A Tale of Two Diagnoses in Probabilistic Systems

Nathalie Bertrand, Serge Haddad, Engel Lefaucheux

A Tale of Two Diagnoses in Probabilistic Systems

Nathalie Bertrand[a,], Serge Haddad[b,1,], Engel Lefaucheux[a,b,]

[a] *Inria, Campus Universitaire de Beaulieu, Rennes, France*
[b] *LSV, ENS Cachan, Université Paris-Saclay & 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: exact diagnosability requires that almost surely a fault is detected and that no fault is erroneously claimed; approximate diagnosability tolerates a small error probability when claiming a fault; last, accurate approximate diagnosability guarantees that the error probability can be chosen arbitrarily small. In this article, we first refine the specification of diagnosability by identifying three criteria: (1) detecting faulty runs or providing information for all runs (2) considering finite or infinite runs, and (3) requiring or not a uniform detection delay. We then give a complete picture of relations between the different diagnosability specifications for probabilistic systems and establish characterisations for most of them in the finite-state case. Based on these characterisations, we develop decision procedures, study their complexity and prove their optimality. We also design synthesis algorithms to construct diagnosers and we analyse their memory requirements. Finally we establish undecidability of the diagnosability problems for which we provided no characterisation.

Keywords: Fault diagnosis, Partial observation, Probabilistic transition systems, Markov chains

1. Introduction

Diagnosis and diagnosability. 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 [1]. In this work, we concentrate on diagnosis as studied in control of discrete event systems, where it is applied to partially observable systems prone to faults. In this context, it consists in designing fast automatic detection of malfunctions. Diagnosis raises two important issues: deciding whether the

*Corresponding authors
Email addresses: nathalie.bertrand@inria.fr (Nathalie Bertrand), serge.haddad@lsv.fr (Serge Haddad), engel.lefaucheux@lsv.fr (Engel Lefaucheux)

[1] The work of S. Haddad has been supported by ERC project EQualIS (FP7-308087).

system is *diagnosable* and, in the positive case, synthesising a *diagnoser* possibly satisfying additional requirements about memory size, implementability, etc.

Diagnosis of discrete event systems. One of the proposed approaches consists in modelling these systems by partially observable labelled transition systems (LTS) [2]. In such a framework, diagnosability requires that the occurrence of unobservable faults can be deduced from the sequence of observable events occurring before and after the fault. Formally, an LTS is diagnosable if there exists a diagnoser that satisfies two properties: *reactivity* and *correctness*. Reactivity requires that whenever 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. A polynomial time algorithm for diagnosability of LTS was provided [3], however the diagnoser itself can be of exponential size w.r.t. the size of the LTS. Diagnosis has been extended to numerous models (*e.g.* Petri nets [4], pushdown systems [5], etc.) and settings (*e.g.* centralised, decentralised, distributed). It has had an impact on important application areas, such as telecommunication network failure diagnosis. Also, several contributions, gathered under the generic name of active diagnosis, focus on enforcing the diagnosability of a system [6, 7, 8, 9, 10].

Diagnosis of stochastic systems. Beyond LTS, diagnosis was also considered in a quantitative setting, and namely for probabilistic labelled transition systems (pLTS) [11, 12]. Probabilistic LTS can be seen as Markov chains in which the transitions are labelled with events, therefore, one can define a probability measure over their infinite runs. In that context, the specification of the reactivity and correctness properties can be relaxed. More precisely, reactivity now 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 specifications can be considered. A first option, called *A-diagnosability*, is to stick to strong correctness and therefore ask the diagnoser to only claim fault occurrences when a fault is certain. In contrast, given an error threshold ε, *ε-diagnosability* tolerates small errors, allowing to claim a fault if the conditional probability that no fault occurred does not exceed the threshold. Last, so-called *AA-diagnosability* requires the pLTS to be ε-diagnosable for every ε, thus allowing the designer to tune the threshold according to the criticality of the system. The two notions of A-diagnosability and AA-diagnosability were introduced in [11].

Remaining issues. A few semantical and algorithmical issues remained untouched in the above line of work. First, diagnosability was only considered with respect to finite faulty runs. It seems to us as important to also consider diagnosability of correct runs, which requires to introduce the notion of *ambiguity*: a faulty (resp. correct) run is ambiguous if its observed sequence is identical to the one of a correct (resp. faulty) run. Second, we observed that reactivity can be strengthened by requiring that the probability that a run remains ambiguous after a fault occurrence does not depend on the precise run and rather

decreases uniformly. This is motivated by the analogy with non probabilistic LTS, for which the detection delay is uniform. Last the decidability and the exact complexity of the different diagnosability problems and of the diagnosers synthesis were left open. In particular while A-diagnosability was claimed to be in PTIME [13], for what concerns approximate diagnosability (*i.e.* ε and AA-diagnosability), up to our knowledge, a (PTIME-checkable) sufficient condition for AA-diagnosability has been given [11], but no decidability result is known.

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

- In order to give a solid semantical classification of diagnosability notions, we define criteria for diagnosability in probabilistic systems, depending on (1) whether the diagnoser provides information for faulty runs only or for all runs, (2) whether ambiguity is defined at the level of infinite runs, or for longer and longer finite prefixes, and (3) whether the finite delay for fault detection is uniform or may depend on the faulty run. These three dimensions combined with approximate versus exact diagnosis yield several meaningful specifications of diagnosability. Under our terminology, A-diagnosability is renamed uniform FF-diagnosability, where the first F stands for finite prefixes and the second for faulty runs, and we make explicit the fact that the detection delay is uniform. Also, AA-diagnosability corresponds to uniform AFF-diagnosability, where A means approximate, and we call it uniform accurate approximate diagnosability. Last, non uniform AFF-diagnosability corresponds to monitorability for hidden Markov chains, introduced in [14] and further studied in [15], although the two notions first look very different. Beyond the formalisation of all these diagnosability notions, we establish the precise connections between them.

- For finite state probabilistic systems, we show that the notions of exact diagnosability can be characterised by a structural property of a synchronised product of the pLTS with a deterministic (finite or Büchi) automaton acting as an observer. We also characterise accurate approximate diagnosability as a separation property between labelled Markov chains (LMC), precisely a *distance* 1 between appropriate pairs of LMCs built from the pLTS.

- The previous characterisations yield PSPACE procedures for exact diagnosability and a PTIME algorithm for accurate approximate diagnosability thanks to some recent result [16], recovering the PTIME decidability result for monitorability [15].

- Afterwards, we design algorithms for the synthesis of exact diagnosers and prove that their size $2^{\Theta(n)}$ (where n is the number of states of the pLTS model) is optimal. On the contrary, approximate diagnosers may require infinite memory.

- Finally we show that all approximate diagnosability problems except for AFF-diagnosability are undecidable. We also establish a matching complexity lower bound (PSPACE-hardness) for all exact diagnosability problems, disproving the polynomial time result for FF-diagnosability [13]. Note that the complexity of FF-diagnosability was also established in [14], where it corresponds to strong monitorability of an invariant property.

Coming back to the seminal works, while AA-diagnosability is a more intricate notion than A-diagnosability as witnessed by their decidability status (undecidable vs PSPACE), perhaps surprisingly their non uniform variants exhibit an opposite relation (PTIME vs PSPACE). However approximate diagnosers may require infinite memory in contrast to exact diagnosers. In a sense, our contributions highlight the surprises in the story about exact and approximate probabilistic diagnosis, hence the title.

Organisation. In Section 2, we introduce probabilistic LTS, define the possible diagnosability specifications and establish their connections. In Section 3, we provide characterisations for most of these specifications when pLTS are finite. In Section 4, we establish decidability procedures for the diagnosability problem based on these characterisations and we study the complexity of these procedures. In Section 5, we design algorithms for synthesis of diagnosers with optimal size. In Section 6, we prove undecidability and hardness results. Finally, we conclude and give some perspectives to this work in Section 7. For readability concerns, the most technical proofs are deferred to the appendix.

This article extends two conference papers on fault diagnosis for probabilistic systems, respectively on exact diagnosability [12] and approximate diagnosability [17].

2. Diagnosability specification

2.1. Probabilistic labelled transition systems

Our model of stochastic discrete event systems is a transition system labelled with events and where transitions outgoing a state are randomly chosen.

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 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;
- $\mathbf{P}$ is the transition matrix from T to $\mathbb{Q}_{>0}$ fulfilling for all $q \in Q$: $\sum_{(q,a,q') \in T} \mathbf{P}[q, a, q'] = 1$.

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 state q. By definition, in every state q of the pLTS at least one transition is enabled, *i.e.* a pLTS is *live*. We assume all pLTS we consider to be 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 $\rho = q_0 a_0 q_1 \dots$ such that for all $i \geq 0$, $q_i \in Q$, $a_i \in \Sigma$ and when q_{i+1} is defined, $q_i \xrightarrow{a_i} q_{i+1}$. The notion of run can be generalised, starting from an arbitrary state q. We write Ω for the set of all infinite runs starting from q_0, assuming the pLTS $\mathcal{A}$ is clear from context. When it is finite, ρ ends in a state and its *length*, denoted by $|\rho|$, is the number of events occurring in it. Given a finite run $\rho = q_0 a_0 q_1 \dots q_n$ and a (finite or infinite) run $\rho' = q_n a_n q_{n+1} \dots$ starting in the last state of ρ, we call concatenation of ρ and ρ' the run $\rho\rho' = q_0 a_0 q_1 \dots q_n a_n q_{n+1} \dots$. The run ρ is then a *prefix* of $\rho\rho'$, which we denote by $\rho \preceq \rho\rho'$. The *cylinder* generated by a finite run ρ consists of all the infinite runs that extend ρ: $\mathsf{Cyl}(\rho) = \{\rho' \in \Omega \mid \rho \preceq \rho'\}$. The sequence associated with $\rho = q a_0 q_1 \dots$ is the word $\sigma_\rho = a_0 a_1 \dots$, and we write indifferently $q \stackrel{\rho}{\Rightarrow}$ or $q \stackrel{\sigma_\rho}{\Longrightarrow}$ (resp. $q \stackrel{\rho}{\Rightarrow} q'$ or $q \stackrel{\sigma_\rho}{\Longrightarrow} q'$) for an infinite (resp. finite) run ρ. A state q is *reachable* (from the initial state q_0) if there exists a run ρ such that $q_0 \stackrel{\rho}{\Rightarrow} q$, 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 $\mathcal{L}^\omega(\mathcal{A}) = \{\, \sigma \in \Sigma^\omega \mid \exists\ q_0 \stackrel{\sigma}{\Rightarrow} \,\}$.

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:

$$\mathbb{P}_{\mathcal{A}}(\mathsf{Cyl}(q_0 a_0 q_1 \dots q_n)) = \mathbf{P}[q_0, a_0, q_1] \cdots \mathbf{P}[q_{n-1}, a_{n-1}, q_n] \ .$$

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 $\sum_{\rho\in R} \mathbb{P}(\rho)$ which is consistent since all intersections of associated cylinders are empty.

2.2. Partial observation and ambiguity

In order to formalise problems related to fault diagnosis, we partition the set of events Σ into two disjoint sets Σ_o and Σ_u, the sets of *observable* and *unobservable events*, respectively. Moreover, we distinguish a special *fault* event $\mathbf{f} \in \Sigma_u$ that is unobservable. For σ a finite word over Σ, its length is denoted by $|\sigma|$. The projection of words from Σ^* onto the observable event alphabet Σ_o is defined inductively by: $\pi(\varepsilon) = \varepsilon$; for $a \in \Sigma_o$, $\pi(\sigma a) = \pi(\sigma)a$; and for $a \notin \Sigma_o$,

$\pi(\sigma a) = \pi(\sigma)$. We write $|\sigma|_o$ for the observable length of σ, that is $|\pi(\sigma)|$. When σ is an infinite word over Σ, its projection is the limit of the projections of its finite prefixes, and by convention $|\sigma|_o = \infty$. As usual the projection mapping π is extended to languages: for $L \subseteq \Sigma^*$, $\pi(L) = \{\pi(\sigma) \mid \sigma \in L\}$. With respect to the partition $\Sigma = \Sigma_o \uplus \Sigma_u$, a pLTS $\mathcal{A}$ is said *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 onto Σ_o is thus an observed sequence.

The *observable length* of a run ρ denoted by $|\rho|_o \in \mathbb{N} \cup \{\infty\}$, is the number of observable events that occur in it: $|\rho|_o = |\sigma_\rho|_o$. A *signalling run* is a finite run $q_0 a_0 q_1 \cdots a_{n-1} q_n$ such that a_{n-1} is an observable event. Signalling runs are precisely the relevant runs w.r.t. partial observation issues since each observable event provides 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 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 $0 < n \leq |\rho|_o$, $\rho_{\downarrow n}$ denotes the unique prefix of ρ that belongs to SR_n. For convenience, the empty run q_0 is defined as the single signalling run of null length. For an observed sequence $\sigma \in \Sigma_o^*$, we define its cylinder $\mathsf{Cyl}(\sigma) = \sigma\Sigma_o^\omega$ and the associated probability $\mathbb{P}(\mathsf{Cyl}(\sigma)) = \mathbb{P}(\{\rho \in \Omega \mid \pi(\rho_{\downarrow|\sigma|}) = \sigma\}) = \mathbb{P}(\{\rho \in \mathsf{SR}_{|\sigma|} \mid \pi(\rho) = \sigma\})$, often shortened as $\mathbb{P}(\sigma)$.

Let us now classify runs depending on whether they contain a fault or not. A run ρ is *faulty* if its associated sequence 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 infinite runs such that their signalling subrun of observable length n is faulty (resp. correct). We further define the sets of all finite faulty and correct signalling runs F and C and the sets of infinite faulty and correct runs F_∞ and C_∞. A run ρ is a *minimal faulty run* if it is a faulty signalling run and there does not exists a prefix ρ' of ρ that is a faulty signalling run. We write for all $n \in \mathbb{N}$, minF_n for the set of minimal faulty runs of length n and $\mathsf{minF} = \bigcup_{n\in\mathbb{N}} \mathsf{minF}_n$ for the set of all minimal faulty runs. W.l.o.g., by considering two copies of each state of the pLTS, we assume that the state space Q of $\mathcal{A}$ 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 $\pi(\rho) = \pi(\rho') = \sigma$. Otherwise, it is either *surely faulty*, or *surely correct* depending on whether $\pi^{-1}(\sigma) \cap \mathsf{SR} \subseteq \mathsf{F}$ or $\pi^{-1}(\sigma) \cap \mathsf{SR} \subseteq \mathsf{C}$. A run is ambiguous, surely correct or surely faulty if its observed sequence is ambiguous, surely correct or surely faulty respectively. We write Sf_∞ for the set of infinite surely faulty runs. In addition Sf_n is the set of infinite runs whose signalling subrun of observable length n is surely faulty.

2.3. Which diagnosis for pLTS?

For nonprobabilistic systems, diagnosis is defined by the existence of a diagnoser that presents three main features: *verdict*, *correctness* and *reactivity*. Verdict specifies the nature of the information the diagnoser provides along the run: it may only be related to detection of faults or may also assert that (some prefix of) the run does not include a fault. Correctness specifies that when the diagnoser outputs a verdict, this verdict holds. Reactivity asserts that after a fault occurred or a longer prefix of the run does not include fault, the diagnoser will detect it after a finite delay.

The aim of this section is to define appropriate verdict, correctness and reactivity requirements for probabilistic systems. So we start with informal explanations that also motivate the need of considering different versions of diagnosis.

In seminal works about probabilistic systems, the verdict is limited to fault detections and the reactivity is usually relaxed by requiring that when a fault occurs, a diagnoser *almost surely* detects it after a finite delay [11]. Let us look at the pLTS of Figure 1. Considered as a LTS, one cannot detect that the run $q_0\mathbf{f}(f_1a)^\omega$ is faulty due to the correct run $q_0u(q_1a)^\omega$ with same observed sequence a^ω. However with probability 1, a faulty run will produce a 'b' and thus faults are almost surely detected in this pLTS. On the other hand, one cannot provide any information about the single correct run $q_0u(q_1a)^\omega$ since its observed sequence is ambiguous as well as any of its prefix.

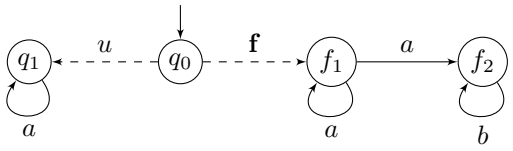


Figure 1: Detecting faults but not correct runs. When probabilities are not specified, we assume a uniform distribution.

In order to examine which information could be provided about correct runs, let us look at the pLTS of Figure 2. A sequence a^n is ambiguous. However up to the $n-1^{th}$ observation, all the runs that correspond this observed sequence were correct which is a useful information for instance to restart later the system from a correct state. Along the (surely correct) observed sequence a^ω, the observer can always deduce that longer and longer prefixes of the run were correct while never being able to assert that the current run is correct.

The correctness requirement may be specified in different ways. For an exact diagnosis, we ask that a fault can be claimed only when a fault surely happened (as it is the case in non probabilistic systems). However it may be necessary to weaken the correctness requirement as illustrated by the pLTS of Figure 3. Since all observed sequences are ambiguous no exact diagnosis can be provided. However it is clear that when in an enough long observed sequence the ratio between occurrences of 'b' and 'a' is close to 3, the probability that

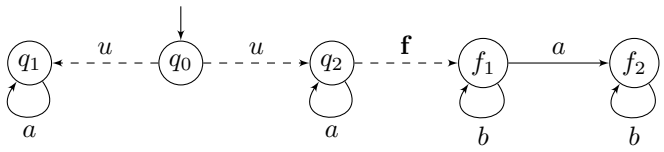


Figure 2: Detecting correctness for longer and longer prefixes of correct runs.

the corresponding run is faulty is close to 1. Let us fix any $\varepsilon > 0$ and only require that the probability of the verdict is erroneous should be less than ε. Then using the strong law of large numbers, (approximate) fault detection is possible in this pLTS.

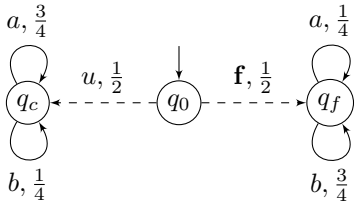


Figure 3: When approximate diagnosis is necessary.

To formalise later the quality of an approximate diagnosis, with every observed sequence $\sigma \in \Sigma_o^*$ we associate a *correctness proportion*

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

which is the conditional probability that a signalling run is correct given that its observed sequence is σ.

The standard way to specify reactivity in probabilistic systems for fault detection is to require that whatever the minimal faulty run, almost surely the diagnoser will output its (faulty) verdict. We may also consider *uniform reactivity* which strengthens reactivity by requiring that the (random) delay is "reasonable" whatever the minimal faulty run. More formally, uniform reactivity ensures that given any positive probability threshold $\alpha > 0$ there exists a delay n_α independent of the minimal faulty run considered such that the probability to exceed this detection delay is bounded by α.

Let us illustrate these reactivity features with the pLTS of Figure 4 for which only approximate diagnosis is possible. Fix some $\varepsilon > 0$ and consider the minimal faulty run $q_0 u q_1 (a q_1)^m \mathbf{f} q_f$. After some occurrences of 'b' (say n), the correctness proportion of the observed sequence $a^m b^n$ is less than ε and thus the diagnoser can output its verdict. However due to the probabilities of an occurrence of 'a' in correct and faulty runs respectively equal to 3/4 and 1/16, n must depend on m and so this reactivity cannot be uniform.

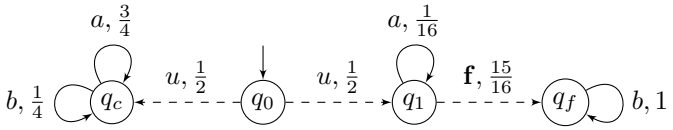


Figure 4: When reactivity cannot be uniform.

