EVALUATING SOME GAIN SCHEDULING STRATEGIES IN DIAGNOSIS OF A TANK SYSTEM

Marcus Klein and Lars Nielsen

Department of Electrical Engineering
Linköping University
Sweden
email:klein@isy.liu.se, lars@isy.liu.se

Abstract: In model-based diagnosis the problem of finding all the relations that can be used to detect and isolate different faults, is solved for linear systems. However, for nonlinear systems the situation is more complicated. Here an approach will be taken using a linear method together with gain scheduling. Linear residual generators are designed at a number of stationary points. The approach is based on using a nominal selector matrix, using null-space redesign dependent on the scheduling variable, and using a proposed optimization method. Two different gain scheduling strategies are applied to form the residual generators between design points, namely nearest neighbor approximation and linear interpolation. The approach is applied to a simple nonlinear system consisting of two coupled water tanks. The simulations show that the performance of the residual generators are good under steady state conditions. It is also shown that linear interpolation has better performance than nearest neighbor approximation.

Keywords: Nonlinear diagnosis, Minimal polynomial basis, Water-tank system, Residual generation, Null-space redesign.

1. INTRODUCTION

In safety critical and environmentally critical products and processes, it has become more and more important to detect a fault and to isolate a faulty component before it leads into severe failure. Obvious examples are aircraft engines or chemical industries. One important class of subsystems in these artifacts are fluid systems, that e.g. could be air dynamics in the air-intake system of a jet engine, or liquid dynamics in a chemical process. These systems are typically non-linear, and this paper is an attempt to handle the type of problems that arise in this important class of systems. Our approach is to use model-based diagnosis.

In model-based diagnosis a model of the system is used to find the relations that can be used to detect and isolate different faults. Methods for finding these relations, called residuals, are used. For linear systems all these relations can be found, but not yet for nonlinear systems. In many control applications linear control methods together with *gain scheduling* has been successfully used for nonlinear systems. This is the approach taken here to design a diagnosis system.

2. GAIN SCHEDULING STRATEGIES

The gain scheduling approach taken here relies on several local linear approximations of the nonlinear system and design of linear residual generators for the linearization points locally. Then these residual generators are combined using a scheduling strategy into a nonlinear residual generator.

2.1 *Model*

A general state space description of a nonlinear system can be parameterized as a Linear Parameter Varying (LPV) system (Lagerberg 1996),

$$\dot{x}(t) = A(\alpha)x(t) + B(\alpha)u(t) \tag{1a}$$

$$y(t) = C(\alpha)x(t) + D(\alpha)u(t) \tag{1b}$$

$$\alpha = \alpha(x(t), u(t)) \tag{1c}$$

where we for each value of α have a linear system. A parameter that indicate the operating conditions of the system is denoted *scheduling variable*.

In gain scheduling the linear matrices in (1) are calculated for some specified values of the scheduling variable α, denoted design points. Let the design points be indexed by j, and denote the design points α^j. The linear matrices are then used used together with the scheduling variable α and the chosen gain scheduling strategy to approximate the nonlinear model at every operating point.

2.2 *Residuals*

The residuals are used to detect misbehavior of the system, and are generated by a residual generator, from the input u and the output y of the system. A residual evaluator is then used to determine the present fault mode, i.e. to give a diagnosis.
A general linear residual generator can be written as:

$$r = \frac{A(p)y(t) + B(p)u(t)}{C(p)} = Q(p)\begin{pmatrix} y \\ u \end{pmatrix} \tag{2a}$$

$$A(p) = a_n p^n + \cdots + a_1 p + a_0 = \sum_{i=0}^{n} a_i p^i \tag{2b}$$

where A, B and C are polynomials of appropriate dimensions, Q is a rational function and p is the derivative operator. The output y of the system is a function of the input u, the faults f that can occur and the disturbances d, which can be described as $y = G_{uy}(p)u + G_{fy}(p)f + G_{dy}(p)d$. The residual generator then becomes:

$$r = Q\begin{bmatrix} G_{uy} & G_{dy} \\ I & 0 \end{bmatrix}\begin{pmatrix} u \\ d \end{pmatrix} + Q\begin{bmatrix} G_{fy} \\ 0 \end{bmatrix} f \tag{3}$$

