Automatica 45 (2009) 364–371

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Sensor placement for fault isolation in linear differential-algebraic systems☆

Erik Frisk*, Mattias Krysander, Jan Åslund

Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden

ARTICLE INFO

Article history:
Received 27 November 2007
Received in revised form
22 May 2008
Accepted 28 August 2008
Available online 17 December 2008

Keywords:
Fault isolation
Diagnosis
Sensor placement
Linear differential-algebraic equations

ABSTRACT

An algorithm is proposed for computing which sensor additions make a diagnosis requirement specification regarding fault detectability and isolability attainable for a given linear differential-algebraic model. Restrictions on possible sensor locations can be given, and if the diagnosis specification is not attainable with any available sensor addition, the algorithm provides the solutions that maximize specification fulfillment. Previous approaches with similar objectives have been based on the model structure only. Since the proposed algorithm utilizes the analytical expressions, it can handle models where structural approaches fail.

1. Introduction

Systematic methods for fault diagnosis and process supervision are important in many industrial applications. To be able to perform model based supervision, some redundancy is needed and this redundancy can be provided by mounting sensors on the process together with a model description of the process behavior. Scientific attention has mainly been devoted to design of a diagnosis system given a model of a process equipped with a fixed set of sensors. Not as much attention has been devoted to decide which sensors to include in the process. The topic of this paper is to decide where to put sensors so that a given fault isolation performance specification is attainable, based on a differential-algebraic model.

An example of related previous work is Basseville, Benveniste, Moustakides, and Rougé (1987) where sensor location for optimal detection performance is studied. In Debouk, Lafortune, and Teneketzis (2002) a minimal cost solution is sought, given a pre-specified algorithm for determining if a set of sensors achieves a desired fault isolation performance specification. The objective in this paper is to compute all sensor additions that achieve a specified isolability requirement from a linear differential-algebraic model, and this is the main contribution compared to previous works. Other related works are Commault, Dion, and Agha (2006), Raghuraj, Bhushan, and Rengaswamy (1999) and Travé-Massuyès, Escobet, and Olive (2006) who all have a similar objective but, contrary to this paper, utilize a structural description of the model instead of the analytical equations. In Wang, Song, and Wang (2002), the structural method for the sensor selection proposed in Raghuraj et al. (1999) is used together with monitoring techniques based on Principal Component Analysis (PCA). Since the algorithm developed in this paper utilizes the analytical model, it can handle models where structural approaches do not give a complete answer and provide a weaker result.

The basic principles of the algorithm developed in this paper are the same as in the structural algorithm presented in Frisk and Krysander (2007) and Krysander and Frisk (2008). The objective here is the same, but since we now consider analytical models, other theoretical tools have to be applied and basic algorithmic steps are fundamentally different. The motive for this analytical approach is that structural methods might give incorrect answers for some models. The example in Section 5 is taken from Murota (2000) where the reasons for shortcomings of structural methods are investigated. References Frisk and Krysander (2007) and Krysander and Frisk (2008) includes a discussion on how different structural approaches for sensor placement relate to each other. This is also relevant here due to the relationship between the structural work in Frisk and Krysander (2007) and Krysander and Frisk (2008) and the analytical work here.

The proposed solution is straightforward to implement in software packages like Mathematica, Maple, or Matlab, and a Mathematica implementation can be downloaded from the authors' website http://www.fs.isy.liu.se/Software/LinSensPlaceTool/.

☆ A brief version covering parts of the material in this paper has been submitted to IFAC World Congress 2008. This paper was recommended for publication in revised form by Associate Editor Michele Basseville under the direction of Editor Torsten Söderström.

* Corresponding author. Tel.: +46 13285714; fax: +46 13282035.
E-mail addresses: frisk@isy.liu.se (E. Frisk), matkr@isy.liu.se (M. Krysander), jaasl@isy.liu.se (J. Åslund).

0005-1098/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2008.08.013

2. Problem formulation

Before the main objective of the paper is formally presented, a small example is discussed that illustrates fundamental problems in sensor placement for fault diagnosis. The example is modeled by a fifth order linear system of ordinary differential equations. This example will be used throughout the paper and consists of the following equations:

$$\begin{aligned} e_1&: \quad \dot{x}_1 = -x_1 + x_2 + x_5 \\ e_2&: \quad \dot{x}_2 = -2x_2 + x_3 + x_4 \\ e_3&: \quad \dot{x}_3 = -3x_3 + x_5 + f_1 + f_2 \\ e_4&: \quad \dot{x}_4 = -4x_4 + x_5 + f_3 \\ e_5&: \quad \dot{x}_5 = -5x_5 + u + f_4 \end{aligned}$$

where x_i are the state variables, u a known control signal, and f_i the faults we want to detect and isolate.

Faults are modeled by fault signals that are included in the model equations and $f_i \neq 0$ indicates a fault. From now on f_i will be used to denote both the fault signal and the corresponding fault mode. Let $\mathcal{F}$ denote the set of faults. A detectability performance specification is then a set $\mathcal{F}_{det} \subseteq \mathcal{F}$ specifying the detectability requirement and an isolability requirement is a set $\mathcal{I}$ of ordered pairs $(f_i, f_j) \in \mathcal{F} \times \mathcal{F}$, meaning that f_i is isolable from f_j. If a fault f_i is isolable from f_j then f_i is also detectable. Thus, without loss of generality, it is assumed that

$$(f_i, f_j) \in \mathcal{I} \Rightarrow f_i \in \mathcal{F}_{det}.$$

To fulfill a fault isolability specification, we later show that it may be necessary to add more than one sensor measuring the same variable. Therefore, when sets of sensors are considered, multisets are used. A multiset is a set where multiple instances of a member are allowed. Generalizations of the standard set operations like union and intersection are straightforward.

