Optimizing Multihypothesis Diagnosis of Control-Actuator Failures in Linear Systems

Yakov Ben-Haim
Technion—Israel Institute of Technology, Haifa, Israel

This paper describes a method for designing a maximum-likelihood multihypothesis algorithm for diagnosing control-actuator failures in linear systems. Uncertainty in the temporal behavior of a malfunctioning actuator is represented by employing the set theoretic technique called "convex modeling" whereby each class of malfunctions is represented by a convex set of vector-valued failure functions. The hypothesized malfunctions upon which the diagnosis is based are chosen so that specified sets of failures are correctly diagnosed. It is shown that, in some cases, the selection of a robust and efficient set of hypothesized malfunctions can be based on the solution of linear optimization problems. This is applied to the steady open-loop flight of an AFTI/F16 aircraft. These results are then extended to the case where the malfunctioning actuators are incorporated in a feedback loop. In this case nonlinear equations are formulated and some properties studied.

I. Introduction

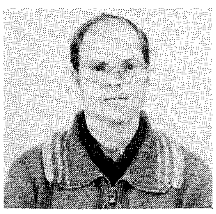
A MULTITUDE of concepts have been employed for the diagnosis of malfunctions in linear systems. A common approach is based on hypothesizing a set of possible malfunctions and then subjecting measurements of the system to a maximum likelihood test to decide which hypothesized malfunction is most likely to have given rise to the measurements.[1-8] This approach is appealing for several reasons. The concept of maximum likelihood is intuitively satisfying as a criterion of optimality. It has been asserted, however, that the likelihood ratio test is not a uniformly most powerful test when the failure occurs as a random event at an unknown or random time.[9] In addition, prior information about the system can be exploited by judicious selection of the hypothesized malfunctions.

The performance of a multihypothesis algorithm for malfunction diagnosis is limited by the disparity between its finite set of hypothesized malfunctions and the infinity of possible failures. A large number of hypothesized malfunctions are usually deemed necessary for reliable diagnosis in the presence of the substantial uncertainty that accompanies the occurrence of failures. However, real-time implementation of a multihypothesis algorithm of high multiplicity is problematical. The aim of this paper is to propose a plausible representation for the uncertainty associated with malfunction and to develop a method for evaluating the performance of a multihypothesis algorithm with respect to this uncertainty. This performance evaluation forms the basis for selecting a robust and efficient collection of hypothesized malfunctions.

II. Maximum Likelihood Malfunction Diagnosis

In this section, we state the maximum-likelihood multihypothesis approach to diagnosing additive failures in linear dynamic systems and formulate the problem to be studied. Let $f(t)$ be a vector function representing a specific control-actuator malfunction, and let $y_f(t)$ represent the average measured system response to $f(t)$. Because the system is linear and the failure is additive, $y_f(t)$ is an affine transformation of $f(t)$. (The specific form that $y_f(t)$ assumes for control actuator failure will be discussed later.) Throughout the paper we let E^L represent a Euclidean space of dimension L to which measurement vectors y belong. Let $p(y|f)$ be the conditional probability density of the system response given a malfunction f. We shall assume that $p(y|f)$ decreases monotonically with a norm of $y - y_f$. This requirement is fulfilled, for example, if $p(y|f)$ is a multivariate Gaussian density and if the square of the norm of y is $y^T V_f^{-1} y$, where V_f is the covariance matrix of y given malfunction f. The superscript T implies matrix transposition. Different norms can be defined with respect to different malfunctions, for example, if the covariance matrix depends on the malfunction. We denote the various norms as follows. An inner product of elements x and y in E^L, with respect to the malfunction f, is denoted $[x,y]_f$. Our only assumption regarding this inner product is that $[x,x]_f^{1/2}$ is a norm, which will be denoted $\|x\|_f$.

Many distinct classes of actuator failures can occur: single or multiple failures, locked surfaces, or widely varying surface deflections. In an important class of malfunctions, the affected control surfaces fail to trail the control commands.



Yakov Ben-Haim received his Ph.D. degree from the University of California at Berkeley in 1978. Since then he has been at the Technion—Israel Institute of Technology, where he is currently an Associate Professor in the mechanical engineering faculty. His research is in the field of optimization in the presence of complex sources of uncertainty. He has studied problems in malfunction control, applied mechanics, and nuclear radiation measurements. Professor Ben-Haim has published two books: *The Assay of Spatially Random Material* (Kluwer, 1985) and, with I. Elishakoff, *Convex Models of Uncertainty in Applied Mechanics* (Elsevier, 1990).

Received Oct. 20, 1988; revision received Feb. 7, 1989.

Instead, these control surfaces deflect autonomously. The failure vectors $f(t)$ are assumed to belong to a set of uniformly bounded but otherwise freely varying functions. The failure sets are defined as

$$F(p)=\Big\{f=(f_1,\ldots,f_M):\quad \tilde{p}_m\le f_m(t)\le \hat{p}_m \quad t\in[0,\infty),\quad m=1,\ldots,M\Big\} \tag{1}$$

where $p=(\tilde{p}_1,\hat{p}_1,\ldots,\tilde{p}_M,\hat{p}_M)$. Thus the autonomous value of the mth control function varies arbitrarily in time between $\tilde{p}_m$ and $\hat{p}_m$. Usually the number of actuator failures is less than the dimension of the control vector. This is represented by choosing $\tilde{p}_m=\hat{p}_m=0$ for each of the functioning actuators. The $F(p)$ will be referred to as the *failure set* for malfunctions of type p. The set $F(p)$ is convex. Convex sets have been used for modeling uncertain phenomena in a wide range of engineering applications.[10-13]

