Architectures for Real-Time Sequential Detection

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

Introduction

In this paper we discuss an application and implementation of a nonlinear signal processor. This processor implements the natural mathematical generalization of the well known Kalman filter. 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 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 signal model. In many cases, it is possible to obtain this signal model from the physics or apriori knowledge of the system. In other cases, powerful statistical techniques must be used to determine the signal model. (The latter case is still an area of active research.)

Applications of these results to radar problems have been considered and implementations are currently being pursued. For example, if we sample the continuous time return at a sampling period equal to the pulsed radars interpulse interval, models like (1.1) and (1.2) can be used to represent the received samples from ships or chaff-clouds (Baras [1978]). Actually, such models have even been implemented in hardware at NRL simulation facilities. One such application is described in (Baras [1978]). This is an example of the applicability of our result to real time target discrimination in radar systems.

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 scalar-valued signal x_t which satisfies the stochastic differential equation

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

where n_t is a white noise signal. Unfortunately, we cannot observe x_t directly, instead we only observe y_t, a scalar-valued stochastic process. Under each hypothesis the observed data is the output of a stochastic differential equation, i.e.,

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

where $\bar{n}_t$ is another white noise signal which is independent of n_t. Notice that if $f(\cdot)$ and $h(\cdot)$ are linear and $g(\cdot)$ is constant then this becomes the standard problem solved by the Kalman filter.

Data is observed continuously starting at an initial time which is taken for convenience to be zero. At each time $t > 0$, the decision-maker can either stop and declare one of the hypotheses to be true or can continue collecting data. The decision-maker selects his decision based on the data collected up to time t, so as to minimize an appropriate cost function. To summarize, the decision maker makes two decisions, when to stop (represented by τ) and once stopped, which hypothesis is true (represented by δ).

* 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.

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 $\{y_t,\ t \geq 0\}$. Let E_1 denote the expectation when P_1 is used; similarly for E_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).

More precisely, a decision policy involves the selection of a termination time τ, and of a binary valued decision δ. If $\delta = 1$, we shall accept hypothesis H_1; if $\delta = 0$ we shall accept hypothesis H_0. An *admissible decision policy* is any pair $u = (\tau, \delta)$ of RV's where τ and δ depend only on past observations. The collection of all admissible decision policies will be denoted by $\mathcal{U}$.

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

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

and

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

Pictorially, this means as Λ_t is computed for each time t it is compared to the interval (A, B). The process is stopped the first time Λ_t is larger than B or smaller than A (See Figure 1).

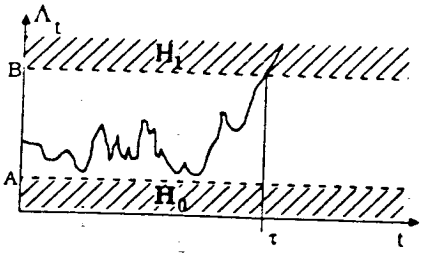

Figure 1. *Typical plot of the likelihood ratio*

A threshold policy will be described as above and will be identified with the threshold constants (A, B).

The sequential likelihood ratio, Λ_t has continuous sample paths, therefore for threshold policies, Λ_τ takes on the values A or B. Together with Girsanov's theorem this implies that the false alarm and miss probabilities of a threshold policy are explicit functions of A and B (Shiryayev [1977]).

This result has several technical consequences, which we will state without proof. The proofs are contained in (Shiryayev [1977]).

(1) 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

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

(2) Let $u = (\tau, \delta)$ be any policy in $\mathcal{U}$ 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 in (1.7). Then

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

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

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

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

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* $\mathcal{U}(\alpha, \beta)$,

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

The term in the expectation above represents the expected energy present. Usually, the observation time is minimized, subject to the error probability constraints (2.1). However, 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" 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).

The proof of this theorem follows from (2) above and some tedious inequalities.

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, i.e.,

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

For every $A \in \mathcal{F}$, we have

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

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 $\{\hat{h}_t,\ t \geq 0\}$ 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 average cost. If $u = (\tau, \delta)$ is any admissible policy, then the corresponding average 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,*

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

It can be shown that for any admissible policy u there exists a corresponding threshold policy which has no greater cost, 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 a threshold policy u^* in Γ 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 $0 < a^* < b^* < 1$.

The theorem above tells us how to construct the optimal sequential detector.

