Accurate approximate diagnosability of stochastic systems

Nathalie Bertrand, Serge Haddad, Engel Lefaucheux

Accurate approximate diagnosability of stochastic systems

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

[1] Inria Rennes, France
[2] LSV, ENS Cachan & CNRS & Inria, France

Abstract. Diagnosis of partially observable stochastic systems prone to faults was introduced in the late nineties. Diagnosability, *i.e.* the existence of a diagnoser, may be specified in different ways: (1) exact diagnosability (called A-diagnosability) requires that almost surely a fault is detected and that no fault is erroneously claimed while (2) approximate diagnosability (called ε-diagnosability) allows a small probability of error when claiming a fault and (3) accurate approximate diagnosability (called AA-diagnosability) requires that this error threshold may be chosen arbitrarily small. Here we mainly focus on approximate diagnoses. We first refine the almost sure requirement about finite delay introducing a uniform version and showing that while it does not discriminate between the two versions of exact diagnosability this is no more the case in approximate diagnosis. Then we establish a complete picture for the decidability status of the diagnosability problems: (uniform) ε-diagnosability and uniform AA-diagnosability are undecidable while AA-diagnosability is decidable in PTIME, answering a longstanding open question.

1 Introduction

Diagnosis and diagnosability. The increasing use of software systems for critical operations motivates the design of fast automatic detection of malfunctions. In general, diagnosis raises two important issues: deciding whether the system is *diagnosable* and, in the positive case, synthesizing a *diagnoser* possibly satisfying additional requirements about memory size, implementability, etc. One of the proposed approaches consists in modelling these systems by partially observable labelled transition systems (LTS) [10]. In such a framework, diagnosability requires that the occurrence of unobservable faults can be deduced from the previous and subsequent observable events. Formally, an LTS is diagnosable if there exists a diagnoser that satisfies *reactivity* and *correctness* contraints. Reactivity requires that if a fault occurred, the diagnoser eventually detects it. Correctness asks that the diagnoser only claims the existence of a fault when there actually was one. Diagnosability for LTS was shown to be decidable in PTIME [6] while the diagnoser itself could be of size exponential

w.r.t. the size of the LTS. Diagnosis has been extended to numerous models (Petri nets [2], pushdown systems [7], etc.) and settings (centralized, decentralized, distributed), and have had an impact on important application areas, *e.g.* for telecommunication network failure diagnosis. Also, several contributions, gathered under the generic name of active diagnosis, focus on enforcing the diagnosability of a system [3, 4, 9, 12].

Diagnosis of stochastic systems. Diagnosis was also considered in a quantitative setting, and namely for probabilistic labelled transition systems (pLTS) [1, 11], that can be seen as Markov chains in which the transitions are labelled with events. Therefore, one can define a probability measure over infinite runs. In that context, the specification of reactivity and correctness can be relaxed. Here, reactivity only asks to detect faults almost surely (*i.e.* with probability 1). This weaker reactivity constraint takes advantage of probabilities to rule out negligible behaviours. For what concerns correctness, three natural variants can be considered. *A-diagnosability* sticks to strong correctness and therefore asks the diagnoser to only claim fault occurrences when a fault is certain. *ε-diagnosability* tolerates small errors, allowing to claim a fault if the conditional probability that no fault occurred does not exceed ε. *AA-diagnosability* requires the pLTS to be ε-diagnosable for all positive ε, allowing the designer to select a threshold according to the criticality of the system. A-diagnosability and AA-diagnosability were introduced in [11]. Recently, we focused on semantical and algorithmic issues related to A-diagnosability, and in particular we established that A-diagnosability is PSPACE-complete [1]. When it comes to approximate diagnosability (*i.e.* ε and AA-diagnosability), up to our knowledge, a (PTIME-checkable) sufficient condition for AA-diagnosability [11] has been given, but no decidability result is known.

Contributions. Our contributions are twofold. From a semantical point of view, we investigate the specification of reactivity, introducing *uniform reactivity* which requires that once a fault occurs, the probability of detection when time elapses converges to 1 uniformly w.r.t. faulty runs. Uniformity provides the user with a stronger guarantee about the delay before detection. We show that uniform A-diagnosability and A-diagnosability coincide while this is no longer the case for approximate diagnosability. From an algorithmic point of view, we first show that ε-diagnosability and its uniform version are undecidable. Then we characterize AA-diagnosability as a separation property between labelled Markov chains (LMC), precisely a *distance* 1 between appropriate pairs of LMCs built from the pLTS. Thanks to [5], this yields a polynomial time algo-

rithm for AA-diagnosability. Surprisingly, while an AA-diagnoser may require infinite memory contrary to the case of A-diagnosers, AA-diagnosability can be checked more efficiently than A-diagnosability (PTIME vs PSPACE). Finally, we show that uniform AA-diagnosability is undecidable.

Organization. In Section 2, we introduce the different variants of diagnosability and establish the full hierarchy between these specifications. In Section 3, we address the decidability and complexity issues related to approximate diagnosis. Full proofs are postponed to the Appendix.

2 Specification of diagnosability

Probabilistic labelled transition systems.

To represent stochastic discrete event systems, we use transition systems labelled with events and in which the transition function is probabilistic.

Definition 1. *A* probabilistic labelled transition system *(pLTS) is a tuple* $\mathcal{A} = \langle Q, q_0, \Sigma, T, \mathbf{P} \rangle$ *where:*

- Q *is a finite 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 probability function fulfilling for every* $q \in Q$*:* $\sum_{(q,a,q') \in T} \mathbf{P}[q, a, q'] = 1$.

Observe that a pLTS is a labelled 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.

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 $\rho = q_0 a_0 q_1 \dots$ 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 $\rho = q_0 a_0 q_1 \dots q_n$ and a (finite or infinite) run $\rho' = q_n a_n q_{n+1} \dots$, we call concatenation of ρ and ρ' and we write $\rho\rho'$ for the run $q_0 a_0 q_1 \dots q_n a_n q_{n+1} \dots$; the run ρ is then a *prefix* of $\rho\rho'$, which we denote $\rho \preceq \rho\rho'$. The *cylinder* generated by a finite run ρ consists of all infinite runs that extend ρ: $\mathsf{Cyl}(\rho) = \{\rho' \in \Omega \mid \rho \preceq \rho'\}$. The

sequence associated with $\rho = qa_0q_1\ldots$ is the word $\sigma_\rho = a_0a_1\ldots$, and we write indifferently $q \overset{\rho}{\Rightarrow}$ 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:

$$\mathbb{P}_{\mathcal{A}}(\mathsf{Cyl}(q_0a_0q_1\ldots q_n)) = \mathbf{P}[q_0, a_1, q_1]\cdots\mathbf{P}[q_{n-1}, a_{n-1}, q_n] \ .$$

When $\mathcal{A}$ is fixed, we may omit the subscript. To simplify, for ρ a finite run, we will sometimes abuse notation and write $\mathbb{P}(\rho)$ for $\mathbb{P}(\mathsf{Cyl}(\rho))$. If R is a (denumerable) set of finite runs (such that no run is a prefix of another one), we write $\mathbb{P}(R)$ for $\sum_{\rho\in R}\mathbb{P}(\rho)$.

Partial observation and ambiguity

Beyond the pLTS model for stochastic discrete event systems, in order to formalize problems related to fault diagnosis, we partition 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 σ be a finite word over Σ; its length is denoted $|\sigma|$. The projection of words onto Σ_o is defined inductively by: $\pi(\varepsilon) = \varepsilon$; for $a \in \Sigma_o$, $\pi(\sigma a) = \pi(\sigma)a$; and $\pi(\sigma a) = \pi(\sigma)$ for $a \notin \Sigma_o$. We write $|\sigma|_o$ for $|\pi(\sigma)|$. When σ is an infinite word, its projection is the limit of the projections of its finite prefixes. As usual the projection mapping is extended to languages: for $L \subseteq \Sigma^*$, $\pi(L) = \{\pi(\sigma) \mid \sigma \in L\}$. With respect to the partition of $\Sigma = \Sigma_o \uplus \Sigma_u$, a pLTS $\mathcal{A}$ is *convergent* if, from any reachable state, there is no infinite sequence of unobservable events: $\mathcal{L}^\omega(\mathcal{A}) \cap \Sigma^*\Sigma_u^\omega = \emptyset$. When $\mathcal{A}$ is convergent, for every $\sigma \in \mathcal{L}^\omega(\mathcal{A})$, $\pi(\sigma) \in \Sigma_o^\omega$. In the rest of the paper we assume that pLTS are convergent. We will use the terminology *sequence* for a word $\sigma \in \Sigma^* \cup \Sigma^\omega$, and an *observed sequence* for a word $\sigma \in \Sigma_o^* \cup \Sigma_o^\omega$. The projection of a sequence to Σ_o is thus an observed sequence.

The *observable length* of a run ρ denoted $|\rho|_o \in \mathbb{N} \cup \{\infty\}$, is the number of observable events that occur in it: $|\rho_o| = |\sigma_\rho|_o$. A *signalling run* is a finite run ending with and observable event. Signalling runs are precisely the relevant runs w.r.t. partial observation issues since each observable

event provides an external observer additional information about the execution. 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 pLTS to be convergent, for every $n > 0$, SR_n is equipped with a probability distribution defined by assigning measure $\mathbb{P}(\rho)$ 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. For an observed sequence $\sigma \in \Sigma_o^*$, we define its cylinder $\mathsf{Cyl}(\sigma) = \sigma\Sigma_o^*$ and the associated probability $\mathbb{P}(\mathsf{Cyl}(\sigma)) = \mathbb{P}(\{\rho \in \mathsf{SR}_{|\sigma|} \mid \pi(\rho) = \sigma\})$, often shortened as $\mathbb{P}(\sigma)$.

Let us now partition runs depending on whether they contain a fault or not. A run ρ is *faulty* if σ_ρ contains $\mathbf{f}$, otherwise it is *correct*. For $n \in \mathbb{N}$, we write F_n (resp. C_n) for the set of faulty (resp. correct) signalling runs of length n, and further define the set of all faulty and correct signalling runs $\mathsf{F} = \cup_{n\in\mathbb{N}}\mathsf{F}_n$ and $\mathsf{C} = \cup_{n\in\mathbb{N}}\mathsf{C}_n$. W.l.o.g., by considering two copies of each state, we assume that the state space Q is partitioned into correct states and faulty states: $Q = Q_f \uplus Q_c$ such that faulty (resp. correct) states, *i.e.* states in Q_f (resp. Q_c) are only reachable by faulty (resp. correct) runs. An infinite (resp. finite) observed sequence $\sigma \in \Sigma_o^\omega$ (resp. Σ_o^*) is *ambiguous* if there exists a correct infinite (resp. signalling) run ρ and a faulty infinite (resp. signalling) run ρ' such that $\pi(\rho) = \pi(\rho') = \sigma$.

Fault diagnosis

Whatever the considered notion of diagnosis in probabilistic systems, *reactivity* requires that when a fault occurs, a diagnoser almost surely will detect it after a finite delay. We refine this requirement by also considering *uniform reactivity* ensuring that given any positive probability threshold α there exists a delay n_α such that the probability to exceed this delay is less or equal than α. Here uniformity means "independently of the faulty run".

