



**HAL**  
open science

## Critical Observations in a Diagnostic Problem

James Christopher Cody, Marie-Odile Cordier, Alban Grastien

► **To cite this version:**

James Christopher Cody, Marie-Odile Cordier, Alban Grastien. Critical Observations in a Diagnostic Problem. IEEE Conference on Decision and Control (CDC-14), Dec 2014, Los Angeles, United States. hal-01100627

**HAL Id: hal-01100627**

**<https://inria.hal.science/hal-01100627v1>**

Submitted on 7 Jan 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Critical Observations in a Diagnostic Problem

Cody James Christopher<sup>1,2</sup>

Marie-Odile Cordier<sup>3</sup>

Alban Grastien<sup>2,1</sup>

**Abstract**—We claim that presenting a human operator in charge of repairing a faulty system with a small subset of observations relevant to the failure improves awareness and confidence of the operator. Consequently, we introduce the problem of finding a set of relevant observations (called the critical observations) that can be used to derive the same diagnosis as the full problem. We show how this problem can be solved and illustrate its benefits on a real diagnostic problem.

## I. INTRODUCTION

In the context of diagnosis, more observations of the system under consideration generally improves the quality and precision of the diagnosis. A greater number of observations means a reduction in the number of unknown variables, and consequently a reduction in the complexity of the problem.

However, there are disadvantages that arise with an increase in observations. Pertinent to this paper is the particular context where a human operator is involved in the monitoring loop. We seek to present the operator with not only the diagnosis, but also a justification or rational explanation as to how the diagnosis was arrived at. The full set of observations in many cases is too large and varied to serve as an effective solution to this requirement, so the question becomes one of determining the most relevant observations. The seminal work in cognitive psychology makes the argument that the average working memory capacity in humans is only  $7 \pm 2$  distinct items [1]. Thus, the smaller this set of observations is, the better it serves in convincing an operator of the rationality of the presented diagnosis.

As further motivation, consider the following examples. In the state of Victoria, Australia, the government has mandated interval smart meters for 2.6 million electricity customers. Each of these meters provides information on electricity consumption every thirty minutes, for an average of 1.4 thousand per second. NASA provide a diagnostic benchmark system called ADAPT [2], used for the Diagnosis Competition, that involves 87 sensors operating at reporting frequencies of 1 or 2 Hz, and occasionally 10 Hz. In either scenario the raw number of observations is much too great for a human operator to process, In both, however, we posit that a small subset of observations is generally enough to convince an operator of the validity of the diagnosis.

For this work we use a consistency- and model- based approach, where a diagnosis is a set of faults that does

not contradict the observations as applied to the model. We assume that operators are only concerned with minimal diagnoses (as detailed in §III), which implies that observations that do not deviate from the norm can generally be ignored. We also take that the minimal diagnoses for a problem have already been computed. We show that under these assumptions, a subset of observations can be used to prove two related results on the problem: (1) the minimality of a presented diagnosis, and (2) the completeness of the presented set of minimal diagnoses. We then present a procedure to compute the *critical* subset of observations — the minimal set of observations that allows us to prove either of the aforementioned results.

The method presented is developed in the context of steady state systems, but can be extended to dynamic state-driven systems. The method can also be extended to event-based observations; this is more complicated however, and is discussed in the conclusion.

This paper is structured as follows: We initially give a simple worked example before providing the basic definitions of diagnosis. We then present the necessary theory to establish formally the problem of finding what we all the *critical observations*, and provide a procedure for this problem. We illustrate the results on the ADAPT-lite benchmark and conclude with a discussion on related works and possible extensions.

## II. WORKED EXAMPLE

We provide a simple example that illustrates the problem, and will refer back to this example multiple times throughout. Figure 1 shows a simplified version of a power network. Electricity flows from the root through each of the buses ( $b_0, b_1, \dots$ ) through to components ( $x_0, x_1, \dots$ ) at the bottom of the tree. Each component has an associated sensor ( $s_0, s_1, \dots$ ) which, for simplicity, cannot itself be faulty and only indicate whether the flow of power to the component is nominal or not.

The outputs of any given node in the network are *normal* provided that the input on that node is normal and the node itself is not faulty. We assume that the input to *root* is normal. The diagnoses for this network will be then therefore be the set of buses that precisely cover the set of abnormal sensors. It is important to note that no sensor is redundant, as a fault on  $x_i$  will only be detectable by  $s_i$ .

We assume that the two sensors  $s_2$  and  $s_3$  return abnormal observations whilst all other observations are nominal. Under this assumption it is obvious that the fault(s) originate either from  $b_3$  or from both  $x_2$  and  $x_3$ . Bus  $b_0$  and *root* can be exonerated as only sensors  $s_2$  and  $s_3$  are affected,

<sup>1</sup>Artificial Intelligence Group, The Australian National University.