Since the fault isolability capability always increases when adding new sensors, there are minimal elements in the family of sensor sets that achieve a certain level of fault isolability. Therefore, we define *minimal sensor set* as a minimal set of sensors to add, to achieve a specified performance specification.

Definition 1 (*Minimal Sensor Set*). Let $\mathcal{P}$ be the set of possible sensor locations (i.e. the set of measurable variables) and let S be a multiset defined on $\mathcal{P}$. Given a detectability and isolability specification, S is a minimal sensor set if the specification is fulfilled when the sensors in S are added, but not fulfilled when any proper subset of S is added.

These minimal sensor sets are of interest for at least two reasons. It is quite naturally of interest (e.g. for economic reasons) to minimize the number of sensors to mount on a process. Also, since all supersets of a minimal sensor set also achieve the specified isolability performance, the minimal sensor sets characterize all sensor sets for a given specification. This may be of interest since a minimal sensor set may not, when evaluating a design on a real process, give the required false alarm or detection probability. Thus, more sensors, in addition to a minimal sensor set, may be required to increase, for example, detection performance, to a required level.

Returning to the example, a first question is, then, what are the minimal sensor sets achieving detectability of all faults? Here it is assumed that sensors measure a state-variable or a function thereof. It can be shown, using conditions for fault detectability in linear systems (see e.g. Nyberg (2002)) that $\{x_1\}$, $\{x_2\}$, $\{x_3, x_4\}$ are minimal sensor sets achieving detectability.

Another requirement is to not only require detectability, but also isolability properties. Here, isolability refers to isolability as it is commonly used in FDI and the consistency based diagnosis AI community, see e.g. Cordier et al. (2004). See Section 3 for details on how isolability is defined in this paper. It can be shown that there are 5 minimal sensor sets that achieve maximal fault isolation: $\{x_1, x_3\}$, $\{x_1, x_4\}$, $\{x_2, x_3\}$, $\{x_2, x_4\}$, and $\{x_3, x_4\}$. Thus, adding sensors measuring all the variables in any of these sets, or a superset of the variables, achieves maximum fault isolability.

Now, it is of course the case that the new sensors may also become faulty. If we also want faults in the new sensors to be isolable from the other faults, we may have to add additional sensors. In this case, if maximum fault isolability is also desired for faults in the new sensors, there are 9 minimal sensor sets where one sensor set is two sensors measuring x_1 and one for x_3, i.e. the multiset $S = \{x_1, x_1, x_3\}$ is a minimal sensor set. The example has illustrated some aspects of the sensor placement problem which leads to the problem formulation of the paper which is stated as:

> Given a model, possible sensor locations, and a detectability/isolability performance specification, find all minimal sensor sets with respect to the required specification. In case the specification is not feasible, minimal sensor sets that achieve maximum performance should be found.

Note that the specification may also specify if faults in new sensors should be detectable and isolable. The method developed in the sections that now follow, addresses this problem for general linear differential-algebraic models.

3. Detectability and isolability analysis

This section will formally introduce the model class used in the paper and state some basic results on fault detectability and fault isolability for linear systems that will be used in the development of the algorithm. The results in this section are primarily based on the presentation in Nyberg and Frisk (2006) that uses a model description similar to what is used in the behavioral approach to systems theory (Polderman & Willems, 1997). However, equivalent results can be derived for any other linear model description.

3.1. The model

The class of models considered, is written as

$$H(p)x + L(p)z + F(p)f = 0 \tag{1}$$

where $x(t) \in \mathbb{R}^{n_x}$, $z(t) \in \mathbb{R}^{n_z}$, $f(t) \in \mathbb{R}^{n_f}$. The matrices $H(p)$, $L(p)$, and $F(p)$ are polynomial matrices in the differentiation operator p. If discrete time systems are considered, the differentiation operator can be replaced by the time shift operator. The vector x contains all unknown signals, which include internal system states and unknown inputs. The vector z contains all known signals such as control signals and measured signals, and the vector f contains the fault-signals. Let the sets $\mathcal{X}$, $\mathcal{Z}$, and $\mathcal{F}$ represent the set of unknown variables, known variables, and fault variables, respectively.

The theoretical development in this paper will be done under two mild assumptions on the model (1). The first assumption states that if there exists a solution $x(t)$ to the model Eq. (1), given a fault $f(t)$ and an observation $z(t)$, then $x(t)$ is unique. In polynomial algebra this translates into that matrix $H(s)$ has full column rank where s is a complex variable. Throughout the text, p will be used in the matrices when they are considered as operators and s is used when operations from polynomial matrix theory is used. It is not restrictive to assume that $H(s)$ has full column rank, since any complete physical model will, given an initial condition, have a unique solution. The second assumption is that, for any given fault signal $f(t)$ there exist signals $z(t)$ and $x(t)$ consistent with the model Eq. (1), i.e. the model imposes no restrictions on feasible $f(t)$. Formally this is equivalent to that for all columns $F_i(s)$ in $F(s)$, it holds that

$$F_i(s) \in \operatorname{Im}\begin{bmatrix} H(s) & L(s) \end{bmatrix}. \tag{2}$$

Example 1. As an example, consider a model given by the following descriptor equations:

$$E\dot{w} = Aw + B_u u + B_d d + B_f f \tag{3a}$$

$$y = Cw + D_u u + D_d d + D_f f \tag{3b}$$

where y is the vector of existing outputs, u the inputs, w the unknown state–space variable, d unknown disturbances to be decoupled, and f the faults. Letting $E = I$ in the equations above, an ordinary state–space description is obtained. In general, E can be singular and even non-square. Also, there is no requirement that (3a) is regular, i.e. that the matrix pencil $sE - A$ is invertible.

