LOWERING ORDERS OF DERIVATIVES IN NON-LINEAR CONSISTENCY RELATIONS - THEORY AND SIMULATION EXAMPLES

ERIK FRISK AND JAN ÅSLUND

Abstract. Consistency relations are often used to design residual generators based on non-linear process models. A main difficulty is that they generally include time differentiated versions of known or measured signals which are difficult to estimate in a noisy environment, especially higher order derivatives. The main results of this paper show how to lower, or if possible avoid, the need to estimate derivatives of known signals in order to compute the residual. This is achieved by rewriting the problem into an integrability problem using state-space realization theory. An attractive feature of the approach is that general differential algebraic system descriptions can be handled in the same way as for example ordinary differential equations and also that stability of the residual generator is always guaranteed.

1. Introduction

Residual generation for fault diagnosis based on non-linear dynamical models has in the control community mainly been performed in two ways, either via non-linear state observation or via non-linear consistency relations. Both approaches have their own set of advantages and disadvantages. For state observation techniques, stable estimation has been studied for many years. However, to facilitate not only fault detection but also fault isolation using structured residuals, *decoupling* of fault signals in the residual is needed. Results for such unknown input decoupling non-linear observers are not as widespread as in the linear case. A notable contribution is [PI00; PI01] which gives a rather general non-linear extension of the geometric approach for linear system proposed in [Mas86; MVW89] for control-affine non-linear systems.

If instead of observers, non-linear consistency relations is used for designing residual generators to fit in a structured residuals framework, the decoupling problem is transformed into an elimination problem. For classes of non-linear model descriptions this elimination problem is generally solvable. For example when the system is described by polynomial differential algebraic equations, constructive elimination theory using Gröbner bases [CLO96] can be used to derive consistency relations with desired decoupling properties. The systems considered need not be on control-affine form but can be of any form and also include algebraic constraints. Consistency relations has proven beneficial in the linear case, but when generalizing to the non-linear problem some additional difficulties arise. A main difficulty with consistency relations is that for the case of dynamic systems, the consistency relation generally include time differentiated versions of known signals and these are not normally known. One way to still use the consistency relation to generate a residual is to resolve to approximations and estimate the time derivatives. This can be done by simple schemes as derivation of low-pass filtered signals or more elaborate schemes using for example spline interpolation techniques. For linear systems, this problem is easily circumvented and no approximations of time derivatives are needed. Adding residual generator dynamics of high enough order and finding a state-space realization using linear realization theory makes it possible to compute a residual without any approximations. This report aims at extending this linear approach also for non-linear systems. This is done by transforming the problem into an integrability problem.

Section 2 give a more formal description of the approach objective and describe the class of residual generators considered. In Section 3, a small example is used to illustrate the basic principle behind the proposed solution which in Section 4 is formalized and theoretical properties are given for first order consistency relations. Finally a simulation example is included in Section 5 where also some aspects of the solution is compared to the solution derived in [PI01].

2. Problem Formulation

Let $y(t)$ and $u(t)$ denote the measurements and controller outputs, i.e. the signals that are considered known. A consistency relation is then a relation

$$c(y(t), u(t)) = 0 \tag{1}$$

that holds in the fault free case. Therefore, testing if $c(y, u) = 0$ is a test on consistency between measured data and the equations/assumptions used when deriving $c(y, u)$. The task of the diagnosis system designer is then to find a suitable set of such relations where different subsets of relations are invalidated by different subsets of faults. In this way, fault isolation can be achieved. This is often referred to as structured residuals [Ger91]. A fault is detectable in (1) if the relation does not hold in case of a fault.

Throughout this report $u^{<\delta>}$ is used to denote the variable u and all its time derivatives up to order δ, i.e. $u^{<\delta>} = \{u, \dot{u}, \ldots, u^{(\delta)}\}$. Also, for notational convenience all known signals will be denoted u. Let $c(u_1^{<\alpha_1>}, \ldots, u_m^{<\alpha_m>}) = 0$ be a consistency relation. If $c(u^{<\alpha>})$ was linear we could add any stable linear residual generator dynamics of high enough order (at least as large as the highest input derivative):

$$r^{(n)} + a_n r^{(n-1)} + \cdots + a_1 r = c(u^{<\alpha>}).$$

This expression can then easily be transformed, using linear realization theory, into a state-space form where no derivatives need to be estimated to be able to compute the residual. For example, if $u + \dot{u} + \ddot{u} = 0$ is a consistency relation. Adding second order residual generator dynamics and writing the residual generator on operator form gives

$$r(t) = \frac{p^2 + p + 1}{p^2 + a_2 p + a_1} u(t)$$

It is clear that this can easily be written on explicit state-space form and the residual r can be computed without the need to estimate any derivatives. This is the approach that in the following sections will be mimicked for also the non-linear case. Unfortunately, but hardly surprising, is that the realization step that was immediate in the linear case is not direct in the non-linear case.

A generalization of the above approach for the non-linear case would be to find residual generator dynamics g and a transformation Ψ such that

$$g(r^{<n>}) = \Psi(c(u^{<\alpha>}), u^{<\alpha>})$$

can be realized on state-space form with no differentiated inputs. The functions g and Ψ must quite naturally satisfy the constraints that $r = 0$ is a stable locus of the differential equation

$$g(r^{<n>}) = 0$$

and that $\Psi(0, u^{<\alpha>}) = 0$ for all u. This means that the residual generator is stable and that if $c(u)$ is a consistency relation so is $\Psi(c(u^{<\alpha>}), u^{<\alpha>})$. Note that if there exists u such that $\Psi(c, u^{<\alpha>}) = 0$ for $c \neq 0$, the transformation of the original consistency relation may reduce the detectability performance of the residual compared to the detectability of the original consistency relation.

