Observers Data Only Fault Detection ⋆

B. Wahlberg * **A. C. Bittencourt** **

* *Automatic Control, KTH, SE 100 44 Stockholm, Sweden*
bo.wahlberg@ee.kth.se
** *Automatic Control, KTH, SE 100 44 Stockholm, Sweden*
andrecb@ee.kth.se

Abstract: Most fault detection algorithms are based on residuals, i.e.the difference between a measured signal and the corresponding model based prediction. However, in many more advanced sensors the raw measurements are internally processed before refined information is provided to the user. The contribution of this paper is to study the problem of fault detection when only the state estimate from an observer/Kalman filter is available and not the direct measured quantities. The idea is to look at an extended state space model where the true states and the observer states are combined. This extended model is then used to generate residuals viewing the observer outputs as measurements. Results for fault observability of such extended models are given. The approach is rather straightforward in case the internal structure of the observer is exactly known. For the Kalman filter this corresponds to knowing the observer gain. If this is not the case certain model approximations can be done to generate a simplified model to be used for standard fault detection. The corresponding methods are evaluated on a DC motor example. The next step is a real data robotics demonstrator.

1. INTRODUCTION

Sensors and observers/estimators are often closely integrated in intelligent sensor systems. This is common in distributed sensor processing applications. It may be very difficult or even impossible to access the raw sensor data since the sensor and state estimator/observer often are integrated and encapsulated. An important application of sensor based systems is model based fault detection, where the sensor information is used to detect abnormal behavior. The typical approach is to study the size of certain residuals, that should be small in case of no fault, and large in case of faults. Most of these methods rely on the direct sensor measurements. The problem when only state estimates are available is less studied. In Sundvall [2006] the problem of fault detection for such a system in mobile robotics is discussed from mainly an experimental point of view. The objective of this paper is to investigate the theoretical foundation of observer data only fault detection, where it is not possible to directly access the raw measured data.

Study the system state space description

$$x(t+1) = Ax(t) + B_u u(t) + B_v v(t) + B_f f(t).$$

Here $x(t)$ denotes the state vector, $u(t)$ is a known input signal, $v(t)$ is process disturbances and $f(t)$ is the unknown fault input. It is common to assume that $f(t)$ is either zero (no fault) or proportional to the the i:th unit vector $f(t) = f_i e_i$ in case of fault number i. Hence B_f is a matrix that determines how different faults affect the state. This covers, for example, faults in actuators.

⋆ This work was partially supported by the Swedish Research Council and the Linnaeus Center ACCESS at KTH.

The behavior of the system can be observed from different sensors j. To simplify the analysis we assume that sensors are integrated with standard observers/Kalman filters.

$$\begin{aligned} y_j(t) &= C_j x(t) + e_j(t) \\ \hat{x}_j(t+1) &= A\hat{x}_j(t) + B_u u(t) + K_j\left(y_j(t) - C_j\hat{x}_j(t)\right) \end{aligned}$$

The input to observer j is the measured output signal $y_j(t)$ and the input $u(t)$. The term $e_j(t)$ represents measurement noise. The output from the observer is $\hat{x}_j(t)$, i.e., an estimate of the state. If for example the Kalman filter is used this could come with a corresponding error covariance matrix

$$\mathrm{Cov}(\hat{x}_j(t) - x(t)) = P_j$$

Problem: We will study the problem when it is only possible to obtain $\hat{x}_j(t)$, and **not** the raw data $y_j(t)$. This seems to be a severe restriction, but from a practical point of view the measurement process could be integrated in the sensor system. One common example is standard GPS, where the measurement is based on satellite tracking and triangularization based techniques. In many applications the state estimate is obtained by more sophisticated methods then a simple linear observer. We will however use this structure for analysis and design purposes, so that the problem can be approached through well studied/standard FDI techniques.

Sofar we have not taken the fault contribution $f(t)$ into account. One possibility is to also estimate $f(t)$ by for example extending the state vector to $\bar{x}(t) = [x(t)\ f(t)]^T$ and apply the Kalman filter or another observer method to estimate the extended state vector $\bar{x}(t)$. Recently, there has been quite a lot of progress in the area of input estimation using Kalman filtering, see Gillijns et al. [2007].