In a sensor placement analysis, there is a need to define possible sensor locations. Here the convention is used that possible sensors measure single variables in the set of unknown variables $\mathcal{X}$. For cases where there are possible sensors that measure a linear function of more than one variable, include the equation

$$y_p = C_p w$$

and add y_p to the set of unknown variables. In matrix form, the model equations become

$$\begin{bmatrix} 0 & -(pE - A) & B_d \\ 0 & C & D_d \\ I & -C_p & 0 \end{bmatrix} \begin{pmatrix} y_p \\ w \\ d \end{pmatrix} + \begin{bmatrix} 0 & B_u \\ -I & D_u \\ 0 & 0 \end{bmatrix} \begin{pmatrix} y \\ u \end{pmatrix} + \begin{bmatrix} B_f \\ D_f \\ 0 \end{bmatrix} f = 0$$

where $\mathcal{X}$ is the set of variables in (y_p, w, d) and possible sensor locations are a subset of these variables. ◇

3.2. Basic results on detectability and isolability

It will be convenient to define the set of observations z that is consistent with different behavioral modes. For example, the set of observations consistent with the fault-free model is written as

$$\mathcal{O}(NF) = \{z | \exists x : H(p)x + L(p)z = 0\}. \tag{4}$$

The observations consistent with the case of fault mode f_i are the observations where there exists a fault signal representing fault i, here denoted by g to avoid notational mix-up with the fault mode f_i, and a signal x such that the model is consistent, i.e.

$$\mathcal{O}(f_i) = \{z | \exists x, g : H(p)x + L(p)z + F_i(p)g = 0\}.$$

With this notation, a definition on detectability is immediate.

Definition 2. Fault f_i is detectable in (1) if

$$\mathcal{O}(f_i) \not\subseteq \mathcal{O}(NF). \tag{5}$$

Although intuitive, a detectability condition directly related to the model matrices is given next.

Theorem 1. *Fault f_i is detectable in (1) if, and only if,*

$$F_i(s) \notin \operatorname{Im} H(s).$$

This result is proved in Nyberg and Frisk (2006) and this is the formal step where condition (2) is needed.

Detection is a special case of isolation, i.e. a fault is detectable if the fault is isolable from the no-fault mode. By noting this similarity, the following definition is natural.

Definition 3. Fault f_i is isolable from fault f_j in (1) if

$$\mathcal{O}(f_i) \not\subseteq \mathcal{O}(f_j). \tag{6}$$

Similarly as for detectability, a condition for fault isolability directly related to the model matrices is given by

Theorem 2. *Fault f_i is isolable from fault f_j in (1) if, and only if,*

$$F_i(s) \notin \operatorname{Im}\begin{bmatrix} H(s) & F_j(s) \end{bmatrix}. \tag{7}$$

Proof. The result follows from Theorem 1 and observing that

$$\mathcal{O}(f_j) = \left\{ z | \exists x, g : [H(p)\ F_j(p)] \begin{pmatrix} x \\ g \end{pmatrix} + L(p)z = 0 \right\}$$

which is in the form (4) with $H(p)$ replaced by $[H(p)\ F_j(p)]$. □

Note that both detectability and isolability are defined as model properties and not as properties of a given set of residual generators. Later in the paper, we will use the notion that fault isolability on the set of detectable single faults is a symmetric relation, and this is proved next.

Corollary 1. *Let fault f_i and f_j be two detectable faults. Fault f_i is isolable from fault f_j if, and only if, fault f_j is isolable from fault f_i.*

Proof. Assume that $F_i(s) \in \operatorname{Im}[H(s)\ F_j(s)]$, i.e. there exist rational functions $x_1(s)$ and $x_2(s)$ such that

$$F_i(s) = H(s)x_1(s) + F_j(s)x_2(s).$$

Since f_i is detectable, $x_2(s) \not\equiv 0$ according to Theorem 1 and

$$F_j(s) = F_i(s)x_2^{-1}(s) - H(s)x_2^{-1}(s)x_1(s).$$

The above proves that $F_i(s) \in \operatorname{Im}[H(s)\ F_j(s)]$ implies that $F_j(s) \in \operatorname{Im}[H(s)\ F_i(s)]$ and, by symmetry, the converse implication follows analogously. □

Note that this result also implies that the relation that a fault is *not* isolable from another fault, is a symmetric relation. The relationship is also reflexive since a fault is trivially not isolable from itself. Below we show that this relation is also a transitive relationship. This means that the relation is an equivalence relation and that the faults can be partitioned into sets such that two faults are isolable if, and only if, they belong to different sets. For related results for structural models, see Krysander, Åslund, and Nyberg (2008).

Corollary 2. *Let fault f_i, f_j, and f_k be detectable faults. If f_i is not isolable from f_j and f_j is not isolable from f_k then f_i is not isolable from f_k.*

Proof. The result follows from Theorem 2 and the observation that if $F_i(s) \in \operatorname{Im}[H(s)\ F_j(s)]$ and $F_j(s) \in \operatorname{Im}[H(s)\ F_k(s)]$ then $F_i(s) \in \operatorname{Im}[H(s)\ F_k(s)]$. □

4. Sensor placement analysis

Theoretical results and an algorithm to solve the problem posed in Section 2 are formulated here. In Sections 4.1–4.3 sensor placement for achieving maximum detectability and isolability is considered, not taking into consideration that the added sensors may fail. The approach is extended in Section 4.4 to handle the possibility that new sensors also may become faulty. In Section 4.5 the maximum isolability requirement is replaced by a specification of a desired fault isolation performance.

