Worst-Case False Alarm Analysis of Fault Detection Systems

Bin Hu and Peter Seiler

***Abstract*— Model-based fault detection methods can be used to reduce the size, weight, and cost of safety-critical aerospace systems. However, the implementation of these methods is based on models. Therefore, disturbance and model uncertainty must be considered in order to certify the fault detection system. This paper considers the worst-case false alarm probability over a class of stochastic disturbances and model uncertainty. This is one analysis needed to assess the overall system reliability. The single step, worst-case false alarm probability is shown to be equivalent to a robust $\mathcal{H}_2$ analysis problem. Hence known results from the robust $\mathcal{H}_2$ literature can be used to upper bound this worst-case probability. Next, bounds are derived for the worst-case false alarm probability over multiple time steps. The multi-step analysis is important because reliability requirements for aerospace systems are typically specified over a time window, e.g. per hour. The bounds derived for the multi-step analysis account for the time correlations introduced by the system dynamics and fault detection filters. Finally, a numerical example is presented to demonstrate the proposed technique.**

I. Introduction

The reliability of safety-critical aerospace systems must be certified with aviation authorities, e.g. the Federal Aviation Administration in the United States or the European Aviation Safety Agency. The system reliability and safety requirements for commercial flight control electronics are typically no more than 10^{-9} catastrophic failures per flight hour [3]. The aircraft industry uses designs that are based almost exclusively on physical redundancy, whose performance is relatively straightforward to certify using fault trees [12].

Replacing some physically redundant components with model-based fault detection and isolation (FDI) algorithms [2], [11], [5], [10] would lead to a dramatic reduction in the system size, weight, and power consumption. In addition, model-based methods could significantly improve the reliability of smaller unmanned aerial vehicles which cannot carry the payload associated with physical redundancy. The recent AddSafe project in Europe [1] dealt with the future green aircraft and assessed the suitability of these more advanced fault detection methods for optimizing the aircraft design. However, despite the benefits of FDI methods, new challenges come up in the certification phase of these model-based methods. Rigorously certifying the performance of model-based FDI would require a worst-case reliability analysis of fault detection systems due to nonlinear, time-varying and uncertain aircraft dynamics. The worst case reliability is also important for understanding the trade off between the robustness against uncertainty and good attenuation for disturbance. For example, the full-state observer-based FDI design can completely reject the disturbance when model is known perfectly [15]. However this method can be sensitive to model uncertainty. A worst-case analysis is needed to understand the overall performance of this FDI design method.

A direct worst-case reliability analysis is difficult to perform. In [8], the overall reliability is decoupled into false alarm analysis and missed detection analysis under reasonable assumptions. The focus of this paper is on the worst-case false alarm analysis problem which is required to determine the overall reliability of an analytically redundant system. Monte Carlo simulations provide a general solution to estimate the worst-case false alarm probability. However, numerous simulations may be needed due to the model uncertainty. An analytical method is provided in this paper to complement the Monte Carlo approach.

This paper formulates the worst-case false alarm analysis problem for a typical model-based fault detection system (Section II). In Section III the solution of worst-case single-step false alarm probability is connected to the robust $\mathcal{H}_2$ performance analysis problem. The main contribution of this paper is presented in Section IV. Two different upper bounds are developed for the worst-case false alarm probability over a time window. The first upper bound is also based on the robust $\mathcal{H}_2$ performance analysis problem. The second upper bound is better in the stochastic sense but may be conservative for some model uncertainty. A numerical example is given in Section V to demonstrate the proposed method.

II. Problem Formulation

Consider the uncertain aircraft model shown in Figure 1 where u_k and y_k are the control inputs and measurements at the discrete time k, respectively. The signals w_k and v_k represent stochastic process and sensors noises. A multiplicative uncertainty set [20] is used to describe the uncertain (healthy) aircraft dynamics linearized at some flight condition:

$$\mathcal{S}_M := \{G_0(I + \Delta W_u) \,:\, \|\Delta\|_\infty \leq 1\} \tag{1}$$

G_0 represents the nominal (and healthy) aircraft dynamics and W_u is a stable minimum phase transfer function whose magnitude specifies the uncertainty at each frequency. The true aircraft dynamics, when healthy, is assumed to be in this uncertainty set $\mathcal{S}_M$. An additive signal f_k is used to model the effects of aircraft faults.

A parity-equation approach, shown in Figure 2, can be used to detect aircraft faults. The aircraft is assumed to be operating in closed-loop with controller K used to track reference commands h_k. The parity equation compares the expected response, $\hat{y}_k$, obtained from the nominal model G_0 to the measured value, y_k. A residual signal is generated from

Bin Hu and Peter Seiler are with the Aerospace Engineering and Mechanics Department, University of Minnesota, Email: huxxx221@umn.edu; seiler@aem.umn.edu.

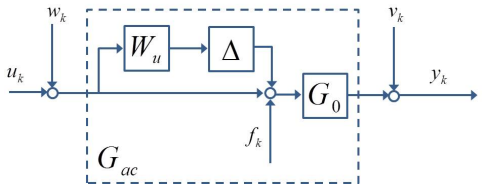

Fig. 1. Uncertain aircraft model with additive fault

this comparison as $r_k = y_k - \hat{y}_k$. The residual is typically small when $f_k = 0$ and large when a fault occurs. Based on r_k, the decision logic generates a signal d_k to indicate the health status, i.e. $d_k = 1$ if a fault has been detected and $d_k = 0$ otherwise. Fixed thresholding is considered here:

$$d_k := \begin{cases} 0 & \text{if } |r_k| \le T \\ 1 & \text{else} \end{cases} \tag{2}$$

where T is the decision logic threshold. A fault is declared when r_k exceeds the threshold T.

