Threshold Selection for Timely Fault Detection in Feedback Control Systems

Tesheng Hsiao and Masayoshi Tomizuka
Department of Mechanical Engineering,
University of California at Berkeley, Berkeley, CA94720-1740
tshsiao@berkeley.edu tomizuka@me.berkeley.edu

Abstract— **The inevitable time delay in fault detection, i.e. the time between the occurrence and detection of the fault, can cause stability problems in feedback control systems. The state of the art technologies, however, provide neither an analysis tool to evaluate the adverse influence of the detection delay on the closed-loop system's stability and performance nor guidelines for designers to overcome the instability introduced by the detection delay. In this paper we conduct a theoretical analysis of *timely fault detection* problem which concerns the detection of faults *before* the closed-loop system's performance deteriorates to an unacceptable extent. We first give a formal definition of the "timely" fault detection problem. Then we derive the upper and lower bounds of thresholds of fault detectors as well as a set of detectable faults. If the threshold is selected within the bounds, faults belonging to the detectable set can be detected in a timely manner and an acceptable level of performance is guaranteed.**

1. Introduction

Modern safety-critical systems such as aircrafts, space stations and nuclear plants must meet increasingly stringent performance requirements during normal operation while assuring stability in the event of malfunctions in actuators, sensors or other components of the system. Fault tolerant control systems are thus gaining increasing attention for enhancing safety and reliability of complex systems. Among all available techniques, a mixture of reconfigurable controllers and fault detectors provides considerable design flexibility and has drawn a lot of attention in the research community [13]. In this structure, stability of the closed loop system after any fault has taken place is guaranteed if (i) faults can be detected in a "timely" manner; in other words, the control system must be aware of the existence of faults *before* the system crashes and (ii) the reconfigurable controller takes immediate and effective actions to accommodate faults right after they have been detected. While the latter has been the topic of many research efforts, the former which we call the *timely fault detection* (TFD) problem remains an open question.

A series of questions can be raised with regard to the TFD problem. Firstly, since the detection delay is inevitable, how do we characterize it in terms of stability and performance degradation? In other words, how long can the delay be in order to assure an acceptable level of performance during the detection delay? To answer this question, we must be able to measure (i) the acceptable level of deterioration of performance, (ii) relationship between faults and performance degradation, and (iii) obscurity of the fault detector's sensitivity to faults by the robust controller. Further questions are listed below. It is impractical to expect the fault detector to detect all types of faults. Then in what sets the faults are guaranteed to be detected in a timely manner? Is the set specified in the previous question broad enough to contain faults that are expected to strike the system? Given a fault detector, how do we evaluate its TFD ability? Given possible types of faults and performance requirements, how do we synthesize a fault detector which has the desired TFD property?

In this paper we focus on sensor and actuator failures and provide partial answers to some of these questions. First of all, a formal definition of TFD will be given which concerns questions (i)~(iii). Instead of proposing a complete design methodology to synthesize a fault detector with desired properties, we assume that the fault detector is given and that the only design parameter is the threshold. (The structure of the fault detector will be discussed in the next section.) Then we illustrate analytically how to select the thresholds to meet the TFD requirement. The proposed approach is applicable to complex systems with LTI robust *residual generators* (see Section 2 and 4).

This paper is organized as follows: Section 2 gives a brief discussion of the fault detector design problem. Section 3 introduces the notational conventions that will be adopted throughout this paper. We give a precise definition of TFD in Section 4 and derive upper and lower bounds of the thresholds in Section 5. The last section concludes this paper.

2. Design of Fault Detector: Overview

Whatever methodology we follow, the fault detector can be cast in a general framework consisting of two stages as shown in Figure 1 [1].

The residual generator takes sensor measurements as inputs and generates *residuals* which are small, ideally zero, when there is no fault and significantly large when fault has taken place. Due to the effect of disturbances, model uncertainties and measurement noise, the residuals are not zero even when there are no faults. A robust residual generator should alleviate these effects while maintaining sensitive to faults.

