SIMULTANEOUS DETECTION AND ESTIMATION FOR DIFFUSION TYPE SIGNALS

by

Kyu San Han

Dissertation submitted to the Faculty of the Gradute School

of the University of Maryland in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

1986

ABSTRACT

Title of Dissertation : Simultaneous Detection and Estimation for Diffusion Type Signals

Kyu San Han, Doctor of Philosophy, 1986

Dissertation Directed by : Dr. John S. Baras
Professor
Department of Electrical Engineering

We consider a problem of combined detection and estimation when signals corresponding to M-ary hypotheses can be represented as outputs of M distinct parameterized stochastic dynamical systems of the Ito type. For a general synthesis of detector and estimator we take a jointly Bayesian approach to the combined detection and estimation under a suitable class of joint cost functions that combine a cost of misdetection with a cost of estimation error, and then utilizing the nonlinear filtering theory we derive subsequent jointly Bayesian receiver structures.

Secondly, based on the Chernoff bounds we analyze detection performances for both full and partial observation of signals generated by several kinds of finite dimensional nonlinear stochastic dynamical systems. Also performance evaluation in the parameter estimation is treated from a detection point of view.

ACKNOWLEDGEMENTS

I wish to express my thanks to Dr. John S. Baras for supporting my graduate studies at the University of Maryland. I also want to say thanks to the faculty members in the department of the Electrical Engineering for their valuable guidance. In addition, appreciation must be expressed for the support and aid of my fellow graduate students.

TABLE OF CONTENTS

1. Introduction

In the first part of the thesis we consider a problem of combined detection and estimation when the signals corresponding to M different hypotheses can be modeled as outputs of M distinct parameterized stochastic dynamical systems of the Ito type. For a general synthesis of detector and estimator we take a jointly Bayesian approach to the combined detection and estimation under a suitable class of joint cost functions which combine cost of misdetection with a cost of estimation error [1] [2] .

The jointly Bayesian approach to the problem of combined detection and estimation was apparently first suggested in the work of Nilsson [3] where radar range resolution in a multi-target environment was treated. And its theoretical formulation in the language of statistical decision theory has been developed by Middleton and Esposito [1]. Although from the mathematical point of view estimation may be regarded as a generalized detection problem, from an operational point of view the two procedures are different: one usually selects different cost functions for each and obtains a different processors as a result. It is then correctly argued in [1] that it is practically appropriate to retain the usual distinction between detection and estimation. They point out that both detection and estimation are being performed under each other's uncertainty, and therefore propose a jointly Bayesian approach leading to a specific mutual coupling between the resulting detector and estimator. In [1] there are various ways that the detector and estimator can be coupled, corresponding to a hierarchy of complex processors. It is obvious that the selected shape of joint cost functions will have a substantial effect on the structure of such a system. Nevertheless systems synthesized have some common features which are inherent of such multifunctional systems and which distinguishes them from the systems that relate the operations of detection and estimation based on intuition or reasoning, for instance, decision based on the maximum likelihood estimates.

In [1],[2] and [4] it is shown that under reasonable conditions the jointly optimal rule of detection and estimation belongs also to the class of nonrandomized statistical decisions and that joint optimization can be replaced by a sequential optimization, where the jointly optimal rule is based on some preceding set of estimates. The fact that jointly optimal rule can be based upon the set of some preceding estimates leads to an interesting implication that for binary hypotheses problems decision is based on the threshold, which partially depends on the observation through the proper a posteriori loss related to an estimation error.

Several other authors have treated the combined detection and estimation problems. Jaffer and Gupta[5] , and Sosulin[6] consider the recursive filtering of Gauss Markov process and diffusion type signals, respectively , in the presence of hypotheses uncertainty under a combined cost function (quadratic cost function for the estimation part). Lainiotis[7] extended the results of [1] to the case of continuous observation on a single interval and to the problem of system identification. In this reference the problem of synthesis of jointly optimal detection, filtering, and identification is formulated as the nonlinear problem of estimation of an augmented composite state vector.

Also Birdsall and Gobien's work is worthy of mentioning [10]. In [10] detectors are the usual generalized likelihood ratio test for composite binary hypotheses. The authors treat the case where the a priori probability densities of parameters under both hypotheses belong to the reproducing classes with respect to the observation models which allow the finite dimensional sufficient statistics for the parameter. In this case the a posteriori densities of parameters, as a whole function, can be identified by use of the recursively updated sufficient statistic, and also no integration of conditional likelihood ratios is necessary to obtain the usual generalized likelihood ratio. However their result is applicable only to certain classes of a priori

densities and observation models. The important result of [10] shows that through a Bayesian approach(i.e. here, by using the a posteriori densities of parameters) both detection and estimation occur simultaneously, with the detector using the a posteriori densities generated by two separate estimators, one for each hypothesis.
It is the objective of chapter 2 to consider the simultaneous detection and estimation and the related previous results in a general perspective by employing a continuous time dynamical signal models, and thereby to incorporate a sequential nature into the jointly optimal algorithm [9].

Major criticisms against the jointly Bayesian approach are the facts that complete knowledge of a priori statistics and criterion for the choice of joint cost functions are required , and also that the resulting receiver structures are complex.
It is known that for a class of cost functions and a class of a priori distributions on parameters to be estimated, the risks associated with the Bayes and maximum likelihood estimates of parameters become equal as the number of independent observations becomes large[11]. This fact may provide the motivation for an asymptotically optimum jointly Bayesian receiver. Whether such asymptotic properties can be retained in our observation model or not is beyond the scope of proposed research. Instead we will seek the possible suboptimal structures based on approximation of the jointly Bayesian system.

The second research objective is performance evaluation of the jointly optimal rule. However evaluation of a Bayes cost is not feasible to carry out due to complexities of our observation model, and also due to a mutual coupling between a jointly optimal detector and estimator. Moreover the jointly Bayesian system seems to possess more significance when the a posteriori accuracy becomes low, and to be asymptotically equivalent to other systems. This necessitates performance analysis for the finite observation time or under low signal to noise ratios. Due to the above difficulties we will instead focus on the performance bounds in simple binary hypotheses detection, and also on the performance bounds in estimation under no hypothesis uncertainty.

Evans[12] considers the case when the observations are sample continuous processes, the signals are nongaussian diffusion or finite state Markov process ,and noises are Brownian motions. Through the measure transformation[8] technique he developed an expression for the exponent in the Chernoff bound under a new measure. Then by introducing a quasi-transition function, he derives a Fokker-Planck type equation for the time evolution of a quantity closely related to the Chernoff-bound exponent. Hibey, Snyder and van Schuppen[13] extended the quasi-transition function method to a discontinuous observation that contains a rate process associated with a counting process. Hibey points out in [14] that the usual conditions for the justification of measure transformation are difficult to verify in some practical cases.

Performance bounds in parameter estimation problems will be treated from a detection point of view by discretizing the parameter space, and by adapting the result for a binary hypotheses detection problem. Past work related to this approach includes that of Liporace[15] who treated the case of parameterized independent and identical observations and also works of Hawkes and Moore[16] who extended the convergence result of [15] to the case of dependent observations characterized by a parametrized linear discrete gaussian systems.

2. Simultaneous detection and estimation for diffusion model

2.1) Description of observation model

In this section we present an observation model when signals of interest can be represented by outputs of diffusion type processes. The statistics of y($\cdot$) are not completely known. More specifically they depend on some parameters and hypotheses. We shall denote the ith hypothesis by H_i for i = 0,...,M. Under each hypothesis H_i , the received data can be represented as

$$dx^i(t) = f^i(x^i(t),\theta^i)dt + g^i(x^i(t),\theta^i)dw^i(t) \qquad ,x^i(0) = x_0^i$$
$$dy(t) = h^i(t,x^i(t),\theta^i)dt + \sqrt{r}\,dv^i(t) \qquad 0 \le t \le T,\ y(0) = 0 \tag{1}$$

1) Null hypothesis H_0 is fixed by an observation model, $dy(t) = \sqrt{r}\ dv^0(t)$. By simply adding a common diffusion model, we can consider also a common diffusion interference present under all hypotheses. Formal time derivative of $v^i(t)$ is ,as usual, a standard gaussian white noise. Time dependence of $h^i(t,x^i(t),\theta^i)$ is explicitly made to account for deterministic modulations frequently appearing in communication literatures. For example $h^i(\cdot)$ can be thought of as a return from a point target whose fluctuating reflectivity causes a doppler-spread effect on the transmitted signal with a certain modulation.

2) For clarity we specify (Ω ,F_T ,P^i) associated with each hypothesis H_i. We construct a sample space Ω as follows.

$$\Omega = (\prod_{j=1}^{M} \Omega^j) \times \Omega_3 = (\prod_{j=1}^{M} \Omega_0^j \times \Omega_1^j \times \Omega_2^j) \times \Omega_3$$

, where Ω_0^j, Ω_1^j are R^{Lj} and R^{Nj} respectively, Ω_2^j= (C[0,T] ;R^{Dj}) ,and Ω_3 is (y $\in$ C[0,T]; R^K , y(0) =0). The element ω=(θ^j, x_0^j ,w^j, y, j =1,,,,M) of Ω satisfy

$$\omega = (\theta^j(\omega), x_0^j(\omega), w^j(t,\omega), y(t,\omega),\ \ 0 \le t \le T,\ \ j = 1,,,M) \tag{2}$$

Let $F(\theta^j)$ be a θ^j associated Borel σ-field on R^{Lj} with $F(x_0^j)$ defined similarly. Let $F_t(w^j)=\sigma(w^j(s)\ ,\ 0\le s\le t)$ and $F_t(y)=\sigma(y(s)\ ,\ 0\le s\le t)$ for $t\in[0,T]$. Let $G_t^j=F(\theta^j)\times F(x_0^j)\times F_t(w^j)$ and $F_t=\sigma(\prod_{j=1}^{M}G_t^j)\times F_t(y)$. The elements of each σ-fields are subset of Ω_0^j, Ω_1^j, Ω_2^j, Ω_3 respectively. However one can regard them also as a subset of Ω with the obvious identifications. For example $A\in F_t(w^i)$ can be regarded with $\prod_{j=1}^{M}\Omega_0^j\times\prod_{j=1}^{M}\Omega_1^j\times A\times\prod_{j\ne i}^{M}\Omega_2^j\times\Omega_3$. Following the latter interpretation we will treat each σ-field as a sub σ-field of F_T.

We have a probability measure P^i defined on F_T satisfying that

(a) θ^i is a parameter random variable having a density $p^i(\theta)$ on R^{Li}. We denote its parameter space by Θ_i to indicate a nature of vector space R^{Li}.

(b) $E^i(x_0^i)^2<\infty$ where E^i denotes an expectation under the probability measure P^i. For most cases of practical interests the density of x_0^i will be a stationary density that a given diffusion model yields at each (θ^i,H_i).

(c) $w^i(t)$ is a standard Wiener process taking values in R^{Di} under ($\sigma(G_t^i\bigcup F_t(y))$, P^i).

(d) $v^i(t)=1/\sqrt{r}\left\{y(t)-\int_0^t h^i(s,x^i(s),\theta^i)ds\right\}$ is a standard Wiener process taking values in R^K under ($\sigma(G_t^i\bigcup F_t(y))$, P^i), where $x^i(t)$ is assumed to be an unique strong solution to the stochastic differential equation of (1) at each $\theta^i\in\Theta_i$. Thus we treat $v^i(t)$ as an observation derived Wiener martingale.

(e) $(\theta^i,x_0^i),w^i(\cdot),v^i(\cdot)$ are independent.

(f) except the above conditions P^i can be arbitrary on F_T. Further details about an assignment of P^i on F_T are irrelevant to both observation models and hypotheses detection parameter estimation problem. For convenience we assume that restriction of P^i to $\sigma(\bigcup_{j=0}^{M}G_T^j)$ are mutually absolutely continuous with respect to that of P^j for all i,j (; we have the identical set of coordinate functions specified by (2) under all hypotheses. However θ^i ,as a random variable , under P^i and under P^j are different random variables having different distributions as $y(\cdot)$ does. Which θ^i , as a random variable , is meant by can be made clear by an associated probability measure or hypothesis) . When certain components of θ^i are of physically common nature with some components of θ^j , there are redundancies in our choice of coordinate functions . We take the present choice to uniformly set the coordinate functions regardless of their detailed natures.

3) f^i ; $R^{Ni} \times R^{Li} \rightarrow R^{Ni}$, g^i ; $R^{Ni} \times R^{Li} \rightarrow R^{Ni \times Di}$, h^i ; $[0,T] \times R^{Ni} \times R^{Li} \rightarrow R^K$ are jointly measurable functions. We assume that $f^i(\cdot,\theta^i), g^i(\cdot,\theta^i)$ are such that an unique Markov solution exists for $x^i(t)$ at each θ^i with $x^i(t)$ being $\sigma(F(x_0^i) \bigcup F_t(w^i)$ measurable and such that for all $t \in [0,T]$, $E^i(|f^i(x^i(t),\theta^i)|^2 \mid \theta^i) < \infty$, $E^i(||g^i(x^i(t),\theta^i)||^2 \mid \theta^i) < \infty$ with $a(x^i,\theta^i) \equiv g^i(x^i,\theta^i) g^i(x^i,\theta^i)^T \geq b \cdot I$ for some positive b , and $E^i(|h^i(t,x^i(t),\theta^i)|^2 \mid \theta^i) < \infty$.

These set of assumptions provide sufficient conditions , based upon which $x^i(t)$ is a semimartingale ar each $\theta^i \in \Theta_i$, and nonlinear filtering theory can be used [17].

Also $h^i(t,x^i,\theta^i)$ is twice continuously differentiable in x^i and once in t at each θ^i .

4) Let $P_Y^{i,\theta}$ be a restriction of probability measure P^i to $F_T(y)$, conditioned on θ^i . The condition that $E^i(\mid h^i(x^i(t),\theta^i)\mid^2 \mid \theta^i) < \infty$ for all $t \in [0,T]$ implies $\int_0^T |h^i(x^i(t),\theta^i)|^2 dt < \infty$ almost surely at each θ^i , and the latter condition implies equivalence between $P_Y^{i,\theta}$ and $P_Y^{j,\theta'}$ where $\theta \in \Theta_i, \theta' \in \Theta_j$ [8].

Since $P^i(B) = \int_{\Theta_i} P_Y^{i,\theta}(B)\, p^i(\theta)\, d\theta$ for $B \in F_T(y)$, restriction of P^i to $F_T(y)$, denoted as P_Y^i , is equivalent to that of P_Y^j . And from our construction of (Ω, F_T, P^i) P^i and P^j are equivalent on F_T . Henceforth we assume that F_t is complete for all $t \in [0,T]$.

Also we assume that there exists an unique conditional probability densities of $(x^i(t),\theta^i)$,given $F_t(y)$ for all $t \in [0,T]$ under each H_i .

5) We assume an existence of unique solution to Zakai equation associated with each (θ^i, H_i) [18] [19] .

2.2) Simultaneous detection and estimation

2.2.1) Problem formulation

Now we formulate the problem of simultaneous detection and estimation. Given an observation $(y(s), 0 \le s \le t,\ t \le T)$, we want to find a jointly optimal rule minimizing an average risk of the form given in (3) at each time t. Thus we restrict the class of admissible rules to a class of nonanticipative rules.

$$R = E\ (\ C\ (\ u\ ,\theta\ ;\ u_d\ ,\theta_e\ )\ ) \qquad (3)$$

u is a hypothesis index random variable taking a value i with a probability $p(i)$, and θ a parameter random variable taking values on Θ_i with the previous density $p^i(\theta)$ when $u = i$. If all the θ^i s are of physically same nature then one may use one coordinate function θ instead of the set given by $(\theta^1 ,,, \theta^M)$,and in this case one only needs a joint distribution of $(u\ ,\theta)$ to identify the random variable θ . In our observation model θ^i may have a different nature depending on i ,and thus a value space of θ depends on u . We will explain a nature of parameter space Θ_0 later. u_d and θ_e are random variables representing a detector and an estimator respectively. For consistency an estimator random variable θ_e takes values on Θ_i when $u_d = i$. Value space for θ_e is same as that of θ . We will later augment θ by the derived parameter random variables such as an envelope of signal process with the corresponding change in θ_e also.

In (3) $C(u\ ,\theta\ ;\ u_d\ ,\ \theta_e\ )$ is a joint cost function assumed to be known , and represents a combined cost of presenting θ_e with a classification of $\theta_e \in \Theta_{u_d}$ when θ actually comes from Θ_u .

It is obvious that the selected shapes of the joint cost functions will have a substantial effect on the jointly optimal decision and estimation systems. Nevertheless, systems synthesized for a different cost functions have a some common features which are inherent of such multifunctional systems and which distinguish them from systems that realize the operations of detection and estimation based on intuition or reasoning ,for instance, decision directed estimation or estimation directed decision.

Now we make precise the definition of (3). At present we will not make any prior assumption on the detailed nature of joint cost assignment except that it is legitimately defined. Since the value spaces of θ and θ_e depend on u and u_d respectively, it is not feasible to justify them as random variables in an usual way. Therefore we temporarily restrict the jointly Bayesian approach to a set of parameters whose nature can be defined independently of hypotheses. Later we will extrapolate the resulting expression for the average risk R to our general case. Let

$$\Theta_i = \overline{\Theta} \times \tilde{\Theta}_i$$

The parameter space $\overline{\Theta}$ is of same physical nature independently of hypotheses. $\tilde{\Theta}_i'$ s are of different nature.

Accordingly

$$\theta^i = (\overline{\theta}^i , \tilde{\theta}^i)$$

Define $\overline{\theta}$ by

$$\overline{\theta} = \sum_{i=0}^{M} 1_{(u=i)} \times \overline{\theta}^i$$

where $1_{(u=i)}$ is an indicator function which is 1 when $i = u$, and otherwise 0 . Then the unconditional density of $\overline{\theta}$ is given by

$$\sum_{i=0}^{M} p^i(\overline{\theta})\, p(i)$$

where $p^i(\overline{\theta}) = \int_{\tilde{\Theta}_i} p^i(\theta)\, d\tilde{\theta}^i$ with $p^i(\theta)$ being defined in section 2.1 . The estimator for the random variable $\overline{\theta}$ will be denoted by $\overline{\theta}_e$.

In the sequel we make some account on the common parameter space. When a null hypothesis , $dy(t) = \sqrt{r}\ dv^0(t)$, is included in a test , only an energy or nonnuisance type parameter can be a component of $\overline{\Theta}$. For other types of parameters can't be physically defined under the null hypothesis. As examples of energy type parameters mean value of a stationary process , scale factor, or duration of a signal if uncertain within an observation interval are usually considered [1,2]. Envelope of a signal process can be considered as an energy type parameter by treating $h_t\,(\equiv \sum_{i=0}^{M} 1_{(*=i)}\, h^i(t, x^i(t), \theta^i)\,)$ as a parameter to estimate. And in this case of envelope estimation under the uncertainty of compound hypotheses, we make a decision, filtering, and estimation continuously, based on the observation up to current time t [5, 6, 7]. At this point we augment $\overline{\theta}$ and $\overline{\theta}_e$ by $[\overline{\theta}^T, h_t{}^T]^T$, $[\overline{\theta}_e{}^T, h_e{}^T]^T$ respectively, where T denotes a transpose of a vector.
Henceforth we denote $[\overline{\theta}^T , h_t^T]^T$ and $[\overline{\theta}_e^T , h_e^T]$ as $\overline{\theta}$ and $\overline{\theta}_e$ respectively. Also $\overline{\Theta}$ denotes either a value space of $\overline{\theta}$ or a value space of the augmented parameter vector $\overline{\theta}$, depending on the context.

When a common interference occurs, a whole set of interference parameters can be components of the common parameters. If no null hypothesis is present, more flexibility in a choice of $\overline{\Theta}$ can be given depending on a particular problem situation.

Parameter space under H_0 is set by $\overline{\Theta}$ and a random variable θ^0 has a density given by $\delta(\theta)$. This treatment is in agreement with a Neymann-Pearson type compound binary hypotheses detection problem , where 0 value of a parameter characterizes a probability distribution of an observation under H_0 .

Accordingly we modify $\prod_{j=1}^{M} \Omega_0^j$ of section (2.1) by $\prod_{j=0}^{M} \Omega_0^j$ with a corresponding change in the definition of F_t , where $\Omega_0^0 = \overline{\Theta}$.

Now we redefine (3) temporarily as

$$R = E(\, C(\, u ,\overline{\theta} ;\, u_d ,\overline{\theta}_e \,)\,) \tag{4}$$

Since $\overline{\theta}$ and $\overline{\theta}_e$ are defined by their characters, $C(\, u ,\overline{\theta} : u_d ,\overline{\theta}_e \,)$ gives a flexibility to a resulting optimal estimate $\overline{\theta}_e^*$ in the sense that it will be an estimate taking into account of hypotheses uncertainty. We define a probability space $(\tilde{\Omega},\tilde{F}_T ,P)$, associated with (4) as follows. Let $\tilde{\Omega}$ be

$$\tilde{\Omega} = U \times \Omega \times S \tag{5}$$

,where $U' = \left\{0,1,,,M\right\}$ and $S = U \times \overline{\Theta}$.

The element $\tilde{\omega} = (u ,\theta^0,\theta^j ,x_0^j ,w^j ,y ,u_d , \overline{\theta}_e ,j = 1,,,M)$ satisfy

$$\tilde{\omega} = (u\,(\tilde{\omega}),\theta^0(\tilde{\omega}),\theta^j\,(\tilde{\omega}),x_0^j\,(\tilde{\omega}),w^j\,(t\,,\tilde{\omega}),y\,(t\,,\tilde{\omega}),u_d\,(\tilde{\omega}),\, \overline{\theta}_e\,(\tilde{\omega}),\;\; 0 \le t \le T\,,\;\; j = 1,,,M)$$

We define a corresponding σ -field $\tilde{F}_t$ as

$$\tilde{F}_t = F(u) \times F_t \times F(S)$$

As before we treat each σ field as a sub σ -field of $\tilde{F}_T$.

Probability measure P on $\tilde{F}_T$ is stated in the sequel. Distribution of u and that of θ^0 are already specified . Given $u = i$, P on F_T coincides with the previous probability measure P^i and thus the restriction of P to $F_T(y)$ is given by $\sum_{i=0}^{M} p\,(i) \times P_Y^i(B)$ for $B \in F_T(y)$. Assignment of P is completed if we specify a distribution of random variables u_d , $\overline{\theta}_e$. u_d and $\overline{\theta}_e$ are, by their natures , conditionally independent of other random variables or processes given $F_t(y)$, and thus specified by the joint conditional probabilities , $P(u_d = i\,, \overline{\theta}_e \le \overline{\theta}' \mid B)$, defined for all B where $B \in F_t(y)$ and $\overline{\theta}' \in \overline{\Theta}$. We note that u_d and $\overline{\theta}_e$ are a given pair of nonanticipative random variables whose statistics depends on y^t . An average risk R depends on a conditional statistics of (u_d , $\overline{\theta}_e$) , which represents a randomized decision and estimation rule in general, and as such, will be denoted by R (P_D ,P_E) .

2.2.2) Jointly Bayesian solution

A conditional statistics of a pair of random variable , $(u_d, \overline{\theta}_e)$, is specified by a following pair of conditional probabilities , $[P_D(u_d = j \mid y^t), P_{Ej}(\overline{\theta}_e \mid y^t)]$ defined for all y^t , where $\overline{\theta}_e \in \overline{\Theta}$, and $P_D(u_d = j \mid y^t)$ is the conditional probability of accepting the hypothesis H_j for a given observation y^t , and $P_{Ej}(\overline{\theta}_e \mid y^t)$ the probability distribution function of estimator random variable $\overline{\theta}_e$ for the given observation y^t on the condition that the hypothesis H_j has been accepted. For an arbitrary observation y^t we must clearly satisfy

$$\int_{\overline{\Theta}} dP_{Ej}(\overline{\theta}_e \mid y^t) = 1 \quad ,0 \le P_{Ej}(\overline{\theta}_e \mid y^t) \le 1 \quad j = 0,,M$$

$$\sum_{k=0}^{M} P_D(u_d = k \mid y^t) = 1 \quad ,0 \le P_D(u_d = k \mid y^t) \le 1 \tag{6}$$

Thus a given estimation rule , P_E , consists of $(P_{Ej})_{j=0}^{j=M}$. We denote , by a pair (P_D, P_E) , a given decision and estimation rule satisfying (6)

Let Ξ be the class of all randomized decision and estimation rules, $(P_D, P_E)'$ s. The jointly optimal decision and estimation rule , (P_D^*, P_E^*) , if exists , is given by

$$(P_D^*, P_E^*) = Arg\ (\underset{\Xi}{Min}\ R(P_D, P_E)) \tag{7}$$

Now we evaluate $R(P_D, P_E)$ by a standard procedure.

$$\begin{aligned} & R(P_D, P_E) \\ &= \sum_{i=0}^{M} E(\,C(u, \overline{\theta}\,; u_d, \overline{\theta}_e) \mid u = i)\ p(i) \\ &= \sum_{i} E(\,E(\,C(u, \overline{\theta}\,; u_d, \overline{\theta}_e) \mid y^t, u = i\ , u_d = j\,, \overline{\theta}_e = \overline{\theta}_e'\ ) \mid u = i\ ) \cdot p(i) \end{aligned} \tag{8}$$

,where $E(\,C(u, \overline{\theta}\,; u_d, \overline{\theta}_e) \mid y^t, u = i\,, u_d = j\,, \overline{\theta}_e = \overline{\theta}_e'\ )$ is defined to be $E(\ \cdots\ \mid \sigma\left\{ F_t(y), F(u), F(u_d), F(\overline{\theta}_e) \right\})|_{y^t, i, j, \overline{\theta}_e'}$.

Since u and $\overline{\theta}$ are conditionally independent of u_d and $\overline{\theta}_e$ given $F_t(y)$, and thus sample values of u_d and $\overline{\theta}_e$ only play a role of fixing the corresponding arguments in the cost functional and nothing more, it follows that

$$E(\, C(u\,,\overline{\theta}\,;\,u_d\,,\overline{\theta}_e\,)\mid\, y^t\,,u\,=i\,,u_d\,=j\,,\overline{\theta}_e\,=\overline{\theta}_e'\,)$$
$$=E(\, C(i\,,\overline{\theta}^i\,;\,j\,,\overline{\theta}_e'\,)\mid y^t\,,u\,=i\,) \tag{9}$$
$$=\int_{R^{L_i+N_i}} C(i\,,\overline{\theta}^i\,;j\,,\overline{\theta}_e'\,)\, p\,(t\,,i\,;x^i\,,\theta^i\,)\, dx^i\; d\,\theta^i$$

with an obvious abuse of notation , where $p\,(t\,,i\,:x^i\,,\theta^i\,)$ is the a posteriori density of $(x^i(t\,),\theta^i\,)$ given $y^t\,,H_i$. Substituting (9) into (8) yields,

$$R(\,P_D\,,P_E\,)$$
$$=\sum_i \int_{\Omega_3}\left\{\sum_{j=0}^{M} P_D(u_d\,=j\;\mid y^t\,)\int_{R^{L+K}}\Big(\int_{R^{L_i+N_i}} C(i\,,\overline{\theta}^i\,;\,j\,,\overline{\theta}_e'\,)\, p\,(t\,,i\,:x^i\,,\theta^i\,)dx^i\; d\,\theta^i\,\Big)dP_{Ej}(\overline{\theta}_e'\;\mid y^t\,)\right\}$$
$$\cdot\, dP_Y^i\; p\,(i\,)$$

where P_Y^i here stands for an induced probability measure on $C\,[0,T\,]$ by $y\,(\,\cdot\,)$ given $u\,=i$, and $L\,+K$ is a dimension of the augmented parameter vector $\overline{\theta}$.

Now we consistently extrapolate the expression of $R\,(P_D\,,P_E\,)$ given above to our general case.

$$R(\,P_D\,,P_E\,)$$
$$=\sum_i \int_{\Omega_3}\left\{\sum_{j=0}^{M} P_D(u_d\,=j\;\mid y^t\,)\int_{R^{L_j+K}}\Big(\int_{R^{L_i+N_i}} C(i\,,\theta^i\,;\,j\,,\theta_e^j)\, p\,(t\,,i\,:x^i\,,\theta^i\,)dx^i\; d\,\theta^i\,\Big)dP_{Ej}(\theta_e^j\mid y^t\,)\right\}$$
$$\cdot\, dP_Y^i\; p\,(i\,) \tag{10}$$

where $\theta^i\,=(\,\overline{\theta}^i\,,\tilde{\theta}^i\,)$, $\theta_e^j=(\,\overline{\theta}_e^j,\tilde{\theta}^j\,)$, and thus $\theta^i\in\Theta_i\;\;\theta_e^j\in\Theta_j$.

We now assume that (6) is accordingly redefined. Now we use the mutual absolute continuity between P^i and P^0 of section 2.1 on $F_T(y\,)$.

$$\int_{\Omega_3}\left\{\sum_{j=0}^{M} P_D(u_d\,=j\;\mid y^t\,)\int_{R^{L_j+K}}\Big(\int_{R^{L_i+N_i}} C(i\,,\theta^i\,;j\,,\theta_e^j)\, p\,(t\,,i\,:x^i\,,\theta^i\,)dx^i\; d\,\theta^i\,\Big)dP_{Ej}(\theta_e^j\;\mid y^t\,)\right\} dP_Y^i$$
$$=\int_{\Omega_3}\left\{\sum_{j=0}^{M}\quad \cdots\cdots\quad\right\} E^0\Big(\frac{dP^i}{dP^0}\;\mid\, y^t\,)\;\; dP_Y^0 \tag{11}$$

,where $E^0(\frac{dP^i}{dP^0}\mid F_t(y))$ is a likelihood ratio random variable at time t between H_i and H_0 , and equals to a Radon-Nikodym derivative between the restrictions of P^i and P^0 to $F_t(y)$ [20]. Let

$$\Phi_{P_{E_j}}(y^t)$$
$$= \int_{R^{L_j+K}} \left\{ \sum_{i=0}^{M} p(i) E^0(\frac{dP^i}{dP^0} \mid y^t) \int_{R^{L_i+N_i}} C(i,\theta^i ; j,\theta_e^j) p(t,i:x^i,\theta^i) dx^i \, d\theta^i \right\} dP_{Ej}(\theta_e^j \mid y^t) \tag{12}$$

Substituting (11) (12) into (10) gives

$$R(P_D, P_E)$$
$$= \int_{\Omega} \left\{ \sum_{j=0}^{M} P_D(u_d = j \mid y^t) \Phi_{P_{E_j}}(y^t) \right\} dP_Y^0 \tag{13}$$

$\Phi_{P_{E_j}}(y^t)$ is a kind of the a posteriori loss incurred under an estimation rule P_E, given $u_d = j$. In section (2.3) we will evaluate the integrand of $\Phi_{P_{E_j}}(y^t)$, at least in theory, by use of the nonlinear filtering formula. Here we focus on characterizing a jointly optimal decision and estimation rule. We extend the development in [4], where a case of discrete sequence of observations is treated in a nonsequential setting.

