October, 1980 LIDS-SR-1145

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139

Status Report Number Five, On the Development of a Methodology for the Detection of System Failures and for the Design of Fault-Tolerant Control Systems

ONR Contract No. N00014-77-C-0224

October 1, 1980 through September 30, 1981

To: Dr. Charles Holland
Mathematics Program (Code 432)
Office of Naval Research
800 North Quincy Boulevard
Arlington, Virginia 22217

SUMMARY

A brief description of the research carried out by faculty, staff, and students of the M.I.T. Laboratory for Information and Decision Systems under ONR Contract N00014-77-C-0224 is described. The period covered in this status report is from October 1, 1980 through September 30, 1981.

The scope of this contract is the development of an overall failure detection system design methodology and of methods for fault-tolerant control. In the following sections we overview the research that has been performed in these areas during the indicated time period. We have also included a list of the papers and reports that have been and are being written as a result of research performed under this contract. In addition, during the period mentioned above, Prof. Alan S. Willsky, principal investigator for this contract, visited the People's Republic of China and Japan. A trip report was submitted to the Mathematics Program (Code 432), and a copy of that report is included as Appendix A to this status report.

I. Robust Comparison Signals for Failure Detection

As discussed in the preceding progress report [19] for this project, a key problem is the development of methods for generating comparison signals that can be used reliably to detect failures given the presence of system parameter uncertainties. Previously we have made some initial progress in this area, as is described in [19] and in more detail in the Ph.D. dissertation of E.Y. Chow [8] and in the paper [16]. In this work we used the idea of redundancy relations which are simply algebraic relationships among system outputs at several times. Using this concept, we proposed an analytic method for determining the parameters defining a comparison signal that is optimum in the sense of minimizing the influence of parameter uncertainties on the value of the signal under normal operation.

This research represented a significant step in increasing our understanding of robust failure detection and in development a useful, complete methodology. There were, however, several key limitations to this earlier work. Specifically,

(a) No algorithmic method existed for identifying and constructing all possible redundancy relations for a given system.

(b) No method existed for constructing the set of redundancy relations which can be used to detect a given failure. More generally, no method existed for finding all sets of redundancy relations which allow one to detect each of a set of specified failures and to distinguish among them.

(c) The optimization formulation developed is complex and its use for systems of moderate size seemed prohibitive. Also, the method leads to a choice of comparison signals that depends upon the system state and input. While this may be appropriate in some problems, it is not in many others. Furthermore the formulation dealt primarily with minimizing the influence of parameter uncertainties on comparison signals under normal operation. No single, satisfactory formulation existed for incorporating the performance of the particular signal choice under both unfailed and failed system conditions.

(d) No coherent picture existed for describing the full range of possible methods for using a particular redundancy relationship and for quantitatively relating performance as measured by the optimization criterion to an actual failure detection algorithm based on the redundancy relation.

During this past year we have initiated a new, related research project aimed at developing algebraic and geometric approaches to overcoming these limitations. We have identified and begun to develop an extremely promising approach. Our work to date will be described in some detail in the forthcoming S.M. thesis proposal of Mr. Xi-Chang Lon [20]. In this section we will briefly outline the main ideas.

Consider a linear system of the form

$$x(k+1) = Ax(k) \tag{1.1}$$

$$y(k) = Cx(k) \tag{1.2}$$

For simplicity in our initial discussion here and in the first part of our research we will not include inputs (and hence will focus on sensor failures and not on actuator failures). Let $y_p(k)$ denote an extended observation vector of length p+l:

$$y_p'(k) = [y'(k), y'(k+1), \ldots, y'(k+p)] \tag{1.3}$$

Any vector

$$a'_p = [a'_{p0}, a'_{p1}, \ldots, a'_{pp}]$$

which satisfies

$$a'_p y_p(k) = \sum_{i=0}^{p} a'_{pi} y(k+i) = 0 \tag{1.4}$$

for all $k \geq 0$ and all possible $x(0)$ is called a parity vector of length p and (1.4) is called a parity check of length p.

A first key problem is to identify all possible parity vectors and to develop algorithms for generating them. The key to this is the following: define the vector of polynomials

$$p(z) = \sum_{i=0}^{P} a_{pi} z^i \tag{1.5}$$

Then a_p is a parity vector if and only if there is an nxl vector of polynomials (n is the dimension of x) q(z) so that

$$p'(z)C(zI-A)^{-1} = q'(z) \tag{1.6}$$

or, equivalently, if and only if

$$[p'(z), -q'(z)]$$

is in the left null space of the matrix

$$\begin{bmatrix} C \\ zI-A \end{bmatrix}$$

The importance of this result is that the last characterization identifies the set of parity relations with the left nullspace of a particular polynomial matrix, and, in fact, this allows us to use some of the powerful tools of the algebraic theory of linear systems to construct all possible redundancy relations and, in fact, to find a basis consisting of parity checks of minimal length. As length directly corresponds to the amount of memory involved in a parity check one intuitively would prefer short checks, in order to minimize the effects of parameter uncertainties.

Work is presently continuing in developing algorithms for constructing parity checks and for finding parity vectors that are useful for particular failure modes. Specifically, suppose that in addition to the normal operation model (1.1), (1.2) we also have a set of possible failure models

$$x(k+1) = A_i x(k) \tag{1.7}$$

$$y(k) = C_i x(k) \tag{1.8}$$

$i=1,\dots,N$. Suppose that a vector a_p is a valid parity vector, i.e. there is a $q(z)$ so that $[p'(z), -q'(z)]$ is the left nullspace of

$$\begin{bmatrix} C \\ zI-A \end{bmatrix} \tag{1.9}$$

Suppose also that there is no polynomial $q_i(z)$ so that $[p'(z), -q_i'(z)]$ is in the left nullspace of

$$\begin{bmatrix} C_i \\ zI-A_i \end{bmatrix} \tag{1.10}$$

In this case the parity check (1.4) will give a value of zero if there is no failure but will generally give a nonzero value if failure mode i occurs. Clearly then what we wish to identify are the intersections of the left nullspaces of the matrices in (1.9) and (1.10). As discussed in

[20] this can also be used to determine sets of parity checks which can distinguish among a set of failures. Work is continuing on obtaining algorithmic solutions.

The research described above is aimed directly at several of the limitations mentioned earlier. Using the results of this research we have also initiated research of a more geometric nature that is aimed at overcoming the remaining limitations. Specifically, it can be seen that the set of all parity checks of order $\leq p$ is equivalent to the orthogonal projection in $\mathbb{R}^{(p+1)m}$ $(m = \dim(y))$ onto the orthogonal complement of the range of the matrix

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^p \end{bmatrix} \tag{1.11}$$

For example, if $y' = (y_1, y_2)$ and $y_2 = \alpha y_1$, then the geometric picture is as is illustrated in Figure 1.1.