The decision making stage returns Positive/Negative answer to the question: is there any fault existing in the

system? The answer is given based on the results of comparing the sizes of residuals with the thresholds which are either fixed or adaptive [4][10][15][17]. Setting the thresholds low results in high *false positive* rates (alarms are issued under no fault conditions). On the other hand setting the thresholds high increases the *false negative* rates (alarms are missed when faults have taken place.) Clearly the selection of the thresholds is closely related to robustness (w.r.t. disturbances) and sensitivity (to faults) of the residual generator. Considerable research efforts has been devoted to the design of a robust residual generator ([1][2][5][14]) whereas existing methods for threshold selection are mostly heuristic. For example, the thresholds are determined by observing the experimental data. A well-designed residual generator facilitates the selection of thresholds and heuristic methods are reasonable in many cases; however, the heuristic approaches may lead to an uncertain result when unforeseen faults have taken place. This is unacceptable for safety-critical systems. We prefer a guaranteed and consistent performance based on an analytic method for threshold selection.

Analytical method for selection of the thresholds has been explored by several researchers. Emami-Naeini et al investigate the threshold selector which causes no false alarm [4]. Stoustrup et al propose an optimal threshold which results in equal false positive and false negative rates in statistical hypotheses testing [17]. Rank et al give conditions for no false alarms and no missed alarms [15]. Estimation of the fault occurrence time through Markov process has also been proposed by Mahmoud et cl[12]. However, few of these works addressed the question of *when* faults can be detected because most of them considered the problem in the frequency domain. In order to answer the TFD questions posed in Section 1, we study the problem in the time domain. To simplify the presentation, we first introduce the notational conventions in the next section.

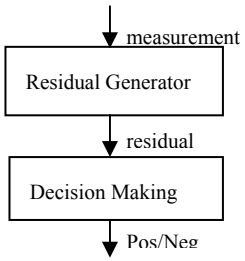

Figure 1 General framework of fault detectors

3. Notational Conventions

Let $f : \mathbf{R} \to \mathbf{R}$. $\|f\|_p$ denotes the L^p *norm* of f for p=1,2,… ∞.

The *truncated signal* of f is defined as $f_t(\tau) = \begin{cases} f(\tau) & \tau \le t \\ 0 & \tau > t \end{cases}$. $f \in L^p_e$ if $f_t \in L^p$ for all t.

L^p_e is called the extended L^p space [3]. If $f \in L^1$, we denote its Laplace transform as $\hat{f}$.

The input-output relation of an LTI system has several equivalent representations. Throughout this paper, we use uppercase letters for the time-domain operator which maps input signals to output signals. Lowercase letters represent its impulse response, e.g. if G is an LTI system, then g and $\hat{g}$ are its impulse response and transfer function respectively.

4. Problem Formulation

4.1 General form of a residual generator

As mentioned in Section 1, we assume that the residual generator is given and we will select the thresholds. Therefore we start from a general form of the residual generator by expressing the residuals $\boldsymbol{r}$ as follows:

$$\boldsymbol{r} = \boldsymbol{G}_f \boldsymbol{f} + \boldsymbol{G}_d \boldsymbol{d} = \boldsymbol{v} + \boldsymbol{G}_d \boldsymbol{d} , \qquad (1)$$

where $\boldsymbol{f}$ and $\boldsymbol{d}$ are faults and disturbances respectively and $\boldsymbol{v}=\boldsymbol{G}_f\boldsymbol{f}$. $\boldsymbol{G}_f$ and $\boldsymbol{G}_d$ are given while $\boldsymbol{f}$ and $\boldsymbol{d}$ are unknown. For each entry r_i of $\boldsymbol{r}$, we set up a threshold T_i. An alarm is issued if for any residual r_i, there exists t such that

$$|r_i(t)| > T_i \qquad (2)$$

We make the following assumptions on the given residual generator to facilitate the discussion.

(A1) G_f is LTI, SISO, stable, causal and strictly proper.

