Proceedings of the 26th Conference
on Decision and Control
Los Angeles, CA • December 1987

Real-Time Sequential Detection for Diffusion Signals

John S. Baras* and Anthony LaVigna**
Electrical Engineering and Systems Research Center
University of Maryland, College Park MD 20742

Abstract

We present the sequential detection problem for diffusion type signals in both the fixed probability of error and the Bayesian formulations. The optimal strategy in both cases is a threshold policy with explicitly computable thresholds. We provide numerical schemes for approximating the relevant likelihood ratio and provide an architecture for real time signal processing.

Introduction

In this paper we discuss the design and implementation of a fundamental nonlinear signal processor. Specifically, we present the problem of sequential hypothesis testing for signals that can be modeled with nonlinear dynamics; we show the optimal detector for the Bayesian and fixed probability of error costs; and, we show a real-time implementation of the optimal detector using systolic arrays.

The techniques above differ from other nonlinear signal processing techniques in that they implement the mathematically optimal detector. They do not make the standard linear-Gaussian assumptions of Kalman filtering. However, they do assume a given nonlinear signal model. In many cases, it is possible to obtain this signal model from the physics or a priori knowledge of the system. In other cases, statistical techniques must be used to determine the signal model. (The latter case is still an area of active research.)

This paper is organized as follows: the basic problem and notation is presented in Section 1; both the Bayesian and fixed probability of error problems are solved in Sections 2 and 3, repectively; the numerical techniques used to solve both problems are presented in Section 4; and, a systolic array implementation is presented in Section 5.

1. Basic Sequential Detection Problem

Throughout this paper, we consider the binary sequential hypothesis testing problem. Here, we are given a vector-valued signal $x_t, x_t \in \mathbb{R}^n$, which satisfies the stochastic differential equation

$$\begin{aligned} dx_t &= f(x_t)\,dt + g(x_t)\,dw_t \\ x_0 &= \nu \end{aligned} \tag{1.1}$$

where w_t is a vector standard Brownian motion. Unfortunately, we cannot observe x_t directly, instead we only observe y_t, a vector-valued stochastic process, $y_t \in \mathbb{R}^P$. Under each hypothesis the observed data is the output of a stochastic differential equation, i.e.,

$$\begin{aligned} &\text{Under } H_1: \quad dy_t = h(x_t)\,dt + dv_t \\ &\text{Under } H_0: \quad dy_t = dv_t \end{aligned} \tag{1.2}$$

where v_t is another standard Brownian motion which is independent of w_t. Notice that if $f(\cdot)$ and $h(\cdot)$ are linear and $g(\cdot)$ is constant then this becomes a standard problem which can be solved by the Kalman filter.

Data is observed continuously starting at an initial time which is taken for convenience to be zero. We let $\mathcal{F}_t^y$ represent the information collected up to time t. At each time t, the decision-maker can either stop and declare one of the hypotheses to be true or can continue collecting data. We let τ represent the termination time and δ represent the decision. The decision-maker selects his decision based on the current information ,$\mathcal{F}_t^y$, so as to minimize an appropriate cost function.

For both the fixed probability of error and the Bayesian formulations of this problem, we have a measurable space $(\Omega, \mathcal{F})$, on which we are given two probability measures P_0, P_1, and the random process $\{y_t,\ t \geq 0\}$. When hypothesis H_0 (respectively H_1) is valid the statistics of the observed process $\{y_t,\ t \geq 0\}$ are governed by measure P_0 (respectively P_1). We let E_1 (respectively E_0) denote the expectation under P_1 (respectively P_0).

More precisely, an *admissible decision policy* is any pair $u = (\tau, \delta)$ of RV's where τ is a $\mathcal{F}_t^y$ stopping time and δ is a $\{0,1\}$-valued $\mathcal{F}_\tau^y$ measurable RV.