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 processes. Under each hypothesis the observed data, y_t, is the output of a stochastic differential equation,

$$\begin{aligned} &\text{Under } H_1: && \frac{dy_t}{dt} = h^1(x_t^1) + \bar{n}_t \\ &\text{Under } H_0: && \frac{dy_t}{dt} = h^2(x_t^2) + \bar{n}_t. \end{aligned} \tag{3.6}$$

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

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

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 a good approximation to $u(x,t)$ and then use it to approximate Λ_t.

From the theory of nonlinear filtering, it is known (Liptser & Shiryayev [1977]) that the unnormalized density of x given the observations y satisfies the *linear* stochastic partial differential equation, known as the *Zakai equation*, given below for the scalar case

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

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

$$L^*u(x,t) = \frac{1}{2}\frac{d^2}{dx^2}\left[g^2(x)\,u(x,t)\right] - \frac{d}{dx}\left[f(x)\,u(x,t)\right] \tag{4.3}$$

where $p_0(x)$ is the initial density of x.

In general, it is not possible to solve (4.1). Therefore, we will use finite difference methods to approximate its solution. Furthermore, since (4.1) is a parabolic equation we use an implicit discretization for x-derivatives.

Using the chain rule on (4.1) we get,

$$\begin{aligned} L^*u(x,t) &= a(x)\,u_{xx}(x,t) + b(x)\,u_x(x,t) + c(x)u(x,t) \\ &= A^*u(x,t) + c(x)u(x,t) \end{aligned} \tag{4.4}$$

where

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

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

Let v_i^k represent the value of $u(x,t)$ at the grid point (x_i, t_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. (Schröder [1978, Corollary 1.6b, p221]).

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 uniformly convergent (Kushner [1977]; LaVigna [1986]; Pardoux & Talay [1983]).

A nice property of the above approximation is that it is positive perserving. Regardless of the relationship of Δt and Δx, the solution V^k is always positive. This is important

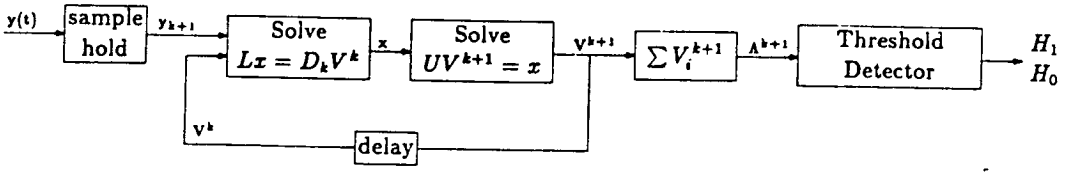

Figure 2. *Overall layout for sequential detector*

since we are approximating a probability density which we know can never be negative.

Schemes similar to (4.9) have been discussed in (Kushner [1977]; Pardoux & Talay [1983]). Furthermore, numerical studies have been performed in (Yavin [1985]) 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

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

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. Our goal is to give a multiprocessor design to efficiently solve (3.1). This means that:

(1) the time necessary to compute V^{k+1}, given V^k, A, and y_k, should be below a problem dependent threshold; and

(2) the control structure should be simple and regular.

We have chosen the *systolic array architecture* of (Kung & Leiserson [1980]) because it can satisfying these goals.

We are interested in systolic processors which perform linear algebra operations. The basic component of these arrays is the *inner product processor* (IPP). At each clock pulse the IPP takes the inputs x, y and a and computes $ax + y$ (the inner product step). This value is output on the y-output line, and the x and a values pass through to their respective output lines untouched.

The number of processors in a systolic array performing matrix-vector multiplication 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 Gaussian elimination without pivoting is used. As is standard with Gaussian elimination, once the factors L and U are found, (5.1) is solved by finding x such that

$$Lx = D_k V^k \tag{5.3}$$

and then by solving

$$U V^{k+1} = x \tag{5.4}$$

to get V^{k+1}.

Let L be the lower triangular matrix resulting from the factorization of a band matrix. We are interested in solving $Lx = b$ for the vector x. This is solved by a simple back substitution algorithm. It is well known that this can be accomplished by systolic arrays (Kung & Leiserson [1980]).

Since $(I - \Delta t A_n)$ is independent of k so are L and U so that this factorization only has to be done at the time of filter design. Therefore, the filter must compute the solution to (5.1) quicker than the problem dependent threshold and does so via (5.3) and (5.4).

It is now possible to give the layout for the sequential detector. Figure 2 shows the overall layout.

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 (Baras & Holley [1987]) where multigrid methods are used and a proposed multi-level architecture results.

References

J. S. Baras [1978], "Ship RCS Scintillation Simulation," Naval Research Laboratory, Secret report, Unclassified title, Technical Report 8189.

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

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

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

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

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

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

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

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

Y. Yavin [1985], "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.