Approximate Diagnosis and Opacity of Stochastic Systems
Engel Lefaucheux

To cite this version:
Engel Lefaucheux. Approximate Diagnosis and Opacity of Stochastic Systems. 2019. hal-02424744

HAL Id: hal-02424744
https://hal.science/hal-02424744
Preprint submitted on 28 Dec 2019

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Approximate Diagnosis and Opacity of Stochastic Systems

Engel Lefauheux
Max-Planck Institute for Software Systems, Germany
elefauch@mpi-sws.org

Abstract. We consider information control questions in a stochastic setting where an observation function provides to an external observer a view of the states along paths and relevant paths are those visiting some state from a fixed subset. Exact disclosure occurs when the observer can deduce from a finite observation that the path is relevant, the approximate disclosure variant corresponding to the execution being identified as relevant with arbitrarily high accuracy. We consider the problems of diagnosability and opacity, which corresponds, in spirit, to the cases where one wants to disclose all the information or hide as much of it as possible. While these problems have already been studied for the exact disclosure notion, there are very few works using the approximate disclosure. We establish that opacity of Markov chains is in \text{EXPTIME} and \text{PSPACE}-hard while nearly every opacity question is undecidable in an active setting. Moreover, we show that diagnosability is \text{EXPTIME}-complete for controllable systems.

1 Introduction

\textit{Diagnosis and Opacity} Due to the omnipresence of communicating devices, controlling the information produced by a system has seen an increasing importance. This control mainly takes two direction. First, it can be done in order to detect some internal behaviour, such as malfunctions, of the system. This direction has been formalised under the name diagnosis for systems modelled by partially observable labelled transition systems (LTS) \cite{22}. In such a framework, diagnosability requires that the occurrence of unobservable faults can be deduced accurately from the previous and subsequent observable events. Diagnosability for LTS was shown to be decidable in \text{PTIME} \cite{17}. Also, several contributions, gathered under the generic name of active diagnosis, focus on enforcing the diagnosability of a system \cite{12,21,24,13}. The second direction of information control aims at hiding a secret behaviour of the system. This property, called opacity, is motivated by security: an external user should not, by observing an execution of a system, acquire the guarantee that it is a secret one. This property was formalised for LTS \cite{11} by specifying a subset of secret paths and requiring that, for any secret path, there is a non-secret one with the same observation. Both diagnosability and opacity thus distinguish a set of relevant path. The \textit{disclosure set} of a system is then the set of (relevant) paths that can be identified as such.
In order to quantify the size of the leak, various measures for the disclosure set, called probabilistic disclosure, were introduced \[20, 1, 5, 3\]. Moreover, in stochastic systems, the notion of disclosure of a relevance of a path can depend on the probabilities. There are three natural variants: exact disclosure, which as in the non-stochastic case require that no non-relevant path share the same observation, $\varepsilon$-disclosure for $\varepsilon > 0$ which tolerates small errors, allowing to claim the relevance of a path if the conditional probability that the path is relevant exceeds $1 - \varepsilon$, and Accurate Approximate disclosure (AA-disclosure) which is satisfied when the accuracy of the relevance guess can be selected arbitrarily high. Under the exact notion of disclosure, both diagnosability and opacity have been studied extensively for stochastic systems \[23, 7, 6, 4, 2\]. In particular, the various exact notions of diagnosability have been shown to be PSPACE-complete for observable Markov chains (oMC). The study of the approximate notions of disclosure has however been more limited, especially the notion of AA-disclosure. The most notable result showed that diagnosability under AA-disclosure is decidable in PTIME \[8\] for oMC.

**Contribution** In this paper, we study diagnosability and opacity in stochastic systems under AA-disclosure.

- we formally introduce a notion of accurate approximate opacity that mirrors the existing diagnosability notion (Definition 7);
- we show that opacity with AA-disclosure for oMC is in EXPSPACE and PSPACE-hard (Theorem 11);
- we establish that diagnosability with AA-disclosure for weighted Markov chains, a controllable setting, is EXPSPACE-complete (Theorem 16);
- we prove the undecidability of most notions of opacity with AA-disclosure for observable Markov decision processes (Theorem 25 and ??).

**Organisation** In Section 2, we define and discuss the notions of disclosure, diagnosability as well as opacity. We also gather and complete the results on approximate diagnosability and opacity in a passive framework. Then, in Section 3, we consider diagnosability for weighted Markov chains, a framework giving a partial external control on the system. Similarly, in Section 4, we study opacity for observable Markov decision processes, a setting where the control is more powerful than the one in weighted Markov chains due to being internal to the system. Finally, in Section 5 we conclude and give some perspectives to this work. For readability concerns, the most technical proofs are deferred to the appendix.

## 2 Diagnosis and Opacity for Markov Chains

For a finite alphabet $\Sigma$, we denote by $\Sigma^*$ (resp. $\Sigma^\omega$) the set of finite (resp. infinite) words over $\Sigma$, $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$ and $\varepsilon$ the empty word. The length of a word $w$ is denoted by $|w| \in \mathbb{N} \cup \{\infty\}$ and for $n \in \mathbb{N}$, $\Sigma^n$ is the set of words of length $n$. A word $u \in \Sigma^*$ is a prefix of $v \in \Sigma^\infty$, written $u \preceq v$, if $v = uw$ for some $w \in \Sigma^\infty$. The prefix is strict if $w \neq \varepsilon$. For $n \leq |w|$, we write $w_{1,n}$ for the prefix of length $n$. 
of \( w \). Given a countable set \( S \), a distribution on \( S \) is a mapping \( \mu : S \to [0,1] \)
such that \( \sum_{s \in S} \mu(s) = 1 \). The support of \( \mu \) is \( \text{Supp}(\mu) = \{ s \in S \mid \mu(s) > 0 \} \). If \( \text{Supp}(\mu) = \{ s \} \) is a single element, \( \mu \) is a Dirac distribution on \( s \) written \( \delta_s \). We denote by \( \text{Dist}(S) \) the set of distributions on \( S \).

For the purpose of information control questions, the model must be equipped with an \textit{observation function} describing what an external observer can see. The observation function can be obtained via a labelling of states or transitions, both options being known to be equivalent. We thus define observable Markov chains (see Figure 1).

**Definition 1 (Observable Markov chains).** An observable Markov chain (oMC) over alphabet \( \Sigma \) is a tuple \( \mathcal{M} = (S, p, O) \) where \( S \) is a countable set of states, \( p : S \to \text{Dist}(S) \) is the transition function, and \( O : S \to \Sigma \) is the observation function.

We write \( p(s'|s) \) instead of \( p(s) (s') \) to emphasise the probability of going to state \( s' \) conditioned by being in state \( s \). Given a distribution \( \mu_0 \) on \( S \), we denote by \( \mathcal{M}(\mu_0) \) the oMC with initial distribution \( \mu_0 \). For decidability and complexity results, we assume that all probabilities occurring in the model (transition probabilities and initial distribution) are rationals. A (finite or infinite) path of \( \mathcal{M}(\mu_0) \) is a sequence of states \( \rho = s_0 s_1 \ldots \in S^\omega \) such that \( \mu_0(s_0) > 0 \) and for each \( i \geq 0 \), \( p(s_{i+1}|s_i) > 0 \). For a finite path, \( \rho = s_0 s_1 \ldots s_n \), we call \( n \) its length and denote its ending state by \( \text{last}(\rho) = s_n \). A finite path \( \rho_1 \) prefixes a finite or infinite path \( \rho \) if there exists a path \( \rho_2 \) such that \( \rho = \rho_1 \rho_2 \). We write \( \text{Cyl}(\rho) \) for the set of infinite paths prefixed by \( \rho \). We denote by \( \text{Path}(\mathcal{M}(\mu_0)) \) (resp. \( \text{FPath}(\mathcal{M}(\mu_0)) \)) the set of infinite (finite) paths of \( \mathcal{M}(\mu_0) \). The \textit{observation sequence} of the path \( \rho = s_0 s_1 \ldots \) is the word \( O(\rho) = O(s_0) O(s_1) \ldots \in \Sigma^\omega \). For a set \( R \) of paths, \( O(R) = \{ O(\rho) \mid \rho \in R \} \) and for a set \( W \) of observation sequences, \( O^{-1}(W) = \{ \rho \mid O(\rho) \in W \} \). We assume that \( \mathcal{M} = (S, p, O) \) is \textit{convergent}: each infinite path \( \rho \) has an infinite observation sequence \( O(\rho) \in \Sigma^\omega \).

![Fig. 1. An observable Markov chain with disclosure \( \frac{1}{4} \). The arrow entering the leftmost state means that the initial distribution is a Dirac on this state. Relevant states are circled twice.](image-url)
Forgetting the labels, an oMC with an initial distribution \( \mu_0 \) becomes a discrete time Markov chain (DTMC). In a DTMC, the set of infinite paths is the support of a probability measure extended from the probabilities of the cylinders by the Caratheodory’s extension theorem:

\[
P_{\mathcal{M}(\mu_0)}(\text{Cyl}(s_0s_1 \ldots s_n)) = \mu_0(s_0)p(s_1|s_0) \ldots p(s_n|s_{n-1})
\]

When \( \mathcal{M}(\mu_0) \) is clear from context, we will sometimes omit the subscript, and write \( P \) for \( P_{\mathcal{M}(\mu_0)} \). Let \( \rho \in \text{FPath}(\mathcal{M}) \), \( w \in \Sigma^* \) and \( E \subseteq \Sigma^\omega \), with a small abuse of notation we write \( P(\rho) \) for \( P(\text{Cyl}(\rho)) \), \( P(w) \) instead of \( P(\bigcup_{\rho \in O^{-1}(w)} \text{Cyl}(\rho)) \) and \( P(E) \) instead of \( P(\{\rho \in \text{Path}(\mathcal{M}(\mu_0)) \mid \rho \in O^{-1}(E)\}) \).

In this paper, we consider diagnosability and opacity problems; in both cases, one needs to identify a set of paths of the system which carries a hidden information (the paths are a faulty behavior of the system for diagnosability, and they are secret for opacity). We focus on the particular case where the relevant behavior of the system is given by a subset of states \( S' \subseteq S \), called relevant states, of the model: a (finite or infinite) path \( s_0s_1 \ldots \) is relevant if \( s_i \in S' \) for some \( i \). The set of infinite relevant paths is denoted \( \text{Rel} \).

**Remark 2.** Without loss of generality, we can assume that once a secret state has been reached by an execution, all subsequent states remain secret (see [2] for example)

With probabilistic systems, the set of paths disclosing that they are relevant depends on the level of confidence that the observer wants. To measure this, we define the proportion of relevant paths among those having the same observation sequence as follow:

**Definition 3 (Proportion of relevant paths).** Given an oMC \( \mathcal{M} = (S, p, O) \), an initial distribution \( \mu_0 \), \( S' \subseteq S \) and an observation sequence \( w \in \Sigma^* \), the proportion of relevant paths associated with the observation sequence \( w \) is:

\[
\rho_{\mathcal{M}(\mu_0)}^{\text{rel}}(w) = \frac{P(\{\rho \in O^{-1}(w) \mid \rho \in \text{Rel}\})}{P(w)}
\]

**Example 4.** Consider the oMC of Figure 1 and the observation sequences \( a^k, a^kb^n \) and \( a^kcm^m \), \( a^k \), for \( k > 1 \), can be produced by a non-relevant path with probability \( 1/2^{k-1} \) and by a relevant path with probability \( 1/2 \times 1/3^{k-2} \). Therefore, \( \rho_{\mathcal{M}(\mu_0)}^{\text{rel}}(a^k) = \frac{1/2^{k-1}}{1/2 + 1/3^{k-2}} \) which converges to 0 when \( k \) grows to infinity. The proportion of relevant paths for the observation \( a^kb^n \) with \( k > 1 \) and \( n \geq 1 \) is similarly \( \rho_{\mathcal{M}(\mu_0)}^{\text{rel}}(a^kb^n) = \frac{1/2^{k-1}}{1/2^{k-1} + 1/3^{k-2}} \) which remains constant for extensions of \( a^kb^n \) as it does not depend on \( n \). Finally, if \( m \geq 1 \), \( \rho_{\mathcal{M}(\mu_0)}^{\text{rel}}(a^kcm^m) = 1 \) as no non-relevant path can produce a 'c'.