2. RESIDUAL BASED FAULT DETECTION

There are in principle two paradigms for residual based fault detection. We will start with the standard problem formulation, with a direct measurable output.

$$
\begin{aligned}
x(t+1) &= Ax(t) + B_u u(t) + B_v v(t) + B_f f(t) \\
y(t) &= Cx(t) + e(t) + D_f f(t)
\end{aligned} \tag{1}
$$

Here we also have the possibility to model and detect sensor faults via D_f. The dimension of $x(t)$ is n_x and the dimension of $y(t)$ is n_y.

The so-called *parity space* approach, recently reviewed in Gustafsson [2007], is based on a sliding window formulation of the state space equations

$$Y(t) = \mathcal{O}x(t-L+1) + H_u U(t) + H_v V(t) + H_f F(t) + E(t)$$

where $Y(t) = \left[y^T(t-L+1) \dots y^T(t)\right]^T$ and similar for the other signal vectors. The matrices are given by

$$
\mathcal{O} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{pmatrix}, \quad H_f = \begin{pmatrix} D_f & 0 & \dots & 0 \\ CB_f & D_f & \dots & 0 \\ \vdots & & \ddots & \vdots \\ CA^{L-2}B_f & \dots & CB_f & D_f \end{pmatrix}
$$

and H_v and H_u are constructed in the same way as H_f, i.e. from the corresponding impulse response coefficients. The residual is then defined by

$$R(t) = W^T \left(Y(t) - H_u U(t)\right)$$

where W is an $L \times n_r$ projection matrix such that $W^T \mathcal{O} = 0$. The dimension $n_r = Ln_y - n_x$, where n_y is the dimension of the output vector $y(t)$ and n_x is the state dimension. This still leaves freedom in the choice of W, which can be used to obtain more structured residuals. In Gustafsson [2007] gives detailed insights of the design and analysis of parity space based methods in case of stochastic disturbances and noise.

An alternative approach is to use an *observer* or a Kalman filter to estimate $x(t)$ and then study the size of the residuals

$$R(t) = \left(Y(t) - \mathcal{O}\hat{x}(t-L+1) - H_u U(t)\right)$$

The choice $L = 1$ just gives the standard innovation process $r(t) = y(t) - C\hat{x}(t)$ used in the observer and in the Kalman filter to update the state estimate.

It is most important that the effects of the faults are visible in the residual vector. A fault is detectable if the transfer function from fault to residual is non-zero, this condition holds even for faults that disappear in the residual after some transient and a stronger condition is that this transfer function is non-zero also in steady-state. Another way to check if certain faults are detectable is to calculate if the extended state space model $\bar{x}(t) = [x(t) \ f(t)]^T$ is observable in classical state space sense. It is also closely related to input estimation, for which conditions are given in Gillijns et al. [2007].

3. RESIDUAL BASED FAULT DETECTION USING OBSERVER DATA ONLY

Residual based techniques are all based on comparing a predicted output $\hat{y}_j(t)$, based on a model, with the *observed* output $y_j(t)$ from a sensor. In case of a systematic difference we will alarm. If only the observer states $\hat{x}_j(t)$ are available the first two ideas for fault detection ideas would be:

- Try to reconstruct $y_j(t)$ using a model of the observer, e.g.
  $$K_j y_j(t) = \hat{x}_j(t+1) - (A - K_j C_j)\hat{x}_j(t) - B_u u(t)$$
  Here we need a very accurate model and the internal structure of the observer, e.g. the gain K_j, otherwise the estimations will be easily biased. In many practical cases this would be difficult. Notice also that if K_j is not full rank, it allows for multiple solutions.
- Assume that there are at least two observers providing $\hat{x}_1(t)$ and $\hat{x}_2(t)$. Define the residual vector
  $$\varepsilon(t) = \hat{x}_1(t) - \hat{x}_2(t)$$
  which should be sensible to fault that affects the two observers in different ways, e.g. sensor faults. This approach does not, however, make direct use of the model of the system and requires at least one redundant sensor.

We will start by analyzing the case with only one observer.

