A comparative analysis of AI and control theory approaches to model-based diagnosis

M-O. Cordier and **P. Dague** and **M. Dumas** and **F. Lévy** and **J. Montmain** and **M. Staroswiecki** and **L. Travé-Massuyès** [1]
(as part of the French IMALAIA group – imalaia@laas.fr)

Abstract. Two distinct and parallel research communities have been working along the lines of the Model-Based Diagnosis approach: the FDI community and the DX community that have evolved in the fields of Automatic Control and Artificial Intelligence, respectively. This paper clarifies and links the concepts that underlie the FDI analytical redundancy approach and the DX logical approach. The formal match of the two approaches is demonstrated and the proof of their equivalence is provided under various assumptions.

1 Introduction

Diagnosis is an active research topic which can be approached from different perspectives according to the type of knowledge available. The so-called Model-Based Diagnosis (MBD) approach rests on the use of an explicit model of the system to be diagnosed. Two distinct and parallel research communities have been using the MBD approach. The Fault Detection and Isolation (FDI) community uses techniques from control theory and statistical analysis. It has now reached a mature state and a number of very good surveys exist in this field [9, 6, 8]. The DX community emerged more recently, with foundations in the fields of Computer Science and Artificial Intelligence [11, 5, 7].

The goals of the IMALAIA group are to agree upon a common FDI/DX terminology, to identify similarities and complementarities in the FDI and DX methods, and to contribute towards a unifying framework, thus taking advantage of the synergy of techniques from the two communities.

This paper clarifies the link between *parity equations or analytical redundancy relations* (ARR for short) and *conflicts* by introducing the notion of *potential conflicts* or *ARR supports.* The formal match of the two approaches is thus shown. The FDI and DX approaches used for fault localization are then analyzed from the two perspectives. The *exoneration* and *no-compensation* assumptions which are implicit in FDI are made clear and the theoretical proof of equivalence of the two approaches is included, according to adopted assumptions. For the sake of clarity, the study is carried out in a pure consistency-based framework, i.e. without fault models.

The example that has been chosen to support the comparative analysis throughout the paper is the well-known system from [3] composed of three multipliers M1, M2, M3 and two adders A1, A2 (see Figure 1). This choice and the fact that the system is assumed to operate in an ideal non-noisy and non-disturbed environment has been made on purpose to focus on the main features of each approach, without being overburdened neither with modeling details, nor with detection criteria. Let us emphasize that this discrete static example has been chosen for sake of clarity, but that the conclusions stemming from the comparison are quite general. In particular both approaches can deal with continuous dynamic systems by basing the methods on differential or recurrent models. On the other side, the problems related to temporal diagnosis [1] involve many open issues in both approaches and are only evoked in the final discussion.

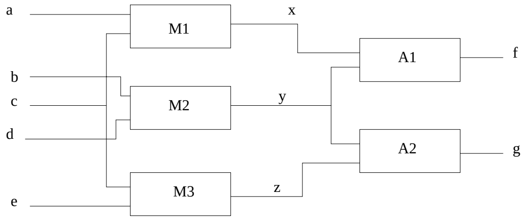

Figure 1. The system

The paper is organized as follows. Sections 2 and 3 present the FDI analytical redundancy approach and the DX logical approach, respectively. Section 4 proposes a unified representation and proves the equivalence of the two approaches. This proof is given under specific assumptions corresponding to two classical cases which are the cases by default assumed in FDI and DX respectively. The general case and a more thorough analysis can be found in the long paper [2]. Finally, Section 5 discusses the results and outlines several interesting directions for future investigation.

2 Redundancy-based diagnosis: the FDI approach

The behavioral model BM of a system is derived from its structure, which shows the links between its components *(structural model*), and the behavior model of each component.

Definition 2.1 The *system model* SM is defined as the *behavioral model* BM, i.e. the set of relations defining the system behavior, together with the *observation model* OM, i.e. the set of relations between the variables X of the system and the observed variables O acquired by the sensors.