4.1. Sensor placement for detectability

A basic building block in the final algorithm will be to find all minimal sensor sets that achieve detectability of faults in a set of equations where the matrix $H(s)$ in (1) has full column rank. A key step in determining which sensors to add is formalized in the following lemma in a constructive and algorithmic fashion.

Corollary 3. *Make the same assumptions as in Lemma 1, let* $W(s) = [H(s) \;\; -F_i(s)] = \sum_{i=0}^{n} W_i s^i$ *and* ρ_j *denote the column degree of the j:th column. Then there exists an integer* $k \leq \sum \rho_j$ *such that the solution to*

$$\begin{bmatrix} W_0 & 0 & \dots & 0 \\ W_1 & W_0 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ W_n & \vdots & \ddots & 0 \\ 0 & W_n & \ddots & W_0 \\ \vdots & 0 & & \vdots \\ 0 & \cdots & 0 & W_n \end{bmatrix} \begin{bmatrix} b_0 \\ a_0 \\ \vdots \\ b_k \\ a_k \end{bmatrix} := \tilde{W}_k \tilde{X}_k = 0 \tag{11}$$

gives the polynomials $a(s) = \sum a_i s^i$ *and* $b(s) = \sum b_i s^i$ *that define the unique solution* $X(s) = b(s)/a(s)$ *to* $H(s)X(s) = F_i(s)$. *Measuring any of the variables in the set* $\{x_j \in \mathcal{X} | X_j(s) \neq 0\}$ *then gives detectability of fault* f_i.

Proof. If the numerator of the solution $X(s)$ is $b(s)$ and the denominator $a(s)$, the equation $H(s)X(s) = F_i(s)$ can be reformulated as

$$W(s) \begin{bmatrix} b(s) \\ a(s) \end{bmatrix} = 0. \tag{12}$$

By assumption, we know that there exists a unique solution $b(s)$ and $a(s)$ and it is easily shown that $b(s)$ and $a(s)$ has degree $k \leq \sum \rho_j$. If n is the degree of matrix $W(s)$, equation (12) can be written as

$$[I \; sI \; \cdots s^{k+n} I] \tilde{W}_k \tilde{X}_k = 0.$$

Since this should hold for all s, the coefficients b_i and a_i (i.e. vector $\tilde{X}_k$) can be directly computed using only constant matrix operations. The non-zero elements in $X(s)$ then directly correspond to non-zero elements in $b(s)$ and thereby to non-zero elements in the coefficient matrices b_i. Direct application of Lemma 1 then ends the proof. □

For a typical case, for example a descriptor model in the form (3), the degree n is 1 and the upper limit of k is given by the model order. Since the sum of column-degrees in the matrix $W(s)$ is an upper limit on the degree of the solution, the solution may have a significantly lower degree. Therefore it may numerically be a good idea to start with $k = 0$ and increase k until a non-empty null-space in (11) is found.

4.2. Sensor placement for isolability of detectable faults

The previous section derived conditions and algorithms for how to find sensor sets that make undetectable faults detectable. This section continues by describing the basic ideas of how to find the minimal sensor sets such that maximum single fault isolability is obtained under the assumption that all faults are detectable. In the next section this assumption will be removed.

The problem of achieving maximum isolability of the set of single faults $\mathcal{F}$ can be divided into $|\mathcal{F}|$ sub-problems, one for each fault, as follows. For each fault $f_j \in \mathcal{F}$, find all measurements that make the maximum possible number of faults isolable from f_j. The solution to the isolability problem will then be obtained by combining the results from all sub-problems. The following example will illustrate the main principle.

Example 3. In Section 4.1 it was shown that $\{x_1\}$ is one of the minimal sensor sets that achieves detectability of all faults in the example from Section 2. Thus, by adding the equation

$$e_6 : \; y_1 = x_1$$

to the model, all faults become detectable. However, with only this sensor, none of the faults are isolable from each other. This example will illustrate a procedure of how to use the results in Section 4.1 to also achieve fault isolability for the model consisting of equations $e_1, \dots, e_6$.

As stated above, the analysis can be divided into $|\mathcal{F}|$ sub-problems, where each sub-problem is to make as many faults as possible, isolable from a specified fault. This procedure can then be iterated for each fault to achieve maximum fault isolability.

The symmetry result in Corollary 1 can be used to simplify the procedure since there is no need to compute detectability sets for faults already treated in previously handled sub-problems. Thus, in this case with 4 faults, the first sub-problem is to isolate f_2, f_3, and f_4 from f_1. The symmetry now gives that isolability of f_1 from f_2 is already treated and thus the second sub-problem is to isolate f_3 and f_4 from f_2. The third and final sub-problem is to isolate f_4 from f_3. The detectability sets in each sub-problem are then collected to compute the minimal sensor sets.

Now, let us consider the first sub-problem; to find sensors that achieve maximum fault isolability from fault f_1. Based on Theorem 2, this is done by achieving detectability of the maximum number of faults when matrix $H(s)$ is redefined as $[H(s) \; F_1(s)]$. Thus, for the first sub-problem we have

$$H(s) = \begin{bmatrix} s+1 & -1 & 0 & 0 & -1 & 0 \\ 0 & s+2 & -1 & -1 & 0 & 0 \\ 0 & 0 & s+3 & 0 & -1 & -1 \\ 0 & 0 & 0 & s+4 & -1 & 0 \\ 0 & 0 & 0 & 0 & s+5 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{13}$$

For the remaining faults $\{f_2, f_3, f_4\}$, the detectability sets are computed using the algorithm in Section 4.1 as

$$D(f_2) = \emptyset, \qquad D(f_3) = \{x_3, x_4\}, \qquad D(f_4) = \{x_2, x_3, x_4, x_5\}.$$

The detectability set for f_2 is empty because no addition of sensors will make f_2 isolable from f_1 which is due to both faults influencing the model in exactly the same way. This also implies that the second sub-problem (i.e. finding sensors that achieve maximum fault isolability from fault f_2) gives identical detectability sets for f_3 and f_4.