Using this proportion, we define a measure on the quantity of information disclosed by a system. We first introduce a notion of approximate disclosure where one considers that a path reveals its relevance if the proportion of relevant paths of its observation sequence is greater than \( 1 - \varepsilon \) for some given \( \varepsilon > 0 \).
**Definition 5 (Approximate information control).** Given an oMC $\mathcal{M} = (S, p, O)$, an initial distribution $\mu_0$, $S' \subseteq S$ and $\varepsilon > 0$, an observation sequence $w \in \Sigma^*$ is $\varepsilon$-disclosing if $P^\rel_{\mathcal{M}(\mu_0)}(w) > 1 - \varepsilon$. It is $\varepsilon$-min-disclosing if it is $\varepsilon$-disclosing and no strict prefix of $w$ is $\varepsilon$-disclosing. Writing $D_{\min}^\varepsilon$ for the set of $\varepsilon$-min-disclosing observation sequences, the $\varepsilon$-disclosure is defined by

$$\text{Disc}^\varepsilon(\mathcal{M}(\mu_0)) = \sum_{w \in D_{\min}^\varepsilon} P\{\{\rho \in \text{Rel} | \exists \rho' \leq \rho, O(\rho') = w\}\}$$

This definition raises the two following decision problems for any $0 \leq \varepsilon < 1$:

- **For opacity:** the $\varepsilon$-disclosure problem consists in, given $\lambda \in [0; 1]$, deciding whether $\text{Disc}^\varepsilon(\mathcal{M}(\mu_0)) > \lambda$.
- **For diagnosis:** the $\varepsilon$-diagnosability problem consists in deciding whether $\text{Disc}^\varepsilon(\mathcal{M}(\mu_0)) = P(\text{Rel})$.

We can see an asymmetry between the problems introduced for opacity and for diagnosis here: in the former the threshold the $\varepsilon$-disclosure is compared to is given by the user while in the latter it is derived from the system. Fortunately, it is known that these problems are undecidable for $\varepsilon \neq 0$.

**Theorem 6.** Given $0 < \varepsilon < 1$, the positive $\varepsilon$-disclosure problem [2] and the $\varepsilon$-diagnosability problem [8] are undecidable for MCs.

In order to regain decidability one can consider slightly more qualitative notions of approximate information control, that we call accurate approximate. Instead of deeming the relevance to be revealed when the proportion of relevant path goes above a given threshold, an infinite observation sequence $w \in \Sigma^\omega$ is $\AA$-disclosing if this proportion converges toward 1. In other words, when observing an $\AA$-disclosing observation sequence, by waiting, one can get an arbitrarily high confidence that the path is relevant.

**Definition 7 (Accurate approximate information control).** Given an oMC $\mathcal{M} = (S, p, O)$, an initial distribution $\mu_0$, and $S' \subseteq S$, an observation sequence $w \in \Sigma^\omega$ is $\AA$-disclosing if $\lim_{n \to \infty} P^\rel_{\mathcal{M}(\mu_0)}(w_{\downarrow n}) = 1$. Writing $D^{\AA}$ for the set of $\AA$-disclosing observation sequences, the $\AA$-disclosure is defined by

$$\text{Disc}^{\AA}(\mathcal{M}(\mu_0)) = \sum_{w \in D^{\AA}} P\{\{\rho \in \text{Rel} | O(\rho) = w\}\}$$

As before, this definition raises two decision problems:

- **For opacity:** the $\AA$-disclosure problem consists in, given $\lambda \in [0; 1]$, deciding if $\text{Disc}^{\AA}(\mathcal{M}(\mu_0)) > \lambda$.
- **For diagnosis:** the $\AA$-diagnosability problem consists in deciding if $\text{Disc}^{\AA}(\mathcal{M}(\mu_0)) = P(\text{Rel})$.

1 The case $\varepsilon = 0$, with a non-strict inequality, is a form of exact disclosure for which some problems are decidable [2, 9].
AA-diagnosability was initially defined in [23] slightly differently: a system was called AA-diagnosable if it was $\varepsilon$-diagnosable for all $\varepsilon > 0$. We introduced the new definition with the study of active systems in mind. However, the two definitions are in fact equivalent.

**Proposition 8.** An oMC is AA-diagnosable iff it is $\varepsilon$-diagnosable for all $\varepsilon > 0$.

With the accurate approximate approach to information control, one gets decidability back. The AA-diagnosability problem for finite oMC was shown to be in PTIME in [8]. This result relies on the notion of distance between two oMC introduced in [15] and defined in the following way: the distance between two oMC $M_1$ and $M_2$ with initial distribution $\mu_1$ and $\mu_2$ is

$$d(M_1(\mu_1), M_2(\mu_2)) = \max_{E \subseteq \Sigma^*} P_{M_1(\mu_1)}(E) - P_{M_2(\mu_2)}(E).$$

The authors of [15] show how to decide in PTIME if the distance between two oMC is 1 thanks to the following characterisation.

**Proposition 9 ([15]).** Given two oMC $M_1$ and $M_2$ and two initial distributions $\mu_1$ and $\mu_2$, $d(M_1(\mu_1), M_2(\mu_2)) < 1$ iff there exists $w \in \Sigma^*$ and two distributions $\pi_1$ and $\pi_2$ such that, denoting $\mu^w_1$ and $\mu^w_2$ the probability distributions reached after observing $w$ in $M_1(\mu_1)$ and $M_2(\mu_2)$ respectively, we have, for $i \in \{1, 2\}$, $\text{Supp}(\pi_i) \subseteq \text{Supp}(\mu^w_i)$ and $d(M_1(\pi_1), M_2(\pi_2)) = 0$ (i.e. $\forall w' \in \Sigma^*, P_{M_1(\pi_1)}(w') = P_{M_2(\pi_2)}(w')$).

Finally, the link between the distance 1 of two oMC and AA-diagnosability was established in [8], giving the PTIME algorithm:

**Theorem 10 ([8]).** Let $M$ be a finite oMC and $\mu_0$ be an initial distribution. $M(\mu_0)$ is not AA-diagnosable iff there exist two states $s \in S'$ and $s' \in S \setminus S'$ with $s'$ belonging to a bottom strongly connected component (BSCC) of $M$ and there exist two finite paths $\rho$ and $\rho'$ of $\text{FPPath}(M(\mu_0))$ such that $\text{last}(\rho) = s$, $\text{last}(\rho') = s'$, $O(\rho) = O(\rho')$ and $d(M(1_s), M(1_{s'})) < 1$.

Considering only the sufficient condition, a stronger result, which we will need later, was in fact proven in [8]: Let $M$ be a (potentially infinite) oMC, $\mu_0$ be an initial distribution, two states $s \in S'$ and $s' \in S \setminus S'$ with $s'$ such that no relevant state can be reached from $s'$ and two finite paths $\rho$ and $\rho'$ of $\text{FPPath}(M(\mu_0))$ such that $\text{last}(\rho) = s$, $\text{last}(\rho') = s'$, $O(\rho) = O(\rho')$. Then $M(\mu_0)$ is AA-diagnosable implies that $d(M(1_s), M(1_{s'})) = 1$.

While AA-diagnosability can be decided in polynomial time, the AA-disclosure problem is a bit more complicated.

**Theorem 11.** The AA-disclosure problem for finite oMC is decidable in EXPTIME. It is PSPACE-hard.

**Proof.** Let us show how to solve the AA-disclosure problem in EXPTIME. We will first build an exponential size oMC which contains additional information compared to the original one, then show that for each BSCC, either almost
We define the initial distribution $\mu_0$. Therefore, computing the $\rho$-partition 9 and the correspondence between $M$ and $s$ can be in, given the produced observation sequence:

- $S' = S \times 2^S$;
- For $(s, B)$, $(s', B') \in S'$, $p'((s', B') \mid (s, B)) = p(s' \mid s)$ if $B' = \cup_{q \in B} \text{Supp}(p(q)) \cap O^{-1}(O(s'))$ else, $p'((s', B') \mid (s, B)) = 0$;
- For $(s, B) \in Q'$, $O(s, B) = O(s)$.

We define the initial distribution $\mu'_0$ for $M'$ by $\mu'_0(s, \text{Supp}(\mu_0) \cap O^{-1}(O(s))) = \mu_0(s)$ for all $s \in S$. There is a one-to-one correspondence between the paths of $\mathcal{M}(\mu_0)$ and $\mathcal{M}'(\mu'_0)$: every path $\rho = s_0s_1 \cdots s_n$ of $\mathcal{M}(\mu_0)$ is associated to an unique path $\rho' = (s_0, B_0)(s_1, B_1) \cdots (s_n, B_n)$ with $O(\rho) = O(\rho')$, $P_{\mathcal{M}(\mu_0)}(\rho) = P_{\mathcal{M}'(\mu'_0)}(\rho')$ and $B_n$ contains the set of states of $S$ that can be reached with a path of observation $O(\rho)$. Due to the latter property, $B_n$ only depends on $O(\rho)$ and is called the belief associated to $O(\rho)$.

Let $(s, B) \in S'$ such that $s \in S'$ and $(s, B)$ belongs to a BSCC of $M'$. Let us show that either for every path $\rho$ ending in $(s, B)$, $P\{\rho' \in \text{Path}(\mathcal{M}'(\mu'_0)) \mid \rho \preceq \rho' \wedge O(\rho') \in D^{AA} \} = 0$ or for every path $\rho$ ending in $(s, B)$, $P\{\rho' \in \text{Path}(\mathcal{M}'(\mu'_0)) \mid \rho \preceq \rho' \wedge O(\rho') \in D^{AA} \} = P(\rho)$. In other words, there are two categories of BSCC composed of relevant states: the ones that almost surely accurately disclose the relevance and the ones that do not accurately approximately disclose the relevance at all.

Let us assume for now that $B$ does not contain any relevant state beside $s$ and that $P\{\rho' \in \text{Path}(\mathcal{M}'(\mu'_0)) \mid \rho \preceq \rho' \wedge O(\rho') \in D^{AA} \} < P(\rho)$. As the characterisation of AA-diagnosability given in Theorem 10 depend on the support of the initial distribution but not on the exact values of the initial distribution, we know that for all distribution $\mu_1$ of support $B$, $\mathcal{M}(\mu_1)$ is not AA-diagnosable. Moreover, as $(s, B)$ belongs to a BSCC and by construction of $B$, the characterisation of AA-diagnosability also implies that there exists $s' \in S \setminus S'$ such that $s'$ belongs to a BSCC of $M'$ and $d(\mathcal{M}(1_s), \mathcal{M}(1_s')) < 1$. Using Proposition 9 and the correspondence between $M$ and $M'$, one deduces that there exists $\rho(s, B) \in \text{FPath}(\mathcal{M}(1_s, B))$ and $\alpha > 0$ such that for all $w \in \Sigma^*$ with $O(\rho) \leq w$

$$P_{\mathcal{M}'(1_{s, B})}(\{\rho' \in \text{FPath}(\mathcal{M}'(1_{s, B})) \mid \rho(s, B) \preceq \rho' \wedge O(\rho') = w\}) \leq \alpha P_{\mathcal{M}'(1_{s', B})}(\{\rho' \in \text{FPath}(\mathcal{M}'(1_{s', B})) \mid O(\rho') = w\}).$$
Therefore, for all $w \in \Sigma^*$ and initial distribution $\mu_1$ of support $B$ we have:

$$p_{\text{rel}}^{\text{M}^\prime}(\mu_1)(w) \leq \frac{P_{\text{M}^\prime(1_s,B)}(w)}{P_{\text{M}^\prime(1_s,B)}(w) + \frac{\mu_1(s)}{\mu_1(B)}P_{\text{M}^\prime(1_s,B)}(w)}$$

(1)

$$\epsilon_w + \sum_{\rho \in O(\rho \rightarrow s,B) \leq w} \frac{\alpha P_{\text{M}^\prime(1_s,B)}(\rho)}{P_{\text{M}^\prime(1_s,B)}(\rho) + \frac{\mu_1(s)}{\mu_1(B)}P_{\text{M}^\prime(1_s,B)}(\rho)} \leq \frac{P_{\text{M}^\prime(1_s,B)}(w) + \frac{\mu_1(s)}{\mu_1(B)}P_{\text{M}^\prime(1_s,B)}(w)}{P_{\text{M}^\prime(1_s,B)}(w) + \frac{\mu_1(s)}{\mu_1(B)}P_{\text{M}^\prime(1_s,B)}(w)}$$

(2)

where $\epsilon_w = P_{\text{M}^\prime(1_s,B)}(\{\rho \in \text{Path}(\text{M}(1_s,B)) \mid \exists \rho_1, \rho_2, \rho = \rho_1 \rho_2, O(\rho) = w\})$ and $w^\rho$ is such that $w = O(\rho)w^\rho$. As with probability 1, a run of $\text{M}^\prime(1_s,B)$ visits $(s,B)$ infinitely often, it will almost surely contain a $\rho_1(s,B)$ subrun, more precisely: the value $P_{\text{M}^\prime(1_s,B)}(w)$ almost surely converges to 0 when $|w|$ diverges to $\infty$. Let $w \in \Sigma^\omega$, if $P_{\text{rel}}^{\text{M}^\prime(\mu_1)}(w_{\downarrow n}) \xrightarrow{n \to \infty} 1$ then, for all $\rho$ such that $O(\rho \rightarrow s,B) \leq w$ we have that $P_{\text{M}^\prime(1_s,B)}(w_{\downarrow n})$ converges to 0, thus, due to Equation 2, $\epsilon_{w_{\downarrow n}}$ does not converge to 0. Therefore $P_{\text{rel}}^{\text{M}^\prime(\mu_1)}(w_{\downarrow n})$ almost surely does not converge to 1. This implies that $P\{\rho' \in \text{Path}(\text{M}^\prime(\mu_0)) \mid \rho \preceq \rho' \wedge O(\rho') \in D^{\text{AA}}\} = 0$.