In order to formalise the different requirements discussed above, we first define several sets of runs related to ambiguity.

Definition 2 (Ambiguous runs). Let $\mathcal{A}$ be a pLTS, $\varepsilon \geq 0$ and $n \in \mathbb{N}_{>0}$. 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.
- $\mathsf{FAmb}_n^\varepsilon$ is the set of infinite faulty ambiguous runs of $\mathcal{A}$ whose observed sequence of length n, σ fulfils: $\mathsf{CorP}(\sigma) > \varepsilon$.

By definition, for all $n \in \mathbb{N}$, $\mathsf{FAmb}_n^0 = \mathsf{FAmb}_n$. Observe that for all $n \in \mathbb{N}, \varepsilon \geq 0$, $\mathsf{CAmb}_n, \mathsf{FAmb}_n$ and $\mathsf{FAmb}_n^\varepsilon$ are open sets, thus measurable, and CAmb_∞ and FAmb_∞ are analytic, thus measurable for the complete measure. Moreover, we have the following link between the family $\{\mathsf{FAmb}_n\}_{n\in\mathbb{N}_{>0}}$ and the set FAmb_∞.

Lemma 1. *Let $\mathcal{A}$ be a pLTS. Then* $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_\infty \setminus \mathsf{FAmb}_n) = 0$. *Moreover, if $\mathcal{A}$ is finitely branching, then* $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \setminus \mathsf{FAmb}_\infty) = 0$.

We propose five specifications of exact diagnosability for probabilistic systems based on three discriminating criteria: whether the unambiguity requirement holds for faulty runs only or for all runs, whether ambiguity is defined at the level of infinite runs or for longer and longer finite signalling prefixes, and whether the delay before detection of minimal faulty runs is uniform. Those notions are summarised in Figure 5 except for uniformity delayed to next figure.

Definition 3 (Exact diagnosability). Let $\mathcal{A}$ be a pLTS.

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

- $\mathcal{A}$ is FA-*diagnosable* if $\limsup_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \uplus \mathsf{CAmb}_n) = 0$.
- $\mathcal{A}$ is *uniformly* FF-*diagnosable* if for all $\alpha > 0$ there exists $n_\alpha \in \mathbb{N}$ such that for all $n \geq n_\alpha$ and all minimal faulty run $\rho \in \mathsf{minF}$

$$\mathbb{P}(\{\rho' \in \mathsf{FAmb}_{n+|\rho|_o} \mid \rho \preceq \rho'\}) \leq \alpha \cdot \mathbb{P}(\rho) \ .$$

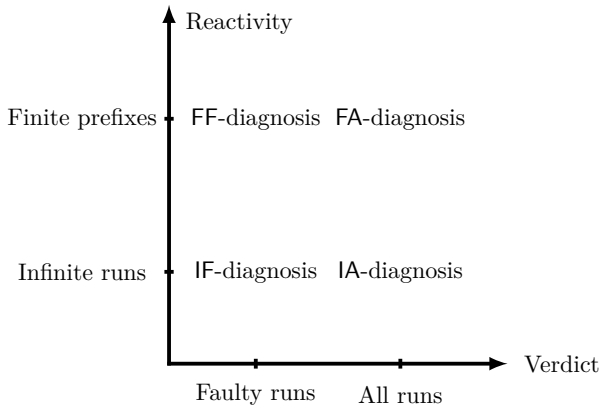


Figure 5: Summarising the variants of exact diagnosis.

Uniform and/or approximate diagnoses are defined for FF-diagnosis as it corresponds to the classical notion of diagnosis. Moreover there is no clear intuition on what would be the meaning of uniformity and approximation for the other variants. So we focus on uniformity and approximation for FF-diagnosis as summarised in Figure 6. $\varepsilon\mathsf{FF}$-*diagnosability* allows the diagnoser to claim a fault when the correctness proportion does not exceed ε, and accurate approximate diagnosability denoted by AFF-*diagnosability* ensures $\varepsilon\mathsf{FF}$-diagnosability for arbitrary $\varepsilon > 0$.

Definition 4 (Approximate diagnosability). Let $\mathcal{A}$ be a pLTS, and $\varepsilon \geq 0$.

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

$$\mathbb{P}(\mathsf{Cyl}(\rho) \cap \mathsf{FAmb}^{\varepsilon}_{n+|\rho|_o}) \leq \alpha \cdot \mathbb{P}(\rho).$$

- $\mathcal{A}$ is *uniformly* $\varepsilon\mathsf{FF}$-*diagnosable* if for all $\alpha > 0$ there exists n_α such that for all minimal faulty run $\rho \in \mathsf{minF}$ and all $n \geq n_\alpha$:

$$\mathbb{P}(\mathsf{Cyl}(\rho) \cap \mathsf{FAmb}^{\varepsilon}_{n+|\rho|_o}) \leq \alpha \cdot \mathbb{P}(\rho).$$

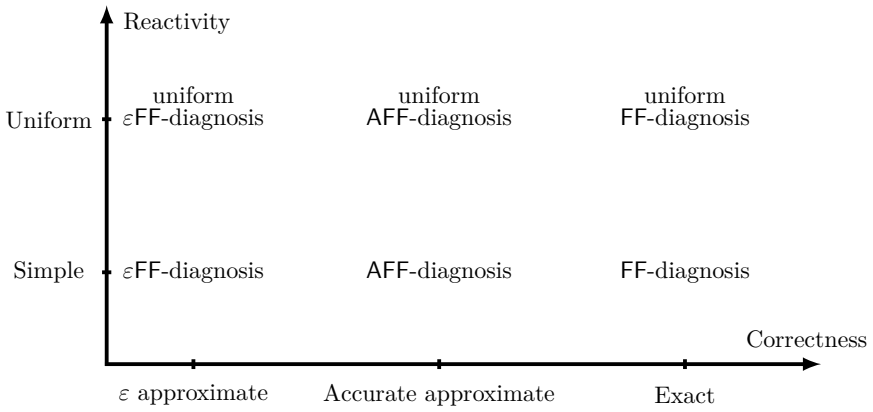
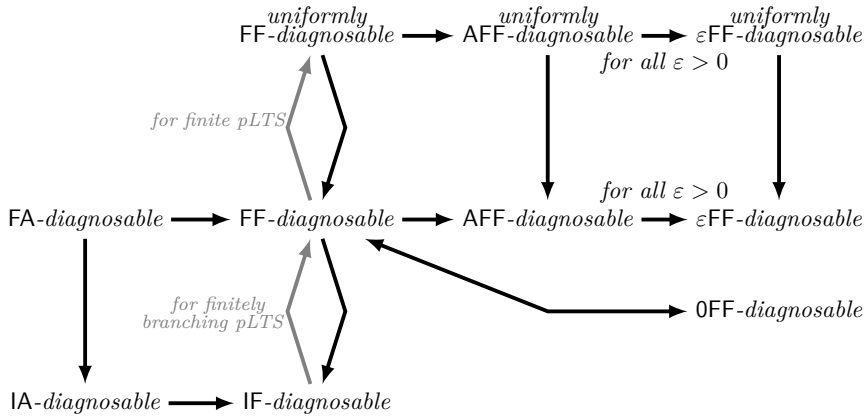


Figure 6: Summarising the approximate variants of FF-diagnosis.

- $\mathcal{A}$ is (resp. uniformly) AFF-*diagnosable* if it is (resp. uniformly) εFF-diagnosable for all $\varepsilon > 0$.

Two variants of diagnosability for stochastic systems were introduced in the original article by Thorsley and Teneketzis [11]: so called *A-diagnosability* and *AA-diagnosability*. In finite pLTS, A-diagnosability corresponds to uniform FF-diagnosability and AA-diagnosability corresponds to uniform AFF-diagnosability. The next theorem summarises the connections between these definitions.

Theorem 1. *The diagnosability notions for pLTS are related according to the diagram below, where arrows represent implications. All implications, except the one from* IF*-diagnosability to* FF*-diagnosability hold for arbitrary infinite-state pLTS. The latter implication holds for finitely branching pLTS. Implications that are not depicted do not hold, already in the case of finite-state pLTS.*



Sketch of proof. Here we only provide the most interesting implications and nonimplications. The remaining proofs are given in appendix.

FF $\Leftrightarrow$ 0FF. Let $\mathcal{A}$ be a 0FF-diagnosable pLTS and $\varepsilon > 0$. Since $(\mathsf{F}_n)_{n\in\mathbb{N}}$ is a nondecreasing sequence converging to F_∞, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0, \mathbb{P}(\mathsf{F}_n \setminus \mathsf{F}_{n_0}) < \varepsilon/2$. By 0FF-diagnosability of $\mathcal{A}$, for all $\rho \in \bigcup_{k\leq n_0} \mathsf{minF}_k$, there exists n_ρ such that for all $n \geq n_\rho$

$$\mathbb{P}(\mathsf{Cyl}(\rho) \cap \mathsf{FAmb}_{n+|\rho|_o}) \leq \frac{\varepsilon}{4} \cdot \mathbb{P}(\rho).$$

Notice that, because the pLTS may be infinitely branching, the set $\bigcup_{k\leq n_0} \mathsf{minF}_k$ may be infinite. We therefore define n_{max} such that $\mathbb{P}(\{\rho \in \bigcup_{k\leq n_0} \mathsf{minF}_k \mid n_\rho > n_{max}\}) \leq \varepsilon/4$. Thus, only a small portion of runs ρ in $\bigcup_{k\leq n_0} \mathsf{minF}_k$ have $n_\rho > n_{max}$. Then for $n \geq n_0 + n_{max}$ we have

$$\begin{aligned}
\mathbb{P}(\mathsf{FAmb}_n) \leq\ & \mathbb{P}(\mathsf{FAmb}_n \setminus \mathsf{F}_{n_0}) + \mathbb{P}(\mathsf{FAmb}_n \cap \mathsf{F}_{n_0}) \\
\leq\ & \mathbb{P}(\mathsf{FAmb}_n \setminus \mathsf{F}_{n_0}) + \mathbb{P}(\{\rho \in \bigcup_{k\leq n_0} \mathsf{minF}_k \mid n_\rho > n_{max}\}) \\
& + \mathbb{P}(\{\rho' \in \mathsf{FAmb}_n \mid \exists \rho \in \bigcup_{k\leq n_0} \mathsf{minF}_k,\ \rho \preceq \rho',\ n_\rho \leq n_{max}\}) \\
\leq\ & \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4}\mathbb{P}(\{\rho \in \bigcup_{k\leq n_0} \mathsf{minF}_k \mid n_\rho \leq n_{max}\}) \leq \varepsilon\ .
\end{aligned}$$

Let $\mathcal{A}$ be a FF-diagnosable pLTS. Consider $\rho \in \mathsf{minF}$ and $\alpha > 0$. There exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $\mathbb{P}(\mathsf{FAmb}_n) \leq \alpha \cdot \mathbb{P}(\rho)$. Thus for all $n \geq n_0$:

$$\mathbb{P}(\mathsf{Cyl}(\rho) \cap \mathsf{FAmb}_{n+|\rho|_o}) \leq \mathbb{P}(\mathsf{FAmb}_{n+|\rho|_o}) \leq \alpha \cdot \mathbb{P}(\rho).$$

IF $\not\Rightarrow$ FF. Consider the infinitely branching pLTS of Figure 7, where the probability of the transition from q_0 to q_{i1} is 2^{-i-1}. There is no infinite ambiguous sequence, so that it is IA-diagnosable, and thus IF-diagnosable. Yet, for all $n \in \mathbb{N}$, the observed sequence a^n is ambiguous and the signalling faulty run $q_0\mathbf{f}f_1a\cdots f_1$ of observable length n has probability $1/2$. Therefore, $\mathbb{P}(\mathsf{FAmb}_n) = 1/2$, so that the pLTS it is not FF-diagnosable.

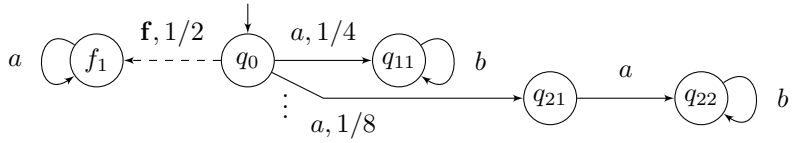


Figure 7: An infinitely branching pLTS that is IA-diagnosable but not FF-diagnosable.

IA $\not\Rightarrow$ FA. Consider the pLTS of Figure 2. Any infinite faulty run contains a b-event, that cannot appear in a correct run, therefore $\mathsf{FAmb}_\infty = \emptyset$. Both infinite correct runs have a^ω as observed sequence, and a^ω cannot be observed during

a faulty run, thus $\mathsf{CAmb}_\infty = \emptyset$. Thus this pLTS is IA-diagnosable. Consider now the infinite correct run $\rho = q_0 u q_1 a q_1 \ldots$. It has probability $1/2$, and all its finite signalling subruns are ambiguous since their observed sequence is a^n, for some $n \in \mathbb{N}$. Thus for all $n \geq 1$, $\mathbb{P}(\mathsf{CAmb}_n) \geq 1/2$, so that this pLTS is not FA-diagnosable.

uniform $\mathsf{AFF} \not\Rightarrow \mathsf{IF}$. Consider the pLTS depicted in Figure 3. All infinite faulty runs are ambiguous, and the probability of faulty runs is $1/2$, thus this pLTS is not IF. Fix some $\varepsilon > 0$ and α. There are two minimal faulty runs $\rho_a = q_0 \mathbf{f} q_f a q_f$ and $\rho_b = q_0 \mathbf{f} q_f b q_f$. Consider first ρ_a and let ρ be the random variable of a signalling run of length n that extends ρ_a. One can express the correctness proportion of ρ in terms of the number of a's in its observed sequence, written $|\rho|_a$:

$$\mathsf{CorP}(\rho) = \frac{(\frac{3}{4})^{|\rho|_a}(\frac{1}{4})^{|\rho|-|\rho|_a}}{(\frac{3}{4})^{|\rho|_a}(\frac{1}{4})^{|\rho|-|\rho|_a} + (\frac{1}{4})^{|\rho|_a}(\frac{3}{4})^{|\rho|-|\rho|_a}}$$

Simplifying this expression, we obtain: $\mathsf{CorP}(\rho) = \frac{1}{1+3^{|\rho|-2|\rho|_a}}$. Now, by the strong law of large numbers, for any $\eta > 0$, there exists n_η such that for every $n \geq n_\eta$, $\mathbb{P}(|4|\rho|_a - |\rho|| > \eta) < \alpha$. So with probability at least $1-\alpha$, the correctness proportion of ρ is bounded by $\frac{1}{1+3^{\frac{\eta+|\rho|}{2}}}$. We can now fix η as a function of ε, so that $\mathbb{P}(\mathsf{CorP}(\rho) \leq \varepsilon) \geq 1-\alpha$.

A similar reasoning applies to ρ_b, and one can then take the maximum of the two integers n_η to prove that the pLTS is uniformly AFF-diagnosable.

$\mathsf{AFF} \not\Rightarrow$ uniform $\varepsilon\mathsf{FF}$. Consider the pLTS of Figure 4. Fix some $0 < \varepsilon < 3/4$, $0 < \alpha < 1$ and n_α. Consider the minimal faulty run $\rho = q_0 u q_1 (a q_1)^{n_\alpha+1} \mathbf{f} q_f b q_f$. Let ρ' be the signalling run of length $2n_\alpha + 2$ such that $\rho \preceq \rho'$. $\pi(\rho') = a^{n_\alpha+1} b^{n_\alpha+1}$. Thus by examination of the pLTS, $\mathsf{CorP}(\pi(\rho')) \geq 3/4$. So

$$\mathbb{P}(\mathsf{Cyl}(\rho) \cap \mathsf{FAmb}^{\varepsilon}_{2n_\alpha+2}) = \mathbb{P}(\rho) > \alpha \cdot \mathbb{P}(\rho).$$

Thus the pLTS is not uniformly $\varepsilon\mathsf{FF}$-diagnosable.

Let ρ be a minimal faulty run. Then $\pi(\rho) = a^{n_0} b$ for some n_0. For all n, let ρ_n be the single the single signalling run of observable length $|\rho|+n$ that extends ρ. It fulfils $\pi(\rho_n) = a^{n_0} b^{n+1}$ and $\mathbb{P}(\rho_n) = \mathbb{P}(\rho)$. The single correct signalling run ρ'_n with $\pi(\rho'_n) = \pi(\rho_n)$ fulfils $\mathbb{P}(\rho'_n) = \frac{3^{n_0}}{2 \cdot 4^{n_0+n+1}}$. Thus $\lim_{n\to\infty} \mathsf{CorP}(\pi(\rho_n)) = 0$. So the pLTS is $\varepsilon\mathsf{FF}$-diagnosable for all $\varepsilon > 0$ and thus AFF-diagnosable. □

3. Characterisations of diagnosability

The goal of this section is to provide characterisations of the different diagnosability notions we introduced. Having algorithmic developments in mind, we focus here on finite-state pLTS. As a consequence, FF-diagnosability and IF-diagnosability coincide, and we consider only IF-diagnosability in the sequel.

For all the exact diagnosability notions, the methodology is similar. We first construct an *ad hoc* deterministic automaton which gathers all the information needed for the diagnosis, by tracking possible correct and faulty executions.

Second, we build the product of the original pLTS with this deterministic automaton, to recover the probabilistic behaviour. Diagnosability can then be characterised on the product by graph-based properties.

As for approximate diagnosability, we show that the diagnosability notions can be characterised relying on the distance 1 problem for labelled Markov chains. This problem, shown to be decidable in PTIME [16] asks for the existence of an event, that is almost sure in one input Markov chain, and has null measure in the other. Labelled Markov chains, the distance 1 problem, and our characterisation are detailed in Subsection 3.2.

3.1. Exact diagnosis

For each notion of exact 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 synchronised 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 diagnosis. Finally, we characterise diagnosability by graph properties on the synchronised product.

The deterministic automata we will build are variants of the deterministic Büchi automaton introduced in [18], that accepts the unambiguous observed sequences. The latter tracks the subsets of possible states reached by signalling runs associated with an observed sequence. It resembles the on-the-fly determinisation of $\mathcal{A}$ viewing unobserved events as silent transitions. However, in view of the forthcoming characterisations, the subsets of correct and faulty states are divided in three sets: U, V and W. A state q belongs to U, if there is a correct signalling run with the current observed sequence ending in q. A state q belongs to $V \cup W$ if there is a faulty signalling run with the current observed sequence ending in q. The partition between V and W ensures that for all $q \in V$, $q' \in W$ and ρ a faulty run ending in q, there exists a faulty run ρ' ending in q' with an earlier fault than the first fault of ρ'. The decomposition between V and W reflects the fact that the automaton tries to resolve the ambiguity between U and W (when both are not empty), while V corresponds to a waiting room of states reached by faulty runs that will be examined when the current ambiguity is resolved. Note also that this construction only considers signalling runs. To simplify the definition of the deterministic automaton, given a set of states S and an observation a, we define $E(a, S) = \{\rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \mid \rho \text{ is a run of } \mathcal{A}, q_{\alpha_0} \in S, \forall i < k\ a_i \in \Sigma_u, a_k = a\}$ the set of signalling run starting in a state of S with observation a. We also define for two sets of states U and V and a set of paths E the set of state $update faulty(U, V, E) = \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ run of } E, q_{\alpha_0} \in V, q_{\alpha_k} = q\} \cup \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k} \text{ faulty run of } E, q_{\alpha_0} \in U, q_{\alpha_k} = q\}$ which contains the states reached by a path of E from V and those reached from U by a faulty run of E. Formally, the states and transitions of the deterministic Büchi automaton $\mathsf{Obs}(\mathcal{A})$ are inductively defined by:

- $s_0 = (\{q_0\}, \emptyset, \emptyset)$ is the initial state of $\mathsf{Obs}(\mathcal{A})$;

- Given (U, V, W) a state of $\mathsf{Obs}(\mathcal{A})$ and $a \in \Sigma_o$, there is a transition $(U, V, W) \xrightarrow{a} (U', V', W')$ as soon as:
  1. $E(a, U \cup V \cup W) \neq \emptyset$,
  2. $U' = \{q \mid \exists \rho = q_{\alpha_0} a_1 \dots a_k q_{\alpha_k}$ correct run of $E(a, U), q_{\alpha_0} \in U, q_{\alpha_k} = q\}$,
  3. If $W = \emptyset$ then $V' = \emptyset$ and $W' = update faulty(U, V, E(a, U \cup V))$,
  4. If $W \neq \emptyset$ then $W' = update faulty(\emptyset, W, E(a, W))$ and $V' = update faulty(U, V, E(a, U \cup V)) \setminus W'$.