In the third sub-problem, where f_3 is considered to be an unknown signal, only the detectability set for f_4 is needed. Calculations give that

$$D(f_4) = \{x_2, x_3, x_4, x_5\}.$$

Now that the final sub-problem has been considered, the results are collected to compute the sensor sets from the detectability sets. The minimal hitting sets for the family of all non-empty detectability sets obtained in all sub-problems are $\{x_3\}$ and $\{x_4\}$ which are also the sensor sets that achieve maximum fault isolability for the system defined by equations $e_1, \dots, e_6$. ◇

Now follows a formalization of the above procedure. For this, let $M(f_j)$ denote the model that is obtained by decoupling fault f_j, i.e. column F_j is moved from matrix $F(s)$ to $H(s)$ as was done in (13) in the example.

Theorem 4. *Assume that all faults in* $\mathcal{F}$ *are detectable in the model* M. *Let* $\mathcal{P} \subseteq \mathcal{X}$ *be the set of possible sensor locations and* M_S *the equations corresponding to adding the set of sensors* S. *For an arbitrary fault* f_j, *the maximum possible number of faults* $f_i \in \mathcal{F} \setminus \{f_j\}$ *are isolable from* f_j *in* $M \cup M_S$ *if, and only if,* S *has a non-empty intersection with all sets in* $\mathcal{D} = \texttt{Detectability}(M(f_j), \mathcal{F} \setminus \{f_j\}, \mathcal{P})$.

Proof. Assume that $D(f_i) \in \mathcal{D}$ and $S \cap D(f_i) = \emptyset$. This means that fault f_i is not isolable from fault f_j. But since $D(f_i) \neq \emptyset$, S can be extended so that $S \cap D(f_i) \neq \emptyset$. Hence, maximal fault isolability from f_j implies that S has non-empty intersection with all sets in $\mathcal{D}$.

Conversely, if S has a non-empty intersection with all elements in $\mathcal{D}$, then according to Theorem 3, maximum number of faults are detectable in $M(f_j) \cup M_S$ which means that maximum number of faults are isolable from f_j in $M \cup M_S$. □

The above result gives the solution for one sub-problem, i.e. how to place sensors such that faults are isolated from a specified fault f_j. How to combine the results from all sub-problems into a solution for the complete problem is summarized in the pseudo-code function, below, that returns the set of minimal sensor sets.

```
function 𝒮 = SensPlaceInDetectable(M, ℱ, 𝒫)
  𝒟 = ∅;
  for f_j ∈ ℱ
    ℱ_d(f_j) := {f_i | i > j};
    𝒟_j = Detectability(M(f_j), ℱ_d(f_j), 𝒫);
    𝒟 := 𝒟 ∪ 𝒟_j
  end
  𝒮 = MinimalHittingSets(𝒟);
```

Remember that here it is assumed that all faults in $\mathcal{F}$ are detectable, and this assumption will be lifted in the next section.

4.3. Sensor placement for both detectability and isolability

Section 4.1 described how to place sensors to achieve detectability and Section 4.2 how to achieve isolability in models where faults are detectable. The algorithms in these two sections will now be combined to achieve maximum isolability in a general model.

Example 4. Consider the example introduced in Section 2. In Section 4.1 it was shown that the minimal sensor sets that achieve detectability are

$$\{x_1\}, \{x_2\}, \{x_3, x_4\}. \tag{14}$$

In Section 4.2 the first set in (14) was chosen and a sensor measuring x_1 was added to the model. In this case, the minimal sensor sets that give maximal isolability are

$$\{x_3\}, \{x_4\}.$$

Noting that a sensor measuring x_1 was first added, the resulting sensor sets are then

$$\{x_1, x_3\}, \{x_1, x_4\}$$

which are minimal sensor sets that fulfill both the detectability and isolability specifications. In (14) it was noted that, in addition to $\{x_1\}$, there are two other sets, $\{x_2\}$ and $\{x_3, x_4\}$, that achieve detectability. The same procedure as for $\{x_1\}$ is thus iterated for these two sets to obtain all minimal sensor sets that fulfill the requirements. ◇

Given a model M that fulfills the assumptions in Section 3.1, the faults $\mathcal{F}$, and the possible sensor locations $\mathcal{P}$, the algorithm below computes the set $\mathcal{S}$ of all minimal sensor sets that achieve maximum isolability.

Faults that cannot be made detectable, cannot be made isolable from other faults and, in addition, all detectable faults are isolable from the non detectable faults. Therefore, to achieve maximum isolability it is sufficient to first achieve maximum detectability and then maximum isolability among the detectable faults.

```
function 𝒮 = SensorPlacement(M, ℱ, 𝒫)
  𝒟 = Detectability(M, ℱ, 𝒫);
  if 𝒟 = ∅
    ℱ_d = detectable faults in M;
    𝒟 = SensPlaceInDetectable(M, ℱ_d, 𝒫);
    𝒮 = MinimalHittingSets(𝒟);
  else
    𝒮 = ∅;
    𝒮_det = MinimalHittingSets(𝒟);
    for s_det ∈ 𝒮_det
      Create the extended model M_e = M ∪ M_{s_det};
      ℱ_e = the detectable faults included in M_e;
      𝒟 = SensPlaceInDetectable(M_e, ℱ_e, 𝒫);
      𝒮_isol = MinimalHittingSets(𝒟);
      𝒮 := 𝒮 ∪ {s_det ∪ s_isol | s_isol ∈ 𝒮_isol};
    end
    Delete non-minimal sensor sets in 𝒮;
  end
```

4.4. Adding sensors with faults

Until now, we have not considered the possibility that new sensors can fail, but this is of course the case in many applications. How to cope with new sensors that may become faulty will be treated next.

