Sensor placement for maximum fault isolability

Erik Frisk and Mattias Krysander
Dept. of Electrical Engineering, Linköping University
SE-581 83 Linköping, Sweden
{frisk,matkr}@isy.liu.se

Abstract

An algorithm is developed for computing which sensors to add to obtain maximum fault detectability and fault isolability. The method is based on only the structural information in a model which means that large and non-linear differential-algebraic models can be handled in an efficient manner. The approach is exemplified on a model of an industrial valve where the benefits and properties of the method is clearly shown.

1 Introduction

Fault diagnosis and process supervision is an increasingly important topic in many industrial applications and there are many publications in this area, see for example [Blanke *et al.*, 2003] and the references therein. To be able to perform model based supervision, some redundancy is needed and this redundancy is typically provided by sensors mounted on the process. Scientific attention has mainly been devoted to the design of a diagnosis system, given a model of a process equipped with a set of sensors. Not as much attention has yet been devoted to decide which sensors to include in the process.

There are many types of performance measures in diagnosis, for example detection performance, false alarm probabilities, time to detection etc. In this paper, sensors are placed such that maximum detectability and isolability is possible, i.e. faults in different components should, as far as possible, be able to be isolated from each other. Since sensor placement is often done early in the design phase, possibly before a reliable process model can be developed, the method developed in this paper is based on a structural process model. This is a coarse model description that can be obtained early and without major engineering efforts. Also, this means that large and non-linear differential-algebraic models can be handled in an efficient manner. The drawback with structural methods is that only best case results are obtained, see [Krysander, 2006] for a more on depth discussion on this.

2 Problem Formulation

Before the main objective of the paper is formally presented, a small example is discussed that illustrates the 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, although the results will be equally applicable to large scale, non-linear, differential-algebraic models. The model 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. Since there are no specified sensors there is no redundancy and the faults are not detectable.

In this example, faults are modeled by fault signals that are included in the model equations and $f_i \neq 0$ indicates a fault. A more general way to include faults is to assign assumptions, or support, to the equations. This type of fault modeling can also easily be used with the approach that will be presented later but for sake of simplicity, fault signal modeling will be used in the paper. Also, from now on only single faults will be considered and f_i will then be used to denote both the fault signal and the fault mode.

Now, define *minimal sensor set* which is a minimal set of sensors to add to achieve maximum fault isolability.

Definition 1 (Minimal sensor set) *Let $\mathcal{S}$ be the set of possible sensor locations, i.e. the set of measurable variables, and let S be a multiset defined on $\mathcal{S}$. Then S is a minimal sensor set if adding the sensors in S give maximum fault isolability and all proper subsets of S do not.*

Note that S is a multiset, which is similar to a set but allows multiple instances of a member. Generalizations of the standard set operations like union and intersection are straightforward. Multisets are used instead of regular sets since it may be necessary to add more than one sensor measuring the same variable.

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, that $\{x_1\}$, $\{x_2\}$, $\{x_3, x_4\}$ are all minimal sensor sets achieving detectability of all faults.

A second step 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]. For details on how isolability is defined in this paper, see Sections 3 and 4. 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 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 want also 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 desired also 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 $S = \{x_1, x_1, x_3\}$.

The problem formulation can now be stated as:

> Given a model and possible sensor locations, find all minimal sensor sets with the maximum possible fault isolability.

The methods developed in sections that now follow aim at addressing this problem for general, non-linear and differential-algebraic models. Doing this analytically is difficult since then inference concerning solutions to the model equations is needed. Instead, a method based on utilizing only the structure of the model is employed. This gives generic results that hold in a best-case situation. An advantage is that large models can be handled in an efficient manner. See Section 3 for some further results on the relation between structural and analytical properties of a model. See also [Krysander, 2006] for an in depth discussion on this topic.

3 Theoretical Background

The sensor placement problem will here be solved using a structural representation of the model. The structural representation of a set of equations M with unknown variables X is a bipartite graph with variables and equations as node sets. The known variables are, in this paper, omitted in the structure because they will not be needed for the analysis. There is an edge in the graph between a node representing an equation $e \in M$ and node representing an unknown variable $x \in X$ if the variable x is contained in e. A bipartite graph can be described by a biadjacency matrix where the rows and columns correspond to the node sets and the position (i, j) is one if there is an edge between node i and j, and a zero otherwise.

Structural methods can be applied to dynamical systems and the structure of the dynamic example formulated in Section 2 is shown in Figure 1 as a biadjacency matrix of the bipartite graph. The position (e_i, x_j) is one if x_j or any time-derivative appear in equation e_i. This structural representation of dynamical systems has been used in for example [Frisk *et al.*, 2003] and [Ploix *et al.*, 2005]. There exist other structural representations of dynamical systems, but the one used here is a compact representation suitable for the sensor placement problem [Krysander, 2006].

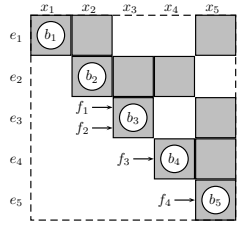


Figure 1: Structure of the linear example in Section 2. Gray areas indicate non-zero elements.

