AI AND AUTOMATIC CONTROL APPROACHES OF MODEL-BASED DIAGNOSIS : LINKS AND UNDERLYING HYPOTHESES

M-O. Cordier[1], P. Dague[2], M. Dumas[3], F. Lévy[2], J. Montmain[4], M. Staroswiecki[5], L. Travé-Massuyès[6]

1. INRA, 65 rte de Saint-Brieuc, 35042 Rennes cedex,
on leave from IRISA, Campus de Beaulieu, F-35000 Rennes,
cordier@irisa.fr
2. LIPN-UPRESA 7030, Avenue J.B. Clément, F-93430 Villetaneuse,
dague, francois.levy@lipn.univ-paris13.fr
3. CEA Saclay, F-91191 Gif-sur-Yvette cedex
michel@soleil.serma.cea.fr
4. EMA-CEA, Site EERIE - Parc George Besse, F-30035 Nîmes cedex1
jacky.montmain@site-eerie.ema.fr
5. LAIL URA CNRS, EUDIL, Université Lille I
F- 59655 Villeneuve d'Ascq cedex
marcel.staroswiecki@univ-lille1.fr
6. LAAS-CNRS, 7 Avenue du Colonel-Roche, F-31077 Toulouse cedex
louise@laas.fr

Abstract: Two distinct communities have been working along the Model-Based Diagnosis approach. This paper clarifies and links the concepts and hypotheses that underly the FDI analytical redundancy approach and the DX consistency-based logical approach. This work results from the collaboration existing within the French IMALAIA group supported by the French National Programs on Automatic Control *GDR Automatique* and on Artificial Intelligence *GDR I3. Copyright © 2000 IFAC*

Keywords: Model-Based Diagnosis, Fault Detection and Isolation, Analytical Redundancy Approach versus Consistency-Based Logical Approach.

1. INTRODUCTION

Diagnosis is an increasingly 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. More specifically, the consistency-based approach only requires knowledge about the normal operation of the system, which is a definite advantage of this approach with respect to others, such as the relational or the pattern recognition approach. In this framework, the occurrence of a fault is captured by discrepancies between the observed behavior and the behavior that is predicted by the model. Fault isolation then rests on interlining the groups of components that are involved in each of the detected discrepancies.

Two distinct and parallel research communities have been working along the MBD approach. The FDI community has evolved in the Automatic Control field from the seventies and 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 (Gertler, 1991) (Frank, 1996) (CEP, 1997). The DX community emerged more recently, with foundations in the fields of Computer Science and Artificial Intelligence (Reiter, 1987) (de Kleer and Williams, 1987) (Hamscher *et* *al.,* 1992). Although the foundations are supported by the same principles, each community has developed its own concepts, tools and techniques, guided by their different modeling backgrounds. The modeling formalisms call indeed for very different technical fields; roughly speaking analytical models and linear algebra on one hand and symbolic and qualitative models with logic on the other hand.

This paper clarifies and links the concepts that underly the FDI analytical redundancy approach and the DX consistency-based logical approach (the DX abductive approach, which rests on fault models, is not considered here, but notice that the consistency-based framework can be generalized to include it (de Kleer *et al.,* 1992)). In particular, the link between *parity equations or analytic redundancy relations* (ARR for short) and *conflicts* is clarified by introducing the notion of *potential conflict*. The FDI and DX approaches used for fault isolation are then analyzed from the two perspectives. It is shown that the first one, based on fault signatures, proceeds to a column interpretation of the signature matrix whereas the latter, based on conflicts, proceeds to a row interpretation. The *exoneration* and the *no-compensation* assumptions which play an important part in both approaches are made clear and interpreted according to the fact that logical soundness or structural properties are seeked.

The well-known polybox example from (de Kleer and Williams, 1987) has been chosen for illustration through the paper. It refers to the simplest diagnosis problem (static system, no incremental diagnosis, no choice of the best next test, no fault models, ideal non-noisy and non disturbed environment).