Similarly, *correctness* of the diagnosis may be specified in different ways. Since we focus on approximate diagnosis, a fault can be claimed after an ambiguous observed sequence. This implies that ambiguity should be quantified in order to assess the quality of the diagnosis. To formalise this idea, with every observed sequence $\sigma \in \Sigma_o^*$ we associate a *correctness proportion*

$$\mathsf{CorP}(\sigma) = \frac{\mathbb{P}(\{\rho \in \mathsf{C}_{|\sigma|} \mid \pi(\rho) = \sigma\})}{\mathbb{P}(\{\rho \in \mathsf{C}_{|\sigma|} \cup \mathsf{F}_{|\sigma|} \mid \pi(\rho) = \sigma\})} \ ,$$

which is the conditional probability that a signalling run is correct given that its observed sequence is σ. Thus approximate diagnosability also denoted *ε-diagnosability* allows the diagnoser to claim a fault when the correctness proportion does not exceed ε while accurate approximate diagnosability denoted *AA-diagnosability* ensures that ε can be chosen as small as desired but still positive.

Definition 2 (Diagnosability notions). *Let $\mathcal{A}$ be a pLTS and $\varepsilon \geq 0$.*

- *$\mathcal{A}$ is* ε-diagnosable *if for all faulty run $\rho \in \mathsf{F}$ and all $\alpha > 0$ there exists $n_{\rho,\alpha}$ such that for all $n \geq n_{\rho,\alpha}$:*

$$\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > \varepsilon\}) \leq \alpha \mathbb{P}(\rho).$$

  $\mathcal{A}$ is uniformly ε-diagnosable *if $n_{\rho,\alpha}$ does not depend on ρ.*
- *$\mathcal{A}$ is* (uniformly) AA-diagnosable *if it is (uniformly) ε-diagnosable for all $\varepsilon > 0$.*

Two variants of diagnosability for stochastic systems were introduced in [11]: AA-diagnosability and A-diagnosability. *A-diagnosability*, which corresponds to exact diagnosis, is nothing else but 0-diagnosability in Definition 2 wording. By definition, A-diagnosability implies AA-diagnosability which implies ε-diagnosability for all $\varepsilon > 0$. Observe also that since ρ (and so $\mathbb{P}(\rho)$) is fixed, ε-diagnosability can be rewritten:

$$\lim_{n \to \infty} \mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > \varepsilon\}) = 0.$$

We now provide examples that illustrate these notions. Consider $\mathcal{A}_1$, the pLTS represented on Figure 1. We claim that $\mathcal{A}_1$ is AA-diagnosable but neither A-diagnosable, nor uniformly AA-diagnosable. We only give here intuitions on these claims, and refer the reader to the proof of Proposition 1 for details (see Appendix A.1). First an ε-diagnoser will look at the proportion of b occurrences and if the sequence is "long" enough and the proportion is "close" to $\frac{3}{4}$, it will claim a fault. However, the delay $n_{\alpha,\rho}$ before claiming a fault cannot be selected independently of the faulty run. Indeed, given the faulty run $\rho_n = q_0 \mathbf{f} q_f (a q_f)^n$, we let $p_{n,m}$ for the probability of extensions of ρ_n by m observable events and with correctness proportion below ε. In order for $p_{n,m}$ to exceed $1-\alpha$, m must depend on n. So $\mathcal{A}_1$ is not uniformly AA-diagnosable. $\mathcal{A}_1$ is neither A-diagnosable since all observed sequences of faulty runs are ambiguous.

Consider now the pLTS $\mathcal{A}_2$ depicted in Figure 2, for which we consider a uniform distribution on the outgoing edges from q_0. First note that every

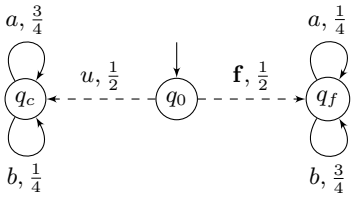


Fig. 1. An AA-diagnosable pLTS $\mathcal{A}_1$, that is neither A-diagnosable, nor uniformly AA-diagnosable.

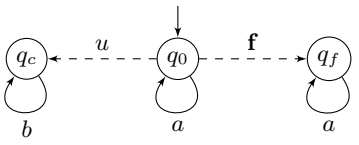


Fig. 2. A uniformly AA-diagnosable pLTS $\mathcal{A}_2$, that is not A-diagnosable.

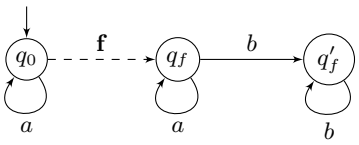


Fig. 3. An A-diagnosable pLTS $\mathcal{A}_3$.

faulty run $(q_0a)^iq_0\mathbf{f}(q_fa)^jq_f$ has a correct run, namely $q_0(aq_0)^{i+j}$ with the same observed sequence. So $\mathcal{A}_2$ is not A-diagnosable. Yet, we argue that it is uniformly AA-diagnosable. The correctness proportion of a faulty run (exponentially) decreases with respect to its length. So the worst run to be considered for the diagnoser is $q_0\mathbf{f}q_faq_f$ implying uniformity.

Consider the pLTS $\mathcal{A}_3$ from Figure 3. Viewed as an LTS, it is not diagnosable, since the observed sequence a^ω is ambiguous and forbids the diagnosis of faulty runs without any occurrence of b. On the contrary, let $\rho = q_0(aq_0)^x\mathbf{f}q_f(aq_f)^y(bq'_f)^z$ be an arbitrary faulty run. If $z > 0$ then $\mathsf{CorP}(\pi(\rho)) = 0$. Otherwise $\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > 0\}) = \frac{1}{2^n}\mathbb{P}(\rho)$ and so $\mathcal{A}_3$ is A-diagnosable.

The next proposition establishes the exact relations between the different specifications. Observe that uniform AA-diagnosability is strictly stronger than AA-diagnosability while A-diagnosability and uniform A-diagnosability are equivalent.

Proposition 1.

- *A pLTS is A-diagnosable if and only if it is uniformly A-diagnosable.*
- *There exists an AA-diagnosable pLTS, not uniformly AA-diagnosable.*
- *There exists a uniformly AA-diagnosable pLTS, not A-diagnosable.*

We have not explicitely defined diagnosers for diagnosable pLTS, as a simple diagnoser could be defined as follows: monitoring the sequence of observed events σ, the diagnoser simply computes the current correctness proportion and outputs "faulty" if $\mathsf{CorP}(\sigma)$ is below the threshold.

However such a diagnoser may need an infinite memory. It is well-known that A-diagnosable pLTS admit finite-memory A-diagnosers [11]. This is no longer the case for AA-diagnosability.

Proposition 2. *There exists an AA-diagnosable pLTS that admits no finite memory diagnoser for any threshold* $0 < \varepsilon \leq \frac{1}{2}$.

Proof. Consider $\mathcal{A}_1$ the AA-diagnosable pLTS of Figure 1 and assume there exists a diagnoser with m states for some threshold $0 < \varepsilon \leq \frac{1}{2}$. After any sequence a^n, it cannot claim a fault. So there exist $1 \leq i < j \leq m+1$ such that the diagnoser is in the same state after observing a^i and a^j. Consider the faulty run $\rho = q_0 \mathbf{f} q_f (aq_f)^i$. Due to the reactivity requirement, there must be a run $\rho\rho'$ for which the diagnoser claims a fault. This implies that for all n, the diagnoser claims a fault after $\rho_n = \rho(aq_f)^{n(j-i)}\rho'$ but $\lim_{n\to\infty} \mathsf{CorP}(\pi(\rho_n)) = 1$, which contradicts the correctness requirement. □

3 Analysis of approximate diagnosability

A-diagnosability was proved to be a PSPACE-complete problem [1]. We now focus on the other notions of approximate diagnosability introduced in Definition 2, and study their decidability and complexity.

Reducing the emptiness problem for probabilistic automata [8] (PA), we obtain the following first result:

Theorem 1. *For any rational* $0 < \varepsilon < 1$*, the* ε*-diagnosability and uniform* ε*-diagnosability problems are undecidable for pLTS.*

We now turn to the decidability status of AA-diagnosability and uniform AA-diagnosability. We prove that AA-diagnosability can be solved in polynomial time by establishing a characterization in terms of distance on labelled Markov chains; this constitutes the most technical contribution of this section.

A *labelled Markov chain* (LMC) is a pLTS where every event is observable: $\Sigma = \Sigma_o$. In order to exploit results of [5] on LMC in our context of pLTS, we introduce the mapping $\mathcal{M}$ that performs (in polynomial time) the probabilistic closure of a pLTS w.r.t. the unobservable events and produces an LMC. For sake of simplicity, we denote by $\mathcal{A}_q$, the pLTS $\mathcal{A}$ where the initial state has been substituted by q.

Definition 3. *Given a pLTS* $\mathcal{A} = \langle Q, q_0, \Sigma, T, \mathbf{P}\rangle$ *with* $\Sigma = \Sigma_o \uplus \Sigma_u$*, the labelled Markov chain* $\mathcal{M}(\mathcal{A}) = \langle Q, q_0, \Sigma_o, T', \mathbf{P}'\rangle$ *is defined by:*

– $T' = \{(q, a, q') \mid \exists \rho \in \mathsf{SR}_1(\mathcal{A}_q)\ \rho = q \cdots aq'\}$.
– *For all* $(q, a, q') \in T', \mathbf{P}'(q, a, q') = \mathbb{P}\{\rho \in \mathsf{SR}_1(\mathcal{A}_q) \mid \rho = q \cdots aq'\}$.

Let E be an *event* of Σ^ω (*i.e.* a measurable subset of Σ^ω for the standard measure), we denote by $\mathbb{P}^{\mathcal{M}}(E)$ the probability that event E occurs in the LMC $\mathcal{M}$. Given two LMC $\mathcal{M}_1$ and $\mathcal{M}_2$, the (probabilistic) distance between $\mathcal{M}_1$ and $\mathcal{M}_2$ generalizes the concept of distance for distributions. Given an event E, $|\mathbb{P}^{\mathcal{M}_1}(E) - \mathbb{P}^{\mathcal{M}_2}(E)|$ expresses the absolute difference between the probabilities that E occurs in $\mathcal{M}_1$ and in $\mathcal{M}_1$. The distance is obtained by getting the supremum over the events.

Definition 4. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two LMC over the same alphabet Σ. Then $d(\mathcal{M}_1, \mathcal{M}_2)$ the* distance between $\mathcal{M}_1$ and $\mathcal{M}_2$ *is defined by:*

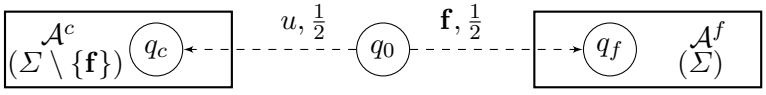
$$d(\mathcal{M}_1, \mathcal{M}_2) = \sup(\mathbb{P}^{\mathcal{M}_1}(E) - \mathbb{P}^{\mathcal{M}_2}(E) \mid E \text{ event of } \Sigma^\omega).$$

The *distance* 1 *problem* asks, given labelled Markov chains $\mathcal{M}_1$ and $\mathcal{M}_2$, whether $d(\mathcal{M}_1, \mathcal{M}_2) = 1$. We summarize in the next proposition, the results by Chen and Kiefer on LMC that we use later.

Proposition 3 ([5]).