Example 5. Consider the example from Section 2. If new sensors are fault-free, it has been shown in Section 4.3 that a minimal sensor set achieving maximum fault isolability is $\{x_1, x_3\}$.

However, if the sensors measuring x_1 and x_3 have faults f_5 and f_6 respectively, maximum fault isolability is not achieved when considering the faults $f_1, \ldots f_4$ in the original model, and also the faults f_5 and f_6 introduced by new sensors. For example f_3 is not isolable from f_5.

By adding another sensor measuring x_1, and thereby introducing a new sensor fault f_7, maximum fault isolability is achieved when considering all faults $f_1, \ldots, f_7$. The sensor set $\{x_1, x_1, x_3\}$ is thus a minimal sensor set achieving maximum isolability when new sensors may become faulty. ◇

The following two theorems concerning detectability and isolability properties of faults in new sensors will be sufficient results for extending the algorithm to include these faults.

Theorem 5. *Let $\mathcal{X}$ be the set of unknown variables in the model M and $x_i \in \mathcal{X}$ measured with a sensor described by an equation $e \notin M$. Then, a fault in the new sensor will be detectable in $M \cup \{e\}$.*

Proof. Let $H_e(s)$ correspond to the $H(s)$ matrix for $M \cup \{e\}$ and F_e the column vector corresponding to the new sensor fault. It is straightforward to show that $M \cup \{e\}$ fulfills condition (2) and it follows from Theorem 1 that e is detectable if, and only if, $F_e \notin \operatorname{Im} H_e(s)$, i.e. the equation

$$H(s)\xi(s) = 0$$
$$\xi_i(s) = 1$$

has no solution. The result follows immediately since $H(s)$ has full column rank. □

A consequence of this result is that we need not consider sensor faults related to sensors s_{det} in the detectability step in the function `SensorPlacement` when we extend the algorithm to include faults in new sensors.

Theorem 6. *Let $\mathcal{X}$ be the set of unknown variables and $\mathcal{F}$ a set of detectable faults in the model M. Furthermore, let M_S be a set of equations describing additional sensors and $\mathcal{F}_S$ the associated set of sensor faults. Then for any sensor fault $f_i \in \mathcal{F}_S$ and for any fault $f_j \in (\mathcal{F} \cup \mathcal{F}_S) \setminus \{f_i\}$, it holds that f_i is isolable from f_j in $M \cup M_S$.*

Proof. Let x_i be a variable measured by a new sensor described by equation e and with a fault f_i. Furthermore, let f_j be an arbitrary fault in $M \cup M_S$ such that $f_j \neq f_i$ and $H(s)$ and $F_j(s)$ matrices corresponding to the equations $M \cup M_S \setminus \{e\}$. Then Theorem 2 gives that a fault $f_i \in \mathcal{F}_S$ in a new sensor is isolable from $f_j \in \mathcal{F} \cup \mathcal{F}_S \setminus \{f_i\}$ if, and only if, the set of equations

$$H(s)\xi(s) + F_j(s)f_j(s) = 0 \tag{15}$$

$$\xi_i(s) = 1 \tag{16}$$

has no solution in $\xi(s)$ and $f_j(s)$. Fault f_j is detectable since, by assumption, all faults in $\mathcal{F}$ are detectable and by Theorem 5 all faults in $\mathcal{F}_S$ are detectable in $M \cup M_S \setminus \{e\}$. It then follows that $F_j(s) \notin \operatorname{Im} H(s)$ which, together with (15), yields that $\xi(s) = 0$. This contradicts (16) which ends the proof. □

For the function `SensorPlacement`, this theorem implies that full isolability is achieved for all sensor faults introduced by the new sensors s_{isol} in the isolability step for free. Of the new sensor faults, only the faults introduced in the detectability step (i.e. faults in the sensors in s_{det}) have to be considered in the isolability step. Next, a summary of the modified procedure is given.

First the detectability step is performed as before, then new faults introduced by sensors s_{det} in the detectability step are included in the model, and finally the isolability step is performed as before. The only needed modification of `SensorPlacement` is that the new faults introduced by sensors s_{det} are included in the creation of the extended model M_e on line 11. The new faults introduced by sensors s_{det} will, in this way, be considered in the isolability step on line 13.

4.5. Fault isolability performance specification

We have discussed sensor placement for achieving detectability and maximum isolability. Since fault isolability performance is gained at the expense of adding more sensors, it is important that the algorithm can handle more precise fault isolability specifications. In Section 2 it was stated that a detectability requirement is a set $\mathcal{F}_{det} \subseteq \mathcal{F}$ and an isolability requirement is a set $\mathcal{I}$ of ordered pairs $(f_i, f_j) \in \mathcal{F} \times \mathcal{F}$, meaning that f_i is required to be isolable from f_j.

It is straightforward to modify the proposed algorithm with a detectability and isolability specification. Two modifications have to be made, one for each specification. First, on lines 1 and 2 in function `SensorPlacement`, change $\mathcal{F}$ to $\mathcal{F}_{det}$. Second, on line 4 in function `SensPlaceInDetectable`, change

$$\mathcal{F}_d(f_j) := \{f_i \in \mathcal{F} | i > j, (f_i, f_j) \in \mathcal{I} \vee (f_j, f_i) \in \mathcal{I}\}; \tag{17}$$

Note that only pairs of detectable faults are considered in `SensPlaceInDetectable`. With the same reasoning as in Section 4.3, it can be shown that this gives maximal solutions in case the specification is not feasible.

