False Alarm and Correct Detection Probabilities over a Time Interval for Restricted Classes of Failure Detection Algorithms

THOMAS H. KERR, MEMBER, IEEE

Abstract—The statistical analysis of failure detection decisions in terms of the *instantaneous* probabilities of false alarm and correct detection for a specified failure magnitude at each check-time have previously been performed for several different failure detection techniques that utilize a Kalman filter. By performing a discrete-time specialization of a result of Gallager and Helstrom on a tightened upper bound for continuous-time level-crossing probabilities, upper bounds on the probabilities of false alarm and correct detection *over a time interval* have been obtained for the specific technique of CR2 failure detection (to allow an accounting for the effect of time correlations of the filter estimates). When these upper bounds are optimized to be as tight as possible to the desired probabilities, the resulting optimization problem for discrete-time is a collection of quadratic programming (QP) problems, which may easily be solved exactly without recourse to approximate solutions as were resorted to in the continuous-time formulation. This technique for evaluating tightened upper bounds on the false alarm and correct detection probabilities may be of general interest, since it can be applied to any failure detection technique or signal detection technique that can relate an exceeding of the deterministic decision threshold by the test statistic directly to a deterministic level being exceeded by a scalar Gaussian random process.

Manuscript received October 4, 1979; revised February 3, 1981. This work was supported by the Department of the Navy, through the Strategic Systems Project Office, SP-24, Contract N00030-76-C-0026.

The author was with The Analytic Sciences Corporation (TASC), Reading, MA 01867. He is now with Intermetrics, Inc., 733 Concord Ave., Cambridge, MA 02138.

I. INTRODUCTION

THE STATISTICAL analysis of a specific failure detection technique, the two confidence regions (CR2)

0018-9448/82/0700-0619$00.75

decision test [1] and [2], in terms of the trade-off of instantaneous probability of false alarm versus probability of correct detection as used in selecting the appropriate decision threshold to catch failures having a specified signal-to-noise ratio response, was completed in [3]. The long-term effect of CR2 failure detection on imperfect switching decisions, which affect availability/reliability, is analyzed for a precision inertial navigation system (INS) complex in [4] using the discrete-state Markov probability models of [5].[1]

Since the CR2 test statistic has a weighted chi-square distribution [3] and may be easily related to an underlying Gaussian distribution, the *instantaneous* probability that the test statistic falls above or below a specified threshold level at any check-time may be conveniently calculated under both failure and nominal conditions. However, the effect of time correlations of the Kalman estimate (that are also inherited by the test statistics by being a function of this estimate) may be evaluated by calculating the probabilities of false alarm and correct detection *over a time interval*. These are useful to quantify false alarm *rates* and to reflect expected *delays to detect*.

By using level-crossing theory, the probabilities of false alarm and correct detection, respectively, are related to the tendency of the CR2 test statistic to cross above the time-varying CR2 decision threshold over a time interval under hypotheses H_0 and H_1. The calculation of the exact probabilities associated with level-crossing problems are intractable in general [6, p. 264], [40], however, it is sometimes possible to prescribe tractable upper bounds on these elusive probabilities. In [7] a continuous-time bound and the associated optimization problem for making the bound as tight as possible are presented. The bound, which can be easily derived using only the rudiments of probability theory, is stated as [7, eq. 2]. The purpose of this paper is: to translate all of the important continuous-time ideas of [7] into their discrete-time analogs to more closely represent the intended computer implementation; to offer a solution procedure for the easier discrete-time optimization problem (demonstrated to be equivalent to a collection of standard quadratic programming problems) that arises naturally in making the bound tight; and to cast the results in the framework of a failure detection algorithm to match theory to application as tight upper bounds on the probability of false alarm and correct detection over a time interval are derived.

A summary of the CR2 implementation equations is presented in Section II to: provide a quick survey of the current status[2] of the CR2 algorithm; summarize how CR2 is used in failure detection; and demonstrate how the problem of interpreting the CR2 test statistic's crossing above the CR2 decision threshold over a time interval may be related to the level-crossing problem in making failure/no-failure decisions. A level-crossing technique for obtaining a tightened upper bound on the probabilities of false alarm and correct detection over a time interval is presented in Section III. Section IV highlights the inner problem structure which can be exploited to a computational advantage in performing calculations of probability of false alarm and correct detection over a time interval. A simple numerical example is given in Section IV to demonstrate the procedure of this paper and to show that there is *no* barrier to applying it to more complex situations of higher dimensions as has been done in contractor reports [43]. The conclusions are summarized in Section V. Apparent extentions of this approach to two other failure detection techniques that claim to possess the requisite Gaussian statistics are indicated in Section V, and applicability of the technique of this paper to two other more general signal detection techniques are discussed at the end of Section III.

II. Summary of CR2 Implementation Equations

The two confidence region approach to failure detection may be applied to systems that have either truth models or error models (as characteristic of nagivation systems [8]) that may be represented in the following linear state-variable form

$$x(k+1) = \Phi(k+1,k)x(k) + w(k) + \nu\delta_{k,\theta} \quad (1)$$

$$z(k) = H(k)x(k) + v(k), \quad (2)$$

where $w(k)$ and $v(k)$ are independent zero-mean Gaussian white noises having covariances of intensity $Q(k)$ and $R(k)$, respectively, and $x(0)$ is a Gaussian random vector initial condition, independent of the noises, of mean x_0 and variance P_0. The failure models that can be monitored using this CR2 approach must be modeled as states of the system of (1) (e.g., it is common practice to model soft or subtle failure candidates such as unwanted deleterious ramp and bias gyro drift rates and accelerometer biases as states in the linear error model of an INS [8]). then a failure may be represented[3] as a vector ν (magnitude and direction) that occurs at the unknown time θ as reflected in using the Kronecker delta $\delta_{k,\theta}$ which is one for $k=\theta$ and zero otherwise.

A Kalman filter, modeled on the unfailed system of (1) with the term $\nu\delta_{k,\theta}$ absent, is implemented to track the system that is to be monitored for failures. The Kalman filter uses the measurements $z(k)$ as inputs and the estimate $\hat{x}(k)$ evolves from a recursive equation which has the

[1]While other researchers first indicated that the standard reliability framework and techniques were not general enough [49]–[51] to accommodate analyses of the effect of failure detection on the systems management (of a control problem) more general than considered here, recent 1980 results [47] use exactly the same discrete-state Markov reliability techniques [5] as used in [4] in 1976 for the same type of problem.

[2]A discussion is provided on [46, p. 51] of the successful performance of a failure detection algorithm that evolved from the scalar CR2 approach and was implemented for Trident Submarine ESGM navigation [56] by Sperry Systems Management, Great Neck, NY.

[3]As discussed in [3], the two confidence region approach can also be used to detect more general time-varying failures, while [44] considers only detectability conditions for time invariant systems.

following form[4] [9]:

$$\hat{x}(k+1) = \Phi(k+1,k)\hat{x}(k) + K(k+1)[z(k+1) - H\Phi(k+1,k)\hat{x}(k)]$$
$$\hat{x}(0) = x_0 \tag{3}$$

with the covariance of error in estimation being provided from

$$P_1(k+1) = \Phi(k+1,k)[I - K(k)H(k)] \cdot P_1(k)\Phi^T(k+1,k) + Q(k+1)$$
$$P_1(0) = P_0, \tag{4}$$

and with the Kalman gain

$$K(k) = P_1(k)H^T(k)[H(k)P_1(k)H^T(k) + R(k)]^{-1}. \tag{5}$$

In order to use CR2 for failure detection, the expected value of the unfailed system and the associated variance available, without using the measurements, are needed and are provided, respectively, as solutions of the following two equations:

$$\bar{x}(k+1) = E[x(k+1) \mid H_0] = \Phi(k+1,k)\bar{x}(k)$$
$$\bar{x}(0) = x_0 \tag{6}$$

and

$$P_2(k+1) = \Phi(k+1,k)P_2(k)\Phi^T(k+1,k) + Q(k+1)$$
$$P_2(0) = P_0. \tag{7}$$

The underlying alternative hypotheses for the failure mode state x_i are

$$\text{(no-failure)} \quad H_0\colon\ \hat{x}_i(k,\omega) \sim N(0, [P_{\tilde{x}\tilde{x}}(k)]_{ii}) \quad \text{(zero mean)} \tag{8}$$

and

$$\text{(failure)} \quad H_1\colon\ \hat{x}_i(k,\omega) \sim N(d(k), [P_{\tilde{x}\tilde{x}}(k)]_{ii}) \quad \text{(nonzero mean)}, \tag{9}$$

where $d(k)$ is defined as

$d(k) = i$th component of the deterministic response of the filter to an assumed specific failure mode $\bar{\nu}$ that is to be detected[5] (10)

and

$P_{\tilde{x}_i\tilde{x}_i}(k)$ is the (i,i)th component $[P_{\tilde{x}\tilde{x}}(k)]_{ii}$ of
$P_{\tilde{x}\tilde{x}}(k) = P_2(k) - P_1(k) > 0 \quad$ for $k > 0$. (11)

Equation (11) can be obtained for the above scalar case by squaring the identity $x_i = (x_i - \hat{x}_i) + \hat{x}_i$, taking unconditional expectations throughout, and noting that the cross term is zero as a consequence of the projection theorem.

