Analysis of LTV Fault Detection Schemes with Additive Faults

Timothy J. Wheeler, Peter Seiler, Andrew K. Packard, and Gary J. Balas

***Abstract*— This paper considers the problem of certifying the performance of a class of model-based fault detection schemes. The underlying plant is assumed to be a linear time-varying (LTV) system subject to a Markov-switching fault input. The fault detection scheme consists of two parts: an LTV component that produces a scalar residual and a static nonlinear function that infers the presence of a fault based on this residual. Probabilistic performance metrics are presented and the complexity of computing these metrics is analyzed. It is shown that under a set of realistic assumptions, this complexity is reduced to polynomial time. An aerospace example, involving a pitot-static probe subject to random bias faults, is used to demonstrate the usefulness of this analysis.**

I. INTRODUCTION

In safety critical applications, a system must not only be highly reliable, but that reliability must be certifiable. This is particularly true in civil aviation, where the FAA requires fly-by-wire control systems to have fewer than 10^{-9} catastrophic failures per flight-hour [1]. Such system-wide failures can occur if the system is rendered inoperable by a critical component failure or if the system performs poorly because of an undetected component failure. One approach, commonly found in avionics systems, is to use parallel redundant components, which ensures the availability of the system, even in the presence of component failures [1], [2], [3]. A failed component is detected by directly comparing the behavior of each redundant component. Hence, these schemes tend to detect faults accurately, and their performance is simple to certify using fault trees [4].

However, in some applications, such as Unmanned Aerial Vehicles (UAVs), the designer cannot afford the extra weight, size, and power needed to support identical redundant components. To prevent system-wide failures due to an undetected component failure, the analytical redundancies between components can be exploited to detect faults. For example, if three measurements m_1, m_2, and m_3 are available, and these quantities are known to satisfy the analytical relations $m_1 = f_1(m_2,m_3)$, $m_2 = f_2(m_1,m_3)$, and $m_3 = f_3(m_1,m_2)$, then the residuals $r_1 = m_1 - f_1(m_2,m_3)$, $r_2 = m_2 - f_2(m_1,m_3)$, etc. can be used to detect failures in the components that produce these measurements. This approach certainly reduces the number of individual components needed; however, there are two main drawbacks to consider. First, merely identifying a fault cannot prevent system-wide failure if the failed component is indispensable (i.e. no other components can perform the same critical function). Second, the performance of fault detection schemes based on analytical redundancy can be difficult to certify if the analytical relationships are dynamic or nonlinear. While the first difficulty is unavoidable, this paper addresses the second difficulty.

Although there is a vast body of literature on model-based fault detection and identification (FDI) (e.g., [5], [6], [7]), little attention is given to the rigorous evaluation of the performance and reliability metrics required to certify safety-critical aerospace systems. Monte Carlo methods [8] provide a statistically rigorous approach to performance analysis, but it can difficult to quantify the error present in the results. The goal of this paper is to define a set of performance metrics that can be computed analytically and to provide a class of systems for which these metrics can be computed efficiently.

The problem of fault detection is modeled by the interconnection of systems shown in Fig. 1. Section II provides a detailed description of the plant P, the fault detection scheme (F,δ), and the associated input and output signals. The framework presented here is a generalization of the fault detection problem considered in [9]. To quantify the performance of a fault detection scheme that falls within this framework, a set of performance metrics is defined in Section II-C. The analysis in Section III demonstrates that, in general, computing these metrics is an intractable problem, but for a specific class of fault detection problems these metrics can be computed in polynomial time. Section III-C presents conditions under which this complexity can be further reduced to quadratic or linear time. In Section IV, this analysis is applied to a fault detection problem involving a pitot-static probe subject to randomly occurring bias faults. Finally, Section V discusses the conclusions of this work, as well as avenues of future research.

II. PROBLEM STATEMENT

First, we establish some notation. Let $(\Omega,\mathscr{F},\mathbb{P})$ be the underlying probability space, and let $\mathscr{K} := \{k \in \mathbb{Z} \,|\, k \geq 0\}$ be the discrete time index set. For any stochastic process $v\colon \Omega\times\mathscr{K} \to \mathbb{R}^n$, the notation $\{v_k\}_{k\in\mathscr{K}}$ or $\{v_k\}$ represents the entire process. For each $k \in \mathscr{K}$, the notation v_k represents the random variable $v_k\colon \omega \mapsto v_k(\omega)$. For $i, j \in \mathscr{K}$ with $i \leq j$, define the partial sequence $v_{i:j} := \{v_i, v_{i+1}, \ldots, v_j\}$.

A. Plant Model

Assume that the plant, labeled P in Fig. 1, is of the form

$$\begin{aligned} x_{k+1} &= \hat{A}_k x_k + \hat{B}_{u,k} u_k + \hat{B}_{v,k} v_k + f_k(\theta_{0:k}) \\ y_k &= \hat{C}_k x_k + \hat{D}_{u,k} u_k + \hat{D}_{v,k} v_k + f_k(\theta_{0:k}), \end{aligned} \tag{1}$$

T. J. Wheeler and A. K. Packard are with the Dept. of Mechanical Engineering, University of California, Berkeley, CA 94708.

P. Seiler and G. J. Balas are with the Dept. of Aerospace Engineering & Mechanics, University of Minnesota, Minneapolis, MN 55455.

Corresponding author: T. J. Wheeler (twheeler@berkeley.edu).

