JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 22, No. 4, AUGUST 1977

Real-Time Failure Detection: A Nonlinear Optimization Problem That Yields a Two-Ellipsoid Overlap Test[1,2]

T. H. KERR[3]

Communicated by A. V. Balakrishnan

Abstract. Real-time failure detection for systems having linear stochastic dynamical truth models is posed in terms of two confidence region sheaths. One confidence region sheath is about the expected no-failure trajectory; the other is about the Kalman estimate. If these two confidence regions of ellipsoidal cross section are disjoint at any time instant, a failure is declared.

A test for two-ellipsoid overlap is developed which involves finding a single point x^* whose presence in both ellipsoids is necessary and sufficient for overlap. Thus, the overlap test is contorted into a search for x^*, shown to be the solution of a nonlinear optimization problem that is easily solved once an associated scalar Lagrange multiplier is known. A successive approximations iteration equation for λ is obtained and is shown to converge as a contraction mapping. The method was developed to detect failures in an inertial navigation system that appear as uncompensated gyroscopic drift rate. For simulated gyroscopic failures, the iterations converged very quickly, easily allowing real-time failure detection.

Key Words. Detection theory, function minimization, Lagrange problems, contraction mapping principle, Kalman filter.

1. Introduction

Failure detection and failure isolation are common problems in engineering systems. In general, failure detection requires continuous

[1] This work was supported by the Department of the Navy, Strategic Systems Project Office, SP-24.

[2] This paper is based on an earlier paper (Ref. 1) which was presented at the IEEE Conference on Decision and Control, Phoenix, Arizona, 1974. A stronger convergence proof is presented in the present paper.

[3] Member of the Technical Staff, The Analytic Sciences Corporation (TASC), Reading, Massachusetts.

This journal is copyrighted by Plenum. Each article is available for $7.50 from Plenum Publishing Corporation, 227 West 17th Street, New York, N.Y. 10011.

vigilant monitoring of the observable output variables of the system. Under normal conditions, the output variables follow certain known patterns of evolution within certain limits of uncertainty introduced by slight random system disturbances and measurement noise in the sensors. When failures occur, the observable output variables deviate from their nominal state-space trajectories or evolutionary pattern. Most failure detection techniques are based on spotting these deviations from the usual in the observable output variables.

Whereas the detection of an unknown signal at a known time or the detection of a known signal at an unknown time are standard problems in communication theory, the problem in failure detection is to detect a signal of unknown magnitude which occurs at an unknown time. Failure detection is a more difficult problem that has only recently received attention in the literature. Mehra and Peschon have suggested several failure detection approaches (Ref. 2) which are based on the innovations properties of the Kalman filter residual. Jones has approached the failure detection problem using a reference model, such as those that exist in observer theory or in a suboptimal Kalman filter (Ref. 3). Willsky has approached the failure detection problem using a generalized likelihood ratio (Ref. 4). Chien has approached the problem using a generalization of the Wald sequential likelihood ratio test (Ref. 5).

This paper reports a different philosophical approach to the failure detection problem. It places a confidence region about the nominal unfailed trajectory corresponding to the H_0-hypothesis and a second confidence region about the Kalman filter estimate based on processing the actual measurements. When these two confidence regions are disjoint, impling a non-H_0 situation, a failure is declared. This approach was motivated by the computational constraints that were imposed for on-line, real-time failure detection in an inertial navigation system (INS). There was a high-dimensional system truth model, but only a reduced-order Kalman filter would be allowed because of the limited computer memory available. This confidence region approach appears to achieve its objective of detecting failures without being susceptible to some of the same ailments that the other four residual-based approaches would experience under similar computational constraints (i.e., the other four approaches were derived using systems or Kalman filters having the same dimensionality as the truth or error models). Residuals can be nonwhite or biased either because a failure occurred or because the Kalman filter is a reduced-order suboptimal filter (Refs. 6–8). The calculated confidence regions of this paper are still exact when used with a particular formulation of a reduced-order filter (Ref. 9) which results in a known covariance of error. The effect of using a reduced-order filter is discussed in a companion article (Ref. 10) which also discusses

other details of implementing for failure detection. Another advantage of this confidence-region approach is that the arguments are geometric in nature and may be easily visualized, as will be seen in the figures.

2. Analytical Theory of Failure Detection Using Two Confidence Regions

Consider a system having a linear state variable truth model or error model of the following form:

$$x(k+1)=\Phi(k+1, k)x(k)+w(k), \tag{1}$$

$$z(k)=H(k)x(k)+v(k), \tag{2}$$

where $w(k)$ and $v(k)$ are independent, zero-mean, white noises having covariances of intensity $Q(k)$ and $R(k)$, respectively. There is a Gaussian random vector initial condition $x(0)$ of mean $\bar{x}_0$ and covariance P_0, and the failure modes of the system are included as states of the above model [e.g., unwanted ramp and bias gyroscopic drift-rates are states in the linear error model of an INS (Ref. 11)].

The solution of the associated Kushner partial differential equation for the conditional probability density function (p.d.f.) of $x(k)$, given the measurements, is a Gaussian p.d.f. having the Kalman estimate $\hat{x}(k)$ and covariance of error $P_1(k)$ as mean and variance, respectively (Ref. 12), where[4]

$$\hat{x}(k+1)=\Phi(k+1, k)\hat{x}(k)+K(k+1)\gamma(k+1), \tag{3}$$

$$\hat{x}(0)=x_0, \tag{4}$$

$$\gamma(k) \triangleq z(k)-H(k)\Phi(k, k-1)\hat{x}(k-1), \tag{5}$$

$$K(k)=P_1(k)H^T(k)[H(k)P_1(k)H^T(k)+R(k)]^{-1}, \tag{6}$$

$$P_1(k+1)=\Phi(k+1, k)[I-K(k)H(k)]P_1(k)\Phi^T(k+1, k)+Q(k+1), \tag{7}$$

$$P_1(0)=P_0. \tag{8}$$

Hypothesis H_0. The system is in the unfailed condition, so that the failure states of the system error model (1) are zero.

We shall use the symbol H_1 to denote the case where Hypothesis H_0 is not satisfied. Thus, H_1 is non-H_0. A particular example is Eq. (1) of Ref. 4.

At a particular fixed time k, the p.d.f. of $x(k)$ under H_0 is a Gaussian having mean

$$\bar{x}(k) \triangleq E[x(k)|H_0]=\Phi(k, 0)E[x(0)|H_0] \tag{9}$$

[4] The propagate and update equations of the Kalman filter have been combined to facilitate manipulation in the proof of Lemma 5.1.

and a variance that is the following solution of the associated linear matrix equation (Ref. 13):

$$P_2(k+1)=\Phi(k+1, k)P_2(k)\Phi^T(k+1, k)+Q(k+1), \tag{10}$$

$$P_2(0)=P_0. \tag{11}$$

This time-varying variance represents the uncertainty in $x(k)$ introduced by the system noise and random initial condition, undiminished by the use of the measurements. Therefore, at a particular time k, there are two Gaussian p.d.f.'s associated with $x(k)$. One is $p_{x(k)|H_0}$ and the other is $p_{x(k)|Z(k)}$. Figure 1 conceptually depicts these two p.d.f.'s for the random variable $x(k)$ at the fixed time k.

The first moments of the two p.d.f.'s, $E[x(k)|H_0]$ and $\hat{x}(k)$, may be considered to be point estimates of the random variable $x(k)$ having uncertainty $P_2(k)$ and $P_1(k)$, respectively. Conservatism is attained by using confidence regions about these estimates, instead of only the estimates themselves. An α_1-confidence region about $\hat{x}(k)$ and an α_2-confidence region about $E[x(k)|H_0]$ are depicted in Fig. 1 as the confidence regions $R_1(k)$ and $R_2(k)$, respectively. The probability of finding the realization of the true state $x(k)$ within these two confidence regions may be calculated analytically, and the probability statements are

$$\text{Prob}[x(k)\in R_1(k)|Z(k)]=\alpha_1, \tag{12}$$

$$\text{Prob}[x(k)\in R_2(k)|H_0]=\alpha_2. \tag{13}$$

In one dimension, these confidence regions are confidence intervals; in higher dimensions, these confidence regions (as usual, taken to have the