Idea: View $\hat{x}(t)$ as the output from the extended system

$$
\begin{aligned}
x(t+1) &= Ax(t) + B_u u(t) + B_f f(t) \\
\hat{x}(t+1) &= A\hat{x}(t) + B_u u(t) \\
&\quad + K\left(Cx(t) + D_f f(t) - C\hat{x}(t)\right) \\
\hat{y}(t) &= C^* \hat{x}(t)
\end{aligned} \tag{2}
$$

where C^* depicts which estimates are available. By using the extended state $\bar{x}(t) = [x^T(t) \ \hat{x}^T(t)]^T$ and the corresponding state space matrices we can interpret this a standard fault detection problem, which could be approached by a parity space method or a Kalman filter based method. The problem when we have m different observers can be approached by augmenting the state space with all observer states $\hat{x}_j(t)$, $j = 1, \dots m$.

There are some basic questions that need to be addressed

- Are the faults detectable using this model?
- What to do if the observer gain K is unknown?
- How to compare/validate the performance of different methods?

3.1 Fault Observability

As described in Kalata et al. [1995], stochastic biases in linear time invariant systems can be identified by augmenting the system state with a bias and implement a Kalman filter. The author utilizes this technique to identify biases in noisy measurements. In Chapter 3 in Tornqvist [2006] these results are extended to check the observability of additive faults with the constrain that $f(t+1) = f(t)$. An important characteristic explored by both authors is the observability issue.

Considering a system as in Equation (2), where only estimated states $\hat{x}(t)$ are available but not $x(t)$, we can augment the faults in the states as $\bar{x}(t) = [x(t) \quad \hat{x}(t) \quad f(t)]$ and analyze the observability with the pair

$$\bar{C} = [\,0 \;\; C^{*} \;\; 0\,], \quad \bar{A} = \begin{bmatrix} A & 0 & B_f \\ KC & (A-KC) & KD_f \\ 0 & 0 & I \end{bmatrix} \tag{3}$$

Observability conditions: With a similar approach as in Tornqvist [2006] Appendix A shows that if the original system is observable (pair (A, C) observable) and if

- All estimates are directly available at the output (i.e. $C^{*} = I$).
- K is full column rank, such that
$$Kz = 0 \quad \Rightarrow \quad z = 0$$

then the extended system will be observable under the same conditions for sensor and process faults as when the actual output $y(t)$ is available. In other words, we will have the same information as when $y(t)$ is directly accessible. The conditions can be summarized as:

- For measurement faults only, if the system has no integrator dynamics (modes with eigenvalue equals to 1), the faults will be observable as long as D_f is full rank. If there is an integrator, then the faults are not observable if D_f is full rank and the states through which the fault propagates should be orthogonal to the measured integrating part of the system.
- For process faults only, the faults should be orthogonal to the contribution of the non-measured part of the system.

3.2 Unknown Observer Structure

If the observer structure is not given, for example if K is unknown, we cannot directly use the extended state space model for fault detection. To overcome this problem, let us make two approximations. The first one is to approximate

$$K\left(Cx_f(t) + D_f f(t)\right) \approx \bar{F}_f \bar{f}(t)$$

Here

$$x_f(t+1) = Ax_f(t) + B_f f(t) \tag{4}$$

and $x_f(t)$ is the fault contribution in $x(t)$. Since $f(t)$ is zero or a constant vector it is most important that this approximation holds stationary i.e. after transients.

The second approximation is

$$K\left(y(t) - C\hat{x}(t)\right) \approx \bar{v}(t)$$

where $\bar{v}(t)$ is a white noise process with a certain covariance matrix. This makes sense from a Kalman filter point of view where the innovations can be viewed as process noise, $\bar{v}(t) = K\epsilon(t)$ where $\epsilon(t) = y(t) - C\hat{x}(t)$ is a white innovation process.

This leads to the simplified model

$$\begin{aligned} \hat{x}(t+1) &= A\hat{x}(t) + B_u u(t) + \bar{F}_f \bar{f}(t) + \bar{v}(t) \\ \bar{y}(t) &= \hat{x}(t) + \bar{e}(t) \end{aligned} \tag{5}$$

With such structure, actuator and sensor faults are mixed and the fault isolation step could be more difficult in this setting.

