AD-A059 977 TEXAS TECH UNIV LUBBOCK INST FOR ELECTRONICS SCIENCE F/G 9/5
ANNUAL REVIEW OF RESEARCH UNDER THE JOINT SERVICES ELECTRONICS --ETC(U)
OCT 78 R SAEKS, K S CHAO, R B ASHER, J WALKUP N00014-76-C-1136

UNCLASSIFIED NL

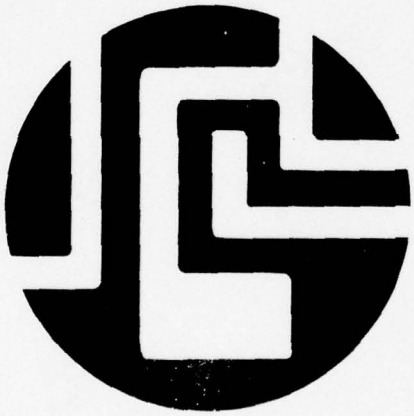
1 OF 2
ADA
059977

ANNUAL REVIEW OF RESEARCH

under the

JOINT SERVICES ELECTRONICS PROGRAM

October 1978



Institute for Electronics Science

TEXAS TECH UNIVERSITY
Lubbock, Texas 79409

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) ANNUAL REVIEW OF RESEARCH UNDER THE JOINT SERVICES ELECTRONICS PROGRAM,	5. TYPE OF REPORT & PERIOD COVERED (9) INTERIM Rept.; 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) R. SAEKS, K.S. CHAO, R.B. ASHER, J. WALKUP, T. NEWMAN, L.R. HUNT, G.J. MURRAY	8. CONTRACT OR GRANT NUMBER(s) (15) N00014-76-C-1136
9. PERFORMING ORGANIZATION NAME AND ADDRESS TEXAS TECH UNIVERSITY INSTITUTE FOR ELECTRONIC SCIENCE LUBBOCK, TEXAS 79409	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 122105
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE 10/1/78; 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) OFFICE OF NAVAL RESEARCH 800 N. QUINCY AVE. ARLINGTON, VA.	15. SECURITY CLASS. (of this report) UNCLASSIFIED; 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE-DISTRIBUTION UNLIMITED (11) 1 Oct 78 (12) 140 p.	
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) FAULT ANALYSIS; STOCHASTIC CONTROL AND ESTIMATION; NONLINEAR CONTROL; MULTIDIMENSIONAL SYSTEM THEORY; OPTICAL NOISE; PATTERN RECOGNITION; LARGE-SCALE SYSTEMS	
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) THE ~~FOLLOWING~~ REPORT REPRESENTS THE THIRD REVIEW OF RESEARCH CONDUCTED UNDER THE AUSPICES OF THE JOINT SERVICES ELECTRONICS PROGRAM AT THE INSTITUTE FOR ELECTRONIC SCIENCE AT TEXAS TECH UNIVERSITY. SPECIFIC TOPICS COVERED INCLUDE FAULT ANALYSIS, STOCHASTIC CONTROL AND ESTIMATION, NONLINEAR CONTROL, MULTI-DIMENSIONAL SYSTEM THEORY, OPTICAL NOISE, AND PATTERN RECOGNITION.	

78 10 10 090

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

410 354

REVIEW OF RESEARCH

under the

JOINT SERVICES ELECTRONICS PROGRAM

at the

INSTITUTE FOR ELECTRONIC SCIENCE

TEXAS TECH UNIVERSITY

October 1978

Lubbock, Texas 79409

PREFACE

The following report represents the third review of research conducted under the auspices of the Joint Services Electronics Program at the Institute for Electronics Science at Texas Tech University. Specific topics covered include, fault analysis, large-scale systems, stochastic control and estimation, nonlinear control, multidimensional system theory, optical noise, and pattern recognition.

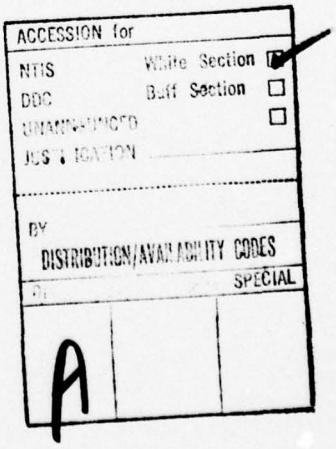


Table of Contents

RESEARCH

on

FAULT ANALYSIS

N. Sen* and R. Saeks

DEPARTMENT OF ELECTRICAL ENGINEERING

TEXAS TECH UNIVERSITY

*Presently with the Datapoint Corp. San Antonio, Texas

ABSTRACT

The fault diagnosis problem for a linear system whose transfer function matrix is measured at a discrete set of frequencies is formalized. A measure of solvability for the resultant equations and a measure of testability for the unit under test is developed. These, in turn, are used as the basis of algorithms for choosing test points and test frequencies.

INTRODUCTION

Conceptually, the fault analysis problem for an analog circuit or system amounts to the measurement of a set of externally accessible parameters of the system from which one desires to determine the internal system parameters or equivalently locate the failed components as illustrated in Figure 1.

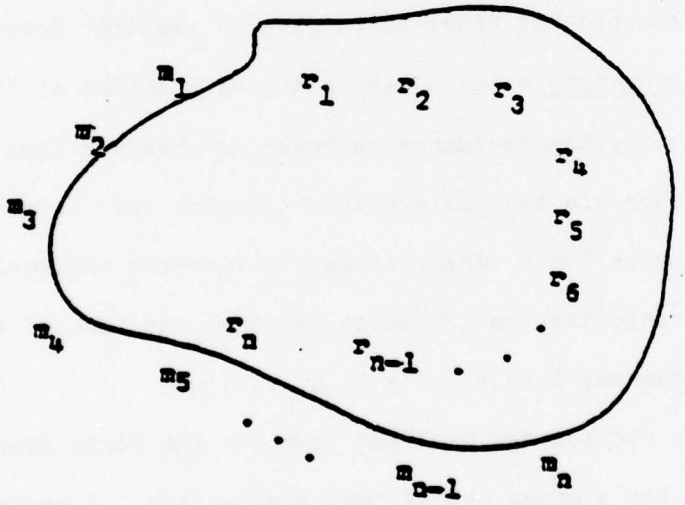


Figure 1. Conceptual Model of Fault Diagnosis Problem.

Here, the measurements, m_i, may represent data taken at distinct test points or alternatively, data taken at a fixed test point under different stimuli. Similarly, the r_i represent parameters characterizing the various internal system components. Here, a single parameter may characterize an entire component, say a resistance, capacitance or inductance. Alternatively, a component may be

represented by several parameters: the h-parameters of a transistor, the poles and gain of an op-amp, etc. In general, one models a system component by the minimum number of parameters which will allow the failure to be isolated up to a shop replaceable assembly (SRA) with all "allowed" system failures manifesting themselves in the form of some parameter change.

To solve the fault diagnosis problem, one then measures $m = \text{col}(m_i)$ and solves a nonlinear algebraic equation

$$m = F(r) \tag{1}$$

for $r = \text{col}(r_i)$ to diagnose the fault. The parameters in the resultant r vector which are out of tolerance then indicate the faulty component.[6]

The purpose of the present paper is to give an explicit formulation of the fault diagnosis equations which arise in the maintanence of linear systems. Here, one measures the system frequency response as observed from a specified set of externally accessible test points at a discrete set of frequencies and it is desired to solve for a vector of internal system parameters, r, which completely characterize the frequency response matrices of the individual system components; $Z_i(s,r)$, $i = 1, 2, \ldots, q$.

In the following section the explicit form for the fault diagnosis equations is derived for a given set of test frequencies. A measure of solvability of these equations is then developed in section3 and empolyed in section 4 in an algorithm for optimally selecting test frequencies. The measure of solvability for the fault analysis equations, given an optimal choice of test frequencies, is then taken as a measure of testability[1,2,5] for the unit under test (UUT) and is used as the basis of an algorithm for the optical choice of test points.[3,4,5] Finally, a number of examples are presented in secion 5.

EXPLICIT FORM OF THE FAULT DIAGNOSIS EQUATIONS

In the case of a linear time-invariant circuit or system, the fault diagnosis equations may be expressed in analytical form.[6] Since the fault diagnosis equations deal with the relationship between the externally measureable system parameters, m, and the internal component parameters, r, we adopt a component connection model as the starting point for the derivation of the fault diagnosis equations.[7,8] This is one of several commonly employed large scale system models in which the components and connections in a circuit or system are modeled by distinct equations, thereby permitting one to explicitally deal with the relationship between the individual component parameters and the composite system parameters.

Since the present study is restricted to linear time-invariant systems, we assume that each component is characterized by a transfer function matrix which is dependent on the potentially variable component parameters, $Z_i(s,r)$. For the classical RLC components $Z_i(s,r)$ may take the form R, Ls, or 1/sC for the case of a resistor, inductor, or capacitor, respectively. More generally, one may model an op-amp by the transfer function $k/(s-p_1)(s-p_2)$ where the parameter vector, r, now represents the three potentially variable component parameters; k, p_1, p_2; or a delay by ke^{sT}, etc. Although the symbol Z is used, the components are not assumed to be represented by impedance matrices. Indeed, hybrid models are used in most of our examples. For the purpose of analysis, it is assumed that all faults manifest themselves in the form of changes, possibly catastophic, in the parameter vector, r, with the frequency characteristics of the components unchanged. Although not universal, this fault hypothesis covers the most commonly encountered situations and subsumes the common industrial practice of assuming that all failures in analog circuits and systems take the form of open and short circuited components.[9]

Our system components are thus characterized by a set of simultaneous equations

$$2. \qquad b_i = Z_i(s,r)a_i \qquad i = 1, 2, \ldots, q$$

where a_i and b_i denote the component input and output vectors, respectively. For notational brevity, these component equations may be combined into a single block diagonal matrix equation

$$3. \qquad b = Z(s,r)a$$

where $b = \text{col}(b_i)$, $a = \text{col}(a_i)$ and $Z(s,r) = \text{diag}\,(Z_i(s,r))$.

Although there are many ways to represent the connection in a circuit or system; say a block diagram, linear graph or signal flow graph, any such representation is simply a graphical means for displaying a set of connection equations: Kirchoff laws, adder equations, etc. As such, for our component connection model we adopt a purely algebraic connection model in which the connection equations are displayed explicitally without the intermediary of some kind of graphical connection diagram. This takes the form

$$4. \qquad \begin{aligned} a &= L_{11}b + L_{12}u \\ y &= L_{21}b + L_{22}u \end{aligned}$$

where u and y represent the vectors of accessible inputs and outputs which are available to the test system. In simple systems, the connection matrices, L_{ij}, are usually obtainable by inspection, whereas, in more complex systems, computer codes have been developed for their derivation.[7] Moreover, they are assured to exist in all but the most pathalogical systems.[8]

It is the pair of simultaneous matrix equations 3 and 4 which are termed the component connection model. By combining equations 3 and 4 to eliminate the component input and output variables, a and b, one may derive[6,7] an expression for the transfer function matrix observable by the test system between the test

input and output vectors, u and y, obtaining

$$5. \qquad S(s,r) = L_{22} + L_{21}(1 - Z(s,r)L_{11})^{-1}Z(s,r)L_{12}$$

where

$$6. \qquad y = S(s,r)u$$

For a linear time-invariant system the transfer function S(s,r) is a complete description of the measurable data about the unit under test available to the test system. Moreover, being rational it is completely determined by its value at a finite number of frequencies. As such, without loss of generality, we may take our measured data to be of the form

$$7. \qquad \mathrm{col}[S(s_1,r),\ S(s_2,r),\ \ldots,\ S(s_k,r)]$$

The <u>fault diagnosis equations</u> then take the form

$$8. \qquad \begin{bmatrix} S(s_1,r) \\ S(s_2,r) \\ \vdots \\ S(s_k,r) \end{bmatrix} \quad \begin{bmatrix} L_{22} + L_{21}(1-Z(s_1,r)L_{11})^{-1}Z(s_1,r)L_{12} \\ L_{22} + L_{21}(1-Z(s_2,r)L_{11})^{-1}Z(s_2,r)L_{12} \\ \vdots \\ L_{22} + L_{21}(1-Z(s_k,r)L_{11})^{-1}Z(s_k,r)L_{12} \end{bmatrix}$$

Since S(s,r) is, in general, a matrix, the fault diagnosis equations as derived above take the form of a matrix (col[$S(s_i,r)$])valued function of a vector valued variable, r. Computationally, however, we prefer to work with a vector valued function of a vector valued variable and hence, we transform S(s,r) into a column vector via

9. $$\mathrm{vec}[S(s,r)] = \mathrm{Col}\,[S^i(s,r)]$$

where $S^i(s,r)$ denotes the ith column of the matrix, $S(s,r)$. With the aid of the identity $\mathrm{vec}[XYZ] = [Z^t \otimes X]\mathrm{vec}[Y]$ equation 8. then transforms into[7,12]

10. $$M = \begin{bmatrix} \mathrm{vec}[S(s_1,r)] = \mathrm{vec}[L_{22}] + [L_{12}^t \otimes L_{21}(1-Z(s_1,r)L_{11})^{-1}]\mathrm{vec}[Z(s_1,r)] \\ \mathrm{vec}[S(s_2,r)] = \mathrm{vec}[L_{22}] + [L_{12}^t \otimes L_{21}(1-Z(s_2,r)L_{11})^{-1}]\mathrm{vec}[Z(s_2,r)] \\ \vdots \\ \mathrm{vec}[S(s_k,r)] = \mathrm{vec}[L_{22} + L_{12}^t \otimes L_{21}(1-Z(s_k r)L_{11})^{-1}]\,\mathrm{vec}[Z(s_k,r)] \end{bmatrix} = F(r)$$

which is the form of the fault diagnosis equations with which we desire to work.

SOLVABILITY OF THE FAULT DIAGNOSIS EQUATIONS

For the fault diagnosis equations derived above to be a viable tool of circuit and system diagnosis two fundamental questions remain to be answered: "What test frequencies should be employed to optimize the solvability of the equations?"and "How solvable are the equations given an optimal choice of test frequencies?" Both of these questions, in turn, hinge on the development of some type of measure of solvability for the fault diagnosis equations.

For a set of linear equations

11. $$m = Fr$$

where r is an n-vector, m is a p-vector and F is a p by m matrix one may characterize the solvability of the equations in terms of the number of arbitrary parameters in its solution (if a solution exists). As such, $\delta = n - \mathrm{rank}(F)$ is a natural measure of the solvability for equation 11. Here, $\delta = 0$ implies that the equation has a unique solution, $\delta = 1$ implies that the solution is determined up

to one arbitrary parameter and so on, with increasing values of δ representing decreasing degrees of solvability.

Unfortunately, the fault diagnosis equations are nonlinear even for linear systems and hence we must resort to the implicit function theorem to obtain a measure of solvability analogous to the above.[13] Indeed, if r_f is a solution to the fault diagnosis equations, then r_f is determined up to a

$$12. \qquad \delta(r_f) = n - \operatorname{rank}\left[\left[\frac{dF}{dr}(r_f)\right]\right]$$

dimensional manifold (of arbitrary parameters) in a neighborhood of r_f. Here dF/dr is the Jacobian matrix of partial derivatives of F with respect to r. With the aid of the matrix identity $d(M^{-1})/dr = -M^{-1}[dM/dr]M^{-1}$, dF/dr can be computed explicitally from equations 8. and 10. yielding

$$13. \qquad \frac{dF}{dr}(r_f) = \begin{bmatrix} \{([1+L_{11}(1-Z(s_1,r_f)L_{11})^{-1}Z(s_1,r_f)]L_{12})^t \otimes (L_{21}(1-Z(s_1,r_f)L_{11})^{-1}\}[(\text{dvec } Z(s_1,r_f)]/dr \\ \{([1+L_{11}(1-Z(s_2,r_f)L_{11})^{-1}Z(s_2,r_f)]L_{12})^t \otimes (L_{21}(1-Z(s_2,r_f)L_{11})^{-1}\}[(\text{dvec } Z(s_2,r_f)]/dr \\ \vdots \\ \{([1+L_{11}(1-Z(s_k,r_f)L_{11})^{-1}Z(s_k,r_f)]L_{12})^t \otimes (L_{21}(1-Z(s_k,r_f)L_{11})^{-1}\}[(\text{dvec } Z(s_k,r_f)]/dr \end{bmatrix}$$

where "t" denotes matrix transposition and $\otimes$ denotes the matrix Kronecker (or tensor) product.

The difficulty with the implicit function theorem is that it only yields local information valid in a neighborhood of a solution. Fortunately, however, given the special nature of the Jacobian matrix of equation 13. coupled with an assumption that the component transfer function matrices $Z_i(s,r)$ are rational in r, it is possible to show that the rank of the Jacobian matrix is "almost constant." This,

in turn, allows us to transform the local measure of solvability of equation 12. into a global measure of solvability. For this purpose we adopt the algebraic geometric definition for the term "almost constant." I.e. we say that a function of r_f is almost constant if it is constant except possibly for those values of r_f lying in an algebraic variety (the solution space of a finite set of non-zero simultaneous polynomial equations in n variables). More generally, we say that a property holds "almost everywhere" (a.e.) or for almost all r_f in n-space if it is true for all values of r_f except possibly those lying in an algebraic variety. Since the Lebesque measure of an algebraic variety is zero, this definition for the concept "almost everywhere" is consistent with the more common measure theoretic definition and is more natural in the context of our application.[14]

Theorem 1: Let $Z_i(s,r)$; $i = 1, 2, \ldots, q$; be rational in r. Then $\delta(r_f)$ is almost constant.

Note, the assumption that $Z_i(s,r)$ is rational in r is quite minor being satisfied by all of the examples given in section II except for the delay (which can be approximated by a function which is rational in r). In practice, the component transfer function matrices will also be rational in s though this is not required for the present theorem since F and dF/dr are formulated in terms of specific test frequencies, $s_1, s_2, \ldots, s_k$. Given our assumption on the $Z_i(s,r)$, together with equation 13., it then follows that $\frac{dF}{dr}(r_f)$ is also rational in r_f.

Proof of Theorem 1: We begin by showing that an arbitrary polynomial matrix in r, P(r), has almost constant rank. Since rank P(r) is restricted to the finite set of integers (0, 1, 2, ..., j; where j is the minimum of the number of rows and columns in P(r)), there exists an r_m which maximizes the rank of P(r)

$$14. \qquad \text{rank}[P(r_m)] \geq \text{rank}[P(r)]$$

Now, the rank of a matrix is the dimension of its largest non-singular square sub-matrix. As such, P(r) admits a square sub-matrix, M(r), whose dimension is

equal to the rank $P(r_m)$ and for which

15. $$\det M(r_m) \neq 0.$$

Now, $\det[M(r)]$ is a polynomial in r which is not identically zero (from equation 15.) and hence, it is non-zero a.e. As such,

16. $$\text{rank}[P(r_m)] \geq \text{rank}[P(r)] \geq \text{rank}[M(r)] = \text{rank}[P(r_m)] \text{ a.e.}$$

showing that $\text{rank}[P(r)] = \text{rank}[P(r_m)]$ almost everywhere. As such, $\text{rank}[P(r)]$ is almost constant.

Now, to verify that rank $[\frac{dF}{dr}(r_f)]$ is constant we decompose this matrix as

17. $$\frac{dF}{dr}(r_f) = \frac{P(r_f)}{d(r_f)}$$

where $P(r_f)$ is a polynomial matrix and $d(r_f)$ is a non-zero common denominator. $P(r_f)$ has almost constant rank while $d(r_f)$ is non-zero almost everywhere and hence can effect the rank of $P(r_f)$ only on an algebraic variety (since the division of a matrix by a non-zero scalar does not effect its rank.) As such, our Jacobian matrix has almost constant rank implying that

18. $$\delta(r_f) = n - \text{rank}\,[\frac{dF}{dr}(r_f)]$$

is also almost constant. The proof of the Theorem is therefore complete.

Given the theorem, we may now define a <u>global measure of solvability</u> for the fault diagnosis equation, δ, as the <u>generic value</u> of $\delta(r_f)$. I.e. the value $\delta(r_f)$ takes on for almost all r_f. This proves to be a natural measure of solvability since it indicates the ambiguity which will result from an attempt to solve the fault diagnosis equations in a neighborhood of almost any failures. Of course, one requires some sort of equation solving algorithm[10,11] to locate a neighborhood of an actual failure. The δ parameter, however, represents a bound on the

performance of any such algorithm. Finally, we note that since δ is independent of r_f, the solution of the fault diagnosis equations, it can be computed at the time the system and its test algorithm are developed by evaluating $\delta(r)$ at a randomly chosen generic point, say r_o. In turn, this parameter may then be employed as an aid in the choice of test frequencies and test points.

TEST FREQUENCY SELECTION

Adopting the measure of solvability, δ, formulated in the preceeding section, it remains to develop an algorithm for choosing a set of test frequencies; $s_1, s_2, \ldots s_k$; which maximize the solvability of the fault diagnosis equations (i.e. minimize δ). To this end, let δ_{min} denote the minimum value achieved by δ for any set of test frequencies; $s_1, s_2, \ldots, s_k$; $k = 1, 2, \ldots, .$ Since the possible values for δ are restricted to the finite set; $\delta = 0, 1, \ldots, n$; such a minimum is assured to exist.

The following theorem gives an explicit formula for computing δ_{min} while its proof yields an algorithm for choosing a set of test points which achieve δ_{min}. Since the purpose of this theorem is to formulate an algorithm for choosing test frequencies, the theorem is expressed in terms of

$$19. \qquad \mathrm{vec}[S(s,r)] = \mathrm{vec}[L_{22}] + [L_{12}^t \otimes L_{21}(1-Z(s,r)L_{11})^{-1}]\,\mathrm{vec}[Z(s,r)]$$

and

$$20. \qquad \frac{d\mathrm{vec}[S(s,r)]}{dr} = \{[(1 + L_{11}(1-Z(s,r)L_{11})^{-1}Z(s,r)]L_{12})^t \otimes (L_{21}(1-Z(s,r)L_{11})^{-1})\} \cdot \{d\mathrm{vec}[Z(s,r)]/dr\}$$

viewed as rational functions in s rather than in terms of the function $F(r)$ which is formulated in terms of an a-priori choice of test frequencies.

Theorem 2: Let $Z_i(s,r)$; $i = 1, 2, \ldots, q$; be rational in s and r. Then

$$\delta_{min} = n - \text{col-rank}\left[\frac{d\mathrm{vec}[S(s,r)]}{dr}\right]$$

where n is the dimension of the parameter vector, r, and "col-rank" denotes the generic number of linearly independent columns of the rational matrix [dev[S(s,r)]/dr] over the field of complex numbers. Moreover, δ_{min} is achieved by almost any choice of $n-\delta_{min}$ distinct complex frequencies.

Proof: For the sake of brevity, we will prove the theorem only for the special case where S(s,r) is a scalar transfer function (allowing us to drop the "vec" transformation) though essentially the same proof goes through in the general case modulo some notational complexities.[5] Also, since the rank of the Jacobian matrix is almost constant it sufficies to fix the parameter vector, r, at any generic point, say r_o. This then reduces [dvec[S(s,r)]/dr] to a row vector of rational functions

21. $$R(s) = [R_1(s) \quad R_2(s) \quad \dots \quad R_n(s)]$$

where

22. $$R_i(s) = [\text{dvec}[S(s,r_o)]/dr_i]$$

and our problem reduces to the verifications of the fact that the number of linearly independent columns of R(s) over the field of complex scalars is equal to the maximum possible rank of the complex matrix

23. $$\begin{bmatrix} R(s_1) \\ \hline R(s_2) \\ \hline \vdots \\ \hline R(s_k) \end{bmatrix} = \left[\begin{array}{c|c|c|c} R_1(s_1) & R_2(s_1) & \dots & R_n(s_1) \\ \hline R_1(s_2) & R_2(s_2) & & R_n(s_2) \\ \hline \vdots & \vdots & & \vdots \\ \hline R_1(s_k) & R_2(s_k) & \dots & R_n(s_k) \end{array}\right] = \text{col}\,(R(s_i))$$

over all possible choices of the complex frequencies; $s_1, s_2, \dots s_k$; $k = 1, 2, \dots$.