An admissible policy $u = (\tau, \delta)$ is a *threshold policy* or *of threshold type* if there exists constants A and B, with $0 < A \leq 1 \leq B < \infty$ and $A \neq B$, such that

$$\tau = \inf(t \geq 0 \mid \Lambda_t \notin (A, B)) \tag{1.3}$$

$$\delta = \begin{cases} 1, & \Lambda_\tau \geq B \\ 0, & \Lambda_\tau \leq A \end{cases} \tag{1.4}$$

Here Λ_t is the likelihood ratio associated with this problem, namely

$$\Lambda_t = \exp\Big(\int_0^t \hat{h}_s^T\,dy_s - \frac{1}{2}\int_0^t \|\hat{h}_s\|^2\,ds\Big), \tag{1.5}$$

and

$$\hat{h}_t = E_1(h(x) \mid \mathcal{F}_t^y). \tag{1.6}$$

Threshold policies will be described as above and will be identified with their threshold constants (A, B).

We prefer threshold policies because of their simple computational structure. The likelihood ratio, Λ_t, is computed and compared to the interval (A, B) at each time t. This process is stopped the first time Λ_t is larger than B or smaller than A.

The sequential likelihood ratio, Λ_t, has continuous sample paths, therefore for threshold policies, Λ_τ takes on the values A

or B (P_0- and P_1-as). Using this fact, together with Girsanov's theorem, it can be shown [8] that for threshold policies

$$P_0(\delta=1)=\frac{1-A}{B-A} \qquad P_1(\delta=0)=\frac{A(B-1)}{B-A}.$$

Some algebra gives the following result.

Let u be a threshold policy with A and B defined by

$$A=\frac{\beta}{1-\alpha} \qquad B=\frac{1-\beta}{\alpha} \tag{1.7}$$

where $\alpha+\beta<1$, then

$$P_0(\delta=1)=\alpha \qquad P_1(\delta=0)=\beta. \tag{1.8}$$

Hence, given desired false alarm and miss probabilities, α,β, it is possible to find thresholds, (A,B), so that the resulting threshold policy has the required probabilities.

The following can also be shown [8].

Let $u=(\tau,\delta)$ be any admissible policy with $P_0(\delta=1)+P_1(\delta=0)<1$ and let

$$\alpha:=P_0(\delta=1) \qquad \beta:=P_1(\delta=0). \tag{1.9}$$

Define $u^*=(\tau^*,\delta^*)$ to be the threshold policy with parameters (A^*,B^*) that correspond to the pair (α,β) as defined by (1.7). Then

$$E_0\Big(\int_0^{\tau^*}\|\hat{h}_s\|^2\,ds\Big)=2w(\alpha,\beta) \tag{1.10}$$

$$E_1\Big(\int_0^{\tau^*}\|\hat{h}_s\|^2\,ds\Big)=2w(\beta,\alpha) \tag{1.11}$$

where

$$w(x,y):=(1-x)\log\frac{1-x}{y}+x\log\frac{x}{1-y}. \tag{1.12}$$

2. Fixed Probability of Error Formulation

Given $0<\alpha,\beta<1$ with $\alpha+\beta<1$, let $\mathcal{U}(\alpha,\beta)$ be the set of all admissible policies u such that

$$P_0(\delta=1)\le\alpha \qquad P_1(\delta=0)\le\beta. \tag{2.1}$$

The fixed probability of error formulation to the sequential hypothesis testing problem requires the solution of the following.

Problem (P_F): *Find u^* in $\mathcal{U}(\alpha,\beta)$ such that for all u in $\mathcal{U}(\alpha,\beta)$,*

$$E_i\Big(\int_0^{\tau}\|\hat{h}_s\|^2\,ds\Big)\ge E_i\Big(\int_0^{\tau^*}\|\hat{h}_s\|^2\,ds\Big), \quad i=0,1. \tag{2.2}$$

The term in the expectation above represents the expected signal energy present. Usually, the observation time is minimized, subject to the error probability constraints (2.1). However, in this problem you cannot always be sure that you receive "good" data because $h(x_t)$ is itself a random process. It is clear that the longer you observe y the more energy signal you receive. Therefore, trying to decide "faster" is related to trying to decide while receiving the "minimum" signal energy. This intuitive idea is captured in (2.2).

