

Optimal Threshold Functions for Fault Detection and Isolation

Jakob Stoustrup

Department of Control Engineering
Aalborg University
Fr. Bajers Vej 7C
9220 Aalborg, Denmark
E-mail: jakob@control.auc.dk
www.control.auc.dk/~jakob/

Henrik Niemann

Ørsted•DTU, Automation
Technical University of Denmark
Building 326
DK-2800 Lyngby, Denmark
E-mail: hhn@oersted.dtu.dk
www.oersted.dtu.dk/staff/all/hhn/

Anders la Cour-Harbo

Department of Control Engineering
Aalborg University
Fr. Bajers Vej 7C
9220 Aalborg, Denmark
E-mail: alc@control.auc.dk
www.control.auc.dk/~alc/

Abstract—Fault diagnosis systems usually comprises two parts: a filtering part and a decision part, the latter typically based on threshold functions. In this paper, a systematic way to choose the threshold values is proposed. A test function for the filtered signals is proposed and a method is described for determining the optimal threshold value for this test function in order to detect presence or absence of a certain fault in the filtered signal. Optimal is in this context taken to mean minimizing the larger of the probabilities for making a false positive decision and for making a false negative decision, which is equivalent to balancing these two probabilities.

I. INTRODUCTION

Most of the emphasis in the literature on fault diagnostic systems is made on designing filters for fault detection or fault estimation, [1], [2], [3], [4], [5], [6], [7], [8], [9]. The output of such filters are signals that take values in a continuous range. Usually, the approaches which have been suggested assumes that the final diagnostic can be designed by applying thresholds to the output of the designed filters. Often it is left to the system designer to choose appropriate values for these thresholds. In this paper, a method for choosing such thresholds based on system knowledge will be proposed. In contrast, [2] contains a thorough treatment of an overall design based on statistical hypothesis test theory for detecting abrupt changes. In the present paper, both abrupt and incipient faults will be treated. A typical structure of a fault diagnostic system is shown in Fig. 1.

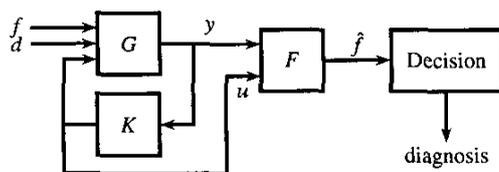


Fig. 1. A typical structure for fault diagnostic systems, where G is the plant, K a controller, F a fault estimation filter, and the 'Decision' box generates the actual diagnosis. f is a vector of fault signals, d is a vector of disturbance signals, \hat{f} is an estimate of the faults, and 'diagnosis' is a set of logical signals, indicating which (if any) fault has occurred.

Fig. 1 illustrates a system G with a controller K for which the known inputs u (control signals) and outputs y (measurements) are fed to a fault estimation filter F . The system is subjected to two unknown sets of inputs: faults f and disturbances/noise d . The output of F is an estimate \hat{f} of the actual faults f which should be designed in such a way, that the effects of the disturbances d are as small as possible.

This paper will only focus on the design on the decision block, so some simplifying assumptions will be made. First of all,

it will be assumed that the system is invertible, such that the transfer function from f to \hat{f} can be made constant or at least diagonal. If this is not satisfied in practice, there will be cross-couplings that will compromise optimality in the results below. It is possible to extend the approach suggested to the non-invertible case. However, in many cases, the methods proposed will give reasonable results even for non-invertible systems even without modification. The second assumption is that the fault estimation errors $e = \hat{f} - f$ are entrywise independent and normally distributed. This might also not apply in many practical cases, but as with the other assumption, it is possible to generalize the concept, and the methods would work in many cases without modification.

The purpose of the decision logic is to decide for each fault whether it has occurred at some time instance, [2], [3], [5], [7], [10]. The common design objective is to make this decision with a predetermined error rate. Making an error means making the wrong decision. The error rate might be set once and for all, [3], [5], it might be run-time adjustable (adaptive tuning), [3], [11], or it might be given by user input. In any case, it is important to distinguish between the two basic types of errors, false positive (FP), i.e. deciding that a fault has occurred in a healthy situation, and false negative (FN), i.e. deciding that a certain situation is healthy, although a fault has actually occurred. In other words, a FP decision is a false alarm, and a FN decision is an actual fault, which was ignored.

It is easy to see that the two objectives of making the probability of making a FP decision small resp. of making the FN probability small are always contrary. Which probability should be emphasized is truly application specific. In the sequel, we shall suggest to minimize the larger of the two. This is obviously equivalent to making the two probabilities equal. However, it is straightforward to introduce another weighting between the two probabilities in the methods proposed below.

The results derived in this paper is based on methods derived in [12], where a general principle for design of active sensors has been given.