Now, clearly if some column of R(s), say the nth, is dependent on the remaining columns, then

$$24. \qquad R_n(s) = \sum_{j=1}^{n-1} c_j R_j(s)$$

for all s. Then by applying 24. individually for each s_i

$$25. \qquad \text{col}\,(R_n(s_i)) = \sum_{i=1}^{n-1} c_j \text{col}(R_j(s_i))$$

for any possible number or choice of the s_i. The rank of the matrix of Equation 23 is therefore less than or equal to the number of linearly independent columns of R(s) over the field of complex numbers.

To prove that equality can be achieved with an appropriate choice of $n-\delta_{min}$ complex test frequencies, s_i, we invoke our assumption that S(s,r) is a scalar transfer function. Without loss of generality, we may assume that $R_1(s)$ through $R_q(s)$ are the linearly independent entries in R(s) over the field of complex numbers in which case we must show that there exists complex frequencies $s_1, s_2, \ldots, s_k$ (k = q in this case) which make the first q columns of the matrix of equation 23. linearly independent.

If q = 1, $R_1(s)$ is not identically zero (since otherwise it would be linearly dependent) and hence for almost all s_1, $R_1(s_1) \neq 0$. As such, the columns in this trivial one by one matrix are linearly independent. With this as a starting point, we will use an inductive argument to show that the theorem holds for all values of q. We, therefore, assume that it has been shown that for q = p there exist complex frequencies; $s_1, s_2, \ldots, s_p$; such that the matrix

$$26. \qquad R_p = \begin{bmatrix} R_1(s_1) & R_2(s_1) & \cdots & R_p(s_1) \\ R_1(s_2) & R_2(s_2) & \cdots & R_p(s_2) \\ \vdots & & & \\ R_1(s_p) & R_2(s_p) & \cdots & R_p(s_p) \end{bmatrix}$$

has linearly independent columns and we desire to show that there exists an s_{p+1} such that the matrix

27.
$$\underline{R}_{p+1}(s) = \begin{bmatrix} R_1(s_1) & R_2(s_1) & \cdots & R_p(s_1) & R_{p+1}(s_1) \\ R_1(s_2) & R_2(s_2) & \cdots & R_p(s_2) & R_{p+1}(s_2) \\ & \vdots & & & \\ R_1(s_p) & R_2(s_p) & \cdots & R_p(s_p) & R_{p+1}(s_p) \\ R_1(s) & R_2(s) & & R_p(s) & R_{p+1}(s) \end{bmatrix}$$

has linearly independent columns for $s = s_{p+1}$. By virture of our assumption that $S(s,r)$ a scalar both $\underline{R}_p$ and $\underline{R}_{p+1}(s)$ are square and we may test for linear independence of the columns of $\underline{R}_{p+1}(s)$ by computing its determinent. Expanding 27. in cofactors along its bottom row, we obtain

28.
$$\det(\underline{R}_{p+1}(s)) = \sum_{j=1}^{p+1} (-1)^{p+j+1} \Delta_{p+1,\,j} R_j(s)$$

Since $\underline{R}_p$ has linearly independent columns $\Delta_{p+1,p+1} \neq 0$, hence, the coefficiencts in the summation of equation 28. are not all zero and thus by the linear independence of the $R_i(s)$ the summation is not identically zero. As such, one can choose almost any s_{p+1} which will make the determinant of $\underline{R}_{p+1}(s_{p+1})$ non-zero thus assuring the $\underline{R}_{p+1}$ has linearly independent columns when its rows are evaluated at the complex frequencies $s_1, s_2, \ldots, s_{p+1}$. The proof of the theorem is thus complete.

Note that the proof of the theorem yields a natural sequential algorithm for choosing test frequencies. Moreover, for the scalar case we have shown that the number of required test frequencies is exactly $n-\delta_{min}$ (equal to the column rank of the Jacobian matrix). In the general case where $S(s,r)$ is not a scalar, the number of required test frequencies is less than or equal to $n-\delta_{min}$.[5]

Although the theorem implies that one can randomly choose almost any $n-\delta_{min}$ test frequencies to maximize the solvability of the fault diagnosis equations, the result does not take cognizence of numerical considerations. Although no theory yet exists for choosing test points with numerical considerations in mind, it has been our experience that the "well posedness" of the fault diagnosis equations is quite sensitive to the choice of test frequencies.[5] In most of our experiments, we have worked with real test frequencies to eliminate the necessity of working in the complex plane. On the other hand, m is most easily measured when values of s_i on the jw axis are employed whereas it has been suggested that test frequencies symetrically spaced around a circle in the complex plane might yield numerically "well posed" equations.

Although the measure of solvability, δ, for the fault diagnosis equations is dependent on the choice of test frequencies, as well as the properties of the unit under test, δ_{min} is determined entirely by the UUT; its components, connections and accessible test points; and is completely independent of the test algorithm employed. As such, δ_{min} may be taken as a natural <u>measure of testability</u>[1] for the UUT which characterizes the degree to which the fault analysis equations can be solved given an optimal choice of test frequencies and solution algorithm. Moreover, δ_{min} may be used as an aid for the optimal selection of test points.[3,4,5] To this end we may choose a set of test points, from several options, so as to minimize δ_{min}. Alternatively, we may attribute a cost to each input and output test point and then choose the least cost combination of test points which yield a specified δ_{min}. This latter process reduces to a rather straighforward integer programming problem and is thus readily automated.[4,5] The technique is illustrated in the examples of the following section.

EXAMPLES

An an initial illustration of the theory consider the RC coupled amplifier with inductive load shown in Figure 2. Here we will take E_i to be the only

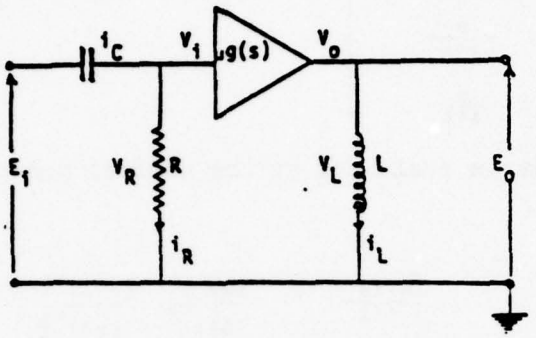


Figure 2: RC coupled amplifier with inductive load.

test input but we will initially allow E_o, i_L, i_C, and V_i to all be taken as test outputs with the measure of testability, δ_{min}, being used to extract a reduced set of test outputs from these options. A component connection model for this circuit is given by

28.
$$\begin{bmatrix} v_o \\ i_L \\ v_C \\ i_R \end{bmatrix} = \begin{bmatrix} \mu g(s) & 0 & 0 & 0 \\ 0 & 1/LS & & \\ 0 & 0 & 1/CS & \\ 0 & 0 & 0 & 1/R \end{bmatrix} \begin{bmatrix} V_i \\ V_L \\ i_C \\ V_R \end{bmatrix}$$

and

29.
$$\begin{bmatrix} V_i \\ V_L \\ i_C \\ V_R \\ -- \\ E_o \\ i_o \\ i_R \\ V_i \end{bmatrix} = \left[\begin{array}{cccc:c} 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{array}\right] \begin{bmatrix} V_o \\ i_L \\ V_C \\ i_R \\ -- \\ E_i \end{bmatrix}$$

Taking our vector of potentially variable component parameters to be $r = \mathrm{col}(\mu, L, C, R)$ each with unity nominal value, we obtain a nominal transfer function matrix

30. $$S(s,r) = \begin{bmatrix} \frac{s(q(s)+1)+1}{s+1} \\ \frac{q(s)}{s+1} \\ \frac{s}{s+1} \\ \frac{s}{s+1} \end{bmatrix}$$

whereas our Jacobain matrix evaluated at the nominal parameter values is given by

31. $$\frac{d\mathrm{vec}[S(s,r)]}{dr} = \begin{bmatrix} \frac{sq(s)}{s+1} & 0 & \frac{sq(s)}{(s+1)^2} & \frac{sq(s)}{(s+1)^2} \\ \frac{q(s)}{s+1} & \frac{-q(s)}{s+1} & \frac{q(s)}{(s+1)^2} & \frac{q(s)}{(s+1)^2} \\ 0 & 0 & \frac{s}{(s+1)^2} & \frac{-s^2}{(s+1)^2} \\ 0 & 0 & \frac{s}{(s+1)^2} & \frac{s}{(s+1)^2} \end{bmatrix}$$

Now, an inspection of this matrix will reveal that it has four independent columns over the field of complex numbers and hence if all four possible outputs are used, we will have $\delta_{min} = 0$ implying that the fault diagnosis equations have locally unique solutions. On the other hand, if only two outputs, E_o and i_C, are measured, our modified Jacobian matrix will reduce to the first and third rows of the matrix shown in equation 31. which has column rank 3. As such, if we only use these two test outputs, we obtain $\delta_{min} = 1$ and hence the solution to the fault diagnosis equations will be characterized by a single arbitrary parameter.

In this latter case, with only E_o and i_C taken as test outputs, theorem 2 implies that dF/dr will have rank 3 for almost any choice of $3 = n - \delta_{min}$ test frequencies. Chocsing $s_1 = 1$, $s_2 = 2$, and $s_3 = 3$, we obtain

32.

$$\frac{dF}{dr}(r_o) = \begin{bmatrix} g(1)/2 & 0 & g(1)/4 & g(1)/4 \\ 0 & 0 & 1/4 & -1/4 \\ \hline 2g(2)/3 & 0 & 2g(2)/9 & 2g(2)/9 \\ 0 & 0 & 2/9 & -2/9 \\ \hline 3g(3)/4 & 0 & 3g(3)/16 & 3g(3)/16 \\ 0 & 0 & 3/16 & -3/16 \end{bmatrix}$$

which has three linearly independent columns as long as $g(1) \neq 0$, $g(2) \neq 0$ and $g(3) \neq 0$. Indeed, in this example, any two of the three frequencies would have sufficed to yield three linearly independent columns. Note, for scalar transfer functions, theorem 2 implies that $n-\delta_{min}$ frequencies are actually required but for matrix transfer functions fewer frequencies may suffice.

Of course, for the circuit of Figure 2, we have a choice of some 15 combinations of the four outputs with which we may choose to work for the diagnosis of the circuit. The resultant δ_{min}'s for the various combinations of outputs are given in table 1.[5]

Finally, with the aid of Table 1, one may readily develop a test point selection algorithm for our circuit.[4,5] For instance, if we desire to find the smallest set of outputs which yield a $\delta_{min} \leq 1$ an inspection of the table will reveal that E_o and i_L, i_L and i_C, or E_o and i_C are the optimal choices. Of course, if one attributes a cost to the various outputs (determined by the convenience of making the required measurements), then we may further distinguish between these three possibilities. For instance, if voltage measurements are deemed to be easier than current measurements, the combination of i_L and i_C may be excluded with the decision between the remaining two options being dependent on whether it is easier to measure the circuit's input current (i_C) or its load current (i_L).

Outputs	δ_{min}
E_o, i_L, i_C, V_i	0
E_o, i_L, i_C	0
i_L, i_C, V_i	1
i_L, V_i, E_o	1
V_i, E_o, i_C	1
E_o, i_L	1
i_L, i_C	1
i_L, V_i	2
V_i, E_o	2
E_o, i_C	1
i_C, V_i	2
E_o	2
i_L	2
i_C	2
V_i	3

Table 1: Measure of testability for the circuit of Figure 2 using various combinations of test outputs.

As a second example, consider the one stage transistor amplifier shown in Figure 3 with the AC equivalent circuit of Figure 4. Since it is clearly impossible to distinguish between failures in the two parallel bias resistors, R_a and R_b, these two resistors have been combined into the single resistor, R'_a in the component connection model of equations 33. and 34. Taking all of the component parameters as potentiall faulty, r becomes a 12 vector composed of C_1, r_x, ..., R_L and as before, we take all parameters to have the nominal value of unity.

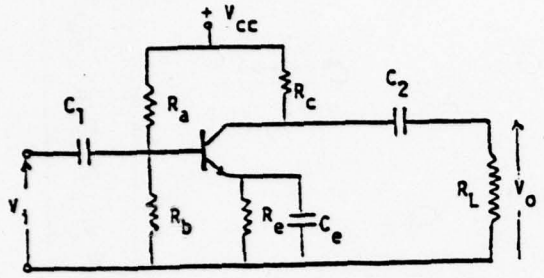


Figure 3. One Stage Transistor Amplifier

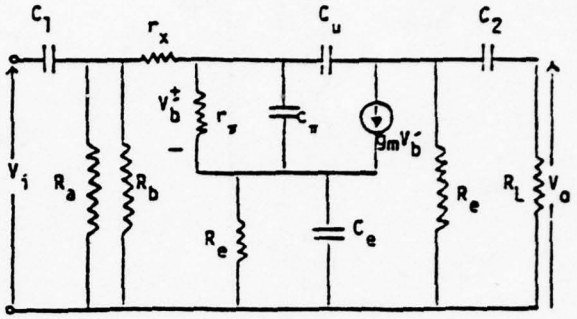


Figure 4. Amplifier Equivalent Circuit

33.

$$
\begin{bmatrix} I_{C_1} \\ I_{r_x} \\ I_{r_\pi} \\ I_{C_u} \\ I_{C_2} \\ V_{R_e'} \\ V_{R_e} \\ V_{C_\pi} \\ V_{C_e} \\ V_{g_m} \\ V_{R_c} \\ V_{R_l} \\ \hline V_o \\ I_{C_1} \\ V_{R_a'} \\ I_e \end{bmatrix} =
\left[\begin{array}{cccccccccccc|c}
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
-1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
-1 & -1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\hline
-1 & -1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0
\end{array}\right]
\begin{bmatrix} V_{C_1} \\ V_{r_x} \\ V_{r_\pi} \\ V_{C_u} \\ V_{C_2} \\ I_{R_a'} \\ I_{R_e} \\ I_{C_\pi} \\ I_{C_e} \\ I_{g_m} \\ I_{R_c} \\ I_{R_l} \\ \hline V_i \end{bmatrix}
$$

Once again we let the input voltage be the only test input for the system and we take V_o, I_{C_1}, V_{R_a}, and I_e, to be possible output test points. The resultant δ_{min} for each of the 15 possible combinations of these output terminals is tabulated in in Table 2.[5]

34.

$$
\begin{bmatrix} V_{C_1} \\ V_{r_x} \\ V_{r_\pi} \\ V_{C_\mu} \\ V_{C_2} \\ I_{R'_a} \\ I_{R_e} \\ I_{C_\pi} \\ I_{C_e} \\ I_{g_m} \\ I_{R_c} \\ I_{R_L} \end{bmatrix} = \begin{bmatrix} 1/C_1 s & & & & & & & & & & & \\ & r_x & & & & & & & & & & \\ & & r_\pi & & & & & & & & & \\ & & & 1/C_\mu s & & & & & & & & \\ & & & & 1/C_2 s & & & & & & & \\ & & & & & 1/R'_a & & & & & & \\ & & & & & & 1/R_e & & & & & \\ & & & & & & & C_\pi s & & & & \\ & & & & & & & & C_e s & & & \\ & & 0 & & & & & & & g_m & 0 & \\ & & & & & & & & & & 1/R_c & \\ & & & & & & & & & & & 1/R_L \end{bmatrix} \begin{bmatrix} I_{C_1} \\ I_{r_x} \\ I_{r_\pi} \\ I_{C_\mu} \\ I_{C_2} \\ V_{R'_a} \\ V_{R_e} \\ V_{C_\pi} \\ V_{C_e} \\ V_{g_m} \\ V_{R_c} \\ V_{R_L} \end{bmatrix}
$$

Outputs	δ_{min}
V_o	3
I_{C_1}	2
$V_{R'_a}$	2
I_e	3
V_o, I_{C_1}	0
V_o, $V_{R'_a}$	1
V_o, I_e	0
I_{C_1}, $V_{R'_a}$	2
I_{C_1}, I_e	1
$V_{R'_a}$, I_e	0
V_o, I_{C_1}, $V_{R'_a}$	0
V_o, I_{C_1}, I_e	0
V_o, $V_{R'_a}$, I_e	0
I_{C_1}, $V_{R'_a}$, I_e	0
V_o, I_{C_1}, $V_{R'_a}$, I_e	0

Table 2: Measure of Testability for the circuit of Figure 3 using various test outputs.

From the table it is apparent that no single test output suffices to yield a $\delta_{min} = 0$ (perfect testability) though $\delta_{min} = 0$ can be achieved using two test outputs; V_o and I_{C_1} or V_o and I_e.

CONCLUSIONS

Our purpose in the preceeding has been to formulate an analytic theory in support of the intuitive art usually associated with the design of a test algorithm. With the aid of the techniques developed above, we believe that it will be possible to develop an automated test program generation (ATPG) algorithm for linear systems.[4,5] Indeed, such an algorithm could be readily combined with the same computer-aided design (CAD) algorithm used in the system design process.[9] Given the component connection equations such an algorithm could be employed to automatically (or interactively) choose test points and test frequencies and generate the required set of fault diagnosis equations. These could then be stored on tape and supplied to the automatic test equipment (ATE) in which a faulty system would be tested and the fault diagnosis equations solved.

Although we do not propose to discuss the actual solution of the fault diagnosis equations here, it should be pointed out that by assuming that relatively few components have failed, say $p << n$, it is possible to develop specialized algorithms for the solution of the fault diagnosis equations which are far more efficient than standard equation solvers in this application.[7,11,12] These are typically derived from the fault simulation algorithms used in the diagnosis of digital systems and may naturally be classified into "simulation before test" and "simulation after test" algorithms. Some of the algorithms are discussed in references 7, 9, 10 and 11.

Finally, we note that as formulated above, the measure of testability, δ_{min}, assumes that any combination component failues is possible. If, however, we assume that at most $p << n$ components fail simultaneously, the ambiguity in the

solution of the fault diagnosis equations may actually be less than δ_{min}. For instance, in the example of Figure 3, with only V_o taken as an output $\delta_{min} = 3$, yet the fault diagnosis equations can be solved exactly if we assume that only one parameter is out of tolerence.[10] The point, here, is that even though the solution of the fault diagnosis equations in n-space has three arbitrary parameters when the solution is restricted to the one dimensional manifold of parameter vectors in which all but one coordinant are nominal it is unique.

REFERENCES

1. Dejka, W.J., "Measure of Testability in Device and System Design," Proc. of the 20th Midwest Symp. on Circuits and Systems, Lubbock, Texas, August, 1977, pp. 39-52.
2. Dejka, W.J., "A Review of Measurements of Testability for Analog Systems," Proc. of the 1977 AUTOTESTCON, Hyannis, Mass., November 1977, pp. 279-284.
3. Sen, N. and R. Saeks, "A Measure of Testability and its Application to Test Point Selection - Theory," Proc. of the 20th Midwest Symp. on Circuits and Systems, Lubbock, Tx., August 1977, pp. 576-583.
4. Sen, N. and R. Saeks, "A Measure of Testability and its Application to Test Point Selection - Computation," Proc. of the 1977 AUTOTESTCON, Hyannis, Mass., November 1977, pp. 212-219.
5. Sen, N., M.S. Thesis, Texas Tech Univ., Lubbock, TX, 1977.
6. Ransom, M.N., and R. Saeks, "A Functional Approach to Fault Analysis in Linear Systems," in Rational Fault Analysis, New York, Marcel Dekker, 1977, pp. 124-134.
7. Saeks, R. and R.A. DeCarlo, Interconnected Dynamical Systems, New York, Marcel Dekker, (to appear).
8. Singh, S.P. and R. -W. Liu, "Existence of State Equation Representation of Linear Large-Scale Dynamical Systems," IEEE Trans. on Circuits and Systems, Vol. CAS-20, pp. 239-246, (1973).
9. Proc. of the Workshop on Automatic Test Technology, NSIA, San Diego, April, 1978, (to appear).

10. Chen, H.M.S., and R. Saeks, "A Search Algorithm for the Solution of the Fault Diagnosis Equations," Unpublished Notes, Texas Tech Univ., 1978.

11. Chen, H.M.S., M.S. Thesis, Texas Tech Univ., 1977.

12. Ransom, M.N. and R. Saeks, "The Connection Function - Theory and Application," Int. J. on Circuit Theory and its Applications, Vol. 3, pp. 5-21, (1975).

13. Fleming, W., Functions of Several Variables, Addison- Wesley, Reading, 1965.

14. Spivak, M., Calculus on Manifolds, Benjamin, Amsterdam, 1966.

RESEARCH

on

LARGE-SCALE SYSTEMS

R. Saeks and K.S. Chao

DEPARTMENT OF ELECTRICAL ENGINEERING

TEXAS TECH UNIVERSITY

ABSTRACT

An algorithm for the inversion of a continuously parameterized family of sparse matricies is formulated in terms of a differential equation characterizing the evolution of the sparse L and U factors of the given family of matrices.

INTRODUCTION

In the various algorithms used for the analysis and design of large-scale circuits and systems, the problem of inverting a continuously parameterized family of sparse matrices, $M(r)$, is often encountered.[1-5] In frequency domain analysis, this might represent a transfer function matrix which one must invert over a specified frequency range[3] while in time domain analysis, such an $M(r)$ arises in the form of the Jacobian matrix for the system equations[1] which is dependent on some potentially variable parameter, r. Typically, one inverts $M(r)$ at a discrete set of points; r_i, $i = 1, 2, \ldots, n$; using a sparse matrix algorithm. Indeed, the more efficient algorithms exploit the fact that the matrices $M(r_i)$ have a common sparsity structure allowing much of the computational overhead to be shared by the n inversions.[1]

An alternative to repeated inversion is the continuations algorithm[5] wherein one integrates the differential equation

$$1. \qquad \dot{Z}(r) = -Z(r)(dM/dr)Z(r) \; ; \; Z(0) = M(0)^{-1}$$

to obtain $M(r)^{-1} = Z(r)$. While the integration of Equation 1 is far more efficient than repeated matrix inversion for small matrices, it fails to take advantage of the sparseness of $M(r)$, thereby rendering the technique inapplicable in a large-scale systems context. The purpose of the present note is to present an alternative continuation algorithm which combines the LU factorization technique of sparse matrix inversion with Equation 1.

LU FACTOR DYNAMICS

Recall the standard sparse matrix inversion technique[6] wherein one factors a matrix into the form $M = LU$ where L is lower triangular and U is upper triangular with ones along the diagonal. We then represent the inverse matrix in the form $M^{-1} = U^{-1}L^{-1}$. The key to the technique is that both L and U and their inverses will be sparse if M is sparse though, in general, M^{-1} is not sparse. As such, one may store and manipulate the inverse of a sparse matrix via its sparse upper and lower triangular factors, U^{-1} and L^{-1}, even though the inverse matrix, itself, is non-sparse. These ideas are combined with the continuation algorithm concept in the following theorem.[7] Here, the notation $^{u}[M]$ is used to denote the strictly upper triangular matrix obtained from M by setting all of the entries of M on or below the diagonal to zero. Similarly, $^{l}[M]$ denotes the lower triangular matrix obtained from M by setting all of the entries above the diagonal to zero.

THEOREM: Let X(r) and Y(r) be solutions of the matrix differential equation

$$\dot{X} = -X^{u}[Y(dM/dr)X] \quad ; \quad X(0) = U(0)^{-1}$$

2.

$$\dot{Y} = -^{l}[Y(dM/dr)X]Y \quad ; \quad Y(0) = L(0)^{-1}$$

Then, $X(r) = U(r)^{-1}$ and $Y(r) = L(r)^{-1}$ where $M(r)^{-1} = U(r)^{-1}L(r)^{-1}$ is the LU factored form of $M(r)^{-1}$. Note, if M(r) and dM/dr are sparse then every matrix involved in the integration of Equation 2 will be sparse. Moreover, the integration may be carried out with the aid only of a matrix multiplication algorithm plus a simple procedure for extracting the upper and lower triangular sub-matrices of Y(dM/dr)X.

