PREPRINTS

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

9TH WORLD CONGRESS

A BRIDGE BETWEEN CONTROL SCIENCE AND TECHNOLOGY

VOLUME VII
COLLOQUIUM 14.4

BUDAPEST, HUNGARY
JULY 2-6 1984

SIMULTANEOUS DETECTION AND ESTIMATION FOR DIFFUSION PROCESS SIGNALS

John S. Baras

Electrical Engineering Department
University of Maryland
College Park, MD 20742

ABSTRACT

We consider the problem of simultaneous detection and estimation when the signals corresponding to the M different hypotheses can be modelled as outputs of M distinct stochastic dynamical systems of the Ito type. Under very mild assumptions on the models and on the cost structure we show that there exist a set of sufficient statistics for the simultaneous detection-estimation problem that can be computed recursively by linear equations. Furthermore we show that the structure of the detector and estimator is completely determined by the cost structure. The methodology used employes recent advances in nonlinear filtering and stochastic control of partially observed stochastic systems of the Ito type. Specific examples and applications in radar tracking and discrimination problems are discussed.

KEYWORDS

diffusion processes, simultaneous detection and estimation, Zakai equations, radar target discrimination

INTRODUCTION

In a typical present day radar environment, the radar receiver is subjected to radiation from various sources. A very important function of the radar receiver is its ability to discriminate between the various waveforms received and select the desired one for further processing. Furthermore an equally important function of the receiver is to estimate important parameters of the radiating source from the received waveforms. Thus the receiver is required often to perform a "combined detection and estimation" function.

An abstract formulation of the combined detection and estimation problem in the language of statistical decision theory has been developed by Middleton and Esposito in [12]. They correctly point out that optimal processing in such problems often requires the mutual coupling of the detection and estimation algorithms. Although from the mathematical point of view estimation may be considered as a generalized detection problem, from an operational point of view the two procedures are different: e.g. one usually selects different cost functions for each and obtains different data processors as a result. It is then correctly argued in [12] that it is practically appropriate to retain the usual distinction between detection and estimation. Pictorially the simultaneous detection-estimation problem is described by figure 1 from [12].

There are various ways that the detector and estimator can be coupled leading to a hierarchy of complex processors. We describe here some important cases.

Detection-directed estimation

Here the detection operation is optimized with a priori knowledge of the existence of an estimator following it. The estimator is dependent on the detector's decision by being gated on only if

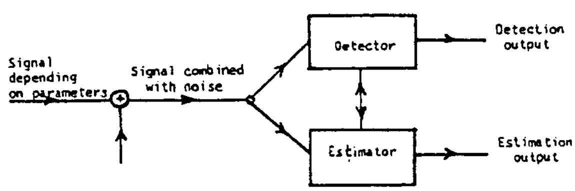

Fig. 1. General block diagram for simultaneous detection-estimation, from [12].

the detector decides that the desired signal is present. The problems of interest to us need a significant extension beyond the results in [12] because here we seek to discriminate between various signals and not simply between "signal plus noise" and "noise alone". The block diagram for this problem is shown in Figure 2 below. Here the coupling is via cost terms that assess the performance

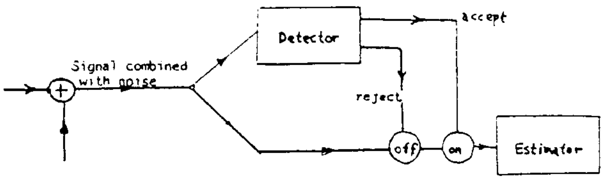

Fig. 2. Detection-directed estimation.

deterioration when the estimator is turned off while the signal is present $C_{e,1}$, or the estimator is turned on when the signal is not present $C_{e,0}$.

Therefore the average risks corresponding to the operations of detection and estimation can be minimized separately. This leads to a detection test that is a modified generalized likelihood test. If the cost terms $C_{e,1}$, $C_{e,0}$ are constant the coupling just reduces to a modification of the threshold [12]. Since the detector's decision rule does not depend on the estimate, the structure of the optimal estimator is not a function of the data region specified by the decision rule of the detector's operation, when the detector's decision is to accept the signal. In practical terms this means that we can choose to estimate only when the detector has decided that the desired signal is present.

