Critical Observations in a Diagnostic Problem

Cody James Christopher[1,2], Marie-Odile Cordier[3], and Alban Grastien[2,1]

[1]Artificial Intelligence Group, The Australian National University.
[2]Optimisation Research Group, NICTA*
[3]Projet Dream, Irisa, Université Rennes 1.

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 and formalise the notion of a set of relevant observations that derive the same diagnosis as the full observations – we call these the critical observations. We show how this set can be identified algorithmically and illustrate its benefits on an instance of a real diagnostic problem.

1 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. We make the assumption that the smaller this set of observations is, the better it serves the purpose of 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 [1], 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.

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

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 §3), 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 notion of what we call the *critical observations*, and provide a procedure for the identification of said observations. We illustrate the results on the ADAPT-lite benchmark and conclude with a discussion on related works and possible extensions.

2 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 $(b_0, b_1, \ldots)$ through to components $(x_0, x_1, \ldots)$ at the bottom of the tree. Each component has an associated sensor $(s_0, s_1, \ldots)$ which, for simplicity, cannot itself be faulty and only indicates 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 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.

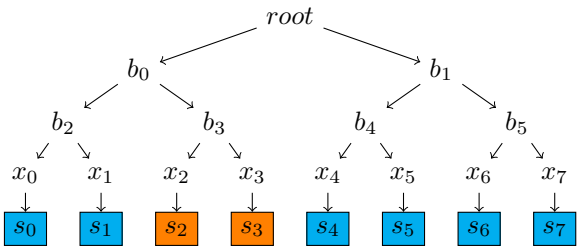

Figure 1: A Simple Power Network. s_2 and s_3 are reporting "abnormal"

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, 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 issue is in computationally identifying the relevant observations that could be used to produce an identical diagnosis.

3 Diagnosis Framework

We make use of the standard general framework for model-based diagnosis [2; 3].

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*,* δ*, is a subset of Comps that implicitly defines the following conjunction:*

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

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,* δ*, that are logically consistent with the system and the observations. We define a function,* Δ*, over problems that returns the set of the diagnoses for* $\mathcal{P}$*:*

$$\Delta(\mathcal{P}) = \left\{\delta \in 2^{Comps} \mid SD, Obs, \Phi(\delta) \nvDash \bot\right\} \quad (2)$$

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

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

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

Running Example

To model our example network (presented in §2), 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 α is normal would be modelled by $i(\alpha, N)$. The buses are then modelled by:

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

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 δ_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 δ_1 are not members of $\Delta_{\min}$.

4 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.

We first remind readers of the motivation driving this section, and support it by example. The concept of a diagnostic subproblem – a reduced version of a given diagnostic problem – is then introduced and forms the crux of the approach we use to compute the critical observations. Three lemmas are presented that concretely relate properties of the subproblem to that of the parent problem. These lemmas form the foundation on which the formal definitions of both sufficient and critical observations are built on, which are then presented.

4.1 Motivation

Understanding the implications of a given diagnoses is important in determining the appropriate reparation 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.

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 explaning hypothesis δ_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.

4.2 Diagnostic 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 write $\mathcal{P} \sqsupseteq \mathcal{P}'$.*

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. We take for granted the monotonicity of entailment — whereby adding logical statements (state-based observations) cannot make an already invalid explanation valid.

The first lemma establishes that any diagnosis of a problem, $\mathcal{P}$, must also be a diagnosis for the subproblem, $\mathcal{P}'$.

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 $(SD, Obs', \delta) \models \bot$.
But, by monotonicity of entailment: $(SD, Obs', \delta, \varphi) \models \bot$,
This implies $\delta \notin \Delta(\mathcal{P})$, contradicting the initial premise. □

Lemma 4.1 is an important formal statement that we cannot lose diagnoses by reducing the set of observations, and is needed for Lemma 4.2. An important edge case is the terminal where $Obs' = \emptyset$ — in the absence of any observation, all possible diagnoses become feasible. That is to say, a consistency checker will not find an inconsistency with any diagnosis if there is no information from the system to claim otherwise.

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 no strictly better solutions than δ 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.