In terms of this perspective, parameter uncertainties manifest themselves as perturbations in the range of the matrix (1.11). For our example, if $\alpha_{min} \leq \alpha \leq \alpha_{max}$, we have the picture depicted in Figure 1.2. For this example it intuitively makes sense to use as a parity check the projection onto a line which is "as orthogonal as possible" to the cone of possible observation subspaces. One logical criterion is to choose a line which makes the largest possible angle with the cone -- i.e. which maximizes the smallest angle of the chosen line with any line in the cone. The idea just described can be extended to the general case, and the

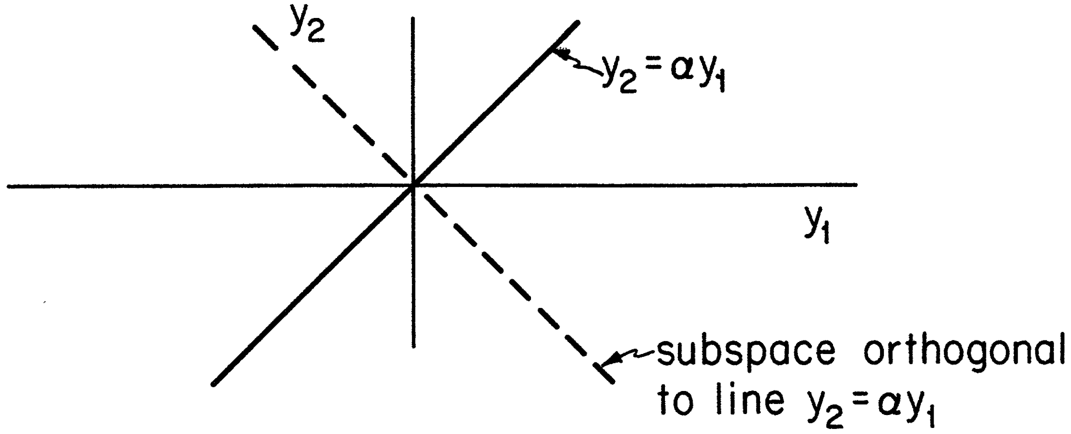

Figure 1.1

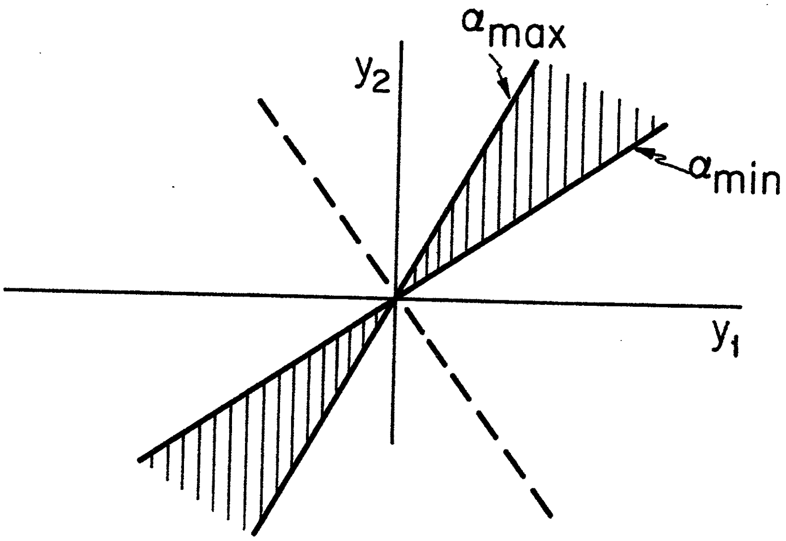

Figure 1.2

optimization problem can be stated in terms of singular values of a particular matrix. Also, this approach can be viewed as a modification of that of Chow in that it overcomes the state-dependent nature of the optimum parity check of Chow. Furthermore, this geometric approach can also be used to formulate problems which allow one to choose the optimum parity checks subject not only to performance constraints under normal operation but also when specific failures occur.

To illustrate the point mentioned at the end of the preceding paragraph, consider our simple example and suppose that a failure results in a shift in α. When there are uncertainties in α, this results in a picture as illustrated in Figure 1.3. Intuitively, we would like to use a parity check consisting of the orthogonal projection onto a line which makes a large angle with lines in the unfailed cone and a small angle with lines in the failed cone. We have obtained a "Neyman-Pearon-like" optimization formulation for this problem and are presently studying the algorithmic solution of this problem and the formulation and solution of problems of distinguishing among a set of possible failures.

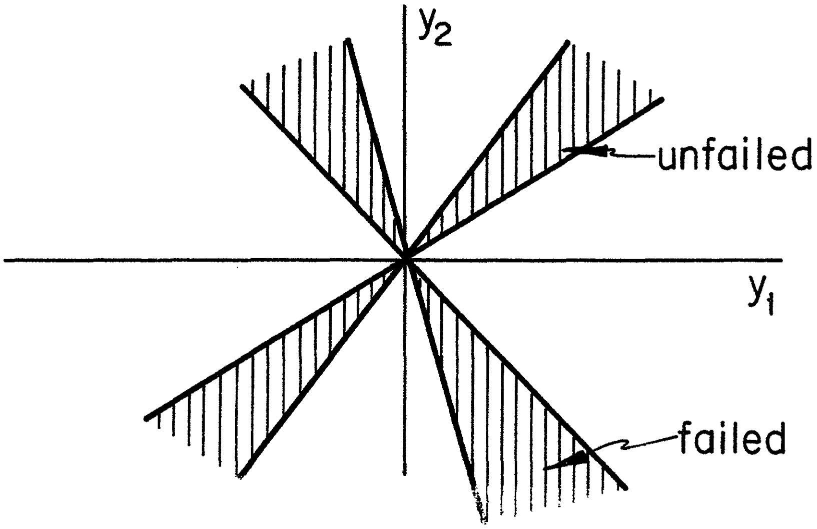

Figure 1.3

II. Fault-Tolerant Control Systems

In the preceding progress report [19] we outlined several classes of discrete-time stochastic control problems that are aimed at providing a framework for gaining an understanding of fault-tolerant optimal control. These problems involve a finite-state jump process denoting the operational status of the system. The system state x evolves according to a linear stochastic equation parametrized by the finite state process. During the past year significant progress has been made on the problems described in [19]. These results will be described in detail in the forthcoming Ph.D. thesis of Mr. H.J. Chizeck [18]. Specifically we have accomplished the following:

(1) As mentioned in [19], the problem is straightforward when ρ is independent of x. However, the qualitative properties of the solution and of the closed-loop system are suprisingly complex, and a wide variety of types of behavior can be obtained. We have now derived a series of results and constructed a set of examples which allow us to understand the possibilities in more detail.

(2) When the transition probabilities of ρ depend on x the problem becomes one of nonlinear stochastic control. This problem reveals many of the critical properties of fault-tolerant systems, including hedging and risk-avoidance. In much of our work in this area we have focussed on the scalar problem where the dependence of ρ on x is piecewise-constant. A cursory glance at this problem indicates that with this formulation the problem can be solved (via dynamic programming) by examining a growing (as we go back in time) number of constrained linear-quadratic problems.