<sup>2</sup>Optimisation Research Group, NICTA\*.

<sup>3</sup>Dream, Université Rennes 1.

\*NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

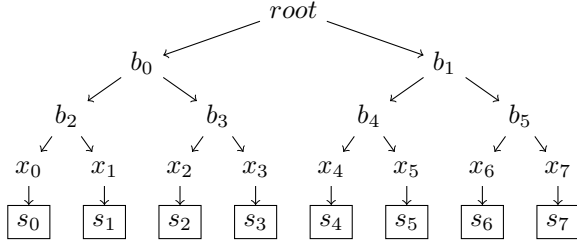


Fig. 1. A Simple Power Network

and we would expect other descendents would also report abnormality if either were the cause.

In this instance, it is clear that not every observation (the readings on all sensors) is needed to produce a useful diagnosis. The problem here is in computationally isolating the relevant observations that could be used to produce an identical diagnosis.

### III. DIAGNOSIS FRAMEWORK

We make use of the standard general framework for model-based diagnosis [3], [4].

*Definition 1:* The *system model*,  $Mod$ , is a tuple  $\langle Comps, SD \rangle$  where  $Comps$  is a set of components and  $SD$  is a statement in first order logic encoding the system behaviour. We use  $Ab(c)$  to specify that component  $c \in Comps$  is behaving abnormally.

*Definition 2:* The *observations*,  $Obs$ , are a set of logical statements.

*Definition 3:* A *diagnostic problem*,  $\mathcal{P}$ , is a tuple  $\langle Mod, Obs \rangle$ .

*Definition 4:* A *diagnostic hypothesis*,  $\delta$ , is a subset of  $Comps$  that implicitly defines the following conjunction:

$$\delta \equiv \left( \bigwedge_{c \in \delta} Ab(c) \right) \wedge \left( \bigwedge_{c \in Comps \setminus \delta} \neg Ab(c) \right). \quad (1)$$

That is, the components  $c \in \delta$  are those believed to be behaving abnormally (thus  $Ab(c)$ ), and all  $c \notin \delta$  are those believed to be behaving normally (thus  $\neg Ab(c)$ ).

*Definition 5:* The *diagnoses* of a problem  $\mathcal{P}$  are the hypotheses,  $\delta$ , that are logically consistent with the system and the observations. We define a function,  $\Delta$ , over problems that returns the set of the diagnoses for  $\mathcal{P}$ :

$$\Delta(\mathcal{P}) = \{ \delta \in 2^{Comps} \mid SD \wedge Obs \wedge \delta \not\equiv \perp \} \quad (2)$$

A diagnosis ( $\delta \in \Delta$ ) is *minimal* if no proper subset of  $\delta$  is also a diagnosis. We define another function,  $\Delta_{min}$ , over problems that returns the set of minimal diagnoses for  $\mathcal{P}$ .

$$\Delta_{min}(\mathcal{P}) = \{ \delta \in \Delta(\mathcal{P}) \mid \nexists \delta' \in \Delta(\mathcal{P}) : (\delta' \subset \delta) \}. \quad (3)$$

We use  $\mathcal{D}$  to represent a set of minimal diagnoses and  $\mathcal{D}^C$  to represent a complete set of minimal diagnoses.

The goal of diagnosis is to determine which components in a given system are faulty. We consider *consistency-based* diagnosis, where a diagnosis is any hypothesis that

is consistent with the observations; that is, it is possible to assign the system variables in a way that agrees with the model and the observations.

#### A. Running Example

To model our example network (presented in §II), we define three binary predicates,  $i, o_1, o_2$ , representing input and (multiple) outputs of each component respectively, with two symbols  $N, A$  representing *normal* and *abnormal*. Under such definitions the fact that the input on  $\alpha$  is normal would be modelled by  $i(\alpha, N)$ . The buses are then modelled by:

$$o_j(\alpha, N) \longleftrightarrow (\neg Ab(\alpha) \wedge i(\alpha, N)). \quad (4)$$

The connections between buses are then modelled by:

$$o_j(\alpha_1, N) \longleftrightarrow i(\alpha_2, N) \quad (5)$$

such that the connection between  $b_0$  and  $b_2$  is represented by  $o_1(b_0, N) \longleftrightarrow i(b_2, N)$ .

Consider the hypothesis  $\delta_0 = \emptyset$  which posits that no component is faulty. We say that  $\delta_0$  is not consistent with the observations  $i(s_2, A)$  and  $i(s_3, A)$ , as our model predicts nominal observations. The possible diagnoses are then:

$$\Delta(\mathcal{P}) = \{ \{b_3\}, \{b_3, x_2\}, \{b_3, x_3\}, \{x_2, x_3\}, \{b_3, x_2, x_3\} \}. \quad (6)$$

Taking  $\delta_1 = \{b_3\}$  and  $\delta_2 = \{x_2, x_3\}$ , we then have  $\Delta_{min}(\mathcal{P}) = \{\delta_1, \delta_2\}$ . This should be interpreted as follows: either  $b_3$  is faulty or both  $x_2$  and  $x_3$  are. In the event that  $b_3$  is faulty, it is possible that there are other faulty components ( $x_2, x_3$ ), but there is no a priori reason to assume this, thus supersets of  $\delta_1$  are not members of  $\Delta_{min}$ .