The scalar or one-dimensional CR2 decision rule based on the overlap or nonoverlap of confidence intervals, as derived in [3], is

NO OVERLAP (DECLARE FAILURE) WHEN

$$|\hat{x}_i(k)| > n(k) \cdot \left(\sqrt{[P_2(k)]_{ii}} + \sqrt{[P_1(k)]_{ii}}\right), \tag{12a}$$

OVERLAP (NO-FAILURE) WHEN

$$|\hat{x}_i(k)| \le n(k) \cdot \left(\sqrt{[P_2(k)]_{ii}} + \sqrt{[P_1(k)]_{ii}}\right), \tag{12b}$$

where $n(k)$ is a time-varying scaling factor for the standard deviations in (12).

An *alternate test* for the overlap of the two confidence intervals results from the artifice obtained by considering the associated parabolas of [3]. The CR2 decision threshold at time $= k$ is

$$K_1(k) \triangleq [n(k)]^2. \tag{13}$$

From [3, eq. 23], the one-dimensional CR2 test statistic at time $= k$ is

$$l(k) \triangleq [\hat{x}_i(k)]^2 / \left(\sqrt{[P_2(k)]_{ii}} + \sqrt{[P_1(k)]_{ii}}\right)^2 \tag{14}$$

and the *equivalent* alternative CR2 decision rule is

NO OVERLAP (DECLARE FAILURE) WHEN

$$l(k) > K_1(k) \tag{15a}$$

OVERLAP (NO-FAILURE) WHEN

$$l(k) \le K_1(k). \tag{15b}$$

While the obvious decision rule of (12) is easy to obtain as a test for the overlap/nonoverlap of confidence intervals, *it is the alternate formulation of the decision rule of* (15) *that generalizes to two or more dimensions* as discussed in [2] and [3]. For simplicity, the derivation of the probability of false alarm and correct detection over a time interval will be confined to the one-dimensional CR2 in this paper, without any severe loss of generality as indicated in Section V.

An equivalent representation for the one-dimensional CR2 decision threshold of (13) is given as [3, eq. (39b)] to be

$$K_1(k) = b^2 \cdot \left[\left(\sqrt{P_2(k)} - \sqrt{P_1(k)}\right) / \left(\sqrt{P_2(k)} + \sqrt{P_1(k)}\right)\right], \tag{16}$$

where

$b \triangleq$ the deterministic, pre-specified, constant[6] scale factor in the time-varying decision threshold of the CR2 algorithm. (17)

[4] The propagate and update equations of the discrete Kalman filter have been combined for brevity and conciseness.

[5] The time history of the parameter $d(k)$ may be evaluated through a simulation using the truth model with the system and measurement noise sample functions zeroed out and only the $\bar{\nu}$ failure mode activated.

[6] The policy of keeping the scale factor b constant is required to maintain a constant instantaneous probability of false alarm at each check-time $= k$, a policy that is routinely maintained in CFAR detectors [41] where the interference characteristics are either unknown *a priori* or vary with time, and conveniently facilitates later scaling in Section IV-B.

The effect of time-correlated estimates under H_0 is incorporated in

$$c_{km} \triangleq E[\hat{x}_i(k)\hat{x}_i(m)] = \begin{cases} [\Phi^{(k-m)}P_{\hat{x}\hat{x}}(m)]_{ii}, & \text{for } k > m \\ [P_{\hat{x}\hat{x}}(m)]_{ii}, & \text{for } k = m \\ [P_{\hat{x}\hat{x}}(k)\Phi^{T(k-m)}]_{ii}, & \text{for } k < m, \end{cases} \quad (18)$$

where Φ is the transition matrix over one check-time step of the associated filter, and the numbers c_{km} are elements of the following $(N_2 - N_1 + 1) \times (N_2 - N_1 + 1)$ matrix

$$C \triangleq \{c_{km}\}_{k=N_1 \quad m=N_1}^{k=N_2 \quad m=N_2}. \quad (19)$$

The expression of (18) for the serial time correlations of the estimates under H_0 is obtained by postmultiplying the iteration of (3) by $\hat{x}^T(k)$, taking expectations throughout, and utilizing the well-known property that the innovations term of zero mean within the bracket is uncorrelated with the estimate of the prior time-step k. The following time-varying threshold is defined for later convenience in the derivation of Section III:

$$s(k) \triangleq b\sqrt{[P_{\hat{x}\hat{x}}(k)]_{ii}}, \quad (20)$$

and the positive numbers $s(k)$ are elements of the following vector

$$s \triangleq [s(N_1), s(N_1+1), \cdots, s(N_2-1), s(N_2)]^T, \quad (21)$$

while the numbers $d(k)$ of (10) are elements of the following vector of time histories of the mean filter response to a specific failure

$$d = [d(N_1), d(N_1+1), \cdots, d(N_2-1), d(N_2)]^T. \quad (22)$$

III. A Level-Crossing Technique for a Tightened Upper Bound on P_{fa} and P_d over a Time Interval

The probabilities of false alarm and correct detection over a time interval will now be upper-bounded using a bounding technique from level-crossing theory. These upper bounds will be subsequently tightened through an optimization procedure.

The idea of approaching failure detection using level-crossing theory also appears in [10] where a geometric moving average algorithm is investigated. However, only the calculation of the expected level-crossing times, using many approximations, is attempted in [10] and a rough calculation of P_{fa} is attempted in the refinement of [11] which did not explore level-crossing bounds. Level-crossing theory is also used here in the derivation of expressions for the probabilities of false alarm and correct detection over a time interval for the CR2 failure detection algorithm.

The principal event that relates the scalar instantaneous decision rule of (12) to the crossing of the threshold level $s(k)$, defined in (20), over the time interval from time corresponding to N_1 to time corresponding to N_2 is[7]

$$A(\omega) = \{\omega \text{ such that for at least one } j, |\hat{x}_i(j,\omega)| \geq s(j), N_1 \leq j \leq N_2\}. \quad (23)$$

The probability of false alarm and probability of correct detection over the time interval from time corresponding to N_1 to time corresponding to N_2, respectively, are

$$P_{fa}[N_1, N_2] \triangleq \Pr[A(\omega) \mid H_0] \quad (24)$$

and

$$P_d[N_1, N_2] \triangleq \Pr[A(\omega) \mid H_1]. \quad (25)$$

However, all attention here will be focused on obtaining a tightened upper bound on $P_{fa}[N_1, N_2]$ since a lower bound on $P_d[N_1, N_2]$ is more appropriate as the respective worst case design consideration.

As suggested by Professor Alan S. Willsky of M.I.T., the principal event $A(\omega)$ may be decomposed into mutually exclusive and exhaustive constituent events (in a manner similar to what is done in the proof of Kolmogorov's inequality [12, p. 29]) as

$$A(\omega) = A_{N_1}(\omega) \cup \left\{ \bigcup_{j=N_1+1}^{N_2} A_j(\omega) \right\}, \quad (26)$$

where

$$A_{N_1}(\omega) \triangleq \{\omega \text{ such that } |\hat{x}_i(N_1,\omega)| > s(N_1)\} \quad (27)$$

and

$$A_j(\omega) \triangleq \{\omega \text{ such that } |\hat{x}_i(m,\omega)| \leq s(m) \text{ for } N_1 \leq m \leq j-1 \text{ and } |\hat{x}_i(j,\omega)| > s(j)\}, \quad (28)$$

(i.e., the set where j is the *first* time that $\hat{x}_i(\cdot,\omega)$ exceeds the threshold in the time interval from N_1 to N_2) and

$$A_j(\omega) \cap A_k(\omega) = \varnothing, \quad (29)$$

the null or empty set for $j \neq k$. It is useful to further decompose the event $A_j(\omega)$ more finely into positive and negative first crossings as

$$A_j(\omega) = A_j^+(\omega) \cup A_j^-(\omega), \quad (30)$$

where

$$A_j^+(\omega) \triangleq \{\omega \text{ such that } |\hat{x}_i(m,\omega)| \leq s(m) \text{ for } N_1 \leq m \leq j-1 \text{ and } \hat{x}_i(j,\omega) > s(j)\} \quad (31)$$

and

$$A_j^-(\omega) \triangleq \{\omega \text{ such that } |\hat{x}_i(m,\omega)| \leq s(m) \text{ for } N_1 \leq m \leq j-1 \text{ and } \hat{x}_i(j,\omega) < -s(j)\} \quad (32)$$

and, additionally,

$$A_j^+(\omega) \cap A_k^-(\omega) = \varnothing, \quad \text{for all } j \text{ and } k. \quad (33)$$

[7]Common notation, as in [7], is to suppress the ω, which represents a simple event in the underlying probability space, as it appears in $\hat{x}(k,\omega)$ and denote it as just $\hat{x}(k)$; but the ω will not be suppressed here because it is of fundamental importance in the supporting rigor.