Proof of the Theorem: First, we observe that if Y(0) is lower triangular, then $\dot{Y}$ will be lower triangular and so will Y(r) for all r. Similarly, if X(0) is upper triangular with ones on the diagonal, then $\dot{X}$, being the product of an upper triangular and strictly upper triangular matrix, will be strictly upper

triangular. As such, $X(r)$ will be upper triangular with ones on the diagonal for all r. Thus, $X(r)$ and $Y(r)$ have the correct form and it remains to verify the equality $M(r)^{-1} = X(r)Y(r)$. Here,

$$X(r)Y(r) = X(0)Y(0) + \int_0^r [X(q)\dot{Y}(q)]dq = X(0)Y(0) + \int_0^r [\dot{X}(q)Y(q) + X(q)\dot{Y}(q)]dq$$

$$= X(0)Y(0) + \int_0^r \{-X(q)^{u}[Y(q)(dM/dq)X(q)]Y(q) - X(q)^{l}[Y(q)(dM/dq)X(q)]Y(q)\}dq$$

$$3. \quad = X(0)Y(0) + \int_0^r \{-X(q)[Y(q)(dM/dq)X(q)]Y(q)\}dq$$

$$= X(0)Y(0) + \int_0^r [X(q)Y(q)](dM/dq)[X(q)Y(q)]dq$$

Differentiation of both sides of Equation 3 with respect to r then results in

$$4. \quad [X(r)\dot{Y}(r)] = [X(r)Y(r)](dM/dr)[X(r)Y(r)]$$

Finally, a comparison of Equations 4 and 1 reveals that $X(r)Y(r) = M(r)^{-1}$ since both $X(r)Y(r)$ and $M(r)^{-1}$ satisfy the same differential equation.

EXAMPLE

Consider the family of matrices

$$5. \quad M(r) = \begin{bmatrix} 1 & r \\ -1 & 1 \end{bmatrix}$$

Here, $M(0)$ is lower triangular and hence has the trivial LU-factorization

$$6. \quad M(0) = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = L(0)U(0)$$

while

$$7. \quad \dot{M}(r) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

As such, we have

8. $$L(0)^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

and

9. $$U(0)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Now, upon using an Euler integration formula [$Z(h) = Z(0) + h\dot{Z}(0)$] we may estimate $U(.1)^{-1}$ and $L(.1)^{-1}$ via the equalities

$$U(.1)^{-1} = U(0)^{-1} + (.1)U(\dot{0})^{-1}$$

$$= U(0)^{-1} - (.1)U(0)^{-1}\,u_{[L(0)^{-1}\dot{M}(0)U(0)^{-1}]}$$

10.
$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & .1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1/10 \\ 0 & 1 \end{bmatrix}$$

and

$$L(.1)^{-1} = L(0)^{-1} + (.1)L(\dot{0})^{-1}$$

$$= L(0)^{-1} - {}^{1}[L(0)^{-1}\dot{M}(0)U(0)^{-1}]L(0)^{-1}$$

11.
$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ .1 & .1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 9/10 & 9/10 \end{bmatrix}$$

Multiplying these estimates then yields

12. $$M(.1)^{-1} = U(.1)^{-1}L(.1)^{-1} = \begin{bmatrix} 91/100 & -9/100 \\ 9/10 & 9/10 \end{bmatrix}$$

which compares favorably with the exact inverse

13. $$M(.1)^{-1} = \begin{bmatrix} 10/11 & -1/11 \\ 10/11 & 10/11 \end{bmatrix}$$

The error here is due to the approximation inherent in the numerical integration process and can be reduced by use of a more accurate integration procedure. Of course, the result of the theorem is exact and the computed value for $M(r)^{-1}$ will be as accurate as the integration process employed.

REFERENCES

1. Hachtel, G., Brayton, R.K., and F. Gustavson, "The Sparse Tabeau Approach to Network Analysis and Design," IEEE Trans. on Circuit Theory, Vol. CT-18, pp. 101-113, (1971).

2. Chua, L.O. and P.M. Lin, Computer-Aided Analysis of Electronic Circuits, Prentice-Hall, Englewood Cliffs, 1975.

3. Erisman, A.M. and G. Spies, "Exploiting Problem Characteristic in Sparse Matrix Approach to Frequency Domain Analysis," IEEE Trans. on Circuit Theory, Vol. CT-19, pp. 260-264, (1972).

4. Saeks, R. and K.S. Chao, "Continuations Approach to Large-Change Sensitivity Analysis," IEE(London) Jour. on Electronic Circuits and Systems, Vol. 1, pp. 11-16, (1976).

5. Chao, K.S. and R. Saeks, "Continuation Methods in Circuit Analysis," IEEE Proc., Vol. 65, pp. 1187-1194, (1977).

6. Tewerson, R.P., Sparse Matrices, Academic Press, New York, 1972.

7. Saeks, R. and R.A. DeCarlo, Interconnected Dynamical Systems, Marcel Dekker, New York, (to appear).

RESEARCH

on

STOCHASTIC CONTROL AND ESTIMATION

R.B. ASHER

DEPARTMENT OF ELECTRICAL ENGINEERING

TEXAS TECH UNIVERSITY

* Presently with the Drincon Corp., San Diego, California

ABSTRACT

There has been considerable work dealing with the topic of filtering for problems with state dependent noise [1-3]. As well as being of theoretical interest, the topic is of some practical importance since many systems are better modeled as having multiplicative disturbances instead of additive. One example occurs in the momentum exchange method for regulating the angular procession of a rotating space craft [4]. There is a disturbance which depends on the procession rates. Another example occurs in the design of phase lock loops [2]. The phase instability of an oscillator described in rectangular coordinates appears as white, state dependent noise. If one received a signal which consisted of a large number of sinusoids of various frequencies, each having phase distortion, then one would have to build a high order filter to recover the signal using existing methods.

INTRODUCTION

The design of high order filters is often problematic from the viewpoint of on-line computation. Therefore, a number of researchers have been interested in designing filters of reduced order [5-8]. It often happens that one is only interested in estimating a lower order linear transformation of a state vector, and it seems reasonable to attempt to do this with a lower order filter. Design of the filter parameters is a fixed configuration optimization problem [8-10]. In such problems, the structure is not necessarily optimal, but given the structural constraints, the parameters are selected optimally. It is interesting to note that these problems often have non-unique solutions because there are too many free parameters. This feature can be used to obtain filters which are easier to implement than well-known techniques such as Kalman filtering,

even when the fixed configuration filter is of full order [8]. In some cases, there is no performance loss associated with the alternative linear filter, [8], [11].

In this paper we seek to extend the reduced order filtering results developed in [8] to problems with state dependent noise. The problem is similar to that considered in [12], however, in [12] a discrete system model was considered, and only a single stage/optimization was performed. Here a continuous time problem is considered, and the matrix minimum principle [13] is used to obtain a solution. Because we allow a driving term in the filter to remove any a-priori bias, it turns out that the problem has singular arcs, which is not surprising considering previous works [8], [11] in the area. A very nice feature of the work is that in some cases only linear two-point boundary value problems are obtained. These can be solved either by a direct use of linear systems theory or by a Riccati equation technique. Under certain circumstances only a single-point boundary-value problem must be solved.

PROBLEM STATEMENT

The system of interest is assumed to be modeled by the Ito stochastic differential equation

$$dx(t) = A(t)x(t)dt + dw(t) + \sum_{i=1}^{n} \left[x_i(t)-\mu_i(t)\right] G_i(t)d\nu(t) \qquad (1)$$

where $x(t)$ is the state vector of dimension n and $\mu(t)$ is the mean value of the state vector. The disturbances are zero mean incremental Wiener processes with covariances

$$\begin{aligned} E\{dw(t)dw^T(t)\} &= Q(t)dt \\ E\{d\nu(t)d\nu^T(t)\} &= \Xi(t)dt \end{aligned} \qquad (2)$$

It is not hard to show [14] that the mean value vector, μ satisfies

$$d\mu(t) = A(t)\mu(t)dt \tag{3}$$

The initial condition for (1) is random with known mean and variance

$$E\{x(t_o)\} = \mu_o \tag{4}$$

$$Var\{x(t_o)\} = P_o \tag{5}$$

Equation (4) is obviously the initial condition for (3).

The observation vector is also corrupted by state dependent noise.

$$dy(t) = C(t)\,x(t)\,dt + dv(t) + \sum_{i=1}^{n}\left[x_i(t) - \mu_i(t)\right] M_i(t)d\nu \tag{6}$$

In (6), $y(t)$ is the observation vector of dimension m, $dv(t)$ is the additive measurement disturbance, and $d\nu(t)$ is the multiplicative disdurbance. The vector $\nu(t)$ may be large, and some of its elements affect the dynamic model through the terms G_i, while others affect the observational model through the terms M_i. The additive disturbance, $dv(t)$ is a zero mean incremental Wiener process with covariance

$$E\{dv(t)dv^T(t)\} = R(t)dt \tag{7}$$

The terms $w(t)$, $v(t)$, $\nu(t)$ and $x(t_o)$ are uncorrelated.

Only a linear transformation of $x(t)$ is to be estimated, i.e., it is desired to estimate

$$z(t) = N(t)x(t) \tag{8}$$

where $z(t)$ is a vector of dimension $\ell \leq n$.

The estimate of $z(t)$, which we call $\hat{z}(t)$ is constrained to be obtained by the filter equation

$$d\hat{z}(t) = \left[F(t)\hat{z}(t) + g(t)\right] dt + K(t)dy \tag{9}$$

The vector $g(t)$ and the initial condition, $\hat{z}(t_o)$ are to be selected so that

$$E\{e(t)\} = 0 \quad \forall t = \left[t_o, t_f\right] \tag{10}$$

where e(t) is the error vector

$$e(t) = z(t) - \hat{z}(t) \tag{11}$$

The matrices F(t) and K(t) are then to be selected so that a quadratic performance measure

$$J = E\left\{ \int_{t_o}^{t_f} e^T(t)\, \bar{Q} e(t) dt + e^T(t_f) S e(t_f) \right\} \tag{12}$$

is minimized. The weighting matrix S is assumed to be positive definite symmetric. The weighting matrix $\bar{Q}$ may be positive definite or zero and is critically important to the solution.

GENERAL SOLUTION

In order to proceed, it is convenient to develop an equation for the error. From the Ito differential rule [15], it is seen that

$$dz(t) = N(t)dx(t) + \dot{N}(t)x(t)dt \tag{13}$$

Using (6), (9), and (13) it is seen that the differential equation of the error is

$$de(t) = dz(t) - d\hat{z}(t)$$

or

$$de = \left[(NA-FN-KC+\dot{N})\, x - g\right] dt + N\, dw - K\, dv + Fe\, dt + \left[N \sum_{i=1}^{n} \bar{x}_i G_i - K \sum_{i=1}^{n} \bar{x}_i M_i\right] dv \tag{14}$$

In (14) we have introduced the notation, $x = \bar{x} - \mu$. From (14) it is seen that

$$\frac{d}{dt} E\{e(t)\} = F(t) E\{e(t)\} \tag{15}$$

provided that

$$g(t) = (NA-FN-KC+\dot{N})\mu(t) \tag{16}$$

If furthermore

$$\hat{z}(t_o) = N(t_o)\mu(t_o) \tag{17}$$

it is clear that

$$E\{e(t_o)\} = 0 \tag{18}$$

From (15) and (18), one can see that (10) is satisfied so that (16) and (17) are appropriate selections. If g(t) is selected according to (16), the error differential equation can be written as

$$de = (NA-FN-KC+\dot{N})\bar{x}dt + Ndw-Kdv + Fedt + \sum_{i=1}^{n} \bar{x}_i (NG_i-KM_i)dv \tag{19}$$

The equation for $\bar{x}$ is

$$d\bar{x} = A\bar{x}dt + dw + \sum_{i=1}^{n} \bar{x}_i G_i dv \tag{20}$$

Clearly $\bar{x}$ and e are both zero mean processes.

If (19) and (20) are put in 1 equation, it is easy to see how the second moment matrix defined as

$$P(t) \triangleq \left[\begin{array}{c|c} P_{xx}(t) & P_{xe}(t) \\ \hline P_{ex}(t) & P_{ee}(t) \end{array}\right] \triangleq \left[\begin{array}{c|c} E\{\bar{x}(t)\bar{x}^T(t)\} & E\{\bar{x}(t)e^T(t)\} \\ \hline E\{e(t)\bar{x}^T(t)\} & E\{e(t)e^T(t)\} \end{array}\right] \tag{21}$$

propagates. This is useful since the performance measure (12) may be written as

$$J = tr\{\int_{t_o}^{t_f} \bar{Q}P_{ee}(t)dt + SP_{ee}(t_f)\} \tag{22}$$

If one has the appropriate constraint equation, the optimal selection of F(t) and K(t) may thus be solved with deterministic theory using the matrix minimum principle.

Equations (19) and (20) may be written as

$$\begin{bmatrix} d\bar{x}(t) \\ \hline de(t) \end{bmatrix} = \left[\begin{array}{c|c} A & 0 \\ \hline (NA-FN-KC+\dot{N}) & F \end{array}\right] \begin{bmatrix} \bar{x} \\ \hline e \end{bmatrix} dt + \begin{bmatrix} dw \\ \hline Ndw - Kdv \end{bmatrix}$$

$$+ \sum_{i=1}^{n} \bar{x}_i \Gamma_i \, d\nu \tag{23}$$

where

$$\Gamma_i \triangleq \begin{bmatrix} G_i \\ \hline NG_i - KM_i \end{bmatrix} \tag{24}$$

The second moment matrix associated with (23) satisfies [4],

$$\dot{P} = GP + PG^T + \hat{Q} + \hat{\psi} \tag{25}$$

where

$$G \triangleq \left[\begin{array}{c|c} A & 0 \\ \hline (NA-FN-KC+\dot{N}) & F \end{array}\right] \tag{26}$$

$$\hat{Q} \triangleq \left[\begin{array}{c|c} Q & QN^T \\ \hline NQ & KRK^T + NQN^T \end{array}\right] \quad : \tag{27}$$

and

$$\hat{\psi} = \sum_{\substack{i=1 \\ j=1}}^{n} P_{xx_{ij}} \, \Gamma_i \Xi \Gamma_j^T \tag{28}$$

Partitioning P in (25) we obtain the individual equations,

$$\dot{P}_{xx} = AP_{xx} + P_{xx}A^T + Q + \Psi_1 \tag{29}$$

$$\dot{P}_{ee} = \left[NA-FN-KC+\dot{N}\right] P_{xe} + P_{ex} \left[NA-FN-KC+\dot{N}\right]^T + FP_{ee} + P_{ee}F^T + NQN^T + KRK^T + K\Psi_3K^T - N\Psi_2K^T - K\Psi_2{}^TN^T + N\Psi_1N^T \tag{30}$$

and

$$\dot{P}_{xe} = AP_{xe} + P_{xx}\,(NA-FN-KC+\dot{N})^T + P_{xe}F^T + QN^T + \Psi_1N^T - \Psi_2K^T \tag{31}$$

In (29), (30), and (31), the terms Ψ_1, Ψ_2, and Ψ_3 are defined as

$$\Psi_1 = \sum_{\substack{i=1\\j=1}}^{n} P_{xx_{ij}}\, G_i \Xi G_j{}^T \tag{32}$$

$$\Psi_2 = \sum_{\substack{i=1\\j=1}}^{n} P_{xx_{ij}}\, G_i \Xi M_j{}^T \tag{33}$$

$$\Psi_3 = \sum_{\substack{i=1\\j=1}}^{n} P_{xx_{ij}}\, M_i \Xi M_j{}^T \tag{34}$$

The term P_{ex} is simply the transpose of P_{xe}. Clearly P_{xx} can be calculated independently, and can thus be regarded as a known quantity. The problem is to select K and F so that (22) is minimized subject to the constraints imposed by (30) and (31).

The Hamiltonian for this problem is then

$$H = tr\left\{\bar{Q}P_{ee} + \dot{P}_{ee}\Lambda^T{}_{ee} + \dot{P}_{xe}\Lambda^T{}_{xe} + \dot{P}_{ex}\Lambda^T{}_{ex}\right\} \tag{35}$$

where Λ_{ee}, Λ_{xe}, and Λ_{ex} are Lagrange multiplier matrices associated with P_{ee}, P_{xe}, and P_{ex} respectively. The constraint equation for P_{ex} is incuded for symmetry.

The optimal solution for the gain K(t) is obtained by setting the gradient of H with respect to K equal to zero. This leads to the expression for K.

$$K = \Lambda_{ee}^{-1}\left[\Lambda_{ee}\ (P_{ex}C^T + N\Psi_2) + \Lambda_{ex}\ (P_{xx}C^T + \Psi_2)\right]\left[R + \Psi_3\right]^{-1} \tag{36}$$

where the required inverses are assumed to exist. The Lagrange multiplier matrices satisfy the equations

$$\dot{\Lambda}_{ee} = -\frac{\partial H}{\partial P_{ee}} = -\left\{\bar{Q} + \Lambda_{ee}F + F^T\Lambda_{ee}\right\} \tag{37}$$

and

$$\dot{\Lambda}_{xe} = -\frac{\partial H}{\partial P_{xe}} = -\left\{(NA-FN-KC+\dot{N})^T\Lambda_{ee} + A^T\Lambda_{xe} + \Lambda_{xe}F\right\} \tag{38}$$

The matrix Λ_{ex} is just the transpose of Λ_{xe}. The initial conditions for (29), (30), and (31) are

$$P_{xx}(t_o) = \mathrm{Var}\left\{x(t_o)\right\} = P_o \tag{39}$$

and

$$P_{xe}(t_o) = P_o\,N(t_o)^T;\ P_{ee}(t_o) = N(t_o)\,P_o\,N(t_o)^T \tag{40}$$

The terminal values for (37) and (38) are as required by the transversality condition applied at the terminal time

$$\Lambda_{ee}\ (t_f) = S \tag{41}$$

and

$$\Lambda_{xe}\ (t_f) = 0 \tag{42}$$

Notice that $\Lambda_{ee}(t)$ can be computed separately without solving the rest of the problem if F is known beforehand. However at this point, we have not yet determined how F should be selected. It will be seen that this depends in a critical way on the nature of $\bar{Q}$. We will consider two different classes of problems.

CASE I.

In this case, we assume that $\bar{Q} = 0$. The meaning of this is that the quality of the estimation algorithm is only important at the terminal time. This may make sense for rather a large class of problems. The reason that this case is of particular interest is that the selection of F does not affect the Hamiltonian, so that we are free to select its value based on other considerations.

Consider that part of the Hamiltonian which depends explicitly on F.

$$H^* = \mathrm{tr}\left\{F\Theta + \Theta^T F^T\right\} \tag{43}$$

where

$$\Theta = (P_{ee} - NP_{xe})\Lambda_{ee} + (P_{ex} - NP_{xx})\Lambda_{xe} \tag{44}$$

From (39) and (40) it is clear that $\Theta(t_o) = 0$. If it can be shown that $\Theta(t) = 0$ for all t in the interval of interest, then a singular arc exists. The Hamiltonian is independent of F. In this case, one does not need to specify F to stay on the singular arc. Differentiating Θ gives

$$\dot{\Theta} = F\Theta - \Theta F + K\left[RK^T\Lambda_{ee} + \Psi_3 K^T \Lambda_{ee} - CP_{xe}\Lambda_{ee} - CP_{xx}\Lambda_{xe} - \Psi_2^T\Lambda_{ee} - \Psi_2^T N^T \Lambda_{ee}\right] \tag{45}$$

The bracketed term in the above is zero whenever K is chosen optimally, i.e., according to (36). Hence

$$\dot{\Theta}(t) = F(t)\Theta(t) - \Theta(t)F(t) \tag{46}$$

and (46) implies that $\Theta(t) = 0$ for all $t \geq t_o$ since $\Theta(t_o) = 0$. The selection of F is thus not a performance factor. It may be selected a-priori so that $\Lambda_{ee}(t)$ can be precomputed. It may be selected so as to achieve some other objective such as reduced sensitivity, computational convenience or to minimize some alternative performance measure specifically involving F.

When one thinks about it, the singularity with respect to F is not particularly surprising. Clearly two different filters can even produce the same output at a particular time, given the same input. What is interesting, is that this fact is generally overlooked, and as the example problem will show, that an alternative filter structure can be relatively easily implemented.

CASE II.

In this case the weighting matrix, $\bar{Q}$, is a positive definite symmetric matrix. When one develops an expression for $\dot{\Theta}$, the result is

$$\dot{\Theta} = F\Theta - \Theta F + \Omega\bar{Q} \tag{47}$$

instead of (46), where

$$\Omega = NP_{xe} - P_{ee} \tag{48}$$

Thus unless Ω is zero, a singular arc does not exist.

It is easily seen that $\Omega(t)$ does not equal zero unless F is slected appropriately. From the initial conditions, $\Omega(t_o) = 0$. Taking the time derivative of Ω we get

$$\dot{\Omega} = F\Omega+\Omega F^T + (NP_{xx}-P_{ex})\,(NA-FN-KC+\dot{N})^T - K\left[RK^T-\Psi_2{}^TN^T+\Psi_3K^T-CP_{xe}\right] \tag{49}$$

Examining the last equations we see that if

$$(NA-FN-KC+\dot{N}) = 0 \tag{50}$$

then

$$\dot{\Omega} = F\Omega+\Omega F^T \tag{51}$$

This follows from the fact that when (50) holds, $\Lambda_{xe}(t)$ is zero for all t in the interval. Consequently the expression for the gain becomes

$$K = \left[P_{ex}C^T + N\Psi_2\right]\left[R + \Psi_3\right]^{-1} \tag{52}$$

and (52) is sufficient to have the last term in (49) be zero. In view of (51) and the fact that $\Omega(t_o)$ is zero, it is clear that $\Omega(t)$ is zero for all $t\in\left[t_o,\ t_f\right]$ provided that (50) holds and that the gain is selected optimally.

When $\Omega(t)$ is zero, it may be seen that the orthogonality requirement is met in a reduced state space, i.e.

$$\Omega(t) = N(t)P_{xe}(t)-P_{ee}(t) = E\left\{\left[z(t)-e(t)\right]e^T(t)\right\} = 0 \tag{53}$$

Since $\hat{z} = z-e$, (53) may be written as

$$E\left\{\hat{z}(t)\ e^T(t)\right\} = 0 \tag{54}$$

so that what we have required for singularity is that the error and the estimate be orthogonal.

When N is the identity matrix and there is no state dependent noise, the result is the Kalman filter, with the requirement (50) that

$$F(t) = A(t)-K(t)C(t) \tag{55}$$

which of course means that the filter is of full order. When the filter is of reduced order, and N is constant, what we have is the observer constraint equation [16]

$$NA-FN-KC = 0 \tag{56}$$

In general, when $\bar{Q} > 0$, (50) is a necessary condition for a singular arc. Clearly it is not always possible to select F to satisfy (50). In such cases, the problem needs to be reformulated so that an unbounded F is not indicated. Alternatively a suboptimal solution can be accepted. We will examine this topic in the next section.

A necessary and sufficient condition that (50) have a solution F, is that

$$\left[NA-KC+\dot{N}\right] N^{\dagger}N = \left[NA-KC+\dot{N}\right] \forall t \in \{t_o, t_f\} \tag{57}$$

If (57) holds then a solution is

$$F = \left[NA-KC+\dot{N}\right] N^{\dagger} + \Gamma\left[I-NN^{\dagger}\right] \tag{58}$$

where $N^{\dagger}$ is the pseudo inverse of N and where Γ is an arbitrary $\ell x \ell$ matrix [17]. When the matrix (NN^T) is nonsingular then the solution (58) can be written as

$$F = \left[NA-KC+\dot{N}\right] N^T \left[NN^T\right]^{-1} \tag{59}$$

SPECIFIC F SOLUTIONS

In the preceding section we have shown that when one is only interested in estimation at a particular time, the selection of F may be based on considerations other than optimality, so that one may pick it prior to optimization. Furthermore, when $\bar{Q} > 0$, it may not be possible to find an F which results in a singular arc. In that case one may opt to select F prior to optimization. In this section, we will see that when F is selected a priori, the two point boundary value problem which must be solved for the selection of K is linear, and hence relatively easy to solve.

