On Abstraction-Based Controller Design With Output Feedback

Rupak Majumdar  
MPI-SWS, Germany  
rupak@mpi-sws.org

Necmiye Ozay  
Univ. of Michigan, Ann Arbor, USA  
necmiye@umich.edu

Anne-Kathrin Schmuck  
MPI-SWS, Germany  
akschmuck@mpi-sws.org

ABSTRACT

We consider abstraction-based design of output-feedback controllers for dynamical systems with a finite set of inputs and outputs against specifications in linear-time temporal logic. The usual procedure for abstraction-based controller design (ABCD) first constructs a finite-state abstraction of the underlying dynamical system, and second, uses reactive synthesis techniques to compute an abstract state-feedback controller on the abstraction. In this context, our contribution is two-fold: (I) we define a suitable relation between the original system and its abstraction which characterizes the soundness and completeness conditions for an abstract state-feedback controller to be refined to a concrete output-feedback controller for the original system, and (II) we provide an algorithm to compute a sound finite-state abstraction fulfilling this relation.

Our relation generalizes feedback-refinement relations from ABCD with state-feedback. Our algorithm for constructing sound finite-state abstractions is inspired by the simultaneous reachability and bisimulation minimization algorithm of Lee and Yannakakis. We lift their idea to the computation of an observation-equivalent system and show how sound abstractions can be obtained by stopping this algorithm at any point. Additionally, our new algorithm produces a realization of the topological closure of the input/output behavior of the original system if it is finite-state realizable.

CSC CONCEPTS

- Computer systems organization → Sensors and actuators; Robotic control  
- Theory of computation → Abstraction;

ACM Reference Format:

1 INTRODUCTION

Controller synthesis for dynamical systems against specifications in linear temporal logic is a core problem in correct-by-construction design of cyber-physical systems. One way to solve this problem relies on abstracting the state space to a finite-state system, followed by algorithmic techniques from reactive synthesis to compute an abstract controller which is then refined to a concrete one for the original system [1, 7, 22, 26]. Most algorithms, and certainly most state-of-the-art synthesis tools such as SCOTS [23], pFaces [11], or Mascot [10], implement this abstraction-based control design (ABCD) workflow while assuming the entire state of the underlying system to be observable. In this paper, we relax the condition of full state observation. We consider ABCD when the system has a finite number of observable outputs and a controller must decide its input choice (from a finite set) based solely on the history of applied inputs and observed outputs. Such output-feedback control is common in control design, as the observation of the state is usually limited by the availability and precision of the sensors.

As an example, consider the tank reactor shown in Fig. 1. It has a finite number of water level sensors ($l_0, \ldots, l_5$) which indicate whether the current water level touches the sensor or not by returning true or false. Further, it can be observed (but not controlled) whether the outlet valve is open ($o = \text{true}$) or closed ($o = \text{false}$). The controller can set the inlet valve open (by applying $u = +$) or closed (by applying $u = 0$). The actual state of the system, i.e., the precise value of the water level, is not observable. In this example, a given input/output sequence of observed true sensor values and applied inputs (e.g., $v = \{l_0\} \{+\} \{l_0\} \{+\} \{l_0, l_1, o\} \{0\} \{l_0, o\} \{+\} \ldots$) provides certain knowledge about the current state of the tank system, which might be sufficient to implement a controller ensuring the satisfaction of a specification over the observables. For example, one might want to ensure that the tank never overflows (i.e., $l_5$ never becomes true) while still containing a limited amount of water (i.e., $l_1$ is always true). We show how finite-state abstractions of the input/output behavior of such an infinite state dynamical system can be constructed for the purpose of ABCD with output-feedback.

There is a rich history of output-feedback control design for continuous dynamical systems w.r.t. classical control objectives (such as stability or tracking) based on observer design [13, 25], with recent extensions to systems with finite external alphabets [6] and estimator-based abstractions for control with partial-information [5, 8, 16]. In the context of temporal-logic control of finite-state systems, output-feedback control gives rise to games of incomplete

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{tank.png}
\caption{Tank reactor modeled as a dynamical system $S$ over an infinite bounded state space $X \subseteq \mathbb{R}^3$ with finite input space $U \subseteq \{+, 0\}$ and finite output space $Y \subseteq 2^\sigma$ denoting the set of sensors $\sigma = \{l_0, \ldots, l_5, o\}$ which are currently ‘true.’}
\end{figure}
We tackle the termination problem similar to the \( l \)-complete abstraction framework \cite{Lee96}. Since KAM always constructs sound finite-state abstractions of the original system, we can run a synthesis procedure at any point to see if an abstract controller ensuring the specification exists. If a controller can be found, the abstraction construction can stop. If not, the construction continues until we try again after a future iteration. This iterative ABCD procedure is sound and relatively complete—i.e., a topologically closed finite-state abstraction that allows to construct an abstract controller for the given specification exists, our procedure will eventually find it.