– *Given two LMC $\mathcal{M}_1, \mathcal{M}_2$, there exists an event E such that:*
$$d(\mathcal{M}_1, \mathcal{M}_2) = \mathbb{P}^{\mathcal{M}_1}(E) - \mathbb{P}^{\mathcal{M}_2}(E).$$
– *The distance 1 problem is decidable in polynomial time for LMC.*

Towards the decidability of AA-diagnosability, let us first explain how to solve the problem on a subclass of pLTS called *initial-fault pLTS*. Informally, an initial-fault pLTS $\mathcal{A}$ consists of two disjoint pLTS $\mathcal{A}^f$ and $\mathcal{A}^c$ and an initial state q_0 with an outgoing unobservable correct transition leading to $\mathcal{A}^c$ and a transition labelled by $\mathbf{f}$ leading to $\mathcal{A}^f$ (see the figure below). Moreover no faulty transitions occur in $\mathcal{A}^c$. We denote such a pLTS by $\mathcal{A} = \langle q_0, \mathcal{A}^f, \mathcal{A}^c \rangle$; the formal definition is in Appendix A.3.



The next lemma establishes a strong connection between distance of LMC and diagnosability of initial-fault pLTS.

Lemma 1. *Let $\mathcal{A} = \langle q_0, \mathcal{A}^f, \mathcal{A}^c \rangle$ be an initial-fault pLTS. Then $\mathcal{A}$ is AA-diagnosable if and only if $d(\mathcal{M}(\mathcal{A}^f), \mathcal{M}(\mathcal{A}^c)) = 1$.*

Let us sketch the proof, which is detailed in Appendix A.3. Assuming $\mathcal{A}$ is AA-diagnosable, we pick an arbitrary ε and apply ε-diagnosability to the faulty run $q_0\mathbf{f}q_f$: we choose for the convergence threshold some $\alpha = O(\varepsilon)$ and consider a random observed sequence of length $n_{\rho,\alpha}$. Then we prove that the event E_ε, expressing that the correctness proportion of such a random observed sequence does not exceed ε, fulfills: $\mathbb{P}^{\mathcal{M}(\mathcal{A}^f)}(E_\varepsilon) \geq 1 - O(\varepsilon)$ and $\mathbb{P}^{\mathcal{M}(\mathcal{A}^c)}(E_\varepsilon) \leq O(\varepsilon)$. Therefore, $d(\mathcal{M}(\mathcal{A}^f), \mathcal{M}(\mathcal{A}^c)) = 1$.
Assuming now that $d(\mathcal{M}(\mathcal{A}^f), \mathcal{M}(\mathcal{A}^c)) = 1$, we let E be an event such that $|\mathbb{P}^{\mathcal{M}(\mathcal{A}^f)}(E) - \mathbb{P}^{\mathcal{M}(\mathcal{A}^c)}(E)| = 1$. We pick an arbitrary threshold $\varepsilon > 0$ and proceed in two steps. First we consider for $n \in \mathbb{N}$, a random observed sequence of length n, that can be extended to an infinite sequence in E triggered by a faulty run, and have a correctness proportion greater than ε. By a probabilistic reasoning, the probability of such random sequences converges to 0 when n goes to infinity. Second, we consider a faulty run ρ_f and a convergence threshold α. Note that the observed sequences of *all* finite faulty runs that extend ρ_f can be extended into infinite sequences in E. Then applying the first result on convergence, we choose an appropriate n_α and show that $\mathcal{A}$ is AA-diagnosable.

In order to understand why characterizing AA-diagnosability for general pLTS is more involved, let us study the pLTS $\mathcal{A}_2$ presented in Figure 2 where outgoing transitions of any state are equidistributed. Recall that $\mathcal{A}_2$ is AA-diagnosable (and even uniformly AA-diagnosable).
Let us look at the distance between pairs of a correct and a faulty states of $\mathcal{A}$ that can be reached by runs with the same observed sequence. On the one hand, $d(\mathcal{M}(\mathcal{A}_{q_0}), \mathcal{M}(\mathcal{A}_{q_f})) \leq \frac{1}{2}$ since for any event E either (1) $a^\omega \in E$ implying $\mathbb{P}^{\mathcal{M}(\mathcal{A}_{q_f})}(E) = 1$ and $\mathbb{P}^{\mathcal{M}(\mathcal{A}_{q_0})}(E) \geq \frac{1}{2}$ or (2) $a^\omega \notin E$ implying $\mathbb{P}^{\mathcal{M}(\mathcal{A}_{q_f})}(E) = 0$ and $\mathbb{P}^{\mathcal{M}(\mathcal{A}_{q_0})}(E) \leq \frac{1}{2}$. On the other hand, $d(\mathcal{M}(\mathcal{A}_{q_c}), \mathcal{M}(\mathcal{A}_{q_f})) = 1$ since $\mathbb{P}^{\mathcal{M}(\mathcal{A}_{q_f})}(a^\omega) = 1$ and $\mathbb{P}^{\mathcal{M}(\mathcal{A}_{q_c})}(a^\omega) = 0$.
We claim that the former pair is irrelevant, since the correct state q_0 does not belong to a bottom strongly connected component (BSCC) of the pLTS, while the latter one is relevant since q_c belongs to a BSCC.

The next theorem characterizes AA-diagnosability, establishing the soundness of this intuition. Moreover, it states the complexity of deciding AA-diagnosability.

Theorem 2. *Let $\mathcal{A}$ be a pLTS. Then, $\mathcal{A}$ is AA-diagnosable if and only if for every $q_c \in Q_c$ belonging to a BSCC and $q_f \in Q_f$ reachable by runs with the same observed sequence, $d(\mathcal{M}(\mathcal{A}_{q_c}), \mathcal{M}(\mathcal{A}_{q_f})) = 1$.*
The AA-diagnosability problem is decidable in polynomial time for pLTS.

The full proof of Theorem 2 is given in Appendix A.4. Let us sketch the key ideas to establish the characterization of AA-diagnosability in terms of the distance 1 problem.

The left-to-right implication is the easiest one, and is proved by contraposition. Assume there exist two states in $\mathcal{A}$, $q_c \in Q_c$ belonging to a BSCC and $q_f \in Q_f$ reachable resp. by ρ_c and ρ_f with $\pi(\rho_c) = \pi(\rho_f)$, and with $d(\mathcal{M}(\mathcal{A}_{q_c}), \mathcal{M}(\mathcal{A}_{q_f})) < 1$. Applying Lemma 1 to the initial-fault pLTS $\mathcal{A}' = \langle q'_0, \mathcal{A}_{q_f}, \mathcal{A}_{q_c} \rangle$, one deduces that $\mathcal{A}'$ is not AA-diagnosable. First we relate the probabilities of runs in $\mathcal{A}$ and $\mathcal{A}'$. Then we show that considering the additional faulty runs with same observed sequence as ρ_f does not make $\mathcal{A}$ AA-diagnosable.

The right-to-left implication is harder to establish. For ρ_0 a faulty run, $\alpha > 0, \varepsilon > 0$, $\sigma_0 = \pi(\rho_0)$ and $n_0 = |\sigma_0|$, we start by extending the runs with observed sequences σ_0 by n_b observable events where n_b is chosen in order to get a high probability that the runs end in a BSCC. For such an observed sequence $\sigma \in \Sigma_o^{n_b}$, we partition the possible runs with observed sequence $\sigma_0\sigma$ into three sets: $\mathfrak{R}_\sigma^F$ is the subset of faulty runs; $\mathfrak{R}_\sigma^C$ (resp. $\mathfrak{R}_\sigma^T$) is the set of correct runs ending (resp. not ending) in a BSCC. At first, we do not take into account the "transient" runs in $\mathfrak{R}_\sigma^T$. We apply Lemma 1 to obtain an integer n_σ such that from $\mathfrak{R}_\sigma^F$ and $\mathfrak{R}_\sigma^C$ we can diagnose with (appropriate) high probability and low correctness proportion after n_σ observations. Among the runs that trigger diagnosable observed sequences, some exceed the correctness proportion ε, when taking into account the runs from $\mathfrak{R}_\sigma^T$. Yet, we show that the probability of such runs is small, when cumulated over all extensions σ, leading to the required upper bound α.

Using the characterization, one can easily establish the complexity of AA-diagnosability. Indeed, reachability of a pair of states with the same observed sequence is decidable in polynomial time by an appropriate "self-synchronized product" of the pLTS. Since there are at most a quadratic number of pairs to check, and given that the distance 1 problem can be decided in polynomial time, the PTIME upper-bound follows.

In constrast, uniform AA-diagnosability is shown to be undecidable by a reduction from the emptiness problem for probabilistic automata, that is more involved than the one for Theorem 1.

Theorem 3. *The uniform AA-diagnosability problem is undecidable for pLTS.*

4 Conclusion

This paper completes our previous work [1] on diagnosability of stochastic systems, by giving here a full picture on approximate diagnosis. On the one hand, we performed a semantical study: we have refined the reactivity specification by introducing a uniform requirement about detection delay w.r.t. faults and studied its impact on both the exact and approximate case. On the other hand, we established decidability and complexity of all notions of approximate diagnosis: we have shown that (uniform) ε-diagnosability and uniform AA-diagnosability are undecidable while AA-diagnosability can be solved in polynomial time.

There are still interesting issues to be tackled, to continue our work on monitoring of stochastic systems. For example, prediction and prediagnosis, which are closely related to diagnosis and were analyzed in the exact case in [1], should be studied in the approximate framework.

References

1. N. Bertrand, S. Haddad, and E. Lefaucheux. Foundation of diagnosis and predictability in probabilistic systems. In *Proceedings of FSTTCS 2014*, volume 29 of *LIPIcs*, pages 417–429. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2014.
2. M.P. Cabasino, A. Giua, S. Lafortune, and C. Seatzu. Diagnosability analysis of unbounded Petri nets. In *Proceedings of CDC 2009*, pages 1267–1272. IEEE, 2009.
3. F. Cassez and S. Tripakis. Fault diagnosis with static and dynamic observers. *Fundamenta Informaticae*, 88:497–540, 2008.
4. E. Chanthery and Y. Pencolé. Monitoring and active diagnosis for discrete-event systems. In *Proceedings of SP 2009*, pages 1545–1550. Elsevier, 2009.
5. T. Chen and S. Kiefer. On the total variation distance of labelled Markov chains. In *Proceedings of CSL-LICS 2014*, pages 33:1–33:10. ACM, 2014.
6. 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.
7. C. Morvan and S. Pinchinat. Diagnosability of pushdown systems. In *Proceedings of HVC 2009*, volume 6405 of *LNCS*, pages 21–33. Springer, 2009.
8. A. Paz. *Introduction to Probabilistic Automata*. Academic Press, 1971.
9. M. Sampath, S. Lafortune, and D. Teneketzis. Active diagnosis of discrete-event systems. *IEEE Transactions on Automatic Control*, 43(7):908–929, 1998.
10. 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.
11. D. Thorsley and D. Teneketzis. Diagnosability of stochastic discrete-event systems. *IEEE Transactions on Automatic Control*, 50(4):476–492, 2005.
12. D. Thorsley and D. Teneketzis. Active acquisition of information for diagnosis and supervisory control of discrete-event systems. *Journal of Discrete Event Dynamic Systems*, 17:531–583, 2007.

A Appendix

This appendix contains proofs that are omitted in the core of the paper.

A.1 Proof of Proposition 1

Proposition 1.

