AD-A102 025 MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR INFORMA--ETC F/G 12/1
SEQUENTIAL DECISION RULES FOR FAILURE DETECTION.(U)
JUL 81 E Y CHOW, A S WILLSKY N00014-77-C-0224
UNCLASSIFIED LIDS-P-1109 NL

END
DATE FILMED
8-81
DTIC

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL II

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A102025
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) SEQUENTIAL DECISION RULES FOR FAILURE DETECTION	5. TYPE OF REPORT & PERIOD COVERED Technical Paper
	6. PERFORMING ORG. REPORT NUMBER LIDS-P-1109
7. AUTHOR(s) Edward Y. Chow, Alan S. Willsky	8. CONTRACT OR GRANT NUMBER(s) ONR Contract N0014-77-C-0224 (NR 041-516)
9. PERFORMING ORGANIZATION NAME AND ADDRESS MIT Laboratory for Information and Decision Systems Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 North Quincy St. Arlington, VA 22217	12. REPORT DATE July 1981
	13. NUMBER OF PAGES 11 pages
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE JUL 27 1981 S D E

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The formulation of the decision making of a failure detection process as a Bayes sequential decision problem (BSDP) provides a simple conceptualization of the decision rule design problem. As the optimal Bayes rule is not computable, a methodology that is based on the Baysian approach and aimed at a reduced computational requirement is developed for designing suboptimal rules. A numerical algorithm is constructed to facilitate the design and performance evaluation of these suboptimal rules. The result of applying this design methodology to an example shows that this approach is a useful one.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 55 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A102025

DTIC FILE COPY

July, 1981

LIDS-P-1109

SEQUENTIAL DECISION RULES FOR FAILURE DETECTION*

Edward Y. Chow, Schlumberger-Doll Research
Ridgefield, Connecticut 06877

Alan S. Willsky, Laboratory for Information and Decision Systems,
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract

The formulation of the decision making of a failure detection process as a Bayes sequential decision problem (BSDP) provides a simple conceptualization of the decision rule design problem. As the optimal Bayes rule is not computable, a methodology that is based on the Baysian approach and aimed at a reduced computational requirement is developed for designing suboptimal rules. A numerical algorithm is constructed to facilitate the design and performance evaluation of these suboptimal rules. The result of applying this design methodology to an example shows that this approach is a useful one.

* This work was supported in part by the Office of Naval Research under Contract No. N00014-77-C-0224 and in part by NASA Ames Research Center under Grant NGL-22-009-124.

1. INTRODUCTION

The failure detection and identification (FDI) process involves monitoring the sensor measurements or processed measurements known as the residual [1] for changes from its normal (no-fail) behavior. Residual samples are observed in sequence. If a failure is judged to have occurred and sufficient information (from the residual) has been gathered, the monitoring process is stopped. Then, based on the past observations of residual, an identification of the failure is made. If no failure has occurred, or if the information gathered is insufficient, monitoring is not interrupted so that further residual samples may be observed. The decision to interrupt the residual-monitoring to make a failure identification is based on a compromise between the speed and accuracy of the detection, and the failure identification reflects the design tradeoff among the errors in failure classification. Such a decision mechanism belongs to the extensively studied class of sequential tests or sequential decision rules. In this paper, we will employ the Bayesian Approach [2] to design decision rules for FDI systems.

In Section 2, we will describe the Bayes formulation of the FDI decision problem. Although the optimal rule is generally not computable, the structure of the Bayesian approach can be used to derive practical suboptimal rules. We will discuss the design of suboptimal rules based on the Bayes formulation in Section 3. In Section 4, we will report our experience with this approach to designing decision rules through a numerical example and simulation.

2. THE BAYESIAN APPROACH

The BSDP formulation of the FDI problem consists of six elements:

1) Θ: the set of states of nature or failure hypotheses. An element θ of Θ may denote a single type i failure of size ν occurring at time τ($\theta=(i,\tau,\nu)$) or the occurrence of a set of failures (possibly simultaneously), i.e. $\theta=\{(i_1,\tau_1,\nu_1),\ldots,(i_n,\tau_n,\nu_n)\}$. Due to the infrequent nature of failure, we will focus on the case of a single failure.

In many applications it suffices to just identify the failure type without estimating the failure size. Moreover, it is often true that a detection system based on $(i,\tau,\bar{\nu})$ for some appropriate $\bar{\nu}$ can also detect and identify the type of the failure (i,τ,ν) for $\nu>\bar{\nu}$. Thus, we may use $(i,\tau,\bar{\nu})$ to represent (i,τ). In the aircraft sensor FDI problem [3], for instance, excellent results were obtained using this approach. Now we have the discrete nature set

$$\Theta = \{(i,\tau),\ i=1,\ldots,M,\ \ \tau=1,2,\ldots,\}$$

where we assume there are M different failure types of interest.

2) μ: the prior probability mass function (PMF) over the nature set Θ. This PMF represents the a priori information concerning the failure, i.e. how likely it is for each type of failure to occur, and when is a failure likely to occur. Because this information may not be available or accurate in some cases, the need to specify μ is a drawback of the Bayes approach for such cases. Nevertheless, we will see that it can be regarded as a parameter in the design of a Bayes rule.

In general, μ may be arbitrary. Here, we assume the underlying failure process has two properties: i) the M failures of Θ are independent of one another, and ii) the occurrence of each failure i is a Bernoulli process with (discrete) parameter q_i. The Bernoulli process (corresponding to the Poisson process in continuous time) is a common model for failures in physical components; the independence assumption

describes a large class of failures (such as sensor failures) while providing a simple approximation for the others. It is straightforward to show that

$$u(i,\tau)=\sigma(i)\rho(1-\rho)^{\tau-1} \quad i=1,\ldots,M, \quad \tau=1,2,\ldots,$$

where

$$\rho=1-\prod_{j=1}^{M}(1-\rho_j)$$

$$\sigma(i)=\rho_i(1-\rho_i)^{-1}\Big[\sum_{j=1}^{M}\rho_j(1-\rho_j)^{-1}\Big]^{-1}$$

The parameter ρ may be regarded as the parameter of the combined (Bernoulli) failure process - the occurrence of the first failure; $\sigma(i)$can be interpreted as the marginal probability that the first failure is of type i. Note that the present choice of u indicates the arrival of the first failure is memoryless. This property is useful in obtaining time-invariant suboptimal decision rules.

3) $\mathcal{D}(k)$: the discrete set of terminal actions (failure identifications) available to the decision maker when the residual-monitoring is stopped at time k. An element δ of $\mathcal{D}(k)$may denote the pair (j,t), i.e. declaration of a type j failure to have occurred at time t. Alternatively, δ may represent an identification of the j-th failure type without regard for the failure time, or it may signify the presence of a failure without specification of its type or time, i.e. simply an alarm. Since the purpose of FDI is to detect and identify failures that have occurred $\mathcal{D}(k)$ should only contain identifications that either specify failure times at/before k, or do not specify any failure time. As a result, the number of terminal decisions specifying failures times grows with k while the number of decisions not specifying any time will remain the same. In addition, $\mathcal{D}(k)$ does not include the declaration of no failure, since the residual-monitoring is stopped only when a failure appears to have occurred.