A residual should in the fault free case be zero, i.e. Q should be chosen such that

$$0 \equiv Q\begin{bmatrix} G_{uy} & G_{dy} \\ I & 0 \end{bmatrix} =: QM \tag{4}$$

To be able to detect a fault, the transfer function from fault to residual must be non-zero (Nyberg 1999). A feasible method of finding all appropriate Q:s parameterized in a minimal way, is *the Minimal Polynomial Basis Approach* as described in (Nyberg & Frisk 1999). The method can be summarized in two steps:

(1) Find a minimal polynomial basis $N_M(s)$ for the left null space of $M(s)$.
(2) Choose the polynomial matrices $\phi(s)$ and $D_W(s)$ in

$$Q(s) = D_W^{-1}(s)\phi(s)N_M(s) = D_W^{-1}(s)W(s) \tag{5}$$

where $\phi(s)$ is a polynomial matrix used to pick out the appropriate residual generator from the basis $N_M(s)$. To make $Q(s)$ realizable, $D_W(s)$ is an invertible polynomial matrix with row degree higher or equal to $\phi(s)N_M(s)$:s corresponding row degree (Kailath 1980). The poles of $Q(s)$ can be chosen arbitrarily.

2.3 *Interpolation*

The linear residual generators designed at specific operating points work well locally, but not necessarily otherwise. Interpolation of residual generators designed off-line will be the approach taken here. Design linear residual generators in a number of design points α^j. One disadvantage with this approach is that the disturbances and the non-monitored faults might not be decoupled for all operating points, since (4) might hold only for the design points j. Two interpolation methods will be tried here, nearest neighbor and linear interpolation. With the two-step method above, all Q are found and the resulting residual generator r^j in (2a) for every design point α^j can then be written as:

$$r^j = \frac{A^j(p)y(t) + B^j(p)u(t)}{C^j(p)} \tag{6a}$$

$$A^j(p) = a_n^j p^n + \cdots + a_1^j p + a_0^j = \sum_{i=0}^{n} a_i^j p^i \tag{6b}$$

with B^j and C^j defined in a similar way. Here we will let the residuals r^j have the same states when the interpolation is made, by using the same C^j for all design points α^j. This makes it equivalent to interpolate the residuals r^j as to interpolate the coefficients a_i^j and b_i^j. Interpolation between a_i^j and a_i^{j+1} then yields a_i in (2b), determined by the scheduling variable α and the chosen interpolation method.

Gain Switching One gain scheduling approach is to *switch* between the designed residual generators when moving around in the operation range, i.e. to use nearest neighbor approximation. The parameters a_i are calculated as:

$$a_i = \beta^{j+1} a_i^{j+1} + \beta^j a_i^j \tag{7}$$

where

$$\beta^{j+1} = \begin{cases} 1 & \beta \geq 0.5 \\ 0 & \beta < 0.5 \end{cases} \qquad \beta^j = 1 - \beta^{j+1} \tag{8a}$$

$$\beta = \frac{\alpha - \alpha^j}{\alpha^{j+1} - \alpha^j} \qquad \forall \alpha \in [\alpha^j, \alpha^{j+1}[\tag{8b}$$

At a switch point (a point where $\beta = 0.5$) a jump (discontinuity) in the residual can occur, due to switching of parameters in the residual generator. This can be avoided if we find a bump-less transfer function. Another problem, called chatter, can occur when the present operating point moves back and forth over the switch point, causing the diagnosis system to switch back and forth. This can be avoided using hysteresis (Lagerberg 1996).