Consider substituting the gain expression (36) in (31) and (38). The resulting expressions are

$$\begin{aligned}\dot{P}_{xe} = {} & AP_{xe} + P_{xx}(\dot{N}+NA-FN)^T + P_{xe}F^T + QN^T + \Psi_1 N^T \\ & -(P_{xx}C^T+\Psi_2)(R+\Psi_3)^{-1}\left[(P_{xx}C^T+\Psi_2)^T \Lambda_{xe}\Lambda_{ee}^{-1}\right. \\ & \left. + \Psi_2^T N^T + C P_{xe}\right] \end{aligned} \tag{60}$$

and

$$\begin{aligned}\dot{\Lambda}_{xe} = {} & -(NA-FN+\dot{N})^T\Lambda_{ee} - A^T\Lambda_{xe} - \Lambda_{xe}F + C^T(R+\Psi_3)^{-1} \\ & \cdot\left[(P_{xx}C^T+\Psi_2)^T\Lambda_{xe} + (CP_{xe}+\Psi_2^T N^T)\Lambda_{ee}\right] \end{aligned} \tag{61}$$

When F is known a priori, both P_{xx} and Λ_{ee} are known in the sense that they may be precomputed. The above equations are then seen to give a linear TPBVP in the matrices P_{xe} and Λ_{xe}. The solution may be obtained in a straight forward manner using linear systems theory, or alternatively by assuming that the elements of Λ_{xe} are linearly related to those of P_{xe}, and obtaining a solution involving a Riccati equation. The values obtained for P_{xe} and Λ_{xe} may then be used in the gain expression (36).

We cannot overemphasize the importance of the fact that our result is a linear TPBVP, since it is reasonable to expect to solve a linear matrix TPBVP. Often a nonlinear matrix TPBVP is so difficult to solve, that the utility of the result is questionable. We shall explain procedures for solving a linear TPBVP by looking at a particularly easy case in which Λ_{ee} is a scalar times the identity matrix. This results when both F and $\bar{Q}$ are scalars times the identity matrix. When this is true, (60) may be written as

$$\dot{P}_{xe} = L_{11} P_{xe} + L_{12} \Lambda_{xe} + D_1 \tag{62}$$

where

$$L_{11} = A + F - L^* C \tag{63}$$

$$L_{12} = -L^* (C P_{xx} + \Psi_2^T) \Lambda^{-1}_{ee} \tag{64}$$

$$D_1 = P_{xx} (NA - FN + \dot{N})^T + Q N^T + \Psi_1 N^T - L^* \Psi_2^T N^T \tag{65}$$

and where

$$L^* = (P_{xx} C^T + \Psi_2) (R + \Psi_3)^{-1} \tag{66}$$

Equation (62) is of the form

$$\dot{\Lambda}_{xe} = L_{21} P_{xe} + L_{22} \Lambda_{xe} + D_2 \tag{67}$$

where

$$L_{21} = C^T (R + \Psi_3)^{-1} C \Lambda_{ee} \tag{68}$$

$$L_{22} = -A^T - F + C^T L^{*T} \tag{69}$$

and

$$D_2 = -(NA - FN + \dot{N})^T \Lambda_{ee} + C^T (R + \Psi_3)^{-1} \Psi_2^T N^T \Lambda_{ee} \tag{70}$$

Let L be the matrix

$$L = \left[\begin{array}{c|c} L_{11} & L_{12} \\ \hline L_{21} & L_{22} \end{array}\right] \tag{71}$$

and Φ be the associated state transition matrix which can similarly be partitioned

$$\Phi = \left[\begin{array}{c|c} \Phi_{11} & \Phi_{12} \\ \hline \Phi_{21} & \Phi_{22} \end{array}\right] \tag{72}$$

Then the solution to (62) is

$$P_{xe}(t) = \Phi_{11}(t, t_o)P_{xe}(t_o) + \Phi_{12}(t,t_o)\,\Lambda_{xe}(t_o) + \int_{t_o}^{t} \left[\Phi_{11}(t,\tau)\,D_1(\tau) + \Phi_{12}(t,\tau)\,D_2(\tau)\right] d\tau \tag{73}$$

and

$$\Lambda_{xe}(t) = \Phi_{21}(t,t_o)P_{xe}(t_o) + \Phi_{22}(t,t_o)\,\Lambda_{xe}(t_o) + \int_{t_o}^{t} \left[\Phi_{21}(t,\tau)\,D_1(\tau) + \Phi_{22}(t,\tau)D_2(\tau)\right] d\tau \tag{74}$$

Applying (74) at time $t = t_f$ gives

$$\Lambda_{xe}(t_f) = 0 = \Phi_{21}(t_f,t_o)P_{xe}(t_o) + \Phi_{22}(t_f,t_o)\,\Lambda_{xe}(t_o) + \int_{t_o}^{t_f} \left[\Phi_{21}(t_f,\tau)D_1(\tau) + \Phi_{22}(t_f,\tau)D_2(\tau)\right] d\tau \tag{75}$$

We can solve (75) for $\Lambda_{xe}(t_o)$ and substitute the results in (73) and (74) to obtain the solution for all $t \in [t_o, t_f]$.

There is another approach which is probably preferable in most cases. We assume that Λ_{xe} is linearly related to P_{xe} by the relationship

$$\Lambda_{xe}(t) = U(t)P_{xe}(t) + B(t) \tag{76}$$

Differenatiating (76), one obtains the differential equation

$$\dot{\Lambda}_{xe} = \dot{U}P_{xe} + U\left[L_{11}P_{xe} + D_1 + L_{12}UP_{xe} + L_{12}B\right] + \dot{B} \tag{77}$$

Alternatively, from (67)

$$\dot{\Lambda}_{xe} = L_{21}P_{xe} + L_{22}UP_{xe} + L_{22}B + D_2 \tag{78}$$

Equating (77) and (78), we get for U

$$\dot{U} + UL_{11} + UL_{12}U = L_{21} + L_{22}U \tag{79}$$

and for B

$$UL_{12}B + \dot{B} + UD_1 = L_{22}B + D_2 \tag{80}$$

Since $\Lambda_{xe}(t_f) = 0$, the terminal conditions for U and for B are

$$U(t_f) = 0 \tag{81}$$

$$B(t_f) = 0 \tag{82}$$

The Riccati equation (79) and equation (80) can be solved backwards in time from the above terminal conditions. The optimal gain may then be expressed as

$$K = \left[P_{xe}\bar{T}C^T + N\Psi_2 + \Lambda_{ee}^{-1}(UP_{xe}+B)^T(P_{xx}C^T+\Psi_2)\right]\left[R+\Psi_3\right]^{-1} \tag{83}$$

and P_{xe} is evaluated as

$$\dot{P}_{xe} = L_{11}P_{xe} + L_{12}\left[UP_{xe} + B\right] + D_1 \tag{84}$$

The matrices Λ_{ee}, P_{xx}, U, and B must be evaluated off line, however, P_{xe} and K can be evaluated on line if this is desired. Most likely these would also be evaluated off line and K stored for on line calculation of $\hat{z}(t)$ using (9).

EXAMPLES

The first example we shall consider is of the category discussed in Case II. We assume that A(t) is zero, N(t) = C(t), and that there exists an F such that

$$FC = \dot{C} - KC \quad \forall t \in \left[t_o, t_f\right] \tag{85}$$

then if CC^T is nonsingular

$$F = \left[\dot{C}C^T - KCC^T\right](CC^T)^{-1} \tag{86}$$

The filter equation is

$$d\hat{z} = \dot{C}C^T(CC^T)^{-1}\hat{z}dt + K\left[dy - \hat{z}dt\right] \tag{87}$$

The initial condition for (87) is

$$\hat{z}(t_o) = C(t_o)\,\mu_o \tag{88}$$

The gain is of the form

$$K(t) = \left[P_{xe}{}^T C^T + C\Psi_2\right]\left[R + \Psi_3\right]^{-1} \tag{89}$$

where P_{xe} is the solution to

$$\dot{P}_{xe} = P_{xe}(CC^T)^{-1}\left[\dot{C}C^T - CC^T K^T\right] + (Q+\Psi_1)\, C^T - \Psi_2\, K^T \tag{90}$$

Alternatively since $CP_{xe} = P_{ee}$, it may be desirable to evaluate (89) as

$$K(t) = \left[P_{ee} + C\Psi_2\right]\left[R + \Psi_3\right]^{-1} \tag{91}$$

where P_{ee} is the solution to

$$\begin{aligned}\dot{P}_{ee}(t) = &\left[\dot{C}C^T(CC^T)^{-1} - K\right] P_{ee} + P_{ee}\left[\dot{C}C^T(CC^T)^{-1} - K\right]^T \\ &+ CQC^T + KRK^T + K\Psi_3K^T - C\Psi_2K^T \\ &- K\Psi_2{}^TC^T + C\Psi_1C^T\end{aligned} \tag{92}$$

The reason (92) is appealing is that P_{ee} has fewer elements to calculate then P_{xe}.

The next example is concerned with the very simple problem of estimating a constant having zero mean and variance 1 prior to observations. The observation is of the form

$$dy = x\, dt + dv + Mx\, d\upsilon \tag{93}$$

where v and υ are zero mean white noise with covariance parameter 1 and M is constant. We are interested in estimating the value of x at time, T. Hence

$$J = E\left\{e^2\,(T)\right\} \tag{94}$$

and this is a problem of the category referred to as Case 1. For computational convenience we select F = 0. The TPBVP then is

$$\dot{P}_{xe} = -\gamma\,(P_{xe} + \Lambda_{xe}) \tag{95}$$

$$\dot{\Lambda}_{xe} = \gamma\,(P_{xe} + \Lambda_{xe}) \tag{96}$$

with $P_{xe}(0) = 1$ and $\Lambda_{xe}(T) = 0$, where

$$\gamma \triangleq (1 + M^2)^{-1} \tag{97}$$

The solution is

$$P_{xe}(t) = \frac{1 - \gamma(t-T)}{1 + \gamma T} \tag{98}$$

and

$$\Lambda_{xe}(t) = \frac{\gamma (t - T)}{1 + \gamma T} \tag{99}$$

Interestingly, because of the complimentary nature of $P_{xe}(t)$ and $\Lambda_{xe}(t)$, the gain is a constant,

$$k(t) = k = \frac{\gamma}{1} \cdot \frac{}{\gamma T} \tag{100}$$

The filter is simply

$$d\hat{x}(t) = \frac{\gamma}{1 + \gamma T}\, dy(t) \tag{101}$$

and the error variance at time $t = T$ is

$$P_{ee}(T) = \frac{1}{1 + \gamma T} = \frac{1 + M^2}{1 + M^2 + T} \tag{102}$$

The filter (101) is simpler to construct than the choice which would require $F = -K$, i.e., one of the form

$$d\hat{x}(t) = k(t) \left[dy(t) - \hat{x}(t)dt\right] \tag{103}$$

even though it is obviously a full order filter. The authors feel that the nonunique property of optimal linear filters for certain cases is a feature which one should take advantage of.

REMARKS AND CONCLUSIONS

We have extended the results of [8] to problems having state dependent noise in the observation and dynamical equation. Control theoretic methods have been used to solve the problem, and optimal solutions have been shown to correspond to singular arcs. Different solutions result when there is an integral performance measure than when only estimation at the terminal time is important. In some cases, we have seen that it makes sense to select the filter matrix ahead of time and then optimize the gain. The computational algorithms associated with such prior selection are particularly convenient. There are no terribly difficult TPBVP's in this approach and that is why the authors feel that it is practical and useful, both for full order and reduced order filters. The amount of off line calculation necessary to simplify on line filtering appears to be quite realistic.

REFERENCES

[1] McLane, P. J., "Optimal Linear Filtering for Linear Systems with State Dependent Noise," International Journal of Control, Vol. 10, No. 1, pp. 41-51, 1969.

[2] Gustafson, E. G., "On Optimal Estimation and Control of Linear Systems with State Dependent Noise," Ph.D. Thesis, MIT, June, 1972.

[3] Sims, C. S., "Linear Estimation in Discrete Random Parameter Systems," SWIEEECO Record of Technical Papers, April 4-6, 1973, Houston, Texas.

[4] McLane, P. J., "Optimal Stochastic Control of Linear Systems with State and Control Dependent Disturbances," IEEE Trans. Aut. Control, Vol. AC-16, No. 6, pp. 793-798.

[5] Hutchinson, C. E., and J. A. D'Appotito, "Minimum Variance Reduced State Filter," presented at 1972 IEEE Conference on Decision and Control, New Orleans, Dec. 1972.

[6] Damiani, T. R., "A General Partial State Estimator," presented at 14th Midwest Symp. Circuit Theory, Denver, Colorado, May, 1971.

[7] Sims, C. S., "An Algorithm for Estimating a Portion of a State Vector," IEEE Trans. Aut. Control, Vol. AC-19, No. 4, August, 1974.

[8] Sims, C. S., and R. B. Asher, "Optimal and Suboptimal Linear Filtering," Proceeding of 1977 JACC, San Francisco, California.

[9] Sims, C. S., and J. L. Melsa, "A Fixed Configuration Approach to the Stochastic Linear Regulator Problem," Proceedings 1970 Joint Automatic Control Conference, June 24-26, 1970, Atlanta, Georgia.

[10] Sims, C. S., "Fixed Configuration Control of Systems with State and Control Dependent Noise," Proceedings 1971 Joint Automatic Control Conference, August 11-13, St. Louis, Missouri.

[11] Lee, J. S., "On the Relation of Singular Control to Linear Estimation," Proceedings 1970 Joint Automatic Control Conference, June 24-26, 1970, Atlanta, Georgia.

[12] Sims, C. S., "Observer-Estimators for Discrete Random Parameter Systems," Proceedings of the Eleventh Annual Allerton Conference on Circuit and Systems Theory, Monticello, Illinois, October 3-5, 1973.

[13] Athans, M., "The Matrix Minimum Principle," Information and Control, Vol. 11, pp. 592-606, 1968.

[14] Cumming, J. G., "Derivation of Moments of a Continuous Stochastic System," Int. Journal Control, Vol. 5, No. 1, pp. 85-90.

[15] Jazwinski, A. H., Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

[16] Luenberger, D. G., "Observers for Multivariable Systems," IEEE Trans. Automatic Control., Vol. AC-11, pp. 190-197, April 1966.

[17] Boullion, T. L., and P. L. O'Dell, Generalized Inverse Matrices, Wiley-Interscience, New York, 1971, pp. 4-5.

RESEARCH

on

NONLINEAR CONTROL

R.L. Hunt

DEPARTMENT OF MATHEMATICS

TEXAS TECH UNIVERSITY

ABSTRACT

Consider the nonlinear system

$$\dot{x}(t) = f(x(t)) + \sum_{i=1}^{n-1} u_i(t) g_i(x(t)),\ x(0) = x_0 \in M,$$

where M is a connected C^∞ real n-dimensional manifold, $f, g_1, \ldots, g_{n-1}$ are C^∞ linearly independent vector fields on M, and $u_1, \ldots, u_{n-1}$ are real-valued controls. Since the integral manifolds of $g_1, \ldots, g_{n-1}$, if any, are real (n-1)-dimensional submanifolds of M, such a system is called a hypersurface system. Suppose U is the largest open subset of M which is reachable from x_0 and suppose $U \neq M$. It is shown that the boundary of U is a C^∞ real(n-1)-dimensional submanifold N of M and N is an integral manifold of the vector fields $g_1, \ldots, g_{n-1}$. Moreover the restriction of f to N must assign vectors pointing in the direction of U. Such a U is called the region of reachability for the system with initial point x_0. Many ideas here parallel those used in several complex variables for the studies of regions of holomorphy and of uniqueness of analytic continuation for the CR-functions on a C^∞ real hypersurface in $\mathbb{C}^n$, $n > 1$.

INTRODUCTION

Let M be a connected C^∞ real n-dimensional manifold,

$f, g_1,\ldots,g_{n-1}$ be C^∞ linearly independent vector fields on M, and $u_1,\ldots,u_{n-1}$ be real-valued controls. The system

$$\dot{x}(t) = f(x(t)) + \sum_{i=1}^{n-1} u_i(t)g_i(x(t)),\ x(0) = x_0 \in M,$$

is called a hypersurface system since the number of g vector fields is n-1. We know that the reachable set of this system contains an open set in M (see arguments in [8]), and we denote by U the largest open subset of M which is reachable from x_0. This set U is called the region of reachability from x_0, and if U = M, the system is controllable from x_0. If $U \neq M$ then we prove that the boundary of U is a C^∞ real(n-1)-dimensional submanifold N of M and N is an integral manifold of the vector fields $g_1,\ldots,g_{n-1}$. In addition, the restriction of the vector field f to N must point in the direction of U.

This article is arranged in the following way. In section 2 we give definitions and relevant examples. Section 3 contains a local theory concerning the boundary of U under the assumption that this boundary is C^1 near one of its points. In section 4 we state a theorem from [6] concerning a subbundle of the tangent bundle to M and allowing us to remove the C^1 restriction. Then we prove our main result, Theorem 4.2, and give several applications.

DEFINITIONS AND EXAMPLES

We shall use the classical Frobenius Theorem and Chow's Theorem [2]. For a statement of these results and their applications to control theory we refer the reader to [1].

Of interest to us is the controllability of the system

$$(2.1)\quad \dot{x}(t) = f(x(t)) + \sum_{i=1}^{n-1} u_i(t) g_i(x(t)),\ x(0) = x_0 \in M,$$

with M, f, $g_1,\ldots,g_{n-1}$, and $u_1,\ldots,u_{n-1}$ as in the introduction. We let T(M) denote the tangent bundle of M with fiber $T_x(M)$ for $x \in M$.

Recall that if X is a vector yield on M, then α is an integral curve of X if α is a C^∞ mapping from a closed interval $I \subset \mathbb{R}$ into M such that

$$\frac{d\alpha(t)}{dt} = X(\alpha(t)) \text{ for all } t \in I.$$

Definition 2.1 [8]. If D is a subset of T(M), then an integral curve of D is a mapping α from a real interval $[t,t']$ into M such that there exist $t = t_0 < t_1 < \ldots < t_k = t'$ and vector fields $X_1,\ldots,X_k$ in D with the restriction of α to $[t_{i-1},t_i]$ being an integral curve of X_i, for each $i = 1,2,\ldots,k$.

Definition 2.2. Let D be a subset of T(M) and let $x_0 \in M$. A point $x \in M$ is D-reachable from x_0 if there is an integral curve α of D and some $T \geq 0$ in the interval for α such that $\alpha(0) = x_0$ and $\alpha(T) = x$. A subset S of M is D-reachable from x_0 if every point $x \in S$ is reachable from x_0.

For the remainder of this article D is the subset of T(M) given by the vector fields $f(x(t)) + \sum_{i=1}^{n-1} u_i(t)g_i(x(t))$. Therefore we drop the D from D-reachable.

For complete vector fields (i.e. vector fields which can be defined for all $-\infty < t < \infty$) it can be seen from arguments in [8] (and not difficult to prove for the special case of hypersurface systems like (2.1) even if the vector fields may not be complete) that the reachable set from x_0 contains a nonempty open subset of M, containing x_0 in its closure.

Definition 2.3. The largest open subset U of M which is reachable from x_0 is called the region of reachability from x_0. If U = M, we say that the system is controllable from x_0.

We make two final comments before introducing some illustrative examples. Since we are considering unbounded controls (both positive and negative), no generality will be lost in assuming that we can move along the integral curves of $g_1,\ldots,g_{n-1}$. If we define the Lie bracket of two linearly independent vector fields f and g in $\mathbb{R}^2$ by $[g,f] = \frac{dg}{dx}f - \frac{\partial f}{\partial x}g$, then we find that [g,f] is a linear combination of f and g. This means that there are no "new" directions in which a solution to the system

$$\dot{x}(t) = f(x(t)) + u(t)g(x(t)) \ , \ x(0) = x_0 \in \mathbb{R}^2$$

can move (see [1]).

Example 1. Consider the linear system

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + u(t)\begin{bmatrix}0\\ 1\end{bmatrix} = f(x(t)) + u(t)g,$$

where M is the open right half plane in $\mathbb{R}^2$. By the well known test of Kalman [7] this system is not controllable for all of $\mathbb{R}^2$ from any point x_0 in $\mathbb{R}^2$ since $\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}$ is equal to $\begin{bmatrix}0\\1\end{bmatrix}$. Let $x_0 = (x_1^0, x_2^0)$ be an arbitrary point in M. The integral curve of g through (x_1^0, x_2^0) is given by the vertical line $x_1 = x_1^0$. Suppose U is the open set in M defined by $\{(x_1, x_2) \mid x_1 > x_1^0\}$. Since we can reach any point on $x_1 = x_1^0$, our only hope in escaping from the set U is that there is some point on the line $x_1 = x_1^0$ at which f is in the direction of the complement of U. However, $f(x(t)) = \begin{bmatrix}x_1^0\\ x_2\end{bmatrix}$ on this line and $x_1^0 > 0$, implying U contains the reachable set from x_0. Letting $u(t) = 0$, we reach every point on the line $x_2 = \frac{x_2^0}{x_1^0} x_1$ to the right of (x_1^0, x_2^0). Using infinite controls and the interval curves of g, we find that the region of reachability of our system from x_0 is the set U.

Example 2. Consider the linear system

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + u(t)\begin{bmatrix}0\\1\end{bmatrix} = f(x(t)) + u(t)g,$$

where M is the open upper half plane in $\mathbb{R}^2$. From Kalman [7] we see that this system is controllable for all of $\mathbb{R}^2$ from any point $x_0 \in \mathbb{R}^2$ since $\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix}$ is not a linear combination of $\begin{bmatrix}0\\1\end{bmatrix}$. Let $x_0 = (x_1^0, x_2^0)$ be any point in M. Again, the integral curve of g through (x_1^0, x_2^0) is given by the vertical line

$x_1 = x_1^0$, and we must restrict x_2 to be positive for points on this line in order to stay in M. Let U be the open set in M given by $\{(x_1,x_2) | x_1 > x_1^0, x_2 > 0\}$. The vectors that f assigns at each point of the line $x_1 = x_1^0$ in M are in the direction of U since $x_2^0 > 0$. This is also true for each vertical line in M to the right of $x_1 = x_1^0$. Hence the integral curve given by $f(x(t))$ with initial value (x_1^0,x_2^0) must move forever to the right. Those integral curves of g including and to the right of $x_1 = x_1^0$ intersect the integral curves of f transversally (defined in section 3) at each of its points. Thus the reachable set from x_0 in M is U. Since we cannot move outside of M, U is the region of reachability from x_0.

We remark that we are assuming f and g are linearly independent on M in our theory. Suppose that we relax this temporarily and let $M = \mathbb{R}^2$ in our Example 2. Starting at any point $x_0 = (x_1^0,x_2^0)$ in $\mathbb{R}^2$, we can always move vertically (both up and down) along $x_1 = x_1^0$, so we can assume $x_2^0 > 0$. An argument as in our previous discussion shows we can reach the set $\{(x_1,x_2) | x_1 \geq x_1^0\}$ since we no longer must remain in the open upper half plane. We take a new point (x_1',x_2') with $x_1' = x_1^0$ and $x_2' < 0$. Since the vectors f assigns along the line $x_1 = x_1^0$ are negative in the x_1 sense when (x_1^0,x_2)

is in the open lower half plane, we can also reach the region $\{(x_1,x_2) | x_1 \leq x_1^0, x_2 < 0\}$. Moving along the integral curves of g will give us the entire plane, as predicted by the Kalman theory.