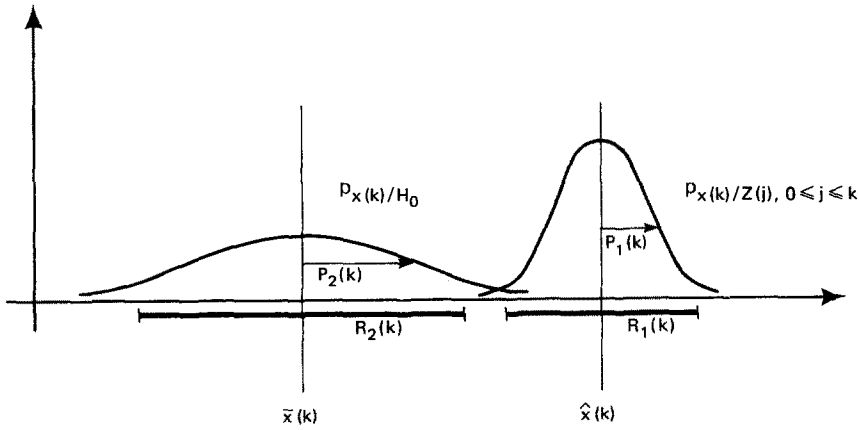

Fig. 1. Gaussian densities used in two-confidence-region failure detection.

same boundaries as the levels of constant p.d.f.) are ellipsoids. The probability of finding the realization of the true state $x(k)$ within these confidence region ellipsoids may be calculated analytically, and the probability statements are

$$\text{Prob}[(x(k)-\hat{x}(k))^T P_1^{-1}(k)(x(k)-\hat{x}(k)) \leq K_1 | Z(k)] = \alpha_1, \tag{14}$$

$$\text{Prob}[(x(k)-\bar{x}(k))^T P_2^{-1}(k)(x(k)-\bar{x}(k)) \leq K_2 | \text{H}_0] = \alpha_2. \tag{15}$$

The constant $K_1(K_2)$ corresponding to the probability level $\alpha_1(\alpha_2)$ is the normalized score value associated with the α_1 probability level of a chi-squared random variable having n degrees of freedom, where n is the number of rows of $x(k)$. The theoretical justification for the calculation of $K_1(K_2)$ using chi-squared is given in Ref. 14; and the details of handling a time-varying confidence region sheath are discussed on pp. 281–291 of Ref. 15 for a different application.

At each decision time k, a confidence region about $\hat{x}(k)$, having α_1 probability of containing the true state $x(k)$, and a confidence region about

$$\bar{x}(k) \triangleq E[x(k)|\text{H}_0],$$

having α_2 probability of containing the true state $x(k)$, may be constructed. These confidence regions are ellipsoids having centers and variances that vary with time to define the two sheaths that are depicted in Fig. 2. The confidence region about $\bar{x}(k)$ represents nominal unfailed behavior (H_0), within the uncertainty introduced by the system noise. The confidence region about $\hat{x}(k)$ reflects the current information of the measurements which indicate the actual situation (either H_0 or H_1) within the uncertainty of the system and measurement noise. As long as the two sheaths overlap, the true state may be in both confidence regions. However, when the ellipsoids are disjoint, the true state cannot be in both confidence regions, and a failure is declared. Declaring a failure corresponds to a failure mode state being judged to be different from the nominal, while taking into account the

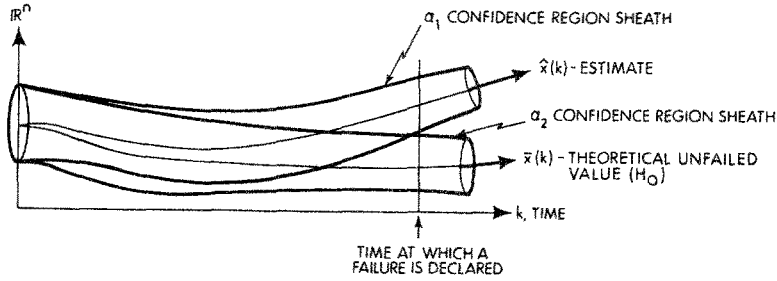

Fig. 2. Two nonoverlapping confidence region ellipsoids indicate failures.

uncertainty introduced by system and measurement noise. In the INS application, *soft* or subtle failures, such as intolerable uncompensated gyroscopic drift, are to be detected, where gyroscopic drifts are failure-mode states of the linear filter model.

3. Two-Confidence-Region Failure Detection: Analytical Basis for the Two-Ellipsoid Overlap Test

A test for the overlap of two ellipsoids having the same confidence level ($\alpha_1 = \alpha_2$ or, equivalently, $K_1 = K_2$) is presented which involves finding a single point x^* whose *presence* in both ellipsoids is necessary and sufficient for overlap and whose *absence* from both ellipsoids is necessary and sufficient for disjointness. Basing an overlap test on the containment or noncontainment of this special point x^* within the ellipsoid is desirable from the point of view of performing real-time numerical calculations. The point $x^*(k)$ may be examined for containment within the α_1-ellipsoid about $\hat{x}(k)$ by forming the inner product

$$(x^*(k)-\hat{x}(k))^T P_1^{-1}(k)(x^*(k)-\hat{x}(k)). \tag{16}$$

There is containment if this scalar quantity is below the threshold K_1. An analogous procedure is followed to check for containment of the point $x^*(k)$ in the other ellipsoid.

Consider the two ellipsoids depicted in Fig. 3, representing a cross section of Fig. 2 at a fixed time. The boundaries of the two ellipsoids C_1 and C_2 are

$$(x-\hat{x})^T P_1^{-1}(x-\hat{x}) = K_1, \tag{17}$$

$$(x-\bar{x})^T P_2^{-1}(x-\bar{x}) = K_2, \tag{18}$$

where the interiors of C_1 and C_2 correspond to x such that the scalar inner product of the left-hand sides of (17) and (18) are, respectively, $\leq K_1$ and

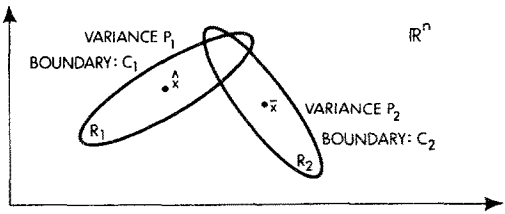

Fig. 3. Overlap of ellipsoids.

$\leq K_2$. For $K_1 = K_2$, these ellipsoids are equal levels of the following two strictly convex (since P_1 and P_2 are positive definite) quadratic forms:

$$y_1 = (x-\hat{x})^T P_1^{-1}(x-\hat{x}), \tag{19}$$

$$y_2 = (x-\bar{x})^T P_2^{-1}(x-\bar{x}). \tag{20}$$

These two quadratic forms are pictured in Fig. 4. The two quadratic forms eventually intersect in a closed curve[5] as depicted in Fig. 5.

The implicit expression for the x-projection of the curve of intersection of the two ellipsoidal parabolas in the domain R^n is

$$y_1 = y_2$$

or, equivalently,

$$G(x) \triangleq -(x-\hat{x})^T P_1^{-1}(x-\hat{x}) + (x-\bar{x})^T P_2^{-1}(x-\bar{x}) = 0. \tag{21}$$

The maximum and minimum points on the curve of intersection are y^{**} and y^*, respectively (see Fig. 5). The x-projection of y^* and y^{**} are x^* and x^{**}, respectively. Naturally, x^{**} and x^* lie on the curve

$$G(x) = 0.$$

It is the pair (x^*, y^*) that is of interest in the two-ellipsoid overlap problem.

When

$$K_1 = K_2,$$

and if there is overlap, the point x^* is considered to be the *most interior point* in the intersection of the two ellipsoids described by (17) and (18). The

[5] Degenerate forms: If $P_1 = P_2$, then the x-projection into R^n of the curve of intersection is a straight line; if $\hat{x} = \bar{x}$ and $P_1 < P_2$, then the curve is a single point.

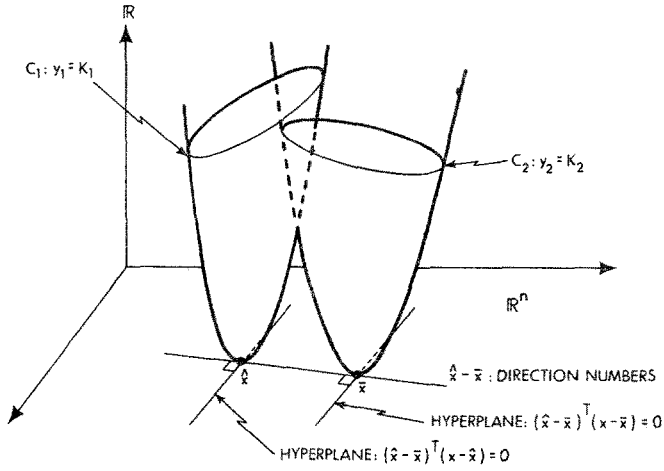

Fig. 4. Ellipsoids are specific levels of two quadratic forms.