To be able to proceed, the problem is not solved in its full generality and two restrictions are imposed on g and Ψ:

(1) It is assumed that the residual generator dynamics is on the form
$$r^{(n)} + g(r^{<n-1>})$$
i.e. that the highest order derivative enters linearly. This seems like a minor restriction since the order of residual generator dynamics always can be increased and linear dynamics can be appended.

(2) The transformation Ψ is assumed to be on the form
$$\Psi(c(u), u^{<\alpha>}) = \alpha(u^{<\alpha>})c(u^{<\alpha>})$$
where $\alpha(u^{<\alpha>}) \neq 0$. Here it is clear that the requirement $\Psi(0, u^{<\alpha>}) = 0$ is trivially fulfilled. This is the more restrictive restriction of the two. See [FÅ03] for a more thorough discussion on this topic.

This means that the following class of residual generators is considered in this report
$$r^{(n)} + g(r^{<n-1>}) = \alpha(u^{<\alpha>})c(u^{<\alpha>}) \tag{2}$$
The function α is called an *integrating* factor. Why α is called an integrating factor will be made clear in the following sections along with a motivation for this residual generator structure.

The problem can now be stated as follows: Find stable residual generator dynamics g and an integrating factor α such that there exists state variables such that (2) can be written on explicit state-space form
$$\dot{z} = f(z, u)$$
$$r = h(z, u)$$
If this is not possible, try to find a solution where input derivatives are lowered as much as possible. This report only provide a first step in the solution of the above problem where input derivatives are lowered at most one step. This means that only first order consistency relations are fully covered by the results in this report. Discussions on how to proceed for higher order relations are discussed further in [FÅ03].

2.1. **Fault response.** It is interesting to see what the *internal form*, or the fault response, of a residual on the form (2) looks like and also to analyze the influence of the integrating factor on the fault response. To explore this and derive an expression for the fault response, let $c(u^{<\alpha>}, f^{<\gamma>})$ (for short here written $c(u, f)$) be the consistency relation including also the faults to be detected. Here it is, without loss of generality, assumed that $f = 0$ corresponds to the fault-free case. Since it is a consistency relation it holds that $c(u, f) = 0$ for all $u(t)$ and $f(t)$ that are consistent with the original model equations. The following expression can then easily be derived
$$0 = c(u, f) = c(u, 0) - (c(u, 0) - c(u, f)) = c(u, 0) - c_f(u, f)$$
where $c_f(u, f) = 0$ if $f = 0$. A fault f_i is detectable in the consistency relation $c(u, 0) = 0$ if and only if
$$\frac{\partial}{\partial f_i} c_f(u, f) \neq 0$$
The fault response in the residual can then be written as the solution to the differential equation
$$r^{(n)} + g(r^{<n-1>}) = \alpha(u)c_f(u, f) \tag{3}$$

which means that fault f_i will be detectable only if

$$\frac{\partial}{\partial f_i}\{\alpha(u)c_f(u,f)\} = \alpha(u)\frac{\partial}{\partial f_i}c_f(u,f) \neq 0$$

This condition will also be a sufficient condition for detectability if the residual generator dynamics is such that $r = 0$ is not a stable locus of

$$r^{(n)} + g(r^{<n-1>}) = \alpha(u)c_f(u,f) \neq 0$$

Thus, detectability can be lost when α is 0 (or small), but in any other case the detectability of the residual (2) is the same as in the original consistency relation.

3. Introductory Example

In this section, a design procedure is outlined and also the residual generator structure (2) is motivated. For this, consider the small first order non-linear consistency relation

$$c(u) = u_1\dot{u}_2 + A\dot{u}_1u_2 - \theta_2u_3 \tag{4}$$

Assume linear first order residual generator dynamics, for example

$$\dot{r} + a_1r = u_1\dot{u}_2 + A\dot{u}_1u_2 - \theta_2u_3$$

With $A = 1$ it is straightforward to verify that with the state variable $z = r - u_1u_2$ we get an explicit state-space realization of the residual generator as:

$$\begin{aligned}\dot{z} &= -a_1z - a_1u_1u_2 - \theta_2u_3\\ r &= z + u_1u_2\end{aligned}$$

Here, no estimates of derivatives are needed to compute the residual r. The reason this worked was that

$$\frac{d}{dt}(u_1u_2) = u_1\dot{u}_2 + \dot{u}_1u_2 \tag{5}$$

That is, with a suitable change of variables we obtained a linear problem. Now, consider a case where $A \neq 1$, then there exists no state variable like above such that all input derivatives disappear. This is because

$$u_1\dot{u}_2 + A\dot{u}_1u_2$$

is not a total derivative for $A \neq 1$, i.e. there exists no function like u_1u_2 in (5) for this case. A solution in this case is to introduce an integrating factor. Multiply the consistency relation with for example $\alpha(u_1,u_2) = u_1^{A-1}$ and we obtain a new consistency relation (the price is possible loss of detectability when $u_1 = 0$)

$$c'(u_1,u_2,u_3) = \alpha(u)c(u) = u_1^A\dot{u}_2 + Au_1^{A-1}\dot{u}_1u_2 - u_1^{A-1}\theta_2u_3$$

and here the nice property (5) of the $A = 1$ case is retrieved since

$$c'(u_1,u_2,u_3) = \frac{d}{dt}(u_1^Au_2) - u_1^{A-1}\theta_2u_3$$

Then $\dot{r} + a_1r = c'(u_1,u_2,u_3)$ can be realized by the state variable $z = r - u_1^Au_2$ which give a realization

$$\begin{aligned}\dot{z} &= -a_1z - a_1u_1^Au_2 - u_1^{A-1}\theta_2u_3\\ r &= z + u_1^Au_2\end{aligned} \tag{6}$$

For this to be useful, a constructive procedure to compute the integrating factor and the state variable is needed. A procedure for this will now be outlined for this example and then in Section 4, the approach is formalized for general first order consistency relations. It is also proven that the success of the procedure is a necessary and sufficient condition for the existence of residual generators on the form (2) that can be transformed into a state-space description with no derivated inputs.