Now let us consider the case where there exists another relevant state $s'' \in B$. Either we can deal with it similarly to what was done for $s$, or it is irrelevant to the computation:

- If $d(\text{M}(1_s), \text{M}(1_{s''})) < 1$, then as $(s,B)$ belongs to a BSCC, we can deduce that $d(\text{M}(1_s), \text{M}(1_{s''})) < 1$ and thus we can do the previous reasoning on both state simultaneously.
- If $d(\text{M}(1_s), \text{M}(1_{s''})) = 1$, for any distribution $\mu_1$ with support $B$, we have

$$P_{\text{M}^\prime(1_s,B)}(\{w \in \Sigma^\omega \mid \frac{P_{\text{M}^\prime(1_s,B)}(w_{\downarrow n})}{P_{\text{M}^\prime(1_s,B)}(w_{\downarrow n}) + \frac{\mu_1(s)}{\mu_1(B)}P_{\text{M}^\prime(1_s,B)}(w_{\downarrow n})} \xrightarrow{n \to \infty} 1\}) = 0.$$ 

Therefore, they almost surely do not impact the reasoning above.

This proves that there is a partition on the BSCC of $\text{M}^\prime$ between the ones that accurately approximately and almost surely disclose the relevance and the ones that do not accurately approximately disclose it at all. Moreover, using the PTIME AA-diagnosability test on the BSCC of $\text{M}^\prime$ one can effectively determine this partition. Therefore, one can obtain the value of $\text{Disc}^{\text{AA}}(\text{M}(\mu_0))$ by computing the probability to reach the disclosing BSCC, which is known to be in PTIME ($\text{M}^\prime$ being exponential in the size of $\text{M}$, this yields an EXPTIME algorithm). As $\text{Disc}^{\text{AA}}(\text{M}(\mu_0)) = \text{Disc}^{\text{AA}}(\text{M}(\mu_0'))$, this allows us to solve the AA-diagnosis problem.

The hardness is obtained by reduction from the universality problem for non-deterministic finite automaton (NFA), which is known to be PSPACE-complete [18].
3 Active Approximate Diagnosis

We will now consider an active setting where a controller can modify the behaviour of the system. Exact notions of diagnosis [6] and opacity [2] have been studied in an active setting. The framework used for each notion is not equivalent however as they do not give the same power to the controller. This difference comes from an intrinsic distinction between the two problems:

- Diagnosability corresponds to situations where one wants to obtain information from the system through exterior control. Therefore the controller is supposed to have the same amount of information as the diagnoser.

- For opacity on the contrary, the control either aims to diffuse an information outside of the system (case of a virus for example) or is implemented in the system during the design to protect it. In these two cases, the controller knows the exact state of the system.

Therefore we consider Weighted Markov Chains to study diagnosability and in the next section we will use Markov Decision Processes for opacity.

3.1 Diagnosis for Weighted Markov chains

**Definition 12 (WMC).** A weighted Markov Chain (WMC) over alphabet $\Sigma$ is a tuple $M = (S, T, O)$ where $S$ is a finite set of states, $T: S \times S \to \mathbb{N}$ is the transition function labelling transitions with integer weights and $O: S \to \Sigma$ is the observation function.

The alphabet is partitioned into controllable and uncontrollable events $\Sigma = \Sigma_c \cup \Sigma_e$. A set $\Sigma_0 \subseteq \Sigma$ of allowed events in a state $s \in S$ is a set of observations such that $\Sigma_e \subseteq \Sigma_0$ and $\{s' \in S | T(s, s') > 0 \land O(s') \in \Sigma_0\} \neq \emptyset$. Given a state $s$ and a set of allowed events $\Sigma_0$, we define the transition probability $p(s, \Sigma_0)$ such that for all $s'$ with $O(s') \in \Sigma_0$, $p(s, \Sigma_0)(s') = \frac{T(s, s')}{\sum_{s'' \in S, O(s'') \in \Sigma_0} T(s, s'')}$. As before, we write $p(s'|s, \Sigma_0)$ instead of $p(s, \Sigma_0)(s')$. Given an initial distribution $\mu_0$, an infinite path of a WMC $\rho = s_0 \Sigma_0 s_1 \Sigma_1 \ldots s_i$ is $K(\rho) = O(s_0) \Sigma_0 O(s_1) \Sigma_1 \ldots O(s_i)$.

The nondeterministic choice of the set of allowed events is resolved by strategies.

**Definition 13 (Strategy for WMC).** A strategy of WMC $M$ with initial distribution $\mu_0$ is a mapping $\sigma: (\Sigma \times 2^\Sigma)^* \Sigma \to \text{Dist}(2^\Sigma)$ associating to any knowledge sequence a distribution on sets of events.

We will only consider here strategies that do not generate a deadlock, i.e. strategies $\sigma$ such that in a state $s$ reached after a knowledge $b$, $\sigma(b)$ is a set of allowed
events for \( s \). Given a strategy \( \sigma \), a path \( \rho = s_0 \Sigma_0 s_1 \Sigma_1 \ldots \) of \( \mathcal{M} \) is \( \sigma \)-compatible if for all \( i \), \( \Sigma_i \in \text{Supp}(\sigma(K(s_0 \Sigma_0 s_1 \Sigma_1 \ldots s_i))) \). A strategy \( \sigma \) is deterministic if \( \sigma(b) \) is a Dirac distribution for each knowledge sequence \( b \). In this case, we denote by \( \sigma(b) \) the set of allowed actions \( \Sigma_m \in 2^{\Sigma} \) such that \( \sigma(b) = 1_{\Sigma_m} \). Let \( b \) be a knowledge sequence. We define \( B_{\mathcal{M}(\mu_0)}(b) \) the belief about states corresponding to \( b \) as follows:

\[
B_{\mathcal{M}(\mu_0)}(b) = \{ s \mid \exists \rho \in \text{FPath}(\mathcal{M}(\mu_0)), \ K(\rho) = b \land s = \text{last}(\rho) \}
\]

A strategy \( \sigma \) is belief-based if for all \( b \), \( \sigma(b) \) only depends on its belief \( B_{\mathcal{M}(\mu_0)}(b) \) (i.e. given two knowledge sequence \( b \) and \( b' \) if \( B_{\mathcal{M}(\mu_0)}(b) = B_{\mathcal{M}(\mu_0)}(b') \) then \( \sigma(b) = \sigma(b') \)). For belief-based strategies, we will sometimes write \( \sigma(B) \) for the choice of the strategy made for knowledge sequences producing the belief \( B \).

As for oMC, the secret is defined by the reachability of a set \( S' \subseteq S \) of secret states of the WMC and note that the construction ensuring that once a secret state is visited, the path remains secret forever, extends naturally from oMC to WMC. We consider only WMC of this form in the following.

A strategy \( \sigma \) on \( \mathcal{M}(\mu_0) \) defines an infinite Markov chain \( \mathcal{M}_\sigma(\mu_0) \) with set of states the finite \( \sigma \)-compatible paths, that can be equipped with the observation function associating \( \Sigma_{n-1} \mathcal{O}(s_n) \) with the state associated to the finite path \( \rho = s_0 \Sigma_0 \ldots \Sigma_{n-1} s_n \) (\( \Sigma_{n-1} \) being omitted if \( n = 0 \)). The transition function \( p_\sigma \) is defined for \( \rho \) a \( \sigma \)-compatible path and \( \rho' = \rho \Sigma_n s' \) by 

\[
p_\sigma(\rho' | \rho) = \sigma(\mathcal{K}(\rho))(\Sigma_n) p(s'|s, \Sigma_n) \text{ and we denote by } P_{\mathcal{M}_\sigma(\mu_0)} \text{ the associated probability measure.}
\]

When the strategy possesses some good regularity properties, this oMC is equivalent to a finite one (i.e. there is a one-to-one correspondence between the paths of each oMC, it preserves the knowledge sequence and the probability. The two oMC have therefore the same disclosure properties). For instance given a deterministic belief based strategy \( \sigma \), one can define the oMC \( \mathcal{M}'_\sigma \) with set of states \( S \times 2^\Sigma \times 2^S \), observation \( \mathcal{O}'_\sigma(s, \Sigma', B) = (\mathcal{O}(s), \Sigma') \), initial distribution \( \mu_0'(s, \emptyset, \text{Supp}(\mu_0) \cap \mathcal{O}^{-1}(\mathcal{O}(s))) = \mu_0(s) \) and transition function \( p'_\sigma((s_1, \Sigma_1, B_1) \mid (s_2, \Sigma_2, B_2)) = p(s_2 | s_1, \Sigma_2) \) if \( \sigma(B_1) = \Sigma_2 \) and \( B_2 = B_{\mathcal{M}(\mu_0)}(\mathcal{O}(s_2)) \) for \( \mu_1 \) a distribution of support \( B_1 \), \( p'_\sigma((s_1, \Sigma_1, B_1) \mid (s_2, \Sigma_2, B_2)) = 0 \) otherwise. The oMC \( \mathcal{M}'_\sigma \) is exponential in the size of \( \mathcal{M} \) and is equivalent to \( \mathcal{M}_\sigma \). When considering belief-based strategies, we will call \( \mathcal{M}_\sigma \) the finite equivalent oMC.

Writing \( \mathcal{V}_{\mathcal{M}_\sigma(\mu_0)} \) for the set of infinite paths corresponding to AA-disclosing observation sequences in \( \mathcal{M}_\sigma(\mu_0) \), we have \( \text{Disc}^{\text{AA}}(\mathcal{M}_\sigma(\mu_0)) = \mathcal{P}_{\mathcal{M}_\sigma(\mu_0)}(\mathcal{V}_{\mathcal{M}_\sigma(\mu_0)}) \). Remark that an observation sequence of the oMC induced by a WMC and a strategy contains both the observation of the state of the WMC and the choices of allowed events done by the strategy. The observation sequence of a path in the induced oMC is therefore equal to the knowledge sequence of the corresponding path in the WMC. This choice of observation was done to express that the choices made by the strategy are known to the observer. An important consequence of this decision is that the strategy does not modify which observation sequences are AA-disclosing.
Lemma 14. Given $\mathcal{M}$ a WMC, $\mu_0$ an initial distribution, $S' \subseteq S$, $\sigma, \sigma'$ two strategies and $w$ an observation sequence produced by at least one path of $\mathcal{M}_\sigma(\mu_0)$ and one path of $\mathcal{M}_{\sigma'}(\mu_0)$, then $p_{\text{rel}}^{\mathcal{M}_\sigma(\mu_0)}(b) = p_{\text{rel}}^{\mathcal{M}_{\sigma'}(\mu_0)}(b)$.

