INFORMATION AND CONTROL 19, 75–92 (1971)

Joint Detection, Estimation and System Identification*

D. G. LAINIOTIS

Department of Electrical Engineering and Electronics Research Center, The University of Texas at Austin, Austin, Texas 78712

Recent results of Middleton and Esposito (1968) and Lainiotis (1969) on single-shot joing detection-estimation for discrete data are extended to the single-shot continuous data case and generalized to joint Bayesian detection-estimation-system identification. Moreover, previous results were generalized to the case of causal estimator. Specifically, it is shown that the above problem constitutes a class of nonlinear mse estimation problems, with the attendant difficulties in realizing the optimal nonlinear estimators. However, by utilizing the adaptive approach, closed form integral expressions are given. These are given in terms of the generalized likelihood ratio $\Lambda(t)$, which is a sufficient statistic for Bayes-optimal compound detection. The latter in turn is specified by a continuum (for continuous θ) θ-conditional likelihood ratious $\Lambda(t/\theta)$ each of which is the LR for testing for the model specified by the parameter value θ. The latter LR's are, moreover, given in terms of optimal mse causal estimators. In essence then, it has been shown that system identification is equivalent to multihypothesis testing, with a continuum or finite sequence of hypotheses, respectively, for continuous or finite discrete range of θ.

I. INTRODUCTION

The problem considered here is the single-shot, compound detection-estimation problem specified by the following equations

$$\begin{aligned} z(t) &= s(t) + v(t) \\ &= \beta y(t) + v(t), \end{aligned} \tag{1}$$

where $\{z(t)\}$ is the m-vector observable random process, $v(t)$ is the white gaussian, zero-mean observation noise random process with covariance matrix $R(t)$, and β is the so-called indicator variable which takes values 1 or 0

* This work was supported by the Air Force Office of Scientific Research under Grant AFOSR 69-1764 and Joint Services Electronics Program Grant AFOSR 69-1792.

depending on whether H_1 (signal present) or H_0 (signal absent) is true, with a priori probability p_1, p_0, respectively ($\sum_0^1 p_i = 1$).

Moreover, the signal random process $\{y(t)\}$ is assumed to be "adequately" modeled by the state-variable model

$$y(t) = h(x(t), t; \theta) \tag{2a}$$

and

$$\frac{dx(t)}{dt} = f(x(t), t; \theta) + g(x(t), t; \theta)\, u(t), \tag{2b}$$

where $\{x(t)\}$ is the n-vector signal "state" random process, and $\{u(t)\}$ is a zero-mean q-vector white gaussian process, independent of $\{v(t)\}$, whose covariance matrix is the identity matrix. The functionals $h(\cdot)$, $f(\cdot)$, and $g(\cdot)$ are time-varying nonlinear functionals of the state vector $x(t)$, and of θ. The s-vector parameter θ is time-invariant and, if known, specifies the above model completely. However, in the compound problem considered, θ is unknown and following the Bayesian approach, it is considered a random variable with known or assumed a priori density $p(\theta)$.

The initial state-vector $x(t_0)$ is independent of $\{v(t)\}$ and $\{u(t)\}$ for $t \geqslant t_0$ and has a known θ-conditional gaussian a priori density, denoted by $p(x(t_0)/t_0, \theta)$ with means $\hat{x}(t_0/t_0, \theta)$ and covariance matrix $P(t_0/t_0, \theta)$.

Given the continuous record $\nu_t = \{z(\tau); \tau \in (t_0, t)$ for the current baud, for which a decision is to be made, the Bayes-risk minimizing decision rule as well as the minimum mean-square estimates of the signal random process $s(t)$, and the unknown parameter θ are to be derived.

Middleton and Esposito (1968) considered the single-shot joint detection-estimation problem for discrete data. Specifically, for both the strong- and weak-coupling case and for quandratic-cost function (mse) estimation of $s(t)$, they obtained the following fundamental relationship

$$\hat{s}(t/\nu_n) = \frac{\Lambda_n}{1 + \Lambda_n} \hat{s}(t/\nu_n, \beta = 1), \tag{3}$$

where $\hat{s}(t/t) \equiv E[s(t)/\nu_n]$, $\hat{s}(t/\nu_n, \beta = 1) \equiv E[s(t)/\nu_n, \beta = 1]$ and Λ_n is the generalized likelihood ratio Middleton and Esposito (1968) defined as

$$\Lambda_n \equiv \frac{\rho \int p(\nu_n/\beta = 1, \theta)\, p(\theta)\, d\theta}{p(\nu_n/\beta = 0)}, \qquad \rho \equiv \frac{p_1}{p_0}, \tag{4}$$

where $\nu_n = \{z(t_1), z(t_2), \ldots, z(t_n); t_i \in (t_0, t)\}$, and $p(\nu_n/\beta = 1, \theta)$ is the θ-parameter conditional density of ν_n under H_1, and $p(\nu_n/\beta = 0)$ is the probability density of ν_n under H_0. Note θ is connected to the signal only. No other restrictions were imposed on the model.

Lainiotis (1969), moreover, recently established the fundamental fact that detection is indeed mean-square estimation of the indicator variable β for the model of Eq. (1). Specifically, he established that the Bayes-optimal decision procedure is to decide H_1 or H_0 depending on whether $\hat{\beta}(\lambda_n) \gtrless c_0$, respectively, where c_0 depends on the a priori costs for each decision, and $\hat{\beta}(\nu_n) \equiv E[\beta/\nu_n]$. Moreover, Lainiotis (1969) showed that

$$\begin{aligned}\hat{\beta}(\nu_n) &= p(\beta = 1/\nu_n) \\ &= \frac{\Lambda_n}{1 + \Lambda_n}.\end{aligned} \tag{5}$$

Note that Eq. (5) above is a special case, Esposito (1969), of Middleton and Esposito's formuls, since $\hat{\beta}(\nu_n/\beta = 1)$ is obviously one.