Now, continue with the example. The problem when $A \neq 1$ is that

$$u_1\dot{u}_2 + A\dot{u}_1 u_2$$

is not a total derivative. This is equivalent to that

$$F = [Au_2 \ u_1]$$

is not an exact differential, i.e. the partial differential equation $d\lambda = F$ has no solution. But, if instead of requiring $d\lambda = F$ we settle for the less restrictive $d\lambda \in \text{span } F$, or equivalently, the existence of an integrating factor α such that the partial differential equation

$$d\lambda = \alpha F \tag{7}$$

has a solution. Multiply the above equation from the right with $F^{\perp}$ and we get the equivalent partial differential equation

$$d\lambda \, F^{\perp} = 0 \tag{8}$$

This is a special type of equation that is discussed in Section A.1 for which necessary and sufficient solvability conditions exists and also a constructive procedure of how to compute λ. In this example we get

$$d\lambda(u_1, u_2) \begin{bmatrix} u_1 \\ -Au_2 \end{bmatrix} = 0$$

According to Theorem A.1, this has a solution if and only if the vector field $[u_1(t) \quad -Au_2]^T$ is involutive. The involutive property is immediate since the vector field is non-singular and of dimension 1. A solution is given by $\lambda = u_1^A u_2$. Note that the solution is not unique and that there exists infinitely many solutions. In fact it is easy to verify that if λ is a solution, so is $\gamma(\lambda)$ where $\gamma(\cdot)$ is an arbitrary function. See [FÅ03] how this fact can be used to make implementation of the residual generator easier.

The corresponding integrating factor can then be computed using (7) as

$$\alpha = \frac{\frac{\partial\lambda}{\partial u_1}}{Au_2} \left(= \frac{\frac{\partial\lambda}{\partial u_2}}{u_1} \right) = \frac{Au_1^{A-1}u_2}{Au_2} = u_1^{A-1}$$

Once again, given F it is completely (almost true) automatic to compute λ and α and is easily implemented in a computer algebraic tool. Now that α and λ have been computed, the residual generator description becomes (on form (2)):

$$\dot{r} + a_1 r = \alpha(u_1, u_2)c(u) = -a_1 x_1 + \dot{\lambda}(u_1, u_2) - \theta_2\alpha(u_1, u_2)u_3$$

Again, with the state variable $z = r - \lambda(u_1, u_2)$ we get the state-space realization

$$\begin{aligned} \dot{z} &= -a_1 z - a_1\lambda(u_1, u_2) - \theta_2\alpha(u_1, u_2)u_3 \\ r &= z + \lambda(u_1, u_2) \end{aligned}$$

with $\alpha = u_1^{A-1}$ and $\lambda = u_1^A u_2$ which is identical to (6).

Note that when A is not an integer the residual is a complex quantity if $y_1 < 0$. This is because the integrating factor is not a mapping from $\mathbb{R} \to \mathbb{R}$ but rather from $\mathbb{R} \to \mathbb{C}$. Although this situation is not desirable, it is a problem that often can be avoided by choosing a suitable solution to PDE (8). In this case it is straightforward to verify that instead of using $\lambda = u_1^A u_2$, the solution $\lambda = |u_1|^A u_2$ also satisfies (8) and does not produce complex solutions in case $u_1 < 0$. The corresponding integrating factor becomes $\alpha = |u_1|^{A-1}\text{sgn } u_1$. Systematic procedures to avoid complex solutions has not yet been explored.

4. General First Order Consistency Relations

Next, a result will be presented that formalizes the procedure from last section for first order consistency relations. The objective is to, given a consistency relation $c(u) = 0$, find an integrating factor α and residual generator dynamics $g(r)$ such that

$$\dot{r} + g(r) = \alpha(u)c(u, \dot{u}) \tag{9}$$

can be written on state space form. Some remarks are also given along with a complete and constructive design procedure.

Theorem 4.1. *Let $c(\dot{u}, u)$ be a consistency relation of order 1. There exists, in a neighborhood of a point u_0, an integrating factor $\alpha(u)$ and a change of coordinates such that* (9) *is transformed into a state-space realization with no input derivatives if and only if $c(\dot{u}, u)$ can be written on the form*

$$c(\dot{u}, u) = \sum_{i=1}^{n} h_i(u)\dot{u}_i + v(u) \tag{10}$$

where the functions h_i satisfy the condition

$$h_i\left(\frac{\partial h_k}{\partial u_j} - \frac{\partial h_j}{\partial u_k}\right) + h_j\left(\frac{\partial h_i}{\partial u_k} - \frac{\partial h_k}{\partial u_i}\right) + h_k\left(\frac{\partial h_j}{\partial u_i} - \frac{\partial h_i}{\partial u_j}\right) = 0 \tag{11}$$

for all i, j, k and $h \neq 0$ in a neighborhood of u_0.

Proof. See [FÅ03] for a complete proof. ∎

Remark 4.1. From the proof it is interesting to see that the residual generator dynamics g can be chosen *arbitrarily* as long as it has order greater or equal than 1. Therefore, the residual generator dynamics can *always be chosen linearly* which makes it easy to guarantee stability. This is a highly attractive feature of the approach.