### IV. CRITICAL OBSERVATIONS

We now present key results that allow us to formalize the problem of finding a subset of observations that are sufficient to infer the same diagnoses as those of a given diagnosis problem.

#### A. Motivation

Understanding the implications of a given diagnoses is important in determining the appropriate repairation or workaround actions. One way to improve awareness of, as well as trust in, a given diagnoses is for the human operator in charge to be able to relate it to the observations that produced it. In a real system with thousands of observation inputs however, this is not a trivial task. Providing the operator the relevant observations therefore becomes a very useful tool for decision support.

*1) Running Example:* Returning to our running example, we previously computed the minimal diagnoses as  $\delta_1 = \{b_3\}$  and  $\delta_2 = \{x_2, x_3\}$ . In this scenario, it is obvious that the observations provided by  $s_2$  and  $s_3$  are crucial to explaining hypothesis  $\delta_2$ . Furthermore, the observations provided by  $s_0$  or  $s_1$  allow us to disregard  $b_0$  and  $root$  as potential candidates, so keeping one of the nominal observations is considered useful. The other nominal observations, however, are not similarly useful except to say that no other incident is taking place in the rest of the network, and the assumption of which was already used in the computation of the minimal diagnoses.

## B. Subproblems

We now present the notion of diagnostic subproblems and the properties that link the diagnoses of problems to their subproblems. We assume at this stage that (some or all of) the minimal diagnoses of the problem,  $\mathcal{P}$ , have been computed.

*Definition 6:* Take diagnosis problem  $\mathcal{P} = \langle Mod, Obs \rangle$ . A *subproblem* of  $\mathcal{P}$  is a problem  $\mathcal{P}' = \langle Mod', Obs' \rangle$  such that  $Mod' = Mod$  and  $Obs' \subseteq Obs$ . That is, some observations are no longer considered.

We now establish three lemmas that are crucial to this work. For the remainder of this section we take  $\mathcal{P}$  to be a diagnosis problem, and  $\mathcal{P}'$  to be a subproblem of  $\mathcal{P}$  as given by Definition 6. Thus:

*Lemma 4.1:*  $\Delta(\mathcal{P}) \subseteq \Delta(\mathcal{P}')$

*Proof:* By contradiction.

Take some  $\delta \in \Delta(\mathcal{P})$  and assume that  $\delta \notin \Delta(\mathcal{P}')$ . Set  $\varphi = Obs \setminus Obs'$ .

From assumption,  $\delta \notin \Delta(\mathcal{P}')$  implies  $(Mod, Obs', \delta) \models \perp$ . But, by monotonicity of entailment:  $(Mod, Obs', \delta, \varphi) \models \perp$ . This implies  $\delta \notin \Delta(\mathcal{P})$ , contradicting the initial premise. ■

This first lemma establishes that any diagnosis of a problem,  $\mathcal{P}$ , must also be a diagnosis for the subproblem,  $\mathcal{P}'$ . It is an important formal statement that we cannot lose diagnoses by reducing the set of observations, and is needed for Lemma 4.2.

*Lemma 4.2:*  $\Delta_{\min}(\mathcal{P}') \cap \Delta(\mathcal{P}) \subseteq \Delta_{\min}(\mathcal{P})$

*Proof:* By contradiction.

Take some  $\delta \in \Delta_{\min}(\mathcal{P}') \cap \Delta(\mathcal{P})$ , and assume  $\delta \notin \Delta_{\min}(\mathcal{P})$ . From assumption,  $\exists \delta_o \subsetneq \delta$ , with  $\delta_o \in \Delta(\mathcal{P})$ .

From Lemma 4.1,  $\delta_o \in \Delta(\mathcal{P})$  implies that  $\delta_o \in \Delta(\mathcal{P}')$ . However since  $\delta_o \subsetneq \delta$ , we have  $\delta \notin \Delta_{\min}(\mathcal{P}')$  (definition 5), which contradicts the initial premise. ■

Lemma 4.2 shows that any diagnosis for the original instance,  $\mathcal{P}$ , that is minimal for  $\mathcal{P}'$  must then also be minimal for  $\mathcal{P}$ . Keeping in mind that the minimality of a diagnosis provides a way of distinguishing whether one solution is better than another, this lemma provides an important consequence — if a subset of observations allows the claim that that no strictly better solutions than  $\delta$  exists, then the claim is also allowed with the original set of observations. In other words, we have shown that using a subset of the original observations is a legitimate way of proving the minimality of a diagnosis.