Let $F(p^1),\ldots,F(p^K)$ be disjoint failure sets and let H_k be a finite collection of malfunctions chosen from $F(p^k)$, for $k=1,\ldots,K$. Let

$$H=\bigcup_{k=1}^{K}H_k$$

A maximum likelihood multihypothesis algorithm for malfunction diagnosis is based on the collection H of vector functions representing hypothesized malfunctions. Having obtained a measurement y, the algorithm seeks a hypothesized malfunction $h_{ml}\in H$, which satisfies

$$\|y_{h_{ml}}-y\|^2_{h_{ml}}=\min_{h\in H}\|y_h-y\|^2_h \tag{2}$$

The function h_{ml} is most likely to be the system condition that caused the measurement y because $p(y\,|\,h)$ decreases monotonically with $\|y-y_h\|_h$.

Given failure sets $F(p^1),\ldots,F(p^K)$, and given sets of hypothesized malfunctions $H_1,\ldots,H_K$, we will say that failures of type-p^k are *correctly* diagnosed if every failure in $F(p^k)$ is ascribed by the multihypothesis algorithm to a hypothesized failure in H_k. A collection

$$H=\bigcup_{k=1}^{K}H_k$$

of malfunction hypotheses is robust if the failure sets $F(p^1),\ldots,F(p^K)$ are correctly diagnosed. A robust collection H of malfunction hypotheses is efficient if no smaller set of hypotheses is robust. The problem to be studied here is the development of a computationally feasible method for determining whether or not a given set of hypothesized malfunctions is robust. This determination forms the basis for searching for an efficient collection of hypotheses.

An important simplification occurs when the norms $\|\cdot\|_{h_k}$ are the same for all hypothesized malfunctions. An example is developed in Sec. IV for actuator failures in an open-loop linear system.

III. Representing Uniformly Bounded Control-Actuator Failures

Our aim in this section is to develop a convenient formalism for representing the measurements of a closed-loop linear system with uniformly bounded control-actuator failure. The main result of this section is Eq. (13), which is an expression for the complete response set.

Consider the failure-free dynamic system

$$\frac{dx}{dt}=Ax(t)+Bu(t)+\nu_1(t) \tag{3}$$

$$y(t)=Gx(t)+\nu_2(t) \tag{4}$$

$$u(t)=S(t)x(t) \tag{5}$$

where x, y, and u are state, measurement, and control vectors of dimension N, L, and M, respectively, ν_1 and ν_2 are zero-mean white Gaussian noise vectors with known, constant covariance matrices, and A, B, and G are known constant matrices. The choice of the feedback gain matrix $S(t)$ is immaterial to our discussion.

Let us now consider the representation of J control actuator failures. The indices of the failed actuators are $j=(j_1,\ldots,j_J)$. When a malfunction occurs in the j_kth control actuator, its normal control function, $u_{j_k}(t)$, is replaced by an autonomous expression, $f_{j_k}(t)$. Let $f(t)$ be an M-element vector whose j_kth element is the autonomous behavior of the failed j_kth actuator, for $k=1,\ldots,J$, and whose other elements are zero. Let I_j be the matrix obtained from the $M\times M$ identity matrix by removing each of the J rows $j_1,\ldots,j_J$. Thus $I_ju(t)$ is a vector obtained by removing the elements $j_1,\ldots,j_J$ from the nominal control vector $u(t)$. Similarly, BI_j^T is an $N\times(M-J)$ matrix obtained by removing the columns $j_1,\ldots,j_J$ from the matrix B. Using this notation, the dynamic response of the system to failure of J actuators whose indices are j is described by

$$\frac{dx}{dt}=Ax(t)=BI_j^TI_ju(t)+Bf(t)+\nu_1(t) \tag{6}$$

The normal control algorithm still calculates the feedback control vector from Eq. (5). However, $f_{j_k}(t)$ is implemented rather than $u_{j_k}(t)$. Combining Eqs. (5) and (6) yields

$$\frac{dx}{dt}=\Big[A+BI_j^TI_jS(t)\Big]x(t)+Bf(t)+\nu_1(t) \tag{7}$$

The state vector $x(t)$ can be expressed in terms of a transition matrix $X_j(t)$, which is the solution of the following differential equation[14]

$$\frac{dX_j}{dt}=\Big[A+BI_j^TI_jS(t)\Big]X_j(t),\qquad X_j(0)=I \tag{8}$$

Finally, the measurement vector (with noise) in response to failure vector $f(t)$ is

$$\tilde{y}_f(t)=GX_j(t)x(0)+G\int_0^t X_j(t)X_j^{-1}(\tau)\Big[Bf(\tau)+\nu_1(t)\Big]\,d\tau+\nu_2(t) \tag{9}$$

Unless $S(t)=0$ (the open-loop case) the transition matrix, X_j, depends on which actuators are malfunctioning, so the covariance matrix of $\tilde{y}$ depends on the failure. Consequently the quadratic norm, based on the covariance matrix of the measurement, varies with the failure.

The failure set for malfunctions of type p is $F(p)$, as in Eq. (1). Each failure $f(t)$ in $F(p)$ is mapped to an average measurement vector $y_f(t)$ (without noise) in measurement space [Eq. (9) with $\nu_1=\nu_2=0$]. Let $C(p)$ be the set of all the average measurement vectors obtained from failures in the set $F(p)$. That is

$$C(p)=\Big\{y:\quad y(t)=y_f(t)\quad \text{for all}\quad f\in F(p)\Big\} \tag{10}$$

We will call $C(p)$ the *complete response set* for failures of type p. Since the failure set $F(p)$ is convex, the response set $C(p)$ is likewise convex because $y_f(t)$ is an affine transformation of f.