Remark 4.2. In the design procedure, the integrating factor is computed indirectly, by first solving the partial differential equation (8) for λ and then using (7) for computing the integrating α. The advantage of solving for λ instead of α directly can be illustrated in the small example from Section 3. It can be shown that the partial differential equation when solving for α directly is the equation

$$(1 - A)\alpha(u_1, u_2) - Au_2\frac{\partial \alpha(u_1, u_2)}{\partial u_2} + u_1\frac{\partial \alpha(u_1, u_2)}{\partial u_1} = 0$$

which has solutions

$$\alpha(u_1, u_2) = u_1^{A-1}\gamma(u_1^A u_2)$$

where γ is any function. Note that this equation is not on the special form discussed in Theorem A.1 which in general makes solving for α more difficult than solving for λ.

Remark 4.3. Theorem 4.1 only provide local solutions. To provide global conditions, additional completeness conditions of involved distributions. These requirements are directly given by Theorem 2 in [DR95].

4.1. Summary of design procedure.

The design procedure is summarized in the following steps

(1) Assure that the consistency relation is on the form

$$c(u, \dot{u}) = \sum_{i=1}^{n} h_i(u)\dot{u}_i + v(u), \tag{12}$$

and denote $\Delta = \text{span}\ \{h_1, \ldots, h_r\}$.

(2) Compute the annihilator of Δ, i.e. $\Omega = \Delta^{\perp}$.
(3) Solve the partial differential equation $d\lambda(u)\Omega(u) = 0$ for $\lambda(u)$. A computer algebraic tool can, in benign cases, be used to automatically compute the solutions by solving a set of ordinary differential equations, see for example the proof of Frobenius theorem in [Isi95] for details and a constructuve algorithm.
(4) Compute the coefficients $\alpha(u)$ by solving the *linear* equation

$$d\lambda(u) = \alpha(u)h(u)$$

The new consistency relation can now be formed as

$$c'(u, \dot{u}) = \alpha(u)c(u, \dot{u}) \tag{13}$$

which directly can be rewritten as

$$c'(u, \dot{u}) = \frac{d}{dt}\lambda(u) + \alpha(u)v(u) \tag{14}$$

This is on a linear form which makes it direct to transform

$$\dot{r} + \beta r = c'(u, \dot{u})$$

into an explicit state-space form using the state variable $z = r - \lambda$:

$$\begin{aligned}
\dot{z} &= -\beta z - \beta\lambda(u) + \alpha(u)v(u) \\
r &= z + \lambda(u)
\end{aligned}$$

5. Simulation Example

In this section, simulation of a point-mass model of a satellite is used to illustrate the procedure and some of its properties. Also, a comparison of some aspects of the solution obtained with the geometric approach to non-linear fault detection [PI00; PI01] is included.

The model is taken from [PI01] and can be stated as

$$\begin{aligned}
\dot{x}_1(t) &= x_2(t) \\
\dot{x}_2(t) &= x_1(t)x_4^2(t) - \theta_1 \frac{1}{x_1^2(t)} + \theta_2 u_1(t) + d(t) \\
\dot{x}_3(t) &= x_4(t) \\
\dot{x}_4(t) &= -2\frac{x_2(t)x_4(t)}{x_1(t)} + \theta_2 \frac{u_2(t)}{x_1(t)} + \theta_2 \frac{f(t)}{x_1(t)} \\
y_1(t) &= x_1(t) \\
y_2(t) &= x_3(t) \\
y_3(t) &= x_4(t)
\end{aligned}$$

where $(x_1,\, x_3)$ are the polar coordinates of the satellite (radius,angle) and x_2, x_4 are the tangential respectively the radial velocity of the satellite. There are two actuator signals, u_1 and u_2 which are the radial and tangential thrusters. There is a fault f acting on u_2 and detecting this fault is the objective. A matter which makes this difficult is the unknown disturbance d which must be decoupled in the residual.

For this, a consistency relation is derived that is sensitive to the fault f where the disturbance d is decoupled. Since the model after trivial manipulations can be stated on polynomial form, a consistency relation can be automatically computed using Gröbner basis techniques. The resulting consistency relation including the fault signal is

$$y_1\dot{y}_3 + 2\dot{y}_1 y_3 - \theta_2 u_2 - \theta_2 f = 0 \tag{15}$$

See [Fri01] for more details. With 3 measurements and 1 disturbance to decouple, one might suspect at least one more consistency relation. One such relation, also found by the Gröbner basis machinery, is $\dot{y}_2 - y_3 = 0$ but this one is not sensitive to the fault. Therefore, only the first relation is explored further. This means that the consistency relation used for detecting the fault is

$$y_1\dot{y}_3 + 2\dot{y}_1 y_3 - \theta_2 u_2 = 0 \tag{16}$$

and the internal form is given by (3) with $c_f(u, f) = \theta_2 f$. It is immediate that the highest order derivatives $\dot{y}_1$ and $\dot{y}_3$ enters affinely and the conditions of Theorem 4.1 is fulfilled. Equation (16) can easily be identified as (4) with $A = 2$ and obvious renaming of the signals. Thus, following the example in Section 3 we know that an integrating factor is given by $\alpha = y_1$ and that the residual generator on form (2) is given by

$$\frac{1}{a_1}\dot{r} + r = y_1^2\dot{y}_3 + 2y_1\dot{y}_1 y_3 - \theta_2 y_1 u_2, \quad a_1 > 0$$

A first order realization of the above residual generator is given by

$$\begin{aligned} \dot{z} &= -a_1 z - a_1 y_1^2 y_3 - \theta_2 y_1 u_2 \\ r &= a_1(z + y_1^2 y_3) \end{aligned} \tag{17}$$

The residual generator dynamics has the form $\frac{1}{a_1}\dot{r} + r$ instead of $\dot{r} + a_1 r$ to achieve unit DC-gain of the residual generator. Note that the residual is always real in this case even if the A parameter is non-integer. This is because the radius of the satellite position is always positive. In the simulations[1] both thrusters are set to full reverse ($u_i = -1$) at $t = 20$. The disturbance is a white and Gaussian random process. Figure 1 show the satellite position in a fault free simulation.