The next lemma improves this result by showing that there are no other minimal diagnoses other than those implied by the subproblem.

*Lemma 4.3:*  $\Delta_{\min}(\mathcal{P}') \subseteq \Delta(\mathcal{P}) \Rightarrow \Delta_{\min}(\mathcal{P}) \subseteq \Delta_{\min}(\mathcal{P}')$

*Proof:* By contradiction.

Take  $\Delta_{\min}(\mathcal{P}') \subseteq \Delta(\mathcal{P})$ , and assume  $\Delta_{\min}(\mathcal{P}) \not\subseteq \Delta_{\min}(\mathcal{P}')$ . From assumption, there exists  $\delta \in \Delta_{\min}(\mathcal{P}) \setminus \Delta_{\min}(\mathcal{P}')$ .

By definition  $\delta \in \Delta(\mathcal{P})$ , and from Lemma 4.1,  $\delta \in \Delta(\mathcal{P}')$ . As  $\delta \notin \Delta_{\min}(\mathcal{P}')$ , there exists  $\delta_o \in \Delta_{\min}(\mathcal{P}')$  such that  $\delta_o \subsetneq \delta$ . From the premise,  $\delta_o \in \Delta_{\min}(\mathcal{P}') \Rightarrow \delta_o \in \Delta(\mathcal{P})$ . However this contradicts the consequence of the assumption,  $\delta \in \Delta_{\min}(\mathcal{P})$ . ■

The final lemma establishes that if all minimal diagnoses of  $\mathcal{P}'$  are also diagnoses of  $\mathcal{P}$ , then all minimal diagnoses of  $\mathcal{P}$  are minimal diagnoses of  $\mathcal{P}'$ . We can combine Lemma 4.2 with Lemma 4.3 to prove that if all minimal diagnoses of  $\mathcal{P}'$  are diagnoses of  $\mathcal{P}$ , then  $\Delta_{\min}(\mathcal{P}) = \Delta_{\min}(\mathcal{P}')$ .

Consequently, if a subset of observations projects that a system is experiencing all the faults from at least one of the set of hypotheses  $\{\delta_1, \dots, \delta_k\}$ , then this claim is true with the original set of observations. In other words, we are proving that there is no alternative explanation to the observations than  $\Delta_{\min}(\mathcal{P}')$  (though the reality could indeed be worse than a minimal diagnosis).

There are some caveats, however. Whilst a subset of observations is sufficient to disprove the validity of a diagnosis (Lemma 4.1), it is not sufficient in general to prove its validity. Indeed, notice that both Lemma 4.2 and Lemma 4.3 apply only to the minimal diagnoses of  $\mathcal{P}'$  that are also diagnoses of  $\mathcal{P}$ .

Futhermore, these results hold in a consistency-based diagnostic framework, but not in a probabilistic one. Logical consistency-based diagnosis ( $\neq \perp$ ) enjoys the monotonicity of entailment — whereby adding logical statements (state-based observations) cannot make an already invalid explanation valid — that probabilistic frameworks unfortunately do not, as an explanation which was unlikely (compared to others explanations) can suddenly become highly probable if added observations support this explanation and contradict others.

## C. Critical Sets

We now present the concept of sufficient observations and formalise the problem of finding a minimal sufficient subproblem.

Take a set  $\mathcal{D}$  of diagnoses for the problem  $\mathcal{P}$  (i.e.,  $\mathcal{D} \subseteq \Delta(\mathcal{P})$ ). We are interested in proving two properties about  $\mathcal{D}$ :

- *Minimality:*  $\mathcal{D} \subseteq \Delta_{\min}(\mathcal{P})$ , i.e., there are no strictly better explanations than those of  $\mathcal{D}$ ;
- *Completeness:*  $\mathcal{D} = \Delta_{\min}(\mathcal{P})$ , i.e., there are no alternative explanations than those of  $\mathcal{D}$ .

*Definition 7:* Take diagnosis problem  $\mathcal{P}$  and its subproblem  $\mathcal{P}'$ . Take a set  $\mathcal{D}$  of diagnoses of  $\mathcal{P}$  and a property of  $\mathcal{D}$  with respect to  $\mathcal{P}$ . Subproblem  $\mathcal{P}'$  is *sufficient* for this property if the property of  $\mathcal{D}$  can be proved using only  $\mathcal{P}'$ .

From Lemma 4.2, a subproblem is sufficient for minimality if and only if the following holds:  $\mathcal{D} \subseteq \Delta_{\min}(\mathcal{P}')$ . Similarly from Lemma 4.3, a subproblem is sufficient for completeness if and only if the following holds:  $\mathcal{D} = \Delta_{\min}(\mathcal{P}')$ .