Additional proofs can be found in the extended version \cite{Majumdar14}.

## 2 PRELIMINARIES

### Notation

We use the symbols \( \mathbb{N}, \mathbb{Z}, \mathbb{R}, \) and \( \mathbb{R}_{\geq 0} \) to denote the sets of natural numbers, integers, reals, and positive reals, respectively. Given \( a, b \in \mathbb{R} \) s.t. \( a \leq b \), we denote by \([a, b]\) a closed interval and define \([a, b] = \{a, b\} \cap \mathbb{Z} \) as its integer counterpart. For a set \( W \), we write \( W^+ \) and \( W^w \) for the sets of finite and infinite sequences over \( W \), respectively. For \( w \in W^+ \), we write \(|w| \) for the length of \( w \) and \( \varepsilon \) for the empty string with \(|\varepsilon| = 0 \); the length of \( w \in W^w \) is \( \infty \). We define \( \text{dom}(w) = \{0, \ldots, |w| - 1\} \) if \( w \in W^+ \), and \( \text{dom}(w) = \mathbb{N} \) if \( w \in W^w \). For \( k \in \text{dom}(w) \) we write \( w(k) \) for the \( k \)-th symbol of \( w \) and \( w[0:k] \) for the restriction of \( w \) to the domain \([0:k]\). Given two sets \( A \) and \( B \), \( f : A \rightarrow B \) and \( f : A \rightarrow B \) denote a set-valued and ordinary map, respectively. \( f \) is called strict if \( f(a) \neq \emptyset \) for all \( a \in A \). The inverse mapping \( f^{-1} : B \rightarrow A \) is defined via its respective binary relation: \( f^{-1}(b) = \{a \in A \mid b \in f(a)\} \). By slightly abusing notation, we lift maps to subsets of their domain in the usual way, i.e., for a set-valued map \( f : A \rightarrow B \) and \( a \subseteq A \) we have \( f(a) = \{b \mid \exists a \in a \cdot b \in f(a)\} \), and similarly for ordinary maps.

### Systems

A system \( S = (X, X_0, U, F, Y, H) \) consists of a state space \( X \), a set of initial states \( X_0 \subseteq X \), a finite input space \( U \), a strict set-valued transition function \( F : X \times U \rightarrow X \), a finite output space \( Y \), and an output function \( H : X \rightarrow Y \). To simplify notation, we assume that \( H \) respects \( X_0 \), that is, if \( H^{-1}(y) \cap X_0 \neq \emptyset \) we have \( H^{-1}(y) \subseteq X_0 \). The system \( S \) is called finite state if \( X \) is finite.

### Trace Semantics

A path of \( S \) is an infinite sequence \( \pi = x_0; x_1; u_1 \ldots u_n \ldots \) such that \( x_0 \in X_0 \) and for all \( k \in \mathbb{N} \) we have \( x_{k+1} \in F(x_k, u_k) \). The set of all paths over \( S \) is denoted by \( \text{Paths}(S) \). The prefix up to \( x_n \) of a path \( \pi \) over \( S \) is denoted by \( \pi[0:n] \) with length \( |\pi[0:n]| = n + 1 \) and last element \( \text{Last}(\pi[0:n]) = x_n \). The set of all such prefixes is denoted by \( \text{Pref}(S) \).

The unique external sequence of a path \( \pi \) of \( S \) is defined as \( \text{Ext}(\pi) = y_0; y_0y_1u_1 \ldots \), where \( y_k = H(x_k) \) for all \( k \in \mathbb{N} \). The sets of all external sequences over \( S \) are denoted by \( \text{Ext}(S) \) and we define \( \text{EPref}(S) := \text{Ext}(\text{Pref}(S)) \). The set \( \text{Ext}(S) \) is called topologically closed (or closed for short) if for any infinite sequence \( v = y_0; y_0y_1u_1 \ldots \in Y(U)^\omega \), whenever \( v[0:k] \in \text{EPref}(S) \) for all \( k \in \mathbb{N} \) it holds that \( v \in \text{Ext}(S) \). We say that \( S \) has closed external behavior if \( \text{Ext}(S) \) is closed (see, e.g., \cite{Sifakis90} for details).