Continuous Gain Scheduling To avoid the problem with discontinuities in the residual at switch points, continuous gain scheduling can be used. The residuals then becomes continuous over the entire operating range. Linear interpolation is an example of a continuous interpolation, which yields a_i as:

$$a_i = \beta a_i^{j+1} + (1 - \beta) a_i^j \tag{9}$$

where β is determined by (8b).

3. THE WATER TANK SYSTEM

The different gain scheduling strategies are to be implemented and simulated on a water tank system.

Figure 1 shows a schematic picture of the process. The water tank system consists of two coupled tanks, where the water level of the upper tank is controlled by the pump. Each tank has a sensor measuring the water level and a flow sensor measuring the flow out of the tank. Sensor and actuator faults can also be introduced.

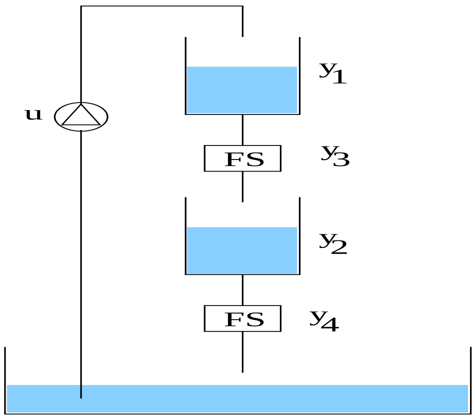

Fig. 1. A schematic figure of the water tank system, where FS are flow sensors. Each tank has a water level sensor and the water level in the upper tank is controlled by the pump.

Modeling Mass balance for one tank gives

$$A\frac{dh}{dt} = Q_{in} - Q_{out} \tag{10}$$

where A is the cross section, h is the water level and $Q_{in/out}$ is the flow in/out of the tank. The flow out of the tank can be described by Bernoulli's law $Q_{out} = a\sqrt{2gh}$, where a is the effective cross section of the outlet hole and g is the acceleration of gravity. The input to the process is v (input voltage to the pump [0-10 V]) and the outputs are $y_1 \dots y_4$ (voltage from sensors[0-10 V]). Assume that the flow generated by the pump is proportional to the applied voltage, i.e. $Q_{in} = k_a v$. Introducing the constants $c_i = \frac{a_i}{A_i}\sqrt{2g}$ and $b_i = \frac{1}{A_i}$ and modeling the sensor and actuator faults in Table 1 as additive yields the model:

$$\frac{dh_1}{dt} = k_a b_1(v + f_a) - c_1\sqrt{h_1} \tag{11a}$$

$$\frac{dh_2}{dt} = \frac{b_2}{b_1}c_1\sqrt{h_1} - c_2\sqrt{h_2} \tag{11b}$$

$$y = \begin{pmatrix} h_1 + f_{h1} \\ h_2 + f_{h2} \\ \frac{c_1}{b_1}\sqrt{h_1} + f_{f1} \\ \frac{c_2}{b_2}\sqrt{h_2} + f_{f2} \end{pmatrix} \tag{11c}$$

where subscript i corresponds to tank i. The stationary points (h_1^0, h_2^0, v^0) for the water tank system are given by $h_1^0 = \frac{k_a^2}{2ga_1^2}(v^0)^2$ and $h_2^0 = \frac{a_1^2}{a_2^2}h_1^0 = \frac{k_a^2}{2ga_2^2}(v^0)^2$ parameterized by e.g. the stationary water level h_1^0. The linearization at a stationary point (h_1^0, h_2^0, v^0) for $h_1^0 > 0$ is then given by:

$$\dot{x} = \begin{pmatrix} -\frac{c_1}{2\sqrt{h_1^0}} & 0 \\ \frac{b_2 c_1}{2b_1\sqrt{h_1^0}} & -\frac{c_2^2 b_1}{2c_1 b_2\sqrt{h_1^0}} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} k_a b_1 \\ 0 \end{pmatrix} u \tag{12a}$$

$$y - y^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{c_1}{2b_1\sqrt{h_1^0}} & 0 \\ 0 & \frac{c_2^2 b_1}{2c_1 b_2^2\sqrt{h_1^0}} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{12b}$$