(A2) There exists a positive constant D such that $\|\boldsymbol{G}_d\boldsymbol{d}\|_\infty \le D$ for all $\mathbf{d}$

Note that Assumption (A2) implies that the residual generator is robust w.r.t. $\boldsymbol{d}$. This assumption is quite restrictive and may not be satisfied by all residual generators.

4.2 Definition of timely fault detection

In this subsection, we give a formal definition of timely fault detection (TFD). A *bounded detection delay with a fixed upper bound* looks like, at first glance, a reasonable definition, i.e. if a fault detector can detect all possible faults within a fixed time we say the fault detector has the TFD property. This definition will lead to a statement like "this fault detector can detect faults within 5 seconds once they have taken place," which is itself incomplete and flawed. The flaws are three folds. (i) The start time of incipient faults is ambiguous and so is the detection delay in this case. (ii) When a robust controller and a fault detector cooperate in a feedback loop, the controller is designed to make all states and signals in the loop insensitive to possible perturbations, including faults; however the fault detector's sensitivity relies on the information conveyed by these signals to monitor the abnormality of the system. The controller and the fault detector have conflicting goals from this point of view. Thus the fixed detection delay property is incomplete

without specifying the controller in use. (iii) The fixed delay time reveals little information about the stability and performance of the closed loop system. The performance degradation depends on its bandwidth and the types of faults taken place in the system. A 5-second delay is negligible for a chemical process with a 10-hour time constant but is detrimental to a vehicle running on the highway.

The concept of TFD in the context of feedback control system is entangled with the following factors: the types of faults, the interaction with the feedback controller and the allowable performance degradation before faults are detected. A reasonable definition must take all these factors into consideration and remain as simple as possible so that it can be applied easily to evaluate the fault detector's performance. The bounded delay definition is too restrictive. A more flexible definition is based on the system's performance: if the faults can be detected, no matter how long it takes, before the controller fails to sustain an acceptable level of performance, then we say the faults are detected in a timely manner. A natural question that would be raised at this point is how to quantify the system's performance under the influence of faults?

Several performance indices are widely used in the control community, e.g. quadratic performance (LQG controller design), H^∞-norm criterion, or ratio of current output covariance to minimum output covariance [6]. However these indices are either proposed for controller design or for a stochastic setting, which may not be suitable for our own purpose. To make our definition simple we use the size (w.r.t. a predefined norm) of the fault as a performance index. In most cases, the fault signal is a function of the state and time. Its size increases as the state deviates further away from the origin. Large (sensor/ actuator) faults usually imply poor regulation control. Therefore the size of fault is used as a criterion to determine TFD property. We further classify the faults according to their sizes into three levels each of which corresponds to a different degree of performance degradation.

1. If the size of the fault is less than some positive number K_{min}, the robust controller can handle it well and there is no need to issue alarms. Actually the alarm should be forbidden since after the alarm is issued, the controller switches to a degraded mode and results in a worse performance because the degraded mode controller uses only parts of the sensor information and actuator power.
2. Let the positive number K_{max} be the worst performance level that is tolerable for the normal mode controller. If the size of the fault is greater than K_{max}, the controller must switch to the degraded mode in order to sustain an acceptable level of performance.
3. If the size of the fault is between K_{min} and K_{max}, it does not matter whether the alarm is issued or not because the normal mode and the degraded mode controller may achieve similar performance.

Now we can give our definition of TFD. From the discussion above, TFD means that the residual should exceed the threshold T (hence the alarm is issued) before the size of the fault grows beyond K_{max}. We also give the definition of *no false alarm*: if the size of the fault is less then K_{min}, the residual should be less than the threshold T (hence no alarm will be issued). From (2) L^∞-norm is a natural choice to measure the size of the residual. Therefore we choose L^1-norm as a measure of the size of faults because of mathematical tractability (e.g. Hölder's inequality can be applied to the pair of L^∞-norm and L^1-norm [16]). We express these ideas formally as follows:

No false alarms:

$$\exists T>0 \text{ s.t. } \|f_t\|_1 < K_{min} \Rightarrow \|r_t\|_\infty < T \quad \text{for all } t \qquad (3)$$

No missed alarms (timely fault detection):

$$\exists T>0 \text{ s.t. } \|f_t\|_1 \geq K_{max} \Rightarrow \|r_t\|_\infty \geq T \quad \text{for all } t \qquad (4)$$

Notice that the two positive numbers K_{min} and K_{max} are given based on the controller's ability and the performance requirement prior to the analysis. Hence the sophisticated relations among controller, plant and fault detector are condensed into these two parameters. A simple example is given below to illustrate the idea of choosing $\|f\|$ as the performance index and the choices of K_{min} and K_{max}.

