Foundation of Diagnosis and Predictability in Probabilistic Systems

Nathalie Bertrand, Serge Haddad, Engel Lefaucheux

Foundation of Diagnosis and Predictability in Probabilistic Systems*

Nathalie Bertrand[1], Serge Haddad[2], and Engel Lefaucheux[1,2]

1 **Inria, France** nathalie.bertrand@inria.fr
2 **LSV, ENS Cachan & CNRS & Inria, France**
{serge.haddad,engel.lefaucheux}@ens-cachan.fr

Abstract

In discrete event systems prone to unobservable faults, a diagnoser must eventually detect fault occurrences. The diagnosability problem consists in deciding whether such a diagnoser exists. Here we investigate diagnosis for probabilistic systems modelled by partially observed Markov chains also called probabilistic labeled transition systems (pLTS). First we study different specifications of diagnosability and establish their relations both in finite and infinite pLTS. Then we analyze the complexity of the diagnosability problem for finite pLTS: we show that the polynomial time procedure earlier proposed is erroneous and that in fact for all considered specifications, the problem is PSPACE-complete. We also establish tight bounds for the size of diagnosers. Afterwards we consider the dual notion of predictability which consists in predicting that in a safe run, a fault will eventually occur. Predictability is an easier problem than diagnosability: it is NLOGSPACE-complete. Yet the predictor synthesis is as hard as the diagnoser synthesis. Finally we introduce and study the more flexible notion of *prediagnosability* that generalizes predictability and diagnosability.

1 Introduction

Diagnosis. In computer science, diagnosis may refer to different kinds of activities. For instance, in artificial intelligence it can describe the process of identifying a disease from its symptoms, as performed by the expert system MYCIN [3]. In this work, we concentrate on diagnosis as studied in control theory, where it is applied to partially observable systems prone to faults. A sequence of observations of such a system is said to be surely correct (respectively surely faulty) if all possible runs corresponding to this sequence are correct (respectively faulty); otherwise the observed sequence is ambiguous. While monitoring the system, the *diagnoser* should rule out ambiguities, and in particular detect that a fault occurred; and the problem of existence of such a diagnoser is refered to as *diagnosability* [12]. In order to anticipate problems triggered by fault occurrences, one can also be interested in *predictors* that detect that a fault will eventually occur, and the *predictability* problem [6] is concerned with the existence of a predictor.

Diagnosis of discrete event systems. Diagnosability and predictability were first defined and studied in the framework of finite discrete event systems modelled by labeled transition systems (LTS), and the problems were shown to be solvable in PTIME (see [8] and [6], respectively). Despite the polynomial time complexity of the decision problems, for diagnosable (respectively predictable) LTS, the size of the diagnoser (respectively predictor) constructed by the algorithms may be exponential. Diagnosers as well as predictors must ensure two

* This work has been supported by project ImpRo ANR-2010-BLAN-0317 and the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement 257462 HYCON2 NOE.

requirements: *correctness*, meaning that the information provided by the diagnoser/predictor is accurate, and *reactivity*, ensuring that a fault will eventually be detected.

Diagnosis of probabilistic systems. Building on the work for LTS, the notion of diagnosability was later extended to Markov chains with labels on transitions, also called probabilitic labeled transition systems (pLTS) [13]. In a probabilistic context, the reactivity requirement now asks that faults will be almost surely eventually detected. Regarding correctness, two specifications have been proposed: either one sticks to the original definition and requires that the provided information is accurate, defining *A-diagnosability*; or one weakens the correctness by admitting errors in the provided information that should, however, have an arbitrary small probability when the delay before the diagnostic is long enough, defining *AA-diagnosability*. From a computational viewpoint, PTIME algorithms have been proposed to solve these two specifications of probabilistic diagnosability [4]. Predictability in pLTS with arbitrary small probability of erroneous information has also been studied in [5].

In case a system is not diagnosable, one may be able to control it, by forbidding some controllable actions, so that is becomes diagnosable. This property of *active diagnosability* has been studied for probabilistic systems in [1] pursuing the work of [11, 7] for discrete-event systems. Decidability and complexity issues are considered and optimal size diagnosers are synthesized. Interestingly, the diagnosability notion from [1] slightly differs from the original one in [13].

Remaining issues. Some issues remained untouched in the above line of work. First, diagnosability was only considered w.r.t. finite faulty runs. It seems as important to consider diagnosability of correct runs, and ambiguity can also be defined for infinite computations. Second, in most work, the complexity of the varied diagnosability problems and of the diagnosers synthesis were left open. Moreover, optimizing the delay between the fault occurrence and its detection is an important issue. Yet the search for diagnosers (or predictors) with optimal reactivity was not even considered. Last predictability and diagnosability were independently studied while combining them is obviously a fruitful direction.

Contributions. In this paper, we address the above mentioned gaps, and revisit diagnosability and predictability for probabilistic systems, from a semantical as well as a computational perspectives.

- In constrast to existing work, we define diagnosability directly on the ambiguity triggered by the behaviours of the system, and then establish that it is equivalent to the existence of a diagnoser.
- In order to give a firm semantical classification of diagnosability notions, we define criteria for diagnosability in probabilistic systems, depending on (1) whether the ambiguity is related to faulty runs only or to all runs and, (2) whether ambiguity is defined at the level of infinite runs, or for longer and longer finite subruns. A priori these two dimensions yield four specifications. We prove that two of them coincide leading to three main specifications: FF-diagnosability, IA-diagnosability used in [1] and FA-diagnosability, and we establish the connections between them. In addition we show that FF-diagnosability is equivalent to the A-diagnosability of [13] for finite pLTS and that this hypothesis is necessary.
- For finite state probabilistic systems, we show that these three notions of diagnosability can be characterized based on deterministic (finite or Büchi) automata acting as *monitors*, and synchronized with the pLTS. We further prove that the diagnosability problem (for all three specifications) is PSPACE-complete, contradicting the polynomial time result for FF-diagnosability [4], and identify the error in their algorithm.
- Afterwards, we design algorithms for the synthesis of finite-memory diagnosers and prove