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, since U tracks the possible states after a correct run. When $W = \emptyset$ the current signalling run may be ambiguous (if $V \neq \emptyset$) but the "oldest" possible faulty runs under scrutiny have been discarded. Hence, any infinite observed sequence of $\mathcal{A}$ passing infinitely often through F is not ambiguous (either it is surely faulty, or ambiguities are resolved one after another).

The next proposition recalls the property of this automaton.

Proposition 1 ([18])**.** *Let $\mathcal{A}$ be a finite pLTS. Then the deterministic Büchi automaton $\mathsf{Obs}(\mathcal{A})$ accepts the infinite unambiguous observed sequences of $\mathcal{A}$.*

3.1.1. IF*-diagnosability*

As explained earlier, for each diagnosability notion, we will consider a variant of $\mathsf{Obs}(\mathcal{A})$. For IF-diagnosability, we can omit the faulty sets of states V and W. We write $\mathsf{IF}(\mathcal{A})$ for the resulting simplified automaton, obtained from $\mathsf{Obs}(\mathcal{A})$ by only considering the U-component of states.

Figure 8 illustrates this construction on the pLTS of Figure 2. This automaton reflects that, after observing at least one a, and as long as b is not observed, the current signalling run is surely correct leading to either q_1 or q_2 (state s_1), and once b happens, the current signalling run is surely faulty, thus the set of possible correct states is empty (state s_2).

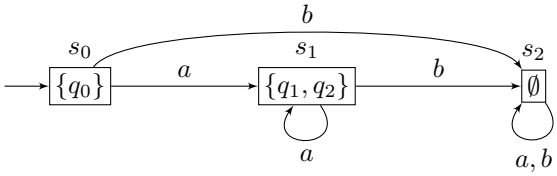


Figure 8: The IF-automaton of pLTS of Figure 2.

To recover the stochastic behaviour of $\mathcal{A}$ which is not reflected in $\mathsf{IF}(\mathcal{A})$, we now define the pLTS $\mathcal{A}_{\mathsf{IF}} = \mathcal{A} \times \mathsf{IF}(\mathcal{A})$ as the product of $\mathcal{A}$ and $\mathsf{IF}(\mathcal{A})$ synchronised over observed events. Since $\mathsf{IF}(\mathcal{A})$ is deterministic and complete, $\mathcal{A}_{\mathsf{IF}}$ is still a

pLTS, with the same stochastic behaviour as $\mathcal{A}$. In addition, the U-component of a state (q, U) of $\mathcal{A}_{\mathsf{IF}}$ stores the relevant information w.r.t IF-diagnosability of the observed sequence so far.

Carrying on with the example pLTS of Figure 2, Figure 9 shows the resulting product pLTS. Observe that it has two bottom strongly connected components (BSCC), each consisting of one of the absorbing states (q_1, s_1) and (f_2, s_2).

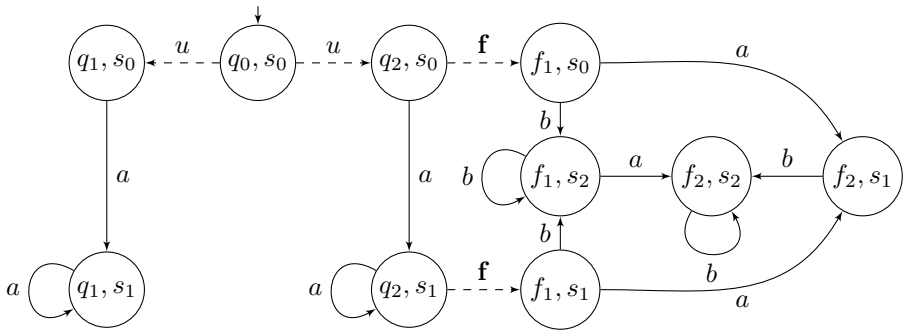


Figure 9: The synchronised product of pLTS of Figure 2 and its IF-automaton.

In finite DTMC every run almost surely ends in a BSCC, and IF-diagnosability is concerned with (faulty) ambiguous infinite runs. Unsurprisingly, our characterisation of IF-diagnosability is thus based on the BSCC of $\mathcal{A}_{\mathsf{IF}}$.

Proposition 2. *Let $\mathcal{A}$ be a finite pLTS. Then $\mathcal{A}$ is IF-diagnosable if and only if $\mathcal{A}_{\mathsf{IF}}$ has no BSCC containing a state (q, U) with $q \in Q_f$ and $U \neq \emptyset$.*

Proof. Suppose first that there exists a reachable BSCC C of $\mathcal{A}_{\mathsf{IF}}$ and a state $s = (q, U)$ in C such that $q \in Q_f$ and $U \neq \emptyset$. Let ρ be a signalling run leading from the initial state s_0 of $\mathcal{A}_{\mathsf{IF}}$ to s. Now, for every state $s' = (q', U') \in C$, necessarily $q' \in Q_f$ and $U' \neq \emptyset$, because C is strongly connected. So for every signalling run ρ' that extends ρ, writing $s' = (q', U')$ for the state ρ' leads to, there exists a correct signalling run ρ'' such that $\pi(\rho'') = \pi(\rho')$ and $q_0 \xrightarrow{\rho''} q''$ with $q'' \in U'$. As a consequence the observed sequence $\pi(\rho'')$ is ambiguous, and for every $n \geq |\rho|_o$, $\mathbb{P}(\mathsf{FAmb}_n) \geq \mathbb{P}(\rho)$, so that $\mathcal{A}$ is not IF-diagnosable.

Suppose now that for every state $s = (q, U)$ of a BSCC C, either $q \in Q_c$, or $U = \emptyset$. This property is in fact uniform by BSCC: for every BSCC C, either for every state $(q, U) \in C$, $q \in Q_c$, or, for every state $(q, U) \in C$, $U = \emptyset$. This is a straightforward consequence of C being strongly connected. Moreover, if a run ρ reaches a pair (q, U) then $q \in Q_c$ implies $U \neq \emptyset$. Indeed, let ρ' be the greatest signalling run prefix of ρ. ρ' ends in a pair (q', U') where $U' = U$ as $\pi(\rho) = \pi(\rho')$. Moreover if $q \in Q_c$, then $q' \in Q_c$, therefore $q' \in U$ implying that $U \neq \emptyset$. Therefore in $\mathcal{A}_{\mathsf{IF}}$ the BSCC are partitioned in non-faulty ones, in case all q-components of states in C are non-faulty, and faulty ones, in case all U-components of states in C are empty ensuring unambiguity of faulty runs ending

in a BSCC. Thus an infinite faulty ambiguous run must only visit transient states. Since almost surely runs leave the transient states and reach a BSCC, this implies that $\mathbb{P}(\mathsf{FAmb}_\infty) = 0$. □

3.1.2. FA-*diagnosability*

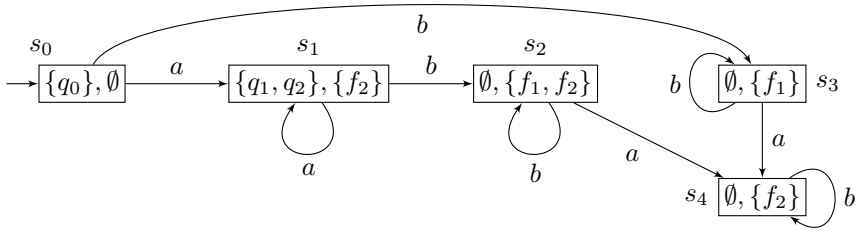


Figure 10: The FA-automaton of pLTS of Figure 2.

For FA-diagnosability, we again start from $\mathsf{Obs}(\mathcal{A})$ and gather the V and W components into a unique set, that we again call V. The resulting simplified automaton is denoted by $\mathsf{FA}(\mathcal{A})$.

Figure 10 illustrates this construction on the pLTS of Figure 2. As expected, the FA-automaton is a refinement of the IF-automaton: the U-component of a state in $\mathsf{FA}(\mathcal{A})$ corresponds to a state in $\mathsf{IF}(\mathcal{A})$. For instance, state s_2 of Figure 8 is split here into s_2, s_3 and s_4.

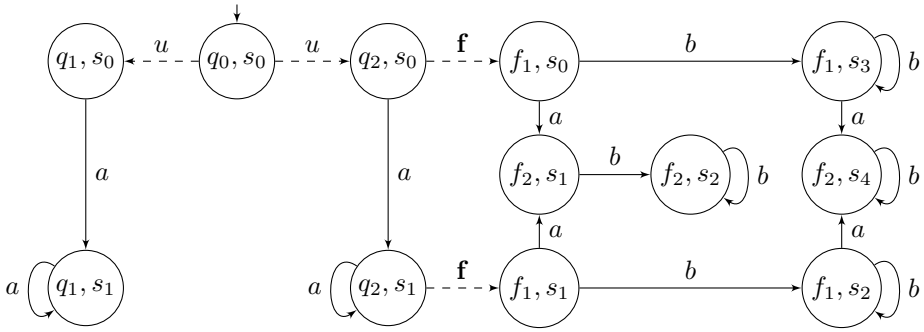


Figure 11: The synchronised product of pLTS of Figure 2 and its FA-automaton.

We now define the pLTS $\mathcal{A}_{\mathsf{FA}} = \mathcal{A} \times \mathsf{FA}(\mathcal{A})$ as the product of $\mathcal{A}$ and $\mathsf{FA}(\mathcal{A})$ synchronised over observed events. $\mathcal{A}_{\mathsf{FA}}$ is still a pLTS with same stochastic behaviour as $\mathcal{A}$ augmented with the relevant information of the observed sequence w.r.t FA-diagnosability. Figure 11 continues our example and shows the synchronised product for the pLTS of Figure 2.

Again, FA-diagnosability is characterised through the BSCC of $\mathcal{A}_{\mathsf{FA}}$.

Proposition 3. *Let $\mathcal{A}$ be a finite pLTS. $\mathcal{A}$ is* FA*-diagnosable if and only if $\mathcal{A}_{\mathsf{FA}}$ has no BSCC that:*

- *either contains a state (q, U, V) with $q \in Q_f$ and $U \neq \emptyset$;*
- *or contains a state (q, U, V) with $q \in Q_c$ and $V \neq \emptyset$.*

Note that the characterisation of FA-diagnosability is symmetric for correct states and V-component (resp. faulty states and U-component). This reflects that the definition of FA-diagnosability itself is symmetric.

3.1.3. IA*-diagnosability*

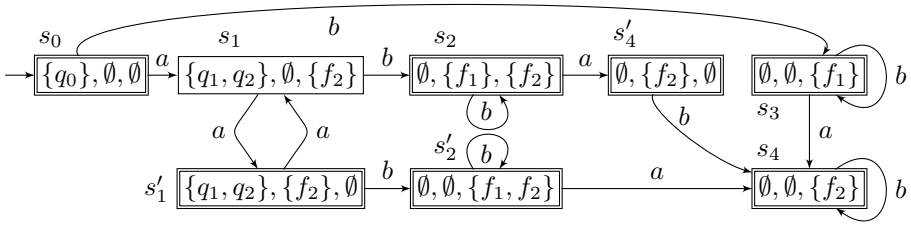


Figure 12: The IA-automaton of pLTS of Figure 2.

For IA-diagnosability, we use $\mathsf{Obs}(\mathcal{A})$ with no transformation, however, to stick to the presentation for the other diagnosability notions, we write here $\mathsf{IA}(\mathcal{A})$ for $\mathsf{Obs}(\mathcal{A})$.

Figure 12 shows the IA-automaton of the pLTS depicted in Figure 2, where accepting states for the Büchi condition are doubly framed. Observe that the state s_1 of its FA-automaton (see Figure 10) has been split into two states s_1 and s'_1, and s'_1 is an accepting state for $\mathsf{Obs}(\mathcal{A})$. The infinite observed sequence a^ω, which is indeed unambiguous, is thus accepted.

As before, to come up with a characterisation, one builds $\mathcal{A}_{\mathsf{IA}} = \mathcal{A} \times \mathsf{IA}(\mathcal{A})$, the product of $\mathcal{A}$ and $\mathsf{IA}(\mathcal{A})$ synchronised over observed events. Figure 13 shows the synchronised product corresponding to the pLTS depicted in Figure 2. Among the BSCC, all the faulty ones (*i.e.* the ones reached after a faulty event) have $U = \emptyset$, while $\{(q_1, s_1), (q_1, s'_1)\}$, the single one that is reached by a correct run, has a state (q_1, s'_1) with $W = \emptyset$.

Using Proposition 1 we get the following:

Proposition 4. *Let $\mathcal{A}$ be a finite pLTS. $\mathcal{A}$ is* IA*-diagnosable if and only if $\mathcal{A}_{\mathsf{IA}}$ has no BSCC such that:*

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

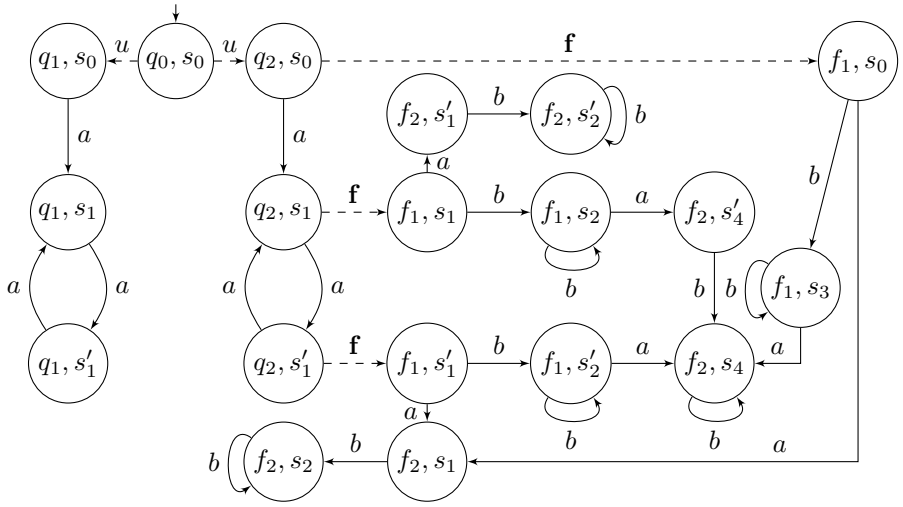


Figure 13: The synchronised product of pLTS of Figure 2 and its IA-automaton.

3.2. Approximate diagnosis

We now turn to the characterisation of approximate diagnosis and particularly of AFF-diagnosability. The reason why we only consider AFF-diagnosability here will become clear in Subsection 6.1 where we show that all other approximate diagnosability notions are undecidable. Our characterisation of AFF-diagnosability uses the notion of distance between two Markov chains with labels on the transitions. A *labelled Markov chain* (LMC) is a pLTS where every event is observable: $\Sigma = \Sigma_o$. In order to exploit results of [16] on LMC in our context of pLTS, we introduce the mapping $\mathcal{M}$ that computes *in polynomial time* the probabilistic closure of a pLTS w.r.t. the unobservable events and produces an LMC. Informally, the probabilities of all paths of $\mathcal{A}$ from state q to state q' with same observed sequence $a \in \Sigma_o$ are gathered to obtain the probability in $\mathcal{M}(\mathcal{A})$ to move from q to q' with label a. The transformation is formally defined below. For sake of simplicity, we denote by $\mathcal{A}_q$, the pLTS $\mathcal{A}$ where the initial state has been substituted by q.

Definition 5. 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:

- $T' = \{(q, a, q') \mid \exists \rho \in \mathsf{SR}_1(\mathcal{A}_q)\ \rho = q \cdots a q'\}$ (and so a is observable).
- for every $(q, a, q') \in T'$, $\mathbf{P}'(q, a, q') = \mathbb{P}(\{\rho \in \mathsf{SR}_1(\mathcal{A}_q) \mid \rho = q \cdots a q'\})$.

Let E be an *event* of Σ_o^ω (*i.e.* a measurable subset of Σ_o^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 $\mathcal{M}_2$ generalises 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 between $\mathcal{M}_1$ and $\mathcal{M}_2$ is defined as the supremum over the events:

Definition 6. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two LMC over the same alphabet Σ_o. Then $d(\mathcal{M}_1, \mathcal{M}_2)$ the *distance between* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ is:

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

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

Proposition 5 ([16])**.**

- *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 for LMC is decidable in polynomial time.*

Let us first explain how to characterise AFF-diagnosability 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 Figure 14). Moreover no faulty transitions occur in $\mathcal{A}^c$.

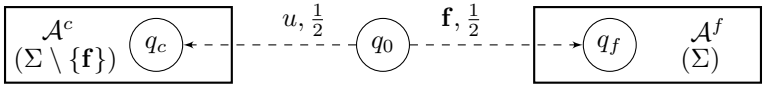


Figure 14: Initial-fault pLTS.

Definition 7 (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 every $t \in T_f$, $\mathbf{P}(t) = \mathbf{P}_f(t)$, for every $t \in T_c$, $\mathbf{P}(t) = \mathbf{P}_c(t)$, and for every $t \in T \setminus (T_c \cup T_f)$ $\mathbf{P}(t) = 1/2$.

We denote such a pLTS by $\mathcal{A} = \langle q_0, \mathcal{A}^f, \mathcal{A}^c\rangle$.

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

Lemma 2. *Let $\mathcal{A} = \langle q_0, \mathcal{A}^f, \mathcal{A}^c \rangle$ be an initial-fault pLTS. Then $\mathcal{A}$ is* AFF*-diagnosable if and only if $d(\mathcal{M}(\mathcal{A}^f), \mathcal{M}(\mathcal{A}^c)) = 1$.*

In order to understand why characterising AFF-diagnosability for general pLTS is more involved, consider again the pLTS $\mathcal{A}_2$ presented in Figure 3, on page 8, where outgoing transitions of any state are uniform. Recall that $\mathcal{A}_2$ is AFF-diagnosable (and even uniformly AFF-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, $d(\mathcal{M}(\mathcal{A}_{q_0}), \mathcal{M}(\mathcal{A}_{q_f})) \leq 1/2$ since for any event E either (1) $a^\omega \in E$ implying $\mathbb{P}^{\mathcal{M}(\mathcal{A}_{q_f})}(E) = 1$ and $\mathbb{P}^{\mathcal{M}(\mathcal{A}_{q_0})}(E) \geq 1/2$ or (2) $a^\omega \notin E$ implying $\mathbb{P}^{\mathcal{M}(\mathcal{A}_{q_f})}(E) = 0$ and $\mathbb{P}^{\mathcal{M}(\mathcal{A}_{q_0})}(E) \leq 1/2$. 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 pair (q_0, q_f) is irrelevant, since the correct state q_0 does not belong to a bottom strongly connected component (BSCC) of the pLTS, while (q_c, q_f) is relevant since q_c belongs to a BSCC triggering a "recurrent" ambiguity. The next theorem characterises AFF-diagnosability, establishing the soundness of this intuition.