4) $L(k;\theta,\delta)$: the terminal decision cost function at time k. $L(k;\theta,\delta)$ denotes the penalty for deciding $\delta\in\mathcal{D}(k)$ at time k when the true state of nature is $\theta=(i,\tau)$. It is assumed to be bounded and non-negative and have the structure:

$$L(k;(i,\tau),\delta)=\begin{cases}L((i,\tau),\delta) & \tau\le k, \quad \delta\in\mathcal{D}(k)\\ L_F & \tau>k \quad \delta\in\mathcal{D}(k)\end{cases}$$

where $L(\theta,\delta)$ is the underlying cost function that is independent of k; L_F denotes the penalty for a false alarm, and it may be generalized to be dependent on δ. It is only meaningful for a terminal action (identification) that indicates the correct failure (and/or time) to receive a lower decision cost than one that indicates the wrong failure (and/or time). We further assume that the penalty due to an incorrect identification of the failure time is only dependent on the error of such an identification. That is for $\delta=(j,t)$,

$$L((i,\tau),(j,t)) = L(i,j,(t-\tau))$$

and for δ with no time specification

$$L((i,\tau),\delta)= L(i,\delta)$$

5) $r(k)$: the m-dimensional residual (observation) sequence. We shall let $p(r(1),\ldots,r(k)|(i,\tau))$ denote their joint conditional density. Assuming that the residual is affected by the failure in a causal manner, its conditional density has the property

$$p(r(1),\ldots,r(k)|(i,\tau))=p(r(1),\ldots,r(k)|(0,-)) \qquad i=1,\ldots,M, \quad \tau>k$$

where (0,-) is used to denote the no-fail condition. For the design of suboptimal rules, we will assume that the residual is an independent Gaussian sequence with V(mxm matrix) as the time-independent covariance function and $g_i(k-\tau)$ as the mean given that the failure (i,τ) has occurred. With the covariance assumed to be the same for all failures, the mean function $g_i(k-\tau)$, characterizes the effect of the failure (i,τ), and it is henceforth called the <u>signature</u> of (i,τ) (with $g_i(k-\tau)=0$, for i=0, or $\tau>k$). We have chosen to study this type of residuals because its special structure facilitates the development of insights into the design of decision rules. Moreover, the Gaussian assumption is reasonable in many problems and has met with success in a wide variety of applications, e.g., [3] [4]. (It should be noted that the use of more general probability densities for the residual will not add any conceptual difficulty.)

6) $c(k,(i,\tau))$: the delay cost function having the properties:

$$c(k,(i,\tau)) = \begin{cases}c(i,k-\tau) & >0 \quad \tau<k\\ 0 & \tau\ge k\end{cases}$$

$$c(i,k_1-\tau)>c(i,k_2-\tau) \qquad k_1>k_2>\tau$$

After a failure has occurred at τ, there is a penalty for delaying the terminal decision until time $k>\tau$ with the penalty an increasing function of the delay $(k-\tau)$. In the absence of a failure, no penalty is imposed on the sampling. In this study we will consider a delay cost function that is linear in the delay, i.e. $c(i,k-\tau)=c(i)(k-\tau)$, where c(i) is a positive function of the failure type i, and may be used to provide different delay penalties for different types of failures.

A sequential decision rule naturally consists of two parts: a stopping rule (or sampling plan) and a terminal decision rule. The stopping rule, denoted by $\Phi=(\phi(0),\phi(1;r(1)),\ldots,\phi(k;r(1),\ldots,r(k)),\ldots)$ is a sequence of functions of the observed residual samples, with $\phi(k;r(1),\ldots,r(k))=1$, or 0. When $\phi(k;r(1),\ldots,r(k))=1$, (0), residual-monitoring or sampling is stopped (continued) after the k residual samples, r(1),...,r(k) are observed. Alternatively, the stopping rule may be defined by another sequence of functions $\Psi=(\psi(0),\psi(1;r(1)),\ldots,\psi(k;r(1),\ldots,r(k)),\ldots)$, where $\psi(k;r(1),\ldots,r(k))=1$ (0) indicates that residual-monitoring has been carried on up to and including time (k-1) and will (not) be stopped after time k when residual samples, r(1),...,r(k) are observed. The functions ϕ and ψ are related to each other in the following way

$$\psi(k;r(1),\ldots,r(k)) = \phi(k;r(1),\ldots,r(k))\cdot\prod_{s=0}^{k-1}[1-\phi(s,r(1),\ldots,r(s))]$$

with $\psi(0)=\phi(0)$.

The terminal decision rule is a sequence of functions, $D=(d(0),d(1;r(1)),\ldots,d(k;r(1),\ldots,r(k)),\ldots)$, mapping residual samples, r(1),...,r(k) into the terminal action set $\mathcal{D}(k)$. The function $d(k;r(1),\ldots,r(k))$ represents the decision rule used to arrive at an action (identification) if sampling

is stopped at time k and the residual samples, r(1), ...,r(k) are observed.

As a result of using the sequential decision rule (ϕ,D), given (i,τ) is the true state of nature, the total expected cost is:

$$U_0[(i,\tau),(\phi,D)]=\sum_{k=0}^{\infty} E_{i,\tau}\{\psi(k;r(1),\ldots,r(k))[c(k,(i,\tau))+ L(k;(i,\tau),d(k;r(1),\ldots,r(k)))]\}$$

The BSDP is defined as: determine a sequential decision rule (ϕ^*,D^*) so that the sequential Bayes risk U_s is minimized, where

$$U_s(\phi,D)=EU_0[(i,\tau),(\phi,D)]=\sum_{i=1}^{M}\sum_{\tau=1}^{\infty} u(i,\tau)U_0[(i,\tau),(\phi,D)]$$

(ϕ^*,D^*) is called the Bayes Sequential Decision Rule (BSDR) with respect to u, and it is optimal in the sense that it minimizes the sequential Bayes risk.

In the following we will discuss an interpretation of the sequential risk for the FDI problem. Let us define the following notation

$$P_F(\tau)=\sum_{k=1}^{\tau-1} E_{0,-}\psi(k;r(1),\ldots,r(k))$$

$$\mathcal{D}=\bigcup_{k=0}^{\infty}\mathcal{D}(k)$$

$$\bar{S}(k,\delta)=\{[r(1),\ldots,r(k)]: \psi(k;r(1),\ldots,r(k))=1, d(k,r(1),\ldots,r(k))=\delta\},\ \delta\in\mathcal{D}$$

$$P_r\{\bar{S}(k,\delta)|i,\tau\}=\int_{\bar{S}(k,\delta)} p(r(1),\ldots,r(k)|i,\tau)dr(1)\ldots dr(k)$$

$$\bar{\tau}(i,\tau)=\sum_{k=\tau}^{\infty}(k-\tau)(1-P_F(\tau))^{-1}E_{i,\tau}\psi(k;r(1),\ldots,r(k))$$

$$P((i,\tau),\delta)=\sum_{k=\tau}^{\infty} P_r\{\bar{S}(k,\delta)|i,\tau\}(1-P_F)^{-1}$$

where $P_F(\tau)$ is the probability of stopping to declare a failure before the failure occurs at τ, i.e, the probability of false alarm when a failure occurs at time τ; $\mathcal{D}$ is the set of terminal actions for all times; $\bar{S}(k,\delta)$ is the region in the sample space of the first k residuals where the sequential rule (ϕ,D) yields the terminal decision δ. Clearly, the $\bar{S}(k,\delta)$'s are disjoint sets with respect to both k and δ. The expressions $\bar{\tau}(i,\tau)$ and $P((i,\tau),\delta)$ are the conditional expected delay in decision (i.e. stopping sampling and making a failure identification) and the conditional probability of eventually declaring δ, given a type i failure has occurred at time τ and no false alarm has been signalled before this time respectively. $P((i,\tau),\delta)$ is the generalized cross-detection probability. Finally, the sequential Bayes risk U_s can be written as

$$U_s(\phi,D)=\sum_{i=1}^{M}\sum_{\tau=1}^{\infty} u(i,\tau)\{L_F P_F(\tau)+(1-P_F(\tau))[c(i)\bar{\tau}(i,\tau)+\sum_{\delta\in\mathcal{D}} L((i,\tau),\delta)P((i,\tau),\delta)]\} \qquad (1)$$

Equation (1) indicates that the sequential Bayes risk is a weighted combination of the conditional false alarm probability, expected delay to decision and cross-detection probabilities, and the optimal sequential rule (ϕ^*,D^*) minimizes such a combination. From this vantage point, the cost functions (L and c) and the prior distribution (u) provide for the weighting, hence, a basis for indirectly specifying the tradeoff relationships among the various performance issues. The advantage of the indirect approach is that only the total expected cost instead of every individual performance issue needs to be considered explicitly in designing a sequential rule. The drawback of the approach, however, lies in the choice of a set of appropriate cost functions (and sometimes the prior distribution) when the physical problem does not have a natural set, as it doesn't in general. In this case, the Bayes approach is most useful with the cost functions (and the prior distribution) considered as design parameters that may be adjusted to obtain an acceptable design.