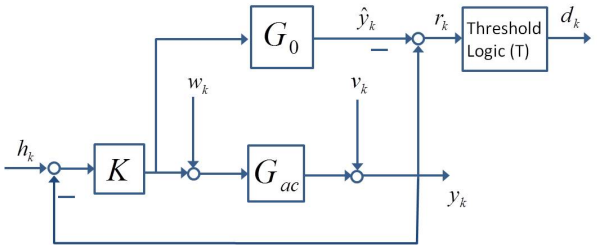

Fig. 2. Simple parity equation fault detection system

A false alarm is generated if a fault is declared ($d_k = 0$) when the aircraft is healthy ($f_k = 0$). This may occur due to the process and/or sensor noises. The false alarm probability depends on the aircraft dynamics and hence it depends on the model uncertainty. This paper presents analytical methods to bound the worst-case false alarm probability achieved by any model in the uncertainty set $\mathcal{S}_M$.

A formal statement of the analysis problem is now provided. Consider the case of steady flight and hence the linearized reference commands satisfy $h_k = 0$. In this case, the parity equation system in Figure 2 can be redrawn, via block diagram manipulation, to be in the form shown in Figure 3. The system M can be easily computed from the parity equation diagram and depends on K, W_u, and G_0. It is assumed that the input n_k is an IID, zero-mean, Gaussian stochastic process. When w_k and v_k are colored noises driven by linear time-invariant (LTI) models, the dynamics part can always be absorbed into the system M so that n_k only represents the innovation of these noises. It can further be assumed that the process is unit variance, $n_k \sim \mathcal{N}(0, I)$. The unit variance assumption is without loss of generality as the variance of n_k can be absorbed into the system M. The uncertainty is shown entering in a feedback fashion. Let $\Delta \star M$ denote the (upper) linear fractional transformation [20] that relates input n to output r:

$$r = \left(M_{22} + M_{21}(I - M_{11}\Delta)^{-1}M_{12}\right) n \tag{3}$$

where M_{ij} denote the blocks of M partitioned according to the dimensions of Δ. The discussion focused on the specific case of multiplicative uncertainty and a parity-equation fault detection system. To generalize the discussion, let M be any LTI, discrete-time system of appropriate dimensions and $\mathbf{\Delta}$ a set of (possibly structured) parametric and dynamic LTI uncertainty. It is assumed that the system is robustly stable, i.e. $\Delta \star M$ is assumed to be stable for all $\Delta \in \mathbf{\Delta}$. Finally for false alarm analysis, it is reasonable to assume the system is in steady state and hence the residual r_k is a stationary zero-mean Gaussian process. The worst-case, single-step false alarm analysis problem is now formally defined.

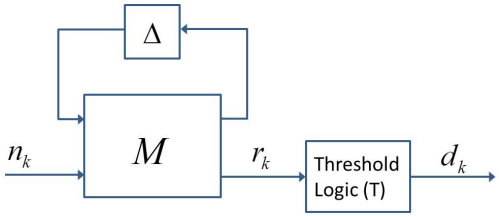

Fig. 3. Uncertain system for analysis of worst-case false alarm probability

Definition 1: For any fixed $\Delta \in \mathbf{\Delta}$, the per-frame false alarm probability, denoted $P_1(\Delta)$, is the conditional probability that $d_k = 1$ given that $f_k = 0$. The worst-case per-frame false alarm probability is $P_1^* := \max_{\Delta \in \mathbf{\Delta}} P_1(\Delta)$.

The per-frame false alarm probability $P_1(\Delta)$ is also called the false alarm rate (FAR). For safety-critical aerospace systems, system reliability requirements are typically specified over a time window. For example, flight control systems certified with the FAA are required to have less than 10^{-9} catastrophic failures per flight hour [3]. These system level requirements indicate that the false alarm probability should also be specified over a time window. This motivates the following definition of a multi-step false alarm probability:

Definition 2: For any fixed $\Delta \in \mathbf{\Delta}$, the N-step false alarm probability, denoted $P_N(\Delta)$, is the conditional probability that $d_k = 1$ for some k in $1 \le k \le N$, given that $f_k = 0$ for all k in $1 \le k \le N$. The worst-case N-step false alarm probability is $P_N^* := \max_{\Delta \in \mathbf{\Delta}} P_N(\Delta)$.

The problem formulation contains several assumptions. First, the presentation assumed a simple parity-equation fault detection system. There are more advanced approaches to design the fault detection system, e.g. observers, parameter estimators, and robust filtering [2], [11], [5], [10]. Most of these advanced methods fit within the general framework in Figure 3 as long as the fault detection filter is LTI. Next, the discussion focused on an uncertain aircraft model with a single multiplicative uncertainty. As noted above, the problem formulation is sufficiently general to handle structured LTI uncertainties. Simple constant thresholding was assumed for the threshold logic. Thresholding is widely used in commercial aerospace applications due to its simplicity. There are many other approaches for designing the decision logic, e.g. time-varying thresholds, statistical testing methods, and fuzzy logic [10], [11], [5]. The restriction to

constant thresholds can be viewed as a steady-state approximation for time-varying thresholds. The analysis in this paper forms a foundation to investigate more complicated decision functions. Finally, additive faults were considered in Figure 1. In fact, the form of the fault is unimportant because false alarm analysis only considers the fault free case.