A. Thresholds on the Fault Estimates

The easiest way of detecting a fault is to fix a threshold above which a fault estimate is considered to specify a signal level above which a fault is assumed and below which the signal is assumed to be the cause of disturbances/noise, [3], [13], [5], [14]. In most fault diagnostic systems, this approach is used to provide input to the supervisory level of the system. Typically, the threshold is set to a level which empirically yields a predetermined probability of FP under given conditions. If the distribution and variance of the noise is known, it is easy to

determine the threshold. This method employs a fixed threshold and is therefore easy to implement.

Although applying an FP-based threshold directly to the fault estimates is straightforward and simple to do, this approach does not yield the true error rate since it does not include the FN probability. The challenge when including the FN probability is that it depends on the noise level as well as the desired maximum sensitivity to small fault signal amplitudes whereas the FP probability depends on the noise level only. This is illustrated in the following example.

Assume that the noise in a given control system is normally distributed with zero mean and standard deviation 10, and that the error rate is specified to 10^{-6} . This corresponds to 4.75 times the standard deviation, and the threshold should be set to 48 accordingly (assuming that only integers are allowed). This ensures that the probability of detecting a fault when there is none is 10^{-6} . However, if a fault is present and results in a true fault estimate of 48 half of the estimates will be below the threshold, and thus the FN probability is 0.5. When the fault signal develops to larger amplitudes and the true estimate increases to 96 the FN probability drops to 10^{-6} . In both cases the FP probability is 10^{-6} . If the diagnostic system is specified to have an error rate of (at most) 10^{-6} this is only valid for the faults which causes a true estimate of (at least) 96.

Note that while the fixed threshold on the measurements works fine in a white noise scenario (when the above consideration are taken into account) the method lacks the ability to properly distinguish between large estimates caused by a large amplitude fault and by powerful noise.

B. Adaptive Threshold Levels

In the sequel, only the decision part of the diagnostic system will be considered. Thus, for notational reasons, the fault estimates will be referred to simply as y rather than $\hat{f}(t)$. Thus, time dependencies will be suppressed, although it will be assumed that the estimates are averaged over a small time window. The procedure will be described for a specific fault. The individual faults are assumed to be independent. Thus, the procedure should be repeated for each fault. Methods for designing estimation based fault detectors are described in e.g. [15], [16], [17], [18], [19]. Methods for exact and optimal design of observers for signal estimation can be found in [20].

The deficiency of the fixed FP-based threshold detection demonstrates the need for an adaptive decision method, [3], [21], [11], [22]. In the following subsections an adaptive method is reported. It is based on the principle of regular signal-to-noise ratios (SNR),

$$10 \log_{10} \frac{y_k^2}{\sum_{n=M}^N y_n^2} \quad k = 0, 1, \dots, M-1,$$

where M is the number of possible faults and $N+1$ the number of samples in the time window considered. This means that instead of detecting the faults by a threshold on the estimates, the detection is based on some sort of SNR, called a detection function. The function in the two detection methods reported here is

$$\Theta(y) \equiv \frac{y_0^2}{\sum_{k=0}^N y_k^2} \quad (1)$$

where $N+1$ is the length of the y signal. This implicitly assumes that y_0 is the fault under consideration and the other entries y_1 through y_N are noise. Having information about the noise, it is natural to compare the (potential) fault signal to the noise directly as in (1).

The detection scheme takes the following form:

$$\mathcal{T}(y, \alpha) = \begin{cases} \text{false} & \Theta(y) < \alpha, \\ \text{true} & \text{otherwise.} \end{cases}$$

The purpose of using this method is to be able to properly detect faults in the case of severe noise. At the same time it must be able to provide a detection scheme with a predetermined error rate for normally distributed noise. This goes for FP as well as FN errors.

II. DETECTION METHOD

Two parameters have to be determined in order to use the detection method. They can obviously be determined empirically by trial and error. However, if one wants to quantify the error rates it is necessary to know the relation between instances of the (almost) stochastic process y , and α . As argued in the beginning of Section I it is reasonable to require the FP and FN error rates to be equal in the random-noise worst-case scenario, i.e. in the case where the signal is the weakest possible and yet still useful.

In the following a statistical model for balancing the probability of a FP decision and a FN decision is presented. The entire exercise is about determining the correct α . Obviously, the probability P_{FP} of detecting a signal when none was received decreases with large values of α . And vice versa, when α is small, the probability P_{FN} of ignoring a signal that was actually received is also small. Thus, choosing α can be seen as a compromise between FP and FN risks.

The purpose of the statistical model is to determine the optimal threshold α . We define an optimal threshold as that value of α for which the probability for a worst-case FN decision based on $\mathcal{T}(\alpha)$ equals the probability for an FP decision based on $\mathcal{T}(\alpha)$. A worst-case FN decision is understood as an FN decision in the presence of the faintest received signal which is to be considered useful. It is straightforward to modify the approach below in order to meet this compromise with a preference to either a low FP or a low FN probability.

Note that in the following model Θ has been divided by N since this makes the denominator resemble the variance of the signal.