- *A pLTS is A-diagnosable if and only if it is uniformly A-diagnosable.*
- *There exists an AA-diagnosable pLTS, not uniformly AA-diagnosable.*
- *There exists a uniformly AA-diagnosable pLTS, not A-diagnosable.*

Proof. The first item states the equivalence of uniform A-diagnosability and A-diagnosability. We first recall the two notions. A pLTS $\mathcal{A}$ is

- *A-diagnosable* if for every $\alpha > 0$ and every faulty run $\rho \in \mathsf{F}$ there exists $n_{\rho,\alpha} \in \mathbb{N}$ such that for every $n \geq n_{\rho,\alpha}$

$$\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > 0\}) \leq \alpha.$$

- *uniformly A-diagnosable* if for every $\alpha > 0$ there exists $n_\alpha \in \mathbb{N}$ such that for every $n \geq n_\alpha$ and every faulty run $\rho \in \mathsf{F}$

$$\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > 0\}) \leq \alpha \cdot \mathbb{P}(\rho)$$

In order to establish the equivalence of the two notions, we recall a characterization of A-diagnosability, established in [1]. From a pLTS $\mathcal{A} = \langle Q, q_0, \Sigma, T, \mathbf{P}\rangle$, we contruct $\mathcal{B}l(\mathcal{A})$, its *belief automaton*, a complete automaton that tracks the subset of states reached by a correct signalling run associated with a given observed sequence. The sets S of states and $\Delta_{\mathcal{B}l}$ of transitions of $\mathcal{B}l(\mathcal{A})$ are inductively defined by:

- $s_0 = \{q_0\}$ is the initial state of $\mathcal{B}l(\mathcal{A})$;
- for every U state of $\mathcal{B}l(\mathcal{A})$ and every $a \in \Sigma_o$, letting

$$U' = \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \in \mathsf{SR}(\mathcal{A}), \\ q_{\alpha_0} \in U, \forall i < k \ a_i \in \Sigma_u \setminus \{\mathbf{f}\}, a_k = a, q_{\alpha_k} = q\},$$

  there exists a transition $U \xrightarrow{a} U'$ in $\mathcal{B}l(\mathcal{A})$.

$\mathcal{B}l(\mathcal{A})$ consists of a deterministic version of the correct behaviours of $\mathcal{A}$. From $\mathcal{A}$ and $\mathcal{B}l(\mathcal{A})$, one can build the product pLTS $\mathcal{A}_{\mathcal{B}l} = \mathcal{A} \times \mathcal{B}l(\mathcal{A})$ by synchronization on observable events. Since $\mathcal{B}l(\mathcal{A})$ is deterministic and complete, $\mathcal{A}_{\mathcal{B}l}$ is still a pLTS, with same stochastic behaviour as $\mathcal{A}$. The following characterization of A-diagnosability of $\mathcal{A}$ considers bottom strongly connected components (BSCC) of $\mathcal{A}_{\mathcal{B}l}$ and was proved in [1]:

Theorem A (Characterization of A-diagnosability [1]). *Let $\mathcal{A}$ be a finite pLTS. Then $\mathcal{A}$ is A-diagnosable if and only if $\mathcal{A}_{\mathcal{B}l}$ has no BSCC containing a state (q, U) with $q \in Q_f$ and $U \neq \emptyset$.*

• Let us show that A-diagnosability implies uniform A-diagnosability. Let $\mathcal{A}$ be a pLTS, and assume it is A-diagnosable. Given a run ρ of $\mathcal{A}$, we write $\rho_{\mathcal{B}l}$ for its associated run in $\mathcal{A}_{\mathcal{B}l}$: $\rho_{\mathcal{B}l}$ extends the states of ρ by subsets of possible correct states after the prefixes of the observed sequence $\pi(\rho)$. We let S_{BSCC} denote the set of states of $\mathcal{A}_{\mathcal{B}l}$ that belong to a BSCC. Last, for every state (q, U) of $\mathcal{A}_{\mathcal{B}l}$ and every $n \in \mathbb{N}$, we denote by $\mathsf{SR}_n^{q,U}$ the set of signalling runs of length n starting in (q, U).
Let $\alpha > 0$. Our objective is to define n_α such that for every $n \geq n_\alpha$ and every faulty run $\rho \in \mathsf{F}$:

$$\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho)) > 0\}) \leq \alpha \cdot \mathbb{P}(\rho).$$

We first exploit the transience of states of $\mathcal{A}_{\mathcal{B}l}$ that do not belong to a BSCC, and the almost sure convergence towards the BSCC. For every reachable state (q, U) of $\mathcal{A}_{\mathcal{B}l}$, there exists $n_{q,U} \in \mathbb{N}$ such that the measure of the runs that start in (q, U) and do not reach a BSCC within $n_{q,U}$ steps is smaller than α. Formally, for every (q, U), there exists $n_{q,U} \in \mathbb{N}$ such that for every $n \geq n_{q,U}$, $\mathbb{P}(\{\rho'_{\mathcal{B}l} \in \mathsf{SR}_n^{q,U} \mid \mathsf{last}(\rho'_{\mathcal{B}l}) \notin S_{\mathsf{BSCC}}\}) \leq \alpha$. We let n_α be the maximum value $n_{q,U}$ over states (q, U) of $\mathcal{A}_{\mathcal{B}l}$.
Now let us consider a faulty run $\rho \in \mathsf{F}$ of $\mathcal{A}$, and let $(q, U) = \mathsf{last}(\rho_{\mathcal{B}l})$. Since $n_\alpha \geq n_{q,U}$, we deduce that $\mathbb{P}(\{\rho'_{\mathcal{B}l} \in \mathsf{SR}_{n_\alpha}^{q,U} \mid \mathsf{last}(\rho'_{\mathcal{B}l}) \notin S_{\mathsf{BSCC}}\}) \leq \alpha$. Therefore, in $\mathcal{A}$

$$\mathbb{P}(\{\rho' \in \mathsf{SR}_{n_\alpha+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{last}(\rho'_{\mathcal{B}l}) \notin S_{\mathsf{BSCC}}\}) \leq \alpha \cdot \mathbb{P}(\rho).$$

Thanks to the characterization recalled in Theorem A, the only BSCCs reachable from (q, U) in $\mathcal{A}_{\mathcal{B}l}$ necessarily have an empty second component. This means that once a run $\rho'_{\mathcal{B}l}$ reaches such a BSCC, $\rho'_{\mathcal{B}l}$ admits no correct run with same observed sequence, and hence $\mathsf{CorP}(\pi(\rho'_{\mathcal{B}l})) = 0$. Equivalently, $\mathsf{CorP}(\pi(\rho')) > 0$ implies $\mathsf{last}(\rho'_{\mathcal{B}l}) \notin S_{\mathsf{BSCC}}$. Thus

$$\mathbb{P}(\{\rho' \in \mathsf{SR}_{n_\alpha+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > 0\}) \leq \alpha \cdot \mathbb{P}(\rho)$$

which shows that $\mathcal{A}$ is uniformly A-diagnosable.

• Consider the pLTS of Figure 1. Observe that for a random run ρ starting from q_c (resp. q_f), by the strong law of large numbers, the following holds almost surely $lim_{n\to\infty} = \frac{|\pi(\rho)_{\downarrow n}|_a}{n} = \frac{3}{4}$ (resp. $\frac{1}{4}$) where $|\sigma|_a$ is the number of occurrences of a in σ. So the event $\limsup_{n\to\infty} = \frac{|\pi(\rho)_{\downarrow n}|_a}{n} > \frac{1}{2}$ has

probability 1 (resp. 0) for a run ρ starting from q_c (resp. q_f). Applying Lemma 1, this pLTS is AA-diagnosable.

Pick some arbitrary n and consider the faulty run $\rho = q_0\mathbf{f}(q_f a)^{n+1}q_f$. For all signalling run ρ' extending ρ, of observable length $2n+1$, $\mathsf{CorP}(\pi(\rho')) > \frac{1}{2}$ implying $\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > \frac{1}{2}\}) = \mathbb{P}(\rho)$. Thus pLTS of Figure 1 is not uniformly $\frac{1}{2}$-diagnosable.

• Consider the pLTS of Figure 2. and an arbitrary faulty signalling run $\rho = q_0\mathbf{f}q_f(aq_f)^{n+1}$. For every faulty run $\rho' = q_0\mathbf{f}q_f(aq_f)^m$ with $m > n$, its observed sequence $\pi(\rho') = a^m$ and $\mathsf{CorP}(a^m) = \frac{2}{3^m+1}$. So given $\varepsilon > 0$, let us define n_ε such that $\frac{2}{3^{n_\varepsilon}+1} \leq \varepsilon$. Then for all $n \geq n_\varepsilon$, $\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > \varepsilon\}) = 0$. Thus the pLTS of Figure 2 is uniformly AA-diagnosable.

On the contrary, since $\mathsf{CorP}(a^m) > 0$ for all m. For all n, $\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > 0\}) = \mathbb{P}(\rho)$. Thus the pLTS of Figure 2 is not A-diagnosable. □

A.2 Undecidability of (uniform) ε-diagnosability

A probabilistic automaton (PA) $\mathcal{A}$ is defined by an alphabet Σ, a set of states Q including an initial state q_0 and a subset of final states F, and for all $a \in \Sigma$ a stochastic matrix, $\mathbf{P}_a$, indexed by $Q \times Q$. When $\mathbf{P}_a[q, q'] > 0$, there is a transition from q to q' labelled by a and $\mathbf{P}_a[q, q']$. Given a word $w = a_1 \dots a_n \in \Sigma^*$, the acceptance probability of w, $\mathbf{Pr}_{\mathcal{A}}(w)$ is defined by $\mathbf{Pr}_{\mathcal{A}}(w) = \sum_{q \in F} \mathbf{P}_w[q_0, q]$ where $\mathbf{P}_w = \mathbf{P}_{a_1} \cdots \mathbf{P}_{a_n}$. Given a rational threshold $0 < \varepsilon < 1$, the language $\mathcal{L}_{\mathcal{A},\varepsilon}$ is defined by $\mathcal{L}_{\mathcal{A},\varepsilon} = \{w \in \Sigma^* \mid \mathbf{Pr}_{\mathcal{A}}(w) > \varepsilon\}$. Given a probabilistic automaton $\mathcal{A}$ and a threshold ε, the emptiness problem asks whether $\mathcal{L}_{\mathcal{A},\varepsilon} = \emptyset$. This problem is undecidable even for a fixed ε and when considering automata such that there is no word w with $\mathbf{Pr}_{\mathcal{A}}(w) = 1$ [8]. We are now in position to prove the following undecidability result.

Theorem 1. *For any rational $0 < \varepsilon < 1$, the ε-diagnosability and uniform ε-diagnosability problems are undecidable for pLTS.*

Proof. The proof is by reduction from the emptiness problem for probabilistic automata. This reduction will establish the result for the two versions of ε-diagnosability. Let $\mathcal{A}$ be a probabilistic automaton. Define the pLTS $\mathcal{A}' = \langle Q', q_0, \Sigma', T', \mathbf{P}' \rangle$ as follows.

- $\Sigma' = \Sigma \uplus \{\sharp, \mathbf{f}\}$, $\Sigma'_{uo} = \{\mathbf{f}\}$;
- $Q' = Q \cup \{q_c^\sharp, q_f^\sharp, f^\sharp\}$;

– $T' = \{(q, a, q) \mid q, q' \in Q, a \in \Sigma, \mathbf{P}_a[q, q'] > 0\}$
$\cup \{(q, \sharp, q_c^\sharp \mid q \in F\} \cup \{(q, \sharp, q_f^\sharp \mid q \in Q \setminus F\}$
$\cup \{q_c^\sharp, \sharp, q_c^\sharp\} \cup \{q_f^\sharp, \mathbf{f}, f^\sharp\} \cup \{f^\sharp, \sharp, f^\sharp\}$
– $\mathbf{P}'$ is defined by:
  - For all $q \in Q$ and $a \in \Sigma$, $\mathbf{P}'(q, a, q') = \frac{\mathbf{P}_a[q,q']}{1+|\Sigma|}$;
  - For all $q \in F$, $\mathbf{P}'(q, \sharp, q_c^\sharp) = \frac{1}{1+|\Sigma|}$;
  - For all $q \in Q \setminus F$, $\mathbf{P}'(q, \sharp, q_f^\sharp) = \frac{1}{1+|\Sigma|}$;
  - $\mathbf{P}'(q_f^\sharp, \mathbf{f}, f^\sharp) = \mathbf{P}'(f^\sharp, \sharp, f^\sharp) = \mathbf{P}'(q_c^\sharp, \sharp, q_c^\sharp) = 1$.