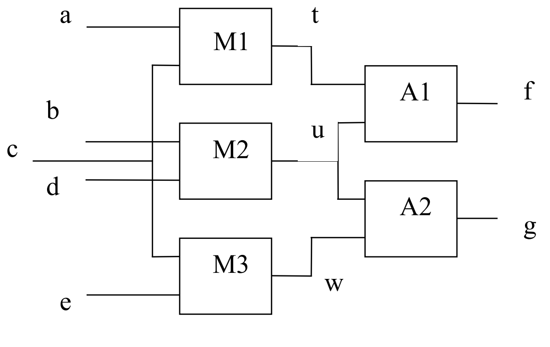

Figure 1 – *The system*

The paper is organized as follows. Sections 2 and 3 present the FDI analytical redundancy approach and the DX consistency-based logical approach. The underlying concepts and the common assumptions adopted by the FDI and DX communities, are outlined and discussed in section 4. Finally, some conclusions are given in section 5.

2. ANALYTICAL REDUNDANCY

2.1. The system model

Definition 2.1. The system is a set of interconnected components. Each component operates according to a set of (static or dynamic) constraints between its input and output variables. The system behavioral model (BM) is the set of the constraints which describe the component behaviors taking into account the component interconnections. The observation model (OM) enumerates the subset of the variables which are known to the engineer. The system model (SM) is the pair (BM, OM).

Example Elementary components are the adders A1, A2, the multipliers M1, M2, M3 together with the set of sensors. The system model SM is given by the following:

BM	OM
RM1 $t = a.c$	RSa $a = a_{obs}$
RM2 $u = b.d$	RSb $b= b_{obs}$
RM3 $w = c.e$	RSc $c = c_{obs}$
RA1 $f = u + t$	RSd $d = d_{obs}$
RA2 $g = u + w$	RSe $e = e_{obs}$
	RSf $f = f_{obs}$
	RSg $g = g_{obs}$

Definition 2.2. A diagnosis problem is defined by the system model SM, a set of observations OBS assigning values to the observed variables and a set of faults □. Note that without any loss of generality, a fault can be seen as a faulty component (or a set of faulty components). It can also be seen as one (or several) of the corresponding constraints of BM being not satisfied. In the following, a fault is denoted by a set of faulty component(s).

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 components is COMPS = {{A1}, {A2}, {M1}, {M2}, {M3}} and the set of faults is □ = 2^{COMPS}.

2.2. Analytic redundancy relations

Definition 2.3. An analytic redundancy relation (ARR) is a relation entailed by SM which contains only observed variables. It is generally written under the form: □(OBS) = 0.

This means that under normal operating conditions, the observed values have to satisfy the ARR. When faults are present, the ARR is expected not to be satisfied. A *residual* r is thus introduced such that: □(OBS) = r and r = 0 in the non-faulty case.

ARRs are obtained from SM by eliminating the unknown variables. The problem can be formalized in a graph theoretic framework (Staroswiecki and Declerck, 1989) or in the framework of numerical models (Patton and Chen, 1991) (Staroswiecki and Comtet-Varga, 2000).

Example The following ARRs can be obtained from the polybox system:

ARR1 $r_1 = 0$ where $r_1 \equiv f_{obs} - a_{obs}.c_{obs} - b_{obs}.d_{obs}$
ARR2 $r_2 = 0$ where $r_2 \equiv g_{obs} - b_{obs}.d_{obs} - c_{obs}.e_{obs}$
ARR3 $r_3 = 0$ where $r_3 \equiv f_{obs} - g_{obs} - c_{obs}.(a_{obs} - e_{obs})$

Each ARR is characterized by its *structure*, i.e. by the list of constraints which have to be satisfied in order for the ARR to be satisfied (that is the minimal list of constraints required to obtain the ARR by an elimination process). Each constraint being associated with a component, the structure of an ARR is in the following denoted by the corresponding set of components. The structure of ARR1 is {A1, M1, M2}; the structure of ARR2 is {A2, M2, M3} and the structure of ARR3 is {A1, A2, M1, M3}. Notice that even if r_3 is a linear algebraic combination of r_1 and r_2, the structure of ARR3 is not the union of the structures of ARR1 and ARR2.

