ELSEVIER

Information Sciences 121 (1999) 349–365

INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL

www.elsevier.com/locate/ins

Comments on the theory of measurement in diagnosis from first principles

Benjamin Han, Shie-Jue Lee *, Hsin-Tai Yang

Department of Electrical Engineering, National Sun Yat-Sen University, Kaohsiung 80424, Taiwan, ROC

Received 12 March 1997; received in revised form 17 January 1999; accepted 25 June 1999

Abstract

When finding diagnoses from first principles, one needs to refine possible diagnoses by making measurements from the system being diagnosed. Based on the work of Reiter, Hou has developed and formalized an efficient incremental method for computing all diagnoses upon measurement. However, we feel that some points in Hou's paper need clarifications. In this paper, we describe an elaborate picture of the relationships among measurements, conflict sets and diagnoses. We also present some comments on the equivalence relation and Hou's procedure for conflict recognition. © 1999 Elsevier Science Inc. All rights reserved.

Keywords: Propositional logic; Observations; Measurements; Conflict sets; Diagnosis; Equivalence class

1. Introduction

Under the assumption that the description of a system is consistent, if the inputs and the outputs, i.e., the observation, of the system conflict with the way the system is meant to behave, the diagnostic problem is to pinpoint the possible diagnoses, i.e., the possible sets of faulty components that cause the misbehavior. Many researchers have proposed various kinds of approaches to

* Corresponding author. Fax: +886-7 525 4012.
E-mail address: leesj@ee.nsysu.edu.tw (S.-J. Lee).

0020-0255/99/$ - see front matter
PII: S 0 0 2 0 - 0 2 5 5 (9 9) 0 0 0 3 4 - 1

tackle the problem; among them, Reiter [4] has built and formalized the major theorems of diagnosis from first principles upon the work of de Kleer [1] and Genesereth [2]. In the paper, Reiter gave the following definitions:

Definition 1.1 ([4, *Definition* 2.1]). A *system* is a pair (SD, COMP) where
1. SD, the *system description*, is a set of first-order sentences describing
   (a) the functionality of a component within the system; e.g., let A be an AND gate with two inputs, the sentence describing it is:

   $$\neg AB(A) \supset \text{out}(A) = \text{and}(\text{in1}(A),\ \text{in2}(A)),$$

   where the predicate AB stands for *abnormal*, which is *true iff* gate A is malfunctioning, and 'and' is a function returning *true iff* both inputs of gate A are *true*.
   (b) the connections between the components of the system; e.g., $\text{out}(A) = \text{in1}(B)$ means that the output of component A is connected with the first input of component B.
2. COMP is a finite set of constants denoting the collection of components of the system.

Definition 1.2 ([4, *Proposition* 3.4]). Δ is a *diagnosis* for (SD, COMP, OBS) *iff* Δ is a minimal set such that SD $\cup$ OBS $\cup$ $\{\neg\text{AB}(c)|c \in \text{COMP} - \Delta\}$ is consistent, where OBS stands for the observation of the system (SD, COMP).

Definition 1.3 ([4, *Definition* 4.1]). $C \subseteq$ COMP is a *conflict set* (CS) for (SD, COMP, OBS) *iff* SD $\cup$ OBS $\cup$ $\{\neg\text{AB}(c)|c \in C\}$ is inconsistent. A *minimal* conflict set (MCS) is a CS such that none of its subsets is a CS. [1]

Definition 1.4 ([4, *Definition* 4.3]). H is a *hitting set* (HS) for a collection of sets S *iff* $H \subseteq \bigcup_{C\in S} C$ such that $\text{H} \cap C \neq \emptyset$ for each $C \in S$. A *minimal* hitting set (MHS) is an HS such that none of its subsets is an HS.

To locate all possible diagnoses, Reiter showed that all we have to do is find all MHSs for the collection of MCSs, as stated in the following theorem.

Theorem 1.5 ([4, Corollary 4.5]). *$\Delta \subseteq$ COMP is a diagnosis for* (*SD*, *COMP*, *OBS*) *iff Δ is an MHS for the collection of MCSs for* (*SD*, *COMP*, *OBS*).

[1] For brevity, through out the rest of this paper we will use the initials CS for *conflict set*, MCS for *minimal conflict set*, HS for *hitting set*, and MHS for *minimal hitting set*.

However, to refine further the set of diagnoses, one needs to take measurements from the system. Let Π be a measurement. Each diagnosis then predicts either Π or $\neg\Pi$.

Definition 1.6 ([4, *Proposition* 5.3]). A diagnosis Δ for (SD, COMP, OBS) predicts Π *iff*

$$\text{SD} \cup \text{OBS} \cup \{\neg AB(c) \mid c \in \text{COMP} - \Delta\} \models \Pi.$$

It turns out that each diagnosis is either preserved or rejected according to the prediction, as stated in the following theorem.

Theorem 1.7 ([4, Theorem 5.7]). *Suppose that every diagnosis for* (*SD*, *COMP*, *OBS*) *either predicts* Π *or* $\neg\Pi$, *then*:
1. *Every diagnosis for* (*SD*, *COMP*, *OBS*) *which predicts* Π *is a diagnosis for* (*SD*, *COMP*, $OBS \cup \{\Pi\}$).
2. *A diagnosis for* (*SD*, *COMP*, *OBS*) *which predicts* $\neg\Pi$ *is not a diagnosis for* (*SD*, *COMP*, $OBS \cup \{\Pi\}$).
3. *Any diagnosis for* (*SD*, *COMP*, $OBS \cup \{\Pi\}$) *which is not a diagnosis for* (*SD*, *COMP*, *OBS*) *is a strict superset of some diagnosis for* (*SD*, *COMP*, *OBS*) *which predicts* $\neg\Pi$; *i.e.*, *any new diagnosis resulting from the new measurement* Π *will be a strict superset of some old diagnosis which predicts* $\neg\Pi$.