It is more convenient, however, to define $C(p)$ in terms of its boundary. Define the constant failure vector $\bar{p}=(\bar{p}_1,\ldots,\bar{p}_M)$, where $\bar{p}_m=(\hat{p}_m+\tilde{p}_m)/2$ for $m=1,\ldots,M$. Let $\bar{y}(t)$ be the average response to the constant failure $\bar{p}$, and so

$\bar{y}(t)=y_{\bar{p}}(t)$. That is

$$\bar{y}(t)=GX_j(t)x(0)+G\int_0^t X_j(t)X_j^{-1}(\tau)B\bar{p}\,d\tau$$

Let $F^*(p)$ be the set

$$F^*(p)=\left\{f=(f_1,\ldots,f_M):\quad |f_m(t)|\leq\frac{\hat{p}_m-\tilde{p}_m}{2}\right\} \tag{11}$$

Every element g in $F(p)$ can be expressed as $g=\bar{p}+f$ where f belongs to $F^*(p)$. Thus the response to g can be expressed as the sum of the response to $\bar{p}$ and the response to f. Let $\Psi(t,\tau)=GX_j(t)X_j^{-1}(\tau)B$. Now the response set $C(p)$ can be expressed as

$$C(p)=\left\{y:\quad y=\bar{y}(t)+\int_0^t\Psi(t,\tau)f(\tau)\,d\tau\quad\text{for}\quad f\epsilon F^*(p)\right\} \tag{12}$$

From this expression, it is evident that $C(p)$ is convex, contains the point $\bar{y}(t)$, and is symmetric with respect to inversion through $\bar{y}(t)$. Also, every element of $C(p)$ can be expressed as $y=\bar{y}(t)+\alpha\rho(\omega)\omega$ where ω is a unit vector in the direction from $\bar{y}$ to y, $\rho(\omega)$ is the distance along ω from $\bar{y}$ to the boundary of $C(p)$ and $0\leq\alpha\leq1$. That is, the complete response set can be represented as

$$C(p)=\left\{y:\quad y=\bar{y}(t)+\alpha\rho(\omega)\omega,\quad 0\leq\alpha\leq1,\quad \omega^T\omega=1\right\} \tag{13}$$

To evaluate the radius function $\rho(\omega)$ we must first identify the elements of F^* that generate the boundary points of $C(p)$. Let ϕ be a vector in E^L. For a given $f\epsilon F^*(p)$, the set of points z that satisfy

$$\phi^Tz=\phi^T\left(\bar{y}(t)+\int_0^t\Psi(t,\tau)f(\tau)\,d\tau\right) \tag{14}$$

constitutes a plane in E^L through the point y_f and perpendicular to ϕ, as shown by the line L_1 in Fig. 1. The distance of this plane from $\bar{y}$ is

$$\text{dis}(y_f,\bar{y})=\frac{1}{\sqrt{\phi^T\phi}}\left|\phi^T\int_0^t\Psi(t,\tau)f(\tau)\,d\tau\right|$$

This distance varies as f varies on the set F^*. That element of F^* that maximizes $\text{dis}(y_f,\bar{y})$ defines a boundary point of $C(p)$, denoted BP in Fig. 1. Let $\psi^m(t,\tau)$ represent the mth column of $\Psi(t,\tau)$. Then $\text{dis}(y_f,\bar{y})$ is maximized on F^* when the elements of the vector f are chosen as

$$f_m(\tau;\phi)=\frac{\hat{p}_m-\tilde{p}_m}{2}\,\text{sgn}\left[\phi^T\psi^m(t,\tau)\right],\qquad m=1,\ldots,M \tag{15}$$

where $\text{sgn}(x)=\pm1$, matching the sign of x. [A similar maximization problem is discussed in Eqs. (21-24), to which the reader is referred for justification of Eq. (15).] Boundary points of $C(p)$ are now represented as

$$y(t;\phi)=\bar{y}(t)+\int_0^t\Psi(t,\tau)f(\tau;\phi)\,d\tau \tag{16}$$

where $f(\tau;\phi)$ in this expression is defined in Eq. (15). Distinct boundary points are obtained by varying ϕ. Each boundary point in turn defines a value of the radius vector. For each ϕ, the radius of $C(p)$ along direction $\omega=y(t;\phi)-\bar{y}(t)$ is $\sqrt{\omega^T\omega}$, which can be tabulated numerically as a function of the direction ω. Let $\rho(\omega)$ represent this tabulation. The argument of ρ need not be a normalized vector, but we will adopt the convention that, for any scalar α, $\rho(\alpha\omega)=|\alpha|\rho(\omega)$ and that $\rho(\omega)$ precisely equals the radius of $C(p)$ along ω when ω is a unit vector.

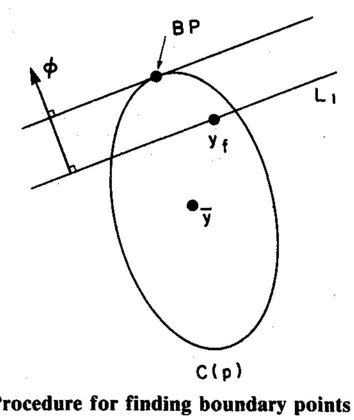


Fig. 1 Procedure for finding boundary points of $C(p)$.