3.1 Dulmage-Mendelsohn decomposition

We will frequently use the Dulmage-Mendelsohn decomposition [Dulmage and Mendelsohn, 1958] which is illustrated in Figure 2. The decomposition defines a partition $(M_0, M_1, \ldots, M_n, M_\infty)$ of the set of equations M, a similar partition of the set of unknowns X, and a partial order on the sets M_i. If the rows and columns are rearranged according to this order, the biadjacency matrix has the form shown in Figure 2. There are zero entries in the white parts of the matrix and there might be ones in the gray-shaded parts. Three main parts of M can be identified in the partition, M_0 is called the structurally underdetermined part, $\cup_{i=1}^{n} M_i$ is the structurally just-determined part, and M_∞ is the structurally overdetermined part. In the figure, each pair (M_i, X_i) is related to a block which is denoted by b_i.

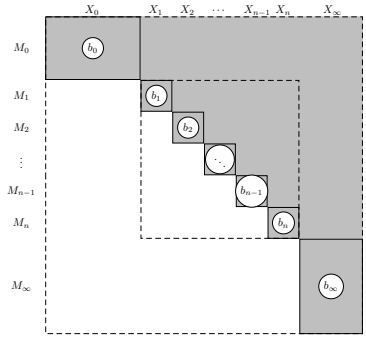


Figure 2: Dulmage-Mendelsohn decomposition

Diagnosis related treatments of Dulmage-Mendelsohn decomposition can be found in e.g. [Blanke *et al.*, 2003; Krysander, 2006].

3.2 Structural formulation of fault diagnosis

In this section, we will give structural characterizations of fault diagnosis properties. By doing this, the sensor placement problem can be formulated as a graph theoretical problem.

Let M denote a set of equations and F a set of single faults. Without loss of generality, it is possible to assume that a single fault can only violate one equation. If a fault signal f appears in more than one equation, we simply replace f in the equations with a new variable x_f and add equation $f = x_f$ which then will be the only equation violated by this fault. An example of this procedure is also given in the example in Section 5. Let $e_f \in M$ be the equation that might be violated

by a fault $f \in F$. For the example introduced in Section 2, $e_{f_1} = e_{f_2} = e_3$, $e_{f_3} = e_4$, and $e_{f_4} = e_5$.

An equation is, in the generic case, monitorable if it is contained in the structurally overdetermined part of M [Blanke *et al.*, 2003]. If the structurally overdetermined part of a set of equations M is denoted by M^+, then the structural characterization of detectability can be defined as follows.

Definition 2 *A fault f is structurally detectable in a model M if $e_f \in M^+$.*

Returning to the example and illustrating the correspondence between detectable faults and structurally detectable faults, assume that a sensor y measuring x_4 has been added to the process and included in the model by $e_6 : y = x_4$. The faults f_3 and f_4 are the detectable faults and a residual capable of detecting them is

$$r = 20y + 9\dot{y} + \ddot{y} - u = 5f_3 + \dot{f}_3 + f_4$$

which in fact is the only residual generator for this model modulo post filtering. Thus, faults f_1 and f_2 are not detectable. The structurally overdetermined part M^+ of the model $M = \{e_1, e_2, e_3, e_4, e_5, e_6\}$ is equal to $\{e_4, e_5, e_6\}$. The equations $e_{f_3} = e_4$ and $e_{f_4} = e_5$ corresponding to the detectable faults f_3 and f_4 belong to M^+, but not the equations corresponding to the other faults. This implies according to Definition 2 that the detectable faults, f_3 and f_4, are the structurally detectable faults in M which is in accordance with the analytical result above.

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, it holds that a fault f_i, isolable from f_j, can violate a monitorable equation in the model describing the behavior of the process having a fault f_j. The equations valid with a fault f_j is $M \setminus \{e_{f_j}\}$ and the monitorable part of these equations is, in the generic case, equal to $(M \setminus \{e_{f_j}\})^+$. This motivates the following structural characterization of isolability.

Definition 3 *A fault f_i is structurally isolable from f_j in a model M if $e_{f_i} \in (M \setminus \{e_{f_j}\})^+$.*

The structural detectability and isolability definitions will next be used in a structural approach for solving the sensor placement problem.

4 A Structural Approach

Theoretical results and an algorithm to solve the problem posed in Section 2 is here formulated using the theory in Section 3. Due to space limitations, all proofs are omitted but can be found in [Krysander and Frisk, 2007].

A general assumption of the approach is that the model does not contain any underdetermined part. This is not a restrictive assumption since any complete physical model will, given an initial condition, have a unique solution and thereby no underdetermined part. Without loss of generality, it is also assumed that no fault affects more than one equation and that possible sensors measure a function of one unknown variable. In case there are possible sensors that measure some function h of more than one unknown variable, include a new equation $x_{new} = h(x)$ in the model.

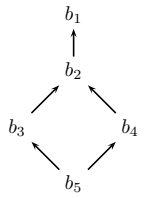


Figure 3: Hasse diagram of the partial order over the set of strongly connected components B.