Coupled detection-estimation with decision rejection

Here detection and estimation run in parallel and are followed by rejection of the estimate if the detector's decision is not to accept the signal. Here the detector's cost depends on the value of the estimate. The block diagram is shown in figure 3 below.

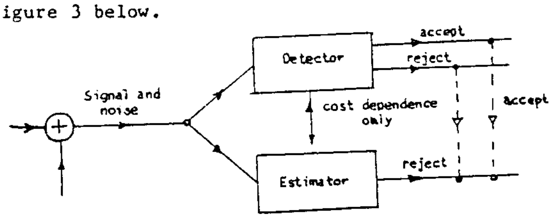

Fig. 3. Coupled detection-estimation with decision rejection.

Typically, one solves the detection problem knowing the estimator. Then a second optimization is performed over all estimators. This case usually results in relatively simple estimators and complex highly nonlinear detectors [12].

Motivation for the problems studied stems from distributed target discrimination [2]-[3]. We concentrate in this paper on a two hypothesis detection formulation, but it is clear that the methods can be easily extended to M-ary detection problems. The two hypotheses are H_0 = the received signal is a process y_{0t} plus noise, H_1 = the received signal is a process y_{1t} (different from y_{0t}) plus noise. Both processes are modeled as outputs of stochastic dynamical systems of the diffusion type. The noise is the same in both cases. Due to this fact we can assume that noise is eliminated from the mathematical formulation of the problem of detection, while as we shall see its presence may be crucial for the estimation problem.

We did not study detectors with "learning" and we suggest this as a promising extension of the results reported here. We note however, that our formalism includes general "learning" algorithms. Most of the work on detectors with "learning" is problem specific and does not utilize dynamical system models for the signals as we do.

The major criticism for the work of Middleton and Esposito [12], is that although they used a Bayesia approach to the estimation problem, they considered nonrecursive solutions and detection was coupled to estimation through cost structure which explicitly considers coupling of the detection and estimation costs. Clearly nonrecursive solutions are not appropriate for advanced sensors employed in guided platforms. Furthermore it would be unrealistic to assume that the designer has such explicit knowledge of the functional couplings between detection and estimation costs.

Several other authors have analyzed the problem. Scharf and Lytle [13] studied detection problems involving Gaussian noise of unknown level, thus including noise parameters in the problem. As in [12], their solution is also nonrecursive, and focuses on the existence of uniformly most powerful tests. Jaffer and Gupta [10], [11] consider the recursive Bayesian problem using a quadratic cost, Gauss-Markov processes and estimating only signal parameters.

Birdsall and Gobien [6] considered the problem of simultaneous detection and estimation from a Bayesian viewpoint. This work is close in spirit with our approach, although the class of problems we can analyze by our methods is significantly wider. We also follow a Bayesian methodology during the initial phase of analysis. It becomes clear that using Bayesian methods one can analyze the problems under consideration in an inherently intuitive, simple conceptually manner which can be easily obscured in highly structured methodologies utilizing specific detector structures and cost relationships. As a result one can analyze the special problems described in figures 2 and 3 as specializations of a wider picture and framework. The results reported in [6] are limited by two important assumptions: (a) the observed data have densities that display finite dimensional sufficient statistics under both hypotheses for the unknown parameters and (b) the unknown parameters form a finite-dimensional vector. Both nonsequential and sequential problems are analyzed in [6]. The most important result of [6] is the proof that through a Bayesian approach both estimation and detection occur simultaneously, with the detector using the a posteriori densities generated by two separate estimators, one for each hypothesis. A particularly attractive feature is that no assumptions are made on the estimation criterion and very flexible assumptions are made on the detecction criterion. When finite-dimensional sufficient statistics exist the optimum processor partitions naturally into three parts: a "primary" processor which is totally independent of a priori distributions on the parameters, a "secondary" processor which modifies the output according to the priors and solves the detection problem, and an estimator which uses the output of the other two in estimating the unknown parameters. Only the estimator structure depends on cost functionals.