that their size $2^{\Theta(n)}$ (where n is the number of states of the pLTS model) is optimal.
- Since predictability is an interesting alternative to diagnosability, we introduce two possible specifications for predictability in probabilistic systems, and show that in both cases the predictability problem is NLOGSPACE-complete. Yet, as for diagnosers, the optimal size of predictors is in $2^{\Theta(n)}$.
- Last, we introduce and study *prediagnosability* that combines the benefits of predictability and diagnosability: depending on the observations, a prediagnoser behaves as a diagnoser or a predictor. Prediagnosability is of interest since predictability is more difficult to achieve than diagnosability, also prediagnosers can be seen as "as soon as possible" diagnosers. For the varied notions of prediagnosability, we establish that the prediagnosability problem is PSPACE-complete and design prediagnosers with optimal size.
- Summarizing we provide a full picture of the hierarchy for the different notions and the frontier between NLOGSPACE and PSPACE-complete problems.

Organization. In Section 2, we introduce probabilistic LTS, define the possible diagnosability specifications, establish their connection. In Section 3, we provide characterizations for diagnosability of finite pLTS and we determine the exact complexity of the diagnosability problems. In Section 4, we design algorithms for synthesis of diagnosers with optimal size. In Section 5, we study predictability and prediagnosis, and focus on optimal diagnosers. All the proofs and additional results can be found in the companion research report [2].

2 Diagnosability specification

In the context of stochastic discrete event systems diagnosis, systems are often modeled using labeled transition systems.

▶ **Definition 1.** A *probabilistic labeled transition system* (pLTS) is a tuple $\mathcal{A} = \langle Q, q_0, \Sigma, T, \mathbf{P}\rangle$ where:
- Q is a set of states with $q_0 \in Q$ the initial state;
- Σ is a finite set of events;
- $T \subseteq Q \times \Sigma \times Q$ is a set of transitions;
- $\mathbf{P} : T \to \mathbb{Q}_{>0}$ is the probabilistic transition function fulfilling for all $q \in Q$: $\sum_{(q,a,q')\in T} \mathbf{P}(q,a,q') = 1$.

Observe that a pLTS is a labeled transition system (LTS) equipped with transition probabilities. The transition relation of the underlying LTS is defined by: $q \xrightarrow{a} q'$ for $(q,a,q') \in T$; this transition is then said to be *enabled* in q. A pLTS is said to be *live* if in every state q of the pLTS, a transition is enabled. We assume the pLTS we consider are countably branching, *i.e.*, in every state q, only countably many transitions are enabled, so that the summation $\sum_{(q,a,q')\in T} \mathbf{P}(q,a,q')$ is well-defined.

Let us now introduce some important notions and notations that will be used throughout the paper. A *run* ρ of a pLTS $\mathcal{A}$ is a (finite or infinite) sequence $\rho = q_0a_0q_1\dots$ such that for all i, $q_i \in Q$, $a_i \in \Sigma$ and when q_{i+1} is defined, $q_i \xrightarrow{a_i} q_{i+1}$. The notion of run can be generalized, starting from an arbitrary state q. We write Ω for the set of all infinite runs of $\mathcal{A}$ starting from q_0, assuming the pLTS is clear from context. When it is finite, ρ ends in a state q and its *length*, denoted $|\rho|$, is the number of actions occurring in it. Given a finite run $\rho = q_0a_0q_1\dots q_n$ and a (finite or infinite) run $\rho' = q_na_nq_{n+1}\dots$, we call concatenation of ρ and ρ' and we write $\rho\rho'$ the run $q_0a_0q_1\dots q_na_nq_{n+1}\dots$; the run ρ is then a *prefix* of $\rho\rho'$, which we denote $\rho \preceq \rho\rho'$. The *cylinder* defined by a finite run ρ is the set of all infinite

runs that extend ρ: $C(\rho) = \{\rho' \in \Omega \mid \rho \preceq \rho'\}$. The sequence associated with $\rho = qa_0q_1\ldots$ is the word $\sigma_\rho = a_0a_1\ldots$, and we write equally $q \overset{\rho}{\Rightarrow}$ or $q \overset{\sigma_\rho}{\Longrightarrow}$ (resp. $q \overset{\rho}{\Rightarrow} q'$ or $q \overset{\sigma_\rho}{\Longrightarrow} q'$) for an infinite (resp. finite) run ρ. A state q is *reachable* (from q_0) if there exists a run such that $q_0 \overset{\rho}{\Rightarrow} q$, which we alternatively write $q_0 \Rightarrow q$. The language of pLTS $\mathcal{A}$ consists of all infinite words that label runs of $\mathcal{A}$ and is formally defined as $\mathcal{L}^\omega(\mathcal{A}) = \{\, \sigma \in \Sigma^\omega \mid q_0 \overset{\sigma}{\Rightarrow} \,\}$.

Forgetting the labels and merging (and summing the probabilities of) the transitions with same source and target, a pLTS yields a discrete time Markov chain (DTMC). As usual for DTMC, the set of infinite runs of $\mathcal{A}$ is the support of a probability measure defined by Caratheodory's extension theorem from the probabilities of the cylinders:

$$\mathbb{P}(C(q_0a_0q_1\ldots q_n)) = \mathbf{P}(q_0, a_1, q_1)\cdots\mathbf{P}(q_{n-1}, a_{n-1}, q_n) \ .$$