The artificial measurement noise $\bar{e}(t)$ can be used to cope with unmodeled characteristics of the system. For example, for sensors over a network or with a weak real-time performance, one can use $\bar{e}(t)$ to include jitter, missed samples, delays, etc. or to cope with sensor/system unknown dynamics.

After defining $\bar{e}(t)$, it can be used to tune a Kalman filter observer for the system as in Equation (5) and a standard parity space method or Kalman filter based method can be used to design a fault detection algorithm. It is easy to extend this approach to several observers by combining the observer states as

$$\begin{bmatrix} \hat{x}_1(t+1) \\ \hat{x}_2(t+1) \\ \vdots \\ \hat{x}_j(t+1) \end{bmatrix} = \mathbf{A} \begin{bmatrix} \hat{x}_1(t) \\ \hat{x}_2(t) \\ \vdots \\ \hat{x}_j(t) \end{bmatrix} + \mathbf{B_u} u(t) + \bar{\mathbf{F}}_{\mathbf{f}} \bar{f}(t) + \bar{v}(t)$$

where $\mathbf{A}$, $\bar{\mathbf{F}}_{\mathbf{f}}$ and $\mathbf{B_u}$ are diagonal matrices relating observers and fault states and actuator dynamics, and $u(t)$,$\bar{f}(t)$ and $v(t)$ are the extended control, fault and noise inputs.

3.3 Performance evaluation methods

We are interested in analyzing the quality of the residuals generated by the different methods rather than for a full detection scheme. Different factors influence a residual success in terms of detection, including noise/model disturbances decoupling and sensitivity to faults. These qualities are often in contradiction and can be seem as an optimization problem. Some recent results for optimal residual generation can be found in Liu [2008], where the author shows closed-form solutions for some sets of the problem, the solutions however, require that the faults are directly visible at the output (full column rank D_f), which will never be the case for sensors integrated with observer/Kalman filters since the faults travel through the observer dynamics before they appear in the output.

For the analysis of our proposed methods, we will use a definition of a good residual as one that resembles to white gaussian noise with no peaks or abrupt changes in a fault-free case (possibly decreasing the false-alarm rates) and that bias under a fault as great as possible (possibly increasing the detection rate). We depict two empirical quantities to relate these qualities.

The kurtosis statistic is a measure of the peakedness of a signal, we can use it to analyze the resemblance of the residuals *over a fault-free scenario* (NF) to a gaussian distribution. It is defined as

$$\kappa = \frac{E\left[\varepsilon(t)|_{NF} - \mu_{\varepsilon(t)_{NF}}\right]^4}{\sigma^4_{\varepsilon(t)_{NF}}} - 3$$

where $\mu_{\varepsilon(t)_{NF}}$ and $\sigma_{\varepsilon(t)_{NF}}$ are the mean and standard deviation of the residual under no-fault while $E[z]$ is the expected value of z. For a gaussian residual the test approximates to zero, and it is usually used to detect abrupt variations over a gaussian signal. In Hadjileontiadis et al. [2005], for instance, it is used for crack detection over a beam with vibration analysis. Basically, a κ close to 0 will relate to a white gaussian distribution while higher κ means more of the variance is due to big sporadic deviations or biases.

The fault-to-noise ratio as defined in Gustafsson [2001] is a measure of a fault sensitivity relative to noise and is defined as a ratio between the expected value of a fault influence in the system output and the noise variance, it is, in fact, a similar concept as the signal-to-noise-ratio (SNR) but applied to a fault. For a known FNR, we can define the measure

$$\delta = \left\| \frac{[\varepsilon(t)|_F]/\sigma_{\varepsilon(t)_F}}{FNR} \right\|$$

where $E[\varepsilon(t)_F]$ and $\sigma_{\varepsilon(t)_F}$ are the expected value and standard deviation of the residual *under a fault hypothesis* (F). In this manner, δ will vary from 0 to 1 (best possible residual, which will have the same quality as the direct fault influence to the output, only possible if we use a perfect simulator of the system to generate the redundancy).