Example 1. A linear time invariant system suffering from sensor and actuator faults can be modeled as follows:

$$\dot{\boldsymbol{x}}(t) = \mathbf{A}_p \boldsymbol{x}(t) + \mathbf{B}_p \boldsymbol{u}(t) + \mathbf{E}_p \boldsymbol{w}(t) + \mathbf{f}_a(t)$$
$$\boldsymbol{y}(t) = \mathbf{C}_p \boldsymbol{x}(t) + \mathbf{f}_s(t) + \boldsymbol{n}(t)$$

where $\boldsymbol{x}$, $\boldsymbol{u}$, $\boldsymbol{y}$, $\boldsymbol{w}$, $\boldsymbol{n}$, $\mathbf{f}_s$, and $\mathbf{f}_a$ are state, input, output, disturbance, measurement noise, sensor fault and actuator fault respectively. In the feedback control case, $\mathbf{u}$ is a function of $\boldsymbol{y}$ or $\boldsymbol{x}$. If the sensors are broken, the fault signals can be expressed as $\mathbf{f}_s = -\mathbf{C}_p \boldsymbol{x}$. Hence $\|\mathbf{f}_s\|$ is the same as the output regulation error in this case. To determine K_{min} and K_{max}, we can regard $\mathbf{f}_a$ and $\mathbf{f}_s$ as disturbance and measurement noise and simulate the performance of the controller. An alternative is to incorporate K_{mn} and K_{max} into the controller design phase, e.g. H^∞ controller design method, and find out the controller for the worst case scenario.

■ ■ ■

5. Admissible region of the threshold

In this section, we assume that the residual generator (1) is given. We will explore the upper and lower bounds of the thresholds such that (3) and (4) are satisfied. By Assumption (A1) G_f is stable and strictly proper; thus

$g_f \in L^1 \cap L^\infty$. We also assume that the fault f belongs to both extended L^1 and L^∞ spaces, i.e. $f \in L^1_e \cap L^\infty_e$

5.1 No false alarms case

Since we have assumed that the residual generator is robust (Assumption (A2)), it is easy to find a lower bound of the threshold for no false alarm case.

Assume for any time t, $\|f_t\|_1 < K_{\min}$. Then

$$\|r_t\|_\infty \le \|g_f * f_t\|_\infty + \|G_d d\|_\infty \le \|g_f\|_\infty \|f_t\|_1 + D < K_{\min}\|g_f\|_\infty + D =: L$$

If we choose $T \ge L$, then (3) is satisfied, i.e. there will be no false alarm for all $f \in L^1_e \cap L^\infty_e$

5.2 No missed alarms – minimum phase case

In this and next sections, we will derive the upper bound of the threshold satisfying (4). The no missed alarm condition is related to the sensitivity of the residual generator. We have to consider the worst case scenario which corresponds to the smallest gain of the residual generator [7][11]. However the smallest gain is zero among arbitrary faults: e.g. G_f is strictly proper and faults are sinusoidal functions with arbitrarily high frequencies. Then the residual becomes arbitrarily small and cannot be detected. Thus the selection of the threshold depends on the types of faults we are interested in. We therefore make the following assumptions on the types of faults we want to detect:

(F1) The bandwidth of the faults is less than Ω

(F2) $\exists \gamma > 0$ such that $\gamma \|f_t\|_1 \le \|f_t\|_\infty$ for all t

Remark: Consider a constant fault with a small magnitude ε, $0 < \varepsilon << 1$. The L^1-norm of the truncated fault signal will eventually exceed K_{max} after a sufficiently long time. But the corresponding residual could be smaller than any fixed threshold as $\varepsilon \to 0$. This implies that faults with small L^∞-norm cannot be detected. Assumption (F2) excludes this situation by assuming that the L^∞-norm of the detectable faults is lower bounded by its L^1-norm.