In order to formalize problems related to fault diagnosis, we partition the event set Σ into two disjoint sets Σ_o and Σ_u, the sets of *observable* and of *unobservable events*, respectively. Moreover, we distinguish a special *fault* event $\mathbf{f} \in \Sigma_u$. Let $\sigma \in \Sigma^*$ be a finite word; its length is denoted $|\sigma|$. The projection of σ onto Σ_o is defined inductively by: $\mathcal{P}(\varepsilon) = \varepsilon$; for $a \in \Sigma_o$, $\mathcal{P}(\sigma a) = \mathcal{P}(\sigma)a$; and $\mathcal{P}(\sigma a) = \mathcal{P}(\sigma)$ for $a \notin \Sigma_o$. Write $|\sigma|_o$ for $|\mathcal{P}(\sigma)|$. When σ is an infinite word, its projection is the limit of the projections of its finite prefixes. This projection is applicable to runs via their associated sequence; it can be either finite or infinite. As usual the projection is extended to languages. With respect to the partition of $\Sigma = \Sigma_o \uplus \Sigma_u$, a pLTS $\mathcal{A}$ is *convergent* if there is no infinite sequence of unobservable events from any reachable state: $\mathcal{L}^\omega(\mathcal{A}) \cap \Sigma^*\Sigma_u^\omega = \emptyset$. When $\mathcal{A}$ is convergent, for every $\sigma \in \mathcal{L}^\omega(\mathcal{A})$, $\mathcal{P}(\sigma) \in \Sigma_o^\omega$. In the rest of the paper we assume that pLTS are convergent. We will refer to a *sequence* for a finite or infinite word over Σ, and an *observed sequence* for a finite or infinite sequence over Σ_o. Clearly, the projection onto Σ_o of a sequence yields an observed sequence.

The *observable length* of a run ρ denoted $|\rho|_o \in \mathbb{N} \cup \{\infty\}$, is the number of observable actions that occur in it: $|\rho|_o = |\sigma_\rho|_o$. A *signalling* run is a finite run whose last action is observable. Signalling runs are precisely the relevant runs w.r.t. partial observation issues since each observable event provides an additional information about the execution to an external observer. In the sequel, SR denotes the set of signalling runs, and SR_n the set of signalling runs of observable length n. Since we assume that the pLTS are convergent, for all $n > 0$, SR_n is equipped with a probability distribution defined by assigning measure $\mathbb{P}(\rho) = \mathbb{P}(C(\rho))$ to each $\rho \in \mathsf{SR}_n$. Given ρ a finite or infinite run, and $n \leq |\rho|_o$, $\rho_{\downarrow n}$ denotes the signalling subrun of ρ of observable length n. For convenience, we consider the empty run q_0 to be the single signalling run, of null length.

Let $\mathcal{A}$ be a pLTS. A run ρ is *faulty* if σ_ρ contains $\mathbf{f}$, otherwise it is *correct*. W.l.o.g., by considering two copies of each state, we assume that the states of $\mathcal{A}$ are partitioned into correct states and faulty states: $Q = Q_f \uplus Q_c$ where Q_f are faulty states, and Q_c correct states. Faulty (resp. correct) states are only reachable by faulty (resp. correct) runs. An observed sequence $\sigma \in \Sigma_o^\omega$ is *surely correct* if $\mathcal{P}^{-1}(\sigma) \cap \mathcal{L}^\omega(\mathcal{A}) \subseteq (\Sigma \setminus \mathbf{f})^\omega$; it is *surely faulty* if $\mathcal{P}^{-1}(\sigma) \cap \mathcal{L}^\omega(\mathcal{A}) \subseteq \Sigma^*\mathbf{f}\Sigma^\omega$; otherwise, it is *ambiguous*. For finite sequences, we need to rely on signalling runs: a finite observed sequence $\sigma \in \Sigma_o^*$ is *surely faulty* (resp. *surely correct*) if for every signalling run ρ with $\mathcal{P}(\sigma_\rho) = \sigma$, ρ is faulty (resp. correct); otherwise it is ambiguous. A (finite signalling or infinite) run ρ is *surely faulty* (resp. *surely correct*, *ambiguous*) if $\mathcal{P}(\rho)$ is surely faulty (resp. surely correct, ambiguous).

In order to introduce diagnosability, we define different subsets of infinite runs.

▶ **Definition 2** (Ambiguous runs)**.** Let $\mathcal{A}$ be a pLTS and $n \in \mathbb{N}$ with $n \geq 1$. Then:

- FAmb_∞ is the set of infinite faulty ambiguous runs of $\mathcal{A}$;

- CAmb_∞ is the set of infinite correct ambiguous runs of $\mathcal{A}$;
- FAmb_n is the set of infinite runs of $\mathcal{A}$ whose signalling subrun of observable length n is faulty and ambiguous;
- CAmb_n is the set of infinite runs of $\mathcal{A}$ whose signalling subrun of observable length n is correct and ambiguous.

We propose four possible specifications of diagnosability for probabilistic systems. There are two discriminating criteria: whether the non ambiguity requirement holds for faulty runs only (_F) or for all runs (_A), and whether ambiguity is defined at the infinite run level (I_) or for longer and longer finite signalling subruns (F_).

▶ **Definition 3** (Diagnosability specifications)**.** Let $\mathcal{A}$ be a pLTS. Then:

- A pLTS $\mathcal{A}$ is IF-diagnosable if $\mathbb{P}(\mathsf{FAmb}_\infty) = 0$.
- A pLTS $\mathcal{A}$ is IA-diagnosable if $\mathbb{P}(\mathsf{FAmb}_\infty \uplus \mathsf{CAmb}_\infty) = 0$.
- A pLTS $\mathcal{A}$ is FF-diagnosable if $\limsup_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n) = 0$.
- A pLTS $\mathcal{A}$ is FA-diagnosable if $\limsup_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \uplus \mathsf{CAmb}_n) = 0$.