2.3. The fault signature matrix

Definition 2. 4. Given a set R of n ARRs and a set □ of m single faults, the signature of a fault $□_j$ is given by the binary vector $FS_j = (s_{ij}, i = 1, ..n)^T$ where:

R□□ □ {0, 1}
$(ARR_i, □_j)$ □ $s_{ij} = 1$ iff $□_i$ belongs to the structure of ARR_i.

The interpretation of some s_{ij} being 0 is that the occurrence of the fault $□_j$ does not affect ARR_i meaning that residual $r_i = 0$ in the presence of that fault. Note that $s_{ij} = 1$ means that residual r_i *is expected* to be affected by fault $□_j$, but it is *not guaranteed* that it will really be (the fault might non be detectable by this ARR, see discussion in section 4.2).

Definition 2.5. Given a set R of n ARRs and a set □ of m single faults, the *signature matrix* gathers all the fault signatures. The 1s in each row of the signature matrix give the structure of the corresponding ARR.

Example. The single faults signature matrix is:

	A1	A2	M1	M2	M3
ARR1	1	0	1	1	0
ARR2	0	1	0	1	1
ARR3	1	1	1	0	1

Considering multiple faults leads to expand the number of columns of the signature matrix up to a total of 2^{m-1} columns if all possible combinations are considered. Let $□_J$ be the multiple fault □ $_{i□J}$ $□_i$. Its signature is generally obtained as:

$$FS(□_J) = \ _{i□J} FS(□_i)$$

2.4. The diagnosis

The generation of the diagnostic set is based on a column interpretation of the signature matrix and consists in comparing the *observation signature* with the fault signatures.

Definition 2.6. The signature of a given observation OBS is a binary vector :

$$OS = (OS_1, \ldots OS_n)^T \text{ where } OS_i = 0 □ \ r_i = 0.$$

The first step is to build the observation signature, i.e. to decide whether a residual value is zero or not. This problem has been thoroughly investigated in the FDI community (Basseville and Nikiforov, 1993). It can be stated in statistical or in geometric terms, making use of the available noise and disturbance models.

Example With OBS as in section 2.1, $OS = (1, 0, 1)^T$. 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 is to compare the observation signature with the fault signatures. A solution to this decision-making problem is to define a *consistency criterion* as follows:

Definition 2.7. An observation signature OS is consistent with a fault signature FS if and only if $OS_i = FS_i$ for all i.

Definition 2.8. The *diagnostic sets* are given by the faults whose signatures are consistent with the observation signature.

Example

OS	Diagnoses
$(1, 0, 1)^T$	{A1}, {M1}, {A1, M1}
$(1, 1, 0)^T$	{M2}
$(1, 1, 1)^T$	any multiple fault except {A1, M1} and {A2, M3}

Note that, since noise or perturbations may cause decision errors, 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 DIAGNOSIS

Reiter (1987) proposed a logical theory of diagnosis, which is often referred as diagnosis from first principles. Given a description of a system together with observations which conflict with the way the system is meant to behave, the problem is to determine those components of the system whose abnormal functioning could explain the discrepancy between the observed and correct behaviors. This approach, referred also as the consistency-based approach, was later extended and formalized in (de Kleer *et al.*, 1992). In the following we refer to these basic definitions without considering posterior extensions and refinements.

3.1. The system model

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 system components, is a finite set of constants. SD uses a distinguished predicate AB, interpreted to mean abnormal : for a given c of COMPS, ¬AB(c) describes the case where component c behaves correctly.

Note that with the AB predicate, the DX approach makes explicit the fact that a formula in SD describes the normal behavior of a given component.