As mentioned, we assume that a more concise set of observations better serves the purpose of providing a rational explanation to a human operator in a diagnosis loop. A critical subproblem is therefore a sufficient problem from which no observation can be removed:

*Definition 8:* Take a diagnosis problem  $\mathcal{P}$  and its subproblem  $\mathcal{P}'$ . Take a set  $\mathcal{D}$  of diagnoses of  $\mathcal{P}$  and a property of  $\mathcal{D}$  with respect to  $\mathcal{P}$ . Subproblem  $\mathcal{P}'$  is *critical* for this property

if it is sufficient and no strict subproblem of  $\mathcal{P}'$  is sufficient (minimal).

## V. FINDING CRITICAL OBSERVATIONS

Having defined critical subproblems — a minimal subproblem that remains sufficient — we wish to now find one.

It is easily shown that sufficiency as defined is a monotonic property; if a subproblem is not sufficient then neither are its subproblems. Therefore, an approach to finding a critical subproblem consists of iteratively testing whether removing a specific observation maintains the required properties or not.

As explained in § IV, we can determine the sufficiency of a subproblem by computing the set of minimal diagnoses of this subproblem. This operation, however, is potentially expensive and we seek a more efficient approach.

We instead present an alternative approach for testing for sufficiency with respect to either property that is based on a single consistency check. One of the consequence of Lemma 4.1, is that removing observations can only increase the total number of diagnoses; therefore, we need to make sure that these added diagnoses do not affect the relation between  $\mathcal{D}$  and  $\Delta_{\min}(\mathcal{P}')$ .

It is possible to identify which hypotheses,  $\nabla$ , should *not* ever be added to the set of diagnoses. The consistency check verifies whether any of these hypotheses are consistent with the model and the observations. The set  $\nabla$  can be quite large and it should not be enumerated; we present at the end of this section a method of representing this set in a logical and compact formulation.

### A. Excluded Hypotheses

With respect to minimality, we need to prevent diagnoses that would invalidate the minimality property of those already in  $\mathcal{D}$ , so we define:

$$\nabla(\mathcal{D}) = \{\delta_o \in 2^{Comps} \mid \exists \delta \in \mathcal{D} . \delta_o \subset \delta\} \quad (7)$$

and present a companion lemma (proof omitted):

*Lemma 5.1:*  $\mathcal{D} \subseteq \Delta(\mathcal{P}) \wedge (\nabla(\mathcal{D}) \cap \Delta(\mathcal{P}') = \emptyset) \iff \mathcal{D} \subseteq \Delta_{\min}(\mathcal{P}')$ .

This lemma demonstrates that we can show  $\mathcal{D}$  is a set of minimal diagnoses for  $\mathcal{P}'$  if  $\nabla$  does not contain any diagnoses for  $\mathcal{P}'$ . Combined with Lemma 4.2, this implies that  $\mathcal{D}$  is a set of minimal diagnoses for the parent problem,  $\mathcal{P}$ , as well.

However, this still does not prevent the adding of new minimal diagnoses that are disjoint to the existing members of  $\mathcal{D}$  (with respect to the components). To preserve the completeness of the original set, we need to exclude all remaining hypotheses:

$$\nabla^C(\mathcal{D}) = \{\delta_o \in 2^{Comps} \mid \nexists \delta \in \mathcal{D} . \delta \subseteq \delta_o\} \quad (8)$$

and present a companion lemma (proof omitted):

*Lemma 5.2:*  $\mathcal{D} \subseteq \Delta(\mathcal{P}) \wedge (\nabla^C(\mathcal{D}) \cap \Delta(\mathcal{P}') = \emptyset) \iff \mathcal{D} = \Delta_{\min}(\mathcal{P}')$ .

Again, this lemma demonstrates that completeness of  $\mathcal{D}$  can be proved by showing that none of the excluded hypotheses in  $\nabla^C$  contradict the subset of observations in  $\mathcal{P}'$ .

The key characteristic of the solution,  $Obs'$ , is that it defines a diagnostic problem to which the set of diagnoses does not intersect either  $\nabla$  or  $\nabla^C$ . We can interpret  $\nabla$  as a disjunction and evaluate the consistency of:

$$SD \wedge Obs' \wedge \nabla \stackrel{?}{\models} \perp. \quad (9)$$

If a contradiction is derived, then  $Obs'$  is sufficient to prove the required property on  $\mathcal{D}$ . Notice that sufficiency is proved when an inconsistency is found, while the validity of a hypothesis is proved when there is no inconsistency.

1) *Running Example:* Referring back to our example in §II, we can compute a sufficient (and indeed, critical) set of observations.