The optimal terminal decision rule D* can be easily shown to be a sequence of fixed-sample-size tests [2]. The determination of the optimal stopping rule ϕ^* is a dynamic programming problem [1]. The immense storage and computation required make ϕ^* impossible to compute, and suboptimal rules must be used.

Despite the impractical nature of its solution, the BSDP provides a useful framework for designing suboptimal decision rules for the FDI problem because of its inherent characteristic of explicitly weighing the tradeoffs between detection speed and accuracy (in terms of the cost structure). A sequential decision rule defines a set of sequential decision regions $\bar{S}(k,\delta)$, and the decision regions corresponding to the BSDR yield the minimum risk. From this vantage point, the design of a suboptimal rule can be viewed as the problem of choosing a set of decision regions that would yield a reasonably small risk. This is the essence of the approach to suboptimal rule design that we will describe next.

3. SUBOPTIMAL RULES

The Sliding Window Approximation

The immense computation associated with the BSDR is partly due to the increasing number of failure hypotheses as time progresses. The remedy for this problem is the use of a sliding window to limit the number of failure hypotheses to be considered at each time. The assumption made under the sliding window approximation is that essentially all failures can be detected within W time steps after they have occurred, or that if a failure is not detected within this time it will not be detected in the future. Here, the window size W is a design parameter, and it should be chosen long enough so that detection and identification of failures are possible, but short enough so that implementation is feasible [1].

The sliding window rule (ϕ^W,d^W) divides the sample space of the sliding window of residuals {r(k-W+1), ...,r(k)}, or equivalently, the space of vectors of posterior probabilities, likelihood ratios, or log likelihood ratios ($\mathcal{L}$) of the sliding window of failure hypotheses into disjoint time-independent sequential decision regions $\{S_0,S_1,\ldots,S_N\}$. Because the residuals are assumed to be Gaussian variables, it is simpler to work with $\mathcal{L}$ (which is related to $\bar{L}$ by a constant):

$$\mathcal{L}(k)=[\mathcal{L}_0'(k),\ldots,\mathcal{L}_{N-1}'(k)]'$$

where

$$\mathcal{L}_\tau(k)=[\mathcal{L}(k;1,\bar{\tau}),\ldots,\mathcal{L}(k;M,\bar{\tau})]'$$

$$\mathcal{L}(k;i,\tau)=\sum_{s=0}^{\bar{\tau}} g_i'(s)W^{-1}r(k-\tau-s) \qquad (2)$$

Then, the sliding window rule states: At each time $k\geq W$, we form the decision statistics $\mathcal{L}(k)$ from the window of residual samples. If $\mathcal{L}(k)\in S_i$, for i=1,..., or N, we will stop sampling to declare H_i; otherwise,

$L(k)\varepsilon S_0$, and we will proceed to take one more observation of the residual. The Bayes design problem is to determine a set of regions $\{S_0^*, S_1^*, \ldots, S_N^*\}$ that minimizes the sequential risk $U_s^W(\{S_i\})$. This represents a functional minimization problem for which a solution is generally very difficult to determine. A simpler alternative to this problem is to constrain the decision regions to take on special shapes, $\{S_i(f)\}$, that are parameterized by a fixed dimensional vector, f, of design variables. Then the resulting design problem involves the determination of a set of parameter values f^* that minimizes the risk $U_s^W(f)$. We will focus our attention on a special set of parametrized sequential decision regions, because they are simple and they serve well to illustrate that the Bayes formulation can be exploited, in a systematic fashion, to obtain simple suboptimal rules that are capable of delivering good performance. These decision regions are:

$$S(j,\bar{\tau}) = \{L(k) : L(k;j,\bar{\tau}) > f(j,\bar{\tau}),\ \varepsilon^{-1}(j,\bar{\tau})[L(k;j,\bar{\tau}) - f(j,\bar{\tau})] > \varepsilon^{-1}(i,\bar{\tau})[L(k;i,\bar{\tau}) - f(i,\bar{\tau})],\ (i,\bar{\tau}) \neq (j,\bar{\tau})\} \quad (3a)$$

$$S(0,-) = \{L(k) : L(k;i,\bar{\tau}) \leq f(i,\bar{\tau}),\ i=1,\ldots,M,\ \bar{\tau}=0,\ldots,W-1\} \quad (3b)$$

where $S(j,\bar{\tau})$ is the stop-to-declare $(j,k-\bar{\tau})$ region and $S(0,-)$ is the continue region (see Fig. 1). Generally the ε's may be regarded as design parameters, but here, $\varepsilon(j,\bar{\tau})$ is simply taken to be the standard deviation of $L(k,j,\bar{\tau})$.

To evaluate $U_s^W(f)$, we need to determine the set of probabilities, $\{\Pr(L(k)\varepsilon S(j,\bar{\tau}), L(k-1)\varepsilon S(0,-), \ldots, L(W)\varepsilon S(0,-) | i,\tau),\ k>W,\ j=0,1,\ldots,M,\ \bar{\tau}=0,\ldots,W-1\}$, which, indeed, is the goal of many research efforts in the so-called level-crossing problem [5]. Unfortunately, useful results (bounds and approximations of such probabilities) are only available for the scalar case [6],[7],[8]. As it stands, each of the probabilities is an integral of a kMW-dimensional Gaussian density over the compound region $S(0,-)x\ldots xS(0,-)xS(j,\bar{\tau})$, which, for large kMW, becomes extremely unwieldy and difficult to evaluate.

The MW-dimensional vector of decision statistics $L(k)$ corresponds to the MW failure hypotheses, and they provide the information necessary for the simultaneous identification of both failure type and failure time. In most applications, such as the aircraft sensor FDI problem [3] and the detection of freeway traffic incidents [4], where the failure time need not be explicitly identified, the f lure time resolution power provided by the full window of decision statistics is not needed. Instead, decision rules that employ a few components of $L(k)$ may be used. The decision rule of this type considered here consists of sequential decision regions that are similar to (3) but are only defined in terms of M components of $L(k)$:

$$S_j = \{L_{W-1}(k) : L(k;j,W-1) > f_j,\ \varepsilon^{-1}(j,W-1)[L(k;j,W-1) - f_j] > \varepsilon^{-1}(i,W-1)[L(k,i,W-1) - f_j],\ \forall i \neq j\} \quad (4a)$$

$$S_0 = \{L_{W-1}(k) : L(k,j,W-1) \leq f_j \quad j=1,\ldots,M\} \quad (4b)$$

where S_j is the stop-to-declare-failure-j region and S_0 is the continue region. It should be noted that the use of (4) is effective if cross-correlations of signatures among hypotheses of the same failure type at different times are smaller than those among hypotheses of different failure types.

The risk for using (4) is

$$U_s^W(f) = L_F \sum_{i=1}^{M} \sum_{\tau=W+1}^{\infty} \nu(i,\tau) \sum_{k=W}^{\tau-1} \sum_{j=1}^{M} P_r\{L_{W-1}(k)\varepsilon S_j, S_0(k-1) | 0,-\} + \sum_{i=1}^{M} \sum_{\tau=1}^{\infty} \nu(i,\tau) \sum_{k=\max[W,\tau]}^{\infty} \sum_{j=1}^{M} [c(i)(k-\tau) + L(i,j)]$$
$$\times \Pr\{L_{W-1}(k)\varepsilon S_j,\ S_0(k-1) | i,\tau\}$$

where

$$S_0(k) = \{L_{W-1}(k)\varepsilon S_0, \ldots, L_{W-1}(W)\varepsilon S_0\}$$

The probabilities required for calculating the risk are given by the recursion:

$$p(L_{W-1}(k+1) | S_0(k), i, \tau) = \left[\int_{S_0} p(L_{W-1}(k) | S_0(k-1), i, \tau) dL_{W-1}(k)\right]^{-1} \times \int_{S_0} p(L_{W-1}(k+1) | L_{W-1}(k), S_0(k-1), i, \tau) \cdot p(L_{W-1}(k) | S_0(k-1), i, \tau) dL_{W-1}(k) \quad k>W \quad (5)$$

$$\Pr\{L_{W-1}(k)\varepsilon S_j,\ S_0(k-1) | i,\tau\} = \Pr\{S_0(k-1) | i,\tau\} \cdot \int_{S_j} p(L_{W-1}(k) | S_0(k-1), i, \tau) dL_{W-1}(k),\ j=0,1,\ldots,M \quad (6)$$

with

$$\Pr\{L_{W-1}(W)\varepsilon S_j | i,\tau\} = \int_{S_j} p(L_{W-1}(W) | i,\tau) dL_{W-1}(W) \quad (7)$$

For M small, numerical integration of (5)-(7) becomes manageable.