IV. Designing the Multihypothesis Diagnosis of Open-Loop Malfunctions

In the absence of feedback in the control loop [$S(t)=0$ in Eq. (5) and $u(t)$ is independent of x] the transition matrix, Eq. (8), is independent of the malfunction. Consequently the quadratic norm based on the covariance matrix of the measurement does not depend on the failure. Determination of the robustness of a collection H of hypothesized malfunctions can be based on the solution of a sequence of linear optimization problems, as shown in this section.

As before, let

$$H=\bigcup_{k=1}^{K}H_k$$

be the complete set of hypothesized malfunctions. Let g and h belong to H, and define the *minimum relative norm* on $C(p^k)$ with respect to g and h as

$$D_k(g,h)=\min_{y\epsilon C(p^k)}(\|y_g-y\|^2-\|y_h-y\|^2) \tag{17}$$

If $D_k(g,h)$ is positive, then every occurrence of failure of type p^k will be ascribed to hypothesized malunction h rather than to g. It is evident from the definition of correct diagnosis that failures of type p^k are correctly diagnosed if, for each $g\epsilon H-H_k$, there is an element $h\epsilon H_k$ such that

$$D_k(g,h)\geq0 \tag{18}$$

This means that, for every failure in $F(p^k)$, no hypothesis outside H_k will be chosen by the multihypothesis algorithm. Consequently type-p^k failures will be correctly diagnosed.

Expanding the norms in Eq. (17) in terms of the inner product, one finds

$$D_k(g,h)=\|y_g\|^2-\|y_h\|^2-2\max_{y\epsilon C(p^k)}[y_g-y_h,y] \tag{19}$$

The maximum on the right-hand side does in fact exist since $[y_g-y_h,y]$ is a linear (and thus continuous) function from the compact set $C(p^k)$ to the real numbers.[15] Consequently, determination of the correct diagnosis of failure type p^k is based on evaluating the maximum of the linear function $[y_g-y_h,y]$ on $C(p^k)$, for each g and h in H. Equation (2) indicates that the multihypothesis algorithm itself evaluates a quadratic expression in y. The adequacy of a linear expression for determining correct diagnosis derives from the fact, expressed in Eq. (17), that correct diagnosis is established by comparing norms that are independent of the hypothesized malfunctions.

To illustrate this analysis, we consider part of the design process for constructing a maximum-likelihood, multihypothesis algorithm for diagnosing control actuator failures in AFTI/F16 aircraft in steady open-loop flight at 0.9 Mach and 20,000 ft altitude. The dynamic behavior and measurements of the failure-free linear system are represented by Eqs. (3-5) with $S(t)=0$. The eight state variables, in order of their appearance in x, are pitch angle, forward velocity, angle of attack, pitch rate, bank angle, sideslip angle, roll rate, and yaw rate. The six control variables, in order of their appearance in u, are right and left horizontal tails (elevators), right and left wing flaps, canards (operated symmetrically), and rudder. These control variables are zero in steady open-loop flight but vary autonomously after failure. G is the 8×8 identity matrix and the values of A (Table 1) and B (Table 2) are taken from Ref. 16.

We will now develop an explicit expression for the maximum in Eq. (19). Let the initial state vector be $x(0)=0$. From Eqs. (3) and (4), one finds the average response to the malfunctioning control vector u to be

$$y_u(t)=\int_0^t Ge^{A(t-\tau)}Bu(\tau)\,\mathrm{d}\tau \tag{20}$$

Let the inner product take the form $[x,y]=x^TV^{-1}y$, where V is the covariance matrix of the response vector. Also, let $\lambda^m(t,\tau)$ be the mth column of the matrix $V^{-1}Ge^{A(t-\tau)}B$. Let $\delta(t)=y_g(t)-y_h(t)$. Then one finds

$$\left[y_g(t)-y_h(t),y_u(t)\right]=\sum_{m=1}^{M}\int_0^t \delta(t)^T\lambda^m(t,\tau)u_m(\tau)\,\mathrm{d}\tau \tag{21}$$

Examination of Eq. (21) shows that the mth integral achieves its maximum when $u_m(\tau)$ is chosen to switch between its extremal values as $\delta(t)^T\lambda^m(t,\tau)$ changes sign. Specifically, Eq. (21) is maximized by choosing the elements of u as

$$\begin{aligned} u_m(\tau)&=\hat{p}_m \quad \text{for} \quad \delta(t)^T\lambda^m(t,\tau)\geq 0\\ &=\check{p}_m \quad \text{for} \quad \delta(t)^T\lambda^m(t,\tau)<0 \end{aligned} \tag{22}$$

Let D_{m+} and D_{m-} denote the subsets of the interval $[0,t]$ for which $\delta(t)^T\lambda^m(t,\tau)$ is nonnegative and negative, respectively. Thus the maximum value of the inner product becomes

$$\max_{u\in F(p^k)}\left[y_g(t)-y_h(t),y_u(t)\right]=\sum_{m=1}^{M}\left\{\hat{p}_m\int_{D_{m+}}\delta(t)^T\lambda^m(t,\tau)\,\mathrm{d}\tau + \check{p}_m\int_{D_{m-}}\delta(t)^T\lambda^m(t,\tau)\,\mathrm{d}\tau\right\} \tag{23}$$

$$=\sum_{m=1}^{M}\left\{\frac{\hat{p}_m+\check{p}_m}{2}\int_0^t\delta(t)^T\lambda^m(t,\tau)\,\mathrm{d}\tau + \frac{\hat{p}_m-\check{p}_m}{2}\int_0^t\left|\delta(t)^T\lambda^m(t,\tau)\right|\,\mathrm{d}\tau\right\} \tag{24}$$

The minimum relative norm on $C(p^k)$ with respect to g and h is obtained by substituting Eq. (24) in Eq. (19). We are now able to determine whether or not a given collection of hypothesized malfunctions is robust.