4.1 Sensor placement for detectability

A basic building block in the final algorithm will be to find minimal sensor sets that achieve structural detectability of faults in an exactly determined set of equations. This section will be devoted to solving this sub-problem by first outlining the solution for the example system from Section 2 and then formally stating the solution. Although the example is given by analytical equations, all results in this section are based on the structural model only.

The example model is, without any additional sensors, an exactly determined set of equations with 5 equations and 5 unknown variables x_i, i.e. all faults are undetectable. Consider first the fault f_3. To make this fault detectable, according to Definition 2, an additional sensor is needed such that equation $e_{f_3} = e_4$ becomes a member of the overdetermined part of the model.

It is straightforward to verify that f_3 becomes detectable if and only if any of the variables $\{x_1, x_2, x_4\}$ are measured. For example, measuring x_4 makes the new measurement equation together with equations e_4 and e_5 an overdetermined set of equations. For this set of equations, a residual generator which is sensitive to fault f_3 can easily be derived. A similar line of reasoning can be made when measuring x_1 or x_2.

Then, why are x_1, x_2, and x_4 exactly those measurements that give detectability of f_3? The explanation can be seen in Figure 1 where it can be noted that block b_1 is connected with b_2 via a non-zero element in position $(1, 2)$ and in a similar fashion is b_2 connected to b_4. Thus, there is a connection between variables x_1, x_2, and x_4, which is precisely the variable in block b_4 including fault f_3. Measuring x_3, i.e. the variable in b_3, do not give detectability of f_3 since there is no connection between b_3 and block b_4.

The above reasoning indicates that some order between the strongly connected components might be useful. Let the exactly determined part of Figure 2 be the adjacency matrix of a directed graph on the set of strongly connected components b_i. A non-zero element in block (i, j) indicates a directed edge from b_i to b_j. A partial order on the blocks can then be defined as $b_i \leq b_j$ if and only if there is a path from b_i to b_j. Figure 3 shows the Hasse diagram of the ordering of the strongly connected components for the example.

With this ordering one can state exactly which parts of an exactly determined model that becomes overdetermined when adding a sensor. The following lemma formalizes the discussion above. This also gives, according to Definition 2, which faults that become detectable as a result of adding a sensor.

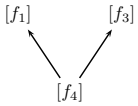


Figure 4: Hasse diagram of the partial order for the linear example over the set of fault classes. In Figure 1 it can be seen that $[f_1] = [f_2]$. Classes $[f_1]$ and $[f_3]$ are the maximal elements of the partial order.

Lemma 1 *Let M be an exactly determined set of equations, b_i a strongly connected component in M, and $e \notin M$ an equation corresponding to measuring any variable in b_i. Then*

$$(M \cup \{e\})^+ = \{e\} \cup (\cup_{j:b_j \leq b_i} M_j)$$

Achieving detectability of one fault affecting a strongly connected component immediately implies detectability of all faults affecting the same component. Therefore it makes sense to define an equivalence relation on the set of faults where all faults influencing the same strongly connected component are equivalent. A set of equivalent faults is denoted as $[f_i]$ where f_i is one element in the equivalence class. Now, based on Lemma 1, it is clear that measuring a variable in a block ordered higher than the block the fault enters achieves detectability. Let $P \subseteq X$ be a set of possible sensor locations and introduce the set

$$D([f_i]) = \{x | b_j \in B \wedge b_i \leq b_j \wedge x \in X_j \cap P\}$$

where X_j is the set of variables corresponding to block b_j, B the set of strongly connected components, and b_i the block that is influenced by the faults in $[f_i]$. The set $D([f_i])$ is thus the set of variables such that measuring *any* variable in the set achieves detectability of all faults in the equivalence class $[f_i]$.

Returning to the example and utilizing the result above, one can see that detectability of f_4 comes automatically when adding sensors to achieve detectability of either the faults in $[f_1]$ or $[f_3]$. This is because b_5 is less or equal than both b_3 and b_4 and according to Lemma 1, block b_5 is automatically included in any overdetermined set of equations when $[f_1]$ or $[f_3]$ are made detectable. This means that it is only necessary to ensure detectability for a subset of the fault classes to ensure detectability of all faults. To illustrate exactly which classes, introduce an order on the equivalence classes of F, defined as $[f_i] \leq [f_j]$ if $b_i \leq b_j$ where b_k is the block where the faults in $[f_k]$ enters the model. Figure 4 shows the Hasse diagram of the partial order for the example model. Here one can see that in the example it is necessary and sufficient to ensure detectability of the maximal elements of the partial order. In the example the set of possible sensor locations is X, but with a P that is a proper subset of X one might have the case where a maximal fault class is not detectable regardless of which sensors in P that is added. In such a case, one need to consider the maximal elements among the detectable fault classes.

The following theorem proves the general result and summarizes the discussion of this section.

Theorem 1 *Let M be an exactly determined set of equations, F the corresponding set of faults, $P \subseteq X$ the set of possible sensor locations, and M_S the equations corresponding to adding a set of sensors S. Then maximal detectability of faults F in $M \cup M_S$ are obtained if and only if S has a non-empty intersection with $D([f])$ for all $[f] \in F_m$ where F_m is the set of maximal fault classes among the fault classes with $D([f]) \neq \emptyset$.*

The above result can be summarized in an algorithm that given a model M, faults F, and a set of possible sensor locations P, computes the family of detectability sets $\mathcal{D}$.