Theorem 2. *Let $\mathcal{A}$ be a pLTS. Then, $\mathcal{A}$ is* AFF*-diagnosable if and only if for every correct state q_c belonging to a BSCC and every faulty state q_f reachable by a faulty run ρ_f such that q_c is reachable by a run with same observed sequence, $d(\mathcal{M}(\mathcal{A}_{q_c}), \mathcal{M}(\mathcal{A}_{q_f})) = 1$.*

The proof of Theorem 2 is given in appendix. Let us sketch the key ideas to establish the characterization of AFF–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 2 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 AFF–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 $\mathcal{A}$ AFF–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 2 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.

As an alternative to the proof of Theorem 2, one could mimic the approach by Kiefer and Sistla [15] for monitorability. The idea would be, from a pLTS $\mathcal{A}$ to derive two hidden Markov chains, say $\mathcal{H}_c$ and $\mathcal{H}_f$ representing respectively the observation sequences for correct and faulty runs of $\mathcal{A}$. However, to establish that distinguishability of $\mathcal{H}_c$ and $\mathcal{H}_f$ corresponds to AFF-diagnosability essentially relies on the same arguments we used in the above proof (and so this alternative approach would not simplify it). The difficulty lies in that the events one conditions by to obtain $\mathcal{H}_c$ and $\mathcal{H}_f$, namely always correct or eventually faulty, anticipate on the future behaviour of the system; in contrast, the correctness proportion appearing in the definition of AFF-diagnosability only speaks about the possible behaviours up to the last observation.

4. Decidability results

We first consider exact diagnosability notions, and establish, thanks to the characterisations of the previous section, that they can all be solved in PSPACE. In all cases, to obtain the PSPACE upper-bound, we avoid building explicitly the exponential size product pLTS (that are used in the characterisations) and only explore it on-the-fly.

Proposition 6. *The* IF*-diagnosability problem is decidable in* PSPACE*.*

Proof. We use the characterisation of IF-diagnosability given in Proposition 2. To obtain a PSPACE algorithm, we avoid building explicitly the product pLTS $\mathcal{A}_{\mathsf{IF}}$, which is exponential in the size of $\mathcal{A}$. Given two states s, s' of $\mathcal{A}_{\mathsf{IF}}$, one can check in polynomial space in the size of $\mathcal{A}$ whether s' can be reached from s. Using this procedure, we can check whether a state s is not in a BSCC by guessing another state s' such that s' is reachable from s but s is not reachable from s'. Here we use Savitch's theorem.
Thus the procedure that decides whether $\mathcal{A}$ is not IF-diagnosable consists in guessing a state $s = (q, U)$ with $q \in Q_f$ and $U \neq \emptyset$, checking that it is reachable from s_0 and whether s is in a BSCC (here again, we use Savitch's theorem). □

We state below similar results for FA- and IA-diagnosability problems.

Proposition 7. *The* FA- *and* IA*-diagnosability problems are decidable in* PSPACE*.*

For approximate diagnosability, we concentrate on AFF-diagnosability and establish a complexity upper-bound, relying on the characterisation from the previous section.

Theorem 3. *The* AFF*-diagnosability problem is decidable in* PTIME *for pLTS.*

Proof. The decidability and complexity results rely on the characterisation of AFF-diagnosability, see Theorem 2. Reachability of a pair of states with the

same observed sequence is decidable in polynomial time by an appropriate "self-synchronised 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 Chen and Kiefer (as recalled in Proposition 5), the decidability and PTIME upper-bound follow. □

5. Diagnoser construction

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

For implementation considerations, we introduce finite memory diagnosers. A finite memory diagnoser is given by a tuple $(M, \Sigma, m_0, \mathsf{up}, D_{fm})$ 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: *correctness* and *reactivity*. Correctness ensures that the information provided is accurate and reactivity specifies which pieces of information the diagnoser must provide. The precise correctness and reactivity requirements will depend on the considered diagnosability notions. An exact diagnoser has an additional property, *commitment*, meaning that when it announces a fault it will announce a fault forever.

5.1. IF*-diagnoser*

We start with IF-diagnosers of pLTS. These diagnosers only provide information about faulty runs. In the sequel we fix $\mathcal{A}$ a finite pLTS. Given $w \in \Sigma_o^\omega$, $w_{\leq n}$ denotes the prefix of w with length n.

Definition 8. An IF*-diagnoser* for $\mathcal{A}$ is a function $D : \Sigma_o^* \to \{\top, ?\}$ such that:

commitment. For all $w \preceq w' \in \Sigma_o^*$, if $D(w) = \top$ then $D(w') = \top$.

correctness. For all $w \in \Sigma_o^*$, if $D(w) = \top$ then w is surely faulty.

reactivity. For every minimal finite faulty run ρ,
$\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\pi(\rho')) = ?\}) = 0$
where for $w \in \Sigma_o^\omega$, $D(w) = \lim_{n\to\infty} D(w_{\leq n})$.

In the above definition, due to commitment the limit is well-defined.

Proposition 8. *A finite pLTS is* IF*-diagnosable if and only if it admits an* IF*-diagnoser.*

Proof. Let $\mathcal{A}$ be a pLTS, and assume there exists an IF-diagnoser D for $\mathcal{A}$. Let $\varepsilon > 0$. Using Lemma 1, first select n_0 such that for all $n \geq n_0$

$$\mathbb{P}(\mathsf{FAmb}_n \Delta \mathsf{FAmb}_\infty) < \varepsilon \ ,$$

where Δ stands for the symmetric difference.
So for $n \geq n_0$ one has $\mathbb{P}(\mathsf{FAmb}_n \Delta \mathsf{FAmb}_{n_0}) < 2\varepsilon$, hence $\mathbb{P}(FAmb_n \setminus FAmb_{n_0}) < 2\varepsilon$.
Since D is correct, for any minimal faulty run $\rho \in \mathsf{SR}_{n_0}$, and for any $n \geq n_0$

$$\{\rho' \in \mathsf{FAmb}_n \mid \rho \preceq \rho'\} \subseteq \{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge D(\pi(\rho')) =?\} \ .$$

By the reactivity condition, D almost surely detects faults, and because the number of signalling runs of fixed observable length is finite (since $\mathcal{A}$ is convergent by hypothesis), there exists $N \in \mathbb{N}$ such that for every $n \geq N + n_0$ and every minimal faulty run $\rho \in \cup_{m \leq n_0} \mathsf{minF}_m$,

$$\mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge D(\pi(\rho')) =?\}) < \varepsilon \cdot \mathbb{P}(\rho) \ .$$

Thus for every $n \geq N + n_0$,

$$\mathbb{P}(\mathsf{FAmb}_n \cap \mathsf{FAmb}_{n_0}) = \mathbb{P}\Big(\biguplus_{\rho \in \mathsf{FAmb}_{n_0}} \{\rho' \in \mathsf{FAmb}_n, \rho \preceq \rho'\} \Big) =$$

$$\sum_{\rho \in \cup_{m \leq n_0} \mathsf{minF}_m} \mathbb{P}(\{\rho' \in \mathsf{FAmb}_n, \rho \preceq \rho'\}) < \varepsilon \sum_{\rho \in \cup_{m \leq n_0} \mathsf{minF}_m} \mathbb{P}(\rho) = \varepsilon \mathbb{P}(\mathsf{F}_{n_0}) \leq \varepsilon \ .$$

Thus $\mathbb{P}(\mathsf{FAmb}_n) < 3\varepsilon$ for every $n \geq N + n_0$, which proves that $\mathcal{A}$ is IF-diagnosable.

Assume that $\mathcal{A}$ is IF-diagnosable. We define the function $D : \Sigma_o^* \to \{\top, ?\}$ by $D(w) = \top$ if and only if w is a surely faulty observed sequence. Let us check that D is an IF-diagnoser. By definition, D fulfils the commitment property. Since $D(w) = \top$ iff w is a surely faulty sequence, D is correct. Now, let ρ be a minimal faulty run.

$$\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\pi(\rho')) =?\}) = \lim_{n \to \infty} \mathbb{P}(\{\rho' \in \mathsf{FAmb}_{n+|\rho|_o} \mid \rho \preceq \rho'\}) \ .$$

For every $n \in \mathbb{N}$, we have $\{\rho' \in \mathsf{FAmb}_{n+|\rho|_o} \mid \rho \preceq \rho'\} \subseteq \mathsf{FAmb}_{n+|\rho|_o}$ and $\lim_{n \to \infty} \mathbb{P}(\mathsf{FAmb}_n) = 0$. Therefore $\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\pi(\rho')) =?\}) = 0$ and D is reactive. □

We now study the size of IF-diagnosers.

Proposition 9. *If $\mathcal{A}$ is an* IF*-diagnosable pLTS with n correct states, one can build an* IF*-diagnoser with at most 2^n memory states where $n = |Q_c|$.*

Proof. For an IF-diagnosable pLTS $\mathcal{A}$ with $\mathsf{IF}(\mathcal{A}) = (Q^*, \Sigma_o, T^*, \{q_0\})$ its deterministic and complete IF-automaton, we define the finite memory diagnoser $(M, \Sigma, \mathsf{up}, m_0, D_{fm})$ with $M = Q^*$ and $m_0 = \{q_0\}$, $\mathsf{up}(q, a) = T^*(q, a)$ and $D_{fm}(U) = \top$ iff $U = \emptyset$. Let us show that the induced diagnoser D is indeed an IF-diagnoser, and that it has at most 2^n memory states, where n is the number of correct states of $\mathcal{A}$.

commitment When U is empty it remains empty forever which implies commitment.

correctness When D outputs the verdict $\top$, $\mathsf{IF}(\mathcal{A})$ is in the state associated with $\emptyset$. Thus the observed sequence is surely faulty.

reactivity If an infinite faulty run ρ is such that $D(\pi(\rho)) =?$ then, by construction of $\mathsf{IF}(\mathcal{A})$ and definition of D, for every length $n \in \mathbb{N}$, there exists a finite correct signalling run $\rho_n \in \mathsf{SR}_n$ such that $\pi(\rho_n) = \pi(\rho_{\downarrow n})$. Using König's lemma, since $\mathcal{A}$ is finitely branching, one can extract an infinite correct run ρ_∞ such that $\pi(\rho_\infty) = \pi(\rho)$, so that $\rho \in \mathsf{FAmb}_\infty$. Moreover $\mathbb{P}(\mathsf{FAmb}_\infty) = 0$ as $\mathcal{A}$ is IF-diagnosable. Putting everything together, for every minimal faulty run ρ, $\mathbb{P}(\{\rho' \in \Omega \mid \rho \preceq \rho' \wedge D(\pi(\rho')) =?\}) = 0$.

size The memory states are states of $\mathsf{IF}(\mathcal{A})$, which are themselves subsets of correct states of $\mathcal{A}$. Therefore, D uses at most 2^n memory states, with $n = |Q_c|$.

□

We show that the size order of the previous IF-diagnoser is optimal.

Proposition 10. *There is a family $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ of IF-diagnosable pLTS such that $\mathcal{A}_n$ has $n + 1$ correct states and it admits no IF-diagnoser with less than 2^n states.*

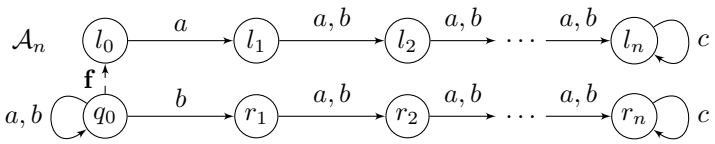


Figure 15: An IF-diagnosable pLTS requiring an IF-diagnoser with exponential size.

Proof. Consider the example of Figure 15 where $\Sigma_o = \{a, b, c\}$ and the initial state is q_0. Consider a finite faulty run including a c event. Its observed sequence belongs to $\mathcal{L} = \{a, b\}^* b \{a, b\}^{n-1} c^+$. Since any finite correct run has an observed sequence belonging to $\mathcal{L}' = \{a, b\}^* \cup \{a, b\}^* a \{a, b\}^{n-1} c^+$ and $\mathcal{L} \cap \mathcal{L}' = \emptyset$, $\mathsf{FAmb}_n \uplus$

$\mathsf{CAmb}_n \subseteq \{\rho \mid \pi(\rho) \in \{a,b\}^n\}$. Since $\lim_{n\to\infty} \mathbb{P}(\{\rho \mid \pi(\rho) \in \{a,b\}^n\}) = 0$, the pLTS is FA-diagnosable and so IA-diagnosable and IF-diagnosable.
Intuitively, when a c is observed, any IF-diagnoser must have remembered the observable event that happened n steps earlier to know if the run is faulty or not. Thus, it must remember the last n observed events, in case a c event occurs. More formally, assume there exists a diagnoser $D = (M, \Sigma, m_0, \mathsf{up}, D_{fm})$ with less than 2^n memory states. Then there exist two distinct words $w_1 \in \{a,b\}^n$ and $w_2 \in \{a,b\}^n$ leading to the same memory state: $\mathsf{up}(m_0, w_1) = \mathsf{up}(m_0, w_2)$. The words w_1 and w_2 differ at least from one letter say $w_1[i] = a$ and $w_2[i] = b$. Consider for $k \geq 1$, the signalling correct run $\rho_{1,k}$ corresponding to observed sequence $w_1 a^{i-1} c^k$ whose sequence of visited states is $q_0^i r_1 \dots r_n^{k+1}$ and the signalling faulty run $\rho_{2,k}$ corresponding to observed sequence $w_2 a^{i-1} c^k$ whose sequence of visited states is $q_0^i l_0 l_1 \dots l_n^{k+1}$. They also lead to the same memory state. By correctness, $D(w_1 a^{i-1} c^k) =?$. Thus for all suffix ρ of $\rho_{2,1}$, $D(\rho) =?$ contradicting the reactivity of D. □

5.2. FA*-diagnoser*

FA-diagnosability and IA-diagnosability not only consider the diagnosis of faults but also of correct runs. Contrary to IF-diagnosers, FA- and IA-diagnosers have three possible verdicts $\top$, related to faulty sequences, $\bot$, linked with correctness, and ? when no information can be derived from the observation.

Definition 9. An FA-diagnoser for $\mathcal{A}$ is a function $D : \Sigma_o^* \to \{\top, \bot, ?\}$ such that:

commitment. For all $w \preceq w' \in \Sigma_o^*$, if $D(w) = \top$ then $D(w') = \top$.

correctness. For every $w \in \Sigma_o^*$,

- if $D(w) = \top$ then w is surely faulty;
- if $D(w) = \bot$ then w is surely correct.

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

Proposition 11. *A finite pLTS $\mathcal{A}$ is* FA*-diagnosable if and only if it admits an* FA*-diagnoser. Furthermore when $\mathcal{A}$ is* FA*-diagnosable, one can build an* FA*-diagnoser with at most 2^n memory states.*

As the pLTS of Figure 15 is FA-diagnosable, and since any FA-diagnoser is also an A-diagnoser, using Proposition 10 we obtain the following lower bound for the size of FA-diagnosers.

Proposition 12. *There is a family $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ of* FA*-diagnosable pLTS such that $\mathcal{A}_n$ has $2n+2$ states and it admits no* FA*-diagnoser with less than 2^n memory states.*

5.3. IA*-diagnoser*

We now introduce IA-diagnosers, that mostly differ from FA-diagnosers on the correctness requirement. Intuitively, IA-diagnosers provide a weaker information about correct runs than the one of FA-diagnosers.

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

commitment. For all $w \preceq w' \in \Sigma_o^*$, if $D(w) = \top$ then $D(w') = \top$.

correctness. 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 $\pi(\rho) = w$, $\rho_{\downarrow |D(w)|_\bot}$ is correct.

reactivity. $\mathbb{P}(\{\rho \in \Omega \mid D_{\sup}(\pi(\rho)) =?\}) = 0$
where for $w \in \Sigma_o^\omega$, $D_{\sup}(w) = \limsup_{n\to\infty} D(w_{\leq n})$.
(due to commitment, $D_{\sup}$ *is well-defined.)*

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 16 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. Let D be defined by: for $w \in \{a, b\}^*(ab \cup aa)$, $D(w) = \bot$, for $w \in \{a, b, c\}^*c$, $D(w) = \top$ and otherwise $D(w) =?$. Then D is an IA-diagnoser.

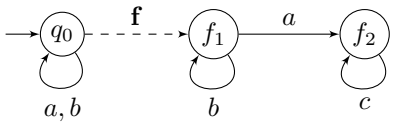


Figure 16: A pLTS which is IA-diagnosable.

Proposition 13. *A finite pLTS* $\mathcal{A}$ *is* IA*-diagnosable if and only if it admits an* IA*-diagnoser. Furthermore when* $\mathcal{A}$ *has* n_c *correct states,* n_f *faulty states and is* IA*-diagnosable, one can build an* IA*-diagnoser with at most* $2^{n_c}3^{n_f}$ *states.*

The following lower bound can be derived from the proof of Proposition 10, since the pLTS of Figure 15 is IA-diagnosable, and because any IA-diagnoser is also an IF-diagnoser.

Proposition 14. *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.*

5.4. εFF-diagnoser

Given a fixed threshold $\varepsilon > 0$, an εFF-diagnoser monitors the sequence of observed events w, computes the current correctness proportion, and outputs $\top$ if $\mathsf{CorP}(w)$ is below ε.

Definition 11. Let $\varepsilon > 0$. An εFF-diagnoser for $\mathcal{A}$ is a function $D : \Sigma_o^* \to \{\top, ?\}$ such that:

correctness. For all $w \in \Sigma_o^*$, if $D(w) = \top$ then $\mathsf{CorP}(w) \leq \varepsilon$;

reactivity. $\lim_{n\to\infty} \mathbb{P}(\{\rho \in \mathsf{F} \cap \mathsf{SR}_n \mid D(\pi(\rho)) =?\}) = 0$.

Proposition 15. *Let $\varepsilon > 0$. A finite pLTS $\mathcal{A}$ is εFF-diagnosable if and only if it admits an εFF-diagnoser.*

Contrary to exact diagnosers, εFF-diagnosers may need infinite memory.

Proposition 16. *There exists an AFF-diagnosable pLTS, thus εFF-diagnosable for every $\varepsilon > 0$, that admits no finite-memory diagnoser when $0 < \varepsilon \leq 1/2$.*

Proof. Consider the AFF-diagnosable pLTS of Figure 4 and assume there exists a εFF-diagnoser with m states for some threshold $0 < \varepsilon \leq 1/2$. 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 $\rho = q_0 u q_1 (aq_1)^i \mathbf{f} q_f b q_f$. Due to the reactivity requirement, there must be a run $\rho\rho'$ for which the diagnoser claims a fault. Thus for all n, the diagnoser claims a fault after $\rho_n = q_0 u q_1 (aq_1)^{i+n(j-i)} \mathbf{f} q_f b q_f \rho'$ but $\lim_{n\to\infty} \mathsf{CorP}(\pi(\rho_n)) = 1$, which contradicts the correctness requirement. □

Note that the definition of εFF-diagnosers differs from the one of monitors for distinguishability of hidden Markov chains [14, 15]. Indeed, while monitors have a prediction power, diagnosers do not. Perhaps surprisingly, one can prove that the two notions though coincide for finite-state models.

6. Hardness of diagnosis

We gave upper bounds on the complexity of diagnosability in Section 4. In this section, we provide lower bounds: on the one hand undecidability of the remaining approximate diagnosability notions, and on the other hand a matching lower bound for the exact diagnosis.