Since dynamical system models are not utilized to represent signals in [6], there is great difficulty in analyzing the far more interesting sequential problem. It is for this reason that one is forced to make the limiting assumptions mentioned above. In our approach we consider diffusion type models for the signals and we utilize modern methods from nonlinear filtering and modern methods from nonlinear filtering and stochastic control to analyze the problem [7]-[8]. Corresponding results for Markov chain models can be easily obtained, but we only give brief comments for such problems here.

NOMENCLATURE AND FORMULATION OF THE SEQUENTIAL PROBLEM

In this section we present a general formulation for the continuous time, sequential, simultaneous detection and estimation problem when the signals can be represented as outputs of diffusion type processes [14]. To simplify notation, terminology and subsequent computations we consider only the scalar observation case here. All results extend to vector observations in a straight-forward manner. The observed data y(t) constitute therefore a real-valued scalar stochastic process.

The statistics of y(·) are not completely known. More specifically they depend on some parameters and some hypotheses. For simplicity we shall consider here only the binary hypotheses detection problem. Extensions to M-ary detection are trivial. We shall denote by H_0, H_1 the two mutually exclusive and exhaustive hypotheses.

Under hypothese H_0, the received data y(t) can be represented as

$$\begin{aligned} dy(t) &= h^0(x^0(t),\theta^0)dt + dv(t) \\ dx^0(t) &= f^0(x^0(t),\theta^0)dt + g^0(x^0(t),\theta^0)dw^0(t) \end{aligned} \tag{1}$$

where θ^0 is a vector-valued unknown parameter that may be assumed fixed or random throughout the problem. Here v(·), w(·) are independent, 1-dimensional and n_0-dimensional respectively standard Wiener processes [18]. In other words when hypothesis H_0 is true the received data can be thought of as the output of a stochastic dynamical system, corrupted by white Gaussian noise. h^0, f^0, g^0, θ^0 parameterize the nonlinear stochastic system.

Similarly when hypothesis H_1 is true, the received data y(t) can be modeled as

$$\begin{aligned} dy(t) &= h^1(x^1(t),\theta^1)dt + dv(t) \\ dx^1(t) &= f^1(t),\theta^1)dt + g^1(x^1(t),\theta^1)dw^1(t) \end{aligned} \tag{2}$$

where now x^1 is n_1-dimensional. The vector parameters θ^0,θ^1 may have common components. For instance, in the classical "noise or signal-plus-noise" problem any noise parameters clearly appear in both hypotheses and would thus be common to θ^0,θ^1.

We note that we have the same "observation noise" v(·) under both hypotheses. This is clearly the case in radar applications. On the other hand when one is faced with state and parameter dependent observation noises, a simple transformation translates the two models in the form (1) (2). We shall assume that h^i,f^i,g^i, i=0,1, have sufficient properties to guarantee existence and uniqueness of probability distribution functions for y(·) under either hypothesis. As a minimal hypothesis we assume that the martingale problems for (1) and (2) are well posed [14] for all values of θ^0,θ^1 in appropriate compact sets Θ^0,Θ^1 respectively. Furthermore neither (1) nor (2) exhibit esplosions [14] for any value of the parameters. Often we shall make stronger assumptions such as existence of strong solutions to (1) (2), or smoothness of f^i,g^i,h^i, i=0,1, or existence of classical probability densities for y under either hypothesis.

We shall denote by $p^i_y(\cdot,t|\theta^i)$, i=0,1, the probability density of y(t) under hypothesis H^i and when the parameter obtains the value θ^i, i=0,1. We shall denote the probability measures corresponding to y under H^0 or H^1 by μ^0 and μ^1 respectively. As is well known these are measures on the space of continuous functions [14].

Finally we note that although we have assumed time invariant stochastic models in (1), (2) the results extend easily to the time varying case.

Following a Bayesian approach we assume a priori densities for the two parameters θ^0,θ^1 which will be denoted by $p^i(\cdot,0)$, i=0,1 respectively. Similarly initial densities for $x^0(0)$ and $x^1(0)$ are assumed known and independent of θ^0,θ^1 respectively They will be denoted by $p^i_x(\cdot,0)$. The choice of these a priori densities, is frequently a very interesting problem in applications, as they represent the designer's a priori knowledge about the models used.