Unfortunately, the transition density, $p(L_{W-1}(k+1) | L_{W-1}(k), S_0(k-1), i, \tau)$, required in (5) is difficult to calculate, because $L_{W-1}(k)$ is not a Markov process. In order to facilitate computation of the probabilities, we need to approximate the transition density. In approximating the required transition density for $L_{W-1}(k)$ we are, in fact, approximating the behavior of L_{W-1}. A simple approximation is a Gauss-Markov process $\ell(k)$ that is defined by

$$\ell(k+1) = A\ell(k) + \xi(k+1)$$

$$E\{\xi(k)\xi'(t)\} = BB' u_0(k-t)$$

where A and B are MxM constant matrices and ξ is a white Gaussian sequence with covariance equal to the (MxM) matrix BB'. The reason for choosing this model is twofold. Firstly, just as $L_{W-1}(k)$, $\ell(k)$ is Gaussian. Secondly, $\ell(k)$ is Markov so that its transition density can be readily determined. In order to have $\ell(k)$ behave like $L_{W-1}(k)$, we set the matrices A and B and the mean of ℓ such that

$$E_{i,\tau}\{\ell(k)\} = E_{i,\tau}\{L_{W-1}(k)\} \quad (8)$$

$$E_{0,-}\{\ell(k)\ell'(k)\} = E_{0,-}\{L_{W-1}(k) L'_{W-1}(k)\} \quad (9)$$

$$E_{0,-}\{\ell(k)\ell'(k+1)\} = E_{0,-}\{L_{W-1}(k) L_{W-1}(k+1)\} \quad (10)$$

That is, we have matched the marginal density and the one-step cross-covariance of $\ell(k)$ to those of $L_{W-1}(k)$. It can be shown that (8)-(10) uniquely specify

$$A = \Sigma_1' \Sigma_0^{-1}$$

$$BB' = \Sigma_0 - \Sigma_1' \Sigma_0^{-1} \Sigma_1$$

$$E_{i,\tau}(\xi(k+1)) = E_{i,\tau}\{L_{W-1}(k+1)\} - A\,E\{L_{W-1}(k)\}$$

where

$$\Sigma_0 = E\{L_{W-1}(k)L'_{W-1}(k)\} = \sum_{t=0}^{W-1} G_t V^{-1} G'_t$$

$$\Sigma_1 = E\{L_{W-1}(k)L'_{W-1}(k+1)\} = \sum_{t=0}^{W-2} G_{t-1} V^{-1} G'_t$$

$$E_{i,\tau}\{L_{W-1}(k)\} = \begin{cases} 0 & \tau > k \\ \sum_{t=0}^{k-\tau} G_{t-k_0} V^{-1} G'_t & k_0 = k-W+1-\tau \le 0 \\ \sum_{t=0}^{W-1} G_t V^{-1} G'_{t+k_0} & k_0 = k-W+1-\tau > 0 \end{cases}$$

$$G_t = [g_1(t),\ldots,g_M(t)]'$$

Moreover, the matrix A is stable, i.e. the magnitudes of all of the eigenvalues of A are less than unity, and B is invertible if G_0 or G_{W-1} is of rank M. Because ξ is an artificial process (i.e. ξ is not a direct function of the residuals r(k)) $\ell(k)$ can never be implemented for use in (4).

We may choose other Markov approximations of $L_{W-1}(k)$ that match the n-step cross-covariance ($1<n<W$) instead of matching the one-step cross-covariance as in (10). The suitability of a criterion for choosing the matrices A and B, such as (9) and (10), depends directly on the failure signatures under consideration and may be examined as an issue separate from the decision rule design problem. Also, a higher order Markov process may be used to approximate L_{W-1}. However, the increase in the computational complexity may negate the benefits of the approximation.

Now we can approximate the required probabilities in the risk calculation as

$$P_r\{L_{W-1}(k)\in S_j, S_0(k-1)|i,\tau\} \approx P_r\{\ell(k)\in S_j, S_0(k-1)|i,\tau\}$$

$$j=0,1,\ldots,M \quad k\ge W$$

and

$$P_r\{\ell(k)\in S_j,\ S_0(k-1)|i,\tau\}$$

$$= P_r\{S_0(k-1)|i,\tau\} \int_{S_j} p(\ell(k)|S_0(k-1),i,\tau)\,d\ell(k) \qquad (11)$$

where we have applied the same decision rule to $\ell(k)$ as $L_{W-1}(k)$. Therefore, S_j and $S_0(k-1)$ denote the decision regions and the event of continued sampling up to time k for both L_{W-1} and ℓ. Assuming B^{-1} exists, we have

$$p(\ell(k+1)|S_0(k),i,\tau) = \left[\int_{S_0} p(\ell(k)|S_0(k-1),i,\tau)\,d\ell(k)\right]^{-1}$$
$$\times \int_{S_0} p(\xi(k+1) = [\ell(k+1)-A\ell(k)]|i,\tau)$$
$$p(\ell(k)|S_0(k-1),i,\tau)\,d\ell(k), \quad k>W \qquad (12)$$

where $p(\xi(k)|i,\tau)$ is the Gaussian density of $\xi(k)$ under the failure (i,τ). Now the integrals (11) and (12) represent more tractable numerical problems.

In the event that B is not invertible, the transition density is degenerate and (12) is very difficult to evaluate. Very often this problem can be circumvented by batch processing the residuals. That is, we may consider the modified residual sequence: $\bar{r}(k) = [r'(vk-v+1), r'(vk-v+2),\ldots,r'(vk)]'$ for some batch size $v>0$ with $k=1,2,\ldots$ as the new time index. In using $\bar{r}(k)$ we have to augment the signatures as: $[g_i'(0),\ldots,g_i'(v-1)]'$, $i=1,\ldots,M$. By a proper choice of v, the rank of G_0 can be increased to M and B will be invertible.

Non-Window Sequential Decision Rules

Here we will describe another simple decision rule that has the same decision regions as the simplified sliding window rule (4), but the vector (z) of M decision statistics is obtained differently as follows:

$$z(k+1) = \bar{A}\,z(k) + \bar{B}\,r(k+1) \qquad (13)$$

where $\bar{A}$ is a constant stable MxM matrix, and $\bar{B}$ is a Mxm constant matrix of rank M. Unlike the Markov model $\ell(k)$ that approximates $L_{W-1}(k)$, z(k) is a realizable Markov process driven by the residual. The advantages of using z as the decision statistic are: 1) less storage is required, because residual samples need not be stored as necessary in the sliding window scheme, and 2) since z is Markov, the required probability integrals are of the form (11) and (12) so that the same integration algorithm can be directly applied to evaluate such integrals. (It is possible to use a higher order z, but the added complexity will negate the advantages.)

In order to form the statistics z, we need to choose the matrices $\bar{A}$ and $\bar{B}$. When the failure signatures under consideration are constant biases, $\bar{B}$ can simply be set to equal G_0, and $\bar{A}$ can be chosen to be αI, where $0<\alpha<1$. Then, the term $\bar{B}r$ in (13) resembles $g'V^{-1}r$ of (2), and it provides the correlation of the residual with the signatures. The time constant $(1/1-\alpha)$ of z characterizes the memory span of z just as W characterizes that of the sliding window rules.