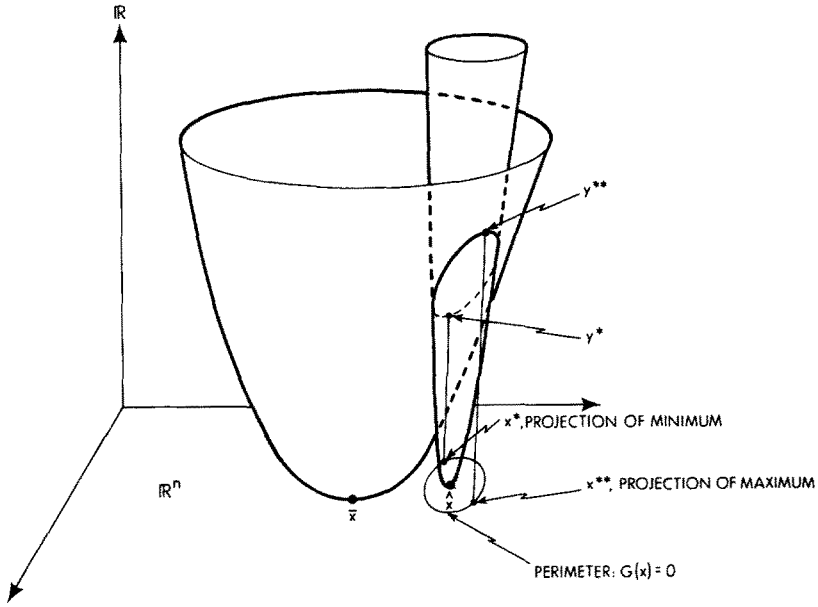

Fig. 5. Projection of intersection of two quadratic forms.

presence of x^* in both ellipsoids is necessary and sufficient for overlap, while the absence of x^* in both ellipsoids is necessary and sufficient for disjointness.

Without going into great detail, the value of the point x^* in determining ellipsoidal overlap may be seen from Fig. 6, which is a true-scale drawing using projective geometry. The upper portion of Fig. 6 shows a top view of two elliptic paraboloids, with the elliptical contours of constant height labeled. The point x^* occurs where two ellipses from two different elliptical families but of the same height are just tangent (i.e., the minimum point of the intersection of the two parabolas). A ray tangent to the two equal level ellipses through their point of common tangency x^* is extended and a folding line is erected perpendicular to this ray. The front view, drawn below the folding line, shows the ellipses of the top view as constant levels. This is just the perspective needed to visually verify the presence or absence of the point x^* in both ellipses as necessary and sufficient for intersection or disjointness, respectively. All constant-level ellipsoids of the two families which are below y^* (i.e., neither contains x^*) are disjoint; all which overlap are above y^* (i.e., both contain x^*). A general method for computationally obtaining this decisive point x^* will now be given.

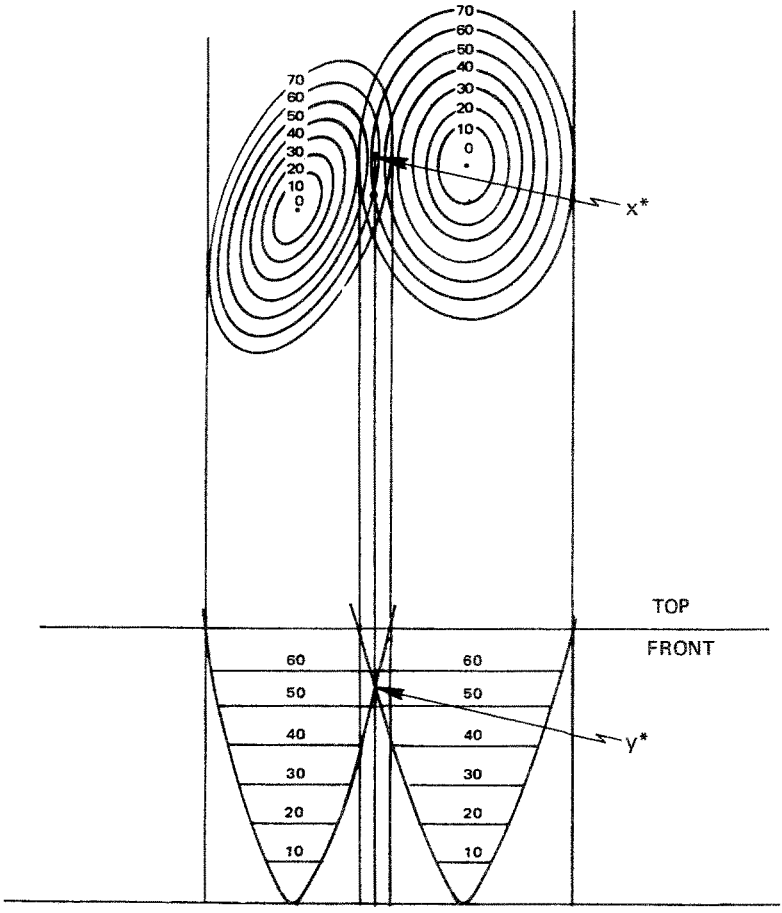

Fig. 6. Ellipsoidal levels above y^* contain x^* (necessary and sufficient condition).

4. Nonlinear Constrained Optimization Problem

The vector x^* is the solution of the following nonlinear optimization problem: minimize (19), subject to the constraint (21). This problem may be recast as an unconstrained optimization problem by using a scalar Lagrange multiplier λ, according to the Kuhn–Tucker theorem (Ref. 16), and finding x^* and λ^* that minimize the Lagrangian

$$l(x, \lambda) = (x - \hat{x})^T P_1^{-1} (x - \hat{x}) + \lambda G(x). \tag{22}$$

By first differentiating (22) with respect to x and setting the result equal to zero and then differentiating (22) with respect to λ and setting the result

equal to zero, the following solution to the minimization problem is obtained.

$$x^* = [(1-\lambda^*)P_1^{-1} + \lambda^* P_2^{-1}]^{-1}\{(1-\lambda^*)P_1^{-1}\hat{x} + \lambda^* P_2^{-1}\bar{x}\}, \tag{23}$$

where λ^* satisfies

$$\{1-[w^T A^{-1}(\lambda)P_1 A^{-1}(\lambda)w/w^T A^{-1}(\lambda)P_2 A^{-1}(\lambda)w]\}\lambda^2 - 2\lambda + 1 = 0. \tag{24}$$

Note that (24) is almost a quadratic equation, where

$$A(\lambda) \triangleq [(1-\lambda)P_2 + \lambda P_1], \tag{25}$$

$$w \triangleq (\hat{x} - \bar{x}). \tag{26}$$

On simplification, (23) reduces to

$$x^* = (1-\lambda^*)P_2 A^{-1}(\lambda^*)\hat{x} + \lambda^* P_1 A^{-1}(\lambda^*)\bar{x}, \tag{27}$$

which is easily evaluated with only one matrix inversion once the Lagrange multiplier λ^* is known.

5. Lagrangian Multiplier

Substituting x^* of (27) back into (22) yields

$$l(x^*, \lambda) = (1-\lambda)(x^* - \hat{x})^T P_1^{-1}(x^* - \hat{x}) + \lambda(x^* - \bar{x})^T P_2^{-1}(x^* - \bar{x}), \tag{28}$$

which must be minimized only over the scalar λ. Once the proper λ is found, (27) yields the proper x^* to minimize the original constrained problem. Substituting the identities

$$x^* - \hat{x} \equiv \lambda P_1 A^{-1}(\lambda)(\bar{x} - \hat{x}), \tag{29}$$

$$x^* - \bar{x} \equiv (1-\lambda)P_2 A^{-1}(\lambda)(\hat{x} - \bar{x}) \tag{30}$$

into (28) results in the following simplification:

$$l(\lambda) = \lambda(1-\lambda)w^T A^{-1}(\lambda)w, \tag{31}$$

where $A(\lambda)$ and w are defined in (25) and (26), respectively. A necessary condition for a minimum, obtained by differentiating (31) with respect to λ, setting the result equal to zero, adding and subtracting the same term, and simplifying, is that

$$\dot{l}(\lambda) = (1-\lambda)w^T A^{-1}(\lambda)w - \lambda w^T A^{-1}(\lambda)P_1 A^{-1}(\lambda)w = 0, \tag{32}$$

where the dot denotes differentiation with respect to λ; upon rearranging, (32) becomes

$$\lambda = 1/[1 + (w^T A^{-1}(\lambda)P_1 A^{-1}(\lambda)w/w^T A^{-1}(\lambda)w)]. \tag{33}$$