The problem has, however, a significant amount more structure which we have now characterized. This characterization has allowed us to pinpoint the nature of hedging and risk-avoidance for these systems, to reduce the computational complexity of the solution by a substantial amount, and to obtain a finite look-ahead approximation.

(3) We have also completed an investigation of the problem described in (2) above in the presence of bounded process noise. In this case the piecewise-quadratic nature of the solution of (2) is lost in some regions, but the insight from (2) allows us to obtain an approximation to the cost-to-go which reduces the problem to one much like that without process noise.

(4) We have also obtained some initial results for the vector version of the problem of (2). In this case the situation becomes far more complex, as the regions into which one must divide the state space at each stage of the algorithm have complex shapes. Work is continuing on obtaining approximation methods for these regions much as we did for the costs-to-go in (3).

In addition to these problems we have also made progress in a fault-tolerant optimal control when we have noisy observations of the state. Specifically, we have been examining a problem in which a system may switch from normal operation to a failed condition and where our controller must decide if and when to switch from a control law optimal for normal operation (with a criterion specific to normal operation) to one optimal under failed conditions (perhaps with a different criterion). This is a novel but exceedingly important sequential decision problem. Specifically, standard statistical decision problems are aimed at providing a tradeoff between incorrect decision probabilities and decision delay. For control problems, these are only indirect performance indicators - e.g. the effect of a false alarm depends on the performance loss resulting from switching from the normal control law and the effect of detection delay depends on the performance loss from using the normal law after the system has failed.

At this point we have obtained the form of the solution, but much work remains in developing algorithms and in understanding the nature of the solution.

III. Additional Problems in Detection

During the past year we have continued and initiated work along several directions. Brief descriptions follow:

(1) Decision rules. In the work of Chow described in [8, 17, 19] we describe an algorithm for computing optimum decision rules. This algorithm was complex computationally, and extensions to more involved detection problems using this approach are prohibitively complex. The reason for this is that optimum algorithms attempt to partition the space of possible conditional probability vectors for the given set of hypotheses into decision regions. The boundaries of these regions are the points where two decisions yield *exactly* equal performance. It is our opinion that most of the computational complexity is due to this goal of finding the precise boundaries, which involves obtaining precise statistical predictions of the evolution of the conditional probabilities under each hypothesis. We have recently initiated the investigation of suboptimum algorithms based on *approximate* descriptions of the evolution of conditional probabilities. This formulation offers the possibility of solving far larger problems at reduced computational cost and with small and perhaps negligible performance loss. These possibilities remain to be examined.

(2) Complex decision problems. As discussed in [19] there is an exceedingly large and rich class of problems that involve continuous processes coupled together with discrete processes whose transitions represent events in the observed signals or the underlying systems. The methods we have developed and are developing for failure detection represent in some sense a first step in attacking the simplest problems of this type, i.e., ones in which we must detected *isolated* and *sporadic* events. We have also initiated investigations of problems in which we wish to detect and identify *sequences* of events. Such

problems are of significance for the reliable control of large scale systems and, in our opinion, hold the key for solving many complex signal processing problems. In the preceding progress report [19] we outlined a generic problem formulation for event-driven signal generation. During this past year we have built on this formulation to develop a structure for signal processing algorithms for event-driven signals. The building blocks for these algorithms are specialized detection algorithms of the type one uses for failure detection, and the key problem is one of developing decision mechanisms based on the outputs of these simple algorithms. As discussed in [19], the major issue is one of pruning the tree of possible sequences of events in an optimum manner. The approximate methods described in (1) above are potentially of great value for this problem. In addition to our analytical work, we are also working on several specific applications. This experience is exceedingly useful in providing insight into the nature of problems of this type. At this time we are working on problems of electrocardiogram analysis based on an event-driven model, efficient edge detection in images, the detection of objects given remote integral data (which is of direct application to problems of tomographic tracking of cold-temperature regions in the ocean), and optimum closed-loop strategies for searching for objects. The fact that such a wide variety of problems can be approached essentially from one unified perspective indicates, we feel, the central importance of this research effort.

(3) Event-Driven Models for Dynamic Systems. Based on the perspective in (2), we have initiated a more mathematical aspect of our research based on the development of simplified event-driven models for nonlinear systems affected by small amounts of noise and/or rare events. The motivation for this research is that the exact analysis of such models or the solutions of problems of estimation and control for such models may be considerably more complicated (often these problems are intractable) than

those for simplified models obtained through asymptotic analysis. As an example, consider the scalar stochastic system described by the stochastic differential equation

$$dx(t) = f(x(t))dt + \varepsilon dw(t) \tag{3.1}$$

where

$$f(x) = \begin{cases} -(x-1) & x > 0 \\ -(x+1) & x < 0 \end{cases} \tag{3.2}$$

This system is characterized by the property that for time intervals that are small the process behaves like a linear process near one equilibrium or another, while for long times the aggregate process sgn(x(t)) converges (as $\varepsilon\to 0$) to a Markov jump process. Consequently, one might expect that estimation of x(t) based on measurements of the form

$$dy(t) = x(t)dt + dv(t) \tag{3.3}$$

might be accomplished based on viewing the process as the state of an event-driven linear system. More generally, one can consider analogous models for other nonlinear systems possessing multiple equilibria and subject to small noise. We already have some results along the lines indicated for simple examples, and we are continuing to investigate more general situations. Note that the estimation algorithms that result are of precisely the form considered in (2). It is our feeling that this research direction represents a very promising approach to obtaining a substantial extension to the class of estimation problems for which tractable solutions can be found.

PERSONNEL

During this time period Prof. Alan S. Willsky (principal investigator), Dr. Stanley B. Gershwin, Prof. B.C. Levy, Prof. R.R. Tenney, Dr. David Castanon, Prof. Shankar Sastry, and students X.-L. Lou, H.J. Chizeck, P.C. Doerschuk, D. Rossi, C. Bunks, and M. Coderch have been involved in research outlined in this status report. Of these people Prof. Willsky, Dr. Gershwin, and Dr. Castanon have received financial support under this contract, and Mr. Lon and Mr. Chizeck have been research assistants supported by this contract.

REFERENCES

(Reference number with asterisks report work performed in part under this contract).

*1. C.S. Greene, "An Analysis of the Multiple Model Adaptive Control Algorithm," Report ESL-TH-843, Ph.D. Thesis, M.I.T., August 1978.

*2. H. Chizeck and A.S. Willsky, "Towards Fault-Tolerant Optimal Control," Proc. IEEE Conf. on Decision and Control, San Diego, Calif., Jan. 1979.

*3. C.S. Greene and A.S. Willsky, "Deterministic Stability Analysis of the Multiple Model Adaptive Control Algorithm," Proc. of the 19th IEEE Conf. on Dec. and Cont., Albuquerque, N.M., Dec. 1980; extended version to be submitted to IEEE Trans. Aut. Control.

*4. "Status Report Number One, on the Development of a Methodology for the Detection on System Failures and for the Design of Fault-Tolerant Control Systems," Rept. ESL-SR-781, Nov. 15, 1977.