Running Example
To illustrate this lemma, we refer back to our running example. Take a subproblem $\mathcal{P}'_\alpha$ with reduced observations $Obs' = \{o(s_2, A), o(s_5, N), o(s_7, N)\}$. Using only Obs', we arrive at possible minimal diagnoses of $\{x_2\}$, $\{b_3\}$, and $\{b_0\}$. Note that $\Delta_{\min}(\mathcal{P}'_\alpha) \cap \Delta(\mathcal{P}) = \{b_3\}$ and from Lemma 4.2, $\{b_3\} \in \Delta_{\min}(\mathcal{P})$. Conversely, $\{x_2\}$ and $\{b_0\}$ are therefore not minimal diagnoses of $\mathcal{P}$.

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 $\{\delta_1, \ldots, \delta_k\}$, 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 ($\not\models \bot$) enjoys the monotonicity of entailment that probabilistic frameworks unfortunately do not, as an explanation which was unlikely (compared to other explanations) can suddenly become highly probable if added observations support this explanation and contradict others.

4.3 Sufficient & 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 on $\mathcal{D}$:

- *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 $\mathcal{D}$.

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 $\mathcal{P}'$. Formally $suff(\mathcal{D}, \mathcal{P}')$ is a predicate that satisfies:

$$\begin{array}{l} \forall \mathcal{D} \subseteq \Delta(\mathcal{P}) : \\ suff_{\sharp}(\mathcal{D}, \mathcal{P}') \Leftrightarrow \\ \forall \mathcal{P}'' \sqsupseteq \mathcal{P}' \ . \ [\mathcal{D} \subseteq \Delta(\mathcal{P}'') \Rightarrow \mathcal{D} \sharp \Delta_{\min}(\mathcal{P}'')] \, , \end{array} \tag{7}$$

where $\sharp$ is either $\subseteq$ or $=$ depending on whether we want to prove minimality or completeness. $\mathcal{P}''$ in the equation above is introduced as the only details we recall from $\mathcal{P}$ are that $\mathcal{P} \sqsupseteq \mathcal{P}'$ and $\mathcal{D} \subseteq \Delta(\mathcal{P})$, thus $\mathcal{P}'$ will be sufficient *for all* parent problems, $\mathcal{P}''$ (including $\mathcal{P}$), where $\mathcal{D} \subseteq \Delta(\mathcal{P}'')$.

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 further observation can be removed:

Definition 7. *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).*

We now demonstrate what the sufficiency predicate practically represents and illustrate these results on the running example.

Theorem 4.4. *The following two equivalences hold:*

$$suff_{\subseteq}(\mathcal{D}, \mathcal{P}') \equiv (\mathcal{D} \subseteq \Delta_{\min}(\mathcal{P}')) \tag{8}$$

$$suff_{=}(\mathcal{D}, \mathcal{P}') \equiv (\mathcal{D} = \Delta_{\min}(\mathcal{P}')) \tag{9}$$

Proof. We concentrate on $suff_{\subseteq}$ as the same argument can be used for $suff_{=}$.
[$\Rightarrow$] By construction.
Assume $\mathcal{P}'' \sqsupseteq \mathcal{P}'$, $\mathcal{D} \subseteq \Delta(\mathcal{P}'')$, and $\mathcal{D} \subseteq \Delta_{\min}(\mathcal{P}')$, then from Lemma 4.2, $\mathcal{D} \subseteq \Delta_{\min}(\mathcal{P}'')$.
[$\Leftarrow$] By contrapositive.
Assume $\mathcal{D} \not\subseteq \Delta_{\min}(\mathcal{P}')$. Take $\mathcal{P}'' = \mathcal{P}'$; then $\mathcal{P}'' \sqsupseteq \mathcal{P}'$ and $\mathcal{D} \subseteq \Delta(\mathcal{P}'')$ (since $\mathcal{D} \subseteq \Delta(\mathcal{P})$). Clearly $\mathcal{D} \not\subseteq \Delta_{\min}(\mathcal{P}'')$. □

In other words, under the assumption that $\mathcal{D}$ is a set of diagnoses of problem $\mathcal{P}$, minimality (or completeness respectively) of $\mathcal{D}$ can be proved by demonstrating that $\mathcal{D}$ is minimal (or complete respectively) for problem $\mathcal{P}'$.