Example: Elementary components are the adders A1, A2, the multipliers M1, M2, M3 together with the set of sensors. The system model SM is hence given by the following:
BM: **RM1**: $x = a \times c$ **RM2**: $y = b \times d$
RM3: $z = c \times e$ **RA1**: $f = x + y$ **RA2**: $g = y + z$
OM: **RSa**: $a = a_{\text{obs}}$ **RSb**: $b = b_{\text{obs}}$ **RSc**: $c = c_{\text{obs}}$
RSd: $d = d_{\text{obs}}$ **RSe**: $e = e_{\text{obs}}$ **RSf**: $f = f_{\text{obs}}$ **RSg**: $g = g_{\text{obs}}$

[1] respectively IRISA Rennes, LIPN Paris 13, CEA Saclay, LIPN Paris 13, EERIE Nîmes, LAIL Lille, LAAS Toulouse

Definition 2.2 A *diagnosis problem* is defined by the system model SM, a set of observations OBS assigning values to observed variables, and a set of faults F [2].

Example: OBS = $\{a_{obs} = 2, b_{obs} = 2, c_{obs} = 3, d_{obs} = 3, e_{obs} = 2, f_{obs} = 10, g_{obs} = 12\}$. The set of single faults is SF = $\{F_{A1}, F_{A2}, F_{M1}, F_{M2}, F_{M3}\}$ and the set of faults is F = 2^{SF}.

Definition 2.3 The *system structure* is defined through a binary application s: $SM \times V \rightarrow \{0,1\}$, where $V = X \bigcup O$ is the set of variables and $s(rel,v) = 1$ if and only if v appears in relation *rel*.

Definition 2.4 An *analytical redundancy relation* (ARR) is a relation entailed by SM (and the components whose behavior model is used by this entailment are said to be *involved* in the ARR) which contains only observed variables, and which can therefore be evaluated from OBS. It is noted $r = 0$, where r is called the *residual* of the ARR. For a given OBS, the instantiation of the residual is noted *val(r,OBS)*, abbreviated as *val(r)* when not ambiguous. Thus, *val(r,OBS)=0* if the observations satisfy the ARR.

ARRs can be obtained from the system model by eliminating the unknown variables. This problem can be formalized in a graph theoretical framework, which comes down to the well-known problem of finding a complete matching w.r.t. the unknown variables X in the bipartite graph whose incidence matrix is the matrix associated to the application s. In this system structure matrix representation, a complete matching appears as a selection of one and only one entry per column, corresponding to an unknown variable, and per row, corresponding to a SM relation.

Example: A complete matching leads to the following ARRs:
ARR1: $r_1 = 0$ where $r_1 \equiv f_{obs} - a_{obs} \times c_{obs} - b_{obs} \times d_{obs}$
ARR2: $r_2 = 0$ where $r_2 \equiv g_{obs} - b_{obs} \times d_{obs} - c_{obs} \times e_{obs}$
If we assume that the sensors are not faulty, the ARRs can be rewritten as:
ARR1: $f - (a \times c + b \times d) = 0$ ARR2: $g - (b \times d + c \times e) = 0$

Let us call the ARRs that are obtained from a given complete matching *elementary ARRs*. Given a set of elementary ARRs, additional redundancy relations can be obtained by combining the elementary ones.

Example: A third redundancy relation ARR3: $f - g - a \times c + c \times e = 0$ can be obtained. The components involved in ARR3 are A1, A2, M1, M3. Notice that it is not the union of the components involved in ARR1 (A1, M1, M2) and in ARR2 (A2, M2, M3).

Besides analytical redundancy relations, a fundamental concept in the FDI approach is that of *fault signature*.

Definition 2.5 Given a set R = $\{ARR_1, \ldots, ARR_n\}$ of n ARRs and a set F = $\{F_1, \ldots, F_m\}$ of m faults, the signature of a fault F_j is given by the binary vector $FS_j = [s_{1j}, \ldots, s_{nj}]^T$ in which s_{ij} is given by:
$(ARR_i, F_j) \mapsto s_{ij} = 1$ if some components involved in F_j are involved in ARR_i
$\mapsto s_{ij} = 0$ otherwise.

The interpretation of some s_{ij} being 0 is that the occurrence of the fault F_j does not affect ARR_i, meaning that $val(r_i) = 0$. The interpretation of some s_{ij} being equal to 1 is that the occurrence of the fault F_j is expected to affect ARR_i, meaning that $val(r_i)$ is now expected to be different from 0.

Definition 2.6 Given a set R of n ARRs, the signatures of a set F of m faults all put together constitute the so-called *signature matrix*.