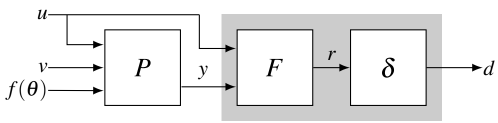


Fig. 1. Block diagram showing the plant P subject to a deterministic input u, a noise process v, and a random fault signal $f(\theta)$. The shaded region is the fault detection scheme (F,δ), and the output d is the decision or inference made by (F,δ).

where $\{u_k\}$ is a known deterministic input and $\{v_k\}$ is an i.i.d., Gaussian stochastic process with $v_k \sim \mathcal{N}(0,I)$, for all k. The random occurrence of faults is modeled by a time-homogeneous Markov chain $\{\theta_k\}$ with a finite state space $\mathscr{M} := \{0,1,\dots,m\}$, transition probability matrix

$$\Pi_{ij} := \mathbb{P}(\theta_{k+1} = j \,|\, \theta_k = i), \qquad i,j \in \mathscr{M},$$

and initial distribution π_0 [10]. The random variable θ_k is called the *mode* of the system at time k. The signal $\{f_k\}$ is a sequence of deterministic functions that map the mode sequence $\theta_{0:k}$ to an additive fault input $f_k(\theta_{0:k})$, such that for all k, $\theta_k = 0$ implies that $f_k(\theta_{0:k}) = 0$. That is, the event $\{\theta_k = 0\}$ occurs when the system (1) is in the nominal mode (i.e., no faults) at time k. Conditional on the event $\{\theta_{0:k} = \hat{\theta}_{0:k}\}$, the system (1) is a linear time-varying (LTV) Gaussian system driven by the known deterministic inputs $\{u_k\}$ and $\{f_k(\hat{\theta}_{0:k})\}$ and the random input $\{v_k\}$. Hence, the system (1) belongs to the class of *conditionally linear-Gaussian systems* [11].

B. Fault Detection Scheme

Since the random occurrence of faults is modeled by the Markov chain $\{\theta_k\}$, the general purpose of fault detection and identification is to infer something about the current value of $\{\theta_k\}$. More precisely, if for some s the set $\mathscr{M}$ is partitioned into s disjoint subsets $\mathscr{M}_0, \mathscr{M}_1, \dots, \mathscr{M}_{s-1}$ and we define $\mathscr{D} = \{0,1,\dots,s-1\}$, then the goal is to determine for which $d \in \mathscr{D}$ is the event $\{\theta_k \in \mathscr{M}_d\}$ most likely at time k. In this paper, we consider the simplest case, called *fault detection*, where $\mathscr{M}_0 := \{0\}$, $\mathscr{M}_1 := \{1,2,\dots,m\}$, and $\mathscr{D} := \{0,1\}$. In other words, the objective is to determine if the system is in the nominal mode or in *some* fault mode.

Assume that fault detection scheme, labeled F and δ in Fig. 1, is modeled as a deterministic LTV system F given by

$$\begin{aligned} \xi_{k+1} &= \tilde{A}_k \xi_k + \tilde{B}_{u,k} u_k + \tilde{B}_{y,k} y_k, \\ r_k &= \tilde{C}_k \xi_k + \tilde{D}_{u,k} u_k + \tilde{D}_{y,k} y_k, \end{aligned} \tag{2}$$

and a sequence of static, memoryless, deterministic functions $\{\delta_k\}$, where each $\delta_k \colon \mathbb{R} \to \mathscr{D}$ is called a *decision function*. The output $\{d_k\}$, given by $d_k = \delta_k(r_k) \in \mathscr{D}$, indicates the decision or inference made by the fault detection scheme. The role of the system F is to generate an output $\{r_k\}$, known as the *residual*, that has small mean and variance when $\{\theta_k\}$ is in $\mathscr{M}_0$ and has large mean and large variance otherwise. The complementary role of each decision function δ_k is to determine when the residual r_k is large enough to indicate that the event $\{\theta_k \in \mathscr{M}_1\}$ is likely at time k. A commonly-used decision function is

$$\delta(r) = \mathbb{I}(|r| > \varepsilon),$$

where $\mathbb{I}$ is the indicator function and $\varepsilon > 0$ is the threshold.

C. Probabilistic Analysis

For each $k \in \mathscr{K}$, define the events

$$\begin{aligned} H_{0,k} &:= \{\theta_k \in \mathscr{M}_0\}, \quad R_{0,k} := \{d_k = 0\}, \\ H_{1,k} &:= \{\theta_k \in \mathscr{M}_1\}, \quad R_{1,k} := \{d_k = 1\}. \end{aligned}$$

The reliability of the plant is given by $H_{0,k}$ and $H_{1,k}$, while the behavior of the fault detection scheme is given by $R_{0,k}$ and $R_{1,k}$ [12]. Note that $\{H_{0,k}, H_{1,k}\}$ and $\{R_{0,k}, R_{1,k}\}$ both form partitions of the sample space Ω. For each $k \in \mathscr{K}$, the performance of the fault detection scheme, with respect to the plant, is given by the following four events: a *true negative*, $R_{0,k} \cap H_{0,k}$; a *false positive*, $R_{1,k} \cap H_{0,k}$; a *false negative*, $R_{0,k} \cap H_{1,k}$; and a *true positive*, $R_{1,k} \cap H_{1,k}$ [12]. These events also form a partition of the sample space Ω, and their corresponding probabilities are denoted

$$P_k^{\mathrm{TN}} = \mathbb{P}(R_{0,k} \cap H_{0,k}), \tag{3}$$

$$P_k^{\mathrm{FP}} = \mathbb{P}(R_{1,k} \cap H_{0,k}), \tag{4}$$

$$P_k^{\mathrm{FN}} = \mathbb{P}(R_{0,k} \cap H_{1,k}), \tag{5}$$

$$P_k^{\mathrm{TP}} = \mathbb{P}(R_{1,k} \cap H_{1,k}). \tag{6}$$

These probabilities provide all the necessary information, because their values can be used to compute the probability of any event in the σ-algebra generated by the collection $\{R_{0,k}, R_{1,k}, H_{0,k}, H_{1,k}\}$. Since the values of (3)–(6) sum to one, only three of the four quantities must be computed.