The smallest gain of a system can be derived from the operator norm of its inverse system provided that the inverse system exists and is bounded. In this subsection, we assume that G_f is minimum phase. Thus $\hat{g}^{-1}$ is stable but improper. Suppose the relative degree of G_f is p; then define

$$\hat{\varphi}(s) = \prod_{i=1}^{p+1} \frac{\omega_i}{s + \omega_i}, \text{ for } \omega_1 > \omega_2 > \ldots \omega_{p+1} >> \Omega > 0 \qquad (5)$$

Hence $\hat{h} = \hat{\varphi}\hat{g}_f^{-1}$ is stable and proper. For $\omega < \Omega$ we have $\hat{h}(j\omega) \approx \hat{g}_f^{-1}(j\omega)$. To be more precise [16], let $v = G_f f$ and $\hat{\bar{f}} = \hat{h}\hat{v} = \hat{\varphi}\hat{f}$. Then given any $\varepsilon > 0$, we can choose ω_1, $\omega_2, \ldots \omega_{p+1}$ in (5) such that

$$\|f - \bar{f}\|_1 < \varepsilon . \qquad (6)$$

If $\|f_t\|_1 > K_{\max}$, then

$$\|r_t\|_\infty \ge \|v_t\|_\infty - D \ge \frac{\|\bar{f}_t\|_\infty}{\|h\|_1} - D \ge \frac{\gamma(K_{\max} - \varepsilon)}{\|h\|_1} - D =: U_1$$

If we choose $T \le U_1$, faults satisfying Assumptions (F1) and (F2) will always be detected in a timely manner.

Remark: If $U_1 \le 0$, it does not necessarily mean that we cannot find a positive threshold to meet (4) since the condition given here is only sufficient. To obtain a positive U_1, we may either redesign the residual generator to reduce D or redesign the controller to increase K_{max}.

5.3 No missed alarms – non-minimum phase case

In this subsection, we assume that G_f is non-minimum phase. Hence $\hat{g}_f^{-1}$ is unstable and improper.

Let $z > 0$ be a non-minimum phase zero of G_f and $f(t) = e^{zt}$. Then $v = G_f f \to 0$ as $t \to \infty$, but $f \to \infty$ as $t \to \infty$, i.e. f cannot be detected. Therefore given a residual generator G_f, we have to further restrict the set of detectable faults with respect to G_f. One way to do this is to require that the detectable set consists of faults whose corresponding residuals are lower bounded by the state of G_f. To describe this idea formally, we first express G_f in a state space form:

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}f \\ v &= \boldsymbol{C}\boldsymbol{x} \end{aligned} \qquad (7)$$

For simplicity, we assume that G_f has distinct poles. Let $\boldsymbol{A} \in \mathbf{C}^{n \times n}$ be diagonal and $\boldsymbol{C} = [1,1,\ldots 1]$. Then for each fault f, we define:

$$\alpha_f = \sup\left\{ \frac{|\boldsymbol{x}(t)|}{\|v_t\|_\infty};\ \forall t \text{ s.t.} \|f_t\|_1 > K_{\max} \right\} \qquad (8)$$

where $|\bullet|$ denotes the vector norm in the Euclidean space while $\|\bullet\|$ denotes the L^p-norm in the functional space. In the definition (8) we are only interested in the time interval at which the L^1-norm of f_t exceeds K_{max} since (8) will be applied to derive the upper bound of the threshold which satisfies (4).