Due to the M-ary nature of hypotheses detection, there is always a nonrandomized jointly optimal decision rule, which can be shown as follows. For any arbitrary fixed estimation rule P_E, the optimum decision rule P_D^* has the following form.

$$P_D^*(u_d = k \mid y^t) = 1 \quad , \quad P_D^*(u_d = j \mid y^t) = 0 \quad \text{for all} j \neq k$$
$$\text{iff} \quad \Phi_{P_{Ek}}(y^t) \leq \Phi_{P_{E_j}}(y^t) \quad \text{for all} j \neq k \tag{14}$$

Therefore we can confine ourselves to the class of nonrandomized decision rules. Let us now consider an arbitrary fixed disjoint subdivision, $\left\{ \Gamma_j \right\}_{j=0}^{M}$, of Ω_3 characterized by

$$\bigcup_{j=0}^{M} \Gamma_j = \Omega_3, \quad \Gamma_j \bigcap \Gamma_k = \phi \text{ for } j \neq k.$$

An optimal estimation rule P_E^* under an arbitrary nonrandomized decision rule P_D corresponding to some $\left\{ \Gamma_j \right\}_{j=0}^{M}$,if exists, satisfies

$$R(P_D, P_E^*) = \underset{P_E}{Min} \sum_{j=0}^{M} \int_{\Gamma_j} \Phi_{P_{Ej}}(y^t) dP_Y^0 = \sum_{j=0}^{M} \underset{P_{Ej}}{Min} \int_{\Gamma_j} \Phi_{P_{Ej}}(y^t) dP_Y^0 \tag{15}$$

In (15) we implicitly assume that minimum exists in the class Ξ . If exists in Ξ , the minimum should occur at an estimation rule , P_E^* , satisfying that for $j = 0,,,M$

$$P_{Ej}^*(\theta_e^j \mid y^t) = Arg(\underset{P_{Ej}}{Min} \ \Phi_{P_{Ej}}(y^t)) \quad ,y \in \Gamma_j \ . \tag{16}$$

We note from (16) that $P_{Ej}^*(\theta_e^j \mid y^t)$ is functionally independent of subdivision $\left\{\Gamma_j\right\}_{j=0}^{M}$ given $y \in \Gamma_j$,and hence also of the given decision rule. Only the optimal estimation rule, P_E^* , as a whole, depends on the subdivision $\left\{\Gamma_j\right\}_{j=0}^{M}$.

Now we further assume that for all $j \in \left\{0,,,M\right\}$ there is $\left\{P_{Ej}^*\right\}_{j=0}^{M}$ satisfying

$$P_{Ej}^*(\theta_e^j \mid y^t) = Arg(\underset{P_{Ej}}{Min} \ \Phi_{P_{Ej}}(y^t)), \quad y \in \Omega_3 \tag{17}$$

with probability 1.

Since we have only $M+1$ choices of decisions at each y^t (through the restriction to the class of nonrandomized decision rules), and also due to the independence of $P_{Ej}^*(\theta_e^j \mid y^t)$ on the form of subdivision(i.e. (16)), the jointly optimal rule is specified by

$$\begin{cases} \text{For an arbitrary observation } y^t, \\ \text{select } \ (k \ ,P_{Ek}^*(\theta_e^k \mid y^t)) \ \text{ if and only if} \\ \Phi_{P_{Ek}^*}(y^t) \ \le \ \Phi_{P_{Ej}^*}(y^t) \quad \text{for some } k \text{ and all } j \ : \\ P_{Ej}^*(y^t) = Arg(\underset{P_{Ej}}{Min} \quad \Phi_{P_{Ej}}(y^t)) \end{cases} \tag{18}$$

All the rules satisfying (18) with probability 1 are equivalent. Let

$$A_j(\theta_e^j, y^t) = \sum_{i=0}^{M} p(i) E^0(\frac{dP^i}{dP^0} \mid y^t) \int_{R^{L_i+N_i}} C(i, \theta^i \ ; j, \theta_e^j) \, p(t, i : x^i, \theta^i) dx^i \ d\theta^i$$

(19)

If $A_j(\theta_e^j, y^t)$ is continuous and lower bounded in θ_e^j (e.g. $C(i, \theta^i\ ;\ j, \theta_e^j) \geq 0$) at each y^t and for all j , then a nonrandomized estimation rule under the temporal decision of H_j exists, and is given by taking exclusively a $\theta_e^{j*} \in \Theta_j$ satisfying

$$\theta_e^{j*}(y^t) = Arg\ (\underset{\Theta_j}{Min}\ A_j(\theta_e^j, y^t)) \tag{20}$$

From (20) it follows that

$$A_j(\theta_e^{j*}, y^t) = \underset{P_{E_j}}{Min}\ \Phi_{P_{E_j}}(y^t)$$

,and the jointly optimal rule is given by

$$\begin{cases} \text{Select } (k, \theta_e^{k*}) \text{ iff} \\ A_k(\theta_e^{k*}, y^t) \leq A_j(\theta_e^{j*}, y^t) \quad \text{for all } j \neq k\ : \\ \theta_e^{j*}(y^t) = Arg\ (\underset{\Theta_j}{Min}\ A_j(\theta_e^j, y^t)) \end{cases} \tag{21}$$

According to the structure of the jointly optimal rule (21), estimations of $\theta_e^{j*}(y^t)$ under the temporary decision of H_j for all j precede the jointly optimal detection and estimation. Due to the assumptions made on $A_j(\theta_e^j, y^t)$, joint optimization can be replaced by the sequential optimization procedure, which simply exhausts all the possible $F_t(y)$ measurable rules in an obvious way. The implications of the rule (21) on detection and estimation operations will be shown in section (2.5).

2.3) Evaluation of the a posteriori loss

In this section we express $A_j(\theta_e^j, y^t)$ based on the well known results in non-linear filtering theory. Throughout this section we explicitly partition the augmented vector θ^i by $[\theta^i\ , h_t^i]$, where the latter θ^i is the system parameter vector of (1).
In section (2.1), we have defined the probability measure P^i on the σ -field F_T . Define a new measure P_0^i on the sub σ -field , $\sigma(G_T^i \bigcup F_T(y))$, of F_T by Girsanov transformation given below [8][20] .

$$\frac{dP_0^i}{dP^i} = \exp\left(-\frac{1}{r}\int_0^T h_t^{iT}\, dy(t) + \frac{1}{2r}\int_0^T |h_t^i|^2\, dt \right) \tag{22}$$

,where $h_t^i = h^i(t, x^i(t), \theta^i)$, and h_t^{iT} denotes a transpose.

Then P_0^i has the following properties[20].

a) P_0^i is a probability measure on $(\Omega, \sigma(G_T^i \bigcup F_T(y)))$

b) Under P_0^i, $(y(t), \sigma(G_t^i \bigcup F_t(y)))$ is a Wiener process of quadratic variation $r \cdot t$.

c) Under P_0^i, the processes $(y(t), 0 \le t \le T)$ and $(x^i(t), w^i(t), \theta^i, 0 \le t \le T)$ are independent.

d) The restriction of P_0^i to G_T^i is the same as the corresponding restriction of P^i.

e) P^i is absolutely continuous with respect to P_0^i, and

$$\frac{dP^i}{dP_0^i} = \exp\left(\frac{1}{r}\int_0^T h_t^{i\,T}\, dy(t) - \frac{1}{2r}\int_0^T |h_t^i|^2\, dt \right) \equiv Z_T^i$$

,and thus Z_t^i is a martingale under $(\sigma(G_t^i \bigcup F_t(y)), P_0^i)$.

Under each hypothesis H_i the following fundamental result ,known as Bayes formula formula [20] in the nonlinear filtering theory, holds in the present formulation also.

$$E^i(\, \phi(x^i(t), \theta^i) \mid F_t(y)\,) = \frac{\sigma_t^i(\phi)}{\sigma_t^i(1)} \tag{23}$$

,where $\sigma_t^i(\phi) = E_0^i(\phi(x^i(t), \theta^i)\, Z_t^i \mid F_t(y))$, and $\phi(\cdot\,,\cdot)$ is a measurable function such that $E^i|\phi(x^i(t), \theta^i)| < \infty$, and E^i, E_0^i denote the expectation taken under P^i and P_0^i respectively.

By definition $A_j(\theta_e^j, y^t)$ equals to

$$A_j(\theta_e^j, y^t) = \sum_{i=0}^{M} p(i)\, E^0\!\left(\frac{dP^i}{dP^0} \mid y^t\right) E^i(\, C(i, \theta^i, h_t^i; j, \theta_e^j) \mid y^t\,) \tag{24}$$

From (23) we can identify

$$E^i(\, C(i, \theta^i, h_t^i; j, \theta_e^j) \mid F_t(y)\,) = \frac{\sigma_t^i(\phi')}{\sigma_t^i(1)} \tag{25}$$

,where $\phi'(x^i(t), \theta^i) = C(i, \theta^i, h_t^i; j, \theta_e^j)$.

Also we note that

$$\sigma_t^i(1) \equiv E_0^i(Z_t^i \mid F_t(y)) = E^0(\frac{dP^i}{dP^0} \mid F_t(y)) \quad P_0^i \ , P^0 \ \ a.s. \tag{26}$$

Since

$$E_0^i(\ Z_t^i \mid F_t(y)) = \frac{dP_Y^i}{dP_{0,Y}^i}|_{F_t(y)} \quad ,P_0^i \ \ a.s.$$

$$E^0(\ \frac{dP^i}{dP^0} \mid F_T(y)) = \frac{dP_Y^i}{dP_Y^0}|_{F_t(y)} \quad ,P^0 \ \ a.s.$$

,and since $P_{0,Y}^i$, the restriction of P_0^i to $F_T(y)$, is identical to P_Y^0 from Girsanov's theorem (; | denotes a restriction to a sub σ -field) .

From (25) and (26) we have

$$E^i(\ C(i,\theta^i,h_t^i\ ;\ j,\theta_e^j) \mid F_t(y)) \cdot E^0(\frac{dP^i}{dP^0}) \mid F_t(y)) = E_0^i(\ C(i,\theta^i,h_t^i\ ;\ j,\theta_e^j)\ Z_t^i \mid F_t(y)\ ) \quad a.s. \tag{27}$$

Define $\rho(t,i;x^i,\theta^i)$ such that for an arbitrary measurable function ϕ satisfying $E^i|\phi(x^i(t),\theta^i)| < \infty$,

$$\sigma_t^i(\phi) = \int_{R^{L_i}} \int_{R^{N_i}} \phi(x^i,\theta^i)\,\rho(t,i;x^i,\theta^i)dx^i \ d\theta^i \qquad \text{a.s.} \tag{28}$$

Then from (23) and from the arbitrariness of $\phi(\cdot\ ,\cdot)$, $\rho(t,i;x^i,\theta^i)$ can be identified as the unnormalized a posteriori density satisfying

$$\frac{\rho(t,i;x^i,\theta^i)}{\int_{R^{L_i}}\int_{R^{N_i}}\rho(t,i;x^i,\theta^i)dx^i \ d\theta^i} = p\,(x^i(t),\theta^i \mid y^t,H_i\ )$$

,where equality is understood in a proper sense.

Let $\rho(t,i;\theta^i) = \int_{R^{N_i}}\rho(t,i;x^i,\theta^i)dx^i$.

We can identify $\rho(t,i;\theta^i)$ through the following steps.

$$\begin{aligned}&\int_{R^{L_i}}\phi(\theta^i)\,\rho(t,i;\theta^i)\ d\theta^i \\ &= E_0^i(\ \phi(\theta^i)\ Z_t^i \mid y^t) \\ &= E_0^i[\ \phi(\theta^i)\ E_0^i(\ Z_t^i \mid F_t(y),F(\theta^i)\ ) \mid y^t\ ]\end{aligned} \tag{29}$$

We note that

$$\begin{aligned}
&E_0^i(\, Z_t^i \mid F_t(y), F(\theta^i)\,) \\
&= E_0^i(\, Z_T^i \mid F_t(y), F(\theta^i)\,) \\
&= \frac{dP^i}{dP_0^i}\Big|_{F_t(y)\bigcup F(\theta^i)} \\
&= \frac{p^i(\theta)}{p^i(\theta)} \cdot \frac{dP_Y^{i,\theta}}{dP_{0,Y}^i}\Big|_{F_t(y)}
\end{aligned} \tag{30}$$

since θ^i and the process y are independent under P_0^i .

Substituting (30) into (29) yields,

$$\begin{aligned}
&\int_{R^{Li}} \phi(\theta^i)\, \rho(t\,,i\,;\theta^i)\, d\,\theta^i \\
&= E_0^i(\, \phi(\theta^i)\, \frac{dP_Y^{i,\theta}}{dP_{0,Y}^i}\Big|_{F_t(y)} \mid y^t\,) \\
&= \int_{R^{Li}} \phi(\theta^i)\, \frac{dP_Y^{i,\theta}}{dP_{0,Y}^i}\Big|_{y^t} p^i(\theta)\, d\,\theta^i
\end{aligned} \tag{31}$$

,due to the same reason following (30).

Since $\phi(\,\cdot\,)$ is an arbitrary integrable function, (31) implies that

$$\begin{aligned}
&\rho(t\,,i\,;\theta^i) \\
&= p^i(\theta)\, \frac{dP_Y^{i,\theta}}{dP_{0,Y}^i}\Big|_{y^t} \\
&= p^i(\theta)\, \frac{dP_Y^{i,\theta}}{dP_Y^0}\Big|_{y^t} \\
&= p^i(\theta)\, E^0(\frac{dP^{i,\theta}}{dP^0} \mid y^t\,) \\
&= \Lambda(\theta^i)\cdot p^i(\theta) \\
&= \exp\,(\,\frac{1}{r}\int_0^t \hat{h}_s^i(\theta^i)^T\, dy\,(s\,) - \frac{1}{2r}\int_0^t |\,\hat{h}_s^i(\theta^i)\,|^2 ds\ )\ p^i(\theta)
\end{aligned} \tag{32}$$

,where $\hat{h}_t^i(\theta^i) = E^i(\, h_t^i \mid y^t, \theta^i\,)$.

In (32) $\Lambda(\theta^i)$ is the conditional likelihood ratio between (θ^i, H_i) and H_0 , and is expressed by the causal representation [21] .

We now derive the evolution equation that $\rho(t\,,i\,;x^i\,,\theta^i)$ satisfies.

Since $x^i(t\,)$ is a θ^i conditionally Markov process, $\sigma_t^i(\phi(x^i(t\,),\theta^i\,))$ satisfies the stochastic integral equation given below [20].

$$\sigma_t^i(\phi) = \sigma_0^i(\phi) + \int_0^t \sigma_s^i(\, L_{\theta^i}\phi\,)ds\ + \frac{1}{r}\int_0^t \sigma_s^i(h_s^{i\,T}\ \phi)dy\,(s\,) \tag{33}$$

,where L_{θ^i} is the infinitesmal generator for a diffusion process $x^i(t)$ corresponding to the parameter value θ^i under the hypothesis H_i, and is given by

$$L_{\theta^i} = \sum_{j=0}^{N_i} f_j^i(x^i,\theta^i)\,\frac{\partial}{\partial x_j} + \frac{1}{2}\sum_{j,k}[\,g^i g^{i\,T}\,]_{j,k}\,\frac{\partial^2}{\partial x_j\,\partial x_k}$$

,where x_j is a component of state vector x^i .

In (33) $\phi(\cdot,\theta^i)$ belongs to domain of the operator L_{θ^i} and is such that the stochastic integration of (33) can be justified as a local martingale under P_0^i.

From (28) and (33) ,and also by applying integration by parts in x^i under certain differentiability conditions, we formally obtain the following stochastic partial differential equation that $\rho(t,i;x^i,\theta^i)$ satisfies.

$$\begin{aligned} d\,\rho(t,i;x^i,\theta^i) &= L_{\theta^i}^{*}\,\rho(t,i;x^i,\theta^i)dt \;+\; \frac{1}{r}\,h^i(t,x^i,\theta^i)^T\,\rho(t,i;x^i,\theta^i)\,dy(t) \\ \rho(0,i;x^i,\theta^i) &= p^i(\theta)\,p^i(x_0 \mid \theta) \end{aligned} \qquad (34)$$

, where $L_{\theta^i}^{*}$ is the formal adjoint of L_{θ^i} ,and the initial condition is specified by (23). (34) is a bilinear stochastic partial differential equation for the unnormalized joint density $\rho(t,i;x^i,\theta^i)$. The use of (34) as a way to provide an analytical approximation of ρ , jointly in (x^i,θ^i), is beyond a scope of the present work.

A typical approach to avoid PDE formulation is the use of approximating finite dimensional nonlinear filtering obtained by treating $[x^{i\,T},\theta^{i\,T}]^T$ as a joint state, such as the jointly gaussian approximation of the a posteriori density of $(x^i(t),\theta^i)$. However convergence properties of implemented conditional means of (x^i,θ^i) are of questionable merits and difficult to check. In a related work Ljung studied the extended Kalman filter as a parameter estimator in the discrete parameterized linear gaussian system. He treated $[x^i(t)^T,\,\theta^{i\,T}]^T$ as an augmented state vector, and modified the gain factor of extended Kalman filter algorithm properly so as to ensure global convergence of parameter estimates [22] . A similar work may also be expected for the parameterized diffusion cases. However such fundamental knowledges as to identifiability of a given parameterized system should be prerequisites for a similar analysis [23] .

At each fixed θ^i, (34) is Zakai equation associated with (θ^i, H_i) ,and thus $\rho(t,i;x^i,\theta^i)$ of (34) at each fixed θ^i is equal to the solution to the Zakai equation weighted by $p^i(\theta)$. Approximating the parameter space by a finite set A_i belonging to Θ_i, we have a bank of the unnormalized a posteriori density calculators , each corresponding to $\theta^i \in A_i$. At each θ^i one can directly implement Zakai equation numerically [19] [24], or may use the gaussian approximation of the a posteriori density of $x^i(t)$ especially when signal to noise ratios are high or very rapid real time implementations are required. When only the estimation of θ^i is of interest, an approximate realization of the causal conditional mean of the signal in the likelihood ratio of (32) may also be used as a simple suboptimal structure, as suggested in [21], with a due attention on the implementation of stochastic integral (such as the Stratonovich version of causal likelihood ratio [25]).

In section (2.4) we adapt the Bucy-Senne's numerical filtering algorithm [27] to the present situation by treating (x^i,θ^i) partially jointly.

In theory the a posteriori loss $A_j(\theta_e^j, y^t)$ associated with a temporary decision and estimation ,(j, θ_e^j), can be given by

$$\begin{aligned} &A_j(\theta_e^j, y^t) \\ &= \sum_{j=0}^{M} p(i)\, \sigma_t^i(\phi(x^i(t), \theta^i)) \\ &= \sum_{j=0}^{M} p(i) \int_{R^{Li}} \int_{R^{Ni}} \phi(x^i, \theta^i)\, \rho(t,i;x^i,\theta^i)\, dx^i\, d\theta^i \end{aligned} \qquad (35)$$

, where ρ satisfies (34) ,and $\phi(x^i(t),\theta^i) = C(i, \theta^i, h_t^i; j, \theta_e^j)$.

When directly implemented, (35) requires discretizations of the parameter and state spaces. The structure of simultaneously optimal receiver is drawn in Fig.1 from a functional point of view.

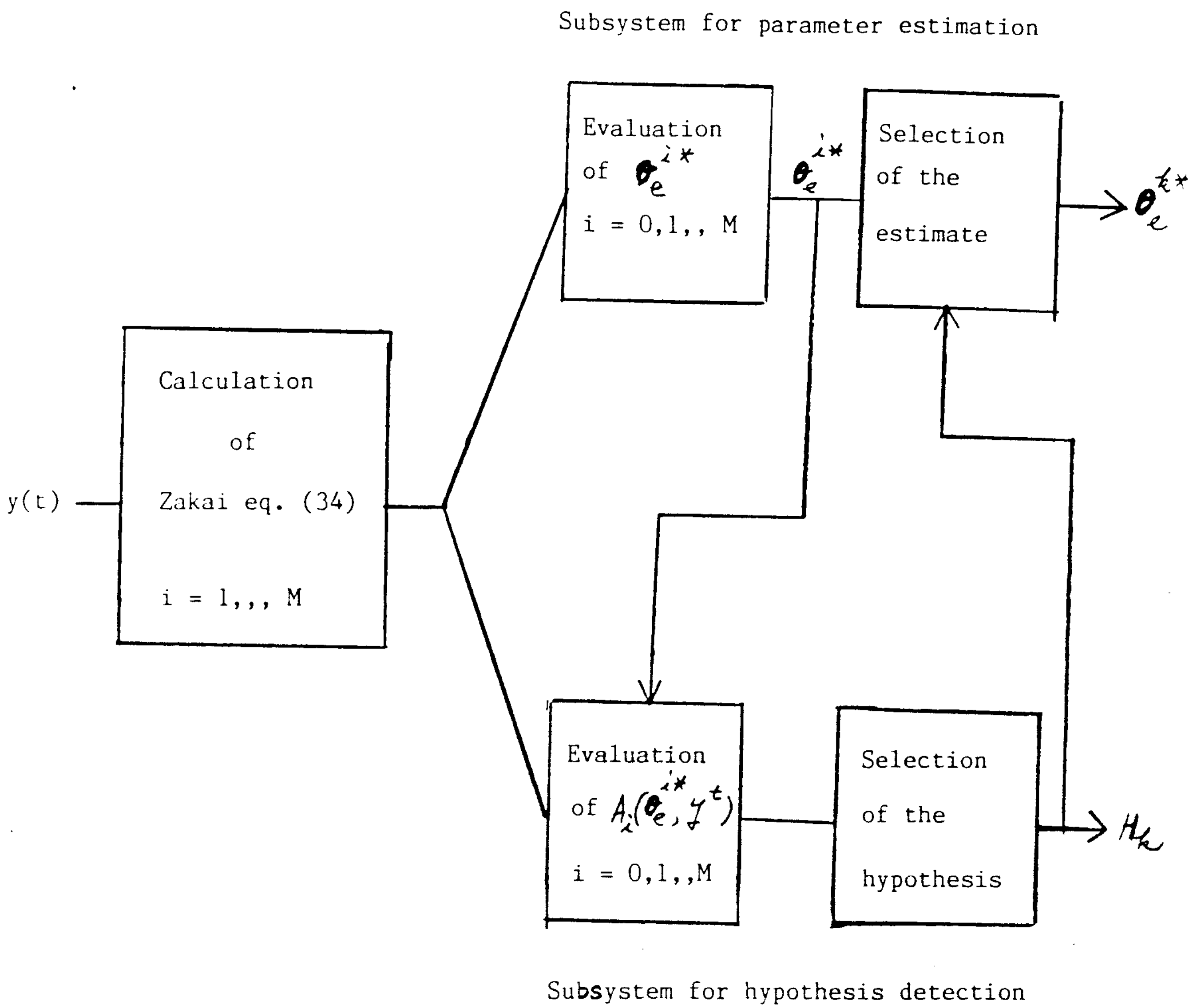
Subsystem for parameter estimation
Evaluation of θ_e^{i*} i = 0,1,, M
θ_e^{i*}
Selection of the estimate
θ_e^{k*}
y(t)
Calculation of Zakai eq. (34) i = 1,,, M
Evaluation of $A_i(\theta_e^{i*}, y^t)$ i = 0,1,,M
Selection of the hypothesis
H_k
Subsystem for hypothesis detection

Fig.1

2.4) Numerical Approximation

In this section we treat a discrete version of model (1) . The resulting discrete time observation model may be either a description of originally discrete Markov process or a time discretized version of continuous observation model through the following standard procedure. Let the continuous observation model be given by

$$\begin{aligned} d\,\mathbf{x}(t) &= \mathbf{f}(\mathbf{x}(t))\,dt + \mathbf{g}(\mathbf{x}(t))\,d\,\mathbf{w}(t) \\ d\,\mathbf{y}(t) &= h(t, \mathbf{x}(t))\,dt + \sqrt{r}\;d\,\mathbf{v}(t) \end{aligned} \tag{36}$$

Suppose the time axis is discretized by Δ. Then we obtain

$$\mathbf{x}((k+1)\Delta) - \mathbf{x}((k\Delta)) = \int_{k\Delta}^{(k+1)\Delta} \mathbf{f}(\mathbf{x}(t))\,dt + \int_{k\Delta}^{(k+1)\Delta} \mathbf{g}(\mathbf{x}(t))\,d\,\mathbf{w}(t)$$

Assume that Δ is sufficiently small such that

$$\begin{aligned} &\int_{k\Delta}^{(k+1)\Delta} \mathbf{f}(\mathbf{x}(t))\,dt \sim \mathbf{f}(\mathbf{x}(k\Delta))\,\Delta \\ &\int_{k\Delta}^{(k+1)\Delta} \mathbf{g}(\mathbf{x}(t))\,d\,\mathbf{w}(t) \sim \mathbf{g}(\mathbf{x}(k\Delta)) \left\{ \mathbf{w}((k+1)\Delta) - \mathbf{w}(k\Delta) \right\} \end{aligned} \tag{37}$$

Define $x(k)$, $f(x(k))$ as

$$\begin{aligned} x(k) &= \mathbf{x}(k\Delta) \\ f(x(k)) &= x(k) + \mathbf{f}(x(k))\,\Delta \end{aligned} \tag{38}$$

, and $w(k)$, $g(x(k))$ as

$$\begin{aligned} w(k) &= \frac{1}{\sqrt{\Delta}} \left\{ \mathbf{w}((k+1)\Delta) - \mathbf{w}(k\Delta) \right\} \\ g(x(k)) &= \sqrt{\Delta}\;\mathbf{g}(x(k)) \end{aligned} \tag{39}$$

Then $w(k)$ is a zero mean white gaussian vector with covariance

$$E(w(k)\,w(k)^T) = I$$

Thus the discrete state equation becomes

$$x(k+1) = f(x(k)) + g(x(k))\,w(k) \tag{40}$$

Similarly for the measurement model,

$$d\,\mathbf{y}(t) = h(t, \mathbf{x}(t))\,dt + \sqrt{r}\;d\,\mathbf{v}(t) \;,$$

we obtain

$$\mathbf{y}((k+1)) - \mathbf{y}(k\Delta) \sim h(k\Delta, \mathbf{x}(k\Delta))\,\Delta + \sqrt{r}\left\{\mathbf{v}((k+1)\Delta) - \mathbf{v}(k\Delta)\right\} \tag{41}$$

Now define $y(k)$ as

$$y(k) = \frac{1}{\Delta}\left\{\mathbf{y}((k+1)\Delta) - \mathbf{y}(k\Delta)\right\} \tag{42}$$

, and $v(k)$ as

$$v(k) = \frac{1}{\sqrt{\Delta}}\left\{\mathbf{v}((k+1)\Delta) - \mathbf{v}(k\Delta)\right\} \tag{43}$$

Then the discrete measurement equation is

$$y(k) = h(k, x(k)) + \sqrt{\frac{r}{\Delta}}\, v(k) \tag{44}$$

, where $v(k)$ is a zero mean white gaussian noise vector with covariance

$$E(v(k)\, v(k)^T) = I$$

Equations (37) through (44) yields a discrete time Markov sequence. Quality of approximations, (37) (41) , depends on a bandwidth of the given diffusion. In [20] a bandwidth , B , is intuitively defined by

$$B = \sup_x \left\{ \frac{|\mathbf{f}(x)|}{1+|x|} + \frac{\|\mathbf{g}(x)\|^2}{1+|x|^2} \right\}$$

under the assumption that

$$|\mathbf{f}(x)| + \|\mathbf{g}(x)\| \leq k\sqrt{1+|x|^2} \quad .$$

It follows that $\Delta << \frac{1}{B}$ for a good approximation.

In the linear system $f(x(k))$ can be exactly identified by

$$f(x(k)) = \Phi(\Delta)\cdot x(k)$$

, where $\Phi(t, t_0)$ is a state transition matrix for the given linear system. If considering the previous parametric uncertainty, we have the following observation model under each hypothesis H_i ,

$$x^i(t+1) = f^i(x^i(t),\theta^i) + g^i(x^i(t),\theta^i)\, w^i(t) \qquad x^i(0) = x_0^i$$
$$y(t) = h^i(t, x^i(t), \theta^i) + \sqrt{r'}\, v^i(t) \tag{45}$$
$$t = 0,1,2,,,T$$

,where $r' = \sqrt{\frac{r}{\Delta}}$ when (45) is an approximation to (1). In connection with r', we mention that $r' \to \infty$ as $\Delta \to 0$. Hence if going from (45) to the continuous counter part , one should take care of order of limiting procedures. The advantage of the present discretization lies in that one can use the floating grid method without much complication to redefine the quantization of state space [27] adaptively , and that the resulting numerical filtering algorithm is inherently stable.

$f^i(\cdot,\cdot), g^i(\cdot,\cdot)$, and $h^i(\cdot,\cdot,\cdot)$ are measurable vector-valued or matrix-valued functions respectively, and also they are integrable when considered as random processes at each θ^i. As before the null hypothesis is given by $y(t) = \sqrt{r'}\, v^0(t)$. Construction of (Ω, F_T, P^i) is similar to that of section (2.1) . P^i yields a joint density function,

$$p^i(\theta^i, x_0^i, w^i, y) = p^i(\theta)\, p^i(x_0 \mid \theta) \prod_{t=0}^{T} p_{g1}(w^i(t)) \prod_{t=0}^{T} p_{g2}(y(t) \mid \theta^i, x_0^i, w^i(s), s \le t\text{-}1)$$

, where p_{g1} ,and p_{g2} are jointly normal distributions with proper dimensions of identity covariance matrices respectively.

Define a probability measure P_0^i on the sub σ-field of F_T ,$\sigma(G_T^i \bigcup F_T(y))$, by

$$\frac{dP_0^i}{dP^i} = \exp(-\frac{1}{r'} \sum_{t=0}^{T} h_t^{iT}\, y(t) + \frac{1}{2r'} \sum_{t=0}^{T} | h_t^i |^2) \tag{46}$$

$y(T) = \frac{1}{\Delta}\left\{ \mathbf{y}((T+1)\Delta) - \mathbf{y}(T\Delta) \right\}$, if considered as the discrete sampling of continuous observation. Then the new probability measure ,P_0^i, yields the following probability density,

$$p_0^i(\theta^i, x_0^i, w^i, y) = p^i(\theta)\, p^i(x_0 \mid \theta) \prod_{t=0}^{T} p_{g1}(w^i(t)) \prod_{t=0}^{T} p_{g2}(y(t)) \tag{47}$$

Let

$$Z_t^i = \exp\left(\frac{1}{r'} \sum_{s=0}^{t} h_s^{iT} \; y(s) - \frac{1}{2r'} \sum_{s=0}^{t} | h_s^i |^2 \right)$$

Then

$$E_0^i (Z_T^i \mid \sigma(G_t^i \cup F_t(y))) = Z_t^i \tag{48}$$

We state below a representation theorem ,due to Bucy [26]. For an arbitrary integrable function $\phi(\cdot,\cdot)$,

$$E^i (\phi(x^i(t'), \theta^i) \mid F_t(y)) = \frac{E_0^i (Z_t^i \phi(x^i(t'), \theta^i) \mid F_t(y))}{E_0^i (Z_t^i \mid F_t(y))}, \quad t' \leq t \tag{49}$$

The proof appearing in [20] for Bayes formula can also be used here. We note that $w^i(\cdot),v^i(\cdot)$ need not be gaussian processes for (49) and the subsequent result to hold.

Let $t = t'$ in (49) , and define $\rho(t,i;x^i,\theta^i)$ as in section (2.3) .

$$E_0^i (Z_t^i \phi(x^i(t), \theta^i) \mid y^t) = \int_{R^{L_i}} \int_{R^{N_i}} \phi(x^i,\theta^i) \rho(t,i; x^i,\theta^i) \, dx^i \, d\theta^i \quad \text{a.s.} \tag{50}$$

$\rho(t,i;x^i,\theta^i)$ is an unnormalized conditional density of (x^i,θ^i) given y^t .

From (50) ρ can be identified in the following way.

$$\begin{aligned} & E_0^i (Z_t^i \phi(x^i(t),\theta^i) \mid y^t) \\ & = \int \int E_0^i (Z_t^i \phi(x^i(t),\theta^i) \mid y^t , x^i(t) = x^i ,\theta^i) \, p_0^i (t , x^i , \theta^i \mid y^t) \, dx^i \, d\theta^i \\ & = \int \int E_0^i (Z_t^i \phi(x^i(t),\theta^i) \mid y^t , x^i , \theta^i) \, p^i(t, x^i, \theta^i) \, dx^i \, d\theta^i \quad , \end{aligned}$$

due to (47) . Thus from (50) and arbitrariness of ϕ ,

$$\rho(t,i;x^i,\theta^i) = E_0^i (Z_t^i \mid y^t , x^i , \theta^i) \, p^i(t, x^i, \theta^i) \tag{51}$$

, where equality is understood in a proper sense.