Running Example
Considering our previous example, taking $\mathcal{D}_1 = \{\delta_1\} = \{\{b_3\}\}$, a critical subproblem that proves minimality of $\mathcal{D}_1$ is formed by selecting a critical (minimal) subset of observations for which δ_1 is a minimal diagnosis. The singleton $\{o(s_2, A)\}$ is sufficient to prove minimality of $\mathcal{D}_1$, which is quite intuitive: considering only the information that i) the output of s_2 is abnormal and ii) that the other (hidden) observations do not contradict the diagnosis δ_1, then the fact that δ_1 is a minimal diagnosis is obvious.

Similarly a critical subproblem proving completeness of $\mathcal{D}_{1,2} = \{\delta_1, \delta_2\}$, where $\delta_2 = \{x_2, x_3\}$, necessitates the identification of a subset of observations whose corresponding set of minimal diagnoses is exactly $\mathcal{D}_{1,2}$. One such critical subset is $\{o(s_1, N), o(s_2, A), o(s_3, A)\}$. Again, this is quite intuitive: the two abnormal observations alone leave that at least one of $root$, b_0, b_3, or both x_2 and x_3 are faulty, and adding the nominal observation of s_1 shows that the first two options ($root$ and b_0) are inconsistent with the observations, thus invalidating them.

Notice that the observations on the right side of the network are not part of any critical set of observations. This is because they exonerate the components on the right side of the network that we have no reason to suspect. One could claim that they do exonerate the suspect root, and this is interesting information for completeness, however, this information is strictly subsumed by that of the observations from s_0 and s_1, which also exonerates bus b_0.

5 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 outlined in § 4, crucial to determining the sufficiency of a subproblem is verifying that $\mathcal{D} \subseteq \Delta_{\min}(\mathcal{P}')$ or $\mathcal{D} = \Delta_{\min}(\mathcal{P}')$. The naïve way of achieving this is by computing the set of minimal diagnoses of this subproblem and explicitly confirming the relation between $\mathcal{D}$ and $\Delta_{\min}(\mathcal{P}')$. This operation, however, is potentially expensive and we instead present an alternative that avoids this expense.

One of the consequences 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 should *not* ever be added to the set of diagnoses. We call this set, ∇, the Excluded Hypotheses.

Using ∇, we propose a consistency check that verifies whether any of these excluded hypotheses are consistent with the model and the observations of the subproblem such that they can be accounted for. ∇ 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.

5.1 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:

$$\nabla(\mathcal{D}) = \left\{ \delta_o \in 2^{Comps} \mid \exists \delta \in \mathcal{D} \ . \ \delta_o \subsetneq \delta \right\} \tag{10}$$

and present a companion lemma:

Lemma 5.1.
$\mathcal{D} \subseteq \Delta(\mathcal{P}) \Rightarrow ((\nabla(\mathcal{D}) \cap \Delta(\mathcal{P}')) = \emptyset \Leftrightarrow \mathcal{D} \subseteq \Delta_{\min}(\mathcal{P}'))$

Proof. Take $\mathcal{D} \subseteq \Delta(\mathcal{P})$. From Lemma 4.1, $\mathcal{D} \subseteq \Delta(\mathcal{P}')$.

[⇒] By contradiction.
Take $(\nabla(\mathcal{D}) \cap \Delta(\mathcal{P}')) = \emptyset$ and assume $\mathcal{D} \nsubseteq \Delta_{\min}(\mathcal{P}')$.
From assumption, $\exists \delta \in \mathcal{D}$ *such that* $\delta \notin \Delta_{\min}(\mathcal{P}')$.
Therefore, $\exists \delta_o \subsetneq \delta$ *such that* $\delta_o \in \Delta_{\min}(\mathcal{P}')$.
But $\delta_o \in \nabla(\mathcal{D}), \delta_o \in \Delta(\mathcal{P}')$ by definition.
Therefore $(\nabla(\mathcal{D}) \cap \Delta(\mathcal{P}')) \neq \emptyset$, contradicting the second premise.