Although the probabilities (3)–(6) provide all the necessary information, their numerical values can be difficult to interpret. For example, suppose that $\mathbb{P}(H_{1,k}) \approx 0$ for $k = 0,1,\dots,T$. This implies that

$$\mathbb{P}(H_{1,k}) = P_k^{\mathrm{FN}} + P_k^{\mathrm{TP}} \approx 0.$$

Since both P_k^{FN} and P_k^{TP} are small, it is difficult to get a sense of how well the fault detection scheme will perform in the presence of a fault at times $k \in \{0,1,\dots,T\}$. In this case, it is beneficial to consider the relative magnitudes of P_k^{FN} and P_k^{TP}. This approach gives rise to two conditional probabilities: the probability of *detection*

$$P_k^{\mathrm{D}} := \mathbb{P}(R_{1,k} \,|\, H_{1,k}) = \frac{P_k^{\mathrm{TP}}}{P_k^{\mathrm{FN}} + P_k^{\mathrm{TP}}}, \tag{7}$$

and the probability of a *false alarm*

$$P_k^{\mathrm{F}} := \mathbb{P}(R_{1,k} \,|\, H_{0,k}) = \frac{P_k^{\mathrm{FP}}}{P_k^{\mathrm{TN}} + P_k^{\mathrm{FP}}}. \tag{8}$$

If the probability $\mathbb{P}(H_{1,k})$ is known, equations (7) and (8) can be rearranged to compute the probabilities (3)–(6) from P_k^{D}, P_k^{F}, and $\mathbb{P}(H_{1,k})$.

III. COMPUTATIONAL ISSUES

For each $k \in \mathscr{K}$, the joint density of the residual r_k and the mode θ_k is given by

$$p(r_k, \theta_k) = \sum_{\theta_{0:k-1} \in \mathscr{M}^k} p(r_k, \theta_{0:k}) = \sum_{\theta_{0:k-1} \in \mathscr{M}^k} p(r_k \,|\, \theta_{0:k}) p(\theta_{0:k}). \tag{9}$$

This density can be used to compute probabilities (3)–(6); for instance,

$$P_k^{\mathrm{TP}} = \sum_{\theta_k=1}^{m} \int_{\delta_k^{-1}(1)} \Bigg(\sum_{\theta_{0:k-1}\in\mathscr{M}^k} p(r_k \,|\, \theta_{0:k}) p(\theta_{0:k}) \Bigg) \mathrm{d}r_k. \quad (10)$$

There are two difficulties inherent in computing this formula. The first is the number of mode sequences $\theta_{0:k}$ that must be considered; in total, there are $m \cdot |\mathscr{M}|^k = \mathscr{O}(m^{k+1})$ terms in the sum. Second, even though the conditional density $p(r_k \,|\, \theta_{0:k})$ is Gaussian, the sets $\delta_k^{-1}(0)$ and $\delta_k^{-1}(1)$ may be too complex to compute the necessary integrals. Thus, in the general case, computing the performance metrics (3)–(6) is intractable. To ameliorate this complexity, we establish a sufficient set of restrictions on the structure of the Markov chain $\{\theta_k\}$ and the form of the decision functions $\{\delta_k\}$, so that the performance metrics can be computed efficiently.

A. Sufficient Restrictions

First, we address the complexity issues arising from Markov chain $\{\theta_k\}$. There are many approximate methods, such as the Interacting Multiple Model and Generalized Pseudo-Bayesian algorithms [13], that reduce the complexity by effectively consolidating information about the past and considering only a fixed number of terms in the joint density (9). In our approach, we compute the exact solution for a special class of Markov chains that are well-suited to fault detection problems.

Assume that $\mathbb{P}(\theta_0 = 0) = 1$, i.e., the system almost surely starts in the nominal mode. Suppose that the system has ℓ components that fail independently, and assume that once a component fails, it remains in a failed state indefinitely. The set of components that have failed at or before a given time can be encoded by a binary string of length ℓ, where a 1 indicates a failure. Note that there are 2^ℓ such binary strings, and each can be identified with a member of the state space $\mathscr{M} := \{0, 1, \dots, 2^\ell - 1\}$. The assumption that a failed component must remain in a failed state is captured by the transition probability matrix Π. For example, suppose $\ell = 2$. If the binary strings are labeled as

$$\mathscr{M} = \{0,1,2,3\} \leftrightarrow \{00,10,01,11\},$$

then Π must be of the form

$$\Pi = \begin{bmatrix} * & * & * & * \\ 0 & * & 0 & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where $*$ indicates a possibly nonzero entry. Note that relabeling the binary sequences would just permute the rows and columns of Π. Under these assumptions, a fault model consisting of ℓ components has only $\mathscr{O}(k^\ell)$ mode sequences $\theta_{0:k} \in \mathscr{M}^{k+1}$ with nonzero probability.