5. R.V. Beard, "Failure Accomodation in Linear Systems Through Self-Reorganization," Rept. MVL-71-1, Man-Vehicle Lab., M.I.T., Cambridge, Mass., Feb. 1971.

6. H.L. Jones, "Failure Detection in Linear Systems," Ph.D. Thesis, Dept. of Aeronautics and Astronautics, M.I.T., Camb., Mass., Sept. 1973.

*7. "Status Report Number Two, on the Development of a Methodology for the Detection of System Failures and for the Design of Fault-Tolerant Control Systems," Rept. LIDS-SR-873, Dec. 27, 1978.

*8. E.Y. Chow, "A Failure Detection System Design Methdology," Ph.D. Thesis, Dept. of Elec. Eng. and Comp. Sci., M.I.T., Nov. 1980.

9. R.B. Washburn, "The Optimal Sampling Theorem for Partially Ordered Time Processes and Multiparameter Stochastic Calculus," Ph.D. Thesis, Dept. of Mathematics, M.I.T., Feb. 1979.

*10. A.S. Willsky, "Failure Detection in Dynamic Systems," paper for AGARD Lecture Series NO. 109 on Fault Tolerance Design and Redundancy-Management Techniques," Athens, Rome and London, October 1980.

*11. "Status Report Number Three, On the Development of a Methodology for the Detection of System Failures and for the Design of Fault-Tolerant Control Systems," Oct. 25, 1979.

12. B.R. Walker, "A Semi-Markov Approach to Quantifying the Performance of Fault-Tolerant Systems," Ph.D. Dissertation, M.I.T., Dept. of Aero. and Astro., Cambridge, Mass., July 1980.

*13. H.R. Shomber, "An Extended Analysis of the Multiple Model Adaptive Control Algorithm" S.M. Dissertation, M.I.T., Dept. of Elec. Eng. and Comp. Sci., also L.I.D.S. Rept. LIDS-TH-973, Feb. 1980.

*14. D.A. Castanon, H.J. Chizeck, and A.S. Willsky, "Discrete-Time Control of Hybrid Systems," Proc. 1980 JACC, Aug. 1980, San Francisco.

*15. H.J. Chizeck and A.S. Willsky, "Jump-Linear Quadratic Problems with State-Independent Rater," Rept. No. LIDS-R-1053, M.I.T., Laboratory for Information and Decision Systems, Oct. 1980.

*16. E.Y. Chow and A.S. Willsky, "Issues in the Development of a General Design Algorithm forr Reliable Failure Detection," Proc. 19th IEEE Conf. on Dec. and Control, Albuquerque, New Mexico, December 1980; extended version to be submitted for review for publication.

*17. E.Y. Chow and A.S. Willsky, "Sequential Decision Rules for Failure Detection," Proc. 1981 JACC, June 1981, Charlottesville, Virginia.

*18. H.J. Chizeck, "Fault-Tolerant Optimal Control", Ph.D. Thesis, Dept. of Elec. Eng. and Comp. Sci., M.I.T., to be completed in Nov. 1981.

*19. "Status Report Number Four, On the Development of a Methodology for the Detection of System Failures and for the Design of Fault-Tolerant Control Systems," Oct. 1980.

*20. X.-L. Lou, "Robust Failure Detection," S.M. Thesis proposal, Dept. of Elec. Eng. and Comp. Sci., M.I.T., to be completed in Nov. 1981.

APPENDIX A

Copy of the report on Prof. Alan S. Willsky's visit to the People's Republic of China and Japan.

Massachusetts Institute of Technology
Laboratory for Information and Decision Systems
Cambridge, MA 02139

REPORT ON TRAVEL TO THE PEOPLE'S REPUBLIC OF CHINA AND JAPAN

Supported in part by Contract ONR/N00014-77-C-0224

September 21, 1981

Prepared by: Alan S. Willsky

Submitted to: Dr. Charles Holland
Mathematics Program (Code 432)
Office of Naval Research
800 North Quincy Boulevard
Arlington, Virginia 22217

This report summarizes the trip of Prof. Alan S. Willsky to the People's Republic of China and Japan. The primary purposes of this trip were to participate in the Bilateral Seminar on Control Systems held in Shanghai, China and the Seventh Triennial World Congress (held in Kyoto, Japan) of the International Federation of Automatic Control. The following is the itinerary followed by Prof. Willsky:

August 9-12	Shanghai, China
August 13-16	Xian, China
August 16-19	Beijing, China
August 19-22	Tokyo, Japan
August 22-29	Kyoto, Japan

Prof. Willsky served as technical program chairman for the meeting in Shanghai and as one of the organizers of the activities of the official IEEE Control Systems Society delegation during the entire visit to China. In addition, Prof. Willsky was one of three plenary speakers during the Bilateral Seminar. The subject of his talk was an introduction to and survey of methods for the detection of abrupt changes in signals and systems. Prof. Willsky's research in this field has been and is presently supported in part by ONR.

The basic purpose of the visit by the IEEE delegation was to establish ties between the Control Systems Society and the Chinese Association of Automation and to provide an opportunity for discussion among researchers from both organizations. To achieve these objectives, the delegation organizers structured the visit to allow for ample opportunity for discussion and for members of the IEEE delegation to gain knowledge and understanding about China, the Chinese people, and research in China. In addition to the 3-day meeting in Shanghai, there were also visits to Xian and Beijing. Cultural, social, and technical activities were organized in both of these cities. In Xian the delegation visited Xian Jiaotong University, and Prof. Willsky was involved in a discussion of implementation issues for digital control systems. Also involved in this discussion was Dr. Stuart L. Brodsky of ONR. In Beijing a technical interchange was held at The Great Hall of the People.

Overall the visit to China was exceedingly worthwhile. The meeting in Shanghai was a significant success, and the contacts made there will allow for continued interaction. In particular, a number of Chinese researchers expressed great interest in Prof. Willsky's lecture and provided him with information and publications concerning research on failure detection and adaptive control in China. The visit to Xian Jiaotong University was also valuable, as it provided the opportunity to see one of China's leading and fastest growing technical universities. Beyond these specific scheduled events the many informal, unscheduled discussions at banquets provided further information about research at institutions that were not visited.

The other major portion of this trip was the IFAC World Congress, the largest (approximately 1500 attendees) meeting of researchers in automatic control. In addition to presenting a paper on implementation issues in digital control and attending various technical sessions, Prof. Willsky also had the opportunity to discuss research topics with researchers from many countries. In particular, Prof. Willsky engaged in numerous discussions on problems of abrupt changes, failure detection and fault-tolerant control. Prof. Willsky spoke with Prof. K. Åström of Sweden, Prof. L. Ljung of Sweden, Dr. F. Pau of France, Prof. V. Utkin of the Soviet Union, and Prof. A. Halme of Finland, among others. These discussions were of great value in updating Prof. Willsky's knowledge of related research around the world. In addition, Prof. Willsky also was able to learn much about the status and direction of robotics research in Japan. As this represents an important and promising direction for future research, the opportunity provided by this visit to Jpan was a significant one.

