

$$\frac{d}{dt} C_t(x, H) = -(K-1)A(H)C_t(x, H) - A(H)x \frac{d}{dx} C_t(x, H) + b^2(H) \frac{d}{dx} C_t(x, H) \cdot \sum_{i=1}^{K} \frac{-(x - \hat{x}_t^i(H))}{\sigma_t^i(H)^2} + \frac{b^2(H)}{2} \frac{d^2}{dx^2} C_t(x, H) - \frac{b^2(H)}{C_t(x, H)} \left(\frac{d}{dx} C_t(x, H) \right)^2 - b^2(H) C_t(x, H) \sum_{i=1}^{K} \sum_{j=1}^{i-1} \frac{(x - \hat{x}_t^i(H))(x - \hat{x}_t^j(H)}{\sigma_t^i(H)^2 \sigma_t^j(H)^2}. \quad \text{(A-5)}$$

Equation (A-5) admits a solution of the form

$$C_t(x, H) = f_1(t, H) e^{-(x - f_2(t,H))^2 / 2f_3(t,H)}. \quad \text{(A-6)}$$

Substituting (A-6) into (A-5) and grouping like powers of $(x - f_2(t, H))$ yields (4-17), (4-19), and (4-21). Matching the initial condition for $C_0(x, H)$ yields (4-18), (4-20), and (4-22).

Now, consider the equation for $\sigma_t^i(H)^2$:

$$\frac{d}{dt} \frac{1}{\sigma_t^i(H)^2} = \frac{-2A(H)}{\sigma_t^i(H)^2} - \frac{b^2(H)}{\sigma_t^i(H)^4} + C_t^i(H)^2. \quad \text{(A-7)}$$

Similarly, straightforward filtering arguments yield

$$\frac{d}{dt} \frac{1}{\sigma_t(H)^?} = \frac{-2A(H)}{\sigma_t(H)^2} - \frac{b^2(H)}{\sigma_t(H)^4} + \sum_{i=1}^{K} C_t^i(H)^2 \quad \text{(A-7)}$$

with initial conditions

$$\sigma(H)^2 = \sigma_0^i(H)^2 = \sigma_0(H)^2. \quad \text{(A-9)}$$

Combining (4-17) and (4-18) with (A-6)–(A-8) yields (4-24). Similarly, consider the equation for $\sigma_t^i(H)^{-2} \hat{x}_t^i(H)$:

$$d \frac{\hat{x}_t^i(H)}{\sigma_t^i(H)^2} = \left(-A(H) - \frac{b^2(H)}{\sigma_t^i(H)^2} \right) \frac{\hat{x}_t^i(H) dt}{\sigma_t^i(H)^2} + C_t^i(H) dy_t^i \quad \text{(A-10)}$$

and

$$d \frac{\hat{x}_t(H)}{\sigma_t(H)^2} = \left(-A(H) - \frac{b^2(H)}{\sigma_t(H)^2} \right) \frac{\hat{x}_t(H) dt}{\sigma_t(H)^2} + \sum_{i=1}^{K} + C_t^i(H) dy_t^i \quad \text{(A-11)}$$

with initial conditions

$$\frac{\hat{x}_0^i(H)}{\sigma_0(H)^2} = \frac{x_0^i(H)}{\sigma_0^i(H)^2} = \frac{\hat{x}_0(H)}{\sigma_0(H)^2}. \quad \text{(A-12)}$$

Combining (4-19), (4-20) and (A-10)–(A-12) yields (4-23).

The final result which must be established in Theorem 3 is (4-15) and (4-16). From standard filtering results (e.g., [8]),

$$p_t(x, H) = \frac{1}{\sqrt{2\pi}\sigma_t(H)} P_t(H) \exp \left[\frac{1}{2} - (x - \hat{x}_t(H))^2 / \sigma_t(H)^2 \right]. \quad \text{(A-13)}$$

Combining (A-13), (A-11), (A-12), and (A-6) gives

$$P_t(H) = \frac{\sqrt{2\pi}}{D} \sigma_t(H) \cdot \exp \left[\frac{1}{2} (x - \hat{x}_t(H))^2 / \sigma_t(H)^2 \right] \cdot \prod_{i=1}^{K} \frac{P_t^i(H)}{\sqrt{2\pi}\sigma_t^i(H)} \cdot \frac{\exp \left[\sum_{i=1}^{K} -\frac{1}{2} (x - \hat{x}_t^i(H))^2 / \sigma_t^i(H)^2 \right]}{f_1(t, H)} \cdot e^{(x - f_2(t,H))^2 / 2f_3(t,H)}. \quad \text{(A-14)}$$