Bounds on the Subevents—Bounds on each constituent subevent may be obtained by using the following two set inclusions

$$A_j^+(\omega) \subset \left\{\omega \text{ such that } \sum_{m=N_1}^{j} g(m)[\hat{x}_i(m,\omega) - s(m)] \geq 0\right\} \quad (34)$$

$$A_j^-(\omega) \subset \left\{\omega \text{ such that } \sum_{m=N_1}^{j} g(m)[\hat{x}_i(m,\omega) + s(m)] \leq 0\right\}, \quad (35)$$

where $g(m)$ will be defined[8] in (50). By the monotone property of positive probability measures [13], the following upper bounds hold:

$$\Pr\left[A_j^+(\omega) \mid H_0\right] \leq \Pr\left[\left\{\omega \text{ such that } \sum_{m=N_1}^{j} g(m)[\hat{x}_i(m,\omega) - s(m)] \geq 0 \middle| H_0\right\}\right] \quad (36)$$

and

$$\Pr\left[A_j^-(\omega) \mid H_0\right] \leq \Pr\left[\left\{\omega \text{ such that } \sum_{m=N_1}^{j} g(m)[\hat{x}_i(m,\omega) + s(m)] \leq 0 \middle| H_0\right\}\right]. \quad (37)$$

Paralleling the steps in discrete time that were followed in continuous time in [7] results in

$$y_j(\omega) \triangleq \sum_{m=N_1}^{j} g(m)\hat{x}_i(m,\omega), \quad (38)$$

being a Gaussian random variable as the deterministically weighted sum of samples of a Gaussian process with the mean being

$$E[y_j] \triangleq \sum_{m=N_1}^{j} g(m)E[\hat{x}_i(m,\omega)] = \begin{cases} 0, & \text{under } H_0 \\ d^T g, & \text{under } H_1, \end{cases} \quad (39)$$

and the variance under both H_0 and H_1 being

$$\sigma^2(g) \triangleq \text{var}[y_j] = \sum_{m=N_1}^{j} \sum_{n=N_1}^{j} g(m)E[\hat{x}_i(m)\hat{x}_i(n)]g(n) = g^T C g, \quad (40)$$

where C is defined in (18). Now let the parameter to be indirectly optimized through the specification of g be defined as

$$\eta_j(g) \triangleq s^T g/\sigma(g) \quad (41)$$

then

$$\Pr\left[\sum_{m=N_1}^{j} g(m)[\hat{x}_i(m) - s(m)] \geq 0 \mid H_0\right] = \Pr\left[y_j - \sigma(g)\cdot\eta_j \geq 0 \mid H_0\right] = \frac{1}{2}\text{erfc}\left[\eta_j/\sqrt{2}\right] \quad (42)$$

and

$$\Pr\left[\sum_{m=N_1}^{j} g(m)[\hat{x}_i(m) - s(m)] \geq 0 \mid H_1\right] = \Pr\left[y_j - \sigma(g)\cdot\eta_j \geq 0 \mid H_1\right] = \frac{1}{2}\text{erfc}\left[\left(\eta_j/\sqrt{2}\right) - d^T g/\sqrt{2}\,\sigma(g)\right]. \quad (43)$$

Upper-Bounding the Principal Event—A bound on the principal event, $A(\omega)$, in terms of upper bounds on all of the constituent pieces or subevents of which the principal event is comprised is obtained using the properties of probability measures on disjoint subsets and on nested sets as

$$P_{fa}[N_1, N_2] \triangleq \Pr[A(\omega) \mid H_0]$$

$$= \Pr\left[A_{N_1}(\omega) \cup \left\{\bigcup_{j=N_1+1}^{N_2} A_j(\omega)\right\} \middle| H_0\right] \quad (44a)$$

$$= \Pr\left[A_{N_1}(\omega) \mid H_0\right] + \Pr\left[\bigcup_{j=N_1+1}^{N_2} A_j(\omega) \middle| H_0\right] \quad (44b)$$

$$= \Pr\left[A_{N_1}(\omega) \mid H_0\right] + \Pr\left[\left(\bigcup_{j=N_1+1}^{N_2} A_j^+(\omega)\right) \cup \left(\bigcup_{j=N_1+1}^{N_2} A_j^-(\omega)\right) \middle| H_0\right] \quad (44c)$$

$$= \Pr\left[A_{N_1}(\omega) \mid H_0\right] + \sum_{j=N_1+1}^{N_2} \Pr\left[A_j^+(\omega) \mid H_0\right] + \sum_{j=N_1+1}^{N_2} \Pr\left[A_j^-(\omega) \mid H_0\right] \quad (44d)$$

$$\leq \text{erfc}\left[b/\sqrt{2}\right] + \sum_{j=N_1+1}^{N_2} \frac{1}{2}\text{erfc}\left[\eta_j/\sqrt{2}\right] + \sum_{j=N_1+1}^{N_2} \frac{1}{2}\text{erfc}\left[\eta_j/\sqrt{2}\right] \quad (44e)$$

$$= \text{erfc}\left[b/\sqrt{2}\right] + \sum_{j=N_1+1}^{N_2} \text{erfc}\left[\eta_j/\sqrt{2}\right]. \quad (44f)$$

[8] Notation for the ith component of the deterministic vectors s and g are $s(i)$ and $g(i)$, respectively.

Tightening the Upper Bound—Given $\underset{(j\times j)}{C}$ and $\underset{(j\times 1)}{s}$ of (18)–(21), respectively, the upper bound of (44) may be made tight by optimizing the deterministic vector quantity g for each j-dimensional subproblem to specify the deterministic parameter η_j through (41). The optimization problem, which solves for each of the η_j that minimize the right side of (44), may be equivalently reformulated analagous to the continuous-time reformation of [7] as several individual optimization problems of the following form (see Appendix III for verification): find the deterministic vector f such that

$$f_j \geq 0, \qquad \text{for all components of } f \tag{45}$$

and

$$f^T s' = h \tag{46}$$

(for arbitrary specified positive scalar h)[9] that minimizes

$$\sigma^2(f) = f^T C' f, \tag{47}$$

where C' is related to the original time-correlation effect of the Kalman estimates as summarized in the C of (18) by

$$c'_{mn} = \begin{cases} c_{mn} & \text{for } N_1 \leq m \leq j-1,\ N_1 \leq n \leq j-1 \\ -c_{mj}, & \text{for } N_1 \leq m \leq j-1,\ n = j \\ -c_{jn}, & \text{for } m = j,\ N_1 \leq n \leq j-1 \\ c_{jj}, & \text{for } m = j,\ n = j \end{cases} \tag{48}$$

and s' is related to the original effect of decision threshold level and instantaneous covariance as summarized in s of (20), (21) by

$$s'_m = \begin{cases} -s_m, & \text{for } 1 \leq m \leq j-1 \\ s_j, & \text{for } m = j. \end{cases} \tag{49}$$

The original g is related to f by

$$g(m) \triangleq \begin{cases} -f(m), & \text{for } N_1 \leq m \leq j-1 \\ f(j), & \text{for } m = j, \end{cases} \tag{50}$$

where f is arbitrary but nonnegative. Under the assumption that $\hat{x}_i(\cdot, \omega)$ does *not* have an improper Gaussian distribution [14, p. 15], then

$$C > 0. \tag{51}$$

Consequently, C' as defined in (48), which is just C with the signs reversed in the last row and column excepting the diagonal element, may easily be shown (as done in Appendix I) to also satisfy

$$C' > 0. \tag{52}$$

The problem of minimizing the strictly (via (52)) quadratic cost function of (47), subject to a single linear equality constraint of (46) and the nonnegativity constraint of (45), is recognized to be of the type known as quadratic programming (QP) problems.[10] There are several algorithms that may be used to solve this QP problem [15]–[17], [19], but only a well-documented M.I.T. program [18] has been extensively used by the author because of the convenience of its accessibility.[11] This M.I.T. program is essentially a revision of an earlier RAND Corporation implementation of the Dantzig algorithm [19].