APPENDIX B

Copy of the paper

Sequential Decision Rules for Failure Detection

by

Edward Y. Chow
Alan S. Willsky

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 decision 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.

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) behavipr. 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 $\tau(\theta=(i,\tau,\nu))$ 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 accurrate 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 (success) parameter ρ_i. The Bernoulli process (corresponding to the Poisson process in continuous time) is a common model for failures in physical components; the independence assumption

* 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.

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

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

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 μ 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 $\delta\epsilon\mathcal{D}(k)$ 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:

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

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:

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

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

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

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

(Φ^*,D^*) is called the Bayes Sequential Decision Rule (BSDR) with respect to μ, 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

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

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

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

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

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

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

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 $\bar{t}(i,\tau)$ 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

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

Equation (1) indicates that the sequential Bayes risk is a weighted combination of the condtional 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 (μ) 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 mst 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 $\{r(k-W+1),\ldots,r(k)\}$, or equivalently, the space of vectors of posterior probabilities, likelihood ratios, or log likelihood ratios ($\bar{L}$) 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 L (which is related to $\bar{L}$ by a constant):

$$L(k)=[L_0'(k),\ldots,L_{W-1}'(k)]'$$

where

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

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

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

$\mathcal{L}(k)\varepsilon S_0$, 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:

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

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

where $S(j,\bar{t})$ is the stop-to-declare $(j,k-\bar{t})$ region and $S(0,-)$ is the continue region (see Fig. 1). Generally the ε's may be regarded as design parameters, but here, $\varepsilon(j,\bar{t})$ is simply taken to be the standard deviation of $\mathcal{L}(k,j,\bar{t})$.

To evaluate $U_s^W(f)$, we need to determine the set of probabilities, $\{Pr\{\mathcal{L}(k)\varepsilon S(j,\bar{t}),\mathcal{L}(k-1)\varepsilon S(0,-),\ldots,\mathcal{L}(W)\varepsilon S(0,-)|i,\tau\},\ k\geq W,\ j=0,1,\ldots,M,\ \bar{t}=0,\ldots,W-1\}$, 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 $S(0,-)x\ldots xS(0,-)xS(j,\bar{t})$, which, for large kMW, becomes extremely unwieldy and difficult to evaluate.

The MW-dimensional vector of decision statistics $\mathcal{L}(k)$ 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 failure time resolution power provided by the full window of decision statistics is not needed. Instead, decision rules that employ a few components of $\mathcal{L}(k)$ 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 $\mathcal{L}(k)$:

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

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

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

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

where

$$S_0(k)=\{\mathcal{L}_{W-1}(k)\varepsilon S_0,\ldots,\mathcal{L}_{W-1}(W)\varepsilon S_0\}$$

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

$$p(\mathcal{L}_{W-1}(k+1)|S_0(k),i,\tau)=$$
$$[\int_{S_0}p(\mathcal{L}_{W-1}(k)|S_0(k-1),i,\tau)d\mathcal{L}_{W-1}(k)]^{-1}$$
$$x\int_{S_0}p(\mathcal{L}_{W-1}(k+1)|\mathcal{L}_{W-1}(k),S_0(k-1),i,\tau)\cdot$$
$$p(\mathcal{L}_{W-1}(k)|S_0(k-1),i,\tau)d\mathcal{L}_{W-1}(k) \quad k>W \qquad (5)$$

$$Pr\{\mathcal{L}_{W-1}(k)\varepsilon S_j,\ S_0(k-1)|i,\tau\}=Pr\{S_0(k-1)|i,\tau\}\cdot$$
$$\int_{S_j}p(\mathcal{L}_{W-1}(k)|S_0(k-1),i,\tau)d\mathcal{L}_{W-1}(k),\ j=0,1,\ldots,M \qquad (6)$$

with

$$Pr\{\mathcal{L}_{W-1}(W)\varepsilon S_j|i,\tau\}=\int_{S_j}p(\mathcal{L}_{W-1}(W)|i,\tau)d\mathcal{L}_{W-1}(W) \qquad (7)$$

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

Unfortunately, the transition density, $p(\mathcal{L}_{W-1}(k+1)|\mathcal{L}_{W-1}(k),S_0(k-1),i,\tau)$, required in (5) is difficult to calculate, because $\mathcal{L}_{W-1}(k)$ 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 $\mathcal{L}_{W-1}(k)$ we are, in fact, approximating the behavior of $\mathcal{L}_{W-1}$. 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 $\mathcal{L}_{W-1}(k)$, $\ell(k)$ is Gaussian. Secondly, $\ell(k)$ is Markov so that its transition density can be readily determined. In order to have $\ell(k)$ behave like $\mathcal{L}_{W-1}(k)$, we set the matrices A and B and the mean of ℓ such that

$$E_{i,\tau}\{\ell(k)\}=E_{i,\tau}\{\mathcal{L}_{W-1}(k)\} \qquad (8)$$

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

$$E_{0,-}\{\ell(k)\ell'(k+1)\}=E_{0,-}\{\mathcal{L}_{W-1}(k)\mathcal{L}_{W-1}(k+1)\} \qquad (10)$$

That is, we have matched the marginal density and the one-step cross-covariance of $\ell(k)$ to those of $\mathcal{L}_{W-1}(k)$. 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

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

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

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

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

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

$$P_r\{L_{W-1}(k)\varepsilon S_j, S_0(k-1)|i,\tau\} \approx P_r\{\ell(k)\varepsilon S_j, S_0(k-1)|i,\tau\}$$
$$j=0,1,\ldots,M \quad k\geq W$$

and

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

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

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

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: $r(\tilde{k}) = [r'(v\tilde{k}-v+1), r'(v\tilde{k}-v+2),\ldots,r'(v\tilde{k})]'$ for some batch size $v>0$ with $\tilde{k}=1,2,\ldots$ as the new time index. In using $\tilde{r}(\tilde{k})$ 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:

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

where $\tilde{A}$ is a constant stable MxM matrix, and $\tilde{B}$ 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 $\tilde{A}$ and $\tilde{B}$. When the failure signatures under consideration are constant biases, $\tilde{B}$ can simply be set to equal G_0, and $\tilde{A}$ can be chosen to be αI, where $0<\alpha<1$. Then, the term Br 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 $\tilde{A}$ may still be handled in the same way as in the constant-bias case, but the selection of a $\tilde{B}$ matrix is more involved. With some insights into the nature of the signatures, a reasonable choice of $\tilde{B}$ 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 $\tilde{B}$ is available:

$$\tilde{B} = \begin{bmatrix} \beta'_1 \\ \beta'_2 \end{bmatrix}.$$

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 $\tilde{B}$ 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 $r(\tilde{k})=[r'(2\tilde{k}-1),r'(2\tilde{k})]'$, $\tilde{k}=1,2,\ldots$. Then we can set $\tilde{B}$ to be

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

Thus, the first and second rows of $\tilde{B}$ capture the constant-bias and ramp nature g_1 and g_2, respectively (and this $\tilde{B}$ has rank two). The use of the modified resudual $\tilde{r}(\tilde{k})$ 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

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

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

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

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

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:

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

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

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

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 $P_t(j|i)$ 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

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

where

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

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

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 $\Pr\{S_0(T)|0,-\} \approx 1$ 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.

$$L_F = 9$$

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

$$c_1=c_2=1$$

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

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

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, μ, 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 [8.85, 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 ℓ 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 Randon 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).