Now we derive the recursive equation that $\rho(t,i;x^i,\theta^i)$ satisfies. From (51) we have

$$\rho(t+1,i;x^i,\theta^i) = E_0^i (Z_{t+1}^i \mid y^{t+1}, x^i(t+1) = x^i , \theta^i) \, p^i(t+1, x^i,\theta^i) \quad . \tag{52}$$

Let

$$\begin{aligned} & \tilde{x}^i(t) = [x^i(t)^T , \theta^{iT}]^T \\ & H_t (\tilde{x}^i(t), y(t)) = \exp\left(\frac{1}{r'} h_t^{iT} y(t) - \frac{1}{2r'} | h_t^i |^2 \right). \end{aligned}$$

Then $\tilde{x}^i(t)$ is Markov. And from (52) ,

$$\begin{aligned}
&\rho(t+1,i;u) \\
&= E_0^i\left[Z_t^i H_{t+1}(\tilde{x}^i(t+1),y(t+1)) \mid y^{t+1}, \tilde{x}^i(t+1)=u \right] \cdot p^i(t+1,u) \\
&= H_{t+1}(u,y(t+1))\, E_0^i\left[E_0^i(Z_t^i \mid y^{t+1}, \tilde{x}^i(t+1)=u, \tilde{x}^i(t)=v) \right. \\
&\qquad\qquad \left. \mid y^{t+1}, \tilde{x}^i(t+1)=u \right] \cdot p^i(t+1,u) \\
&= H_{t+1}(u,y(t+1))\, E_0^i\left[E_0^i(Z_t^i \mid y^t, v) \mid y^{t+1}, u \right] \cdot p^i(t+1,u)
\end{aligned}$$

, since under P_0^i $\tilde{x}^i(\cdot)$ and $y(\cdot)$ are independent , $y(\cdot)$ a white gaussian process , and $\tilde{x}^i(\cdot)$ a Markov process having the same distribution as under P^i .

From the above we obtain

$$\begin{aligned}
&\rho(t+1,i;u) \\
&= H_{t+1}(u,y(t+1)) \left\{ \int E_0^i(Z_t^i \mid y^t, v) \frac{p_{t+1\mid t}^i(u \mid v)\, p^i(t,v)}{p^i(t+1,u)}\, dv \right\} \cdot p^i(t+1,u) \\
&= H_{t+1}(u,y(t+1)) \int \rho(t,i;v)\, p_{t+1\mid t}^i(u \mid v)\, dv
\end{aligned}$$

Since

$$p_{t+1\mid t}^i(u \mid v) = p_{t+1\mid t}^i(x^i(t+1)=x^i \mid x^i(t)=x^{i'}, \theta^{i'})\, \delta(\theta^i - \theta^{i'}) \quad ,$$

the desired recursive equation is

$$\rho(t+1,i;x^i,\theta^i) = H_{t+1}(x^i,\theta^i, y(t+1)) \int_{R^{N_i}} \rho(t,i;x^{i'}, \theta^i)\, p_{t+1\mid t}^i(x^i \mid x^{i'}, \theta^i)\, dx^{i'} \quad (53)$$

with the initial condition ,

$$\rho(0,i;x^i,\theta^i) = E_0^i(Z_0^i \mid y(0),x^i,\theta^i)\, p^i(0;\theta^i,x^i) .$$

The jointly optimal detection and estimation rule for the discrete observation model (36) is obviously the same as before ,and $\Phi_{P_{E_j}}(y^t)$ under the temporal decision of j and an estimation rule P_{Ej} is given by

$$\Phi_{P_{E_j}} = \int_{R^{L_j+K}} \left\{ \sum_{j=0}^{M} p(u=j) \int_{R^{L_i+N_i}} C(i,\theta^i ; j,\theta_e^j)\, \rho(t,i;x^i,\theta^i)\, dx^i\, d\theta^i \right\} dP_{Ej}(\theta_e^j \mid y^t)$$

,where θ^i and θ_e^i are defined as the previous augmented vectors of section (2.2). Depending on assumptions , the jointly optimal rule is given by either eq (18) or (21).

Now we consider an numerical algorithm to calculate $\rho(t,i;x^i,\theta^i)$.

We adapt the Bucy-Senne's numerical filtering algorithm to the present observation model where parameter uncertainties are present. The essence of Bucy- Senne's algorithm lies on the use of floating grid, and construction of floating grid is based upon the implemented a posteriori conditional error covariance and conditional mean of signal. Thereby their algorithm incorporates the good feature of gaussian approximation technique into the nonlinear filtering , while it does not involve any numerical instability in calculating the error covariance unlike the gaussian approximation , where a precautionary measure is necessary to make the conditional error covariance positive definite.

Henceforth we neglect the hypothesis index i , assuming a some fixed value. We use an usual point mass approximation of $\rho(t;x^i, \theta^i)$. Let

$$q(t;x,\theta) \equiv \sum_{i_1, i_2,, i_N = 1}^{2M+1} q(t; g_t(i_1, i_2,,, i_N), \theta)\, \delta(x - g_t(i_1, i_2,,, i_N)) \ , \theta \in A_i \quad (54)$$

, where δ is the Dirac delta function , and $2M+1$ a number of grid points in each coordinate x_j^i , and A_i a finite subset of Θ_i .

In (54) $g_t(\cdot\cdot)$ is the floating grid map defined by

$$K^N \rightarrow R^N \qquad , K = \left\{ 1,2,,,, 2M+1 \right\}$$

Define $q(t,\theta)\ (\in R^{2M+1^N})$ as

$$q(t,\theta) = [\, q(t; g_t(1, 1,,, 1), \theta),,, q(t; g_t(2M+1,,, 2M+1), \theta)\,]^T \quad (55)$$

The straightforward extension of Bucy-Senne's algorithm requires that the grid map g_t depends on each parameter values. Here we take the approach that the grid map of the state space is independent of θ, and is obtained by considering the a posteriori uncertainty of parameter random variable. Calculation of $q(t;x,\theta)$ is based on the evolution eq (53) . The advantage of the present approach is that we can reduce the computational burden and storage space , which are otherwise required to determine the grid map at each θ .

The inherent drawback of this scheme is explained in the following. Suppose that a true value of the parameter is θ_0 under H_i and that $\theta_0 \in A_i$ for simplicity. Consider

$q(t;x,\theta)$ under H_i . Value range of the implemented version of a pseudo conditional mean ($\approx E^i(\,x(t)\mid\theta,y^t\,)$) has to be limited by the present grid map favorable for the implemented version of the conditional mean of the state ($\approx E^i(x(t)\mid y^t\,)$) , and therefore $q(t;x,\theta)$ should also be affected in the way that the actual functional shape of $\rho(t;x,\theta)$ is smoothed out in $q(t;x,\theta)$. The degree of the drawback will depend on the range of A_i chosen , given a common floating grid for the state space. If the a posteriori accuracy about θ is low, the above will not be a matter of concerns , since the grid map is not dominantly determined by $q(t;x,\theta)$ associated with a particular θ_0 . With the increased observation period the adversary effect of present quantization scheme will begin to appear . Whether this would be a matter of minor importance or not depends on a shape of $q(t;x,\theta)$ during the intermediate observation period , and accordingly we are not sure of convergence properties of $q(t;x,\theta)$, in terms of the implemented conditional means, with the present method of gridding for the state space.

Since we use the point mass approximation, and also from the evolution eq (53) , the grid A_i of the parameter space has to be fixed unless a proper interpolation of $q(t;x,\theta)$ is to be used .

With this approximation the state becomes a pair ($q(t,\theta), g_t(\cdot\,\cdot)$; $\theta\in A_i$) , the first a $(2M+1)^N$ vector at each θ , and the second a $(N)\times(2M+1)^N$ matrix, the total collection of actual quantized points in R^N .

Substituting (54) into the evolution equation (53) at each θ , we arrive at the state update equation .

$$
\begin{aligned}
&q(t+1;i_1,,,,,i_N,\theta) \\
&= H_{t+1}(\,g_{t+1}(i_1,,,\,i_N\,),\theta,\,y(t+1)\,) \\
&\quad\times\sum_{j_1,,,\,j_N=1}^{2M+1} T(\,g_{t+1}(i_1,,,\,i_N\,),\,g_t(j_1,,,\,j_N\,),\theta\,)\,q(t;\,j_1,,,j_N\,,\theta\,)
\end{aligned}
\tag{56}
$$

, where

$$T(\, g_{t+1}, g_t\,, \theta\,)$$
$$\equiv p_{t+1\,|\,t}(\, g_{t+1} \mid g_t\,, \theta\,)$$
$$= N\,[\, g_{t+1} - f\,(\, g_t\,,\theta\,),\; a\,(\, g_t\,,\theta\,)\,]$$

with $a\,(\,x\,,\theta\,) = g\,(\,x\,,\theta\,)\,g\,(\,x\,,\theta\,)^T$. In (2.1) we assume that $a\,(\,x\,,\theta\,)$ is positive definite for all x and θ.

(56) can be sequentially computed , once the gridding is determined. However we must evaluate $(2M+1)^{2N}$ elements of matrix T ,and $(2M+1)^N$ elements of diagonal matrix H_{t+1} , and perform multiplications at each $\theta \in A_i$. In order to select the gridding most effectively under the present constraint on g_t , we centers the grid for $[\, q\,(t+1,\theta\,) : \theta \in A_i\,]$ at the best estimate of $x\,(t+1)$ given y^t (i.e. $\hat{x}\,(t+1 \mid t\,)\,(\equiv E(\,x\,(t+1) \mid y^t\,)$ since the grid g_{t+1} must be given before $q\,(t+1,\theta\,)$ is computed, and only $q\,(t\,,\theta\,)$ is known at time t) .

If the given model (36) is derived from the continuous model (1) , $\hat{x}\,(\,t+1 \mid t)$ is actually the filtered estimate of $\mathbf{x}(\,(t+1)\,\Delta\,)$ given $(\,\mathbf{y}^{(t+1)\,\Delta}\,)$ as can be seen from the definition of $\mathbf{y}(t\,)$ of (42).

Similarly the mesh size and directions are given in terms of eigenvalues and eigenvectors of conditional error covariance $\sum(\,t+1 \mid t\,)$ of $x\,(t+1)$ given y^t .

$$\begin{aligned}
&\hat{x}\,(\,t+1 \mid t\,) \\
&= E(\,f\,(x\,(t\,),\theta\,) + g\,(x\,(t\,),\theta\,)\,w\,(t\,) \mid y^t\,) \\
&= E(\,f\,(x\,(t\,),\theta\,) \mid y^t\,) + E\,[\,E(\,g\,(x\,(t\,),\theta)\,w\,(t\,) \mid \sigma(x\,(t\,),\theta\,)\,,F_t\,(y\,)\,) \mid y^t\,) \\
&= \int f\,(x\,(t\,),\theta\,)\,P\,(x\,(t\,),\theta \mid y^t\,)\,dx\;d\,\theta \\
&\to \frac{1}{C\,(t\,)} \sum_{j_1 .. j_N,\,\theta \in A_i}^{2M+1} f\,(\,g_t(\,j_1\,,, j_N\,)\,,\theta\,)\,q\,(t\,;j_1\,,,, j_N\,,\theta\,)\,,
\end{aligned} \tag{57}$$

, where $C\,(t\,)$ is the normalization constant for q , and is an approximation to $E^0(\,\frac{dP^i}{dP^0} \mid y^t\,)$.

$$\begin{aligned}
&\Sigma(t+1 \mid t) \\
&= E[(x(t+1) - \hat{x}(t+1 \mid t))(x(t+1) - \hat{x}(t+1 \mid t))^T \mid y^t] \\
&= E[x(t+1)\,x(t+1)^T \mid y^t] - \hat{x}(t+1 \mid t)\,\hat{x}(t+1 \mid t)^T \\
&= E[(f(x(t),\theta) + g(x(t),\theta)\,w(t))(f^T(x(t),\theta) + w^T g^T(x(t),\theta)) \mid y^t] - \hat{x}\,\hat{x}^T \\
&= E[f(x(t),\theta)\,f^T(x(t),\theta) \mid y^t] + E[g(x(t),\theta)\,w(t)\,w^T(t)\,g^T(x(t),\theta) \mid y^t] - \hat{x}\,\hat{x}^T \\
&= E[f(x(t),\theta)\,f^T(x(t),\theta) \mid y^t] + E[g(x(t),\theta)\,g^T(x(t),\theta) \mid y^T] - \hat{x}\,\hat{x}^T \\
&= \int\int \left\{ f(x,\theta)\,f^T(x,\theta) + g(x,\theta)\,g^T(x,\theta) \right\} \cdot P(x(t),\theta \mid y^t)\,dx\,d\theta - \hat{x}\,\hat{x}^T \\
&= \frac{1}{C(t)} [\Sigma(\, f(g_t(j_1,, j_N), \theta)\, f^T(g_t(j_1,,, j_N),\theta) \qquad (58) \\
&\qquad + g(g_t(j_1,, j_N), \theta)\, g^T(g_t(j_1,, j_N), \theta)\,) \\
&\qquad \times q(t, j_1,, j_N, \theta)] - \hat{x}(t+1 \mid t)\,\hat{x}(t+1 \mid t)^T
\end{aligned}$$

Let $\lambda_{t+1}^1,, \lambda_{t+1}^N$, $e_{t+1}^1,, e_{t+1}^N$ be the eigenvalues and the eigenvectors of $\Sigma(t+1 \mid t)$. Define a new coordinate system

$$\tilde{x}(t+1) = [\, e_{t+1}^1 \cdots \cdots e_{t+1}^N \,]^{-1} [\, x(t+1) - \hat{x}(t+1 \mid t) \,] \qquad (59)$$

Then

$$E[\,\tilde{x}(t+1)\,\tilde{x}(t+1)^T\,] = D(\lambda) \qquad (60)$$

, where $D(\lambda)$ is a diagonal matrix with $d_{ii} = \lambda_{t+1}^i$.

Define a grid on the new coordinate system by

$$\begin{bmatrix} K_{t+1}^1 \\ \cdot \\ \cdot \\ K_{t+1}^l \\ \cdot \\ \cdot \\ K_{t+1}^N \end{bmatrix}$$

,where $K_{t+1}^l = n_g\,(\lambda_{t+1}^l)^{\frac{1}{2}}\,(\frac{i_l - 1}{M} - 1)$,$i_l \in K$ and n_g is a constant parameter controlling the step size.

Update g_{t+1} by

$$g_{t+1}(i_1,,, i_N) = [\, e_{t+1}^1 \cdots e_{t+1}^N \,] \begin{bmatrix} K_{t+1}^1 \\ \cdot \\ \cdot \\ \cdot \\ K_{t+1}^N \end{bmatrix} + \hat{x}(t+1 \mid t) \qquad (61)$$

(61) defines a desired update of the floating grid at each time. The above is only an

outline of the numerical algorithm ,and further simplifying will be possible depending on the problems.

2.5) Implication on the operations of detection and estimation

In this section we first examine the implication of the jointly Bayesian approach on the operations of detection and estimation. Also the unified formulation of detection and estimation as a nonlinear estimation problem [7] is given for a compound multiple hypotheses problem and treated as a special case of the jointly Bayesian approach . Extension to the multiple observation intervals of the jointly optimal rule is also presented.

2.5.1) Effect of joint cost function on the coupling policy

Throughout this subsection we will assume the cost assignment given below as our basis for the joint cost function.

$$
\begin{aligned}
C(i,\theta^i : j,\theta_e^j) &= C_d(i,j) + C_e(i,\overline{\theta}^i : j,\overline{\theta}_e^j) \quad \text{for } i \neq j \text{ or } i = j = 0 \\
C(i,\theta^i : j,\theta_e^i) &= C_d(i,i) + C_e(i,\overline{\theta}^i : i,\overline{\theta}_e^i) + C_e(i,\tilde{\theta}^i : i,\tilde{\theta}_e^i) \quad \text{for } i \neq 0
\end{aligned}
\tag{62}
$$

where $\overline{\theta}^i \in \overline{\Theta}$, $\overline{\theta}_e^j \in \overline{\Theta}$, $\tilde{\theta}^i \in \tilde{\Theta}_i$, and $\tilde{\theta}_e^j \in \tilde{\Theta}_j$.

In (62) $C_d(i,j)$ is a cost due to misdetection and the remaining terms due to estimation error. The present cost assignment emphasizes the operational difference between the two decision processes of detection and estimation [1] . Since it is quite likely that action corresponding to detection is different from the action corresponding to estimation, it seems appropriate to anatomize the selection of cost functions as indicated in (62). (62) is a generalized version of the cost function proposed for the strong coupling by Middleton and Esposito [1] , and the difference is that (62) takes into account of different natures of parameter spaces under each hypotheses whereas Middleton restricts the parameters of interest to the common energy type parameters. In [28] Levin used a joint cost function defined by dropping the consideration of common nature of parameters, and in this case $\tilde{\theta}^i = \theta^i$ also $\tilde{\theta}_e^j = \theta_e^j$.

We now examine the implications of the jointly Bayesian approach by taking a compound binary hypotheses problem,

$$H_1 : \begin{cases} dx^1(t) = f^1(x^1(t),\theta^1)dt + g^1(x^1(t),\theta^1)dw^1(t) \\ dy(t) = h^1(t,x^1(t),\theta^1)dt + \sqrt{r}\,dv^1(t) \end{cases}$$

$$H_0 : \quad dy(t) = \sqrt{r}\,dv^0(t)\,.$$

Suppose the common energy type parameter to estimate is the signal envelope. Recall the definition of h_t as $1_{(*=1)}\, h(t,x^1(t),\theta^1)$. Thus the parameter of interest under H_0 is h_t $(\equiv 0)$ and the parameters under H_1 are h_t and θ^1. The jointly optimal rule (21) for the above problem is

$$\begin{cases} \text{Decide } (H_1,\theta_e^{1*}) \ \text{ if } \ A_1(\theta_e^{1*},y^t) \le A_0(\theta_e^{0*},y^t) \\ \text{otherwise, decide } (H_0,\theta_e^{0*}) \\ \theta_e^{j*} = Arg(\underset{\Theta_j}{Min}\, A_j(\theta_e^j,y^t)) \end{cases} \tag{63}$$

, where $A_j(\theta_e^j,y^t)$ is defined by (19).

Specifying the meaning of (62) as it is applied to the present case,

a) $C_e(0,h_t^0\,;1,h_e^1) = C_e(0,0\,;1,h_e^1)$: estimation error when H_0 is true and decision is false.

b) $C_e(0,0\,;0,h_e^0)$: estimation error when H_0 is true and decision is correct.

c) $C_e(1,h_t^1\,;0,h_e^0)$: estimation error when H_1 is true and decision is false.

d) $C_e(1,h_t^1\,;1,h_e^1) + C_e(1;\theta^1,\theta_e^1)$: estimation error when H_1 is true and decision is correct.

The above cost assignment can also be understood as a parametrization of classical cost functions (for detection only), according to the signal envelope and its estimate. That an estimate appears under the decision of H_0 reflects the consideration that an estimate need not always be rejected by a decision of H_0 (i.e. there is a positive probability of error in detection) . From the general cost functions given we will go through special cases , which yield various levels of complexities and coupling policies in receiver implementation. Let $\Lambda_t \equiv E_0(\frac{dP^1}{dP^0} \mid y^t)$. From (19) we obtain,

$$
\begin{aligned}
&A_0(h_e^0, y^t) \\
&= p(0) [C_d(0,0) + C_e(0,0;0,h_e^0)] \\
&+ p(1) \Lambda_t [C_d(1,0) + \int C_e(1,h_t^1 ; 0,h_e^0) P(t,1;x^1,\theta^1) dx^1 d\theta^1] \\
&A_1(h_e^1, y^t) \\
&= p(0) [C_d(0,1) + C_e(0,0 ; 1,h_e^1)] \\
&+ p(1) \Lambda_t [C_d(1,1) + \int C_e(1,h_t^1 ; 1,h_e^1) P(t,1;x^1,\theta^1) dx^1 d\theta^1 \\
&+ \int C_e(1,\theta^1;1,\theta_e^1) P(t,1;\theta^1) d\theta^1]
\end{aligned}
\tag{64}
$$

Then

$$
\begin{aligned}
&h_e^{0*}(y^t) \\
&= Arg (\underset{R^K}{Min} (p(0) C_e(0,0;0,h_e^0) + p(1) \Lambda_t \int C_e(1,h_t^1 ;0,h_e^0) P(t,1;x^1,\theta^1) dx^1 d\theta^1)) \\
&h_e^{1*}(y^t) \\
&= Arg (\underset{R^K}{Min} (p(0) C_e(0,0;1,h_e^1) + p(1) \Lambda_t \int C_e(1,h_t^1;1,h_e^1) P(t,1;x^1,\theta^1) dx^1 d\theta^1)) \\
&\theta_e^{1*}(y^t) = Arg (\underset{\Theta_1}{Min} \int C_e(1,\theta^1;1,\theta_e^1) P(t,1;\theta^1) d\theta^1)
\end{aligned}
\tag{65}
$$

The jointly optimal decision rule , written in an usual form , is

$$
\Lambda_t \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p(0)}{p(1)} \frac{A(0,1) - A(0,0)}{A(1,0) - A(1,1)} \tag{66}
$$

, if $(A(1,0) - A(1,1)) \geq 0$. Otherwise choice of H_1,H_0 is reversed. In (66),

$$
\begin{aligned}
A(0,0) &= C_d(0,0) + C_e(0,0;0,h_e^{0*}) \\
A(0,1) &= C_d(0,1) + C_e(0,0;1,h_e^{1*}) \\
A(1,0) &= C_d(1,0) + \int C_e(1,h_t^1;0,h_e^{0*}) P(t,1;x^1,\theta^1) dx^1 d\theta^1
\end{aligned}
$$

, and

$$A(1,1) = C_d(1,1) + \int C_e(1,h_t^{\,1};1,h_e^{\,1*})\, P(t,1;x^1,\theta^1)\, dx^1\, d\theta^1$$
$$+ \int C_e(\,1,\theta^1;1,\theta_e^{1*}\,)\, P(t,1;\theta^1)\, d\theta^1$$

That the denominator of right side of (66) can be negative depending on observations does not imply any contradiction within the frame of the jointly Bayesian approach, since (66) is a version of (63). However this suggests a possible defect in using the joint cost function of the form (62) in general. From a practical point oriented to detection the following inequalities are usually desired ,

$$A(0,1) \geq A(0,0) \quad , \quad A(1,0) \geq A(1,1).$$

But it is not guaranteed to retain the above relations a.s. in observations ,since now $A(i,j)$ depends on estimates , and this implies that there can be a significant unbalance between detection and estimation operations unless the joint cost functions are properly weighted. We note that when H_1 is true, the estimation error associated with a temporary decision $u_d = 1$ will decrease as the observation interval increases, and thereby the second part of the above inequalities tends to be satisfied.

We indicated a pathological effect caused by a general coupling policy between detection and estimation. We heuristically examine the jointly optimal decision rule from a detection point of view. Suppose the hypothesis H_0 is true. Then the a posteriori estimation loss associated with a decision of H_1 tends to be relatively larger , and the threshold increases and consequently is favorable for a decision of H_0. Suppose H_1 is true. Then the a posteriori estimation error associated with a decision of H_1 tends to be relatively smaller, and the threshold decreases and consequently is favorable for a selection of H_1.

We now look at the nature of estimates (65) obtained under the jointly Bayesian approach. For this purpose we use the following quadratic cost functions,

$$\begin{cases} C_e(0,0;0,h_e^0) = C_{00}\cdot(h_e^0)^2 \\ C_e(0,0;1,h_e^1) = C_{01}\cdot(h_e^1)^2 \\ C_e(1,h_t^1;0,h_e^0) = C_{10}\cdot(h_t^1 - h_e^0)^2 . \\ C_e(1,h_t^1;1,h_e^1) = C_{11}\cdot(h_t^1 - h_e^1)^2 \\ C_e(1,\theta^1;1,\theta_e^1) = C_\theta(\theta^1 - \theta_e^1)^2 \end{cases} \tag{67}$$

Then,

$$\begin{aligned} h_e^{0*} &= C_{10}\Lambda_t\ \hat{h}_t^1 / [\ \frac{p(0)}{p(1)} C_{00} + C_{10}\Lambda_t\ ] \\ h_e^{1*} &= C_{11}\Lambda_t\ \hat{h}_t^1 / [\ \frac{p(0)}{p(1)} C_{01} + C_{11}\Lambda_t\ ] \\ \theta_e^{1*} &= E^1(\theta^1 \mid y^t) \end{aligned} \tag{68}$$

,where $\hat{h}_t^1 = \int h(t,x^1,\theta^1)\,P(t,1;x^1,\theta^1)\,dx^1\,d\theta^1$, conditional mean of the signal envelope under the assumption that $P(H_1) = 1$. The form of estimate , h_e^{j*}, is due to the fact that h_t has an unconditional density given by

$$p(0)\,\delta(h_t) + p(1)\,p^1(h_t)$$

, where $p^1(h_t)$ is a conditional density of h_t at t given H_1 .

When $C_{00} = C_{01} = C_{10} = C_{11} = k$ (68) reduces to

$$h_e^{0*} = h_e^{1*} = h_e^* = \frac{p(1)\Lambda_t\ \hat{h}_t^1}{p(0) + p(1)\Lambda_t}$$

,which can also be put into a more revealing form

$$h_e^* = P(H_1 \mid y^t)\,\hat{h}_t^1 \tag{69}$$

The estimate (69) reflects the fact that $0 < P(H_1) < 1$, and is a convex combination of $\hat{h}_t^1$,and 0 by the corresponding a posteriori probabilities. The estimate θ_e^{1*} is simply a Bayes conditional conditional mean under the assumption that $P(H_1) = 1$. The jointly Bayesian approach does not exert any influence on θ_e^{1*} except that θ_e^{1*} contributes to the determination of threshold and then is validated or invalidated depending on the results of detection.

Though Levin [28] included parameters like θ_e^{1*} in the jointly Bayesian approach, actually their use may be controversial. Thus if we restrict the jointly Bayesian approach to the common energy type parameters only, the jointly optimal decision rule is given by (66) with $A(1,1)$ being replaced by

$$A(1,1) = C_d(1,1) + \int C_e(1,h_t^1;1,h_e^{1*}) P(t,1;x^1,\theta^1)\, dx^1\, d\theta^1 .$$

Under the cost functions (67) and the assumption on C_{ij} , the jointly optimal estimate is independent of the detection , and is given by (69). Decision rule under the same condition is

$$\Lambda_t \gtrless \frac{p(1)}{p(0)} \frac{C_d(0,1) + k\, (h_e^*)^2 - C_d(0,0) - k\, (h_e^*)^2}{A(1,0) - A(1,1)}$$

,where

$$A(1,0) = C_d(1,0) + \int k\, (h_t^1 - h_e^*)^2 P(t,1;x^1)\, dx^1$$
$$A(1,1) = C_d(1,1) + \int k\, (h_t^1 - h_e^*)^2 P(t,1;x^1)\, dx^1$$

And thus the jointly optimal decision rule is an usual optimal decision rule given by

$$\Lambda_t > \frac{p(1)}{p(1)} \frac{C_d(0,1) - C_d(0,0)}{C_d(1,0) - C_d(1,1)} \tag{70}$$

According to (69) and (70), there is no coupling between the detection and estimation operations , which is termed as no coupling policy in [1] . Generalizing the above, we set

$$\begin{aligned} C_e(0,0;0,h_e^0) &= C_e(0,h_e^0) \\ C_e(0,0;1,h_e^1) &= C_e(0,h_e^1) \\ C_e(1,h_t^1;0,h_e^0) &= C_e(\, h_t^1,h_e^0\,) \\ C_e(1,h_t^1;1,h_e^1) &= C_e(\, h_t^1,h_e^1\,) . \end{aligned} \tag{71}$$

Then we have the cost function of the form

$$C(i,h_t^i\, ;\, j,h_e^j) = C_d(i,j) + C_e(\, h_t^i,h_e^j\,)$$

, and

$$h_e^{j*} = Arg\ (\, Min\ [\, p(0)\, C_e(0,h_e^j) + p(1)\, \Lambda_t \int C_e(h_t^1,h_e^j)\, P(t,1;x^1,\theta^1)\, dx^1\, d\theta^1\,]\,) .$$

Thus the estimates h_e^{1*} and h_e^{0*} are independent of decisions and the detection is given by the usual optimal detection rule (70). Schematics for the no coupling policy is depicted in Fig-2 .

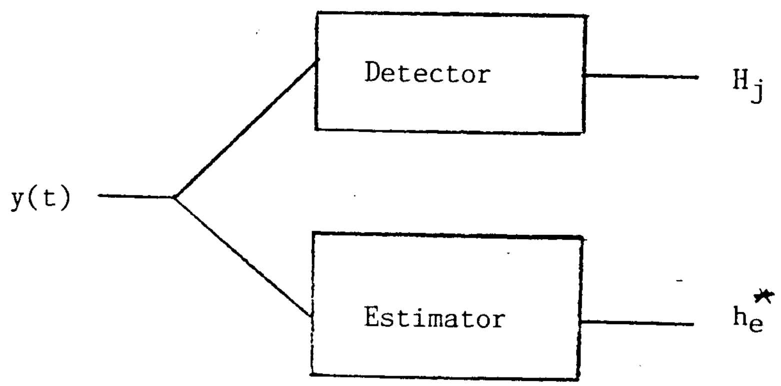

Fig 2

If modifying $C_e(1,h_t^1;0,h_e^0)$ of (71) by

$$C_e(1,h_t^1;0,h_e^0) = C_e(h_t^1,h_e^0) + f(h_t^1), \tag{72}$$

then,

$$\begin{aligned} &A_0(h_e^0,y^t) \\ &= p(0)\,[\,C_d(0,0) + C_e(0,h_e^0)\,] \\ &\quad + p(1)\Lambda_t\,[\,C_d(1,0) + \int (\,C_e(h_t^1,h_e^0) + f(h_t^1)\,)\,P(t,1;x^1,\theta^1)\,dx^1\,d\theta^1\,] \end{aligned}$$

$$\begin{aligned} &A_1(h_e^1,y^t) \\ &= p(0)\,[\,C_d(0,1) + C_e(0,h_e^1)\,] \\ &\quad + p(1)\Lambda_t\,[\,C_d(1,1) + \int C_e(h_t^1,h_e^1)\,P(t,1;x^1,\theta^1)\,dx^1\,d\theta^1\,] \end{aligned}$$

Obviously estimation is independent of decision and coincides with the one for (71). The jointly optimal decision rule is given by

$$\Lambda_t \gtrless \frac{p(1)}{p(0)}\,\frac{C_d(0,1) - C_d(0,0)}{C_d(1,0) + \int f(h_t^1)\,P(t,1;x^1,\theta^1)\,dx^1\,d\theta^1 - C_d(1,1).} \tag{73}$$

, and the detector is independent of the estimator. The effective cost functions for the decision rule (73) are $C_d(0,0)$, $C_d(0,1)$, $C_d(1,0) + f(h_t)$, $C_d(1,1)$. And the effective cost function for the estimation rule is $C_e(h_t, h_e)$.

If we take a policy of decision rejection(i.e. invalidate the estimate when H_0 is decided), the coupling between the detector and estimator is termed as the weak coupling [1] [6].

Finally we consider the strong coupling policy , a version of general coupling where no estimation operation is necessary when H_0 is decided [1] [2]. This is done by setting

$$C(i,\theta^i ; 0,\theta_e^0) = C(i,\theta^i ; 0). \tag{74}$$

Thus the conditional distribution of θ_e^0 under $u_d = 0$ or its value are not matters of concern.

A schematic for the strong coupling policy is depicted in Fig-3.