Examples 1 and 2 suggest a possible solution to our problem for 2-dimensional linear systems. If U is the region of reachability from x_0 in M, then the boundary of U in M should be the integral curves of g on which the vectors given by f point in the direction of U.

Since we are most interested in nonlinear systems, we examine the following bilinear system.

Example 3. Consider the system

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + u(t) \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= f(x(t)) + u(t)g(x(t)),$$

where M is the set $\mathbb{R}^2-\{(0,0)\}$. Brockett [1] states that if $x_0 = (x_1^0, x_2^0)$ is in the positive quadrant, then the region of reachability U from x_0 is contained in this quadrant. The integral curve of g through a point on $x_1 = 0, x_2 > 0$ is the line $(0,x_2)$ with $x_2 > 0$. Moreover, the integral curve of g through a point on $x_2 = 0, x_1 > 0$ is the line $(x_1,0)$ with $x_1 > 0$. These together form the boundary of the first quadrant in M. The vector field f assigns vectors to this boundary which point toward the first quadrant. Thus there is no hope of a solution starting in this quadrant to leave it.

We could give many more examples at this time, but they would all hint at the same conclusion. In the system (2.1), the important items to check appear to be the integral manifolds of $g_1,\ldots,g_{n-1}$, if any exist, and the direction of the vector field f on these integral manifolds. We next examine these conditions for regions of reachability with C^1 boundaries.

C^1 BOUNDARIES

Let U be the region of reachability of the hypersurface system given in (2.1). Let x be an element of the boundary of U (denoted by ∂U), and assume ∂U is C^1 in some open neighborhood W of x in M. As just mentioned we need to consider the directions of f on $W \cap \partial U$ and the possibility of having an integral manifold of $g_1,\ldots,g_{n-1}$ through x. Recall that a differentiable submanifold S of M is an <u>integral manifold</u> of $g_1,\ldots,g_{n-1}$ if $T_y(S)$ is the space spanned by $g_1,\ldots,g_{n-1}$ at y for each $y \in S$.

For a more thorough discussion of integral manifolds we must consider the Lie bracket which we defined earlier for $\mathbb{R}^2$ (a very special case). Let g_i and g_j be different vector fields on M and define $[g_i,g_j] = \frac{\partial g_i}{\partial x} g_j - \frac{\partial g_j}{\partial x} g_i$. This may or may not give us a "new" direction in which to move (see [1]). Let L_A be the smallest Lie algebra generated by taking successive Lie brackets of the $g_1,\ldots,g_{n-1}$ given in equation (2.1). If we get a vector space of the same dimension at each point of M, then L_A is a fiber bundle with constant fiber dimension n-1

or n. Moreover, L_A is a fiber subbundle of the tangent bundle to M.

The following definition is essential to our work. Let S_1 and S_2 be C^1 submanifolds of M of dimensions k and n-k respectively.

Definition 3.1. The manifolds S_1 and S_2 intersect transversally at a point $y \in S_1 \cap S_2$ if and only if $T_y(S_1) \oplus T_y(S_2) = T_y(M)$. Here $\oplus$ denotes the direct sum.

We now prove a result under the assumption that locally our open set has a C^1 boundary.

Theorem 3.2. Let O be an open set in M which is reachable from x_0 for system (2.1), and let x be an arbitrary point in ∂O. Suppose there is an open neighborhood W of x in M such that $W \cap \partial O$ is a C^1 real(n-1)-dimensional submanifold of M. If any one of the following conditions holds, then O is not the region of reachability from x_0.

i) the fiber dimension of L_A at x is n,

ii) the integral curve of some g_i, $1 \leq i \leq n-1$, is transversal to ∂O at x,

iii) f assigns at x a vector pointing in the direction of the complement of O.

Proof. If i) holds then the fiber dimension of L_A at all points in some open neighborhood of x in M must be n (since n is maximal). Thus ∂O cannot be an integrable manifold of $g_1, \ldots, g_{n-1}$ near x by the Frobenius Theorem, and there exist a point $y \in \partial O$

arbitrarily close to x and a g_i, $1 \leq i \leq n-1$, such that the integral curve of g_i is transversal to ∂O at y. Hence, i) reduces to ii).

Next we assume that ii) is true. If the integral curve of g_1, chosen arbitrarily from the set $g_1,\ldots,g_{n-1}$, is transversal to ∂O at x, then it is transversal to ∂O in $W \cap \partial O$, W being an open neighborhood of x in M (this W may be a smaller set than our original W). Following the integral curves of g_1 that start in $W \cap O$, a reachable set from x_0, and continuing past $W \cap \partial O$, we have that O cannot be the region of reachability from x_0.

If iii) holds at x, then it holds for all points in $W \cap \partial O$, and the argument given in ii) with g_1 replaced by f implies the desired result. Q.E.D.

It is interesting to note that condition i) does not depend on $W \cap \partial O$ being a C^1 manifold.

We seek a minimum number of necessary conditions that an open set $U \subset M$ be the region of reachability from x_0.

Theorem 3.3. Let U be the region of reachability from x_0 of the system (2.1). Suppose ∂U is a C^1 manifold for an open neighborhhood W of $x \in \partial U$ in M. Then $W \cap \partial U$ is an integral manifold of $g_1,\ldots,g_{n-1}$, and the vector field f assigns to $W \cap \partial U$ vectors pointing in the direction of U.

Proof. It follows from part ii) of the preceding theorem that $W \cap \partial U$ is an integral manifold of $g_1,\ldots,g_{n-1}$. Hence $W \cap \partial U$ is actually a C^∞ submanifold of M. Since f, $g_1,\ldots,g_{n-1}$ form a

linearly independent set on M and $W \cap \partial U$ is an integral manifold of $g_1,\ldots,g_{n-1}$, part iii) implies the statement concerning f. Q.E.D.

We shall prove in the next section that the hypothesis ∂U is C^1 near x is superfluous.

THE MAIN RESULT

The following theorem was proved in [6] for use in the uniqueness of analytic continuation problem for CR-distributions on CR-hypersurfaces in $\mathbb{C}^n$, $n > 1$. The statement concerning a C^2 boundary can be relaxed to C^1, or we can simply replace C^1 by C^2 everywhere in the preceding section.

Theorem 4.1 [6]. Let M be a C^∞ manifold of dimension n, and let H be a subbundle of the tangent bundle of M with fiber dimension n-1. Suppose $U \subset M$ is an open set with the property that if $O \subset U$ is an open set having a C^2 boundary, then for each $x \in \partial O \cap \partial U$ we have $T_x(\partial O) = H_x$ (the fiber of H at x). Then for each point $x \in \partial U$, there is a neighborhood V of x, a real-valued function $h \in C^\infty(V)$ with nonzero differential for all point in V, and a closed nowhere dense set $E \subset \mathbb{R}$ such that

(1) $\partial U \cap V = \{x \in V | h(x) \in E\}$,

(2) for each $\ell \in E$, $S_\ell = \{x \in V | h(x) = \ell\}$ is an integral manifold of H; i.e. the boundary of U is foliated by integral manifolds of H.

We now restate Theorem 3.3 under more general conditions.

Theorem 4.2. Let U be the region of reachability from x_0 of the system (2.1). Then $\partial U = N$ is a C^∞ integral manifold of

$g_1,\ldots,g_{n-1}$ (or more generally, is foliated by such integral manifolds) and f assigns vectors on N which point in the direction of U.

Proof. Let H be the subbundle of T(M) spanned by $g_1,\ldots,g_{n-1}$. If O is an open subset of U with a C^2 boundary, then an application of Theorem 3.2 and Theorem 4.1 give us the stated conclusion. Q.E.D.

We have the following important corollary, the proof of which is obvious.

Corollary 4.3. Suppose M is connected and contains no integral manifolds of $g_1,\ldots,g_{n-1}$ for which both of the following statements hold:

a) The closure of the integral manifold is foliated by integral manifolds.

b) The vectors assigned by f on this integral manifold always point in the same direction relative to the integral manifold (i.e. if this manifold divides M into two components, the vectors must point toward the same component). Then the system (2.1) is controllable from any $x_0 \in M$.

Let Λ^d denote Hausdorff measure (see [3]) in dimension d on M. Suppose L is the set of points on which the Lie algebra L_A has dimension n-1. Then L is a closed set in M, and the Frobenius Theorem implies that L contains the integral manifolds of $g_1,\ldots,g_{n-1}$, if any exist. For such an integral manifold we must have $\Lambda^{n-1}(L) > 0$, and we have proved our next result.

Theorem 4.4. If $\Lambda^{n-1}(L) = 0$ and M is connected, then the system (2.1) is controllable from any $x_0 \in M$.

Notice that if M is of dimension 2, we always have integral curves of g for the system $\dot{x}(t) = f(x(t)) + u(t)g(x(t))$, $x(0) = x_0$. Thus Theorem 4.4 does not apply in this case.

We state two theorems from [1] and indicate in a rather superficial way the relation of these theorems to this present work. We restrict our attention to dimension 2 and to a hypersurface system.

Theorem 4.5 [1]. Suppose f and g are vector fields on a C^∞ real 2-dimensional manifold M. Suppose that $\{f,g\}$ meet the conditions of Chow's Theorem for C^∞ vector fields, and suppose that for each initial condition x_0 the solution of $\dot{x}(t) = f(x(t))$ is periodic with a least period $T(x_0)$. Then the reachable set from x_0 of the system $\dot{x}(t) = f(x(t)) + u(t)g(x(t))$ is the set given by Chow's Theorem.

We start at $x_0 \in M$ and take the integral curve of g through x_0. Suppose this curve divides M into two connected components M^+ and M^-. If the solution of $\dot{x}(t) = f(x(t))$ starts at x_0 in the direction of M^+, then since the solution is periodic, there is some point on the integral curve of g through x_0 at which the vector of f is in the M^- direction. Of course, this is in keeping with Theorem 4.2.

In [5] is proved a very nice generalization of the following result, which we state in dimension 2.

Theorem 4.6 [5]. Consider the system

$$\dot{x}(t) = f(x(t)) + u(t)g(x(t)),\ x(0) = x_0,$$

for a C^∞ real 2-dimensional manifold M. Suppose $[f,g] = \alpha g$ on M, where α is a C^∞ function on M. Then the reachable set from x_0 is obtained by taking the integral curve α of f through x_0 (in the positive time sense) and then all integral curves of g interesecting α.

This 2-dimensional version can be seen in light of the following result found in [4]. The one-parameter group of transformations generated by f permutes the integral curves of g with a change of parametrization if $[f,g] = \alpha g$ for some C^∞ function α on M. Interpreted freely, once an integral curve of f passes through an integral curve of g it can never return. This seems to be in agreement with Theorem 4.2.

An obvious question to ask is if the necessary conditions of Theorem 4.2 are also sufficient.

Theorem 4.7. Let $x_0 \in M$ and suppose U is the smallest open subset of M with $x_0 \in \bar{U}$ satisfying $\partial U = N$ is an integral manifold of $g_1,\ldots,g_{n-1}$ and f assigns vectors to N in the direction of U. Then U is the region of reachability from x_0 for the system (2.1).

Proof. We know that we can reach an open set and by the theory developed in this paper we have that we can reach U. The important fact to remember is that to leave U we must break through N near some point $x_1 \in N$. In the system

$\dot{x}(t) = f(x,t) + \sum_{i=1}^{n-1} u_i(t)g_i(x(t))$ at the point x_1 we can move in the directions $f, g_1, \ldots, g_{n-1}, -g_1, \ldots, -g_{n-1}$, and $f + \sum_{i=1}^{n-1} u_i(t)g_i$ for the appropriate finite u_i's. Since N is an integral manifold of $g_1, \ldots, g_{n-1}$, Lie brackets like $[g_i, g_j]$ with $i \neq j$ will give us no "new" directions in which to move from x_1. Also, since $f, g_1, \ldots, g_{n-1}$ span T(M), the brackets $[f, g_i]$, $i=1,\ldots,n-1$, will yield vector fields which are linear combinations of $f, g_1, \ldots, g_{n-1}$ (the same is also true for successive Lie brackets). The only linear combinations here which we can use at x_1 are those already indicated by the system. Q.E.D.

The proof of Theorem 4.7 is applicable only for hypersurface systems (or certain general systems that behave like hypersurface systems). We shall consider results as in this paper for general systems of the form

$\dot{x}(t) = f(x(t)) + \sum_{i=1}^{m} u_i(t)g_i(x(t))$ elsewhere. Such a system with $m < n-1$ seems much more difficult to handle than a hypersurface system.

REFERENCES

1. R. W. Brockett, Nonlinear systems and differential geometry, Proc. IEEE 64(1976), 61-72.

2. W. L. Chow, Über Systems Von Linearen Partiellen Differentialgleichungen erster Ordinung, Math. Ann. 177(1939), 98-105.

3. H. Federer, Geometric Measure Theory, Springer-Verlag, New York, 1969.

4. R. Hermann, Differential Geometry and the Calculus of Variations, Academic Press, New York, 1968.

5. R. Hirschorn, Topological semigroups, sets of generators, and controllability, Duke Math J. 40(1973), 937-947.

6. L. R. Hunt, J. C. Polking, and M. J. Strauss, Unique continuation for solutions to the induced Cauchy-Riemann equations, J. Differential Equations 23(1977), 436-447.

7. R. E. Kalman, P. L. Falb, and M. A. Arbib, Topics in Mathematical Systems Theory, McGraw-Hill, New York, 1969.

8. H. Sussmann and V. Jurdjevic, Controllability of nonlinear systems, J. Differential Equations 12(1972), 95-116.

RESEARCH

on

MULTIDIMENSIONAL SYSTEM THEORY

J. Murray

DEPARTMENT OF ELECTRICAL ENGINEERING

TEXAS TECH UNIVERSITY

ABSTRACT

A class of two-dimensional recursive digital filters called symmetric half-plane filters is discussed; some properties of these filters are derived and it is shown that in certain situations these properties may give the symmetric half-plane filters both theoretical and practical advantages over previously proposed filters. In particular, they are ideally suited to highly parallel processing.

INTRODUCTION

In the literature on 2-dimensional recursive digital filters, two main types of filter have been studied; these are the quarter-plane filter (e.g.[1], [2]) and the asymmetric half-plane filter [3]. Basically, the two correspond to two different concepts of causality. The general stability conditions for a wide class of filters (including symmetric half-plane) were discussed in [4]; unfortunately, however, those filters are not recursively implementable in general. Here we will consider a class of filters which are recursively implementable, and satisfy the same stability conditions as those in [4].

SYMMETRIC HALF-PLANE FILTERS

By a symmetric half-plane filter we will mean a (causal, recursive) 2-dimensional digital filter, the denominator of whose transfer function is of the form

$$A(Z_1, Z_2) = 1 + \sum_{m=1}^{M} \sum_{n=-N}^{N} a_{mn} Z_1^n Z_2^m \qquad (1)$$

This differs from the filters in [4] in that m goes from 1, rather than 0, to M; i.e., this filter omits all of the row m=0 except for the constant term; the asymmetric half-plane filters omit half of this row. The filter (1) is recursively realizable, since the computation of the output at any point depends only on the outputs in previously computed rows; looked at from another point of view, each row of output depends only on previous rows of output. This has two effects; firstly, it focuses attention on the row as the basic element in the filter; secondly, it implies that all the outputs in a given row may be computed in parallel, since each output in a row depends only on outputs in previous rows, and not on any of the outputs of the same row. This is the main practical advantage of this class of filters - it would be of significance, however, only in real-time hardware applications of 2-dimensional filtering, and these seem to be few.

SOME PROPERTIES

Using the methods in [4], one can easily derive the following:

The filter (1) (i.e., the all-pole filter whose denominator is $A(Z_1,Z_2)$) is stable if $A(Z_1,Z_2) \neq 0$ for all (Z_1, Z_2) such that $|Z_1| = 1$ and $|Z_2| \leq 1$.

We note that this set is the same as that for the symmetric half-plane filter in [4]; it is smaller than that for the asymmetric half-plane filters [3]. It is the smallest "instability set" (known to the author) of any recursively implementable class of filters.

However, there is a price to be paid; the amplitude response of the filter is restricted as follows:

If $A(Z_1,Z_2)$ is of the form (1), and if $\frac{1}{A(Z_1,Z_2)}$ is the transfer function of a stable filter, then

$$\int_0^{2\pi} \log \; |A(e^{j\theta_1},e^{j\theta_2})| \; d\theta_2 = 0 \qquad (2)$$

independently of θ_1.

Thus, the average gain along any line of length 2π parallel to the θ_2-axis is constant; or in other words the filter cannot have variations in the θ_1-direction in overall (average) gain. Equivalently, if the cepstrum of $|A(e^{j\theta_1}, e^{j\theta_2})|$ is given by $\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \tilde{a}_{mn} Z_1^n Z_2^m$,

then $\tilde{a}_{on} = 0$, for all n.

This follows immediately from (2) and the definition of the cepstrum; (2) will be proved in a forthcoming paper.

This implies that in order to realize an arbitrary magnitude function, the filter must either have a (nonminimum-phase) numerator polynomial, or the filter must be cascaded with a 1-dimensional filter in Z_1. It is very easy to calculate the ideal amplitude response of this filter.

DESIGN AND IMPLEMENTATION CONSIDERATIONS

It is conceptually convenient (and in a large number of cases, computationally efficient) to implement the convolution in the Z_1-direction by means of the Fourier Transform. (It is assumed from the beginning that the dimension of the array to be filtered is a known fixed constant in the Z_1-direction, i.e., each row is of the same fixed width). From this point of view, and regarding each row as a single entity described by its 1-dimensional Z-transform, the coefficients a_{mn} in (1) are irrelevant; what matters are the M functions

$$\hat{a}_m(e^{j\theta_1}) = \sum_{n=-N}^{N} a_{mn}(e^{j\theta_1})^N.$$

Further, the stability requirement for the filter is equivalent to the requirement that for each fixed θ_1, the filter defined by

$$A_{\theta_1}(Z_2) = \frac{1}{\sum_{m=1}^{M} \hat{a}_m(e^{j\theta_1}) Z_2^m + 1}$$

is 1-dimensionally stable. Finally, the functions $\hat{a}_m(e^{j\theta_1})$ do not have to be

analytic or meromorphic functions; this is seen by letting $N \to \infty$. In other words, the roots of $\hat{a}_m(e^{j\theta_1})$ can vary quite arbitrarily with θ_1. Thus, we can design the one-variable filter $A_{\theta_1}(Z_2)$ by any one of the usual one-variable design methods we choose (yielding a stable filter) for each θ_1: the result will be a stable two-variable filter: further, if our one-variable method gives poles and zeros explicitly, we have the same for our two variable filter, which can therefore be expressed as a cascade of filters of degree 1 in Z_2. Finally, if one desires a filter of finite degree in Z_1, one can solve the following approximation problem (for each m, $1 \leq m \leq M$); minimize (over b_{mn})

$$\left\| \sum_{n=-N}^{N} b_{mn} e^{jn\theta_1} - \hat{b}_m(e^{j\theta_1}) \right\| \text{ subject to}$$

$$\left| \sum_{n=-N}^{N} b_{mn} e^{jn\theta_1} \right| < 1 \text{ for all } \theta_1, \text{ where}$$

$\hat{b}_m(e^{j\theta_1})$ denotes the m-th pole of $A_{\theta_1}(Z_2)$ as a function of θ_1, and $\| \ \|$ denotes some error norm.

Hopefully this will become clearer on consideration of the following example.

EXAMPLE

We wish to design a filter with second-order Butterworth response in Z_2 to approximate the fan filter whose passband is the set $|\theta_2| < |\theta_1|$. For fixed θ_1, therefore, the filter is a 1-dimensional filter whose passband is the set $|\theta| < |\theta_1|$. Using the bilinear transform technique, we find the second-order continuous Butterworth filter

$$\frac{1}{1 + \sqrt{2}\, s/\omega_c + s^2/\omega_c^2}$$

transforms into

$$\frac{\omega_c^2 (1 + Z_2)^2}{\omega_c^2 - \sqrt{2}\,\omega_c + 1 + (2\omega_c^2 - 2) Z_2 + (\omega_c^2 + \sqrt{2}\,\omega_c + 1) Z_2^2} \tag{3}$$

while its (stable) poles transform into

$$\{1 + \omega_c(\tfrac{-1+j}{\sqrt{2}})\}/\{1 - \omega_c(\tfrac{-1+j}{\sqrt{2}})\}$$

and

$$\{1 + \omega_c(\tfrac{-1-j}{\sqrt{2}})\}/\{1 - \omega_c(\tfrac{-1-j}{\sqrt{2}})\}$$

for $\omega_c \geq 0$.

In accordance with the usual frequency warping, we take $\omega_c = |\tan \theta_1/2|$; however, we note that this causes stability problems at $\theta_1=0$ and $\theta_1=\pi$; we therefore take a small perturbation of ω_c, e.g.,

$$\omega_c = \sqrt{\frac{(\sin \theta_1/2)^2 + \varepsilon}{(\cos \theta_1/2)^2 + \varepsilon}} \tag{4}$$

The filter can now be directly implemented by multiplying the Fourier transform of the previous output rows by the appropriate functions according to (3) and (4) and performing the recursion from row to row directly.

Alternatively, if a finite-order (in Z_1) filter is desired, we must solve the problem: Minimize (over b_{1n})

$$\int_0^{2\pi} \left| \frac{\sqrt{2}+(j-1)|\tan \theta_1/2|}{\sqrt{2}+(1-j)|\tan \theta_1/2|} - \sum_{n=-N}^{N} b_{1n} e^{jn\theta_1} \right| d\theta_1$$

$$\text{subject to} \quad \left| \sum_{n=-N}^{N} b_{1n} e^{jn\theta_1} \right| < 1 \quad \text{for all} \quad \theta_1$$

and similarly for the other root. The resulting filter may then be implemented in

cascade. It should be pointed out that while the above optimization problem is not simple, it is one-dimensional.

Finally, the above example was chosen for simplicity and convenience rather than realism. Clearly a Butterworth filter is not optimal for this problem, especially when it yields a design which is not all-pole; if we have to store input rows (as well as output rows) we may as well use them, and design an elliptic type filter; alternatively, we might use a filter which is all-pole in its discrete form.

REFERENCES

[1] R.M. Mersereau and D.E. Dudgeon, "Two-Dimensional Digital Filtering", Proc. IEEE, Vol. 63, pp. 610-623, Apr. 1975.

[2] J.L. Shanks, S. Treitel, and J.H. Justice, "Stabilty and Synthesis of Two-Dimensional Recursive Filters", IEEE Trans. Audio Electroacoust, Vol. AU-20, pp. 115-128, Oct. 1972.

[3] M.P. Ekstrom and J.W. Woods, "Two-Dimensional Spectral Factorization with Application to Recursive Digital Filtering", IEEE Trans. Acoust., Speech and Signal Processing, Vol. ASSP-24, pp. 115-128, April 1976.

[4] J.H. Justice and J.L. Shanks, "Stability Criterion for N-Dimensional Digital Filters", IEEE Trans. Autom. Contr., Vol. AC-18, pp. 284-286, June 1973.

RESEARCH

on

OPTICAL NOISE

G. Froehlich and J. Walkup

DEPARTMENT OF ELECTRICAL ENGINEERING

TEXAS TECH UNIVERSITY

ABSTRACT

Many physical noise processes are signal-dependent. One well known example is film-grain noise (1-3). In this note, an example of the application of optimal estimators for images in signal-dependent film-grain noise is presented.

THE MODEL

A versatile model incorporating both signal-independent additive noise and signal-dependent noise is utilized. This model is given in Eq. (1),

$$r = s + kf(s)n_1 + n_2, \tag{1}$$

where r is the observed photographic density, s is the original uncorrupted image density, k is the scanning constant, f(s) is some function of s, and n_1 and n_2 are signal-independent noise processes. Thus, the middle term on the right-hand side of Eq. (1) is the signal-dependent noise term.

It is assumed that n_1, n_2, and s are mutually statistically independent. To apply the model to film-grain noise problems, let $f(s) = s^p$, where p is ususally taken to be 1/2 or 1/3 (1-3).