Example COMPS = {A1, A2, M1, M2, M3}. SD = {ADD(x) ∧ ¬AB(x) ⇒ Output(x) = Input1(x) + Input2(x), MULT(x) ∧ ¬AB(x) ⇒ Output(x) = Input1(x) × 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)}.

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

Example 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 is a set of observations.

Definition 3.4. A *diagnosis* for (SD, COMPS, OBS) is a set of components Δ ⊆ COMPS such that SD ∪ OBS ∪ {AB(c) | c ∈ Δ} ∪ {¬AB(c) | c ∈ COMPS \ Δ} is satisfiable. A *minimal diagnosis* is a diagnosis Δ such that ∀ Δ' ⊂ Δ, Δ' is not a diagnosis.

Following the parcimony principle, minimal diagnoses are preferred. A method based upon the concept of conflict set has been proposed in Reiter (1987) to generate minimal diagnoses. It is at the basis of most of the implemented DX algorithms.

3.2. Conflicts and diagnosis

Definition 3.5. An *R-conflict* for (SD, COMPS, OBS) is a set of components C ⊆ COMPS such that SD ∪ OBS ∪ {¬AB(c) | c ∈ 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

f	g	Minimal R-conflicts
10	12	{A1, M1, M2}, {A1, A2, M1, M3}
10	10	{A1, M1, M2}, {A2, M2, M3}
10	14	{A1, M1, M2}, {A2, M2, M3}, {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 (Reiter, 1987).

Proposition 3.1. Δ is a minimal diagnosis for (SD, COMPS, OBS) if and only if Δ is a minimal hitting set for (i.e. intersects any set of) the collection of (minimal) R-conflicts for (SD, COMPS, OBS).

Example

f	g	Minimal diagnoses
10	12	{A1}, {M1}, {A2, M2}, {M2, M3}
10	10	{M2}, {A1, A2}, {A1, M3}, {A2, M1}, {M1, M3}
10	14	{A1, A2}, {A1, M2}, {A1, M3}, {A2, M1}, {A2, M2}, {M1, M2}, {M1, M3}, {M2, M3}

4. A COMPARISON FRAMEWORK

4.1. The basic notions: ARR and Conflicts

In both approaches, diagnosis is triggered when discrepancies occur between the modeled (correct) behavior and the observations (OBS). The DX notion of R-conflict is related to the FDI notion of ARR structure. If the components which correspond to some ARR structure behave correctly, the ARR is satisfied by any OBS. This is a consequence of the way ARRs are built from the models of these components. The violation of some ARR by one OBS incriminates at least one of its structure components. It is thus clear that the structure of each ARR is a *potential R-conflict*, in the sense that, each time the ARR is violated by a given OBS, it becomes an R-conflict for (SD, OBS).

Given the signature matrix of the FDI approach which crosses ARRs in rows and sets of components in columns, the FDI theory compares the observation signature to the fault signatures whereas DX considers separately each line corresponding to a violated ARR, isolating R-conflicts before searching for a common explanation. These approaches can be referred as the *column view* and the *line view* respectively. From the computational point of view, the main difference between the FDI and DX approaches is that in FDI most of the computational work is done off-line. From the only knowledge of observed variables, i.e. sensor locations, modeling knowledge is compiled: ARRs are obtained by combining model equations or constraints, and eliminating unknown variables. The only thing that has to be done on-line, i.e. when a given OBS is acquired, is to compute the truth value (w.r.t. OBS) of each ARR and to compare the obtained observation signature with the faults theoretical signatures (columns of the signature matrix). In terms of R-conflicts, it means that potential R-conflicts are compiled and that, for any given OBS, R-conflicts are exactly those potential R-conflicts which are the structures of ARRs which are not satisfied by OBS.

4.2. The basic assumptions

The originality and the power of both FDI and consistency-based DX approaches result from the fact that they use only the correct behavior of the components: no model of faulty behavior is needed. Nevertheless, different assumptions are by default adopted by each approach, leading to different computations of the diagnoses, which explains the different results obtained on the example (compare sections 2 and 3). These assumptions concern: 1) the manifestations of the faults through observations and 2) the case of simultaneous faults and of their interaction.