The following successive-approximation equation for λ is obtained from (33):

$$\lambda_{n+1} = 1/[1 + (w^T A^{-1}(\lambda_n) P_1 A^{-1}(\lambda_n) w / w^T A^{-1}(\lambda_n) w)]. \tag{34}$$

We return now to address the question of whether (34) gives the maximum x^{**} or the minimum[6] x^*, as depicted in Fig. 5. Consider (27). For $\lambda = 0$ in (27), the result is

$$x^* = P_1 P_1^{-1} \hat{x} = \hat{x}; \tag{35}$$

for $\lambda = 1$ in (27), the result is

$$x^* = P_2 P_2^{-1} \bar{x} = \bar{x}. \tag{36}$$

In (35)–(36), $\hat{x}$ and $\bar{x}$ are the centers of the two ellipsoidal parabolas where they are just tangent to the horizontal plane representing R^n in Fig. 5. Notice that the minimum occurs on the continuous curve in R^n for $x^*(\lambda)$, as a function of λ, where it also intersects the constraint set

$$G(x) = 0$$

of (21) [a condition embodied in satisfying (33)]. Since $x^*(\lambda^*)$, the true constrained minimum, occurs between

$$x^*(0) = \hat{x} \qquad \text{and} \qquad x^*(1) = \bar{x};$$

the correct value of λ satisfies the condition

$$0 < \lambda^* < 1. \tag{37}$$

Notice that the maximum occurs for $\lambda > 1$ in (27), with (33) also satisfied. The following Lemmas 5.1 and 5.2 and the associated Corollary 5.1 are easily proved. All proofs are given in the Appendix.

Lemma 5.1. The following result holds:

$$P_2(k) > P_1(k) \geq 0 \qquad \text{for} \quad k > 1. \tag{38}$$

It is easy to establish this fact rigorously, since it is intuitively obvious that the covariance of error of the filter with no measurements is *greater than* the covariance of error given the measurements (here, *greater than* is taken in the matrix sense as the positive definiteness of $P_2 - P_1$).

[6] A standard computational technique for solving the optimization problem, such as the Fletcher–Powell method (Ref. 17), was not used: It was seen that both x^{**} and x^* satisfy the necessary conditions for a minimum. It must be guaranteed that the computational method is going to the minimum at each iteration and not oscillating back and forth between going to a maximum and going to a minimum.

Lemma 5.2. If

$$0<\lambda_n<1, \tag{39}$$

$$P_2>P_1>0, \tag{40}$$

then

$$1/2<\lambda_{n+1}<1. \tag{41}$$

Corollary 5.1. Under the condition (40), if

$$0<\lambda_0<1, \tag{42}$$

then

$$1/2<\lambda_n<1 \tag{43}$$

for all $n \geq 1$.

The purpose of Lemma 5.1, Lemma 5.2, and Corollary 5.1 is to aid in establishing rigorously that, if the successive iteration formula (34) is reasonably initialized with a value of λ_0 in the interval $(1/2, 1)$ (say, $\lambda_0 = 0.75$), then the successive iterates also lie within this interval, where the correct value of λ for attaining the minimum also lies. The existence of this correct λ-value, which may be argued geometrically from Fig. 5, is established rigorously in the following lemma.

Lemma 5.3. There exists a fixed point λ^* in $(0, 1)$ such that

$$\lambda^* = g(\lambda^*), \tag{44}$$

where

$$g(\lambda) \triangleq 1/(1+(s(\lambda)/b(\lambda))), \tag{45}$$

with

$$s(\lambda) \triangleq w^T A^{-1}(\lambda) P_1 A^{-1}(\lambda) w, \tag{46}$$

$$b(\lambda) \triangleq w^T A^{-1}(\lambda) w. \tag{47}$$

Note that Lemma 5.3 is consistent with (33).

Now, the convergence of the successive iterations equation to the correct value λ^* will be investigated. The successive iterations equation (34) is of the form

$$\lambda_{n+1} = g(\lambda_n). \tag{48}$$

The objective is to establish that $g(\cdot)$ is a contraction mapping, i.e.,

$$|g(\lambda_{n+1})-g(\lambda_n)|<\xi|\lambda_{n+1}-\lambda_n|<\xi^n|\lambda_1-\lambda_0| \tag{49}$$

for some ξ such that

$$0<\xi<1. \tag{50}$$

It is sufficient to show that (see p. 32 of Ref. 18)

$$|\dot{g}(\lambda)|<1 \tag{51}$$

or, equivalently, that (see p. 32 of Ref. 18)

$$\dot{g}(\lambda)<1, \tag{52}$$

$$-1<\dot{g}(\lambda). \tag{53}$$

From the form of (45), the following expression for the derivative of $g(\cdot)$ with respect to λ is obtained:

$$\dot{g}(\lambda)=(s(\lambda)\dot{b}(\lambda)-b(\lambda)\dot{s}(\lambda))/(s(\lambda)+b(\lambda))^2. \tag{54}$$

Using (54), we can see that the conditions (52)–(53) are satisfied, or equivalently (51) is satisfied, as asserted in the following theorem.

Theorem 5.1. If the condition (38) holds, then

$$s(\lambda)\dot{b}(\lambda)-b(\lambda)\dot{s}(\lambda)<(s(\lambda)+b(\lambda))^2, \tag{55}$$

$$-(s(\lambda)+b(\lambda))^2<s(\lambda)\dot{b}(\lambda)-b(\lambda)\dot{s}(\lambda); \tag{56}$$

hence, $g(\cdot)$ of (45) is a contraction mapping by Theorem II.2.2 of Ref. 18. Consequently, λ^* of Lemma 5.3 is unique by Theorem II.1.3 of Ref. 18; and it is permissible to take the limit in (43) as $n\to\infty$ (since it exists) to yield

$$1/2<\lambda^*<1, \tag{57}$$

which is consistent with (37). In other words, the iteration algorithm (34), when properly initialized, converges to the correct λ^*.

The condition (38), that is sufficient for the conclusion of Theorem 5.1, is proved in Lemma 5.1. This is exactly the condition that exists when solving this associated Lagrange optimization problem in the context of failure detection using confidence regions.

A minimum rate of convergence for the successive approximations algorithm will now be established.

Theorem 5.2. The iteration algorithm of Eq. (34) converges at least linearly to λ^*.

In actual implementation, the performance of this algorithm indicated that the rate of convergence was greater than just linear. Since a linear rate of convergence has been established, Steffensen's method (Ref. 19) may be used to accelerate the rate of convergence.

There is also an interesting numerical problem (and a remedy) when the ideal mathematical conclusions associated with a contraction mapping (i.e., the conclusion that a unique fixed point as the limit of the successive approximations is the *solution* of the optimization problem) are altered slightly in doing numerical calculations on a computer having computer words composed of a finite bit size and having the consequent roundoff errors. In computer calculations, the successive iterations converge to a point that is within a *ball* centered about the *ideal fixed-point solution*, where the ball has a radius that is the roundoff error (Ref. 19). In the ellipsoid overlap computations, use of double-precision calculations reduces the occurrence of roundoff error enough to make the calculated solution for x^* *close enough* to the theoretical solution for practical purposes. The verification that this proposed remedy is sufficient was to evaluate (19) and (20) using $x = x^*$ and to check that the evaluations were identical; whereas, when there was a roundoff problem, the two separate evaluations did not agree exactly.

6. Summary

The main contribution of this paper is considered to be the two-ellipsoid overlap test which is fast enough for possible implementation in an INS real-time failure detection application. A scalar iteration equation was obtained for an associated Lagrange multiplier λ, used in obtaining x^*, the cornerstone of the CR2 ellipsoid overlap test.[7] The iteration equation (34) involved the inversion of a matrix sum [Eq. (25)] and was initially rather unwieldly analytically as convergence was investigated. This difficulty was surmounted, and a proof of convergence and the rate of convergence were established. The table in Fig. 7 summarizes the mechanization equations needed to implement this CR2 ellipsoid overlap test for failure detection in real time.

A summarizing overview is given in Fig. 8 on how this failure detection approach works. The theoretical basis of the CR2 failure detection approach is a generalization of the use of confidence intervals. The three main ideas that serve as the foundation for CR2 failure detection are shown as three different diagrams in Fig. 8 and are discussed below. These diagrams are shown in juxtaposition to facilitate a comparison of how the relative overlapping of the confidence regions affects the scalar test statistic at three specific check times (t_1, t_2, t_3). These confidence regions are portrayed in Fig. 8(a). At each check time, these confidence regions are elliptical. A *failure* is declared when the two confidence regions *do not overlap.*

[7] CR2 is an acronym for two confidence regions.