A. Statistical Model for Deterministic Fault Signal

The reader is reminded that if N stochastic variables $\{Y_k\}_{k=1 \dots N}$ are normally distributed, $Y_k \in N(0, 1)$, then $\sum_{k=1}^N Y_k^2$ belongs to the $\chi^2(N)$ distribution, which is a special case of the Γ distribution, $\chi^2(N) = \Gamma(\frac{N}{2}, 2)$. This means that the probability of an FP decision is

$$P_{FP}(\alpha) = P(Z_0 > \alpha \bar{Z})$$

where

$$Z_0 = Y_0^2 \in \Gamma\left(\frac{1}{2}, 2\right)$$

and

$$\bar{Z} = \frac{1}{N} \sum_{k=1}^N Y_k^2 \in \Gamma\left(\frac{N}{2}, 2\right),$$

and the probability of an FN decision is

$$P_{FN}(\alpha, R_{\max}) = P(R_{\max} \sigma^2 < \alpha \bar{Z}),$$

where σ^2 is the true (and unknown) variance of the noise. R_{\max} is the worst-case SNR, i.e. $R_{\max} = y_{\text{low}}/\sigma$, where y_{low} is the lowest detectable signal level of y_0 . Note that this makes R_{\max} the 'real' SNR since y_0 is (for the time being) assumed to be deterministic. R_{\max} would typically be a design parameter

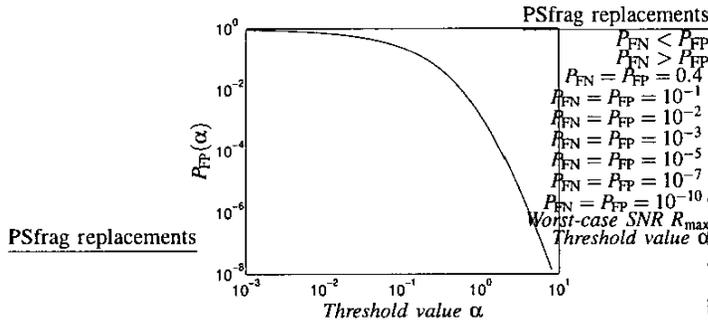


Fig. 2. The probability for making an FP decision for a given SNR threshold α and for $N = 13$.

or adjustable by the user of the diagnostic system. Now, the probability distribution function P_{FP} can be computed by PSfrag replacements

$$P_{FP}(\alpha) = \iint_A f_{z_0} f_{\bar{z}} dz_0 d\bar{z}, A = \{(z_0, \bar{z}) : z_0, \bar{z}, z_0 - \alpha \bar{z} > 0\}$$

$$= \iint_A \frac{z_0^{-\frac{1}{2}} e^{-\frac{z_0}{2}}}{\Gamma(\frac{1}{2})\sqrt{2}} \times \frac{\bar{z}^{-\frac{N-2}{2}} e^{-\frac{\bar{z}}{2}}}{\Gamma(\frac{N}{2})2^{\frac{N}{2}}} dz_0 d\bar{z} \quad (2)$$

Introducing polar coordinates, the integral in (2) becomes

$$\frac{2^{-\frac{N+1}{2}}}{\Gamma(\frac{1}{2})\Gamma(\frac{N}{2})} \int_0^{\arctan \frac{1}{\alpha}} \int_0^{\infty} \left(\frac{r^{N-1} e^{-r(\cos\theta + \sin\theta)}}{\cos\theta \sin^{N-2}\theta} \right)^{\frac{1}{2}} dr d\theta \quad (3)$$

The double integral can be separated into two single integrals by the substitution

$$r = \frac{\bar{r}}{\cos\theta + \sin\theta},$$

that is,

$$P_{FP}(\alpha) = K_1 \int_0^{\arctan \frac{1}{\alpha}} \left(\frac{(\cos\theta + \sin\theta)^{3-N}}{\cos\theta \sin^{N-2}\theta} \right)^{\frac{1}{2}} d\theta \quad (4)$$

$$K_1 = \frac{2^{-\frac{N+1}{2}}}{\Gamma(\frac{1}{2})\Gamma(\frac{N}{2})} \int_0^{\infty} \bar{r}^{\frac{N-1}{2}} e^{-\frac{\bar{r}}{2}} d\bar{r}.$$

Finally, it can be shown that the standard substitution $t = \tan \frac{\theta}{2}$ leads to the following algebraic integrand

$$P_{FP}(\alpha) = 2^{\frac{N}{2}} K_1 \int_0^{\sqrt{1+\alpha^2}-\alpha} \frac{t^{\frac{N-2}{2}}}{(2-(t-1)^2)^{\frac{N-1}{2}} \sqrt{1-t^2}} dt. \quad (5)$$

Even though the integral in (5) is algebraic, it can not be resolved analytically. However, numerical experiments show that e.g. an adaptive recursive Newton Cotes 8 panel rule performs better on (5) than on (4). The resulting probability function for $N = 13$ is shown in Fig. 2 ($N = 13$ is chosen to match the experimental signals in Section III).