$$y^0 = \begin{pmatrix} h_1^0 & \frac{c_1^2 b_2^2}{c_2^2 b_1^2}h_1^0 & \frac{c_1}{b_1}\sqrt{h_1^0} & \frac{c_1}{b_1}\sqrt{h_1^0} \end{pmatrix}^T \tag{12c}$$

where the states are $x_i = h_i - h_i^0$ and the input is $u = v - v^0$. The LPV in (12) is parameterized by h_1^0, therefore the variable h_1 will be used as the scheduling variable. Another choice could be v or h_2 since the linearization is also parameterized by v^0 or h_2^0. However, the input v will not reflect the current state of the system, and therefore it would not be a good choice. Using h_2 instead of h_1 would be equivalent with this configuration of the water tanks. The parameters in (11) are identified

f_{h1}	Sensor fault in the water level sensor in the upper tank (tank 1).
f_{h2}	Sensor fault in the water level sensor in the lower tank (tank 2).
f_{f1}	Fault in flow sensor 1 (flow out of tank 1).
f_{f2}	Fault in flow sensor 2 (flow out of tank 2).
f_a	Actuator fault.

Table 1. Faults that can be introduced to the water tank system.

(Klein 1999) and summarized in table 3. This model will be used for simulation with the assumption that the gain factor k_a for the pump is 1.

Parameter	Value	Standard deviation
c_1	0.0638	$6.28{\cdot}10^{-5}$
c_2	0.0878	$2.22{\cdot}10^{-4}$
b_1	0.0452	$5.73{\cdot}10^{-5}$
b_2	0.0419	$2.30{\cdot}10^{-5}$

Table 2. The values and standard deviations for the parameters in (11).

4. DIAGNOSIS SYSTEM

In this section the design of the diagnosis system for the coupled water tanks will be described.

4.1 *Structure*

To determine the present fault mode, residuals that are sensitive to certain faults (monitored faults) and insensitive to other faults (non-monitored faults) are needed. To systematize this a residual structure is used, where for every residual the monitored faults are denoted with a '1' and the non-monitored faults are denoted with a '0'. A diagonal residual structure is chosen and shown in Table 3. Since the decoupling of the non-monitored fault can not be guaranteed when the actual operating point is not a design point,it seems reasonable to decouple as few faults as possible in each residual to make the residual more robust. By choosing a diagonal

	f_{h1}	f_{h2}	f_{f1}	f_{f2}	f_a
r_1	0	1	1	1	1
r_2	1	0	1	1	1
r_3	1	1	0	1	1
r_4	1	1	1	0	1
r_5	1	1	1	1	0

Table 3. A diagonal residual structure.

structure, a larger freedom in forming the residuals is also given, compared to if more faults are to be decoupled in each residual. When no faults are decoupled, the dimension of the null-space is equal to the number of measurements in the non-disturbance case (Frisk 2000), i.e. 4 for the water tank system. Therefore by decoupling one fault per residual, we will end up with a basis for the null space with dimension 3, assuming the fault is detectable. For example the polynomial basis $N_M(s)$ for r_1 at design point $h_1 = 3$ is:

$$N_M(s) = \begin{bmatrix} 0 & 0.033+s & -0.042 & 0 & 0 \\ 0 & 0 & 0.016+s & 0 & -0.016 \\ 0 & -0.78 & 0 & 1 & 0 \end{bmatrix} \quad (13)$$

From the polynomial basis $N_M(s)$ all residual generators $Q(s)$ decoupling the fault f_{h1} at the design point $h_1 = 3$ can be found through (5). Constant faults are often important to detect and isolate, and therefore a low pass filter $D_W(s) = s + 1$ is chosen.