Although Theorem 1.7 sheds some light on the practical way for computing all diagnoses after adding Π, it is Hou [3] who has developed and formalized an efficient incremental approach for computing them. However, we feel that some points in Hou's paper need clarifications. In this paper, we supplement and revise some theorems and procedures described in [3]. In Section 2, we provide a concise review on the relationships between measurements and CSs described in [3]. Then in Section 3, we show an elaborate picture of the relationships among measurements, CSs and diagnoses by supplementing the theorems in [3]. In Section 4, we give a concise review of Hou's consequences about conflict recognition. Then in Section 5, we discuss the equivalence relation and give comments on Hou's conflict recognition.

2. Hou's consequences about measurements, conflict sets, and diagnoses

To refine further the set of possible diagnoses by a measurement Π, Hou defined a CS for $(\text{SD}, \text{COMP}, \text{OBS} \cup \{\Pi\})$ *resulting from the measurement* Π as follows.

Definition 2.1 ([3, *Definition* 3.2]). A *CS* for $(\text{SD}, \text{COMP}, \text{OBS} \cup \{\Pi\})$ *resulting from a measurement* Π is a set $\{c_1, c_2, \ldots, c_k\} \subseteq \text{COMP}$ such that

$$\text{SD} \cup \text{OBS} \cup \{\neg AB(c_i) | i = 1, \ldots, k\}$$

is *consistent*, and

$$\text{SD} \cup \text{OBS} \cup \{\Pi\} \cup \{\neg AB(c_i) | i = 1, \ldots, k\}$$

is *inconsistent*.

By combining Definitions 2.1 with 1.3 and 1.6, Hou presented the following important relationship between the diagnoses for (SD, COMP, OBS) predicting $\neg\Pi$ and the CSs for $(\text{SD}, \text{COMP}, \text{OBS} \cup \{\Pi\})$ resulting from Π.

Corollary 2.2 ([3, Proposition. 3.3]). *If* $\Delta \subseteq \text{COMP}$ *is a diagnosis for* (*SD*, *COMP*, *OBS*) *which predicts* $\neg\Pi$, *then* COMP-Δ *is a CS for* (*SD*, *COMP*, *OBS* $\cup \{\Pi\}$) *resulting from* Π.

On the other hand, every MCS for (SD, COMP, OBS $\cup \{\Pi\}$) resulting from Π is a subset of COMP-Δ, where Δ is a diagnosis for (SD, COMP, OBS) predicting $\neg\Pi$.

Theorem 2.3 ([3, Theorem 3.6]). *Every MCS for* (*SD*, *COMP*, *OBS* $\cup \{\Pi\}$) *resulting from* Π *is one of the MCSs derived from COMP-*Δ_i, *which is a CS for* (*SD*, *COMP*, *OBS* $\cup \{\Pi\}$) *resulting from* Π, *with* Δ_i *being a diagnosis for* (*SD*, *COMP*, *OBS*) *predicting* $\neg\Pi$.

Corollary 2.2 and Theorem 2.3 together show that by deriving all MCSs from COMP-Δ_i, where Δ_i is a diagnosis for (SD, COMP, OBS) predicting $\neg\Pi$, one can get all, nothing more than all, MCSs for $(\text{SD}, \text{COMP}, \text{OBS} \cup \{\Pi\})$ resulting from Π. Moreover, the set of MCSs for $(\text{SD}, \text{COMP}, \text{OBS} \cup \{\Pi\})$ can be divided into two collections, as stated below.

Theorem 2.4 ([3, Theorem 3.7]). *Any MCS for* (*SD*, *COMP*, *OBS* $\cup \{\Pi\}$) *is either an MCS for* (*SD*, *COMP*, *OBS*) *or an MCS for* (*SD*, *COMP*, *OBS* $\cup \{\Pi\}$) *resulting from* Π.

This theorem gives an interesting glimpse about the relationships between the MCSs for (SD, COMP, OBS) and the MCSs for $(\text{SD}, \text{COMP}, \text{OBS} \cup \{\Pi\})$. However, in the rest of [3], no further explanation regarding the relationships is given.

3. Relationships among measurements, conflict sets, and diagnoses

First we show that the MCSs for (SD, COMP, OBS) can also be divided into two collections.

Theorem 3.1. *Any MCS for (SD, COMP, OBS) is either an MCS for (SD, COMP,* OBS $\cup \{\Pi\}$) *or a strict superset of some MCS for (SD, COMP, OBS* $\cup \{\Pi\}$) *resulting from* Π.

Proof. Let $P = \{c_1, \ldots, c_k\}$ be an MCS for (SD, COMP, OBS) and $P_i = P - \{c_i\}$. From Definition. 1.3 we know that

$$\text{SD} \cup \text{OBS} \cup \{\neg\text{AB}(c) \mid c \in P\} \text{ is inconsistent,}$$

and

$$\text{SD} \cup \text{OBS} \cup \{\neg\text{AB}(c) \mid c \in P_i\} \text{ is consistent.}$$

1. If SD $\cup$ OBS $\cup$ $\{\Pi\}$ $\cup$ $\{\neg\text{AB}(c) \mid c \in P_i\}$ is inconsistent, then by Definition 2.1, P_i is a CS for (SD, COMP, OBS $\cup$ $\{\Pi\}$) resulting from Π. Any MCS derived from P_{i} then is a strict subset of P.
2. If

   $$\text{SD} \cup \text{OBS} \cup \{\Pi\} \cup \{\neg AB(c) | c \in P_i\}$$

   is consistent, since

   $$\text{SD} \cup \text{OBS} \cup \{\Pi\} \cup \{\neg\text{AB}(c) | c \in P\}$$

   is inconsistent, then by Definition 1.3 P is an MCS for (SD, COMP, OBS $\cup$ $\{\Pi\}$). □

Given both Theorems 2.4 and 3.1, one may wonder what relationships exist between the two collections of MCSs for (SD, COMP, OBS) and the two collections of MCSs for (SD, COMP, OBS $\cup$ $\{\Pi\}$). Fig. 1 shows the relationships depicted by Theorems 2.4 and 3.1.