4. ILLUSTRATIVE EXAMPLE

In order to explore the different configurations presented, we consider a simple linear DC motor, as shown in Figure 1. Non-linearities such as flexibilities and friction are simplified. The applied voltage in the motor terminals is the controlled input to the system V_{app} while angular speed is taken as output. The states are current and angular

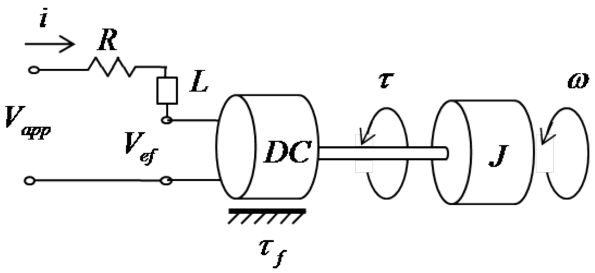

Fig. 1. DC motor model.

speed $x(t) = [i(t) \quad \omega(t)]^T$ and the governing matrices

$$A = \begin{bmatrix} \dfrac{-R}{L} & -\dfrac{k_b}{L} \\ \dfrac{k_m}{J} & \dfrac{-k_f}{J} \end{bmatrix}, \quad B = \begin{bmatrix} \dfrac{1}{L} & 0 \end{bmatrix}^T \tag{6}$$
$$C = [\,0 \ 1\,], \qquad D = [\,0\,]$$

where k_m, k_b and k_f are armature, emf and friction constants. Process noise with variance Q, is considered to affect the system as random oscillations in the current $i(t)$. While measurement noise with variance R, appears in the angular speed measurements. So that $B_w = [\,1\ 0\,]^T$ and $D_v = [1]$.

Integrated sensor model. We are interested in studying the residuals when only the state estimate from an observer/Kalman filter is available and not the direct measured quantity. For this purpose we use a sensor model integrated with an observer as in Equation (2). Taking $C^* = C$ and the observer gain K as the stationary Kalman filter gain for the given process and noise covariances, Q and R. During our example, this model should be referred to whenever we use the notation $\hat{y}(t)$.

Faults. Two step faults with $FNR = 10$ (that produces a 10 times greater amplitude in the sensor output $y(t)$ than the present noise) are considered: process faults, $f_p(t)$, appearing as a step torque opposing to the system; sensor faults, $f_s(t)$, an offset measurement deviation. So that $B_f = \begin{bmatrix} 0 & -\dfrac{1}{J} \end{bmatrix}^T$ and $D_f = [1]$.

In the following Sections we discuss the performance of different methods to generate the residual considering the output of such integrated sensors.

4.1 Augmented system observer

When K is known, a possible approach to generate a residual is to augment system and sensor states $\bar{x}(t) = [x(t) \quad \hat{x}(t)]$ and use the augmented model

$$\bar{A} = \begin{bmatrix} A & 0 \\ KC & (A - KC) \end{bmatrix}, \ \bar{B}_u = \begin{bmatrix} B_u \\ B_u \end{bmatrix}, \ \bar{C} = \begin{bmatrix} 0 \\ C \end{bmatrix}^T$$

to design an observer/Kalman filter. With the redundant output $\bar{y}(t)$ from this model we can set a residual as $\bar{\varepsilon}(t) = \hat{y}(t) - \bar{y}(t)$ to detect faults. Considering our DC motor example, when we configure our observer as a Kalman filter with gain $\bar{L}$, we have for process and sensor faults a response as shown in Figure 2 from which is easy to

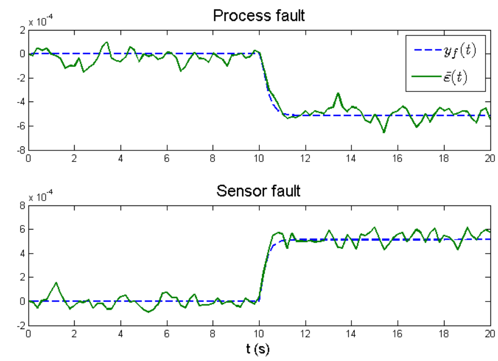

Fig. 2. Residual $\bar{y}(t)$ for process and sensor fault (solid line) together with direct fault contribution to the sensor output $y_f(t)$.

depict that the residual performance is good. In fact, for process faults we had $(\delta,\ \kappa) = (0.88,\ \approx 0)$ and sensor faults $(0.85,\ \approx 0)$.