Theorem 2. *If u^* is the threshold policy with constants (A^*,B^*) defined by*

$$A^*=\frac{\beta}{1-\alpha}, \qquad B^*=\frac{1-\beta}{\alpha}, \tag{2.3}$$

then u^* solves problem (P_F).

Proof: (Sketch, for details see [4] or [10]).

1. $E_1\big(\int_0^{\tau}\|\hat{h}_s\|^2\,ds\big)=-2\,E_1(\log\Lambda_\tau^{-1})$
2. From Jensens inequality we have that

$$\begin{aligned}E_1\Big(\frac{1}{2}\int_0^{\tau}\|\hat{h}_s\|^2\,ds\Big) &= -P_1(\delta=1)\,E_1(\log\Lambda_\tau^{-1}\mid\delta=1)- \\ &\quad -P_1(\delta=0)\,E_1(\log\Lambda_\tau^{-1}\mid\delta=0) \\ &\ge -P_1(\delta=1)\log E_1(\Lambda_\tau^{-1}\mid\delta=1)- \\ &\quad -P_1(\delta=0)\log E_1(\Lambda_\tau^{-1}\mid\delta=0)\end{aligned}$$

3. After some algebra,

$$\begin{aligned}E_1\Big(\frac{1}{2}\int_0^{\tau}\|\hat{h}_s\|^2\,ds\Big) &\ge (1-\beta)\log\frac{1-\beta}{\alpha}+\beta\log\frac{\beta}{1-\alpha} \\ &= E_1\Big(\frac{1}{2}\int_0^{\tau^*}\|\hat{h}_s\|^2\,ds\Big)\end{aligned}$$

4. A similar technique works for the case with E_0.

3. Bayesian Formulation

For the Bayesian formulation, let H be an $\{0,1\}$-valued RV indicating the true hypothesis. By φ we denote the *a priori* probability that hypothesis H_1 is true. We consider a probability measure P on $(\Omega,\mathcal{F})$ such that

$$P(H=1)=\varphi \qquad P(H=0)=1-\varphi, \tag{3.1}$$

and such that for every $A\in\mathcal{F}$

$$P(A)=\varphi P_1(A)+(1-\varphi)P_0(A), \tag{3.2}$$

where P_1 and P_0 are the measures defined in Section 1.

We shall assume, two costs are incurred. The first cost is for observation and is accrued according to $k\int_0^t\|\hat{h}_s\|^2\,ds$, where $k>0$ and $\{\hat{h}_t,\ t\ge0\}$ is defined by (1.6). The second cost is associated with the final decision δ and is given by

$$C(H,\delta)=\begin{cases}c_1, & \text{when } H=1 \text{ and } \delta=0;\\ c_2, & \text{when } H=0 \text{ and } \delta=1;\\ 0, & \text{otherwise,}\end{cases} \tag{3.3}$$

where $c_1>0$ and $c_2>0$.

We are interested in minimizing the expected cost. If $u=(\tau,\delta)$ is any admissible policy, then the corresponding expected cost is

$$J(u)=E\Big(k\int_0^{\tau}\|\hat{h}_s\|^2\,ds+C(H,\delta)\Big). \tag{3.4}$$

The Bayesian approach to sequential detection seeks to find the solution to the following problem.

Problem (P_B): *Given* $\varphi \in (0,1)$, *find* u^* *such that,*

$$J(u^*) = \inf_{u \in \mathcal{U}} J(u). \tag{3.5}$$

It can be shown that to any admissible policy u there corresponds a threshold policy which has no greater cost, $J(u)$, therefore, the infimum in (3.5) need only be computed over threshold policies. In fact, it can be shown that the infimum is obtained and the following theorem results.