The starting point for selecting hypothesized failures is specification of the failure sets which must be correctly diagnosed. Identification of a robust and efficient set of hypothesized malfunctions is then an iterative process. At least one hypothesis must be included in H for each failure set that is to be correctly diagnosed. Given an initial choice of H, Eqs. (19) and (24) are used to determine whether or not the required failure sets are correctly diagnosed. Elements of H are then modified, and new elements are included until correct diagnosis is attained for each specified failure set.

The procedure for determining the robustness of a given set of hypotheses can be inverted, in part, to aid in the search for hypothesized malfunctions. A simple numerical example will illustrate this analysis. Suppose it is desired to correctly diagnose malfunctions of failures in the second and fifth control functions (left elevator and canards), when these control surfaces are deflecting autonomously. For graphical simplicity, we will select hypothesized malfunctions h_i, which are constant in time and nonzero only in the second and fifth elements. Thus hypothesized malfunctions can be represented as points in the plane, where the horizontal and vertical coordinates are the second and fifth elements of the failure vector, h_{i2} and h_{i5}, respectively. Three hypothesized malfunctions, h_1, h_2, and h_3 have been included in H to diagnose other failures, as shown in Fig. 2a. It is now desired to select the minimum set of hypotheses needed to assure correct diagnosis of left elevator and canard deflections between, for example, 0.6 and 0.8 deg. Let us denote this failure set $F(0.6,0.8)$.

Each point in the square region of Fig. 2a represents a constant failure in $F(0.6,0.8)$. However, not each such point, if used as a hypothesized malfunction, would yield correct diagnosis of the malfunctions in $F(0.6,0.8)$. Let h be a point in the square region of Fig. 2a, and consider the maximum likelihood comparison between h and h_1. Equations (19) and (24) are used to evaluate $D(h_1,h)$, the minimum relative norm on $C(0.6,0.8)$ with respect to h_1 and h. The minimum relative norm for each point h below the curve in Fig. 2b is found to be positive, which indicates that these hypotheses yield correct diagnosis of the failures in question when compared with hypothesis h_1. The minimum relative norm of all points above the curve in Fig. 2b is negative, which means that hypothesized failures above the curve will not yield correct diagnosis. Figure 2c shows a similar analysis based on comparison with h_2. Again the minimum relative norm $D(h_2,h)$, is positive for

Table 1 Matrix A[a]

0	0	0	1.00000	0	0	0	0
−32.1830	0.012075	38.2906	−30.1376	0	0	0	0
−0.00112	−0.000022	−1.48446	0.994789	0	0	0	0
−0.000309	−0.00013	4.27171	−0.777221	0	0	0	0
0	0	0	0	0	0	1.00000	0
0	0	0	0	0.03449	−0.343554	0.0326360	−0.997556
0	0	0	0	0	−55.2526	−2.80004	0.145674
0	0	0	0	0	7.23700	−0.0231840	−0.362530

[a]Units of the state variables are rads, rads/s, or ft/s.

Table 2 Matrix B[a]

0	0	0	0	0	0
1.00296	1.00296	1.15840	1.15840	0	0
−0.0746135	−0.0746135	−0.122462	−0.122462	0	0
−12.0291	−12.0291	−3.23635	−3.23635	0	0
0	0	0	0	0	0
0.0133045	−0.0133045	−0.0006855	0.0006855	0.0267340	0.0370320
−25.3645	25.3645	−25.5251	25.5251	5.53185	10.3955
−2.56855	2.56855	−0.625030	0.625030	5.89254	−5.80890

[a]Units of the state and control variables are rads, rads/s, or ft/s.

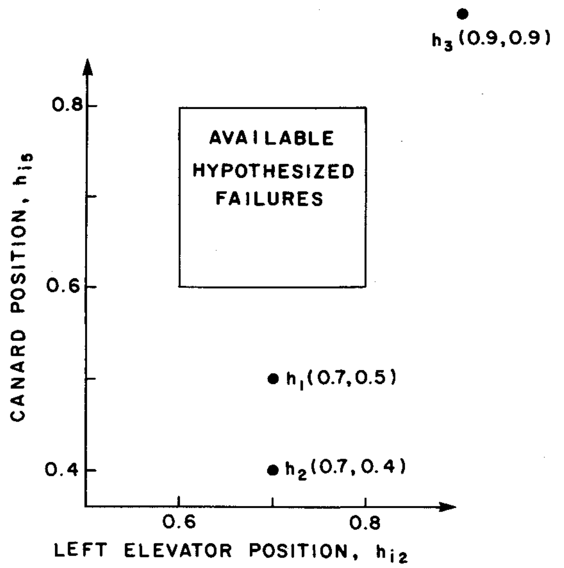


Fig. 2a Hypothesized malfunctions represented by points in the h_{i2} vs h_{i5} plane.