Second, we address the problem of computing integrals of the form

$$\int_{\delta_k^{-1}(d_k)} p(r_k \,|\, \theta_{0:k}) \,\mathrm{d}r_k,$$

where $k \in \mathscr{K}$, $d_k \in \mathscr{D}$ and $\theta_{0:k} \in \mathscr{M}^{k+1}$. Assume that each decision function δ_k is a threshold function

$$\delta_k(r_k) := \mathbb{I}(|r_k| > \varepsilon_k), \qquad \varepsilon_k > 0.$$

If z is some Gaussian random variable with $z \sim \mathscr{N}(\mu, \sigma^2)$ and $\varepsilon > 0$, then $\mathbb{P}(-\varepsilon < z < \varepsilon)$ is given by

$$\int_{-\varepsilon}^{\varepsilon} p(z)\,\mathrm{d}z = \frac{1}{2}\left[\operatorname{erf}\left(\frac{\varepsilon - \mu}{\sqrt{2}\sigma} \right) - \operatorname{erf}\left(\frac{-\varepsilon - \mu}{\sqrt{2}\sigma} \right) \right],$$

where $\operatorname{erf}\colon \mathbb{R} \to [-1,1]$ is the error function. Analytically, $\operatorname{erf}(\cdot)$ can be approximated by a rational function with maximum relative error less than 10^{-19} [14]. Evaluating the rational approximation requires a fixed number of floating point operations. Thus, the probability $\mathbb{P}(-\varepsilon < z < \varepsilon)$ can be computed accurately in $\mathscr{O}(1)$ time. We conclude that once the densities $p(r_k \,|\, \theta_0 : k)$ and $p(\theta_{0:k})$ have been calculated (see Section III-B), the performance metrics can be evaluated with $\mathscr{O}(k^\ell)$ calls to $\operatorname{erf}(\cdot)$ in $\mathscr{O}(k^\ell)$ time.

B. Computational Procedure

Because of their linear structure, systems (1) and (2) can be combined into a single system

$$\begin{aligned} \eta_{k+1} &= A_k \eta_k + B_{u,k} u_k + B_{v,k} v_k + B_{f,k} f_k(\theta_{0:k}), \\ r_k &= C_k \eta_k + D_{u,k} u_k + D_{v,k} v_k + D_{f,k} f_k(\theta_{0:k}), \end{aligned} \quad (11)$$

where $\eta_k := (x_k, \xi_k)$ is the combined state. For each $k \in \mathscr{K}$, define $\hat{\eta}_k := \mathbb{E}(\eta_k \,|\, \theta_{0:k} = \hat{\theta}_{0:k})$ and $\hat{r}_k := \mathbb{E}(r_k \,|\, \theta_{0:k} = \hat{\theta}_{0:k})$. Strictly speaking, we should write $\hat{\eta}_k(\hat{\theta}_{0:k})$ and $\hat{r}_k(\hat{\theta}_{0:k})$, but we omit this argument when the sequence $\hat{\theta}_{0:k}$ is clear from context. The sequences $\{\hat{\eta}_k\}$ and $\{\hat{r}_k\}$ are given by

$$\begin{aligned} \hat{\eta}_{k+1} &= A_k \hat{\eta}_k + B_{u,k} u_k + B_{f,k} f_k(\hat{\theta}_{0:k}), \\ \hat{r}_k &= C_k \hat{\eta}_k + D_{u,k} u_k + D_{f,k} f_k(\hat{\theta}_{0:k}). \end{aligned} \quad (12)$$

Similarly, for each k, define the conditional covariance matrices $P_k := \operatorname{var}(\eta_k \,|\, \theta_{0:k} = \hat{\theta}_{0:k})$ and $Q_k := \operatorname{var}(r_k \,|\, \theta_{0:k} = \hat{\theta}_{0:k})$. Then, $\{P_k\}$ and $\{Q_k\}$ are given by

$$\begin{aligned} P_{k+1} &= A_k P_k A_k^* + B_{v,k} B_{v,k}^*, \\ Q_k &= C_k P_k C_k^* + D_{v,k} D_{v,k}^*. \end{aligned} \quad (13)$$

Each update (from k to $k+1$) of equations (12) and (13) takes constant time, so for a fixed final time $T \in \mathscr{K}$ and a given mode sequence $\hat{\theta}_{0:T}$, the sequences $\hat{r}_{0:T}$ and $Q_{0:T}$ can be computed in $\mathscr{O}(T)$ time. Since $\{\theta_k\}$ is a Markov process,

$$p(\theta_{0:k}) = p(\theta_k \,|\, \theta_{k-1}) p(\theta_{0:k-1}) = \Pi_{\theta_{k-1}\theta_k} p(\theta_{0:k-1}),$$

and $\mathbb{P}(\theta_{0:k} = \hat{\theta}_{0:k})$ can be recursively computed from $\mathbb{P}(\theta_{0:k-1} = \hat{\theta}_{0:k-1})$ with a single multiplication. Since there are $\mathscr{O}(T^\ell)$ mode sequences, the entire joint density (9) can be computed in $\mathscr{O}(T^{\ell+1})$ time.

C. Special Case

Recall that the Markov chain $\{\theta_k\}$ is interpreted as the status of ℓ components that fail randomly. We consider a special class of fault signals $\{f_k(\cdot)\}$ of the form

$$f_k(\theta_{0:k}) = \sum_{i=1}^{\ell} \lambda_i \Big(k - T_i(\theta_{0:k}) \Big), \qquad k \in \mathscr{K}, \quad (14)$$

TABLE I

TIME-COMPLEXITY OF COMPUTING THE PERFORMANCE METRICS.