More generally, if we consider failure signatures that are not constant biases, the choice of $\bar{A}$ may still be handled in the same way as in the constant-bias case, but the selection of a $\bar{B}$ matrix is more involved. With some insights into the nature of the signatures, a reasonable choice of $\bar{B}$ can often be made. To illustrate how this may be accomplished, we will consider an example with two failure modes and an m-dimensional residual vector. Let

$$g_1(k-\tau) = \beta_1$$

$$g_2(k-\tau) = \beta_2(k-\tau+1)$$

That is, g_1 is a constant bias, and g_2 is a ramp. If β_1 and β_2 are not multiples of each other a simple choice of $\bar{B}$ is available:

$$\bar{B} = \begin{bmatrix} \beta_1' \\ \beta_2' \end{bmatrix}$$

If $\beta_1=\alpha_1\beta$ and $\beta_2=\alpha_2\beta$, where α_1 and α_2 are scalar constants, the above choice of $\bar{B}$ has rank one and is not useful for identifying either signature. Suppose we batch process every two residual samples together, i.e. we use the residual sequences $\bar{r}(k)=[r'(2k-1),r'(2k)]'$, $k=1,2,\ldots$. Then we can set $\bar{B}$ to be

$$\bar{B} = \begin{bmatrix} \beta' & \beta' \\ \beta' & 2\beta' \end{bmatrix}$$

Thus, the first and second rows of $\bar{B}$ capture the constant-bias and ramp nature β_1 and β_2, respectively

(and this $\tilde{B}$ has rank two). The use of the modified resudual $\tilde{r}(k)$ in this case causes no adverse effect, since it only lengthens slightly the interval between times when terminal decisions may be made. A big increase in such intervals i.e., the batch processing of $r(k),\ldots,r(k+v)$ simultaneously for large v, may however, be undesirable. For problems where the signatures vary drastically as a function of the elapsed time, or the distinguishability among failures depends essentially on these variations, the effectiveness of using z diminishes. In such cases the sliding window decision rule should provide better performance because of its inherent nature to look for a full window's worth of signature.

Probability Calculation

An algorithm based on 1-dimensional Gaussian quadrature formulas [9] has been developed to compute the probability integrals of (11) and (12) for the case M=2. (It can be extended to higher dimension with an increase in computation.) The details of this quadrature algorithm is described in [1]. Its accuracy has been assessed via comparison with Monte Carlo simulations (see the numerical example). With this algorithm we can evaluate the performance probabilities and risks associated with the suboptimal decision rules described above.

Risk Calculation

In the absence of a failure, the conditional density has been observed to essentially reach a steady state at some finite time T>W.[1] Then, for $k \geq T$ we have

$$\Pr\{\ell(k)\varepsilon S_j \mid S_0(k-1),0-\} = \bar{b}_j \tag{14}$$

$$\Pr\{\ell(k)\varepsilon S_j,\ell(k-1)\varepsilon S_0,\ldots,\ell(\tau)\varepsilon S_0 \mid S(\tau-1),i,\tau\} = b_j(k-\tau \mid i) \qquad k \geq \tau \geq T \tag{15}$$

That is, once steady state is reached, only the relative time (elapsed time) is important. Generally, failures occur infrequently, and decision rule with low false alarm probabilities are employed. Thus, it is reasonalbe to assume 1) $\rho \ll 1$ ($(1-\rho)^T \approx 1$), and 2) $\Pr\{S_0(T) \mid 0,-\} \approx 1$. The sequential risk associated with (4) for M=2 can be approximated by

$$U_s^W(f) = P_F L_F + (1-P_F) \sum_{i=1}^{2} \sigma(i) \sum_{j=1}^{2} \sum_{t=0}^{\infty} [c(i)t + L(i,j)] b_j(t \mid i) \tag{16}$$

where

$$P_F = \frac{(1-\rho)(1-\bar{b}_0)}{1-\bar{b}(1-\rho)}$$

Next, we seek to replace the infinite sum over t in (16) by the finite sum up to $t=\Delta$ plus a term approximating the remainder of the infinite sum. Suppose we have been sampling for Δ steps since the failure occurred. Define:

$$P_t(j \mid i) = \Pr\{\ell(t)\varepsilon S_j \mid S_0(t-1),i,0\} \qquad j=0,1,2$$

If we stop computing the probabilities after Δ, we may approximate

$$P_t(j \mid i) \approx P_\Delta(j \mid i) \qquad j=0,1,2, \quad t>\Delta \tag{17}$$

When the signature of the failure model is a constant (including the no-fail case), the reasoning behind (14) holds, and we can see that $P_t(j \mid i)$ will reach a steady state value as t (the elaspsed time) increases. Then, (17) is a valid approximation for a large Δ. For the case where failure signatures are not constants, the probability of continuing after Δ time steps (for sufficiently large Δ) may be arbitrarily small. The error introduced by (17) in the risk (and performance probability) calculation is, consequently, small.

Substituting (17) in (16), we get

$$U_s^W(f) = P_F L_F + (1-P_F) \sum_{i=1}^{2} \sigma(i) [c(i)\bar{t}_i + \sum_{j=1}^{2} L(i,j) P(i,j)] \tag{18}$$

where

$$\bar{t}_i = \sum_{j=1}^{2} \sum_{t=0}^{\Delta} t\, b_j(t \mid i) + b_0(\Delta \mid i)\,\Delta + \frac{1}{1-P_\Delta(0 \mid i)} \tag{19}$$

$$P(i,j) = \sum_{t=0}^{\Delta} b_j(t \mid i) + b_0(\Delta \mid i)\,\frac{P_\Delta(j \mid i)}{1-P_\Delta(0 \mid i)} \tag{20}$$

P_F is the unconditional false alarm probability, i.e. the probability of one false alarm over all time, $\bar{t}_i$ is the conditional expected delay to decision, given that a type i failure has occurred, and P(i,j) is the conditional probability of declaring a type j failure, given that failure i has occurred. From the assumption that $\Pr\{S_0(T) \mid 0,-\} \approx 1$ and the steady condition (14), it can be shown that the mean time between false alarms is simply $(1-\bar{b}_0)^{-1}$. Now all the probabilities in (18)-(20) can be computed by using the quadrature algorithm. Note that the risk expression (18) consists only of finite sums and it can be evaluated with a reasonable amount of computational effort. With such an approximation of the sequential risk, we will be able to consider the problem of determining the decision regions (the thresholds f) that minimize the risk.

It should be noted that we could consider choosing a set of thresholds that minimize a weighted combination of certain detection probabilities (P(i,j)), the expected detection delay ($\bar{t}_i$), and the mean time between false alarms ($(1-\bar{b}_0)^{-1}$). Although such an objective function will not result in a Bayesian design in general, it is a valid design criterion that may be useful for some application.

Risk Minimization

The risk minimization problem has two features that deserve special attention. Firstly, the sequentail risk is not a simple function of the threshold f, and the derivative with respect to f is not readily available. Secondly, calculating the risk is a costly task. Therefore, the minimum-seeking procedure to be used must require few function (risk) evaluations, and it must not require derivatives. The sequence-of-quadratic-programs (SQP) algorithm studied by Winfield [10] has been chosen to solve this problem, because it does not need any derivative information and it appears to require fewer function evaluations than other well-known algorithms [10]. Furthermore, the SOP is simple, and it has quadratic convergence. Very briefly, the algorithm consists of the following. At each iteration, a quadratic surface is fitted to the risk function locally, then the quadratic model is minimized over a constraint region (hence the name SQP). The risk function is evaluated at this minimum and is used in the surface fitting of the next iteration. The details of the application of SQP to risk minimization

[1] Unfortunately, we have not been able to prove such convergence behavior using elementary techniques. More advanced function-theoretic methods may be necessary.

is reported in [1].

4. NUMERICAL EXAMPLE

Here, we will discuss an application of the suboptimal rule design methodology described above to a numerical example. We will consider the detection and identification of two possible failure modes (without identifying the failure times). We assume that the residual is a 2-dimensional vector, and the vector failure signatures, $g_1(t)$, i=1,2, as functions of the elapsed time t are shown in Table 1. The signature of the first failure mode is simply a constant vector. The first component of $g_2(t)$ is a constant, while the second component is a ramp. We have chosen to examine these two types of signature behavior (constant bias and ramp) because they are simple and describe a large variety of failure signatures that are commonly seen in practice. For simplicity, we have chosen V, the covariance of r, to be the identity matrix.