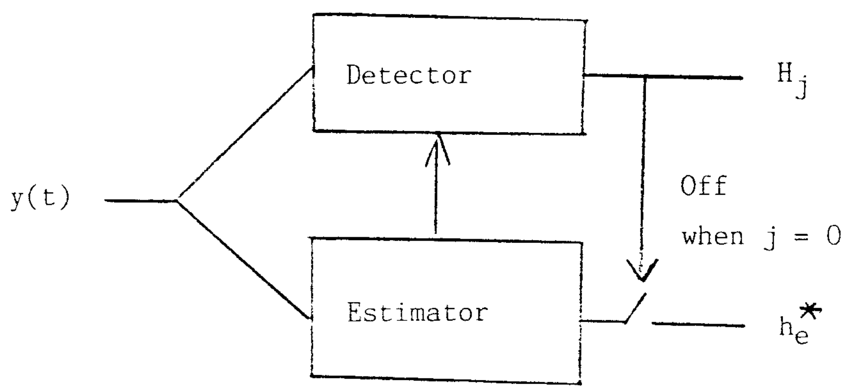

Fig.3

we will take a different look at the observation model (1) , and accordingly θ_e^{1*} will also be modified properly.

We now consider two special cases of jointly optimal detection and estimation under a strong coupling policy such that $C(i,\theta^i\ ;0,\theta_e^0) = C_d(i\ ,0)$. Then from (19) ,

$$A_0(y^t) = \sum_{i=1}^{M} p(i)\, C(i\ ,0)\, \Lambda_t^i + p(0)\, C(0\ ,0)$$

, where $\Lambda_t^i = E^0(\frac{dP^i}{dP^0} \mid y^t)$.

Throughout the remainder of this subsection $\overline{\Theta} = R^{L+K}$, where L is the number of common energy type system parameters and K the dimension of the signal envelope.

1.$C(i,\theta^i\ ; j,\theta_e^j) = C_d(i,j) + (\overline{\theta}^i - \overline{\theta}_e^j)^T D_{1i}\, (\overline{\theta}^i - \overline{\theta}_e^j)$ for $i \neq j$ or $i=j=0$
$C(i,\theta^i\ ; i,\theta_e^i) = C_d(i,i) + (\overline{\theta}^i - \overline{\theta}_e^i)^T D_{1i}\, (\overline{\theta}^i - \overline{\theta}_e^i) + (\tilde{\theta}^i - \tilde{\theta}_e^i)^T D_{2i}\, (\tilde{\theta}^i - \tilde{\theta}_e^i)$ for $i \neq 0$
,where D_{1i} and D_{2i} are positive definite matrices.

By letting the gradient (with respect to θ_e^j) of $A_j(\theta_e^j, y^t) = 0$ for $j \neq 0$, we obtain

$$h_e^{j*}(y^t) = \sum_{i=1}^{M} \frac{p(i)\, \Lambda_t^i\, E^i(h_t^i \mid y^t)}{p(0) + \sum_{k=1}^{M} p(k)\, \Lambda_t^k}$$

$$\overline{\theta}_e^{j*}(y^t) = \sum_{i=1}^{M} \frac{p(i)\, \Lambda_t^i\, E^i(\theta^i \mid y^t)}{p(0) + \sum_{k=1}^{M} p(k)\, \Lambda_t^k}$$

$$\tilde{\theta}_e^{j*}(y^t) = E^j(\tilde{\theta}^{j*} \mid y^t)$$

Since

$$P(H_i \mid y^t) = \frac{p(i)\, \Lambda_t^i}{p(0) + \sum_{k=1}^{M} p(k)\, \Lambda_t^k},$$

$\overline{\theta}_e^{j*}$ can be expressed as

$$\overline{\theta}_e^{j*} = \sum_{i=1}^{M} E^i(\theta^i \mid y^t)\, P(H_i \mid y^t).$$

For the cost function at hand, the estimates $\overline{\theta}_e^{j*}$ and h_e^{j*} are independent of j (

independent of decision except being validated) due to the independence of D_{1i} on j , and are given by the convex combination of each conditional means obtained under the assumption that $P(H_i) = 1$. When a signal involved has a low signal to noise ratio, this character will be a significant improving factor in the sense of the ensemble risk associated with estimation, compared to the classical conditional mean estimate validated by the decision of some H_i, which is not really the conditional mean of the parameter concerned in the presence of hypotheses uncertainty.

The jointly optimal decision and estimation is based on (21) and the estimates given above.

2. $C(i,\theta^i\ ;j,\theta_e^j) = C_d(i,j) + b_{1i}\ (\ c_{1i} - \delta(\overline{\theta}^i - \overline{\theta}_e^j)\ )$ for $i \neq j$ or $i = j = 0$

$C(i,\theta^i\ ;i,\theta_e^i) = C_d(i,i) + b_{1i}\ (\ c_{1i} - \delta(\overline{\theta}^i - \overline{\theta}_e^i)\ ) + b_{2i}(c_{2i} - \delta(\tilde{\theta}^i - \tilde{\theta}_e^i)\ )$ for $i \neq 0$

, where $\delta(\cdot)$ is a Dirac delta function ,and b_{ij}, c_{ij} are cost controlling factors . The prototype of the above cost function was proposed in [1] [2] , and called as a simple cost function. We obtain the following estimates.

$$\overline{\theta}_e^{j*} = Arg\ (\ \underset{R^{L+K}}{Max} \sum_{i=1}^{M} b_i\ p(i)\, \Lambda_t^i\, P(t,i\,;\overline{\theta}_e^j, h_e^j)\ )$$

$$\tilde{\theta}_e^{j*} = Arg\,(\ \underset{\Theta_j}{Max}\, P(t,j\,;\tilde{\theta}_e^j)\ \ )$$

, where $P(t,i\ ;\cdot,\cdot)$ is the a posteriori density of $(\overline{\theta}^i\ , h_t^i)$ obtained from $P(t,i\ ;\overline{\theta}^i, x^i\ )$. The estimate $\overline{\theta}_e^{j*}$ can be interpreted as a generalized MAP estimate.

2.5.2) Nonlinear estimation approach

In this subsection we view the original observation models from a different perspective. For convenience we rewrite the model (1).

$$dx^i(t) = f^i(x^i(t),\theta^i)\,dt + g^i(x^i(t),\theta^i)\,dw^i(t)\quad ,\ x^i(0) = x_0^i$$

$$dy(t) = h^i(t,x^i(t),\theta^i)\,dt + \sqrt{r}\,dv^i(t)\quad ,\ y(0) = 0\ ,0 \le t \le T \qquad (1)$$

We recall the remark (2-f) of section (2.1) following the observation model (1) , where , under H_i , all the coordinate functions are assumed to have arbitrary probability distribution functions except $(\theta^i, x_0^i, w^i, y)$. The construction is based on the point that (θ^j, x_0^j, w^j) for $j \neq i$ are irrelevant to the observation model defined under H_i.

Now we specialize remark (2-f) on the observation model (1) by imposing the following conditions under P^i .

a) (θ^j, x_0^j, w^j) are independent with each other.

b) (θ^j, x_0^j, w^j) for $j \neq i$ are independent of y , and have the same distribution as under P^j.

This approach to the observation model leads to the following description.

$$\begin{bmatrix} dx^1(t) \\ \cdot \\ \cdot \\ dx^M(t) \\ d\theta^1(t) \\ \cdot \\ \cdot \\ d\theta^M(t) \\ d\,1_{(u=1)}(t) \\ \cdot \\ \cdot \\ d\,1_{(u=M)}(t) \end{bmatrix} = \begin{bmatrix} f^1(x^1(t),\theta^1) \\ \cdot \\ \cdot \\ f^M(x^M(t),\theta^M) \\ 0 \\ \cdot \\ \cdot \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} dt + \begin{bmatrix} g^1(x^1(t),\theta^1) & 0 & \cdot & 0 \\ \cdot & & \cdot\ \cdot & \cdot \\ \cdot & & \cdot\ \cdot & \cdot \\ 0 & & \cdot\ \cdot & g^M(x^M(t),\theta^M) \\ 0 & & \cdot\ \cdot & 0 \\ \cdot & & \cdot\ \cdot & \cdot \\ \cdot & & \cdot\ \cdot & \cdot \\ 0 & & \cdot\ \cdot & 0 \\ 0 & & \cdot\ \cdot & 0 \\ \cdot & & \cdot\ \cdot & \cdot \\ \cdot & & \cdot\ \cdot & \cdot \\ 0 & & \cdot\ \cdot & 0 \end{bmatrix} \begin{bmatrix} dw^1(t) \\ \cdot \\ \cdot \\ dw^M(t) \end{bmatrix}$$

$$dy(t) = \sum_{i=1}^{M} h^i(t, x^i(t), \theta^i)\, 1_{(u=i)}\, dt + \sqrt{r}\; dv(t) \quad y(0) = 0, \; 0 \leq t \leq T \qquad (75)$$

,where $1_{(u=i)}$ is an indicator random variable taking 1 when $u = i$.

Lainiotis [7], Kadota [29] takes the above unified nonlinear estimation approach to detection and estimation for a compound binary hypotheses problem , and (75) is a straightforward extension to multiple hypotheses. According to the above formulation (θ^i, w^i, x^i) for all i are defined uniquely and exist independently of hypotheses. Define the augmented state vector by

$$x_a(t) = [\; x^i(t)^T, \theta^{i\,T}, 1_{(u=i)} \quad , i = 1,,M \;]^T \quad .$$

Let $F_t(x_a) = \sigma(x_a(s) , 0 \le s \le t)$,and let $F_t = \sigma(F_t(x_a) \bigcup F_t(y))$ for the formulation (75). Probability measure P is defined in accordance with the explanation so far given. And $v(t) \equiv \frac{1}{\sqrt{r}} \left\{ y(t) - \int_0^t h_s \, ds \right\}$ is a (F_t, P) standard Wiener process , where $h_t = \sum_{i=1}^{M} 1_{(u=i)} h^i(t, x^i(t), \theta^i)$.

Following [7] , the joint detection and estimation is now formulated as an unified nonlinear estimation problem , where the following quadratic loss function is minimized at each time t ,

$$E \, [(z_a(t) - \hat{a}(t))^T D (z_a(t) - \hat{a}(t))]$$

where $z_a(t) = [h_t , \theta^{i\,T} , 1_{(u=i)} , i = 1,,,M]^T$. As usual D is a positive definite matrix , and $\hat{a}(t)$ is an estimator. Later we will treat the present nonlinear estimation formulation as a special case of the previous jointly Bayesian approach under the additional conditions (a) and (b) above. Before proceeding consider the quadratic cost function of the form , $E \, (1_{(u=i)} - \hat{a}(t))^2$,where $\hat{a}$ is an arbitrary estimator. We can identify the a posteriori mean of $1_{(u=i)}$ as MMSE estimate through the following standard way.

$$\int (1_{(u=i)} - \hat{a}(y^t))^2 dP(1_{(u=i)} \mid y^t)$$
$$= \sum_{1_{(u=i)}=0}^{1} (1_{(u=i)} - \hat{a}(y^t))^2 P(1_{(u=i)} \mid y^t)$$
$$= \sum (1_{(u=i)} - P(u = i \mid y^t) + P(u = i \mid y^t) - \hat{a}(y^t))^2 P(1_{(u=i)} \mid y^t)$$
$$= \sum (1_{(u=i)} - P(u = i \mid y^t))^2 P(1_{(u=i)} \mid y^t) + \sum (P(u = i \mid y^t) - \hat{a}(y^t))^2 P(1_{(u=i)} \mid y^t)$$
$$\ge \sum (1_{(u=i)} - P(u = i \mid y^t))^2 P(1_{(u=i)} \mid y^t)$$

, with equality if and only if $\hat{a}(y^t) = P(u = i \mid y^t)$.

Due to the conditions (a) and (b),

$$p(x^j(t), \theta^j , j = 1 ,, M \mid y^t , u = i) P(u = i \mid y^t)$$
$$= p(x^i(t), \theta^i \mid y^t , u = i) \cdot \prod_{j \ne i} p(x^j(t), \theta^j) \cdot P(u = i \mid y^t)$$

We may use Girsanov transformation of P on F_T and nonlinear filtering formula to derive the evolution equation for $p(x^i(t), \theta^i \mid y^t , u = i)$. Here we only refer to

(34) of section (2-3) that $\rho(t,i;x^i,\theta^i)$ satisfies.

We state directly the MMSE estimates below.

$$E(1_{(u=i)} \mid y^t) = \frac{\Lambda_t^i p(i)}{p(0) + \sum_{k=1}^{M} p(k)\Lambda_t^k}$$

$$\begin{aligned} &E(h_t \mid y^t) \\ &= E[E(h_t \mid F(u), F_t(y)) \mid y^t] \\ &= \sum_{i=1}^{M} E^i(h_t \mid y^t) P(u = i \mid y^t) \end{aligned}$$

$$\begin{aligned} &E(\theta^i \mid y^t) \\ &= E[E(\theta^i \mid F_{(u)}, F_t(y)) \mid y^t] \\ &= \sum_{j=0}^{M} E^j(\theta^i \mid y^t) P(u = j \mid y^t) \\ &= E(\theta^i) P(u \neq i \mid y^t) + E^i(\theta^i \mid y^t) P(u = i \mid y^t) \\ &i = 1,, M \end{aligned} \tag{76}$$

,where $E^j(\theta^i \mid y^t)$ is the conditional mean of θ^i given y^t when $u = j$ with similar interpretation for the h_t part . Binary version of (76) was proposed in [7]. The estimate $E(\theta^i \mid y^t)$ is actually a consequence of the mentioned view point on the hypotheses detection and parameter estimation, and we think that it may be interpreted as a conservative estimate ,in practice , given by the convex combination of the a posteriori estimate with the a priori estimate.

We now look at (76) from the previous jointly Bayesian approach. If detection is based on the MAP estimate, the estimates of (76) is the result of the following joint cost function,

$$(1 - \delta_{ij}) + C_{e1}(\theta, \theta_e) + C_{e2}(h_t; h_e)$$

, where δ_{ij} is a Kronecker delta , $\theta = [\theta^{1T}, \cdots, \theta^{MT}]^T$, and $C_{ei}(\cdot,\cdot)$ a quadratic cost function. Obviously $(\theta_e^{j*}, h_e^{j*})$ is independent of j , and there is no coupling between detector and estimator. Notice that we now have only common parameters due to the conditions (a) and (b) , and also due to the estimation of all θ^j s .

2.5.3) Extension to multiple observations

The one shot receiver of Fig.1 has a straightforward extension to the case of multiple observations , where each hypotheses of (1) are active during the intervals, $nT' \leq t \leq nT' + T$ ($T' \geq T$, $n = 0,1,2,\cdots$). Here T' is a repetition interval of the active observations, and sequential decisions and estimations are made during each active observation interval .

We take the following adaptive version of (21) . The set of the priors for Fig.1 for each new period of observations are to be updated by using the results of previous observations. Let the set of the priors at $t = nT'$ be denoted by $p(i;nT')$, $p^i(\theta\,;nT')$, and $p^i(x_{nT'} \mid \theta)$ respectively. The unnormalized density $\rho(t,i;x^i,\theta^i)$ at $t \in [\,nT' \,, nT' + T\,]$ is evaluated by using (34) and an observation $y_{nT'}^t$ with the above set of the priors. A version of the jointly optimal rule is set by (21) with y^t being replaced by $y_{nT'}^t$. Hence the extended rule for multiple observations is partitioned into a fixed nonadaptive part and and an adaptive part. The adaptive part updates the set of the priors in the following way.

$$
\begin{aligned}
p(0\,;(n+1)\,T') &= \frac{p(0)}{p(0) + \sum_{j=1}^{M} p(j) \int\!\!\int \rho(nT' + T, j\,; x^j, \theta^j)\, dx^j\, d\theta^j} \\
p(i\,;(n+1)\,T') &= \frac{p(i) \int\!\!\int \rho(nT' + T, i\,; x^i, \theta^i)\, dx^i\, d\theta^i}{p(0) + \sum_{j=1}^{M} p(j) \int\!\!\int \rho(nT' + T, j\,; x^j, \theta^j)\, dx^j\, d\theta^j} \qquad (77) \\
p^i(\theta;(n+1)\,T') &= \frac{\rho(nT' + T, i\,;\theta^i)}{\int \rho(nT' + T, i\,;\theta^i)\, d\theta^i}
\end{aligned}
$$

For $p^i(x_{(n+1)\,T'} \mid \theta)$, we consider two possible cases. When the diffusions $x^i(t)$ exist for the entire observation interval and only the active observations are made intermittently, one can use the extrapolation equation given by Fokker- Plank equation,

$$\frac{\partial}{\partial t} p^i(t, x \mid \theta) = L^*_{\theta^i} \; p^i(t, x \mid \theta) \ , \ nT' + T \le t \le (n+1) \, T'$$

$$p^i(x_{nT' + T} \mid \theta) = \frac{\rho(t, i; x^i, \theta^i) \cdot \int \rho(t, i; \theta^i) \, d\theta^i}{\int\int \rho(t, i; x^i, \theta^i) \, dx^i \, d\theta^i \cdot \rho(t, i; \theta^i)} \Big|_{t = nT' + T}$$

In this case the present rule is truly a jointly optimal rule at each $t \in [\, nT' \ , nT' + T \,]$ for all n . The observation for the case is represented by ,

$$dy(t) = 1_{(t \in [\, nT', nT' + T \,])} \, h^i(t, x^i, \theta^i) \, dt + \sqrt{r} \ dv^i(t), \ \ n = 0, 1 \cdots$$

under each H_i .

When a diffusion models (1) are valid only for the active observation intervals, we choose to fix $p^i(x_{(n+1)T'} \mid \theta)$ by $p^i(x_0 \mid \theta)$ for all n , provided that the diffusions yield stationary densities.

3. Performance evaluation

Evaluation of the joint Bayes cost for the simultaneously optimal rule is not feasible to carry out, due to complexities of our observation model, and also due to the coupling between the jointly optimal detector and estimator. As shown in the previous chapter the jointly optimal Bayesian system seems to have a growing significance when the a posteriori accuracy becomes low, and to be asymptotically equivalent to other systems. This necessitates performance analysis for the finite observation interval or under low signal to noise ratio, where the coupling between the detector and estimator plays an important role and at the same time makes general performance analysis difficult.

In this chapter we focus on performance bounds in simple binary hypotheses detection ,and also on performance bounds of parameter estimation under no hypothesis uncertainty.

3.1) Performance bounds in binary nongaussian detection problems

Suppose we are given a some measurable space (Ω,F) and two probability measures, P^0 and P^1 , are defined on it . Also we are given a (right-continuous) increasing sub σ -fields of F , $\left\{F_t\right\}$, with respect to which all the processes involved in a binary hypotheses detection problems are adapted. We assume that $F=\sigma(\bigcup_{t\geq 0} F_t)$ where time index t may be either real or integer. As before let $F_t(y)$ be the σ -field generated by the observation up to a current time. For convenience we assume that P^0 and P^1 are mutually absolutely continuous on F_t at each $t\in[0,\infty)$ though we only need their restrictions to $F_t(y)$ are mutually absolutely continuous. F is assumed to be complete under P^0 with the corresponding completion of F_0 and $F_0(y)$, and thus F_t , $F_t(y)$ are complete under either probability measures for all $t\in[0,\infty)$ due to the equivalence between P^0 and P^1 at each t .

The optimal test statistic for detection problem is the likelihood ratio defined by

$$\Lambda_t = E^0\left(\frac{dP^1}{dP^0} \mid F_t(y) \right) = \frac{dP^1}{dP^0} \Big|_{F_t(y)}$$

,and for a threshold γ likelihood ratio test is

$$\Lambda_t \gtrless \gamma$$

Henceforth we denote the restrictions of P^i to $F_t(y)$ and $F(y)\,(\equiv \sigma(\bigcup_{t\geq 0} F_t(y)))$ by $P_Y^{i,t}$ and P_Y^i respectively .

Chernoff bound has been frequently used to bound the probabilities of errors , first and second kind , in hypotheses detection problems [37]. When the likelihood ratio test is used for a binary hypotheses problem the probability of false alarm is ,in general, bounded by

$$\begin{aligned} &P_F(t) \\ &= \text{Prob(accept } H_1 \mid H_0 \text{ is true)}\,|_t \\ &\leq \gamma^{-s}\, E^0(\Lambda_t^s) \end{aligned} \tag{1}$$

, where the real variable s is such that $E^0(\Lambda_t^s)$ exists. For $0 < s < 1$, $E^0(\Lambda_t^s)$ exists. Therefore the tightest bound on $s \in (0,1)$ is obtained by

$$P_F(t) \leq \operatorname*{Inf}_{0<s<1} \gamma^{-s}\, E^0(\Lambda_t^s)$$

The Chernoff bound on the probability of miss , P_M , is likewise given by

$$\begin{aligned} &P_M(t) = \text{Prob(accept } H_0 \mid H_1 \text{ is true)}\,|_t \\ &\leq \operatorname*{Inf}_{-1<s<0} \gamma^{-s}\, E^1(\Lambda_t^s) \end{aligned} \tag{2}$$

Also we note the following frequently used relation

$$E^1(\Lambda_t^s) = E^0(\Lambda_t^{s+1}). \tag{3}$$

Hence if we can upperbound P_F by computing $E^0(\Lambda_t^{s+1})$,then we know $E^1(\Lambda_t^s)$ immediately.

When we use a suboptimal statistic T_t instead of Λ_t , in the comparison with the threshold γ , the relations (1) and (2) remain valid , however (3) is no longer valid. Before proceeding with the evaluation of Chernoff bounds we examine the limit behavior of $E^0(\Lambda_t^s)$ as $t \to \infty$.

Proposition 3.1

$\lim_{t\to\infty} E^0(\Lambda_t^s) = 0$ for $0 < s < 1$ if only if $\Lambda_\infty = 0$ P^0 a.s.

Proof: First we note that Λ_∞ defined by

$$\Lambda_\infty = \lim_{t\to\infty} \Lambda_t$$

exists, since Λ_t is a $(F_t(y), P^0)$ (right continuous) nonnegative martingale and also due to the corresponding supermartingale convergence theorem [8] . By Jensen's inequality on concave function we have

$$E^0(\Lambda_t^s) \leq [E^0(\Lambda_t)]^s \quad , \text{for } 0 < s < 1 .$$

Consider an arbitrary sequence of $\left\{ t_i \right\}$, $t_i \uparrow \infty$ as $i \to \infty$. From the above inequality there is $\epsilon > 0$ at each s such that

$$\underset{i}{Sup}\, E^0(\Lambda_{t_i}^s)^{1+\epsilon} \leq 1 \quad , \text{for } 0 < s < 1 . \tag{4}$$

(4) is a sufficient condition for the uniform integrability of $\Lambda_{t_i}^s$ (e.g. [31] , pp 188) . Since $\left\{ \Lambda_{t_i}^s \right\}_{i \geq 1}$ is an uniformly integrable family and $\lim_{i\to\infty} \Lambda_{t_i}^s = 0$ a.s. by hypothesis, $\Lambda_{t_i}^s \to 0$ in L^1 sense also (e.g. [31] , pp 186) . Since the above is true for all the sequences $\left\{ t_i \right\}$ such that $t_i \uparrow \infty$,

$$\lim_{t\to\infty} E^0(\Lambda_t^s) = 0 \quad , 0 < s \leq 1$$

Thus Λ_t^s is a potential for $0 < s < 1$.

Converse is also true since then Λ_t^s is a potential for $0 < s < 1$.

□

And in the above case we have

$$P_F(\infty) = \lim_{t\to\infty} P_F(t) \leq \underset{s}{Min}\ \gamma^{-s} \lim_{t\to\infty} E^0(\Lambda_t^s) = 0 .$$

Also for the probability of miss,

$$P_M(\infty) = \lim_{t \to \infty} P_M(t)$$
$$\leq \min_{s \in (-1,0)} \gamma^{-s} \lim_{t \to \infty} E^1(\Lambda_t^{\ s})$$
$$= \min_{s} \gamma^{-s} \lim_{t \to \infty} E^0(\Lambda_t^{s+1}) = 0$$

Proposition 3.1 simply means that the Chernoff bounds become sufficiently tight as $t \to \infty$ for $0 < s < 1$ if $\Lambda_\infty = 0$.

Now we consider the use of suboptimal statistic T_t. Chernoff bound $\hat{P}_F(t)$ for the suboptimal test becomes

$$\hat{P}_F(t) \leq \inf_{s>0} \gamma^{-s} E^0(T_t^s).$$

Suboptimal tests are justified , among other complicated aspects such as the asymptotic relative efficiency , in terms of asymptotical convergences to correct decisions. In analogy with proposition 3.1 we directly set a general condition to ensure such convergences .

Corollary 3.2

If (1) $T_\infty = \lim_{t \to \infty} T_t$ exists P^0 a.s., and $T_\infty = 0$ (2) there is $s' > 0$ such that $E^0(T_t^{s'}) \leq K$ uniformly in $t \geq 0$ for some finite K, then $E^0(T_t^s) \to 0$ for $s \in (0, s')$ and $\hat{P}_F(t) \to 0$ as $t \to \infty$.

[]

We note that $P_F(t) \to 0$ does not imply $P_M(t) \to 0$, since the relation (3) does not hold for a suboptimal statistic (e.g. a rule deciding exclusively H_0), and thus we need a similar condition to ensure $P_M(t) \to 0$.

Since

$$T_t \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \rightleftarrows T_t^{-1} \underset{H_1}{\overset{H_0}{\gtrless}} \gamma^{-1}$$

$$E^1(T_t^s), -1 < s < 0 \rightleftarrows E^1([T_t^{-1}]^s), 0 < s < 1,$$

we directly state

Corollary 3.3

If (1) $\lim_{t \to \infty} T_t^{-1}$ exists and is 0 P^1 a.s. (2) there is $s' > 0$ such that $E^1([\,T_t^{-1}\,]^{s'}) \leq K$ uniformly in $t \geq 0$ for some finite K, then $\hat{P}_M(t) \to 0$.

[]

Now we look at the sufficient condition to guarantee $\Lambda_\infty = 0$, P^0 a.s. . We quote the following series of basic theorems from [30] [32] .

Theorem ([30])

Let $P_Y^1 \underset{loc}{<<} P_Y^0$. Then

$$
\begin{aligned}
P_Y^1 << P_Y^0 &\rightleftarrows E^0(\Lambda_\infty) = 1 \rightleftarrows P_Y^1(\Lambda_\infty < \infty) = 1 \\
P_Y^1 \perp P_Y^0 &\rightleftarrows E^0(\Lambda_\infty) = 0 \rightleftarrows P_Y^1(\Lambda_\infty = \infty) = 1
\end{aligned}
\tag{5}
$$

[]

, where $\perp$ denotes the singular relation between P_Y^1 and P_Y^0 on $F(y)$.

In ref [30] the proof of this theorem is based on Lebesgue decomposition of P_Y^1 with respect to P_Y^0 under the condition that $P_Y^1 \underset{loc}{<<} P_Y^0$, and also it is proved that Λ_∞ exists under P_Y^1 also ([30] ,pp 493-495) .

Let t be integer ,and assume that we have a discrete sequence of observations.

Theorem (Kabanov , Lipster , Shiryayev : [30] pp 496-498)

Let $P_Y^1 \underset{loc}{<<} P_Y^0$ and let $\alpha_t = \Lambda_t \Lambda_{t-1}^{+}$, $t \geq 1$ with $\Lambda_0 = 1$.

Then (with $F_0(y) = \left\{ \phi , \Omega \right\}$ such as the case when $y(0) = 0$)

$$
\begin{aligned}
P_Y^1 << P_Y^0 &\rightleftarrows P_Y^1 \left\{ \sum_{t=1}^{\infty} [\,1 - E^0(\sqrt{\alpha_t} \mid F_{t-1}(y))\,] < \infty \right\} = 1 \\
P_Y^1 \perp P_Y^0 &\rightleftarrows P_Y^1 \left\{ \sum_{t=1}^{\infty} [\,1 - E^0(\sqrt{\alpha_t} \mid F_{t-1}(y))\,] = \infty \right\} = 1
\end{aligned}
\tag{6}
$$

[]

,where

$$\Lambda_{t-1}^{+} = \begin{cases} \Lambda_{t-1}^{-1} & \text{if } \Lambda_{t-1} \neq 0 \\ 0 & \text{if } \Lambda_{t-1} = 0 \end{cases} .$$

We note that

$$P_Y^{1,t}(\Lambda_t = 0) = \int_{\Lambda_t = 0} \Lambda_t \; dP_Y^0 = 0$$

,and also $P_Y^{0,t}(\Lambda_t = 0) = 0$ due to the local mutual absolute continuity in our case. The above theorem is called as the predictable criterion ,and leads to to the Kakutani dichotomy for independent observations. The continuous counterpart of the above theorem is now quoted from [32] for the case when $y(t)$ has a continuous trajectory (i.e. no jumps are involved) .

From the local mutual absolute continuity between P^1 and P^0, there exists a continuous local martingale M_t ($\in M_{loc}^c(F_t(y), P^0)$), $M_0 = 0$ such that

$$\Lambda_t = \exp(M_t - \frac{1}{2} < M,M >_t)$$

where $<M,M>_t$ is the quadratic variation of M_t . Or equivalently

$$d\Lambda_t = \Lambda_t \; dM_t \; ,$$

which is known as Doleans-Dade equation.

Theorem (Kabanov , Lipster , Shiryayev : [32])

Suppose $P_Y^1 \underset{loc}{<<} P_Y^0$ and let M_t be defined as above. Then

$$\begin{aligned} P_Y^1 << P_Y^0 &\rightleftarrows P_Y^1 (< M,M >_t < \infty) = 1 \\ P_Y^1 \perp P_Y^0 &\rightleftarrows P_Y^1 (< M,M >_t = \infty) = 1 \end{aligned} \qquad (7)$$

[]

The conditions (5)-(7) provide sufficient conditions to ensure that $\Lambda_t \to 0$ P^0 a.s. as $t \to \infty$. Conditions for $T_\infty \to 0$ will depend on a specific suboptimal statistic , and later we will treat one such example.

We have discussed the limit behavior of the Chernoff bound under the mentioned condition. We now consider the convergence rate of the Chernoff exponents when the underlying signal or state process is Markov. Evans proposed to use the so-called quasi-transition function method [12] that was originally developed by Kac to evaluate the expectation of a multiplicative functional of Markov process. Zakai also mentioned the quasi-transition function as his motivation in deriving the non-linear filtering formula [33] . Evans considers signal-in-noise type observations when the signals are nongaussian diffusions or finite state Markov process , and noises are Brownian motions. Through the measure transformation technique he developed an expression for the exponent in the Chernoff bound under a new measure. Then by use of a quasi-transition function, he derives a Fokker-Planck type equations for the time evolution of a quantity closely related to the Chernoff exponent. The method of analysis is itself quite general. Hibey, Snyder , and van Schuppen [13] apply the quasi-transition function method to a discontinuous observation that contains a rate process associated with a counting observation. However Hibey points out in [14] that the usual conditions for the justifications of measure transformation are difficult to check in some practical cases such as a linear finite dimensional gaussian system. Here we apply the quasi-transition function method to several kinds of Markov signals with some justification of the measure transformation technique.

3.1.1) Full observation of state

Throughout this subsection $x(t)$ stands for an observation.

A. Discrete Markov Chain

Let $x^0(t)$, $x^1(t)$ be time-homogeneous Markov processes taking values in a finite state space , S ($\equiv \left\{ 1 , 2 , \cdots , M \right\}$), at discrete time instants. Let $p_0^0(\cdot)$,$p_0^1(\cdot)$ be their arbitrary initial probability vectors.

$$\begin{cases} H_0 : \text{transition probability } p^0(i,j) = p^0(x(t+1)=j \mid x(t)=i) \\ H_1 : \text{transition probability } p^1(i,j) \end{cases}$$

Likelihood ratio between two hypotheses is given by

$$\Lambda_t = \frac{p_0^1(x(0))}{p_0^0(x(0))} \prod_{n=1}^{t} \frac{p^1(x(n-1),x(n))}{p^0(x(n-1),x(n))} \tag{8}$$

Suppose H_0 is true. The quantity of interest is $E^0(\Lambda_t^s)$.

$$\begin{aligned} E^0(\Lambda_t^s) &= E^0[E^0(\Lambda_t^s \mid x(t))] \\ &= \sum_{i=1}^{M} E^0(\Lambda_t^s \mid x(t)=i)\, p_t^0(i) . \end{aligned} \tag{9}$$

,where $p_t^0(i)$ is a probability that $x(t)=i$.