[⇐] By contradiction.
Take $\mathcal{D} \subseteq \Delta_{\min}(\mathcal{P}')$ and assume $(\nabla(\mathcal{D}) \cap \Delta(\mathcal{P}')) \neq \emptyset$.
Then $\exists \delta_o \in (\nabla(\mathcal{D}) \cap \Delta(\mathcal{P}'))$.
However, $\delta_o \in \nabla(\mathcal{D}) \Rightarrow \exists \delta \in \mathcal{D}$ *such that* $\delta_o \subsetneq \delta$, and $\delta_o \in \Delta(\mathcal{P}')$.
But $((\delta_o \subsetneq \delta) \wedge (\delta_o \in \Delta(\mathcal{P})) \wedge (\delta \in \mathcal{D})) \Rightarrow \mathcal{D} \nsubseteq \Delta_{\min}(\mathcal{P}')$, contradicting the second premise.
Both directions of the bi-implication are satisfied under the initial premise. □

This lemma demonstrates that we can show $\mathcal{D}$ is a set of minimal diagnoses for $\mathcal{P}'$ if ∇ 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 (in addition to ∇):

$$\nabla^C(\mathcal{D}) = \left\{\delta_o \in 2^{Comps} \mid \nexists \delta \in \mathcal{D} \,.\, \delta \subseteq \delta_o\right\} \tag{11}$$

and present a companion lemma:

Lemma 5.2.
$\mathcal{D} \subseteq \Delta(\mathcal{P}) \Rightarrow \left((\nabla^C(\mathcal{D}) \cap \Delta(\mathcal{P}')) = \emptyset \Leftrightarrow \mathcal{D} = \Delta_{\min}(\mathcal{P}')\right)$

Proof. Take $\mathcal{D} \subseteq \Delta(\mathcal{P})$. From Lemma 4.1, $\mathcal{D} \subseteq \Delta(\mathcal{P}')$.

[⇒] By contradiction.
Take $(\nabla^C(\mathcal{D}) \cap \Delta(\mathcal{P}')) = \emptyset$ and assume $\mathcal{D} \neq \Delta_{\min}(\mathcal{P}')$.
There are two cases:
[1]: $\exists \delta \in \mathcal{D}$ *such that* $\delta \notin \Delta_{\min}(\mathcal{P}')$.
However, $\delta \in \Delta(\mathcal{P}')$, and therefore $\exists \delta_o \subseteq \delta$ *such that* $\delta_o \in \nabla^C \wedge \delta_o \in \Delta_{\min}(\mathcal{P}')$.
Therefore $(\nabla^C(\mathcal{D}) \cap \Delta(\mathcal{P}')) \neq \emptyset$, contradicting the second premise.

[2]: $\exists \delta \in \Delta_{\min}(\mathcal{P}')$ *such that* $\delta \notin \mathcal{D}$.
There are two further subcases:
[2*a*]: $\exists \delta_o \in \mathcal{D}$ *such that* $\delta_o \supsetneq \delta$ and thus $\delta \in \nabla^C(\mathcal{D})$ giving $(\nabla^C(\mathcal{D}) \cap \Delta(\mathcal{P}')) \neq \emptyset$, or
[2*b*]: δ is disjoint to every element in $\mathcal{D}$, and thus $\delta \in \nabla^C(\mathcal{D})$ giving $(\nabla^C(\mathcal{D}) \cap \Delta(\mathcal{P}')) \neq \emptyset$.
In either case, this contradicts the second premise.

[⇐] By contradiction.
Take $\mathcal{D} = \Delta_{\min}(\mathcal{P}')$ and assume $(\nabla^C(\mathcal{D}) \cap \Delta(\mathcal{P}')) \neq \emptyset$.
Then $\exists \delta_o \in (\nabla^C(\mathcal{D}) \cap \Delta(\mathcal{P}'))$.
However, this δ_o is a valid diagnosis ($\delta_o \in \Delta(\mathcal{P}')$) and is disjoint from all diagnoses in $\mathcal{D}$ and as such $\delta_o \in \nabla^C(\mathcal{D})$
Therefore $\exists \delta_q \subseteq \delta_o$ *such that* $\delta_q \in \Delta_{\min}(\mathcal{P}')$.
Therefore $\delta_q \in \mathcal{D}$, but $\delta_q \in \mathcal{D} \Rightarrow \delta_o \notin \nabla^C(\mathcal{D})$, contradicting the premise. Both directions of the bi-implication are satisfied under the initial premise. □

Again, this lemma demonstrates that completeness of $\mathcal{D}$ can be proved by showing that none of the excluded hypotheses in ∇^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 ∇ or ∇^C. We can interpret ∇ as a disjunction and evaluate the consistency of:

$$SD, Obs', \Phi(\nabla) \overset{?}{\models} \bot. \tag{12}$$

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.