Theorem 3. *There exists an admissible threshold policy* u^* *that solves problem* (P_B). *The optimal thresholds* $0 < A^* \leq 1 \leq B^* < \infty$ *with* $A^* \neq B^*$, *are given by the relations*

$$A^* = \left(\frac{1-\varphi}{\varphi}\right)\left(\frac{a^*}{1-a^*}\right), \qquad B^* = \left(\frac{1-\varphi}{\varphi}\right)\left(\frac{b^*}{1-b^*}\right).$$

where a^* *and* b^* *are the unique solutions of the transcendental equations*

$$c_2 + c_1 = k(\Psi'(a^*) - \Psi'(b^*))$$

$$c_2(1-b^*) = c_1 a^* + (b^* - a^*)(c_1 - k\Psi'(a^*)) + k(\Psi(b^*) - \Psi(a^*)),$$

with

$$\Psi(x) = (1-2x)\log\frac{x}{1-x}.$$

satisfying $0 < a^* < b^* < 1$.

Here again the thresholds are unique functions of the cost parameters. For a detailed proof we refer to [4] or [10].

The results for the hypothesis testing problem when H_0 contains a signal, in addition to noise, require a little modification. Here, $\{x_t^1,\ t \geq 0\}$ and $\{x_t^2,\ t \geq 0\}$ represent different signal vector-valued processes. Under each hypothesis the observed data, y_t, is the output of a stochastic differential equation,

$$\begin{aligned} &\text{Under } H_1: \quad dy_t = h^1(x_t^1)\,dt + dv_t \\ &\text{Under } H_0: \quad dy_t = h^2(x_t^2)\,dt + dv_t \end{aligned} \tag{3.6}$$

The solutions to the Bayesian and fixed probability of error formulations are still valid with Λ_t defined by

$$\Lambda_t = \exp\Big(\int_0^t (\hat{h}_s^1 - \hat{h}_s^2)^T\,dy_s - \frac{1}{2}\int_0^t (\|\hat{h}_s^1\|^2 - \|\hat{h}_s^2\|^2)\,ds\Big). \tag{3.7}$$

4. Numerical Solution

In this section we will discuss the numerical method used to approximate Λ_t. It can be shown that $\Lambda_t = \int_{\mathbb{R}} u(x,t)\,dx$ where $u(x,t)$ is the solution to the *Zakai equation*. Our strategy will be to find an approximation to $u(x,t)$ which results in a good approximation to Λ_t.

From the theory of nonlinear filtering, it is known [5] that under the appropriate conditions on f, g, and h, the unnormalized density of x given the observations y satisfies the *linear* stochastic partial differential equation, known as the *Zakai equation*, given below

$$du(x,t) = L^* u(x,t)\,dt + u(x,t)\,h(x)^T\,dy_t \tag{4.1}$$

$$u(x,0) = p_0(x). \tag{4.2}$$

When x is scalar

$$\begin{aligned} L^* u(x,t) &= \frac{1}{2}\frac{d^2}{dx^2}\left[g^2(x)\,u(x,t)\right] - \frac{d}{dx}\left[f(x)\,u(x,t)\right] \qquad (4.3)\\ &= a(x)\,u_{xx}(x,t) + b(x)\,u_x(x,t) + c(x)u(x,t) \\ &= A^* u(x,t) + c(x)u(x,t) \qquad (4.4) \end{aligned}$$

and

$$\begin{aligned} a(x) &= \frac{1}{2}g^2(x) \\ b(x) &= g(x)\,g'(x) - f(x) \\ c(x) &= g(x)\,g''(x) + (g'(x))^2 + g(x)\,g'(x) - f'(x) \end{aligned} \tag{4.5}$$

with $p_0(x)$ the initial density of x.