Exploiting the probability function $P_{FP}(\alpha)$ which can be evaluated numerically by (5) it is straightforward to get a calibrating curve for α under the constraint that $P_{FP} = P_{FN}$. To that end, we start with a value of α , and numerically determine $P_{FP}(\alpha)$. Then the inverse of the $\chi^2(N)$ distribution function applied to $P_{FP}(\alpha)$ yields the ratio between R_{max} and α . This relationship is shown in the top plot of Fig. 3.

B. Statistical Model for Stochastic Fault Signals

In the analysis above, we have modeled the false negative situation as receiving a noise-free signal y_0 of a certain magnitude, which is incorrectly classified as noise, since the real noise signal y_1 through y_N happens to be large at the same time. This is of course unphysical to some extent, but was done in order to simplify the expressions.

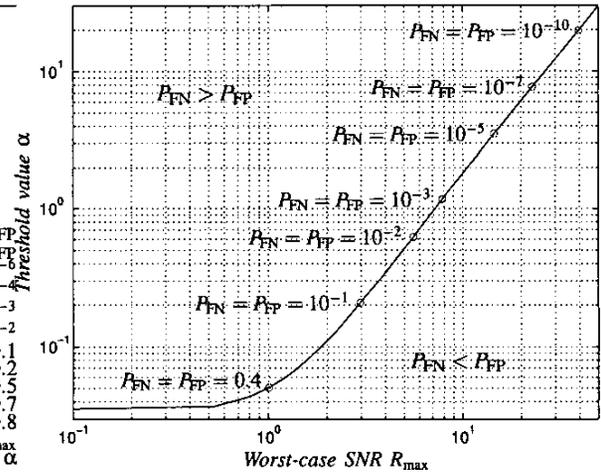
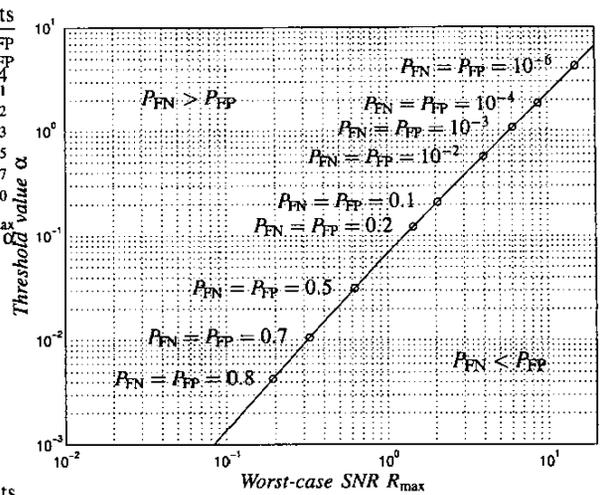


Fig. 3. The curves show the relationship between the worst-case SNR value R_{max} and the threshold value α in the test $\mathcal{T}(\alpha)$. The top plot for a deterministic y_0 and the bottom plot for a stochastic y_0 . Here $N = 13$.

A more realistic model is obtained by assuming that y_0 is an outcome of a stochastic variable, also in the false negative decision case. The analysis in principle involves the same steps as above, but the algorithm to compute the calibration curve now becomes a bit more involved.

With respect to the false positive decision case, nothing is changed and the false positive probabilities as a function of the threshold value α can be precomputed. In order to determine the calibration case, the best approach is to choose a grid of values for the threshold value, α . Then the task is to determine for each value of α , a corresponding value of the signal-to-noise-ratio (SNR) which leads to the same probability for a false negative decision as for a false positive decision for that α . It is obvious that for a fixed value of α , the false negative probabilities are monotone (non-decreasing) functions of the SNR. Thus, the right values of SNR can be found for instance by a simple bisection approach with SNR as the independent variable. For fixed values of α and SNR, the false negative probabilities can be computed as

$$P_{FN}(\alpha, SNR) = P\left\{ \frac{y_0^2}{\sum_{i=1}^N y_i^2} < \alpha : y_0 \in N(y_{min}, \sigma), y_i \in N(0, \sigma), i = 1 \dots N \right\}$$

where y_{min} denotes the smallest possible (mean) signal received

(at least the smallest for which the algorithm is guaranteed to fulfill the specified probabilities). Introducing $\text{SNR} = \frac{y_0}{\sigma}$, $\xi_0 = \frac{y_0}{\sigma}$, and $\bar{\xi} = \frac{1}{\sigma^2} \sum_{n=1}^N y_n^2$ (which is then $\chi^2(N)$ -distributed), we obtain