The definition of α_f is based on the state space realization (7). If $\alpha_f < \infty$ and we perform the state transformation: $\boldsymbol{z} = \boldsymbol{N}\boldsymbol{x}$, where $\boldsymbol{N} \in \mathbf{C}^{n \times n}$ is invertible. Then $|z(t)|_\infty \le \|\boldsymbol{N}\|_\infty |\boldsymbol{x}(t)|_\infty \le \alpha_f \|\boldsymbol{N}\|_\infty \|v_t\|_\infty$ Hence we can choose arbitrary state realizations in (8). The choice of diagonal matrix $\boldsymbol{A}$ is for simplicity.

Given $0 < \alpha < \infty$, we define the set F_α as follows:

$$F_\alpha = \left\{ f \in L^1_e \cap L^\infty_e ; \alpha_f \le \alpha \right\} \qquad (9)$$

We made the following assumption about the detectable faults:

(F3) $f \in F_\alpha$ for some given α, $0 < \alpha < \infty$

Example 2 below illustrates how to compute α_f for faults with the form $e^{\sigma t}\cos(\omega t)$.

Example 2 Consider $f = \mathrm{Re}(e^{\beta t})$, where $\beta=\sigma+j\omega$, $\omega>0$ and $\hat{g}_f(\beta) \neq 0$. Suppose $A=\mathrm{diag}(\lambda_1, \lambda_2,\ldots\lambda_n)$ in (7) and $\beta<<\lambda_i$, i=1,2,…n. Then by direct computation we have

$$x_i(t) = \frac{b_i}{2}\left(\frac{1}{\beta-\lambda_i}e^{\beta t} + \frac{1}{\overline{\beta}-\lambda_i}e^{\overline{\beta} t} - e^{\lambda_i t}\left(\frac{1}{\beta-\lambda_i}+\frac{1}{\overline{\beta}-\lambda_i}\right)\right)$$

where $\overline{\beta}$ denotes the complex conjugate of β

Therefore $\alpha_f \approx \frac{e^{\pi\sigma}}{2}\frac{|\beta|}{|\hat{g}(\beta)|\omega}\max_{1\le i\le n}\left\{\frac{1}{|\beta-\lambda_i|}+\frac{1}{|\overline{\beta}-\lambda_i|}\right\}$

If $\omega=0$, the formula can be even simpler:

$$\alpha_f \approx \frac{1}{|\hat{g}(\beta)|}\max_{1\le i\le n}\left\{\frac{1}{|\beta-\lambda_i|}\right\}$$

■ ■ ■

Remark:

(1) The requirement $\beta<<\lambda_i$, i=1,2,…n. implies that the response of the fault detector is much faster than the fault and therefore the transient response does not affect the detection.

(2) For $\omega>0$ and $\sigma>0$, α_f increases exponentially as σ increases. For a fixed α, we can always find a fault f with sufficient large σ such that $f\notin F_\alpha$. This implies that we cannot detect arbitrarily fast faults in a timely manner. This result is consistent with Assumption (F1).

Because $\hat{g}_f$ has one or more non-minimum phase zeros, the inverse system is unstable. To overcome the instability problem, we utilize the *non-causal* stable impulse response of $\hat{g}_f^{-1}$ For LTI systems, multiple impulse responses may be transformed to the same transfer function with different regions of convergence (ROC) [9]. If the LTI system does not have pure imaginary poles, then we can always find a stable (maybe non-causal) impulse response.

Example 3: Let $\hat{g}(s) = \frac{1}{s-1}$ If the ROC is $\mathrm{Re}\{s\}>1$, then the corresponding impulse response is $g_1(t) = e^t 1(t)$, where $1(t) = \begin{cases} 1 & t \ge 0 \\ 0 & t < 0 \end{cases}$. g_1 is causal and unstable.

On the other hand if the ROC is $\mathrm{Re}\{s\}<1$, the corresponding impulse response is $g_2(t) = -e^t 1(-t)$, which is anti-causal and stable ($g_2 \in L^1$)

■ ■ ■

We make another assumption on the residual generator:

(A3) $\hat{g}_f$ does not have pure imaginary zeros.