1 **function** $\mathcal{D}$ = `Detectability`(M,F,P)
2 Compute block and fault class orders using M;
3 F_m = set of maximal fault classes $[f]$ s.t. $D([f]) \neq \emptyset$;
4 $\mathcal{D}$=$\{D([f]) | [f] \in F_m\}$;

Our objective was not to compute the set of detectability sets $\mathcal{D}$, but rather minimal sensor sets. For this, note that a hitting set for a family of sets is a set that has non-empty intersection with each set in the family. Thus, a minimal hitting set algorithm [Reiter, 1987; de Kleer, 1987] applied to the family of sets $\mathcal{D}$ can be used to efficiently find all minimal sensor sets.

For the example model, as previously noted, the maximal fault classes are $[f_1]$ and $[f_3]$ and the corresponding detectability sets are $D([f_1]) = \{x_1, x_2, x_3\}$ and $D([f_3]) = \{x_1, x_2, x_4\}$. Theorem 1 gives that the minimal sensor sets that achieve detectability of all faults are $\{x_1\}$, $\{x_2\}$, and $\{x_3, x_4\}$, which are the same sensor sets as was determined in Section 2.

4.2 Sensor placement for isolability of detectable faults

This section describes 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 structurally detectable. In the next section this assumption will be removed.

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

Each sub-problem can be formulated as a detectability problem, as will be shown next. Assume that M is a model, including sensors such that all faults are detectable, and M_S represents a set of equations describing an additional sensor set S. Given the sensor set S, a fault f_i is isolable from f_j in the model $M \cup M_S$ if $e_{f_i} \in ((M \cup M_S) \setminus \{e_{f_j}\})^+$ according to Definition 3. By introducing $M' = M \setminus \{e_{f_j}\}$, this can be written as

$$e_{f_i} \in (M' \cup M_S)^+ \quad (1)$$

which according to Definition 2 means that f_i is structurally detectable in $M' \cup M_S$. Hence the maximum possible number of faults $f_i \in F \setminus \{f_j\}$ are isolable from f_j in $M \cup M_S$ if the maximum possible number of faults $f_i \in F \setminus \{f_j\}$ are structurally detectable in the model $(M \cup M_S) \setminus \{e_{f_j}\}$. This shows that each sub-problem can be formulated as a detectability problem.

Next, we use the example formulated in Section 2 to outline the solution of one sub-problem before formally stating

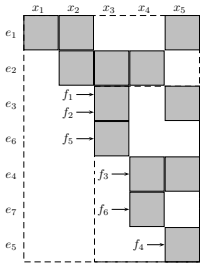


Figure 5: Block structure of the example in Section 2 extended with measurements of x_3 and x_4.

the solution. Assume that we have added sensors measuring $\{x_3, x_4\}$ such that all faults are detectable. Furthermore, assume that these sensors can be faulty and denote these faults f_5 and f_6 respectively. A row permuted structure of the obtained model $M = \{e_1, e_2, \ldots, e_7\}$ is shown in Figure 5.

Consider the sub-problem associated with fault f_1. The set M' in (1) is equal to $M \setminus \{e_{f_1}\} = M \setminus \{e_3\}$. The sub-problem is, given the model $M \setminus \{e_3\}$, to find the minimal additional sensor sets S such that as many of the faults $f_2, f_3, \ldots, f_6$ as possible become detectable in $M' \cup M_S$.

The fault f_2 is not included in M' and cannot be structurally detectable in $M' \cup M_S$ for any sensor set S. This implies that f_2 is not isolable from f_1 with any sensor addition and this also follows from the fact that these to faults violate the same equation. The faults f_3, f_4, and f_6 in the structurally overdetermined part $(M')^+ = \{e_4, e_5, e_7\}$ are according to Definition 2 structurally detectable in M' and require no additional measurements. Fault f_5 in the just-determined part $\{e_1, e_2, e_6\}$ is not detectable, but f_5 can become detectable in $M' \cup M_S$ if S is appropriate selected.

Sufficient and necessary requirements on S can be computed by the function `Detectability` described in Section 4.1. By applying this function to the structurally just-determined part of M', i.e. the sub-graph of M' defined by the node sets $\{e_1, e_2, e_6\}$ and $\{x_1, x_2, x_3\}$, we get that $D([f_5]) = \{x_1, x_2, x_3\}$. Hence, one of the variables in the detectability set $\{x_1, x_2, x_3\}$ must be measured to make the faults $F \setminus \{f_1, f_2\}$ detectable in $M' \cup M_S$ and this implies that all faults in $F \setminus \{f_1, f_2\}$ are isolable from f_1 in $M \cup M_S$. The solution to the sub-problem related to fault f_1 will be the computed detectability set. The next theorem formalizes the solution of a sub-problem like the one discussed above.