In this paper, the above results of Middleton and Esposito (1968) and Lainiotis (1969) are extended to the continuous data case and generalized to joint detection-estimation-system identification, as well as to the case of causal mse estimators.

II. Optimal Joint Detection, Estimation and System Identification

The problem stated in the introduction constitutes joint detection (i.e., β), signal estimation (i.e., s(t)), and system identification (i.e., θ). It may be shown to be simply a nonlinear mean-square estimation problem. To see this, augment the state-vector $x(t)$ with θ and β, so that the augmented state-vector $x_a(t) \equiv [x^T(t) \vdots \theta^T \vdots \beta]^T$. Then the model defining equations become

$$\frac{dx_a(t)}{dt} = f_a(x_a(t), t) + g_a(x_a(t), t)\, u(t), \tag{6a}$$

$$z(t) = h_a(x_a(t), t) + v(t), \tag{6b}$$

where

$$f_a(x_a(t), t) \equiv [f^T(x(t), t; \theta) \vdots 0 \vdots 0]^T, \qquad g_a(x_a(t), t) \equiv [g^T(x(t), t; \theta) \vdots 0 \vdots 0]^T$$

and

$$h_a(x_a(t), t) \equiv \beta h(x(t), t; \theta).$$

It is apparent from their definitions that $f_a(\cdot)$, $g_a(\cdot)$, and $h_a(\cdot)$ are, in general, nonlinear functionals of $x_a(t)$.

In view of the discussion in the introduction, it is readily seen that the optimal mean-square estimate (mse) of $x_a(t)$ given ν_t, denoted $\hat{x}_a(t/t) \equiv [\hat{x}^T(t/t) \vdots \hat{\theta}^T(t) \vdots \hat{\beta}(t)]^T$, contains all the quantities for the solution of the joint minimum Bayes-risk detection and mse estimation and system identification problem. Thus, the above problem is equivalent to mse nonlinear estimation, with the attendant difficulties in realizing nonlinear estimators, Jazwinski (1970). ∎

In this paper, by utilizing the adaptive approach, closed form formulas are obtained. The desired adaptive realization is obtained by considering both θ and β as an augmented parameter vector $\alpha = [\theta^T \vdots \beta]^T$. The adaptive realization is given in the following:

THEOREM I. (*Partition Theorem*).

$$\hat{\beta}(t) = \frac{\rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}, \tag{7a}$$

$$\hat{s}(t/t) = \frac{\rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}\, \hat{s}_1(t/t), \tag{7b}$$

$$\hat{\theta}(t) = \hat{\theta}(t_0)\, p(\beta = 0/t) + \frac{\rho \int \theta \Lambda(t/\theta)\, p(\theta)\, d\theta}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}, \tag{7c}$$

where

$$\begin{aligned} \hat{\beta}(t) &\equiv E[\beta/t] \\ &= p(\beta = 1/t) = 1 - p(\beta = 0/t), \\ \hat{s}(t/t) &\equiv E[s(t)/t], \\ \hat{s}_1(t/t) &\equiv E[s(t)/t, \beta = 1], \\ \hat{\theta}(t) &\equiv E[\theta/t], \end{aligned}$$

and where

$$\begin{aligned} p(\theta/t) &= \frac{1 + \rho\Lambda(t/\theta)}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}\, p(\theta), \\ \hat{\theta}(t_0) &= E[\theta/t_0] = \textit{the a priori mean}, \end{aligned} \tag{7d}$$

and

$$\Lambda(t/\theta) \equiv \exp\left\{\int_{t_0}^{t} \hat{h}_1^T(\sigma/\sigma, \theta)\, R^{-1}(\sigma)\, z(\sigma)\, d\sigma - \tfrac{1}{2}\int_{t_0}^{t} \| \hat{h}_1(\sigma/\sigma, \theta)\|^2_{R^{-1}(\sigma)}\, d\sigma\right\}, \tag{7e}$$

where $\hat{h}_1(\sigma/\sigma, \theta) \equiv E[h(x(\sigma), \sigma; \theta)/\sigma, \theta, \beta = 1]$.

Proof. The proof is given in the appendix.

The following remarks on the above results are pertinent:

It is seen from Eq. (7a) that the statistic sufficient for Bayes-optimal detection is $\Lambda(t) \equiv \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta$. $\Lambda(t)$ is the continuous data generalized likelihood ratio for compound detection. In terms of $\Lambda(t)$, Eq. (7a) is the same as the one obtained by Lainiotis (1969) for discrete data.

Moreover, we note that the generalized likelihood ratio is specified by a continuum of θ-conditional likelihood ratios (LR) $\Lambda(t/\theta)$, each of which is the LR for the detection problem

$$H_1 : z(t) = h(x(t), t; \theta) + v(t),$$
$$H_0 : z(t) = v(t),$$

for θ a specified admissible value (admissible in the sense of $p(\theta)$). In other words, $\Lambda(t/\theta)$ is the LR for testing whether the signal generated by the model specified by parameter value θ is present. This leads us to the conclusion that, essentially, system identification is equivalent to multihypothesis testing with a continuum of hypotheses (for continuous θ) corresponding to each possible model indexed by θ.

In addition, we note from Eq. (7e) that $\Lambda(t/\theta)$ is in the canonical estimator-correlator form of Kailath (1969), which is particularly well-suited to interpretation and approximation. However, it must be pointed out that the generalized likelihood ratio for continuous data $\Lambda(t)$ is applicable to Bayes-optimal compound detection and as such is more general than Kailath's (1969) LR, the latter being applicable to classical hypothesis testing without consideration of prior probabilities.