Let us illustrate these specifications on the two pLTS of Figure 1 where $\{u, \mathbf{f}\}$ is the set of unobservable events, represented by dashed arrows. Here and later on, unless mentioned, the transitions outgoing a state are uniformly distributed. On the left, a faulty run will almost surely produce a b-event that cannot be mimicked by the single correct run. Thus this pLTS is IF–diagnosable. The unique correct run $\rho = q_0 u q_1 (a q_1)^\omega$ has probability $\frac{1}{2}$ and its corresponding observed sequence a^ω is ambiguous. Thus this pLTS is not IA-diagnosable. On the right, any infinite faulty run will contain a b-event, and cannot be mimicked by a correct run, therefore $\mathsf{FAmb}_\infty = \emptyset$. The two infinite correct runs have a^ω as observed sequence, and cannot be mimicked by a faulty run, thus $\mathsf{CAmb}_\infty = \emptyset$. As a consequence, this pLTS is IA-diagnosable. Consider now the infinite correct run $\rho = q_0 u q_1 (a q_1)^\omega$. It has probability $\frac{1}{2}$, and all its finite signalling subruns are ambiguous since their observed sequence is a^n, for some $n \in \mathbb{N}$. Thus for all $n \geq 1$, $\mathbb{P}(\mathsf{CAmb}_n) \geq \frac{1}{2}$, so that this pLTS is not FA-diagnosable.

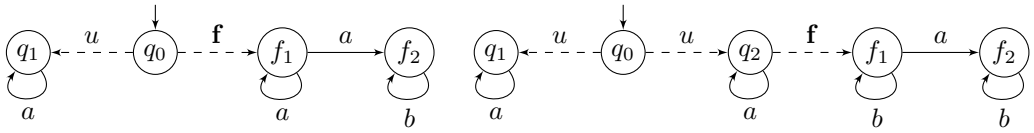

Figure 1 Left: a pLTS that is IF-diagnosable but not IA-diagnosable. Right: a pLTS that is IA-diagnosable but not FA-diagnosable.

The next theorem establishes the connections between these definitions.

▶ **Theorem 4.** *The different diagnosability notions for pLTS relate according to the table below. Moreover, all implications hold for infinite-state pLTS, and non implications already hold for finite-state pLTS. (The implication marked with * requires finitely branching pLTS.)*

Diagnosability	*All runs*		*Faulty runs*
Signalling runs	FA	$\Rightarrow$ $\not\Leftarrow$	FF
	$\Downarrow\not\Uparrow$		$\Downarrow\Uparrow^*$
Infinite runs	IA	$\Rightarrow$ $\not\Leftarrow$	IF

To conclude this section, we compare IF-diagnosability with A-diagnosability from [13].

▶ **Theorem 5.** *A finite pLTS $\mathcal{A}$ is* IF*-diagnosable if and only if it is A-diagnosable, that is:* $\forall \varepsilon > 0$, $\exists N_\varepsilon \in \mathbb{N}$, *for every faulty signalling run* ρ *and every* $n \geq N_\varepsilon$, $\mathbb{P}(\{\rho' \in \mathsf{FAmb}_{n+|\rho|_o} \mid \rho \preceq \rho'\}) < \varepsilon \mathbb{P}(\rho)$. *This condition is only sufficient for finitely branching infinite pLTS.*

3 Complexity of diagnosability

In this section, we establish the complexity of diagnosability stated in the next theorem.

▶ **Theorem 6.** *The* IF*-diagnosability,* IA*-diagnosability, and* FA*-diagnosability problems for finite pLTS are* PSPACE*-complete.*

To prove membership in PSPACE, we provide characterizations of the different diagnosability notions we introduced. For each notion of diagnosability, we proceed similarly. First, given a pLTS $\mathcal{A}$ we design a deterministic automaton that accepts some (finite or infinite) observed sequences of $\mathcal{A}$. Then we build the synchronized product of this automaton with $\mathcal{A}$, to obtain another pLTS with the same stochastic behaviour as $\mathcal{A}$ but augmented with additional information about the current run, that will be useful for diagnosability. Finally, we characterize diagnosability by graph properties on the synchronized product.

Here we only detail the procedure for IA-diagnosability. Its automaton $\mathsf{IA}(\mathcal{A})$ is the deterministic Büchi automaton introduced in [7]. Its states are triples of disjoint subsets of states (U, V, W) where given some observed sequence, U is the set of possible correct states and V and W are possible faulty states. The decomposition between V and W reflects the fact that the IA-automaton tries to resolve the ambiguity between U and W (when both are non empty), while V corresponds to a waiting room of states reached by faulty runs that will be examined when the current ambiguity is resolved. The set F of accepting states consists of all triples (U, V, W) with $U = \emptyset$ or $W = \emptyset$. When $U = \emptyset$, the current signalling run is surely faulty. When $W = \emptyset$ the current signalling run may be ambiguous (if $V \neq \emptyset$) but the "oldest" possible faulty runs have been discarded. Hence, any infinite observed sequence of $\mathcal{A}$ passing infinitely often through F is not ambiguous (ambiguities are resolved one after another).

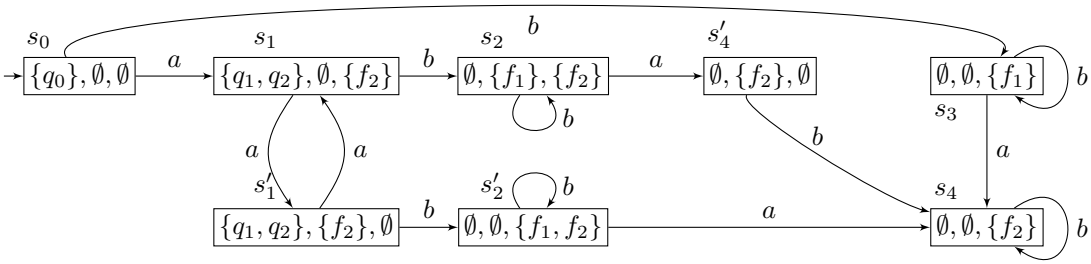

Figure 2 The IA-automaton of pLTS depicted on the right of Figure 1.

Figure 2 shows the IA-automaton of the pLTS depicted on the right of Figure 1. Observe that, despite the fact that all observed sequences a^n are ambiguous as witnessed by the possible faulty state f_2, a^ω, which is indeed unambiguous, is accepted by the IA-automaton since its execution infinitely often visits state $(\{q_1, q_2\}, \{f_2\}, \emptyset)$.