Running Example
Referring back to our example in §2, we can compute a sufficient (and indeed, critical) set of observations.

Taking $\mathcal{D}_{1,2} = \{\delta_1, \delta_2\}$, we can compute $\nabla(\mathcal{D})$:

$$\nabla(\mathcal{D}_{1,2}) = \{\emptyset, \{x_2\}, \{x_3\}\} \tag{13}$$

This result shows that a subset of observations is sufficient to prove minimality of $\mathcal{D}_{1,2}$ if it excludes three hypotheses: $\emptyset$ (no component is faulty), $\{x_2\}$ (only component x_2 is faulty) and $\{x_3\}$. There is only one critical subset of observations that achieves this: $\{o(s_2, A), o(s_3, A)\}$. Our example is small enough that we can intuitively grasp which observations are required, but as mentioned earlier, we can compute it by iteratively testing whether removing a specific observation effects the required properties.

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 $\mathcal{D}_{1,2}$. The set $\nabla^C(\mathcal{D}_{1,2})$ is defined by:

$$\nabla^C(\mathcal{D}_1) = \{\{b_0\}, \{x_0\}, \ldots, \{b_0, x_0\}, \{b_0, x_2\}, \ldots\} \tag{14}$$

In words, this is all hypotheses that do not include b_3, and/or x_2 and x_3 together, 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. Therefore we must include the observations $o(s_2, A)$ and $o(s_3, A)$, which has the effect of ruling out every hypothesis that does not mention $root$, b_0, b_3, x_2 or x_3. Adding the observation $o(s_0, N)$ has the effect of removing the hypotheses that mention b_0 or $root$, leaving only hypotheses that consistent with the model — those that include b_3 or both x_2 and x_3 — none of which belongs to $\nabla^C(\mathcal{D}_1)$. This is visualized in Figure 2.

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 δ_2. As δ_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.

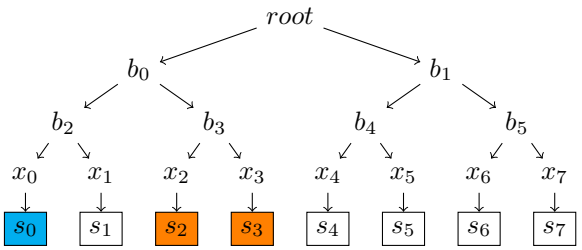

Figure 2: A Simple Power Network. Retaining only the critical observations on s_0, s_2, and s_3

5.2 Symbolic Representation of ∇

In the small running example with less than twenty components, the set ∇^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 δ may contain several components. The set $\nabla(\mathcal{D}) = \{\delta\}$ can be represented symbolically as follows:

$$\Gamma(\nabla(\mathcal{D}_\delta)) \equiv \left(\bigwedge_{c \in Comps \setminus \delta} \neg \mathrm{Ab}(c) \right) \wedge \left(\bigvee_{c \in \delta} \neg \mathrm{Ab}(c) \right). \tag{15}$$

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 δ_2 is ensured by proving that the joint assumptions: (1) no component outside δ_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 ∇^C. The symbolical representation of $\nabla^C(\mathcal{D})$ is:

$$\Gamma(\nabla^C(\mathcal{D})) \equiv \bigwedge_{\delta \in \mathcal{D}} \left(\bigvee_{c \in \delta} \neg \mathrm{Ab}(c) \right). \tag{16}$$

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 the size $\mathcal{D}$ may be exponential in the size of *Comps*).

Running Example

Back to the example, recall that $\mathcal{D}_{1,2} = \{\delta_1, \delta_2\}$. Completeness of $\mathcal{D}_{1,2}$ is ensured by proving that the following joint assumptions contradict the model and the observations:

- component b_3 is not faulty, and
- not both of components x_2 and x_3 are faulty (i.e. only one or zero are).

6 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 [1]. 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 3. 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 [4].

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 ∇ sets, is very similar in construction.

Figure 4 shows a reduced example of observation trace on the ADAPT-lite system. The system has a number of observation providing sensors: voltage and current sensors on the line (in blue and yellow), temperature sensors on the battery, position sensors on ciruit breakers and relays, (in purple) and a speed transmitter on the fan. 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 4, 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.