In the figure, a circle stands for a set of MCSs. The lower circle is the set of MCSs after adding Π. The section denoted by 'NEW' is the collection of MCSs for (SD, COMP, OBS $\cup$ $\{\Pi\}$) resulting from Π. Fig. 1(c) in effect states the following corollary.

Corollary 3.2. *Let* C_1 *be the collection of MCSs for (SD, COMP, OBS) and* C_2 *be the collection of MCSs for (SD, COMP, OBS* $\cup \{\Pi\}$). *Then* C_1 *can be partitioned into* C' *and* C_{11}, *and* C_2 *can be partitioned into* C' *and* C_{21}, *where* C_{21} *is the collection of MCSs for (SD, COMP, OBS* $\cup \{\Pi\}$) *resulting from* Π and $\forall c \in C_{11}$, *c is a strict superset of some* c' *in* C_{21}.

It is easy to see from Fig. 1(c) that a diagnosis for (SD, COMP, OBS), if it intersects all MCSs for (SD, COMP, OBS $\cup$ $\{\Pi\}$) resulting from Π, must be an HS for the collection of MCSs for (SD, COMP, OBS $\cup$ $\{\Pi\}$). Furthermore, Hou [3, Corollary 3.12] showed that such a diagnosis for (SD, COMP, OBS) is also a diagnosis for (SD, COMP, OBS $\cup$ $\{\Pi\}$). On the other hand,

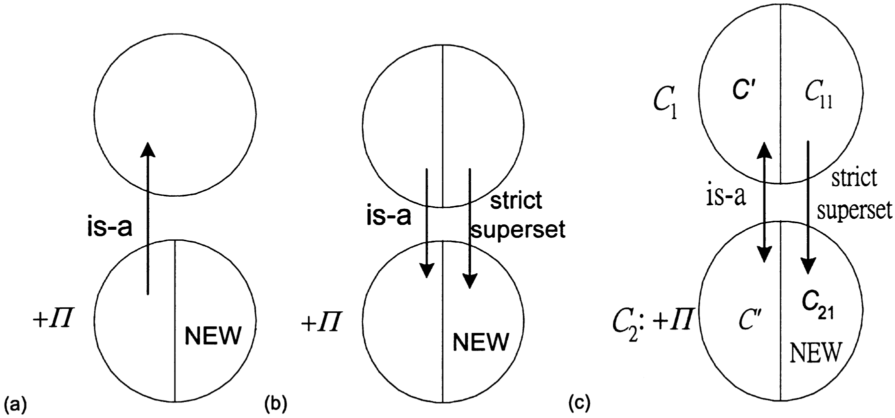

Fig. 1. The relationships between the MCSs before adding Π and the MCSs after adding Π depicted by: (a) Theorem 2.4; (b) Theorem 3.1; (c). Theorem 2.4 and Theorem 3.1.

in Theorem 1.7 (1), Reiter has shown that if a diagnosis for (SD, COMP, OBS) predicts Π, then it is also a diagnosis for (SD, COMP, OBS $\cup$ $\{\Pi\}$). To relate these two viewpoints, we have the following theorem:

Theorem 3.3. *Let Δ be a diagnosis for (SD, COMP, OBS). Then Δ predicts Π iff it intersects each MCS for (SD, COMP, OBS $\cup \{\Pi\}$) resulting from Π.*

Proof. ($\Rightarrow$) Since Δ predicts Π, from Definition 1.6 we have

$$\text{SD} \cup \text{OBS} \cup \{\Pi\} \cup \{\neg\text{AB}(c) \mid c \in \text{COMP} - \Delta\} \text{ is consistent,}$$

i.e., COMP-Δ is not a CS for (SD, COMP, OBS $\cup$ $\{\Pi\}$). Assume that P is an MCS for (SD, COMP, OBS $\cup \{\Pi\}$) resulting from Π and $P \cap \Delta = \emptyset$, then $P \subseteq$ COMP-Δ, i.e., COMP-Δ is a CS for (SD, COMP, OBS $\cup$ $\{\Pi\}$), a contradiction.

($\Leftarrow$) Assume that Δ predicts $\neg\Pi$, then we have

$$\text{SD} \cup \text{OBS} \cup \{\Pi\} \cup \{\neg\text{AB}(c) \mid c \in \text{COMP} - \Delta\} \text{ is inconsistent,}$$

i.e., COMP-Δ is a CS for (SD, COMP, OBS $\cup$ $\{\Pi\}$). Since all MCSs for (SD, COMP, OBS $\cup$ $\{\Pi\}$) derived from COMP-Δ do not intersect Δ, we have a contradiction. □

Therefore the above result shows that Theorem 1.7 is in fact equivalent to Corollary 3.12 in [3].

The picture of the relationships between the diagnoses before adding Π and the diagnoses after adding Π can be drawn directly from Theorem 1.7 (See Fig. 2).