Fault Model	Simulations	Calls to erf$(\cdot)$
General Markov Chain	$\mathscr{O}(m^{T+1})$	$\mathscr{O}(Tm^{T+1})$
Restricted Markov Chain	$\mathscr{O}(T^{\ell})$	$\mathscr{O}(T^{\ell+1})$
Special case (LTV)	$\mathscr{O}(\ell T)$	$\mathscr{O}(T^{\ell+1})$
Special case (LTI)	$\mathscr{O}(\ell)$	$\mathscr{O}(T^{\ell+1})$

where each λ_i is a deterministic function and $T_i(\theta_{0:k})$ is the time at which the ith component fails in the mode sequence $\theta_{0:k}$. If the ith component does not fail, we take $T_i(\theta_{0:k}) = \infty$. Assume that $\lambda_i(s) = 0$ for $s < 0$. In other words, λ_i "switches on" when component i fails. Since system (11) has linear structure, we can use superposition to significantly reduce the amount of computation needed to compute $\{\hat{r}_k\}$ and $\{Q_k\}$.

For $i = 1, 2, \ldots, \ell$ and $\tau = 1, 2, \ldots, T$, define $\theta_{0:T}^{(i,\tau)}$ to be the mode sequence for which $T_i = \tau$ and $T_j = \infty$, for $i \neq j$. Suppose that for all such (i,τ) pairs, we set $\hat{\eta}_0 = 0$, $u_k = 0$, for all k, and simulate equation (12) with the input $\{f_k(\theta_{0:k}^{(i,\tau)})\}$ to obtain the corresponding conditional mean $\hat{r}_{0:T}^{(i,\tau)}$. Also, let $\hat{r}_{0:T}^{(0,0)}$ be the result of simulating equation (12) with the original values of $\hat{\eta}_0$ and $u_{0:T}$ but with no faults. Then, the value of $\hat{r}_{0:T}$ corresponding to an arbitrary mode sequence $\hat{\theta}_{0:T}$ can be obtained by superposing $\hat{r}_{0:T}^{(i,\tau)}$ for the appropriate pairs (i,τ). Using superposition the number of simulations needed is reduced from $\mathscr{O}(T^{\ell})$ to $\mathscr{O}(\ell T)$. Moreover, if the system (11) is LTI then for each i and j, $\hat{r}_{0:T}^{(i,1)}$ can be shifted τ time-steps to obtain $\hat{r}_{0:T}^{(i,\tau)}$, which further reduces the number of simulations to $\mathscr{O}(\ell)$. However, in these special cases, the error function must still be evaluated $\mathscr{O}(T^{\ell+1})$ times to compute the performance metrics. These time-complexity results are summarized in Table I.

IV. APPLICATION: PITOT-STATIC PROBE

Nearly all aircraft use a pitot-static probe to determine airspeed V and altitude h. Because these data are essential for flying, the pitot-static probe is integrated into the flight control feedback loop. These sensors are prone to a number of failures, such as icing and blockage, that cause them to produce incorrect values. If such a failure goes undetected, the autopilot system or the pilot may use the erroneous values to issue commands that cause the aircraft to crash. To avoid such disasters, large commercial aircraft, such as the Boeing 777 [2], [3], have multiple pitot-static probes in different locations. However, most aircraft designers have developed a set of standard operating procedures that allow safe recovery of the aircraft when a pitot-static probe failure is detected [15]. In this application we explore the detection of such faults by exploiting the analytical redundancy between airspeed, altitude, and flight path angle. Hansen *et al.* [16] present a similar example, which uses the methods of statistical change-point detection [17], [18] to model sensor faults.

A. System Description

Consider the fault detection problem shown in Fig. 2. The pitot tube measures the total pressure p_t, and the static port

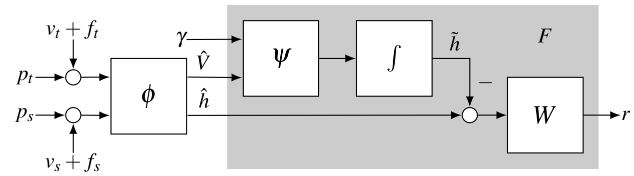


Fig. 2. Block diagram of a pitot-static probe with a fault detection scheme based on analytical redundancy. The map ϕ (shown graphically in Fig. 3) represents the plant P, while the shaded region, labeled F, is the dynamic portion of the fault detection scheme.

measures the static pressure p_s. These measurements are corrupted by Gaussian white noise processes, v_t and v_s, and randomly occurring bias faults, f_t and f_s. The airspeed and altitude are derived using the relations (see Fig. 3)

$$\begin{bmatrix} V \\ h \end{bmatrix} = \phi(p_t, p_s) := \begin{bmatrix} \operatorname{sign}(p_t - p_s) c_3 \left(\left| \left(\frac{p_t - p_s}{p_0} + 1 \right)^{c_4} - 1 \right| \right)^{\frac{1}{2}} \\ c_1 \left(1 - \left(\frac{p_s}{p_0} \right)^{c_2} \right) \end{bmatrix},$$

where the constants $c_1 = 44.331\,\mathrm{km}$, $c_2 = 0.1903$, $c_3 = 760.427\,\mathrm{m/s}$, $c_4 = 2/7$, and $p_0 = 101.325\,\mathrm{kPa}$ model the troposphere ($h \leq 17\,\mathrm{km}$) [1]. We use the notation $\hat{V}$ for the derived airspeed and $\hat{h}$ for the derived altitude to indicate that these quantities are corrupted by random disturbances and faults. Note ϕ actually gives the *indicated* airspeed, but we ignore this issue for the sake of simplicity.

The fault signals are defined as $f_t(t) := b_t \mathbb{I}(t \geq T_t)$ and $f_s(t) := b_s \mathbb{I}(t \geq T_s)$, for $t \geq 0$, where b_t and b_s are known, fixed biases and T_t and T_s are independent exponential random variables $T_t \sim \operatorname{Exp}(\lambda_t)$ and $T_s \sim \operatorname{Exp}(\lambda_s)$.