With these preliminaries we can now formulate the problem. Let y^t denote as usual the portion of the observed sample path "up to time t", i.e. $y^t = \{y(s),s\leq t\}$. Given the observed data y^t, we wish to design a processor which at time t will optimally select simultaneously which of the two hypotheses H_0 or H_1 is true, and optimal estimates for the parameters θ^0 and θ^1. Moreover the processor should operate recursively so as to permit real-time implementation.

To complete the problem formulation we need to specify costs for detection and estimation. Let $c_i(\hat{\theta}^i(t),\theta^i)$, i=0,1 be the penalty for "estimating" θ^i, by $\hat{\theta}^i(t)$ at time t. If c_i is quadratic we have the well known minimum variance estimates. Similarly let $\gamma(t)$ denote the decision, at time t, of whether we declare hypothesis H_0 or H_1 to hold. Then $k(\gamma(t), i)$, i=0,1 will denote the penalty when the true hypothesis is H_i and we decide $\gamma(t)$, at time t. Obviously there are infinitely many variations on the possible choice for a cost function. We shall consider only two possibilities in this report. Finite time average integral cost

$$\begin{aligned} J_f = E\{\int_0^T \lambda_e[c_0(\hat{\theta}^0(t),\theta^0)\chi\{t,\gamma(t)=0\} + \\ c_1(\hat{\theta}^1(t),\theta^1)\chi\{t,\gamma(t)=1\}dt + \lambda_d k(\gamma(t),i)dt\} \end{aligned} \tag{3}$$

and infinite time average discounted cost.

$$J_d = E\{\int_0 C(\gamma,\hat\theta^0,\hat\theta^1,x)e^{-\alpha t}dt\} \quad (4)$$

where $C(\gamma,\hat\theta^0,\hat\theta^1,x)$ is the integrand in (3) and α the discount rate. λ_e,λ_d are weights. The reasons for the characteristic functions appearing in (3), (4) are rather obvious. The estimator will contribute cost only when utilized, and it will be utilized for θ^0 only when $\gamma(t)=0$. We would like to point out that this does not preclude both estimators from running continuously. This scheme is used only to assess costs properly.

The appropriate formulation of the problem is as a partially observable stochastic control problem. The admissible controls are

$$\begin{aligned} &\gamma : R \to \{0,1\} \\ &\hat\theta^0: R \to \Theta^0 \\ &\hat\theta^1: R \to \Theta^1 \end{aligned} \quad (5)$$

where all functions are nonanticipative with respect to y; i.e. measurable w.r. to F_t^y:

$$\gamma(\cdot),\hat\theta^0(\cdot),\hat\theta^1(\cdot)\varepsilon F_t^y \quad (6)$$

The cost is either (3) or (4). For the system dynamics we proceed as follows. The state equations are mixed consisting of the continuous components

$$\begin{aligned} dx^0(t) &= f^0(x^0(t),\theta^0(t))dt + g^0(x^0(t),\theta^0(t))dw^0(t) \\ dx^1(t) &= f^1(x^1(t),\theta^1(t))dt + g^1(x^1(t),\theta^1(t))dw^1(t) \\ d\theta^0(t) &= 0 \\ d\theta^1(t) &= 0 \end{aligned} \quad (7)$$

and the discrete component z(t) which can take only the values 0 or 1 and is constant. The initial densities for $x^0,x^1,\theta^0,\theta^1$ have already been described. The initial probability vector for z(t) (which tracks which hypothesis is true) is

$$\begin{aligned} &Pr\{z(0) = 0\} = P_0 \\ &Pr\{z(0) = 1\} = P_1 \end{aligned} \quad (8)$$

The observations are

$$dy(t)=(1-z(t)h^0(x^0(t),\theta^0)dt+z(t)h^1(x^1(t),\theta^1)dt+dv(t) \quad (9)$$

Since (7) are degenerate, there are some technical minor difficulties, which can be circumvented however using recent techniques. This completes the formulation of the problem.

STRUCTURE OF THE OPTIMAL PROCESSOR

Following recent results [7]-[8] in stochastic optimal control theory we have obtained first the following results that reduce the partially observed stochastic control problem described in section 2 to an equivalent, infinite dimensional fully observed problem.