4.2 *Design of Residuals from the Null Basis* $N_M(s)$

How should a residual from the null basis $N_M(s)$ be chosen? By making the response from all the monitored faults f_j to the residual r_i as large as possible, i.e. the transfer function $G_{r_i f_j}$ should in some sense be maximized. In this approach only the constant gain from f_j to r_i will be used. Further only constant selection matrices $\phi(s)$ will be considered. Then finding the $\phi_i(s)$ in (14) gives our $Q(s)$. Thus, the optimization criterion

$$\phi_i(s)^j = \arg \max_{\|\phi_i^j\|_2=1} \min_k \|G_{r_i f_k}(0)\|_2 \quad (14)$$

will yield the selection matrix ϕ_i that maximizes the smallest DC-gain from any of the monitored faults k at the design point j. The optimization criterion yields different ϕ_i^j for every design point j. Choosing different ϕ_i^j (for j) could result in worse performance, than if the same ϕ_i^j had been chosen for all j, because ϕ_i^j and ϕ_i^{j+1} could represent different permutations leading to improper interpolation. Although in our simulations this problem has not occured, the same ϕ_i^j will be used for all j. Summing up, we are using a fix selection matrix based on a design point $h_1 = 3$ in the middle of the operating range, but there are different null-spaces $W^j(s)$ for the various design points.
Using the optimization criterion in (14) at $h_1^0 = 3$ yields $\phi_1 = [\,0\ 0.98\ 0.02\,]$, which together with (13) gives us the first row of the polynomial matrix W^j as:

$$W^j(1,:) = [0\ \ -0.018\ \ 0.018 + 0.98s\ \ 0.02\ \ -0.018] \quad (15)$$

All components of $W^j(s)$ can be written as $a_1^j s + a_0^j$ (compare (2b)). In this application, a_1^j are constant but the absolute value of a_0^j decreases monotonically with the design point j (and scheduling variable), as depicted for the coefficient $W^j(1,2)$ in figure 2. The nonlinearities are apparently more

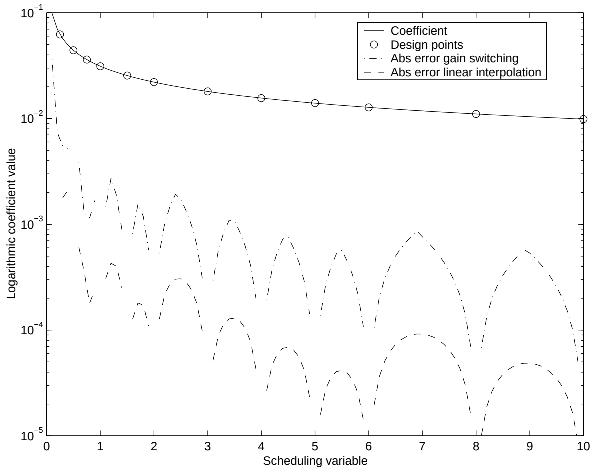

Fig. 2. Typical coefficient dependence of the scheduling variable. The design points are marked, together with the absolute errors for the two gain scheduling methods.

significant for small values of the scheduling variable h_1. Therefore a denser distribution of design points for lower h_1-values is chosen as $h_1 = [\,0.25\ 0.5\ 0.75\ 1\ 1.5\ 2\ 3\ 4\ 5\ 6\ 8\ 10\,]$. With the two interpolation strategies, the absolute error of the coefficients dependence of the scheduling variable h_1 is also shown in Figure 2. Not surprisingly, linear interpolation captures the coefficients dependence of the scheduling variable better.
Now the monitored faults response in the residuals can be calculated as the last term in (3), and as an example r_1 is shown in Figure 3 for design point $\alpha = 3$. The DC-gain from the monitored faults to

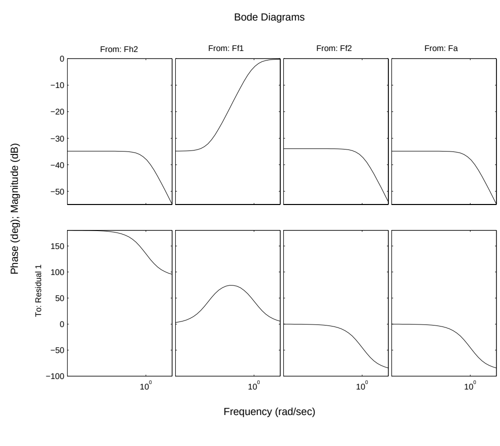

Fig. 3. The monitored faults response in residual r_1 for design point $h_1 = 3$.

residual r_1 are approximately the same, which was the goal of the optimization criteria in (14).