In this note, we let p = 1/2 and we assume n_1 and n_2 are zero mean Gaussian random variables, with variances σ_1^2 and σ_2^2, respectively. Further, s is assumed to be a Gaussian random variable with mean μ_s and variance σ_s^2.

THE ESTIMATOR STRUCTURES

The maximum likelihood (ML) estimate is found by maximizing p(r/s) over s (3). For the model of Eq. (1), the estimate is found

to be

$$\hat{s}_{ML} = \left[r^2 + \left(\frac{k^2\sigma_1^2}{2}\right)^2 + \frac{2r\sigma_2^2}{k^2\sigma_1^2} + \left(\frac{\sigma_2^2}{k^2\sigma_1^2}\right)^2\right]^{1/2} - \frac{k^2\sigma_1^2}{2} - \frac{\sigma_2^2}{k^2\sigma_1^2}, \quad (2)$$

as compared to the simple estimate

$$\hat{s}_{ML} = r \quad (3)$$

which results when the signal-dependent noise term of Eq. (1) is omitted.

The maximum a posteriori probability (MAP) estimate is found by maximizing p(s/r) over s(3). For the model of Eq. (1) and the above assumptions, the estimate $\hat{s}_{map}$ is found to be the solution of

$$\left[\frac{2k^4\sigma_1^4}{\sigma_s^2}\right]\hat{s}^3 + \left[\frac{4k^2\sigma_1^2\sigma_2^2 - 2k^4\sigma_1^4\mu_s}{\sigma_s^2} + 2k^2\sigma_1^2\right]\hat{s}^2$$
$$+ \left[\frac{2\sigma_2^4 - 4k^2\sigma_1^2\sigma_2^2\mu_s}{\sigma_s^2} + k^4\sigma_1^4 + 2\sigma_2^2\right]\hat{s} \quad (4)$$
$$+ \left[k^2\sigma_1^2(\sigma_2^2 - r^2) - 2\sigma_2^2 r - \frac{2\sigma_2^4\mu_s}{\sigma_s^2}\right] = 0.$$

Again, omission of the signal-dependent noise term in Eq. (1) results in a comparatively simplified estimate,

$$\hat{s}_{MAP} = \frac{\sigma_s^2}{\sigma_s^2+\sigma_2^2} r + \frac{\sigma_2^2}{\sigma_s^2+\sigma_2^2} \mu_s. \quad (5)$$

Because this MAP estimate includes prior information about the image, it should give superior performance. In fact, under the above assumptions it can be shown that the MAP estimator minimizes the mean square estimation error (3).

RESULTS

Figure 1 is the original image of an archer. Figure 2 is the noisy image generated digitally according to the model of Eq. (1). The image in Figure 3 is the estimate found by the solution of the MAP equation, Eq. (4), with μ_s and σ_s^2 taken to be the sample mean and variance of the original image.

One factor severaly affecting estimator performance is violation of the assumption that the image statistics are Gaussian. For a discussion of this, see the paper by Froehlich et.al. (3).

ACKNOWLEDGEMENTS

This work was supported in part by the Joint Services Electronics Program at Texas Tech University under Office of Naval Research Contract No. 76-C-1136.

The authors are also grateful for the assistance of Gus Oliver, who contributed to this note, while an Undergraduate Research Participant sponsored by the National Science Foundation at Texas Tech University.

REFERENCES

1. Walkup, J. and R. Choens, Optical Engineering, Vol. 13, pp. 258-266, (1974).
2. Naderi, F. and A. Sawchuk, Appl. Optics, Vol. 17, pp. 1228-1237, (1978).

3. Froehlich, G., Walkup, J., and R. Asher, to appear in J. Opt. Soc. Am., October 1978.

Figure 1: The original image, k = 0.5.

Figure 2: The noisy image, $\sigma_1 = 0.4$, $\sigma_2 = 0.1$, $\sigma_2 = 0.2$.

Figure 3: The Estimate.

RESEARCH

on

PATTERN RECOGNITION

T. Newman

DEPARTMENT OF MATHEMATICS

TEXAS TECH UNIVERSITY

ABSTRACT

In the following we consider a one-parameter group G acting on a set Ω of patterns. We show that under extremely mild assumptions the problem of recognition of equivalence of two patterns reduces to the recognition of translated functions defined on the real line. Thus, the recognition problem is reduced to one of a practical nature. In fact, the well known properties of the Fourier transform with respect to translation can then be applied in most instances to provide a tractable solution.

FOURIER TRANSFORMS

Let K denote the set of real numbers and C denote the set of complex numbers. We follow Bremerman [1] in defining the Fourier transform of a function $f: K \to K$ (or C) as

$$\hat{f}(\alpha) = \int_{-\infty}^{\infty} e^{j\alpha s} f(s)\,ds. \qquad (1)$$

Of course, the integral may fail to exist, even for well-behaved functions f. Thus, $\hat{f}$ may be a distribution [1] (i.e., a linear functional on a suitable test space). In any event, let us consider the one-parameter group of translations acting on $f: K \to K$ in which a point $t \in K$ corresponds to a translation $\lambda_t: f(s) \to f(s-t)$. By change of variable in equation (1) above we have

$$\widehat{\lambda_t f}(\alpha) = e^{j\alpha t}\hat{f}(\alpha) \qquad (2)$$

In the terminology of our previous reports [6,7], we see that for each $\alpha \in C$ the mapping $\hat{E}_\alpha: f \to \hat{f}(\alpha)$ is a relative invariant whose

modulus is the exponential map ρ_α: K → {z ∈ C| |z|=1} given by $\rho_\alpha(t) = e^{j\alpha t}$. This is a well-known property of the Fourier transform.

Now let f_1, f_2: K → K and let us seek t ∈ K such that $f_2 = \lambda_t f_1$. Provided $\hat{f}_1$ and $\hat{f}_2$ both exist, we see that a necessary condition is that for all α ∈ C we have

$$\hat{f}_2(\alpha) = e^{j\alpha t}\hat{f}_1(\alpha) \tag{3}$$

Determination of t now becomes a matter of practicality. Observe that the condition expressed in equation (2) is sufficient as well since it implies that $f_2 = \lambda_t f_1$ a.e. (almost everywhere). Thus we see that the values of the Fourier transform give a complete set of relative invariants. Now, from (2) and the relation $|e^{j\alpha t}| = 1$ we deduce that the values $|\hat{f}(\alpha)|$ are invariant. Consequently, a necessary condition that $f_2 = \lambda_t f_1$ for some t ∈ K is again expressed by the requirement that for all α ∈ C we have

$$|\hat{f}_2(\alpha)| = |\hat{f}_1(\alpha)| . \tag{4}$$

In the following, we will assume that a recognition problem is solved when it is reduced to the translation problem for functions defined on K.

ONE-PARAMETER GROUPS

By a one-parameter group of transformations acting in K^n we main a family $G = \{\varphi_t\}$ of coordinate transformations in K^n which depend analytically on a real parameter t and form a group with respect to composition (see [4] and [5]). Note that each φ_t: $K^n \to K^n$. It is always possible to choose the parameterization in such a way that $\varphi_t \cdot \varphi_s = \varphi_{t+s}$. Such a system of coordinates in G are called canonical coordinates. For $x \in K^n$ and $t \in K$ we use $\varphi(x,t)$ and $\varphi_t(x)$ interchangeably. The vector field of the group is defined to be

$$\xi(x) = \frac{\partial}{\partial t}\Big|_{t=0}. \tag{5}$$

Note that also ξ: $K^n \to K^n$, say $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ where each ξ_i: $K^n \to K$. Finally, the infinitesimal generator of G is the differential operator

$$U = \sum_{i=1}^{n} \xi_i(x) \frac{\partial}{\partial x_i}, \tag{6}$$

where $x = (x_1, \ldots, x_n)$.

Now the equations for canonical coordinates are expressed as

$$\varphi(\varphi(x,t),s) = \varphi(x,t+s) \tag{7}$$

for all $s,t \in K$ and $x \in K^n$. Application of $\frac{\partial}{\partial t}|_{t=0}$ and $\frac{\partial}{\partial s}|_{s=0}$ to both sides of equation (7) and the use of equations (5) and

(6) yield the results

$$U\varphi = \frac{\partial \varphi}{\partial t}$$

and

$$\frac{\partial \varphi}{\partial t} = \xi(\varphi).$$

The first of these shows that $\varphi(x,t)$ may be reconstructed from the infinitesimal generator as the unique solution to the partial differential equation

$$U\varphi = \frac{\partial \varphi}{\partial t}, \quad \varphi(x,0) = x, \tag{8}$$

while the second shows that $\varphi(x,t)$ may be reconstructed from the vector field ξ as the solution of

$$\frac{\partial \varphi}{\partial t} = \xi(\varphi), \quad \varphi(x,0) = x, \tag{9}$$

which, for a given x, is actually an ordinary differential equation in t. Equation (8) is typically solved [ ] by obtaining solutions to the system of ordinary differential equations

$$\frac{dx_1}{\xi_1(x)} = \frac{dx_2}{\xi_2(x)} = \ldots = \frac{dx_n}{\xi(x_n)} = dt. \tag{10}$$

Equation (9) may be expressed in integral form as

$$\varphi(x,t) = x + \int_0^t \xi(\varphi(x,s))ds. \tag{11}$$

The latter formulation suggests an interesting physical realization. In figure (1) we see a feedback realization of equation (11). The output of an integrator is input as argument to a generator for ξ values. The resulting value is input to the integrator. Starting with the input x to the integrator

at time t = 0, the output is the trajectory $\varphi(x,t)$, $t \geq 0$. Replacing ξ by $-\xi$ gives the trajectory backward in time, $\varphi(x,-t)$, $t \geq 0$. In effect, one is able to obtain group trajectories in a straightforward manner in analog form.

This observation merits further study and will be actively pursued in the future research. Comparison with the work of Brockett [2,3] amplifies the belief that further efforts along this line can reduce many recognition problems

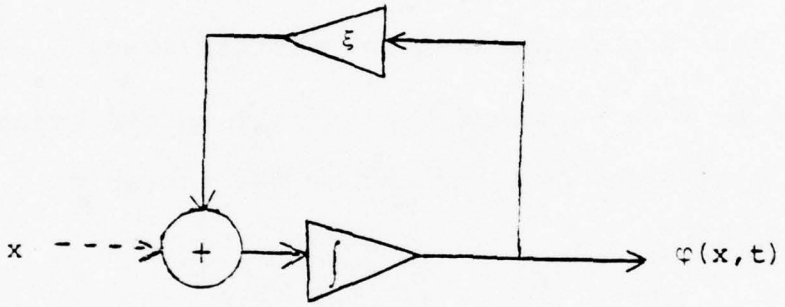


Figure 1. Realization of equation (11).

to problems in control theory. Initial efforts, not complete at this time, suggest that similar methods can be used in more general n-parameter groups and that a control may be introduced to select among the (infinitely) many one-parameter subgroups.

INVARIANTS FOR ONE PARAMETER GROUPS

As before, let G be a one-parameter group of transformations in K^n. It is shown in [5] that a map $h: K^n \to K$ is G-invariant if and only if $Uh = 0$. Furthermore, there are n-1 linearly independent solutions $h_1, h_2, \ldots, h_{n-1}$ such that any invariant h is expressible as $h = H(h_1, h_2, \ldots, h_{n-1})$.

Now, we may select $y_1 = h_1(x), \ldots, y_{n-1} = h_{n-1}(x)$ and an additional coordinate y_n so that in the coordinate system $y_1, \ldots, y_n$ the action of G becomes

$$\varphi_i(y,t) = y_i, \quad i < n$$
$$\text{and} \qquad \varphi_n(y,t) = y_n + t. \tag{12}$$

That is, in suitable coordinates, the action of G is translation in the last coordinate.

Now let us suppose that the pattern space Ω is a set of functions $f: K^n \to V$, where V is a real vector space and that the action of G on Ω is induced by the action of G in K^n. Thus, for $\varphi_t \in G$, $f \in \Omega$

$$(\varphi_t f)(x) = f(\varphi_t^{-1}(x)). \tag{13}$$

Following the change of coordinates, we see that the problem of recognition of equivalence of $f_1, f_2 \in \Omega$ is reduced to translation in the n-th coordinate. Thus, we may consider this problem reduced to the Fourier transform methods previously discussed.

THE GENERAL PROBLEM FOR ONE-PARAMETER GROUPS

Let Ω be an arbitrary set of patterns and let G be a one-parameter group which acts on Ω. Also, let $R: \Omega \to V$ be a measurement function on Ω with values in some real vector space. As in earlier reports we obtain a representation of $w \in \Omega$ as a function $w^r: G \to V$ by defining $w^r(g) = R(g^{-1}w)$. In terms of canonical coordinates in G we may define a new representation

$\bar{w}$: K → V by the formula

$$\bar{w}(s) = w^r(\varphi_s), \quad s \in K. \tag{14}$$

Now for $t \in K$ we have

$$\overline{\varphi_t w}(s) = (\varphi_t w)^r(\varphi_s) = w^r(\varphi_t{}^{-1}\varphi_s)$$

$$= w^r(\varphi_{s-t}) = \bar{w}(s-t) = \lambda_t \bar{w}(s).$$

That is, we have successfully represented Ω as a class of functions $\bar{w}$: K → V is such a way that the action of G becomes translation. The question of equivalence of patterns may now be resolved by the use of the Fourier transform methods previously discussed.

CONCLUSIONS

We have shown that the recognition of equivalence of patterns under a one-parameter group is always reducible to one-dimensional translation of functions of a real variable. The recognition problem is thus one of a practical nature which may be resolved by known methods, such as the use of Fourier transforms.

Further work will be done in the attempt to generalize the above to general n-parameter groups. This will be complicated, in general, by the lack of a transform analogous to the Fourier transform, which fails to be readily available except in the Abelian case.

We have also given a means of analog generation of trajectories which suggests a control theory approach to recognition problems. This technique will be extended to n-parameter groups and the relation to general system theory will be explored.

AD-A059 977 TEXAS TECH UNIV LUBBOCK INST FOR ELECTRONICS SCIENCE F/G 9/5
ANNUAL REVIEW OF RESEARCH UNDER THE JOINT SERVICES ELECTRONICS --ETC(U)
OCT 78 R SAEKS, K S CHAO, R B ASHER, J WALKUP N00014-76-C-1136

UNCLASSIFIED NL

2 OF 2
ADA
059977

END
DATE
FILMED
12-78
DDC

REFERENCES

[1] Bremerman, H., Distributions, Complex Variables and Fourier Transforms, Addison-Wesley, Reading, Mass., 1965.

[2] Brockett, R. W., Non-linear Systems Theory, Proceedings IEEE, Vol. 64, No. 1, 1976, pp. 61-72.

[3] Brockett, R. W., Lie Theory and Control Systems Defined on Spheres, Siam. J. Appl. Math., Vol. 25, No. 2, 1973, pp. 213-225.

[4] Cohn, P. M., Lie Groups, Cambridge University Press, London, 1957.

[5] Ince, E. L., Ordinary Differential Equations, Dover, New York, 1956.

[6] Newman, T. G., Research on Pattern Recognition, in Semiannual Review of Research under the Associate Joint Services Electronics Program, Contract N0014-76-C-1136, April, 1977, pp. 96-103.

[7] Newman, T. G., Pattern Recognition, in Semiannual Review of Research Under the Associate Joint Services Electronics Program, Contract N0014-76-C-1136, October, 1977, pp. 143-160.

RESEARCH

on

BUILT-IN TESTING

R. Saeks

DEPARTMENT OF ELECTRICAL ENGINEERING

TEXAS TECH UNIVERSITY

ABSTRACT

A failure prediction algorithm for application in a periodic on-line maintenence system operating in a Poisson shock environment is described. The system under test is measured at periodic maintenence intervals with the data derived therefrom, being used to estimate system lifetime and determine an optimal replacement time. The resultant algorithm is simulated and compared with various fixed replacement schedules.

INTRODUCTION

Fault analysis processes, have been and will continue to be very significant factors in the safety and reliability of electrical systems. This is especially true due to the following facts: a rapid advancement in the complexity and size of modern systems, increased availability and capabilities of computers, and rapidly changing technologies in integrated circuit fabrication. Due to this, fault analysis has become much more than an academic research topic. Fault analysis is applicable in an industrial environment to minimize cost, extend the lifetime of the overall system, control maintenance schedules, and effectively plan manpower needs.

Although considerable effort has been expended during the past decade to develope techniques for fault detection and diagnosis in both analog and digital electronic circuits[1] little attention has been given to the possibility of formulating algorithms for fault prediction. To accurately predict a fault, a device must be tested at periodic maintenance intervals. If the device fails or does not operate correctly, it is replaced

immediately. The device may be assumed good if its characteristics are in tolerence. However, if the characteristics are slightly off nominal, but the device still operates correctly, one can attempt to predict if the device will fail before the next scheduled maintenance interval. If device failure is predicted, it can be replaced before failure occurs as part of planned preventative maintenance.

With the advent of the low-cost microprocessor, on-line fault prediction is possible and practical.[2] A curve fitting algorithm for on-line fault prediction was first introduced by Saeks, Liberty and Tung[3,4,5] in 1975. It was assumed that prior life-time statistics for the system under test were known. Also, performance data of the system at each maintenance interval were collected. The application of these data to a second order polynomial equation resulted in an estimation of the time at which the component under test would exceed tolerance limits. Based on a criterion of simultaneously minimizing on-line failures and maximizing component lifetime, a decision as to whether or not the component should be replaced is made at each maintenance interval.

The disadvantages of this curve fitting algorithm are: the application is limited to failures due to permanent overstress, the second order polynomial is too simple to describe the performance of the component, and the prior lifetime statistics for the component are often not available.

Another area where an extensive research effort is being applied is shock models and wear processes. Esary, Marshall and Proschan[6,7,8] introduced a shock model for the life distribution of a component subjected to a sequence of shocks randomly occuring in time according to a homogeneous Poisson process. They also

considered the related shock models in which each shock caused a random amount of damage and failure occured when the accumulated damage exceeded a specified threshold. This failure model is well known in modern reliability theory.

Employing the Poisson-Shock model, another curve fitting fault prediction algorithm which will overcome the disadvantages of the Saeks-Liberty-Tung algorithm will be discussed in the present paper.[9] In the following section, a model for the failure dynamics of a system component parameter is formulated. Here, it is assumed that the faulure is due to the component being subjected to a sequence of Poisson distributed shocks[10,11] with the measurable parameter being controlled by an unknown difference equation whose underlying discrete "component time" process is defined by the number of shock to which the component has been subjected. Since both the failure dynamics (i.e. the difference equation) and the relationship between "component time" and real time are unknown, our failure model is doubly stochastic. The third section of the paper is devoted to the formulation of an algorithm for estimating the component failure dynamics and its "lifetime", defined to be the number of shocks required to cause component failure. This is followed by the formulation of an "optimal" replacement theory wherein the optimal real time at which to replace a component is computed in terms of its estimated "lifetime". Finally, the results of a simulation of the algorithm in both an ideal and noisy environment are presented and compared with the simulated performance for several fixed replacement schedules.

FAILURE DYNAMICS

The performance of an analog device subject to a series of discrete shocks (switching process, improper operation, etc...) may

drift due to the shock damage. Let C(N) represent values of a particular component parameter, where the "component time", N, denotes the number of shocks the component has received. It is assumed that the drifting parameters can be described by a first order† difference equation of the form:

$$1. \qquad C(N+1) = C(N) - a_o - a_1 N - a_2 N^2 - \ldots - a_h N^h \quad C(0) = 1$$

Here, the coeficients and order of the "forcing polynomial" are assumed to be unknown and must be estimated as part of the fault prediction process. A little algebra together with the standard recursive formula for solving a difference equations will reveal that

$$2. \qquad C(N) = 1 - \sum_{j=0}^{N-1} \sum_{i=0}^{h} a_i j^i$$

Now, if the tolerence limit for the component parameter is taken to be C = 0, we may define the lifetime of the component to be the smallest integer, N, for which $C(N) \leq 0$. This integer which we denote by L then represents the number of shocks necessary to cause the component to fail.

Consider a simple example where the "forcing polynomial" is taken to be of the first order with positive coeficients. Then 1. reduces to

$$C(N+1) = C(N) - a_o - a_1 N \; ; \; C(0) = 1$$

From equation 2. C(N) can thus be expressed as

$$C(N) = 1 - a_0 N - \frac{N(N-1)}{2} a_1$$

†The concepts described herein carry over without modification to the case where the failure model is characterized by higher order difference equations. The first order model, however, suffices to illustrate the theory and his hence used throughout the present paper.

Then the lifetime of this component is the smallest integer satisfying the equation

$$1 - a_0 L - \frac{L(L-1)}{2} a_1 \geq 0$$

That is, L is the smalles integer such that

$$L \geq \frac{(2a_0 - a_1)^2 + 8a_1 - (2a_0 - a_1)}{2a_1}$$

Since the failure model of equation 1. is dependent on "component time", i.e. the number of shocks the component has received, rather than real time, it remains to define the relationship between "component time" and real time. Following common practice in reliability theory[6], we assume that this relationship is determined by a Poisson process. Indeed, this is the unique point process which has the scaling properties required for such an application.[11] Here, the probability of N shocks occuring in the time interval t is:

$$3. \qquad P_N(t) = e^{-kt} \frac{(kt)^N}{N!} \qquad N = 0,1,2, \ldots$$

Where k is a given constant representing the average number of shocks per unit time. Therefore, (kt) is the average number of shocks in the time interval t.

If a component with lifetime L is subjected to Poisson shock with constant k the probability that it will fail (i.e. receive at least L shocks) by time t, is then given by the formula[9]

$$F(t) = \sum_{n=0}^{L-1} P_n(t)$$

$$= \sum_{n=0}^{L-1} e^{-kt} \frac{(kt)^n}{n!}$$

Thus, even though the lifetime of our component is integer valued, in our model the actual failure time is continuously distributed since the time at which the component receives the Lth shock is continuously distributed.

ESTIMATION OF FAILURE DYNAMICS AND LIFETIME

In a periodic maintenance system, the performance of a component is measured at each maintenance interval nT. That is to say, $(C_1, C_2, \ldots, C_g)$ is the performance data taken at maintenance times $(T, 2T, \ldots, gT)$. The estimation problem can be stated as:

"Given performance data $(C_1, C_2, \ldots, C_g)$, T and k, estimate the unknown constants $(a_0, a_1, \ldots, a_h)$ of the failure dynamics." Since it is assumed that the system is subjected to Poisson Shock with constant k, the expected number of shocks in each maintenance interval is kT.† As such, if we assume that C_m is the value of the component parameter at N = mkT, then upon substituting $C_m = C(mkT)$ into equation 2. we obtain

$$\sum_{j=0}^{mkT-1} a_0 j^0 + \sum_{j=0}^{mkT-1} a_1 j^1 + \ldots\ldots + \sum_{j=0}^{mkT-1} a_h j^h = 1 - C_m$$

where m = 1, 2, 3, ..., g or in the matrix form:

†Although not theoretically necessary, we assume that kT is an integer.

$$4.\quad JA \triangleq \begin{pmatrix} \sum_{j=0}^{kT-1} j^0 & \sum_{j=0}^{kT-1} j^1 & \cdots & \sum_{j=0}^{kT-1} j^h \\ \sum_{j=0}^{2kT-1} j^0 & \sum_{j=0}^{2kT-1} j^1 & \cdots & \sum_{j=0}^{2kT-1} j^h \\ \vdots & \vdots & & \vdots \\ \sum_{j=0}^{gkT-1} j^0 & \sum_{j=0}^{gkT-1} j^1 & & \sum_{j=0}^{gkT-1} j^h \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_h \end{pmatrix} = \begin{pmatrix} 1-C_1 \\ 1-C_2 \\ \vdots \\ 1-C_g \end{pmatrix} \triangleq Z$$

Since the number of data points, g, is typically much greater than the order of the polynomial assumed in the failure model, h, it is not expected that equation 4. admits an exact solution. Rather, we attempt to solve for a coefficeint vector, A, which minimizes the error between JA and Z. In particular, if one adopts a least squares error criterion the optimal A is given by

$$5.\qquad A^o = J^{-G}Z$$

where J^{-G} denotes the generalized inverse of J.[12] Indeed, if as is typically the case J has full column rank than $J^{-G} = (J^tJ)^{-1}J^t$ where "t" denotes matrix transposition. As such, we take the $A^o = \text{col}(a_o^o, a_1^o, \ldots, a_h^o)$ as our estimate of the coefficients of the difference equation characterizing the failure dynamics of our drifting parameter, C, as per equation 1.