Is is important to analyze how robust this method is to model errors introduced through errors in the sensor gain K. We analyze the sensitivity of the residual to K by varying it with a scaling factor as $K \times \alpha$ while $\bar{L}$ remains constant. The pair $(\delta,\ \kappa)$ is computed for α varying within $[10^{-1} - -10^5]$. The result is a pair $(\delta,\ \kappa) = (0.85,\ 0.30)$ in the worst case showing the robustness of the approach in this example.

4.2 Simplified system observer

As discussed in Section 3.2, we can simplify the sensor dynamics yielding a model as in Equation (5), with faults and internal observer dynamics appearing in the terms $\bar{F}_f \bar{f}(t)$ and $\bar{v}(t)$. With this simplification, we can design an observer to generate a redundancy $\breve{y}(t)$ and a residual $\breve{\varepsilon}(t) = \hat{y}(t) - \breve{y}(t)$ sensible to faults.

In our example, we set the observer gain equal to the sensor gain, $\breve{L} = K$. In this setting, the observer has similar dynamics to the sensor, in fact, if we could use the direct measured output $y(t)$ as input to our observer, sensor and observer would be equivalent. For a process fault, we have $(\delta,\ \kappa) = (0.87,\ \approx 0)$ and sensor fault $(0.84,\ \approx 0)$ which is a slightly worsened result when compared to the results in the earlier Section, with the augmented system observer.

To analyze the robustness on the gain selection, $\breve{L}$, we again, vary it with a scaling factor as $\breve{L} = K \times \alpha$ and plot the pair $(\delta,\ \kappa)$. The result, shown in Figure 3, indicates

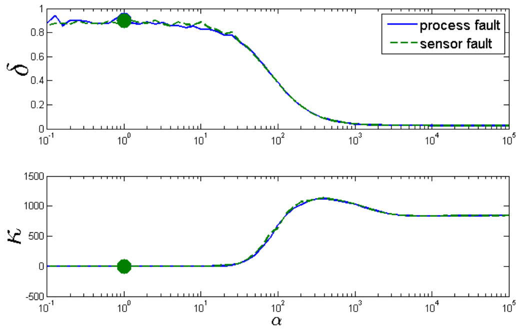

Fig. 3. δ and κ versus scaling factor α for process and sensor faults. The circle depicts the case when $\breve{L} = K$.

that the residual is worsened with the overestimation of K. This result is to be expected since, the larger the gain, the more relevance is given to the measurements and therefore, the residual will be less sensitive to faults (δ decrease). As well, it also increases the observer speed, with the observer trying to reach the signal faster and consequently increasing transient errors (κ increase).

4.3 Multiple sensors

A common approach to fault detection is to take a residual as the direct difference between two redundant sensors $\varepsilon(t) = y_i(t) - y_j(t)$. We study this case for our example comparing its performance with model-based generated residuals.

So far, our sensor estimates the states through angular speed measurements, $\hat{y}_\omega(t)$. To provide a redundancy, we depict a sensor that estimates the states through position measurements $\theta(t)$, $\hat{y}_\theta(t)$. Notice that the subscript in $\hat{y}_\theta(t)$ and $\hat{y}_\omega(t)$ denotes directly what is the measured quantity. For such redundant sensor, we have the states $[i(t) \quad \omega(t) \quad \theta(t)]$ and model,

$$A = \begin{bmatrix} \frac{-R}{L} & -\frac{k_b}{L} & 0 \\ \frac{k_m}{J} & \frac{-k_f}{J} & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{L} \\ 0 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}^T \tag{7}$$

we suppose this sensor is also integrated with a Kalman filter and outputs an estimate of $\omega(t)$. The sensor noise variance is set to produce the same error order in the output as for the first sensor so that both have similar qualities.