In our example, the signature matrix for the set of single faults corresponding to components A1, A2, M1, M2 and M3 is given by:

	F_{A1}	F_{A2}	F_{M1}	F_{M2}	F_{M3}
ARR1	1	0	1	1	0
ARR2	0	1	0	1	1
ARR3	1	1	1	0	1

The case of multiple faults can be dealt with by expanding the number of columns of the signature matrix, leading to a total number of 2^m-1 columns with m the number of single faults, if all the possible multiple faults are considered. Let F_J be a multiple fault corresponding to the occurrence of k single faults $F_{j1}, \ldots, F_{jk}$, then the entries of the signature vector of F_J are given by:
$s_{ij} = 0$ if $s_{i\,j1} = \ldots = s_{i\,jk} = 0$
$s_{ij} = 1$ if $\exists l\ 1 \leq l \leq k$ such that $s_{i\,jl} = 1$

Example: Extending the matrix above, the 26 additional columns have a $[1, 1, 1]^T$ signature, except for $F_{\{A1, M1\}}$ which has a $[1, 0, 1]^T$ signature, and for $F_{\{A2, M3\}}$ which has a $[0, 1, 1]^T$ signature.

The diagnostic sets in the FDI approach are given in terms of the faults accounted for in the signature matrix. The generation of the diagnostic sets is based on a column interpretation of the signature matrix and consists in comparing the *observation signature* with the fault signatures. This comparison is stated as a decision-making problem.

Definition 2.7 The signature of an observation OBS is a binary vector OS = $[OS_1, \ldots, OS_n]^T$ where $OS_i = 0$ iff $val(r_i,OBS) = 0$.

The first step (the *detection* task) is to build the observation signature, i.e. to decide whether a residual value is zero or not, in the presence of noises and disturbances. This problem has been thoroughly investigated within the FDI community. It is generally stated as a statistical decision-making problem, making use of the available noise and disturbance models.

Example: With OBS as above, OS = $[1, 0, 1]^T$. In the case $f = 10$ and $g = 10$, OS = $[1, 1, 0]^T$ and in the case $f = 10$ and $g = 14$, OS = $[1, 1, 1]^T$.

The second step (the *isolation* task) is to actually compare the observation signature with the fault signatures. A solution to this decision-making problem is to define a *consistency criterion* as follows:

Definition 2.8 An observation signature OS = $[OS_1, \ldots, OS_n]^T$ is consistent with a fault signature $FS_j = [s_{1j}, \ldots, s_{nj}]^T$ if and only if $OS_i = s_{ij}$ for all i.

Definition 2.9 The *diagnostic sets* are given by the faults whose signatures are consistent with the observation signature.

Example: The diagnostic sets got for the following observation signatures are:
OS = $[1, 0, 1]^T \Leftrightarrow F_{A1}$ or F_{M1} or $F_{\{A1, M1\}}$
OS = $[1, 1, 0]^T \Leftrightarrow F_{M2}$
OS = $[1, 1, 1]^T \Leftrightarrow$ any multiple fault except $F_{\{A1, M1\}}$ and $F_{\{A2, M3\}}$

Note that the FDI community generally uses a *similarity-based consistency criterion* arising from the definition of a distance rather than the equality-based criterion defined above.

3 Logical-based diagnosis: the DX approach

Reiter [11] proposed a logical theory of diagnosis. This approach, also referred to as consistency-based diagnosis, was later extended

[2] In order to facilitate the comparison with DX, and without loss of generality, a fault can be seen as a set of faulty components.

and formalized in [4]. In the following we refer to the basic definitions of [11] without considering posterior extensions and refinements. The description of the behavior of the system is component-oriented and rests on first-order logic.

Definition 3.1 A *system model* is a pair (SD, COMPS) where SD, the *system description*, is a set of first-order logic formulas with equality and COMPS, the components of the system, is a finite set of constants. SD uses a distinguished predicate AB, interpreted to mean abnormal. $\neg$AB(c) with c belonging to COMPS hence describes the case where the component c is behaving correctly.