The dynamic portion of the fault detection scheme F is contained in the shaded region of Fig. 2. The input γ is the flight path angle of the aircraft, which we assume is measured exactly with no noises or faults. Consider the following analytical relationship between V, h, and γ:

$$h(t) = h(0) + \int_0^t \psi\big(V(\tau), \gamma(\tau)\big)\,\mathrm{d}\tau = \int_0^t V(\tau) \sin\gamma(\tau)\,\mathrm{d}\tau,$$

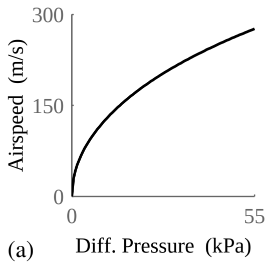
which is used to derive $\tilde{h}$ from γ and $\hat{V}$. The fault detection scheme attempts to detect the faults f_t and f_s by analyzing the difference $\hat{h} - \tilde{h}$. However, as the noisy signal $\psi(\hat{V}, \gamma)$ passes through the integrator, the noise accumulates and $\tilde{h}$ diverges from $\hat{h}$. To counteract this effect, a high-pass or "washout" filter of the form

$$W(s) = \frac{s}{s+a}, \qquad a > 0,$$

is applied to the difference $\hat{h} - \tilde{h}$ to produce the residual r. The drawback of using this filter is that it removes the DC component from the signal $\hat{h} - \tilde{h}$. The decision function (not depicted in Fig. 2) is a threshold function with threshold $\varepsilon > 0$, and the same decision function is used at each $k \in \mathscr{K}$.

B. Applying the Framework

To apply the framework developed in Sections II and III, the plant P must be LTV. As shown in Fig. 3, the map ϕ is only mildly nonlinear for modest changes in differential



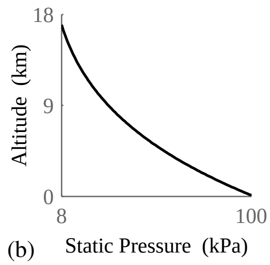


Fig. 3. Plot of (a) the (indicated) airspeed V as a function of differential pressure $p_d = p_t - p_s$ and (b) the altitude h as a function of static pressure p_s. The values plotted here are typical for subsonic flight in the troposphere.

pressure $p_d := p_t - p_s$ and static pressure p_s, so we take the first-order approximation

$$\phi(p_t + v_t + f_t, p_s + v_s + f_s) \approx \phi(p_t, p_s) + \Phi \begin{bmatrix} v_t \\ v_s \end{bmatrix} + \Phi \begin{bmatrix} f_t \\ f_s \end{bmatrix},$$

where $\Phi := (D\phi)(p_t, p_s)$ is the Jacobian linearization of ϕ at (p_t, p_s). Using this linearization, the continuous-time version of the combined system (11) is

$$\begin{aligned} \dot{\eta} &= -a\eta + \begin{bmatrix} \sin(\gamma) & a \end{bmatrix} \big(u + \Phi(t)(v + f)\big) \\ r &= -\eta + \begin{bmatrix} 0 & 1 \end{bmatrix} \big(u + \Phi(t)(v + f)\big), \end{aligned} \tag{15}$$

where $u := \phi(p_t, p_s)$, $v := [v_t \quad v_s]^T$, and

$$f := \begin{bmatrix} f_t \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ f_s \end{bmatrix}.$$

At time $t = 0$, the residual has zero mean if $\eta(0) = h(0)$.

The system (15) is discretized as follows: fix a sample time Δ_t; then, for all $k \in \mathscr{K}$, sample a trajectory $(p_{t,k}, p_{s,k}, \gamma_k)$, compute the input $u_k = \phi(p_{t,k}, p_{s,k})$, and compute the state-space data $\Phi_k = (D\phi)(p_{t,k}, p_{s,k})$ and $\sin(\gamma_k)$. In discrete-time, T_t and T_s are replaced by geometric random variables $\tau_t \sim \mathrm{Geo}(q_t)$ and $\tau_s \sim \mathrm{Geo}(q_s)$, where $q_t = 1 - \exp(-\lambda_t \Delta_t)$ and $q_s = 1 - \exp(-\lambda_s \Delta_s)$. This ensures that $\mathbb{P}(T_t < k\Delta_t) = \mathbb{P}(\tau_t < k)$ and $\mathbb{P}(T_s < k\Delta_t) = \mathbb{P}(\tau_s < k)$. The corresponding Markov chain has state space $\mathscr{M} = \{0, 1, 2, 3\}$, initial distribution $\pi_0 = [1\ 0\ 0\ 0]$, and transition probability matrix

$$\Pi = \begin{bmatrix} (1-q_t)(1-q_s) & q_t(1-q_s) & q_s(1-q_t) & q_t q_s \\ 0 & (1-q_s) & 0 & q_s \\ 0 & 0 & (1-q_t) & q_t \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

C. Numerical Results

For this analysis, we use the following parameter values: $\Delta_t = 0.05\,\mathrm{s}$, $V = 45\,\mathrm{m/s}$, $\gamma = 0.5^\circ$, $h_0 = 200\,\mathrm{m}$, $b_t = -40\,\mathrm{Pa}$, $b_s = 50\,\mathrm{Pa}$, and $v_{t,k} \sim \mathscr{N}(0, 0.297\,\mathrm{Pa}^2)$ and $v_{s,k} \sim \mathscr{N}(0, 0.297\,\mathrm{Pa}^2)$, for all $k \in \mathscr{K}$. The component failure probabilities are $q_t = q_s = 1.389 \cdot 10^{-7}$, which corresponds to a mean time-to-failure (MTTF) of 1000 hrs [4].