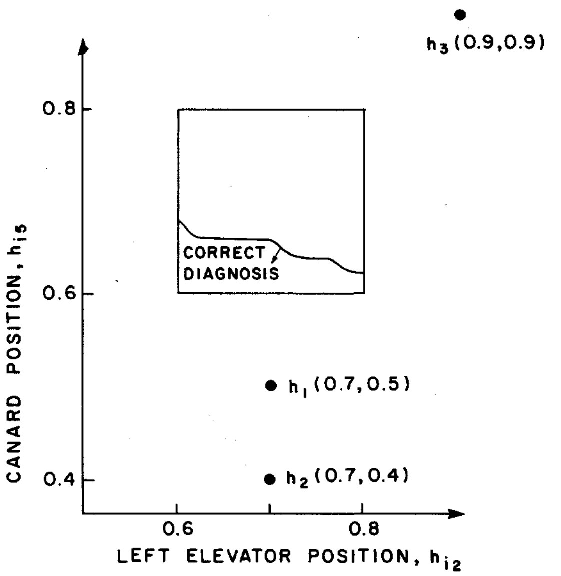


Fig. 2b Malfunctions in the square region and below the curve yield correct diagnosis in comparison with h_1.

points below the curve and negative for points above the curve. Thus, correct diagnosis in comparison with h_2 can be achieved only if a point below the curve in Fig. 2c is included in H. Comparison of Figs. 2b and 2c shows that correct diagnosis with respect to h_1 ensures correct diagnosis with respect to h_2. The analysis is repeated to determine the hypothesized malfunctions that yield correct diagnosis in comparison with h_3, and the results appear in Fig. 2d. Points above the curve yield correct diagnosis of all failures in $F(0.6,0.8)$, whereas points below the curve do not. Overlaying Figs. 2b–d, as in Fig. 2e, shows that two hypothesized malfunctions are necessary and sufficient to achieve correct diagnosis of all failures in $F(0.6,0.8)$. One hypothesis must lie between the intermediate and upper curves, and one must lie below the lowest curve. Correct diagnosis of the failure set $F(0.6,0.8)$ requires that two such hypotheses be included in H as long as h_1, h_2, and h_3 are in H. Likewise, unless additional hypotheses are added to H for diagnosis of different failure sets, the two hypotheses which have been identified are sufficient to assure correct diagnosis of $F(0.6,0.8)$. This analysis is continued until conditions are established for defining the smallest set of hypothesized malfunctions that ensure correct diagnosis for each of the specified failure sets.

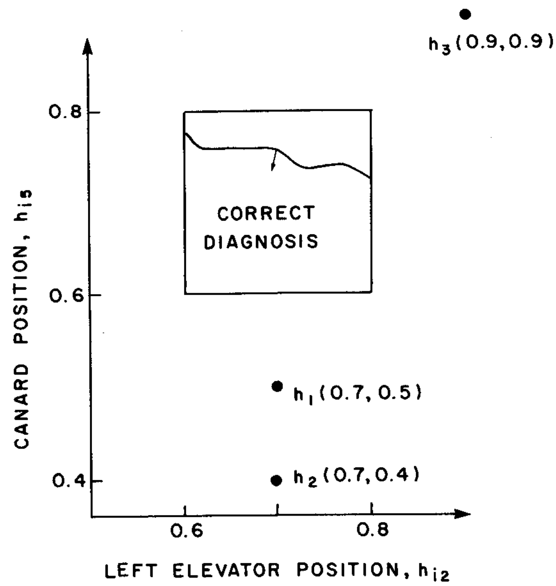


Fig. 2c Malfunctions in the square region and below the curve yield correct diagnosis in comparison with h_2.

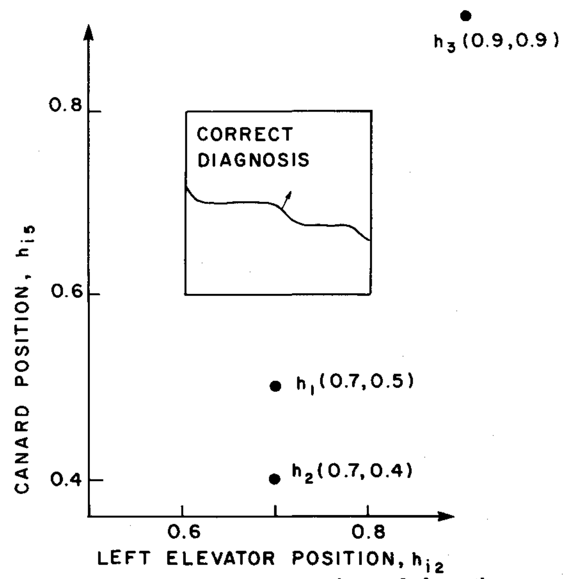


Fig. 2d Malfunctions in the square region and above the curve yield correct diagnosis in comparison with h_3.

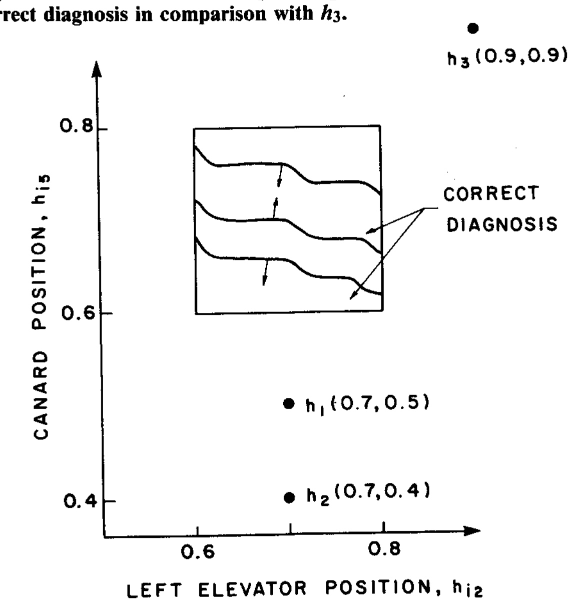


Fig. 2e Overlay of Figs. 2b–d, showing necessity of two malfunctions for correct diagnosis in comparison with h_1, h_2, h_3.