Single fault exoneration assumption In the DX approach, absolutely no assumption is made about how a component may behave when it is faulty. It is in particular admitted that faults may be not detected by some ARRs in which they are involved. This ensures the fundamental property of the DX approach, i.e. its logical soundness. In the matrix framework, this means that, for any given OBS, only those rows (ARRs) which are not satisfied by OBS are considered. For each one, the singleton columns (components) having a 1 in this row constitute the associated R-conflict. Possible diagnoses are built from these R-conflicts.

Conversely, the FDI approach is based on a direct reasoning about the effects of a fault (column) - viewed as a violation of the correct behavioral model of the corresponding component - on the ARRs (rows). In addition to the obvious fact that a fault cannot affect an ARR in which it is not involved, i.e. a row having a 0 in the corresponding column, which is the direct form of the reasoning used in DX, the idea is that a fault is assumed to affect the ARRs in which it is involved, i.e. the rows having a 1 in the corresponding column, causing them to be violated by any given OBS. Hence, given OBS, not only, as in DX, any column involved in a violated row is a fault candidate, but also any column involved in a satisfied ARR is implicitly *exonerated* (satisfied rows are thus also used in the reasoning).

Example With f = 10 and g = 12, A2, M2 and M3 are exonerated. With f = 10 and g = 10, A1, A2, M1 and M3 are exonerated.

In fact this result is *not sound* from a logical point of view. However, the exoneration assumption which is implicitly made in the FDI approach is justified as far as structural properties, i.e. properties which hold for *almost every single fault,* are considered. Indeed, *non-detectable single faults* (i.e single faults which do not violate some ARRs in which they are involved, if equality-based consistency criterion is used) are highly singular, in the sense they satisfy some specific property.

Example From this point of view, the polybox is a standard example: in any of the three observation cases, single fault diagnoses obtained by FDI and DX are identical, i.e. single fault exoneration assumption is licit here.

It has to be emphasized that results concerning non-detectability are available in the FDI community (Basseville and Nikiforov, 1993) (Staroswiecki and Comtet-Varga, 1999).

Multiple faults and the no-compensation assumption. In the DX approach, by default, there is no limitation on the number of possible simultaneous faults: minimal diagnoses are built as minimal hitting sets of the collection of minimal R-conflicts and are not limited in size. Single and multiple faults are thus dealt with in exactly the same framework. This means also that absolutely no assumption is made about how simultaneous faults may interact. It is in particular admitted that several faults may compensate each other, resulting in some ARRs in which they are involved to be satisfied.

Conversely, the single fault assumption is frequently adopted in many FDI applications. This results from practical considerations, since it happens frequently that multiple faults have a very low probability, and not considering them drastically simplifies the computation. When this hypothesis is not realistic, the FDI approach requires possible multiple faults to be identified and the way they combine their effects in the ARRs to be anticipated. The theoretical signatures of multiple faults are generally obtained from the signatures of single faults (cf. section 2.3) according to the intuitive idea that a multiple fault may affect an ARR if and only if at least one of the single faults it is made up of may affect this ARR. The new column J corresponding to the multiple fault C_J = {C_{j1}, ..., C_{jk}} must thus have a 1 in any row i if and only if at least one of the {C_{jl}} columns has a 1 in i, i.e. J has to verify for all i : $s_{i\ J} = 1$ iff □l $1 \leq l \leq k$ such that $s_{i\ jl} = 1$. Clearly, this way of defining the extended signature matrix matches the line view and the hitting set property of DX. Notice that within this framework the implicit exoneration assumption used in FDI means not only that single faults are detectable but also that several faults can never compensate each other.

Example In the two cases f = 10, g = 12 and f = 10, g = 10, all minimal multiple fault diagnoses obtained by the DX approach are exonerated by the FDI approach, due to the no-compensation assumption.