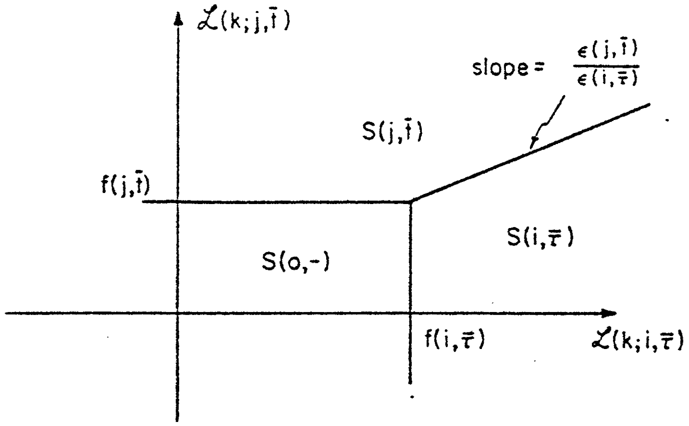
$\mathcal{L}(k;j,\bar{1})$
slope = $\frac{\epsilon(j,\bar{1})}{\epsilon(i,\bar{\tau})}$
$S(j,\bar{1})$
$f(j,\bar{1})$
$S(i,\bar{\tau})$
$S(o,-)$
$f(i,\bar{\tau})$
$\mathcal{L}(k;i,\bar{\tau})$