A few basic facts are required before proceeding. Based on the assumptions, the residual r_k is a stationary zero-mean Gaussian process. For fixed $\Delta \in \mathbf{\Delta}$, the autocovariance function is defined as $\beta_l(\Delta) := E[r_k r_{k+l}]$. Note that $\beta_0(\Delta) := E[r_k^2]$ is the variance of the stationary residual process. The autocovariance sequence $\beta_l(\Delta)$ can be computed based on a frequency domain approach. The transfer function from input n_k to the FDI residual r_k is given by $T_{n\to r}(z,\Delta) := \Delta \star M(z)$. Thus the spectrum of the residual r_k is given by

$$\Phi_R(\omega,\Delta) = T_{n\to r}(e^{j\omega},\Delta) T_{n\to r}^T(e^{-j\omega},\Delta) \tag{4}$$

It is well-known [13] that the spectrum is related to the autocovariance coefficients by:

$$\Phi_R(\omega,\Delta) = \sum_{l=-\infty}^{\infty} \beta_l(\Delta) e^{-jl\omega} \tag{5}$$

Hence the autocovariance coefficients can be extracted as:

$$\begin{aligned} \beta_l(\Delta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_R(\omega,\Delta) e^{jl\omega} d\omega \\ &= \frac{1}{\pi}\int_{0}^{\pi} \Phi_R(\omega,\Delta)\cos(l\omega) d\omega \end{aligned} \tag{6}$$

III. Worst-case False Alarm Rate

This section considers the worst-case single step false alarm probability, i.e. the worst-case FAR P_1^*. The main result is that a robust $\mathcal{H}_2$ performance analysis problem can be used to obtain an upper bound on P_1^*.

For fixed $\Delta \in \mathbf{\Delta}$ the variance of the residual process is given by $\beta_0(\Delta) := E[r_k^2]$. Thus the FAR can be computed via a one-dimensional Gaussian integral:

$$\begin{aligned} P_1(\Delta) &= P[|r_1| > T] \\ &= 1 - \frac{1}{\sqrt{2\pi\beta_0^2(\Delta)}} \int_{-T}^{T} e^{-\frac{r_1^2}{2\beta_0^2(\Delta)}} dr_1 \end{aligned} \tag{7}$$

The worst-case FAR is given by

$$P_1^* = \max_{\Delta\in\mathbf{\Delta}} \left[1 - \frac{1}{\sqrt{2\pi\beta_0^2(\Delta)}} \int_{-T}^{T} e^{-\frac{r_1^2}{2\beta_0^2(\Delta)}} dr_1 \right] \tag{8}$$

The next lemma is useful for solving this optimization.

Lemma 1: Define

$$Q_1 := \frac{1}{\sqrt{2\pi}\beta_0} \int_{-T}^{T} e^{-\frac{r_1^2}{2\beta_0^2}} dr_1 \tag{9}$$

Q_1 monotonically decreases with $\beta_0 \in (0, +\infty)$.

Proof: By a change of variables Q_1 can be written as

$$Q_1 = \frac{1}{\sqrt{2\pi}} \int_{-\frac{T}{\beta_0}}^{\frac{T}{\beta_0}} e^{-\frac{r_1^2}{2}} dr_1 \tag{10}$$

$\frac{T}{\beta_0}$ monotonically decreases with β_0 and hence so does Q_1. ■

By Lemma 1 the worst-case false alarm probability is

$$P_1^* = 1 - \frac{1}{\sqrt{2\pi}} \int_{-T/\beta_0^*}^{T/\beta_0^*} e^{-\frac{r_1^2}{2}} dr_1 \tag{11}$$

where β_0^* is the worst-case variance:

$$\beta_0^* = \max_{\Delta\in\mathbf{\Delta}} \beta(\Delta) \tag{12}$$

Next note that Equation 6 with $l = 0$ implies that the variance can be computed as

$$\beta_0(\Delta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} T_{n\to r}(e^{j\omega},\Delta) T_{n\to r}^T(e^{-j\omega},\Delta) d\omega \tag{13}$$

The integral in this equation is precisely the discrete time $\mathcal{H}_2$ norm of $T_{n\to r}$, i.e. the variance is given by $\beta_0(\Delta) = \|T_{n\to r}(\Delta)\|_2$. Thus the worst-case variance optimization in Equation 12 is equivalent to a robust $\mathcal{H}_2$ analysis problem:

$$\beta_0^* = \max_{\Delta\in\mathbf{\Delta}} \|T_{n\to r}(\Delta)\|_2 \tag{14}$$

Previous results on robust $\mathcal{H}_2$ performance analysis can be found in [6], [14], [18] and the references contained therein. Standard multiplier techniques have been applied to compute upper bounds on the worst-case $\mathcal{H}_2$ norm for uncertain systems. This ultimately leads to convex, Semidefinite Programs (SDP) to compute the bounds. These existing methods to compute upper bounds on the robust $\mathcal{H}_2$ upper bound for the worst-case, single-step false alarm probability P_1^*.

IV. Worst-case False Alarm Probability

As noted previously, it is important in aerospace applications to compute upper bounds on P_N^*. This is the worst case false alarm probability over a fixed N-step window. This section presents two tractable upper bounds for P_N^*. For notational convenience define $Q_N(\Delta) := 1 - P_N(\Delta)$. For fixed $\Delta \in \mathbf{\Delta}$, $Q_N(\Delta)$ is the probability that no alarm is declared within the N-step window conditioned on the absence of a fault. For $N = 1$, this definition reduces to $Q_1(\Delta) = 1 - P_1(\Delta) = P[|r_k| \le T]$. The worst-case, N-step false alarm probability can thus be expressed as

$$P_N^* = 1 - \min_{\Delta\in\mathbf{\Delta}} Q_N(\Delta) \tag{15}$$

A. Worst-case: Sidak's Bound

Based on Definition 2, $P_N(\Delta)$ can be expressed as

$$\begin{aligned} P_N(\Delta) &= P[\cup_{k=1}^N \{|r_k| > T\}] \\ &= 1 - P[\cap_{k=1}^N \{|r_k| \le T\}] \end{aligned} \tag{16}$$