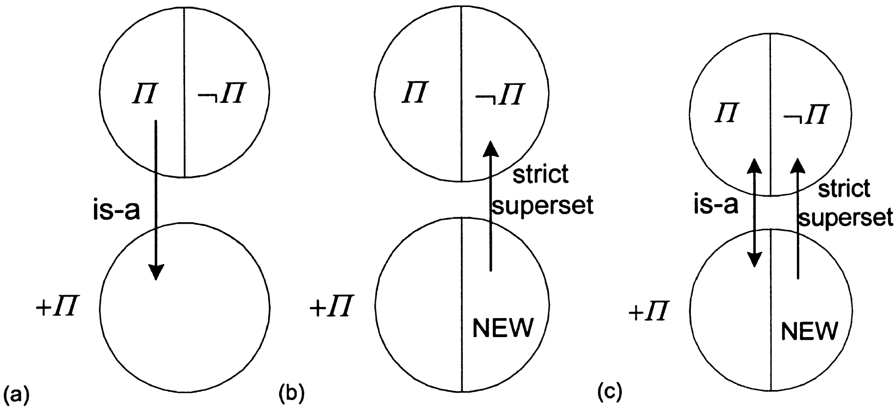

Fig. 2. The relationships between the diagnoses before adding Π and the diagnoses after adding Π depicted by: (a) Theorem 1.7 (1); (b) Theorem 1.7 (2) and (3); (c) Theorem 1.7 as a whole.

In the figure, a circle stands for a set of diagnoses. The upper diagnoses are divided into those predicting Π and those predicting $\neg\Pi$. The lower circle is the set of diagnoses after adding Π. The section denoted by 'NEW' is the collection of diagnoses for (SD, COMP, OBS $\cup$ $\{\Pi\}$) but not for (SD, COMP, OBS).

Fig. 2 essentially shows the way to compute all diagnoses for (SD, COMP, OBS $\cup$ $\{\Pi\}$). What we need to do is preserve all diagnoses for (SD, COMP, OBS) predicting Π, and then compute all 'new' diagnoses from the diagnoses for (SD, COMP, OBS) predicting $\neg\Pi$. To compute the new diagnoses, Hou introduced the idea of *candidates* as follows: [2]

Definition 3.4. A *candidate* k for (SD, COMP, OBS $\cup$ $\{\Pi\}$) is a set of components such that $k = \Delta \cup h$, where Δ is a diagnosis for (SD, COMP, OBS) predicting $\neg\Pi$ and h is an MHS for C, which is a collection of MCSs for (SD, COMP, OBS $\cup$ $\{\Pi\}$) resulting from Π such that $\forall P \in C$, P does not intersect Π.

An immediate consequence follows from the definition above:

Corollary 3.5. *A candidate k for* (*SD*, *COMP*, *OBS* $\cup$ $\{\Pi\}$) *is an HS for the collection of MCSs for* (*SD*, *COMP*, *OBS* $\cup$ $\{\Pi\}$).

[2] In [3], Hou did not give a formal definition for *candidates*, but he adopted the term 'candidate' for the same use. However, we find it helpful to single out the definition.

Proof. Let $k = \Delta \cup h$. By Theorem 3.3 we know that Δ does not intersect some MCSs for $(\mathrm{SD}, \mathrm{COMP}, \mathrm{OBS} \cup \{\Pi\})$ resulting from Π. Let C denote such collection of MCSs. But we know that h is an HS for C. □

Next we show that every new diagnosis can be generated by some candidate.

Theorem 3.6. *Let Δ be a diagnosis for $(SD, COMP, OBS \cup \{\Pi\})$ but not for $(SD, COMP, OBS)$. Then there exists a candidate k for $(SD, COMP, OBS \cup \{\Pi\})$ such that $k = \Delta$.*

Proof. By Theorem 1.7 (3), we know that $\Delta = \Delta_i \cup \Delta_j$, where $\Delta_i \cap \Delta_j = \emptyset$ and Δ_i is a diagnosis for (SD, COMP, OBS) predicting $\neg\Pi$. Since there must exist some candidate k such that $k = \Delta_i \cup h$, where h is an MHS for C, the collection of MCSs derived from COMP-Δ_i, all we have to do is prove that $\Delta_j = h$.

Assume that Δ_j is not an HS for C, then there must exist some $P \in C$ such that $P \cap \Delta_j = \emptyset$. Since Δ_i does not intersect any $P \in C$, Δ could not intersect every MCS for $(\mathrm{SD}, \mathrm{COMP}, \mathrm{OBS} \cup \{\Pi\})$, a contradiction. Δ_j is also minimal, for if not, Δ could not be minimal. □

As a matter of fact, the above result is equivalent to the theorem Hou [3, Theorem 3.13] suggested for computing all new diagnoses based on a measurement Π: (1) if a diagnosis for (SD, COMP, OBS) intersects all MCSs for $(\mathrm{SD}, \mathrm{COMP}, \mathrm{OBS} \cup \{\Pi\})$ resulting from Π, then it must be a diagnosis for $(\mathrm{SD}, \mathrm{COMP}, \mathrm{OBS} \cup \{\Pi\})$ as well; (2) if a diagnosis for (SD, COMP, OBS) does not hit some MCSs, which form a collection C, for $(\mathrm{SD}, \mathrm{COMP}, \mathrm{OBS} \cup \{\Pi\})$ resulting from Π, then we simply compute the candidates, i.e., the union of the diagnosis and the MHSs for the collection C, and replace the diagnosis with all candidates which are not subsumed or duplicated by another candidate or diagnosis already generated in case (1). However, in the proof of the theorem, Hou did not show that every new diagnosis can be generated by at least one candidate.

It is interesting to know why a candidate could be non-minimal. Let Δ be a diagnosis for (SD, COMP, OBS) predicting $\neg\Pi$, and h be an MHS for C, the collection of MCSs which Δ does not intersect. Intuitively h patches the 'hole' that Δ misses. However, it is possible that h not only makes candidate $k(k = \Delta \cup h)$ intersect each MCS in C, but also introduce for Δ more 'contact' with some MCS that Δ already touches. The following theorem states a sufficient condition under which the candidate must be *minimal*.