Example: COMPS = {A1, A2, M1, M2, M3}
SD = { ADD(x) $\wedge$ $\neg$AB(x) $\Rightarrow$ Output(x) = Input1(x) + Input2(x),
MULT(x) $\wedge$ $\neg$AB(x) $\Rightarrow$ Output(x) = Input1(x) $\times$ Input2(x),
ADD(A1), ADD(A2), MULT(M1), MULT(M2), MULT(M3),
Output(M1) = Input1(A1), Output(M2) = Input2(A1),
Output(M2) = Input1(A2), Output(M3) = Input2(A2),
Input2(M1) = Input1(M3) }

Let us note one point which differs somewhat from the description of the system in the FDI approach: with the distinguished predicate AB it is possible to make explicit the fact that a formula in SD describes the normal behavior of a given component. The description can easily be extended to include faulty behaviors.

A diagnosis problem results from the discrepancy between the normal behavior of a system as described by the system model and a set of observations.

Definition 3.2 A set of observations OBS is a set of first-order formulas.

Example: An example of observations for our system is OBS = {Input1(M1) = 2, Input2(M1) = 3, Input1(M2) = 2, Input2(M2) = 3, Input2(M3) = 2, Output(A1) = 10, Output(A2) = 12}.

Definition 3.3 A *diagnosis problem* is a triple (SD, COMPS, OBS) where (SD, COMPS) is a system model and OBS a set of observations.

A diagnosis is a conjecture that certain components of the system are behaving abnormally. This conjecture has to be consistent with what is known about the system and with the observations.

Definition 3.4 A *diagnosis* for (SD, COMPS, OBS) is a set of components $\Delta \subseteq$ COMPS such that SD $\bigcup$ OBS $\bigcup$ {AB(c) | c $\in \Delta$} $\bigcup$ {$\neg$AB(c) | c $\in$ COMPS – Δ} is satisfiable. A *minimal diagnosis* is a diagnosis Δ such that $\forall\Delta' \subset \Delta$, Δ' is not a diagnosis.

Following the principle of parsimony, minimal diagnoses are often the preferred ones. For the sake of simplicity, we will limit ourselves to minimal diagnoses. A method based upon the concept of conflict set has been proposed in [11] to generate minimal diagnoses and is at the basis of most of implemented DX algorithms.

Definition 3.5 An *R-conflict* for (SD, COMPS, OBS) is a set of components C = {c1, ..., ck} $\subseteq$ COMPS such that SD $\bigcup$ OBS $\bigcup$ {$\neg$AB(c) | c $\in$ C} is inconsistent. A *minimal R-conflict* is an R-conflict which does not include any R-conflict.

An R-conflict can be interpreted as follows: one at least of the components in the R-conflict is faulty in order to account for the observations.

Example: The system with the observations as seen above has the following minimal R-conflicts: {A1, M1, M2} and {A1, A2, M1, M3} due to the abnormal value of 10 for *f*. In the case *f* = 10 and *g* = 10, the two minimal R-conflicts are: {A1, M1, M2} and {A2, M2, M3}. In the case *f* = 10 and *g* = 14, there are three minimal R-conflicts: {A1, M1, M2}, {A2, M2, M3} and {A1, A2, M1, M3}.

Using these minimal R-conflicts, it is possible to give a characterization of minimal diagnoses which provides a basis for computing them [11].

Proposition 3.1 Δ is a minimal diagnosis for (SD, COMPS, OBS) if and only if Δ is a minimal hitting set[3] for the collection of (minimal) R-conflicts for (SD, COMPS, OBS).

Example: With *f* = 10 and *g* = 12, there are four minimal diagnoses given by the minimal hitting sets for {{A1, M1, M2}, {A1, A2, M1, M3}} which are: Δ1 = {A1}, Δ2 = {M1}, Δ3 = {A2, M2}, Δ4 = {M2, M3}. With *f* = 10 and *g* = 10, there are five minimal diagnoses given by the minimal hitting sets for {{A1, M1, M2}, {A2, M2, M3}} which are: Δ1 = {M2}, Δ2 = {A1, A2}, Δ3 = {A1, M3}, Δ4 = {A2, M1}, Δ5 = {M1, M3}. With *f* = 10 and *g* = 14, there are eight minimal diagnoses given by the minimal hitting sets for {{A1, M1, M2}, {A2, M2, M3}, {A1, A2, M1, M3}} which are: Δ1 = {A1, A2}, Δ2 = {A1, M2}, Δ3 = {A1, M3}, Δ4 = {A2, M1}, Δ5 = {A2, M2}, Δ6 = {M1, M2}, Δ7 = {M1, M3}, Δ8 = {M2, M3}.