Processing Step	Calculations To Be Performed At Decision Time k
#1	Update: $P_2(k)=\Phi(k,k-1)P_2(k-1)\Phi^T(k,k-1) + Q(k)$ $\bar{x}(k) \equiv \Phi(k,k-1)\,\bar{x}(k-1)$
#2	Read $\hat{x}(k)$ and $P_1(k)$ [From Kalman Filter]
#3	Solve the Optimization Problem: (Actually use only the failure mode states) Form: $W \triangleq \hat{x}(k) - \bar{x}(k)$ $\lambda_o = \frac{3}{4}$ Iterate to Convergence: $\begin{cases} A_n \triangleq \left[(1-\lambda_n)P_2(k) + \lambda_n P_1(k)\right] \\ \lambda_{n+1} = \dfrac{1}{1 + \dfrac{W^T A_n^{-1} P_1(k) A_n^{-1} W}{W^T A_n^{-1} W}} \end{cases}$ Test: Stop When $\lvert\lambda_{n+1} - \lambda_n\rvert < 10^{-6}\lvert\lambda_n\rvert$ Fix Maximum Number of Iterations at 30. Solution is λ^*.
#4	Calculate Solution: $A^* = \left[(1-\lambda^*)P_2(k) + \lambda^* P_1(k)\right]$ $x^* = (1-\lambda^*)P_2 A^{*-1}\hat{x} + \lambda^* P_1 A^{*-1}\bar{x}$
#5	Perform Overlap Test: For $\ell(k) \triangleq (x^* - \hat{x})^T P_1^{-1}(x^* - \hat{x})$ If $\ell(k) > K_1$, Declare a failure If $\ell(k) \leq K_1$, Continue
#6	Proceed to Next Decision Time, k+1, Repeat steps #1-#6.

Fig. 7. Mechanization equations for CR2 failure detection.

At each check time t_i, the two elliptical cross sections of the confidence regions, shown in Fig. 8(a), are fixed levels of two parabolas, shown in Fig. 8(b). The problem is to determine whether these two ellipses overlap. In developing the real-time detection algorithm, the test for the presence or

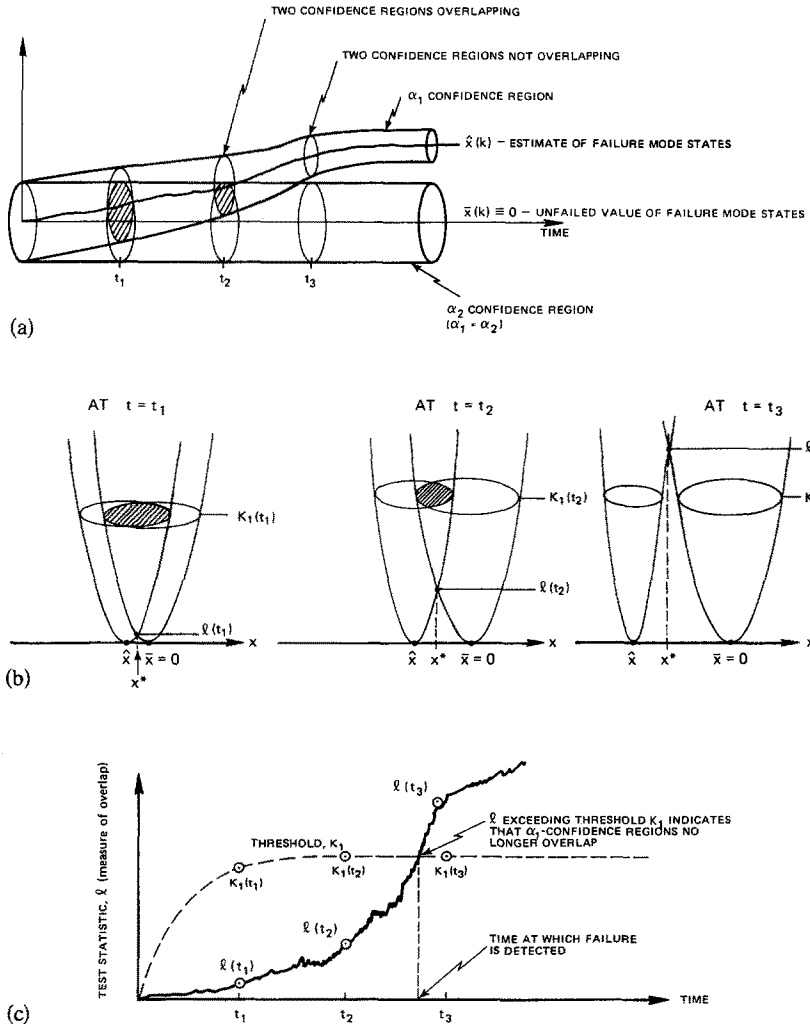

Fig. 8. Overview of procedure: (a) two confidence regions; (b) new optimization at each time instant; (c) CR2 failure detector.

absence of overlap was formulated as the solution of a minimization problem. The relative position of $l(t_i)$, the minimum point of the intersection of the two parabolas, to $K_1(t_i)$, the level that corresponds to the elliptical cross section of the confidence regions, determines if there is overlap (and, if so, the amount of overlap). As long as $l(t_i)$ is below $K_1(t_i)$, there is overlap; but, when $l(t_i)$ exceeds $K_1(t_i)$, then the confidence regions are disjoint and a

failure is declared. The relationship between the test statistic $l(t_i)$ and the decision threshold $K_1(t_i)$ is summarized in Fig. 8(c). It is sufficient to observe only the test statistic $l(t_i)$ and to declare failures when $l(t_i)$ exceeds $K_1(t_i)$.

In the CR2 failure detector, a higher level of the threshold K_1, to which the test statistic $l(t_i)$ is compared, effectively raises the heights of the ellipsoids in the associated optimization problem; this corresponds to stouter confidence regions. Analytic expressions which are used for prespecifying the time-varying decision threshold K_1 and the expressions for the instantaneous probabilities under H_0 and H_1 are derived in Ref. 10. The expressions are used in the setting of the threshold K_1 in a characteristic trade-off of instantaneous probability of false alarm versus the probability of correct detection associated with hypothesis-testing detection decisions. The probabilities of false alarm and correct detection that characterize the CR2 test over a time interval have been calculated using the level-crossing probability upper bound that is optimized in Ref. 20. The result of applying these upper bounds to the discrete-time CR2 technique is shown in Ref. 21.

It is hoped that this test for ellipsoid overlap will also be useful in areas other than failure detection (such as comparing simulation model results against experimental data when ellipsoidal confidence regions exist for both, Ref. 22).

7. Appendix

The following propositions constitute the rigorous working tools that were used both in the proofs of the lemmas and in the main convergence proof (Theorem 5.1).

Proposition 7.1. For symmetric, positive-definite matrices H and J, if

$$H < J, \tag{58}$$

then

$$H^{-1} > J^{-1}. \tag{59}$$

Proposition 7.2. For the conditions of Proposition 7.1.,

$$\alpha H < \alpha J, \tag{60}$$

for all scalar α such that $\alpha > 0$.

Proposition 7.3. For H as in Proposition 7.1, $H > 0$, then

$$H^{-1} > 0. \tag{61}$$

Proposition 7.4. For the conditions of Proposition 7.1 and M of full rank, if condition (58) holds, then

$$0 < M^T H M < M^T J M. \tag{62}$$

Proposition 7.5. For symmetric H and J as in Proposition 7.1, it follows that

$$[H+J]^T = [H+J] > 0. \tag{63}$$

Proposition 7.6. For H and J as in Proposition 7.1,

$$([H+J]^{-1})^T = [H+J]^{-1} > 0. \tag{64}$$

The above familiar propositions are easily proved and can be found in Ref. 23.

Proof of Lemma 5.1. Now, P_2 is the solution of (10) and P_1 is the solution of (7); hence, upon subtracting (7) from (10) and letting

$$N(k) \triangleq P_2(k) - P_1(k), \tag{65}$$

the following result is obtained:

$$\begin{aligned} N(k+1) = {} & \Phi(k+1, k) N(k) \Phi^T(k+1, k) \\ & + \Phi(k+1, k) P_1(k) H^T(k) [H(k) P_1(k) H^T(k) + R(k)]^{-1} \\ & \times H(k) P_1(k) \Phi^T(k+1, k). \end{aligned} \tag{66}$$

By the condition of observability and controllability (Ref. 24), it follows that

$$P_1(k) > 0. \tag{67}$$

Now, $H(k)$ is assumed to be of full rank; so, by Proposition 7.4,

$$H(k) P_1(k) H^T(k) > 0. \tag{68}$$

Adding the positive-definite matrix $R(k)$ to (68), inverting, and using Proposition 7.6 yields

$$[H(k) P_1(k) H^T(k) + R(k)]^{-1} > 0. \tag{69}$$

Note that, by the weaker (but more generally met) condition of detectability and stabilizability (Ref. 25), condition (69) directly holds without the requirement of proceeding from condition (67) to (69).