Theorem 3.7. *Let k be a candidate for $(SD, COMP, OBS \cup \{\Pi\})$ such that $k = \Delta_i \cup h$, where Δ_i is a diagnosis for $(SD, COMP, OBS)$ predicting $\neg\Pi$ and h is an MHS for C, the collection of MCSs derived from COMP-Δ_i. If $P \cap h = \emptyset$, where P is an MCS for $(SD, COMP, OBS)$, then k is minimal.*

Proof. Since Δ_i is a diagnosis for (SD, COMP, OBS), $\Delta_i - \{c\}$ does not intersect some MCS P for (SD, COMP, OBS), where $c \in \Delta_i$. If P is also an MCS for $(\text{SD}, \text{COMP}, \text{OBS} \cup \{\Pi\})$ and $P \cap h = \emptyset, (\Delta_i - \{c\}) \cup h$ cannot intersect P. Therefore $k - \{c\}$ cannot intersect P.

If P is not an MCS for (SD, COMP, OBS $\cup \{\Pi\}$), then by Theorem 3.1, P must be a strict superset of some P', where P' is an MCS for $(\text{SD}, \text{COMP}, \text{OBS} \cup \{\Pi\})$ resulting from Π. Since $\Delta_i - \{c\}$ does not intersect P' and $P' \cap h = \emptyset, (\Delta_i - \{c\}) \cup h$ cannot intersect P'. Therefore $k - \{c\}$ cannot intersect any such P'. □

4. Hou's enhanced conflict recognition

In [3], Hou showed that to compute all new diagnoses for $(\text{SD}, \text{COMP}, \text{OBS} \cup \{\Pi\})$, one has to derive all MCSs from COMP-Δ_i, where Δ_i is a diagnosis for (SD, COMP, OBS) predicting $\neg\Pi$. The process of the derivation requires a test for each subset of COMP-Δ_i by calling a propositional calculus prover, therefore it needs substantial computation power. To ease the situation, Hou showed that some Δ_i can be safely ignored from the derivation process by first introducing the notion of *equivalence* between two components and *homogeneity* between two diagnoses.

Definition 4.1 ([3, *Definition* 4.1]). For the system (SD, COMP, OBS), $\forall c_1, c_2 \in \text{COMP}$, c_1 and c_2 are defined to be *equivalent* if $c_1 \in P \leftrightarrow c_2 \in P$, where P is an MCS for (SD, COMP, OBS). This relation is indicated by $c_1 \sim c_2$.

Definition 4.2 ([3, *Definition* 4.2]). Two diagnoses Δ_1 and Δ_2 for (SD, COMP, OBS) are defined to be *homogeneous* if $\forall c_i \in \Delta_1$, there must exist some $c_j \in \Delta_2$ such that $c_i \sim c_j$, and $\forall c_j \in \Delta_2$, there must exist some $c_i \in \Delta_1$ such that $c_j \sim c_i$. This relation is indicated by $\Delta_1 \sim \Delta_2$.

Note that both the equivalence relation and the homogeneity relation are symmetric, transitive and reflexive. Next Hou showed that under certain condition, a homogeneous diagnosis predicting $\neg\Pi$ can be safely removed from the derivation process without losing any MCSs.

Theorem 4.3 ([3, Theorem 4.5]). *Suppose that Δ_1 and Δ_2 are two diagnoses for (SD, COMP, OBS) which predict $\neg\Pi$ and $\Delta_1 \sim \Delta_2$. If all MCSs for (SD, COMP, OBS $\cup \{\Pi\}$) derived from COMP-Δ_1 do not intersect Δ_2, then all MCSs derived from COMP-Δ_1 and COMP-Δ_2, respectively, are the same ones.*

The above theorem provides an efficiency enhancement: when two diagnoses Δ_1 and Δ_2 predicting $\neg\Pi$ are homogeneous, if all MCSs derived from COMP-Δ_1 do not intersect Δ_2, one can safely ignore Δ_2, i.e., one does not need to derive any MCS from COMP-Δ_2. However, to exploit Theorem 4.3, one has to maintain equivalence classes of components and homogeneity classes of diagnoses when adding more measurements. The following theorem points out the change of the equivalence relation when adding measurement Π.

Theorem 4.4 ([3, Theorem 4.7]). *For every* $\alpha \nsim \beta$ *for* (*SD*, *COMP*, *OBS*), *we still have* $\alpha \nsim \beta$ *for* (*SD*, *COMP*, *OBS* $\cup \{\Pi\}$).

In other words, to maintain equivalence classes of components, one follows the procedure listed below. [3]

Procedure 4.5 [3, p. 307]. Update equivalence classes of components as follows:
1. For every $\alpha \sim \beta$ for (SD, COMP, OBS), if one determines $\alpha \sim \beta$ based on all MCSs $P_1, \ldots, P_m$ for (SD, COMP, OBS $\cup$ $\{\Pi\}$) resulting from Π, then one has $\alpha \sim \beta$ for (SD, COMP, OBS $\cup$ $\{\Pi\}$). If one determines $\alpha \nsim \beta$ based on $P_1, \ldots, P_m$, then one has $\alpha \nsim \beta$ for (SD, COMP, OBS $\cup$ $\{\Pi\}$).
2. For some $\alpha \sim \beta$ for (SD, COMP, OBS), if all $P_1, \ldots, P_m$ do not include both α and β, then one keeps $\alpha \sim \beta$ for (SD, COMP, OBS $\cup$ $\{\Pi\}$).
3. For every $\alpha \nsim \beta$ for (SD, COMP, OBS), one must have $\alpha \nsim \beta$ for (SD, COMP, OBS $\cup$ $\{\Pi\}$), and pay no particular attention to whether $\alpha \sim \beta$ or $\alpha \nsim \beta$ based on $P_1, \ldots, P_m$.
4. For every new $\alpha \sim \beta$ determined by all MCSs $P_1, \ldots, P_m$ resulting from Π, if α and/or β do not occur in any previous equivalence class, namely α and/or β are not the members of any old MCS, then one keeps $\alpha \sim \beta$ for (SD, COMP, OBS $\cup$ $\{\Pi\}$).