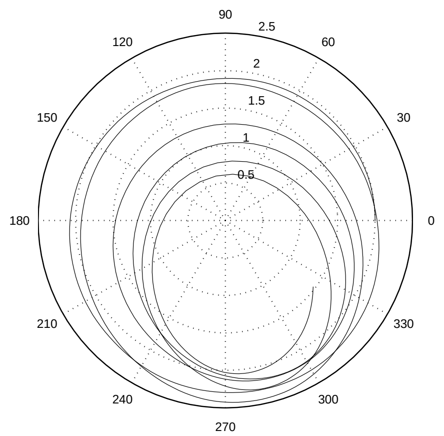
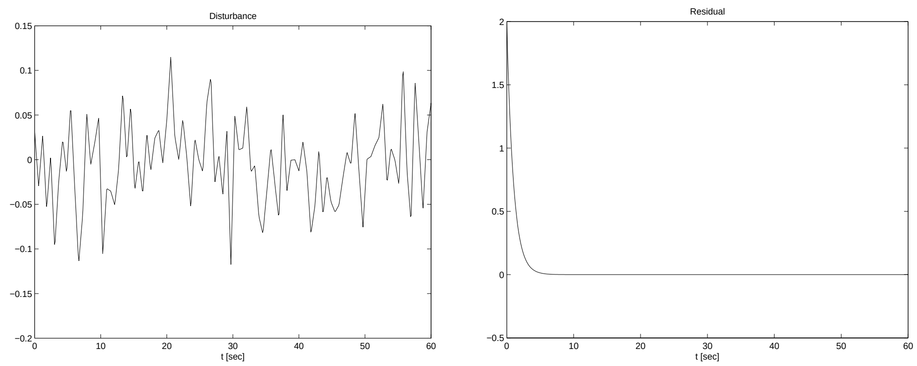


FIGURE 1. Satellite position in a fault free simulation.

The corresponding disturbance realization and residual is shown in Figure 2. It is clear that the residual is 0 after influence from the unknown initial conditions has vanished, i.e. the influence of the disturbance d is completely decoupled. The rate of decay of the influence from unknown initial conditions is completely controlled by the residual generator dynamics. In this simulation the poles of the residual generator dynamics is placed at $s = -1$ by setting $a_1 = 1$.

The solid line in Figure 3-a show the residual in case of a 10% decrease of tangential thruster power starting at $t = 30$, i.e. $f = -0.1\,u_2$ when $t \geq 30$ and $f = 0$ otherwise. The fault is clearly visible in the residual and can be detected

[1]In the simulations the parameter values $\theta_1 = 2$ and $\theta_2 = 10^{-2}$ are used.



(a) Disturbance.

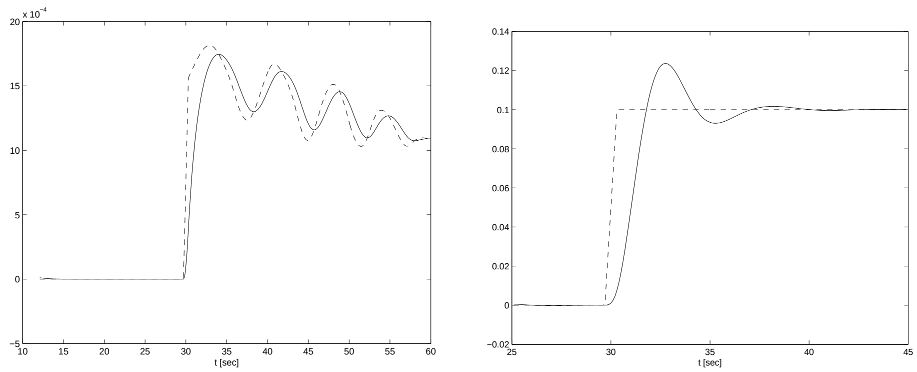
(b) Fault-free residual

FIGURE 2. The disturbance d and the residual r in fault free simulation.

by thresholding of the residual. Next, we verify that the fault response obeys expression (3). Also in Figure 3-a, the signal $\alpha(u)c_f(u)$ is plotted as a dashed line. According to (3), the residual should be a low-pass filtered version of this signal, i.e.

$$r = \frac{a_1}{p + a_1}\left(\alpha(u)c_f(u)\right) = \frac{a_1}{p + a_1}\left(\theta_2 y_1 f\right) \qquad (18)$$

and when looking at the plots this seems reasonable. A more detailed computation show that (18) holds exactly.



(a) The solid line is the residual and the dashed is the signal $\alpha(u)c_f(u, f)$.

(b) The solid line is the fault estimation $\hat{f}$ and the dashed line is the true fault f.

FIGURE 3. The residual and a fault estimate in the case of a 10% fault in the tangential thruster.

5.1. **Estimating the Fault Size.** Consider the case when the fault is constant or slowly varying. Then we can use expression (18) to estimate the fault size. Below two ways of estimating the fault size is explored where one is a standard observer based technique and one is based on direct computation of the fault size.