Taking  $\delta_1 = \{b_3\}$ ,  $\delta_2 = \{x_2, x_3\}$ , with  $\mathcal{D} = \{\delta_1, \delta_2\}$ , we can compute  $\nabla(\mathcal{D})$ :

$$\nabla(\mathcal{D}) = \{\emptyset, \{x_2\}, \{x_3\}\} \quad (10)$$

This result show that a subset of observations is sufficient to prove minimality of  $\mathcal{D}$  if it excludes the three hypotheses that would be considered strictly better than those in  $\mathcal{D}$ :  $\emptyset$  (no component is faulty),  $\{x_2\}$  (only  $x_2$  is faulty) and  $\{x_3\}$ . There is only one critical subset of observations that achieves this:  $\{o(s_2, A), o(s_3, A)\}$  — both  $s_2$  and  $s_3$  are abnormal.

If we only consider  $\mathcal{D}_1 = \{\delta_1\}$ , we obtain  $\nabla(\mathcal{D}_1) = \{\emptyset\}$ , indicating that a subset of observations is sufficient to prove minimality of  $\mathcal{D}_1$  if it excludes only the hypothesis that no component is faulty, and thus only a single observation (e.g.,  $o(s_2, A)$ ) is necessary to achieve this.

Consider now the completeness of  $\mathcal{D}$ :

$$\nabla^C(\mathcal{D}) = \{\{b_0\}, \{x_0\}, \dots, \{b_0, x_0\}, \{b_0, x_2\}, \dots\} \quad (11)$$

This is all hypotheses that do not include  $x_2$  and  $x_3$  together and/or  $b_3$  — more than 10,000 elements. As  $\nabla^C(\mathcal{D})$  will always be a superset of  $\nabla(\mathcal{D})$ , a critical subset for completeness is always a superset of a critical subset for minimality. Thus we must include the previously indicated observations  $o(s_2, A)$  and  $o(s_3, A)$ , which has the effect of excluding all hypotheses that do not suspect *root*,  $b_0$ ,  $b_3$ ,  $x_2$  or  $x_3$ . To demonstrate completeness, hypotheses that suspect *root* or  $b_0$  must also be excluded — adding  $o(s_0, N)$  has this effect, leaving only hypotheses that are consistent with the model (those suspecting  $b_3$  or both  $x_2$  and  $x_3$ ), none of which belongs to  $\nabla^C(\mathcal{D})$ .

If we tried to prove completeness of (the incomplete)  $\mathcal{D}_1$ , we would end up with a set  $\nabla^C(\mathcal{D}_1)$  containing  $\delta_2$ . As  $\delta_2$  is a diagnosis, it is also a diagnosis of all subproblems and there is no critical set of observations.

Notice that all the critical sets presented above consider the observation  $o(s_4, N)$  as irrelevant, among other things it indicates that component  $x_4$  is nominal. Keep in mind, however, that we are only interested in the minimal diagnoses. The established minimal diagnoses for the problem do not say anything about the state of component  $x_4$  except that there is no reason to suspect  $x_4$  of being faulty.

## B. Symbolic Representation of $\nabla$

In the small running example with less than twenty components, the set  $\nabla^C$  already contains over 10,000 elements. Since the size of this set increases exponentially with the number of components, it is impractical to enumerate it. Fortunately the consistency checker does not need an explicit enumeration, but can use the symbolic representation that we now present.

Assume that  $\mathcal{D}$  is a singleton hypothesis,  $\{\delta\}$ , where  $\delta$  may contain several components. The set  $\nabla(\mathcal{D})$  can be represented symbolically as follows:

$$\nabla(\mathcal{D}) \equiv \left( \bigwedge_{c \in \text{Comps} \setminus \delta} \neg \text{Ab}(c) \right) \wedge \left( \bigvee_{c \in \delta} \neg \text{Ab}(c) \right). \quad (12)$$

The symbolic representation for a non-singleton is simply the disjunction of the singleton representations for each of hypotheses in  $\mathcal{D}$ .

In our running example,  $\mathcal{D}_2 = \{\delta_2\}$ , the minimality of  $\delta_2$  is ensured by proving that the joint assumptions: (1) no component outside  $\delta_2$  is faulty, and (2) not both of components  $x_2$  and  $x_3$  are faulty, contradict the model and the observations. This representation is linear in the size of *Comps*.

We now turn to  $\nabla^C$ . The symbolical representation of  $\nabla^C(\mathcal{D})$  is:

$$\nabla^C(\mathcal{D}) \equiv \bigwedge_{\delta \in \mathcal{D}} \left( \bigvee_{c \in \delta} \neg \text{Ab}(c) \right). \quad (13)$$

Back to the example,  $\mathcal{D}_{1,2} = \{\delta_1, \delta_2\}$ . Completeness of  $\mathcal{D}_{1,2}$  is ensured by proving that the joint assumptions: (1) component  $b_3$  is not faulty, and (2) not both of components  $x_2$  and  $x_3$  are faulty, contradict the model and the observations.