From the convexity of each of the two constraints of (45) and (46) separately, it may be inferred that the intersection (satisfying both constraints simultaneously) is convex [20, p. 333]. Consequently, it may be concluded that the *local QP solution exists* and is a *global* solution since the convex cost function of (47) is being minimized over a convex region [20, p. 119].

The solution to each j-dimensional QP problem f^* may be converted to the corresponding g^* via (50) and finally used to calculate $\eta_j^*(g^*)$ via (41). When these values of $\eta_j^*(g^*)$ are used in the upper bound of (44), the result is the tightest upper bound *of this form* and in the CR2 failure detection application, has been very useful. However the techniques of deriving and evaluating these tightened upper bounds on the probability of false alarm over a time interval may be conveniently applied to any other failure detection or signal detection technique that can relate the exceeding of the decision threshold by the test statistic directly to a deterministic level being exceeded by a scalar Gaussian random process (viz., [53], [54]). Apparently, none of the other failure detection approaches of [10], [11], [21]–[36], [48], [53] have yet addressed or solved the problem of specifying or approximating probability of false alarm over a time interval.

IV. Performing Interval Probability Calculations

The expression for evaluating an upper bound on the probability of false alarm over a time interval was derived in Section III as (44).[12] The true probabilities over a time interval may be bracketed above and below[13] as

$$\max_{N_1 \leq k \leq N_2} P_{fa}(k) \leq P_{fa}[N_1, N_2] \leq \operatorname{erfc}\left[b/\sqrt{2}\right] + \sum_{j=N_1+1}^{N_2} \operatorname{erfc}\left[\eta_j/\sqrt{2}\right] \tag{53}$$

[9] That an arbitrary h yields the same unique solution to the optimization problem is asserted in [7] and explicitly demonstrated graphically here for two dimensions in Appendix II.

[10] In [7], the equivalent continuous-time problem involves integral equation inequalities and only alternative approximate procedures could be suggested as an approach to a solution.

[11] This software package was recommended by Dr. N. R. Sandell of Alphatech, Inc. Personal validation of [18] proceeded by utilizing several QP test problems with known solutions as provided in [15].

[12] The upper bound on probability of correct detection over a time interval, given in (54) is obtained in a manner completely analagous to the derivation of $P_{fa}[N_1, N_2]$ except that the argument of the error function includes terms due to the nonzero mean failure response of (39) under H_1. The results of the QP optimizations apply simultaneously to both (53) and (54).

[13] The inequality of the lower bound is an equality for $N_2 = N_1$.

and

$$\max_{N_1 \le k \le N_2} P_d(k) \le P_d[N_1, N_2] \le \frac{1}{2}\operatorname{erfc}\left[\left(b + d(N_1)/\sqrt{P_{\tilde{x}\tilde{x}}(N_1)}\right)/\sqrt{2}\right] + \frac{1}{2}\operatorname{erfc}\left[\left(b - d(N_1)/\sqrt{P_{\tilde{x}\tilde{x}}(N_1)}\right)/\sqrt{2}\right] + \frac{1}{2}\sum_{j=N_1+1}^{N_2} \operatorname{erfc}\left[\left(\eta_j/\sqrt{2}\right) - \boldsymbol{d}^T\boldsymbol{g}/\sqrt{2}\,\sigma(\boldsymbol{g})\right] + \frac{1}{2}\sum_{j=N_1+1}^{N_2} \operatorname{erfc}\left[\left(\eta_j/\sqrt{2}\right) + \boldsymbol{d}^T\boldsymbol{g}/\sqrt{2}\,\sigma(\boldsymbol{g})\right], \tag{54}$$

where $P_{fa}[N_1, N_2]$ and $P_d[N_1, N_2]$ are symbolic representations for the probability of false alarm and correct detection, respectively, over the time interval from time corresponding to N_1 to time corresponding to N_2. How the upper bound calculation was implemented for off-line evaluation is discussed in Section IV-A. The upper bounds, which only require use of the error function complement (erfc), are then optimized as discussed in detail in Section III, with implementation details presented in Section IV-A, to make the upper bound as tight as possible for a close approximation to the desired probabilities over a time interval. In Section V-B, it is shown that the probabilities over a time interval for any other choice of decision threshold and/or different failure magnitudes of the same failure mode may be obtained by just scaling the results of one stacked case optimization run. Finally, sample calculations of the form encountered in the specific inertial navigation system application of [3] and [4] are provided in Section IV-C.

A. Algorithmic Implementation of Optimization Solution

An algorithm to compute the solution of the QP problem, associated with the minimization of (47) while satisfying (45), (46), may be easily incorporated within a computer algorithm for calculating tight upper bounds as approximations to $P_{fa}[N_1, N_2]$ and $P_d[N_1, N_2]$ from the evaluation of the right sides of (53) and (54), respectively. Notice that the interval probabilities are calculated as functions of: C, the matrix of time correlations of the estimates; $\boldsymbol{s}$, a function of the threshold for the mechanized algorithm and covariances of the estimate; and $\boldsymbol{d}$, the mean of the signal response of the Kalman filter to a failure. Additionally, the underlying theoretical structure can be exploited to avoid unnecessary duplication of common computations by using the scaling properties that are discussed in Section IV-B.

B. Exploiting Parameterization to Avoid Unnecessary Additional Optimizations

1) *Use of Scaling in Calculating* $P_{fa}[N_1, N_2]$: it will now be shown that the results of the optimization of Section III are naturally scaled by the decision threshold b. This allows the results of optimizing for any b to be obtained from the results of optimizing for a single fixed value of b, chosen to be $b \equiv 1$ for convenience. The significance of the structural observations of Section IV-B1) are minor from a strictly theoretical point of view but are of tremendous practical value in demonstrating the otherwise nonobvious conclusion that the number of computer runs (i.e., expense) required to set b for a specified $P_{fa}[N_1, N_2]$ can be limited to just one.

By the definition of $\boldsymbol{s}$ in (20), (21), it is seen that $\boldsymbol{s}$ incorporates the quantity b, consequently, any arbitrary $\boldsymbol{s}$ may be represented as

$$\boldsymbol{s} = b \cdot \boldsymbol{s}'', \tag{55}$$

where

$$\boldsymbol{s}'' \text{ is what is obtained with } b \equiv 1. \tag{56}$$

Consequently, the unique solution for η_j obtained from the optimization algorithm, as in (41), may be represented for arbitrary b as

$$\eta_j = \boldsymbol{g}^T\boldsymbol{s}'/\sigma(\boldsymbol{g}) = b \cdot \boldsymbol{g}^T\boldsymbol{s}''/\sigma(\boldsymbol{g}) = b \cdot \eta_j', \tag{57}$$

where

$$\eta_j' \text{ is uniquely determined by the optimization algorithm for } b \equiv 1. \tag{58}$$

Using the representation of (57), the bound of (44) may be rewritten as

$$P_{fa}[N_1, N_2] \le \operatorname{erfc}\left[b/\sqrt{2}\right] + \sum_{j=N_1+1}^{N_2} \operatorname{erfc}\left[b \cdot \eta_j'/\sqrt{2}\right]. \tag{59}$$

When the parameters η_j are optimized via the quadratic programming algorithm, the following conservative[14] close approximation is obtained

$$P_{fa}[N_1, N_2] \approx \operatorname{erfc}\left[b/\sqrt{2}\right] + \sum_{j=N_1+1}^{N_2} \operatorname{erfc}\left[b \cdot \eta_j'/\sqrt{2}\right]. \tag{60}$$

2) *Use of Scaling in Calculating* $P_d[N_1, N_2]$: in a manner similar to the way η_j is reexpressed in (57), the optimal $\boldsymbol{f}$ that solves the system of (45), (46), and (47) may be represented as

$$\boldsymbol{f} = b \cdot \boldsymbol{f}', \tag{61}$$

where $\boldsymbol{f}'$ is the unique solution of the optimization problem for $b \equiv 1$. By the definition of $\boldsymbol{g}$ in (50), the relationship that is inherited by $\boldsymbol{g}$ from $\boldsymbol{f}$ is

$$\boldsymbol{g} = b \cdot \boldsymbol{g}', \tag{62}$$

[14]Any undesirable propensity to false alarm will never be evaluated as being less than what is actually present by using this tightened upper bound because the true probability of false alarm over the time interval can be no greater than what is given in (60). In providing P_{fa} and P_M (i.e., $1 - P_d$) expressions for the Schweppe likelihood ratio [42, part 3], optimized Chernoff upper bounds are taken to be equalities in a manner analagous to the approach of (60).

where g' is the solution of the optimization problem for $b \equiv 1$.