4 Unified framework for DX and FDI approaches

4.1 ARRs vs R-conflicts

In both approaches, diagnosis is triggered when discrepancies occur between the modeled (correct) behavior and the observations (OBS). In the ARR framework, discrepancies come from ARRs which are not satisfied by OBS. In DX, discrepancies allow the identification of R-conflicts, where an R-conflict is a set of components the correctness of which supports a discrepancy. An analogous concept can be defined in FDI.

Definition 4.1 The *support* of an ARR is the set of components involved in this ARR, i.e. columns of the signature matrix with a non zero element in the row corresponding to this ARR. It is also called a *potential R-conflict*. This name is justified by the following result.

Proposition 4.1 Let OBS be a set of observations for a system modeled by SM (resp. SD). There is an identity between the set of minimal R-conflicts for OBS and the set of minimal potential R-conflicts associated to the ARRs which are not satisfied by OBS (proof in [2]).

Example: The potential R-conflicts are: C1 = {A1, M1, M2} (support of ARR1), C2 = {A2, M2, M3} (support of ARR2) and C3 = {A1, A2, M1, M3} (support of ARR3). With *f* = 10 and *g* = 12, ARR1 and ARR3 are not satisfied, which gives rise to the minimal R-conflicts C1 and C3. With *f* = 10 and *g* = 10, ARR1 and ARR2 are not satisfied, which gives rise to the minimal R-conflicts C1 and C2. With *f* = 10 and *g* = 14, ARR1, ARR2 and ARR3 are not satisfied, which gives rise to the minimal R-conflicts C1, C2 and C3.

Let us now analyze the relationship between potential R-conflicts and R-conflicts. From the computational point of view, the main difference between the FDI and DX approaches is that in FDI most of the work is done off-line. Using just the knowledge of observed variables, i.e. sensor locations, modeling knowledge is compiled: ARRs are obtained by combining model constraints and eliminating unobserved variables. The only thing that has to be done on-line, i.e. when a given OBS is acquired, is to compute the falsity value (w.r.t. OBS) of each ARR and to compare the observation signature obtained with the fault signatures. In terms of R-conflicts, this means that potential

[3] A hitting set for a collection of sets is a set that intersects any set of the collection.

R-conflicts are compiled and that, for any OBS, R-conflicts are exactly those potential R-conflicts which are supports of those ARRs which are not satisfied by OBS.

4.2 The matrix framework

The FDI approach uses the signature matrix crossing ARRs in rows and sets of components in columns. It was shown in section 2 that, given an observation OBS, diagnosis is achieved by identifying those columns which are identical (or closest w.r.t. a distance function) to the observation signature column.

In the DX approach, it has been seen in section 3 that minimal diagnoses are obtained as minimal hitting sets of the collection of (OBS-) R-conflicts. From proposition 4.1 above, such R-conflicts can be viewed as the supports of those ARRs which are not satisfied by OBS, i.e. by looking at the corresponding set of rows I. A minimal hitting set of the collection of R-conflicts is then a minimal set J of singleton columns such that each of the rows of I intersects at least one column of J (i.e. has a non zero element in this column).

It is thus quite natural to adopt this matrix framework as a formal basis on which to compare the two approaches. The following notations are used:

- Let R = $\{ARR_i$ / $i = 1 \ldots n\}$ be the set of ARRs and COMPS = $\{C_j$ / $j = 1 \ldots m\}$ the set of components of the system. FS = $[s_{ij}]_{i = 1 \ldots n,\, j = 1 \ldots m}$ is the signature matrix. The j^{th} column of FS is the signature of a fault in C_j and is noted FS_j. For J = $\{j_1, ..., j_k\} \subseteq \{1,...,m\}$, let us note C_J the subset $\{C_j$ / $j \in J\}$, and s_{iJ} the element of the extended matrix FS at line i and column J.
- Any observation OBS splits the set R into two subsets:
  - the subset R_{false} of ARRs it is inconsistent with, i.e. R_{false} = $\{ARR_i \equiv (r_i = 0)$ / $val(r_i,OBS) \neq 0\}$.
  - the subset R_{true} = ARR – R_{false} of ARRs it is consistent with, i.e. R_{true} = $\{ARR_i \equiv (r_i = 0)$ / $val(r_i,OBS) = 0\}$.