This reduction is represented in Figure 4.

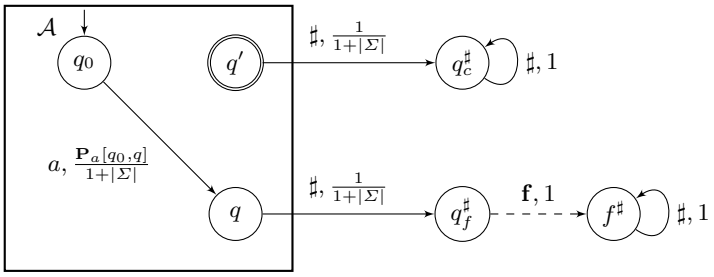


Fig. 4. From probabilistic automata to pLTS.

We claim that $\mathcal{A}'$ is ε-diagnosable and uniformly ε-diagnosable if and only if $\mathcal{L}_{\mathcal{A},\varepsilon} = \emptyset$.

• Assume that there exists a word $w \in \Sigma^*$ such that $\mathbf{Pr}_{\mathcal{A}}(w) > \varepsilon$. Consider the set of signalling correct runs with observed sequence $w\sharp^{n+2}$. By construction, its probability is $\frac{\mathbf{Pr}_{\mathcal{A}}(w)}{(1+|\Sigma|)^{|w|+1}}$. Consider the set of signalling faulty runs with observed sequence $w\sharp^{n+2}$. By construction, its probability is $\frac{1-\mathbf{Pr}_{\mathcal{A}}(w)}{(1+|\Sigma|)^{|w|+1}}$. By hypothesis on $\mathcal{A}$, $\mathbf{Pr}_{\mathcal{A}}(w) < 1$. So this set of faulty runs is non empty. Let us choose some faulty run ρ with observed sequence $w\sharp\sharp$. Using our preliminary remarks, for all n:

$$\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > \varepsilon\}) = \mathbb{P}(\rho)$$

Thus $\mathcal{A}'$ is not ε-diagnosable.

• Assume that for all word $w \in \Sigma^*$, $\mathbf{Pr}_{\mathcal{A}}(w) \leq \varepsilon$. Let ρ be a faulty run. By construction of $\mathcal{A}'$, its observed sequence is $w\sharp^{n+2}$ with $w \in \Sigma^*$. Using the same reasoning as before, for all $\rho \preceq \rho'$:

$$\mathsf{CorP}(\pi(\rho')) = \mathbf{Pr}_{\mathcal{A}}(w) \leq \varepsilon$$

Thus for any $\alpha > 0$, choosing $n_\alpha = 0$, one gets:

$$\mathbb{P}(\{\rho' \in \mathsf{SR}_{n_\alpha+|\rho|} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > \varepsilon\}) = 0$$

So $\mathcal{A}'$ is uniformly ε-diagnosable. □

A.3 Decidability of AA-diagnosability for initial-faut pLTS

Definition A (Initial-fault pLTS). *A pLTS $\mathcal{A} = \langle Q, q_0, \Sigma, T, \mathbf{P}\rangle$ is an* initial fault pLTS *if there exist two disjoint pLTS $\mathcal{A}^f = \langle Q_f, q_f, \Sigma, T_f, \mathbf{P}_f\rangle$ and $\mathcal{A}^c = \langle Q_c, q_c, \Sigma \setminus \{\mathbf{f}\}, T_c, \mathbf{P}_c\rangle$ such that:*

- $Q = \{q_0\} \uplus Q_f \uplus Q_c$;
- $T = T_f \uplus T_c \uplus \{(q_0, u, q_c), (q_0, \mathbf{f}, q_f)\}$ *with* $u \in \Sigma_u$;
- *for all* $t \in T_f$, $\mathbf{P}(t) = \mathbf{P}_f(t)$, *for all* $t \in T_c$, $\mathbf{P}(t) = \mathbf{P}_c(t)$ *for all* $t \in T \setminus (T_c \cup T_f)$ $\mathbf{P}(t) = \frac{1}{2}$.

Lemma 1. *Let $\mathcal{A} = \langle q_0, \mathcal{A}^f, \mathcal{A}^c\rangle$ be an initial-fault pLTS. Then $\mathcal{A}$ is AA-diagnosable if and only if $d(\mathcal{M}(\mathcal{A}^f), \mathcal{M}(\mathcal{A}^c)) = 1$.*

Proof. We write $\mathbb{P}$, $\mathbb{P}_f$ and $\mathbb{P}_c$ for the probability distributions of pLTS $\mathcal{A}$, $\mathcal{A}^f$ and $\mathcal{A}^c$. By construction of $\mathcal{M}(\mathcal{A}^f)$ and $\mathcal{M}(\mathcal{A}^c)$, for every observed sequence σ, $\mathbb{P}^{\mathcal{M}(\mathcal{A}^f)}(\sigma) = \mathbb{P}_f(\sigma)$ and similarly $\mathbb{P}^{\mathcal{M}(\mathcal{A}^c)}(\sigma) = \mathbb{P}_c(\sigma)$. In words, the mapping $\mathcal{M}$ lets unchanged the probability of occurrence of an observed sequence.

• Assume that $\mathcal{A}$ is AA-diagnosable. Then for every $\varepsilon > 0$ and every faulty run ρ, the following equation holds:

$$\lim_{n\to\infty} \mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > \varepsilon\}) = 0. \tag{1}$$

Pick some $0 < \varepsilon < 1$. By applying Equation (1) on the faulty run $\rho_f = q_0\mathbf{f}q_f$ with $|\pi(\rho_f)| = 0$, there exists some $n \in \mathbb{N}$ such that:

$$\mathbb{P}(\{\rho \in \mathsf{SR}_n \mid \rho_f \preceq \rho \wedge \mathsf{CorP}(\pi(\rho)) > \varepsilon\}) \leq \varepsilon.$$

Let $\mathfrak{S}$ be the set of observed sequences of faulty runs with length n and correctness proportion not exceeding theshold ε:

$$\mathfrak{S} = \{\sigma \in \Sigma_o^n \mid \exists \rho \in \mathsf{SR}_n, \pi(\rho) = \sigma \wedge \rho_f \preceq \rho \wedge \mathsf{CorP}(\sigma) \leq \varepsilon\}.$$

$E = Cyl(\mathfrak{S})$ is the event consisting of the infinite suffixes of those sequences. Let us show that $\mathbb{P}_c(E) \leq \frac{\varepsilon}{1-\varepsilon}$ and $\mathbb{P}_f(E) \geq 1 - 2\varepsilon$.

$$\mathbb{P}_f(E) = 1 - 2\mathbb{P}(\{\rho \in \mathsf{SR}_n \mid \rho_f \preceq \rho \wedge \mathsf{CorP}(\pi(\rho)) > \varepsilon\}) \geq 1 - 2\varepsilon.$$

Moreover, for every observed sequence $\sigma \in \mathfrak{S}$, there exists a faulty run ρ such that $\pi(\rho) = \sigma$. Thus, $\mathsf{CorP}(\sigma) \leq \varepsilon$. Using the definition of CorP:

$$\mathsf{CorP}(\sigma) = \frac{\mathbb{P}(\{\rho \in \mathsf{C}_n \mid \pi(\rho) = \sigma\})}{\mathbb{P}(\{\rho \in \mathsf{SR}_n \mid \pi(\rho) = \sigma\})} = \frac{\mathbb{P}_c(\sigma)}{\mathbb{P}_c(\sigma) + \mathbb{P}_f(\sigma)} \leq \varepsilon.$$

Thus, $\mathbb{P}_c(\sigma) \leq \frac{\varepsilon}{1-\varepsilon}\mathbb{P}_f(\sigma)$. Hence:

$$\mathbb{P}_c(E) = \sum_{\sigma \in \mathfrak{S}} \mathbb{P}_c(\sigma) \leq \sum_{\sigma \in \mathfrak{S}} \frac{\varepsilon}{1-\varepsilon}\mathbb{P}_f(\sigma) = \frac{\varepsilon}{1-\varepsilon}\mathbb{P}_f(E) \leq \frac{\varepsilon}{1-\varepsilon}.$$

Therefore $d(\mathcal{M}(\mathcal{A}^c), \mathcal{M}(\mathcal{A}^f)) \geq \mathbb{P}_f(E) - \mathbb{P}_c(E) \geq 1 - \varepsilon(2 + \frac{1}{1-\varepsilon})$. Letting ε go to 0, we obtain $d(\mathcal{M}(\mathcal{A}^c), \mathcal{M}(\mathcal{A}^f)) = 1$.

• Conversely assume that $d(\mathcal{M}(\mathcal{A}^f), \mathcal{M}(\mathcal{A}^c)) = 1$. Due to Proposition 3, there exists an event $E \subseteq \Sigma_o^\omega$ such that $\mathbb{P}_f(E) = 1$ and $\mathbb{P}_c(E) = 0$.
For all $n \in \mathbb{N}$, let $\mathfrak{S}_n$ be the set of prefixes of length n of the observed sequences of E: $\mathfrak{S}_n = \{\sigma \in \Sigma_o^n \mid \exists \sigma' \in E, \sigma \preceq \sigma'\}$.
For all $\varepsilon > 0$, let $\mathfrak{S}_n^\varepsilon$ be the subset of sequences of $\mathfrak{S}_n$ whose correctness proportion exceeds threshold ε: $\mathfrak{S}_n^\varepsilon = \{\sigma \in \mathfrak{S}_n \mid \mathsf{CorP}(\sigma) > \varepsilon\}$.
As $\bigcap_{n\in\mathbb{N}} \mathfrak{S}_n = E$, $\lim_{n\to\infty} \mathbb{P}_c(\mathfrak{S}_n) = \mathbb{P}_c(E) = 0$. So $\lim_{n\to\infty} \mathbb{P}_c(\mathfrak{S}_n^\varepsilon) = 0$.
On the other hand for all $n \in \mathbb{N}$,

$$\mathbb{P}_c(\mathfrak{S}_n^\varepsilon) = \sum_{\sigma \in \mathfrak{S}_n^\varepsilon} \mathbb{P}_c(\sigma) > \sum_{\sigma \in \mathfrak{S}_n^\varepsilon} \frac{\varepsilon}{1-\varepsilon}\mathbb{P}_f(\sigma) = \frac{\varepsilon}{1-\varepsilon}\mathbb{P}_f(\mathfrak{S}_n^\varepsilon).$$

Therefore we have $\lim_{n\to\infty} \mathbb{P}_f(\mathfrak{S}_n^\varepsilon) = 0$.