To get a sense of how each fault affects system (15), we plot the change in the residual's statistics due to a particular fault. Fig. 4 shows the change due to f_t occurring at $T_f = 1\,\mathrm{min}$, and Fig. 5 shows the residual due to f_s occurring at $T_s = 1\,\mathrm{min}$. Three cases are plotted: nominal (black), fault with $a = 0.003$ (blue), and fault with $a = 0.0075$ (green). In each case, the region within three standard deviations of the mean is shaded in the appropriate color. Note that the green lines reach steady-state more quickly, but the blue lines achieve more separation between the nominal and faulty cases. Hence, there is a trade-off between how quickly the faults can be detected and how easily the fault is distinguished from the nominal case. The instantaneous jump shown in Fig. 5 is due to the direct feedthrough from f_s to r in (15). This makes it difficult to detect the fault between 3 and 8 min, when the residual changes sign.

In Fig. 6, the performance metrics $\{P_k^{\mathrm{TN}}\}$ (solid lines) and $\{P_k^{\mathrm{FP}}\}$ (dashed lines) are plotted for $a = 0.003$ and $\varepsilon = 2\,\mathrm{m}, 3.5\,\mathrm{m}, 5\,\mathrm{m}$ (blue, green, and red, respectively). The curves $\{P_k^{\mathrm{FN}}\}$ and $\{P_k^{\mathrm{TP}}\}$ are omitted, because their values are approximately zero over the time window plotted. Hence, this plot does not provide information about the relative performance of each scheme when a fault occurs.

In Fig. 7, the performance metrics $\{P_k^{\mathrm{D}}\}$ (solid lines) and $\{P_k^{\mathrm{F}}\}$ (dashed lines) are plotted for $a = 0.003$ and $\varepsilon = 2\,\mathrm{m}, 3.5\,\mathrm{m}, 5\,\mathrm{m}$ (blue, green, and red, respectively). Recall that P_k^{D} accounts for all possible faults up to time k. For small k, the instantaneous jump shown in Fig. 5 dominates, which causes a peak in the $\{P_k^{\mathrm{D}}\}$ curve at 4 min. However, as k increases, these early faults begin to settle and there are more faults to consider, so the curve $\{P_k^{\mathrm{D}}\}$ becomes smoother. The values plotted in Fig. 6 and Fig. 7 allow us to directly compare the performance of different fault detection schemes and certify the time-varying reliability of the system under each scheme.

V. CONCLUSIONS

The reliability of safety-critical systems must be certified, but there is little work in the literature that rigorously analyzes the performance of model-based fault detection systems. The framework presented in this paper provides a class of model-based fault detection problems for which the performance can be computed analytically. Is is shown that, under a reasonable set of assumptions, this computation can be carried out in polynomial time. This analysis is applied to a model of a pitot-static probe subject to randomly occurring bias faults with known magnitudes, which are detected using analytical redundancy. The data obtained from this analysis (shown in Fig. 6 and Fig. 7) can be used to certify the performance of a fault detection scheme.

Future work on this topic will include the study of more complex decision functions, such as the likelihood ratio test

$$d_k = \delta_k^{\mathrm{LRT}}(r_{0:k}) := \mathbb{I}\left(\frac{p(r_{0:k} \,|\, H_{1,k})}{p(r_{0:k} \,|\, H_{0,k})} > \varepsilon_k \right),$$

and the up-down counter, which is defined by the recurrence

$$\begin{aligned} c_{k+1} &= c_k + C_{\mathrm{up}} \mathbb{I}(|r_k| > \varepsilon_k) - C_{\mathrm{down}} \mathbb{I}(|r_k| \leq \varepsilon_k), \\ d_k &= \mathbb{I}(c_k > \tau_k), \end{aligned}$$

where $c_0 = 0$, $C_{\mathrm{up}} \geq C_{\mathrm{down}} > 0$, and $\varepsilon_k, \tau_k > 0$. The likelihood ratio test has desirable theoretical properties [12], and the up-down counter is commonly used in avionics applications [1].

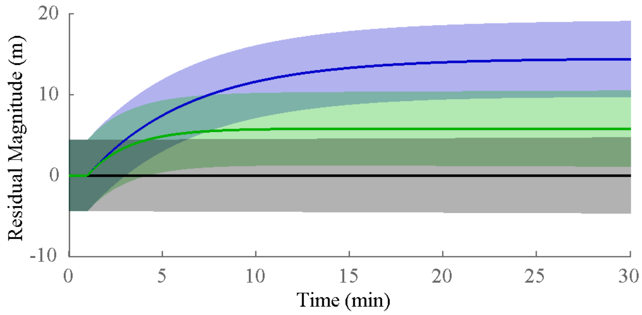


Fig. 4. Plot of the residual due to a fault in the total pressure (p_t) channel. The solid lines are the means, and the shaded region is three standard deviations away from the mean. Here, black represents the residual with no faults, blue corresponds to $a = 0.003$, and green corresponds to $a = 0.0075$.

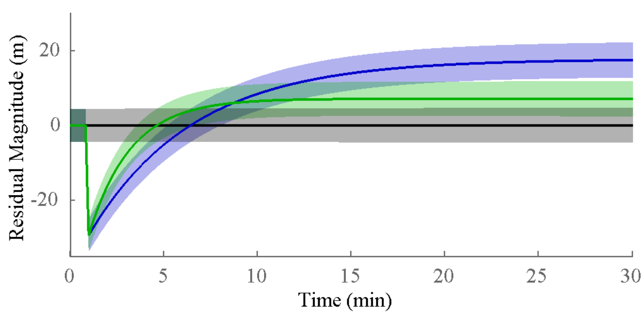


Fig. 5. Plot of the residual due to a fault in the static pressure (p_s) channel. The solid lines are the means, and the shaded region is three standard deviations away from the mean. Here, black represents the residual with no faults, blue corresponds to $a = 0.003$, and green corresponds to $a = 0.0075$.