5. SIMULATIONS

In this section the gain scheduled diagnosis system will be tested through simulations of the water tank system. As a test case a 480 seconds long test cycle has been defined, shown in Figure 4. The test case can be divided into several interesting regions, all of them showing specific problems for the residual generator when using gain scheduling. These problems are:

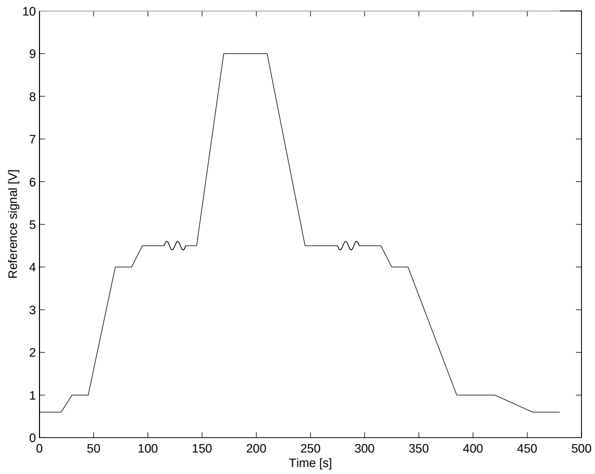

Fig. 4. The reference signal for h_1 in the test case.

(1) Transient behavior of the residual generator.
(2) Chatter, as described in Section 2.3.
(3) Decoupling when faults are introduced in the scheduling variable.
(4) Decoupling when the actual operation point is not a design point.

The transients are less steep on the second half of the test cycle, due to the slow dynamics when emptying the tanks. The test case is more thoroughly described in (Klein 1999). Simulations are performed with the test cycle using gain switching or linear interpolation as gain scheduling strategy, in both the fault free case and in all the cases of a single fault of size 0.5 occuring at $t = 15$ s. The resulting residuals of these simulations can be found in the complete report (Klein 1999). Some results regarding the problems above are now presented in the following.

Transient behavior When the water tank system changes operating point, a transient behavior for the residual generator may occur. Since the linearizations are made in equilibrium, they will not correspond too well with the nonlinear system during transients, and this model error can cause residuals to become large even in the fault free case. This is exemplified in Figure 5, where residual 5 is shown in the fault free case. During the time intervals when transients are present the amplitude of the residual becomes considerably larger, making the evaluation of the residual harder. The peak value during a transient is smaller for linear interpolation, than for gain switching, suggesting that linear interpolation should be used. During transients a weighting factor on the residual could be introduced, being small when the system is subject to transient behavior. This is not investigated any further here.

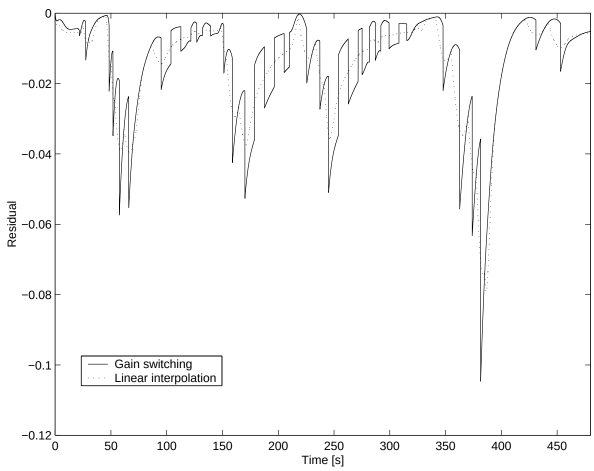

Fig. 5. Residual 5 in the fault free case , when using linear interpolation and gain switching as gain scheduling strategies.