Let $r_s(t,i) = E^0(\Lambda_t^s \mid x(t)=i)\, p_t^0(i)$. $r_s(t,i)$ is called as a quasi-transition function due to its evolution property. Once $r_s(t,i)$ is known for all i , we obtain $E^0(\Lambda_t^s)$ simply by (9) . Below we derive the evolution equation that $r_s(t,i)$ satisfies by use of Markov property of $x(t)$.

$$\begin{aligned} r_s(t+1,j) &= E^0(\Lambda_{t+1}^s \mid x(t+1)=j)\, p_{t+1}^0(j) \\ &= E^0\left(\Lambda_t^s \left[\frac{p^1(x(t),j)}{p^0(x(t),j)}\right]^s \mid x(t+1)=j\right) p_{t+1}^0(j) \\ &= E^0\left[E^0\left(\Lambda_t^s \left[\frac{p^1(x(t),j)}{p^0(x(t),j)}\right]^s \mid x(t+1)=j, x(t)=i\right) \mid x(t+1)=j\right] p_{t+1}^0(j) \\ &= E^0\left[\left[\frac{p^1(i,j)}{p^0(i,j)}\right]^s E^0(\Lambda_t^s \mid x(t)=i) \mid x(t+1)=j\right] p_{t+1}^0(j) \end{aligned}$$

because x^t is conditionally independent of $x(t+1)$ given $x(t)$, due to Markov property of $x(t)$, and so is Λ_s^t .

$$\begin{aligned} & r_s(t+1,j) \\ &= \sum_{i=1}^{M} \left[\frac{p^1(i,j)}{p^0(i,j)}\right]^s E^0(\Lambda_t^s \mid x(t)=i)\, p_t^0(i)\, p^0(i,j) \\ &= \sum_{i=1}^{M} \left[\frac{p^1(i,j)}{p^0(i,j)}\right]^s r_s(t,i)\, p^0(i,j) \\ &= \sum_{i=1}^{M} (p^1(i,j))^s (p^0(i,j))^{1-s} r_s(t,i) \end{aligned}$$

Hence $r_s(t,i)$ satisfies the recursion below.

$$r_s(t+1,j) = \sum_{i=1}^{M} (p^1(i,j))^s (p^0(i,j))^{1-s} \, r_s(t,i)$$

$$\text{with} \quad r_s(0,i) = (p_0^1(i))^s \, (p_0^0(i))^{1-s} \, .$$

Define a matrix $Q(s)$ by

$$Q_{ij} = [\, p^1(i,j)\,]^s \, [\, p^0(i,j)\,]^{1-s}$$

Let $r_s(t) = [\, r_s(t,1), \cdots, r_s(t,M)\,]^T$. Then we have

$$r_s(t+1) = Q^T(s)\, r_s(t)$$
$$\text{and,}$$
$$E^0(\Lambda_t^s) = \sum_{i=0}^{M} r_s(t,i) \tag{10}$$

Newmann and Stuck [34] also used the Chernoff bound for this problem and identifies the matrix $Q(s)$ by directly evaluating $E^0(\Lambda_t^s)$. We can easily identify a role played by the eigenvectors and eigenvalues of $Q(s)$ in determining the asymptotic behavior of $E^0(\Lambda_t^s)$. Typically $Q(s)$ has an unique largest positive real eigenvalue $\lambda_0(s)\,(<1)$, and $E^0(\Lambda_t^s) = O(\lambda_0(s)^t)$ as $t \to \infty$. For convenience, we quote the result in [34] . The spectral radius of a matrix $Q(s)$ with possibly complex eigenvalues $\left\{ \lambda_i \right\}$ is denoted by $r(Q(s))$;

$$r(Q(s)) = \underset{i}{Max} \, | \, \lambda_i \, |$$

Lemma (Stuck and Newmann [35])

1) $\limsup_{t \to \infty} \frac{1}{t} \ln E^0(\Lambda_t^s) \le \ln r(Q(s))$

i.e., $E^0(\Lambda_t^s) \le O(r(Q(s))^t)$ as $t \to \infty$.

2) There is at least one eigenvector w of $Q(s)$ with nonnegative entries and with $r(Q(s))$. If $w^T \, [\, (p_0^1(1))^s \cdot (p_0^0(1))^{1-s}, \cdots, (p_0^1(M))^s \cdot (p_0^1(M))^{1-s} \,]^T = 0$ then

$$\lim_{t \to \infty} \frac{1}{t} \ln E^0(\Lambda_t^s) = \ln r(Q(s)) \qquad \Box$$

B. Discrete Time Markov Signal

We consider an observation model given below,

$$H_1 : x(t+1) = f^{1}(x(t)) + \sqrt{g}\ \ w^{1}(t)$$

$$H_0 : x(t+1) = f^{0}(x(t)) + \sqrt{g}\ \ w^{0}(t) \tag{11}$$

For simplicity we consider only a scalar valued process ,and assume that g is a constant and $w^{i}(t)$ a white gaussian noise with $N(0,1)$, independent of $x(0)$. We assume that mutual absolute continuity holds at each t . As before let $r_s(t,x) = E^0(\Lambda_t^s \mid x(t) = x)\, p_t^0(x)$, where $p_t^i(x)$ is a probability density of $x(t)$ under H_i . By a similar procedure, we obtain

$$r_s(t+1,x) = \int_R [\, p_{t+1\mid t}^{1}(x \mid y)\,]^s\, [\, p_{t+1\mid t}^{0}(x \mid y)\,]^{1-s}\, r_s(t,y)\, dy \tag{12}$$

and $E^0(\Lambda_t^s) = \int r_s(t,x)\, dx$, where $p_{t+1\mid t}^{i}(x \mid y) = p^{i}(x(t+1) = x \mid x(t) = y)$. From (11) , we have

$$p_{t+1\mid t}^{i}(x \mid y) = \frac{1}{\sqrt{2\pi g}} \exp(-\frac{1}{2g}(x - f^{i}(y))^2) .$$

Rewriting (12), we have

$$r_s(t+1,x) = \int H(x,y:s)\, r_s(t,y)\, dy \tag{13}$$

,where

$$H(x,y:s) = \frac{1}{\sqrt{2\pi g}} \exp(-\frac{1}{2g}[\, s(x - f^{1}(y))^2 + (1-s)(x - f^{0}(y))^2 \,])$$

, and the initial condition , $r_s(0,x)$, is given by

$$r_s(0,x) = (\, p_0^1(x)\,)^s \cdot (p_0^0(x))^{1-s} .$$

(13) with the above initial condition yields a recursive solution to $r_s(t,x)$ for all $t \geq 0$. We now rederive (13) , which is somewhat more revealing , based on a measure transformation technique. We express $E^0(\Lambda_t^s)$ by

$$
\begin{aligned}
&E^0(\Lambda_t^s) \\
&= E^1(\Lambda_t^s \frac{dP^0}{dP^1} |_{F_t(x)}) \\
&= E^1(\frac{p^0(x_0)^{1-s}}{p^1(x_0)^{1-s}} \exp(\frac{(s-1)}{g} \sum_{j=1}^{t} [f^1(x_{j-1}) - f^0(x_{j-1})] \cdot x_j - \frac{(s-1)}{2g} \sum_{j=1}^{t} [f^1(x_{j-1})^2 - f^0(x_{j-1})^2])) .
\end{aligned}
\tag{14}
$$

Let $\beta_t = \frac{(s-1)}{g} \sum_{j=1}^{t} (f^1(x_{j-1}) - f^0(x_{j-1})) \cdot (x_j - f^1(x_{j-1}))$. We assume that

$$
[f^1(x_t) - f^0(x_t)]^2 < \infty \quad P^1 \text{ a.s.} \tag{15}
$$

at each t .

Define $\epsilon(\beta_t)$ as below.

$$
\epsilon(\beta_t) = \exp(\beta_t - \frac{1}{2} <\beta,\beta>_t) \quad \text{for } t \geq 1 \tag{16}
$$

, where $<\beta,\beta>_t$ is defined by

$$
<\beta,\beta>_t = \frac{(s-1)^2}{g} \sum_{j=1}^{t} [f^1(x_{j-1}) - f^0(x_{j-1})]^2
$$

, and $\epsilon(\beta_0)$ is defined to be 1 . We define a new probability measure P^2 on $F_T(x)$ by

$$
\frac{dP^2}{dP^1} = \epsilon(\beta_T) , \tag{17}
$$

where T is such that $t \in [0,T]$. That $\epsilon(\beta_t)$ is a $(F_t(x), P^1)$ martingale will be shown later as well as the validity of the measure transformation. The reason for the above measure transformation is to get rid of x_j of the cross product term appearing in (14), and is motivated by [13] [14] . Then $E^0(\Lambda_t^s)$ becomes

$$
\begin{aligned}
&E^0(\Lambda_t^s) \\
&= E^2(\Lambda_t^s \cdot \frac{dP^0}{dP^1} \cdot \frac{dP^1}{dP^2}) \\
&= E^2 \left\{ (\frac{p^0(x_0)}{p^1(x_0)})^{1-s} \exp(\frac{s^2 - s}{2g} \sum_{j=1}^{t} [f^1(x_{j-1}) - f^0(x_{j-1})]^2) \right\}
\end{aligned}
\tag{18}
$$

Define a quasi-transition function $r_s(t,x)$ by

$$
r_s(t,x) = E^2((\frac{p^0(x_0)}{p^1(x_0)})^{1-s} \exp(\frac{s^2 - s}{2g} \sum_{j=1}^{t} [f^1(x_{j-1}) - f^0(x_{j-1})]^2) \mid x(t) = x) \cdot p_t^2(x) \tag{19}
$$

where we need to obtain the stochastic difference equation that x_t satisfies under P^2 .

For a moment we concentrate on the evolution equation that $r_s(t,x)$ satisfies.

$$r_s(t+1,x)$$
$$= E^2\left(\left(\frac{p^0(x_0)}{p^1(x_0)}\right)^{1-s} \exp\left(\frac{s^2-s}{2g} \sum_{j=1}^{t+1} [f^1(x_{j-1}) - f^0(x_{j-1})]^2 \right) \mid x(t+1) = x \right) \cdot p_{t+1}^2(x)$$
$$= \int \exp\left(\frac{s^2-s}{2g} [f^1(y) - f^0(y)]^2 \right) p_{t+1|t}^2(x \mid y)\, r_s(t,y)\, dy \qquad (20)$$

by the Markov property of $x(t)$ under P^2, and also due to the smoothing property of the conditional expectation.

Lemma 3.4

1) $E^1(\epsilon(\beta_t)) = 1$

2) $p_{t+1|t}^2(x \mid y) = \frac{1}{\sqrt{2\pi g}} \exp(-\frac{1}{2g}(x - [s\, f^1(y) + (1-s) f^0(x)])^2)$

Proof ; we show (1) by an elementary calculation . Assume $t \geq 1$.

$$E^1(\epsilon(\beta_t) \mid F_{t-1}(x))$$
$$= \exp\left(-\frac{1}{2g}(s-1)^2 \sum_{j=1}^{t} [f^1(x_{j-1}) - f^0(x_{j-1})]^2 + \frac{(s-1)}{g} \sum_{j=1}^{t-1} [f^1(x_{j-1}) - f^0(x_{j-1})](x_j - f^1(x_{j-1}) \right)$$
$$\cdot E^1\left(\exp\left[\frac{(s-1)}{g} (f^1(x_{t-1}) - f^0(x_{t-1}))(x_t - f^1(x_{t-1})) \right] \mid F_{t-1}(x) \right)$$

But $x_t - f^1(x_{t-1})$ is independent of $F_{t-1}(x)$, and a gaussian with $N(0,g)$ under P^1 , and thus

$$E^1\left(\exp\left[\frac{(s-1)}{g} (f^1(x_{t-1}) - f^0(x_{t-1}))(x_t - f^1(x_{t-1})) \right] \mid F_{t-1}(x) \right)$$
$$= \exp\left(\frac{1}{2g} (s-1)^2 [f^1(x_{t-1}) - f^0(x_{t-1})]^2 \right)$$

,and is finite a.s. by assumption. Combining the above two yields

$$E^1(\epsilon(\beta_t) \mid F_{t-1}(x)) = \epsilon(\beta_{t-1})$$

Also $E^1(\epsilon(\beta_1) \mid x_0) = 1$ independently of x_0 . Since we define $\epsilon(\beta_0) = 1$, $\epsilon(\beta_t)$ is a $(F_t(x), P^1)$ martingale. We showed that $E^1(\epsilon(\beta_t)) = 1$, and thus we can set $\frac{dP^2}{dP^1} = \epsilon(\beta_T)$.

Now it remains to derive the stochastic difference equation that $x(t)$ satisfies under P^2 . Let $B \in F_t(x)$. Then we have

$$E^1(1_B(\omega))$$
$$= \int_B p^1(x^t) \prod_{j=0}^{t} dx_j$$
$$= \int_B \prod_{j=1}^{t} \frac{1}{\sqrt{2\pi g}} \exp(- \frac{1}{2g} [x_j - f^1(x_{j-1})]^2) p^1(x_0) \cdot \prod_{j=0}^{t} dx_j$$

where we make no distinction between $B(\in F_t(x))$ and its image on the observation space for convenience. Also,

$$E^2(1_B(\omega) \frac{dP^1}{dP^2})$$
$$= \int_B \prod_{j=1}^{t} \exp (-\frac{(s-1)}{g} [f^1(x_{j-1}) - f^0(x_{j-1})][x_j - f^1(x_{j-1})]$$
$$+ \frac{1}{2g} (s-1)^2[f^1(x_{j-1}) - f^0(x_{j-1})]^2) p^2(x^t) \prod_{j=0}^{t} dx_j$$

Let $p^2(x^t) = (\frac{1}{\sqrt{2\pi g}})^t \exp (q(x^t)) p^2(x_0)$.

Since B is arbitrary, we must have

$$q(x^t) = -\sum_{j=1}^{t} \frac{1}{2g} [x_j - \left\{ f^1(x_{j-1}) + (s-1) (f^1(x_{j-1}) - f^0(x_{j-1})) \right\}]^2$$

, and $p^1(x_0) = p^2(x_0)$. The above implies that

$$x_j = \left\{ s f^1(x_{j-1}) - (s-1) f^0(x_{j-1}) \right\} + \sqrt{g} \; w^2(j-1)$$

,where $w^2(t)$ is a white gaussian process with N (0, 1) under P^2 . Thus,

$$p^2_{t+1|t}(x \mid y) = \frac{1}{\sqrt{2\pi g}} \exp (-\frac{1}{2g} (x - [s f^1(y) + (1-s) f^0(x)])^2) \qquad (21)$$

[]

(21) is a sort of a discrete counter part of Girsanov's theorem in the present case. Combining (20) and (21), we have

$$r_s(t+1,x) = \int H(x,y:s) r_s(t,y) dy \qquad (22)$$

, where

$$H(x,y:s) = \frac{1}{\sqrt{2\pi g}} \exp (-\frac{1}{2g} (x - [s f^1(y) + (1-s) f^0(y)])^2 + \frac{s^2 - s}{2g} [f^1(y) - f^0(y)]^2)$$

which is identical to (13). Since we want to obtain some analytical result, we assume that the kernel $H(x,y:s)$ can be expressed as

$$H(x,y:s) = \sum_{j=0}^{\infty} \lambda_j \, \phi_j(\frac{x}{\gamma}) \, \phi_j(\frac{y}{\sigma}) \ , \ \lambda_j \in R \tag{23}$$

, where $\left\{ \phi_j(x) \right\}$ are assumed to be a complete orthonormal set for L^2 space , and dependence of the right side side of(23) on the parameter s is implicitly assumed. Motivation for the representation (23) comes from [35] , where series expansions of nongaussian densities are used. If $\gamma = \sigma$, (23) is an usual expansion known as the Mercer's expansion. Clearly a necessary condition for (23) to hold in the mean square sense is that $H(x,y:s)$ satisfy

$$\int \int H(x,y:s)^2 \, dx \, dy = \sigma \gamma \sum_{j=0}^{\infty} \lambda_j^2 < \infty$$

Let $r_s(t,x) = \sum_{i=0}^{\infty} c_i(t) \phi_i(x)$. From (22) and (23) we have

$$r_s(t+1,x) = \int [\sum_{j=0}^{\infty} \lambda_j \, \phi_j(\frac{x}{\gamma}) \, \phi_j(\frac{y}{\sigma})] \, r_s(t,y) \, dy$$

Also we note the following relations,

$$\phi_j(\frac{y}{\sigma}) = \sum_{i=0}^{\infty} d_{ij} \, \phi_i(y)$$

$$\phi_j(\frac{x}{\gamma}) = \sum_{k=0}^{\infty} e_{kj} \, \phi_k(y) .$$

Substituting the above two into $r_s(t+1,x)$ and arranging the terms out , we have

$$r_s(t+1,x) = \sum_{k=0}^{\infty} \left\{ \sum_{i=0}^{\infty} [\sum_{j=0}^{\infty} e_{kj} \, \lambda_j \, d_{ij}] \, c_i(t) \right\} \phi_k(x) .$$

Hence we obtain

$$c_k(t+1) = \sum_{i=0}^{\infty} (\sum_{j=0}^{\infty} e_{kj} \, \lambda_j \, d_{ij}) \, c_i(t)$$

, or

$$c(t) = (E \lambda D^T)^t \, c(0) \tag{24}$$

where $c(t) = [c_0(t), c_1(t), \cdots]^T$, λ is a diagonal matrix with $\lambda_{ij} = \delta_{ij} \lambda_i$,

$E_{kj} = e_{kj}$, and $D_{ij} = d_{ij}$ defined earlier . Also

$$E^0(\Lambda_t^s) = \int r_s(t,x)\, dx = \sum_{i=0}^{\infty} c_i(t)$$

The asymptotic behavior of $c(t)$ as $t \to \infty$ depends on a dominant eigenvalue of $E \lambda D^T$.

Proposition 3.5

If the matrix $E \cdot \lambda \cdot D^T$ of (24) has an unique dominant eigenvalue satisfying $|\tilde{\lambda}_0| < 1$, then $E^0(\Lambda_t^s)$ decreases with a rate $0(|\tilde{\lambda}_0|^t)$ as $t \to \infty$.

[]

At present we were not able to characterize eigenvalues $\left\{ \tilde{\lambda}_j \right\}$ of $E \lambda D^T$.

If Kabanov's condition (6) holds in a given observation model then $E^0(\Lambda_t^s) \to 0$, which in turn implies that spectrum of $E \lambda D^T$ is less than 1 whenever the representation (23) holds.

Example 3.1

We treat here a simple scalar gaussian case given by

$$H_1; \quad x(t+1) = x(t) + w^1(t) ,\ x_0 = 0$$
$$H_0; \quad x(t) = w^0(t) ,\ x_0 = 0$$

The present method can be readily extended to a general stationary vector case also. From (22) and the above observation model,

$$r_s(t+1,x) = \int \exp\left(\frac{s^2 - s}{2} y^2 \right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - s\ y)^2\right) r_s(t,y)\, dy \qquad (25)$$

$$r_s(0,x) = \delta(x)$$

For this case we find a general solution after direct manipulations with (25),

$$r_s(t,x) = \frac{1}{\sqrt{2\pi}} \cdot \prod_{j=1}^{t} \frac{1}{\sqrt{K(j)}} \cdot \exp\left(-\frac{a(t)}{2a(t) + s^2} x^2\right) \qquad (26)$$

, where $a(t)$ is positive for all $t \geq 1$. We have a recursive equation for $a(t)$ given by

$$a(t+1) = \frac{a(t)}{2a(t)+s^2} - \frac{s^2-s}{2} \quad , t \geq 1 \quad , 0 < s < 1 \tag{27}$$

, and

$$\frac{1}{K(t)} = \frac{1}{2a(t)+s^2} \ .$$

(27) yields an unique positive solution $a(t)$ for all $s \in (0,1)$, and also supports a stationary solution a_∞ . Hence $r_s(t,x)$ becomes asymptotically separated in t and x and $E^0(\Lambda_t^s) \rightarrow 0$ with an exponential rate given by $(\frac{1}{\sqrt{K_\infty}})^t$. For $s = \frac{1}{2}$,

$$a_\infty = \frac{33}{32} \ , \ \frac{1}{\sqrt{K_\infty}} = \sqrt{\frac{16}{37}} \ .$$

We note that the recursive equation for $a(t)$ will be replaced by matrix Ricatti equation for a general vector case, and when an unique stationary solution exists an exponential (asymptotically) convergence rate can be obtained and explicitly calculated.

Example 3-2

Consider the example 3-1. The kernel $H(x,y:s)$ was

$$H(x,y:s) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - s\ y)^2 + \frac{s^2-s}{2} y^2\right)$$

For $s = \frac{1}{2}$, it turns out that $H(x,y:\frac{1}{2})$ can be expanded as follows.

$$H(x,y:s) = \sum_{j=0}^{\infty} \sqrt{2}\ \xi^{1/2}\ \xi^j\ \phi_j(\frac{x}{\gamma})\ \phi_j(\frac{y}{\sigma})$$

where

$$\sigma = 8^{1/4} \ , \ \gamma = 2^{1/4} \ , \ \xi = \frac{1}{\sqrt{3}}$$

$$\phi_j(\frac{x}{\sigma}) = (\frac{2^j\ j!}{2\ \sigma\sqrt{\pi}})^{-1/2}\ N(0,\sigma)\ H_j(\frac{x}{\sigma})$$

$$H_j(\frac{x}{\sigma}) = (-\sigma)^j\ N(0,\sigma)^{-2}\ \frac{d^j}{dx^j}\ N(0,\sigma)^2 \ .$$

We note that $H_j(x)$ is the Hermite polynomial satisfying

$$\int H_i(x)\, H_j(x) \exp(-x^2)\, dx = \pi^{-1/2}\, 2^j\, j!\, \delta_{ij} .$$

Then a dominant coefficient of the matrix λ appearing in (22) is

$$\lambda_0 = \sqrt{2\xi} < 1$$

Example 3-3

We now examine a suboptimal statistic T_t for the observation model (11) . Throughout the example $f^0(x) = 0$ will be assumed. Also let $f^1(x) = f(x)$ and $g = 1$. Suppose the suboptimal statistic is given by a direct approximation of Λ_t .

$$T_t = \exp\left(\sum_{j=1}^{t} x_j\, f^*(x_{j-1}) - \frac{1}{2} \sum_{j=1}^{t} f^*(x_{j-1})^2 \right) , \quad t \geq 1 \tag{28}$$

where $f^*(x)$ is a function that simplifies $f(x)$, due to either its complexity or lack of its sufficient knowledge. T_0 is defined to be 1 for simplicity. We first consider $E^0(\Lambda_t^*)$ and prove the conditions of corollary 3.2 for this particular problem.

1) $T_\infty = 0$, P^0 a.s.

Let $M(t) = \sum_{j=1}^{t} x_j\, f^*(x_{j-1})$, and $M_0 = 0$. Let $E^0(\, f^*(x_t)\,)^2 < \infty$ at each t . M_t is a $(F_t(x), P^0)$ square integrable martingale Also let $<M,M>_t$ denote its quadratic variation given by $\sum_{j=1}^{t} f^*(x_{j-1})^2$.

Hence $T_t = \exp(M_t - \frac{1}{2} <M,M>_t)$. We quote the following theorem known as a strong law of square integrable martingales in [30] .

Theorem ([30] pp 487)

If $<M,M>_t$ is the quadratic variation of the square integrable martingale M_t and $<M,M>_t = \infty$ a.s., then with probability 1

$$\frac{M_t}{<M,M>_t} \to 0 \text{ a.s. as } t \to \infty . \qquad \square$$

Suppose $f^*(x)$ is such that $E^0(f^*(x))^2 < \infty$. Then according to the theorem ,

$$\lim_{t \to \infty} \frac{M_t}{<M,M>_t} = 0 \text{ a.s.}$$

Let N be a subset of Ω such that $<M,M>_t \to \infty$ for all $\omega \in \Omega - N$. Then the set N is obviously a P^0 null set under our null hypothesis assumptions. Given $\epsilon > 0$ there is $t_0(\omega)$ at each $\omega \in \Omega - N$ such that $\frac{M_t}{<M,M>_t} < \epsilon$ for all $t \geq t_0(\omega)$.

Let $\epsilon < \frac{1}{2}$. Then $(\frac{M_t}{<M,M>_t} - \frac{1}{2}) < -a < 0$ for all $t \geq t_0(\omega)$, and

$$\limsup_{t \to \infty} T_t = \limsup \exp (<M,M>_t (\frac{M_t}{<M,M>_t} - \frac{1}{2})$$
$$\leq \lim \exp (-a \ <M,M>_t) = 0$$

at each $\omega \in \Omega - N$. Thus $T_\infty = 0 \ P^0$ a.s.

2) $\lim E^0(T_t^s) = 0$ for $s \in (0,1)$.

Notice that T_t can be regarded as a likelihood ratio for the following hypotheses detection problem.

$$\tilde{H}_1 ; \quad x(t) = f^*(x(t-1)) + w^1(t-1)$$
$$H_0 ; \quad x(t) = w^0(t)$$
$$p^1(x_0) = p^0(x_0)$$

By proposition 3.1 and since $T_t \to 0$,we have $\lim E^0(T_t^s) = 0$ for $s \in (0,1)$.

We can use (22) to find a converging rate. Here we use an elementary method to obtain a rate of convergence. To proceed with a specific calculation and for simplicity let

$$f^*(x) = sgn(f(x)) = \begin{cases} 1 \text{ for } f(x) \geq 0 \\ -1 \text{ for } f(x) < 0 \end{cases}$$

Then

$$
\begin{aligned}
&E^0(\,T_t^s) \\
&= E^0(\,\exp(\,s \sum_{j=1}^{t} [\,sgn\,(f\,(x_{j-1})\,)\,x_j - \frac{1}{2}\,]\,)\,) \\
&= E^0(\,\exp(\,s \sum_{j=odd}^{t} [\,sgn\,(f\,(x_{j-1})\,)\,x_j - \frac{1}{2}\,]\,)\exp(\,s \sum_{j=even}^{t} [\,\cdots\,]\,) \\
&\leq [\,E^0(\,\exp(\,2s \sum_{j=odd}^{t} [\,sgn\,(f\,(x_{j-1})\,)x_j - \frac{1}{2}\,]\,)\,)\,]^{1/2} \cdot [\,E^0(\,\exp(\,2s\,\cdots)\,)\,]^{1/2}
\end{aligned}
$$

by Cauchy -Schwarz inequality that will be validated later. Consider a factor of the right side of the above.

$$
\begin{aligned}
&E^0(\,\exp(\,2s \sum_{j=odd}^{t} [\,sgn\,(f\,(x_{j-1})\,)\,x_j - \frac{1}{2}\,]\,)\,) \\
&= \prod_{j=odd}^{t} E^0(\,\exp[\,2s\,\,sgn\,(f\,(x_{j-1})\,)\,x_j - s\,]\,) \\
&= \prod_{j=odd}^{t} \exp(2s^2 - s\,)\,,
\end{aligned}
$$

due to the independence of $\left\{ x_{j-1},x_j \right\}_{j=odd}$ under P^0. Thus Cauchy-Schwarz bound becomes

$$E^0(T_t^s\,) \leq (\,\exp(\,2s^2 - s\,)\,)^{t/2}$$

This Cauchy-Schwarz bound is valid at each t for all s but ensures a converging rate only for $s \in (0,\frac{1}{2})$. For $s = \frac{1}{4}$, $P_F(t\,) \leq \exp(-\frac{1}{16}\,t\,)$.

We now consider $E^1(T_t^s)$.

1) $T_t^{-1} \to 0$, P^1 a.s.

From the definition of a suboptimal statistic (28) and the observation model H_1 we have

$$T_t^{-1} = \exp(\,-\sum_{j=1}^{t} f^{\,*}(x\,(j-1))\,w^1(j-1) - \sum_{j=1}^{t} [f^{\,*}(x\,(j-1))\,x\,(j-1) - \frac{1}{2}(\,f^{\,*}(\,x\,(j-1))\,)^2\,]\,)$$

Suppose $f^{\,*}(x\,) = sgn\,(f\,(x\,)\,)$ for simplicity. Then T_t^{-1} simplifies to

$$T_t^{-1} = \exp(-\sum_{j=1}^{t} sgn\,(f\,(x\,(j-1))\,)\cdot w^1(j-1) - \sum_{j=1}^{t} [\,|\,f\,(x\,(j-1))\,| - \frac{1}{2}\,]\,)$$

Suppose

(1) $x\,(t\,)$ is ergodic under P^1, (2) $E^1(|\,f\,(x\,(t\,))|) > \frac{1}{2}$.

Then $T_t^{-1} \rightarrow 0$ P^1 a.s. can be shown by use of the strong law of square integrable martingale. We note that we assume a too much stringent condition that $x(t)$ is ergodic.

2) there is s such that $E^1(T_t^{-s}) < K$ for $t \geq 0$

We were unable to verify the above uniform integrability condition by some effective means.

As an alternative we directly use (22). We assume that $f(x)$ lies on the first and third quadrants and $f(0) = 0$ for convenience. From (28) we have

$$E^1(T_t^{-s}) = E^1 \exp(-s \sum_{j=1}^{t} sgn(x(j-1))\, x(j) + \frac{s}{2} t)$$

We try a previous type of measure transformation on $F_T(x)$ specified by

$$\frac{dP^2}{dP^1} = \exp(-s \sum_{j=1}^{T} sgn(x(j-1))(x(j) - f(x(j-1))) - \frac{1}{2} s^2 T)$$

Thus $E^1(T_t^{-s})$ becomes

$$E^1(T_t^{-s}) = E^2(\exp(-s \sum_{j=1}^{t} | f(x(j-1)) | + \frac{1}{2}(s^2 + s)t)) \tag{29}$$

As before we obtain the difference equation that $x(t)$ satisfies under P^2. Let $B \in F_t(x)$.

$$\begin{aligned} &E^2(1_B(\omega)) \\ &= \int_B p^2(x^t) \prod_{j=0}^{t} dx_j \\ &= \int_B [\frac{1}{\sqrt{2\pi}}]^t \exp(q(x^t))\, p^2(x_0) \prod_{j=0}^{t} dx_j \end{aligned}$$

Whereas,

$$E^1(1_B(\omega) \frac{dP^2}{dP^1}) = \int_B \exp(-s \sum_{j=1}^{t} sgn(x(j-1))(x(j) - f(x(j-1))) - \frac{1}{2} s^2 t)\, p^1(x^t) \prod_{j=0}^{t} dx_j$$

,where $p^1(x^t)$ is given from the observation model. Proceeding as before we can identify $q(x^t)$,and obtain the following difference equation.

$$x(j) = f(x(j-1)) - s \; sgn(x(j-1)) + w^2(j-1) \tag{30}$$

,where $w^2(j)$ is a white gaussian process. By use of (29) and (30) we obtain a recursive equation given below. Let $r_s(t,x) = E^2(\exp(-s\sum_{j=1}^{t} | f(x(j-1)) |) \mid x(t) = x)\, p_t^2(x)$. Then

$$r_s(t+1,x) = \int \exp(-s\,| f(y) |)\, p_{t+1|t}^2(x \mid y)\, r_s(t,y)\, dy$$

The above combined with (29) gives $E^1(T_t^{-s})$ at each t . It seems that there is no simple way to obtain an analytical estimate but to compute numerically the given recursive equation.