Since the error model of (1), (2) and the filter of (3)–(5) are linear systems, the vector of filter responses to arbitrary failure magnitudes may be represented by

$$d = \text{scale} \cdot d', \tag{63}$$

where d' is the filter response to a particular failure mode for a 100X Budget[15] failure. Now the argument of the erfc in the jth term in the first summation of the upper bound of (54) is

$$\begin{aligned}(\eta_j/\sqrt{2}) &- d^T g/\sqrt{2}\,\sigma(g)\\ &= (b\cdot \eta_j'/\sqrt{2}) - (\text{scale}\cdot d)^T(b\cdot g')/\sqrt{2}\,(b\cdot\eta_j')\\ &= (b\cdot\eta_j'/\sqrt{2}) - \text{scale}(d')^T g'/\sqrt{2}\,\eta_j'.\end{aligned} \tag{64}$$

Using the representation of (64), the upper bound of (54) may be rewritten as

$$\begin{aligned}P_d[N_1, N_2] &\le \frac{1}{2}\operatorname{erfc}\left[\left(b + d(N_1)/\sqrt{P_{\tilde{x}\tilde{x}}(N_1)}\right)/\sqrt{2}\right]\\ &= \frac{1}{2}\operatorname{erfc}\left(\left(b - d(N_1)/\sqrt{P_{\tilde{x}\tilde{x}}(N_1)}\right)/\sqrt{2}\right]\\ &+\frac{1}{2}\sum_{j=N_1+1}^{N_2}\operatorname{erfc}\left[(b\cdot\eta_j'/\sqrt{2})\right.\\ &\left.-\text{scale}\cdot(d')^T(g')/\sqrt{2}\,\sigma(g')\right]\\ &+\frac{1}{2}\sum_{j=N_1+1}^{N_2}\operatorname{erfc}\left[(b\cdot\eta_j'/\sqrt{2})\right.\\ &\left.+\text{scale}\cdot(d')^T(g')/\sqrt{2}\,\sigma(g')\right].\end{aligned} \tag{65}$$

Unlike the situation for $P_{fa}[N_1, N_2]$ in (60), the propensity for correct detection is *not* undesirable and the conservative bound to use for $P_d[N_1, N_2]$ is not the upper bound but the lower bound of (54). Use of the lower bound as a conservative approximation yields

$$P_d[N_1, N_2] \approx \max_{N_1 \le k \le N_2} P_d(k). \tag{66}$$

However, both upper and lower bounds can be effectively used to bracket the actual probability of correct detection. A numerical example for which the optimized upper bound expressions are evaluated is presented in Section IV-C. This example was selected to provide convenient numbers in order to avoid obscuring the evaluation technique that is being illustrated.

C. A Numerical Example

Determination of the constituent elements for an evaluation of the probability of false alarm over a time interval consisting of three equispaced[16] check-times will be demonstrated. The CR2 decision threshold scale factor b will be set to achieve a specified false alarm of 0.05 over the time interval. Given C and s'' as in (19) and (56), respectively, to be

$$\underset{(3\times 3)}{C} = \begin{bmatrix}2 & 1 & 1\\ 1 & 3 & 2\\ 1 & 2 & 4\end{bmatrix}\ (\text{normalized units})^2 \tag{67a}$$

$$s'' = [1 \quad 3 \quad 4]^T\ (\text{normalized units}). \tag{67b}$$

In evaluating (59) using the data of (67), two QP problems of dimension three and two, respectively, must be solved to obtain the proper η_2 and η_3 of (59). Using the transformation rules of (48) and (49), the three-dimensional quadratic program that must be solved minimizes (47) while satisfying (45), (46) with

$$C' = \begin{bmatrix}2 & 1 & -1\\ 1 & 3 & -2\\ -1 & -2 & 4\end{bmatrix} > 0 \tag{68a}$$

and

$$s' = [-1 \quad -3 \quad 4]^T, \tag{68b}$$

while the two-dimensional quadratic program that must be solved again minimizes (47) while satisfying (45), (46) with

$$C' = \begin{bmatrix}2 & -1\\ -1 & 3\end{bmatrix} > 0 \tag{69a}$$

and

$$s' = [-1 \quad 3]^T. \tag{69b}$$

For convenience, take

$$h = 10 > 0 \tag{70}$$

in (46) even though the evaluation of η_j is independent of the specific positive h value in (46) (as demonstrated in Appendix II for the two-dimensional case).

The solution vector of the three-dimensional QP problem corresponding to the f' of (61), for the matrices and vector of (68), is

$$f' = [0 \quad 0 \quad 1],^T \tag{71}$$

while the solution vector for the matrix and vector of the two-dimensional data of (69) is[17]

$$f' = [0 \quad 1]^T. \tag{72}$$

Utilizing (50) and (41) yields

$$\eta_3' = 2 \tag{73}$$

$$\eta_2' = \sqrt{3}. \tag{74}$$

Using the results of (73), (74), the optimally tightened

[15]Any convenient magnitude failure can be used for this purpose of calibration as long as the associated signal-to-noise ratio (SNR) is nondegenerate such as being greater than one.

[16]The procedure is unaltered when applied for check-times that are not equispaced.

[17]The QP solution vectors f' encountered in the actual INS application of [3], [43], while ranging from dimensions 16 to 2, also had the degenerate structure of a single nonzero entry as encountered in the above example. Rather than being faced with an arbitrary general QP problem to solve, the fact that the application described herein involves only QP problems with special internal structural dependencies (i.e., both C and s are functions of $P_{\tilde{x}\tilde{x}}(k)$ in the definitions of (18) and (20)) may perhaps explain the simplified form of the answer. However, the corresponding solution values for η_j are not degenerate simplifications.

upper bound of (59) for the data of (67) is

$$P_{fa}[N_1, N_2] \le \text{erfc}\left[b/\sqrt{2}\right] + \text{erfc}\left[b \cdot 2/\sqrt{2}\right] + \text{erfc}\left[b \cdot \sqrt{3}/\sqrt{2}\right], \quad (75)$$

which may be conservatively taken as a close approximation to $P_{fa}[N_1, N_2]$ via (60). The objective of setting the CR2 decision threshold scale factor b to make

$$P_{fa}[N_1, N_2] = 0.05 \quad (76)$$

may be conservatively met (so that $P_{fa}[N_1, N_2]$ is no more than 0.05) by evaluating the right side of (75) over a range of different selections of b, then interpolating to determine the specific b to achieve equality of the right side of (75) to 0.05. The value of the CR2 decision threshold scale factor b that makes the right side of (75) equal to 0.05 may be verified, using tables of erfc, to be

$$b = 1.98 \quad (77)$$

and for this value of b the upper and lower bounds of (53) form the following tight bracket as

$$0.0478 \le P_{fa}[N_1, N_2] \le 0.05. \quad (78)$$

The above numerical example demonstrated how to calculate the probability of false alarm over a time interval consisting of three check-times and involved solving $(3-1)=2$ QP problems of dimension three and two. The application of the same technique that is demonstrated in Section IV-C to the general problem of calculating the probability of false alarm over a time interval (as (60)) consisting of N check-times involves solving $N-1$ QP problems of dimensions $N, N-1, N-2, \cdots, 2$ corresponding to the proper submatrices of the full $N \times N$ matrix[18] of time correlations of the estimate in (19). The current QPS computer program [18] can easily accommodate these computations in one stacked-case run.