Premultiplying and postmultiplying (69) by

$$\Phi(k+1, k) P_1(k) H^T(k) \tag{70}$$

and its transpose, using Proposition 7.4 to conclude that the result is positive definite, and adding the (at worst) positive semidefinite quantity

$$\Phi(k+1, k)N(k)\Phi^T(k+1, k) \tag{71}$$

yields the right-hand side of (66), which is positive definite. Since the right-hand side of (66) is positive definite, by equality it follows that

$$N(k+1)>0 \tag{72}$$

for every $k>0$. At time $k=0$, the term (71) could only be assumed to be positive semidefinite, since

$$N(0)=P_2(0)-P_1(0)=P_0-P_0=0; \tag{73}$$

but the term (71) is positive definite for $k>0$, by induction on (72). By the definition (65), the result (38) of Lemma 5.1 follows. □

Proof of Lemma 5.2. If (39) holds, then

$$0<(1-\lambda_n)<1; \tag{74}$$

and, since (40) holds, by Proposition 7.2, the following inequality results:

$$(1-\lambda_n)P_2>(1-\lambda_n)P_1. \tag{75}$$

Adding $\lambda_n P_1$ to both sides of (75) yields

$$[(1-\lambda_n)P_2+\lambda_n P_1]>(1-\lambda_n)P_1+\lambda_n P_1=P_1>0. \tag{76}$$

Premultiplying and postmultiplying both sides by

$$w^T[(1-\lambda_n)P_2+\lambda_n P_1]^{-1} \tag{77}$$

and its transpose yields the relation

$$w^T[(1-\lambda_n)P_2+\lambda_n P_1]^{-1}w$$
$$>w^T[(1-\lambda_n)P_2+\lambda_n P_1]^{-1}P_1[(1-\lambda_n)P_2+\lambda_n P_1]^{-1}w>0. \tag{78}$$

Dividing through by the left-hand side, adding 1 to both sides, and taking reciprocals yields

$$1/2<1/(1+s(\lambda_n)/b(\lambda_n))<1, \tag{79}$$

where $s(\cdot)$ and $b(\cdot)$ are defined in (46) and (47). Now, the statement

$$1/2<\lambda_n<1 \tag{80}$$

is equivalent to (79) and is consistent with (37). □

Proof of Corollary 5.1. By Lemma 5.2, if $0<\lambda_0<1$, then

$$1/2<\lambda_1<1;$$

applying Lemma 5.2 again yields

$$1/2<\lambda_2<1.$$

The fact that (43) holds for all n follows by induction and by successively applying Lemma 5.2. □

Proof of Lemma 5.3. Using the definitions (25), (46), (47), and (45), it is seen that $g(\lambda)$ is continuous in λ for fixed P_1, P_2, and w. By Lemma 5.1 and Corollary 5.1, the continuous function $g(\cdot)$ maps the interval $[0, 1]$ into $[0, 1]$; hence, by Ref. 19, there is a fixed point λ^* such that (44) holds. □

Some additional propositions are needed in the proof of Theorem 5.1. Since these propositions are familiar, the proofs will be abbreviated.

Proposition 7.7. For $Z=Z^T>0$, there exists a matrix U such that

$$U^T=U^{-1}, \tag{81}$$

$$U^TZU=\Lambda \triangleq \operatorname{diag}(d_{11}, d_{22}, \ldots, d_{nn})>0. \tag{82}$$

Proof. Take U to be the normalized eigenvector matrix associated with the symmetric, positive-definite matrix Z; then, U has the property indicated in (81) and diagonalizes Z, as shown in (82). □

Proposition 7.8. For

$$Z \triangleq P_1^{\frac{1}{2}}P_2^{-1}P_1^{\frac{1}{2}}, \tag{83}$$

and with the condition (40) holding, then

$$I>\Lambda>0, \tag{84}$$

where Λ is defined in Proposition 7.7; specifically,

$$1>d_{ii}>0, \qquad i=1, \ldots, n. \tag{85}$$

Proof. Using (40) and Proposition 7.1 results in

$$P_1^{-1}>P_2^{-1}. \tag{86}$$

Premultiplying and postmultiplying both sides of (86) by $P_1^{\frac{1}{2}}$ and its transpose and using Proposition 7.4 yields

$$I>P_1^{\frac{1}{2}}P_2^{-1}P_1^{\frac{1}{2}}=Z>0. \tag{87}$$

Premultiplying and postmultiplying (87) by U defined in Proposition 7.7 yields

$$I > \Lambda = \text{diag}(d_{11}, d_{22}, \ldots, d_{nn}) > 0 \tag{88}$$

or, equivalently, (85). □

Proposition 7.9. If

$$0 < \lambda < 1, \tag{89}$$

$$0 < d_{ii} < 1, \tag{90}$$

then

$$0 < 1 + (d_{ii} - 1)\lambda < 1. \tag{91}$$

Proof. Subtracting 1 throughout (90) yields

$$-1 < d_{ii} - 1 < 0. \tag{92}$$

Multiplying (92) by λ, $\lambda > 0$, yields

$$-\lambda < (d_{ii} - 1)\lambda < 0. \tag{93}$$

Adding 1 to both sides yields

$$0 < 1 - \lambda < 1 + (d_{ii} - 1)\lambda < 1, \tag{94}$$

and (91) follows. □

Proposition 7.10. For

$$\Lambda = \text{diag}(d_{11}, \ldots, d_{nn}), \tag{95}$$

then

$$0 < \Lambda \tag{96}$$

iff

$$0 < d_{ii}, \qquad i = 1, \ldots, n. \tag{97}$$

Proof. This is obvious from the definition of positive definiteness. □

Proof of Theorem 5.1. For convenience of notation, let

$$s \triangleq w^T A^{-1} P_1 A^{-1} w \tag{98}$$

$$b \triangleq w^T A^{-1} w, \tag{99}$$

$$c \triangleq w^T A^{-1} P_1 A^{-1} P_1 A^{-1} w, \tag{100}$$

$$d \triangleq w^T A^{-1} P_1 A^{-1} P_2 A^{-1} w, \tag{101}$$

$$e \triangleq w^T A^{-1} P_2 A^{-1} w, \tag{102}$$

$$f \triangleq w^T A^{-1} P_2 A^{-1} P_2 A^{-1} w, \tag{103}$$

where A is defined in (25) and where it is noted that s, b, c, d, e, f are scalars resulting from a vector inner product operation. Define the symmetric, positive-definite, inner product matrices correspondingly to be

$$S \triangleq A^{-1} P_1 A^{-1}, \tag{104}$$

$$B \triangleq A^{-1}, \tag{105}$$

$$C \triangleq A^{-1} P_1 A^{-1} P_1 A^{-1}, \tag{106}$$

$$D \triangleq A^{-1} P_1 A^{-1} P_2 A^{-1}, \tag{107}$$

$$E \triangleq A^{-1} P_2 A^{-1}, \tag{108}$$

$$F \triangleq A^{-1} P_2 A^{-1} P_2 A^{-1}. \tag{109}$$

Notice that the terms that appear in the numerator of (54) may be expressed as

$$\dot{b} = -e + s, \tag{110}$$

$$\begin{aligned} s &= w^T A^{-1} P_1 A^{-1} A A^{-1} w = (1-\lambda) w^T A^{-1} P_1 A^{-1} P_2 A^{-1} w \\ &\quad + \lambda w^T A^{-1} P_1 A^{-1} P_1 A^{-1} w = (1-\lambda) d + \lambda c, \end{aligned} \tag{111}$$

$$\dot{s} = 2d - 2c, \tag{112}$$

$$b = (1-\lambda) e + \lambda s = (1-\lambda) e + \lambda (1-\lambda) d + \lambda^2 c, \tag{113}$$

and (110) can be rewritten, using (111), as

$$\dot{b} = -e + (1-\lambda) d + \lambda c. \tag{114}$$

Therefore, the left-hand side of (55) may be reexpressed, using (111)–(114), as

$$\begin{aligned} s\dot{b} - b\dot{s} &= [-(1-\lambda) de - \lambda ce + (1-\lambda)^2 d^2 + 2\lambda (1-\lambda) dc + \lambda^2 c^2] \\ &\quad - [2(1-\lambda) de - 2(1-\lambda) ce + 2\lambda (1-\lambda) d^2 + 2\lambda (-1+2\lambda) dc - 2\lambda^2 c^2] \\ &= -3(1-\lambda) de + (2-3\lambda) ce + (1-\lambda)(1-3\lambda) d^2 + 2(2-3\lambda) dc + 3\lambda^2 c^2. \end{aligned} \tag{115}$$