Assumption (A3) guarantees that we can always find a stable (but non-causal) impulse response of $\hat{g}_f^{-1}$ Denote it as $\tilde{g}_f^{-1}$. Suppose that $v=G_f f$, then

$$\hat{f} = \hat{g}_f^{-1}\hat{v} \quad \text{or} \quad f = \tilde{g}_f^{-1} * v \qquad (10)$$

To compute f from the convolution equation (10), two conditions must be satisfied: (i) The ROC's of $\hat{g}_f^{-1}$ and $\hat{v}$ must have nonempty intersection and (ii) we must know the whole time history of v. At present time t, the future values of $v(\tau), \tau>t$, is unknown. However we can conceptually construct a fictitious fault $\tilde{f}$ such that $\tilde{f}(\tau) = f(\tau)$ for $\tau\le t$ and assign arbitrary values for $\tilde{f}(\tau)$, $\tau>t$ as long as $\tilde{f} \in L^1 \cap L^\infty$. Consequently we get a fictitious signal $\tilde{v} = g_f * \tilde{f}$. Because g_f is causal and stable, $\tilde{v}(\tau) = v(\tau)$ for $\tau\le t$ and $\tilde{v} \in L^1 \cap L^\infty$. Hence both the ROCs of $\hat{g}_f^{-1}$ and $\hat{v}$ contain the imaginary axis and the whole time history of $\tilde{v}$ is available. Therefore $f(\tau) = \tilde{f}(\tau) = (\tilde{g}_f^{-1} * \tilde{v})(\tau)$ for $\tau\le t$. Since $\tilde{f}(\tau)$ can be assigned arbitrarily for $\tau>t$, $\tilde{f} = f_t$ is a convenient choice.

Now $\tilde{g}_f^{-1}$ is stable but improper. We can apply the same procedure in Section 5.2 to construct a stable proper approximation $\tilde{h}$ of $\tilde{g}_f^{-1}$ such that (6) holds for a given $\varepsilon>0$. If $\|f_t\|_1 > K_{max}$, then

$$\|\tilde{v}\|_\infty \ge \frac{\gamma(K_{max}-\varepsilon)}{\|\tilde{h}\|_1} \qquad (11)$$

Since $\tilde{f} = f_t$, for $s>t$ we have

$$|\tilde{v}(s)| = |\boldsymbol{C}\boldsymbol{x}(s)| \le |\boldsymbol{C}|_2 |\boldsymbol{x}(s)|_2 \le \sqrt{n}\left\|e^{\boldsymbol{A}(s-t)}\right\|_2 |\boldsymbol{x}(t)|_2 \le n|\boldsymbol{x}(t)|_\infty$$

According to Assumption (F3), $|\boldsymbol{x}(t)|_\infty \le \alpha\|v_t\|_\infty$ and $\tilde{v}(\tau) = v(\tau)$, $\tau\le t$. We have

$$\|\tilde{v}\|_\infty = \sup_s |\tilde{v}(s)| = \max\{\|v_t\|_\infty, \sup_{s>t}|\tilde{v}(s)|\} \le M\|v_t\|_\infty$$

where $M=\max\{1,n\alpha\}$. From (11) we obtain the upper bound U2:

$$\|r_t\|_\infty \ge \|v_t\|_\infty - D \ge \frac{\gamma(K_{max}-\varepsilon)}{M\|\tilde{h}\|_1} - D =: U_2$$

If $T<U_2$, then there will be no missed alarms and all faults satisfying (F1)~(F3) can be detected in a timely manner in the sense of definition (4).

5.4 Discussion

From Sections 5.1~5.3, we conclude that if there exists a threshold T such that $L\le T\le U_1$, (or U_2), then there will be no false alarms and no missed alarms for faults satisfying (F1)~(F3). Furthermore, suppose that G_f has non-minimum phase zeros and let $K_{min}=kK_{max}$, $0<k\le 1$. Rearranging the inequality, we have

$$\frac{D}{K_{max}} \le \frac{1}{2}\left(\frac{\gamma}{M\left\|\tilde{h}\right\|_1} - k\left\|g_f\right\|_\infty \right) =: R \qquad (12)$$

A large ratio of D/K_{max} is desirable since it implies that the fault detector is robust w.r.t. large disturbance while sensitive to small faults. (12) also indicates the limitation of a fault detector. On the other hand, suppose that the disturbance ***d*** in (1) has unity L^∞-norm, then $D = \|g_d\|_1$ (12) also serves as a performance index for the design of fault detectors. Namely the systems G_f and G_d in (1) can be selected by maximizing R/D. If (12) holds, we conclude that the fault detector has the TFD property. The design problem will be the subject of future research.