6.1. Undecidability results

After having previously proved that AFF-diagnosability can be solved in polynomial time, we now establish that all other specifications of approximate diagnosability are undecidable. This result could be expected for εFF-diagnosability and uniform εFF-diagnosability since it is often the case for problems mixing probabilities, partial observation and quantitative requirement

(here represented by ε). On the contrary, the undecidability of the uniform AFF-diagnosability problem is at first sight surprising since it is a slight variation of the AFF-diagnosability problem. In fact the reduction for the latter problem is more intricate than the one for the former problems. In all cases, we reduce from the emptiness problem for probabilistic automata [19].

Theorem 4. *For any rational $0 < \varepsilon < 1$, the εFF-diagnosability and uniform εFF-diagnosability problems are undecidable for pLTS.*

Proof. 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 every $a \in \Sigma$ 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 $\mathbf{Pr}_{\mathcal{A}}(w) = 1$ [19].

Let $\mathcal{A}$ be such a probabilistic automaton, with $\mathbf{Pr}_{\mathcal{A}}(w) < 1$ for every $w \in \Sigma^*$. Define the pLTS $\mathcal{A}' = \langle Q', q_0, \Sigma', T', \mathbf{P}' \rangle$ as follows.

- $\Sigma' = \Sigma \uplus \{\sharp, \mathbf{f}\}$, $\Sigma'_u = \{\mathbf{f}\}$;
- $Q' = Q \cup \{q_c^\sharp, q_f^\sharp, f^\sharp\}$;
- $T' = \{(q, a, q) \mid q, q' \in Q, a \in \Sigma, \mathbf{P}_a[q, q'] > 0\}$
  $\cup \{(q, \sharp, q_c^\sharp \mid q \in F\} \cup \{(q, \sharp, q_f^\sharp \mid q \in Q \setminus F\}$
  $\cup \{q_c^\sharp, \sharp, q_c^\sharp\} \cup \{q_f^\sharp, \mathbf{f}, f^\sharp\} \cup \{f^\sharp, \sharp, f^\sharp\}$
- $\mathbf{P}'$ 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 illustrated in Figure 17. $\mathbf{P}'$ fulfils the requirement for pLTS. For instance, let $q \in F$ and $a \in \Sigma$, then $\sum_{q' \in Q} \mathbf{P}_a[q, q'] = 1$, thus:

$$\sum_{(q,a,q') \in T'} \mathbf{P}'(q, a, q') = \sum_{a \in \Sigma} \sum_{q' \in Q} \frac{\mathbf{P}_a[q, q']}{1 + |\Sigma|} + \mathbf{P}'(q, \sharp, q_c^\sharp) = \frac{|\Sigma|}{1 + |\Sigma|} + \frac{1}{1 + |\Sigma|} = 1.$$

We claim that the following three assertions are equivalent:

1. $\mathcal{A}'$ is εFF-diagnosable;
2. $\mathcal{A}'$ is uniformly εFF-diagnosable;

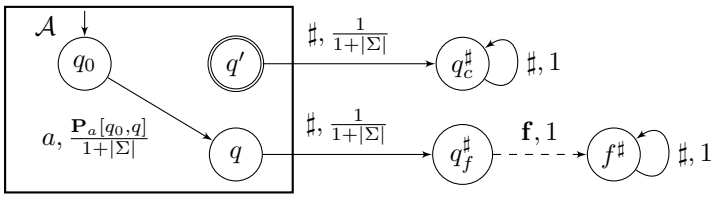


Figure 17: From probabilistic automata to pLTS.

3. $\mathcal{L}_{\mathcal{A},\varepsilon} = \emptyset$.

Given that uniform εFF-diagnosability entails εFF-diagnosability, it suffices to prove that item 1 implies item 3, and item 3 implies item 2. The first implication is proved by contraposition.

1 implies 3 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 $\frac{\mathbf{Pr}_{\mathcal{A}}(w)}{(1+|\Sigma|)^{|w|+1}}$. Similarly, the set of signalling faulty runs with observed sequence $w\sharp^{n+2}$ has probability $\frac{1-\mathbf{Pr}_{\mathcal{A}}(w)}{(1+|\Sigma|)^{|w|+1}}$. Thus $\mathsf{CorP}(w\sharp^{n+2}) = \mathbf{Pr}_{\mathcal{A}}(w) > \varepsilon$. By assumption on $\mathcal{A}$, $\mathbf{Pr}_{\mathcal{A}}(w) < 1$, so that the set of faulty runs with observed sequence $w\sharp^{n+2}$ is non empty. Pick ρ a minimal faulty run with observed sequence $w\sharp\sharp$. Using the above probability values, for every $n \geq 0$:

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

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

3 implies 2 Assume that for every word $w \in \Sigma^*$, $\mathbf{Pr}_{\mathcal{A}}(w) \leq \varepsilon$. Let ρ be a minimal faulty run of $\mathcal{A}'$. By construction, its observed sequence is of the form $w\sharp^2$ with $w \in \Sigma^*$. Using the same reasoning as above, for every $\rho \preceq \rho'$: $\mathsf{CorP}(\pi(\rho')) = \mathbf{Pr}_{\mathcal{A}}(w)$, and thus $\mathsf{CorP}(\pi(\rho')) \leq \varepsilon$. Therefore, 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 εFF-diagnosable.

This finishes the proof that εFF- and uniform εFF-diagnosability are undecidable. □

Uniform AFF-diagnosability is also shown to be undecidable by a reduction from the emptiness problem for probabilistic automata.

Theorem 5. *The uniform* AFF*-diagnosability problem is undecidable for pLTS.*

Sketch of proof. We proceed by a reduction from the emptiness problem of probabilistic automata where w.l.o.g. one assumes that the acceptance probability of any word lies between 1/4 and 3/4. Given such a probabilistic automaton one builds a pLTS as follows.

- With probability 1/2 one enters one of the two copies of the automaton whose probabilities are modified in a similar way as in the previous proof.
- In a nonaccepting state (resp. accepting) state of the first (resp. second) copy, one may exit the automaton outputting a $\flat$ and enter a terminating block. In the second copy, a fault occurs before the $\flat$. In an accepting state (resp. nonaccepting) state of the first (resp. second) copy, one may "restart" the automaton outputting a $\sharp$.
- The terminating block of the first copy iteratively outputs with probability 1/2 $\sharp$ or $\flat$ while the terminating block of the second copy endlessly outputs $\flat$.

Due to the behaviour of the terminating blocks, the correctness proportion of a faulty run goes to 0 as its length increases. Thus the pLTS is AFF-diagnosable.
Observe that the language of observation sequences of minimal faulty runs is $(\Sigma^*\sharp)^*\Sigma^*\flat$.
Assume there exists a word w with acceptance probability strictly greater than 1/2. Then in the pLTS, the correctness proportion of $(w\sharp)^n\flat$ fulfils:
$\lim_{n\to\infty} \mathsf{CorP}((w\sharp)^n\flat) = 1$. Due to this property (and the behaviours of the terminating blocks), the pLTS is not uniformly AFF-diagnosable. If no such word exists, then for any $w = w_1\sharp w_2\sharp \dots w_k\flat$, $\mathsf{CorP}(w) \leq 3/4$. Due to this property (and the behaviours of the terminating blocks), the pLTS is uniformly AFF-diagnosable. □

As uniform AFF-diagnosability is equivalent to the notion of AA-diagnosability introduced in [11] which decidability was left open (only necessary conditions were given), this theorem solves this question.

6.2. PSPACE*-hardness of exact diagnosis*

In order to establish a lower bound for the complexity of exact 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 $v^{-1}\mathcal{L} = \Sigma^*$. Recently, several variants of the universality problem were shown to be PSPACE-complete [20] but, to the best of our knowledge, eventual universality has not been considered.

Because of our diagnosis framework, we focus on live nondeterministic 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 [21] to the eventual universality problem.

Proposition 17. *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.*

Now that we established that universal eventuality is PSPACE-hard, we can use it to establish a complexity lower bound for the different exact diagnosability problems.

Proposition 18. *The* IF*-diagnosability,* FA*-diagnosability and* IA*-diagnosability problem are* PSPACE*-hard.*

Proof. The proof is done by reduction from the eventual universality problem. Let $\mathcal{A}$ be a live NFA over Σ, in which all states are final. One builds in polynomial time the initial-fault pLTS $\mathcal{A}'$ as depicted in Figure 18 where $\Sigma_o = \Sigma \uplus \{\sharp\}$, $\Sigma_u = \{u, \mathbf{f}\}$, and in which all transitions outgoing a state have the same probability. We establish the following two implications:

- $\mathcal{A}'$ is not FA-diagnosable implies $\mathcal{A}$ is eventually universal;
- $\mathcal{A}$ is eventually universal implies $\mathcal{A}'$ is not IF-diagnosable.

Since FA-diagnosability implies IA-diagnosability implies IF-diagnosability, this will achieve the proof that all three notions are at least as hard as eventual universality.

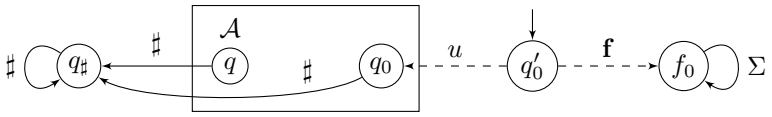


Figure 18: A reduction for PSPACE-hardness of IF-, FA- and IA-diagnosability.

• Assume that $\mathcal{A}'$ is not FA-diagnosable. By Proposition 3, either $\mathcal{A}'_{\mathsf{FA}}$ contains a reachable BSCC $\mathcal{C}$ with some state $s = (q, U, V) \in \mathcal{C}$ such that $q \in Q_f$ and $U \neq \emptyset$ or $\mathcal{A}'_{\mathsf{FA}}$ contains a reachable BSCC $\mathcal{C}$ with some state $s = (q, U, V) \in \mathcal{C}$ such that $q \in Q_c$ and $V \neq \emptyset$. The latter case is excluded since the only correct state belonging to a BSCC is $q_\sharp$ which is only reachable by a transition labelled by $\sharp$. As this observation cannot occur in a faulty run, $q = q_\sharp$ implies $V = \emptyset$. Consider the former case: obviously $q = f_0$. Since $\mathcal{C}$ is a BSCC and f_0 is a sink state in $\mathcal{A}'$, for every state $s' = (q', U', V') \in \mathcal{C}$, one has $q' = f_0$ and $U' \neq \emptyset$. Since in f_0 all events of Σ are enabled, this implies that for all $w \in \Sigma^*$, there is a correct run ρ_1 in $\mathcal{A}'$ starting from some state of $q \in U$ with observed sequence w. Consider an observed sequence $v \in \Sigma^*$ labelling a run in $\mathcal{A}'_{\mathsf{FA}}$ from the initial state to s. Then there is correct run in $\mathcal{A}'$ from q'_0 to q with observed sequence v. So the run $\rho = \rho_0\rho_1$ has vw as observed sequence. Since $\rho = q'_0 u \rho'$ with ρ' a run of $\mathcal{A}$ starting from q_0, $vw \in \mathcal{L}(\mathcal{A})$. This holds for any word w, thus $v^{-1}\mathcal{L}(\mathcal{A}) = \Sigma^*$ and $\mathcal{A}$ is eventually universal.

• Assume that there exists a word $v \in \Sigma^*$ such that $v^{-1}\mathcal{L}(\mathcal{A}) = \Sigma^*$. Of course, any word extending v is also a witness that $\mathcal{A}$ is eventually universal. Let

$v' \in \Sigma^*$ be such that, in $\mathcal{A}'_{\mathsf{FA}}$, a faulty run with observed sequence vv' ends in a BSCC $\mathcal{C}$. Since $(vv')^{-1}\mathcal{L}(\mathcal{A}) = \Sigma^*$, all states of $\mathcal{C}$ are of the form (f_0, U) with $U \neq \emptyset$. Therefore, by Proposition 2, $\mathcal{A}'$ is not IF-diagnosable. □

7. Conclusion

In this work, we settled the foundations of diagnosability for partially observed stochastic systems. In particular, we investigated semantical issues and identified several relevant definitions for diagnosability in a probabilistic context, providing a complete picture of the relations between these notions. We have also established characterisations for all exact diagnosis specifications and for one approximate diagnosis specification. Based on these characterisations, we have designed decision procedures and diagnoser synthesis algorithms. For the remaining cases, we proved the decision problems to be undecidable. The current paper thus rewrites the tale of exact and approximate diagnoses, with some unexpected news, such as the undecidability of the AA-diagnosability of [11].

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 analysed in the exact case in [12], should be studied in the approximate framework. 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. Second, we plan to move to more quantitative versions of diagnosis including optimisation issues. The objective would be to minimise the observational capacities of the monitor, either spatially or timely, by restricting either the observable actions, or the observation time instants, while preserving diagnosability. Last, in [22], we recently studied exact diagnosis for infinite-state probabilistic systems. Such a work could be completed by an analysis of approximate diagnosis for infinite-state probabilistic systems.

We are very grateful to the reviewers for their valuable comments, and in particular for pointing the similarities between AFF-diagnosability and monitorability.

References

[1] B. Buchanan, E. Shortliffe, Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project, Reading, MA: Addison-Wesley, 1984.

[2] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, D. Teneketzis, Diagnosability of discrete-event systems, IEEE Trans. Aut. Cont. 40 (9) (1995) 1555–1575. doi:10.1109/9.412626.

[3] S. Jiang, Z. Huang, V. Chandra, R. Kumar, A polynomial algorithm for testing diagnosability of discrete-event systems, IEEE Transactions on Automatic Control 46 (8) (2001) 1318–1321. doi:10.1109/9.940942.

[4] M. Cabasino, A. Giua, S. Lafortune, C. Seatzu, Diagnosability analysis of unbounded Petri nets, in: Proceedings of CDC'09, IEEE, 2009, pp. 1267–1272. doi:10.1109/CDC.2009.5400608.

[5] C. Morvan, S. Pinchinat, Diagnosability of pushdown systems, in: Proceedings of HVC'09, LNCS 6405, Springer, 2009, pp. 21–33. doi:10.1007/978-3-642-19237-1_6.

[6] F. Cassez, S. Tripakis, Fault diagnosis with static and dynamic observers, Fundamenta Informaticae 88 (2008) 497–540.

[7] M. Sampath, S. Lafortune, D. Teneketzis, Active diagnosis of discrete-event systems, IEEE Transactions on Automatic Control 43 (7) (1998) 908–929. doi:10.1109/9.701089.

[8] D. Thorsley, D. Teneketzis, Active acquisition of information for diagnosis and supervisory control of discrete-event systems, Journal of Discrete Event Dynamic Systems 17 (2007) 531–583. doi:10.1007/s10626-007-0027-y.

[9] E. Chanthery, Y. Pencolé, Monitoring and active diagnosis for discrete-event systems, in: Proceedings of SP'09, Elsevier, 2009, pp. 1545–1550.

[10] N. Bertrand, E. Fabre, S. Haar, S. Haddad, L. Hélouët, Active diagnosis for probabilistic systems, in: Proceedings of FoSSaCS'14, Vol. 8412 of LNCS, Springer, 2014, pp. 29–42. doi:10.1007/978-3-642-54830-7_2.

[11] D. Thorsley, D. Teneketzis, Diagnosability of stochastic discrete-event systems, IEEE Transactions on Automatic Control 50 (4) (2005) 476–492. doi:10.1109/TAC.2005.844722.

[12] N. Bertrand, S. Haddad, E. Lefaucheux, Foundation of diagnosis and predictability in probabilistic systems, in: Proceedings of FSTTCS'14, Vol. 29 of LIPIcs, Leibniz-Zentrum für Informatik, 2014, pp. 417–429. doi:10.4230/LIPIcs.FSTTCS.2014.417.

[13] J. Chen, R. Kumar, Polynomial test for stochastic diagnosability of discrete-event systems, IEEE Transactions on Automation Science and Engineering 10 (4) (2013) 969–979. doi:10.1109/TASE.2013.2251334.

[14] A. P. Sistla, M. Zefran, Y. Feng, Monitorability of stochastic dynamical systems, in: Proceedings of the 23rd International Conference on Computer Aided Verification, CAV'11, Snowbird, UT, USA, July 14-20, 2011, 2011, pp. 720–736.

[15] S. Kiefer, A. P. Sistla, Distinguishing hidden Markov chains, in: Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16, New York, NY, USA, July 5-8, 2016, ACM, 2016, pp. 66–75.

[16] T. Chen, S. Kiefer, On the total variation distance of labelled Markov chains, in: Proceedings of CSL-LICS'14, ACM, 2014, pp. 33:1–33:10. doi:10.1145/2603088.2603099.

[17] N. Bertrand, S. Haddad, E. Lefaucheux, Accurate approximate diagnosability of stochastic systems, in: Proceedings of LATA'16, Vol. 9618 of LNCS, Springer, 2016, pp. 549–561. doi:10.1007/978-3-319-30000-9_42.

[18] S. Haar, S. Haddad, T. Melliti, S. Schwoon, Optimal constructions for active diagnosis, in: Proceedings of FSTTCS'13, Vol. 24 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2013, pp. 527–539. doi:10.4230/LIPIcs.FSTTCS.2013.527.

[19] A. Paz, Introduction to Probabilistic Automata, Academic Press, 1971.

[20] N. Rampersad, J. Shallit, Z. Xu, The computational complexity of universality problems for prefixes, suffixes, factors, and subwords of regular languages, Fundam. Inf. 116 (1-4) (2012) 223–236. doi:10.3233/FI-2012-680.

[21] A. R. Meyer, L. J. Stockmeyer, The equivalence problem for regular expressions with squaring requires exponential space, in: Proceedings of SWAT'72, IEEE Computer Society, Washington, DC, USA, 1972, pp. 125–129. doi:10.1109/SWAT.1972.29.

[22] N. Bertrand, S. Haddad, E. Lefaucheux, Diagnosis in infinite-state probabilistic systems, in: Proceedings of CONCUR'16, Vol. 59 of LIPIcs, Leibniz-Zentrum für Informatik, 2016, pp. 37:1–37:14.

Appendix A.

This appendix contains proofs that are omitted in section 2.

Lemma 1. *Let $\mathcal{A}$ be a pLTS. Then* $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_\infty \setminus \mathsf{FAmb}_n) = 0$. *Moreover, if $\mathcal{A}$ is finitely branching, then* $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \setminus \mathsf{FAmb}_\infty) = 0$.