The residual r_k is a zero-mean Gaussian process for each $\Delta \in \mathbf{\Delta}$. The explicit dependence of r_k on Δ has not been denoted for simplicity. Sidak's probability bound (Theorem 1 in [16]) can be used to obtain the following inequality:

$$P_N(\Delta) \le 1 - \Pi_{k=1}^N P[|r_k| \le T] \tag{17}$$

Since r_k is stationary $P[|r_k| \le T] = Q_1(\Delta)$ for each k and hence $P_N(\Delta) \le 1 - Q_1^N(\Delta)$. This yields an upper bound on

the worst-case false alarm probability:

$$P_N^* \le \max_{\Delta\in\mathbf{\Delta}} \left[1 - Q_1^N(\Delta)\right] \le 1 - \min_{\Delta\in\mathbf{\Delta}} Q_1^N(\Delta) \tag{18}$$

By Lemma 1, a worst-case Δ maximizes the variance. Thus the N-step worst-case false alarm probability is bounded as:

$$P_N^* \le 1 - (1 - P_1^*)^N \tag{19}$$

where P_1^* is the worst-case single-step false alarm probability. Equation 19 provides a simple bound on the N-step worst-case false alarm probability. This bound neglects time-correlations that may exist in the residual process due to FDI filters and/or aircraft dynamics. For comparison with the next section define $\gamma_N^{(1)}(\Delta) = Q_1^N(\Delta)$ so that Sidak's bound is $P_N(\Delta) \le 1 - \gamma_N^{(1)}(\Delta)$.

B. Worst-case: Extended Sidak's Bound

Sidak's bound was recently extended to derive a sequence of monotonically improving bounds on $P_N(\Delta)$ [9]. The bounds apply to zero-mean stationary Gaussian processes. The first order bound used in the previous subsection is $P_N(\Delta) \le 1 - Q_1^N(\Delta)$ and this corresponds to Sidak's original result. For fixed $\Delta \in \mathbf{\Delta}$, $Q_1(\Delta)$ is a one-dimensional Gaussian integral. Given the variance $\beta_0(\Delta)$, this computation is easy, e.g. using the Matlab function `erf`.

The second-order bound derived in [9] is given by:

$$P_N(\Delta) \le 1 - \left(\frac{Q_2(\Delta)}{Q_1(\Delta)}\right)^{N-2} Q_2(\Delta) \tag{20}$$

where $Q_2(\Delta) := P[|r_k| \le T, |r_{k+1}| \le T]$. For fixed $\Delta \in \mathbf{\Delta}$, $Q_2(\Delta)$ is a two-dimensional Gaussian integral. Given the variance and one step covariance, i.e. $\beta_0(\Delta)$ and $\beta_1(\Delta)$, this integral can be efficiently computed in `Matlab`, e.g. using the `mvncdf` function [7]. This second-order bound is no worse than the first-order bound and it typically is significantly better. The improved second-order bound requires increased computation since a two-dimensional integral must be evaluated in addition to a one-dimensional integral. The results in [9] provide a sequence of improving bounds (third-order, etc) that rely on increasingly higher dimensional Gaussian integrals. In this paper only the second order bound defined above will be considered.

Define $\gamma_N^{(2)}(\Delta) := \left(\frac{Q_2(\Delta)}{Q_1(\Delta)}\right)^{N-2} Q_2(\Delta)$. With this notation, a tighter bound of P_N^* is given by:

$$P_N^* \le 1 - \min_{\Delta\in\mathbf{\Delta}} \gamma_N^{(2)}(\Delta) \tag{21}$$

Thus another bound of P_N^* can be obtained by solving:

$$\bar{\gamma}_N^{(2)} := \min_{\Delta\in\mathbf{\Delta}} \gamma_N^{(2)}(\Delta) = \min_{(\beta_0,\beta_1)\in\Upsilon_1} \gamma_N^{(2)}(\beta_0,\beta_1) \tag{22}$$

where Υ_1 is a subset of $\mathbb{R}^2$:

$$\Upsilon_1 = \{(\beta_0,\beta_1) : \beta_l = \frac{1}{\pi}\int_0^\pi \Phi_R(\omega,\Delta)\cos(l\omega)d\omega, \Delta \in \mathbf{\Delta}\}$$

There is a slight abuse of notation at this point. In particular, $\gamma_N^{(2)}$ is a function of the uncertainty Δ. The notation $\gamma_N^{(2)}(\beta_0,\beta_1)$ indicates that this bound can be written in a functional form that only shows its explicit dependence on β_0 and β_1. $\gamma_N^{(2)}$ still depends on Δ implicitly since β_0 and β_1 are functions of Δ. Similar notations such as $Q_1(\beta_0)$ and $Q_2(\beta_0,\beta_1)$ will also be used for simplicity.

Directly solving $\bar{\gamma}_N^{(2)}$ on Υ_1 is difficult. One alternative approach is enlarging Υ_1 to a larger set where the minimization could be easily solved and then obtaining a lower bound for $\bar{\gamma}_N^{(2)}$. Let $\mathbf{\Phi}$ be a set of spectrum functions:

$$\mathbf{\Phi} := \{\Phi_R(\omega,\Delta) : \Delta \in \mathbf{\Delta}\} \tag{23}$$

which is a subset of $\tilde{\mathbf{\Phi}}$:

$$\tilde{\mathbf{\Phi}} := \left\{\Phi_R : \inf_{\Delta\in\mathbf{\Delta}} \Phi_R(\omega,\Delta) \le \Phi_R \le \sup_{\Delta\in\mathbf{\Delta}} \Phi_R(\omega,\Delta)\right\}$$