To come up with a characterization, one builds $\mathcal{A}_{\mathsf{IA}} = \mathcal{A} \times \mathsf{IA}(\mathcal{A})$, the product of $\mathcal{A}$ and $\mathsf{IA}(\mathcal{A})$ synchronized over observed events.

▶ **Proposition 7.** *A finite pLTS $\mathcal{A}$ is* IA*-diagnosable if and only if $\mathcal{A}_{\mathsf{IA}}$ has no bottom strongly connected component (BSCC) such that:*

- *either, all its states (q, U, V, W) fulfill $q \in Q_f$ and $U \neq \emptyset$;*
- *or all its states (q, U, V, W) fulfill $q \in Q_c$ and $W \neq \emptyset$.*

The decision algorithm for IA-diagnosability checks whether the above characterization is satisfied by looking for a state that violates the disjunction and then checking that it belongs to a BSCC. This can be done in polynomial space without explicitly building $\mathcal{A}_{\mathsf{IA}}$, and relying on Savitch's theorem.

In order to establish a lower bound for the complexity of IA-diagnosability, we introduce a variant of language universality. A language $\mathcal{L}$ over an alphabet Σ is said *eventually universal* if there exists a word $v \in \Sigma^*$ such that $v^{-1}\mathcal{L} = \Sigma^*$, where $v^{-1}\mathcal{L}$ denotes the left quotient of $\mathcal{L}$ by v: $v^{-1}\mathcal{L} = \{v' \mid vv' \in \mathcal{L}\}$. Recently, several variants of the universality problem were shown to be PSPACE-complete [10] but, to the best of our knowledge, eventual universality has not yet been considered.

Because of our diagnosis framework, we focus on live non deterministic finite automata (NFA). Similarly to pLTS, an NFA is *live* if from every state there is at least one outgoing transition. The language of an NFA $\mathcal{A}$, denoted $\mathcal{L}(\mathcal{A})$, is defined as the set of finite words that are accepted by $\mathcal{A}$. We reduce the universality problem for NFA, which is known to be PSPACE-complete [9] to the eventual universality problem to obtain the following result.

▶ **Proposition 8.** *Let $\mathcal{A}$ be a live NFA where all states are terminal. Then deciding whether $\mathcal{L}(\mathcal{A})$ is eventually universal is* PSPACE*-hard.*

Let us sketch how we reduce this problem to IA-diagnosability. Given a live NFA $\mathcal{A}$ over Σ where all states are terminal, one builds the pLTS of Figure 3 where $\Sigma \cup \{\sharp\}$ are observable. Since a correct run almost surely "outputs" a $\sharp$, ambiguity may only occur with faulty runs. Since after the fault one observes Σ^*, using the characterization in Proposition 7 one concludes that the pLTS is not IA-diagnosable if and only if $\mathcal{A}$ is eventually universal.

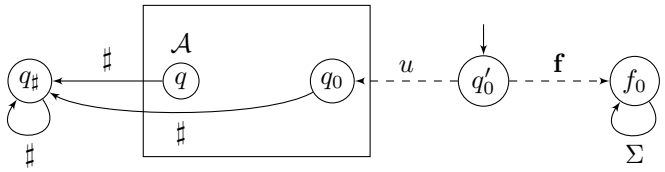

Figure 3 A reduction for PSPACE-hardness of IA-diagnosability.

Theorem 6 seems to contradict the PTIME decision procedure from [4] for A-diagnosability (or, equivalently IF-diagnosability). However, we establish that:

▶ **Fact 9.** *The* PTIME *algorithm of [4] for A-diagnosability is erroneous.*

4 Diagnoser construction

In this section, we focus on the construction of diagnosers. A diagnoser is a function $D : \Sigma_o^* \to \{?, \top, \bot\}$ assigning to every finite observation sequence a verdict. Informally when a diagnoser outputs ? it does not provide any information, while $\top$ ensures that a fault is certain and $\bot$ that some information about correctness has been provided. We consider the natural partial order $\prec$ on these values defined by $? \prec \top$ and $? \prec \bot$.

A finite memory diagnoser is given by a tuple $(M, \Sigma_o, m_0, \mathsf{up}, D_{fm})$ where M is a finite set of memory states, $m_0 \in M$ is the initial memory state, $\mathsf{up} : M \times \Sigma_o \to M$ is a memory update function, and finally $D_{fm} : M \to \{?, \top, \bot\}$ is a diagnoser function. The mapping up is extended into a function $\mathsf{up} : M \times \Sigma_o^* \to M$ defined inductively by $\mathsf{up}(m, \varepsilon) = m$ and $\mathsf{up}(m, wa) = \mathsf{up}(\mathsf{up}(m, w), a)$. A finite memory diagnoser is not a diagnoser as defined above, yet it induces the diagnoser defined by $D(w) = D_{fm}(\mathsf{up}(m_0, w))$.

Diagnosers we define in the sequel will have two important properties: soundness and reactivity. Soundness ensures that the information provided is accurate and reactivity specifies which pieces of information the diagnoser must provide. The precise soundness and reactivity requirements depend on the diagnosability notion of interest. Moreover, we restrict to diagnosers that, once they output $\top$, never change their verdict in the future. Note that any sound diagnoser can be turned into one that is sound and satisfies this commitment property. In this short version, we only introduce IA-diagnosers (the synthesis of FA-diagnosers and IF-diagnosers is similar and even simpler). Intuitively, IA-diagnosers may resolve an ambiguity late, while another one has already been produced.

▶ **Definition 10.** An IA-diagnoser for $\mathcal{A}$ is a function $D : \Sigma_o^* \to \{\top, \bot, ?\}$ such that

soundness For all $w \in \Sigma_o^*$
- if $D(w) = \top$, then w is surely faulty;
- if $D(w) = \bot$, letting $|D(w)|_\bot = |\{0 < n \leq |w| \mid D(w_{\leq n}) = \bot\}|$, then for all signalling run ρ such that $\mathcal{P}(\rho) = w$, $\rho_{\downarrow |D(w)|_\bot}$ is correct.

reactivity $\mathbb{P}(\{\rho \in \Omega \mid D_{\sup}(\mathcal{P}(\rho)) = ?\}) = 0$ where for $w \in \Sigma_o^\omega$, $D_{\sup}(w) = \limsup_{n\to\infty} D(w_{\leq n})$.