We would like to compare the classical approach, $\varepsilon_0(t) = \hat{y}_\omega(t) - \hat{y}_\theta(t)$, with model based generated residuals. We consider one augmented states observer for each available sensor, $\bar{y}_\omega(t)$ and $\bar{y}_\theta(t)$ with the residuals as

$$\begin{aligned} \bar{\varepsilon}_\omega(t) &= \hat{y}_\omega(t) - \bar{y}_\omega(t) \\ \bar{\varepsilon}_\theta(t) &= \hat{y}_\theta(t) - \bar{y}_\theta(t) \\ \bar{\varepsilon}_{\omega,\theta}(t) &= \bar{y}_\omega(t) - \bar{y}_\theta(t) \end{aligned} \tag{8}$$

Because the sensors are different, the same fault may cause different influences in the output for each sensor, but since we are interested in showing the relative performance between model-based generated residuals and $\varepsilon_0(t)$, we compute a relative measure as

$$\Delta = \left\| \frac{E[\varepsilon_i(t)|_F]/\sigma_{\varepsilon_i(t)}}{E[\varepsilon_0(t)|_F]/\sigma_{\varepsilon_0(t)}} \right\|$$

where $\varepsilon_i(t)$ is one of the residuals from Equation 8. $\Delta > 1$ will depict a residual with a larger fault sensitivity than the one generated by $\varepsilon_0(t)$. Table 1 shows Δ for process and sensor faults. The results show that taking

Residual	Process fault	Sensor $\hat{y}_\omega(t)$ fault
$\bar{\varepsilon}_\omega(t)$	0.1498	1.6059
$\bar{\varepsilon}_\theta(t)$	1.3890	≈ 0.00
$\bar{\varepsilon}_{\omega,\theta}(t)$	3.0104	0.02

Table 1. Δ for different faults. In all cases $\kappa \approx 0$.

the residual as the direct difference between sensors, $\varepsilon_0(t)$, may not provide the best residual. It is expected this would be even more significant in case the noise variances differed considerably for each sensor, since the observers also attenuate noise.

4.4 Summary

Different aspects have been analyzed trough our illustrative examples, some important remarks:

- Though this was not fully explored in the example, different tuning configurations have been used and it was noticed that the use of an augmented states observer is likely to improve the fault sensitivity when compared to the observer using a simplified sensor model as presented in Section 4.2.
- The analysis on the observer gain choice in Section 4.2 indicates that the fault sensitivity is improved as smaller we choose the gain. Such result is motivated by the fact that lower gains will thrust more on the model and therefore, the resulting residual will be more sensitive to unmodeled influences, such as faults. Nevertheless, the choice of the gain should actually be seem as a compromise between model uncertainties and the fault sensitivity.
- Finally, the example with multiple sensors depicted that using a model-based residual can improve the fault sensitivity.

5. CONCLUSIONS

The paper analyzed several structures for observers data only fault detection. Section 2 discussed standard approaches for fault detection; Section 3 presented ideas and addressed some basic questions for the problem, including a discussion over fault observability, knowledge on observer structure and residual performance measures. Finally, Section 4 illustrated the problem through a simulated example, covering the approaches for fault detection using redundant sensors and Kalman filter based methods with known and unknown sensor structure. Most of the methods have shown to be useful, with slight improvements when one consider both system and sensor states in the estimation. There are yet some open problems such as more general observability conditions for faults, methods to support the choice of the observer gains, analysis under model uncertainties, etc. which shall be presented in future work, together with example from a real robotics application.

REFERENCES

P. Sundvall. Mobile Robot Fault Detection using Multiple Localization Modules. *Kungliga Tekniska högskolan*, TRITA-EE 2006:044.

F. Gustafsson. Statistical signal processing approaches to fault detection. *Annual Reviews in Control*, 31:41–54, 2007.

F. Gustafsson. Adaptive filtering and change detection. Wiley, 2001.

N. Liu and K. Zhou. Optimal Robust Fault Detection for Linear Discrete Time Systems. *Journal of Control Science and Engineering*, Volume 2008, Issue 1, 2008.

D. Törquvist. Statistical Fault Detection with Applications to IMU Disturbances. *Linköping Studies in Science and Technology*, Thesis No. 1258, 2006.

G. F. Franklin, J. D. Powell and A. Emami-Naeini. Feedback Control of Dynamic Systems. Prentice Hall, 5th edition, 2006.

P. R. Kalata and T. A. Chmielewski Jr. On the Identification of Stochastic Biases in Linear Time Invariant Systems. *In Proceedings of the American Control Conference*, pp 4067-4071, Seatle, Washington, 1995.