Forming an observer to estimate the fault size using expression (18) can be done in many ways. A direct way is for example given by

$$\dot{\hat{r}} = -a_1\hat{r} + a_1\theta_2 y_1\hat{f} + K_1(r - \hat{r})$$
$$\dot{\hat{f}} = K_2(r - \hat{r})$$

Figure 3-b show the output of this observer where the solid line is the observer estimate of the fault and the dashed the true fault. It is seen that the estimate converges nicely to the true fault. Note that this is not a general result, no guarantees of identifiability/observability of f in (18) can be proven. However, in this particular example this simple approach to estimating the fault works.

A second approach is to directly compute the fault. Consider again the internal form of the residual generator

$$\dot{r} + a_1 r = a_1\theta_2 y_1 f$$

Now also consider a signal w which can be computed on-line according to

$$\dot{w} + a_1 w = a_1\theta_2 y_1$$

Using this signal, we can produce a simple fault estimate by

$$\hat{f} = \begin{cases} \frac{r}{w} & \text{if } w \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

To analyze the properties of this fault estimator let $e = wf - r$. We then have

$$\begin{aligned} \dot{e} = \dot{w}f - \dot{r} + w\dot{f} = -a_1 wf + a_1\theta_2 y_1 f + a_1 r - a_1\theta_2 y_1 f + w\dot{f} = \\ = -a_1(wf - r) + w\dot{f} = -a_1 e + w\dot{f} \end{aligned}$$

Now, assume again that f is constant or slowly varying which means that $\dot{f} \approx 0$. Neglecting the influence from $\dot{f}$ we get $e = 0$ is a global and stable locus which means that

$$wf \to r \text{ when } t \to \infty$$

From this it is natural to compute a fault estimate according to (19). Figure 4 shows the result of this fault estimation scheme. Note that the success of an approach

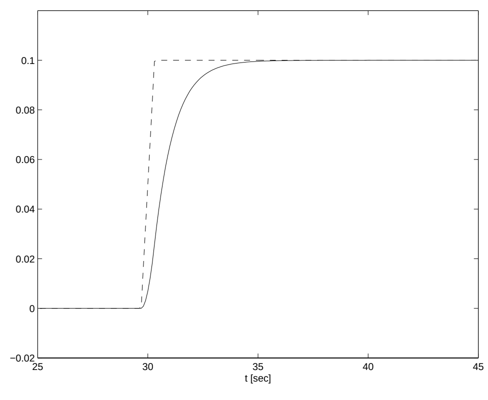


FIGURE 4. The solid line is the fault estimate and the dashed is the actual fault.

like above relies on:

(1) linear residual generator dynamics.
(2) affine fault influence with no time differentiated signals in the coefficient.

Further investigations are needed to evaluate if and how the scheme can be generalized, noise properties of the estimate etc.

5.2. **Comparison with solution by De Persis et.al.** This section consist of a brief comparison with the design just presented and the solution presented in [PI01] where a residual generator is designed for the same problem. Also some more general remarks of differences and similarities are included.

First, a brief outline of the main ideas of the solution proposed in [PI01]. The goal is to find a change of coordinates $\Psi(x)$ and an output transformation $\Phi(y)$ such that the original model is transformed into the following form

$$\begin{aligned}
\dot{z}_1 &= f_1(z_1, z_2) + g_1(z_1, z_2)u + l_1(z_1, z_2)f \\
\dot{z}_2 &= f_2(z_1, z_2, z_3) + g_2(z_1, z_2, z_3)u + l_2(z_1, z_2, z_3)f + p_2(z_1, z_2, z_3)d \\
\dot{z}_3 &= f_3(z_1, z_2, z_3) + g_3(z_1, z_2, z_3)u + l_3(z_1, z_2, z_3)f + p_3(z_1, z_2, z_3)d \\
\tilde{y}_1 &= h_1(z_1) \\
\tilde{y}_2 &= z_2
\end{aligned}$$

Then, using the first dynamic equation and the measurements we get the "quotient" system:

$$\begin{aligned}
\dot{z}_1 &= f_1(z_1, \tilde{y}_2) + g_1(z_1, \tilde{y}_2)u + l_1(z_1, \tilde{y}_2)f \\
\tilde{y}_1 &= h_1(z_1)
\end{aligned}$$

which can be used for diagnosis since the disturbance has been eliminated from the equations. The residual generator can for example be an observer for the quotient system which, under some rather technical assumptions, is observable. The necessary transformations is found using differential geometric tools such as computing the smallest unobservability distribution containing the disturbance. For the satellite example explored in this section, the quotient system is

$$\begin{aligned}
\dot{z}_{11} &= \frac{z_{12}}{\tilde{y}_2^2} \\
\dot{z}_{12} &= \theta_2 \tilde{y}_2 u_2 + \theta_2 \tilde{y}_2 f \\
\tilde{y}_1 &= \begin{pmatrix} z_{11} \\ z_{12} \end{pmatrix}
\end{aligned} \tag{20}$$

where the output transformation $\tilde{y} = \Phi(y)$ and the state transformation is given by

$$\tilde{y}_1 = \begin{pmatrix} y_2 \\ y_1^2 y_3 \end{pmatrix}, \quad \tilde{y}_2 = y_1, \quad z_1 = \begin{bmatrix} x_3 \\ x_1^2 x_4 \end{bmatrix}, \quad z_2 = x_1, \quad z_3 = x_2$$

In this case it can be seen that not the whole quotient system z_1 is needed in the observer design, only z_{12}. A residual generator can then be formed as

$$\begin{aligned}
\dot{\hat{z}}_{12} &= \theta_2 \tilde{y}_2 u_2 + K(\tilde{y}_{12} - \hat{z}_{12}) \\
r &= \tilde{y}_{12} - \hat{z}_{12}
\end{aligned}$$

Using the original outputs rather than the transformed outputs and letting $z = -\hat{z}_{12}$ we get

$$\begin{aligned}
\dot{z} &= -\theta_2 y_1 u_2 - K(y_1^2 y_3 + z) \\
r &= z + y_1^2 y_3
\end{aligned}$$

which is identical to (17) modulo the a_1 scaling factor that was introduced in (17) to achieve unit DC gain. In the consistency relation based design it was stated that with a non-integer A, a complex residual generator was obtained. Careful analysis of the observer based design here shows that exactly the same difficulty appears

[PI01]. But in this example, due to physical reasons (the radius can never be less than 0) this will never happen.