Fig.1 Sequential Decision Regions in 2 Dimensions

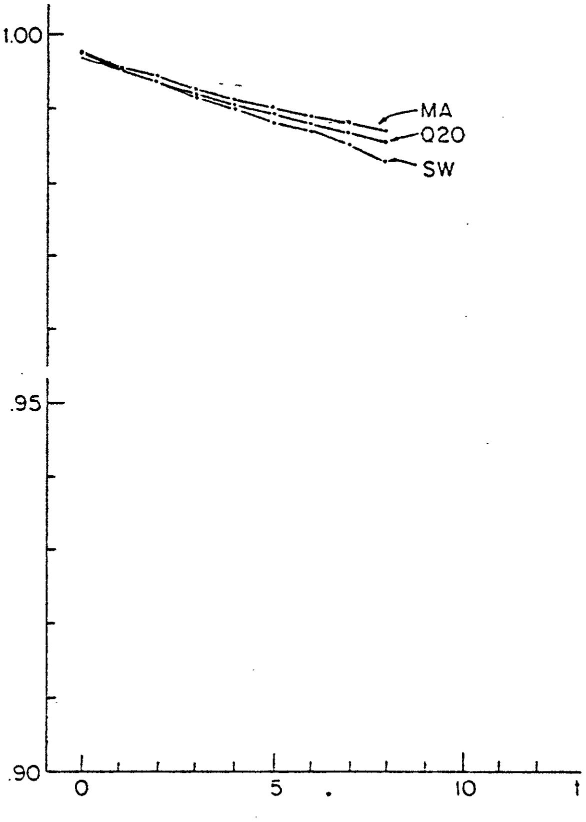

Fig.2 $b_0(t/0)$ - SW, MA, and Q20

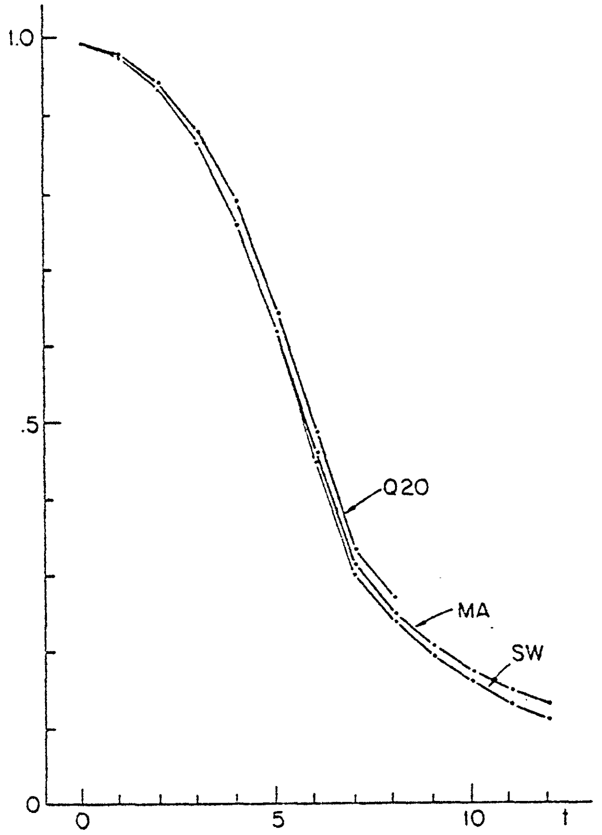

Fig.3 $b_0(t/1)$ - SW, MA, and Q20

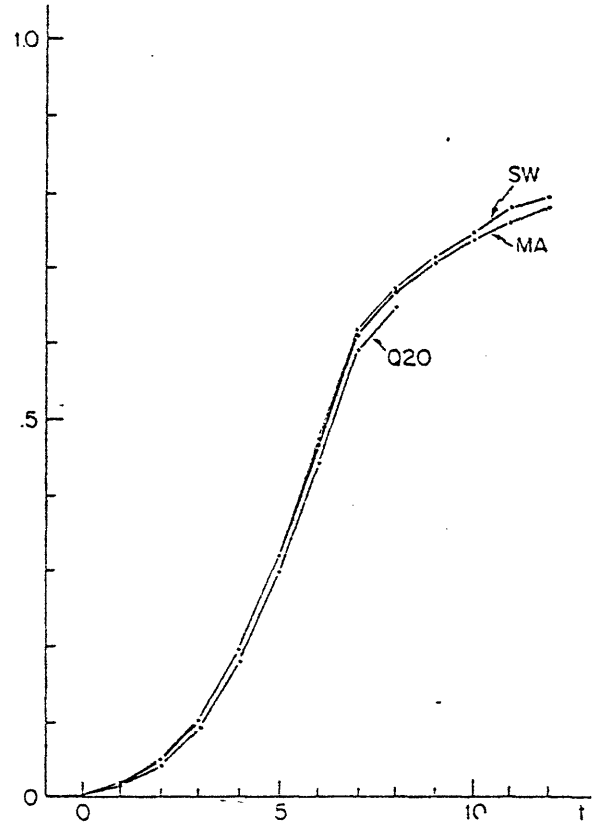

Fig.5 $\sum_{s=0}^{t} b_1(s/1)$ - SW, MA, and Q20

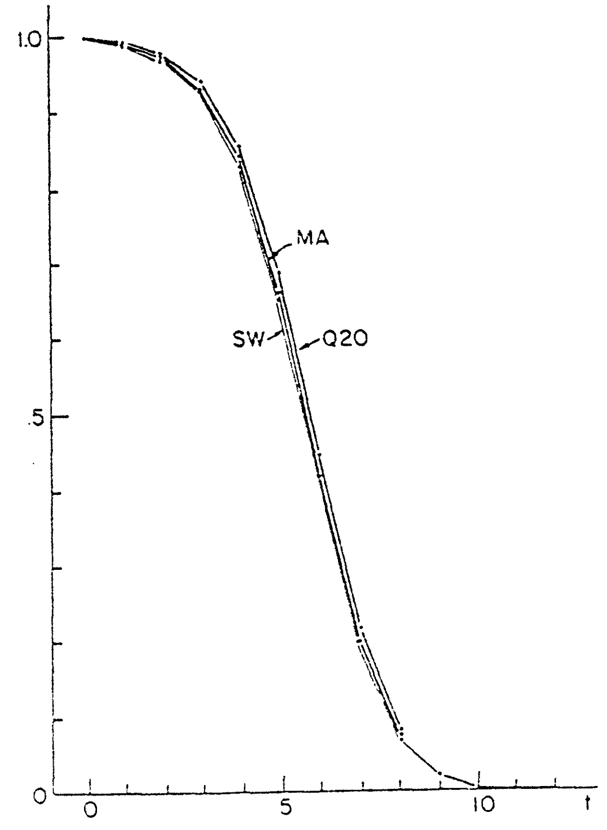

Fig.4 $b_0(t/2)$ - SW, MA, and Q20

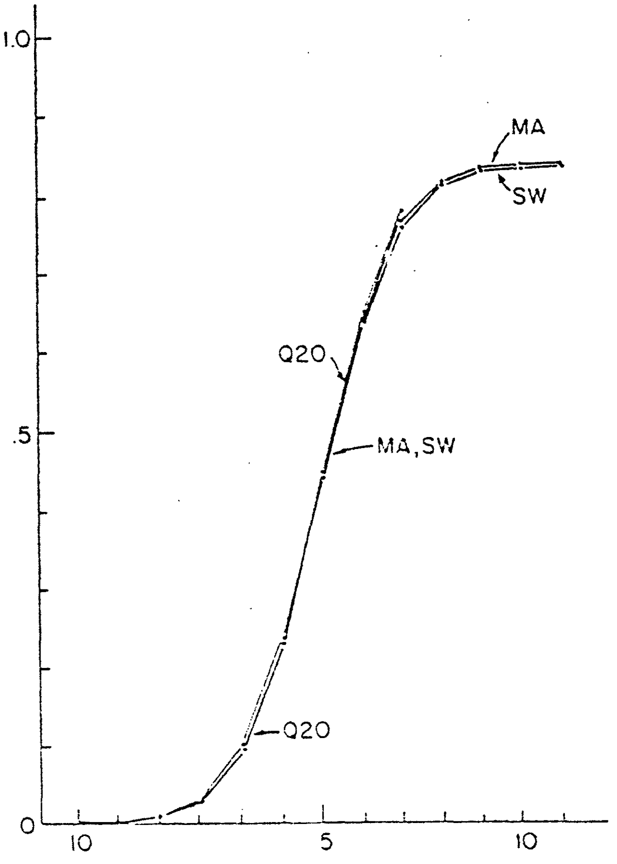

Fig.6 $\sum_{s=0}^{t} b_2(s/2)$ - SW, MA, and Q20