V. Designing the Multihypothesis Diagnosis of Closed-Loop Malfunctions

Let

$$H = \bigcup_{k=1}^{K} H_k$$

where each set H_k contains malfunctions drawn from the set $F(p^k)$ of uniformly bounded failures. The system is described by Eqs. (3) and (4), and the feedback gain in Eq. (5) is nonzero. We wish to determine whether or not malfunctions of type p^k are correctly diagnosed. Equation (17) must be modified to account for the fact that, due to the feedback in the control loop, the quadratic norm depends on the failure. Accordingly, let g and h belong to H and define the minimum relative norm on $C(p^k)$ with respect to g and h as

$$D_k(g,h) = \min_{y \in C(p^k)} \left(\| y_g - y \|_g^2 - \| y_h - y \|_h^2 \right) \tag{25}$$

The main result of this section is the evaluation of this minimum relative norm. Once that is achieved, the hypothesized malfunctions are selected by the iterative procedure illustrated in Sec. IV.

Let g and h be hypothesized malfunctions, and let y_g and y_h be the corresponding average responses. Let $y = \bar{y} + \eta$ be an element of $C(p^k)$, where $\bar{y}$ is defined, with respect to the parameters p^k, as in connection with Eq. (12) and $\eta = \alpha\rho(\omega)\omega$ as in Eq. (13). The expression to be minimized in Eq. (25) becomes

$$\| y_g - y \|_g^2 - \| y_h - y \|_h^2 = (y_g - \bar{y} - \eta)^T V_g^{-1} (y_g - \bar{y} - \eta)$$

$$- (y_h - \bar{y} - \eta)^T V_h^{-1} (y_h - \bar{y} - \eta) \tag{26}$$

$$= \eta^T \Delta \eta - 2\zeta^T \eta + \mu \tag{27}$$

where $\Delta = V_g^{-1} - V_h^{-1}$, $\zeta = V_g^{-1}(y_g - \bar{y}) - V_h^{-1}(y_h - \bar{y})$, and $\mu = \|y_g - \bar{y}\|_g^2 - \|y_h - \bar{y}\|_h^2$. Failures of type p^k are correctly diagnosed if, for each $g \in H - H_k$, there is an element $h \in H_k$ such that

$$D_k(g,h) \geq 0 \tag{28}$$

Referring to Eq. (13), it is evident that η is a vector of arbitrary orientation whose length does not exceed the distance in direction η of $\bar{y}$ from the boundary of $C(p^k)$. Thus η is constrained by

$$|\sqrt{\eta^T \eta}| \leq \rho\left(\frac{\eta}{\sqrt{\eta^T \eta}}\right) = \frac{1}{\sqrt{\eta^T \eta}} \rho(\eta) \tag{29}$$

where $\rho(\omega)$ is determined numerically as explained in Sec. III. This inequality constraint on the maximization of Eq. (27) can be replaced by an equality by introducing an undetermined quantity, β

$$\eta^T \eta + \beta^2 = \rho(\eta) \tag{30}$$

Adjoin the constraint to the expression in Eq. (27) as

$$D^* = \eta^T \Delta \eta - 2\zeta^T \eta + \mu + \lambda \left[\eta^T \eta + \beta^2 - \rho(\eta) \right] \tag{31}$$

Necessary conditions for a stationary point of Eq. (27) are

$$0 = \frac{\partial D^*}{\partial \eta} = 2\Delta\eta - 2\zeta + 2\lambda\eta - \lambda \frac{\partial \rho}{\partial \eta} \tag{32}$$

$$0 = \frac{\partial D^*}{\partial \beta} = 2\lambda\beta \tag{33}$$

Equation (33) together with the constraint implies that $\lambda = 0$ if $\eta^T \eta < \rho(\eta)$. Thus an extremum of Eq. (27) occurs in the interior of $C(p^k)$ if the solution of

$$\Delta \eta = \zeta \tag{34}$$

satisfies $\eta^T \eta < \rho(\eta)$. If not, then the extrema of Eq. (27) occur on the boundary of $C(p^k)$ and must satisfy

$$(\Delta + \lambda I)\eta = \zeta + \frac{1}{2} \lambda \frac{\partial \rho}{\partial \eta} \tag{35}$$

and

$$\eta^T \eta = \rho(\eta) \tag{36}$$

Equations (34-36) determine the constrained extrema of $D_k(g,h)$. Failures of type p^k are correctly diagnosed if the condition in Eq. (28) is satisfied.

The solution of Eqs. (35) and (36) is computationally somewhat cumbersome. It is therefore useful to know that if Δ is a positive definite matrix, then Eq. (27) has precisely one minimum and may have several local maxima. Or, if Δ is negative definite, then Eq. (27) has precisely one maximum and may have several local minima. If Δ is indefinite, then Eq. (27) can have several minima and maxima. The proof of these assertions appears in the Appendix.

VI. Conclusions

This paper has described a method for designing a maximum-likelihood, multihypothesis algorithm for diagnosing control-actuator failures in linear systems. Uncertainty in the temporal behavior of a malfunctioning actuator is represented by employing the set theoretic technique called convex modeling. For open-loop systems (autonomous controllers), the diagnosis algorithm is designed by solving a sequence of linear optimization problems. For closed-loop feedback systems, the design of the diagnosis algorithm requires the solution of nonlinear equations. The resulting diagnosis algorithm is robust and efficient. It is robust in that the diagnosis invariably distinguishes between failure sets that represent complex uncertainty in the temporal form of the malfunctions. It is efficient in that no smaller set of hypothesized malfunctions could achieve correct diagnosis of the required classes of failures. The significance of this result is that design of an algorithm for diagnosis of control actuator failure can be based on a systematic and numerically implementable procedure that yields the best possible algorithm, in the sense of robustness and efficiency defined here.