Let ρ be a faulty run and $\alpha > 0$. There exists $k \geq |\rho|_o$ such that for all $n \geq k$, $\mathbb{P}_f(\mathfrak{S}_n^\varepsilon) \leq \alpha$. Let $n \geq k$, and $\widetilde{\mathfrak{S}}_n$ be the set of observed sequences of length n triggered by a run with prefix ρ and whose correctness proportion exceeds ε:

$$\widetilde{\mathfrak{S}}_n = \{\sigma \in \Sigma_o^n \mid \exists \rho' \in \mathsf{SR}_n, \rho \preceq \rho' \wedge \pi(\rho') = \sigma \wedge \mathsf{CorP}(\sigma) > \varepsilon\}.$$

Let us prove that $\mathbb{P}(\widetilde{\mathfrak{S}}_n) \leq \alpha$.
Since $\mathbb{P}_f(\mathfrak{S}_n) \geq \mathbb{P}_f(E) = 1$, $\mathbb{P}_f(\widetilde{\mathfrak{S}}_n \cap (\Sigma_o^n \setminus \mathfrak{S}_n)) = 0$.
Since $\mathbb{P}_f(\mathfrak{S}_n^\varepsilon) < \alpha$, $\mathbb{P}_f(\widetilde{\mathfrak{S}}_n \cap \mathfrak{S}_n) \leq \mathbb{P}_f(\mathfrak{S}_n^\varepsilon) \leq \alpha$.
Thus $\mathbb{P}_f(\widetilde{\mathfrak{S}}_n) = \mathbb{P}_f(\widetilde{\mathfrak{S}}_n \cap \mathfrak{S}_n) + \mathbb{P}_f(\widetilde{\mathfrak{S}}_n \cap (\Sigma_o^n \setminus \mathfrak{S}_n)) \leq \alpha$.
Since α is arbitrary, we have proven that $\lim_{n\to\infty} \mathbb{P}_f(\widetilde{\mathfrak{S}}_n) = 0$.

Observe now that $\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > \varepsilon\}) = \frac{1}{2}\mathbb{P}_f(\widetilde{\mathfrak{S}}_n)$.
Therefore, $\lim_{n\to\infty} \mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > \varepsilon\}) = 0$.
So $\mathcal{A}$ is AA-diagnosable. □

A.4 Decidability of AA-diagnosability for general pLTS

Theorem 2. *Let $\mathcal{A}$ be a pLTS. Then, $\mathcal{A}$ is AA-diagnosable if and only if for every $q_c \in Q_c$ belonging to a BSCC and $q_f \in Q_f$ reachable by runs with the same observed sequence, $d(\mathcal{M}(\mathcal{A}_{q_c}), \mathcal{M}(\mathcal{A}_{q_f})) = 1$.*
The AA-diagnosability problem is decidable in polynomial time for pLTS.

Proof. The characterization follows from Lemmas A and B given below. Now, for what concerns complexity, reachability of a pair of states with the same observed sequence is decidable in polynomial time by an appropriate "self-synchronized product" of the pLTS. Since there are at most a quadratic number of pairs to check, and given that the distance 1 problem can be decided in polynomial time due to Proposition 3, the PTIME upper-bound follows. □

To establish the first item of Theorem 2, that is, the characterization of AA-diagnosability, we consider each implication in separate lemmas.

Lemma A. *Let $\mathcal{A}$ be a pLTS with $q_c \in Q_c$ belonging to a BSCC, $q_f \in Q_f$, $d(\mathcal{M}(\mathcal{A}_{q_f}), \mathcal{M}(\mathcal{A}_{q_c})) < 1$ and runs $q_0 \stackrel{\rho_c}{\Longrightarrow} q_c$ and $q_0 \stackrel{\rho_f}{\Longrightarrow} q_f$ with $\pi(\rho_c) = \pi(\rho_f)$. Then $\mathcal{A}$ is not AA-diagnosable.*

Proof. Let us introduce some notations.

$$\sigma_0 = \pi(\rho_f) = \pi(\rho_c),\ p_f = \mathbb{P}(\rho_f),\ p_c = \mathbb{P}(\rho_c).$$

Let p_g ($\geq p_f$) be the probability of the faulty runs with projection σ_0:

$$p_g = \mathbb{P}(\{\rho \in \mathsf{SR}_{|\sigma|} \mid \pi(\rho) = \sigma_0, \text{ and } \rho \text{ is faulty}\}).$$

For all $n \geq |\sigma|$, let $\mathfrak{S}_n$ be the set of observed sequences of length n "extending" ρ_f:

$$\mathfrak{S}_n = \{\sigma \in \Sigma_o^n \mid \exists \rho \in \mathsf{SR}_n, \rho_f \preceq \rho \wedge \pi(\rho) = \sigma\}.$$

Given $\sigma \in \mathfrak{S}_n$, we "decompose" p_f, p_c and p_g as follows.

- $p_f^\sigma = \mathbb{P}\{\rho \in \mathsf{SR}_n \mid \rho_f \preceq \rho \wedge \pi(\rho) = \sigma\}$;
- $p_c^\sigma = \mathbb{P}\{\rho \in \mathsf{SR}_n \mid \rho_c \preceq \rho \wedge \pi(\rho) = \sigma\}$;
- $p_g^\sigma = \mathbb{P}\{\rho \in \mathsf{SR}_n \mid \rho \text{ is faulty and } \pi(\rho) = \sigma\}$.

We introduce the initial-fault pLTS $\mathcal{A}' = \langle q_0', \mathcal{A}_{q_f}, \mathcal{A}_{q_c}\rangle$ well-defined as q_c membership of a BSCC implies that $\mathcal{A}_{q_c}$ does not trigger faults. Since $d(\mathcal{M}(\mathcal{A}_{q_f}), \mathcal{M}(\mathcal{A}_{q_c})) < 1$, due to Lemma 1, there exist positive reals $\alpha', \varepsilon' \leq 1$ such that for all $n_0 \in \mathbb{N}$ there exists $n \geq n_0$:

$$\mathbb{P}'\{\rho \in \mathsf{SR}_n \mid q_0' \mathbf{f} q_f \preceq \rho \wedge \mathsf{CorP}(\pi(\rho)) > \varepsilon\} > \alpha'$$

where $\mathbb{P}'$ denotes the probability for $\mathcal{A}'$.

This entails the following inequality for $\mathcal{A}$.

$$\mathbb{P}\{\rho \in \mathsf{SR}_n \mid \rho_f \preceq \rho \wedge \frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)} + \frac{p_c}{p_f} p_f^{\pi(\rho)}} > \varepsilon'\} > 2p_f\alpha'.$$

Indeed in $\mathcal{A}'$, the probability of a faulty (resp. correct) run with observed sequence $\pi(\rho)$ is $\frac{p_f^{\pi(\rho)}}{2p_f}$ (resp. $\frac{p_c^{\pi(\rho)}}{2p_c}$). Finally the $2p_f$ factor of the lower bound takes into account the fact that the probability of reaching q_f is $\frac{1}{2}$ while in $\mathcal{A}$ the probability of ρ is p_f.

Observe that $\frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)} + \frac{p_c}{p_f} p_f^{\pi(\rho)}} > \varepsilon'$ is equivalent to $\frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)} + p_f^{\pi(\rho)}} > \frac{\varepsilon' p_c}{\varepsilon' p_c + (1-\varepsilon') p_f}$.

So defining $\tilde{\varepsilon} = \frac{\varepsilon' p_c}{\varepsilon' p_c + (1-\varepsilon') p_f} \leq 1$ and $\tilde{\alpha} = 2p_f\alpha' \leq 2$, the previous inequality can be rewritten:

$$\mathbb{P}\{\rho \in \mathsf{SR}_n \mid \rho_f \preceq \rho \wedge \frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)} + p_f^{\pi(\rho)}} > \tilde{\varepsilon}\} > \tilde{\alpha}.$$

Let $\mathfrak{S}'_n$ be the subset of observed sequences of $\mathfrak{S}_n$ whose correctness proportion is greater than $\tilde{\varepsilon}$ when only considering extensions of ρ_f, but smaller than $\varepsilon^* = \frac{\tilde{\alpha}\tilde{\varepsilon}}{4}$ when considering all faulty runs:

$$\mathfrak{S}'_n = \{\sigma \in \mathfrak{S}_n \mid \frac{p_c^\sigma}{p_c^\sigma + p_f^\sigma} > \tilde{\varepsilon} \wedge \frac{p_c^\sigma}{p_c^\sigma + p_g^\sigma} \leq \varepsilon^*\}.$$

Let $\sigma \in \mathfrak{S}'_n$, $p_f^\sigma < \frac{1-\tilde{\varepsilon}}{\tilde{\varepsilon}} p_c^\sigma$ and $p_c^\sigma \leq \frac{\varepsilon^*}{1-\varepsilon^*} p_g^\sigma$. Therefore $p_f^\sigma < \frac{(1-\tilde{\varepsilon})\varepsilon^*}{(1-\varepsilon^*)\tilde{\varepsilon}} p_g^\sigma$.

Summing over all sequences of $\mathfrak{S}'_n$: $\sum_{\sigma \in \mathfrak{S}'_n} p_f^\sigma < \frac{(1-\tilde{\varepsilon})\varepsilon^*}{(1-\varepsilon^*)\tilde{\varepsilon}} p_g$.

Since $p_g \leq 1$: $\sum_{\sigma \in \mathfrak{S}'_n} p_f^\sigma \leq \frac{(1-\tilde{\varepsilon})\tilde{\alpha}}{4(1-\frac{\tilde{\alpha}\tilde{\varepsilon}}{4})} \leq \frac{\tilde{\alpha}}{2}$.

Thus,

$$\mathbb{P}\{\rho \in \mathsf{SR}_n \mid \rho_f \preceq \rho \wedge \frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)} + p_g^{\pi(\rho)}} > \varepsilon^*\} \geq$$
$$\mathbb{P}\{\rho \in \mathsf{SR}_n \mid \rho_f \preceq \rho \wedge \frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)} + p_f^{\pi(\rho)}} > \tilde{\varepsilon}\} - \sum_{\sigma' \in \mathfrak{S}'_n} p_f^{\sigma'} > \tilde{\alpha} - \frac{\tilde{\alpha}}{2} = \frac{\tilde{\alpha}}{2}.$$

Observe that given $\sigma \in \mathfrak{S}_n$, $\mathsf{CorP}(\sigma) \geq \frac{p_c^\sigma}{p_c^{\sigma'} + p_g^\sigma}$ since we ignore correct runs ρ with $\pi(\rho) = \sigma$ that do no extend ρ_c. So defining $\varepsilon = \varepsilon^*$ and $\alpha = \frac{\tilde{\alpha}}{2}$, for all $n_0 \in \mathbb{N}$ there exists $n \geq n_0$:

$$\mathbb{P}\{\rho \in \mathsf{SR}_n \mid \rho_f \preceq \rho \wedge \frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)} + p_g^{\pi(\rho)}} > \varepsilon\} > \alpha$$

which establishes that $\mathcal{A}$ is not AA-diagnosable. □

Lemma B. *Let $\mathcal{A}$ be a pLTS such that for all $q_0 \overset{\rho_c}{\Longrightarrow} q_c$ and $q_0 \overset{\rho_f}{\Longrightarrow} q_f$ with $\pi(\rho_c) = \pi(\rho_f)$, $q_f \in Q_f$ and $q_c \in Q_c$ belonging to a BSCC, $d(\mathcal{M}(\mathcal{A}_{q_c}), \mathcal{M}(\mathcal{A}_{q_f}) = 1$. Then $\mathcal{A}$ is AA-diagnosable.*

Proof. Let ρ_0 be a faulty run, $\alpha > 0, \varepsilon > 0$, $\sigma_0 = \pi(\rho_0)$ and $n_0 = |\sigma_0|$.
Before developping the proof, we sketch its structure and illustrate it in Figure 5. First we extend the runs with observed sequences σ_0 by n_b observable events where n_b is chosen in order to get a high probability that the runs end in a BSCC.
Let $\sigma \in \Sigma_o^{n_b}$ be such an observed sequence. We partition the possible runs with observed sequence $\sigma_0\sigma$ into three sets $\mathfrak{R}_\sigma^F$, $\mathfrak{R}_{\sigma'}^C$ and $\mathfrak{R}_\sigma^T$. $\mathfrak{R}_\sigma^F$ is the subset of faulty runs while $\mathfrak{R}_\sigma^C$ (resp. $\mathfrak{R}_\sigma^T$) is the set of correct runs ending (resp. not ending) in a BSCC. At first, we do not take into account the runs in $\mathfrak{R}_\sigma^T$. We apply Lemma 1 to obtain an integer n_σ such that from $\mathfrak{R}_\sigma^F$ and $\mathfrak{R}_\sigma^C$ we can diagnose with (appropriate) high probability and low correctness proportion after n_σ observations. Among the runs that trigger diagnosable observed sequences, some will exceed the correctness proportion, ε, when taking into account the runs from $\mathfrak{R}_\sigma^T$. Yet we will show that the probability of such runs is small when cumulated over all extensions σ leading to the required upper bound α.