We lift the map \( \text{Last} \) to external sequences and write \( x \in \text{LastX}S(\rho) \) if there exists \( \pi \in \text{Pref}(S) \) s.t. \( x = \text{Ext}(\pi) \) and \( x = \text{Last}(\pi) \). For a state \( x \in X \) we define all prefixes of \( S \) that reach \( x \) as \( \text{Hist}(x) = \{ \pi \in \text{Pref}(S) \mid \text{Last}(\pi) = x \} \) and all external sequences generated by such prefixes as \( \text{EHist}(x) = \{ \rho \in \text{EPref}(S) \mid x \in \text{LastX}S(\rho) \} \). If the system \( S \) we are referring to is clear from the context we omit the subscript \( S \) from the maps \( \text{LastX} \) and \( \text{EHist} \).
Control Strategies. We define state-feedback and output-feedback control strategies as functions $C^\dagger : \text{Pref}(S) \to U$ and $C : \text{EPref}(S) \to U$, respectively. We say that a path $x$ of $S$ is compliant with $C$ (resp. $C^\dagger$) if for all $k \in \mathbb{N}$, we have $u(k) = C(\text{Ext}(\tau_{0\cdots k-1}))$ (resp. $u(k) = C^\dagger(\tau_{0\cdots k-1}))$. We denote the set of all paths and prefixes of $S$ compliant with $C$ by $\text{CPrefs}(S, C)$ and $\text{CPrefs}(S, C^\dagger)$, respectively. We further use $\text{Ext}(S, C)$ and $\text{EPref}(S, C)$ to denote the sets $\text{Ext}(\text{CPrefs}(S, C))$ and $\text{Ext}(\text{CPrefs}(S, C^\dagger))$, respectively. For a state-feedback controller $C^\dagger$ all sets are defined analogously.

A sound abstraction of a controller $C$ with a system $S$ over the set of path prefixes, the set $\text{Ext}(S, C)$ is topologically closed if $\text{Ext}(S)$ is.

Control Problem. We consider $\omega$-regular specifications over a finite set of atomic input and output propositions $AP_I$ and $AP_O$. We omit the standard definitions of $\omega$-regular languages (see, e.g., [27, 28]). To simplify notation, we assume that $U = 2^{AP_I}$ and $Y = 2^{AP_O}$. In this setting, an $\omega$-regular specification $\psi$ can be written as a language $\{\psi\} \subseteq \{Y(U)^\omega\}$ of desired external sequences. Given a system $S$ and a specification $\psi$, the output-feedback control problem, written $(\hat{S}, \hat{\psi})$, asks to find an output-feedback control strategy $S$ such that $\text{Ext}(\hat{S}, \hat{\psi}) \subseteq \{\psi\}$. We define $\mathcal{W}(S, \hat{\psi}) = \{C | \text{Ext}(S, C) \subseteq \{\psi\}\}$ as the set of all such output-feedback control strategies. For a state-feedback controller $C^\dagger$, we define analogously the set $\mathcal{W}^\dagger(S, \hat{\psi})$.

3 ABSTRACTION-BASED CONTROLLER DESIGN WITH OUTPUT-FEEDBACK

Abstraction-Based Controller Design (ABCD) is a well-known approach to solving a controller synthesis problem for a dynamical system $S$ against specifications defined by a language $\{\psi\}$. Here, the dynamical system $S$ is first abstracted to a finite-state system $\hat{S}$ and then techniques from reactive synthesis (e.g., [15, 28]) are used to design an abstract controller for $\hat{S}$ ensuring $\psi$.

In this section, we will formalize the required relation between $S$ and $\hat{S}$ to refine an abstract state-feedback controller $C^\dagger$ on $\hat{S}$ to an output-feedback controller $C$ on $S$. We start our formalization by providing a general definition of sound abstractions in Sec. 3.1 which adapts feedback refinement relations [22] to systems with finite input and output sets. We show that for this definition the usual refinement of an abstract state-feedback controller to a concrete state-feedback controller carries over from [22]. As the main contribution of this section, we then show in Sec. 3.2 that the definition of sound abstraction needs to be applied to the external trace semantics of $S$ rather than to its state transitions to allow for ABCD with output feedback control.