A closer look at the quotient system (20) reveals that it describes two consistency relations. The first equation, if substituting the sensor values, simplifies to

$$\dot{y}_2 = \frac{y_1^2 y_3}{y_1^2} = y_3$$

which is one of the consistency relation found in the beginning of this section. More interesting is the second equation in the quotient system which resolves to

$$\frac{d}{dt}\left(y_1^2 y_3\right) = \theta_2 y_1 u_2 + \theta_2 y_1 f$$

which is not the consistency relation (15), but rather the transformed consistency relation $\alpha(u)c(u, f)$. Furthermore, the state variable z_{12} is nothing but the function λ, a solution to the partial differential equation (8). This simple way of using the quotient system to derive consistency relations was of course only possible here because the measurement equation (20) was invertible which is not possible in a general case. However, this rises questions about possible connections between transformed consistency relations $\alpha(u)c(u) = 0$ and the quotient system.

The above comparison indicates that there may exist some links between the consistency relation approach presented here and the observer based approach from [PI00; PI01]. Since only first order consistency relations are fully explored here, further research is needed before a more fundamental comparison can be made. Now, to conclude the comparison a few general remarks, inspired by the simulation study, on differences and similarities between the approaches and also some open questions for further research.

The original plant model in [PI01] is assumed to be control-affine with respect to control inputs, faults, and disturbances. Furthermore it is assumed that no signals to be decoupled appear in the measurement equation of the model. The latter is only a problem when the signal affects *both* the measurement equation and the dynamic equations. Decoupling of a signal that *only* affects a measurement equation is easily achieved by temporarily removing the measurement from the model. For consistency relations, no such restrictions are needed and any model descriptions including differential algebraic descriptions are handled in the same way.

It is desirable that the residual generator does not depend on the accuracy of more model equations than necessary. However, the order of the quotient system is in general the number of states in the original model minus the dimension of the minimal unobservability distribution containing the disturbance. In general this does not render a minimal order residual generator and this is a property shared with its linear counterpart [Mas86; MVW89] which is also not able to produce minimal order residual generators. On the other hand, consistency relations are well suited to find local relationships in the model of considerably lower order than the complete model. For linear systems this is illustrated in for example [FN01].

In addition to the problem with unknown time differentiated signals addressed here, the complexity of the relation tend to grow in size very quickly with model size and the number of signals to decouple. Further research on how the complexity of residual generators grow with increasing model size is an open question and also if this is less of a problem in non-linear observer based approaches or if they also share this trait.

The existence conditions in Theorem 4.1 is straightforward to verify. The corresponding conditions for the observer approach are on the other hand rather technical. This is especially true for the conditions that ensures strong observability of the quotient system. Since the residual generator studied here does not constitute an observer, no observability conditions have to be fulfilled. Also, since the linear residual generator dynamics always can be chosen, stability is easily ensured. Note that this is only proven for first order consistency relations. For higher order consistency relations the residual generator dynamics may not arbitrary and it may very well be so that the requirements on the dynamics for higher order consistency relations are difficult to satisfy when at the same time requiring the dynamics to be stable. This is also a topic for further research.

Perhaps the most important question is to determine if and how restrictive the proposed approach is compared to the observer based approach? Perhaps there exists, possibly for a class of non-linear systems, a similar equivalence result between consistency based and observer based residual generators as in the linear case? This is of course impossible to say before the consistency based approach outlined here is generalized to consistency relations of any order.

6. Concluding Remarks

How to use non-linear consistency relations to compute a residual for fault detection and isolation has been considered. A major problem concerning this is that for dynamic systems the consistency relation generally includes time differentiated versions of known variables and these are generally not known. For linear systems, this is easily circumvented using linear realization theory and a residual can be computed where no derivatives are needed. The aim of this report is to extend the linear approach also for non-linear systems. The restrictions for when this is possible will of course not always be fulfilled, therefore the procedure should in such a case lower the derivatives as much as possible. Also, a main objective has been to rely only on operations for which constructive algorithms exist and for which strong support in computer algebraic tools exists.

This is done by considering a class of residual generators (2), where the original consistency relation can be multiplied by a suitable function $\alpha(u)$ called integrating factor and arbitrary residual generator dynamics are introduced. Here, only the problem of lowering the input derivatives one step is solved and Theorem 4.1 gives necessary and sufficient conditions for when there exists an integrating factor and residual generator dynamics such that the orders of input derivatives are lowered one step. Also, the sufficiency part of the proof provide a constructive algorithm how to compute a solution. It is noteworthy that in the solution, *any* residual generator dynamics is possible. This further means that the residual generator dynamics always can be chosen linearly which greatly simplifies stability analysis and the choice of design freedom.

The restriction of the class of residual generators was to be able to derive a useful and general solution. One can consider more general transformations of the consistency relation than just multiplication of an arbitrary integrating factor. The results here at least give a necessary conditions for a general transformation, i.e. the consistency relation must be transformed such that the highest order input derivatives appear affinely.