  OBS is thus described through its signature OS, which is the binary column vector defined by: for all i = 1 . . . n, $OS_i = 1$ if $ARR_i \in R_{false}$ and $OS_i = 0$ if $ARR_i \in R_{true}$.

The FDI theory compares the observation signature to the fault signatures whereas DX considers separately each line corresponding to an ARR in R_{false}, isolating R-conflicts before searching for a common explanation. In the following, these approaches are called *column view* and *line view* respectively.

4.3 Exoneration and no-compensation assumptions

The originality and the power of both the FDI and DX approaches result from the fact that they are based only on the correct behavior of the components: no model of faulty behavior is needed. Nevertheless, different assumptions concerning the manifestations of the faults through observations are adopted by default by each approach, leading to different computations of the diagnoses, which explains the different results obtained on the example. These assumptions concern: 1) the manifestations of the faults through observations and 2) the case of simultaneous faults and of their interaction.

In addition to the obvious fact that a fault cannot affect an ARR in which it is not involved, which is the direct form of the reasoning used in DX, the idea used in FDI is that a fault necessarily manifests itself by affecting the ARRs in which it is involved, causing them not to be satisfied by any given OBS. Hence not only, as in DX, is any column involved in a not satisfied row a fault candidate, but also any column involved in a satisfied ARR is implicitly exonerated (satisfied rows are thus also used in the reasoning). In fact this result is not sound but rests on an exoneration assumption which is implicitly made in the FDI approach and has to be considered explicitly in order to compare the FDI approach with the DX approach.

Definition 4.2 *(ARR-based exoneration assumption)* A set of faulty components necessarily shows its faulty behavior, i.e. causes any ARR in which it is involved not to be satisfied by any given OBS. Or, equivalently, given OBS, any set of components involved in a satisfied ARR is exonerated, i.e. each component of its support is considered to be behaving correctly.

Note that this exoneration assumption is made up of 1) a *single fault exoneration assumption* (each individual component shows its faulty behavior) and 2) a *no-compensation assumption* (the individual effects of faulty components never compensate each other).

From the matrix viewpoint, the fact that ARR_i exonerates C_j will appear as usual (cf. section 2) in FS as $s_{ij} = 1$, whereas we have chosen to represent the fact that C_j is in the support of ARR_i but that the exoneration is not assumed by s_{ij} = X. The elements of FS can thus take their values in {0,1}, {0,X} or {0,X,1}. The semantics of s_{ij} = X is: a fault in C_j can explain why ARR_i is not satisfied, but ARR_i may happen to be satisfied even when C_j is faulty. The semantics of $s_{ij} = 1$ is: a fault in C_j forces ARR_i not to be satisfied (hence if ARR_i is satisfied then C_j is not faulty - which explains the term "exoneration"). The generalized use of an exoneration assumption for each component will be called the *exoneration and no-compensation case (exo/no-comp)* and corresponds to the assumption by default in the FDI approach, while the total lack of exoneration will be called the *no-exoneration and compensation case (no-exo/comp)* and corresponds to the assumption by default in the DX approach.

4.4 Equivalence in the exo/no-comp case

In this case, fault signatures involve only 0 and 1. As seen in section 2, the signature of the column C_J of the extended matrix is given by the following *fault interaction law* which expresses the no-compensation assumption:

$$s_{iJ} = \sup\{s_{ij} / j \in J\} \text{ for the order } 0 < 1 \qquad \text{(FIenc)}$$

Let $\text{Support}(ARR_i) = \{C_J / s_{iJ} = 1\}$ and $\text{Scope}(C_J) = \{ARR_i / s_{iJ} = 1\}$. The *column view* searches for a perfect match of a fault signature with the observation signature. A set C_J is a diagnosis if and only if:

$$R_{false} = \text{Scope}(C_J). \qquad \text{(CVenc)}$$

In the *line view*, the diagnoses are subsets C_J of COMPS such that:

$$\forall i\, (ARR_i \in R_{false} \Rightarrow \exists j \in J, C_j \in \text{Support}(ARR_i)) \wedge \qquad \text{(LVenc)}$$
$$\forall i\, (ARR_i \in R_{true} \Rightarrow \forall j \in J, C_j \in \text{COMPS} - \text{Support}(ARR_i))$$

Due to(Flenc) this is equivalent to: $\forall i\, (ARR_i \in R_{false} \Leftrightarrow C_J \in \text{Support}(ARR_i))$ which is itself equivalent to (CVenc), which proves the equivalence of the column and line views.