Chatter Residual 1 is simulated using gain switching and linear interpolation in the fault free case and with the fault f_{f1} in figure 6. In the time interval shown, the operating point moves back and forth around a switch point. The gain switched interpolation method has larger discontinuities and higher amplitudes than the linear interpolation method has. The levels of the residuals subjected to the fault have a clearly larger amplitude, so evaluation would not be a problem here. But the ratio between the residual subject to f_{f1} and the residual in the fault free case is much larger in the linearly interpolated case, suggesting that smaller faults could be detected with linear interpolation.

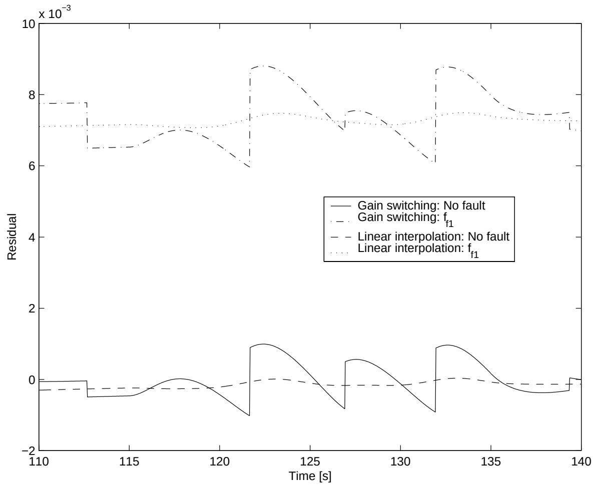

Fig. 6. Residual 1 (time interval 110–140 s) in both the fault free case and in the case with a fault $f_{f1} = 0.5$, using gain switching and linear interpolation.

Decoupling problem for a fault in the scheduling variable A so far neglected problem, is that the use of the measurement signal y_1 as the gain scheduled variable h_1 could severly degrade the performance of the diagnosis system. When the scheduling variable is subject to a fault, the actual operating point is different from the one concluded from the scheduling variable. That situation has occured in

Figure 7, where the linearly interpolated r_1 is shown in the fault free case, together with the residuals subject to the faults f_{h1} and f_{f1} respectively. According to Table 3 f_{h1} should be decoupled, but

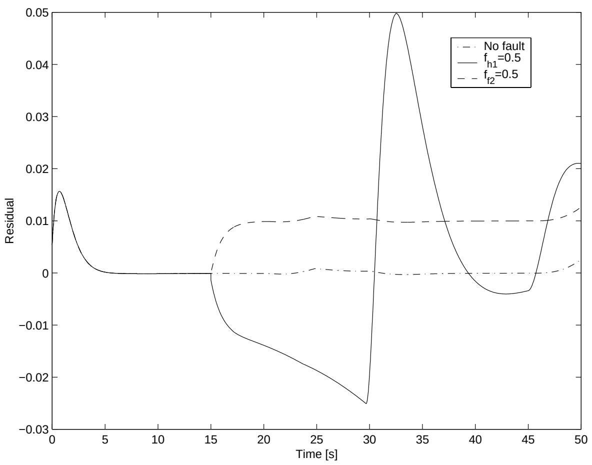

Fig. 7. Residual 1 using linear interpolation as gain scheduling strategy. The fault f_{h1} should be decoupled in r_1, but it is not.

apparently it is not. It even has larger amplitude than r_1 subject to f_{f1}. The fault f_{h1} is therefore not decoupled. This is a problem only for low water levels, since a linearization has smaller validity range in that operating region. As expected, the decoupling of f_{h1} is better for higher water levels (Klein 1999). To solve the problem with the decoupling of a fault in the scheduling variable, we could either put in '1' instead of '0' in the residual structure in Table 3 and thereby limit the isolation abilities of the diagnosis system or we could use e.g. an observer for h_1 not using measurement signal y_1, as the scheduling variable for (at least) residual 1.