Equation (7b) above is the continuous data version of Middleton and Esposito's [1969] result. We note that it has similar form and interpretation as their result, with, however, a very significant difference. Namely, Middleton and Esposito's [1969] results were valid, essentially, for noncausal estimators since in the discrete approach used in their paper the waveform is estimated as a whole given all the data. In contrast the results given herein, generalize

Middleton and Esposito's (1969) results not only to the continuous case but also to the case of causal estimators.[1]

Using Eq. (7a), $\hat{s}(t/t)$ may be given the interesting form

$$\hat{s}(t/t) = \hat{\beta}(t)\, \hat{s}_1(t/t). \tag{8}$$

Similarly, for $\hat{\theta}(t)$, we have

$$\hat{\theta}(t) = \hat{\theta}(t_0)[1 - \hat{\beta}(t)] + \frac{\rho \int \theta \Lambda(t/\theta)\, p(\theta)\, d\theta}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}\,. \tag{9}$$

Equations (8–9) establish once more the fact that optimal estimation of $s(t)$ and θ requires nonlinear processing of the data. This is true even in the case of linear models, namely, when $f(\cdot)$, $g(\cdot)$, and $h(\cdot)$ are linear functionals of the state-vector $x(t)$.

In the following corollary, expressions for the mse estimates of $h(\cdot)$ and $s(t)$ as well as the corresponding conditional error-covariance matrices are given.

COROLLARY 1.

$$\hat{s}(t/t) = \hat{\beta}(t)\, \hat{h}_1(t/t), \tag{10a}$$

$$P_s(t/t) = \hat{\beta}(t)\, P_{h_1}(t/t), \tag{10b}$$

where

$$\hat{h}_1(t/t) \equiv E[h(x(t), t, \theta)/t, \beta = 1],$$

$$P_s(t/t) \equiv E\{[s(t) - \hat{s}(t/t)][s(t) - \hat{s}(t/t)]^T/t\},$$

and

$$P_{h_1}(t/t) \equiv E\{[h(x(t), t, \theta) - \hat{h}_1(t/t)][h(x(t), t, \theta) - \hat{h}_1(t/t)]^T/t, \beta = 1\},$$

$$\hat{h}_1(t/t) = \int \hat{h}_1(t/t, \theta)\, p_1(\theta/t)\, d\theta, \tag{10c}$$

where $\hat{h}_1(t/t, \theta)$ *was defined previously and* $p_1(\theta/t)$ *is given by*

$$p_1(\theta/t) \equiv p(\theta/t, \beta = 1) = \frac{\Lambda(t/\theta)}{\int \Lambda(t/\theta)\, p(\theta)\, d\theta}\, p(\theta), \tag{11}$$

where $p(\theta)$ *is the a priori density of* θ; ρ *and* $\Lambda(t/\theta)$ *were defined previously.*

[1] This important difference was pointed out to the author by the reviewer.

The conditional error-covariance matrix $P_{h_1}(t/t)$ *is given by*

$$P_{h_1}(t/t) = \int \{P_{h_1}(t/t, \theta) + [\hat{h}_1(t/t, \theta) - \hat{h}_1(t/t)][\hat{h}_1(t/t, \theta) - \hat{h}_1(t/t)]^T\} p_1(\theta/t)\, d\theta, \tag{12}$$

where $P_{h_1}(t/t, \theta)$ *and* $\hat{h}_1(t/t, \theta)$ *are given* (*Jazwinski*, 1970), *as the solution of coupled partial differential equations, whose solution is possible only for special cases, e.g., for linear models.*

Proof. The proof is given in the appendix.

At this point we note that the realization of the nonlinear estimator and the associated computational requirements depend essentially on the range of admissible values for θ, viz., whether the range of θ is discrete (finite) or continuous. In most applications, the range is continuous resulting in excessive computational requirements. One approach for alleviating the problem is to quantize the θ-space. Moreover, such quantization is reasonable in view of the fact that quantization occurs naturally in any physical realization of a system. In any case, either because θ is discrete or because of quantization of a continuous range, $p(\theta) \simeq \sum_{i=1}^{n} p(\theta_i)\, \delta(\theta - \theta_i)$. For such $p(\theta)$ all integrations with respect to θ in Theorem I and Corollary 1 become sums of N terms, each term corresponding to a particular value of θ, θ_i, $i = 1, 2,\ldots, N$.

It was shown earlier that system identification constitutes in essence a hypothesis testing problem with a continuum of hypotheses. Parameter space quantization, however, has reduced the problem to one with a finite set of hypotheses. In this context, system identification constitutes a sequence of hypotheses testing problems each corresponding to testing for the model indexed by parameter value θ_i, $i = 1, 2,\ldots, N$.

Applying Itô's, Jazwinski (1970), differentiation rule to Eqs. (7a, 7d, 11), we obtain stochastic differential equations for the temporal evolution of $\hat{\beta}(t)$, $p(\theta/t)$ and $p_1(\theta/t)$. These are given in the following corollary:

COROLLARY 2.

$$d\hat{\beta}(t) = \hat{\beta}(t)[1 - \hat{\beta}(t)]\, \hat{h}_1^T(t/t)\, R^{-1}(t)[dz(t) - \hat{\beta}(t)\, \hat{h}_1(t/t)\, dt]$$

with initial condition $\hat{\beta}(t_0) = p_1$, *and*

$$dp_1(\theta/t) = p_1(\theta/t)[\hat{h}_1(t/t, \theta) - \hat{h}_1(t/t)]\, R^{-1}(t)[dz(t) - \hat{\beta}(t)\, \hat{h}_1(t/t)\, dt]$$

with initial condition $p_1(\theta/t_0) = p(\theta)$

$$dp(\theta/t) = \hat{\beta}(t)[p_1(\theta/t)\, \hat{h}_1^T(t/t, \theta) - p(\theta/t)\, \hat{h}_1^T(t/t)]\, R^{-1}(t)[dz(t) - \hat{\beta}(t)\, \hat{h}_1(t/t)\, dt]$$

with initial condition $p(\theta/t_0) = p(\theta)$.

Proof. The proof consists of straightforward application of Ito's differentiation rule to Eqs. (7a, 7d, 11), and as such it is omitted.

Again it must be emphasized that algorithms for the evaluation of $\hat{h}_1(t/t, \theta)$ are needed. Such algorithms are possible for special cases only, such as the case of linear models. This case is treated in the following section.