Proof. Observe that Ω admits the following partitions $\Omega = \mathsf{FAmb}_\infty \uplus \mathsf{C}_\infty \uplus \mathsf{Sf}_\infty$ and for all $n \in \mathbb{N}$, $\Omega = \mathsf{FAmb}_n \uplus \mathsf{C}_n \uplus \mathsf{Sf}_n$. Thus, for all $n \in \mathbb{N}$,

$$\begin{aligned}\mathsf{FAmb}_\infty \setminus \mathsf{FAmb}_n &= (\mathsf{C}_n \uplus \mathsf{Sf}_n) \cap \mathsf{FAmb}_\infty \\ &= (\mathsf{C}_n \uplus \mathsf{Sf}_n) \setminus (\mathsf{C}_\infty \uplus \mathsf{Sf}_\infty) \subseteq (\mathsf{C}_n \setminus \mathsf{C}_\infty) \uplus (\mathsf{Sf}_n \setminus \mathsf{Sf}_\infty).\end{aligned}$$

Since for all n, $\mathsf{Sf}_n \subseteq \mathsf{Sf}_\infty$, one gets:

$$\mathsf{FAmb}_\infty \setminus \mathsf{FAmb}_n \subseteq \mathsf{C}_n \setminus \mathsf{C}_\infty \ .$$

$\{\mathsf{C}_n\}_{n\in\mathbb{N}}$ is a nonincreasing family of sets and we claim that $\mathsf{C}_\infty = \bigcap_{n\in\mathbb{N}} \mathsf{C}_n$. Indeed an infinite run ρ is correct if and only if $\mathbf{f}$ does not occur in it *i.e.* if and only if all its signalling subruns are correct. Thus,

$$\lim_{n\to\infty} \mathbb{P}(\mathsf{C}_n \setminus \mathsf{C}_\infty) = 0 \text{ implying } \lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_\infty \setminus \mathsf{FAmb}_n) = 0 \ .$$

Using again the two partitions we obtain:

$$\begin{aligned}\mathsf{FAmb}_n \setminus \mathsf{FAmb}_\infty &= (\mathsf{C}_\infty \uplus \mathsf{Sf}_\infty) \cap \mathsf{FAmb}_n \\ &= (\mathsf{C}_\infty \uplus \mathsf{Sf}_\infty) \setminus (\mathsf{C}_n \uplus \mathsf{Sf}_n) \subseteq (\mathsf{C}_\infty \setminus \mathsf{C}_n) \uplus (\mathsf{Sf}_\infty \setminus \mathsf{Sf}_n).\end{aligned}$$

Since for all n, $\mathsf{C}_\infty \subseteq \mathsf{C}_n$, one gets:

$$\mathsf{FAmb}_n \setminus \mathsf{FAmb}_\infty \subseteq \mathsf{Sf}_\infty \setminus \mathsf{Sf}_n$$

Let us show that, under the assumption that $\mathcal{A}$ is finitely branching, $\mathsf{Sf}_\infty \subseteq \bigcup_{n\in\mathbb{N}} \mathsf{Sf}_n$. Let $\rho \notin \bigcup_{n\in\mathbb{N}} \mathsf{Sf}_n$. We build a tree as follows:

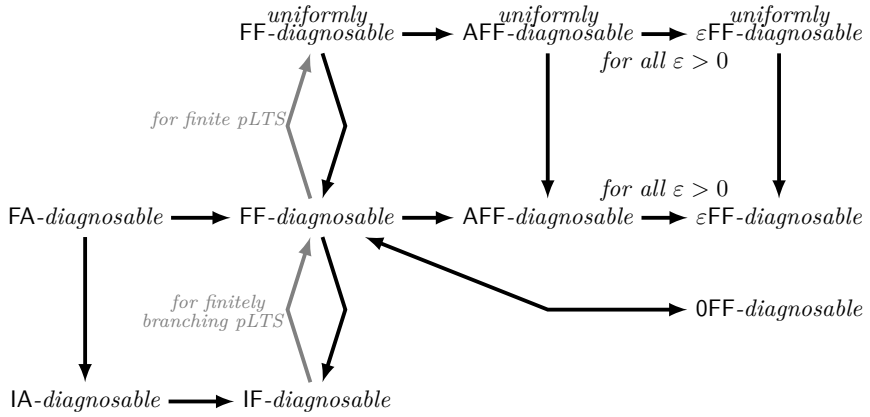
- Nodes at level n correspond to the correct signalling runs whose observed sequence is $\pi(\rho_{\downarrow n})$;
- The node at level $n+1$ associated with ρ' is a child of the node at level n associated with ρ'' if $\rho'' \preceq \rho'$.

Since $\rho \notin \bigcup_{n\in\mathbb{N}} \mathsf{Sf}_n$, for all $n \in \mathbb{N}$, there exists a correct run with observed sequence $\pi(\rho_{\downarrow n})$, so that the above-defined tree is infinite. Since the pLTS is finitely branching and convergent, the tree is also finitely branching. By König's lemma, it must contain an infinite branch, thus there exists an infinite correct run whose observed sequence is $\pi(\rho)$. As a consequence ρ is not surely faulty: $\rho \notin \mathsf{Sf}_\infty$. This establishes that $\mathsf{Sf}_\infty \subseteq \bigcup_{n\in\mathbb{N}} \mathsf{Sf}_n$. Thus:

$$\lim_{n\to\infty} \mathbb{P}(\mathsf{Sf}_\infty \setminus \mathsf{Sf}_n) = 0 \text{ implying } \lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \setminus \mathsf{FAmb}_\infty) = 0$$

which concludes the proof. □

Theorem 1. *The diagnosability notions for pLTS are related according to the diagram below, where arrows represent implications. All implications, except the one from* IF*-diagnosability to* FF*-diagnosability hold for arbitrary infinite-state pLTS. The latter implication holds for finitely branching pLTS. Implications that are not depicted do not hold, already in the case of finite-state pLTS.*



The remainder of this section is devoted to the proof of Theorem 1. When the probabilities are omitted in examples of pLTS, we implicitly assume uniform distributions on outgoing edges for all state.

Implication.. We start by proving all implications.

$\mathsf{FA} \Rightarrow \mathsf{FF}$**.** Immediate since for all n, $\mathsf{FAmb}_n \subseteq \mathsf{FAmb}_n \uplus \mathsf{CAmb}_n$.

$\mathsf{IA} \Rightarrow \mathsf{IF}$**.** Immediate since $\mathsf{FAmb}_\infty \subseteq \mathsf{FAmb}_\infty \uplus \mathsf{CAmb}_\infty$.

$\mathsf{FF} \Rightarrow \mathsf{IF}$**.** This implication is a consequence of Lemma 1.

$\mathsf{IF} \Rightarrow \mathsf{FF}$ **assuming finite branching.**
This implication is a consequence of Lemma 1.

for all $\varepsilon > 0$**, uniform** $\mathsf{AFF} \Rightarrow$ **uniform** $\varepsilon\mathsf{FF}$**.**
By definition of uniform AFF-diagnosability.

for all $\varepsilon > 0$**,** $\mathsf{AFF} \Rightarrow \varepsilon\mathsf{FF}$**.**
By definition of AFF-diagnosability.

for all $\varepsilon \geq 0$**, uniform** $\varepsilon\mathsf{FF} \Rightarrow \varepsilon\mathsf{FF}$**.**
Let $\mathcal{A}$ be a uniform εFF-diagnosable pLTS, $\varepsilon \geq 0$, $\alpha > 0$ and $\rho \in \mathsf{minF}$. By definition of uniform εFF-diagnosability, there exists n_α such that

$$\mathbb{P}(\mathsf{Cyl}(\rho) \cap \mathsf{FAmb}^{\varepsilon}_{n+|\rho|_o}) \leq \alpha \cdot \mathbb{P}(\rho).$$

Thus $\mathcal{A}$ is εFF-diagnosable.

uniform AFF $\Rightarrow$ AFF**.**

For all $\varepsilon > 0$, uniform AFF $\Rightarrow$ uniform εFF and uniform εFF $\Rightarrow$ εFF. Thus uniform AFF-diagnosability implies AFF-diagnosability

uniform FF $\Rightarrow$ FF**.**

By definition, uniform FF-diagnosability is uniform 0FF-diagnosability. Since uniform 0FF-diagnosability implies 0FF-diagnosability which has been shown to be equivalent to FF-diagnosability, one gets the result.

FF $\Rightarrow$ **uniform** FF **assuming finite pLTS.** See Proposition A.

FA $\Rightarrow$ IA**.**

For all $n \in \mathbb{N}$, define $\mathsf{CAmb}_{n,\infty}$ the set of correct ambiguous runs that admit an observationally equivalent run which is faulty before its n^{th} observable event. Observe that the sequence of sets $\{\mathsf{CAmb}_{n,\infty}\}_{n\in\mathbb{N}}$ is non-decreasing and that $\mathsf{CAmb}_\infty = \bigcup_{n\in\mathbb{N}} \mathsf{CAmb}_{n,\infty}$. Moreover, by definition, $\mathsf{CAmb}_{n,\infty} \subseteq \mathsf{CAmb}_n$. Assume that $\limsup_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \uplus \mathsf{CAmb}_n) = 0$. By Lemma 1, $\mathbb{P}(\mathsf{FAmb}_\infty) = 0$. For all $\varepsilon > 0$, there exists $n_1 \in \mathbb{N}$ such that for all $n \geq n_1$, $\mathbb{P}(\mathsf{CAmb}_n) < \varepsilon$ and thus $\mathbb{P}(\mathsf{CAmb}_{n,\infty}) < \varepsilon$. On the other hand, there exists $n_2 \in \mathbb{N}$ such that for all $n \geq n_2$, $\mathbb{P}(\mathsf{CAmb}_\infty) - \mathbb{P}(\mathsf{CAmb}_{n,\infty}) < \varepsilon$. Combining these two inequalities for $n = \max(n_1, n_2)$, one obtains $\mathbb{P}(\mathsf{CAmb}_\infty) < 2\varepsilon$. As ε is arbitrary, $\mathbb{P}(\mathsf{CAmb}_\infty) = 0$.

FF $\Rightarrow$ AFF**.**

FF-diagnosability has been shown to be equivalent to 0FF-diagnosability. By definition, for all $\varepsilon' \geq \varepsilon$, εFF-diagnosability implies ε'FF-diagnosability. Thus fixing $\varepsilon = 0$ and letting ε' arbitrary, one obtains that FF-diagnosability implies AFF-diagnosability.

uniform FF $\Rightarrow$ **uniform** AFF**.**

By definition uniform FF-diagnosability is uniform 0FF-diagnosability and for all $\varepsilon' \geq \varepsilon$, uniform εFF-diagnosability implies uniform ε'FF-diagnosability. Thus fixing $\varepsilon = 0$ and letting ε' arbitrary, one obtains that uniform FF-diagnosability implies uniform AFF-diagnosability.

Non-implications. We now provide counter-examples for the implications that do not hold.

FF $\not\Rightarrow$ FA**.**

Consider the pLTS from Figure 1 where $\Sigma_u = \{u, \mathbf{f}\}$ is the set of unobservable events, represented by dashed transitions. Any faulty run will almost surely contain a b-event that cannot occur in the single infinite correct run. Thus the probability of faulty signalling ambiguous runs with observable length n converges to 0 when n goes to ∞. Thus this pLTS is FF-diagnosable. The infinite correct run $\rho = q_0 u q_1 a q_1 \ldots$ has probability $1/2$ and its observed sequence a^ω is ambiguous. Therefore this pLTS is not IA-diagnosable. Since FA-diagnosability implies IA-diagnosability, it is not FA-diagnosable.

$\mathsf{IF} \not\Rightarrow \mathsf{IA}$.

Consider again the not IA-diagnosable and FF-diagnosable pLTS of Figure 1. Since FF-diagnosability implies IF-diagnosability, it is IF-diagnosable.

$\mathsf{FA} \not\Rightarrow$ **uniform** $\varepsilon\mathsf{FF}$ **when considering infinite pLTS.**

Let us consider the pLTS of Figure A.19. It is FA-diagnosable as almost surely an infinite faulty (resp. correct) run contains a b (resp. c) that can not be mimicked by a correct (resp. faulty) run. We claim that it is not uniformly εFF-diagnosable for all ε such that $0 < \varepsilon < 1/2$. Remark that for all $n \in \mathbb{N}$, $\mathsf{CorP}(a^n) \geq 1/2$. Fix some $0 < \alpha < 1$ and $n_\alpha \in \mathbb{N}$. Consider the minimal faulty run $\rho = q_0 u f_1 a f_2 \dots a f_{n_\alpha} \mathbf{f} f'_{n_\alpha}$. The shortest extension of ρ that is not ambiguous (*i.e.* contains a b) contains $n_\alpha + 1$ observable events more than ρ. Therefore, $\mathbb{P}(\{\rho' \in \mathsf{FAmb}^{\varepsilon}_{n_\alpha+|\rho|_o} \mid \rho \preceq \rho'\}) = \mathbb{P}(\rho) > \alpha \cdot \mathbb{P}(\rho)$.

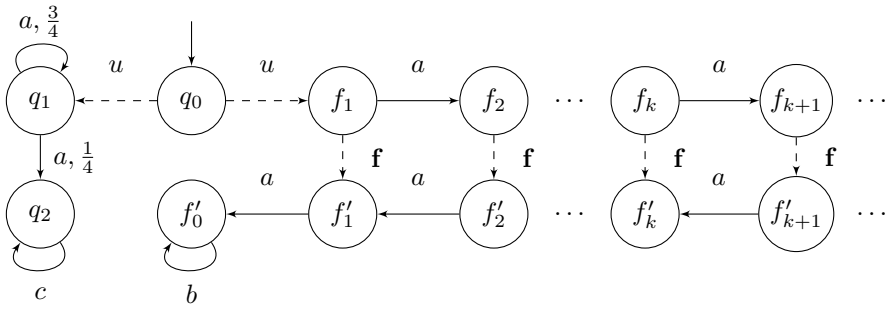


Figure A.19: An infinite FA-diagnosable pLTS that is not uniformly εFF-diagnosable.

uniform $\varepsilon\mathsf{FF} \not\Rightarrow \mathsf{AFF}$.

Consider the pLTS of Figure A.20. There is a single signalling minimal faulty run $q_0 \mathbf{f} q_f a q_f$. Any observed sequence of length at least 1 is ambiguous and corresponds with equal probability to a signalling correct or a faulty run. Consequently it is not AFF-diagnosable, yet it is uniformly εFF-diagnosable for $\varepsilon = 1/2$.

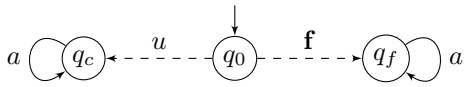


Figure A.20: A uniform 1/2FF-diagnosable pLTS, that is not AFF-diagnosable.

Appendix B.

This appendix contains proofs that are omitted in section 3.

Relying on the characterisations, we are now in a position to establish an implication that was claimed in Theorem 1.

Proposition A. *Let $\mathcal{A}$ be a finite pLTS. If $\mathcal{A}$ is* FF*-diagnosable, then it is uniformly* FF*-diagnosable.*

Proof. Let $\mathcal{A}$ be an FF-diagnosable pLTS. For a run ρ of $\mathcal{A}$, we let ρ_{IF} be its associated run in $\mathcal{A}_{\mathsf{IF}}$: ρ_{IF} extends the states appearing along ρ by subsets of possible correct states after the corresponding prefix of the observed sequence $\pi(\rho)$. We let S_{BSCC} denote the set of states of $\mathcal{A}_{\mathsf{IF}}$ that belong to a BSCC. Last, for every state (q, U) of $\mathcal{A}_{\mathsf{IF}}$ and every $n \in \mathbb{N}$, we denote by $\mathsf{SR}_n^{q,U}$ the set of signalling runs of length n starting at (q, U).

Let $\alpha > 0$. Our objective is to define n_α such that for every $n \geq n_\alpha$ and every minimal faulty run $\rho \in \mathsf{minF}$:

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

We first exploit the almost sure convergence towards BSCC in $\mathcal{A}_{\mathsf{IF}}$. For every state (q, U) of $\mathcal{A}_{\mathsf{IF}}$, the measure of runs starting in (q, U) and avoiding all BSCC during n steps tends to 0, when n goes to infinity. Thus, given α, for every reachable (q, U), there exists $n_{q,U} \in \mathbb{N}$ such that for every $n \geq n_{q,U}$, $\mathbb{P}(\{\rho'_{\mathsf{IF}} \in \mathsf{SR}_n^{q,U} \mid \mathsf{last}(\rho'_{\mathsf{IF}}) \notin S_{\mathsf{BSCC}}\}) \leq \alpha$. We define n_α as the maximum of $n_{q,U}$ over all states (q, U).

Now let ρ be a minimal faulty run of $\mathcal{A}$, and define $(q, U) = \mathsf{last}(\rho_{\mathsf{IF}})$. Since $n_\alpha \geq n_{q,U}$, $\mathbb{P}(\{\rho'_{\mathsf{IF}} \in \mathsf{SR}_{n_\alpha}^{q,U} \mid \mathsf{last}(\rho'_{\mathsf{IF}}) \notin S_{\mathsf{BSCC}}\}) \leq \alpha$. Therefore, as $\mathcal{A}$ and $\mathcal{A}_{\mathsf{IF}}$ have the same probabilistic behaviour,

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

Thanks to the characterisation of Proposition 2, all states in BSCC reachable from (q, U) in $\mathcal{A}_{\mathsf{IF}}$ necessarily are of the form $(q', \emptyset)$. Therefore, if a finite run ρ'_{IF} reaches such a BSCC, ρ'_{IF} admits no correct run with same observed sequence, and hence $\mathsf{CorP}(\pi(\rho'_{\mathsf{IF}})) = 0$. Equivalently, $\mathsf{CorP}(\pi(\rho')) > 0$ implies $\mathsf{last}(\rho'_{\mathsf{IF}}) \notin S_{\mathsf{BSCC}}$. 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 FF-diagnosable. □

Proposition 3. *Let $\mathcal{A}$ be a finite pLTS. $\mathcal{A}$ is* FA*-diagnosable if and only if $\mathcal{A}_{\mathsf{FA}}$ has no BSCC that:*

- *either contains a state (q, U, V) with $q \in Q_f$ and $U \neq \emptyset$;*
- *or contains a state (q, U, V) with $q \in Q_c$ and $V \neq \emptyset$.*

Proof. To prove the left-to-right implication, we proceed by contraposition. If one assumes the first item holds, the same argument as in the proof of Proposition 2 apply. Precisely, suppose that there exists a reachable BSCC C of

$\mathcal{A}_{\mathsf{FA}}$ and a state $s = (q, U, V)$ in C such that $q \in Q_f$ and $U \neq \emptyset$. Let ρ be a signalling run leading from the initial state s_0 of $\mathcal{A}_{\mathsf{FA}}$ to s. Now, for every state $s' = (q', U', V') \in C$, necessarily $q' \in Q_f$ and $U' \neq \emptyset$, because C is strongly connected. So for every signalling run ρ' that extends ρ, writing $s' = (q', U', V')$ for the state ρ' leads to, there exists a correct signalling run ρ'' such that $\pi(\rho'') = \pi(\rho')$ and $q_0 \xrightarrow{\rho''} q''$ with $q'' \in U'$. As a consequence the observed sequence $\pi(\rho'')$ is ambiguous, and for every $n \geq |\rho|_o$, $\mathbb{P}(\mathsf{FAmb}_n) \geq \mathbb{P}(\rho)$, so that $\mathcal{A}$ is not FA-diagnosable.

Suppose now that there exists a reachable BSCC C of $\mathcal{A}_{\mathsf{FA}}$ and a state $s = (q, U, V)$ in C such that $q \in Q_c$ and $V \neq \emptyset$. Since the pair (U, V) is unchanged by unobservable transitions, w.l.o.g we assume that s is the successor of some state of C by an observable event and we denote C' the set of such states.

Observe that a signalling run that reaches s is ambiguous. Denote $\pi_i(s')$ the probability that a random path visits a state s' at instant i. In a finite DTMC, for every state s' of a BSCC the Cesaro-limit $\pi_\infty(s') = \lim_{n\to\infty} 1/(n+1) \sum_{i=0}^{n} \pi_i(s')$ exists and is greater than 0. For $s' \in C'$ denote by $p_{s',s}$ the probability of an observable transition from s' to s. Then $0 < \sum_{s' \in C'} \pi_\infty(s') p_{s',s} \leq \liminf_{n\to\infty} 1/(n+1) \sum_{i=0}^{n} \alpha_i(s)$ where $\alpha_i(s)$ is the probability that a random path at time i is a signalling run visiting s. From time 0 to time n, a run can be a signalling run at most $n+1$ times. Thus:

$$\frac{1}{n+1} \sum_{i=0}^{n} \alpha_i(s) \leq \frac{1}{n+1} \sum_{i=0}^{n} \mathbb{P}(\mathsf{CAmb}_i)$$

which implies that

$$0 < \liminf_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} \mathbb{P}(\mathsf{CAmb}_i) \leq \limsup_{n\to\infty} \mathbb{P}(\mathsf{CAmb}_n) \ .$$

In this case also, we conclude that $\mathcal{A}$ is not FA-diagnosable.