($D_{\sup}$ is well-defined since once the diagnoser outputs $\top$, it always sticks to this verdict.)

The reactivity condition requires that almost surely the diagnoser detects a fault or guarantees that longer and longer subruns of the current run are correct. Soundness of $\top$ verdict implies that indeed the run is faulty. The interpretation of $D(w) = \bot$ is that the diagnoser ensures that any signalling subrun of length $|D(w)|_\bot \leq |w|$ of a signalling run for w is correct. Of course it may deduce this information from the last $|w| - |D(w)|_\bot$ observations. This is illustrated on the example of Figure 4 for which we describe an IA-diagnoser. After observing any sequence $wbaa$, with $w \in \{a, b\}^*$, the diagnoser knows a posteriori that two steps before, that is after the observation of wb, the run was necessarily correct. Indeed, observing the suffix aa is not possible after a fault, yet wba is not surely correct. The function D defined by: for $w \in \{a, b\}^*(ab + aa)$, $D(w) = \bot$, for $w \in \{a, b, c\}^*c$, $D(w) = \top$ and otherwise $D(w) = ?$, is an IA-diagnoser.

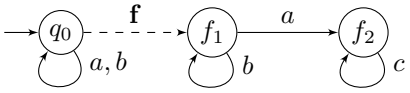

Figure 4 A pLTS which is IA-diagnosable.

The next proposition establishes that this definition of diagnosers is appropriate for IA-diagnosability. Furthermore it provides tight lower and upper bounds for the size of IA-diagnosers. The pLTS of Figure 5 is used to prove the lower bound. Intuitively, every IA-diagnoser for this pLTS must decide, on observing a c, whether the run is faulty or correct. To do so, it must remember whether, n observations earlier, the event was a or b. Due to the self-loop on q_0, it cannot know when a c will occur, and must remenber the n last observations. This requires at least 2^n memory states.

▶ **Proposition 11.** *A finite pLTS $\mathcal{A}$ is* IA*-diagnosable if and only if it admits an* IA*-diagnoser. For every pLTS $\mathcal{A}$ with n_c correct states and n_f faulty states which is* IA*-diagnosable, one can build an* IA*-diagnoser with at most $2^{n_c}3^{n_f}$ states. There is a family $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ of* IA*-diagnosable pLTS such that $\mathcal{A}_n$ has $2n+2$ states and it admits no* IA*-diagnoser with less than 2^n memory states.*

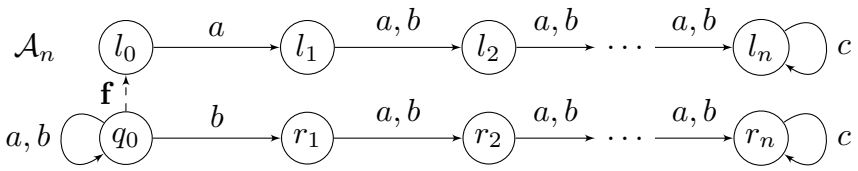

Figure 5 Example of an IA-diagnosable pLTS requiring an IA-diagnoser with exponential size.

5 Predictability and prediagnosis

Predictability. Fault predictability has been first introduced for LTS in [6]: in words, an LTS is predictable (resp. k-predictable) if a fault can be predicted (resp. at least before k observations) whatever the future behavior of the LTS. There are two possible adaptations for pLTS: (1) either one sticks to the original definition and requires that the fault surely occurs or, (2) one relaxes it and only requires that the fault almost surely occurs.

In order to reason about predictability, we introduce some particular prefixes of a run. For a finite run ρ, and $k \in \mathbb{N}$, we define $pre_k(\rho)$, the k-past of ρ, by $pre_k(\rho) = \rho_{\downarrow|\rho|_o - \min(k,|\rho|_o)}$. For example, in the pLTS of Figure 6, $pre_0(q_0bq_1\mathbf{f}q_2) = q_0bq_1$ as $\mathbf{f}$ is unobservable and $pre_1(q_0bq_1\mathbf{f}q_2) = q_0$. In fact for $k \geq 1$, $pre_k(q_0bq_1\mathbf{f}q_2) = q_0$.

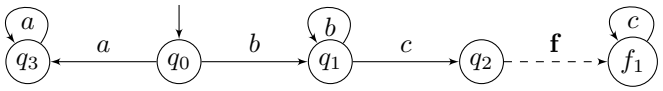

Figure 6 A 0-surely predictable and 1-predictable pLTS.

We also introduce sets of observed sequences defined on their possible future behaviors. In words, an observed sequence σ forbids prediction of a fault when there is still either a correct infinite run where σ is a prefix of its observed sequence (UPC) or a set of positive measure of such runs (UPSC). Thus in order to be k-predictable, k observations before a possible fault the observed sequence should not belong to these sets (see Definition 13).

▶ **Definition 12** (ultimately possibly (significantly) correct)**.** Let σ be a finite observed sequence of a pLTS $\mathcal{A}$. Then:

- σ is *ultimately possibly correct* if $\{\rho' \in \Omega \mid \sigma \preceq \mathcal{P}(\rho')\} \cap \mathsf{C}_\infty \neq \emptyset$. The set of ultimately possibly correct observed sequences is denoted UPC.
- σ is *ultimately possibly significantly correct* if $\mathbb{P}(\{\rho' \in \Omega \mid \sigma \preceq \mathcal{P}(\rho')\} \cap \mathsf{C}_\infty) > 0$. The set of ultimately possibly significantly correct observed sequences is denoted UPSC.

▶ **Definition 13** ((sure) predictability)**.** Let $k \in \mathbb{N}$.