Proof. Let $\mathcal{M}$ be a WMC, $\mu_0$ an initial distribution, $\sigma$ be a strategy and $w = \sigma_0, \sigma_1, \ldots, \sigma_n$ be an observation sequence produced by at least one path of $\mathcal{M}_\sigma(\mu_0)$. By definition of $w$, $p_{\text{rel}}^{\mathcal{M}_\sigma(\mu_0)}(w)$ is defined and in particular $\prod_{i=0}^{n-1} \sigma(K(w_{i+1})) \neq 0$. We have

$$p_{\text{rel}}^{\mathcal{M}_\sigma(\mu_0)}(w) = \frac{P_{\mathcal{M}_\sigma(\mu_0)}(\{(\rho \in O^{-1}(w) \mid \rho \in \text{Rel})\})}{P_{\mathcal{M}_\sigma(\mu_0)}(w)}$$

$$= \frac{\sum_{\rho \in O^{-1}(w)} P_{\mathcal{M}_\sigma(\mu_0)}(\rho)}{\sum_{\rho \in O^{-1}(w)} P_{\mathcal{M}_\sigma(\mu_0)}(\rho)}$$

$$= \frac{\sum_{\rho = \sigma_0, \sigma_1, \ldots, \sigma_n \in O^{-1}(w)} \prod_{i=0}^{n-1} \sigma(K(w_{i+1})) \pi(s_{i+1} \mid s_i, \Sigma_i)}{\sum_{\rho = \sigma_0, \sigma_1, \ldots, \sigma_n \in O^{-1}(w)} \prod_{i=0}^{n-1} \sigma(K(w_{i+1})) \pi(s_{i+1} \mid s_i, \Sigma_i)}$$

which is independent of $\sigma$, therefore for any strategy $\sigma'$ such that at least one path of $\mathcal{M}_{\sigma'}(\mu_0)$ produces $w$, $p_{\text{rel}}^{\mathcal{M}_{\sigma'}(\mu_0)}(w) = p_{\text{rel}}^{\mathcal{M}_\sigma(\mu_0)}(w)$.

We study the two following diagnosability problems over WMC:

- The AA-diagnosability problem consists in, given a WMC $\mathcal{M}$ and an initial distribution $\mu_0$, deciding if there exists a strategy $\sigma$ such that $\mathcal{M}_\sigma(\mu_0)$ is AA-diagnosable.
The strategy problem consists in, given an AA-diagnosable WMC $M$ with initial distribution $\mu_0$, computing the strategy $\sigma$ achieving $\text{Disc}^{AA}(M_\sigma(\mu_0)) = P_{M_\sigma(\mu_0)}(\text{Rel})$.

**Example 15.** Consider the WMC on the left of Figure 2. Without any control (i.e., with a strategy permanently allowing every event), one obtains the oMC of Figure 1, which is not AA-diagnosable. However, assuming ‘b’ is a controllable event, the strategy that always forbid it induces the oMC on the right of Figure 2 which is AA-diagnosable: every relevant path almost surely contains a ‘c’ that can not be generated by a non-relevant path. This oMC is in fact diagnosable exactly as once a ‘c’ occurs the proportion of relevant paths is equal to 1.

### 3.2 Solving AA-diagnosability for WMCs

While approximate diagnosability is simpler than exact diagnosability for oMC (PTIME vs PSPACE)[8, 7], for WMCs, this difference disappears and both are EXPTIME-complete. The EXPTIME-completeness of exact diagnosis for WMC was established in [6]. We will devote this section to the proof of the following theorem:

**Theorem 16.** The AA-diagnosability is EXPTIME-complete.

First, the hardness is established by a reduction from safety games with imperfect information [10]. This result is obtained directly by applying the proof of Proposition 3 of [6].

**Proposition 17.** The AA-diagnosability is EXPTIME-hard.

**Proof.** In the proof of Proposition 3 of [6], a given safety game with imperfect information has a winning strategy iff no path is relevant. Moreover if a path is relevant, then its observation sequence is not AA-disclosing. More precisely, for $\rho$ a relevant path, the proportion of relevant path with observation sequence $K(\rho)$ is equal to $\frac{1}{2}$. Therefore the existence of a winning strategy in this game is equivalent to AA-diagnosability ensuring the EXPTIME-hardness.

The most important step to solve AA-diagnosability for WMC is to develop a good understanding on the strategies optimising AA-disclosure. For starters, with a straightforward adaptation of a proof of [14], we show that one can consider deterministic strategies only.

**Lemma 18.** Deterministic strategies are sufficient to decide AA-diagnosability.

We can further restrict the strategies by limiting ourselves to belief-based strategy. This is not intuitive as it means that although AA-diagnosability of an oMC depends on the exact values of the probabilities in the oMC, a control of the WMC only needs to keep the set of current possible states into account.

**Lemma 19.** Belief-based strategies are sufficient to decide AA-diagnosability.
Proof. Let \( M \) be a WMC, \( \mu_0 \) be an initial distribution and \( \sigma \) be a deterministic strategy such that \( M_\sigma(\mu_0) \) is AA-diagnosable. We define a belief based strategy \( \sigma' \) from \( \sigma \) in the following way. Let \( \rho \in \text{FPath}(M_\sigma(\mu_0)) \). We define by \( E_\rho \) the set of finite path producing the same belief as \( \rho \), i.e. \( E_\rho = \{ \rho' \in \text{FPath}(M_\sigma(\mu_0)) \mid B_{M(\mu_0)}(\text{O}(\rho')) = B_{M(\mu_0)}(\text{O}(\rho)) \} \). If \( E_\rho = \emptyset \), then \( \sigma'(B_{M(\mu_0)}(\text{O}(\rho))) \) is defined arbitrarily. Else, \( \sigma'(B_{M(\mu_0)}(\text{O}(\rho))) = \bigcup_{\rho' \in E_\rho} \sigma(\text{O}(\rho')) \). Let us show that \( M_\sigma(\mu_0) \) is AA-diagnosable.

Let two states \( q = (s, \Sigma^*, B) \in S^* \) and \( q' = (s', \Sigma^*, B) \in S \setminus S^* \) belonging to a BSCC of \( M_\sigma(\mu_0) \) and reached by two finite paths \( \rho \) and \( \rho' \) of \( \text{FPath}(M_\sigma(\mu_0)) \) with \( \text{O}(\rho) = \text{O}(\rho') \). We will show that \( d(M_\sigma(1_q), M_\sigma(1_{q'})) = 1 \) using the characterisation given in Proposition 9. More precisely, for any observations sequence \( w \in \Sigma^* \), and any pair of distributions on the set of states reached from \( q \) and from \( q' \) after observing \( w \) we consider the probabilistic language generated by similar distributions in \( M_\sigma \) (i.e. distributions giving the same weight to the states of the original WMC \( M \)) and rely on the fact that \( M_\sigma \) is AA-diagnosable to show that the generated languages are different. This implies the distance is 1 thanks to Proposition 9.

Let \( w \in \Sigma^* \) such that \( \mathbb{P}_{M_\sigma(1_q)}(w) > 0 \) and \( \mathbb{P}_{M_\sigma(1_{q'})}(w) > 0 \), we denote by \( B_w \), \( B_q \) and \( B_{q'} \) the beliefs reached after observing \( w \) from the beliefs \( B \), \( \{ q \} \) and \( \{ q' \} \) respectively, let two distributions \( \mu_1 \) and \( \mu_2 \) such that \( \text{Supp}(\mu_1') \subseteq B_w \), \( \text{Supp}(\mu_2') \subseteq B_{q'} \). As \( \sigma' \) does not allow events that are never allowed by \( \sigma \) in the same belief, there exists an observation sequence \( w_\sigma \in \Sigma^* \) such that \( \mathbb{P}_{M_\sigma(\mu_0)}(w_\sigma) > 0 \) and the belief reached in \( B_\sigma \) after a path of observation \( w_\sigma \) from the initial distribution is \( B_w \), i.e. \( B_{B_{M(\mu_0)}(w_\sigma)} = B_w \). We can thus define initial distributions \( \mu_1 \) and \( \mu_2 \) on the set of states reached after observing \( w_\sigma \) in \( M_\sigma \) mimicking the distributions \( \mu_1' \) and \( \mu_2' \) (i.e. giving the same probability to configurations associated to the same state of \( M \)). From the remark following Theorem 10 and Proposition 9, there exists a word \( w_d \) such that \( \mathbb{P}_{M_\sigma(\mu_1)}(w_d) \neq \mathbb{P}_{M_\sigma(\mu_2)}(w_d) \). This implies that there exists a word \( w_d' \) such that \( \mathbb{P}_{M_\sigma(\mu_1)}(w_d') \neq \mathbb{P}_{M_\sigma(\mu_2)}(w_d') \). Indeed, let \( E \) be the set of observation sequences of the form \( w'a \) where \( w' \) is a strict prefix of \( w_d \), \( a \in \Sigma \) \( \mathbb{P}_{M_\sigma(\mu_1)}(w'a) > 0 \) and \( \mathbb{P}_{M_\sigma(\mu_2)}(w'a) = 0 \). If \( \mathbb{P}_{M_\sigma(\mu_1)}(E) \neq \mathbb{P}_{M_\sigma(\mu_2)}(E) \), this implies our result. Else, by construction of the strategy \( \sigma' \) we have:

\[
\mathbb{P}_{M_\sigma(\mu_1)}(w_d) = \mathbb{P}_{M_\sigma(\mu_1)}(w_d) \times (1 - \mathbb{P}_{M_\sigma(\mu_1)}(E))
\neq \mathbb{P}_{M_\sigma(\mu_2)}(w_d) \times (1 - \mathbb{P}_{M_\sigma(\mu_2)}(E))
= \mathbb{P}_{M_\sigma(\mu_2)}(w_d) \times (1 - \mathbb{P}_{M_\sigma(\mu_2)}(E))
= \mathbb{P}_{M_\sigma(\mu_2)}(w_d),
\]

in which case we can choose \( w_d' = w_d \). As this holds for any \( w \in \Sigma^* \) and pair of distributions \( \mu_1 \) and \( \mu_2 \), according to Proposition 9 we have \( d(M_\sigma(1_q), M_\sigma(1_{q'})) = 1 \). From Theorem 10, we can thus deduce that \( M_\sigma(\mu_0) \) is AA-diagnosable. Therefore belief-based strategies are sufficient to decide AA-diagnosability.

A naive NEXPTIME algorithm can be obtained from these two lemmas: we guess a belief-based strategy then verify AA-diagnosability of the exponential
oMC generated by the WMC and the strategy. In the following proposition, we show how to build an adequate belief-based strategy directly, which gives you an EXPTIME algorithm.

**Proposition 20.** The AA-diagnosability problem is in EXPTIME.

**Proof.** Let $M$ be a WMC and $\mu_0$ be an initial distribution. To obtain the result, we first show that within a BSCC, the least restrictive a strategy is, the better it is for the purpose of diagnosis. However, a strategy too permissive may lead to the creation of new BSCC which may not be AA-diagnosable. Therefore, we will build an adequate strategy by an iterative procedure starting from the strategy allowing everything, then restricting it at each step to remove the problematic BSCC.

Let $\sigma$ and $\sigma'$ be two deterministic belief-based strategies such that for any belief $B$ of $M$, $\sigma(B) \subseteq \sigma'(B)$, let $q$ be a relevant state associated to the belief $B$ and belonging to a BSCC of both $M_\sigma(\mu_0)$ and $M_{\sigma'}(\mu_0)$. Then, if the path of $M_{\sigma'}(\mu_0)$ visiting $q$ are almost surely associated to AA-disclosing observation sequence, then so are the path of $M_\sigma(\mu_0)$ visiting $q$. This result follows from (1) Proposition 9, (2) due to the inclusion of the set of allowed events, the set of possible observation sequences in $M_{\sigma'}(\mu_0)$ is a subset of the set of possible observation sequences of $M_{\sigma'}(\mu_0)$ and (3) Lemma 14 which ensures that the probability that an observation is given by a relevant path does not depend on the strategy.