Then a bigger set of (β_0,β_1) is defined as:

$$\tilde{\Upsilon}_1 := \left\{(\beta_0,\beta_1) : \beta_l = \frac{1}{\pi}\int_0^\pi \Phi_R(\omega)\cos(l\omega)d\omega, \Phi_R \in \tilde{\mathbf{\Phi}}\right\}$$

Since $\mathbf{\Phi} \subset \tilde{\mathbf{\Phi}}$, hence $\Upsilon_1 \subset \tilde{\Upsilon}_1$. A lower bound of $\bar{\gamma}_N^{(2)}$ is:

$$\tilde{\gamma}_N^{(2)} := \min_{(\beta_0,\beta_1)\in\tilde{\Upsilon}_1} \gamma_N^{(2)}(\beta_0,\beta_1) \le \bar{\gamma}_N^{(2)} \tag{24}$$

$\tilde{\mathbf{\Phi}}$ is a convex set and the mapping from $\tilde{\mathbf{\Phi}}$ to $\tilde{\Upsilon}_1$ is affine, hence $\tilde{\Upsilon}_1$ is a convex set. One can search $\tilde{\gamma}_N^{(2)}$ on a small subset of $\tilde{\mathbf{\Phi}}$ based on the convexity and the following lemma:

Lemma 2: For fixed β_0, $\gamma_N^{(2)}(\beta_0,\beta_1)$ is an even function of β_1 and is monotonically non-decreasing for $\beta_1 \ge 0$.

Proof: It is trivial to check that $Q_2(\beta_0,\beta_1)$ is an even function of β_1 and so is $\gamma_N^{(2)}(\beta_0,\beta_1)$. For fixed β_0, $Q_2(\beta_0,\beta_1)$ is a non-decreasing function of the correlation coefficient $\frac{\beta_1}{\beta_0}$ on the interval $[0,1)$ (Theorem 1 in [17]). Since β_0 is fixed, $Q_2(\beta_0,\beta_1)$ is also non-decreasing with $\beta_1 \ge 0$ and so is $\gamma_N^{(2)}(\beta_0,\beta_1)$. ■

$\tilde{\Upsilon}_1$ is a convex set, hence at least one point satisfying $\beta_1 = 0$ or at the boundary of $\tilde{\Upsilon}_1$ achieves $\tilde{\gamma}_N^{(2)}$ based on Lemma 2. Denote the boundary of $\tilde{\Upsilon}_1$ as $\partial\tilde{\Upsilon}_1$. To be more specific, define the following two classes of boundary points:

$$\partial\tilde{\Upsilon}_1^+ := \{(\beta_0,\beta_1) : \beta_1 = \max_{(\beta_0',\beta_1')\in\tilde{\Upsilon}_1, \beta_0'=\beta_0} \beta_1'\} \tag{25}$$

$$\partial\tilde{\Upsilon}_1^- := \{(\beta_0,\beta_1) : \beta_1 = \min_{(\beta_0',\beta_1')\in\tilde{\Upsilon}_1, \beta_0'=\beta_0} \beta_1'\} \tag{26}$$

The above definitions make perfect sense geometrically due to the convexity of $\tilde{\Upsilon}_1$. It is clear $\partial\tilde{\Upsilon}_1 = \partial\tilde{\Upsilon}_1^+ \bigcup \partial\tilde{\Upsilon}_1^-$. An efficient algorithm searching $\tilde{\gamma}_N^{(2)}$ relies on the next lemma:

Lemma 3: There exists at least one point $(\tilde{\beta}_0,\tilde{\beta}_1) \in \partial\tilde{\Upsilon}_1$ such that $\gamma_N^{(2)}(\tilde{\beta}_0,\tilde{\beta}_1) = \tilde{\gamma}_N^{(2)}$.

Proof: The set of points with $\beta_1 = 0$ is defined as:

$$\mathcal{B}_1(\tilde{\Upsilon}_1) := \{(\beta_0,\beta_1) : (\beta_0,\beta_1) \in \tilde{\Upsilon}_1, \beta_1 = 0\} \tag{27}$$

Then set $\mathcal{B}(\tilde{\Upsilon}_1) := \partial\tilde{\Upsilon}_1 \bigcup \mathcal{B}_1(\tilde{\Upsilon}_1)$. Lemma 2 and the convexity of $\tilde{\Upsilon}_1$ imply that at least one point in $\mathcal{B}(\tilde{\Upsilon}_1)$ achieves $\tilde{\gamma}_N^{(2)}$. Since $\tilde{\Upsilon}_1$ is convex, $\tilde{\Upsilon}_1$ and the β_0-axis can have 0 or 2 intersections (identical or not). When they have 0 intersection, $\mathcal{B}_1(\tilde{\Upsilon}_1)$ is an empty set so there exists at least one point $(\tilde{\beta}_0,\tilde{\beta}_1) \in \partial\tilde{\Upsilon}_1$ such that $\gamma_N^{(2)}(\tilde{\beta}_0,\tilde{\beta}_1) = \tilde{\gamma}_N^{(2)}$.

When they have 2 intersections, denote the intersections as $(\beta_0^{(1)}, 0))$ and $(\beta_0^{(2)}, 0))$ and suppose $\beta_0^{(1)} \leq \beta_0^{(2)}$. Hence $\mathcal{B}_1(\tilde{\Upsilon}_1) = \{(\beta_0, \beta_1) : \beta_0^{(1)} \leq \beta_0 \leq \beta_0^{(2)}, \beta_1 = 0\}$. Then for any $(\beta_0, \beta_1) \in \mathcal{B}_1(\tilde{\Upsilon}_1)$, Lemma 1 implies: $\gamma_N^{(2)}(\beta_0, \beta_1) = \gamma_N^{(2)}(\beta_0, 0) = Q_1^N(\beta_0) \geq Q_1^N(\beta_0^{(2)}) = \gamma_N^{(2)}(\beta_0^{(2)}, 0)$. Since $(\beta_0^{(2)}, 0) \in \partial\tilde{\Upsilon}_1$, hence the stated lemma is true. ■

Based on Lemma 3, a basic idea of solving $\tilde{\gamma}_N^{(2)}$ is finding out $\partial\tilde{\Upsilon}_1$ and then searching for $\tilde{\gamma}_N^{(2)}$ on a finite grid of $\partial\tilde{\Upsilon}_1$. The main task is to find an efficient way to generate a dense grid on the set $\partial\tilde{\Upsilon}_1$. Theory of uniformly continuous operator is now applied to realize this goal.