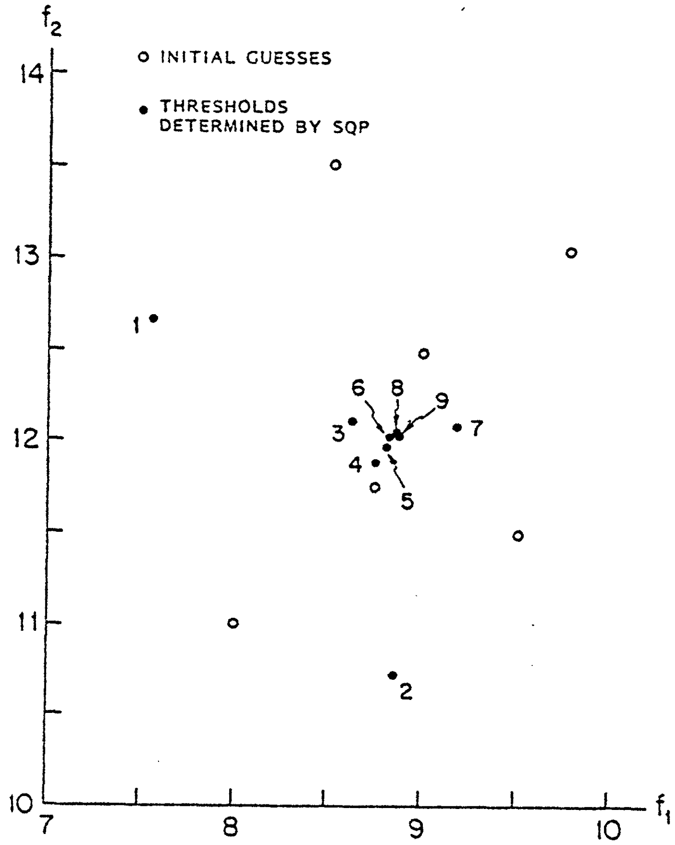

Fig.7 Thresholds Chosen by SQP

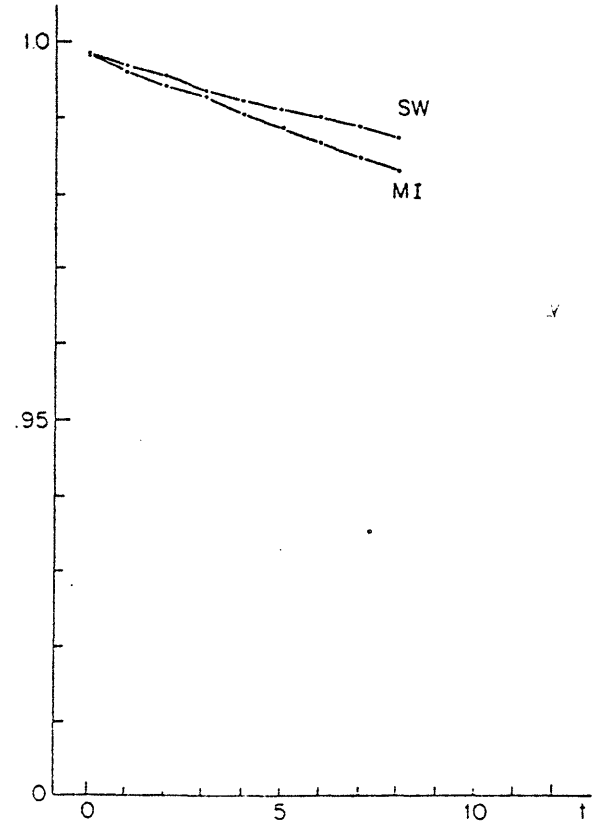

Fig.8 $b_0(t/0)$ - SW and MI

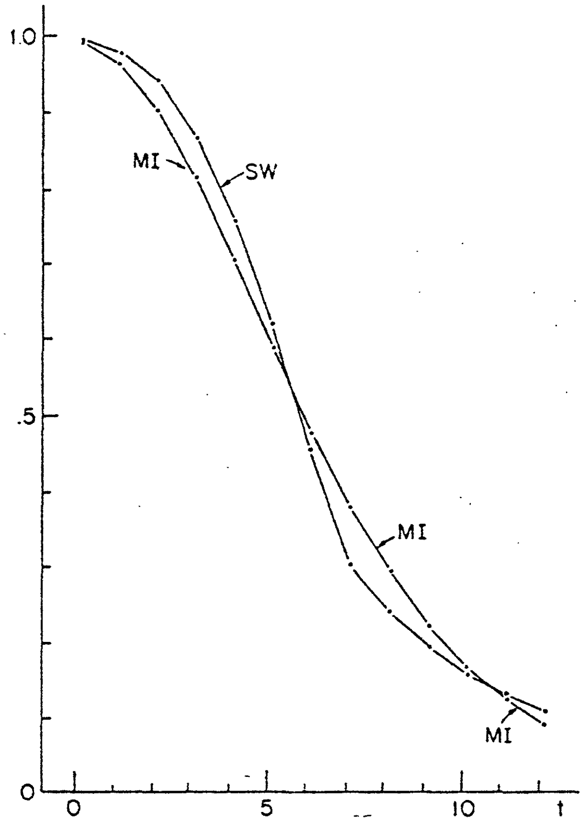

Fig.9 $b_0(t/1)$ - SW and MI

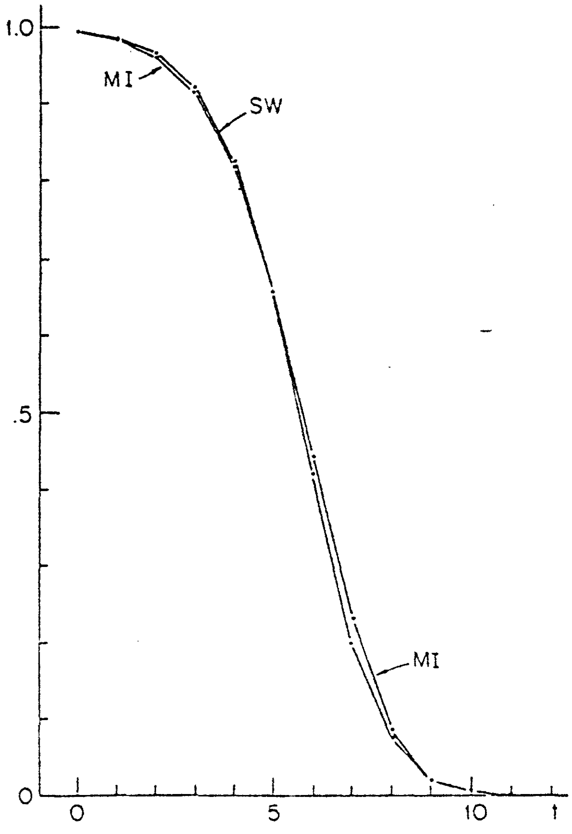

Fig.10 $b_0(t/2)$ - SW and MI

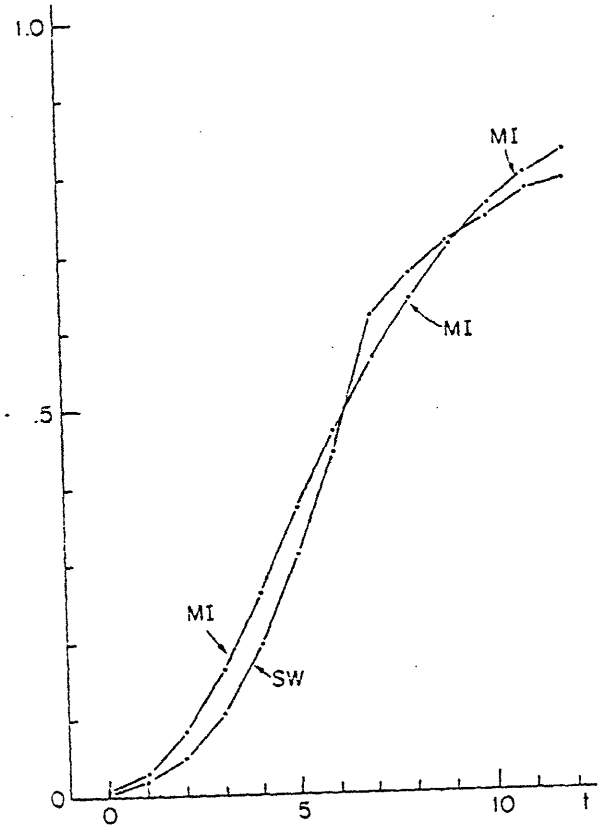

Fig.11 $\sum_{s=0}^{t} b_1(s/1)$ - SW and MI

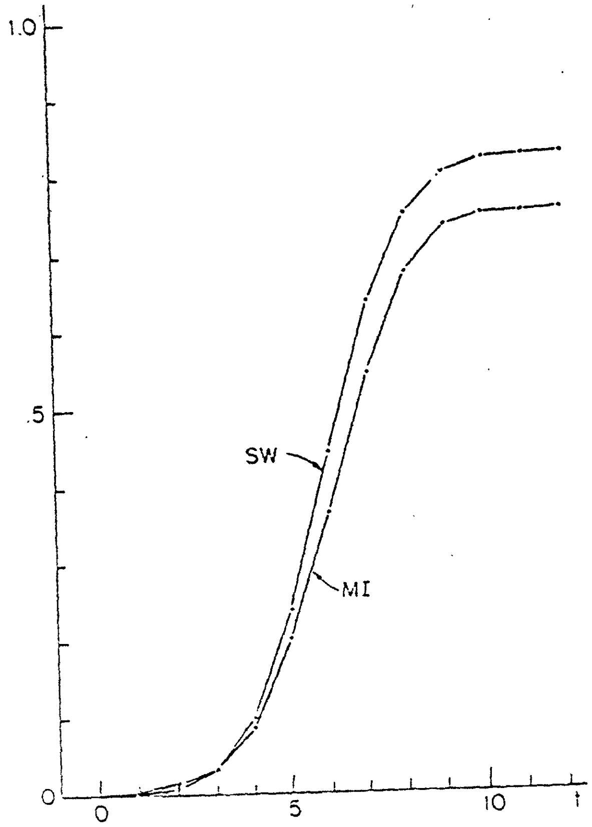

Fig.12 $\sum_{s=0}^{t} b_2(s/2)$ - SW and MI