$$\begin{aligned}
 P_{\text{FN}}(\alpha, \text{SNR}) &= P \left\{ \frac{\sigma^2 \xi_0^2}{\sigma^2 \bar{\xi}} < \alpha : \xi_0 \in N(\text{SNR}, 1), \bar{\xi} \in \chi^2(N) \right\} \\
 &= P \left\{ \xi_0^2 < \alpha \bar{\xi} : \xi_0 \in N(\text{SNR}, 1), \bar{\xi} \in \chi^2(N) \right\} \\
 &= \iint_{\xi_0^2 < \alpha \bar{\xi}} f_{N(\text{SNR}, 1)}(\xi_0) f_{\chi^2(N)}(\bar{\xi}) d\xi_0 d\bar{\xi} \\
 &= \int_{\bar{\xi}=0}^{\infty} \left(F_{N(\text{SNR}, 1)} \left(\sqrt{\alpha \bar{\xi}} \right) - F_{N(\text{SNR}, 1)} \left(-\sqrt{\alpha \bar{\xi}} \right) \right) \times f_{\chi^2(N)}(\bar{\xi}) d\bar{\xi} \quad (6)
 \end{aligned}$$

The calibration curve resulting from applying the bisection algorithm mentioned above based on numerical evaluations of (6), can be seen in the bottom plot of Fig. 3. Comparing the two subplots of Fig. 3, it is easy to see that they are almost identical for large SNRs, whereas the bottom plot suggests larger values of α for small SNR values. In fact, the bottom curve has a left, horizontal asymptote for $\text{SNR} \rightarrow -\infty$, which is easy to verify. Indeed, as $\text{SNR} \rightarrow -\infty$ the measurements y_0 for the false positive and the false negative decisions asymptotically belong to the same distribution, i.e. $N(0, 1)$. Thus, the requirement $P_{\text{FP}} = P_{\text{FN}}$ in the limit leads to

$$\begin{aligned}
 P \left\{ \xi_0^2 < \alpha \bar{\xi} : \xi_0 \in N(0, 1), \bar{\xi} \in \chi^2(N) \right\} \\
 = P \left\{ \xi_0^2 > \alpha \bar{\xi} : \xi_0 \in N(0, 1), \bar{\xi} \in \chi^2(N) \right\}
 \end{aligned}$$

However, since these two probabilities in this case are obviously related also by $P_{\text{FP}} = 1 - P_{\text{FN}}$, we obtain $P_{\text{FP}} = P_{\text{FN}} = \frac{1}{2}$, which corresponds to a unique value of α .

While the P_{FP} is only dependent on α , and decreases with increasing α , the probability of making a FN error is also dependent on the signal level (as described previously). Thus, the P_{FN} curve can be plotted for fixed y_{low} , i.e. fixed SNR. In Fig. 4 seven such curves are plotted for the SNR values corresponding to the seven marked points on the $P_{\text{FP}} = P_{\text{FN}}$ curve in the bottom plot of Fig. 3. As expected P_{FN} increases with increasing α . The points where the P_{FP} curve intersects with the seven P_{FN} curves correspond to the points marked in the bottom plot of Fig. 3.

III. RESULTS

The test setup used is actually just an optical sensor system, so the experiment is designed to demonstrate sensor faults only. It functions by emitting infra red light and receiving reflections from moving objects in front of the system. The setup has been constructed in engineering and financial collaboration with LEGO Engineering, Denmark, and it is shown in Fig. 5. Due to space limitations, a thorough description of the test rig can not be given, but the main point is to give examples of signals which can be analyzed using the approach suggested above.

A. Detection of Faults

This test setup is used mainly for evaluating the detection methods presented in Section I. For that purpose three test signals have been recorded. The signals are all from the same receiver (in Fig. 5 it is the one on the bottom right). Faults have been introduced on the three emitters (from right to left in Fig. 5) (channel 0, 1, and 2, respectively). There are 16 samples

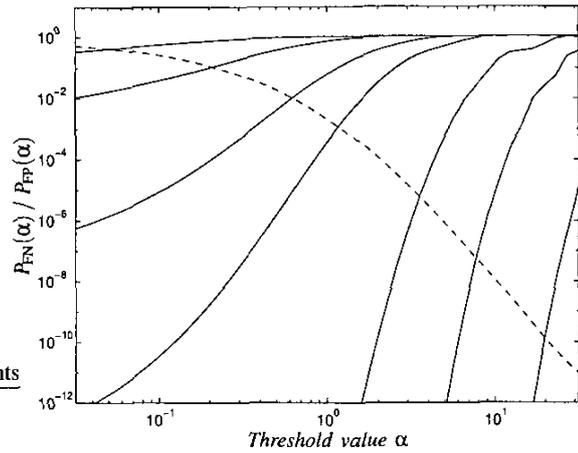


Fig. 4. The one P_{FP} curve (dashed) and the P_{FN} curves (solid) corresponding to the marked points in the top part of Fig. 3. The crossing points gives the α value which gives $P_{\text{FP}} = P_{\text{FN}}$ for the SNR used for drawing each of the seven P_{FN} curves. The slight irregularity of some of the curves are due to numerical instability.