<u>Theorem 1</u>: There exist optimal $\gamma,\hat\theta^0,\hat\theta^1$ for the stochastic optimal control problem (3)-(9).

<u>Proof</u>: This follows from the results of Fleming and Pardoux [8] and Bismut [7]. The only difference is that due to the structure of the dynamics here (i.e. they do not depend on the controls $\gamma,\hat\theta^0,\hat\theta^1$) we can show that optimal controls exist in the class of strict sense controls as specified in section 2 (i.e. $\gamma(t),\hat\theta^0(t),\hat\theta^1(t)$ are measurable with respect to F_t^y).

We then introduce as in Fleming and Pardoux [8] the associated "separated" stochastic control problem. In the separated stochastic control problem the state at time t is a measure Λ_t on R^N (where $N = n_0+n_1+2$), which is an unnormalized conditional distribution of the state $x(t)\triangleq[x_0(t),x_1(t),\theta_0(t),\theta_1(t),z(t)]^T$ of the problem formulated in section 2. The dynamics of the measure-valued process Λ_t obey the Zakai equation of nonlinear filtering [7].

In the sequel we assume that all functions appearing in (1)-(9) are bounded and continuous and that g^0,f^0,g^1,f^1 are Lipschitz in $x^0,\theta^0,x^1,\theta^1$, respectively. Due to the discrete component z(t) of the state x(t) we have to consider a two dimensional measure valued process Λ^0,Λ^1, where Λ^i is the unnormalized conditional distribution of the state $x(t)\triangleq[x_0(t),x_1(t),\theta_0(t),\theta_1(t)]$ (slight abuse of notation here) when hypothesis H_i is true, i=0,1. We further assume that for i=0,1 the corresponding Zakai equation has a unique solution which is absolutely continuous with respect to Lebesque measure; i.e. we assume the existence of conditional unnormalized probability densities for $x(t)\varepsilon R^N$ given y^t. For results on this see [4].

Let $u^i(x,t)$ denote the conditional probability density of x(t) given y^t when hypothesis H_i holds. Then $u^i(\cdot,\cdot)$ satisfies the Zakai equation

$$du^i = L^* u^i dt + dy(t)h^i u^i, \; i=0,1 \quad (10)$$

where L^* is the formal adjoint to the infinitesimal generator of (7)

$$= \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(x)\frac{\partial^2}{x_i x_j} + \sum_{i=1}^{N} b_i(x)\frac{\partial}{\partial x_i} \quad (11)$$

and

$$a = \sigma\sigma^T$$

$$\sigma = \begin{bmatrix} g^0 & & & \\ & g^1 & 0 & \\ 0 & & 0 & \\ & & & 0 \end{bmatrix} \quad (12)$$

$$b = \begin{bmatrix} f^0 & & & \\ & f^1 & 0 & \\ 0 & & 0 & \\ & & & 0 \end{bmatrix} \quad (13)$$

It is clear, due to the diagonal form of a, b and the degeneracy in the driving term of (10), that equivalently we can consider the two stochastic partial differential equations

$$du^i = L_i^* u^i dt + dy(t) h^i u^i, \quad i=0,1 \tag{14}$$

where L^* is the formal adjoint to the infinitesimal generator of the i^{th} component of (7); i.e. it has the form (11) with a,b replaced by a^i, b^i, $i=0,1$. Here

$$a^i = \sigma^i (\sigma^i)^T$$

$$\sigma^i = \begin{bmatrix} g^i & 0 \\ 0 & 0 \end{bmatrix} \tag{15}$$

$$b^i = \begin{bmatrix} f^i & 0 \\ 0 & 0 \end{bmatrix}$$

To complete the description of the "separated" stochastic control problem, let $C(\gamma, \hat{\theta}^0, \hat{\theta}^1, x)$ denote the integrand in the cost definition (3). then if we let

$$u(x,t) = \begin{bmatrix} u^0 \ (x_0, \theta_0, t) \\ u^1 \ (x_1, \theta_1, t) \end{bmatrix} \tag{16}$$

we can rewrite the cost (3) as

$$J_f(\pi) = E_y \{ \int_0^T \int C(\gamma, \hat{\theta}^0, \hat{\theta}^1, x) [(x,t)^T \begin{bmatrix} P_0 \\ P_1 \end{bmatrix}] dx dt \} \tag{17}$$

where π is the policy corresponding to a particular selection of $\gamma(\cdot)$, $\hat{\theta}^0(\cdot)$, $\hat{\theta}^1(\cdot)$, and E_y is expectation with respect to y. Note that u depends explicitly on y.