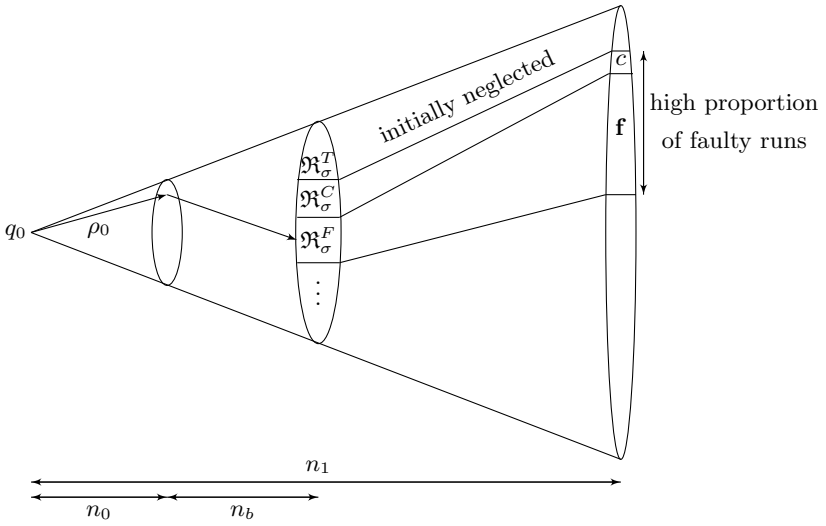


Fig. 5. Illustration of the proof of Lemma B.

Let ε, α be positive reals. Since with probability 1 a random run ends in a BSCC, there exists n_b such that:

$\mathbb{P}\{\rho \in \mathsf{SR}_{n_0+n_b} \mid \sigma_0 \preceq \pi(\rho) \wedge last(\rho) \text{ does not belong to a BSCC}\} < \eta$
where $\eta = \frac{\alpha\varepsilon}{4}$.

Let $\mathfrak{S} = \{\sigma \in \Sigma_o^{n_b} \mid \exists \rho \in \mathsf{SR}_{n_0+n_b}\ \rho_0 \preceq \rho \wedge \pi(\rho) = \sigma_0\sigma\}$. Pick some $\sigma \in \mathfrak{S}$ and define:

- $\mathfrak{R}_\sigma^F = \{\rho \in \mathsf{SR}_{n_0+n_b} \mid \pi(\rho) = \sigma_0\sigma \wedge last(\rho) \in Q_f\}$;
- $\mathfrak{R}_\sigma^C = \{\rho \in \mathsf{SR}_{n_0+n_b} \mid \pi(\rho) = \sigma_0\sigma \wedge last(\rho) \in Q_c$ and belongs to a BSCC$\}$;
- $\mathfrak{R}_\sigma^T = \{\rho \in \mathsf{SR}_{n_0+n_b} \mid \pi(\rho) = \sigma_0\sigma \wedge last(\rho) \in Q_c$ and does not belong to a BSCC$\}$.

Let $Q_c^\sigma = \{last(\rho) \mid \rho \in \mathfrak{R}_\sigma^C\}$ and $Q_f^\sigma = \{last(\rho) \mid \rho \in \mathfrak{R}_\sigma^F\}$. For all pair $(q_f, q_c) \in Q_f^\sigma \times Q_c^\sigma$, consider the initial-fault pLTS $\mathcal{A}' = \langle q_0', \mathcal{A}_{q_f}, \mathcal{A}_{q_c}\rangle$ and denote $\mathbb{P}'$ its associated probability. Due to Lemma 1, for all α', ε' there exists n_{q_f,q_c} such that for all $n \geq n_{q_f,q_c}$:

$$\mathbb{P}'\{\rho \in \mathsf{SR}_n \mid q_0'\mathbf{f}q_f \preceq \rho \wedge \frac{{p'}_c^{\pi(\rho)}}{{p'}_c^{\pi(\rho)} + {p'}_f^{\pi(\rho)}} > \varepsilon'\} \leq \alpha'.$$

where ${p'}_c^{\pi(\rho)}$ (resp. ${p'}_f^{\pi(\rho)}$) is the probability in $\mathcal{A}'$ of a correct (resp. faulty) run with observed sequence $\pi(\rho)$. Define in $\mathcal{A}$, $p_c^{\pi(\rho)}$ (resp. $p_f^{\pi(\rho)}$) the probability of a correct (resp. faulty) run with observed sequence $\pi(\rho)$, $p_f = \min(\mathbb{P}(\rho) \mid \rho \in \mathfrak{R}_\sigma^F)$ and $p_c = \sum_{\rho \in \mathfrak{R}_\sigma^F} \mathbb{P}(\rho)$. By a worst-case reasoning, one gets ${p'}_c^{\pi(\rho)} \geq \frac{2}{p_c} p_c^{\sigma_0\sigma\pi(\rho)}$ and ${p'}_f^{\pi(\rho)} \leq \frac{2}{p_f} p_f^{\sigma_0\sigma\pi(\rho)}$. Thus for all $n \geq n_0 + n_b + \max(n_{q_f,q_c})$:

$$\mathbb{P}\{\rho \in \mathsf{SR}_n \mid \exists \rho' \in R_\sigma^F \wedge \rho' \preceq \rho \wedge \frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)} + \frac{p_c}{p_f} p_f^{\pi(\rho)}} > \varepsilon'\} \leq 2\alpha'.$$

where the factor 2 takes into account the first transition in $\mathcal{A}'$.
Choosing $\varepsilon' = \frac{\varepsilon p_f}{\varepsilon p_f + (2-\varepsilon)p_c}$ and $\alpha' = \frac{\alpha}{4|\mathfrak{S}|}$, after algebraic operations the previous inequality can be rewritten:

$$\mathbb{P}\{\rho \in \mathsf{SR}_n \mid \exists \rho' \in R_\sigma^F \wedge \rho' \preceq \rho \wedge \frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)} + p_f^{\pi(\rho)}} > \frac{\varepsilon}{2}\} \leq \frac{\alpha}{2|\mathfrak{S}|}.$$

Let $n_\sigma = n_0 + n_b + \max(n_{q_f,q_c} \mid (q_f, q_c) \in Q_f^\sigma \times Q_c^\sigma)$ and $n_1 = \max(n_\sigma \mid \sigma \in \mathfrak{S})$ and consider $n \geq n_1$.

We now take into account the runs of $\mathfrak{R}_\sigma^T$. Let $\rho \in \{\rho \in \mathsf{SR}_n \mid \exists \rho' \in \mathfrak{R}_\sigma^F \wedge \rho' \preceq \rho\}$. Define $p_t^{\pi(\rho)}$ be the probability of runs (1) with observed sequence $\pi(\rho)$ and (2) extending runs of $\mathfrak{R}_\sigma^T$. Since a correct run with observed sequence $\pi(\rho)$ must have a prefix in $\mathfrak{R}_\sigma^T$ or in $\mathfrak{R}_\sigma^C$:

$$\mathsf{CorP}(\pi(\rho)) \leq \frac{p_c^{\pi(\rho)} + p_t^{\pi(\rho)}}{p_c^{\pi(\rho)} + p_t^{\pi(\rho)} + p_f^{\pi(\rho)}}.$$

Consider the following set of runs:
$\widetilde{\mathfrak{R}}^n_\sigma = \{\rho \in \mathsf{SR}_n \mid \exists \rho' \in \mathfrak{R}^F_\sigma \wedge \rho' \preceq \rho \wedge \frac{p_c^{\pi(\rho)}+p_t^{\pi(\rho)}}{p_c^{\pi(\rho)}+p_t^{\pi(\rho)}+p_f^{\pi(\rho)}} > \varepsilon \wedge \frac{p_c^{\pi(\rho)}}{p_t^{\pi(\rho)}+p_f^{\pi(\rho)}} \leq \frac{\varepsilon}{2}\}$
For $\rho \in \widetilde{\mathfrak{R}}^n_\sigma$, one gets by algebraic operations, $\frac{2p_t^{\pi(\rho)}}{\varepsilon} > p_f^{\pi(\rho)}$.
Thus $\mathbb{P}(\widetilde{\mathfrak{R}}^n_\sigma) < \frac{2\mathbb{P}(\mathfrak{R}^T_\sigma)}{\varepsilon}$ and $\sum_{\sigma\in\mathfrak{S}} \mathbb{P}(\widetilde{\mathfrak{R}}^n_\sigma) < \frac{2\sum_{\sigma\in\mathfrak{S}} \mathbb{P}(\mathfrak{R}^T_\sigma)}{\varepsilon}$.
Due to the choice of n_b, $\sum_{\sigma\in\mathfrak{S}} \mathbb{P}(\mathfrak{R}^T_\sigma) < \eta$. So $\sum_{\sigma\in\mathfrak{S}} \mathbb{P}(\widetilde{\mathfrak{R}}^n_\sigma) < \frac{2\eta}{\varepsilon} = \frac{\alpha}{2}$.

Summarizing for all $n \geq n_1$:

$$
\begin{aligned}
&\mathbb{P}\{\rho \in \mathsf{SR}_n \mid \rho_0 \preceq \rho \wedge \mathsf{CorP}(\pi(\rho)) > \varepsilon\} \\
=&\sum_{\sigma\in\mathfrak{S}} \mathbb{P}\{\rho \in \mathsf{SR}_n \mid \rho_0 \preceq \rho \wedge \sigma_0\sigma \preceq \pi(\rho) \wedge \mathsf{CorP}(\pi(\rho)) > \varepsilon\} \\
\leq&\sum_{\sigma\in\mathfrak{S}} \mathbb{P}\{\rho \in \mathsf{SR}_n \mid \exists \rho' \in \mathfrak{R}^F_\sigma \wedge \rho' \preceq \rho \wedge \frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)}+p_f^{\pi(\rho)}} > \frac{\varepsilon}{2}\} \\
+&\mathbb{P}\{\rho \in \mathsf{SR}_n \mid \exists \rho' \in \mathfrak{R}^F_\sigma \wedge \rho' \preceq \rho \wedge \frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)}+p_f^{\pi(\rho)}} \leq \frac{\varepsilon}{2} \wedge \frac{p_c^{\pi(\rho)}+p_t^{\pi(\rho)}}{p_c^{\pi(\rho)}+p_t^{\pi(\rho)}+p_f^{\pi(\rho)}} > \varepsilon\} \\
\leq&|\mathfrak{S}|\frac{\alpha}{2|\mathfrak{S}|} + \frac{\alpha}{2} = \alpha
\end{aligned}
$$

which establishes AA-diagnosability of $\mathcal{A}$. □

A.5 Undecidability of uniform AA-diagnosability

The proof is done by a more sophisticated reduction from the emptiness problem for PA than the one of Theorem 1. While it has no connexion with the proof, it is interesting to observe that whatever the original PA, the pLTS of the proof is AA-diagnosable and not A-diagnosable.