- A pLTS $\mathcal{A}$ is *k-surely predictable* if for every run $\rho\mathbf{f}q$ of $\mathcal{A}$, $\mathcal{P}(pre_k(\rho)) \notin$ UPC;

- A pLTS $\mathcal{A}$ is *k-predictable* if for every run $\rho\mathbf{f}q$ of $\mathcal{A}$, $\mathcal{P}(pre_k(\rho)) \notin \mathsf{UPSC}$.

Observe that in the previous definition, one can safely restrict to check the condition on correct runs ρ by considering the first occurrence of a fault in the run $\rho\mathbf{f}q$.

For example, the pLTS of Figure 6 is 0-surely predictable. Every correct run ρ that is followed by $\mathbf{f}$ is such that $\mathcal{P}(\rho) = b^n c$ for some $n \geq 1$. As it is the unique signalling run with such an observed sequence, the fault can be predicted. It is not 1-surely predictable as the 1-past of $\rho = q_0 b q_1 c q_2 \mathbf{f} f_1$ is $pre_1(\rho) = q_0 b q_1$ and the infinite run $\rho' = q_0(bq_1)^\omega$ is correct. However it is 1-predictable as for every signalling run with observed sequence b^n for some $n \geq 1$ (thus ending in q_1) a fault eventually almost surely occurs. Finally it is not 2-predictable since the 2-past of $\rho = q_0 b q_1 c q_2 \mathbf{f} f_1$ is q_0 and the infinite correct run $\rho = q_0(aq_3)^\omega$ has probability $\frac{1}{2}$.

We have established all the relations between the different notions of diagnosability and predicatibility (see Figure 8 in the conclusion). The main result about predicatibility is given in the next theorem, and highlights the complexity gap between predictability and diagnosability for probabilistic systems (recall that their complexity coincide for LTS). Despite this difference, the size of optimal predictors is comparable to the one of optimal diagnosers (see details in our research report [2]).

▶ **Theorem 14.** *Deciding, given $\mathcal{A}$ a pLTS and $k \in \mathbb{N}$, whether $\mathcal{A}$ is k-predictable (resp. surely k-predictable) is an* NLOGSPACE*-complete problem. Moreover, the same complexity applies assuming k is fixed (rather than given as input).*

Prediagnosis. On the one hand, diagnosis is concerned with detection of faults that have occurred: given a sequence of observations a diagnoser tries to detect that a fault has occurred in the *past* of all consistent behaviors. On the other hand, prediction is concerned with anticipation of faults: given a sequence of observations a predictor tries to detect that a fault will eventually occur in the *future* of all consistent behaviors. The notion we introduce now, *prediagnosis*, concerns detection of faults both in the past and in the future.

Let us start by introducing two sets of infinite faulty runs that make prediagnosis impossible. FUPC_∞ is the set of faulty runs that admit for all their finite prefixes a compatible infinite correct run. The condition is strengthened for FUPSC_∞ which gathers the faulty runs that admit for all their finite prefixes, a positive measure of compatible infinite correct runs.

▶ **Definition 15.** Let $\mathcal{A}$ be a pLTS. Then:
- FUPC_∞, the set of *faulty, ultimately possibly correct* runs is defined by:
  $\mathsf{FUPC}_\infty = \{\rho \in \Omega \mid \rho \text{ faulty and } \forall i \in \mathbb{N},\ \mathcal{P}(\rho_{\downarrow i}) \in \mathsf{UPC}\}$
- FUPSC_∞, the set of *faulty, ultimately possibly significantly correct* runs is defined by:
  $\mathsf{FUPSC}_\infty = \{\rho \in \Omega \mid \rho \text{ faulty and } \forall i \in \mathbb{N},\ \mathcal{P}(\rho_{\downarrow i}) \in \mathsf{UPSC}\}$

The reactivity requirement for prediagnosers will impose that these sets are negligible. The difference between these two sets impacts correctness: relying on FUPC_∞ provides a *sure* correctness while relying on FUPSC_∞ only provides an *almost sure* correctness.

▶ **Definition 16** ((Sure) Prediagnosability)**.** Let $\mathcal{A}$ be a pLTS. Then:
- $\mathcal{A}$ is *surely prediagnosable* if $\mathbb{P}(\mathsf{FUPC}_\infty)=0$;
- $\mathcal{A}$ is *prediagnosable* if $\mathbb{P}(\mathsf{FUPSC}_\infty)=0$.

Surprisingly, sure prediagnosability lies strictly between FF-diagnosability and IF-diagnosability with equivalence for finitely branching pLTS. Also (sure) 0-predictability implies (sure) prediagnosability. As expected, the less demanding specification is prediagnosability. All the

relations that we have established between the different notions of diagnosability, predictability and prediagnosability are described by Figure 8 in the conclusion. From a complexity point of view, prediagnosability is equivalent to diagnosability:

▶ **Theorem 17.** *The (sure) prediagnosability problem is* PSPACE*-complete.*

In the research report [2], we formally define and study the notion of *prediagnosers*. Here we informally discuss the interest of prediagnosers. While sure prediagnosability and IF-diagnosability are equivalent for finitely branching pLTS, there are differences between sure prediagnosers and IF-diagnosers. An IF-diagnoser is a sure prediagnoser, but a sure prediagnoser may output a verdict $\top$ even before a fault. This phenomenon occurs even if the pLTS is non predictable. The non predictable pLTS of Figure 7 points out this difference. A diagnoser may output $\top$ only after observing two a's, since then, surely a fault occurred. In contrast, a sure prediagnoser can already output $\top$ after observing the first a. In fact this pLTS is FA-diagnosable since after an occurrence of b, the run is surely correct. Prediagnosers, can be thought of as monitors that emit verdicts as soon as possible, while preserving soundness. In the proof that prediagnosability is equivalent to the existence of a prediagnoser, the prediagnosers we construct are indeed optimal in that sense.