We will design both a simplified sliding window rule (that uses L_{W-1}) and a rule using the Markov statistic z. The parameters associated with the L_{W-1}, ℓ, and z are shown in Table 2, and the cost functions and the prior probabilities are shown in Table 3. To facilitate discussions, we will introduce the following terminology. We will refer to a Monte Carlo simulation of the sliding window rule by SW, a simulation of the rule using the Markov statistic z as Markov implementation (MI), and a simulation of the nonimplementable decision process using the approximation ℓ as Markov approximation (MA). (All simulations are based on 10,000 trajectones.) The notation Q20 refers to the results of applying the quadrature algorithm to the approximation of L_{W-1} by ℓ.

$$g_1(t)=\begin{bmatrix}1\\.5\end{bmatrix} \qquad g_2(t)=\begin{bmatrix}.5\\.25+.25t\end{bmatrix}$$

$$V=\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

Table 1. Failure signatures.

$$W=8$$

$$A=\begin{bmatrix}.826&.058\\.116&.837\end{bmatrix}$$

$$\Sigma_0=\begin{bmatrix}10&8.5\\8.5&14.75\end{bmatrix}$$

$$BB'=\begin{bmatrix}2.32&2.01\\2.01&4.58\end{bmatrix}$$

$$\tilde{A}=\begin{bmatrix}.875&0\\0&.875\end{bmatrix} \qquad \tilde{B}=\begin{bmatrix}1&.5\\.5&2\end{bmatrix}$$

$$\Sigma_z=\begin{bmatrix}5.33&6.40\\6.40&18.13\end{bmatrix} \qquad \tilde{B}V\tilde{B}'=\begin{bmatrix}1.25&1.50\\1.50&4.25\end{bmatrix}$$

Table 2. Parameters for L_{W-1}, ℓ and z.

$$L_F=9$$

$$L(1,2)=L(2,1)=10 \qquad L(1,1)=L(2,2)=0$$

$$c_1=c_2=1$$

$$u(i,\tau)=.5o(1-o)^{\tau-1}, \quad i=1,2$$

$$\rho=.0002 \qquad T=8 \qquad \Delta=8$$

Table 3. Cost Functions and Prior Probability.

The results of SW, MA, and Q20 for the thresholds [8.85, 12.05] are shown in Figs. 2-6 (see (15) for the definition of notations). The quadrature results Q20 are very close to MA, indicating good accuracy of the quadrature algorithm. In comparing SW with MA, it is evident that the Markov approximation (MA) slightly under-estimates the false alarm rate of the sliding window rule (SW). However, the response of the Markov approximation to failures is very close to that of the sliding window rule. In the present example, L_{W-1} is a 7-th order process, while its approximation ℓ is only of first order. In view of this fact, we can conclude that ℓ provides a very reasonable and useful approximation of L_{W-1}.

The successive choices of thresholds by SQP for the sliding window rule are plotted in Fig. 7. Note that we have not carried the SQP algorithm far enough so that the successive choices of thresholds are, say, within .001 of each other. This is because towards later iterations the performance indices become relatively insensitive to small changes of the f's. This together with the fact that we are only computing an approximate Bayes risk means that fine scale optimization is not worthwhile. Therefore, with the approximate risk, the SQP is most efficiently used to locate the zone where the minimum lies. That is, the SQP algorithm is to be terminated when it is evident that it has converged into a reasonably small region. Then we may choose the thresholds that give the smallest risk as the approximate solution of the minimization.

In the event that thresholds that yield the smallest risk do not provide the desired detection performance, the design parameters, L, c, u, and W may be adjusted and the SQP may be repeated to get a new design. A practical alternative method is to make use of the list of performance indices (e.g. P(i,j)) that are generated in the risk calculation, and choose a pair of thresholds that yields the desired performance.

The performance of the decision rules using L_{W-1} and z as determined by SQP are shown in Figs. 8-12. (The thresholds for L_{W-1} are [8.35, 12.05] and those for z are [6.29, 11.69].) We note that MI has a higher false alarm rate than SW. The speed of detection for the two rules is similar. While MI has a slightly higher type-1 correct detection probability than SW, SW has a consistently higher $b_2(t|2)$ (type-2 correct detection probability) than MI. By raising the thresholds of the rule using z appropriately, we can decrease the false alarm rate of MI down to that of SW with an increase in detection delay and slightly improved correct detection probability for the type-2 failure (with ramp signature). Thus, the sliding window rule is slightly superior to the rule using z in the sense that when both are designed to yield a comparable false alarm rate, the latter will have longer detection delays and slightly lower correct detection probability (for type-2 failure). In view of the fact that a decision rule using z is much simpler to implement, it is worthy of being considered as an alternative to the sliding window rule.

In summary, the result of applying our decision rule design method to the present example is very good. The quadrature algorithm has been shown to be useful, and the Markov approximation of L_{W-1} by z is a valid one. The SQP algorithm has demonstrated its simplicity and usefulness through the numerical example. Finally, the Markov decision statistic z has been shown to be a worthy alternative to the sliding window statistic L_{W-1}.

5. CONCLUSION

A methodology based on the Bayesian approach is developed for designing suboptimal sequential decision rules. This methodology is applied to a numerical example, and the results indicate that it is a useful design approach.

REFERENCES

[1] E. Y. Chow, A Failure Detection Design Methodology, Sc.D. Thesis, Dept. of Elec. Eng. and Comp. Sci., M.I.T., Cambridge, Mass., Oct. (1980).

[2] D. Blackwell and M.A. Girshick, Theory of Games and Statistical Decisions, Wiley, New York (1951).

[3] J.C. Deckert, M.N. Desai, J.J. Deyst and A.S. Willsky, "F-8 DFBW Sensor Failure Identification Using Analytic Redundancy," IEEE Trans. Auto. Control, Vol. AC-22, No. 5, pp. 795-803, Oct. (1977).

[4] A.S. Willsky, E.Y. Chow, S.B. Gershwin, C.S. Greene, P.K. Houpt, and A.L. Kurkjian, "Dynamic Model-Based Techniques for the Detection of Incidents on Freeways," IEEE Trans. Auto. Control, Vol. AC-25, No. 3, pp. 347-360, Jun. (1980).

[5] I.F. Blake and W.C. Lindsey, "Level-Crossing Problems for Random Processes," IEEE Trans. Information Theory, Vol. IT-19, No. 3, pp. 295-315, May (1973).

[6] R.G. Gallager and C.W. Helstrom, "A Bound on the Probability that a Gaussian Process Exceeds a Given Function," IEEE Trans. Information Theory, Vol. IT-15, No. 1, pp. 163-166, Jan. (1969).

[7] D.H. Bhati, "Approximations to the Distribution of Sample Size for Sequential Tests, I. Tests of Simple Hypotheses," Biometrika, Vol. 46, pp. 130-138, (1973).

[8] B.K. Chosh, "Moments of the Distribution of Sample Size in a SPRT," American Statistical Association Journal, Vol. 64, pp. 1560-1574, (1969).

[9] P.J. Davis and P. Rabinowitz, Numerical Integration, Blaisdell Publishing Company, Waltham, Mass., (1967).

[10] D.H. Winfield, Function Minimization Without Derivatives by a Sequence of Quadratic Programming Problems, Harvard Univ., Division of Engineering and Applied Physics, Technical Report No. 537, Cambridge, Mass., Aug. (1967).