From now on, denote $\phi_1(\omega) := \inf_{\Delta\in\mathbf{\Delta}} \Phi_R(\omega)$ and $\phi_2(\omega) := \sup_{\Delta\in\mathbf{\Delta}} \Phi_R(\omega)$. Let $M^- : [0, \pi] \to \mathbb{R}^2$ denote the operator that maps c to $(\beta_0, \beta_1) = M^-(c)$ by

$$\beta_0 = \frac{1}{\pi}\int_0^c \phi_1(\omega)d\omega + \frac{1}{\pi}\int_c^\pi \phi_2(\omega)d\omega \tag{28}$$

$$\beta_1 = \frac{1}{\pi}\int_0^c \phi_1(\omega)\cos(\omega)d\omega + \frac{1}{\pi}\int_c^\pi \phi_2(\omega)\cos(\omega)d\omega \tag{29}$$

In a similar manner, define operator $M^+ : [0, \pi] \to \mathbb{R}^2$ mapping c to $(\beta_0, \beta_1) = M^+(c)$ by

$$\beta_0 = \frac{1}{\pi}\int_0^c \phi_2(\omega)d\omega + \frac{1}{\pi}\int_c^\pi \phi_1(\omega)d\omega$$

$$\beta_1 = \frac{1}{\pi}\int_0^c \phi_2(\omega)\cos(w)\omega d\omega + \frac{1}{\pi}\int_c^\pi \phi_1(\omega)\cos(\omega)d\omega$$

One can show M^- and M^+ are both uniformly continuous. The range spaces $\mathcal{I}(M^-) = \partial\tilde{\Upsilon}_1^-$ and $\mathcal{I}(M^+) = \partial\tilde{\Upsilon}_1^+$. For details, see Lemma 4 and its proof in Appendix. Hence a sufficient dense finite grid on $\partial\tilde{\Upsilon}_1$ can be generated by applying M^- and M^+ on a sufficiently dense grid of c on $[0, \pi]$. Then one can compute $\gamma_N^{(2)}(\beta_0, \beta_1)$ for points on the generated grid of $\partial\tilde{\Upsilon}_1$ and search the minimum value as $\tilde{\gamma}_N^{(2)}$.

To sum up, the algorithm is first generating a grid of c on $[0, \pi]$, and then applying the integral operator M^+ and M^- numerically to get a grid of (β_0, β_1) in $\mathbb{R}^2$. The next step is computing $\gamma_N^{(2)}(\beta_0, \beta_1)$ on this resulted grid and searching the minimum value as $\tilde{\gamma}_N^{(2)}$. Finally an upper bound of worst-case N-step false alarm probability P_N^* can be obtained by $\alpha_N^{(2)} := 1 - \tilde{\gamma}_N^{(2)} \geq P_N^*$.

For realistic application, $\phi_1(\omega)$ and $\phi_2(\omega)$ sometimes may be hard to obtain. Then any lower bound of $\phi_1(\omega)$ and upper bound of $\phi_2(\omega)$ can be used as replacements to obtain the set $\tilde{\mathbf{\Phi}}$. If $\tilde{\mathbf{\Phi}}$ happens to contain the worst-case spectrum associated with the worst-case Δ, $\alpha_N^{(2)}$ will always be tighter than the worst-case Sidak's bound in Inequality 19. Otherwise, $\alpha_N^{(2)}$ could become conservative.

V. Numerical Examples

This section presents a numerical example to demonstrate the proposed worst-case false alarm analysis method. A simple fault detection scheme for monitoring additive aileron faults is considered. Consider again the parity-equation scheme in Figure 2. The healthy, nominal dynamics for the aircraft roll mode are modeled by a first order process from aileron to roll-rate: $G_0 = -\frac{0.0161}{z-0.9878}$. This is a discrete-time version of Example 7.2 in [4] assuming a $100Hz$ sample rate. For simplicity the uncertainty is modeled as multiplicative, real gain uncertainty at the input. In particular, $W_u = 0.1$ and $\Delta \in [-1, 1]$ represent a 10% gain uncertainty. Proportional control with a gain $K = -5$ is used to track roll-rate commands. The measurement noise v_k is assumed to be an IID Gaussian process with $v_k \sim \mathcal{N}(0, 1)$. The disturbance, e.g. wind gusts, has slower dynamics and hence w_k is modeled by a first-order autoregressive model with transfer function $\frac{0.5}{z-0.995}$ and an IID Gaussian innovation $\sim \mathcal{N}(0, 1.3)$. The example analyzes the per-hour false alarm probability. For a $100Hz$ sample rate, this corresponds to $N = 3.6 \times 10^5$ sample frames per hour. The decision logic threshold is chosen to be $T = 45$.