Using $\mathcal{F}_{det}$ and $\mathcal{I}$ as above, it is possible to give a detailed specification. However, it is often more natural and convenient to use other representations of the isolability specification. A simpler, but less general, specification is illustrated in the following example.

Example 6. For the example given in Section 2, assume that we want to compute sensor placements such that faults in $\{f_1, f_2\}$ are isolable from faults in $\{f_3, f_4\}$ and vice-versa, but for example fault f_3 need not be isolable from f_4. The family $\{\{f_1, f_2\}, \{f_3, f_4\}\}$ can then be used to represent the isolability specification.

It is straightforward to verify that this specification is equivalent to the isolability and detectability requirements

$$\mathcal{I} = \{(f_1, f_3), (f_1, f_4), (f_2, f_3), (f_2, f_4)\}$$
$$\mathcal{F}_{det} = \{f_1, f_2, f_3, f_4\}. \quad \diamond$$

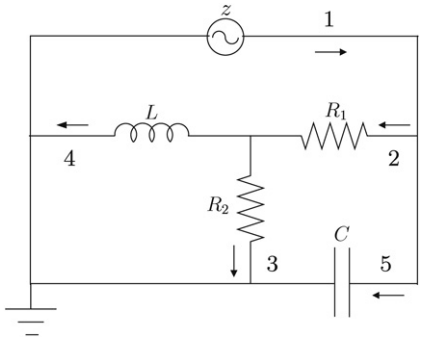

Fig. 1. An electrical circuit.

Given a model and a set of sensors, Corollaries 1 and 2 imply that detectability and isolability properties of this system always can be represented by a family of sets as the specification in the example above. The detectable faults are included in one of the fault sets and the isolability properties are interpreted as in the simplified isolability specification. This means that the simplified isolability specification is a natural way to state the requirements, but note that it is not as expressive as the more general form with pairs of isolable faults. An isolability specification in the form of family of sets of faults is easy to translate into an isolability set with pairs of faults and a corresponding modification of (17) is straightforward.

5. Example

In this section, the sensor placement algorithm will be demonstrated by applying it to the electrical circuit shown in Fig. 1. The circuit has 5 components, a voltage source $z(t)$, two resistors R_1 and R_2, an inductor L, and a capacitor C and they can fail independently of each other. The input signal $z(t)$ is assumed to be known. The branches are enumerated $k = 1, 2, \ldots, 5$ and $f_1, \ldots, f_5$ denote faults in the corresponding components. The current through branch k is i_k and the voltage across is u_k. The behavior of the fault free system is given by

$$u_1 = u_5 \qquad u_5 = u_2 + u_3 \qquad u_3 = u_4$$
$$i_1 = i_2 + i_5 \qquad i_1 = i_3 + i_4 + i_5$$
$$u_1 = z \qquad u_2 = R_1 i_2 \qquad u_3 = R_2 i_3$$
$$u_4 = L\frac{\mathrm{d}}{\mathrm{d}t} i_4 \qquad i_5 = C\frac{\mathrm{d}}{\mathrm{d}t} u_5.$$

The 5 equations, in the last two rows above, model the 5 components that may fail. The equations describe fault free behavior, and there is no equation describing a faulty component, i.e. the behavior of a faulty component is undefined.

In the first run of the sensor placement algorithm, maximum fault isolability is desired among the faults f_1 to f_5 under the condition that all currents i_k and voltages u_k can be measured. The output of the algorithm is that there are 5 minimal sensor sets that achieve full isolability. The sets $\{i_1, i_3\}$ and $\{i_1, i_4\}$ are the only minimal sensor sets with cardinality 2; all other sets contain 3 sensors.

Now, assume that all added sensors can also fail. For this case there are 7 minimal sensor sets achieving full isolability where 4 sensor sets have cardinality 3, and 3 sensor sets have cardinality 4. One of the minimal cardinality sensor sets is $\{i_1, i_1, i_4\}$, i.e. current i_1 is measured twice. For the case above where new sensors cannot fail, $\{i_1, i_4\}$ was a minimal sensor set but this does not give maximum isolability when sensor faults are considered. When only measuring i_1 once, the fault in the sensor measuring i_1 is not isolable from a fault in the capacitor. Interestingly, all minimal sensor sets include only current measurements, meaning that any voltage measurement will be superfluous.

In a third run, all inputs are the same as in the second run, with the exception that only voltages can be measured, because voltage measurements can be performed without disconnecting wires in the circuit. With this restriction, full isolability cannot be achieved as noted above. The maximum isolability is that the voltage source fault can be isolated from all other faults, faults in the resistors and in the inductor are not isolable from each other, and the capacitor fault f_5 cannot even be detected, i.e. the fault classes that can be isolated are described by the set $\{\{f_1\}, \{f_2, f_3, f_4\}\}$. There are 10 minimal sensor sets achieving this isolability specification and $\{u_2, u_3\}$ and $\{u_2, u_4\}$ are the ones with minimal cardinality.

In the fourth and final run, we input the isolability specification $\{\{f_1\}, \{f_2, f_3, f_4\}, \{f_5\}\}$, assume that all voltages and currents can be measured, and sensors do not fail. Thus, we do not require full isolability, which should imply that we may not need as many sensors as in the first run. This time there are 13 minimal sensor sets, all with cardinality 2. In this case the isolability achieved by different minimal sensor sets are not the same. For example, the set $\{i_1, i_4\}$, returned also in the first run, achieves full isolability, but, for instance, the minimal sensor set $\{i_2, i_5\}$ achieves exactly the specified isolability. Hence, some minimal sensor sets might achieve better isolability than that specified, but the retraction of any sensor in any minimal sensor set will take the isolability performance below that specified.