Example: With $f = 10$ and $g = 12$, i.e. observation signature $[1, 0, 1]^T$, there are 2 minimal single fault diagnoses {A1} and {M1} and one superset diagnosis {A1, M1} (the components A2, M2 and M3 are exonerated as members of the support of the satisfied ARR2). With $f = 10$ and $g = 10$, i.e. observation signature $[1, 1, 0]^T$, the only diagnosis is {M2} (the components A1, A2, M1 and M3 are exonerated as members of the support of the satisfied ARR3). With $f = 10$ and $g = 14$, i.e. observation signature $[1, 1, 1]^T$, there are 8 minimal double fault diagnoses (those found in section 3) and 16 superset diagnoses (exoneration plays no role here).

4.5 Equivalence in the no-exo/comp case

In this case, which is the common one in DX, fault signatures involve only 0 and X, and X matches both 0 and 1. From the semantics of X seen in 4.3, it results that columns of the extended matrix are built according to the following rule: a multiple fault can explain that a given ARR is not satisfied if and only if at least one of its faults can explain it, i.e. several faults never produce more than the combination of their separate effects; on the other hand, it is admitted that a faulty component does not necessarily affect an ARR in which it is involved (single fault no-exoneration) and that several faults may always compensate each other (compensation), resulting in a satisfied ARR. The *fault interaction law* can thus be stated as:

$s_{iJ} = \sup\{s_{ij} \mid j \in J\}$ for the order $0 < X$ (FInec)

Let $\mathrm{WeakSupport}(\mathrm{ARR}_i) = \{C_J \mid s_{iJ} \neq 0\}$ and $\mathrm{WeakScope}(C_J) = \{\mathrm{ARR}_i \mid s_{iJ} \neq 0\}$.

In the *column view*, C_J is a diagnosis if and only if:

$R_{\mathrm{false}} \subseteq \mathrm{WeakScope}(C_J)$ (CVnec)

In the *line view* the diagnoses are the sets C_J such that:

$\forall i\ (\mathrm{ARR}_i \in R_{\mathrm{false}} \Rightarrow \exists j \in J, C_j \in \mathrm{WeakSupport}(\mathrm{ARR}_i))$ (LVnec)

Due to (FInec), this translates to: $\forall i\ (\mathrm{ARR}_i \in R_{\mathrm{false}} \Rightarrow C_J \in \mathrm{WeakSupport}(\mathrm{ARR}_i))$ which in turn is the same as $R_{\mathrm{false}} \subseteq \mathrm{WeakScope}(C_J)$, i.e. (CVnec). This proves the equivalence of diagnoses.

Example: The extended signature matrix is obtained from the usual one (see section 2) by replacing each 1 by X. With $f = 10$ and $g = 12$, i.e. observation signature $[1, 0, 1]^T$, there are 4 minimal diagnoses: the 2 single fault diagnoses {A1} and {M1} and the 2 double fault diagnoses {A2, M2} and {M2, M3}, and 22 superset diagnoses. With $f = 10$ and $g = 10$, i.e. observation signature $[1, 1, 0]^T$, there are 5 minimal diagnoses: the single fault diagnosis {M2} and the 4 double fault diagnoses {A1, A2}, {A1, M3}, {A2, M1} and {M1, M3}, and 20 superset diagnoses. With $f = 10$ and $g = 14$, i.e. observation signature $[1, 1, 1]^T$, the results are the same that in 4.4.

4.6 Equivalence in the general case

It is now simple to provide an extension of the framework which allows three-valued fault signatures, involving 0, X and 1. In this case, exoneration applies to some components w.r.t. some ARRs, but not to all. Equivalence can be proved in the same way as above [2].

5 Conclusion and prospects

The first goal of FDI was fault detection and associated decision procedures. Its main interest was to offer sophisticated techniques so as to combine observations such as observers and filters. DX, on the other hand, aimed at localization by recognizing subsets of the system description that conflicted with the observation. Our study proves that a significant part of the two theories fits into a common framework which allows a precise comparison. When they adopt the same hypotheses with respect to how faults manifest themselves, FDI and DX views agree on diagnoses. This opens the possibility of a fruitful cooperation between these two diagnostic approaches, getting the best from each one: compiling modeling knowledge under ARRs form according to sensor locations before any observation has been made, which is the main advantage of the FDI approach and, thanks to explicit correctness assumptions, computing at the same time potential R-conflicts (supports of ARRs) to give rise, given an OBS, to R-conflicts on which the diagnoses generation is based, which is the main advantage of the DX approach.