To determine homogeneity classes of diagnoses, one simply compares each diagnosis with another, and determines if $\Delta_i \sim \Delta_j$ [3, Section 5.1]. [4]

5. Comments on enhanced conflict recognition

First we show some observations of Definition 4.1.

---

[3] Later in Section 5 we will simplify this procedure.

[4] Later in Section 5 we will describe an improved procedure for updating homogeneity classes.

Corollary 5.1. *If α and β are two components such that no MCS for (SD, COMP, OBS) contains them, then $\alpha \sim \beta$ for (SD, COMP, OBS).*

The above corollary in effect states that each and every component is in one equivalence class from the beginning, regardless whether it hits an MCS or not. However, in [3] it is Hou's intention to ignore all the components which do not hit any MCS from being considered in any equivalence class. This implication does not make any difference in deciding if two diagnoses are homogeneous or not since all the components considered in such decisions actually show up in at least one MCS. Nevertheless, if we follow Definition 3.1 strictly Procedure 4.5 (2) will be rendered redundant, and at the same time we have the following corollary derived directly from Definition. 3.1 and Theorem 4.4.

Corollary 5.2. *An equivalence class will only split into small classes; it will never grow.*

Therefore there cannot exist any *new* $\alpha \sim \beta$ for $(\mathrm{SD}, \mathrm{COMP}, \mathrm{OBS} \cup \{\Pi\})$ described in Procedure 4.5 (4). This result makes Procedure 4.5 (4) redundant as well. Consequently, Procedure 4.5 can be simplified to consist of only two rules as follows.

Procedure 5.3. Update equivalence classes of components as follows:
1. For every $\alpha \sim \beta$ for (SD, COMP, OBS), if one determines $\alpha \sim \beta$ based on all MCSs $P_1, \ldots, P_m$ for $(\mathrm{SD}, \mathrm{COMP}, \mathrm{OBS} \cup \{\Pi\})$ resulting from Π, then one has $\alpha \sim \beta$ for $(\mathrm{SD}, \mathrm{COMP}, \mathrm{OBS} \cup \{\Pi\})$. If one determines $\alpha \nsim \beta$ based on $P_1, \ldots, P_m$, then one has $\alpha \nsim \beta$ for $(\mathrm{SD}, \mathrm{COMP}, \mathrm{OBS} \cup \{\Pi\})$.
2. For every $\alpha \nsim \beta$ for (SD, COMP, OBS), one must have $\alpha \nsim \beta$ for $(\mathrm{SD}, \mathrm{COMP}, \mathrm{OBS} \cup \{\Pi\})$, and pay no particular attention to whether $\alpha \sim \beta$ or $\alpha \nsim \beta$ based on $P_1, \ldots, P_m$.

Now we show an observation concerning diagnoses and the equivalence relation.

Corollary 5.4. *Let Δ be a diagnosis for (SD, COMP, OBS). Then for all distinct d_i, $d_j \in \Delta$, $d_i \nsim d_j$.*

Proof. Assume that there exist distinct $d_i, d_j \in \Delta$ such that $d_i \sim d_j$. Then both d_i and d_j must hit the same MCS for (SD, COMP, OBS), for if not, by Definition 4.1 neither of them is a member of any MCS for (SD, COMP, OBS), and Δ cannot contain them. Without the loss of generality, assume that $\Delta - \{d_i\}$ does not intersect some MCS P. But then P does not contain d_j, and $d_i \nsim d_j$, a contradiction. □

The above corollary then leads us to an useful property for updating the homogeneity relation.

Corollary 5.5. *If* $\Delta_1 \sim \Delta_2$, *then* $|\Delta_1| = |\Delta_2|$.

Proof. Assume that $|\Delta_1| < |\Delta_2|$, then by Definition. 4.2 there exist $d_i \in \Delta_1$ and distinct $d_j, d_k \in \Delta_2$ such that $d_i \sim d_j$ and $d_i \sim d_k$. But then we have $d_j \sim d_k$, a contradiction. □

Corollary 5.5 suggests an improvement on finding homogeneous diagnoses: we do not have to compare two diagnoses with different sizes. A flash of reflection then leads us to another would-be-useful corollary.

Corollary 5.6. *Let n be the number of equivalence classes. Then all diagnoses of size n are homogeneous.*

Proof. From Corollary 5.4 we know that for a diagnosis Δ such that $|\Delta| = n$, every $c \in \Delta$ comes from one of n different equivalence classes. □

It seems that Corollary 5.6 could be used to improve the efficiency of conflict recognition, however, it is not the case. It is interesting to see why rarely the size of a diagnosis reaches the magnitude of n. To be more specific, we shall prove that only when we have the same number of equivalence classes and MCSs, and no two MCSs intersect each other, does the cardinality of a diagnosis equal the number of equivalence classes. First we define an equivalence matrix as follows.

Definition 5.7. Let $\mathrm{MCS}_1, \mathrm{MCS}_2, \ldots, \mathrm{MCS}_m$ be m minimal conflict sets and $\mathrm{EC}_1, \mathrm{EC}_2, \ldots, \mathrm{EC}_n$ be n equivalence classes for (SD, COMP, OBS). An equivalence matrix (EM) for (SD, COMP, OBS) at a particular moment is defined as an $n \times m$ matrix such that the element $e_{i,j}$ at the intersection of the ith row and the jth column is 1 if $\mathrm{EC}_i \subseteq \mathrm{MCS}_j$, otherwise $e_{i,j}$ is 0.