For a somewhat trivial case we can show that the Chernoff exponent decreases to 0 faster than a certain exponential convergence rate. Suppose $f(x)$ is monotonically nondecreasing with $f(x) = -f(-x)$, $f(0) = 0$. Then

$$\begin{aligned}
& E^2(\exp(-s\sum_{j=1}^{t+1} | f(x(j-1)) |)) \\
&= E^2[E^2(\exp(-s\sum_{j=1}^{t} | f(x(j-1)) |)\exp(-s\,| f(x(t)) |) \mid x^{t-1})] \\
&= E^2[\exp(-s\sum_{j=1}^{t} | f(x(j-1)) |)\, E^2(\exp(-s\,| f(x(t)) |) \mid x(t-1))] \\
&\leq E^2(\exp(-s\sum_{j=1}^{t} | f(x(j-1)) |)) \cdot E^w(\exp(-s\,| f(x(t)) |))
\end{aligned}$$

, due to the present assumption on $f(x)$. $E^w(\cdot)$ denotes that the expectation is taken with the density of $w(t)^2$ (i.e. $N(0,1)$ here). The above implies that

$$E^2(\exp(-s\sum_{j=1}^{t} | f(x(j-1)) |)) \leq \prod_{j=1}^{t} E^w(\exp(-s\,| f(x(j)) |)$$

Thus

$$E^1(T_t^{-s}) \leq \left\{ \exp(\frac{1}{2}(s^2 + s)) \cdot E^w(\exp(-s\,| f(x) |)) \right\}^t$$

Below we show that there is s' such that the ratio appearing in the above is less than 1 for $s \in (0,s')$ for a certain f satisfying the present assumption. .

$$E^{w}(\exp(-s\,|\,f(x)\,|\,))$$
$$= \frac{1}{\sqrt{2\pi}} \int \exp(-s\,|\,f(x)\,|\,) \exp(-\frac{1}{2}x^2)\, dx$$
$$\equiv \exp(-g(s))$$

,where $g(s) \geq 0$, $g(0) = 0$, and monotonically decreasing and differentiable in s . We want to show that for $s \in (0,s')$

$$q(s) \equiv \frac{1}{2}(s^2 + s\ ) - g(s) < 0$$

Notice that $q(0) = 0$. For the derivative of $q(s)$ at $s\ = 0$ to be negative ,

$$g'(0) = \frac{1}{\sqrt{2\pi}} \int |\,f(x)\,| \exp(-\frac{x^2}{2})\, dx\ > \frac{1}{2}$$

,where $g'(s)$ is a derivative of $g(s)$. For any $f(x)$ satisfying the above under the previously given assumption , there is s' such that for all $s \in (0,s')$ our upperbound on the Chernoff exponent decreases exponentially in t . This fact can also be used to prove the P^1 almost sure convergence of T_t^{-1} to zero by use of the following version of Borel- Cantelli lemma with Markov inequality.

Theorem ([30])

Let $(\epsilon_n)_{n\geq 1}$ be a sequence of positive numbers such that $\epsilon_n \downarrow 0$, $n \rightarrow \infty$. If

$$\sum_{n=1}^{\infty} P\left\{\ |\,\xi_n - \xi\,| \geq \epsilon_n\ \right\} < \infty$$

then $\xi_n \rightarrow \xi$ a.s.

C. Diffusion

Observations under each hypotheses are given by

$$\begin{aligned} H_1\ &;\ dx(t) = f^{1}(x(t))\, dt + \sqrt{g}\ dw^{1}(t)\ \ ,\ x(0) = 0 \\ H_0\ &;\ dx(t) = f^{0}(x(t))\, dt + \sqrt{g}\ dw^{0}(t)\ \ ,\ x(0) = 0 \end{aligned} \qquad (31)$$

For simplicity $x(t) \in R^1$, $w^i(t)$ a real valued standard Wiener process with respect to $(F_t(x),P^i)$, g a constant > 0 . $x(0)$ is assumed to be 0 for convenience, since otherwise one has to take into account of the system response due to an initial data.

Under either hypotheses we assume an existence of unique strong and Markov solution, and that

$$P^i(\omega : \int_0^T | f^i(x(t,\omega)) |^2 dt < \infty) = 1$$
$$P^i(\omega : \int_0^T | f^i(w^i(t,\omega)) |^2 dt < \infty) = 1 \tag{32}$$

As a result restrictions of $P^{i'}$ s to $F_T(x)$ are equivalent to each other. Also probability measures on $C[0,T]$ induced by $x(t)$ under either hypotheses are equivalent to a Wiener measure on $C[0,T]$, that is induced by $\sqrt{g}\ w^i(t)$ (Th. 7.3.2 , [8]). By applying Girsanov's theorem to (31) the likelihood ratio between H_1 and H_0 is given by

$$\Lambda_t = \exp(\frac{1}{g}\int_0^t [f^1(x(s)) - f^0(x(s))] dx(s) - \frac{1}{2g}\int_0^t [f^1(x(s))^2 - f^0(x(s))^2] ds)$$

Let $\beta(t) = \frac{(s-1)}{\sqrt{g}}\int_0^t (f^1(x(s)) - f^0(x(s))) dw^1(s)$. Then $\beta(t)$ is a continuous local martingale (M_{loc}^c) with respect to $(F_t(x),P^1)$. Define $\epsilon(\beta(t))$ as

$$\epsilon(\beta(t)) = \exp(\beta(t) - \frac{1}{2} <\beta,\beta>_t) ,$$

where $<\beta,\beta>_t$ is the quadratic variation of $\beta(t)$. If $E^1\,\epsilon(\beta(T)) = 1$, then we can define a new probability measure P^2 on $(\Omega, F_T(x))$ by

$$\frac{dP^2}{dP^1} = \epsilon(\beta(T)) \tag{33}$$

where P^i is assumed to be restricted to $F_T(x)$. In [14] where partial observations models are treated, Hibey mentioned that conditions to ensure a validity of the above measure transformation are usually difficult to verify unless $f^i(x)$ is bounded. In this connection we state Novikov condition [8] for the present case.

Theorem [Novikov] If

$$E^1[\exp(\frac{(s-1)^2}{2g}\int_0^T [f^1(x(s)) - f^0(x(s))]^2 ds)) < \infty \tag{34}$$

,then $E^1\,\epsilon(\beta(T)) = 1$ at each finite T .

By using a quasi-transition function again we present a way to prove (34) .

Let $b(t) = \exp(\frac{(s-1)^2}{2g} \int_0^t [f^1(x(s)) - f^0(x(s))]^2 ds)$. Then

$$b(t) = 1 + \frac{(s-1)^2}{2g} \int_0^t b(s) [f^1(x(s)) - f^0(x(s))]^2 ds$$

Taking the conditional expectation on both sides of the above

$$\begin{aligned} &E^1(b(t) \mid x(t) = x) \\ &= 1 + \frac{(s-1)^2}{2g} E^1(\int_0^t b(s) [f^1(x(s)) - f^0(x(s))]^2 ds \mid x(t) = x) \\ &= 1 + \frac{(s-1)^2}{2g} \int_0^t E^1(b(s) [f^1(x(s)) - f^0(x(s))]^2 \mid x(t) = x) ds \end{aligned}$$

, where change of order of integrations are valid from the Fubini's theorem on a nonnegative jointly measurable function in (t,ω) under a σ -finite measure [31]. And the above becomes

$$1 + \frac{(s-1)^2}{2g} \int_0^t E^1([f^1(y) - f^0(y)]^2 E^1(b(s) \mid x(s) = y) \mid x(t) = x) ds$$

, since $b(s) = b(x^s)$ by definition , and $x(\cdot)$ is Markov.

Let $u(t,x) = E^1(b(t) \mid x(t) = x) p_t^1(x)$. Then we have

$$u(t,x) = p_t^1(x) + \frac{(s-1)^2}{2g} \int_0^t \int [f^1(y) - f^0(y)]^2 u(s,y) p_{t|s}^1(x \mid y) dy \; ds$$

Differentiating both sides of the above with respect to t yields

$$\begin{aligned} &\frac{\partial}{\partial t} u(t,x) \\ &= L_1^* p_t^1(x) + \frac{(s-1)^2}{2g} (f^1(x) - f^0(x))^2 u(t,x) \\ &+ \frac{(s-1)^2}{2g} L_1^* \int_0^t \int (f^1(y) - f^0(y))^2 u(s,y) p_{t|s}^1(x \mid y) dy \; ds \end{aligned}$$

where L_1^* is an adjoint of the infinitesmal operator for $x(t)$ under the hypothesis H_1 .

From the above two equations we obtain

$$\begin{aligned} &\frac{\partial}{\partial t} u(t,x) = L_1^* u(t,x) + \frac{(s-1)^2}{2g} [f^1(x) - f^0(x)]^2 u(t,x) \\ &u(0,x) = \delta(x) \end{aligned} \qquad (35)$$

(35) is an unstable partial differential equation that $u(t,x)$ satisfies. We assume an existence of unique solution to (35) on any finite interval $[0,T]$. If $u(t,x) \in L^1(R^1)$ at each t, then

$$\int u(t,x)\,dx = E^1[\exp(\frac{(s-1)^2}{2g}\int_o^t [f^1(x(s)) - f^0(x(s))]^2 ds)] < \infty$$

,and Novikov condition (34) is satisfied at each finite interval $[0,T]$. We state the above as a proposition

Proposition 3.6

Consider the diffusion model (31). Let $\epsilon(\beta(t))$ be defined by (33). If the partial differential equation

$$\frac{\partial}{\partial t}u(t,x) = L_1^* u(t,x) + \frac{(s-1)^2}{2g}[f^1(x) - f^0(x)]^2 u(t,x)$$

yields an unique solution on any finite interval $[0,T]$ such that $u(t,x) \in L^1(R^1)$ at each t, then

$$E^1(\epsilon(\beta(T))) = 1$$ []

It remains to obtain a sufficient condition from a PDE theory to ensure an existence of locally unique solution to (35) within $L^1(R^1)$ space. At this point we present an another way to validate the measure transformation (33).

Proposition 3.7

Let $f(x) = sf^1(x) + (1-s)f^0(x)$. Suppose $f^i(x)$ is a continuous function of x, and that

$$E^\mu \exp(\frac{1}{g}\int_0^T f(x(t))\,dx(t) - \frac{1}{2g}\int_0^T f(x(t))^2\,dt) = 1$$

where $x(\cdot) \in C[0,T]$ and μ is a Wiener measure on $C[0,T]$ with respect to which the coordinate process $x(t)$ is a Wiener process having a quadratic variation $g \cdot t$.

Then $E^1\epsilon(\beta(T)) = 1$.

[]

Proof ; (32) implies that

$$P^1(\omega : \int_0^T (f^1(x(t)) - f^0(x(t)))^2 dt < \infty) = 1 .$$

Following a standard stopping time technique ([8] , [38]), define the stopping times

$$\sigma_n(x) = \begin{cases} Inf\ (t \le T ; \int_0^t [f^1(x(s)) - f^0(x(s))]^2 ds \ge n) \\ T \qquad : \text{if the upper set is empty} . \end{cases}$$

Let $\tau_n(\omega) = \sigma_n(x(\omega))$, where $x(\omega)$ is the observation. Consider the measure transformation on $(\Omega, F_T(x))$ given below.

$$\begin{aligned} \epsilon_n(\omega) &= \frac{dP_n^2}{dP^1} \\ &= \exp\Big(\frac{(s-1)}{\sqrt{g}} \int_0^T (f^1(x(t)) - f^0(x(t)))I_{[0,\tau_n)]}(t)\, dw^1(t) \\ &\qquad - \frac{(s-1)^2}{2g} \int_0^T (f^1(x(t)) - f^0(x(t)))^2 I_{[0,\tau_n]}(t)\, dt \Big) \end{aligned} \tag{36}$$

$(f^1(x(t)) - f^0(x(t)))\, I_{[0,\tau_n]}(t)$ is $F_t(x)$ measurable and satisfies the Novikov condition. Thus $E^1(\epsilon_n) = 1$. And from the Girsanov theorem

$$w_n^2(t) = w^1(t) - \frac{(s-1)}{\sqrt{g}} \int_0^t [f^1(x(s)) - f^0(x(s))]\, I_{[0,\tau_n]}(s)\, ds \tag{37}$$

is a $(F_t(x), P_n^2)$ Wiener martingale. Substituting (37) into the argument of the exponent of (36) yields

$$\frac{(s-1)}{\sqrt{g}} \int_0^T [f^1(x(t)) - f^0(x(t))]\, I_{[0,\tau_n]}(t)\, dw_n^2(t) + \frac{(s-1)^2}{2g} \int_0^T [f^1(x(t)) - f^0(x(t))]^2 I_{[0,\tau_n]}(t)\, dt$$

To show the uniform integrability of $(\epsilon_n)_{n \ge 1}$ consider

$$\int_{\epsilon_n \ge N} \epsilon_n\, dP^1 = P_n^2(\epsilon_n \ge N) .$$

Then

$$P_n^2(\epsilon_n \geq N)$$
$$= P_n^2(\log\epsilon_n \geq \log N)$$
$$\leq P_n^2\Big(\frac{(s-1)}{\sqrt{g}} \int_0^T [f^1(x(t)) - f^0(x(t))] I_{[0,\tau_n]}(t)\, dw_n^2(t)$$
$$- \frac{(s-1)^2}{2g} \int_0^T [f^1(x(t)) - f^0(x(t))]^2 I_{[0,\tau_n]}(t)\, dt \geq \frac{1}{2}\log N\Big)$$
$$+ P_n^2\Big(\frac{(s-1)^2}{g} \int_0^T [f^1(x(t)) - f^0(x(t))]^2 I_{[0,\tau_n]}(t)\, dt \geq \frac{1}{2}\log N\Big) \qquad (38)$$

The first term on the right side of (38) is bounded by $\frac{2}{\sqrt{N}}$ independently of n from the well-known supermartingale inequality ([8] , ch.2). We need an uniform convergence of the second term to 0 as $N \to \infty$. Let

$$x^n(t) = \int_0^t f^1(x(s)) I_{[0,\tau_n]}(s)\, ds + \sqrt{g}\; w^1(t)$$

where $x(t)$ is a solution to the stochastic differential equation under H_1. $x^n(t)$ under P_n^2 is given by

$$x^n(t) = \int_0^t [\, s\; f^1(x(u)) + (1-s)\, f^0(x(u))\,]\, I_{[0,\tau_n]}(u)\, du + \sqrt{g}\; w_n^2(t)$$

Since $x^n(t) = x(t)$ for $t \in [0,\tau_n]$ and $I_{[0,\tau_n]}(t) = I_{[0,\tau_n']}(t)$ where $\tau_n'(\omega) = \sigma_n(\, x^n(\omega)\,)$, one may write down

$$x^n(t) = \int_0^t [\, s\; f^1(x^n(u)) + (1-s)\, f^0(x^n(u))\,]\, I_{[0,\tau_n']}(u)\, du + \sqrt{g}\; w_n^2(t) \; .$$

Thus $x^n(t)$ satisfies the stochastic differential equation given below.

$$dz(t) = [\, s\; f^1(z(t)) + (1-s) f^0(z(t))\,]\, I_{[0,\sigma_n(z)]}(t)\, dt + \sqrt{g}\; dw_n^2(t)\, , \; z(0) = 0 \, .$$

Consider the following stochastic differential equation.

$$d\,\xi(t) = [\, s\; f^1(\xi(t)) + (1-s)\, f^0(\xi(t))\,]\, dt + \sqrt{g}\; dw(t) \; , \; \xi(0) = 0 \, . \qquad (39)$$

Let $f(x) = s\; f^1(x) + (1-s)\, f^0(x)$. We assume

$$\mu(x \in C[0,T] : \int_0^T f(x(t))^2\, dt < \infty) = 1 \quad (40\text{–}1)$$

$$E^{\mu}(\exp(\frac{1}{g} \int_0^T f(x(t))\, dx(t) - \frac{1}{2g} \int_0^T f(x(t))^2\, dt) = 1 \quad (40\text{–}2)$$

Then by the theorem (7.4.1) of [8], (39) has a weak solution whose induced measure is equivalent to a Wiener measure under which $\xi(t)$ is a Wiener process having a quadratic variation $g \cdot t$.

We assume that $f(x)$ is continuous in x . Then

$$\int_0^T f(x(t))^2\, dt < \infty \text{ for every } x \in C[0,T]$$

And thus (40-1) is trivially satisfied. Then (40-2) is a necessary and sufficient condition for the existence of a weak solution, and furthermore the weak solution is unique (corollary 7.4.1 , [8]) .

Consider a stochastic differential equation given below.

$$d\xi'(t) = [s\, f^1(\xi'(t)) + (1-s)\, f^0(\xi'(t))]\, I_{[0,\sigma_n(\xi')]}(t)\, dt + \sqrt{g}\ dw'(t)\, ,\ \xi'(0) = 0 \quad (41)$$

It can be shown that (41) has also an unique weak solution by the following. Assume (40). Let

$$y(t) = \exp(\frac{1}{g} \int_0^t f(x(s))\, dx(s) - \frac{1}{2g} \int_0^t f(x(s))^2\, ds)$$

Then

$$E^{\mu} \exp(\frac{1}{g} \int_0^T f(x(t))\, I_{[0,\sigma_n(x)]}(t)\, dx(t) - \frac{1}{2g} \int_0^T f(x(t))^2\, I_{[0,\sigma_n(x)]}(t)\, dt)$$
$$= E^{\mu}\, y(\sigma_n)$$
$$= E^{\mu} [E^{\mu} (y(T) \mid F_{\sigma_n}(x))]$$
$$= E^{\mu}(y(T)) = 1$$

due to Doob's optional sampling theorem on a supermartingale [8].

Now define $\xi^n(t)$ as follows.

$$\xi^n(t) = \int_0^t [s\, f^1(\xi(u)) + (1-s)\, f^0(\xi(u))]\, I_{[0,\sigma_n(\xi)]}(u)\, du + \sqrt{g}\ w(t) \quad (42)$$

where $\xi(t)$ is a solution to (39).

Then $\xi^n(t) = \xi(t)$ for $t \in [0,\sigma_n(\xi)]$. Thus $I_{[0,\sigma_n(\xi)]}(t) = I_{[0,\sigma_n(\xi^n)]}(t)$.

Hence $\xi^n(t)$ is also a solution to (41), and by the uniqueness of weak solution to (41) it characterizes the probability measure induced by ξ'. Finally the second term on the right side of (38) is

$$P_n^2(\omega : \frac{(s-1)^2}{g}\int_0^T [f^1(x(t)) - f^0(x(t))]^2 I_{[0,\tau_n]}(t)\,dt \geq \frac{1}{2}\log N)$$
$$= P_n^2(\omega : \frac{(s-1)^2}{g}\int_0^T [f^1(x^n(t)) - f^0(x^n(t))]^2 I_{[0,\tau_n']}(t)\,dt \geq \frac{1}{2}\log N)$$
$$= P_n^{\xi'}(\xi' \in C : \frac{(s-1)^2}{g}\int_0^T [f^1(\xi'(t)) - f^0(\xi'(t))]^2 I_{[0,\sigma_n(\xi')]}(t)\,dt \geq \frac{1}{2}\log N)$$
$$= P^{\xi}(\xi \in C : \frac{(s-1)^2}{g}\int_0^T [f^1(\xi^n(t)) - f^0(\xi^n(t))]^2 I_{[0,\sigma_n(\xi^n)]}(t)\,dt \geq \frac{1}{2}\log N)$$
$$\leq P^{\xi}(\xi \in C : \frac{(s-1)^2}{g}\int_0^T [f^1(\xi(t)) - f^0(\xi(t))]^2\,dt \geq \frac{1}{2}\log N) \tag{43}$$

Notice that the last line of (43) is independent of n, and goes to 0 since $f^i(x)$ is continuous in x. Hence $(\epsilon_n)_{n\geq 1}$ is P^1 uniformly integrable sequences of random variables. Furthermore

$$\epsilon_n \rightarrow \epsilon(\beta(T))$$
$$= \exp(\frac{s-1}{\sqrt{g}}\int_0^T [f^1(x(t)) - f^0(x(t))]\,dw^1(t) - \frac{(s-1)^2}{2g}\int_0^T [f^1(x(t)) - f^0(x(t))]^2\,dt)$$

P^1 a.s. Therefore $E^1(\epsilon_n) \rightarrow E^1\,\epsilon(\beta(T))$ and $E^1\,\epsilon(\beta(T)) = 1$.

[]

In the above we have proved the uniform integrability of $(\epsilon_n)_{n\geq 1}$, provided that $f^i(x)$ is a continuous function of x and the condition (40-2) holds, which in the present situation is equivalent to an existence of unique weak solution to (39). Therefore the measure transformation (33) can be post-validated if eq.(46) below has an unique weak solution.

The Chernoff exponent $E^0(\Lambda_t^s)$ can be given by

$$
\begin{aligned}
&E^0(\Lambda_t^s) \\
&= E^2(\Lambda_t^s \frac{dP^0}{dP^1}\frac{dP^1}{dP^2}) \\
&= E^2[\exp(\frac{(s^2-s)}{2g}\int_0^t [f^1(x(u)) - f^0(x(u))]^2 du)] \\
&= E^2(E^2[\exp(\frac{(s^2-s)}{2g}\int_0^t [f^1(x(u)) - f^0(x(u))]^2 du) \mid x(t) = x])
\end{aligned}
\tag{44}
$$

Hence it is necessary to obtain a stochastic differential equation that $x(t)$ satisfies under P^2.

$\int_0^t \frac{1}{\sqrt{g}}(dx(s) - f^1(x(s))\,ds)$ is a standard Wiener process under $(F_t(x), P^1)$. From the Girsanov's theorem

$$
w^2(t) = w^1(t) - \int_0^t \frac{(s-1)}{\sqrt{g}}[f^1(x(u)) - f^0(x(u))]\,du \tag{45}
$$

is a standard Wiener process under $(F_t(x), P^2)$. Rewriting (45) we obtain a stochastic differential equation that $x(t)$ satisfies under P^2.

$$
dx(t) = [s\, f^1(x(s)) + (1-s)\, f^0(x(t))]\,dt + \sqrt{g}\,dw^2(t) \tag{46}
$$

with $x(0) = 0$. We assume that $x(t)$ is Markov under P^2. A sufficient condition for this assumption to hold is the Lipshitz continuity of the drift coefficient of (46) with a sublinear growth condition (Th. 5.4.1 , [8]).

Let $r_s(t,x)$ be defined by

$$
r_s(t,x) = E^2[\exp(\frac{(s^2-s)}{2g}\int_0^t [f^1(x(u)) - f^0(x(u))]^2 du) \mid x(t) = x]\, p_t^2(x)
$$

Repeating the similar steps leading to (35) , we obtain

$$
\begin{aligned}
&\frac{\partial}{\partial t} r_s(t,x) = L_2^* r_s(t,x) + \frac{(s^2-s)}{2g}[f^1(x) - f^0(x)]^2 r_s(t,x) \\
&r_s(0,x) = \delta(x)
\end{aligned}
\tag{47}
$$

where L_2^* is an adjoint operator for $x(t)$ under P^2, which is assumed to be well defined. We state the result as a proposition below.

Proposition 3.8

Consider the hypotheses detection problem (31). Let $r_s(t,x) = E^2(\Lambda_t^s \frac{dP^0}{dP^2} \mid x(t) = x)$, where P^2 is defined by (33). Then $r_s(t,x)$

satisfies the partial differential equation given below.

$$\frac{\partial}{\partial t} r_s(t,x) = L_2^* r_s(t,x) + \frac{(s^2-s)}{2g} [f^1(x) - f^0(x)]^2 r_s(t,x)$$
$$r_s(0,x) = \delta(x)$$

where L_2^* is the adjoint operator for a Markov process $x(t)$ given by

$$dx(t) = [\, s f^1(x(t)) + (1-s) f^0(x(t)) \,]\, dt + \sqrt{g}\ dw^2(t) \ .$$ []

Separation of variable technique is typical for the PDE of (47). Attempting a solution of the form $r_s(t,x) = G(t)H(x)$ leads to

$$\frac{d}{dt} G(t) = \lambda\, G(t)$$
$$(L_2^* + \frac{(s^2-s)}{2g} [f^1(x) - f^0(x)]^2) H(x) = \lambda\, H(x)$$

The final solution then becomes

$$r_s(t,x) = \sum_{j=0}^{\infty} c_i \exp(\lambda_i t) H_i(x) \ ,$$

where λ_i and H_i are the eigenvalues and eigenfunctions of the operator $L_2^* + \frac{(s^2-s)}{2g} [f^1(x) - f^0(x)]^2$.

If the Kabanov's condition (7)

$$P^1(\lim_{t\to\infty} \int_0^t [f^1(x(s)) - f^0(x(s))]^2 \, ds = \infty) = 1$$

holds , and PDE of (47) can be solved by a separation of variable technique , then $\lambda_0(s) < 0$ at each $s \in (0,1)$, where λ_0 is a dominant eigenvalue of the operator. We note that the Kabanov's condition for the present case is not really a restriction . A heuristic argument for this is as follows :

Fix ω . Consider an arbitrary sequence $\left\{ t_i \right\}$ such that $t_i \uparrow \infty$. Let $g(t_i) = \int_0^{t_i} [f^1(x(s)) - f^0(x(s))]^2 \, ds$. Then $g(t_i)$ is monotonically increasing in i . Hence

$$\lim_{i\to\infty} g(t_i) = \infty \ \text{ or } \ k(\omega)$$

Suppose $k(\omega)$ is true. This implies that

$$\lim_{i\to\infty}\int_{t_i}^{\infty} (f^{1}(x(s)) - f^{0}(x(s)))^2 \, ds = 0$$

which can only be true on a null set under P^1 , considering our diffusion model unless $f^{1}(x) = f^{0}(x)$.

We note that one can use the backward version of (47) . In partial observation problems Hibey , Snyder ,and Van Schuppen [13] [14] used a quasi-transition function , $q(t_0,x)$, of the form

$$q(t_0,x) = E^{i}(\exp(-\int_{t_0}^{t} V(x(s))\, ds) \mid x(t_0) = x)$$

where $V(x) \geq 0$ for all x , and also a backward partial differential equation that $q(t_0,x)$ satisfies. In [14] mathematical justification for the backward evolution equation is quoted from Dynkin and others' works. Major advantage of the backward partial differential equation is that it remains valid when there does not exist transition probability density for a given Markov process. Since all these aspects are matters of rigorous justification, we simply accept (47) , a forward equation , here for a full observation problem. Later in the partial observation we use the backward version due to difficulties arising when signals are present under both hypotheses. We can also call the backward equation as Feynman-Kac formula [19].

3.1.2) Partial Observation

A. Discrete Time Markov Signal

We consider a binary hypotheses detection. Under each hypothesis H_i the observation is given by

$$\begin{aligned} x^{i}(t+1) &= f^{i}(x^{i}(t)) + \sqrt{g^{i}}\, w^{i}(t) \ , \ x_0^{i}(0) = x_0^{i} \\ y(t) &= h^{i}(x^{i}(t)) + v^{i}(t) \end{aligned} \tag{48}$$

where all the processes involved are scalar for the sake of simplicity, and $w^{i}(t)$, $v^{i}(t)$ are as before. The optimal likelihood ratio statistic for the present observation model is usually not feasible to implement in practice except a few cases

, where a finite dimensional sufficient statistic exists. We start our analysis based on the following type of suboptimal statistic T_t .

$$T_t = \frac{\exp(-\frac{1}{2}\sum_{j=0}^{t}[\,y(j) - u^1(j)\,]^2)}{\exp(-\frac{1}{2}\sum_{j=0}^{t}[\,y(j) - u^0(j)\,]^2)} \tag{49}$$
$$= \exp(\sum_{j=0}^{t}(\,u^1(j) - u^0(j)\,)\,y(j) - \frac{1}{2}\sum_{j=0}^{t}[\,u^1(j)^2 - u^0(j)^2\,]$$

,where $u^i(j)$ is an one step ahead suboptimal prediction of $h^i(x(j))$ given y^{j-1} . We want to evaluate $E^0(T_t^s)$, and thus assume that the hypothesis H_0 is true. We further restrict the nature of u^i as below.

$$u^i(t) = u^i(\hat{x}^i(t)) ,$$

where $\hat{x}^i(t)$ is given by

$$\hat{x}^i(t+1) = \tilde{f}^{\,i}(\hat{x}^i(t)) + \sqrt{k_i(t)}(y(t) - u^i(t))$$

Since $y(t) = h^0(x^0(t)) + v^0(t)$, the above becomes

$$\hat{x}^i(t+1)$$
$$= \left\{\tilde{f}^{\,i}(\hat{x}^i(t)) + \sqrt{k_i(t)}\,h^0(x^0(t)) - \sqrt{k_i(t)}\,u^i(\hat{x}^i(t))\right\} + \sqrt{k_i(t)}\,v^0(t)$$
$$= F^i(\hat{x}^i(t), x^0(t)) + \sqrt{k_i(t)}\,v^0(t)$$

Let $x_a(t) = [\,x^0(t), \hat{x}^0(t), \hat{x}^1(t)\,]^T$. Then a stochastic difference equation for $x_a(t)$ is given by

$$\begin{bmatrix} x^0(t+1) \\ \hat{x}^0(t+1) \\ \hat{x}^1(t+1) \end{bmatrix} = \begin{bmatrix} f^0(x^0(t)) \\ F^0(x^0(t),\hat{x}^0(t)) \\ F^1(x^0(t),\hat{x}^1(t)) \end{bmatrix} + \begin{bmatrix} \sqrt{g_0} & 0 \\ 0 & \sqrt{k_0(t)} \\ 0 & \sqrt{k_1(t)} \end{bmatrix} \begin{bmatrix} w^0(t) \\ v^0(t) \end{bmatrix} \tag{50}$$

with the initial condition

$$x^0(0) = x_0^0 \ , \ \hat{x}^i(0) = E^i(x^i(0)) \ .$$

The augmented state $x_a(t)$ is Markov. Since the matrix of (50) is not a nonsingular square matrix, we can not use a forward evolution equation of quasi-transition function as in the case of full observation of state. Thus we instead consider a backward version. Let

$$Q(t_0,t) = \sum_{t_0}^{t} \left\{ (u^1(j) - u^0(j))(h^0(x^0(j)) + v^0(j)) - \frac{1}{2}[u^1(j)^2 - u^0(j)^2] \right\}$$

Define

$$r_s(t_0,x) = E^0(\exp(s\ Q(t_0,t)) \mid x_a(t_0) = x)$$

, where $x = [x^0(t_0) = x_1,\ \hat{x}^0(t_0) = x_2,\ \hat{x}^1(t_0) = x_3]^T$. Then

$$\begin{aligned} r_s(t_0-1,x) &= E^0(\exp(s\ Q(t_0-1,t)) \mid x_a(t_0-1) = x) \\ &= E^0(E^0(\exp(s\ Q(t_0,t))\exp[s\ (u^1(t_0-1) - u^0(t_0-1))(h^0(x(t_0-1)) + v^0(t_0-1)) \\ &\quad - \frac{s}{2}(u^1(t_0-1)^2 - u^0(t_0-1)^2)] \mid x_a(t_0) = y,\ x_a(t_0-1) = x) \mid x_a(t_0-1) = x) \end{aligned}$$

(51)

Notice that we have the following from (50)

$$v^0(t) = \frac{1}{\sqrt{k_0(t)}}(\hat{x}^0(t+1) - F^0(\hat{x}^0(t),x^0(t))).$$

Also from the Markov property of $x_a(t)$, the last line of (51) becomes

$$E^0\left(\left\{ \exp(s\ f(x,y))E^0(\exp(s\ Q(t_0,t)) \mid x_a(t_0) = y) \right\} \mid x_a(t_0-1) = x\right)$$

, where

$$f(x,y) = (u^1(x_3) - u^0(x_2))(h^0(x_1) + \frac{1}{\sqrt{k_0(t_0-1)}}[y_2 - F^0(x_2,x_1)]) .$$

Therefore

$$r_s(t_0-1,x) = \int \exp(sf(x,y))\, r_s(t_0,y)\, p_{t_0|t_0-1}(y \mid x)\, dy \qquad (52)$$

with the boundary condition given by

$$\begin{aligned} r_s(t,x) &= E^0(\exp(sQ(t,t)) \mid x_a(t) = x) \\ &= E^0(\exp(s[u^1(t) - u^0(t)](h^0(x^0(t)) + v^0(t)) - \frac{s}{2}[u^1(t)^2 - u^0(t)^2]) \mid x_a(t) = x) \\ &= \exp(s[u^1(x_3) - u^0(x_2)]h^0(x_1) - \frac{s}{2}[u^1(x_3)^2 - u^0(x_2)^2] + \frac{s^2}{2}[u^1(x_3) - u^0(x_2)]^2) \end{aligned}$$

(53)

and finally

$$E^0(T_t^s) = E^0(r_s(0,x)) = \int r_s(0,x)\, p(x_s(0) = x)\, dx \quad .$$

An usual defect associated with a backward equation is that one has to repeat the whole procedure as t varies.