V. Summary and Conclusion

The motivation for calculating bounds on the probabilities of false alarm and correct detection over a time interval for a decision test for failure detection was presented in Section I. The implementation equations for the CR2 failure detection decision test were summarized in Section II to facilitate a demonstration of how the upper bound calculations may be implemented for a specific failure detection method. The method *may* also possibly be applicable to the generalized likelihood ratio (GLR) failure detection approach of [36] or any other failure detection or signal detection technique [54] that can relate the test statistic exceeding the decision threshold to a Gaussian random variable exceeding a deterministically specified level as in [29] and as *asserted* in [48, p. 608] and demonstrated [52, pp. 31–33] for simplified GLR. Major results of specifying the underlying theoretical framework and partitioning of events that enable use of a discrete-time formulation of the level-crossing upper bounds of [7] to this problem were discussed in Section III. How a particular QP algorithm has been implemented to evaluate the tightened upper bounds was discussed in Section IV where a representative numerical example is also presented.

So far, the presentation has dealt exclusively with upper-bounding the requisite probabilities over a time interval for the scalar or one-dimensional CR2 algorithm. The general multidimensional CR2 algorithm is specified in [1], [2], and [3] and can be easily handled by applying the techniques presented in this paper for the scalar case to each component. The final constant value of the threshold scaling parameter b that should ultimately be used in the multidimensional CR2 is

$$\bar{b} = \min_i b_i, \quad (79)$$

where i ranges over all of the components and b_i is the maximum scale that makes the probability of false alarm meet a prespecified design objective.

Occasionally, the so-called sampling approximation [37] is used as an approach to obtain a close approximation to the probability of false alarm over a time interval consisting of N consecutive check instants as (e.g., [38, p. D–9]):

$$P_{fa}[N_1, N_2] \simeq 1 - (1 - P_{fa})^N, \quad (80)$$

where $P_{fa} \triangleq$ instantaneous probability of false alarm. However, it is demonstrated in [37] using an analytically tractable level-crossing formulation for a simplified representation of a square-law detection device [39], that the approximation of (80) provides an indicated false alarm rate that is better (i.e., smaller) than is actually present ([37, p. 24]). A recent simulation approach for evaluating P_{fa} is reported in [55].

While at first glance, it may appear desirable to proceed to generalize the approach of event partitioning utilized in Section III to a continuous-time formulation for some other possible applications, the following technical problems will impede such a generalization. Since j in $A_j(\omega)$ of (28) refers to the first instant of time where $\hat{x}(t)$ exceeds the threshold, the corresponding continuous-time formulation of $A_{t_j}(\omega)$ would be indexed by the continuum $[t_0, t_f]$ in the generalization encountered in (44b) of

$$\bigcup_{t_j \in [t_0, t_f]} A_{t_j}(\omega) \quad (81)$$

as an uncountable (rather than countable) union of measurable sets, and consequently not guaranteed to be contained in the underlying sigma-algebra without further investigation. Additionally, the sigma-additive property of a positive finite probability measure (that is utilized in (44d) without fanfare) could not be invoked. Finally, the corresponding continuous-time optimization problem *cannot* be conveniently solved exactly as reported in [7], unlike the discrete-time case reported here.

An analytic model for an existing somewhat heuristic operational procedure used to randomize the external naviad fixes for updating the MK3 Mod6 ships inertial navigation system (SINS) of Poseidon submarines is re-

[18] For consistency, $N = N_2 - N_1 + 1$.

ported by this author in [57]. This navigation example is another neat application and slight extension of a fundamental level-crossing result previously derived by C. W. Helstrom in 1956 ([37], [39]). Recent confirming level-crossing results surprisingly similar to those of Helstrom (differing only in whether or not to employ square roots throughout) are reported in [58]. The derivations of [58] are refreshingly explicit.

Appendix I
Strict Convexity in Quadratic Program Preserved

In Section III the problem of upper-bounding the probabilities of false alarm and correct detection are transformed into problems of the same basic form for which upper bounds were constructed in [7]. The resulting problem should be a strictly convex quadratic program to be easily solved. That the resulting problems achieve this desirable property is demonstrated here for completeness via a straightforward three-line proof.

Lemma 1: For

$$f \neq 0 \tag{82}$$

the condition that

$$C > 0 \tag{83}$$

is equivalent to the conditions that

$$g \neq 0 \tag{84}$$

and

$$C' > 0. \tag{85}$$

Proof: If (82) holds, then (84) holds by the definition of g in (50). Since C is positive definite in (50), (83) holds and may be expanded to establish (85) using the definition of g in terms of f as

$$\begin{aligned} 0 < g^T C g &= \sum_{m=N_1}^{j} \sum_{n=N_1}^{j} g(m) c_{mn} g(n) \\ &= \sum_{m=N_1}^{j-1} \sum_{n=N_1}^{j-1} (-f(m)) c_{mn} (-f(n)) \\ &\quad + 2 \sum_{n=N_1}^{j-1} (-f(j)) c_{jn} f(n) + (-f(j)) c_{jj} (-f(j)) \\ &= \sum_{m=N_1}^{j-1} \sum_{n=N_1}^{j-1} f(m) c_{mn} f(n) + 2 \sum_{n=N_1}^{j-1} f(j)(-c_{jn}) f(n) \\ &\quad + f(j) c_{jj} f(j) = f^T C' f, \end{aligned} \tag{86}$$

where C' is defined as in (48). ■

Appendix II
Invariance of Solution with Respect to Positive Numerator Value h (for Dimension = 2)

Case 1—Ignoring only the nonnegative constraint of (45) in performing the optimization.

Given

$$\underset{(2\times 2)}{C} = C^T > 0 \tag{87}$$

$$\underset{(2\times 1)}{s} \geq 0 \quad \text{(nonnegative components)} \tag{88}$$

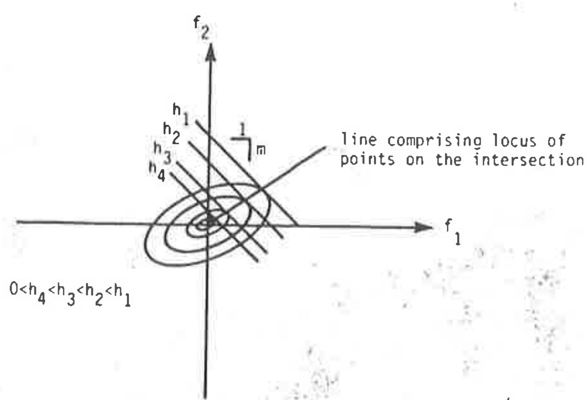

Fig. 1. Locus of points on the elliptical family subject the equality constraint.

as defined, respectively, in (18) and (49). The inner product of (47) as encountered in the optimization defines an elliptical family of the form:

$$f^T C f = K \tag{89a}$$

or

$$c_{11} f_1^2 + 2 c_{12} f_1 f_2 + c_{22} f_2^2 = K. \tag{89b}$$

Differentiating throughout (89b) with respect to f_1 yields

$$c_{11} f_1 + c_{12} f_2 + [c_{22} f_2 + c_{12} f_1] \frac{df_2}{df_1} = 0. \tag{90}$$

Rearranging the above, the equation of tangents to the elliptical family is

$$\frac{df_2}{df_1} = \frac{-c_{11} f_1 - c_{12} f_2}{c_{12} f_1 + c_{22} f_2}. \tag{91}$$

Now consider the equality constraint (46) which is in force as

$$f^T s' = h, \quad \text{(where } h \text{ is an unspecified positive constant)} \tag{92a}$$

or equivalently,

$$s_1' f_1 + s_2' f_2 = h \tag{92b}$$

or

$$f_2 = \frac{h}{s_2'} - \frac{s_1'}{s_2'} f_1 = \frac{h}{s_2'} + m f_1, \tag{93}$$

where

$$\text{slope} = m \triangleq \frac{-s_1'}{s_2'}. \tag{94}$$

Tangency of the family of equality constraints of the same slope to ellipses of the common family occurs when

$$\frac{df_2}{df_1} = \frac{-c_{11} f_1 - c_{12} f_2}{c_{12} f_1 + c_{22} f_2} = m \tag{95}$$

or

$$-c_{11} f_1 - c_{12} f_2 = m(c_{12} f_1 + c_{22} f_2). \tag{96}$$

Rearranging (96) it may be seen that the equation of the *locus* of tangents to the ellipses having constant slope is

$$f_2 = \frac{-[c_{11} + m c_{12}]}{[c_{12} + m c_{22}]} f_1. \tag{97}$$

Each point of the locus represents a solution to the optimization problem for a *different* value of h as indicated in Fig. 1.