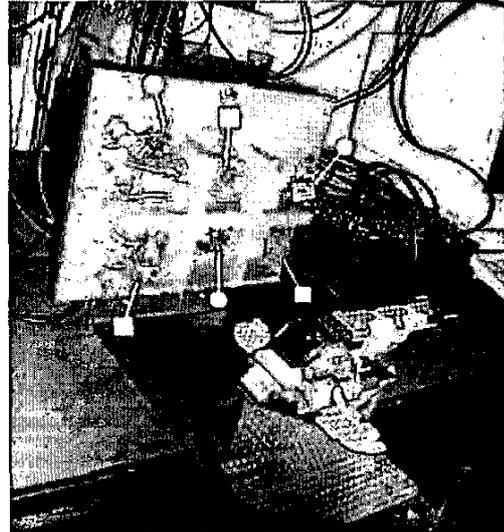


Fig. 5. The setup used for recording the test signals. The squares show the location of the receivers, and the circles show the locations of the emitters. The driver electronics is also located on each of the six small circuit boards. They are individually connected to the computer in the background.

in each block, so there are 3 potentially faulty channels and 13 channels for evaluating the noise at any time. The three signals have been generated in the following way:

Test signal 1: A fault slowly develops receivers, and then, at 3 seconds, disappears. Then it is quickly reintroduced and removed twice, and finally back in. Note, that the fault has different effects on the various emitters.

Test signal 2: The fault now appears at the beginning and disappears again. Then at 4 seconds it reappears. Meanwhile the receiver circuit has been subjected to an electrical disturbance (by quickly touching one of the pins on the photodiode with a screwdriver).

Test signal 3: Again a fault is introduced from the start, this time with larger amplitude than in the previous two signals. For 2.5 seconds the screwdriver has been

touching the photodiode pin.

Clearly, the faults influence more than one output. Thus, a method for joint detection could be used. However, we believe that the benefits of joint detection is small or even negligible in this case. Therefore, we consider only detection for the outputs individually, and since the methods is the same for every output only one, y_1 , is discussed in this section.

The three test signals are shown in Fig. 6. The 13 noise channels are not shown due to space limitations. In all three test signals the noise channels are contain random-noise which is very close to being normally distributed. In the second test signals the noise channels also contains a series of separate spikes, and in the third test signal the samples in the range from 400 to 800 are all spikes.

The detection method uses a threshold on the ratio between signal and noise to classify the SNR. When adapted to the test signals at hand the detection method becomes

$$T(\alpha): \Theta(y) \equiv \frac{y_1^2}{\sum_{k=3}^{15} y_k^2} > \alpha$$

The method has been applied to all three test signals.

B. Applying the Detection Method

The first thing to do is to determine the worst-case SNR, i.e. the weakest detectable signal compared to the expected random-noise level. The weakest signal is chosen to be $y_{low} = 15$. In the present setup this depends on the minimum level at which a fault should be detectable. The first half of the y_1 channel in the first test signal is generated by increasing the fault signal from a small level to the largest admissible and a little further, thus showing the weakest detectable signal to be approximately 15. Since the variance σ^2 is approximately 9.1, $R_{max} = y_{low}/\sigma = 4.8$. Using the curve the lower-most plot in Fig. 6, approximately $\alpha = 0.5$ for which $P_{FP} = P_{FN} \approx 0.02$.

The first test signal subjected to the detection method is shown in Fig. 6. The lowermost graph shows the detection function $\Theta(y)$ and the α is plotted as a dashed line in the same axis. The detection behaves as expected. For the 'no fault' part in the beginning, $T(\alpha)$ is false for almost all samples (approximately 1 out of 50 is expected to be accepted as a useful GM). When the signal level approaches y_{low} , which equals 15 and is shown with a dashed line, the detection shifts in favor increasingly more useful measurements. Again the fraction of signals classified incorrectly is 1/50 for signal level close to y_{low} . Note that the point in time (here measured in samples) at which more situations are considered faulty than healthy is the same as where the average level of the signal is $y_{low}/2$. This happens around sample number 250. This corresponds with the notion that if the signal itself was used for detection (as explained in Section I) and P_{FP} should equal P_{FN} for the weakest detectable signal the threshold should be half the weakest detectable signal level. This level is 48 in the example in Section I.

When the signal level then raises above y_{low} the P_{FN} decreases (but P_{FP} is still the same), and the last 200 samples of the signal shows that every single measurement is considered useful. Here P_{FN} equals approximately 10^{-4} (for signal level 27). Note also that the transient-like measurements are (correctly) considered useful by $T(\alpha)$.