In general, it is not possible to explicitly solve (4.1). Therefore, we will approximate its solution using finite difference methods. Since (4.1) is a linear parabolic equation we use an implicit discretization for x-derivatives, in order to maintain stability. The general vector valued x case is treated in [10]. Here we discuss the scalar valued x case, due to space limitations. The solution is approximated on the internal $D = (a,b)$.

Let $\Delta x > 0$ and define $x_k = a + k\,\Delta x$ and n such that $x_n \leq b$. Consider the collection of points $\{x_k\}_0^n$ in D. Let $\Delta t > 0$ and define $t_k = k\,\Delta t$.

The value of $u(x,t)$ at the grid point (x_i, t_k) is represented by v_i^k. We replace the x-derivatives in A^* with implicit finite difference approximations. To this end,

$$\begin{aligned} a(x)\,u_{xx}(x,t) \quad &\sim \quad a(x_i)\left(\frac{v_{i+1}^{k+1} - 2v_i^{k+1} + v_{i-1}^{k+1}}{(\Delta x)^2}\right) \\ b(x)\,u_x(x,t) \quad &\sim \quad b(x_i)\left(\frac{v_{i+1}^{k+1} - v_i^{k+1}}{\Delta x}\right) \text{ if } b(x_i) > 0 \\ &\qquad\quad b(x_i)\left(\frac{v_i^{k+1} - v_{i-1}^{k+1}}{\Delta x}\right) \text{ if } b(x_i) < 0 \\ u_t(x,t) \quad &\sim \quad \frac{v_i^{k+1} - v_i^k}{\Delta t} \end{aligned} \tag{4.6}$$

Let V^k represent the vector of mesh points at time $k\,\Delta t$. Then the above approximations result in a matrix A_n which approximates A^*. The approximation takes the form

$$(I - \Delta t\,A_n)V^{k+1} = V^k + \text{ other terms.}$$

Note that $a(x_i)$ is always positive and the special way of choosing the first derivative approximation guarantees that the matrix A_n is diagonally dominant and of the form

$$A_n = \begin{pmatrix} - & + & & \\ + & \ddots & \ddots & \\ & \ddots & \ddots & + \\ & & + & - \end{pmatrix}. \tag{4.7}$$

Therefore $(I - \Delta t\, A_n)$ is strictly diagonally dominant. In fact, it is also inverse positive, i.e., every element of the inverse is positive [7].

The final step is to approximate the solution of (4.1) assuming $A^* = 0$. This leads to the matrix

$$D_k = \text{diag}(e^{h(x_i)\,\Delta y_k + (c(x_i) - \frac{1}{2}h(x_i)^2)\,\Delta t}) \tag{4.8}$$

with $\Delta y_k = y_{(k+1)\Delta t} - y_{k\,\Delta t}$.

The overall approximation is

$$(I - \Delta t\, A_n)V^{k+1} = D_k V^k. \tag{4.9}$$

This approximation can be shown to be convergent ([3], [4], and [6]).

A nice property of the above approximation is that it is positive preserving. Regardless of the relationship of Δt and Δx, the solution V^k is always positive. This is important since we are approximating a probability density which we know can never be negative.

Schemes similar to (4.9) have been discussed in ([3], [6]). Furthermore, numerical studies have been performed in [9] using these methods which have produced satisfactory results for approximations to $\hat{h}_t$.

Using V^k defined in (4.9) it is easy to construct a convergent approximation to the likelihood ratio. Let $\{x_k\}_0^n$ be the collection of points defined above and let

$$u_n(x,t) = v_i^k \qquad \text{if} \quad x_i \le x < x_{i+1}$$
$$k\Delta t \le t < (k+1)\Delta t$$

and define

$$\Lambda_t^n = \int_a^b u_n(x,t)\,dx = \sum_{i=0}^{n} v_i^k\,\Delta x$$

then Λ_t^n is a convergent approximation to Λ_t.