From this point we shall denote the ith row in an EM by R_i, the jth column by C_j and the element at their intersection by $e_{i,j}$.

Let's take an example to show a particular EM for a set of MCSs and ECs.

Example 5.8. Assume that we have three minimal conflict sets:

$$\mathrm{MCS}_1 = \{c_1, c_2, c_3\},\ \mathrm{MCS}_2 = \{c_1, c_2, c_4, c_5\},\ \mathrm{MCS}_3 = \{c_4, c_6\}.$$

The corresponding equivalence classes are

$$\mathrm{EC}_1 = \{c_1, c_2\},\quad \mathrm{EC}_2 = \{c_3\},\quad \mathrm{EC}_3 = \{c_4\},\quad \mathrm{EC}_4 = \{c_5\},$$
$$\mathrm{EC}_5 = \{c_6\},\quad \mathrm{E}C_6 = \{c_7\}.$$

Then we have the equivalence matrix EM as shown below.

$$\begin{array}{c} \\ EC_1 \\ EC_2 \\ EC_3 \\ EC_4 \\ EC_5 \\ EC_6 \end{array} \begin{array}{c} \begin{array}{ccc} MCS_1 & MCS_2 & MCS_3 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{array}.$$

An intuition follows immediately after the definition. Since a diagnosis is an MHS for the collection of MCSs, we can first choose a minimal set Ω of ECs such that Ω 'covers' each MCS, i.e. for each MCS it intersects some EC in Ω, then we pick up one element from each EC in Ω to form a diagnosis. The procedure is described below.

Procedure 5.9. Derive diagnoses by simplifying an EM.

Let EM be an $n \times m$ equivalence matrix for $\mathrm{EC}_1, \ldots, \mathrm{EC}_n$ and $\mathrm{MCS}_1, \ldots, \mathrm{MCS}_m$.

1. Let R denote the first row and Ω be an empty set.
2. If R has no entry with value 1, let R be the next row and return to step 2.
3. Add the corresponding EC of row R to set Ω.
4. For each entry with value 1 in R, delete the corresponding column. Also delete row R itself.
5. If the simplified EM still has elements, advance R to the next row and go back to step 2.
6. For each EC in Ω pick an element and collect them to form a new diagnosis.

Note that in the above procedure we follow a strict top-down order to simplify the equivalence matrix and therefore the final set of diagnoses we have is only a subset of the real solution. Though one can obtain the complete set of diagnoses by furnishing the procedure with additional backtracking mechanism, it suffices for our purpose to prove our claim concerning Corollary 5.6.

Now we give an example showing how to derive (partial) diagnoses by simplifying the EM of Example 5.8.

Example 5.10. Let EM denote the same equivalence matrix shown in Example 5.8. First we observe that the first row has two entries with value 1: the first and the second entry. We then delete the first and the second column together with the first row itself. The simplified EM is shown below:

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Next we observe that the second row of the simplified matrix has an entry with value 1, and delete the only remaining column. The corresponding ECs we have chosen are $\{c_1, c_2\}$ and $\{c_4\}$. By picking one element from each of the ECs we have chosen we form two diagnoses $\{c_1, c_4\}$ and $\{c_2, c_4\}$. Note that the diagnoses we have derived are only part of the complete set of diagnoses.

Still we have other useful observations about Definition 5.7.

Theorem 5.11. *Let $EM = \{e_{i,j}\}_{n\times m}$ be an equivalence matrix for* (*SD*, *COMP*, *OBS*). *Then we have the following properties*:
1. *If column C_j contains only one entry $e_{i,j}$ such that $e_{i,j} = 1$, then for all elements $e_{i,k}$ with $k \neq j, e_{j,k}$ must be* 0.
2. *If row R_i has more then one entry with value* 1, *then for each column C_p and C_q which has an entry with value* 1 *in row R_i, there must be element $e_{x,p}$ and $e_{y,q}$ with value* 1, *where $x \neq y \neq i$.*

Proof.
1. The assumption that there is an element $e_{i,k}$ with value 1 implies that $\text{EC}_i \subseteq \text{MCS}_k$. Since we have $\text{MCS}_j = \text{EC}_i$ based on the fact that the column C_j contains only one entry $e_{i,j}$ such that $e_{i,j} = 1$, we conclude $\text{MCS}_j \subseteq \text{MCS}_k$, a contradiction.
2. Let $\Theta_p = \{x | e_{x,p} = 1 \text{ in } C_p \text{ where } x \neq i\}$. Assume that $\Theta_p \subseteq \Theta_q$, then we have $\text{MCS}_p \subseteq \text{MCS}_q$, a contradiction. □

Now we shall show that only when we have equal number of MCSs and ECs and no two MCSs intersect each other do we have a diagnosis with its size being equal to the number of ECs.

Theorem 5.12. *Given m MCSs and n ECs, we have a diagnosis of size n iff $m = n$ and no two MCSs intersect each other.*

Proof. ($\Rightarrow$)Assume $m \neq n$, we have two cases shown below.

Case I. $m < n$. From Procedure 5.9 we know that every selected row must have at least one column deleted. By selecting each row (EC) we then must have at least n columns deleted, which is a contradiction.