Now, applying the detection to the two other test signal and using the same parameters the result is less gratifying. The second test signal seems to be as expected, at least for the non-transient samples. Zooming in on the transients (not shown)

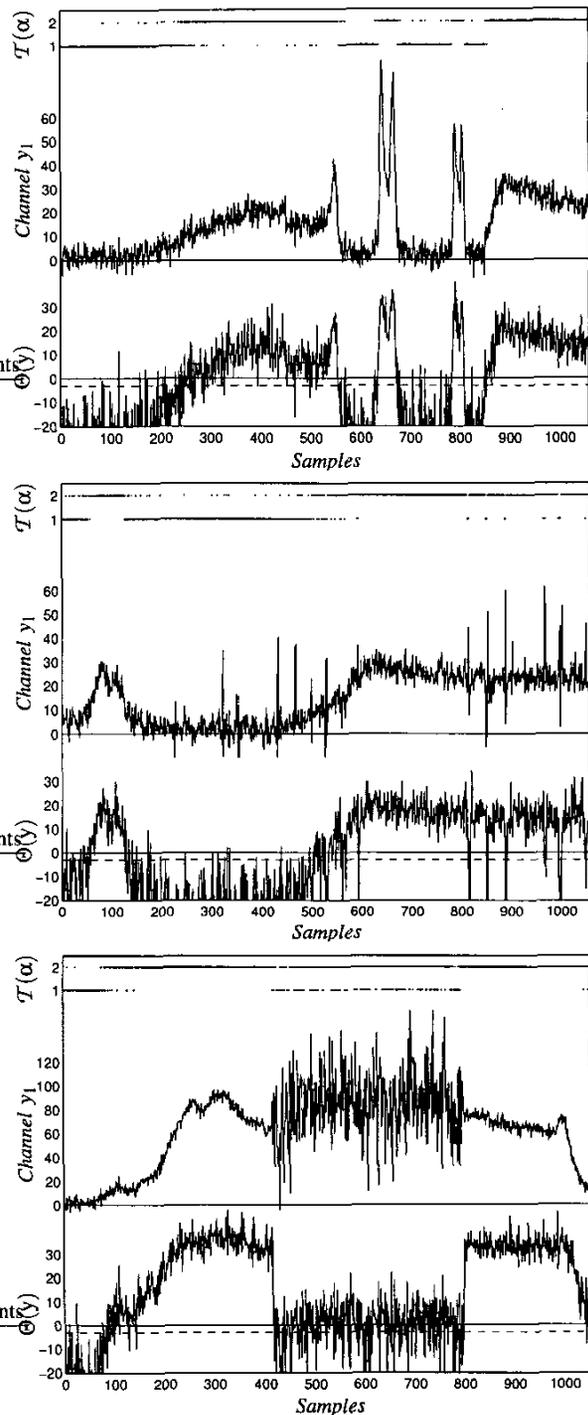


Fig. 6. The first detection method applied to the three test signals. For each plot: In the middle is a graph of the y_1 channel. The lowermost is a graph of the corresponding $\Theta(y)$ (in dB). The two lines on top show for each sample 1) when the test $T(\alpha)$ is false and 2) true. Here $\alpha = 0.5$ (which is -3.0 in the dB scale of the above graph and marked by the dashed line) and $P_{FP} = P_{FN} \approx 0.02$.

reveals that a majority, but not all, of the transients have indeed be classified as useless. The third test signal shows this only too clearly. When not just a few, but 400 consecutive samples are transient noise, the incorrect detection becomes evident as the majority of these samples are classified as useful.

It is worth noting that the detection does not fail because there is no significant difference between $\Theta(y)$ for the faulty and non-faulty parts of the third test signal. It fails because the threshold is wrong. However, the threshold has been determined such that it complies with our notion of proper behavior in a random noise scenario, and accordingly it works fine for the signal without transients. This problem can be solved easily by increasing the threshold to around 20, which seems to separate nicely the faulty and non-faulty parts in the third test signal. However, the result of simply increasing the threshold (which applies to all three test signals) is that virtually all the samples in the first test signal are classified as faulty. Evidently, another detection method is needed to handle this specific problem. In [12] a design method based on the test function:

$$\tilde{\Theta}(y) \equiv \frac{y_0^2}{\sum_{n=1}^N y_n^2 + \beta |y_0|^3}, \quad (7)$$

has been proposed, and a design procedure has been developed in analogy to the method proposed in this paper. In cases like the latter, the detection method based on (7) handles transients and spikes much better than the one presented in this paper at the cost of a slightly more involved design procedure. It has been left out of the present exposition, however, mainly due to space limitations.

IV. CONCLUSIONS

A general principle for designing a decision block for a diagnostic system has been presented. This principle was based on assessing the signal level of disturbances and noise in the system at all times by comparing the various outputs, residuals, or fault estimates. Subsequently, the decision on presence or non-presence of a fault is made by selecting a (dynamical) threshold value for each potential fault residual or estimate. The threshold was chosen in order to balance the risk for making a false alarm (false positive decision) to the risk of ignoring an actual fault (false negative decision). As these risks are always in conflict, obviously balancing them is equivalent to solving a minimax problem, i.e. to minimizing the larger of the two.