Also, since the analysis in Section IV was carried out over a particular flight path, it would be useful to find the input u that gives the worst fault detector performance. Similarly, if there is parametric model uncertainty in the plant P, it would be useful to find the worst-case uncertainty parameter value.

VI. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 0931931 entitled "CPS: Embedded Fault Detection for Low-Cost, Safety-Critical Systems" and by the National Aeronautics and Space Administration under Grant No. NNX07AC40A entitled "Reconfigurable Robust Gain-Scheduled Control for Air-Breathing Hypersonic Vehicles." Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] R. Collinson, *Introduction to Avionics Systems*, 2nd ed. Boston: Kluwer Academic, 2003.

[2] Y. Yeh, "Triple-triple redundant 777 primary flight computer," in *Proc. of the 1996 IEEE Aerospace Applications Conf.*, 1996, pp. 293–307.

[3] ——, "Safety critical avionics for the 777 primary flight controls system," in *The 20th Digital Avionics Systems Conference*, Daytona Beach, FL, Oct. 2001, pp. 1.C.2.1–1.C.2.11.

[4] M. S. Hamada, A. G. Wilson, C. S. Reese, and H. F. Martz, *Bayesian Reliability*. New York: Springer, 2008.

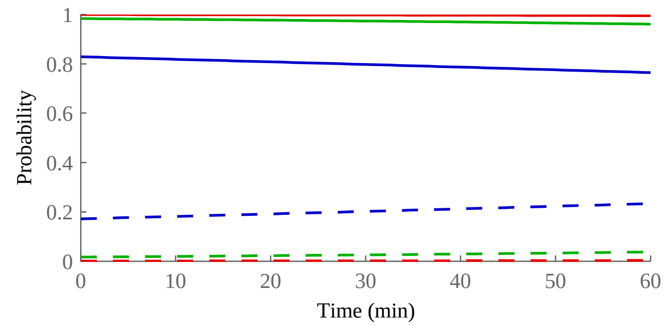


Fig. 6. Plot of $\{P_k^{\mathrm{D}}\}$ (solid lines) and $\{P_k^{\mathrm{F}}\}$ (dashed lines) for thresholds $\varepsilon = 2\,\mathrm{m}$ (blue), $\varepsilon = 3.5\,\mathrm{m}$ (green), and $\varepsilon = 5\,\mathrm{m}$ (red). Both $\{P_k^{\mathrm{FN}}\}$ and $\{P_k^{\mathrm{TP}}\}$ are omitted, because their values are essentially zero over this time window.

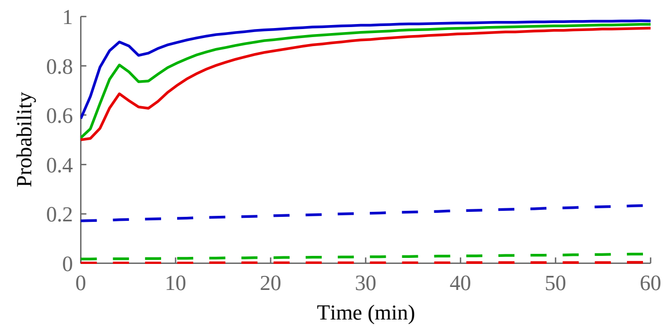


Fig. 7. Plot of $\{P_k^{\mathrm{D}}\}$ (solid lines) and $\{P_k^{\mathrm{F}}\}$ (dashed lines) for thresholds $\varepsilon = 2\,\mathrm{m}$ (blue), $\varepsilon = 3.5\,\mathrm{m}$ (green), and $\varepsilon = 5\,\mathrm{m}$ (red).

[5] J. Chen and R. J. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Boston: Kluwer Academic, 1999.

[6] S. X. Ding, *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*. Berlin: Springer-Verlag, Jan. 2008.

[7] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Berlin: Springer-Verlag, 2006.

[8] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer, 2004.

[9] T. J. Wheeler, P. Seiler, A. Packard, and G. J. Balas, "Performance analysis of fault detection systems based on analytically redundant linear time-invariant dynamics," in *Proc. of the American Control Conf.*, San Francisco, CA, 2011, pp. 214–219.

[10] J. R. Norris, *Markov Chains*. Cambridge: Cambridge University Press, 1997.

[11] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*. New York: Springer, 2005.

[12] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. New York: Springer, 2008.

[13] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York: Wiley, 2001.

[14] W. J. Cody, "Rational Chebyshev approximations for the error function," *Math. of Comp.*, vol. 23, no. 107, pp. 631–637, Sep. 1969.

[15] D. Carbaugh, D. Forsythe, and M. McIntyre, "Erroneous flight instrument information," *AERO Magazine*, vol. 8, pp. 10–21, 1998.

[16] S. Hansen, M. Blanke, and J. Adrian, "Diagnosis of UAV pitot tube failure using statistical change detection," in *Proc. of the 7th IFAC Symp. on Intelligent Autonomous Vehicles*, Lecce, Italy, Sep. 2010.

[17] M. Basseville and A. Benveniste, Eds., *Detection of Abrupt Changes in Signals and Dynamical Systems*. Berlin: Springer-Verlag, 1985.

[18] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge: Cambridge University Press, 2009.