To estimate the failure dynamics of a drifting parameter, the proper choice of the order h is, in general, quite difficult and depends upon physical considerations and engineering experience.

Once h is preselected, however, coefficients to best approximate the failure dynamics can be readily computed via equation 5. The accuracy of the resultant estimate, however, is highly dependent on the choice of the order, h, and on the number of measurements which are taken, g. To find a new set of coefficients for a different combination of h and g, the entire calculation procedure is typically repeated from the very beginning which is impractical in the on-line maintenance system. Fortunately, sequential refinement schemes for obtaining new sets of coefficients without repeating the entire calculation can be developed.[12,13] As such, it is possible to sequentially update one's estimates of the parameters; a_o, a_1, ..., a_h; as additional measurements are taken and/or to increase the order of the model for the failure dynamics without repetitious matrix inversion. Our algorithm for estimation of the failure dynamic underlying the measured data my thus be readily implemented on-line with the computational power presently available in today's microprocessors. The matrix algebraic details of the required sequential refinement schemes are straighforward[12,13] and readily available in the literature. As such, they will not be repeated here.

In practice, given g measurements C_1, C_2, ..., C_g taken at maintenance intervals T, 2T, 3T, ..., gT, one sequentially estimates the coefficients of the failure dynamics; a_o, a_1, ..., a_h; increasing h until no further error reduction is achieved. The resultant set of coefficients is then used in equation 2. to determine the component lifetime, L. Upon solving the equation,

the resultant estimated lifetime is found to be the smallest integer, L, such that

$$6. \quad \sum_{j=0}^{L-1} \sum_{i-0}^{h} a_i j^i \geq 1$$

Of course, if the measured data is not decaying towards zero, i.e. the component isn't failing, this inequality will have no solution in which case we take L to be infinite.[9]

REPLACEMENT THEORY

Although the algorithm outlined in the preceeding secion yields an "optimal" estimate of the number of shocks required to cause failure the time at which the Lth shock takes place is statistical in nature and hence, it still remains to determine the optimal (in an approprate sense) time at which to replace the component. One such criterion is formulated in the following. For this purpose, it is assumed that L has been computed to our satisfaction and we desire to choose a time, T_r, at which to replace the component as a function of L. Given L and T_r we denote the resultant probability of on-line failure (i.e. failure before T_r) by P_f. $P_r = 1 - P_f$ then denotes the probability that the component is replaced at time T_r before it fails. Similarly, we let $\tilde{T}_f$ denote the expected time to failure for those components which fail on-line, we let $\hat{T}$ denote the expected time to failure for all components and we let T* denote the expected time to failure for the components if they were operated to failure without replacement (i.e. $T^* = \hat{T}\big|_{T_r \to \infty}$). Finally, we let $f_L(t)$ denote the probability density function that the component receives the Lth shock at time t, given that the component fails on-line, whereas, $p_i(t)$ represents the density function of the Poisson distribution with parameter (k) and $E_L(t)$ represent the

corresponding distribution function; i.e.

$$7. \quad p_i(t) = \frac{(kt)^i}{i!} e^{-kt} \qquad i = 0, 1, 2, \ldots.$$

$$8. \quad E_L(t) = \sum_{i=0}^{L-1} p_i(t)$$

With the aid of some elementary calculus[9] $P_f, P_r, \tilde{T}_f$, and $\hat{T}$, as well as their derivatives with respect to T_r, can be computed analytically. As such, upon defining an appropriate cost measure an explicit formula for determining an "optimal" T_r given L can be derived. We begin with the derivation of the explicit formula for the various quantities involved in our replacement theory.

Since a component will be replaced by our algorithm if an only if it is still operating at time T_r, i.e. if it has not yet received L shocks at time T_r, the probability of replacement is just the probability of receiving less than L shocks by time T_r. We thus have:

(prop 1) $$P_r = E_L(T_r)$$

$$P_r = \sum_{i=0}^{L-1} \frac{(kT_r)i}{i!} e^{-kT_r}$$

$$= \sum_{i=0}^{L-1} p_i(T_r)$$

$$= E_L(T_r)$$

(prop 2) $$P_f = 1 - E_L(T_r)$$

(prop 3) $$\int_0^{T_r} p_i(t)\, dt = \frac{1}{k}\{1 - E_{i+1}(T_r)\}$$

$$\int_0^{T_r} p_i(t)\, dt = \int_0^{T_r} \frac{(kt)^i}{i!}\, e^{-kt}\, dt$$

$$= \frac{k^i}{i!} \int_0^{T_r} t^i\, e^{-kt}\, dt$$

Using the identity

$$\int x^m\, e^{ax}\, dx = e^{ax} \sum_{r=0}^{m} (-1)^r \frac{m!\, x^{m-r}}{(m-r)!\, a^{r+1}}$$

this becomes

$$\int_0^{T_r} p_i(t)\, dt = \frac{k^i}{i!}\, e^{-kt} \sum_{r=0}^{i} (-1)^r \frac{i!\, t^{i-r}}{(i-r)!\, (-k)^{r+1}} \Bigg]_0^{T_r}$$

$$= \frac{k^i}{i!} \left\{ e^{-k.0} \frac{i!}{k^{i+1}} - e^{-kT_r} \sum_{r=0}^{i} \frac{i!\, T_r^{i-r}}{(i-r)!\, k^{r+1}} \right\}$$

$$= \frac{1}{k} \left\{ 1 - e^{-kT_r} \sum_{r=0}^{i} \frac{(kT_r)^{i-r}}{(i-r)!} \right\}$$

$$= \frac{1}{k} \left\{ 1 - e^{-kT_r} \sum_{j=0}^{i} \frac{(kT_r)^j}{j!} \right\}$$

$$= \frac{1}{k} \left\{ 1 - \sum_{j=0}^{i} p_j(T_r) \right\}$$

$$= \frac{1}{k} \left\{ 1 - E_{i+1}(T_r) \right\}$$

$$\text{(prop 4)} \qquad f_L(t) = \frac{p_{L-1}(t)}{1/k \quad (1-E_L(T_r))}$$

To derive this conditional density function we partition the interval $(0,T_r)$ into N segments of length $\Delta = T_r/N$ and we compute the probability that the Lth shock takes place in the ith time interval, $((i-1)\Delta, i\Delta]$. Since, this can be caused by having L-1 shocks before $(i-1)\Delta$ and at least one shock in the interval $((i-1)\Delta, i\Delta]$ or by having L-2 shocks before $(i-1)\Delta$ and at least two shocks in the interval $((i-1)\Delta, i\Delta]$, etc. the probability of failure in the ith interval is given by

$$\sum_{j=1}^{L} p_{L-j}((i-1)\Delta)[1 - E_j(\Delta)]$$

$$9. \quad = \sum_{j=1}^{L} p_{L-j}((i-1)\Delta) \sum_{q=j}^{\infty} \frac{(\Delta k)^q}{q!} e^{-\Delta k}$$

$$= \sum_{r=1}^{\infty} \frac{1}{r!} \left[\sum_{s=1}^{r} p_{L-s}((i-1)\Delta)\right] (\Delta k)^r e^{-\Delta k}$$

Taking the probability density function at a point t in the interval $((i-1)\Delta, i\Delta]$ to be limiting value of this quanity divided by Δ as Δ goes to zero[10] it is observed that the terms of 9. containing powers of (Δk) greater than 1 go to zero in the limit. As such, the probability density function for the Lth shock to take place at time t is given by

$$10. \quad \lim_{\Delta \to 0} \frac{p_{L-1}((i-1)\Delta)\ (\Delta k)e^{-k}}{\Delta} = kp_{L-1}((i-1)\Delta)$$

Finally, since we are only interested in the conditional probability density function that the Lth shock will take place at time t given that the component fails on-line the quanity of equation 10. must be normalized yielding

$$f_L(t) = \frac{kp_{L-1}((i-1)\Delta)}{P_f} = \frac{p_{L-1}(t)}{\frac{1}{k}(1-E_L(T_r))}$$

as was to be shown.

(Prop 5) $$\tilde{T}_f = \frac{L}{k} \frac{1 - E_{L+1}(T_r)}{1 - E_L(T_r)}$$

Since T_f is the expected lifetime of the components which fail before replacement,

$$T_f = \int_0^{T_r} t\, f_L(t)\, dt$$

$$= \int_0^{T_r} \frac{t\, P_{L-1}(t)}{\frac{1}{k}\{1-E_L(T_r)\}}\, dt$$

$$= \frac{\int_0^{T_r} t\, \frac{(kt)^{L-1}}{(L-1)!}\, e^{-kt}\, dt}{\frac{1}{k}\{1 - E_L(T_r)\}}$$

$$= \frac{\frac{L}{k} \int_0^{T_r} \frac{(kt)^L}{L!}\, e^{-kt}\, dt}{\frac{1}{k}\{1-E_L(T_r)\}}$$

$$= \frac{L\int_0^{T_r} p_L(t)\, dt}{\{1 - E_L(T_r)\}}$$

From (prop 3), equation 11. thus reduces to the desired equality.

(Prop 6) $\hat{T} = \frac{L}{k} \{1 - E_{L+1}(T_r)\} + T_r E_L(T_r)$

$$\hat{T} = P_f \hat{T}_f + P_r T_r$$

$$= \{1 - E_L(T_r)\} \quad \frac{L}{k} \quad \frac{1 - E_{L+1}(T_r)}{1 - E_L(T_r)} + T_r E_L(T_r)$$

$$= \{\frac{L}{k} \quad 1 - E_{L+1}(T_r)\} + T_r E_L(T_r)$$

(Prop 7) $T^* = \frac{L}{k}$

(Prop 8)

$$\frac{d(P_f)}{d(kT_r)} = p_{L-1}(T_r)$$

and

$$\frac{d(P_r)}{d(kT_r)} = -p_{L-1}(T_r)$$

This result follows simply by differentiating the expressions for P_f and P_r of (prop 1) and (prop 2) analytically.

$$P_r = E_L(T_r)$$

$$= \sum_{i=0}^{L-1} \quad \frac{(kT_r)^i}{i!} \quad e^{-kT_r}$$

$$= e^{-kT_r} + \sum_{i=1}^{L-1} \quad \frac{(kT_r)^i}{i!} \quad e^{-kT_r}$$

Thus

$$\frac{d(P_r)}{d(kT_r)} = -e^{-kT_r} + \sum_{i=1}^{L-1} \frac{i(kT_r)^{i-1}}{i!} - \frac{(kT_r)^i}{i!} e^{-kT_r}$$

$$= \sum_{i=1}^{L-1} \frac{(kT_r)^{i-1}}{(i-1)!} e^{-kT_r} - \sum_{i=0}^{L-1} \frac{(kT_r)^i}{i!} e^{-kT_r}$$

$$= E_{L-1} - E_L$$

$$= -p_{L-1}(T_r)$$

Moreover since

$$P_f = 1 - P_r$$

$$\frac{d(P_f)}{d(kT_r)} = \frac{d(1-P_r)}{d(kT_r)} = P_{L-1}(T_r)$$

(Prop 9) $$\frac{d(k\tilde{T}_f)}{d(kT_r)} = L \frac{[1 - E_L(T_r)]\, p_L(T_r) - [1 - E_{L+1}(T_r)]\, p_{L-1}(T_r)}{[1 - E_L(T_r)]^2}$$

From (prop 3)

$$k\hat{T}_f = L \frac{1 - E_{L+1}(T_r)}{1 - E_L(T_r)}$$

Thus by direct differentiation

$$\frac{d(k\tilde{T}_f)}{d(kT_r)} = L \frac{\{1-E_L(T_r)\ p_L(T_r) - \{1-E_{L+1}(T_r)\}\ p_{L-1}(T_r)}{\{1 - E_L(T_r)\}^2}$$

(Prop 10) $$\frac{d(k\hat{T})}{d(kT_r)} = E_L(T_r)$$

From (prop 6)

$$\hat{T} = \frac{L}{k} \{1 - E_{L+1}(T_r)\} + T_r E_L(T_r)$$

hence

$$k\hat{T} = L\{1 - E_{L+1}(T_r)\} + kT_r E_L(T_r)$$

Thus by direct differentiation

$$\frac{d(k\hat{T})}{d(kT_r)} = L\{p_L(T_r)\} + (kT_r(-p_{L-1}(T_r)) + E_L(T_r)$$

$$= L\, p_L(T_r) - kT_r p_{L-1}(T_r) + E_L(T_r)$$

Since

$$L\, p_L(T_r) = L \cdot \frac{(kT_r)^L}{L!}\, e^{-kT_r}$$

$$= (kT_r)\, \frac{(kT_r)^{L-1}}{(L-1)!}\, e^{-kT_r}$$

$$= kT_r p_{L-1}(T_r)$$

this reduces to

$$\frac{d(k\hat{T})}{d(kT_r)} = E_L(T_r)$$

as required.

Given the above statistics for replacement, on-line failure, and expected time to failure of a component with estimated lifetime, L and assumed replacement time T_r we desire to choose T_r (given L) which minimizes some appropriate cost function. Intuitively, this cost function should represent both the cost of on-line failure and the cost of wasted component lifetime due to replacing components be-

fore failure.[5,17] We, therefore, adopt the cost functional

$$\text{Cost} = C_f P_f + C_W(kT^* - k\hat{T})$$

Here, C_f and C_W are appropriate weight factors representing the cost of a on-line failure and the cost of component lifetime wastage, respectively. Thus, the first term in the cost functional represents the expected cost of a failure (i.e. the probability of an on-line failure times the cost of such a failure) whereas, the second term in the cost functional represents the expected cost of wasted component lifetime (i.e. the expected lifetime reduction times the cost per unit time for such a lifetime reduction with k serving as a normalizing factor).

To minimize the cost functional of equation 33. one simply substitutes the values for $P_f(T_r)$, T^*, and $\hat{T}(T_r)$ computed in the preceeding pages, defferentiating the Cost with respect to kT_r and setting it equal to zero. This then results in the equality[9]

12. $$0 = C_f P_{L-1}(T_r) - C_W E_L(T_r)$$

where $d(P_f)/d(kT_r)$ is given by Prop 9 and $d(k\hat{T})/d(kT_r)$ is given by Prop 10. Thus the choice of an optimal T_r given L is reduced to the solution of a single nonlinear equation in one unknown. The solutions of this equation are plotted in Figure 1. for a number of values of L and C_f/C_W. Indeed, it can be readily shown that equation 12. has exactly one solution for $T_r > 0$. Moreover, the function

$$R_L(t) = C_f P_{L-1}(t) - C_W E_L(t)$$

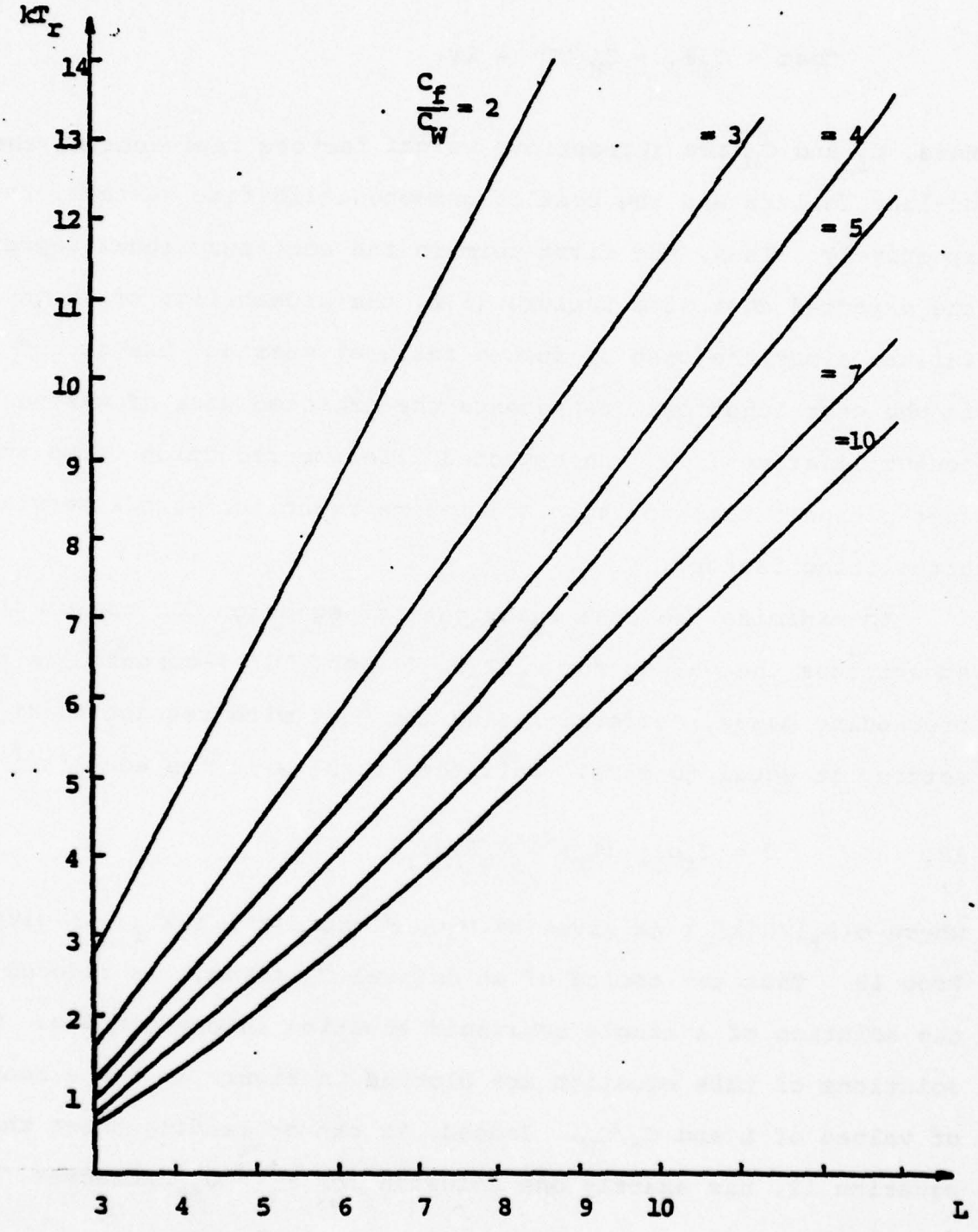


Figure 1. Replacement time (kT_r) v.s. Lifetime L with different weight constant

takes on negative values for $0 < t < T_r$ and positive values for $T_r < t$ hence in an on-line maintenance system one need not even solve equation 12. Rather, one simply evaluates $R_L(t)$ at the time of the next scheduled maintenance. If this results in a negative number, the next scheduled maintenance preceeds the optimal replacement time and hence, we should wait, at least, until the next scheduled maintenance (when we will have more data) to replace the component. On the other hand, if the evaluation of $R_L(t)$ at the next scheduled maintenance time results in a positive value, the optimal replacement time will have passed by the next scheduled maintenance and hence, the component should be replaced at the present maintenance interval.

Summarizing the on-line maintenance algorithm resulting from the above theory takes the following form. At the gth scheduled maintenance interval (at time gT) one measures the component parameter, C_g. If C_g is already out of tolerence, the component is simply replaced and no further analysis is required. If, however, C_g is in tolerence ($C_g > 0$ in our notation) it is used together with the values of the component parameter measured at the previous maintenance intervals to estimate the dynamics of the failure model for the component. Here, sequential refinement schemes may be used both to include the effect of C_g on the estimates made at the (g-1)st maintenance interval and to increase the order of the polynomial used to represent the component failure dynamics. Once the component failure dynamics have been satisfactorily estimated, one solves 6. to estimate whether or not to replace the component. If $R_L((g+1)T)$ 0 the component is replaced, whereas, if

$R_L((g+1)T) < 0$ the component is not replaced, and the system is returned to service until the next scheduled maintenance.

SIMULATIONS

A computer simulation of an on-line periodic maintenance based on the above described algorithm was performed for 600 maintenance intervals with a new component replacing the old component after each replacement decision and/or on-line failure.[9] The system was subjected to computer generated Poisson shocks with constant k = 0.1 shocks per hour and a maintenance interval of T = 20 hours. The simulation was first run using identical components with L = 28 (expected lifetime of 14 maintenance intervals) and then repeated using random components and noisy data measurements.

For the case where identical components were employed, Table 1. gives the total number of replacements and failures resulting from the application of the algorithm over the 600 simulated maintenance intervals with different values of C_f/C_W. For comparison, Table 2. shows the total number of replacements and failures which would have resulted from a fixed replacement strategy ranging from 6 to 12 maintenance intervals. Since the cost function is

$$\text{Cost} = C_f P_f + C_W (kT^* - k\hat{T})$$

the overall cost can be expressed as

$$\text{Cost} = \frac{C_f}{C_W} \quad (\text{No. of failures}) + 0.1 \quad (280^*(\text{No. of components used}) - 12000)$$

The overall costs resulting from the application of our algorithm

and the various fixed replacement schedules may be computed from the data in Tables 1 and 2. The resultant costs for different values of C_f/C_w are given in Table 3.

Note, since L = 28 for each component in this simulation, an optimal replacement strategy of approximately 10 maintenance intervals can be computed from equation 12. without estimating L. As such, it is not surprising that this fixed replacement strategy resulted in lower costs than the algorithm. It should, however, be noted that the algorithm did not have the advantage of a a-priori knowledge of L and yet it sill out-performed all of the fixed replacement strategies except the optimal strategy (that is, optimal once L is known).

In our second simulation, random noise was added to the data to simulate both the effects of imperfect measurements and the effect components with random failure characteristics. Various simulations were run, as before, for 600 maintenance intervals each with k = 0.1 and T = 20, with noise levels ranging between 20 and 60 percent of the tolerence interval. The results of these simulations are given in Tables 4. and 5. Except for a single case, which we believe to be anomolous, the algorithm out-performed any fixed replacement strategy.

CONCLUSION

In the preceeding, we have described a curve fitting algorithm for the prediction of failures in analog devices. The algorithm was tested in a variety of situations and found to be surprisingly effective in predicting failures with relatively little wastage of component lifetime and on-line failure cost.