Theorem 2 *Let M be a set of equations with no structurally underdetermined part, F a set of structurally detectable faults in M, $P \subseteq X$ the set of possible sensor locations, and M_S the equations added by adding the sensor set S. Let M^0 be the just-determined part of $M \setminus \{e_{f_j}\}$, F^0 the faults contained in M^0, and $\mathcal{D} =$ `Detectability`(M^0, F^0, P). Then the maximum possible number of faults $f_i \in F \setminus \{f_j\}$ are structurally isolable from f_j in $M \cup M_S$ if and only if S have a non-empty intersection with all sets in $\mathcal{D}$.*

The result of the theorem can be summarized in a function that given a model M, a set of detectable faults F in M, the possible sensor locations P, and a fault $f \in F$, computes the family of detectability sets $\mathcal{D}$ that solves the isolability sub-problem for f.

function $\mathcal{D}$ = `IsolabilitySubProblem`(M,F,P,f)
  M^0 = just−determined part of $M \setminus \{e_f\}$;
  F^0 = the set of faults F included in M^0;
  $\mathcal{D}$ = `Detectability`(M^0,F^0,P);

An additional sensor set that maximizes the set of fault pairs (f_i, f_j) such that f_i is structurally isolable from f_j must have a non-empty intersection with all detectability sets found in all sub-problems.

function $\mathcal{D}$ = `Isolability`(M,F,P)
  $\mathcal{D} = \varnothing$;
  for $f_i \in F$
    $F' = F \setminus \{f_i\}$;
    $\mathcal{D} = \mathcal{D} \cup$ `IsolabilitySubProblem`(M,F',P,f_i);
  end

The minimal sensor sets that maximize the isolability can be found by applying a minimal hitting set algorithm to the sets in the output $\mathcal{D}$.

For the example shown in Figure 5, the families of detectability sets of the different sub-problems are $\{\{x_1, x_2, x_3\}\}$ for f_1, f_2, and f_5, $\{\{x_1, x_2, x_4\}\}$ for f_3 and f_6, and $\varnothing$ for f_4. We have found two distinct detectability sets and the minimal hitting sets are $\{x_1\}$, $\{x_2\}$, and $\{x_3, x_4\}$. These sets are the minimal additional measurements that achieve maximum single fault isolability.

4.3 Sensor placement for both detectability and isolability

We have shown how isolability can be achieved in a model where all faults are structurally detectable. Next, we will extend the presented solution to models where faults may not be structurally detectable in the original model.

The solution is first outlined for the example described in Section 2. The faults in this model are not detectable and we want to find all minimal sensor sets that maximize fault detectability and isolability. We have shown in Section 4.1 that the minimal sets of measurements to achieve full detectability are $\{x_1\}$, $\{x_2\}$, and $\{x_3, x_4\}$. If we add for example a sensor measuring x_1 described by an equation e_s, we get a new model $M \cup \{e_s\}$ where all faults are detectable. Since all faults are detectable, the previously described method to achieve maximum isolability can be applied to the model $M \cup \{e_s\}$. The minimal sensor sets that solve this problem are $\{x_3\}$ and $\{x_4\}$. By combining this result with the fact that a sensor measuring x_1 has been added to obtain detectability, it follows that $\{x_1, x_3\}$ and $\{x_1, x_4\}$ are two possible sensor sets that achieve maximum detectability and isolability. To compute all minimal sensor sets that achieve maximum isolability, we also have to investigate the solutions when we choose to measure $\{x_2\}$ or $\{x_3, x_4\}$ to obtain full detectability. By solving one isolability problem for each of the minimal sensor sets that achieves full detectability, we get that the minimal sensor sets are $\{x_1, x_3\}$, $\{x_1, x_4\}$, $\{x_2, x_3\}$, $\{x_2, x_4\}$, and $\{x_3, x_4\}$ which are the same sets as in Section 2.

The following description summaries the suggested algorithm that given a model M with no structurally underdetermined part, the faults F, and the possible sensor locations P, computes the family $\mathcal{S}$ of all minimal sensor sets that

achieve maximum isolability. In the algorithm the join operation of two multisets A and B will be used. The join operation is denoted by $A \uplus B$ and is defined as the multiset containing all elements in $A \cup B$ with a multiplicity equal to the sum of the multiplicities in A and B. For example $\{x_1, x_2\} \uplus \{x_1\} = \{x_1, x_1, x_2\}$.

1 **function** $\mathcal{S}$ = `SensorPlacement`(M,F,P)
2 $\mathcal{S} = \varnothing$;
3 M^0 = just−determined part of M;
4 F^0 = the set of faults F included in M^0;
5 $\mathcal{D}$ = `Detectability`(M^0,F^0,P);
6 $\mathcal{S}_d$ = MinimalHittingSets($\mathcal{D}$);
7 **for** $S_i \in \mathcal{S}_d$
8 Create the extended model $M_e = M \cup M_{S_i}$;
9 F_e = the faults included in M_e;
10 $\mathcal{D}$ = `Isolability`(M_e,F_e,P);
11 $\mathcal{S}_i$ = MinimalHittingSets($\mathcal{D}$);
12 $\mathcal{S} = \mathcal{S} \cup \{S_i \uplus S' | S' \in \mathcal{S}_i\}$;
13 **end**
14 Delete non−minimal sensor sets in $\mathcal{S}$;

4.4 Adding sensors with faults

Sensors might have corresponding sensor faults. When adding a sensor, it is possible that a new fault is introduced into the model and in this section it is shown how these additional sensor faults can be handled.