Using (111) and (113), the right-hand side of (55) may be expressed as

$$(s+b)^2=[(1-\lambda)e+(1-\lambda^2)d+\lambda(1+\lambda)c]^2=(1-\lambda)^2e^2+(1-\lambda^2)^2d^2 +\lambda^2(1+\lambda)^2c^2+2(1-\lambda^2)ce+2(1-\lambda)(1-\lambda^2)de +2\lambda(1+\lambda)(1-\lambda^2)dc. \tag{116}$$

On substituting (115)–(116) into (55), it is seen that the verification of the following inequality is equivalent to establishing (55):

$$(2-5\lambda+2\lambda^3)ce+(1-\lambda)(2\lambda^2-5)de+2\lambda(1-4\lambda+\lambda^2+\lambda^3)dc +\lambda^2(2-2\lambda-\lambda^2)c^2+\lambda(1-\lambda)(\lambda^2+\lambda-4)d^2<(1-\lambda)^2e^2. \tag{117}$$

Substituting

$$e=(1-\lambda)f+\lambda d \tag{118}$$

into (117) and moving all terms to the right-hand side yields

$$0<(1-\lambda)^4f^2+(1-\lambda)^2(5+2\lambda-4\lambda^2)fd+\lambda(1-\lambda)(9-4\lambda^2)d^2 -(1-\lambda)(2-5\lambda+2\lambda^3)cf-\lambda(4-13\lambda+2\lambda^2+4\lambda^3)dc+\lambda^2(\lambda^2+2\lambda-2)c^2, \tag{119}$$

which is a scalar relation. Stronger conditions on a matrix relation will now be established: these are sufficient conditions for the scalar relation (119) to hold, which is equivalent to (55) holding.

Factoring A^{-1} into the form

$$A^{-1}(\lambda)=[(1-\lambda)P_2+\lambda P_1]^{-1}=P_1^{-\frac{1}{2}}[(1-\lambda)P_1^{-\frac{1}{2}}P_2P_1^{-\frac{1}{2}}+\lambda I]^{-1}P_1^{-\frac{1}{2}}, \tag{120}$$

the following structure is exhibited by premultiplying and postmultiplying (106), (107), (109) by $P_1^{+\frac{1}{2}}$ to yield:

$$C' \triangleq P_1^{\frac{1}{2}}CP_1^{\frac{1}{2}} =[(1-\lambda)Z^{-1}+\lambda I]^{-1}[(1-\lambda)Z^{-1}+\lambda I]^{-1}[(1-\lambda)Z^{-1}+\lambda I]^{-1}, \tag{121}$$

$$D' \triangleq P_1^{\frac{1}{2}}DP_1^{\frac{1}{2}} =[(1-\lambda)Z^{-1}+\lambda I]^{-1}[(1-\lambda)Z^{-1}+\lambda I]^{-1}Z[(1-\lambda)Z^{-1}+\lambda I]^{-1}, \tag{122}$$

$$F' \triangleq P_1^{\frac{1}{2}}FP_1^{\frac{1}{2}} =[(1-\lambda)Z^{-1}+\lambda I]^{-1}Z^{-1}[(1-\lambda)Z^{-1}+\lambda I]^{-1}Z^{-1}[(1-\lambda)Z^{-1}+\lambda I]^{-1}, \tag{123}$$

where Z is defined in (83). Now premultiply and postmultiply (121)–(123) by

$$[(1-\lambda)Z^{-1}+\lambda I]$$

to yield

$$C'' \triangleq [(1-\lambda)Z^{-1}+\lambda I]^{-1}, \tag{124}$$

$$D'' \triangleq [(1-\lambda)Z^{-1}+\lambda I]^{-1}Z^{-1}, \tag{125}$$

$$F'' \triangleq Z^{-1}[(1-\lambda)Z^{-1}+\lambda I]^{-1}Z^{-1}. \tag{126}$$

Using Proposition 7.7, premultiply and postmultiply (124)–(126) by U in (81) to exhibit the following structure:

$$C''' \triangleq U^TC''U = [(1-\lambda)\Lambda^{-1}+\lambda I]^{-1} = \operatorname{diag}\{d_{ii}/(1+(d_{ii}-1)\lambda)\}, \tag{127}$$

$$D''' \triangleq U^TD''U = [(1-\lambda)\Lambda^{-1}+\lambda I]^{-1}\Lambda^{-1} = \operatorname{diag}\{1/(1+(d_{ii}-1)\lambda)\}, \tag{128}$$

$$F''' \triangleq U^TF''U = \Lambda^{-1}[(1-\lambda)\Lambda^{-1}+\lambda I]^{-1}\Lambda^{-1} = \operatorname{diag}\{1/d_{ii}(1+(d_{ii}-1)\lambda)\}. \tag{129}$$

By Propositions 7.7 and 7.9, the denominators of the elements of the above diagonal matrices are never zero.

A sufficient condition for the scalar relation (119) to hold is for the following stronger matrix condition to hold:

$$\begin{aligned}0 < {} & (1-\lambda)^4[F''']^2+(1-\lambda)^2(5+2\lambda-4\lambda^2)F'''D'''+\lambda(1-\lambda)(9-4\lambda^2)[D''']^2\\ & -(1-\lambda)(2-5\lambda+2\lambda^3)C'''F'''-\lambda(4-13\lambda+2\lambda^2+4\lambda^3)D'''C'''\\ & +\lambda^2(\lambda^2+2\lambda-2)[C''']^2.\end{aligned} \tag{130}$$

Substituting the right-hand sides of (127)–(129) into (130) and using Proposition 7.8 yields

$$\begin{aligned}0 < {} & (1-\lambda)^4\{1/d_{ii}^2(1+(d_{ii}-1)\lambda)^2\}+(1-\lambda)^2(5+2\lambda-4\lambda^2)\\ & \times\{1/d_{ii}(1+(d_{ii}-1)\lambda)^2\}\\ & +\lambda(1-\lambda)(9-4\lambda^2)\{1/(1+(d_{ii}-1)\lambda)^2-(1-\lambda)(2-5\lambda+2\lambda^3)\\ & \times\{1/(1+(d_{ii}-1)\lambda)^2\}\\ & -\lambda(4-13\lambda+2\lambda^2+4\lambda^3)\{d_{ii}/(1+(d_{ii}-1)\lambda)^2\}\\ & +\lambda^2(\lambda^2+2\lambda-2)\{d_{ii}^2/(1+(d_{ii}-1)\lambda)^2\}, \qquad i=1,\ldots,n.\end{aligned} \tag{131}$$

Multiplying (131) by

$$d_{ii}^2(1+(d_{ii}-1)\lambda)^2$$

yields

$$0<(1-\lambda)^4+(1-\lambda)^2(5+2\lambda-4\lambda^2)d_{ii}+(1-\lambda)(-2+14\lambda-6\lambda^3)d_{ii}^2$$
$$-\lambda(4-13\lambda+2\lambda^2+4\lambda^3)d_{ii}^3+\lambda^2(\lambda^2+2\lambda-2)d_{ii}^4, \qquad i=1,\ldots,n, \quad (132)$$

Let h_i be the following function of the real variables λ and d_{ii}:

$$h_i(\lambda, d_{ii}) \triangleq \text{right-hand side of (132)}, \qquad i=1,\ldots,n. \quad (133)$$

Then, (132) holds iff

$$h_i(\lambda, d_{ii})>0 \quad (134)$$

for

$$1/2<\lambda<1, \quad (135)$$

$$0<d_{ii}<1, \qquad i=1,\ldots,n. \quad (136)$$

That (134) holds under the restrictions (135)–(136) will now be established. Notice that

$$h_i(\lambda, 0)=(1-\lambda)^4>0, \quad (137)$$

$$h_i(\lambda, 1)=4>0, \quad (138)$$

under condition (135), $i=1,\ldots,n$. Also notice that

$$h_i(1, d_{ii})=d_{ii}^3(d_{ii}+3)>0, \quad (139)$$

$$h_i(1/2, d_{ii})=(1/16)+(5/4)d_{ii}+(17/8)d_{ii}^2+(3/4)(1-(1/4)d_{ii})d_{ii}^3>0, \quad (140)$$

under condition (136), $i=1,\ldots,n$. Now that the boundaries have been shown to be safe, the interior of the region described by (135)–(136) is examined.