6. Conclusion

In this paper, we identified timely fault detection problem as a key issue for guaranteeing the stability of the fault tolerant feedback control system. Its formal definition was discussed extensively and we derived the upper and lower bounds of the thresholds for no missed and false alarms. The sets of faults that are guaranteed to be detected were described explicitly. The procedure proposed in this paper can serve as an analysis tool to evaluate the performance of a fault detector. Further research efforts are required to relax the restrictive assumptions (e.g. Assumption (A2)) and to develop a guideline for selecting critical parameters (e.g. K_{min} and K_{max}). The synthesis problem of the fault detector satisfying the timely fault detection property is also a future research topic.

Acknowledgement

This work was supported by the California Department of Transportation (CalTrans) under PATH TO4205.

References

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

[2]. Dassanake, S.K, et al, *Using unknown input observers to detect and isolate sensor faults in a turbofan engine*, Proc. of Digital Avionics Systems Conferences, 2000, pp6E5/1 - 6E5/7

[3]. Desore, C. A. and Vidyasagar, M., *Feedback Systems: Input-Output Properties*, Academic Press, 1975

[4]. Emami-Naeini, A., Akhter, M. M. and Rock, S. M., *Effect of Model Uncertainty on Failure Detection: The Threshold Selector*, IEEE Trans. On Automatic Control, Vol. 33, No. 12, pp1106~1115

[5]. Frank, P.M., *Enhancement of Robustness in Observer-based Fault Detection*, International Journal of Control, Vol. 59, No. 4, pp955-981, 1994.

[6]. Harris, T. J., Seppala, C. T., and Desborough, L. D., *A Review of Performance Monitoring and Assessment Techniques for Univariate and Multivariate Control System*s, Journal of Process Control Vol.9, 1999, pp1-17

[7]. Hou, M. and Patton, R. J., *An LMI Approach to H-/H$_\infty$ Fault Detection Observers*, UKACC International Conference on Control, 1996

[8]. Hsiao, T. and Tomizuka, M., *Observer-based Sensor Fault Detection and Identification with Application to Vehicle Lateral Control*, Proc. of the American Control Conference, 2004, pp810~815

[9]. Jackson, L. B., *Signals, Systems, and Transforms*, Addison Wesley, 1991

[10]. Le, K., et al. *Adaptive Thresholding – A Robust Fault Detection Approach,* Proc. of Conference on Decision and Control, 1997, pp4490~4495

[11]. Liu, J., Wang, J. L., Yang, G-H, *An LMI Approach to Worst Case Analysis for Fault Detection Observers*, Proc. of the American Control Conference, 2003.

[12]. Mahmoud, Mufeed, Jiang, Jin, and Zhang, Youmin, Active Fault Tolerant Control Systems, Springer, 2003

[13]. Patton, R. J., *Fault-Tolerant Control Systems: The 1997 Situation*, IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes, 1997

[14]. Patton, R.J. and Chen, J., *Robust Fault Detection Using Eigenstructure Assignment: A Tutorial Consid- eration and Some New Results*, Proc. of Conference on Decision and Control, 1991, pp2242~2246

[15]. Rank, M. L., Niemann H. *Norm Based Design of Fault Detector*, International Journal of Control, Vol 72, No. 9, 1999, pp773~783.

[16]. Rudin, W., *Real and Complex Analysis*, McGraw-Hill, 1987.

[17]. Stoustrup J., Niemann, H. and Cour-Harbo, A. J. *Optimal Threshold Functions for Fault Detection and Isolation*, Proc. of the American Control Conference, 2003, pp1782~1787.