Hence, cancelling the x-dependent term results in

$$P_t(H) = \frac{1}{(\sqrt{2\pi})} \frac{K-1}{D} \cdot \frac{\sigma_t(H)}{\prod_{i=1}^{K} \sigma_t^i(H)} \cdot \frac{\exp \left[-\frac{1}{2} \sum_{i=1}^{K} \frac{\hat{x}_t^i(H)^2}{\sigma^i(H)^2} - \frac{\hat{x}_t(H)^2}{\sigma_t(H)^2} \right]}{f_1(t, H) \exp \left[-\frac{f_2(t, H)^2}{2f_3(t, H)} \right]} \quad \text{(A-15)}$$

where D is chosen such that $\Sigma_{i=1}^N P_t(H_i) = 1$. Equations (4-15) and (4-16) are equivalent to (A.15).

Acknowledgment

The authors wish to thank Prof. A. Willsky and Prof. S. Mitter for their help in discussing these problems and the reviewers for their comments.

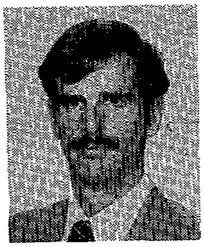
References

[1] J. L. Speyer, "Computation and transmission requirements for a decentralized linear-quadratic-Gaussian control problem," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 266–269, Apr. 1979.

[2] C. Y. Chong, "Hierarchical estimation," presented at the 2nd MIT/ONR Workshop Distributed Inform. Decision Syst. Motivated Naval Command-Contr.-Commun. (C^3) Problems, Naval Postgrad. School, Monterey, CA, July 1979.

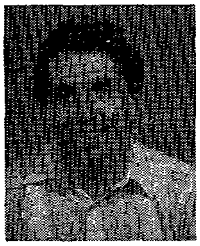
[3] A. S. Willsky, M. Bello, D. A. Castanon, B. C. Levy, and G.

Verghese, "Combining and updating of local estimates and regional maps along sets of one-dimensional tracks," *IEEE Trans. Automat. Contr.*, vol. AC-27, Aug. 1982.
[4] E. Pardoux, "Stochastic partial differential equations and filtering of diffusion processes," *Stochastics*, vol. 2, pp. 127-168, 1979.
[5] R. S. Lipster and A. N. Shiryaev, *Statistics of Random Processes I.* New York: Springer-Verlag, 1977.
[6] D. A. Castanon and D. Teneketzis, "Nonlinear data fusion," in *Proc. 1982 MIT/ONR Workshop Command Contr.*, Monterey, CA, Aug. 1982.
[7] C. A. Striebel, *Optimal Control of Discrete Time Stochastic Systems.* New York: Springer-Verlag, 1975.
[8] M. Loeve, *Probability Theory.* New York: Van Nostrand, 1963.
[9] D. W. Strook and S. R. S. Varadhan, *Multidimensional Diffusion Processes.* New York: Springer-Verlag, 1979.
[10] M. Zakai, "On the optimal filtering of diffusion processes," *Z. Wahr. Ver. Geb.*, vol. 11, 1969.
[11] J. M. C. Clark, "The design of robust approximations to the stochastic differential equations of nonlinear filtering," in *Commun. Syst. and Random Process Theory*, J. K. Skwirzynski, Ed., NATO Advanced Study Inst. Ser., Sijthoff and Noorshoff, Alphen aan den Rijn, 1978.
[12] O. Hijab, "The adaptive LOG problem I," *IEEE Trans. Automat. Contr.*, vol. AC-28, Feb. 1983.



David A. Castanon was born in La Habana, Cuba, on November 24, 1950. He received the B.S. degree in electrical engineering (with high honors) from Tulane University, New Orleans, LA, in 1971, and the Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology, Cambridge, in 1976.

From 1976 to 1981 he was a Research Associate with the Laboratory for Information and Decision Systems at M.I.T. He is presently with ALPHATECH, Inc., Burlington, MA. His research interests include stochastic control and estimation, team theory and dynamic game theory, and qualitative theories of stochastic large-scale systems.



Demosthenis Teneketzis was born in Drama, Greece, in 1951. He received the B.S. degree in electrical engineering from the University of Patras, Greece, in 1974, and did his graduate studies at M.I.T., Cambridge, where he received the S.M. and Ph.D. degrees in electrical engineering in 1976 and 1979, respectively.

From 1979 to 1980 he worked for Systems Control, Inc., Palo Alto, CA, and from 1980 to 1984 he was with ALPHATECH, Inc., Burlington, MA. Currently he is an Assistant Professor of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor. His interests are in stochastic control, estimation, game theory, team theory, and information theory.