Using this result, we build iteratively the most permissive strategy ensuring AA-diagnosability. We start with the strategy $\sigma_0$ allowing everything. Assume we built the strategy $\sigma_k$ such that any less permissive strategy do not ensure AA-diagnosability. If $M_{\sigma_k}(\mu_0)$ is not AA-diagnosable, there exists two states $s$ and $s'$ satisfying the characterisation of Theorem 10. W.l.o.g one can assume that both of these states belong to BSCC of $M_{\sigma_k}(\mu_0)$. From our preliminary result, we know that any strategy that contains the states $s$ and $s'$ in a BSCC does not ensure AA-diagnosability. As any strategy less permissive than $\sigma_k$ does not ensure AA-diagnosability, the only way to deal with these states is to remove them from the oMC. Thus we build $\sigma_{k+1}$ as the most permissive strategy such that $M_{\sigma_{k+1}}(\mu_0)$ does not contain $s$ and $s'$. This can easily be done by belief based strategies as removing the state $s$ is equivalent to removing the belief $B$. This procedure ends when the strategy $\sigma_n$ that is produced either is the most permissive strategy ensuring AA-diagnosability or if one cannot build a strategy removing the problematic states/belief. This algorithm is in EXPTIME as every step of the procedure can be done in exponential time (verification of AA-diagnosability, identification of the pair of problematic states and creation of the new strategy are all steps that can be done in EXPTIME) and there is at most exponentially many steps as each one of them removes at least one belief from the system, and there are exponentially many beliefs. Therefore, the AA-diagnosability problem can be solved in EXPTIME.

The previous proof building the strategy ensuring AA-diagnosability when it exists, this algorithm also solves the strategy problem.
4 Active Approximate Opacity

As discussed at the beginning of Section 3, the framework of the study of active opacity is different from the one used for active diagnosis. While most elements are similar, strategies are given more power in the way they observe and affect the system. Moreover, the goal of the strategies is now either to maximise or to minimise the disclosure of information depending on whether they are deemed adversarial or cooperative.

4.1 Opacity for Observable Markov Decision Processes

Definition 21 (oMDP). An observable Markov Decision Process (oMDP) over alphabet $\Sigma$ is a tuple $M = (S, \text{Act}, p, O)$ where $S$ is a finite set of states, $\text{Act} = \cup_{s \in S}A(s)$ where $A(s)$ is a finite non-empty set of actions for each state $s \in S$, $p : S \times \text{Act} \rightarrow \text{Dist}(S)$ is the (partial) transition function defined for $(s, a)$ when $a \in A(s)$ and $O : S \rightarrow \Sigma$ is the observation function.

As before, we write $p(s' | s, a)$ instead of $p(s, a)(s')$. Given an initial distribution $\mu_0$, an infinite path of $M(\mu_0)$ is a sequence $\rho = s_0a_0s_1a_1 \ldots$ where $\mu_0(s_0) > 0$ and $p(s_{i+1} | s_i, a_i) > 0$, for $s_i \in S$, $a_i \in A(s_i)$, for all $i \geq 0$. Finite paths are defined like for WMC, and we use similar notations for the various sets of paths. Given a path $\rho = s_0a_0s_1a_1 \ldots s_i$ its observation is $O(\rho) = O(s_0)O(s_1) \ldots O(s_i)$.

The nondeterministic choice of the action is resolved by strategies.

Definition 22 (Strategy for oMDP). A strategy for an oMDP $M$ with initial distribution $\mu_0$ is a mapping $\sigma : \text{FPath}(M(\mu_0)) \rightarrow \text{Dist}(\text{Act})$ associating with any finite path $\rho$ a distribution $\sigma(\rho)$ on the actions in $A(\text{last}(\rho))$.

Remark that we do not use the notion of knowledge of a strategy as was done for WMC, this is due to the enhanced strategies used in oMDP which knowledge is in act the path itself. Similarly as for WMC, given a strategy $\sigma$, a path $\rho = s_0a_0s_1a_1 \ldots$ of $M$ is $\sigma$-compatible if for all $i$, $a_i \in \text{Supp}(\sigma(s_0a_0s_1a_1 \ldots s_i))$.

A strategy $\sigma$ is observation-based if for any finite path $\rho$, $\sigma(\rho)$ only depends on the observation $O(\rho)$ and on the last state $\text{last}(\rho)$. We can also adapt the notions of deterministic and belief-based strategies.

A strategy $\sigma$ on $M(\mu_0)$ defines a (possibly infinite) oMC $M_{\sigma}(\mu_0)$ with set of states $\text{FPath}(M_{\sigma}(\mu_0))$ (the finite $\sigma$-compatible paths), that can be equipped with the observation function associating $O(\text{last}(\rho))$ with the finite path $\rho$. The transition function $p_{\sigma}$ is defined for $\rho \in \text{FPath}(M_{\sigma}(\mu_0))$ and $\rho' = \rho a s'$ by $p_{\sigma}(\rho' | \rho) = \sigma(\rho)(a)p(s' | s, a)$ and we denote by $P_{M_{\sigma}(\mu_0)}$ the associated probability measure. The definition of the observation function shows that the observer of the system does not know what action is chosen by the strategy at any step. However, the observer still knows which strategy was selected initially, allowing him to deduce the oMC $M_{\sigma}$.

Disclosure values for oMDP are defined according to the status of the strategies, by considering them as adversarial or cooperative with respect to the system.
Definition 23 (Disclosure of an oMDP). Given an oMDP \( M = (S, \text{Act}, p, O) \), an initial distribution \( \mu_0 \) and a set of relevant states \( S' \subseteq S \), the maximal AA-disclosure of \( S' \) in \( M(\mu_0) \) is \( \text{Disc}^{\text{AA}}_{\text{max}}(M(\mu_0)) = \sup_{\sigma} \text{Disc}^{\text{AA}}(M_{\sigma}(\mu_0)) \) and the minimal AA-disclosure is \( \text{Disc}^{\text{AA}}_{\text{min}}(M(\mu_0)) = \inf_{\sigma} \text{Disc}^{\text{AA}}(M_{\sigma}(\mu_0)) \).

We study the following opacity problems over oMDP:

- **Quantitative decision problems:** The minimal AA-disclosure problem consists in, given an oMDP \( M \) and a threshold \( \theta \in [0, 1] \), deciding if \( \text{Disc}^{\text{AA}}_{\text{min}}(M) \leq \theta \)? The maximal AA-disclosure problem consists in, given an oMDP \( M \) and a threshold \( \theta \in [0, 1] \), deciding if \( \text{Disc}^{\text{AA}}_{\text{max}}(M) \geq \theta \)?

- **Qualitative decision problems:** The limit-sure disclosure problem is the AA-disclosure problem with \( \theta = 1 \) for maximisation and with \( \theta = 0 \) for minimisation and the almost-sure disclosure problem consists in deciding whether there exists a strategy achieving a disclosure of 1 for maximisation and 0 for minimisation.

4.2 Possible Restriction on the Strategies

The decidability result for AA-diagnosability of WMC relied strongly on a restriction to a set of dominant strategies. It is thus natural to take a similar approach for opacity. We can indeed establish some such restriction, for instance to observation-based strategies. This is proven for an exact notion of disclosure in [2], however the very same proof applies to the accurate approximate notion.

Proposition 24 ([2]). Given an oMDP, an initial distribution \( \mu_0 \), \( S' \subseteq S \) and a strategy \( \sigma \), there exists an observation-based strategy \( \sigma' \) such that \( \text{Disc}^{\text{AA}}(M_{\sigma}(\mu_0)) = \text{Disc}^{\text{AA}}(M_{\sigma'}(\mu_0)) \).

However, the restriction cannot be extended to deterministic belief-based strategies. Consider the example on the left of Figure 3 with maximisation of the AA-disclosure in mind. There are three components, in each of them a 'b' is always followed by an 'a', however, the probability that a 'b' occurs after an 'a' varies. This probability is \( \frac{3}{4} \) in the upper component, \( \frac{1}{4} \) in the middle one and depends on the strategy on the one below. A deterministic belief based strategy will either always choose the action 'c' or always the action 'd'. Such a strategy replicates the probabilistic behaviour of one of the other two components, inducing an AA-disclosure of 0. However a randomised strategy, giving for instance a half probability to both actions obtains a \( \frac{1}{2} \) probability to produce a 'b' after an 'a'. This belief-based randomised strategy induces then an AA-disclosure of \( \frac{1}{3} \). One could define a deterministic strategy which is not belief based and obtain a disclosure of \( \frac{1}{3} \) too by alternating the choices of the action 'c' and 'd'. Therefore, maximising strategies require randomisation, more memory than just the belief or both\(^2\).

\(^2\) Using a complicated example, one can show that randomisation cannot always substitute the need for additional memory. This also holds for minimising strategies.
When aiming to minimise the AA-disclosure, we can show that randomisation is necessary. Consider the oMDP depicted on the right of Figure 3. The strategy only has to make a choice between two actions during the first step. Thus, there are only two existing deterministic strategies, choosing respectively 'c' or 'd' in q₀. In both cases, the disclosure is \( \frac{1}{2} \). On the other hand, any randomized strategies \( \sigma_p \) such that \( \sigma_p(q₀) \) activates 'c' with probability \( p \) and 'd' with probability \( (1 - p) \) with \( 0 < p < 1 \), induces an oMC that do not contain any AA-disclosing observation, hence the disclosure is 0.

4.3 (Un)decidability of the Opacity Problems

The examples of the previous subsection point to the idea that the traditional framework for active opacity is more complicated than the one considered for active diagnosability. This is confirmed by the (un)decidability results that we establish below. The undecidability proofs we establish are done by reduction of problems in probabilistic automata (PA). Recall that a PA is a tuple \( \mathfrak{A} = (Q, q₀, \Sigma, T, F) \) where \( Q \) is a finite set of states with \( q₀ \in Q \) the initial state, \( \Sigma \) is a finite alphabet (which cumulates the role of the actions in the oMDP and of the observation), \( T : Q \times \Sigma \to \text{Dist}(Q) \) is the transition function and \( F \subseteq Q \) is the set of final states. We define paths for PA as usual and for a finite path \( \rho = q₀a₁q₁ \ldots aₙqₙ \) of \( \mathfrak{A} \), the word \( a₁ \ldots aₙ \in \Sigma^* \) is called the trace of \( \rho \) and denoted by \( \text{tr}(\rho) \). Writing \( \text{FPath}_{(w,q)}(\mathfrak{A}) = \{ \rho \in \text{FPath}(\mathfrak{A}) \mid \text{tr}(\rho) = w \text{ and last}(\rho) = q \} \) for \( w \in \Sigma^* \) and \( q \in Q \), we define \( P_{\mathfrak{A}}(w, q) = P_{\mathfrak{A}}(\bigcup_{\rho \in \text{FPath}_{(w,q)}(\mathfrak{A})} \text{Cyl}(\rho)) \), \( P_{F_{\mathfrak{A}}}(w, F) = \sum_{q \in F} P_{\mathfrak{A}}(w, q) \) and \( \text{Val}(\mathfrak{A}) = \sup_{w \in \Sigma^*} P_{\mathfrak{A}}(w, F) \).

Given a threshold \( \theta \in (0, 1) \), we set \( L_{\leq \theta}(\mathfrak{A}) = \{ w \in \Sigma^* \mid P_{\mathfrak{A}}(w, F) > \theta \} \). The strict emptiness problem for \( \mathfrak{A} \) consists in asking whether \( L_{> \theta} \) is empty, and is
known to be undecidable for $\theta > 0$ [19]. The value 1 problem, i.e. asking whether $val(A) = 1$, is undecidable as well [16].

Let us first consider the maximisation of $AA$-disclosure.

**Theorem 25.** The maximal $AA$-disclosure problem is undecidable. The maximal limit-sure disclosure problem is undecidable.

As a silver lining, the almost-sure $AA$-disclosure problem is easily decidable.

**Theorem 26.** The maximal almost-sure $AA$-disclosure problem is in PTIME.

These results are not exactly surprising as opacity problems for maximisation had already been shown to be undecidable for exact notions of opacity in [2]. However, while in this same paper the authors show that most opacity problems for minimisation are decidable, these problems become undecidable for the accurate approximate notion of opacity.