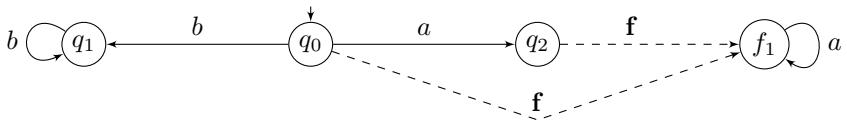

Figure 7 A non-predictable pLTS, for which a sure prediagnoser is quicker than all diagnosers.

6 Conclusion

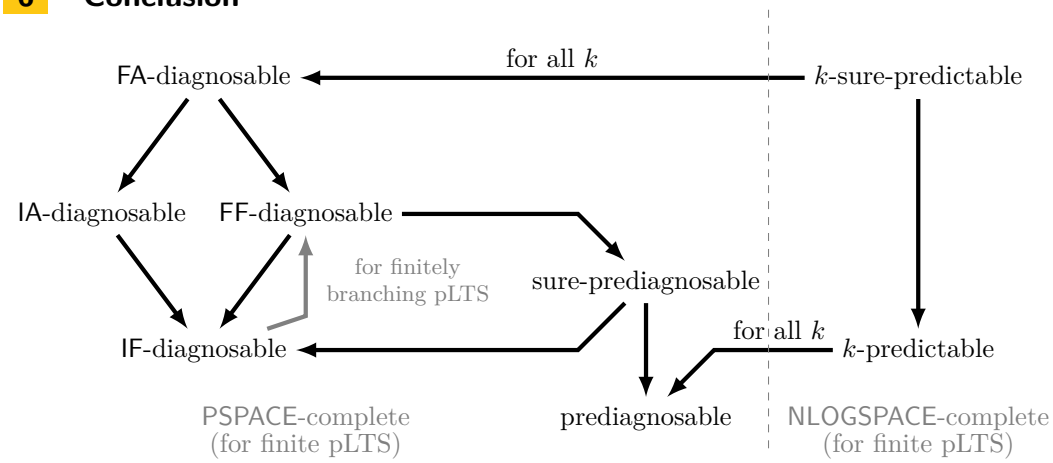

Figure 8 Summarizing relations between specifications, and associated complexities.

In this work, we settled the foundations of diagnosability and predictability for partially observed stochastic systems. In particular, we investigated semantical issues and provided several meaningful definitions for diagnosability and predictability in a probabilistic context. We also introduced prediagnosability, that combines the advantages of diagnosability and predictability. Beyond providing relations between these notions, we obtained tight complexity bounds using graph-based characterizations on the product of the system under scrutiny

and an appropriate monitor. The complexity ranges from NLOGSPACE-completeness for predictability to PSPACE-completeness for diagnosability and prediagnosability, as summarized on Figure 8. Last, we proved exponential almost matching lower and upper bounds for the diagnosers, predictors, and prediagnosers synthesis problems.

The present contribution opens several interesting research perspectives. First of all, the decidability status (and in the positive case, the precise complexity) of the approximate diagnosability (AA-diagnosability) introduced in [13] is still open since we only proved the algorithm from [4] to be erroneous (see [2]). Second, beyond diagnosability and its variants (predictability and prediagnosability), we wish to conduct a systematic study of other paradigms related to partial observability, such as opacity or detectability, in a probabilistic context. Last, we plan to move to more quantitative versions of diagnosis including optimization issues. The objective would be to minimize the observational capacities of the monitor, either spatially or timely by restricting either the observable actions, or the observation time instants, while preserving diagnosability.

References

1 N. Bertrand, E. Fabre, S. Haar, S. Haddad, and L. Hélouët. Active diagnosis for probabilistic systems. In *Proceedings of FoSSaCS'14*, volume 8412 of *Lecture Notes in Computer Science*, pages 29–42. Springer, 2014.

2 N. Bertrand, S. Haddad, and E. Lefaucheux. Foundation of diagnosis and predictability in probabilistic systems. Research Report LSV-14-09, June 2014. Available at `http://www.lsv.ens-cachan.fr/Publis/RAPPORTS_LSV/PDF/rr-lsv-2014-09.pdf`.

3 B.G. Buchanan and E.H. Shortliffe. *Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, 1984.

4 J. Chen and R. Kumar. Polynomial test for stochastic diagnosability of discrete-event systems. *IEEE Transactions on Automation Science and Engineering*, 10(4):969–979, 2013.

5 J. Chen and R. Kumar. Failure prognosability of stochastic discrete event systems. In *Proceedings of ACC'14*, pages 2041–2046. IEEE, 2014.

6 S. Genc and S. Lafortune. Predictability of event occurrences in partially-observed discrete-event systems. *Automatica*, 45(2):301–311, 2009.

7 S. Haar, S. Haddad, T. Melliti, and S. Schwoon. Optimal constructions for active diagnosis. In *Proceedings of FSTTCS'13*, volume 24 of *LIPIcs*, pages 527–539. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2013.

8 S. Jiang, Z. Huang, V. Chandra, and R. Kumar. A polynomial algorithm for testing diagnosability of discrete-event systems. *IEEE Transactions on Automatic Control*, 46(8):1318–1321, 2001.

9 A. R. Meyer and L. J. Stockmeyer. The equivalence problem for regular expressions with squaring requires exponential space. In *Proceedings of SWAT'72*, pages 125–129. IEEE Computer Society, 1972.

10 N. Rampersad, J. Shallit, and Z. Xu. The computational complexity of universality problems for prefixes, suffixes, factors, and subwords of regular languages. *Fundamenta Informaticae*, 116(1-4):223–236, 2012.

11 M. Sampath, S. Lafortune, and D. Teneketzis. Active diagnosis of discrete-event systems. *IEEE Transactions on Automatic Control*, 43(7):908–929, 1998.

12 M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzis. Diagnosability of discrete-event systems. *IEEE Transactions on Automatic Control*, 40(9):1555–1575, 1995.

13 D. Thorsley and D. Teneketzis. Diagnosability of stochastic discrete-event systems. *IEEE Transactions on Automatic Control*, 50(4):476–492, 2005.