The separated problem is to choose a policy π which is a function of u^0, u^1 to minimize (17). This is a fully observed problem since u^0, u^1 satisfy (14) and enter directly into (17). We then have the following very important result:

Theorem 2: Under the above assumptions the optimal $\gamma, \hat{\theta}^0, \hat{\theta}^1$ (which exist according to theorem 1) are functions of u^0, u^1 only. That is they depend on y^t only through the unnormalized conditional densities u^0, u^1.

Proof: The proof follows from appropriate modifications of the results in [7]-[8] and will appear elsewhere.

The significance of the result is that it provides the basic structure of the optimal processor by identifying u^0, u^1 as the sufficient statistics for the original problem. Furthermore the result is free from structural assumptions on the detection and estimation costs and can be established in far greater generality than the results presented here may indicate.

In figure 4 below we give a pictorial illustration of the result. We basically have to run two "filters" in parallel, one for each hypothesis. The output of each filter (which by the way is represented by the bilinear stochastic p.d.e. (14)) is the unnormalized conditional probability density of x^0, θ^0 or

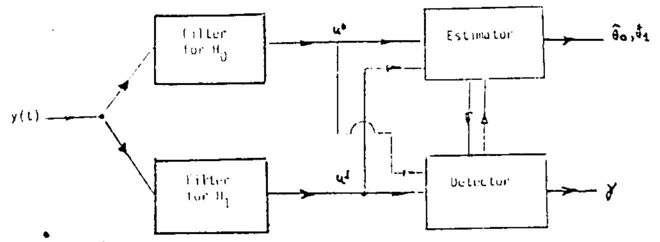

Fig. 4. Illustrating the generic structure of the optimal processor.

x^1, θ^1 given H^0 or H^1. Each filter is driven directly by the observations.

The estimator, detector and their coupling will depend on the explicit cost structure. They are problem dependent. Their explicit functioning can be computed as our final result indicates.

Theorem 3: The explicit dependence of γ (which is discrete valued), $\hat{\theta}_0, \hat{\theta}_1$ on u^0, u^1 can be determined by solving a variational inequality on the space of solutions of (14).

Proof: The result is rather technical. A complete proof will be given elsewhere. It follows by appropriate modifications to the results of [5].

This result opens the way for promising electronic implementation of the optimal processor by the following steps: (1) solve numerically the resulting variational inequality using the methods of [9], (2) implement the resulting numerical algorithm by a special purpose, multiprocessor, VLSI device along the lines of [1]. In simple cost cases explicit solutions of the variational inequality can be obtained of course.

MOTIVATION AND EXAMPLES FROM RADAR TRACKING LOOPS

The primary motivation for the mathematical problem studied in section 3 comes from design consideration of advanced (smart) sensors in guided platforms. To be more specific let us consider radar sensors. The radar return from a scatterer carries (depending on the radar sophistication) significant information about a scatterer. For example range, Doppler extend, shape and extend, motion, of a scatterer can be extracted from a radar return by appropriate processing. In today's dense environment a very important function of an advanced processor is classification of scatterers. This function is required for example by sensors participating in a surveillance network (since threats must be classified, so that appropriate response can be applied), in electronic warfare (since decoys and other countermeasures can be designed to emulate target characteristics) and in tracking radars (since the sensor often must develop a tracking path for a designated priority target).

A related equally important function of a radar receiver is the estimation of parameters embedded in the return signal. For example pulse length, pulse repetition frequency, amplitude scintillation spectrum, conical scan frequency, antenna pointing, surface roughness. The two problems of detection and estimation are indeed closely related, as explained earlier.

In our earlier work [2]-[3] we have developed statistical models for distributed scatterers which can represent accurately phenomena characteristic of distributed scatterer radar returns such as amplitude scintillation and angle noise or glint. In addition we have developed similar statistical models for the effects of multipath on radar returns, for sea clutter returns and for chaff cloud returns.