Appendix: Extrema of D

The complete response set is closed and bounded, and the relative norm is a continuous function. Thus Eq. (27) has a minimum and a maximum on $C(p^k)$. We will show that, if Δ is a positive definite matrix, then Eq. (27) can have exactly one allowed minimum. If Δ is negative definite the argument can be reversed to prove that Eq. (27) can have only one allowed maximum. If Δ is indefinite, then Eq. (27) can have several minima and maxima.

Let d represent the quantity in Eq. (27). Suppose that Δ is positive definite, and that η^1 and η^2 are both allowed local minima of Eq. (27). Then both $\bar{y} + \eta^1$ and $\bar{y} + \eta^2$ belong to $C(p^k)$. Since the response set is convex, every point $y = \bar{y} + \alpha\eta^1 + (1-\alpha)\eta^2$ belongs to $C(p^k)$ for all $0 \leq \alpha \leq 1$. Thus all points along $\eta^1 - \eta^2$ and between η^1 and η^2 are possible extrema. However, since η^1 and η^2 are local minima, the rate of increase of d along $\eta^1 - \eta^2$ must be nonnegative at both η^1 and η^2. That is

$$s_1 \equiv (\eta^2 - \eta^1)^T \frac{\partial d}{\partial \eta} \bigg|_{\eta^1} \geq 0 \tag{A1}$$

$$s_2 \equiv (\eta^1 - \eta^2)^T \frac{\partial d}{\partial \eta}\bigg|_{\eta^2} \geq 0 \tag{A2}$$

Let $\eta^2 = \eta^1 + \xi$. Then

$$s_1 = 2\xi^T(\Delta\eta^1 - \zeta) \tag{A3}$$

$$s_2 = -2\xi^T(\Delta\eta^1 + \Delta\xi - \zeta) \tag{A4}$$

and so,

$$s_1 + s_2 = -2\xi^T\Delta\xi \tag{A5}$$

Since Δ is positive definite and s_1 and s_2 are both nonnegative, Eq. (A5) implies that $\xi = 0$. Consequently, η^1 and η^2 are identical and d has only one allowed minimum.

Acknowledgments

This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, U.S. Air Force, under Grant AFOSR-88-0209. The author is indebted to Osnat Katzanek and to Dror Saddan for invaluable assistance in performing the computer calculations.

References

[1]Deckert, J. C., Desai, M. N., Deyst, J. J., and Willsky, A. E., "F-8 DFBW Sensor Failure Identification Using Analytic Redundancy," *IEEE Transactions,* Vol. AC22, 1977, pp. 795–803.

[2]Chang, C. B., and Dunn, K. P., "ON GLR Detection and Estimation of Unexpected Inputs in Linear Discrete Systems," *IEEE Transactions,* Vol. AC24, 1977, pp. 499–501.

[3]Fiorina, M., and Maffezzioni, C., "A Direct Approach to Jump Detection," *IEEE Transactions,* Vol. AC27, 1979, pp. 428–434.

[4]Friedland, B., "Maximum Likelihood Estimation of a Process with Random Transitions (Failures)," *IEEE Transactions,* Vol. AC24, 1979, pp. 932–937.

[5]Motyka, P. R., and Bell, J. W., "Failure Detection and Isolation for Tactical Aircraft Using the Generalized Likelihood Ratio Test, *Proceedings of the National Aeronautical and Electronics Conference,* Dayton, Ohio, 1980.

[6]Wakabayashi, J., and Fukumoto, A., "Simulation Study of a System for Diagnosis of Nuclear Power Plant Operation," *Nuclear Technology,* Vol. 54, 1981, pp. 19–30.

[7]Tylee, J. L., "Generalized Likelihood Ratio Approach to Detecting and Identifying Failures in Pressurizer Instrumentation," *Nuclear Technology,* Vol. 56, 1982, pp. 484–492.

[8]Mazetis, G. R., and Ebert, D. D., "Pontrygin's Optimal Control Principle Applied to Computer Assisted Accident Diagnostics," *Nuclear Technology,* Vol. 67, 1984, pp. 411–418.

[9]Kerr, T. H., "Statistical Analysis of a 2-Ellipsoid Overlap Test for Real-Time Failure Detection," *IEEE Transactions,* Vol. AC25, 1980, pp. 762–773.

[10]Schweppe, F. C., *Uncertain Dynamic Systems,* Prentice-Hall, London, 1973.

[11]Ben-Haim, Y., *The Assay of Spatially Random Material,* Kluwer, Dordrecht, Holland, 1985.

[12]Ben-Haim, Y., "Convexity Analysis: A Tool for Optimizing Malfunction Isolation," *IEEE Conference on Decision and Control,* Inst. of Electrical and Electronics Engineers, New York, 1986, pp. 1570–1575.

[13]Schmitendorf, W. E., "Design Methodology for Robust Stabilizing Controllers," *Journal of Guidance, Control and Dynamics,* Vol. 10, No. 3, 1987, pp. 250–254.

[14]Bellman, R., *Introduction to Matrix Analysis,* 2nd ed., Tata McGraw-Hill, 1974, p. 173.

[15]Kelly, P. J., and Weiss, M. L., *Geomtry and Convexity,* Wiley, 1979, p. 50.

[16]Schneider, D. L., "QFT Digital Flight Control Design As Applied to the AFTI/F16," M.Sc. Thesis, Air Force Inst. of Technology, Wright-Patterson AFB, OH, AFIT/GE/ENG/86D-4, Dec. 1986.