Decoupling when the actual operating point is not a design point Simulations are made with a reference signal that is constant during longer periods, than in the earlier defined test cycle. In the first time interval (0–60 s), see Figure 8, the residual converges to a nonzero value (although small), showing that perfect decoupling is not achieved when the operating point is not a design point, or rather that we have a model fault. In the second time interval (70–130 s) the residual converges to zero, showing that perfect decoupling can be achieved, since the model now is perfect. This shows that one should not expect the residual generator to work as well between design points as at design points.

6. CONCLUSIONS

Design of residual generators for a coupled water tank system has been done using *gain scheduling*. Simulations of the water tank system have been performed to evaluate the behavior of the residual generators. A set of linear approximations is found to the nonlinear system at various stationary points, then residual generators are designed using the minimal polynomial basis method and a newly proposed optimization method. Two interpolation methods are used as gain scheduling strategies, namely nearest neighbor approximation and linear interpolation for the coefficients in the residual generators. Simulations show that during steady state operation, the residual generators behave well, but during transients the performance of the residual generators decrease significantly. A weighting factor on the residual could therefore be introduced, being small when the system is subject to transient behavior. Simulations also propose that linear interpolation is better than nearest neighbor approximation, as a gain scheduling strategy.

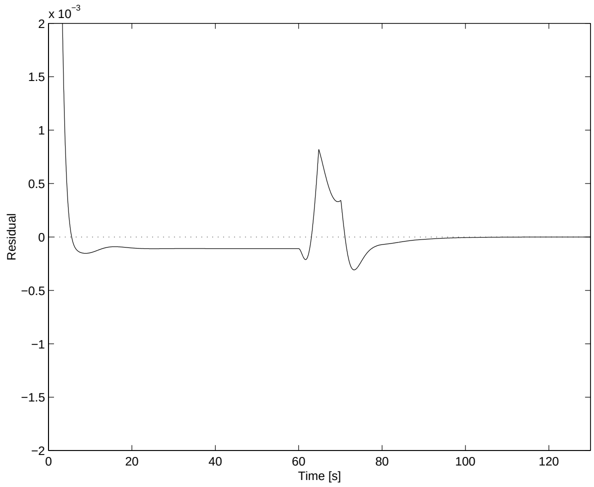

Fig. 8. The steady state behavior for r_1 in the fault free case, for $h_1 = 0.6$ (between design points) in the time interval 0–60 s and for $h_1 = 1$ (design point) in the time interval 70–130 s.

A fault in the scheduling variable causes problems, especially for low water tank levels, since a linearization has smaller validity range in that operating region. It is proposed that different scheduling variables should be used, or rather that they should be calculated in different ways, using different signals from the system. This to make sure that a fault in the scheduling variable can be decoupled.

7. REFERENCES

Frisk, E. (2000). Order of residual generators–bounds and algorithms. IFAC SAFEPROCESS 2000, Budapest, Hungary.

Kailath, T. (1980). *Linear Systems*, Prentice-Hall Inc. p.385, ISBN 0-13-536961-4.

Klein, M. (1999). Evaluating some gain scheduling strategies in diagnosis of a tank system, *Technical Report LiTH-R-2183, ISSN 1400-3902*, Department of Electrical Engineering, Linköping University

Lagerberg, A. (1996). Gain scheduling control and its appliction to a chemical reactor model, Licentiate thesis, Chalmers Tekniska Högskola. ISBN 91-7197-255-2.

Nyberg, M. (1999). *Model Based Fault Diagnosis: Methods, Theory and Automotive Engine Applications*, PhD thesis, Linköping University. ISBN 91-7219-521-5, ISSN 0345-7524.

Nyberg, M. & Frisk, E. (1999). A minimal polynomial basis solution to residual generation for fault diagnosis in linear systems. Proc. of IFAC 1999 World Congress, Beijing, P.R. China. Vol. P pp.61–66.