We note that both integral expressions as well as stochastic differential equations for $\hat{\beta}(t)$, $p_1(\theta/t)$ and $p(\theta/t)$ were given above. However, in the opinion of this author, the integral expressions are far more valuable from a practical standpoint, namely implementation, as well as from an interpretation point of view. These opinions are based on the fact that the stochastic differential equations are nonlinear and coupled, and their solution is, in general, not easily forthcoming. Moreover, approximate solutions of these equations are ad hoc and the effect of the approximations made can not easily be assessed. In contrast, note that the effects of approximations made in evaluating the integral expressions, such as the (finer) quantization of the θ-range, can be easily assessed and improved upon if desired.

III. Special Case: Linear Models

In this section, the case of linear dynamic models is considered. That is, the signal random process is given by

$$y(t) = H(t, \theta)\, x(t), \tag{13a}$$

and

$$\frac{dx(t)}{dt} = F(t, \theta)\, x(t) + G(t, \theta)\, u(t), \tag{13b}$$

where $\{y(t)\}$, $\{u(t)\}$ and $\{x(t)\}$ were defined previously. The results for this case are summarized in the following corollary:

Corollary 3.

$$\hat{s}(t/t) = \hat{\beta}(t)\, \hat{y}_1(t/t), \tag{14a}$$

$$P_s(t/t) = \hat{\beta}(t) \left[\int \{H(t, \theta)\, P_1(t/t, \theta)\, H^T(t, \theta) + [H(t, \theta)\, \hat{x}_1(t/t, \theta) - \hat{y}_1(t/t)] \cdot [H(t, \theta)\, \hat{x}_1(t/t, \theta) - \hat{y}_1(t/t)]^T\}\, p_1(\theta/t)\, d\theta \right], \tag{14b}$$

where

$$\hat{y}_1(t/t) = \int H(t, \theta)\, \hat{x}_1(t/t, \theta)\, p_1(\theta/t)\, d\theta$$

and $\hat{\beta}(t)$ *and* $p_1(\theta/t)$ *are given by the stochastic differential equations*:

$$d\hat{\beta}(t) = \hat{\beta}(t)[1 - \hat{\beta}(t)]\, \hat{y}_1^T(t/t)\, R^{-1}(t)[dz(t) - \hat{\beta}(t)\, \hat{y}_1(t/t)\, dt] \tag{14c}$$

with initial condition $\hat{\beta}(t_0) = p_1$, *and*

$$dp_1(\theta/t) = p_1(\theta/t)[H(t, \theta)\, \hat{x}_1(t/t, \theta) - \hat{y}_1(t/t)]\, R^{-1}(t)[dz(t) - \hat{y}_1(t/t)\, dt], \tag{14d}$$

with initial condition $p_1(\theta/t_0) = p(\theta)$.

$$\begin{aligned} dp(\theta/t) = {} & \hat{\beta}(t)[p_1(\theta/t)\, H(t, \theta)\, \hat{x}_1(t/t, \theta) \\ & - p(\theta/t)\, \hat{y}_1(t/t)]\, R^{-1}(t)[dz(t) - \hat{\beta}(y)\, \hat{y}_1(t/t)\, dt] \end{aligned} \tag{14e}$$

with initial condition $p(\theta/t_0) = p(\theta)$.

Integral expressions for $\hat{\beta}(t)$ and $p_1(\theta/t)$ are given by Eqs. (7a) and (11), where now $\Lambda(t/\theta)$ takes the form

$$\begin{aligned} \Lambda(t/\theta) \equiv \exp \Big\{ & \int_{t_0}^{t} \hat{x}_1^T(\sigma/\sigma, \theta)\, H^T(\sigma, \theta)\, R^{-1}(\sigma)\, z(\sigma)\, d\sigma \\ & - \tfrac{1}{2} \int_{t_0}^{t} \| H(\sigma, \theta)\, \hat{x}_1(\sigma/\sigma, \theta)\|^2_{R^{-1}(\sigma)}\, d\sigma \Big\}. \end{aligned} \tag{15}$$

The signal and model conditional estimate $\hat{x}_1(t/t, \theta)$ and the corresponding error-covariance matrix $P_1(t/t, \theta)$ are now given by the well-known Kalman–Bucy, Jazwinski (1970) equations:

$$\begin{aligned} d\hat{x}_1(t/t, \theta) = {} & F(t, \theta)\, \hat{x}_1(t/t, \theta) \\ & + P_1(t/t, \theta)\, H^T(t)\theta)\, R^{-1}(t)[dz(t) - H(t, \theta)\, \hat{x}_1(t/t, \theta)] \end{aligned} \tag{16a}$$

with initial condition $\hat{x}_1(t/t, \theta) = \hat{x}(t_0/t_0)$, and

$$\begin{aligned} \frac{dP_1(t/t)}{dt} = {} & F(t, \theta)\, P_1(t/t, \theta) + P_1(t/t, \theta) F^T(t, \theta) + G(t, \theta)\, G^T(t, \theta) \\ & - P_1(t/t, \theta)\, H^T(t, \theta)\, R^{-1}(t)\, H(t, \theta)\, P_1(t/t, \theta) \end{aligned} \tag{16b}$$

with initial condition $P_1(t_0/t_0) = P(t_0/t_0)$.

Proof. The proof is based on simple application of Theorem I and Corollary 1 to the linear model given in Eqs. (13), and as such it is omitted.

IV. Suboptimal Nonlinear Estimation Algorithm

It was shown in Section II, that the problem of joint detection-estimation-system identification constitutes a nonlinear estimation problem. But, as was pointed out earlier, the optimal nonlinear filter is specified, in general, by an infinite set of coupled stochastic partial differential equations. Such specification is not useful since it is in general unrealizable. This problem was partially alleviated by using an adaptive approach with quantization of the unknown parameter space. In this section an approximate nonlinear estimation algorithm is used that does not require parameter space quantization. This is based on the so-called relinearized Kalman–Bucy filter (Light, 1970).