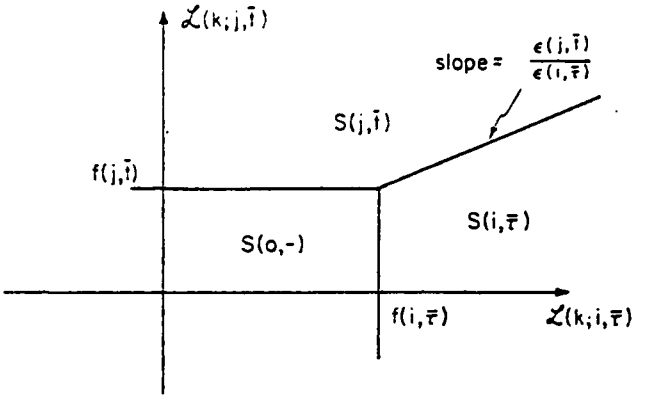

Fig.1 Sequential Decision Regions in 2 Dimensions

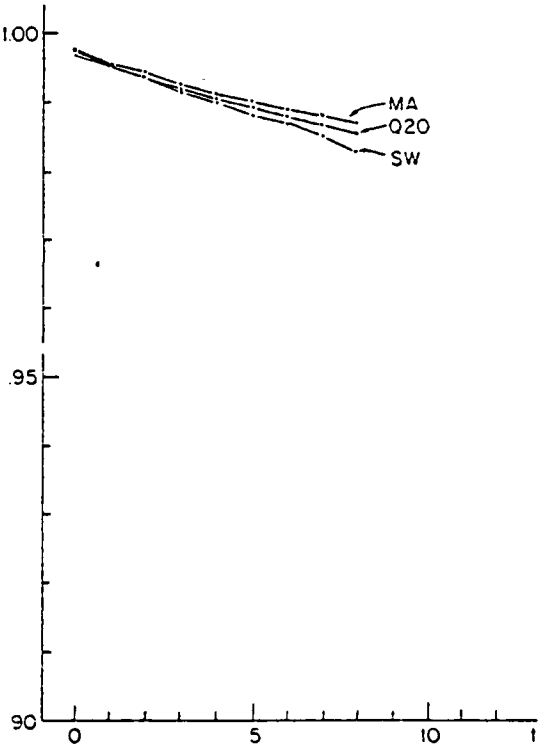

Fig.2 $b_0(t/0)$ - SW, MA, and Q20

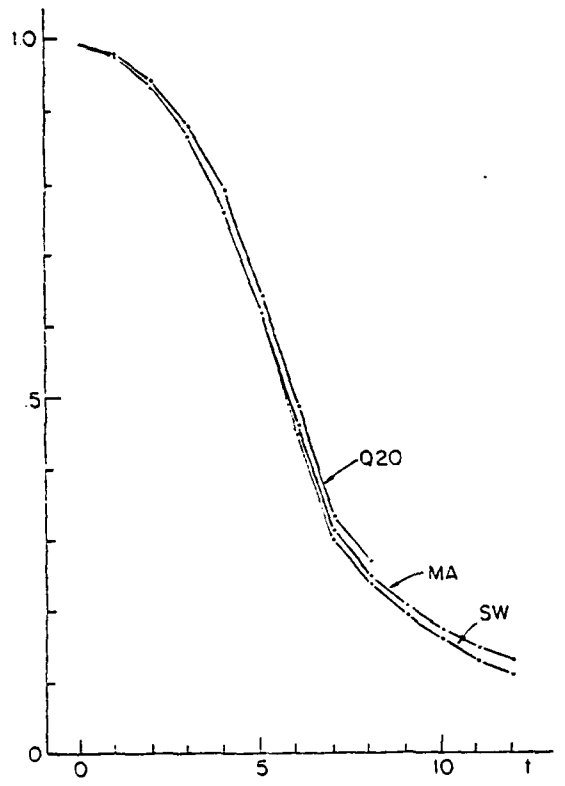

Fig.3 $b_0(t/1)$ - SW, MA, and Q20

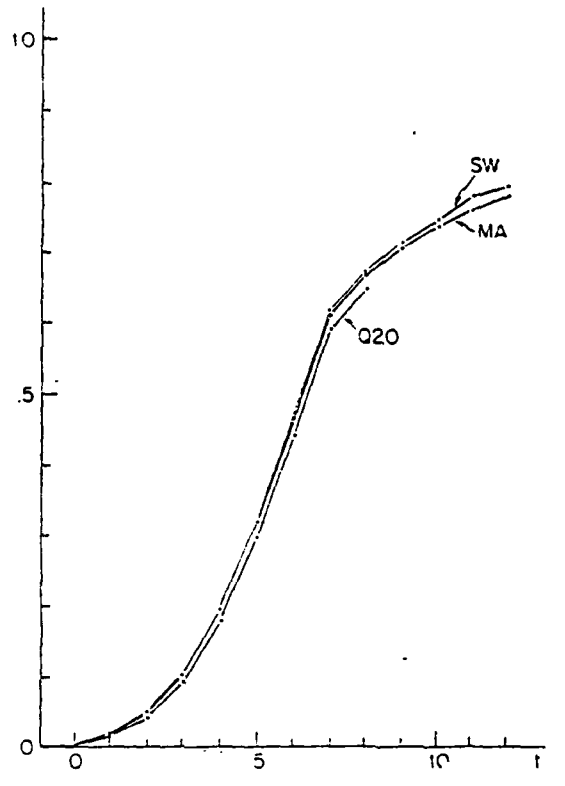

Fig.5 $\frac{t}{s=0}\ b_1(s/1)$ - SW, MA, and Q20

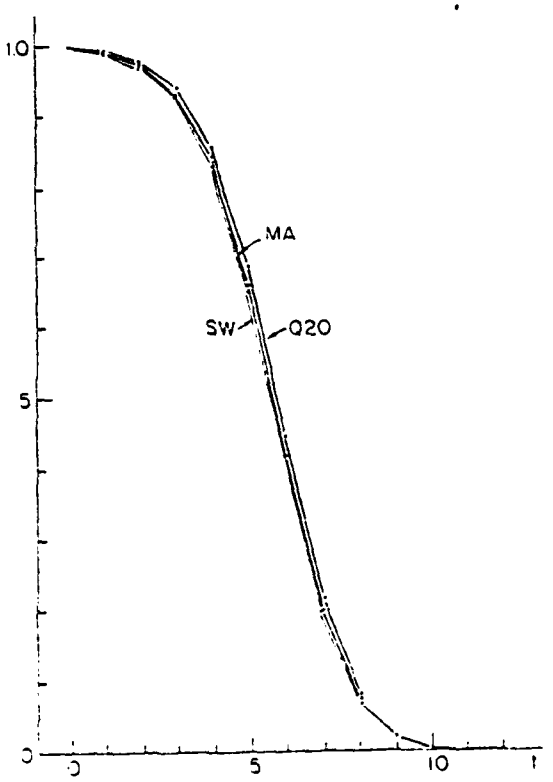

Fig.4 $b_2(t/2)$ - SW, MA, and Q20

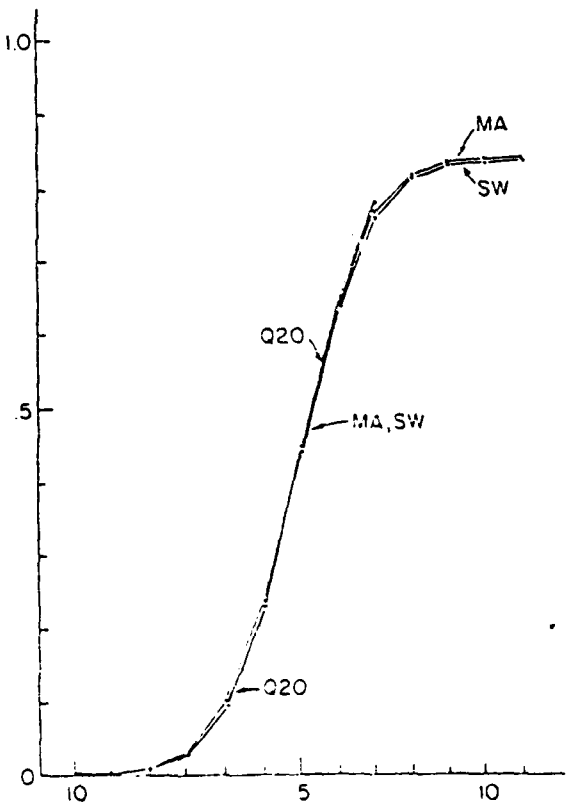

Fig.6 $\frac{t}{s=0}\ b_2(s/2)$ - SW, MA, and Q20