This example has been used in Murota (2000) to illustrate problems with structural approaches for determining the index of a Differential-Algebraic Equation (DAE). Using the structural approach for sensor placement in Frisk and Krysander (2007), a non-trivial reformulation of the model equations is needed to obtain a characterization of all sensor sets.

6. Conclusions

An algorithm has been developed that computes a characterization of all sensor additions that makes a fault isolability specification attainable for a given linear differential-algebraic model. It may be the case that the fault isolation specification is not attainable, for example due to a restriction on possible sensor locations. In such a case, the algorithm then provides solutions that are as close to the specification, with the available sensors.

Due to the exhaustive nature of the problem formulation where all solutions are characterized, the approach might suffer from some combinatorial problems. These problems are mitigated by only considering single fault isolation, and systems with no underdetermined part. It is also possible to control the complexity by limiting the set of possible sensor locations $\mathcal{P}$ or the isolability specification $\mathcal{I}$. It is also straightforward to modify the approach to only compute solutions with a cardinality less than a given threshold.

The new sensors added to make fault isolation possible may also become faulty. These additional sensor faults need to be considered in the analysis, and it has been shown that it might be necessary to add more than one sensor measuring the same variable. Since the approach is analytical, the method can handle models where structural approaches fail.

References

Basseville, M., Benveniste, A., Moustakides, G. V., & Rougé, A. (1987). Optimal sensor location for detecting changes in dynamical behavior. *IEEE Transactions on Automatic Control*, *32*(12), 1067–1975.

Commault, C., Dion, J., & Agha, S.Y. (2006). Structural analysis for the sensor location problem in fault detection and isolation In *Proceedings of IFAC safeprocess'06*.

Cordier, M. O., Dague, P., Levy, F., Montmain, J., Staroswiecki, M., & Travé-Massuyès, L. (2004). Conflicts versus analytical redundancy relations: A comparative analysis of the model based diagnosis approach from the artificial intelligence and automatic control perspectives. *IEEE Transaction on Systems, Man, and Cybernetics – Part B*, *34*(5), 2163–2177.

de Kleer, J. (1987). Diagnosing multiple faults. *Artificial Intelligence*, *32*(1), 97–130.

Debouk, R., Lafortune, S., & Teneketzis, D. (2002). On an optimization problem in sensor selection. *Discrete Event Systems: Theory and Applications*, *12*(4), 417–445.

Frisk, E., & Krysander, M. (2007). Sensor placement for maximum fault isolability. In: *Proceedings of 18th international workshop on principles of diagnosis* (pp. 106–113).

Henrion, D., & Sebek, M. (2000). An algorithm for polynomial matrix factor extraction. *International Journal of Control*, *73*(8), 686–695.

Kailath, Thomas (1980). *Linear systems*. Prentice-Hall.

Krysander, M., Åslund, J., & Nyberg, M. (2008). An efficient algorithm for finding minimal over-constrained sub-systems for model-based diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, *38*(1).

Krysander, M., & Frisk, E. (2008). Sensor placement for fault diagnosis. *IEEE Transaction on Systems, Man, and Cybernetics—Part A*, *38*(6), 1–1410.

Murota, K. (2000). *Matrices and matroids for system analysis*. Springer-Verlag, ISBN: 3-540-66024-0.

Nyberg, M. (2002). Criterions for detectability and strong detectability of faults in linear systems. *International Journal of Control*, *75*(7), 490–501.

Nyberg, Mattias, & Frisk, Erik (2006). Residual generation for fault diagnosis of systems described by linear differential-algebraic equations. *IEEE Transactions on Automatic Control*, *51*(12), 1995–2000.

Polderman, J. W., & Willems, J. C. (1997). *Introduction to mathematical systems theory: A behavioral approach*. New York: Springer Verlag.

Raghuraj, R., Bhushan, M., & Rengaswamy, R. (1999). Locating sensors in complex chemical plants based on fault diagnostic observability criteria. *AIChE*, *45*(2), 310–322.

Reiter, R. (1987). A theory of diagnosis from first principles. *Artificial Intelligence*, *32*(1), 57–95.

Travé-Massuyès, L., Escobet, T., & Olive, X. (2006). Diagnosability analysis based on component-supported analytical redundancy relations. *IEEE Transaction on Systems, Man, and Cybernetics – Part A*, *36*(6), 1146–1160.

Wang, H., Song, Z., & Wang, H. (2002). Statistical process monitoring using improved PCA with optimized sensor locations. *Journal of Process Control*, *12*(6), 735–744.

Erik Frisk was born in Stockholm, Sweden in 1971. He received a M.S. degree in 1996 and a Ph.D. degree in 2001, both from Linköping University, Sweden. Currently he has a position as an associate professor at the Department of Electrical Engineering at Linköping University. Current research interests include observer theory and model-based fault diagnosis based on linear and non-linear process models.

Mattias Krysander was born in Linköping, Sweden 1977. He received a M.S. in electrical engineering in 2000, and a Ph.D. degree in 2006, both from Linköpings University, Sweden. His current research interests in the field of model based diagnosis include fault isolation and sensor placement. Graph-theoretical and structural methods are among his specialties.

Jan Åslund was born in Boden, Sweden in 1971. He received a M.S. in mechanical engineering in 1996 and a Ph.D. in applied mathematics in 2002, both from Linköping University, Sweden. He has a position as assistant professor at the Department of Electrical Engineering at the same university. His current research interests include model-based fault diagnosis and optimal control.