C_f/C_W	No. of replacement	No. of failure
50	48	7
75	56	1
100	52	2
150	54	2
200	54	2

TABLE 1

Total replacements and failures within 600 maintenance intervals for different C_f/C_W

Constant replacement time	No. of replacement	No. of failure
every 6 intervals	100	0
every 7 intervals	85	0
every 8 intervals	75	0
every 9 intervals	65	1
every 10 intervals	59	1
every 11 intervals	48	6
every 12 intervals	39	11

TABLE 2

Total replacements and failures within 600 maintenance intervals for various fixed replacement strategies

Cost / C_f/C_w Methods	50	75	100	150	200
every 6 intervals	1600	1600	1600	1600	1600
every 7 intervals	1096	1096	1096	1096	1096
every 8 intervals	900	900	900	900	900
every 9 intervals	698	723	748	798	848
every 10 intervals	530	555	580	630	680
every 11 intervals	612	762	912	1212	1512
every 12 intervals	750	1025	1300	1850	2400
the algorithm	690	471	512	668	768

TABLE 3

Overall cost with different mathods and different C_f/C_w .

noise level / method	20 %		30 %		40 %		60 %	
	No. of replace	No.of fail	No. of replace	No.of fail	No. of replace	No.of fail	No. of replace	No.of fail
every 6 intervals	100	0	100	0	100	0	94	6
every 7 intervals	85	0	85	0	84	1	78	8
every 8 intervals	75	0	72	3	71	4	64	12
every 9 intervals	64	2	63	3	60	7	52	17
every 10 intervals	56	4	51	9	45	15	45	18
every 11 intervals	45	10	45	10	45	10	39	20
every 12 intervals	36	15	35	16	36	17	31	23
the algorithm	56	3	55	5	55	5	50	11

TABLE 4

Total replacements and failures within 600 maintenance intervals for various fixed replacement strategies and the algorithm at different noise levels.

cost \ noise level / method	0 %	20 %	30 %	40 %	60 %
every 6 intervals	1600	1600	1600	1600	2200
every 7 intervals	1096	1096	1096	1280	2008
every 8 intervals	900	900	1200	1300	2128
every 9 intervals	748	848	948	1376	2432
every 10 intervals	580	880	1380	1980	2364
every 11 intervals	912	1340	1340	1340	2452
every 12 intervals	1300	1728	1828	1984	2612
the algorithm	512	752	980	752	1608

TABLE 5

Overall cost for different methods at different noise levels

REFERENCES

1. Saeks, R., and S.R. Liberty, Rational Fault Analysis, New York, Marcel Dekker, Inc., 1977.

2. R. Saeks, "An Approach to Built-In Testing", Unpublished Notes, Texas Tech University, Lubbock, Texas, 1977.

3. L. Tung, M.S. Thesis, Texas Tech University, Lubbock, Texas, 1975.

4. Tung, L., and R. Saeks, "An Experiment in Fault Prediction", Proc. of the 4th Inter. Symp. on Reliability in Electronics, Budapest, October 1977, pp. 249-257.

5. Tung, L., Liberty, S.R., and R. Saeks, "Fault Prediction Towards a Methematical Theory", in Rational Fault Analysis, New York, Marcel Dekker, pp. 135-142, 1977.

6. Esary, J.D., Marshall, A.W., and F. Proschan, Shock Models and Wear Processes, Ann. Probability, Vol. 1, 1973.

7. J.D. Esary, Ph.D. Dissertation, Univ. of California, Berkeley, 1957.

8. A.W. Marshall, "Some Comments on the Hazard Gradient", J. Stoch. Proc. Applic., Vol. 3, 1975.

9. K.S. Lu, Ph.D. Thesis, Texas Tech University, Lubbock, TX, 1977.

10. A. Papoulis, Probability, Random Variable, and Stochastic Processes, McGraw-Hill Book Comp., New York, 1965.

11. Norman I. Johnson, Discrete Distribution, John Wiley and Sons Inc., New York,

12. Rao, C.R., and S.K. Mitra, Generalized Inverse of Matrices and Its Applications, New York, John Wiley and Sons, 1971.

13. K.S. Lu, MSEE Thesis, Texas Tech University, Lubbock, TX, 1973.

REVIEW OF RESEARCH

in

ELECTRONICS AND RELATED AREAS

RESEARCH FACULTY

in

ELECTRONICS AND RELATED AREAS

SYSTEMS

Name	Title	Area	Office	Phone
R.B. Asher	Assist. Prof. EE	Detection and Estim.	150B-EE	742-3537
K.S. Chao	Assoc. Prof. EE	Nonlinear Circuits and Systems	150A-EE	742-3469
D.L. Gustafson	Assoc. Prof. EE	Microprocessors	150B-EE	742-3530
M.O. Hagler	Prof. EE	Optical Signal Proc.	103B-EE	742-3470
L.R. Hunt	Prof. Math	Math. Syst. Theory	265-Math	742-2566
S.R. Liberty	Assoc. Prof. EE & Stat.	Stochastic Control and Estimation	201B-EE	742-3441
J. Murray	Assist. Prof. EE	Multi-dimensional Digital Filters	258-EE	742-3528
T. Newman	Assoc. Prof. Math & C.S.	Pattern Recognition	1107-BA	742-2571
J. Prabhakar	Assoc. Prof. EE	Communications	104-EE	742-3506
R. Saeks	Prof.-EE and Math	Circuits and Systems	258A-EE	742-3528
L. Tung	Lect.-EE	Math. Systems Theory	258A-EE	742-3528
D.L. Vines	Prof.-EE	Computers	151-EE	742-3536
J. Walkup	Assoc. Prof. EE	Optical Systems & Communications	260B-EE	742-3500

PHYSICAL ELECTRONICS

M. Gundersen	Assoc. Prof. EE	Quantum Electronics	260A-EE	742-3501
A. Kwatra	Lect.-EE	Optics & Quantum Electronics	205-EE	742-3502

W. Portnoy	Prof.-EE	Solid State & Bio-Med. Elec.	152-EE	742-3532
J. Reichert	Assoc. Prof. EE	Optics & Quantum Electronics	203-EE	742-3502
F. Williams	Asst. Prof. EE.	Interaction of Light with Matter	258B-EE	742-3501

ELECTROMAGNETICS

M. Hagler	Prof.-EE	Plasma	103B-EE	742-3470
M. Kristiansen	Horn Prof. EE	Plasma	103A-EE	742-3468
E. Kunhardt	Asst. Prof. EE	Nonlinear Phenomena	260C-EE	742-3545
T. Trost	Assoc. Prof. EE	Antennas & Propagation	102-EE	742-3505

POWER

T. Burkes	Assoc. Prof. EE	Power Conditioning	105C-EE	742-3533
J. Craig	Prof.-EE	Electro-Mech. Devices	101-EE	742-3529
M. Kristiansen	Horn Prof. EE	High Power Switching	103A-EE	742-3468
E. Kunhardt	Asst. Prof. EE	High Power Switching	260C-EE	742-3545
S. Liberty	Assoc. Prof. EE	Solar Energy	201B-EE	742-3441
J. Reichert	Assoc. Prof. EE	Solar Energy	203-EE	742-3502

ACTIVE GRANTS AND CONTRACTS

in

ELECTRONICS AND RELATED AREAS

Systems

Principal Invest.	Agency	Title	Duration	Annual Funding
Saeks	AFOSR	Resolution Space...	1 yr.	19,257
Liberty	ONR	Statistical Performance Analysis...	1 yr.	30,000
Chao	NSF	Continuation Methods	2 yrs.	18,710
Walkup/Hagler	AFOSR	Space-Variant Optical Systems	1 yr.	75,440
Saeks/Levan	AFOSR	Symp. on Oper. Thy. of Networks and Systems	1 yr.	4,468
Asher	AFWL	Estimation in Adaptive Optics	1 yr.	44,330
Gustafson	RADC	Phased Array Antennea Analysis	1 yr.	5,000
Asher	SORF	Nonlinear Estimation and Detection	1 yr.	5,000
Saeks	ONR	Assoc. Joint Services Electronics Prog.	1½ yrs.	173,500
		Total Annual Funding in Systems		$375,705

Physical Electronics

Gundersen	SORF	Studies in Transient Discharges	1 yr.	8,000
Gundersen	ERDA	Laser Research	1 yr.	50,000
Gundersen	NSF	Inovative Infrared Detector	2 yrs.	17,500
Reichert	AFOSR	Analysis of Unstable Optical Resonators	1 yr.	45,219

Portnoy	NSF	Semiconductor Device...	1 yr.	13,800
Williams/ Gundersen	AFOSR	Studies in Transient Discharges	1 yr.	49,995
Williams	Research Corp.	Driven Raman Proc.	1 yr.	10,000
Portnoy	DOE	Switching Device Anal.	1 yr.	9,962
Portnoy	IEEE	Microelectronics Symp.	1 yr.	1,000
Portnoy	SORF	Semiconductor Device...	1 yr.	6,283

Total Annual Funding in Physical Electronics $211,759

Electromagnetics

Kristiansen/ Hagler	NSF	RF Plasma Heating	1 yr.	37,044
Kristiansen	AFOSR	Dense Plasma Heating and Rad. Gen.	1 yr.	99,222
Kristiansen	DOE	Small Toroidal Plasma	1 yr.	2,500
Kristiansen	EPRI	High Power RF Heating	1 yr.	2,525

Total Annual Funding in Electromagnetics $141,291

Power

Craig	TPL	Power System Studies	1 yr.	8,000
Kristiansen	DOE	Laser Triggered Spark Gap	1 yr.	50,088
Burkes	DOE	E Beam Laser Support	1 yr.	15,250
Kristiansen	AFOSR	High Power Switch Dev.	1 yr.	50,382
Burkes	AFWL	Analysis of Pulse Power	1 yr.	63,000
Kristiansen	DOE	Surface Flashover Mech.	1 yr.	59,459

Burkes	NSWC	High Power Switches	1 yr.	39,841
Reichert	ERDA	Crosbyton Solar Power Project	2 yrs.	600,000*
		Total Annual Funding in Power		$886,020

Other Areas

Seacat	SORF	Research and Development	1 yr.	16,775
Kuhnhardt	NSF	Undergraduate Research Part.	1 yr.	17,980
		Total Annual Funding in Other Areas		$34,755
		Total Annual Funding		$1,649,530

*The Department of Electrical Engineering is the prime contractor on the Crosbyton Solar Power Project which is funded at about $1,500,000 annually. Of this amount about $600,000 is spent in the department with the remainder spent in other departments at Texas Tech and/or subcontracted.

RESEARCH LABORATORIES

in

ELECTRONICS AND RELATED AREAS

SYSTEMS

Computer Laboratories:

CDC 1604 facility: hands-on facility for both education and research.....108-EE

CDC 3600 facility: under development.....162-EE

Hybrid Computer facility: minis, micros, and analog facilities.....157-EE

Bio-medical Systems: includes instrumentation and microprocessor application facilities.....215-EE

Circuits and Systems Laboratory: the think tank.....258-EE

Optical Systems Laboratories:

Holographic Optics: primarily used for multiplex holography research.....110 and 103-EE

Optical Signal Processing: research in optical and digital image processing.....216-EE

PHYSICAL ELECTRONICS

Laser Laboratory: infrared laser research.....262-EE

Integrated Circuit Laboratory: fabrication facility for SSI and special purpose devices.....209-EE

Laser Laboratory: interaction of light with matter.....260-EE

ELECTROMAGNETICS

Plasma Laboratories:

Laser/Plasma facility: plasma heating via laser plasma interaction.....113-EE

Tokamak facility: radio frequency heating of toroidal plasmas.....117-EE

Electromagnetics Laboratory: nonlinear wave studies.....111-EE

Antenna Laboratory: radio meteorology and ionspheric studies.....West of the Medical School

POWER

High Voltage Laboratory: pulsed power studies.....North of Textile Building

Solar Energy Laboratory: another think tank.....205-EE

High Power Switching Laboratory: electron beam initiated spark gap.....105-EE

PUBLICATION ACTIVITY

in

ELECTRONICS AND RELATED AREAS

Systems - Refereed Publications

Saeks, R., and S.R. Liberty, Rational Fault Analysis, New York, Marcel Dekker, 1977.

Ransom, M.N., and R. Saeks, "A Functional Approach to Fault Analysis of Linear Systems" in Rational Fault Analysis (ed. R. Saeks and S.R. Liberty), New York, Marcel Dekker Inc., 1977, pp. 124-134.

Liberty, S.R., Tung, L., and R. Saeks, "Fault Prediction - Toward a Mathematical Theory" in Rational Fault Analysis (ed. R. Saeks, and S.R. Liberty), New York, Marcel Dekker, 1977, pp. 135-142.

DeCarlo, R.A., Murray, J., and R. Saeks, "Multivariable Nyquist Theory", Int. Jour. on Cont., Vol. 25, pp. 657-675, (1977).

Chao, K.S., and R. Saeks, "Continuation Methods in Circuit Analysis", IEEE Proc., Vol. 65, pp. 1187-1194, (1977).

DeCarlo, R.A., and R. Saeks, "The Encirclement Condition: An Approach Using Algebraic Topology", Int. Jour. on Cont., Vol. 26, pp. 279-287, (1977).

Decarlo, R.A., Saeks, R., and J. Murray, "A Nyquist-like Test for the Stability of Two-Dimensional Digital Filters", IEEE Proc., Vol. 65, pp. 978-979, (1977).

Asher, R., "Adaptive Estimation of Aberration Coefieients in Adaptive Optics", Information Sciences, Vol. 12, pp. 245-261, (1977).

Asher, R., "Perfect Decoupling of Linear Systems with Discrete Parameter Uncertainties", IEEE Trans. on Automatic Cont., Vol. AC-22, pp. 498-500, (1977).

Marks, R.J., Walkup, J., Hagler, M.O., and T.F. Krile, "Space Variant Processing of 1-D Signals", Appl. Optics, Vol. 16, pp. 739-745, (1977).

Hagler, M.O., Krile, T.F., Marks, R., and J. Walkup, "Holographic Representation of Space-Variant Systems Using Phase Coded Reference Beams", Appl. Optics, Vol. 16, (1977).

Portnoy, W.M., Emergency Medical Care, Lexington, Mass., Lexington Books, 1977.

Marks, R., Walkup, J., and T.F. Krile, "Ambiguity Function Display: An Improved Coherent Processor", Appl. Optics, Vol. 16, pp. 746-750, (1977).

Systems - Conference Papers and Reports

Saeks, R . and K.S. Chao, Proc. of the 20th Midwest Symposium on Circuits and Systems, North Hollywood, Western Periodicals Inc, 1977.

Saeks, R., and N. Levan, Proc. of the 2nd Int. Symp. on the Operator Theory of Networks and Systems, North Hollywood, Western Periodicals, Inc., 1977.

DeCarlo, R.A., Murray, J., and R. Saeks, "Three Graphical Tests for the Stability of Multidimensional Digital Filters", Proc. of the 1977 IEEE Int. Symp. on Circuits and Systems, Pheonix, 1977.

Leake, R.J., and R. Saeks, "On the Computability of Lacunary Sets", Proc. of the 1977 IEEE Int. Symp. on Circuits and Systems, Pheonix, 1977.

Sen, N., and R. Saeks, "A Measure of Testability and its Application to Test Point Selection - Computation", Proc. of the AUTOTESTCON '77, Hyannis, Mass., 1977.

Tavenner, D., Saeks R., and D. Gustafson, "Microprocessor Implementation of a Fault Prediction Algorithm", Proc. of the 20th Midwest Symp. on Circuits and Systems, Lubbock, Tx., 1977.

Sen, N., and R. Saeks, "A Measure of Testability and its Application to Test Point Selection - Theory", Proc. of the 20th Midwest Symp. on Circuits and Systems, Lubbock, TX, 1977.

Tung, L., and R. Saeks, "An Experiment in Fault Prediction", Proc. of the 4th Symp. on Reliability in Electronics, Budapest, 1977.

Tung, L. and R. Saeks, "Wiener-Hopf Techniques in Resolution Space", Proc. of the 2nd Int. Symp. on the Operator Theory of Networks and Systems, Lubbock, Tx., 1977.

Asher, R., "Adaptive Estimation of Aberation Coeficients in Adaptive Optics", Proc. of the 1977 Joint Automatics Control Conf., San Francisco, 1977.

Asher, R., "Optimal and Sub-optimal Results in Full and Reduced Order Linear Filtering", Proc. of the 1977 Joint Automatic Control Conf., San Francisco, 1977.

Asher, R.B., "Adaptive Estimation of Samll Angle Measurements", Proc. of the AIAA Guidance and Cont. Conf., San Diego, 1977.

Asher, R., "Mixed Observable Estimation of Random Solar Electric Propulsion Spacecraft Thrusts", Proc. of the AIAA Guidance and Cont. Conf., San Diego, 1977.

Asher, R., "Fixed Configuration Filters for Image Processing", Proc. of the 1977 Int. Conf. on Cybernetics and Soc., Anahein, Ca., 1977.

Asher, R., "Optical Estimation in Signal Dependent Noise", Proc. of the Optical Soc. of Amer. Annual Meeting, San Diego, 1977.

Asher, R., "Adaptive Estimation of Phase Distortions in Adaptive Optics", Proc. of the Optical Soc. of Amer. Annual Meeting, San Diego, 1977.

Asher, R., "High Angle Attack Flight Cont. Using Stochastic Reference Model Adaptive Control", IEEE Decision and Cont. Conf., Orlando, Fla., 1977.

Asher, R., "Optimal and Sub-optimal Estimation of Mixed Rotational Observables", IEEE Decision and Cont. Conf., Orlando, Fla., 1977.

Chao, K.S., and D.K. Liu, "Transfer Characteristic Plots and Large-Cange Sensitivity of Nonlinear Resistive Networks", Proc. of the 1977 IEEE Int. Symp. on Circuits and Systems, Pheonix, 1977.

Pan, C.T. and K.S. Chao, "Difference Equation Approach for Solving Nonlinear Equations", Proc. of the 20th Midwest Symp. on Circuits and Systems, Lubbock, Tx, 1977.

Gustafson, D.L., "Review of Hughs' Mathematical Description of Linear Systems", Circuits and Systems, Vol. 10, p. 5, 1977.

Gustafson, D., "A Series of Microprocessor Courses", Digest of the Conf. of the ASEE Gulf-Southwest Section, Oklahoma City, 1977.

Prabhakar, J.C., "Synchronization Recovery", Proc. of the NATO Advanced Studies Workshop, London, 1977.

Prabhakar, J.C., "Walsh Transforms - Another Look", Proc. of the 20th Midwest Symp. on Circuits and Systems, Lubbock, TX., 1977.

Marks, R., Walkup, J., and M. Hagler, "Sampling Theorems for Linear Space-Variant Systems", Proc. of the 20th Midwest Symp. on Circuits and Systems, Lubbock, Tx., 1977.

Krile, T.F., Hagler, M., Marks, R., and J. Walkup, "Space Variant Holographic Optics Using Phase Coded Reference Beams", Proc. of the 1977 Int. Optical Comp. Conf., San Diego, 1977.

Bell, S.J., and M. Hagler, "Nyquist Stability Conditions for Laser Operational Amplifiers with Gaussian and Lorentizian Gain Properties", Proc. of the 20th Midwest Symp. on Circuits and Systems, Lubbock, Tx., 1977.

Hagler, M., Krile, T.F., Redis, W.D., and M.I. Jones, "Families of 2-Dimensional Phase-Coded Diffusers for Multiplex Holography", Proc. of the 1977 Meeting of the Optical Soc. of Amer., Toronto, 1977.

Froehlich, G., and J. Walkup, "Some Results on Detection and Estimation in Signal Dependent Noise", Proc. of the 20th Midwest Symp. on Circuits and Systems, Lubbock, Tx., 1977.

Marks, R., and J. Walkup, "Coherent Optical Processor for Ambiguity Function Display and One Dimensional Corrolation/Convolution Operations", SPIE Cong. on Optics in Radar, 1977.

Walkup, J., Froehlich, G., and R. Asher, "Optimal Estimation in Signal Dependent Noise", Proc. of the 1977 Meeting of the Optical Soc. of Amer., Toronto, 1977.

Walkup, J., Marks, R., and M. Irby, "Techniques in One Dimensional Space Variant Processing", Proc. of the 1977 Meeting of the Optical Soc. of Amer., 1977.

Systems - in Press

Gustafson, D., "Review of Peatman's Microcomputer Based Design", IEEE Spectrum, (to appear).

Hagler, M., Froehlich, G., and J. Walkup, "A Set of Optical Information Processing Experiments", IEEE Trans. on Education (to appear).

Marks, R., Walkup, J., and M. Hagler, "Sampling Theorems for Linear Time-Variant Systems", IEEE Trans. on Circuits and Systems, (to appear).

Physical Electronics - Refereed Publications

Williams, P.F., "Experimental Measurement of Dissociative Molecular Potential Functions from Continuous Resonance Raman Spectra", Chem. Physics Lett., Vol. 47, pp. 150-151, (1977).

Williams, P.F., "Evidence for Excited State Interference in Resonance Raman Scattering", Chem. Physics Lett., Vol. 50, pp. 57-59, (1977).

Bushnell, A.H., Gundersen, M., and T.R. Burkes, "Effect of a Small Capacitor in Parallel with a Pulsed CO_2 TEA Laser", IEEE Jour. on Quantum Elec., Vol. QE-12, pp. 447-449, (1977)

Gundersen, M., "Spectroscopy with Visable and Far-Infrared Lasers", Proc. of SPIE, Vol. 82, pp. 7-12, (1976).

Reichert, J.D., "Refraction", in Encyclopedia of Physics, Stroudsburg, Pa., Dowden, Hutchinson, and Ross, Inc., 1977.

Physical Electronics - Conference Papers and Reports

Gundersen, M., "Line Narrowing of CO_2 Lasers Required for Optical Pumping", Report AP-2-77-221, Los Alomos Scientific Laboratory, 1977.

Jones, C.R., Bushnell, A.H. and M. Gundersen, "An Optically Pumped $^{15}NH_3$ Laser", Proc. of the Conf. on Chemical and Molecular Lasers", St. Louis, 1977.

Gundersen, M., "Laser Research", Report of the Conf. on Lasers for Isotope Separation", Albuquerque, 1977.

Gundersen, M., "Spectroscopy with Visible and Far-Infrared Lasers", Proc. of the SPIE Conf., 1976.

Physical Electronics - in Press

Williams, P.F., and D. Russou, "The Resonance Raman Effect", in Topics in Appl. Physics, (to appear).

Jones, C.R., Buchwald, M.I., Gundersen, M., and A.H. Bushnell, "Ammonia Laser Optically Pumped with an HF Laser", Optics Communications, (to appear).

Electromagnetics - Refereed Publications

Hagler, M., and M. Kristiansen, "A Numerical Analysis of High Power Laser Propagation in Magnetized Plasma", in Recent Advances in Plasma Physics, Madras, Indian Acad. of Sci., 1977.

Kunhardt, E., "Propagation of Nonlinear Waves Along Magneto Plasma Columns", Phys. of Fluids, Vol. 20, pp. 1499-1508, (1977).

Kunhardt, E., "Rotating Ball Generator", Rev. Sci. Inst., Dec. 1977.

Electromagnetics - Conference Papers and Reports

Hagler, M., and M. Kristiansen, "Ablation Rates of Sphereical Pellets in a Theta Pinch", IEEE Int. Conf. on Plasma Science, Troy, N.Y., 1977.

Kristiansen, M., and M. Hagler, "An Experimental Arrangement for Laser Beat Heating of Plasma", Proc. of the Engrg. Problems of Fusion Research Conf., Knoxville, 1977.

Hagler, M., and M. Kristiansen, "Description of the Texas Tech Tokomak", APS Plasma Physics Div. Meeting, Atlanta, 1977.

Trost, T., "Radio and Electric Fields Change Measurements on Severe Thunderstorms", Digest of the Amer. Geophysical Union Meeting, 1977.

Electromagnetics - in Press

Kristiansen, M., and M. Hagler, "Ablation Rates for Polystyrene Microspheres in a Theta Pinch Plasma", Jour. of Appl. Phys, (to appear).

Power - Refereed Publications

Kristiansen, M., and M. Hagler, An Introduction to Controlled Thermonuclear Fusion, Lexington Press, Lexington, Mass., 1977.

Power - Conference Papers and Reports

Burkes, T.R. and J.P. Craig, "Power System Analysis", AFWL Tech. Report, Kirtland AFB, N.M., 1977.

Craig, J.P., and S. Tsai, "A Variable Speed Wind Generator", Texas Power and Light Co., Tech Report, Dallas, Tx., 1977.

Craig, J.P., "Evaluation of Projected Effectiveness of the National Energy Plan", Texas Tech Univ. Center for Energy Research Tech. Report, Lubbock, Tx., 1977.

Trost, T., "Pulse Power System Employing Magnetic Energy Storage", Naval Surface Weapons Center Tech. Report, Dahlgren, Va., 1977.

Reichert, J.D., "A Strategy of Calculation of Optical Concentration Distributions for Fixed Mirror Systems", Proc. of the ERDA Solar Energy Workshop on Methods for Optical Mirror Systems, Livermore, Calif., 1977.

Reichert, J.D., "The Crosbyton Solor Power Project: Fixed Sphereical Mirror/Tracking Receiver", Proc. of the ERDA Concentrating Collector Conf., Washington, D.C., 1977,