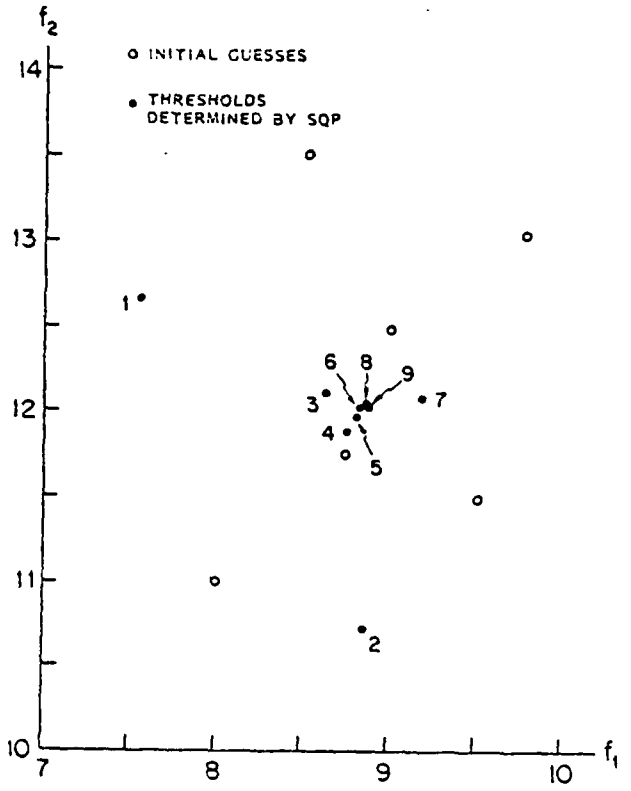

Fig.7 Thresholds Chosen by SQP

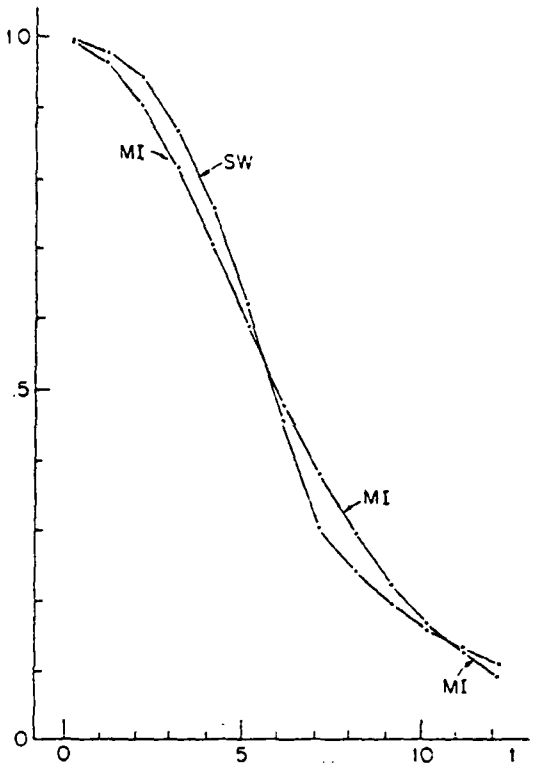

Fig.9 $b_0(t/1)$ - SW and MI

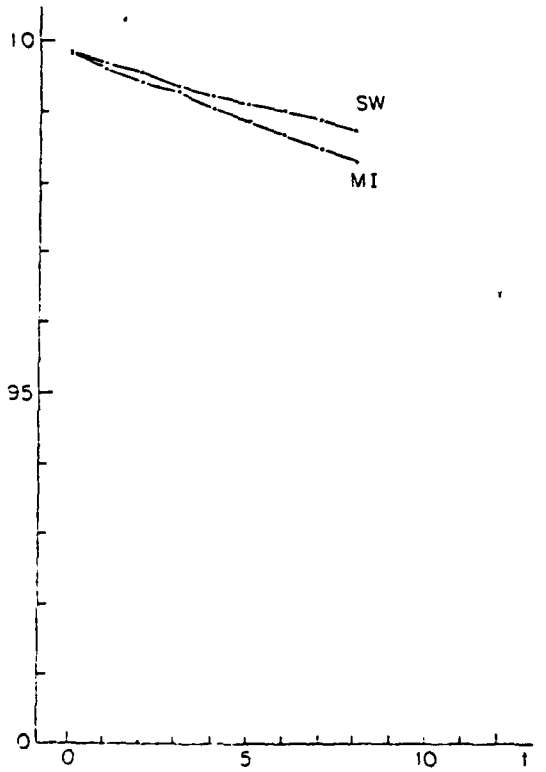

Fig.8 $b_0(t/0)$ - SW and MI

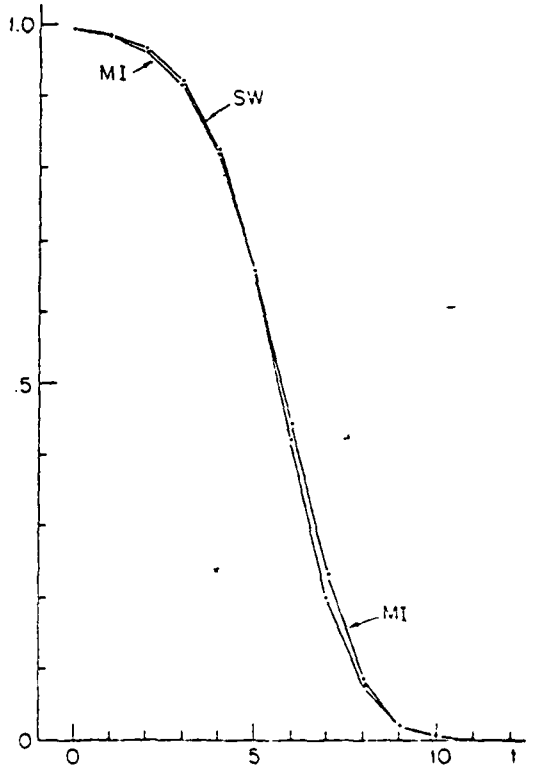

Fig.10 $b_0(t/2)$ - SW and MI

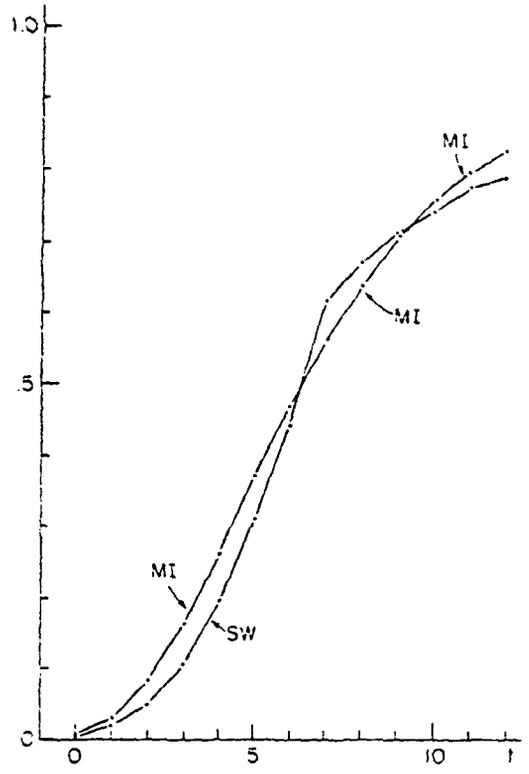

Fig.11 $\sum_{s=0}^{t} b_1(s/1)$ - SW and MI

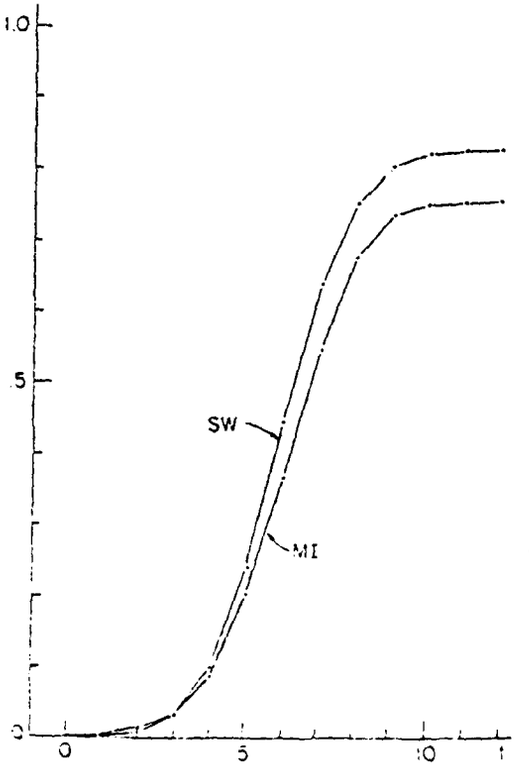

Fig.12 $\sum_{s=0}^{t} b_2(s/2)$ - SW and MI


DAT
FILM