It is widely known that the simplest method of suboptimal nonlinear filtering is that of Kalman–Bucy filtering about a nominal trajectory. Licht (1970) demonstrates that if the optimal estimate instead of fixed a priori trajectory is chosen as the nominal trajectory, the resulting approximate relinearized filter is much more effective than the usual linearized filter about constant nominal trajectory.

The relinearized Kalman–Bucy filter for the model defined by 6(a) and 6(b) is given by the following set of differential equations:

$$\frac{d\hat{x}_a(t/t)}{dt} = f_a(\hat{x}_a(t/t), t) + P(t/t)\,\nabla h_a^T(\hat{x}_a(t/t), t)\,R^{-1}(t)[z(t) - h_a(\hat{x}_a(t/t), t)], \tag{17a}$$

$$\begin{aligned}\frac{dP(t/t)}{dt} = {} & \nabla f_a(\hat{x}_a(t/t), t)\,P(t/t) + P(t/t)\,\nabla f_a^T(\hat{x}_a(t/t), t) + g_a(\hat{x}_a(t/t), t) \\ & \cdot g_a^T(\hat{x}_a(t/t), t) - P(t/t)\,\nabla h_a^T(\hat{x}_a(t/t), t)\,R^{-1}(t)\,\nabla h_a(\hat{x}_a(t/t), t)\,P(t/t),\end{aligned} \tag{17b}$$

where

$$\hat{x}_a(t/t) = E[x_a(t)/t],$$
$$P(t/t) = E[[x_a(t) - \hat{x}_a(t/t)][x_a(t) - \hat{x}_a(t/t)]^T],$$

with initial conditions

$$\hat{x}_a(t_0/t_0) = E[x_a(t_0)],$$
$$P(t_0/t_0) = E[[x_a(t_0) - \hat{x}_a(t_0/t_0)][x_a(t_0) - \hat{x}_a(t_0/t_0)]^T],$$

where $\nabla h_a(x_a(t/t), t)$ and $\nabla f_a(x_a(t/t), t)$ are Jacobian matrices of $h_a(x_a(t/t), t)$ and $f_a(x_a(t/t), t)$, respectively.

In order to illustrate the implementation of the relinearized filter, the linear model in Section III is considered. The results are given in the following corollary:

COROLLARY 4.

$$\begin{bmatrix} d\hat{x}(t/t) \\ d\hat{\theta}(t) \\ d\hat{\beta}(t) \end{bmatrix} = \begin{bmatrix} F(t, \hat{\theta}(t))\, \hat{x}(t/t) \\ 0 \\ 0 \end{bmatrix} dt + P(t/t) \begin{bmatrix} \hat{\beta}(t)\, H^T(t, \hat{\theta}(t)) \\ \hat{\beta}(t)\, \hat{x}^T(t/t)\, \nabla H^T(t, \hat{\theta}(t)) \\ \hat{x}^T(t/t)\, H^T(t, \hat{\theta}(t)) \end{bmatrix} \cdot R^{-1}(t)[dz(t) - \hat{\beta}(t)\, H(t, \hat{\theta}(t))\, \hat{x}(t/t)\, dt], \tag{18a}$$

$$\begin{aligned} P(t/t) = {} & \begin{bmatrix} F(t, \hat{\theta}(t)) & F(t, \hat{\theta}(t))\, \hat{x}(t/t) & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} P(t/t)\, dt \\ & + P(t/t) \begin{bmatrix} F^T(t, \hat{\theta}(t)) & 0 & 0 \\ \hat{x}^T(t/t)\, \nabla F^T(t, \hat{\theta}(t)) & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} dt \\ & + \begin{bmatrix} G(t, \hat{\theta}(t))\, G^T(t, \hat{\theta}(t)) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} dt \\ & - P(t/t) \begin{bmatrix} \hat{\beta}(t)\, H^T(t, \hat{\theta}(t)) \\ \hat{\beta}(t)\, \hat{x}^T(t/t)\, \nabla H^T(t, \hat{\theta}(t)) \\ \hat{x}^T(t/t)\, H^T(t, \hat{\theta}(t)) \end{bmatrix} \\ & \cdot R^{-1}(t) \begin{bmatrix} \hat{\beta}(t)\, H^T(t, \hat{\theta}(t)) \\ \hat{\beta}(t)\, \hat{x}^T(t/t)\, \nabla H^T(t, \hat{\theta}(t)) \\ \hat{x}^T(t/t)\, H^T(t, \hat{\theta}(t)) \end{bmatrix} P(t/t)\, dt \end{aligned} \tag{18b}$$

with initial conditions

$$\hat{x}(t_0/t_0) = E[x(t_0)], \quad \hat{\beta}(t_0) = p_1, \quad \hat{\theta}(t_0) = E[\theta] \quad \textit{and} \quad P(t_0/t_0) = P(t_0),$$

and where $P(t/t)$ *is the error covariance matrix of the estimate, i.e.,*

$$P(t/t) = E\left\{ \begin{bmatrix} x(t) - \hat{x}(t/t) \\ \theta - \hat{\theta}(t) \\ \beta - \hat{\beta}(t) \end{bmatrix} \begin{bmatrix} x(t) - \hat{x}(t/t) \\ \theta - \hat{\theta}(t) \\ \beta - \hat{\beta}(t) \end{bmatrix}^T \Big/ t \right\}.$$

Proof. The proof consists of application of the relinearized equations to the linear model of Section III, after augmenting the state vector as in (6a) and (6b), and will therefore be omitted.

In the following, two special cases of the corollary are given solely for illustrative purpose.