The proof of Proposition 2 has established that a signalling run reaching a BSCC C where for every state $s = (q, U, V)$ q is faulty and $U = \emptyset$ is surely faulty. Similarly a signalling run that reaches a BSCC where for every state $s = (q, U, V)$, q is correct and $V = \emptyset$, is surely correct. Thus an ambiguous signalling run must only visit transient states. Since runs almost surely leave the transient states and reach a BSCC, this implies that:

$$\limsup_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n) + \mathbb{P}(\mathsf{CAmb}_n) = 0 \ ,$$

and therefore, the pLTS is FA-diagnosable. □

Proposition 4. *Let $\mathcal{A}$ be a finite pLTS. $\mathcal{A}$ is IA-diagnosable if and only if $\mathcal{A}_{\mathsf{IA}}$ has no BSCC such that:*

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

Proof. Assume first that $\mathcal{A}_{\mathsf{IA}}$ has a BSCC with (at least) some state (q, U, V, W) with $q \in Q_f$ and $U \neq \emptyset$. Using Proposition 2, $\mathcal{A}$ is not IF-diagnosable and thus not IA-diagnosable either.
If now some BSCC C of $\mathcal{A}_{\mathsf{IA}}$ has all its states (q, U, V, W) with $q \in Q_c$ and $W \neq \emptyset$. In particular none of these states are accepting for the deterministic Büchi automaton $\mathsf{IA}(\mathcal{A})$. Let ρ be a finite signalling run that hits C. By Proposition 1, any infinite run ρ' that extends ρ is ambiguous. From $q \in Q_c$ we deduce that $\mathbb{P}(\mathsf{CAmb}_\infty) \geq \mathbb{P}(\rho) > 0$. Therefore $\mathcal{A}$ is not IA-diagnosable.

Assume now $\mathcal{A}_{\mathsf{IA}}$ has no BSCC such that either, all its states (q, U, V, W) fulfil $q \in Q_f$ and $U \neq \emptyset$, or all its states (q, U, V, W) fulfil $q \in Q_c$ and $W \neq \emptyset$. First observe that in case some BSCC of $\mathcal{A}_{\mathsf{IA}}$ contains some state (q, U, V, W) with $q \in Q_f$ and $U \neq \emptyset$, then all its states satisfy the same constraints. Moreover, if some state (q, U, V, W) of a BSCC has $q \in Q_c$, then all states of this BSCC have their q-component in Q_c. Therefore, the condition can be reformulated as follows: all BSCC C of $\mathcal{A}_{\mathsf{IA}}$ satisfy:

- either all states (q, U, V, W) of C fulfil $q \in Q_f$ and $U = \emptyset$;
- or all states (q, U, V, W) of C fulfil $q \in Q_c$ and some state (q, U, V, W) of C fulfils $W = \emptyset$.

Whatever the case, all contain (at least) an accepting state for the Büchi condition of $\mathsf{IA}(\mathcal{A})$. Since all runs almost surely end in a BSCC and visit each of its states infinitely often, using Proposition 1, almost all runs of $\mathcal{A}_{\mathsf{IA}}$ are unambiguous. This proves that $\mathcal{A}$ is IA-diagnosable. □

Lemma 2. *Let $\mathcal{A} = \langle q_0, \mathcal{A}^f, \mathcal{A}^c \rangle$ be an initial-fault pLTS. Then $\mathcal{A}$ is AFF-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 measures 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 $\mathcal{M}$ leaves unchanged the probability of occurrence of an observed sequence.

Let us now prove the equivalence, starting with the left-to-right implication.

• Assume $\mathcal{A}$ is AFF-diagnosable. Then, for every $\varepsilon > 0$ and every minimal faulty run ρ:

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

Pick some $0 < \varepsilon < 1$. Applying Equation (B.1) on the minimal 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 threshold ε:

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

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

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

The factor 2 comes from the probability $1/2$ in $\mathcal{A}$ to enter $\mathcal{A}^f$ that $\mathbb{P}_f$ does not take into account contrary to $\mathbb{P}$.

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:

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

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 $d(\mathcal{M}(\mathcal{A}^c), \mathcal{M}(\mathcal{A}^f)) \geq \mathbb{P}_f(E) - \mathbb{P}_c(E) \geq 1 - \varepsilon(2 + \frac{1}{1-\varepsilon})$. Since ε was arbitrary, taking the limit when ε goes to 0, we obtain the desired result: $d(\mathcal{M}(\mathcal{A}^c), \mathcal{M}(\mathcal{A}^f)) = 1$.

• Conversely assume that $d(\mathcal{M}(\mathcal{A}^f), \mathcal{M}(\mathcal{A}^c)) = 1$. Thanks to Proposition 5, there exists an event $E \subseteq \Sigma_o^\omega$ such that $\mathbb{P}_f(E) = 1$ and $\mathbb{P}_c(E) = 0$.

For every $n \in \mathbb{N}$, 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 every $\varepsilon > 0$, we also define $\mathfrak{S}_n^\varepsilon$ as the subset of $\mathfrak{S}_n$ consisting of sequences whose correctness proportion exceeds threshold ε: $\mathfrak{S}_n^\varepsilon = \{\sigma \in \mathfrak{S}_n \mid \mathsf{CorP}(\sigma) > \varepsilon\}$.

From $\bigcap_{n\in\mathbb{N}} \mathsf{Cyl}(\mathfrak{S}_n) = E$, we derive that $\lim_{n\to\infty} \mathbb{P}_c(\mathfrak{S}_n) = \mathbb{P}_c(E) = 0$. Thus $\lim_{n\to\infty} \mathbb{P}_c(\mathfrak{S}_n^\varepsilon) = 0$.

On the other hand, for every $n \in \mathbb{N}$,

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

Since ε is fixed, $\mathbb{P}_f(\mathfrak{S}_n^\varepsilon) < \frac{1-\varepsilon}{\varepsilon}\mathbb{P}_c(\mathfrak{S}_n^\varepsilon)$ and $\lim_{n\to\infty} \mathbb{P}_c(\mathfrak{S}_n^\varepsilon) = 0$ imply that $\lim_{n\to\infty} \mathbb{P}_f(\mathfrak{S}_n^\varepsilon) = 0$.

Let ρ be a minimal faulty run and $\alpha > 0$. There exists $n_\alpha \geq |\rho|_o = 1$ such that for all $n \geq n_\alpha$, $\mathbb{P}_f(\mathfrak{S}_n^\varepsilon) \leq \alpha$. Let $n \geq n_\alpha$, 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 $\mathbb{P}(\widetilde{\mathfrak{S}}_n) \leq \alpha$. On the one hand, since $\mathbb{P}_f(\mathfrak{S}_n) \geq \mathbb{P}_f(E) = 1$, $\mathbb{P}_f(\widetilde{\mathfrak{S}}_n \cap (\Sigma_o^n \setminus \mathfrak{S}_n)) = 0$. On the other hand, since $\mathbb{P}_f(\mathfrak{S}_n^\varepsilon) \leq \alpha$, $\mathbb{P}_f(\widetilde{\mathfrak{S}}_n \cap \mathfrak{S}_n) \leq \mathbb{P}_f(\mathfrak{S}_n^\varepsilon) \leq \alpha$. Thus, $\mathbb{P}_f(\widetilde{\mathfrak{S}}_n) = \mathbb{P}_f(\widetilde{\mathfrak{S}}_n \cap \mathfrak{S}_n) + \mathbb{P}_f(\widetilde{\mathfrak{S}}_n \cap (\Sigma_o^n \setminus \mathfrak{S}_n)) \leq \alpha$. Because α was taken arbitrary, we obtain 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, $\lim_{n\to\infty} \mathbb{P}(\{\rho' \in \mathsf{SR}_n \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho')) > \varepsilon\}) = 0$. In conclusion $\mathcal{A}$ is AFF-diagnosable. □

The characterisation from Theorem 2 follows from Lemmas A and B given below, that state each one implication of the equivalence.

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 $q_0 \stackrel{\rho_f}{\Longrightarrow} q_f$ such that $\rho_f \in \mathsf{F}$ and $\pi(\rho_c) = \pi(\rho_f)$. Then $\mathcal{A}$ is not AFF-diagnosable.*

Proof. Let us introduce some notations:

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

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 $\mathcal{A}' = \langle q_0', \mathcal{A}_{q_f}, \mathcal{A}_{q_c}\rangle$. It is well-defined since q_c belongs to a BSCC so that $\mathcal{A}_{q_c}$ does not trigger faults. We write $\mathbb{P}'$ for the probability measure in $\mathcal{A}'$. Since $d(\mathcal{M}(\mathcal{A}_{q_f}), \mathcal{M}(\mathcal{A}_{q_c})) < 1$, due to Lemma 2, there exist positive reals $\alpha', \varepsilon' \leq 1$ such that for all $n_0 \in \mathbb{N}$ there exists $n \geq n_0$:

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

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 $1/2$ 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 $\alpha = \tilde{\alpha}/2$, for all $n_0 \in \mathbb{N}$ there exists $n \geq n_0$:

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

Let ρ_0 be the minimal faulty run such that $\rho_0 \preceq \rho_f$. We observe that $\mathsf{Cyl}(\rho_f) \subseteq \mathsf{Cyl}(\rho_0)$, so that

$$\mathbb{P}(\{\rho \in \mathsf{SR}_n \mid \rho_0 \preceq \rho \wedge \frac{p_c^{\pi(\rho)}}{p_c^{\pi(\rho)} + p_g^{\pi(\rho)}} > \varepsilon\}) > \alpha$$

which establishes that $\mathcal{A}$ is not AFF-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 $\rho_f \in \mathsf{F}$, $\pi(\rho_c) = \pi(\rho_f)$, $q_f \in Q_f$ 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 AFF-diagnosable.*

Proof. Let ρ_0 be a minimal faulty run, $\alpha > 0, \varepsilon > 0$, $\sigma_0 = \pi(\rho_0)$ and $n_0 = |\sigma_0|$.

Before developing the proof, we sketch its structure and illustrate it in Figure B.21. 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 2 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.

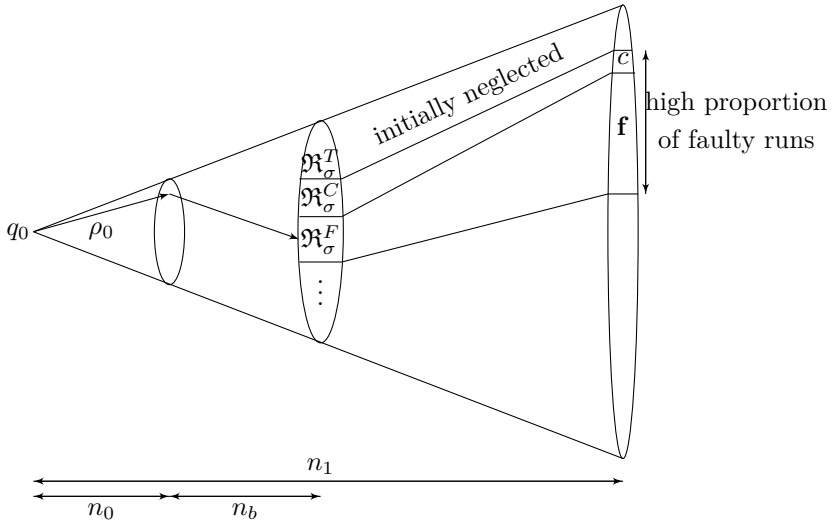


Figure B.21: Illustration of the proof of Lemma B.

Let ε, α be positive reals. Since almost surely a random run ends in a BSCC, there exists n_b such that for $\eta = \frac{\alpha\varepsilon}{4}$

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

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:

- $\mathfrak{R}_\sigma^F = \{\rho \in \mathsf{SR}_{n_0+n_b} \mid \pi(\rho) = \sigma_0\sigma \wedge \mathsf{last}(\rho) \in Q_f\}$;
- $\mathfrak{R}_\sigma^C = \{\rho \in \mathsf{SR}_{n_0+n_b} \mid \pi(\rho) = \sigma_0\sigma \wedge \mathsf{last}(\rho) \in Q_c$ and belongs to a BSCC$\}$;

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

Let $Q_c^\sigma = \{\mathsf{last}(\rho) \mid \rho \in \mathfrak{R}_\sigma^C\}$ and $Q_f^\sigma = \{\mathsf{last}(\rho) \mid \rho \in \mathfrak{R}_\sigma^F\}$. For every 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 measure. Due to Lemma 2, for all $\alpha' > 0$, $\varepsilon' > 0$, 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. $p_f^{\pi(\rho)}$) to be 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 $p_t^{\pi(\rho)}$ to 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:
$\widetilde{\mathfrak{R}}_\sigma^n = \{\rho \in \mathsf{SR}_n \mid \exists \rho' \in \mathfrak{R}_\sigma^F \wedge \rho' \preceq \rho \wedge \frac{p_c^{\pi(\rho)} + p_t^{\pi(\rho)}}{p_c^{\pi(\rho)} + p_t^{\pi(\rho)} + p_f^{\pi(\rho)}} > \varepsilon \wedge \frac{p_c^{\pi(\rho)}}{p_t^{\pi(\rho)} + p_f^{\pi(\rho)}} \leq \frac{\varepsilon}{2}\}$
For $\rho \in \widetilde{\mathfrak{R}}_\sigma^n$, one gets by algebraic operations, $\frac{2p_t^{\pi(\rho)}}{\varepsilon} > p_f^{\pi(\rho)}$.
Thus $\mathbb{P}(\widetilde{\mathfrak{R}}_\sigma^n) < \frac{2\mathbb{P}(\mathfrak{R}_\sigma^T)}{\varepsilon}$ and $\sum_{\sigma \in \mathfrak{S}} \mathbb{P}(\widetilde{\mathfrak{R}}_\sigma^n) < \frac{2\sum_{\sigma \in \mathfrak{S}} \mathbb{P}(\mathfrak{R}_\sigma^T)}{\varepsilon}$.
Due to the choice of n_b, $\sum_{\sigma \in \mathfrak{S}} \mathbb{P}(\mathfrak{R}_\sigma^T) < \eta$, and we derive $\sum_{\sigma \in \mathfrak{S}} \mathbb{P}(\widetilde{\mathfrak{R}}_\sigma^n) < \frac{2\eta}{\varepsilon} = \frac{\alpha}{2}$.

Summarising 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 the AFF-diagnosability of $\mathcal{A}$. □

Appendix C.

This appendix contains proofs that are omitted in section 4.

Proposition 7. *The* FA- *and* IA*-diagnosability problems are decidable in* PSPACE.

Proof. Similarly to the proof of Proposition 6, we use the characterisation of FA-diagnosability given in Proposition 3 without explicitly building the product pLTS $\mathcal{A}_{\mathsf{FA}}$. Here also, we heavily use Savitch's theorem. First given a state (q, U, V) of $\mathcal{A}_{\mathsf{FA}}$ we can check in polynomial space whether it belongs to a BSCC (as in the proof of Proposition 6). We can also check in polynomial space whether it can be reached from some state (q', U', V') with $U' = \emptyset$ or $V' = \emptyset$ by guessing such a state. Combining the two, this provides a polynomial space algorithm to check whether (q, U, V) belongs to a BSCC in which no state (q', U', V') fulfils $U' \neq \emptyset$ and $V' \neq \emptyset$.
Thus the procedure that decides whether $\mathcal{A}$ is not FA-diagnosable consists in guessing a state $s = (q, U, V)$, checking that it is reachable from s_0 and belongs to a BSCC where all states (q', U', V') of the BSCC fulfil $U' \neq \emptyset$ and $V' \neq \emptyset$.

We use the characterisation of IA-diagnosability given in Proposition 4 without building explicitly the product pLTS $\mathcal{A}_{\mathsf{IA}}$, which is exponential in the size of $\mathcal{A}$. First, given a state (q, U, V, W) of $\mathcal{A}_{\mathsf{IA}}$, we can check in polynomial space that it belongs to a BSCC (as in the proof of Proposition 6). We can also check in polynomial space whether it is coreachable from a state (q', U', V', W') that fulfils $U' = \emptyset$ or $W' = \emptyset$ by guessing such a state (we use Savitch theorem here). Combining the two procedures, we can check in polynomial space whether (q, U, V, W) belongs to a BSCC where all states (q', U', V', W') of the BSCC fulfil $U' \neq \emptyset$ and $W' \neq \emptyset$.
Thus the procedure that decides whether $\mathcal{A}$ is not IA-diagnosable consists in guessing a state $s = (q, U, V, W)$, checking that it is reachable from s_0 and belongs to a BSCC where all states (q', U', V', W') of the BSCC fulfil $U' \neq \emptyset$ and $W' \neq \emptyset$. □

Appendix D.

This appendix contains proofs that are omitted in section 5.

Proposition 11. *A finite pLTS $\mathcal{A}$ is* FA*-diagnosable if and only if it admits an* FA*-diagnoser. Furthermore when $\mathcal{A}$ is* FA*-diagnosable, one can build an* FA*-diagnoser with at most 2^n memory states.*

Proof. Assume first that there exists an FA-diagnoser D for $\mathcal{A}$. For every $n \in \mathbb{N}$, we define $\mathsf{FD}_n = \{\rho \in \Omega \mid \forall m \geq n, D(\pi(\rho_{\downarrow m})) = \top\}$ the set of runs that are diagnosed faulty after n observed events, and symmetrically $\mathsf{CD}_n = \{\rho \in \Omega \mid \forall m \geq n, D(\pi(\rho_{\downarrow m})) = \bot\}$ the set of runs that are persistently diagnosed correct after n observed events. The sequences $(\mathsf{CD}_n)_{n\in\mathbb{N}}$ and $(\mathsf{FD}_n)_{n\in\mathbb{N}}$ are non-decreasing. As $? \prec \bot$ and $? \prec \top$, for every run $\rho \in \Omega$, $D_{\inf}(\pi(\rho)) =?$ is equivalent to $\rho \notin \bigcup_n(\mathsf{FD}_n \cup \mathsf{CD}_n)$. Thus $\bigcup_{n\in\mathbb{N}}(\mathsf{FD}_n \cup \mathsf{CD}_n) = \{\rho \in \Omega \mid D_{\inf}(\pi(\rho)) \neq ?\}$. Since D is reactive, $\mathbb{P}(\{\rho \in \Omega \mid D_{\inf}(\pi(\rho)) \neq ?\}) = 1$. Moreover, since D is correct, for every $n \in \mathbb{N}$, $\mathsf{FD}_n \subseteq \mathsf{Sf}_n$ and $\mathsf{CD}_n \subseteq \mathsf{C}_n \setminus \mathsf{CAmb}_n$. Thus for every $n \in \mathbb{N}$, $\mathbb{P}(\mathsf{FAmb}_n \cup \mathsf{CAmb}_n) = 1 - \mathbb{P}(\mathsf{Sf}_n \cup \mathsf{C}_n \setminus \mathsf{CAmb}_n) \leq 1 - \mathbb{P}(\mathsf{FD}_n \cup \mathsf{CD}_n)$ and $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}_n \cup \mathsf{CAmb}_n) \leq 1 - \liminf_{n\to\infty} \mathbb{P}(\{\rho \in \mathsf{SR}_n \mid D(\pi(\rho)) \neq ?\}) = 0$. This shows that $\mathcal{A}$ is FA-diagnosable.