**Theorem 27.** The minimal almost-sure and the minimal limit-sure $AA$-disclosure decision problems are undecidable.

**Corollary 28.** The minimal $AA$-disclosure decision problem is undecidable.

## 5 Conclusion

This paper realises an analysis of approximate information control notions in stochastic systems. We first introduced an approximate notion of opacity that mirrors the existing approximate diagnosability notion. Then we almost fully characterised the decidability and complexity of the diagnosability and opacity notions in a passive and active frameworks. While diagnosability and opacity are closely related, this study highlights how much more complicated opacity (and its associated framework) is.

There remains a complexity gap ($\text{PSPACE}$ versus $\text{EXPTIME}$) for accurate approximate opacity of observable Markov chains that we want to fill. Another interesting research direction is to look for a meaningful decidable notion of approximate opacity in stochastic controllable systems. Finally, for an active system to be accurate approximately diagnosable, we currently request the existence of a control that is able to disclose the relevance of the paths as precisely as one could want. We could modify the problem and instead consider the system diagnosable if, for any given precision, one can find a control detecting relevant paths with this precision. This new definition is obtained by switching two quantifiers. This is risky and can lead to undecidability as was shown in the study of approximate diagnosability for passive systems [8].

## References

3. B. Bérand, O. Kouchmarenko, J. Mullins, and M. Sassolas. Preserving opacity on
4. B. Bérand, O. Kouchmarenko, J. Mullins, and M. Sassolas. Opacity for linear
probabilistic systems. In Proceedings of FoSSaCS’14, volume 8412 of LNCS, pages
7. N. Bertrand, S. Haddad, and E. Lefaucheux. Foundation of diagnosis and predict-
dability in probabilistic systems. In Proceedings of FSTTCS’14, volume 29 of
8. N. Bertrand, S. Haddad, and E. Lefaucheux. Accurate approximate diagnosability
of stochastic systems. In Proceedings of LATA’16, volume 9618 of LNCS, pages
9. Nathalie Bertrand, Serge Haddad, and Engel Lefaucheux. A Tale of Two Diagnoses
of FoSSaCS’08, volume 2 of LIPIcs, pages 73–82. Leibniz-Zentrum für Informatik,
2008.
11. J. W. Bryans, M. Koutny, L. Mazaré, and P. Y. A. Ryan. Opacity generalised to
12. F. Cassez and S. Tripakis. Fault diagnosis with static and dynamic observers.
13. E. Chanthery and Y. Pencolé. Monitoring and active diagnosis for discrete-event
16. H. Gimbert and Y. Ouahadh. Probabilistic automata on finite words: Decidable
and undecidable problems. In Proceedings of ICALP’10, volume 6199 of LNCS,
17. S. Jiang, Z. Huang, V. Chandra, and R. Kumar. A polynomial algorithm for
testing diagnosability of discrete-event systems. Transactions on Automatic Control,
20. A. Saboori and Ch. N. Hadjicostis. Current-state opacity formulations in proba-
Diagnosability of discrete-event systems. Transactions on Automatic Control,

A Equivalence of the AA-diagnosability definitions

Proposition 8. An oMC is AA-diagnosable iff it is ε-diagnosable for all ε > 0.

Proof. Let $\mathcal{M}$ be a finite oMC and $\mu_0$ an initial distribution. Suppose that $\mathcal{M}(\mu_0)$ is AA-diagnosable. By definition, given an AA-disclosing observation sequence $w$, for all $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that $w_{\uparrow n}$ is $\varepsilon$-disclosing. Therefore for all $\varepsilon > 0$, $\text{Disc}^{\varepsilon}(\mathcal{M}(\mu_0)) \leq \text{Disc}^{\varepsilon}(\mathcal{M}(\mu_0))$. Moreover, as $\mathcal{M}$ is AA-diagnosable, $\text{Disc}^{\varepsilon}(\mathcal{M}(\mu_0)) = \mathbf{P}(\text{Rel})$. Thus, $\text{Disc}^{\varepsilon}(\mathcal{M}(\mu_0)) \leq \mathbf{P}(\text{Rel})$. Finally, by definition of $\text{Disc}^{\varepsilon}$, for all $\varepsilon > 0$ $\text{Disc}^{\varepsilon}(\mathcal{M}(\mu_0)) \leq \mathbf{P}(\text{Rel})$. Thus $\text{Disc}^{\varepsilon}(\mathcal{M}(\mu_0)) = \mathbf{P}(\text{Rel})$ and $\mathcal{M}(\mu_0)$ is $\varepsilon$-diagnosable.

Conversely, suppose that $\mathcal{M}(\mu_0)$ is not AA-diagnosable. Let us consider the set of infinite words $D = \cap_{\varepsilon > 0} \text{Disc}^{\varepsilon} \subseteq \text{D}_{\text{max}} \cap \text{D}^{\text{AA}}$. Let us show that $\mathbf{P}(D) = 0$. Let $w \in D$, we have (1) for all $\varepsilon > 0$ there exists $n \in \mathbb{N}$ such that $\mathbf{P}_{\mathcal{M}(\mu_0)}{(w_{\uparrow n})} > 1 - \varepsilon$ and (2) $(\mathbf{P}_{\mathcal{M}(\mu_0)}{(w_{\uparrow n})})_{n \in \mathbb{N}}$ does not converge toward 1. Given $\varepsilon > 0$ and denoting by $E_\varepsilon$ the set of $\varepsilon$-min-disclosing observation sequence, due to (1) we have

$$\mathbf{P}(\{\rho \in \text{O}^{-1}(D) \setminus \text{Rel}\}) < \sum_{w \in E_\varepsilon} \mathbf{P}(\{\rho \in \text{O}^{-1}(w) \setminus \text{Rel}\}) < \sum_{w \in E_\varepsilon} \mathbf{P}(\{\rho \in \text{O}^{-1}(w) \cap \text{Rel}\}) \frac{\varepsilon}{1 - \varepsilon} \leq \frac{\varepsilon}{1 - \varepsilon}. $$

As this holds for all $\varepsilon > 0$, $\mathbf{P}(\{\rho \in \text{O}^{-1}(D) \setminus \text{Rel}\}) = 0$. Moreover, due to (2), there exists $\varepsilon > 0$ such that for infinitely many $n \in \mathbb{N}$ we have $\mathbf{P}_{\mathcal{M}(\mu_0)}{(w_{\uparrow n})} < 1 - \varepsilon$. For all $k \in \mathbb{N}$, we denote by $E_k$ the set of prefixes $w$ of words of $D$ such that $\mathbf{P}_{\mathcal{M}(\mu_0)}{(w_{\uparrow k})} < 1 - \varepsilon$ for the $k$'th time. We then have for all $k$:

$$\mathbf{P}(\{\rho \in \text{O}^{-1}(E_k) \setminus \text{Rel}\}) = \sum_{w \in E_k} \mathbf{P}(\{\rho \in \text{O}^{-1}(w) \setminus \text{Rel}\}) > \sum_{w \in E_k} \mathbf{P}(\{\rho \in \text{O}^{-1}(w) \cap \text{Rel}\}) \frac{\varepsilon}{1 - \varepsilon} \geq \frac{\varepsilon}{1 - \varepsilon} \mathbf{P}(\{\rho \in \text{O}^{-1}(D) \cap \text{Rel}\}).$$

As $(\mathbf{P}(\{\rho \in \text{O}^{-1}(E_k) \setminus \text{Rel}\}))_{k \in \mathbb{N}}$ converges toward $\mathbf{P}(\{\rho \in \text{O}^{-1}(D) \setminus \text{Rel}\})$ which is equal to 0, this implies that $\mathbf{P}(\{\rho \in \text{O}^{-1}(D) \cap \text{Rel}\}) = 0$ and thus that $\mathbf{P}(D) = 0$. As a consequence, $\lim_{\varepsilon \to 0} \mathbf{P}(D_{\text{min}}^{\varepsilon}) = \mathbf{P}(D^{\text{AA}})$. Therefore there exists $\varepsilon > 0$ such that $\mathcal{M}(\mu_0)$ is not $\varepsilon$-diagnosable.
B AA-disclosure problem for oMC

Theorem 11. The AA-disclosure problem for finite oMC is decidable in EXP-TIME. It is PSPACE-hard.

Proof. We establish the hardness by reducing the universality problem for non-deterministic finite automaton (NFA), which is known to be PSPACE-complete [18].

Let \( \mathcal{A} = (Q, \Sigma, T, q_0, F) \) be an NFA (\( Q \) is the set of states, \( q_0 \) the initial one, \( F \) the set of accepting states, \( \Sigma \) the alphabet and \( T \in Q \times \Sigma \times Q \) the transition function). W.l.o.g. we can assume that \( F = Q \) and \( \Sigma = \{a, b\} \). Our first step is to push the observations onto the states (as shown in Figure 4). From \( \mathcal{A} \) we define the incomplete oMC \( \hat{\mathcal{A}} = (S_A, p_A, O_A) \) and the initial distribution \( \mu_0^A \) such that:

- \( S_A = Q \times \Sigma \);
- for \( (q, c), (q', d) \in S_A \), if \( (q, d, q') \in T \), then \( p_A((q', d) \mid (q, c)) = \frac{1}{|S_A|+1} \), else \( p_A((q', d) \mid (q, c)) = 0 \);
- for \( (q, c) \in S_A \), \( O_A(q, c) = c \);
- for \( (q', d) \in S_A \), if \( (q_0, d, q') \in T \), then \( \mu_0^A(q', d) = \frac{1}{|S_A|+1} \), else \( \mu_0^A(q', d) = 0 \).

This oMC is incomplete as none of the distributions \( \mu_0^A \) and \( p_A(\cdot \mid s) \) (for \( s \in S_A \)) sum to 1.

![Fig. 4](image)

We now build the oMC \( \mathcal{M} = (S, p, O) \) represented in Figure 5 where

- \( S = S_A \cup \{s_1, f_a, f_b, f_2\} \);
- given \( s, s' \in S_A \), \( p(s' \mid s) = p_A(s' \mid s) \), \( p(s_2 \mid s) = 1 - \sum_{s' \in S_A} p(s' \mid s) \), for \( h \in \{f_a, f_b\} \) and \( g \in \{f_a, f_b, f_2\} \), \( p(g \mid h) = 1/3 \) and \( p(f_2 \mid f_2) = p(s_2 \mid s_2) = 1 \);
- for \( s \in S_A \), \( O(s) = O_A(s), O(s_2) = O(f_2) = \frac{1}{2} \), \( O(f_a) = a \) and \( O(f_b) = b \).

We also define \( \mu_0 \) as \( \mu_0^A(s) \) for \( s \in S_A \) and \( \mu_0(f_a) = \mu_0(f_b) = \frac{1 - \sum_{s \in S_A} \mu_0(s)}{2} \).

Choosing \( S' = \{f_a, f_b, f_2\} \), let us show that \( \mathcal{A} \) is not universal iff \( Disc^{AA}(M(\mu_0)) > 0 \).
Suppose first that \( \mathcal{A} \) is not universal. There thus exists a word \( w \in \Sigma^* \) such that no path starting in \( S_A \) has observation sequence \( w \). As there exists one relevant path \( \rho \) (starting in either \( f_a \) or \( f_b \)) associated to \( w \), we have \( \Pr_M^{\mathbf{\text{rel}}} (\mu_0) (w) = 1 \). Therefore \( \text{Disc}^{\mathbf{\text{AA}}} (M(\mu_0)) \geq \Pr_M (\mu_0) (\rho) > 0 \).

Conversely, assume that \( \mathcal{A} \) is universal. Let \( \rho \) be a path ending in \( f_2 \) with observation sequence \( \mathbf{O} (\rho) = w \) for some \( w \in \Sigma^* \). As \( \mathcal{A} \) is universal, there exists a finite path \( \rho' \) ending in \( s^\# \) with observation \( w \). Thus, \( \Pr_M^{\mathbf{\text{rel}}} (\mu_0) (w) = 1 \). Moreover, every path ending with a \( \# \) remains with probability 1 in either \( s^\# \) or \( f^\# \), due to this for every \( k \geq 2 \), \( \Pr_M^{\mathbf{\text{rel}}} (\mu_0) (w^k) = \Pr_M^{\mathbf{\text{rel}}} (\mu_0) (w) \). Therefore, \( w \omega \not\in D^{\mathbf{\text{AA}}} \). This implies that no infinite path visiting \( f_2 \) corresponds to an \( \mathbf{\text{AA}} \)-disclosing observation sequence. \( f_2 \) being the only relevant state, \( \text{Disc}^{\mathbf{\text{AA}}} (M(\mu_0)) = 0 \).