In the design procedure, the integrating factor α is computed indirectly via the solution of a partial differential equation from whose solution α can be computed. The advantage of this approach is that instead of solving for α directly using a general partial differential equation, a special type of system of first order partial differential equations is solved. For this class of PDE:s, necessary and sufficient solvability conditions exist together with a constructive procedure to compute solutions.

The approach is illustrated on a 4:th order non-linear model where a residual generator is designed based on a first order consistency relation for the model. In the report, a general expression of the fault response in the residual generator is derived. This fault response is validated and it is also shown how the fault response can be used to estimate the fault size. A small comparison is also included where the design is compared to the geometric approach to fault detection. Possible advantages and disadvantages with respective method it discussed. The comparison also show that, at least in this example, there exists interesting connections between the approaches which are interesting to explore in further research.

Appendix A. Notation and Results Used

A.1. **Frobenius theorem.** Below follows a brief description of a useful result when solving a particular type of partial differential equation. Readers not familiar with basic differential geometric concepts such as distributions are referred to for example [Isi95].

Theorem A.1 (Frobenius). *A nonsingular distribution is completely integrable if and only if it is involutive.*

Proof. See e.g. [Isi95]. ∎

That a non-singular distribution $\Delta(x)$ of dimension d is completely integrable is in [Isi95] stated to be equivalent to the existence of $n-d$ independent solutions to the partial differential equation

$$\frac{\partial \lambda}{\partial x} F(x) = 0 \tag{21}$$

This type of differential equation is shown to be useful when computing state transformations and integrating factors in the main part of this report. Therefore, now follows a brief review of the constructive procedure how to compute the solution to this partial differential equation. This presentation follows [Isi95] closely.

Let the distribution in question be denoted $\Delta(x)$. Since it is non-singular and of dimension d we know that it is (locally) spanned by d smooth vector-fields, i.e.

$$\Delta(x) = \text{span}\ \{f_1(x), \ldots, f_d(x)\}$$

Let $f_{d+1}(x), \ldots, f_n(x)$ be a complementary set of vector fields such that

$$\text{span}\ \{f_1(x), \ldots, f_n(x)\} = \mathbb{R}^n$$

Let $\Phi_t^f(x)$ denote what is called the flow of the vector field $f(x)$. The flow is the function $x(t) = \Phi_t^f(x^0)$ that solves the ordinary differential equation

$$\dot{x} = f(x)$$

with initial condition $x(0) = x^0$. Compute the flows for all vector fields $f_1(x), \ldots, f_n(x)$ and define the mapping:

$$\Psi(z_1, \ldots, z_n; x^0) = \Phi_{z_1}^{f_1} \circ \Phi_{z_2}^{f_2} \circ \cdots \circ \Phi_{z_n}^{f_n}(x^0)$$

Then it can be proven that the inverse of $\Psi(z)$ exists and is smooth, let

$$\phi(x) = \{\phi_i(x)\}_{i=1,n} = \Psi^{-1}(x)$$

Furthermore, it can also be proven that $\phi_{n-d+1}(x), \ldots, \phi_n(x)$ are $n-d$ independent solutions to the differential equation (21). This means that for solving (21) we need to extend the distribution with $n-d$ vector fields, solve n ordinary differential equations, and compute the (proven existing) inverse of $\Psi(z)$. The necessary operations to solve (21) can be easily automated in a symbolic computer algebraic tool. The automated solution of the PDE is of course limited by the ability of the tool to solve ordinary differential equations and some inverses. However, these problems are significantly simpler than solving general PDE:s.

References

[CLO96] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms - an introduction to computational algebraic geometry and commutative algebra*, second ed., Springer Verlag, 1996.

[DR95] E. Delaleau and W. Respondek, *Lowering the orders of derivatives of controls in generalized state space systems*, Journal of Mathematical Systems, Estimation, and Control **5** (1995), no. 3, 1–27.

[FN01] E. Frisk and M. Nyberg, *A minimal polynomial basis solution to residual generation for fault diagnosis in linear systems*, Automatica **37** (2001), 1417–1424.

[Fri01] E. Frisk, *Residual generation for fault diagnosis*, Ph.D. thesis, Linköping University, Sweden, 2001.

[FÅ03] E. Frisk and J. Åslund, *Lowering orders of derivatives in non-linear consistency relations for fault diagnosis using realization theory*, Submitted to Automatica (2003).

[Ger91] J. Gertler, *Analytical redundancy methods in fault detection and isolation; survey and synthesis*, IFAC Fault Detection, Supervision and Safety for Technical Processes (Baden-Baden, Germany), 1991, pp. 9–21.

[Isi95] A. Isidori, *Nonlinear control systems*, 3rd ed., Springer Verlag, 1995.

[Mas86] M.A. Massoumnia, *A geometric approach to the synthesis of failure detection filters*, IEEE Transactions on Automatic Control **31** (1986), 839–846.

[MVW89] M.A. Massoumnia, G.C. Verghese, and A.S. Willsky, *Failure detection and identification*, IEEE Transactions on Automatic Control **34** (1989), 316–321.

[PI00] C. De Persis and A. Isidori, *On the observability codistributions of a nonlinear system*, System and Control Letters **40** (2000), 297–304.

[PI01] ——— , *A geometric approach to nonlinear fault detection and isolation*, IEEE Trans. on Automatic Control **46** (2001), no. 6, 853–865.

Department of Electrical Engineering, Linköping University, SE-584 31 Linköping, Sweden

E-mail address: {frisk,jaasl}@isy.liu.se