The worst-case extended Sidak's bound $\alpha_N^{(2)} = 1 - \tilde{\gamma}_N^{(2)}$ will be computed based on the method proposed in Section IV-B. For this simple problem, one can directly grid $\Delta \in [-1, 1]$ and compute the bounds $\left(1 - \gamma_N^{(2)}(\Delta)\right)$ and $\left(1 - \gamma_N^{(1)}(\Delta)\right)$ over this grid of Δ. This will benchmark the performance of $\alpha_N^{(2)}$. The set $\tilde{\mathbf{\Phi}}$ for this case can be easily specified since $\phi_1(\omega)$ and $\phi_2(\omega)$ can be numerically searched for any fixed frequency over the set $\Delta \in [-1, 1]$. `chebfun` [19] is one `Matlab` toolbox used for computations. Its ability of handling complex variables can be explored to conveniently compute the spectrum of r_k when G_0, K and Δ are all specified. `chebfun` handles numerical integration also well so that the spectrum function can then be directly integrated to get β_0 and β_1 based on Equation 6. The `Matlab` function `mvncdf` [7] is then used to compute probability bounds based on β_0 and β_1.

Figure 4 shows a comparison of $\alpha_N^{(2)}$ and the sampled bounds. One can see $\left(1 - \gamma_N^{(2)}(\Delta)\right)$ are better bounds than $\left(1 - \gamma_N^{(1)}(\Delta)\right)$ since the time correlations introduced by the system dynamics and the FDI filter cannot be ignored in this case. The solution $\alpha_N^{(2)}$ obtained from analysis method proposed in Section IV-B is a very tight worst-case bound for $\left(1 - \gamma_N^{(2)}(\Delta)\right)$ over $\Delta \in [-1, 1]$ and hence should be a good upper bound for the worst-case N-step false alarm probability P_N^*. It meets the expectation since one can check that the set $\tilde{\mathbf{\Phi}}$ contains the worst-case spectrum associated with the worst-case $\Delta = 1$.

This benchmark problem is artificial in the way that $\tilde{\mathbf{\Phi}}$ happens to contain the worst-case spectrum in set $\mathbf{\Phi}$. For realistic problems, the worst-case spectrum in set $\mathbf{\Phi}$ may not be contained in $\tilde{\mathbf{\Phi}}$. Then $\alpha_N^{(2)}$ could be conservative.

VI. Conclusion

This paper analyzed the worst-case false alarm probability of a FDI system over a class of stochastic disturbances and model uncertainty. The single step, worst-case false alarm probability is shown to be equivalent to a robust $\mathcal{H}_2$ analysis problem. Next, two upper bounds are derived for the worst-case false alarm probability over multiple time steps. The

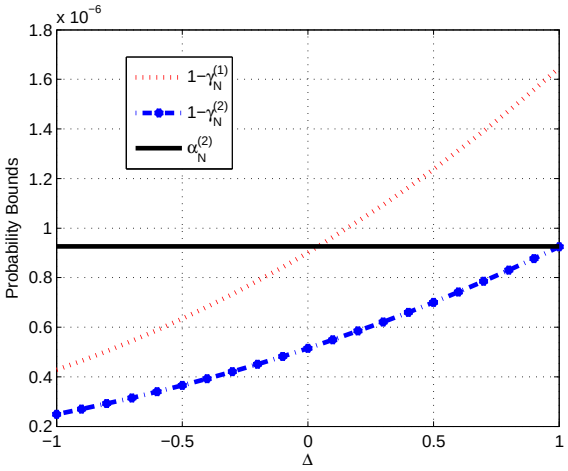

Fig. 4. Comparison of the worst-case analysis solution and the sampled bounds $\gamma_N^{(1)}(\Delta)$ and $\gamma_N^{(2)}(\Delta)$.

worst-case Sidak's bound is also related to the worst-case FAR. The worst-case extended Sidak's bound accounts for the time correlations introduced by the system dynamics and FDI filters. A numerical example is used to demonstrate the proposed technique. In the future, the proposed analysis method will be explored in more complicated applications.

VII. Acknowledgments

This work was supported by NASA under Grant No. NRA NNX12AM55A entitled "Analytical Validation Tools for Safety Critical Systems Under Loss-of-Control Conditions", Dr. Christine Belcastro technical monitor. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of NASA.

References

[1] ADDSAFE: Advanced fault diagnosis for sustainable flight guidance and control. http://addsafe.deimos-space.com/, 2012. European 7th Framework Program.

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

[3] R.P.G. Collinson. *Introduction to Avionic Systems*. Kluwer, 2003.

[4] M. V. Cook. *Flight Dynamics Principles, Second Edition: A Linear Systems Approach to Aircraft Stability and Control*. Butterworth-Heinemann, 2nd edition, 2007.

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

[6] Eric Feron. Analysis of Robust H2 Performance Using Multiplier Theory. *SIAM J. Control Optim.*, 35(1):160–177, January 1997.

[7] A. Genz. Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and Computing*, 14:251–260, 2004.

[8] B. Hu and P. Seiler. A probabilistic method for certification of analytically redundant systems. In *Proceedings of 2nd International Conference on Control and Fault-Tolerant Systems*, pages 13–18, 2013.

[9] B. Hu and P. Seiler. Probability bounds for false alarm analysis of fault detection systems. In *Proceedings of 51st Annual Allerton Conference on Communication, Control, and Computing*, pages 989–995, 2013.

[10] I. Hwang, S. Kim, Y. Kim, and C.E. Seah. A survey of fault detection, isolation, and reconfiguration methods. *IEEE Transactions on Control Systems Technology*, 18(3):636–653, 2010.

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

[12] W.S. Lee, D.L. Grosh, A.F. Tillman, and C.H. Lie. Fault tree analysis, methods, and applications: a review. *IEEE Trans. on Reliability*, 34(3):194–203, 1985.

[13] L. Ljung. *System Identification: Theory for the User*. Pearson Education, 2nd edition, 1998.

[14] Fernando Paganini. Convex methods for robust H2 analysis of continuous-time systems. *IEEE Transactions on Automatic Control*, 44:239–252, 1999.