A method were suggested for designing such a threshold function. The purpose of this is to be able to distinguish between high signal amplitudes caused by the fault signal and by other events. The methods are based on SNR-like functions, and much of the effort was invested in determining the parameters in the methods such that the pure random noise scenario is handled properly and such that the false positive and false negative probabilities of validating measurements incorrectly are evenly balanced.

The method was demonstrated through three recorded test signals to handle the random noise properly. It failed in parts of the transient and for some of the spikes. These problems, however, can to a large extent be fixed by introducing one additional design parameter into the test function, as has been demonstrated in a recent thesis.

ACKNOWLEDGEMENTS

This work is supported by the Danish Technical Science Foundation (STVF) under Grant no. 9701481.

Thanks are due to LEGO Engineering, Denmark, for supporting the construction of the setup for recording the test signals.

REFERENCES

- [1] M. Basseville, "Detection changes in signals and systems - A survey," *Automatica*, vol. 24(3), pp. 309-326, 1988.
- [2] M. Basseville and I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*, Number ISBN 0-13-126780-9 in Information and System Science Series. Prentice Hall, April 1993.
- [3] J. Chen and R. Patton, *Robust model-based fault diagnosis for dynamic systems*, Kluwer Academic Publishers, 1998.
- [4] E.Y. Chow and A.S. Willsky, "Analytical redundancy and the design of robust failure detection systems," *IEEE Transactions on Automatic Control*, vol. 29, pp. 603-614, 1984.
- [5] J. Gertler, *Fault detection and diagnosis in engineering systems*, Marcel Dekker, 1998.
- [6] R. Isermann, "Supervision, fault-detection and fault-diagnosis methods - An introduction," *Control Engineering Practice*, vol. 5(5), pp. 639-652, 1997.
- [7] R.S. Mangoubi, *Robust Estimation and Failure Detection*, Springer-Verlag, 1998.
- [8] M. Massoumnia, "A geometric approach to the synthesis of failure detection filters," *IEEE Transactions on Automatic Control*, vol. 31(9), pp. 839-846, 1986.
- [9] A.S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pp. 601-611, 1976.
- [10] M. Nyberg, *Model based fault diagnosis methods, theory, and automotive engine applications*, Ph.D. thesis, Department of Electrical Engineering, Linköping, Sweden, 1999.
- [11] P.M. Frank, "Analytical and qualitative model-based fault diagnosis - A survey and some new results," *European Journal of Control*, vol. 2, pp. 6-28, 1996.
- [12] A. la Cour-Harbo, *Robust and Low-Cost Active Sensor by means of Signal Processing Algorithms*, Ph.D. thesis, Aalborg University, Dept. of Control Engineering, Fredrik Bajers Vej 7C, DK-9220 Aalborg East, Denmark, August 2002.
- [13] Y.E. Fatah, S. Thapliyal, and C. Kantor, "An LMI approach to the evaluation of alarm thresholds," *International Journal of Robust and Nonlinear Control*, vol. 8, pp. 659-667, 1998.
- [14] M.L. Rank and H.H. Niemann, "Norm based design of fault detectors," *International Journal of Control*, vol. 72(9), pp. 773-783, 1999.
- [15] A. Edelmayer and J. Bokor, "Optimal (∞ scaling for sensitivity optimization of detection filters," *International Journal of Robust and Nonlinear Control*, vol. 12(8), pp. 749-760, 2002.
- [16] H. Niemann, A. Saberi, A. Stoorvogel, and P. Sannuti, "Exact, almost and delayed fault detection - An observer based approach," *International Journal of Robust and Nonlinear Control*, vol. 9, pp. 215-238, 1999.
- [17] A. Saberi, A.A. Stoorvogel, P. Sannuti, and H.H. Niemann, "Fundamental problems in fault detection and identification," *International Journal of Robust and Nonlinear Control*, vol. 10(14), pp. 1209-1236, 2000.
- [18] A.A. Stoorvogel, H.H. Niemann, A. Saberi, and P. Sannuti, "Optimal fault signal estimation," *International Journal of Robust and Nonlinear Control*, vol. 12(8), pp. 697-727, 2002.
- [19] J. Stoustrup and H.H. Niemann, "Fault estimation - A standard problem approach," *International Journal of Robust and Nonlinear Control*, vol. 12(8), pp. 649-673, 2002.
- [20] A. Saberi, A. A. Stoorvogel, and P. Sannuti, "Exact, almost, and optimal input decoupled (delayed) observers," *International Journal of Control*, vol. 73, no. 7, pp. 552-582, 2000.
- [21] A. Emami-Naeini, M.M. Akhter, and S.M. Rock, "Effect of model uncertainty on failure detection: The threshold selector," *IEEE Transactions on Automatic Control*, vol. 33, pp. 1106-1115, 1988.
- [22] P.M. Frank and X. Ding, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *Journal of Process Control*, vol. 7(6), pp. 403-424, 1997.