The next step is to further investigate to see whether different values of h yield different values of η. The objective is to maximize η achieved by keeping numerator constant and minimizing the square of the denominator occurring in

$$\eta = \frac{f^T s'}{\sqrt{f^T C f}}. \tag{98}$$

After removing the radical from the denominator of (98) evaluating along the locus of tangents to the ellipse having constant slope, we have from (97) that:

$$f^T C f$$

$$= \begin{bmatrix} f_1 \\ -\left(\frac{c_{11}+mc_{12}}{c_{12}+mc_{22}}\right) f_1 \end{bmatrix}^T \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix} \begin{bmatrix} f_1 \\ -\left(\frac{c_{11}+mc_{12}}{c_{12}+mc_{22}}\right) f_1 \end{bmatrix} \tag{99a}$$

$$= f_1^2 \left\{ c_{11} - 2\frac{(c_{11}+mc_{12})}{(c_{12}+mc_{22})} c_{12} + c_{22}\frac{(c_{11}+mc_{12})^2}{(c_{12}+mc_{22})^2} \right\} \tag{99b}$$

$$= \frac{f_1^2}{(c_{12}+mc_{22})^2} \{ c_{11}(c_{12}+c_{22}m)^2 - 2c_{12}(c_{11}+c_{12}m) \cdot (c_{12}+c_{22}m) + c_{22}(c_{11}+c_{12}m)^2 \} \tag{99c}$$

$$= \frac{f_1^2}{(c_{12}+c_{22}m)^2} \{ c_{11}^2 c_{22} - c_{11}c_{12}^2 + 2mc_{11}c_{12}c_{22} + m^2 c_{11}c_{22}^2 - 2mc_{12}^3 - m^2 c_{12}^2 c_{22} \} \tag{99d}$$

$$= \frac{f_1^2}{(c_{12}+c_{22}m)^2} \{ (c_{11}c_{22} - c_{12}^2)(c_{11} + 2mc_{12} + m^2 c_{22}) \}. \tag{99e}$$

Now the first term within the braces of (99e) is positive as

$$c_{11}c_{22} - c_{12}^2 > 0 \tag{100}$$

by the positive definite requirement of (51). The nonnegativeness of the second term of (100) is now established by considering the following question: Is $z > 0$?, where

$$z \triangleq m^2 + 2\frac{c_{12}}{c_{22}} m + \frac{c_{11}}{c_{22}}. \tag{101}$$

Further investigation reveals that the roots of the quadratic equation of (101) are

$$m_{1,2} = \frac{-2c_{12} \pm \sqrt{4(c_{12}^2 - c_{11}c_{22})}}{2c_{22}}. \tag{102}$$

There are no real roots (i.e., the parabola does not dip below the horizontal m-axis) when the discriminant is

$$c_{12}^2 - c_{11}c_{22} < 0 \quad \text{or equivalently} \quad c_{12}^2 < c_{11}c_{22} \quad (\text{i.e., } C > 0), \tag{103}$$

so z is always positive for all m by the naturally arising practicality condition of (100).

Using the above results, η can be evaluated from (98) as

$$\eta = \frac{f^T s}{\sqrt{f^T C f}} \tag{104a}$$

$$= \frac{s_1 f_1 + s_2 \left[-\frac{(c_{11}+c_{12}m)}{(c_{12}+c_{22}m)} \right] f_1}{\frac{f_1}{(c_{12}+c_{22}m)} \{ (c_{11}c_{22} - c_{12}^2)(c_{11} + 2mc_{12} + m^2 c_{22}) \}^{1/2}} \tag{104b}$$

$$= \frac{s_2 \left(-m - \frac{c_{11}+c_{12}m}{c_{12}+c_{22}m} \right)}{\frac{f_1}{(c_{12}+c_{22}m)} \{ (c_{11}c_{22} - c_{12}^2)(c_{11} + 2mc_{12} + m^2 c_{22}) \}^{1/2}} \tag{104c}$$

$$= \frac{-s_2(c_{22}m^2 + 2c_{12}m + c_{11})}{\{ (c_{11}c_{22} - c_{12}^2)(c_{11} + 2mc_{12} + m^2 c_{22}) \}^{1/2}}. \tag{104d}$$

Notice that the numerator of (104d) depends only on the C (characterizing the elliptical family) and the slope m. Notice also that (104d) is independent of h as had been asserted and was the goal that has now been analytically demonstrated.

Case 2—Investigate invariance of optimization solution (*wrt* h) while observing the constraint on nonnegativity.

$$\eta = \frac{s^T f}{\sqrt{f^T C f}} = \frac{h}{\sqrt{\begin{bmatrix} 0 & \frac{h}{s_2} \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix} \begin{bmatrix} 0 \\ \frac{h}{s_2} \end{bmatrix}}} \tag{105a}$$

$$= \frac{h}{\sqrt{\frac{c_{22}h^2}{(s_2^2)}}} = \frac{s_2 h}{h\sqrt{c_{22}}} = \frac{s_2}{\sqrt{c_{22}}}. \tag{105b}$$

Notice again, as with (104d), (105b) is independent of the particular value of h as had been asserted here and in [7].

Appendix III
Conversion of Indicated Optimization to a Tractable QP Problem

To make the upper bound terms of (44f) (corresponding to the shaded area under the Gaussian curve of Fig. 2 smaller (i.e., achieving a tight upper bound) choose g to satisfy the following:

$$\max_{g} \eta_j(g) \text{ subject to constraints of (45) on } g \text{ via its definition of (50) in terms of } f, \tag{106}$$

where $\eta_j(g)$ is defined in (41) and is a ratio. As similarly encountered in one particularly lucid alternative derivation of the matched filtering formulation of communications theory [45, p. 140], the ratio of (41) may be maximized by constraining the numerator to be a constant, say h, while minimizing the denominator as

$$\min_{f} \sigma(f) = \sqrt{f^T C' f} \tag{107}$$

subject to the constraints of both (45) and (46).

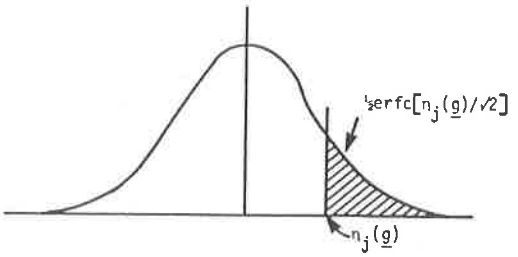

Fig. 2. Maximizing $n_j(g)$ tightens upper bound of (44).

Rather than deal with the radical appearing in (107), an equivalent optimization is

$$\min_f \sigma^2(f) = f^T C f \tag{108}$$

subject to the same constraints of (45) and (46).

While equivalence of the continuous-time analogs of the above three optimization formulations was recognized in [7], the continuous-time formulation of (108) involves the minimization of an iterated integral, subject to both an inequality and *integral* equality constraints. As noted in [7], such infinite-dimensional optimizations are not tractably solvable exactly, while the discrete-time formulation presented in this paper replaces integrals of [7] with summations, then goes further to use vector-matrix notation to recognize the inner product of (108) and the linear equality constraint of (46) that constitute a QP problem, with an easily accessible *exact* solution. Just as the majority of Kalman filtering applications are digitally implemented, it is only the tractable *discrete-time formulation* of the level-crossing problem that is of interest here as has been digitally implemented for the CR2 application, with extensive simulation and real data results in [43], [56].

References

[1] T. H. Kerr, "A two ellipsoid overlap test for real-time failure detection and isolation by confidence regions," *Proc. Conf. Decision and Control*, pp. 735–742, 1974.

[2] ——, "Real-time failure detection: A static nonlinear optimization problem that yields a two ellipsoid overlap test," *J. Optimiz. Theory Appl.*, vol. 22, no. 4, Aug. 1977.

[3] ——, "Statistical analysis of a two ellipsoid overlap test for real time failure detection," *IEEE Trans. Automat. Contr.*, vol. AC-25, no. 4, pp. 762–773, Aug. 1980.

[4] ——, "Failure detection aids for human operator decisions in a precision inertial navigation system complex," *Proc. Symp. Appl. Decision Theory Probl. Diagnosis and Repair*, sponsored by The American Statistical Association, Fairborn, OH, June 2–3, 1976.

[5] M. L. Shooman, *Probabilistic Reliability: An Engineering Approach*. New York: McGraw-Hill, 1968.

[6] E. Wong, "Recent progress in stochastic processes: A survey," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 3, May 1973.

[7] R. G. Gallager and C. W. Helstrom, "A bound on the probability that a Gaussian process exceeds a given function," *IEEE Trans. Inform. Theory*, vol. IT-15, no. 1, Jan. 1969.