Case I. Estimation/System identification. Let $\beta = 1$, θ, a scalar parameter and the model equations have the following form:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \theta x(t) \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t), \qquad z(t) = x(t) + v(t).$$

Then the relinearized Kalman–Bucy filter equations for the optimal estimates, $\hat{x}(t/t)$ and $\hat{\beta}(t)$ and estimation error covariance matrix are given by the following

$$\begin{bmatrix} d\hat{x}(t/t) \\ d\hat{\theta}(t) \end{bmatrix} = \begin{bmatrix} \hat{\theta}(t)\, \hat{x}(t/t) \\ 0 \end{bmatrix} dt + P(t/t) \begin{bmatrix} I \\ 0 \end{bmatrix} R^{-1}(t)[dz(t) - \hat{x}(t/t)\, dt],$$

$$P(t/t) = \begin{bmatrix} \hat{\theta}(t)I & \hat{x}(t/t) \\ 0 & 0 \end{bmatrix} P(t/t)\, dt + P(t/t) \begin{bmatrix} \hat{\theta}(t)I & 0 \\ \hat{x}^T(t/t) & 0 \end{bmatrix} dt$$
$$+ \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} dt - P(t/t) \begin{bmatrix} R^{-1} & 0 \\ 0 & 0 \end{bmatrix} P(t/t)\, dt.$$

As one can easily see upon examining the above equations, the highest conditional moments of $x(t)$ and θ that are required to obtain $x(t/t)$ and $\theta(t)$ are second order moments. Instead, the differential equation for optimal estimates require that the lower order moments equations contain terms which are functions of higher order moments, resulting in infinite dimensional problem.

Case II. Joint detection-estimation. Let $\theta = 1$ and β, unknown, and let this joint detection-estimation problem take the following form:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\beta} \end{bmatrix} = \begin{bmatrix} x(t) \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t), \qquad z(t) = \beta x(t) + v(t),$$

where $\beta = 1$, or 0.

Then the relinearized Kalman–Bucy differential equations for the optimal estimates and error covariance matrix are given by the following equations:

$$\begin{bmatrix} d\hat{x}(t/t) \\ d\hat{\beta}(t) \end{bmatrix} = \begin{bmatrix} \hat{x}(t/t) \\ 0 \end{bmatrix} dt + P(t/t) \begin{bmatrix} \hat{\beta}(t)I \\ \hat{x}^T(t/t) \end{bmatrix} R^{-1}(t)[dz(t) - \hat{\beta}(t)\, \hat{x}(t/t)\, dt],$$

and

$$dP(t/t) = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} P(t/t)\, dt + P(t/t) \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} dt + \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} dt$$
$$- P(t/t) \begin{bmatrix} \hat{\beta}(t) I \\ \hat{\beta}(t)\, \hat{x}^T(t/t) \end{bmatrix} R^{-1} \begin{bmatrix} \hat{\beta}(t) I \\ \hat{\beta}(t)\, \hat{x}(t/t) \end{bmatrix} P(t/t)\, dt.$$

The comments given previously in case I apply here also.

V. Conclusions

Recent results of Middleton and Esposito (1968) and Lainiotis (1969) on single-shot joint detection-estimation for discrete data are extended to the single-shot continuous data case and generalized to joint Bayesian detection-estimation and system identification, as well as to causal estimators. Specifically, it is shown that the above problem constitutes a class of nonlinear mse estimation problems, with the attendant difficulties in realizing the optimal nonlinear estimators. However, by utilizing the adaptive approach, viz., by considering (θ, β) as an unknown parameter to be learned, closed form expressions are given. These are given in terms of the generalized likelihood ratio $\Lambda(t)$, which is a sufficient statistic for Bayes-optimal compound detection. The latter, in turn, is specified by a continuum (for continuous θ) θ-conditional likelihood ratios $\Lambda(t/\theta)$, each of which is the LR for testing for the model specified the parameter value θ. In essence, then, it has been shown that system identification is equivalent to multihypothesis testing, with a continuum or finite sequence of hypotheses, respectively, for continuous or finite discrete range of θ.

In addition to the integral expressions for $\hat{\theta}(t)$, $\hat{s}(t/t)$, and $\hat{\beta}(t)$, stochastic differential equations have been obtained describing the temporal evolution of $\hat{\beta}(t)$, $p_1(\theta/t)$ and $p(\theta/t)$. Moreover, for the special case of linear models stochastic differential equations are obtained for the temporal evolution of $\hat{\beta}(t)$, $p_1(\theta/t)$, $\hat{x}_1(t/t, \theta)$, $P_1(t/t, \theta)$ and $p(\theta/t)$. Finally, a simple, suboptimal nonlinear estimation algorithm is given, based on the relinearized Kalman filter.

Appendix

The proofs of Theorem 1 as well as of the associated corollaries are given below:

Proof of Theorem 1 (*Partition Theorem*). The proof is based on Bucy's representation theorem (Jazwinski, 1970), and the smoothing property of expectations. Specifically, the desired mean-square estimate of β and θ are, as is well known, their conditional means given ν_t. Thus, the a posteriori conditional densities $p(\beta/t)$ and $p(\theta/t)$ must be obtained. These are easily obtained from the joint density $p(\theta, \beta/t)$. Using Bayes's rule, the latter is given by

$$p(\theta, \beta/t) = \frac{p[x(t), \theta, \beta/t]}{p[x(t)/t, \theta, \beta]}. \tag{I-1}$$

Applying Bucy's representation theorem to Eqs. (6), we obtain the a posteriori density of the augmented state vector $x_a(t) = [x^T(t) \vdots \theta^T \vdots \beta]^T$, viz.,

$$\begin{aligned} p[x_a(t)/\nu_t] &\equiv p[x(t), \theta, \beta/t], \\ &= \frac{E^{\nu_t}[\exp \Gamma(t)/x(t), \theta, \beta]\, p[x(t), \theta, \beta]}{E^{\nu_t}[\exp \Gamma(t)]}, \end{aligned} \tag{I-2}$$

where $E^{\nu_t}[\cdot]$ stands for the operation of expectation holding ν_t constant at the measured values, and $\Gamma(t)$ is defined as

$$\Gamma(t) \equiv \int_{t_0}^{t} h_a{}^T(x_a(\sigma), \sigma)\, R^{-1}(\sigma)\, z(\sigma)\, d\sigma - \tfrac{1}{2} \int_{t_0}^{t} \| h_a(x_a)\sigma), \sigma)\|^2_{R^{-1}(\sigma)}\, d\sigma \tag{I-3}$$