Theorem 3. *The uniform AA-diagnosability problem is undecidable for pLTS.*

Proof. The proof is by reduction from the emptiness problem for probabilistic automata. Here we consider a probabilistic automaton with an initial distribution $\mathbf{I}$ so that the emptiness is undecidable even when for all word w, $\frac{1}{4} \leq \mathbf{Pr}_{\mathcal{A}}(w) \leq \frac{3}{4}$. Let $\mathcal{A}$ be such a probabilistic automaton. Define the pLTS $\mathcal{A}' = \langle Q', q_0, \Sigma', T', \mathbf{P}'\rangle$ as follows.

- $\Sigma' = \Sigma \uplus \{\sharp, \flat, \mathbf{f}, u\}$, $\Sigma'_{uo} = \{u, \mathbf{f}\}$;
- $Q' = \{q^u, q^f \mid q \in Q\} \cup \{q_0, b^u, b^f\}$;

– $T' = \{(q_0, u, q^u), (q_0, \mathbf{f}, q^f) \mid q \in Q, \mathbf{I}[q] > 0\} \cup$
$\{(q^u, a, q'^u), (q^f, a, q'^f) \mid q, q' \in Q, a \in \Sigma, \mathbf{P}_a[q, q'] > 0\} \cup$
$\{(q^u, \sharp, q'^u) \mid q \in F, q' \in Q, \mathbf{I}[q'] > 0\} \cup$
$\{(q^f, \sharp, q'^f) \mid q \in Q \setminus F, q' \in Q, \mathbf{I}[q'] > 0\} \cup$
$\{(q^u, \flat, b^u) \mid q \in Q \setminus F\} \cup \{(q^f, \flat, b^f) \mid q \in F\}$
$\cup \{b^u, \sharp, b^u\} \cup \{b^u, \flat, b^u\} \cup \{b^f, \flat, b^f\}$
– $\mathbf{P}'$ is defined by:
- For all $(q_0, u, q^u), (q_0, \mathbf{f}, q^f) \in T$, $\mathbf{P}'(q_0, u, q^u) = \mathbf{P}'(q_0, \mathbf{f}, q^f) = \frac{\mathbf{I}[q]}{2}$;
- For all $(q^u, a, q'^u) \in T$, $\mathbf{P}'(q^u, a, q'^u) = \frac{\mathbf{P}_a[q,q']}{1+|\Sigma|}$;
- For all $(q^f, a, q'^f) \in T$, $\mathbf{P}'(q^f, a, q'^f) = \frac{\mathbf{P}_a[q,q']}{1+|\Sigma|}$;
- For all $(q^u, \sharp, q'^u) \in T$, $\mathbf{P}'(q^u, \sharp, q'^u) = \frac{\mathbf{I}[q']}{1+|\Sigma|}$;
- For all $(q^f, \sharp, q'^f) \in T$, $\mathbf{P}'(q^f, \sharp, q'^f) = \frac{\mathbf{I}[q']}{1+|\Sigma|}$;
- For all $(q^u, \flat, b^u) \in T$, $\mathbf{P}'(q^u, \flat, b^u) = \frac{1}{1+|\Sigma|}$;
- For all $(q^f, \flat, b^f) \in T$, $\mathbf{P}'(q^f, \flat, b^f) = \frac{1}{1+|\Sigma|}$;
- $\mathbf{P}'(b^u, \flat, b^u) = \mathbf{P}'(b^u, \sharp, b^u) = \frac{1}{2}$, $\mathbf{P}'(b^f, \flat, b^f) = 1$.

This reduction is represented in Figure 6.

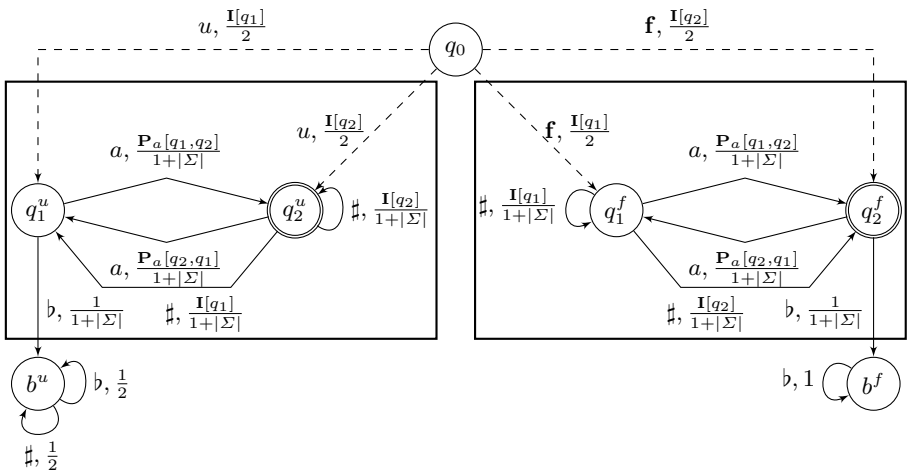


Fig. 6. From probabilistic automata to pLTS: rectangles surround the two copies of Q.

We claim that $\mathcal{A}'$ is uniformly AA-diagnosable if and only if $\mathcal{L}_{\mathcal{A},\frac{1}{2}} = \emptyset$. Observe first that for all $q \in Q$, $\mathcal{L}^\omega(\mathcal{A}'_{q^f}) \subseteq \mathcal{L}^\omega(\mathcal{A}'_{q^u})$ so that all faulty runs are ambiguous.

• Assume that there exists a word $w \in \Sigma^*$ such that $\mathbf{Pr}_{\mathcal{A}}(w) > \frac{1}{2}$. We will prove that $\mathcal{A}'$ is not uniformly $\frac{1}{2}$-diagnosable. So we pick an arbitrary $\alpha < 1$ and n_α.
Consider the observed sequence $\sigma_n = (w\sharp)^n$ for some n to be fixed later. Due to our hypothesis on $\mathcal{A}$, it is ambiguous. Let

$$\gamma_n = \frac{\mathbb{P}(\{\rho' \in \mathsf{C} \mid \pi(\rho') = \sigma_n\})}{\mathbb{P}(\{\rho' \in \mathsf{F} \mid \pi(\rho') = \sigma_n\})}$$

Then γ_n fulfills $\lim_{n\to\infty} \gamma_n = \infty$.
Let ρ_n be a faulty run with $\pi(\rho^n) = \sigma_n$ and ρ be a signalling run extending ρ_n with $|\rho|_o = |\rho_n|_o + n_\alpha$. $\pi(\rho)$ can be decomposed as $\pi(\rho) = \sigma_n \sigma \flat^k$ with $\sigma \in (\Sigma \cup \{\sharp\})^*$. By a straightforward examination of $\mathcal{A}'$ one gets:

$$\frac{\mathbb{P}(\{\rho' \in \mathsf{C} \mid \pi(\rho') = \pi(\rho)\})}{\mathbb{P}(\{\rho' \in \mathsf{F} \mid \pi(\rho') = \pi(\rho)\})} \geq \gamma_n 3^{-n_\alpha}$$

Choosing n such that $\gamma_n 3^{-n_\alpha} > 1$, one gets: $\mathsf{CorP}(\rho) > \frac{1}{2}$. So:

$$\mathbb{P}(\{\rho \in \mathsf{SR}_{n_\alpha + |\rho_n|_o} \mid \rho_n \preceq \rho \wedge \mathsf{CorP}(\pi(\rho)) > \frac{1}{2}\}) = \mathbb{P}(\rho) > \alpha \mathbb{P}(\rho)$$

Thus $\mathcal{A}'$ is not uniformly $\frac{1}{2}$-diagnosable.
• Conversely assume that for all word $w \in \Sigma^*$, $\mathbf{Pr}_{\mathcal{A}}(w) \leq \frac{1}{2}$. Then for all observed sequence $\sigma \in (\Sigma \cup \{\sharp\})^*$, $\mathsf{CorP}(\sigma) \leq \frac{1}{2}$.
Pick any positive ε, α. From any state q^f, there is a path of length at most $|Q|$ to a state q'^f with $q' \in F$ (possibly using a $\sharp$ transition). Let $pmin$ be the minimal positive probability occurring in $\mathcal{A}'$. A path starting from some q^f of length of $|Q|$ has a probability at least $pmin$ to enter state b^f. So a path from some q^f of length $n|Q|$ has a probability at least $1 - (1 - pmin)^n$ to enter b^f. Let n_0 be such that $(1 - pmin)^{n_0} \leq \alpha$ and n_1 be such that $2^{-n_1} \leq \frac{\varepsilon}{3}$. Define $n_\alpha = n_0|Q| + n_1$. Consider any faulty run ρ_f.

$$\mathbb{P}(\{\rho \in \mathsf{SR}_{n_\alpha + |\rho_f|_o} \mid \rho_f \preceq \rho \wedge \mathsf{CorP}(\pi(\rho)) \leq \varepsilon)$$

$$\geq \mathbb{P}(\{\rho \in \mathsf{SR}_{n_\alpha + |\rho_f|_o} \mid \rho_f \preceq \rho \wedge \mathsf{last}(\rho_{\downarrow n_0|Q| + |\rho_f|_o}) = b^f \wedge \mathsf{CorP}(\pi(\rho)) \leq \varepsilon)$$

Let $\sigma \in (\Sigma \cup \{\sharp\})^*$ be an observed sequence. After an occurrence of $\flat$ the fraction between the probability of correct runs with observed sequence $\sigma\flat$ over the probability of faulty runs with observed sequence $\sigma\flat$ is at most multiplied by 3 due to the lower bound $\frac{1}{4}$ for acceptance probability. Let σ be an observed sequence ending by $\flat$. After a new occurrence of $\flat$ the fraction between the probability of correct runs with observed sequence

$\sigma\flat$ over the probability of faulty runs with observed sequence $\sigma\flat$ is divided by 2. So due to our choice of n_0 and n_1:

$$\mathbb{P}(\{\rho \in \mathsf{SR}_{n_\alpha+|\rho_f|_o} \mid \rho_f \preceq \rho \wedge \mathsf{last}(\rho_{\downarrow n_0|Q|+|\rho_f|_o}) = b^f \wedge \mathsf{CorP}(\pi(\rho)) \leq \varepsilon)$$

$$= \mathbb{P}(\{\rho \in \mathsf{SR}_{n_\alpha+|\rho_f|_o} \mid \rho_f \preceq \rho \wedge \mathsf{last}(\rho_{\downarrow n_0|Q|+|\rho_f|_o}) = b^f) \geq (1-\alpha)\mathbb{P}(\rho_f)$$

Thus $\mathcal{A}'$ is uniformly ε-diagnosable. □