Thus, once more, the FDI approach is not logically sound but it is indeed justified from the point of view of structural properties (valid for almost every fault).

Example With f = 10 and g = 12, the double fault DX diagnoses {A2, M2} and {M2, M3} correspond to the following exceptional compensation cases: M2: 2 □ 3 = 4, A2: 4 + 6 = 12 and M2: 2 □ 3 = 4, M3: 3 □ 2 = 8, respectively.

5. CONCLUSION

The first goal of FDI has been fault detection and associated decision procedures. Its main interest was in sophisticated techniques to combine measurements such as observers and filters. On the other hand, DX focused on fault isolation by recognizing subsets of the system description conflicting with the observation. Our study proves that a significant part of both theories fit in a common framework which allows a precise comparison, but that the underlying assumptions are different in both communities, which is explained by the fact that logical soundness is the aim on one hand while structural properties are seeked on the other hand.

Releasing the exoneration and the no-compensation assumptions can be done in the FDI framework by expressing that a fault may or may not affect the ARRs in which it is involved. The symbol X can be used for that, instead of 1. In this way, having no exoneration assumption boils down to have only 0 and X symbols in the signature matrix. A fault has thus a signature column with X in violated rows and 0 or X in satisfied rows. When comparing with the observed signature, X can match either 0 or 1. In this case FDI and DX views agree on diagnoses (Cordier *et al.*, 1999). This opens the possibility of a fruitful cooperation between these two approaches of diagnosis.

Some points have been left out of the boarders of this comparison. On one hand, there is presently no equivalent in DX of the notion of noise. Conversely, the systematic use of fault models which is available in DX has no counterpart in the present FDI view, and has been left out of this framework. Further studies are needed to integrate these aspects, which would be beneficial to both communities.

REFERENCES

Basseville M., Nikiforov I.V. (1993), *Detection of abrupt changes, Theory and applications*, Prentice Hall, Information and System Sciences Series.

Control Engineering Practice (CEP) (1997), Special volume on Supervision, fault detection, and diagnosis of technical systems, 5(5).

Cordier M.-O. *et al.* (1999) Conflicts versus Analytical Redundancy Relations - A comparative analysis of the MBD approach from the AI and Automatic Control perspectives, *Internal report, IMALAIA group.*

De Kleer J., Williams B. C. (1987), Diagnosing multiple faults, *Artificial Intelligence* 32(1), p. 97-130.

De Kleer J., Mackworth A., Reiter R. (1992), Characterizing diagnoses and systems, *Artificial Intelligence* 56(2-3), p. 197-222.

Frank P.M. (1996), Analytical and qualitative model-based fault diagnosis□□A survey and some new results, *European Journal of Control* 2, p. 6-28.

Gertler J. J. (1991), Analytical redundancy methods in fault detection and isolation- a survey and synthesis, *IFAC/IMACS Safeprocess,* Baden-Baden, 1, p. 9-21.

Hamscher W., Console L., de Kleer J. (eds.) (1992), *Readings in Model-Based Diagnosis*, Morgan Kaufmann, San Mateo, CA.

Patton R. J., Chen J. (1991), A review of parity space approaches to fault diagnosis, *IFAC/IMACS Safeprocess*, Baden-Baden.

Reiter R. (1987), A theory of diagnosis from first principles, *Artificial Intelligence* 32(1), p. 57-96.

Staroswiecki M., Declerck Ph. (1989), Analytical redundancy in non-linear interconnected systems by means of structural analysis, *IFAC Symposium on Advanced Information Processing in Automatic Control*, Nancy, II, p. 23-27.

Staroswiecki M., Comtet-Varga G. (1999), Fault Detectability and Isolability in Algebraic Dynamic Systems, *European Control Conference ECC'99*, Karlsruhe.

Staroswiecki M., Comtet-Varga G. (2000), Design of Structured Residuals in Algebraic Dynamic Systems, *IFAC/IMACS Safeprocess*, Budapest.