If we apply the representation theorem once more to Eqs. (I-2) for given (θ, β), we obtain the a posteriori density of the signal state-vector $x(t)$; namely, we obtain

$$\begin{aligned} p[x(t)/\nu_t, \theta, \beta] &\equiv p[x(t)/t, \theta, \beta] \\ &= \frac{E^{\nu_t}[\exp \Gamma(t)/x(t), \theta, \beta]\, p[x(t)/\theta, \beta]}{E[\exp \Gamma(t)/\theta, \beta]}. \end{aligned} \tag{I-4}$$

Combining Eqs. (I-1, I-2, and I-4) yields

$$\begin{aligned} p(\theta, \beta/t) &= \frac{E^{\nu_t}[\exp \Gamma(t)/\theta, \beta]}{E^{\nu_t}[\exp \Gamma(t)]}\, p(\theta, \beta) \\ &= \frac{E^{\nu_t}[\exp \Gamma(t)/\theta, \beta]}{\int E^{\nu_t}[\exp \Gamma(t)/\theta, \beta]\, p(\theta, \beta)\, d\theta\, d\beta}\, p(\theta, \beta). \end{aligned} \tag{I-5}$$

Note that in view of the independence of θ and β, we have $p(\theta, \beta) = p(\theta)\, p(\beta)$, and Eq. (I-5) may be written as

$$p(\theta, \beta/t) = \frac{E^{\nu_t}[\exp \Gamma(t)/\theta, \beta]}{\int E^{\nu_t}[\exp \Gamma(t)/\theta, \beta]\, p(\theta)\, p(\beta)\, d\theta\, d\beta}\, p(\theta)\, p(\beta)$$
$$= \frac{E^{\nu_t}[\exp \Gamma(t)/\theta, \beta]}{\left\{p_0 \int E^{\nu_t}[\exp \Gamma(t)/\theta, \beta = 0]\, p(\theta)\, d\theta + p_1 \int E^{\nu_t}[\exp \Gamma(t)/\theta, \beta = 1]\, p(\theta)\, d\theta\right\}}\, p(\theta)\, p(\beta), \tag{I-6}$$

where the fact that $p(\beta) = p_0\delta(\beta) + p_1\delta(\beta - 1)$, was used in obtaining the latter equality.

Duncan (1967) established that the denominator of the representation theorem is a likelihood function and proved that

$$E^{\nu_t}[\exp \Gamma(t)/\theta, \beta] = \exp\left\{\int_{t_0}^{t} \beta \hat{h}^T(x(\sigma), \sigma; \theta/\nu_0, \theta, \beta)\, R^{-1}(\sigma)\, z(\sigma)\, d\sigma - \tfrac{1}{2}\int_{t_0}^{t} \| \beta \hat{h}^T(x(\sigma), \sigma; \theta/\nu_\sigma, \theta, \beta)\|^2_{R^{-1}(\sigma)}\, d\sigma\right\}, \tag{I-7}$$

where $\hat{h}(x(\sigma), \sigma; \theta/\nu_\sigma, \theta, \beta) \equiv E[h(x(\sigma), \sigma; \theta)/\nu_\sigma, \theta, \beta]$.

Moreover, we note that for $\beta = 0$, $E^{\nu_t}[\exp \Gamma(t)/\theta, \beta = 0] = 1$, while for $\beta = 1$, $E^{\nu_t}[\exp \Gamma(t)/\theta, \beta = 1]$, denoted $\Lambda(t/\theta)$, is given by

$$\Lambda(t/\theta) = \exp\left\{\int_{t_0}^{t} \hat{h}^T(x(\sigma/\sigma, \theta)\, R^{-1}(\sigma)\, z(\sigma)\, d\sigma - \tfrac{1}{2}\int_{t_0}^{t} \| \hat{h}_1(\sigma/\sigma, \theta)\|^2_{R^{-1}(\sigma)}\, d\sigma\right\}, \tag{I-8}$$

where $\hat{h}_1(\sigma/\sigma, \theta) \equiv \hat{h}(x(\sigma), \sigma/\nu_\sigma, \theta, \beta = 1)$.

In view of the above, Eq. (I-6), takes the following form

$$p(\theta, \beta/t) = \frac{\Lambda(t/\theta)}{p_0 + p_1 \int \Lambda(t/\theta)\, p(\theta)\, d\theta}, \tag{I-9}$$

where $p(\beta) = p_0\delta(\beta) + p_1\delta(\beta - 1)$.

The a posteriori probabilities $p(\theta/t)$ and $p(\beta = 1/t)$ are obtained in a straightforward fashion from $p(\theta, \beta/t)$. Namely,

$$p(\theta/t) = \int p(\theta, \beta/t)\, d\beta = \frac{1 + \rho\Lambda(t/\theta)}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}\, p(\theta), \tag{I-10}$$

where $\rho \equiv p_1/p_0$, and

$$\hat{\beta}(t) = p(\beta = 1/t) = \frac{\rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}. \tag{I-11}$$

Using Eq. (I-10), we obtain $\hat{\theta}(t)$ as follows:

$$\begin{aligned} \hat{\theta}(t) &\equiv \int \theta p(\theta/t)\, d\theta \\ &= \frac{\int \theta p(\theta)\, d\theta + \rho \int \theta \Lambda(t/\theta)\, p(\theta)\, d\theta}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta} \\ &= \hat{\theta}(t_0) \frac{1}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta} + \frac{\rho \int \theta \Lambda(t/\theta)\, p(\theta)\, d\theta}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta} \\ &= \hat{\theta}(t_0)[1 - \hat{\beta}(t)] + \frac{\rho \int \theta \Lambda(t/\theta)\, p(\theta)\, d\theta}{1 + \rho \int \Lambda(t/\theta)\, p(\theta)\, d\theta}. \end{aligned} \tag{I-12}$$