B. Diffusion

In this section the observation models are given by

$$\begin{aligned} dx^i(t) &= f^i(x^i(t))dt + g^i(x^i(t))dw^i(t) \qquad ,x^i(0) = x_0^i \\ dy(t) &= h^i(x^i(t))dt + \sqrt{r}\,dv^i(t) \qquad ,y(0) = 0 \end{aligned} \tag{54}$$

For simplicity all the processes involved are real-valued. Assumptions similar to those of section (2-1) are assumed . Likelihood ratio at each t for the above detection problem is

$$\Lambda_t = \exp(\frac{1}{r} \int_0^T (\hat{h}^1(u) - \hat{h}^0(u))dy(u) - \frac{1}{2r} \int_0^t (|\hat{h}^1(u)|^2 - |\hat{h}^0(u)|^2)\, du$$

where $\hat{h}^i(t) = E^i(h^i(x^i(t)) \mid F_t(y))$.

Assume that H_0 is true. As before we use a measure transformation technique to characterize $E^0(\Lambda_t^s)$. We define $\beta(t)$ as follows.

$$\beta(t) = \frac{(s-1)}{\sqrt{r}} \int_0^t (\hat{h}^1(u) - \hat{h}^0(u))\, d\nu^1(u)$$

where $\nu^1(t) = \frac{1}{\sqrt{r}}\left\{ y(t) - \int_0^t \hat{h}^1(u)\, du \right\}$, and is a $(F_t(y),P^1)$ Wiener martingale. Thus $\beta(t) \in M_{loc}^c(F_t(y),P^1)$. Define a new probability measure on $(\Omega,F_T(y))$ by

$$\frac{dP^2}{dP^1} = \epsilon(\beta(T)) = \exp(\beta(T) - \frac{1}{2}<\beta,\beta>_T) \tag{55}$$

, provided that $E^1(\epsilon(\beta(T)) = 1$. In (55) P^i is assumed to be restricted to $F_T(y)$.

As we have said earlier, usual conditions to verify the measure transformation (55) is difficult to use in some practical cases. When $h^i(x^i)$ is bounded, then, of course, Novikov condition holds and (55) is valid. When a finite dimensional optimal filtering is available minor modifications of Proposition 3.6 can be used to check the

Novikov condition. Here we take the approach given in Proposition 3.7 for a general treatment:

Assume $P^1(\int_0^T (\hat{h}^1(t) - \hat{h}^0(t))^2 \, dt < \infty) = 1$. Under H_1 the observation $y(t)$ can be represented by

$$y(t) = \int_0^t \hat{h}^1(u) du + \sqrt{r}\, \nu^1(t)$$

Since $\hat{h}^i(t)$ is $F_t(y)$ measurable there exists a causal functional $\gamma^i : [0,T] \times C \to R$ which is $B([0,T]) \times B(C[0,T])$ measurable and such that for a.e. t (Th. 2.7.2 , [8]) ,

$$\gamma^i(t,\cdot) = \hat{h}^i(t) \, , P^1 \text{ a.s.}$$

Then $y(t)$ under P^1 may be written as

$$y(t) = \int_0^t \gamma^1(u,y) \, du + \sqrt{r}\, \nu^1(t) \tag{56}$$

Define the stopping times

$$\sigma_n(x) = \begin{cases} Inf\ (t \leq T : \int_0^t (\gamma^1(u,x) - \gamma^0(u,x))^2 \, du \geq n) \\ T \qquad : \text{if the upper set is empty} \ . \end{cases}$$

Let $\tau_n(\omega) = \sigma_n(y(\omega))$ under H_1. Consider the measure transformation on $(\Omega, F_T(y))$.

$$\begin{aligned} \frac{dP_n^2}{dP^1}(\omega) &= \epsilon_n(\omega) \\ &= \exp(\frac{(s-1)}{\sqrt{r}} \int_0^T [\gamma^1(u,y(\omega)) - \gamma^0(u,y(\omega))] \, I_{[0,\tau_n(\omega)]}(u) \, d\nu^1(u,\omega) \\ &\quad - \frac{(s-1)^2}{2r} \int_0^T [\gamma^1(u,y(\omega)) - \gamma^0(u,y(\omega))]^2 \, I_{[0,\tau_n(\omega)]}(u) \, du \ ) \end{aligned} \tag{57}$$

$(\gamma^1(t,\omega) - \gamma^0(t,\omega)) \, I_{[0,\tau_n(\omega)]}(t)$ is $F_t(y)$ measurable and satisfies the Novikov condition , and thus $E^1 \epsilon(\beta(T)) = 1$. And from the Girsanov theorem,

$$\nu_n^2(t,\omega) = \nu^1(t,\omega) - \frac{(s-1)}{\sqrt{r}} \int_0^t [\gamma^1(u,y(\omega)) - \gamma^0(u,y(\omega))] \, I_{[0,\tau_n(\omega)]}(u) \, du \tag{58}$$

is a $(F_t(y), P_n^2)$ Wiener martingale.

Substituting (58) into the argument of exponent of (57) yields

$$\frac{(s-1)}{\sqrt{r}}\int_0^T[\gamma^1(u,y(\omega))-\gamma^0(u,y(\omega))]I_{[0,\tau_n(\omega)]}(u)\,d\nu_n^2(t,\omega)$$
$$+\frac{(s-1)^2}{2r}\int_0^T[\gamma^1(u,y(\omega))-\gamma^0(u,y(\omega))]I_{[0,\tau_n(\omega)]}(u)\,du$$

To show the uniform integrability of $(\epsilon_n)_{n\geq 1}$ under P^1, consider the following

$$\int_{\epsilon_n\geq N}\epsilon_n\,dP^1=P_n^2(\epsilon_n(\omega)\geq N)$$

$$\begin{aligned}
&P_n^2(\epsilon_n\geq \log N)\\
&\leq P_n^2\Big(\frac{(s-1)}{\sqrt{r}}\int_0^T[\gamma^1(u,y(\omega))-\gamma^0(u,y(\omega))]\,I_{[0,\tau_n(\omega)]}(u)\,d\nu_n^2(u,\omega)\\
&\quad+\frac{(s-1)^2}{2r}\int_0^T[\gamma^1(u,y(\omega))-\gamma^0(u,y(\omega))]^2\,I_{[0,\tau_n(\omega)]}(u)\,du\geq\frac{1}{2}\log N\Big)\\
&\quad+P_n^2\Big(\frac{(s-1)^2}{r}\int_0^T[\gamma^1(u,y(\omega))-\gamma^0(u,y(\omega))]^2\,I_{[0,\tau_n(\omega)]}(u)\,du\geq\frac{1}{2}\log N\Big)
\end{aligned}\tag{59}$$

The first term on the right-hand side of (59) is bounded by $\frac{2}{\sqrt{N}}$ independently of n by the supermartingale inequality ([8] ch.2). We need an uniform convergence of the second term to 0 as $N\to\infty$. As before define

$$y^n(t,\omega)=\int_0^t\gamma^1(u,y(\omega))\,I_{[0,\tau_n(\omega)]}(u)\,du+\sqrt{r}\;\nu^1(t,\omega)$$

Notice that $y^n(t,\omega)=y(t,\omega)$ on $(\omega:0\leq t\leq\tau_n(\omega))$. $y^n(t,\omega)$ under P_n^2 is an Ito process with respect to $(\nu_n^2(t),F_t(y),P_n^2)$ satisfying

$$\begin{aligned}
y^n(t,\omega)&=\int_0^t[s\,\gamma^1(u,y(\omega))+(1-s)\gamma^0(u,y(\omega))]\,I_{[0,\tau_n(\omega)]}(u)\,du+\sqrt{r}\,\nu_n^2(t,\omega)\\
&=\int_0^T[s\,\gamma^1(u,y^n(\omega))+(1-s)\,\gamma^0(u,y^n(\omega))]I_{[0,\tau_n'(\omega)]}(u)\,du+\sqrt{r}\,\nu_n^2(t,\omega)
\end{aligned}\tag{60}$$

, where $\tau_n'(\omega)=\sigma_n(y^n(\omega))$. The last line of (60) follows from the fact that $y^n(t,\omega)=y(t,\omega)$ on $(\omega:0\leq t\leq\tau_n(\omega))$, and thus $\tau_n(\omega)=\tau_n'(\omega)$.

Consider the following stochastic differential equation.

$$d\,\xi(t)=(s\,\gamma^1(t,\xi)+(1-s)\gamma^0(t,\xi))dt+\sqrt{r}\;dw(t)\qquad,\xi(0)=0\tag{61}$$

Let $\gamma(t,x)=s\,\gamma^1(t,x)+(1-s)\,\gamma^0(t,x)$. We assume that $\gamma(t,x)$ can be chosen such that

$$\int_0^T \gamma^2(t,x)\,dt < \infty \text{ , for every } x \in C[0,T] \quad (62\text{-}1)$$
$$E^{\mu} \exp\left(\frac{1}{r}\int_0^T \gamma(t,x)dx(t) - \frac{1}{2r}\int_0^T \gamma(t,x)^2\,dt\right) = 1 \quad (62\text{-}2) \quad , \qquad (62)$$

where μ is a Wiener measure on $C[0,T]$ under which the coordinate process $x(t)$ is a Wiener process having a quadratic variation $r \cdot t$. Then by the theorem (7.4.1) and corollary (7.4.1) of [8] (61) has an unique weak solution. Consider a stochastic differential equation given below.

$$d\xi'(t) = [s\,\gamma^1(t,\xi) + (1-s)\gamma^0(t,\xi)]\, I_{[0,\sigma_n(\xi')]}(t)\,dt + \sqrt{r}\,dw'(t) \text{ , } \xi'(0) = 0 \,. \qquad (63)$$

By use of Doob's optional sampling theorem one can show that (63) has also an unique weak solution.

Now define $\xi^n(t)$ as follows.

$$\xi^n(t) = \int_0^t [s\,\gamma^1(u,\xi) + (1-s)\gamma^0(u,\xi)\,)I_{[0,\tau_n(\xi)]}(u)\,du + \sqrt{g}\,w(t)$$

, where $\xi(t)$ is a solution to (61), and by the uniqueness of weak solution $\xi^n(t)$ characterizes the probability measure induced by ξ' since $\xi^n(t)$ is also a solution to (63) by definition. Finally the second term on the right-hand side of (59) is bounded by

$$\begin{aligned}
&P_n^2\Big(\frac{(s-1)^2}{r}\int_0^T(\gamma^1(u,y) - \gamma^0(u,y)\,)^2\, I_{[0,\tau_n(\omega)]}(u)\,du \ge \frac{1}{2}\log N\Big) \\
&= P_n^2\Big(\frac{(s-1)^2}{r}\int_0^T(\gamma^1(u,y^n(\omega)) - \gamma^0(u,y^n(\omega))\,)^2\, I_{[0,\tau_n'(\omega)]}(u)\,du \ge \frac{1}{2}\log N\Big) \\
&= P_n^{\xi'}\Big(\xi' \in C : \frac{(s-1)^2}{r}\int_0^T(\gamma^1(u,\xi') - \gamma^0(u,\xi')\,)^2\, I_{[0,\sigma_n(\xi')]}(u)\,du \ge \frac{1}{2}\log N\Big) \qquad (64) \\
&= P^{\xi}\Big(\xi \in C: \frac{(s-1)^2}{r}\int_0^T(\gamma^1(u,\xi^n) - \gamma^0(u,\xi^n)\,)^2 I_{[0,\sigma_n(\xi^n)]}(u)\,du \ge \frac{1}{2}\log N\Big) \\
&\le P^{\xi}\Big(\xi \in C : \frac{(s-1)^2}{r}\int_0^T(\gamma^1(u,\xi) - \gamma^0(u,\xi)\,)^2\,du \ge \frac{1}{2}\log N\Big)
\end{aligned}$$

Notice that the last line of (64) is independent of n , and goes to 0 due to the assumption (62-1) . Hence $(\epsilon_n)_{n\geq 1}$ is P^1 uniformly integrable sequences of random variables. Furthermore

$$\epsilon_n \rightarrow \epsilon(\beta(T))$$
$$= \exp\Big(\frac{s-1}{\sqrt{r}} \int_0^T [\, \gamma^1(u, y(\omega)) - \gamma^0(u, y(\omega))\,]\, d\nu^1(t) - \frac{(s-1)^2}{2r} \int_0^T [\gamma^1(u, y(\omega)) - \gamma^0(u, y(\omega))\,]^2\, du \Big)$$

P^1 a.s. Therefore $E^1\epsilon_n \rightarrow E^1\epsilon(\beta(T))$ and $E^1\epsilon(\beta(T)) = 1$.

[]

In the above we have proved the uniform integrability of $(\epsilon_n)_{n \geq 1}$, provided that (62-1) and (62-2) hold.

$$\text{If } \int_0^T \gamma^i(u, y)^2\, du < \infty \text{ for every } y \in C\ , \tag{65}$$

then (62-1) holds. We were unable to verify (65) for a general situation , and is thinking of the possibility of using the robust version to $E^i(h^i(x^i(t)) \mid F_t(y))$ if such a robust version holds for an unbounded $h^i(x^i)$ also. For a gaussian system one can directly identify a causal functional γ^i as a linear functional of y^t . In this case (61) has an unique strong solution since the causal functional $\gamma(t, x)$ is Lipschitz continuous and satisfies the sublinear growth conditions ([8] ch.5) .

Henceforth we assume that $E^1(\beta(T)) = 1$. Then

$$\begin{aligned} E^0(\Lambda_t^s) &= \int_\Omega \Lambda_t^s(t)\, dP^0 = \int \Lambda_t^s(t)\, \frac{dP^0}{dP^1}\, \frac{dP^1}{dP^2}\, dP^2 \\ &= E^2 \exp\Big(\frac{(s^2 - s)}{2r} \int_0^t (\hat{h}^1(u) - \hat{h}^0(u))^2\, du \Big) \end{aligned} \tag{66}$$

(66) is a result given by Evans [13] [14] under the assumption that $E^1\epsilon(\beta(T)) = 1$. To evaluate (66) we need to characterize $\hat{h}^i(t)$ under P^2 .

From the nonlinear filtering results (e.g. [17]) the conditional means $E^i(h^i(x(t)) \mid F_t(y))$ and $E^i(x^i(t) \mid F_t(y))$ satisfy the following stochastic differential equation .

$$d\hat{h}^i(t) = \hat{H}^i(t)dt + \frac{1}{\sqrt{r}}\textstyle\sum_t^i(h^i,h^i)\, d\nu^i(t)$$
$$d\hat{x}^i(t) = \hat{f}^i(t) + \frac{1}{\sqrt{r}}\textstyle\sum_t^i(x^i,h^i)\, d\nu^i(t) \qquad (67)$$
$$\nu^i(t) = \frac{1}{\sqrt{r}}\int_0^t dy(u) - \hat{h}^i(u)\, du$$

, where

$$\hat{f}^i(t) = E^i(f^i(x^i(t)) \mid F_t(y)) \ , \ \hat{H}^i(t) = E^i(L^i h^i(x^i(t)) \mid F_t(y))$$
$$\textstyle\sum_t^i(x^i,h^i) = E^i[(x^i(t) - \hat{x}^i(t))(h^i(t) - \hat{h}^i(t))] \mid F_t(y)) \ ,$$
$$\textstyle\sum_t^i(h^i,h^i) = E^i[(h^i(t) - \hat{h}^i(t))^2 \mid F_t(y)]$$

here L^i is an infinitesmal generator for a diffusion $x^i(t)$. To obtain the stochastic differential equation that $\hat{h}^i(t)$ satisfies we need to characterize the observation $y(t)$ under P^2. From (55) and Girsanov's theorem

$$\nu^2(t,\omega) = \nu^1(t,\omega) - \frac{(s-1)}{\sqrt{r}}\int_0^t [\hat{h}^1(u,\omega) - \hat{h}^0(u,\omega)]\, du$$

or equivalently (68)

$$y(t) = \int_0^t (s\hat{h}^1(u) + (1-s)\hat{h}^0(u))du + \sqrt{r}\ \nu^2(t)$$

,where $\nu^2(t)$ is a $(F_t(y),P^2)$ Wiener martingale.

Substituting (68) into (67) yields the desired equation that $\hat{h}^i(t)$, $\hat{x}^i(t)$ satisfy under P^2.

$$\begin{bmatrix} d\hat{h}^0(t) \\ d\hat{h}^1(t) \end{bmatrix} = \begin{bmatrix} \hat{H}^0(t) + \frac{s}{r}\sum_t^0(h^0,h^0)[\hat{h}^1(t) - \hat{h}^0(t)] \\ \hat{H}^1(t) + \frac{(1-s)}{r}\sum_t^1(h^1,h^1)[\hat{h}^0(t) - \hat{h}^1(t)] \end{bmatrix} dt + \begin{bmatrix} \frac{1}{\sqrt{r}}\sum_t^0(h^0,h^0) \\ \frac{1}{\sqrt{r}}\sum_t^1(h^1,h^1) \end{bmatrix} d\nu^2(t)$$

$$\begin{bmatrix} d\hat{x}^0(t) \\ d\hat{x}^1(t) \end{bmatrix} = \begin{bmatrix} \hat{f}^0(t) + \frac{s}{r}\sum_t^0(x^0,h^0)[\hat{h}^1(t) - \hat{h}^0(t)] \\ \hat{f}^1(t) + \frac{(1-s)}{r}\sum_t^1(x^1,h^1)[\hat{h}^0(t) - \hat{h}^1(t)] \end{bmatrix} dt + \begin{bmatrix} \frac{1}{\sqrt{r}}\sum_t^0(x^0,h^0) \\ \frac{1}{\sqrt{r}}\sum_t^1(x^1,h^1) \end{bmatrix} d\nu^2(t)$$

(69)

Since $\hat{x}^i(t)$ or $\hat{h}^i(t)$ may or may not be Markov processes , we restrict our attention to the cases where a finite dimensional optimal filtering is available or a finite dimensional suboptimal filtering is used to define a suboptimal test statistic.

As a tool to evaluate (66) or its version for a suboptimal statistic T_t we state a Feynman-Kac formula , but in a heuristic way.

Theorem (Feynman - Kac formula)

Let α be a measurable function defined on R^N , and consider the backward PDE.

$$\frac{\partial}{\partial s} v(s,x) + Lv(s,x) + \alpha(x)\, v(s,x) = 0 \quad , s \leq t \quad , v(t,x) = b(x) \tag{70}$$

where $b \in C_0(R^N)$ and L is an infinitesmal generator of a diffusion process $x(t)$ uniquely satisfying

$$dx(t) = f(t,x(t))dt + g(t,x(t))\, dw(t)$$

Suppose an unique solution exists for (70). Then

$$v(s,x) = E(\, b(x(t)) \exp(\int_s^t \alpha(x(u))\, du \,) \mid x(s) = x \,) \qquad []$$

Proof : the proof will be heuristic and can be found elsewhere.

Let $\Psi(\theta) = v(\theta,x(\theta)) \exp(\int_s^\theta \alpha(x(u))\, du \,)$ where $v(s,x)$ is a solution to (70). Suppose $f(t,x)$ and $g(t,x)$ are sufficiently smooth in both variables. Then $v(\theta,x(\theta))$ is regular enough to apply the Ito's rule to $\Psi(\theta)$.

$$d\,\Psi(\theta) = dv(\theta,x(\theta)) \exp(\int_s^\theta \alpha(x(u))\, du \,) + v(\theta,x(\theta))\, d(\exp(\int_s^\theta \alpha(x(u))\, du \,)$$

$$= \left\{ [\frac{\partial}{\partial \theta} v + Lv\,](\theta,x(\theta))\, d\theta + \nabla v(\theta,x(\theta))^T\, g(\theta,x(\theta))\, dw(\theta) \right\} \exp(\int_s^\theta \alpha(x(u))\, du \,)$$

$$+ v(\theta,x(\theta)) \exp(\int_s^\theta \alpha(x(u))\, du \,)\, \alpha(x(\theta))\, d\theta$$

,or in the integral form

$$\Psi(t) = \Psi(s) + \int_s^t [\frac{\partial}{\partial\theta}v + Lv + \alpha v](\theta,x(\theta)) \exp(\int_s^\theta \alpha(x(u))\,du)\,d\theta$$
$$+ \int_s^t \nabla v(\theta,x(\theta))^T g(\theta,x(\theta)) \exp(\int_s^\theta \alpha(x(u))\,du)\,dw(\theta) \quad (71)$$

Suppose the above Ito integral with $s = 0$ is a martingale . Then taking the conditional expectation on both sides of the above

$$E(\Psi(t) \mid x(s) = x) = E(\Psi(s) \mid x(s) = x)$$

, or equivalently from the definition of $\Psi(t)$,

$$v(s,x) = E(b(x(t)) \exp(\int_s^t \alpha(x(u))\,du) \mid x(s) = x)$$

[]

Let $\phi(t,\omega) = \nabla v(t,x(t,\omega))^T g(t,x(t,\omega)) \exp(\int_0^t \alpha(x(u,\omega))\,du)$. A well-known sufficient condition for the stochastic integral of (71) to be a martingale is that

$$\begin{aligned} &1.\ \phi(t,\omega) \text{ is jointly measurable in } (t,\omega) \\ &2.\ E(\int_0^t |\phi(u,\omega)|^2\,du) < \infty \end{aligned} \quad (72)$$

A general and detailed conditions for the validity of Feynman-Kac formula and (72) is beyond the scope of this work. We will simply assume that Feynman-Kac formula is valid for the problems of our interests. We note the work of Hibey, Snyder and Van Schuppen [13] , where they cited the works of Dynkin and Rosenblatt to justify equations of the type (70). Their basic assumption is that there is $c > 0$ such that $\alpha(x) \leq c$ for all x .

B-1) When a finite dimensional optimal filtering is available

Rewriting (66)

$$E^0(\Lambda_t^s) = E^2(E^2[\exp(\int_0^t \alpha(\hat{h}(u))\,du) \mid \hat{h}_0]) \quad (73)$$

, where $s \in (0,1)$, $\hat{h}(u) = [\hat{h}^0(u),\hat{h}^1(u)]^T$ and $\alpha(x) = \frac{s^2-s}{2r}(x_1 - x_2)^2\ (\leq 0)$. We evaluate the above in a manner similar to the section 3.1 , assuming that a finite dimensional optimal filtering yielding a Markov solution is available. We specialize to the following examples.

Example 3.4

Consider the following hypotheses detection problem

$$dx^i(t) = A^i x^i(t)\,dt + C^i dw^i(t) \quad , x^i(0) = x_0^i$$
$$dy(t) \quad = H^i x^i(t)\,dt + \sqrt{r}\,dv^i(t) \quad , y(0) = 0$$

, where $i = 0,1$. Filtering equations for the problem are given by

$$d\hat{x}^i(t) = A^i \hat{x}^i(t)\,dt + \textstyle\sum_t^i H^i r^{-1}(dy(t) - H^i \hat{x}^i(t)dt\ )$$
$$\hat{x}^i(0) = E^i(x^i(0)) = 0$$

, where $\sum_t^i$ is the error covariance of $\hat{x}^i(t)$, satisfying the well-known Ricatti equation.

Suppose H_0 is true. From (73) we have

$$E^0(\Lambda_t^s) = E^2(\,E^2[\,\exp(\,\frac{s^2-s}{2r}\int_0^t (H^1\hat{x}^1(u) - H^0\hat{x}^0(u))^2\,du\ )\ |\ \hat{x}_0\,]\,)$$

The filtering equations under P^2 is given by (69),

$$\begin{bmatrix} d\hat{x}^0(t) \\ d\hat{x}^1(t) \end{bmatrix} = \begin{bmatrix} A^0\hat{x}^0(t) + s\sum_t^0 H^0 r^{-1}(\,H^1\hat{x}^1(t) - H^0\hat{x}^0(t)\,) \\ A^1\hat{x}^1(t) + (1-s)\sum_t^1 H^1 r^{-1}(\,H^0\hat{x}^0(t) - H^1\hat{x}^1(t)\,) \end{bmatrix} dt + \begin{bmatrix} \sum_t^0 H^0 r^{-1/2} \\ \sum_t^1 H^1 r^{-1/2} \end{bmatrix} d\nu^2(t)$$

Let $r_s(u,x) = E^2(\,\exp(\,\frac{s^2-s}{2r}\int_s^t [H^1\hat{x}^1(\theta) - H^0\hat{x}^0(\theta)\,]^2\,d\theta\,)\ |\ \hat{x}(u) = x\,)$. Recognizing the Feynman-Kac formula, $r_s(u,x)$ is a solution to the backward equation given by

$$-\frac{\partial}{\partial u} r_s(u,x) = L_s\, r_s(u,x) + \frac{s^2-s}{2r}(H^1 x_1 - H^0 x_0)^2$$
$$r_s(t,x) = 1$$

, where L_t is the infinitesmal generator of the vector Markov process $\hat{x}(t)$ under $(F_t(y),P^2)$.

If $dy(t) = \sqrt{r}\,dv^0(t)$ under H_0 , one prefers to use a forward evolution equation rather than its backward version . The forward version, derived through the same procedure used in Prop. 3.8 , is given by

$$\frac{\partial}{\partial t} r_s(t,x) = L_t^* r_s(t,x) + \frac{s^2-s}{2r}[\hat{H}^1 x^1 - \hat{H}^0 x^0]^2 r_s(t,x)$$
$$r_s(0,x) = \delta(x)$$

, where now $r_s(t,x) = E^2(\exp(\int_0^t [H^1\hat{x}^1(u) - H^0\hat{x}^0(u)]^2\, du\ ) \mid \hat{x}(t) = x)\, p^2(t,x)$, and L_t^* is an adjoint operator for $\hat{x}^1(t)$ satisfying

$$d\hat{x}^1(t) = (A^1 - (1-s)\textstyle\sum_t^1 H^1 r^{-1} H^1)\hat{x}^1(t)\, dt + r^{-1/2}\textstyle\sum_t^1 H^1\, d\nu^2(t)$$
$$\hat{x}^1(0) = 0$$

Example 3.5 : Finite-State Markov Process

Let $x(t)$ be a finite state Markov process taking values in $S = \left\{ x_1, \cdots, x_N \right\}$,where $x_i \in R$. Let $p^i(t)$ be the probability that $x(t) = x_i$, and assume that $p(t) = [\, p^1(t), \cdots, p^N(t)\,]^T$ satisfies

$$\frac{d}{dt} p(t) = A\ p(t)$$

, where A is called an intensity matrix.

The observation model is given by

$$H_1 : dy(t) = h(x(t))\, dt + dv^1(t)$$
$$H_0 : dy(t) = dv^0(t)$$

Let $\pi^i(t) = p(x(t) = i \mid F_t(y))$, and let

$$b = [h(x_1), \cdots, h(x_N)]^T$$
$$B = diag(h(x_1), \cdots, h(x_N))$$

Then $\pi(t) = [\pi^1(t), \cdots, \pi^N(t)]^T$ satisfies the following N-dimensional stochastic differential equation .

$$d\,\pi(t) = A\ \pi(t)\, dt + [\, B - b^T \pi(t) I\,]\, \pi(t)\, d\nu^1(t)$$
$$\pi(0) = p(0) \qquad (74)$$

An unique $F_t(y)$ measurable solution of (74) exists , and is Markov with respect to $(F_t(y), P^1)$ ([17] : ch.9 , Theorem 9.2) . We note that Novikov condition is satisfied since $h(x(t))$ is uniformly bounded. Under P^2

$$\nu^2(t) = \nu^1(t) - (s-1)\int_0^t \hat{h}^1(u)\, du$$

is a $(F_t(y), P^2)$ Wiener martingale. Substituting the above into (74) with $\hat{h}^1(t) = b^T \pi(t)$ yields the equation that $\pi(t)$ satisfies.

$$d\,\pi(t) = \left\{ A - (s-1)\, b^T \pi(t)\, [B - b^T \pi(t) I\,] \right\} \pi(t)\, dt + (B - b^T \pi(t) I\,)\pi(t)\, d\nu^2(t) \tag{75}$$
$$\pi(0) = p\,(0)$$

We assume that an unique $F_t(y)$ measurable solution exists and the solution is Markov. Now

$$E^0(\Lambda_t^s) = E^2\left(E^2 \left[\exp\left(\frac{s^2 - s}{2} \int_0^t [\, b^T \pi(u)\,]^2\, du \right) \mid \pi(t) \right] \right)$$

Let $r_s(t,\pi) = E^2\left(\exp\left(\frac{(s^2 - s)}{2} \int_0^t [\, b^T \pi(u)\,]^2\, du \right) \mid \pi(t) = \pi \right) p^2(t,\pi)$, where $p^2(t,\pi)$ is a probability density of $\pi(t)$ under P^2 . Then $r_s(t,\pi)$ is a solution to the forward equation.

$$\frac{\partial}{\partial t} r_s(t,\pi) = L^* r_s(t,\pi) + \frac{s^2 - s}{2} [b^T \pi(t)\,]^2 r_s(t,\pi)$$
$$r_s(0,\pi) = \delta(\pi - p\,(0)\,)$$

, where L^* is the infinitesmal generator for (75).

B-2 Suboptimal Case

Observation models are given by (54). Let the suboptimal statistic be defined by

$$T_t = \exp\left(\frac{1}{r}\int_0^t (h^{*1}(x^{*1}(u)\,) - h^{*0}(x^{*0}(u))\,)dy\,(u) - \frac{1}{2r}\int_0^t [\, h^{*1}(x^{*1}(u)\,)^2 - h^{*0}(x^{*0}(u))^2\,]\, du \right)$$

,where $h^{*i}(x^{*i}(u))$, $x^{*i}(u)$ are respectively suboptimal estimates of $h^i(x^i(t))$ and $x^i(t)$ given y^t under H_i . We replace the filtering equations for the optimal state estimate , (69) , with the following finite dimensional equations.

$$dx^{*i}(t) = f^{*i}(x^{*i}(t)\,)\, dt + r^{-1}\textstyle\sum^{*i}(x^{*i}(t)\,)\,(dy(t) - h^{*i}(x^{*i}(t)\,)\, dt\,) \tag{76}$$

We assume that f^{*i} , $\sum^{*i}$, h^{*i} are such that the solutions to (76) are well defined and $F_t(y)$ adapted. Also we assume that $h^{*i}(x)$ are chosen to be a bounded function of x for a validity of measure transformation technique. Let $h^{*i}(x^{*i}(t)) = h^{*i}(t)$. Define

$$\beta(t) = \frac{s}{\sqrt{r}} \int_0^t (h^{*1}(u) - h^{*0}(u))\, dv^0(u) \in M_{loc}^c(F_t, P^0)$$

$$\epsilon(\beta(t)) = \exp(\beta(t) - \frac{1}{2} <\beta,\beta>_t)$$

, where we use the fact that $v^0(t)$ is a (F_t, P^0) Wiener martingale . Actually $\beta(t)$ is a martingale since $h^{*i}(\cdot)$ is bounded by assumption. As before we shall see that this definition of $\beta(t)$ will lead to a functional that does not depend explicitly on the observations. Since $h^{*i}(\cdot)$ is assumed to be bounded, $E^0\, \epsilon(\beta(t)) = 1$ at each finite t . Define P^2 on F_t by

$$\frac{dP^2}{dP^0} = \epsilon(\beta(t))$$

Representing $E^0(T_t^s)$ as before

$$E^0(T_t^s) = \int T_t^s \frac{dP^0}{dP^2}\, dP^2$$

$$= E^2(E^2(\exp[\int_0^t (\frac{s}{r}(h^{*1}(u) - h^{*0}(u))h^0(x^0(u)) + \frac{s^2}{2r}(h^{*1}(u) - h^{*0}(u))^2 \tag{77}$$

$$- \frac{s}{2r}(h^{*1}(u)^2 - h^{*0}(u)^2))\, du\ ]\ |\ x_0))$$

,where we use the fact that $v^0(t) = y(t) - \int_0^t h^0(u)\, du$ and $x_0 = [x_0^0, x_0^{*0}, x_0^{*1}]^T$ denotes the initial state. We note that the right-hand side does not explicitly depend on the observation processes.