3.1 Sound Abstractions

Given two systems we define a sound abstraction as follows.

**Definition 3.1.** Let $S = (X, X_0, U, F, Y, H)$ and $\hat{S} = (\hat{X}, \hat{X}_0, U, \hat{F}, \hat{Y}, \hat{H})$ be systems. Further, let $\alpha : X \Rightarrow \hat{X}$ and $\gamma : \hat{X} \Rightarrow X$ be two set valued functions s.t. $x \in \gamma(\hat{x})$ iff $\hat{x} \in \alpha(x)$. Then we call $\hat{S}$ a sound abstraction of $S$, written $S \models_\alpha \hat{S}$, if

(A1) $\alpha(x_0) \subseteq \hat{X}_0$,

(A2) $\forall \hat{x} \in \hat{X}, u \in U : \alpha(F(x(u))) \subseteq \hat{F}(\alpha(x), u)$, and

(A3) $\forall \hat{x} \in \hat{X} : H(y(\hat{x})) \subseteq (\hat{H}(\hat{x})$).

$\hat{S}$ is a sound realization of $S$, written $S \models_\alpha \hat{S}$, if $S \models_\alpha \hat{S}$ and $\hat{S} \models_\gamma S$.

As common in abstract interpretation [4], we make $\gamma$ explicit in Def. 3.1 to emphasize that $\{\hat{x}\} \subseteq \alpha(\gamma(\hat{x}))$, where equality may not hold. However, to simplify notation, we often omit $\gamma$ and write $\models_\alpha$ and $\models_\gamma$, as $\gamma$ is fully determined by knowing $\alpha$. Further, we write $\models$ to indicate that there exists $\alpha$ s.t. $\models_\alpha$ holds.

**Remark 1.** Sound abstractions are an adaptation of feedback refinement relations (FRR) [22, Def. V.2] to systems with finite input and output sets in the following sense.

(A1): An FRR is defined for fully initialized systems (i.e., $X_0 = X$), where (A1) follows from the fact that an FRR must be a strict relation.

(A2): To simplify notation, we assume that $F$ is a strict function.

This implies that all inputs are enabled in every state, i.e., $\text{Enab}_S(x) = \{u \in U | F(x, u) \neq \emptyset\} = U$ for all $x \in X$. The definition of FRR makes $\text{Enab}_S$ explicit by replacing (A2) with the two conditions

(A2.1) $\forall x \in X : \text{Enab}_S(x) \subseteq \text{Enab}_S(x)$, and

(A2.2) $\forall x \in X, u \in \text{Enab}_S(x) : \alpha(F(x(u))) \subseteq \hat{F}(\alpha(x), u)$, where $\text{Enab}_S(x)$ coincides with (A2) if $\text{Enab}_S(x) = U$.

(A3): An FRR is defined for systems with full state observation, i.e., $Y = X$, $\hat{Y} = \hat{X}$ and $\hat{H} = H = \text{id}$ with $\text{id}(x) = x$ for all $x \in X$.

This renders $Y$ infinite if $X$ is infinite and does not allow the direct interpretation of an $\omega$-regular specification over $U$ and $Y$. While our condition (A3) enables the use of a common specification for both $S$ and $\hat{S}$ (due to their equivalent finite input/output spaces), this is not possible in [22], due to $Y$ being infinite and $\hat{Y} = \hat{X}$. [22, Def.VI.2] handles this by defining a different abstract specification from the defined FRR and the specification over the original system $S$.

Observe that for a system $S$ and its sound abstraction $\hat{S}$, corresponding states in two runs $x_0 \hat{u}_0 x_1 \ldots$ and $\hat{x}_0 u_0 \hat{x}_1 \ldots$ stay related by $\alpha$ during arbitrarily but finite executions, if they start at related initial states $\hat{x}_0 \in \alpha(x_0)$ (A1) and the same input sequence is applied (A2). In this case (A3) ensures that $\hat{S}$ always produces a subset of the outputs generated by $\hat{S}$ in every instance of the trace. This implies that any arbitrarily but finite external sequence $v$ generated by the original system is contained in $\text{EPref}(S)$. Therefore, any abstract controller solving a given control problem over $\hat{S}$ can be guaranteed to be refinable to a sound controller for $S$, if $\hat{S}$ has closed external behavior. If this is not the case, spurious infinite external traces generated by this controller on $S$ which are not contained in $\text{Ext}(\hat{S})$ might violate the specification. Requiring $\hat{S}$ to have closed external behavior is not with loss of much generality in ABCD: any finite-state system (of the form considered in this paper) has closed external behavior, and we require $S$ to be finite-state in order to apply reactive synthesis techniques for abstract controller design anyways. The next theorem formalizes the above discussion for ABCD with state feedback. The proof uses the same insights as the proof of [22, Thm.VI.3] and is provided in [14].