C Deterministic strategies for \( \mathbf{\text{AA}} \)-diagnosability

Lemma 18. Deterministic strategies are sufficient to decide \( \mathbf{\text{AA}} \)-diagnosability.

Proof. In the proof of Lemma 1 of [14], the authors show that a randomised ‘observation based’ strategy can be seen as an average over a family of deterministic ‘observation based’ strategies. A consequence of their equation (2) in our framework is the following: given a strategy \( \sigma \), for every set of path \( E \), there exists a deterministic strategy \( \sigma^{\mathbf{\text{det}}} \) such that (a) \( \text{Path} (M_{\sigma^{\mathbf{\text{det}}}} (\mu_0)) \subseteq \text{Path} (M_{\sigma} (\mu_0)) \) and (b) \( \Pr_M^{\mathbf{\text{rel}}} (\mu_0) (E) \geq \Pr_M^{\mathbf{\text{rel}}} (\mu_0) (E) \). Using this result with the appropriate set \( E \) we will show that if \( M_{\sigma} (\mu_0) \) is \( \mathbf{\text{AA}} \)-diagnosable then \( M_{\sigma^{\mathbf{\text{det}}}} (\mu_0) \) is \( \mathbf{\text{AA}} \)-diagnosable.

We define \( E_{\sigma} = V_{\text{rel}}^{\mathbf{\text{AA}}} (\mu_0) \cup (\text{Path} (M_{\sigma} (\mu_0)) \setminus \text{Rel}) \) which are the set of \( \sigma \)-compatible paths that are either not relevant or \( \mathbf{\text{AA}} \)-disclosing. Let \( \sigma^{\mathbf{\text{det}}} \) be

---

3 In our framework, by definition, every strategy is ‘observation based’.
the strategy obtained by applying the result of [14] on the set \( E_\sigma \). Suppose \( M_\sigma(\mu_0) \) is AA-diagnosable. By definition, this is equivalent to \( P_{M_\sigma(\mu_0)}(E_\sigma) = 1 \). Due to (b), this implies that \( P_{M_{\sigma,\det}(\mu_0)}(E_\sigma) = 1 \) too. Moreover \( V_{M_{\sigma,\det}(\mu_0)} = V_{M_\sigma(\mu_0)} \cap \text{Path}(M_{\sigma,\det}(\mu_0)) \), thanks to Lemma 14 and (a). Thus

\[
E_\sigma = V_{M_{\sigma,\det}(\mu_0)} \cup (V_{M_\sigma(\mu_0)} \setminus \text{Path}(M_{\sigma,\det}(\mu_0))) \cup (\text{Path}(M_\sigma(\mu_0)) \setminus \text{Rel}) \\
= E_{\sigma,\det} \cup (V_{M_\sigma(\mu_0)} \cup (\text{Path}(M_\sigma(\mu_0)) \setminus \text{Rel}) \setminus \text{Path}(M_{\sigma,\det}(\mu_0))
\]

where \( E_{\sigma,\det} = V_{M_{\sigma,\det}(\mu_0)} \cup (\text{Path}(M_{\sigma,\det}(\mu_0)) \setminus \text{Rel}) \).

Finally, \( P_{M_{\sigma,\det}(\mu_0)}(V_{M_\sigma(\mu_0)} \cup (\text{Path}(M_\sigma(\mu_0)) \setminus \text{Rel}) \setminus \text{Path}(M_{\sigma,\det}(\mu_0)) = 0 \) as no path of this set is \( \sigma\det\)-compatible. Therefore \( P_{M_{\sigma,\det}(\mu_0)}(E_{\sigma,\det}) = 1 \) which implies that \( M_{\sigma,\det}(\mu_0) \) is AA-diagnosable.

\[\text{D Maximisation of the AA-disclosure}\]

**Theorem 25.** The maximal AA-disclosure problem is undecidable. The maximal limit-sure disclosure problem is undecidable.

**Proof.** These results are obtained by reductions from the strict emptiness problem and value 1 problem on probabilistic automata. As a consequence, we first need a method to adapt a given probabilistic automaton to our framework. This transformation bears many similarities with what was done for NFA in the beginning of Theorem 11. For simplicity, we use states without observations (denoted by the observation \( \varepsilon \)), this is without loss of generality as we could remove them using a simple probabilistic closure since no non-deterministic choice occurs within them.

Given a probabilistic automaton \( \mathfrak{A} = (Q, q_0, \{a, b\}, T, F) \) over \( \{a, b\} \) that we suppose complete (i.e. \( T(q, c) \) is defined for all \( q \in Q \) and \( c \in \{a, b\} \)) without loss of generality, we first transform \( \mathfrak{A} \) into an incomplete oMDP \( \mathfrak{M} = (\hat{Q}, \{\varepsilon\}, \hat{p}, \hat{O}) \) over the observation alphabet \( \{a, b\} \) where the observations are pushed from the transitions to the next state (an illustration is given in Fig. 6). The set of states is \( \hat{Q} = Q \cup \{q_c \mid q \in Q \land c \in \{a, b\}\} \). The observation function \( \hat{O} \) is defined by \( \hat{O}(q) = \varepsilon \) and \( \hat{O}(q_c) = c \) for \( q \in Q \) and \( c \in \{a, b\} \). The transition function \( \hat{p} \) is defined for \( q, q' \in \hat{Q} \) and \( c \in \{a, b\} \) by \( \hat{p}(q' \mid q, c) = T(q' \mid q, c) \) and \( \hat{p}(q_0 \mid q, c) = \frac{1}{2} \). This oMDP is incomplete as the probabilities do not sum to 1. Intuitively, a letter to read is chosen at random, and then the transition is taken according to the probabilities induced by the chosen letter. Remark that the strategy do not make any choice here.

From \( \mathfrak{M} \) we build the oMDP \( \mathfrak{M} = (S, \{e, r\}, p, O) \) (represented in Figure 7) where:

\[\begin{align*}
\text{− } S &= \hat{Q} \cup \{s_0, s_1^a, s_1^b, s_0^a, s_0^b\} \cup \{s_2^z \mid z \in \{m, l\}, t \in \{e, a, b, z, \varepsilon\}\}; \\
\text{− } p(q_0 \mid s_0, c) &= p(q_0^a, s_0^a, e) = p(q_0^b, s_0^b, e) = 1/2, \text{ for } q, q' \in \hat{Q}, p(q' \mid q, e) = \hat{p}(q' \mid q, e) \text{, for } q \in F, q', q' \in \hat{Q} \setminus F, p(s_0^a \mid q, e) = p(s_1^a \mid q', e) = 1/2, p(q_0 \mid s_0^a, e) = 1, p(q_0 \mid s_0^b, e) = p(s_1^b \mid s_0^a, e) = 1/2, \text{ for } z \in \{m, l\}, c, d' \in \{a, b\}, p(s_2^z \mid s_0^a, e) = \end{align*}\]
As once reaching a state labelled by which it plays \( e \), the set of relevant states is defined as happens iff for every path \( \rho \) such that any other observation from then on is not useful: after reading a word however, the middle and lower parts are used to make this additional knowledge system is, it could ‘cheat’ and reach the secret almost surely in the upper part. Let us first give the intuition behind this construction. The MDP \( M \) is composed of three parts (upper, middle and lower part of the Figure 7). The upper part mostly imitates the behaviour of the PA \( \mathcal{A} \) on random words, a \( \sharp \) signalling the end of the word. If the run is accepting, i.e. if it ended in a final state of the PA, then the strategy may chose to play \( e \) in order to reach the secret state, otherwise a new word is read. As the strategy knows in which state the system is, it could ‘cheat’ and reach the secret almost surely in the upper part. However, the middle and lower parts are used to make this additional knowledge of the strategy useless: after reading a word \( w \) in the middle part, the strategy chooses between the action \( e \) and \( r \), using the action \( r \) implies that \( w\rho^w \) is not AA-disclosing while using the action \( e \) makes \( w\rho^w \) AA-disclosing but the run also reaches the lower part ensuring that any other observation from then on is not AA-disclosing. In other words, the strategy will have to choose a set of words for which it plays \( e \) simultaneously in the middle and the upper parts.

Formally, let us first identify which relevant paths are disclosing with a strategy \( \sigma \). Let \( \rho \) be a relevant path with observation \( w\rho^k \) for some \( k \in \mathbb{N} \) and \( w \in \{a,b,\sharp\}^* \). As once reaching a state labelled by \( b \), there is no probabilistic behaviour, \( P_{\text{rel}^{\mathcal{A}}(M_{\sigma}(\mu_0))}(w) = P_{\text{rel}^{\mathcal{A}}(M_{\mu_0})}(w) = 1 \). Thus, the infinite observation associated to the unique infinite path extending \( \rho \) is AA-disclosing iff \( P_{\text{rel}^{\mathcal{A}}(M_{\mu_0})}(w) = 1 \). This happens iff for every path \( \rho' \) such that \( \mathcal{O}(\rho') = w \) and \( \text{last}(\rho') = s_{\rho^w} \), \( \sigma(\rho') = e \).
Fig. 7. Reduction from the emptiness problem to the maximal disclosure problem. The square state corresponds to a final state of $\mathfrak{A}$. 
We define the strategy $\sigma$ which is reached with positive probability from any state $s^m_i$. This last condition can also be formulated as the absence of any run which observation is a prefix of $w$, ending in $s^m_i$ and for which $\sigma$ selects the action $e$.

Let $\lambda \in \mathbb{R}$ assume that there exists a word $w \in \{a, b\}^*$ such that $P_A(w, F) > \lambda$. We define the strategy $\sigma$ such that given a path with observation $w_1\sharp w_2\sharp \hdots w_k\sharp$ such that for all $i \leq k$, $w_i \in \{a, b\}^*$, if $w_k = w$ and both $e$ and $r$ are allowed actions in $\text{last}(\rho)$, then $\sigma$ chooses $e$, otherwise it chooses $r$ if possible. This strategy induces a disclosure greater than $\lambda/2$. Indeed, in the upper part of the oMDP, in between two occurrences of $\sharp$ there is a positive probability that $w$ is observed. Thus, with probability 1 a word of the form $w_1\sharp w_2\sharp \hdots w_k\sharp$ such that for all $i \leq k$, $w_i \in \{a, b\}^*$, for all $i < k$, $w_i \neq w$ and $w_k = w$ will be triggered. Moreover, let $v$ be one such word, then, thanks to the choice of the strategy and the remark of the previous paragraph, $P_{\text{rel}M_\sigma(\mu_0)}(v\sharp) = 1$. Finally, the probability that a path of the upper part of the oMDP, with observation $v$ ends in $s^m_i$ (allowing to trigger $b$ on the next step) is $P_A(w, F)$, thus ensuring that $\text{Disc}^{AA}(M_\sigma(\mu_0)) > \lambda/2$.

Conversely, assume that there exists a strategy $\sigma$ such that $\text{Disc}^{AA}(M_\sigma(\mu_0)) > \lambda/2$. We define the set of words $E = \{(w, w') \in \{a, b\}^* \times \{a, b, \sharp\}^* | \exists w \in \{a, b, \sharp\}^*, w' = u\sharp w \wedge \text{last}(\rho) \neq \emptyset \wedge P_{\text{rel}M_\sigma(\mu_0)}(w') = 1\}$. Relying on the earlier remark on which paths are disclosing, we have

$$\text{Disc}^{AA}(M_\sigma(\mu_0)) = \sum_{(w, w') \in E} P_{M_\sigma(\mu_0)}(\{\rho \in \text{FPath}(M_\sigma(\mu_0)) \mid O(\rho) = w'\})$$

$$\leq \sum_{(w, w') \in E} P_{M_\sigma(\mu_0)}(\{\rho \in \text{FPath}(M_\sigma(\mu_0)) \mid O(\rho) = w' \wedge \text{last}(\rho) = s^m_i\})$$

$$= \sum_{(w, w') \in E} P_{M_\sigma(\mu_0)}(w') \cdot 1/2 P_A(w, F)$$

$$\leq 1/2 \max_{(w, w') \in E} P_A(w, F)$$

Therefore, $\text{Disc}^{AA}(M_\sigma(\mu_0)) > \lambda/2$ implies that there exists $w$ such that $P_A(w, F) > \lambda$.