The models developed in [2]-[3] are of the form

$$dx(t) = A(t,\theta)x(t)dt + B(t,\theta)dw(t) \tag{18}$$
$$dy(t) = h(t,x(t),\theta)dt + dv(t)$$

Furthermore A,B,h are piecewise constant with respect to time since the models developed in [2]-[3] are piecewise stationary. For example in [2] we used models like (18) to describe the RCS scintillation for ships. The same type models can be used for other distributed targets such as tanks or armored vehicles. For example when the return appears spiky, indicating higher probability of strong return, an appropriate model is provided by a lognormal process, where x(·) in (18) is scalar and h is chosen to be an exponential function of x. For chaff clouds a more appropriate model is provided by a Rayleigh process, where x(·) is two dimensional, with the two components being identically distributed, independent Gaussian random processes and

$$h(t,x(t),\theta) = \sqrt{x_1^2(t) + x_2^2(t)}.$$

Clearly then in target discrimination problems with distributed targets of this type one encounters problems like those treated in section 3. It is important to note that since the first of (18) is linear the corresponding filtering and stochastic control problems described in section 3 are definitely more tractable. For further examples of this type we refer the reader to [2]-[3].

Further research is needed to apply the powerful results of section 3 to specific problems in order to evaluate current design principles and more importantly in order to suggest new electronic implementations capable of performing in a dense, hostile environment. In particular the methodology developed in 3 can be used to identify the cost structures that lead to the specific hierarchies suggested in figures 2-4.

REFERENCES

[1] Baras, J.S. (1981). Approximate solutions to nonlinear filtering problems by direct implementation of the Zakai equation. Proc. of 1981 IEEE Decision and Control Conference, San Diego, pp. 309-310.

[2] Baras, J.S. (1983 (in publication)). Sea clutter statistical models. Naval Research Laboratory Technical Report.

[3] Baras, J.S. (May 1978). Ship RCS scintillation simulation. Naval Research Laboratory Technical Report 8189.

[4] Baras, J.S., G.L. Blankenship, and W.E. Hopkins, Jr. (Feb. 1983). Existence, uniqueness, and asymptotic behavior of solutions to a class of Zakai equations with unbounded coefficients. IEEE Trans. on Autom. Control, Vol. AC-28, pp.203-214.

[5] Bensoussan, A., and J.L. Lions. (1982). Applications of Variational Inequalities in Stochastic Control, North-Holland.

[6] Birdsall, T.G., and J.O. Gobien. (Nov. 1973). Sufficient statistics and reproducing densities in simultaneous sequential detection and estimation. IEEE Trans. on Inform. Theory, Vol. IT-19, pp. 760-768.

[7] Bismut, Jean-Michel. (1982). Partially observed diffusions and their control. SIAM J. Control and Optim., Vol. 20, pp. 302-309.

[8] Fleming, W.H., and E. Pardoux. (1982). Optimal control for partially observed diffusions. SIAM J. Control and Optim., Vol. 20, pp. 261-286.

[9] Glowinski, R., J.L. Lions, and R. Tremolieres. (1976). Analyse Numberique Des Inequations Variationnelles, Vols. 1, 2. Dunod.

[10] Jaffer, A., and S. Gupta. (Jan. 1972). Coupled detection estimation of Gaussian processes in Gaussian noise. IEEE Trans. Inform. Theory, Vol. IT-18, pp. 106-110.

[11] Jaffer, A., and S. Gupta. (Sept. 1971). Recursive Bayesian estimation with uncertain observation. IEEE Trans. Inform. Theory, Vol. IT-17, pp. 614-616.

[12] Middleton, D., and R. Esposito. (May 1968). Simultaneous optimum detection and estimation of signals in noise. IEEE Trans. on Information Theory, Vol. IT-14, No. 3, pp. 434-444.

[13] Scharf, L., and D. Lytle. (July 1971). Signal detection in Gaussian noise of unknown level: an invariance application. IEEE Trans. Inform. Theory, Vol. IT-17, pp. 404-411.

[14] Stroock, D.W., and S.R.S. Varadhan. (1975). Multidimensional Diffusion Processes. Springer Verlag, Secaucus, NJ.