Moreover, by using the smoothing property of expectations, we have

$$\begin{aligned} \hat{s}(t/t) &= E\{E[s(t)/\nu_t, \beta = i]/\nu_t\} \\ &= \sum_{i=0}^{1} \hat{s}_i(t/t)\, p(\beta = i/t) \\ &= p(\beta = 1/t)\, \hat{s}_1(t/t) \\ &= \hat{\beta}(t)\, \hat{s}_1(t/t), \end{aligned} \tag{I-13}$$

where $\hat{s}_i(t/t) = E[s(t)/\nu_t, \beta = i]$, $i = 0, 1$, and $\hat{s}_0(t/t) = 0$ since $s(t) = \beta y(t)$. In view of the latter, $\hat{s}_1(t/t) = \hat{y}_1(t/t) = \hat{h}(t/t)$. This completes the proof of Theorem 1.

Proof of Corollary I. From Eq. (I-13) above, we have

$$\hat{s}(t/t) = \hat{\beta}(t)\, \hat{h}_1(t/t).$$

Moreover, from the definition of $P_s(t/t)$ and the smoothing property of expectations we have

$$\begin{aligned} P_s(t/t) &= \sum_{i=0}^{1} P_s(t/t, \beta = i)\, p(\beta = i/t) \\ &= P_s(t/t, \beta = 1)\, p(\beta = 1/t) \\ &= \hat{\beta}(t)\, P_{h_1}(t/t), \end{aligned} \tag{I-14}$$

the latter two equalities resulting because, under hypothesis H_1, $s(t) = h(t)$, and under hypothesis H_0, $s(t) = 0$, and $\hat{s}_0(t/t) = 0$, as well as $E[s(t)\, s^T(t)/t, \beta = 0] = 0$.

Again using the smoothing property, we have

$$\begin{aligned} \hat{h}_1(t/t) &= E\{E[h(t)/\nu_t, \theta, \beta = 1]/\nu_t, \beta = 1\} \\ &= \int \hat{h}_1(t/t, \theta)\, p_1(\theta/t)\, d\theta, \end{aligned} \tag{I-15}$$

where $\hat{h}_1(t/t, \theta) \equiv E[h(t)/\nu_t, \theta, \beta = 1]$ and $p_1(\theta/t) \equiv p(\theta/\nu_t, \beta = 1)$ is obtained in a straightforward fashion by $p_1(\theta/t) = p(\theta, \beta = 1/t)/p(\beta = 1/t)$, and given by

$$p_1(\theta/t) = \frac{\Lambda(t/\theta)}{\int \Lambda(t/\theta)\, p(\theta)\, d\theta}\, p(\theta).$$

The conditional error-covariance matrix $P_{\hat{h}_1}(t/t)$ is obtained as follows:

$$\begin{aligned} P_{\hat{h}_1}(t/t) &\equiv E\{[h(x(t), t; \theta) - \hat{h}_1(t/t)][h(x(t), t; \theta) - \hat{h}_1(t/t)]^T/t, \beta = 1\} \\ &= E[E\{[h(x(t), t; \theta) - \hat{h}_1(t/t)][h(x(t), t; \theta) \\ &\quad - \hat{h}_1(t/t)]^T/t, \theta, \beta = 1\}/t, \beta = 1], \end{aligned} \tag{I-16}$$

where

$$\begin{aligned} &E\{[h(x(t), t; \theta) - \hat{h}_1(t/t)][h(x(t), t; \theta) - \hat{h}_1(t/t)]^T/t, \theta, \beta = 1\} \\ &\quad = E[h(x(t), t; \theta)\, h^T(x(t), t; \theta)/t, \theta, \beta = 1] + \hat{h}_1(t/t)\, \hat{h}_1^T(t/t) \\ &\qquad - \hat{h}_1(t/t)\, \hat{h}_1^T(t/t, \theta) - \hat{h}_1(t/t, \theta)\, \hat{h}_1^T(t/t) \\ &\quad = E\{[h(x(t), t; \theta) - \hat{h}_1(t/t, \theta)][h(x(t), t; \theta) - \hat{h}_1(t/t, \theta)]^T/t, \theta, \beta = 1\} \\ &\qquad + [\hat{h}_1(t/t, \theta) - \hat{h}_1(t/t)][\hat{h}_1(t/t, \theta) - \hat{h}_1(t/t)]^T \\ &\quad = P_{\hat{h}_1}(t/t, \theta) + [\hat{h}_1(t/t, \theta) - \hat{h}_1(t/t)][\hat{h}_1(t/t, \theta) - \hat{h}_1(t/t)]^T. \end{aligned} \tag{I-17}$$

Combining Eqs. (I-16) and (I-17), we obtain Eq. (12). This completes the proof of Corollary I.

RECEIVED: November 2, 1970; REVISED: February 11, 1971

References

Duncan, E. E. "Probability Densities for Diffusion Processes," Technical Report 7001-4, Stanford University, Stanford, Calif. 1967.

Esposito, R. (1969), private communication.

Jazwinski, A. H. (1970), "Stochastic Processes and Filtering Theory," Academic Press, New York.

Kailath, T. (1969), A general likelihood-ratio formula for random signals in Gaussian noise, *IEEE Trans. Information Theory* **IT-15**, 350–361.

Lainiotis, D. G. (1969), On a general relationship between estimation, detection and the bhattacharya coefficient, *IEEE Trans. Information Theory* **IT-15**, 504–505.

Licht, B. W. (1970), "Approximations in Optimal Nonlinear Filtering," Ph.D. Thesis, Case-Western Reserve University, ————, 1970.

Middleton, D., and Esposito, E. (1968). Simultaneous optimum detection and estimation of signals in noise, *IEEE Trans. Information Theory* **IT-14**, 434–444.

Printed in Belgium