It is important to notice that the equivalence between the two approaches is obtained either by importing in DX the ARR-based exoneration (enc) assumption implicitly used in FDI or by importing in FDI the no-exoneration (nec) assumption used by default in DX. But another way to express exoneration has been introduced in DX, at the component model level, by assuming that, if the correct behavior model of a component is satisfied by OBS, then this component behaves correctly in the context given by OBS, i.e. by modeling components behavior with bi-conditionals [10]. In [2] this *model-based exoneration* (mbe), which is proved to be weaker than (enc) in the single fault case, is thoroughly compared with (enc). An analog of the proposition 4.1, which relates minimal alibis, i.e. defined Horn AB-clauses entailed by SD $\bigcup$ OBS, with supports of ARRs satisfied by OBS, allows one to prove that any FDI diagnosis with (enc) is a DX diagnosis with (mbe) when SD $\bigcup$ OBS is Horn (but the converse is false). Then the comparison is made between (mbe) and what turns out to be the closest assumption in the FDI framework, i.e. fault exoneration and multiple fault compensation (ec): most of the time the diagnoses obtained are identical (this is the case for the example) but this is not always true.

Some points need future investigation. There is presently no equivalent in DX of the notion of noise and disturbance. Conversely, in the consistency-based extended framework, DX makes a systematic use of fault models, whose counterpart in FDI can be found in assumptions about the additive or multiplicative deviations which model the faults. Fault models have been left out of the framework of the present paper. The conclusions of this work remain valid in case of temporal sequence of observations when the faults do not evolve along time. Such a sequence only provides more observation signatures or more conflicts, allowing diagnoses to be refined by reasoning on each snapshot of the system (state-based approach). Conversely, the incremental diagnosis problem (i.e. when faults occur and evolve along time) is still open on each side: dealing with dynamic residuals and temporal signatures on one side and with simulation-based approach ([12]) on the other side. Further studies are needed to integrate these aspects, which would be beneficial to both communities.

REFERENCES

[1] V. Brusoni, L. Console, P. Terenziani, and D. Theseider Dupré, 'A spectrum of definitions for temporal model-based diagnosis', *Artificial Intelligence*, **102**(1), 39–79, (1998).

[2] M-O. Cordier, P. Dague, M. Dumas, F. Lévy, J. Montmain, M. Staroswiecki, and L. Travé-Massuyès. Conflicts versus analytical redundancy relations. to be submitted, 2000.

[3] R. Davis, 'Diagnostic reasoning based on structure and behavior', *Artificial Intelligence*, **24**, 347–410, (1984).

[4] J. de Kleer, A. Mackworth, and R. Reiter, 'Characterizing diagnoses and systems.', *Artificial Intelligence*, **56**(2-3), 197–222, (1992).

[5] J. de Kleer and B. C. Williams, 'Diagnosing multiple faults', *Artificial Intelligence*, **32**(1), 97–130, (1987).

[6] P.M. Frank, 'Analytical and qualitative model-based fault diagnosis – a survey and some new results', *European Journal of Control*, **2**, 6–28, (1996).

[7] *Readings in Model-Based Diagnosis*, eds., W. Hamscher, L. Console, and J. de Kleer, Morgan Kaufmann, San Mateo, CA, 1992.

[8] R. Iserman, 'Supervision, fault detection and fault-diagnosis methods – an introduction', *Control Engineering Practice*, **5**(5), 639–652, (1997).

[9] R.J. Patton and J. Chen, 'A review of parity space approaches to fault diagnosis', in *IFAC SAFEPROCESS Symposium*, Baden-Baden, (1991).

[10] O. Raiman, 'The alibi principle', In Hamscher et al. [7], 66–70.

[11] R. Reiter, 'A theory of diagnosis from first principles', *Artificial Intelligence*, **32**(1), 57–96, (1987).

[12] P. Struss, 'Fundamentals of model-based diagnosis of dynamic systems', in *15th International Joint Conference on Artificial Intelligence, IJCAI-97*, pp. 480–485, Nagoya, Japan, (1997). Morgan Kaufmann.