While the size of  $\nabla^C(\mathcal{D})$  is exponential in the size of *Comps*, this representation is only linear in the size of  $\mathcal{D}$  and does not directly depend on the size of *Comps* (bearing in mind that  $\mathcal{D}$  may be exponential in the size of *Comps*).

## VI. ILLUSTRATION ON ADAPT-LITE

We now present an example taken from the ADAPT-lite track used as part of the 2009 International Workshop on Principles of Diagnosis (DX) Competition [2]. The hardware system for the DXC-09 Industrial Track is the Electrical Power System testbed in the ADAPT lab at NASA Ames Research Center.

The ADAPT EPS testbed provides a means for evaluating diagnostic algorithms through the controlled insertion of faults in repeatable failure scenarios. The lite version of ADAPT is depicted in Figure 2. The sensors on the ADAPT system return observations at a rate of 1, 2, or 10Hz, which, on the full system, produces nearly one thousand, often ten digit, information inputs per second. We used a model that combines first order logic with linear arithmetics, and we use an SMT solver for the consistency checks [5].

The specific approach used for this problem differs from the one presented in that it does not consider minimality with

respect to set inclusion but with respect to cardinality — diagnoses that minimize the number of faulty components. This change is made as the ADAPT sensors themselves may be faulty, whereas we had previously assumed otherwise. This can lead to unrealistic minimal diagnoses that involve most sensors being faulty, and the minimality of these diagnoses requires at least one observation from each sensor. The extension of our work to minimal cardinality, and in particular the representation of the  $\nabla$  sets, is very similar in construction.

Figure 3 shows a reduced example of observation trace on the ADAPT-lite system. The single minimal-cardinality diagnosis in this problem posits that Sensor IT240 (current flow in amperes) suddenly suffers from an offset fault. The critical observations are identified by our algorithm are indicated in bold in Figure 3, and we can indeed demonstrate that they suffice to prove the diagnosis.

Firstly, notice that IT240 reads a current of 16.3A at time 1500ms, after having read 6.3A prior. This value, according to the system specification and model, is clearly abnormal. Secondly, as the value of IT240 at time 2000ms is different from the former one, we deduce that the problem cannot be that IT240 is stuck at 16.3A. Finally, the only reason (at least, according to our model) for a larger than expected current is that the battery is compensating for a lower than expected voltage; however voltage sensor E235 claims that the voltage is normal (24V is expected upstream of the inverter).

Obviously E235 could be faulty, but that would imply at least one other fault, as a fault from E235 does not explain the abnormal observation from IT240. Such a diagnosis would have a cardinality of two or more, making it less preferred to the cardinality one diagnosis.

## VII. RELATED WORK

The notion of reducing the number of observations in diagnosis problems has been widely studied but with different motivations from this work. Previous work in general aims at reducing the overall cost of observations, which is incurred in multiple different ways: (1) the system must be designed to allow for appropriate and useful observations, (2) sensors must be integrated and additionally powered, (3) observations must be collected, etc.

Optimal diagnosability is concerned with minimizing the number of sensors (or their total cost) while ensuring diagnosability [6] before any observations are considered. The solution to an optimal diagnosability problem works for *every* possible evolution of the system, as opposed to our approach, which is specific to the current evolution and only seeks to provide an explanation for the current circumstances.

Sequential diagnosis [7], and its event-based variant [8], focuses on the problem of deciding which observation should be collected next in order to improve the precision of diagnosis. Similarly to the work presented, sequential diagnosis is an online problem — deciding which observation to collect is not made a priori but with respect to the observations already collected. The difference to this work, however, is