Evaluating (133) at the midpoint of the rectangular region yields

$$h_i(3/4, 1/2)=2577/4096=0.629>0, \qquad i=1,\ldots,n. \quad (141)$$

So far, the function $h_i(\lambda, d_{ii})$ has been shown to be positive on the boundaries [except at $\lambda=1$, $d_{ii}=0$, where it is zero, but these boundaries are excluded anyway by (135)–(136)] and at the midpoint of the region. Since $h_i(\lambda, d_{ii})$ is continuous in λ and d_{ii}, it is possible to infer from the shape of the function $h_i(\lambda, d_{ii})$ that it is positive throughout the region specified by (135)–(136). Even before proceeding with the next analytical step in the proof, the author checked to see that what he wished to prove was true [i.e., that $h_i(\lambda, d_{ii})$ is positive over the region in question] by evaluating $h_i(\lambda, d)$ on the computer and observing that it is indeed positive for a range of values of λ and d spanning in incremental steps the region in question. Rearranging,

the function implicitly defined using (133) may be rewritten as

$$\begin{aligned} h_i(\lambda, d_{ii}) = (1+5d_{ii}-2d_{ii}^2)+(-4-8d_{ii}+16d_{ii}^2-4d_{ii}^3)\lambda \\ +(6-3d_{ii}-14d_{ii}^2+13d_{ii}^3-2d_{ii}^4)\lambda^2 \\ +(-4+10d_{ii}-6d_{ii}^2-2d_{ii}^3+2d_{ii}^4)\lambda^3 \\ +(-4d_{ii}^3+d_{ii}^4)\lambda^4, \qquad i=1,\ldots,n. \end{aligned} \tag{142}$$

The above function of two variables may be shown to be positive throughout the region of interest by a complicated application of Sturm's theorem (Ref. 26), which is normally for only one variable, by considering the other variable to be subsumed in the coefficients. The function (142) was directly shown to be positive by applying Bose's generalization of the Sturm test to a suitable test for the positivity of this function of two variables over the rectangular plate defined by (135)–(136). Therefore, the condition (134) holds.

The fact that the scalar relation (134) is satisfied is equivalent to (135) being satisfied; by Proposition 7.8, this is equivalent to the matrix relation (130) being satisfied. Premultiplying and postmultiplying (130) by

$$w^T P_1^{-\frac{1}{2}}[(1-\lambda)Z^{-1}+\lambda I]^{-1}U \tag{143}$$

and its transpose (this amounts to unraveling what has just been proved and relating it to the original problem) and using Proposition 7.4 results in the scalar relation (119) being satisfied, which was already shown to be equivalent to (117), and ultimately (55), being satisfied. The proof that (56) is satisfied follows by steps analogous to those used in proving (55).

Proof of Theorem 5.2. By Theorem 5.1.3 of Ref. 19, for λ^* such that (44) holds, it follows that

$$|\lambda_{n+1}-\lambda^*| = |g(\lambda_n)-g(\lambda^*)| < \xi|\lambda_n-\lambda^*|, \tag{144}$$

where ξ is the contraction constant obeying the condition (50). Forming the ratio by dividing through by the positive quantity $|\lambda_n-\lambda^*|$ yields

$$0<|\lambda_{n+1}-\lambda^*|/|\lambda_n-\lambda^*| < \xi|\lambda_n-\lambda^*|/|\lambda_n-\lambda^*| = \xi < 1. \tag{145}$$

Upon taking the lim sup (which exists because the sequence is bounded, Ref. 19), the result is

$$\limsup(|\lambda_{n+1}-\lambda^*|/|\lambda_n-\lambda^*|) < \xi < 1 \qquad \text{as } n \to \infty. \tag{146}$$

The fact that the lim sup of this quantity is less than one is just what is needed to prove, by the definition, that the iteration algorithm (34) converges at least linearly to λ^* (Ref. 16). □

References

1. KERR, T. H., *A Two Ellipsoid Overlap Test for Real Time Failure Detection and Isolation by Confidence Regions*, Proceedings of the Conference on Decision and Control, Phoenix, Arizona, 1974.
2. MEHRA, R. K., and PESCHON, J., *An Innovations Approach to Fault Detection and Diagnosis in Dynamic Systems*, Automatica, Vol. 7, No. 5, 1971.
3. JONES, H. L., *Failure Detection in Linear Systems*, MIT, Department of Aeronautics and Astronautics, PhD Thesis, 1973.
4. WILLSKY, A. S., and JONES, H. L., *A Generalized Likelihood Ratio Approach to State Estimation in Linear Systems Subject to Abrupt Changes*, Proceedings of the Conference on Decision and Control, Phoenix, Arizona, 1974.
5. HARRISON, J. V., and CHIEN, T. T., *Failure Isolation for a Minimally Redundant Inertial Sensor System*, IEEE Transactions on Aerospace and Electronics Systems, Vol. AES-11, No. 3, 1975.
6. BOOZER, D. D., and MCDANIEL, W. L., *On Innovation Sequence Testing of the Kalman Filter*, IEEE Transactions on Automatic Control, Vol. AC-17, No. 1, 1972.
7. MARTIN, W. C., and STUBBERUD, A. R., *An Additional Requirement for Innovations Testing in System Identification*, IEEE Transactions on Automatic Control, Vol. AC-19, No. 5, 1974.
8. ATHANS, M., *On the Elimination of Mean Steady-State Errors in Kalman Filters*, Proceedings of the Symposium on Nonlinear Estimation Theory, San Diego, California, pp. 212–214, 1970.
9. D'APPOLITO, J. A., and ROY, K. J., *Reduced Order Filtering with Applications in Hybrid Navigation*,. Proceedings of the IEEE Electronics and Aerospace Systems Conference, Washington, D.C., 1973.
10. KERR, T. H., *Implementing a Two Ellipsoid Overlap Test for Real Time Failure Detection*, IEEE Transactions on Automatic Control (to appear).
11. NASH, R. A., JR., KASPER, J. F., JR., CRAWFORD, B. S., and LEVINE, S. A., *Application of Optimal Smoothing to the Testing and Evaluation of Inertial Navigation Systems and Components*, IEEE Transactions on Automatic Control, Vol. AC-16, No. 6, 1971.
12. JAZWINSKI, A. H., *Stochastic Processes and Filtering Theory*, Academic Press, New York, New York, 1970.
13. SAGE, A. P., *Optimum Systems Control*, Prentice-Hall, Englewood Cliffs, New Jersey, 1968.
14. ANDERSON, T. W., *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, New York, 1958.
15. KERR, T. H., *Applying Stochastic Integral Equations to Solve a Particular Stochastic Modeling Problem*, University of Iowa, PhD Thesis, 1971.
16. LUENBERGER, D. G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1973.
17. FLETCHER, R., and POWELL, M. J. D., *A Rapidly Convergent Descent Method for Minimization*, Computer Journal, Vol. 6, No. 2, 1963.

18. HOLTZMAN, J. M., *Nonlinear System Theory*, Prentice-Hall, Englewood Cliffs, New Jersey, 1970.
19. BLUM, E. K., *Numerical Analysis and Computation Theory and Practice*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1972.
20. GALLAGER, R. G., and HELSTROM, C. W., *A Bound on the Probability that a Gaussian Process Exceeds a Given Function*, IEEE Transactions on Information Theory, Vol. IT-15, No. 1, 1969.
21. KERR, T. H., *False Alarm and Correct Detection Probabilities Over a Time Interval for Failure Detection Algorithms* (to appear).
22. KLEINMAN, D. L., and PERKINS, T. R., *Modeling Human Performance in a Time-Varying Anti-Aircraft Tracking Loop*, IEEE Transactions on Automatic Control, Vol. AC-19, No. 4, 1974.
23. ATHANS, M., and SCHWEPPE, F. C., *Gradient Matrices and Matrix Calculations*, Lincoln Laboratory, Lexington, Massachusetts, 1965.
24. LEE, E. B., and MARKUS, L., *Foundations of Optimal Control Theory*, John Wiley and Sons, New York, New York, 1967.
25. KWAKERNAAK, H., and SIVAN, R., *Linear Optimal Control Systems*, Wiley–Interscience, New York, New York, 1972.
26. VAN VALKENBURG, M. E., *Modern Network Synthesis*, John Wiley and Sons, New York, New York, 1964.
27. BOSE, N. K., *Test for Two-Variables Local Positivity with Applications*, Proceedings of the IEEE, Vol. 64, No. 9, 1976.