Consider again the example introduced in Section 2 and assume now that we want to find all minimal sensor sets that maximize the fault isolability when all sensors introduce new possible faults. To do this, we will follow the algorithm `SensorPlacement` and describe how some of the lines should be modified to cope with additional sensor faults.

The additional sensors that have a corresponding sensor fault have to be specified in the algorithm. This is done by introducing an additional input set $P_f \subseteq P$ where sensors measuring variables in P_f may become faulty and the other sensors may not. For the example P_f is equal to P.

The purpose of line 5 and 6 is to compute all sensor sets that achieves full detectability. In Section 4.1, it was shown that $\{x_1\}$, $\{x_2\}$, or $\{x_3, x_4\}$ are the minimal sensor sets that make the original faults $f_1, \ldots, f_4$ detectable and the next theorem shows that all additional sensors faults will automatically become detectable.

Theorem 3 *Let M be a model with no underdetermined part and let x be a measured variable with a sensor described by an equation $e \notin M$. Then, a sensor fault violating e will be structurally detectable in $M \cup \{e\}$.*

The result of this theorem is for the example that $\{x_1\}$, $\{x_2\}$, and $\{x_3, x_4\}$ are the minimal sensor sets that make all faults, including the new faults introduced by the added sensors, detectable. Hence the inputs to the function `Detectability` described in Section 4.1 do not need to be changed at all.

On line 7, a minimal sensor set S_i that achieves full detectability is selected and on line 8 the equations M_{S_i} are added to the original model to form the extended model M_e. If the sensors may become faulty, i.e. if $s \in S_i$ belongs to P_f, then these faults must be added to the model as done in Figure 5. These faults and the original faults in F are then stored on line 9 in F_e.

The purpose of lines 10-11 is to, given the extended model M_e, find the family $\mathcal{S}_i$ of all minimal additional sensor sets S' achieving maximum isolability among both the faults in F_e and the sensor faults associated with the additional sensors S'. The next result states that if S' achieves maximum isolability among the faults F_e, then S' also achieves the maximum isolability among all faults including the faults introduced by the sensors in S'.

Theorem 4 *Let M be a model with no underdetermined part and F a set of structurally detectable faults in M. Furthermore, let M_S be an equation set describing additional sensors and F_S the associated set of sensor faults. Then for any sensor fault $f \in F_S$ and for any fault $f' \in (F \cup F_S) \setminus \{f\}$, it holds that f is isolable from f' and f' is isolable from f in $M \cup M_S$.*

The theorem shows that once sensors and sensor faults have been added to the original model on line 8, the minimal additional sensor sets to achieve maximum isolability can be computed exactly as before, i.e., the lines 10-14 need not be changed. In conclusion, the only difference in the function `SensorPlacement` when considering sensor faults is to add the additional input P_f that should be used in the creation of the extended model M_e on line 8.

A difference in the result from the case when not considering sensor faults is that the solution might include two sensors measuring the same variable. For the example, the minimal sensor sets when considering sensor faults are $\{x_1, x_1, x_3\}$, $\{x_1, x_1, x_4\}$, $\{x_1, x_2, x_3\}$, $\{x_1, x_2, x_4\}$, $\{x_1, x_3, x_4\}$, $\{x_2, x_2, x_3\}$, $\{x_2, x_2, x_4\}$, $\{x_2, x_3, x_4\}$, and $\{x_3, x_3, x_4, x_4\}$.

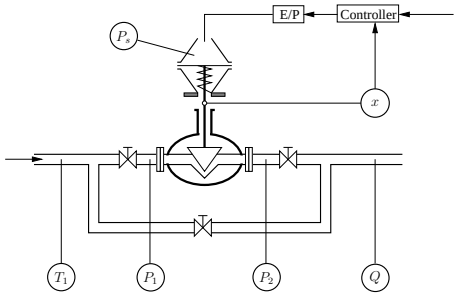
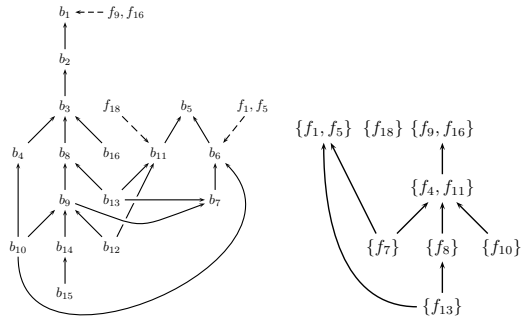


Figure 6: DAMADICS valve

5 Example

The example used to illustrate the results is an industrial valve. A schematic figure of the valve is shown in Figure 6 and consists of three main components: the control valve, a by-pass valve, and a spring-and-diaphragm pneumatic servomotor to operate the valve plug. The figure also shows an internal control loop that is used to increase the accuracy of the valve plug positioning. Details of this model is not included in this presentation and interested readers are referred to e.g. [Syfert *et al.*, 2003] and the references therein. The structure of the model that is used here is derived in [Düştegör *et al.*, 2006].