This equivalence directly shows that the maximal AA-disclosure problem is undecidable. For the maximal limit-sure disclosure, one can use the same reduction with one additional secret state with observation $\sharp$ (thus disclosing) which is reached with positive probability from any state $s^m_i$ with $c \in \{a, b\}$. This means that the longer we wait before selecting a word, the higher the probability that a path that went to the middle part is AA-disclosing. However, the remaining paths are enough to guarantee the same reasoning as before for the paths going to the upper part, thus showing undecidability of maximal limit-sure disclosure.

**Theorem 26.** The maximal almost-sure AA-disclosure problem is in PTIME.

**Proof.** Given $M$ an oMC, let us show that $\text{Disc}^{AA}(M(\mu_0)) = 1$ iff $P_{M(\mu_0)}(\text{Rel}) = 1$.

First, $\text{Disc}^{AA}(M(\mu_0)) \leq P_{M(\mu_0)}(\text{Rel})$ by definition, thus if $\text{Disc}^{AA}(M(\mu_0)) = 1$ then $P_{M(\mu_0)}(\text{Rel}) = 1$. 

Conversely, suppose that $\text{Disc}_{\epsilon}^{\text{AA}}(M(\mu_0)) < 1$, there thus exists a set of infinite observations $E_\alpha$ that are not AA-disclosing and such that $P(\mathcal{M}(\mu_0)(O^{-1}(E_\alpha) \cap \text{Rel}) > 0$. By definition of AA-disclosing, there thus exists $\epsilon > 0$ such that there exists a subset of $E_\alpha$, denoted $E_\epsilon$, of infinite observations for which none of their prefixes are $\epsilon$-disclosing and $P(\mathcal{M}(\mu_0)(O^{-1}(E_\epsilon) \cap \text{Rel}) = \lambda > 0$. By definition of $P_{\text{Rel}}$, this implies that $P_{\mathcal{M}(\mu_0)}(O^{-1}(E_\epsilon) \setminus \text{Rel}) > \frac{\lambda}{\epsilon}$. Therefore, $P(\text{Rel}) < 1 - \frac{\lambda}{\epsilon} < 1$.

Given $\mathcal{M}$ an oMDP, $\mathcal{M}$ is thus almost-surely AA-disclosing iff there exists a strategy $\sigma$ such that $P_{\mathcal{M}(\mu_0)}(\text{Rel}) = 1$. As $\text{Rel}$ is defined by the reachability of a set of states, this is equivalent to almost-sure reachability in MDP which is known to be in PTIME.

### E Minimal disclosure

**Theorem 27.** The minimal almost-sure and the minimal limit-sure AA-disclosure decision problems are undecidable.

**Proof.** Given a probabilistic automaton $\mathfrak{A} = (Q, \{a, b\}, q_0, T, F)$ over $\{a, b\}$ we first transform $\mathfrak{A}$ into an incomplete MDP $\hat{\mathfrak{A}} = (Q, \{e, \}, p, \hat{O})$ as in the proof of Theorem 25.

From $\hat{\mathfrak{A}}$ we build the MDP $M = (S, \{e, c, l\}, p, O)$ (represented in Figure 8) where:

- $S = \hat{Q} \cup \{s_0, s_0^1, s_0^2, s_1, s_1^1, s_1^2, s_2, s_2^1, s_2^2\}$,
- $p(q_0 | q_0, e) = p(s_0^1 | s_0, e) = \frac{1}{2}$. For $q_1, q_2 \in \hat{Q}$, $p(q_2 | q_1, e) = \hat{p}(q_2 | q_1, e)$.
- For $q \in Q$, if $q \in F$ then $p(s_1^1 | q, e) = \frac{1}{2}$ else $p(s_1^2 | q, e) = \frac{1}{2}$. $p(s_0^1 | s_0, e) = p(s_1^1 | s_1^0, e) = 1/2$. $p(s_0^2 | s_0, e) = p(s_1^2 | s_1^0, e) = 1/2$. $p(s_0^2 | s_0, e) = p(s_1^2 | s_1^0, e) = 1$.
- $p(s_2^1 | s_1^1, l) = 1$, $p(s_2^1 | s_1^1, c) = p(s_2^1 | s_1^1, l) = \frac{1}{2}$. $p(q_0 | s_2^2, c) = p(q_0 | s_2^2, l) = \frac{1}{2}$. $p(q_0 | s_2^1, e) = p(q_0 | s_2^1, e) = 1$.
- $p(s_2^1 | s_1^1, l) = 1$, $p(s_2^1 | s_1^1, c) = p(s_2^1 | s_1^1, l) = \frac{1}{2}$. $p(q_0 | s_2^2, c) = p(q_0 | s_2^2, l) = \frac{1}{2}$. $p(q_0 | s_2^1, e) = p(q_0 | s_2^1, e) = 1$.

From $\hat{\mathfrak{A}}$ we build the MDP $\hat{\mathfrak{A}} = (Q, \{e, c, l\}, p, \hat{O})$ (represented in Figure 8) where:

- $S = \hat{Q} \cup \{s_0, s_0^1, s_0^2, s_1, s_1^1, s_1^2, s_2, s_2^1, s_2^2\}$, (i.e. the upper component of the system). We will show that $\text{Disc}_{\epsilon}^{\text{AA}}(M) > 0$ iff there exists a word $w$ such that $P_A(w) > \frac{1}{2}$.

The idea of the proof is the following. During the first transition one goes with some probability in $s_1^0$ or $s_2^0$ (lower and upper systems of the Figure 8). Then, on both side a word $w \in (a + b)^*$ is read with same probability, a $\sharp$ marking the end of the word. On the upper side, this $\sharp$ is followed by a $\flat$ with probability $P_A(w)$ and by a $\flat$ otherwise. On the lower side, a $\flat$ is read with a probability chosen by the controller between 0 and $\frac{1}{2}$ and a $\flat$ otherwise. Therefore, the controller can reproduce the same probability on both side of the system (and thus give no information to the observer) iff the acceptance probability of $w$ in $\mathfrak{A}$ is between
0 and \( \frac{1}{2} \). The execution then starts again from the initial state of both copies of the automaton.

More formally, suppose that there exists a word \( w_d \in \{a, b\}^* \) such that \( \mathbb{P}_A(w_d) > \frac{1}{2} \). Given a finite observation \( w \in \Sigma^* \), we define the value \( \text{ratio}_{w_d}(w) \) as the ratio between the number of occurrence of \((a + b)w_d\) over the number of occurrence of \((a + b)w_{\bar{d}}\) in \( \Sigma^w \). This definition is extended to infinite observations by taking the limit, when defined, of the ratios of its finite prefixes. Let \( \sigma \) be any strategy. We define the set of observations \( E = \{ w \in \Sigma^w \mid \text{ratio}_{w_d}(w) > 1/2 \} \). Thanks to the weak law of large numbers and by choice of \( w_d \), we have that \( \mathbb{P}_{M_{\sigma}}(1_{\sigma_0 = \epsilon_0}) = 1 \) and \( \mathbb{P}_{M_{\sigma}}(1_{\sigma_0 = \epsilon_0}^1) = 0 \). As from the initial state, a path of \( M_{\sigma}(\mu_0) \) goes either in \( s_0 \epsilon q_0 \), becoming a relevant path, or \( s_0 \epsilon q_0^1 \), from which it can never become relevant, this implies that with probability 1, a relevant path has an observation belonging to \( E \). Let us show that these relevant paths are almost surely AA-disclosing which will imply that \( \text{Disc}_{\min}(\text{AA}, M(\mu_0)) = \frac{1}{2} \).

For every \( n \in \mathbb{N} \), let \( \mathcal{S}_n \) be the set of prefixes of length \( n \) of the observations of \( E \): \( \mathcal{S}_n = \{ \sigma \in \Sigma^0_n \mid \exists \sigma' \in E, \sigma \preceq \sigma' \} \). For every \( \varepsilon > 0 \), we also define \( \mathcal{S}_n^\varepsilon \) as the subset of \( \mathcal{S}_n \) consisting of observations whose proportion of relevant paths exceeds threshold \( 1 - \varepsilon \) in \( M(\mu_0) \): \( \mathcal{S}_n^\varepsilon = \{ \sigma \in \mathcal{S}_n \mid \mathbb{P}_{M_{\sigma}}(\varepsilon_0) < 1 - \varepsilon \} \).

From \( n \in \mathbb{N} \), the set \( \mathcal{C}(\mathcal{S}_n) = E \), we derive that \( \lim_{n \to \infty} \mathbb{P}_{M_{\sigma}}(1_{\mu_0 \in \epsilon_0})(\mathcal{S}_n) = \mathbb{P}_{M_{\sigma}}(1_{\mu_0 \in \epsilon_0})(E) = 0 \). Thus \( \lim_{n \to \infty} \mathbb{P}_{M_{\sigma}}(1_{\mu_0 \in \epsilon_0})(\mathcal{S}_n^\varepsilon) = 0 \).

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

\[
\mathbb{P}_{M_{\sigma}}(1_{\mu_0 \in \epsilon_0})(\mathcal{S}_n^\varepsilon) = \sum_{\sigma \in \mathcal{S}_n^\varepsilon} \mathbb{P}_{M_{\sigma}}(1_{\mu_0 \in \epsilon_0})(\sigma) > \sum_{\sigma \in \mathcal{S}_n^\varepsilon} \frac{\varepsilon}{1 - \varepsilon} \mathbb{P}_{M_{\sigma}}(1_{\mu_0 \in \epsilon_0})(\sigma) = \frac{\varepsilon}{1 - \varepsilon} \mathbb{P}_{M_{\sigma}}(1_{\mu_0 \in \epsilon_0})(\mathcal{S}_n^\varepsilon) .
\]

Since \( \varepsilon \) is fixed, \( \mathbb{P}_{M_{\sigma}}(1_{\mu_0 \in \epsilon_0})(\mathcal{S}_n^\varepsilon) < \frac{1}{1 - \varepsilon} \mathbb{P}_{M_{\sigma}}(1_{\mu_0 \in \epsilon_0})(\mathcal{S}_n^\varepsilon) \) and \( \lim_{n \to \infty} \mathbb{P}_{M_{\sigma}}(1_{\mu_0 \in \epsilon_0})(\mathcal{S}_n^\varepsilon) = 0 \). This implies that with probability 1, a path whose observation belongs to \( E \) is \( \varepsilon \)-disclosing. As this holds for every \( \varepsilon > 0 \), from Proposition 8, we deduce that the infinite paths with observations in \( E \) are almost surely AA-disclosing.

Conversely, suppose that every word \( w \in \{a, b\}^* \) verifies \( \mathbb{P}_A(w) = \lambda \leq \frac{1}{2} \). We define the strategy \( \sigma \) such that after a path \( \rho \) ending in \( s_1^2 \) with an observation \( w_1 d_1 w_2 d_2 \cdots w_n d_n \) where \( w_i \) is a word of \( \{a, b\}^* \) and \( d_i \in \{a, b\} \), \( \sigma(\rho)(c) = 2.\mathbb{P}_A(w) \) and \( \sigma(\rho)(l) = 1 - 2.\mathbb{P}_A(w_n) \). For every other path, \( \sigma \) chooses the only available action: \( c \). With this choice, for all \( i \in \mathbb{N} \) the probability that \( d_i \) is equal to \( \bar{c} \) for a secret or a non secret path is equal to \( \mathbb{P}_A(w_n) \). Therefore, for any finite path \( \rho \), \( \mathbb{P}_{M(\mu_0)}(\sigma(\rho)) = 1/2 \). Thus \( \text{Disc}_{\min}(\text{AA}, M(\mu_0)) = 0 \).

Consequently, the minimal almost sure disclosure decision problem is undecidable.
Fig. 8. Reduction from the emptiness problem to the minimal almost-sure disclosure problem. The square state corresponds to a final state of $\tilde{A}$. 