Case II. $m > n$. Since every EC has to be selected during the simplification process, without loss of generality we can rearrange the order of columns so that the equivalence matrix is partitioned into an $n \times n$ diagonal matrix and an $n \times (m-n)$ matrix, as shown below:

$$
n\left\{\left[\begin{array}{cccc|c}
\multicolumn{4}{c|}{\overbrace{\hphantom{1\quad 1\quad \ddots\quad 1}}^{n}} & \overbrace{\hphantom{\qquad\qquad}}^{m-n} \\
1 & & & & \\
 & 1 & & & \\
 & & \ddots & & \\
 & & & 1 &
\end{array}\right]\right. .
$$

For column $C_{n+1}, \ldots, C_m$ each column must have at least one entry with value 1. In fact, by Property (2) in Theorem 5.11 we know that such column, $C_j (n+1 \leqslant j \leqslant m)$, must have an additional entry with value 1, say entry $e_{i,j}$. Then again by Property (2) in Theorem 5.11 we know $C_i (0 \leqslant i \leqslant n)$ must have one additional entry $e_{k,i}$ with value 1, where $k \neq i$. Therefore in the left diagonal matrix at least one row R_k has at least two entries with value 1.

Now consider Procedure 5.9. Since in the diagonal matrix there exists one row which has at least two entries with value 1, by selecting this row at least two columns of the diagonal matrix will be deleted. Therefore during the first $(n-1)$ row selections all columns in the diagonal matrix will be deleted. Now we show that all columns in the right $n \times (m-n)$ matrix will also be deleted before the last row selection could happen. Assume that column $C_j (n+1 \leqslant j \leqslant m)$ remains after the first $(n-1)$ row selections and deletions. Since we know that in the unsimplified matrix element $e_{n,n} = 1$, by Property (2) in Theorem 5.11 we know that at least one entry in C_j, say $e_{k,j}$ must have value 1, where $k < n$. Then C_j must have been deleted when selecting R_k, which is a contradiction.

($\Leftarrow$) Since no two MCSs intersect each other, no two columns can have an entry with value 1 in the same row. Furthermore, since $m = n$ we have entries with value 1 distributed in every row. This in turn implies that each column can only have one entry with value 1, for if not the corresponding MCS intersects another one. By Property (1) in Theorem 5.11 we then know that each row contains only one entry with value 1, which guarantees that a diagnosis of size n exists. □

Theorem 5.12 clearly shows that only when we have the same number of MCSs and ECs and no two MCSs intersect each other could Corollary 5.6 be useful. Note that the condition here is equivalent to that each MCS is an EC by itself. In practical situations this rarely occurs since usually only a small subset of all components is involved in the conflict sets. Therefore our claim concerning Corollary 5.6 is verified.

Having the results above we are now ready to give a complete procedure for updating homogeneity classes of diagnoses. Although in Theorem 4.4 Hou has shown the dynamics of the equivalence relation when adding more measurements, a corresponding result concerning homogeneity classes of diagnoses, as shown below, is not presented.

Corollary 5.13. *For every $\Delta_i \nsim \Delta_j$ for* (SD, $COMP$, OBS), *we still have* $\Delta_i \nsim \Delta_j$ *for* (SD, $COMP$, $OBS \cup \{\Pi\}$), *where* Δ_i *and* Δ_j *are two diagnoses predicting* Π.

Proof. *Case* I. $|\Delta_i| \neq |\Delta_j|$. By Corollary 5.5 we know $\Delta_i \nsim \Delta_j$ for (SD, COMP, OBS $\cup$ $\{\Pi\}$).

Case II. $|\Delta_i| = |\Delta_j|$. Since there must be $\alpha \in \Delta_i$ and $\beta \in \Delta_j$ such that $\alpha \nsim \beta$ for (SD, COMP, OBS), from Theorem 4.4 we know that $\alpha \nsim \beta$ for (SD, COMP, OBS $\cup$ $\{\Pi\}$), which implies $\Delta_i \nsim \Delta_j$ for (SD, COMP, OBS $\cup$ $\{\Pi\}$). □

In summary, to update homogeneity classes of diagnoses, one follows the procedure below:

Procedure 5.14. Let Δ be a new diagnosis after adding a measurement Π. Update homogeneity classes of diagnoses as follows:
1. For each homogeneity class, delete the diagnosis which predicts $\neg\Pi$.
2. For every $\Delta_i \sim \Delta_j$ for (SD, COMP, OBS), determine if $\Delta_i \sim \Delta_j$ or $\Delta_i \nsim \Delta_j$ based on the updated equivalence classes.
3. For every $\Delta_i \nsim \Delta_j$ for (SD, COMP, OBS), one must have $\Delta_i \nsim \Delta_j$ for (SD, COMP, OBS $\cup$ $\{\Pi\}$), and pay no particular attention to whether $\Delta_i \sim \Delta_j$ or $\Delta_i \nsim \Delta_j$ based on the updated equivalence classes.
4. If $|\Delta| = m \neq n$, compare Δ with Δ_m, where $|\Delta_m| = m$, as follows: if $\forall d_i \in \Delta$ there exists $d_j \in \Delta_m$ such that $d_i \sim d_j$, then $\Delta \sim \Delta_m$. If no such Δ_m exists, Δ forms a Singleton homogeneity class.

6. Conclusion

Hou [3] presented an efficient incremental method for computing all diagnoses upon measurement. The method allows old diagnoses to be refined to generate new diagnoses when a measurement is made, avoiding generating new diagnoses from the scratch. We have presented some clarifications to Hou's method. We described an elaborate picture of the relationships among measurements, conflict sets and diagnoses. We also presented some comments on the equivalence relation and Hou's procedure for conflict recognition.

Acknowledgements

Comments from anonymous referees are appreciated. We also thank J.-D. Wu for his helpful insight about equivalent classes.

References

[1] M.R. Genesereth, The use of design descriptions in automated diagnosis, Artificial Intelligence 24 (1984) 411–436.
[2] J. de Kleer, Local methods for localizing faults in electronic circuits, MIT AI Memo, vol. 394, Cambridge, MA, 1976.
[3] A. Hou, A theory of measurement in diagnosis from first principles, Artificial Intelligence 65 (1994) 281–328.
[4] R. Reiter, A theory of diagnosis from first principles, Artificial Intelligence 32 (1987) 57–95.