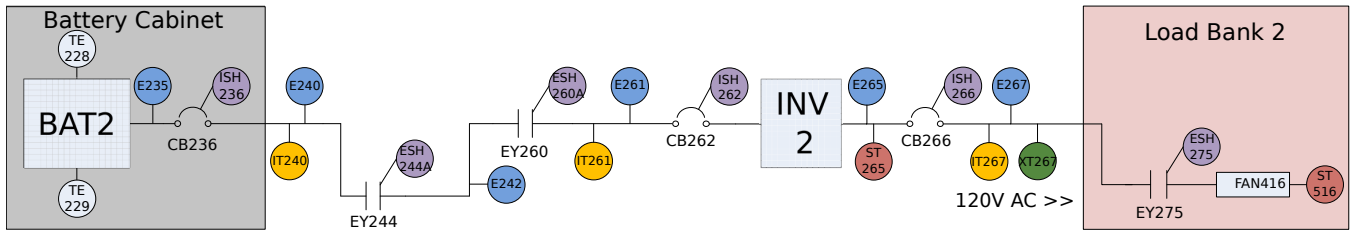


Fig. 2. Schematic for ADAPT-lite

```

sensors @1000 { E235 = 24.4, E240 = 24.4, E242 = 24.3, E261 = 24.4, E265 = 120.8, E267 = 120.9,
ESH244A = true, ESH260A = true, ESH275 = true, ISH236 = true, ISH262 = true, ISH266 = true, IT240 =
6.3, IT261 = 6.3, IT267 = 0.94, ST265 = 60.4, ST516 = 900.0, TE228 = 71.6, TE229 = 72.8 };
sensors @1500 { E235 = 24.4, E240 = 24.4, E242 = 24.3, E261 = 24.3, E265 = 120.8, E267 = 120.9,
ESH244A = true, ESH260A = true, ESH275 = true, ISH236 = true, ISH262 = true, ISH266 = true, IT240 =
16.3, IT261 = 6.3, IT267 = 0.97, ST265 = 60.4, ST516 = 900.0, TE228 = 71.6, TE229 = 72.8 };
sensors @2000 { E235 = 24.4, E240 = 24.4, E242 = 24.3, E261 = 24.3, E265 = 120.8, E267 = 120.9,
ESH244A = true, ESH260A = true, ESH275 = true, ISH236 = true, ISH262 = true, ISH266 = true, IT240 =
16.4, IT261 = 6.3, IT267 = 0.94, ST265 = 60.4, ST516 = 900.0, TE228 = 71.6, TE229 = 72.7 };

```

Fig. 3. Example of Observations - Sensor IT240 offset

that all observations are kept, even when they provide no information or when they are redundant.

### VIII. CONCLUSION

In this paper we presented an approach to provide an operator involved in a diagnosis loop with a manageable subset of observations with the intent of providing a better understanding and appreciation of the results returned by the diagnosis procedure. The intended application of this work is in contexts where the sheer volume of observations is overwhelming, making it difficult for a human operator to verify the validity of the given diagnosis. We made the assumption that the optimality of the solution is tied to the conciseness of the explanation. It is clear then, that the most concise explanation is strictly preferred, which we formalised with respect to the notion of minimality.

This definition of optimality could be further refined. For instance, one might consider that a better solution would be one that involves reasoning about the smallest number of variables (or components) or that involves the simplest rules (for instance, avoiding complex numerical operations). In the ADAPT-lite example from the previous section, it seems more natural to include the observation of IT261 rather than that of E235 because the disagreement between the two observations of the current flow is more obvious than the inconsistency between the current flow and the voltage.

Additionally, it might be interesting to optimize the criterion of confidentiality by abstracting away details: in a power network context, one could report that a specified household has been consuming electricity during the day rather than reporting the precise time and consumption amount.

Another direction of research is the more practical question of computing the critical set of observations fast. One possibility is to analyze the solving of  $Mod \wedge Obs \wedge \nabla$  in order to extract the observations used to prove inconsistency;

although this set of observations may not be minimal, it would provide a good first estimate.

A last interesting extension is to consider event-based observations — extracting the critical information from a large flow of alarms. Event-based observations bring an additional subtlety that is not a factor in state-based observations. Specifically, there is a difference between not observing an event and ignoring an event that has been observed; for instance, the repeated observations that a window is being closed without an observation of it ever being opening is symptomatic of a problem. The critical information may therefore include that certain observations were *not* made.

### REFERENCES

- [1] G. A. Miller, “The magical number seven, plus or minus two: some limits on our capacity for processing information.” *Psychological review*, vol. 63, no. 2, p. 81, 1956.
- [2] T. Kurtoglu, S. Narasimhan, S. Poll, D. Garcia, L. Kuhn, J. de Kleer, A. van Gemund, and A. Feldman, “First international diagnosis competition – DXC’09,” in *20th International Workshop on Principles of Diagnosis (DX-09)*, 2009, pp. 383–396.
- [3] R. Reiter, “A theory of diagnosis from first principles,” *Artificial Intelligence (AIJ)*, vol. 32, no. 1, pp. 57–95, 1987.
- [4] J. de Kleer and B. Williams, “Diagnosing multiple faults,” *Artificial Intelligence (AIJ)*, vol. 32, pp. 97–130, 1987.
- [5] A. Grastien, “Diagnosis of hybrid systems by consistency testing,” in *24th International Workshop on Principles of Diagnosis (DX-13)*, 2013, pp. 9–14.
- [6] L. Brandán Briones, A. Lazovik, and P. Dague, “Optimal observability for diagnosability,” in *Nineteenth International Workshop on Principles of Diagnosis (DX-08)*, 2008, pp. 31–38.
- [7] J. de Kleer, “Using crude probability estimates to guide diagnosis,” *Artificial Intelligence*, vol. 45, no. 3, pp. 381–391, 1990.
- [8] F. Cassez and S. Tripakis, “Fault diagnosis with dynamic observers,” in *Ninth International Workshop on Discrete Event Systems (WODES-08)*, 2008, pp. 212–217.