L. J. Hadjileontiadis, E. Douka, A. Trochidis. Crack detection in beams using kurtosis. *Computers and Structures*, 83:-919, (2005).

S. Gillijns, B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*, 43:-116, (2007).

Appendix A. FAULT OBSERVABILITY

The Popov- Belevitch-Hautus test on the augmented extended systems depicts a pair (A,C) to be observable if

$$\begin{pmatrix} C \\ A - sI \end{pmatrix}$$

has full column rank for all s (see Franklin et al. [2006] for more). The faults modes are $s = 1$, therefore the analysis is separated for $s \neq 1$ and $s = 1$.

Given that our original system is observable (pair (A, C) observable), we analyze the observability for the augmented system as in Equation (3). First, when $s \neq 1$ the observability is given by:

$$\begin{pmatrix} 0 & C^* & 0 \\ (A - sI) & 0 & B_f \\ KC & ((A - KC) - sI) & KD_f \\ 0 & 0 & (1 - s)I \end{pmatrix}$$

The last two columns are full column rank for $s \neq 1$ and $C^* = I$, and the analysis can be simplified to

$$\begin{pmatrix} (A - sI) \\ KC \end{pmatrix} \tag{A.1}$$

Given that for a pair (A, C) to be observable, they should have non-intersecting null-spaces ($\mathcal{N}_{A-sI} \cap \mathcal{N}_C = 0$), the condition above is translated to $\mathcal{N}_{A-sI} \cap \mathcal{N}_{KC} = 0$. Using Theorem 1, it is rewritten as $\mathcal{N}_{A-sI} \cap \mathcal{N}_C = 0$ which is actually the observability condition on the original system, which holds and the condition for $s \neq 1$ is checked.

Theorem 1. Given K is full column rank, for any B we have:

$$\mathcal{N}_{KB} = \mathcal{N}_B$$

Proof: Suppose $\mathcal{N}_K = \emptyset$. Such that

$$Kv = 0 \quad \rightarrow v = 0$$

Now, let us check the null space of KB

$$(KB)w = 0, \quad K(Bw) = 0 \quad \rightarrow Bw = r \in \mathcal{N}_K \quad \rightarrow Bw = 0$$

finally, the solution for $Bw = 0$ is the null-space of B and therefore,

$$\mathcal{N}_{KB} = \mathcal{N}_B$$

When $s = 1$ and $C^* = I$, it is equivalent to analyze

$$\begin{pmatrix} (A - I) & B_f \\ KC & KD_f \end{pmatrix}$$

For *measurement faults* ($B_f = 0$) and considering that full column rank of a matrix A is equivalent to

$$Av = 0 \quad \Leftrightarrow v = 0$$

we have

$$\begin{pmatrix} (A - I) & 0 \\ KC & KD_f \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and $x \in \mathcal{N}_{A-I}$, $KCx + KD_f y = 0$. Notice that if A has no integrators, then the condition is achieved if D_f is rank. If A has integrators, then D_f should not be full rank and we analyze that the first condition is only true when x is zero or an eigenvector of A with eigenvalue equals to 1. Hence, it is sufficient to analyze such x.

Take U as a basis formed by the eigenvectors of A with eigenvalues 1 and rewrite the last condition as $K(CUr + D_f y) = 0$ where r is any vector. For full column rank K, this condition can only be true if CU and D_f share image spaces and the condition on the observability can be rewritten as

$$\mathcal{R}_{CU} \cap \mathcal{R}_{D_f} = \emptyset$$

Which means that *the faults should be orthogonal to the measured integrating part of the system.*

For *process faults* ($D_f = 0$) only we can analyze

$$\begin{pmatrix} (A - I) & B_f \\ KC & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and we have $x \in \mathcal{N}_C$ (Theorem 1) and $(A-I)x + B_f y = 0$. Taking W as a basis for the null-space of C the final condition is $\mathcal{R}_{(A-I)W} \cap \mathcal{R}_{B_f} = \emptyset$.

Which means that the *faults must be orthogonal to the contribution of the non-measured part of the system.*