Continuing as before we describe the stochastic differential equation that the joint state $x(t)\ (\equiv [x^0(t), x^{*0}(t), x^{*1}(t)]^T)$ satisfies under P^2 . From Girsanov's theorem,

$$v^2(t) = v^0(t) - \frac{s}{\sqrt{r}} \int_0^t (h^{*1}(u) - h^{*0}(u))\, du$$

$$w^2(t) = w^0(t)$$

Or equivalently,

$$y(t) = \int_0^t \left\{ h^0(u) + s\ [h^{*1}(u) - h^{*0}(u)] \right\} du + \sqrt{r}\ v^2(t)$$

Substituting the above into (54) and (76) yields

$$\begin{bmatrix} dx^0(t) \\ dx^{*0}(t) \\ dx^{*1}(t) \end{bmatrix}$$

$$= \begin{bmatrix} f^0(x^0(t)) \\ f^{*0}(x^{*0}(t)) + \frac{1}{r}\Sigma_t^{*0}(x^{*0}(t))(h^0(t) - h^{*0}(t)) + \frac{s}{r}\Sigma_t^{*0}(x^{*0}(t))[h^{*1}(t) - h^{*0}(t)] \\ f^{*1}(x^{*1}(t)) + \frac{1}{r}\Sigma_t^{*1}(x^{*1}(t))(h^0(t) - h^{*1}(t)) + \frac{s}{r}\Sigma_t^{*1}(x^{*1}(t))(h^{*1(t)} - h^{*0}(t)) \end{bmatrix} dt$$

$$+ \begin{bmatrix} g^0(x^0(t)) & 0 \\ 0 & r^{-1/2}\Sigma_t^{*0}(x^{*0}(t)) \\ 0 & r^{-1/2}\Sigma_t^{*1}(x^{*1}(t)) \end{bmatrix} \begin{bmatrix} dw^0(t) \\ dv^2(t) \end{bmatrix} \qquad (78)$$

We assume that (78) has an unique strong solution and the solution is Markov.

Define $r_s(u,x)$ as follows.

$$r_s(u,x) = E^2(\exp(\int_s^t [\frac{s}{r}(h^{*1}(\theta) - h^{*0}(\theta))h^0(x^0(\theta)) + \frac{s^2}{2r}(h^{*1}(\theta) - h^{*0}(\theta))^2$$
$$- \frac{s}{2r}(h^{*1}(\theta)^2 - h^{*0}(\theta)^2)]d\theta) \mid x(u) = x)$$

where $x(u)$ is the previously defined augmented state. Recognizing the Feynman-Kac formula, $r_s(u,x)$ is a solution to the backward equation.

$$-\frac{\partial}{\partial u} r_s(u,x) = Lr_s(u,x) + \alpha(x)\, r_s(u,x) \quad , \quad u \le t$$
$$r_s(t,x) = 1$$

where

$$\alpha(x) = \frac{s}{r}(h^{*1}(x^{*1}) - h^{*0}(x^{*0}))\, h^0(x^0) + \frac{s^2}{2r}(h^{*1}(x^{*1}) - h^{*0}(x^{*0}))^2 - \frac{s}{2r}(h^{*1}(x^{*1})^2 - h^{*0}(x^{*0})^2)$$

, and L is the infinitesmal generator of the Markov process (78). $r_s(0,x)$ is given as a special case of $r_s(u,x)$ by letting $u = 0$. and

$$E^0(T_t^s) = E^2\, r_s(0,x) = \int r_s(0,x)\, p^2(x_0^0)\, dx$$

, where $p^2(x_0^0) = p^0(x_0^0)$ is an initial density of x_0^0 .

Example 3.6

We here treat a particular suboptimal test statistic arising from an approximation of diffusion by a finite state Markov process. Observation model of interest is given by

$$\begin{aligned} H_1 :\ & dx(t) = f(x(t))dt + g(x(t))dw(t) \ , x(0) = x_0 \\ & dy(t) = h(x(t))dt + dv^1(t) \ , y(0) = 0 \\ H_0 :\ & dy(t) = dv^0(t) \ , y(0) = 0 \end{aligned}$$

Our suboptimal test statistic of interest is given by

$$T_t = \exp\left(\int_0^t b^T \pi(u)\, dy(u) - \frac{1}{2}\int_0^t (b^T \pi(u))^2\, du\right)$$

where $\pi(t)$ satisfies the optimal filtering equation for example 3.5 ,

$$\begin{aligned} d\pi(t) &= A\ \pi(t)\, dt + (B - b^T\pi(t)I)\pi(t)(dy(t) - b^T\pi(t)\, dt) \\ \pi(0) &= \pi_0 \ . \end{aligned} \tag{79}$$

,where A is some intensity matrix approximating the given diffusion $x(t)$. We assume that an unique $F_t(y)$ measurable solution exists under either hypotheses. Let $\beta(t) = s\int_0^t b^T \pi(u)\, dv^0(u)$. $\beta(t)$ is a $(F_t(y),P^0)$ martingale. Define a measure transformation on $F_t(y)$ by

$$\frac{dP^2}{dP^0} = \epsilon(\beta(t))$$

Then

$$E^0(\,T_t^s\,) = E^2(\,\exp(\,\frac{s^2 - s}{2}\int_0^t (b^T \pi(u))^2\, du\ )$$

The equation that $\pi(t)$ satisfies under P^2 is

$$\begin{aligned} d\pi(t) &= \left\{A\ \pi(t) - (1-s)(B - b^T\pi(t)\, I)\pi(t) b^T \pi(t)\right\}dt + (B - b^T\pi(t)\, I)\pi(t)\, dv^2(t) \\ \pi(0) &= \pi_0 \end{aligned} \tag{80}$$

where $v^2(t)$ is a $(F_t(y),P^2)$ Wiener martingale defined by Girsanov's theorem. We assume that an unique Markov $F_t(y)$ measurable solution exists under P^2. Letting $r_s(u,\pi) = E^2[\exp(\frac{s^2-s}{2}\int_s^t (b^T\pi(\theta))^2\, d\theta) \mid \pi(u)=\pi]$, $r_s(u,\pi)$ is a solution to

$$-\frac{\partial}{\partial u} r_s(u,\pi) = Lr_s(u,\pi) + \frac{s^2-s}{2} b^T \pi\, r_s(u,\pi) \quad , \quad u \le t$$
$$r_s(t,\pi) = 1$$

where L is the infinitesimal generator of (80).

Next we consider $E^1(T_t^s)$. Define $\beta(t)$ as follows.

$$\beta(t) = s\int_0^t b^T\pi(u)\, dv^1(u) \quad , \quad s < 0$$

$$\epsilon(\beta(t)) = \exp(s\int_0^t b^T\pi(u)\, dv^1(u) - \frac{s^2}{2}\int_0^t (b^T\pi(u))^2\, du\ )$$

Define P^2 on F_t by

$$\frac{dP^2}{dP^1} = \epsilon(\beta(t))$$

Characterizing $E^1(T_t^s)$ as before,

$$E^1(T_t^s) = E^2(\exp(\int_0^t [\frac{s^2-s}{2}(b^T\pi(u))^2 + sb^T\pi(u)\, h^1(x^1(u))]du\ ))$$

Under P^2 ,

$$y(t) = \int_0^t (h(x(u)) + sb^T\pi(u))du + v^2(t)$$

where $v^2(t)$ is a (F_t, P^2) Wiener martingale defined by Girsanov's theorem. Substituting the above into the filtering equation (79) , we have

$$d\pi(t) = \left\{A + [h(x(t)) - (1-s)b^T\pi(t)](B - b^T\pi(t)I)\right\}\pi(t)dt + (B - b^T\pi(t)I)\pi(t)dv^2(t)$$
$$\pi(0) = \pi_0$$

(81)

We assume that (81) with the given diffusion equation

$$dx(t) = f(x(t))\, dt + g(x(t))\, dw(t)$$

forms a jointly Markov process under (F_t, P^2) .

Let $x_a(t) = [\pi(t)^T, x(t)]^T$ and let $\alpha(x_a) = \frac{s^2-s}{2}(b^T\pi)^2 + sb^T\pi h(x)$. Letting $r_s(u, x_a) = E^2(\exp(\int_s^t \alpha(x_a(\theta))\, d\theta) \mid x_a(u) = x_a)$, $r_s(u, x_a)$ satisfies

$$-\frac{\partial}{\partial u} r_s(u, x_a) = L r_s(u, x) + \alpha(u, x_a)\, r_s(u, x_a), \quad u \le t$$
$$r_s(t, x_a) = 1$$

where L is the infinitesmal generator for $x_a(t)$ under P^2.

3.2) Performance bounds in parameter estimation

Observation model of interest is given by

$$\begin{aligned} dx(t) &= f(x(t),\theta)\, dt + g(x(t),\theta)\, dw(t), \quad x(0) = x_0 \\ dy(t) &= h(x(t),\theta)\, dt + \sqrt{r}\, dv(t), \quad y(0) = 0. \end{aligned} \tag{82}$$

We assume the conditions stated in remarks of Section 2.1. We assume that θ also contains parameters that parameterizes initial densities of x_0 if necessary. The a posteriori density $p(\theta \mid y^t)$ is given by normalizing (2-32).

$$p(\theta \mid y^t) = \frac{p(\theta)\Lambda_t(\theta)}{\int_\Theta p(\theta)\Lambda_t(\theta)\, d\theta} \tag{83}$$

where $\Lambda_t(\theta)$ is the conditional likelihood ratio between an observation model corresponding to θ and $dy(t) = \sqrt{r}\, dv(t)$.

An implemented version of (83) is set by

$$p^*(\theta_j \mid y^t) = \frac{p(\theta_j)\Lambda_t(\theta_j)}{\sum_{i=0}^{N} p(\theta_i)\Lambda_t(\theta_i)}, \quad \theta_j \in A \tag{84}$$

where $P^*(\theta_j \mid y^t)$ is the implemented version of a posteriori density and $A = \left\{\theta_j\right\}_{j=0}^{j=N}$, $\theta_j \in \Theta$. The approximate Bayes conditional mean is defined by

$$\theta^*(t) = \sum_i \theta_i\, p^*(\theta_i \mid y^t) \tag{85}$$

Performance bounds of the estimator (85) is of interest in this section. Problems of this type was studied previously by Liporace [15] and Hawkes and Moore [16]. Liporace treated a parameter estimation for an i.i.d. sequence of observation random variables , and proves an exponential convergence rate of $E(p(\theta_j \mid y^t) \mid \theta)$ to 0 for $\theta^j \neq \theta_m$, where θ_m is the closest element (assuming an uniqueness) in A to θ (a true parameter value) under the Kullback information function. Hawkes and Moore studied the convergence rate of the estimator in a case of dependent observations characterized by a finite dimensional linear discrete gaussian system with unknown parameters, and prove the asymptotically exponential convergence rate of $E(p(\theta_j \mid y^t) \mid \theta)$ to 0 for θ_j θ by invoking the asymptotic per sample Kullback information function [16] [36].

Here we first treat a case when a true parameter value belongs to a finite set A chosen. Hence the parameter estimation problem is degenerated into a M -ary hypotheses detection , and the method of previous section is readily applied. We note that we fail to treat the case when θ lies outside of the set A , due to the difficulties that will be mentioned later.

Suppose $\theta \in A$ and $\theta = \theta_0$. Denote a conditional mean square error between $\theta^*(t)$ and θ_0 by $\sigma^2(t)$.

$$\sigma^2(t) = E^{\theta_0} \| \theta^*(t) - \theta_0 \|^2 \tag{86}$$

where $\| x \|^2 = x^T x$. Then

$$\begin{aligned} & E^{\theta_0} \| \theta^*(t) - \theta_0 \|^2 \\ & = \sum_{j=1}^{N} \sum_{i=1}^{N} (\theta_j - \theta_0)^T (\theta_i - \theta_0) E^{\theta_0} [p^*(\theta_j \mid y^t) p^*(\theta_i \mid y^t)] \end{aligned} \tag{87}$$

Since $p^*(\theta_j \mid y^t) \leq 1$, we have

$$p^*(\theta_j \mid y^t) p^*(\theta_i \mid y^t) \leq p^*(\theta_j \mid y^t) \leq p^*(\theta_j \mid y^t)^s ,$$

where $0 < s < 1$. Let $R = \underset{j}{Max} \| \theta_j - \theta_0 \|$. Thus the right-hand of (87) is bounded by

$$E^{\theta_0} \| \theta^*(t) - \theta_0 \|^2$$
$$\leq N^2 R^2 \underset{j \neq 0}{Max} E^{\theta_0}(p^*(\theta_j \mid y^t)^s)$$
$$\leq N^2 R^2 \underset{j \neq 0}{Max} \left\{ [\frac{p(\theta_j)}{p(\theta_0)}]^s \; E^{\theta_0} \frac{\Lambda_t^s(\theta_j)}{\Lambda_t^s(\theta_0)} \right\} \tag{88}$$

by deleting in the denominator of the right-hand side of (84) all the terms except the one indexed by θ_0 .

We note that $\dfrac{\Lambda_t(\theta_j)}{\Lambda_t(\theta_0)}$ is the likelihood ratio between observation models corresponding to θ_j and θ_0 . By proposition (3.1) , for $s \in (0,1)$

$$E^{\theta_0} \Lambda_t^s(\theta_j, \theta_0) \rightarrow 0 \;\text{ iff }\; \Lambda_t(\theta_j, \theta_0) \rightarrow 0 \;\; P^{\theta_0} a.s.$$

And according to the Kabanov's criterion , the above is true if and only if

$$P^{\theta_j}(\lim_{t \rightarrow \infty} \int_0^t (\hat{h}_u(\theta_0) - \hat{h}_u(\theta_j))^2 \, du = \infty) = 1 \tag{89}$$

Unfortunately we don't know the implications of (89) on the parameterized nonlinear system (82) .

If (89) holds for any θ and θ_j ($\in \Theta$) then

$$E^{\theta_0} \| \theta^*(t) - \theta_0 \|^2 \rightarrow 0$$

for any choice of A and θ_0 . We note that (88) holds for an approximate implementation of $\Lambda_t(\theta_j)'$ s . That is, if we use $T_t(\theta_j)$ for $j \in (0, \cdots, N)$ in (84) and (85) , then (88) holds with $\Lambda_t(\theta_j)$ being replaced by $T_t(\theta_j)$. The upper bound of (88) as a function of t can , in theory , be evaluated by the method of previous section by identifying the observations corresponding to θ_0 and θ_j as the ones under H_0 , H_1 respectively.

Suppose that the upperbound of (88) goes to 0 for any θ_0 , θ_j belonging to Θ . Then obviously,

$$E^{\theta_0} \frac{\Lambda_t^s(\theta_j)}{\Lambda_t^s(\theta_0)} < 1$$

for a sufficiently large t . Consider the Schwarz inequality.

$$E^{\theta_0}\left[\frac{\Lambda_t(\theta_j)}{\Lambda_t(\theta_0)}\right]^{1/2} \leq \int \Lambda_t(\theta_j)^{1/2}\Lambda_t(\theta)^{1/2}\,dP^0 \leq 1 \tag{90}$$

where $(\frac{1}{\sqrt{r}}y(t),P^0)$ is a Wiener martingale , and the equality holds iff $\Lambda_t(\theta_j) = \Lambda_t(\theta_0)$ P^0 a.s. Based on (90) one can conclude that if the upper bound of (88) goes to 0 for all θ_0 , θ_j ($\in \Theta$) , then for a sufficiently large t

$$\Lambda_t(\theta_j) = \Lambda_t(\theta_0) \;\; P^0 \text{ a.s.} \;\; \text{iff} \;\; \theta_j = \theta_0 \quad , \text{for all } \theta_j \, , \theta_0 \in \Theta \tag{91}$$

(91) alone is a necessary and sufficient condition to ensure a convergence of the upper bound of (88) to 0 in the case of i.i.d. sequence of observations , and is known as an identifiability condition [15]. Of course (91) reduces to a condition on a family of parameterized densities of an observable in the i.i.d. case. In our case we replace the density for i.i.d. observable by a Radon - Nikodym derivative , $\Lambda_t(\theta_j)$. Hence (91) is , at least , a necessary condition for (89) to hold, however we still do not know the implications of (91) on a nonlinear observation system (82) .

Now we consider a case when θ lies outside of A chosen. Notice that θ_0 is not a true value of parameter. Suppose we can implement likelihood ratios exactly. A modified form of (88) can be easily given, provided that $E^{\theta}\frac{\Lambda_t^s(\theta_j)}{\Lambda_t^s(\theta_m)}$ is bounded for $s \in (0,s')$ at each t.

$$E^{\theta}\,||\,\theta^*(t) - \theta_m\,||^2 \leq N^2R^2 \operatorname*{Max}_{j\neq m}\left\{\left[\frac{p(\theta_j)}{p(\theta_m)}\right]^s E^{\theta}\;\frac{\Lambda_t^s(\theta_j)}{\Lambda_t^s(\theta_m)}\right\} \tag{92}$$

where $\theta_m \in A$. For the inequality (92) to be meaningful in the present case, the Chernoff bound factor should go to 0 for all $j \neq m$. Thus θ_m should be such that

$$E^{\theta}\frac{\Lambda_t^s(\theta_j)}{\Lambda_t^s(\theta_m)} \rightarrow 0 \;\text{ for any }\; j \neq m \tag{93}$$

To characterize θ_m in a different way, suppose (93) is true in a given pair of (A , θ) . Then for a sufficiently large t

$$E^{\theta} \frac{\Lambda_t^{\theta}(\theta_j)}{\Lambda_t^{\theta}(\theta_m)} < 1 \text{ for all } j \neq m$$

Thus

$$E^{\theta} \ln(\frac{\Lambda_t^{\theta}(\theta_j)}{\Lambda_t^{\theta}(\theta_m)}) \leq \ln E^{\theta}(\frac{\Lambda_t^{\theta}(\theta_j)}{\Lambda_t^{\theta}(\theta_m)}) < 0$$

for a sufficiently large t . And hence for a sufficiently large t ,

$$E^{\theta} \ln(\Lambda_t(\theta_j)) < E^{\theta} \ln(\Lambda_t(\theta_m))$$

Defining $J(\theta_j : t) = E^{\theta} \ln(\Lambda_t(\theta_j))$ the above is

$$J(\theta_j : t) < J(\theta_m : t) \tag{94}$$

$J(\theta_j : t)$ is an analogous expression of Kullback information function for an i.i.d. observation [15] . That $J(\theta_j : t)$ is a proper measure of similarity is endorsed by the fact that the function is maximized at $\theta_j = \theta$. A simple proof is :

$$J(\theta_j : t) - J(\theta : t) = E^{\theta} \ln(\frac{\Lambda_t(\theta_j)}{\Lambda_t(\theta)})$$
$$\leq E^{\theta}(\frac{\Lambda_t(\theta_j)}{\Lambda_t(\theta)} - 1) = 0$$

with equality iff $\Lambda_t(\theta_j) = \Lambda_t(\theta)\ P^0$ a.s. In the above , we use the fact that $\ln x \leq x - 1$. Once (91) is true, $J(\theta_j : t)$ has an unique maximum at $\theta_j = \theta$ for a sufficiently large t .

Thus if (93) is true, then θ_m is necessarily an unique point that maximizes $J(\theta_j : t)$, for a sufficiently large t , in A . Or more strongly θ_m maximizes $\overline{J}(\theta_j)$ defined by

$$\overline{J}(\theta_j) = \lim_{t \to \infty} \frac{1}{t} J(\theta_j : t) \tag{95}$$

, provided that limit exists for all θ_j ($\in A$) . $\overline{J}(\theta_j)$ can be called as an asymptotic Kullback information function. (94) or (95) is useful only when one can establish inequalities between $J(\theta_j : t)'$ s for a sufficiently large t or $\overline{J}(\theta_j)'$ s . For the i.i.d. observation case or the case considered in [16] , (95) can be examined . Actually the existence of Kullback information function with the identifiability condition ensures that (93) is true for an arbitrary pair (A , θ) for an i.i.d. observation case under some mild condition. Returning to our problem, if one can find a mean to establish inequalities between $\bar{J}(\theta_j)'$ s , an uniquely maximizing element θ_m can be served as a candidate to examine (93). However we were not able to find such means.

For an approximate implementation of $\Lambda_t(\theta_j)$ the situation becomes more difficult to analyze since one can no longer invoke the Kullback information function to identify θ_m .

4. Conclusion

We have presented a general synthesis of the operations of detection and estimation for diffusion type signal models along the way to implement a sequential receiver. The approach we follow is jointly Bayesian in the sense of Middleton and Esposito. At first we have not made any prior assumption on the detailed nature of joint cost assignment except that it can be legitimately defined. For the dynamical signal model of section 2.1 we show that the simultaneously optimal receiver has the following inherent features:

- Jointly optimal rule exists within a class of nonrandomized rules under mild conditions on cost functions and on the observation model.
- The synthesized system is partitioned into two parts: one for solving the Zakai equation (2-34) recursively, the other being a two mutually coupled subsystems for detection and estimation respectively.
- Estimation operations precede the signal discrimination operation ,and the nature of estimates are Bayes estimates which take into account of hypotheses uncertainties.
- Jointly optimal decision rule obtained incorporates the results of estimates under temporary decisions.

Taking the particular cost assignment of the form (2-62) for which costs of misclassification and incorrect estimation are separate and additive, we show that several coupling structures that are operationally appealing can be derived and justified through a jointly Bayesian approach. The present cost assignment is taken to consider different natures of parameters in each signal models. However it seems controversial to naturally justify the application of jointly Bayesian approach ,with the present type of cost function, to a set of nuisance type parameters , since the resulting estimates of such parameters are usual Bayes estimates that are not affected by uncertainties of hypotheses. The nonlinear estimation approach of

Lainiotis deals with a general situation where parameter vector of concern are not restricted to be of same nature. However his approach essentially is a matter of the attitude on the observation model, and accepting the same viewpoint on the observation model we treat the result of Lainiotis as a particular case that can be derived by using a cost function (2-62).

As the second topic we have considered the performance analysis in binary hypotheses tests when signals of interest can be assumed to be generated by several kinds of finite dimensional nonlinear stochastic dynamical systems. Before proceeding with the evaluation of Chernoff bound, we show that Chernoff bound becomes, in general, sufficiently tight for a sufficiently long observation interval if the two measures associated with binary hypotheses become singular in the limit. Though several authors, notably Evans, addressed the use of evolution equation for the quantity closely related to Chernoff exponent, the essential core of this method involving measure transformation of Girsanov remained to be justified for a wide class of practical observation models. We first treat a discrimination of two fully observed diffusions and prove the validity of measure transformation for a sufficiently wide class of observation models by invoking a sufficient condition for the existence of unique weak solution, (3-40) to a stochastic differential equation. Thus, while the fully observed diffusion may be regarded as a somewhat degenerate case of partially observed one, we can obtain a relatively complete result. Our method of analysis is different from the one in [34] in that we take the evolution approach to evaluate the Chernoff exponent. We carry out the same approach to justify measure transformation for a partial observation of diffusion, assuming that an optimal likelihood ratio test is used. Here the stochastic differential equation of concern, (3-61), contains a drift term given by a causal functional, and we are only able to validate the measure transformation for linear systems. Also we present an alternative way to validate by using an evolution equation for an evaluation of the functional appearing in the

Novikov condition. When applied to a hypotheses test involving partial observation where H_0 is a null hypothesis, this method seems to be applicable to the case where a finite dimensional suboptimal filtering is used to implement a suboptimal test statistic having an unbounded $h^{*1}(x)$ of (3-76).

We have also treated performance analysis for hypotheses tests involving two fully or partially observed discrete time Markov signals. In the full observation case we show the asymptotically exponential convergence rate of Chernoff exponent, provided that the function $H(x,y:s)$ of (3-22) allows a series expansion of (3-23). Here we use the measure transformation , though not required, to obtain an evolution equation (3-22) in a way analogous to that for a diffusion. In the partial observation case we derive the backward evolution equation of the quantity governing the performances, assuming a particular suboptimal statistic.

Several examples are given, illustrating the use of Chernoff bound for a discrete time or continuous time observation, and for an optimal or suboptimal test statistic.

Performance analysis in parameter estimation, under no hypothesis uncertainty, is treated from a detection point of view by discretizing a parameter space. We straightforwardly extend the previous work of Liporace for identical and independent observation to the case of a partial observation of diffusion, and bound the quadratic losses by (3-88) or (3-92) respectively, depending on whether a true value of parameter belongs to a given grid discretizing a parameter space. However our present treatment is incomplete due to difficult issues such as the identifiability condition, compared to earlier works by Liporace, Hawkes and Moore.

References

[1] D.Middleton and R.Esposito, "Simultaneous Optimum Detection and Estimation of Signals in Noise," IEEE Trans. on Inform. Theory, Vol.14, No.3, May 1968, pp. 434-444

[2] A.Frederickson, D.Middleton, and D.Vandelinde, "Simultaneous Signal Detection and Estimation under Multiple Hypotheses," IEEE Trans. on Inform. Theory, Vol.18, No.5, Sep. 1972, pp. 607-614

[3] N.Nilsson, "On the Optimum Range Resolution of Radar Signals in Noise," IRE Trans. on Inform. Theory, 1961, pp. 245-253

[4] B.Levin and Y.Shinakov, "Bayesian System for the Simultaneous Discrimination of Several Signals and Estimation of Their Parameters," Telecomm. and Radio Engrg., Vol.26, No.4, 1971, pp. 67-70

[5] A.Jaffer and S.Gupta, "Coupled Detection and Estimation of Gaussian Processes in Gaussian Noise," IEEE Trans. on Inform. Theory, Vol.18, 1972, pp. 106-110

[6] Y.Sosulin and A.Shakuhurin, "Simultaneous Detection and Estimation of Stochastic Signals on a Background Noise," Engrg. Cybernatics, 1976, No.1, pp.109-115

[7] D.Lainiotis, "Joint Detection,Estimation and System Identification," Information and Control, Vol.19, 1971, pp. 75-92

[8] G.Kallianpur, "Stochastic Filtering Theory," Springer-Verlag, 1980

[9] J.Baras, " Simultaneous Detection and Estimation for Diffusion Type Signals," SRR-83, University of Maryland

[10] T.Birdsall and J.Gobien, "Sufficient Statistics and Reproducing Densities in Simultaneous Sequential Detection and Estimation," IEEE Trans. on Inform.

Theory, Vol.19, Nov. 1973, pp. 760-768

[11] B.Levin and Y.Shinakov, "Asypmtototic Properties of Bayes Estimates of Parameters of a Signal Masked by Interference," IEEE. Trans. on Inform. Theory, Vol.18, No-1, Jan. 1972, pp. 102-106

[12] J.Evans, "Chernoff Bounds on the Error Probability for the Detection of Nongaussian Signals," IEEE Trans. on Inform. Theory, Vol.20, No.5, 1974, pp. 569-577

[13] J.Hibey, D.Snyder and J.van Schuppen, "Error Probability Bounds for Continuous Time Decision Problems," IEEE Trans. on Inform. Theory, Vol.24, No.5, Sep. 1978, pp. 608-622

[14] J.Hibey, "Error Probability Bounds for Continuous Time Decision Problems," D.Sc dissertation, Dep. Sys.Sc and Math., Washington Univ. Aug. 1976

[15] L.Liporace, "Variance of Bayes Estimates," IEEE Trans. on Inform. Theory, Vol.17, No.6, Nov. 1971, pp. 665-669

[16] R.Hawkes and J.Moore, "Performance Bounds for Adaptive Estimation ," Proceedings of IEEE, Vol.64, No.8, Aug. 1976, pp.1143-1150

[17] R.Lipster and A.Shiryayev, "Statistics of Random Process I," Springer-Verlag, 1976

[18] J.Baras, G.Blankenship and W.Hopkins, "Existence, Uniqueness, and Asymptotic Behavior of Solutions to a Class of Zakai Equations with Unbounded Coefficients," IEEE Trans. on Auto. Control, Vol.28 No.2, Feb. 1983, pp. 203-214

[19] E.Pardoux, "Nonlinear Filtering, Prediction and Smoothing," in The Mathematics of Filtering and identification and Applications (Nato Advanced Study Institute Series), Dordrecht, The Netherlands: Reidel, 1981, pp. 529-557

[20] E.Wong, "Stochastic Processes in Information and Dynamical Systems," ch.6, McGraw-Hill, 1971

[21] T.Kailath, "Generalized Likelihood Ratio Formula," IEEE Trans. on Inform. Theory, Vol-15, 1969

[22] L.Ljung, "Asymptotic Behavior of the Extended Kalman Filter as a Parameter Estimator for Linear Systems," IEEE Trans. on Auto. Control, Vol.24 No.1, Feb. 1979, pp. 36-50

[23] E.Tse and J.Anton, "On the Identifiability of Parameters," IEEE Trans. on Auto. Control, Vol.17 No.5, Oct. 1972, pp. 637-646

[24] H.Kushner, "Probability Methods for Approximations in Stochastic Control and for Elliptic Equations," Ch.7, Academic Press, 1977

[25] R.Stratonovich, "Detection and Estimation of Signals in Noise when One or Both are Nongaussian," Proceedings of IEEE, Vol.58, No.5, May 1970, pp. 670-679

[26] R.Bucy and P.Joseph, "Filtering for Stochastic Processes with Application to Guidance," Wiley, 1968

[27] R.Bucy and K.Senne, "Digital Synthesis of Nonlinear Filters," Automatica, Vol.7, 1971 , pp. 287-298

[28] B.Levin and Y.Shinakov, "Jointly Optimum Algorithms of Signal Detection and Parameter Estimation (a Review)," Radio Eng. and Electron. Phys., Vol. , 1977, pp. 1-15

[29] T.Kadota, "Optimal Causal Simultaneous Detection and Estimation of Random Signal Fields in a Gaussian Noise Field," IEEE Trans. on Inform. Theory, Vol.24, No.3, May 1978, pp. 637-646

[30] A.Shiryayev, "Probability," Springer-Verlag, 1984

[31] R.Ash, " Real Analysis and Probability," Academic Press, 1972

[32] A.Shiryayev, " Martingales: Recent Developments, Results and Applications ," International Statistical Review, Vol.49, 1981, pp. 199-233

[33] Zakai, "On the Optimal Filtering of Diffusion Processes," Z.Wahrscheinlichkeitstheorie verw. Geb.11, 1969, pp. 230-243

[34] C.Newman and Stuck, "Chernoff Bounds for Discriminating Between Two Markov Processes," Stochastics, Vol.2, 1979, pp. 139-153

[35] S.Schwarz, "A Series Technique for the Optimum Detection of Stochastic Signals in Noise," IEEE Trans. on Inform. Theory, Vol.15, No.3, May 1969, pp. 362-369

[36] R.Hawkes, "Performance Bound for Bayes Estimates," IEEE Trans, on Inform. Theory, Sep. 1976, pp. 610-615

[37] H. Van Trees, "Detection, Estimation, and Modulation," Part 1 and 3, Wiley, 1971

[38] T.Kailath and M.Zakai, "Absolute Continuity and Radon-Nikodym Derivatives for Certain Measures Relative to Wiener Measures," Annals of Mathematical Statistics, Vol.42, No.1, 1971, pp. 130-140