7 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.

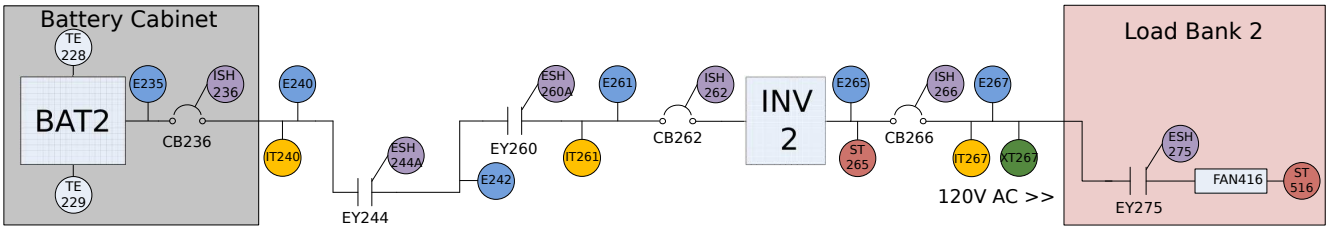

Figure 3: Schematic for ADAPT-lite

```
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 };
```

Figure 4: Example of Observations - Sensor IT240 offset

Optimal diagnosability is concerned with minimizing the number of sensors (or their total cost) while ensuring diagnosability [5] 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 [6], and its event-based variant [7], focuses on the problem of deciding which observation should be collected next in order to improve the precision of diagnosis (sequential diagnosis being generally performed in a context where faults have already been detected albeit not identified yet, while the dynamic observers are used before detection).

Both problems, as well as the one presented in this paper, aim at minimizing a cost associated with the observations (optimize the resources required to collect or to process the observations). The important difference between these tasks is that sequential diagnosis is performed *before* the final diagnoses are available.

In order to minimize the expected cost, sequential diagnosers use a conservative strategy, which selects the observations that split the set of candidates as evenly as possible. In contrast, an optimal choice for critical observations are the observations that isolate the set of final diagnoses from the non-diagnosis ones without regard to the size (or their associated probability mass) of these sets. As a consequence we doubt that the presented theory of critical observations can be used to improve the problem of sequential diagnosis.

An additional consequence is that the critical observations solver will automatically dismiss observations that *might* provide information but do not in the current situation. For instance, assume that the system presented as the running example includes an imperfect sensor that reports the status of bus x_2: either N, F, or U (unknown). If the observation is U, this information will not appear in the critical observations.

8 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 any specifics that may compromise consumer confidentiality, such as the precise time of consumption and the amount consumed.

Further research could be done in regards to the more practical question of computing the critical set of observations fast. One possibility would be to analyze the solving of

$Mod \wedge Obs \wedge \nabla$ in order to extract the specific observations used to prove inconsistency as part of the consistency check. However, whilst this set of observations may not necessarily be minimal, it has the potential to provide a good first approximation.

A last interesting extension is in considering event-based observations as opposed to state-based observations. 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. Identifying and classifying this information is therefore crucial in extracting the critical information from a large flow of alarms when considering event-based observations.

References

[1] T. Kurtoglu, S. Narasimhan, Sc. Poll, D. Garcia, L. Kuhn, J. de Kleer, Ar. 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.

[2] R. Reiter, "A theory of diagnosis from first principles," *Artificial Intelligence (AIJ)*, vol. 32, no. 1, pp. 57–95, 1987.

[3] J. de Kleer and Br. Williams, "Diagnosing multiple faults," *Artificial Intelligence (AIJ)*, vol. 32, pp. 97–130, 1987.

[4] Al. Grastien, "Diagnosis of hybrid systems by consistency testing," in *24th International Workshop on Principles of Diagnosis (DX-13)*, 2013, pp. 9–14.

[5] 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.

[6] J. de Kleer, "Using crude probability estimates to guide diagnosis," *Artificial Intelligence*, vol. 45, no. 3, pp. 381–391, 1990.

[7] F. Cassez and S. Tripakis, "Fault diagnosis with dynamic observers," in *Ninth International Workshop on Discrete Event Systems (WODES-08)*, 2008, pp. 212–217.