5. Architectures for Signal Processor

The finite difference scheme used to approximate the solution of (4.1) involves solving the linear equation

$$(I - \Delta t A_n)\,V^{k+1} = D_k V^k, \tag{5.1}$$

for each time $t = k\Delta t$. Here D_k is a data dependent diagonal matrix. We have chosen the *systolic array architecture* of [2] to implement this scheme.

The number of processors in a systolic array solving the system in (5.1) depends only on the bandwidth of the matrix and not on its dimension. This makes a systolic design ideal for implementing *finite difference* approximations, where the number of mesh points is not known a priori but where the maximum bandwidth is determined by the specific scheme.

To solve (5.1), we use Gaussian elimination without pivoting. This method, which is stable since $(I - \Delta t\, A_n)$ is strictly diagonally dominant, results in matrices L and U such that

$$LU = (I - \Delta t\, A_n) \tag{5.2}$$

where U is an upper triangular matrix and L is a unit lower triangular matrix.

The matrices L and U will be bi-diagonal since no pivoting is used. As is standard with Gaussian elimination, once the factors L and U are found, (5.1) is solved by back substitution.

It is well known that back substitution can be accomplished by systolic arrays [2]. The algorithm takes $2n + 2$ time units to operate. Here n is the dimension of the matrix. Hence we are able to solve (5.1) in $5n+4$ time units. The extra n units result from a necessary buffering of the data between the two back substitution operations.

Notice that the matrix, $(I - \Delta t\, A_n)$, is independent of the received data, y_t, therefore, the LU factorization only has to be done at the time of filter design.

The overall layout for the sequential detector is given in Figure 1. Both solve operations are performed by a systolic array.

Discussion

Work is currently under way on a board level design implementing the above systolic algorithm. This board will use current digital signal processing chips and will be installed in an IBM PC. With current technology, this board will be capable of receiving and processing data at a rate of 20 kHz.

This architecture solves a long standing problem for scalar state and observation models. It can be extended to vector observations of a scalar state. However, the extension to observations of vector states is very hard. We have shown that a systolic architecture works as well for states of dimension 2. For higher state dimensions, a different architecture is needed. Further information is contained in [1], where multigrid methods are used and a proposed multi-level architecture results.

References

[1] J. S. Baras & K. Holley, "Multi-grid methods for the Zakai equation and related real-time architectures," 1987.

[2] H. T. Kung & C. E. Leiserson, "Algorithms for VLSI processor arrays," in *Introduction to VLSI Systems*, C. Mead and L. Conway, ed., Addison-Wesely, Reading, Massachusetts, 1980, 271–292.

[3] H. J. Kushner, *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press, New York, NY, 1977.

[4] A. LaVigna, "Real time sequential hypothesis testing for diffusion signals," Masters Thesis, University of Maryland, 1986.

[5] R. S. Liptser & A. N. Shiryayev, *Statistics of Random Processes I : General Theory* (English Translation), Springer-Verlag, New York–Heidelberg–Berlin, 1977.

[6] E. Pardoux & D. Talay, "Discretization and simulation of stochastic differential equations," in *Publication de Mathematiques Appliquees Marseille-Toulon*, Université de Provence, Marseille, 1983.

[7] J. Schröder, "M-matrices and generalizations using an operator theory approach," *SIAM Review* 20 (1978), 213–244.

[8] A. N. Shiryayev, *Optimal Stopping Rules* (English Translation), Springer-Verlag, New York–Heidelberg–Berlin, 1977.

[9] Y. Yavin, "Numerical Studies in Nonlinear Filtering," in *Lecture Notes in Control and Information Sciences 65*, A. V. Balakrishnan and M. Thoma, ed., Springer-Verlag, New York–Heidelberg–Berlin, 1985.

[10] J. S. Baras & A. LaVigna, "Real-Time Sequential Detection and Estimation for Diffusion Signals," *submitted for publication.*

* The work of this author was supported partially through NSF Grant NSFD CDR-85-00108 and partially through ONR Grant N00014-83-K-0731.

** The work of this author was supported by an ONR Fellowship.