[15] R. J. Patton and J. Chen. Robust fault detection using eigenstructure assignment: A tutorial consideration and some new results. In *Proceedings of the IEEE Conference on Decision and Control*, 1991.

[16] Z. Sidak. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967.

[17] Z. Sidak. On multivariate normal probabilities of rectangles: Their dependence on correlations. *The Annals of Mathematical Statistics*, 39(5):pp. 1425–1434, 1968.

[18] M. Sznaier, T. Amishima, P.A. Parrilo, and J. Tierno. A convex approach to robust H2 performance analysis. *Automatica*, 38(6):957 – 966, 2002.

[19] L. N. Trefethen et al. *Chebfun Version 4.2*. The Chebfun Development Team, 2011. `http://www.maths.ox.ac.uk/chebfun/`.

[20] K. Zhou, J.C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice-Hall, 1996.

Appendix

Lemma 4: M^- and M^+ are uniformly continuous. The range space of M^- is denoted by $\mathcal{I}(M^-)$. Then $\mathcal{I}(M^-) = \partial\tilde{\Upsilon}_1^-$. The range space of M^+ is denoted by $\mathcal{I}(M^+)$. Then $\mathcal{I}(M^+) = \partial\tilde{\Upsilon}_1^+$.

Proof: We will only prove the uniform continuity of M^- and $\mathcal{I}(M^-) = \partial\tilde{\Upsilon}_1^-$. The proof for M^+ is just identical. Due to the fact that $|\phi_1(\omega) - \phi_2(\omega)| \leq \sup_{0\leq\omega\leq\pi} |\phi_1(\omega) - \phi_2(\omega)|$, the uniform continuity of M^- is straightforward to prove by applying an ϵ-δ argument and triangle inequality.

To prove $\partial\tilde{\Upsilon}_1^+ \subset \mathcal{I}(M^-)$, suppose $(\beta_0', \beta_1') \in \partial\tilde{\Upsilon}_1^-$, it suffices to prove $(\beta_0', \beta_1') \in \mathcal{I}(M^-)$. All points in $\tilde{\Upsilon}_1$ satisfy:

$$\underline{\beta_0} = \frac{1}{\pi}\int_0^\pi \phi_1(\omega)d\omega \leq \beta_0 \leq \frac{1}{\pi}\int_0^\pi \phi_2(\omega)d\omega = \bar{\beta_0} \quad (30)$$

And so does β_0'. Notice the mapping from c to β_0 described by Equation 28 is continuous. The mapping also maps $c = 0$ to $\beta_0 = \underline{\beta_0}$ and maps $c = \pi$ to $\beta_0 = \bar{\beta_0}$. Hence there exists $c' \in [0, \pi]$ such that $\beta_0' = \frac{1}{\pi}\int_0^{c'} \phi_1(\omega)d\omega + \frac{1}{\pi}\int_{c'}^\pi \phi_2(\omega)d\omega$. To prove $\partial\tilde{\Upsilon}_1^- \subset \mathcal{I}(M^-)$, it is sufficient to show that $M^-(c') = (\beta_0', \beta_1')$. Hence, one only needs to show that $\beta_1 \geq \frac{1}{\pi}\int_0^{c'} \phi_1(\omega)\cos(\omega)d\omega + \frac{1}{\pi}\int_{c'}^\pi \phi_2(\omega)\cos(\omega)d\omega$ for all points $(\beta_0, \beta_1) \in \tilde{\Upsilon}_1$ satisfying $\beta_0 = \beta_0'$. It is equivalent to show that $\frac{1}{\pi}\int_0^\pi \Phi_R(\omega)\cos(\omega)d\omega \geq \frac{1}{\pi}\int_0^{c'} \phi_1(\omega)\cos(\omega)d\omega + \frac{1}{\pi}\int_{c'}^\pi \phi_2(\omega)\cos(\omega)d\omega$ holds for any $\Phi_R \in \tilde{\mathbf{\Phi}}$ under the constraint $\frac{1}{\pi}\int_0^\pi \Phi_R(\omega)d\omega = \frac{1}{\pi}\int_0^{c'} \phi_1(\omega)d\omega + \frac{1}{\pi}\int_{c'}^\pi \phi_2(\omega)d\omega$. This statement can be easily proved based on the monotonicity of $\cos(\omega)$ for $\omega \in [0, \pi]$ by: $\int_0^{c'} (\Phi_R - \phi_1)\cos(\omega)d\omega \geq \cos(c')\int_0^{c'} (\Phi_R - \phi_1)\, d\omega = cos(c')\int_{c'}^\pi (\phi_2 - \Phi_R)\, d\omega \geq \int_{c'}^\pi (\phi_2 - \Phi_R)\cos(\omega)d\omega$.

Finally to prove $\mathcal{I}(M^-) \subset \partial\tilde{\Upsilon}_1^-$, it suffices to show that $M^-(c) \in \partial\tilde{\Upsilon}_1^-$ for any $c \in [0, \pi]$. It is equivalent to show that $\frac{1}{\pi}\int_0^\pi \Phi_R(\omega)\cos(\omega)d\omega \geq \frac{1}{\pi}\int_0^c \phi_1(\omega)\cos(\omega)d\omega + \frac{1}{\pi}\int_c^\pi \phi_2(\omega)\cos(\omega)d\omega$ holds for any $\Phi_R \in \tilde{\mathbf{\Phi}}$ under the constraint $\frac{1}{\pi}\int_0^\pi \Phi_R(\omega)d\omega = \frac{1}{\pi}\int_0^c \phi_1(\omega)d\omega + \frac{1}{\pi}\int_c^\pi \phi_2(\omega)d\omega$. This has already been proved in last paragraph. ■