(a) Order among strongly connected components. It is also noted, with dashed arrows, where some important faults appear in the model.

(b) Order among the equivalence classes on the set of faults.

Figure 7: Order among strongly connected components and faults for the Damadics valve model.

The original model included a specified set of sensors but since the objective here is to perform sensor placement analysis, almost all sensors has been removed. Three sensors has been kept, measurements of the two ambient pressures P_1, P_2, and the measurement of the valve position x that is used in the internal control loop. A fault, denoted f_{10} in the model, influences 2 equations and therefore a dummy variable x_{f10}, and an equation $x_{f10} = f_{10}$ is introduced to ensure that the assumption that each fault only influences 1 equation holds. In this example, all unknown variables *except* the dummy variable x_{f10} are assumed to be possible sensor locations. Of course no fault variables f_i can be measured. This leaves us with a model, which has no underdetermined part, consisting of 17 equations in 16 unknown variables and 12 different faults.

First, to determine which sensors that are necessary to obtain detectability of all faults, the partial orders on the strongly connected components and the fault equivalence classes are computed. Figure 7 shows the Hasse diagrams for both partial orders. In Figure 7-b it is clear that there are three maximal elements of the order, $\{f_1, f_5\}$, $\{f_{18}\}$, and $\{f_9, f_{16}\}$. Thus, obtaining detectability of these faults will automatically provide detectability of all other faults. It is noted in Figure 7-a which strongly connected components the maximal faults influence. Theorem 1 then gives that a sensor set achieving detectability has a non-empty intersection $D([f])$ for each maximal fault class. The unknown variables that appear in each relevant strongly connected component are $X_1 = \{P_z\}$, $X_5 = \{Q\}$, $X_6 = \{Q_v\}$, $X_{11} = \{Q_{v3}\}$ and then $D(\{f_1, f_5\}) = \{Q, Q_v\}$, $D(\{f_{18}\}) = \{Q, Q_{v3}\}$, and $D(\{f_9, f_{16}\}) = \{P_z\}$. By computing minimal hitting sets for these three sets one obtains two minimal sensor sets $\{P_z, Q\}$ and $\{P_z, Q_v, Q_{v3}\}$ and it can be verified using Definition 2 that all faults then become detectable.

Adding any of the above set of sensors only achieves detectability of the faults and does not give full isolability. Running the algorithm from Section 4.4, computing sensor sets that achieves maximum isolability also for faults in the new sensors, gives 8 minimal sensor sets. The minimal sensor sets has 7 or 8 sensors and one minimal set is to add sensors measuring the variables $\{P_s, P_z, P_z, Q, Q, Q_{v3}, x\}$. Note here that we need to add 2 sensors each for the variables P_z and Q. With these sensors, all faults are isolable from each other except for the pairs $\{f_4, f_{11}\}$, $\{f_1, f_5\}$, and $\{f_9, f_{16}\}$. This is because these faults cannot be isolated by adding more sensors measuring unknown variables since they appear in the same equation in the model. The only solution is to do further fault modeling [Frisk *et al.*, 2003] or, possibly rather unrealistic, include a sensor that measures the fault signal directly as in [Commault *et al.*, 2006].

6 Related Work

Sensor placement for diagnosis and fault detection is a well studied problem considering many different aspects. This discussion on relations to other works will focus on three papers that all have problem formulations with strong similarities to this paper.

In [Commault *et al.*, 2006] the sensor placement problem is addressed using input-output separators in a graph-based representation of the system model. A main difference to our paper is that Commault et.al. aims at adding sensors such that, in the linear case, it is possible to obtain a diagonal transfer matrix from faults to residual. This is often a rather unrealistic goal since this is only possible if there are more sensors than faults and for example if the added sensors may become faulty it is generically not possible to solve the posed problem. In addition it is in the paper assumed that fault signals can be measured which is an unrealistic assumption.

The basic problem formulation in [Raghuraj *et al.*, 1999] is almost identical to our paper but the model description is a little bit different. It is a graph-based description and they do not allow cycles in the graph and this results in loss of isolability performance in the solution. A drawback with their proposed solution is that their algorithm does not find all minimal sensor sets, the result does not even need to be minimal. However, it should be possible to use a minimal hitting set algorithm, instead of their greedy search, to obtain all minimal solutions to their posed problem. Another pair of differences are that they do not consider faults in the added sensors and also that faults entering in more than one equation is treated in a non-standard way. For example, in their approach it is not possible to add sensors such that the faults in the model

$$\dot{x} = Ax + \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} f$$

are isolable which is clearly possible.