Assume now that $\mathcal{A}$ is FA-diagnosable. From $\mathsf{FA}(\mathcal{A}) = (Q^*, \Sigma_o, T^*, (\{q_0\}, \emptyset))$ the FA-automaton of $\mathcal{A}$, we define $D = (M, \Sigma, m_0, \mathsf{up}, D_{fm})$ the finite memory diagnoser where $M = Q^*$, $m_0 = (\{q_0\}, \emptyset)$, $\mathsf{up}(m, a) = T^*(m, a)$, $D_{fm}((U, V)) = \top$ iff $U = \emptyset$ and $D_{fm}((U, V)) = \bot$ iff $V = \emptyset$. Let us check that D is an FA-diagnoser, and that its size is at most 2^n if n denotes the number of states of $\mathcal{A}$.

commitment When U is empty it remains empty forever which implies commitment.

correctness Let $w \in \Sigma_o^*$ be an observed sequence. If (U, V) is the state in $\mathsf{FA}(\mathcal{A})$ reached after reading w, then recall that U (resp. V) is the set of states in $\mathcal{A}$ that can be reached by correct (resp. faulty) signalling runs labelled by w. By construction, if $D(w) = \top$ then w is surely faulty, and if $D(w) = \bot$ then w is surely correct.

reactivity Let ρ be a signalling run such that $D(\pi(\rho)) =?$. Due to the characterisation of Proposition 3, the SCC of $\mathcal{A}_{\mathsf{FA}}$ that ρ has reached cannot be a BSCC. So given some n, $\mathbb{P}(\{\rho \in \Omega \mid \exists m \geq n\ D(\pi(\rho_{\downarrow m}) =?\}) \leq \mathbb{P}(\{\rho \in \Omega \mid \rho_{\downarrow n} \text{ does not reach a BSCC}\})$. Thus $\mathbb{P}(\{\rho \in \Omega \mid D_{\inf}(\pi(\rho)) =?\}) = \lim_{n\to\infty} \mathbb{P}(\{\rho \in \Omega \mid \exists m \geq n\ D(\pi(\rho_{\downarrow m}) =?\}) \leq \limsup_{n\to\infty} \mathbb{P}(\{\rho \in \Omega \mid \rho_{\downarrow n} \text{ does not reach a BSCC}\}) = 0$.

size D has at most 2^n memory states because every state of $\mathsf{FA}(\mathcal{A})$ consists of a pair (U, V) with $U \subseteq Q_c$ and $V \subseteq Q_f$.

□

Proposition 13. *A finite pLTS $\mathcal{A}$ is* IA*-diagnosable if and only if it admits an* IA*-diagnoser. Furthermore when $\mathcal{A}$ has n_c correct states, n_f faulty states and is* IA*-diagnosable, one can build an* IA*-diagnoser with at most $2^{n_c}3^{n_f}$ states.*

Proof. Assume first that there exists an IA-diagnoser D for $\mathcal{A}$, and let ρ be an infinite run. By reactivity, almost surely $D_{\sup}(\pi(\rho)) \in \{\top, \bot\}$. If $D_{\sup}(\pi(\rho)) = \top$ then there exists some n such that $D(\pi(\rho_{\downarrow n})) = \top$. By correctness, $\rho_{\downarrow n}$ is surely faulty and thus ρ is surely faulty. If $D_{\sup}(\pi(\rho)) = \bot$, we claim that ρ is surely correct. Observe that the diagnoser infinitely often outputs $\bot$, so by correctness, for all n, $\pi(\rho_{\downarrow n})$ is surely correct and thus in particular $\rho_{\downarrow n}$ is correct. Assume there exists an infinite faulty run ρ' with $\pi(\rho') = \pi(\rho)$. There exists a n such that for all $m \geq n$, $\rho'_{\downarrow n}$ is faulty. Thus by correctness there can be no more n $\bot$ verdicts for $\pi(\rho)$ contradicting the fact that $D_{\sup}(\pi(\rho)) = \bot$. Thus with probability 1, an infinite run is unambiguous.

Assume now that $\mathcal{A}$ is IA-diagnosable, and denote $\mathsf{IA}(\mathcal{A})$ its IA-automaton. For any word $w \in \Sigma_o^*$, we denote by (U_w, V_w, W_w) the state in $\mathsf{IA}(\mathcal{A})$ reached after reading w. For every finite signalling run ρ of $\mathcal{A}$, we denote by $(U_\rho, V_\rho, W_\rho) = (U_{\pi(\rho)}, V_{\pi(\rho)}, W_{\pi(\rho)})$. The function D is then defined as follows: $D(w) = \top$ iff $U_w = \emptyset$, $D(w) = \bot$ iff $W_w = \emptyset$ and $U_w \neq \emptyset$, and in all other cases $D(w) =?$. Let us prove that D is indeed an IA-diagnoser for $\mathcal{A}$.

commitment. When U_w is empty it remains empty forever which implies commitment.

correctness. For any word w, if $U_w = \emptyset$, by construction of $\mathsf{IA}(\mathcal{A})$, w is surely faulty. Assume now that $W_w = \emptyset$ and $U_w \neq \emptyset$. Let w' the greatest proper prefix of w such that $W_{w'} = \emptyset$. Let ρ be any signalling run with $\pi(\rho) = w$. Assume that $\rho_{\downarrow |w'|}$ is faulty. Thus the states visited by $\rho_{\downarrow n}$ for $|w'| < n \leq |w|$ were always in $W_{\rho_{\downarrow n}}$. Since $W_w = \emptyset$, this not possible and so $\rho_{\downarrow |w'|}$ is correct. Thus every time a state with $W = \emptyset$, the length of the greatest prefix, for which all signalling subruns corresponding to this prefix are correct, is increased. This establishes correctness.

reactivity. Let ρ be an infinite run such that $D_{sup}(\pi(\rho)) =?$. Due to the characterisation of Proposition 4, either (1) the SCC of $\mathcal{A}_{\mathsf{IA}}$ that ρ infinitely often visits is not a BSCC or (2) ρ does not visit infinitely often all the states of this SCC. The probability of such runs is null which establishes the reactivity.

□

Proposition 15. *Let $\varepsilon > 0$. A finite pLTS $\mathcal{A}$ is* $\varepsilon\mathsf{FF}$*-diagnosable if and only if it admits an* $\varepsilon\mathsf{FF}$*-diagnoser.*

Proof. Let $\mathcal{A}$ be a pLTS. Assume that $\mathcal{A}$ is $\varepsilon\mathsf{FF}$-diagnosable. Let D be the diagnoser defined by: for all $w \in \Sigma_o^*$, $D(w) = \top$ iff $\mathsf{CorP}(w) \leq \varepsilon$. Such an $\varepsilon\mathsf{FF}$-diagnoser is correct by definition. Let $\alpha > 0$. Since $(\mathsf{F}_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence converging to F_∞, there exists $n_0 \in \mathbb{N}$ such that for all

$n \geq n_0, \mathbb{P}(\mathsf{F}_n \setminus \mathsf{F}_{n_0}) < \alpha/2$. By εFF-diagnosability of $\mathcal{A}$, for all $\rho \in \bigcup_{k \leq n_0} \mathsf{minF}_k$, there exists n_ρ such that for all $n \geq n_\rho$

$$\mathbb{P}(\mathsf{Cyl}(\rho) \cap \mathsf{FAmb}^{\varepsilon}_{n+|\rho|_o}) \leq \frac{\alpha}{2} \cdot \mathbb{P}(\rho).$$

Define $n_{max} = \max_{\rho \in \bigcup_{n \leq n_0} \mathsf{minF}_n} n_\rho$. Then for $n \geq n_0 + n_{max}$ we have

$$\begin{aligned}
\mathbb{P}(\mathsf{FAmb}^{\varepsilon}_n) &\leq \mathbb{P}(\mathsf{FAmb}^{\varepsilon}_n \cap \mathsf{F}_{n_0}) + \mathbb{P}(\mathsf{FAmb}^{\varepsilon}_n \setminus \mathsf{F}_{n_0}) \\
&\leq \sum_{\rho \in \bigcup_{k \leq n_0} \mathsf{minF}_k} \mathbb{P}(\mathsf{Cyl}(\rho) \cap \mathsf{FAmb}^{\varepsilon}_n) + \mathbb{P}(\mathsf{F}_n \setminus \mathsf{F}_{n_0}) \\
&\leq \sum_{\rho \in \bigcup_{k \leq n_0} \mathsf{minF}_k} \frac{\alpha}{2} \cdot \mathbb{P}(\rho) + \frac{\alpha}{2} \leq \alpha
\end{aligned}$$

So we have established that $\lim_{n\to\infty} \mathbb{P}(\mathsf{FAmb}^{\varepsilon}_n) = 0$.
By definition of D, $\mathsf{FAmb}^{\varepsilon}_n = \{\rho \in \mathsf{F}_n \mid D(\pi(\rho)) = ?\}$. Thus D is reactive.
Conversely assume that there exists an εFF-diagnoser D for $\mathcal{A}$.
Let ρ be a minimal faulty run and $\alpha > 0$.
Since D is reactive, $\lim_{n\to\infty} \mathbb{P}(\{\rho' \in \mathsf{F}_n \mid D(\pi(\rho')) = ?\}) = 0$.
So there exists $n_{\rho,\alpha} \in \mathbb{N}$ such that for all $n \geq n_{\rho,\alpha}$,

$$\mathbb{P}(\{\rho' \in \mathsf{F}_n \mid D(\pi(\rho')) = ?\}) \leq \alpha \cdot \mathbb{P}(\rho) \ .$$

Thus for all $n \geq n_{\rho,\alpha}$:

$$\begin{aligned}
\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid D(\pi(\rho')) = ? \wedge \rho \preceq \rho'\}) \leq \ & \mathbb{P}(\{\rho' \in \mathsf{F}_{n+|\rho|_o} \mid D(\pi(\rho')) = ?\}) \\
\leq \ & \alpha \cdot \mathbb{P}(\rho) \ .
\end{aligned}$$

Since D is correct,
$\mathsf{Cyl}(\rho) \cap \mathsf{FAmb}^{\varepsilon}_{n+|\rho|_o} \subseteq \mathsf{Cyl}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid D(\pi(\rho')) = ? \wedge \rho \preceq \rho'\})$. This establishes that $\mathcal{A}$ is εFF-diagnosable. □

Appendix E.

This appendix contains proofs that are omitted in section 6.

Theorem 5. *The uniform* AFF*-diagnosability problem is undecidable for pLTS.*

Proof. Here we consider a probabilistic automaton with an initial distribution **I**. Observe that the emptiness is undecidable even when for every word w, $1/4 \leq \mathbf{Pr}_{\mathcal{A}}(w) \leq 3/4$. Let $\mathcal{A}$ be such a probabilistic automaton. Define the pLTS $\mathcal{A}' = \langle Q', q_0, \Sigma', T', \mathbf{P}' \rangle$ as follows.

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

- $T' = \{(q_0, u, q^u), (q_0, u, q^f) \mid q \in Q, \mathbf{I}[q] > 0\}$
  $\cup \{(q^u, a, q'^u), (q^f, a, q'^f) \mid q, q' \in Q, a \in \Sigma, \mathbf{P}_a[q, q'] > 0\}$
  $\cup \{(q^u, \sharp, q'^u) \mid q \in F, q' \in Q, \mathbf{I}[q'] > 0\}$
  $\cup \{(q^f, \sharp, q'^f) \mid q \in Q \setminus F, q' \in Q, \mathbf{I}[q'] > 0\}$
  $\cup \{(q^u, \flat, b^u) \mid q \in Q \setminus F\} \cup \{(q^f, \mathbf{f}, b^f) \mid q \in F\}$
  $\cup \{b^u, \sharp, b^u\} \cup \{b^u, \flat, b^u\} \cup \{b^f, \flat, b^f\}$

- $\mathbf{P}'$ is defined by:
  - For all $(q_0, u, q^u), (q_0, u, q^f) \in T'$, $\mathbf{P}'(q_0, u, q^u) = \mathbf{P}'(q_0, u, q^f) = \frac{\mathbf{I}[q]}{2}$;
  - 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 $(q^f, \mathbf{f}, b^f) \in T'$, $\mathbf{P}'(q^f, \mathbf{f}, b^f) = \frac{1}{1+|\Sigma|}$;
  - $\mathbf{P}'(b^u, \flat, b^u) = \mathbf{P}'(b^u, \sharp, b^u) = \frac{1}{2}$, $\mathbf{P}'(b^f, \flat, b^f) = 1$.

This reduction is illustrated in Figure E.22. $\mathbf{P}'$ fulfils the requirement for pLTS. For instance, let $q \in Q$,

$$\sum_{(q^f,a,q') \in T'} \mathbf{P}'(q^f, a, q') = \sum_{a \in \Sigma} \sum_{q' \in Q} \frac{\mathbf{P}_a[q, q']}{1 + |\Sigma|} + \sum_{q' \in Q} \frac{\mathbf{I}[q']}{1 + |\Sigma|} = \frac{|\Sigma|}{1 + |\Sigma|} + \frac{1}{1 + |\Sigma|} = 1.$$

We claim that $\mathcal{A}'$ is uniformly AFF-diagnosable if and only if $\mathcal{L}_{\mathcal{A},1/2} = \emptyset$.
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 $\mathbf{Pr}_{\mathcal{A}}(w) > 1/2$. We will prove that $\mathcal{A}'$ is not uniformly $\frac{1}{2}\mathsf{FF}$-diagnosable. So we pick arbitrary $0 < \alpha < 1$ and n_α.
Consider the observed sequence $\sigma_n = (w\sharp)^n\flat$ for some n to be fixed later. Due to our hypothesis on $\mathcal{A}$, it is ambiguous. Let

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

Since $\mathbf{Pr}_{\mathcal{A}}(w) > 1/2$, γ_n fulfils $\lim_{n\to\infty} \gamma_n = \infty$.
Let ρ_n be a minimal faulty run with $\pi(\rho_n) = \sigma_n$. Let ρ be a signalling run extending ρ_n with $|\rho|_o = |\rho_n|_o + n_\alpha$. Then $\pi(\rho) = \sigma_n \sigma \flat^{n_\alpha}$. By a straightforward examination of $\mathcal{A}'$ one gets:

$$\frac{\mathbb{P}(\{\rho' \in \mathsf{C} \mid \pi(\rho') = \pi(\rho_n)\flat^{n_\alpha}\})}{\mathbb{P}(\{\rho' \in \mathsf{F} \mid \pi(\rho') = \pi(\rho_n)\flat^{n_\alpha}\})} = \frac{\gamma_n 2^{-n_\alpha}}{1 + \gamma_n 2^{-n_\alpha}} \quad .$$

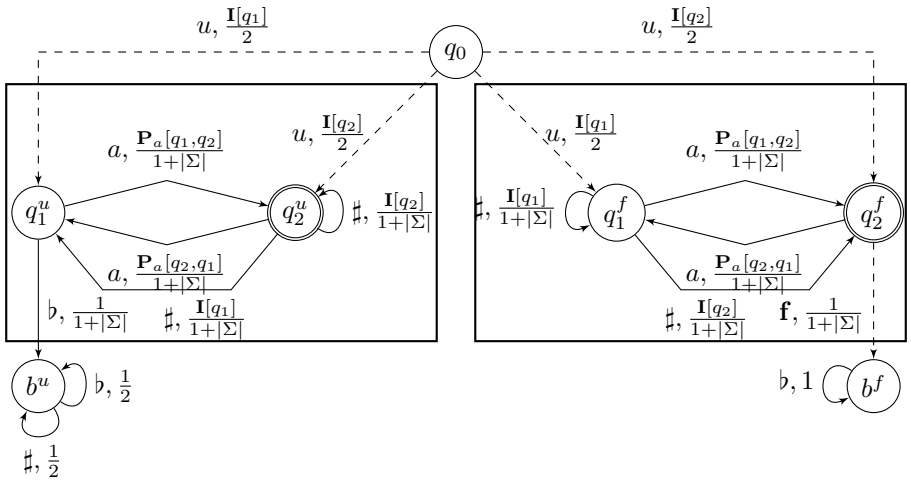


Figure E.22: From probabilistic automata to pLTS: rectangles surround the two copies of Q.

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

• Conversely assume that for every word $w \in \Sigma^*$, $\mathbf{Pr}_{\mathcal{A}}(w) \leq 1/2$. Combining this assumption with the hypothesis that in $\mathcal{A}$, $\mathbf{Pr}_{\mathcal{A}}(w) \geq 1/4$, one deduces that for every observed sequence $\sigma \in (\Sigma \cup \{\sharp\})^*\flat$, $\mathsf{CorP}(\sigma) \leq 3/4$. On the other hand, for every minimal faulty run ρ, $\pi(\rho) \in (\Sigma \cup \{\sharp\})^*\flat$.

Pick any positive ε, α and consider an arbitrary minimal faulty run ρ. The observed sequence σ' of a faulty run ρ' that extends ρ fulfils $\sigma' = \pi(\rho)\flat^n$ for some n. 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 halved. Thus choosing n_α such that $\frac{3 \cdot 2^{-n_\alpha}}{1+3 \cdot 2^{-n_\alpha}} \leq \varepsilon$, for all $n \geq n_\alpha$:

$$\mathbb{P}(\{\rho' \in \mathsf{SR}_{n+|\rho|_o} \mid \rho \preceq \rho' \wedge \mathsf{CorP}(\pi(\rho)) \leq \varepsilon) = \mathbb{P}(\rho) \geq (1-\alpha)\mathbb{P}(\rho) \ .$$

Thus $\mathcal{A}'$ is uniformly εFF-diagnosable and since ε was arbitrarily chosen $\mathcal{A}'$ is uniformly AFF-diagnosable. □

Proposition 17. *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.*

Proof. Let $\mathcal{A} = (Q, \Sigma, T, q_0, F)$ be an NFA. Starting from $\mathcal{A}$, one builds in polynomial time the NFA $\mathcal{A}' = (Q', \Sigma', T', q_0, Q')$ where $\Sigma' = \Sigma \uplus \{\sharp\}$, $Q' = Q \uplus \{s\}$, and

$$T' = T \cup \{(q, \sharp, q_0) \mid q \in F\} \cup \{(s, a, s) \mid a \in \Sigma\} \cup \{(q, a, s) \mid a \in \Sigma,\ q \not\to_{\mathcal{A}}\} \ .$$

• Assume that $\mathcal{L}(\mathcal{A}) = \Sigma^*$. Any word w over the alphabet Σ' can be decomposed into $w = w_1 \sharp w_2 \sharp \dots \sharp w_n$ with $w_i \in \Sigma^*$. For each factor w_i, since $\mathcal{A}$ is universal, there exists a run ρ_i on w ending in some terminal state q_i in $\mathcal{A}$. Therefore w is accepted in $\mathcal{A}'$ by the run $\rho_1 \sharp \rho_2 \sharp \dots \sharp \rho_n$. Hence $\mathcal{A}'$ is universal, and thus eventually universal: $\varepsilon^{-1}\mathcal{L}(\mathcal{A}') = \Sigma'^*$.

• Conversely assume that $\mathcal{A}'$ is eventually universal and let $v \in \Sigma'^*$ be such that $v^{-1}\mathcal{L}(\mathcal{A}') = \Sigma'^*$. Let ρ be a run ending in an accepting state with observation v. Given $w \in \Sigma^*$, we consider the word $w' = v \sharp w \sharp$. Since $\mathcal{A}'$ is eventually universal with witness v, $w' \in \mathcal{L}(\mathcal{A}')$ and there exists an accepting run that can be decomposed as: $\rho \sharp \rho' \sharp q_0$ where run ρ' which corresponds to word w has q_0 as initial state, ends in a final state of $\mathcal{A}$, and only uses transitions of $\mathcal{A}$. So ρ' is a run of $\mathcal{A}$ that accepts w. Therefore $w \in \mathcal{L}(\mathcal{A})$, and $\mathcal{A}$ is universal. □