[8] R. A. Nash, *et al.*, "Application of optimal smoothing to the testing and evaluation of inertial navigation systems and components," *IEEE Trans. Automat. Contr.*, vol. AC-16, no. 6, Dec. 1971.

[9] A. Gelb, Ed., *Applied Optimal Estimation*. Cambridge, MA: M.I.T., 1974.

[10] W. G. S. Hines, "A simple monitor of a system with sudden parameter changes," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 2, Mar. 1976.

[11] ——, "Improving a simple monitor of a system with sudden parameter changes," *IEEE Trans. Inform. Theory*, vol., IT-22, no. 4, July 1976.

[12] J. Lamperti, *Probability: A Survey of the Mathematical Theory*. New York: W. A. Benjamin, 1966.

[13] W. Rudin, *Real and Complex Analysis*. New York: McGraw-Hill, 1966.

[14] I. I. Gikhman and A. V. Skorokhad, *Introduction to the Theory of Random Processes*. Philadelphia, PA: W. B. Saunders, 1969.

[15] J. C. G. Boot, *Quadratic Programming: Algorithms—Anomalies—Applications*. Chicago, IL, Rand McNally, 1964.

[16] R. Fletcher, "A FORTRAN subroutine for general quadratic programming," *Harwell Rep.*, no. AERE-R.6370, United Kingdom Atomic Energy Association, June 1970.

[17] ——, "The calculation of feasible points for linearly constrained optimization problems," *Harwell Rep*, no. AERE-R.6354, United Kingdom Atomic Energy Association, Apr. 1970.

[18] "Quadratic programming system (QPS)," *Applic. Progr. Descript.*, AP-53; Publication of the Massachusetts Institute of Technology Information Processing Center, Apr. 1970.

[19] G. Dantzig, *Linear Programming and Extensions*. Princeton, NJ: Princeton University, 1963, pp. 490–496.

[20] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1973.

[21] P. M. Newbold and Y. C. Ho, "Detection of changes in the characteristics of a Gauss-Markov process," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-4, no. 5, Sept., 1968.

[22] R. K. Mehra and J. Peschon, "An innovations approach to fault detection and diagnosis in dynamic systems," *Automatica*. vol. 7, 1971.

[23] T. T. Chien, "An adaptive technique for a redundant sensor navigation system," *Proc. AIAA Guidance and Control Conf.*, Aug. 1972.

[24] H. L. Jones, "Failure detection in linear systems," Ph.D. dissertation, Dep. Aeronaut. and Astronaut., Massachusetts Institute of Technology, Cambridge, MA, 1973.

[25] H. Rome, "Automated degradation detection and isolation for SINS," *Proc. AIAA Guidance and Control Conf.*, Key Biscayne, FL, Aug., 1973.

[26] R. B. Broen, "A nonlinear voter-estimator for redundant systems," *Proc. 1974 IEEE Conf. Decision and Control*, Phoenix, AZ, 1974.

[27] A. S. Willsky and H. L. Jones, "A generalized likelihood ratio approach to state estimation in linear systems subject to abrupt changes," *Proc. Conf. Decision and Control*, Phoenix, AZ, Nov. 20–22, 1974.

[28] M. H. A. Davis, "The application of nonlinear filtering to failure detection in linear systems," *IEEE Trans. Automat. Contr.*, Apr. 1975.

[29] E. G. Gai and R. E. Curry, "Failure detection by pilots during automatic landing: Models and experiments," *Proc. Eleventh Ann. Conf. Manual Control*, Ames Research Center, Moffett Field, CA, May, 1975.

[30] J. V. Harrison and T. T. Chien, "Failure Isolation for a Minimally Redundant Inertial Sensor System," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-11, no. 3, May 1975.

[31] A. S. Willsky, J. J. Deyst, and B. S. Crawford, "Adaptive filtering and self-test methods for failure detection and compensation," *J. Spacecr. Rockets*, vol. 12, July, 1975.

[32] R. Montgomery, "Failure detection and control system reconfiguration," presentation made at MIT-NASA/Ames Workship on Systems Reliability Issues for Future Aircraft, Massachusetts Institute of Technology, Cambridge, MA, 18–22 Aug. 1975.

[33] R. B. Broen, "Performance of fault tolerant estimators in a noisy environment," *Proc. AIAA Guidance and Control Conf.*, Boston, MA, Aug. 20–22, 1975.

[34] J. J. Deyst and J. C. Deckert, "Application of likelihood ratio methods to failure detection and identification in the NASA F-8 DFBW aircraft," *Proc. Conf. Decision and Control*, Houston, TX, Dec., 1975.

[35] ——, "Maximum likelihood failure detection techniques applied to the shuttle RCS jets," *J. Spacecr. Rockets*, vol. 13, no. 2, Feb., 1976.

[36] A. S. Willsky and H. L. Jones, "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-21, no. 1, Feb., 1976.

[37] C. W. Helstrom, "The accuracy of probability distributions computed by the sampling approximation," Westinghouse Research Rep. 8-1259-R2, May 1956.

[38] K. S. Tait, E. F. Toohey, and C. J. Vahlberg, "Improved guidance study results: 2nd interim report," The Analytic Sciences Corporation, Tech. Rep. TR-836-2, Nov. 1977 (Confidential).

[39] C. W. Helstrom, "Two notes on a Markov envelope process," *IRE Trans.*, vol. IT-5, pp. 139–140, Sept. 1959.

[40] T. Mimaki and T. Munakata, "Experimental results on the level-crossing intervals of Gaussian processes," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 4, pp. 515–519, July 1978.

[41] R. Nitzberg, "Low-loss constant false alarm rate processors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-15, no. 5, pp. 719–723, Sept. 1979.

[42] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Parts I (and 3)*. New York: John Wiley and Sons, 1968 (and 1971).

[43] T. H. Kerr, "Improving ESGM failure detection in the SINS/ESGM system (U)," TASC Tech. Rep. TR-678-3-1, Oct. 1976 (Confidential).

[44] A. K. Caglayan, "Necessary and sufficient conditions for detectability of jumps in linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-25, no. 4, pp. 833–834, Aug. 1980.

[45] J. C. Hancock, *The Principles of Communication Theory*. New York: McGraw-Hill, 1961.

[46] "FBM weapons systems steering task group meeting," SSPO, Dep. of the Navy, Washington, DC, 17 Aug. 1977 (SECRET).

[47] J. V. Harrison, K. C. Daly, and E. Gai, "Reliability and accuracy prediction for a redundant strapdown navigator," *Proc. AIAA Guidance and Control Conf.*, Danvers, MA, pp. 403–413, 11–13 Aug. 1980.

[48] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, Nov. 1976.

[49] J. D. Birdwell and M. Athans, "On the relationship between reliability and linear quadratic optimal control," *Proc. Conf. Decision and Control*, New Orleans, LA, pp. 129–134, 7–9 Dec. 1977.

[50] M. Athans, "The role of operations research in systems theory and applications," Rep. LIDS-R-922, Massachusetts Institute of Technology, 28 June 1979.

[51] H. J. Chizeck and A. S. Willsky, "Towards fault-tolerant optimal control," *Proc. Conf. Decision and Control*, New Orleans, LA, pp. 19–20, 7–9 Dec. 1977.

[52] E. Y. Chow, "Analytical studies of the generalized likelihood ratio technique for failure detection," M.S. thesis, Massachusetts Institute of Technology, Feb. 1976.

[53] I. V. Nikiforov, "Cumulative sums for detection of changes in random process characteristics," *Automat. Remote Contr.*, vol. 40, no. 2, pp. 192–202, translation July 1979 (original Feb. 1979).

[54] R. D. Martin and C. P. McGath, "Robust detection of stochastic signals," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 4, pp. 537–541, July 1974.

[55] Y. K. Chan and R. W. Edsinger, "A correlated random number generator and its use to estimate false alarm rates of airplane sensor failure detection algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-26, no. 3, pp. 676–680, June 1981.

[56] T. H. Kerr, "TASC refinements of SINS/ESGM failure detection procedures," TASC slide presentation SP-889-3-1, for Sperry Systems Management, Feb. 1977.

[57] ——, "Modeling and evaluating an empirical INS difference monitoring procedure used to sequence SSBN naviad fixes," *Navigation: J. Inst. Navigation*, vol. 28, no. 4, Winter 1981–1982.

[58] R. Barakat, "The level-crossing rate and above level durations of the intensity of a Gaussian random process," *Inform. Sci.*, vol. 20, pp. 83–84, 1980.