A third related work is [Travé-Massuyès *et al.*, 2006] where the problem is approached by hypothesizing sensors and then computing the set of analytical redundancy relations (ARR), using all possible causalities, tracing the support of each ARR and then obtain isolability properties of the model. Travé-Massuyès et.al. assumes exoneration, i.e. that a fault always makes the corresponding residuals to exceed their thresholds, which is not assumed in our paper since this is a rather unrealistic assumption. Our approach computes which sensors to add to obtain a certain isolability performance while [Travé-Massuyès *et al.*, 2006] does it the other way around, adding all possible sensors and then removing

sensors until isolability performance decreases. One can expect severe complexity problems with such an approach since the number of ARR:s is exponential in the redundancy of the model [Krysander *et al.*, 2007] and by adding all possible sensors you obtain maximum redundancy. Another difference worth noting is that the performance measure in their paper is a scalar value, the diagnosability degree, equal to the quotient of the number of fault classes by the number of faults. However, different sensor setups may have different isolability properties and still have the same diagnosability degree. This is the reason why the complete isolability relation, rather than e.g. the diagnosability degree, is used as a performance measure in our paper. Similar to our paper, Travé-Massuyès et.al. also includes the case where the new sensors also may become faulty. However, typically this also means that you may have to add more than one sensor to a specific variable and this is not covered in [Travé-Massuyès *et al.*, 2006] indicating possibly incomplete results.

7 Conclusions

The sensor placement problem has been addressed in this paper. The objective is to add sensors such that maximum fault isolability can be achieved. It is often the case that some process variables cannot be measured and this information need to be considered in an analysis. In addition, new sensors may of course also become faulty and these faults must also be included in the analysis. Typically, this means that more than one sensor have to be added measuring a specific signal.

A key contribution is a new algorithm for sensor placement that cope with all aspects mentioned above. Given a model, the possible sensor locations and a specification of which sensors that may be faulty the algorithm computes all minimal sensor sets that make, as far as possible, faults isolable from each other. Typically there is a cost associated with each type of sensor, for example price, maintenance cost, reliability etc. This means that the sensor set with the least number of sensors may not be the best choice. Since the result of the algorithm contain *all* minimal sensor sets, it is straightforward to pose an optimality condition regarding cost to find the best choice of sensors to add.

The algorithm has been applied to a non-trivial industrial valve model with 17 equations and 15 possible sensor positions. The result was 8 minimal sensor sets that achieve maximum isolability also for faults in the added sensors. The minimal sensor sets have 7 or 8 sensors and several of them contain 2 sensors at the same position. A Matlab implementation of the algorithm is available at `http://www.fs.isy.liu.se/Software/SensPlaceTool/`.

References

[Blanke *et al.*, 2003] M. Blanke, M. Kinneart, J. Lunze, and M. Staroswiecki. *Diagnosis and Fault-Tolerant Control*. Springer-Verlag, 2003.

[Commault *et al.*, 2006] C. Commault, J. Dion, and S.Y. Agha. Structural analysis for the sensor location problem in fault detection and isolation. In *Proceedings of IFAC Safeprocess'06*, Beijing, China, 2006.

[Cordier *et al.*, 2004] M.O. Cordier, P. Dague, F. Levy, J. Montmain, M. Staroswiecki, and L. Travé-Massuyès. 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, 2004.

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

[Düştegör *et al.*, 2006] Dilek Düştegör, Erik Frisk, Vincent Coquempot, Mattias Krysander, and Marcel Staroswiecki. Structural analysis of fault isolability in the DAMADICS benchmark. *Control Engineering Practice*, 14(6):597–608, 2006.

[Dulmage and Mendelsohn, 1958] A. L. Dulmage and N. S. Mendelsohn. Coverings of bipartite graphs. *Canadian Journal of Mathematics*, 10:517–534, 1958.

[Frisk *et al.*, 2003] Erik Frisk, Dilek Düştegör, Mattias Krysander, and Vincent Cocquempot. Improving fault isolability properties by structural analysis of faulty behavior models: application to the DAMADICS benchmark problem. In *Proceedings of IFAC Safeprocess 03*, Washington, USA, 2003.

[Krysander and Frisk, 2007] M. Krysander and E. Frisk. Some theoretical results on sensor placement for diagnosis based on fault isolability specifications. Technical Report LiTH-R-2770, ISY, Linköping, Sweden, 2007. `http://www.fs.isy.liu.se/Publications/`.

[Krysander *et al.*, 2007] Mattias Krysander, Jan Åslund, and Mattias Nyberg. An efficient algorithm for finding minimal over-constrained sub-systems for model-based diagnosis. *Accepted for publication in IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 2007.

[Krysander, 2006] Mattias Krysander. *Design and Analysis of Diagnosis Systems Using Structural Methods*. PhD thesis, Linköpings universitet, June 2006.

[Ploix *et al.*, 2005] S. Ploix, M. Desinde, and S. Touaf. Automatic design of detection tests in complex dynamic systems. In *Proceedings of 16th IFAC World Congress, Prague*, Prague, Czech Republic, 2005.

[Raghuraj *et al.*, 1999] R. Raghuraj, M. Bhushan, and R. Rengaswamy. Locationg sensors in complex chemical plants based on fault diagnostic observability criteria. *AIChE*, 45(2):310–322, 1999.

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

[Syfert *et al.*, 2003] M. Syfert, M. Bartys, J. Quevedo, and R.J. Patton. Development and application of methods for actuator diagnosis in industrial control systems (damadics): A benchmark study. In *Proceedings of IFAC Safeprocess 03*, Washington, USA, 2003.

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