**Theorem 3.2.** Let $S$ and $\hat{S}$ be systems s.t. $\hat{S}$ has closed external behavior. If $S \models_\gamma \hat{S}$ and $\hat{S} \models_\gamma S$ then $\text{Ext}(\hat{S}, \hat{\psi})$ then $C^\dagger = C^\dagger \circ \alpha \models_\gamma \text{W}^\dagger(\hat{S}, \hat{\psi})$. Further, if $S$ has closed external behavior and $S \models_\gamma \hat{S}$ then $\text{W}^\dagger(\hat{S}, \hat{\psi}) \models_\gamma \emptyset$ iff $\text{W}^\dagger(\hat{S}, \hat{\psi}) = \emptyset$. \footnote{See Rem. 2 in Sec. 4.1 for a discussion of this choice.}
3.2 Sound Abstractions for Output Feedback

Now we consider the case of output feedback. Here, the only available information about the system $S$ that we can utilize for control are external prefixes $v \in \text{EPrefs}(S)$. With this, however, we usually cannot uniquely determine the current state of the system, i.e., $\text{Last}(X(v))$ is usually a set of states and not a singleton. Further, it is well known that any state of a system (and therefore closed) abstraction $\tilde{S}$ with state-feedback control can be utilized for output-feedback control. We further have

$$\text{Ext}(\tilde{S}) \subseteq \text{Ext}(S^*)$$

for the external trace system $\tilde{S}$ of $S$ in which a state represents a finite external history of $S$, and the transitions extend the external history by one step.

**Definition 3.3.** Given a system $S = (X, X_0, U, F, Y, H)$, its induced external trace system is the system $S^* = (X^*, X_0^*, U, F^*, Y, H^*)$, where $X^* := \text{EPrefs}(S), X_0^* := H(X_0), F^*(\rho, u) := \{py | F(\text{Last}(\tilde{x}), u) \cap H^{-1}(y) \neq \emptyset\}$. It should be noted that, by definition, $S^*$ has closed external behavior. We further have $\text{EPrefs}(S) = \text{EPrefs}(S^*), \text{Ext}(S^*) \subseteq \text{Ext}(S^*)$, and $\text{Ext}(S) = \text{Ext}(S^*)$ iff $S$ has closed external behavior. That is, $\text{Ext}(S^*)$ is the behavioral closure of $\text{Ext}(S)$ [29].

To refine an abstract state-feedback controller to an output-feedback controller for the original system, one needs to relate useful information about the system state $x$ to the so-called external trace system $S^*$ of $S$ in which a state represents a finite external history of $S$, and the transitions extend the external history by one step.

**Algorithm 1 KA: Knowledge-Based Abstraction**

1. $X_0 \leftarrow \{x_0 \cap H^{-1}(y) \in 2^X \mid y \in Y\}$
2. $\tilde{X}_0 \leftarrow \emptyset$ and $\tilde{X} \leftarrow X_0$
3. while $\tilde{X}_0 \neq \tilde{X}$ do
   4. $\tilde{X}_0 \leftarrow \tilde{X}$
5. for $\tilde{x} \in \tilde{X}_0, u \in U, y \in Y$ do
   6. $\tilde{X} \leftarrow F(\tilde{x}, u) \cap H^{-1}(y)$
   7. $\tilde{X} \leftarrow \tilde{X} \cup (\tilde{x})'$ if $\tilde{x}' \neq \emptyset$
8. end for
9. end while
10. Define $\tilde{x}' \in \tilde{F}(\tilde{x}, u)$ iff there exist $y$ s.t. $\tilde{x}' = F(\tilde{x}, u) \cap H^{-1}(y)$
11. Define $H(\tilde{x}) = y$ iff $y \in H(\tilde{x})$
12. return $\tilde{S}^k = (\tilde{X}, \tilde{X}_0, U, Y, \tilde{F}, \tilde{H})$

Cor. 3.4 to obtain a sound ABCD framework for output-feedback control without explicitly computing $S^*$.

4 COMPUTING ABSTRACTIONS

We now turn to the algorithmic problem of computing system abstractions such that designing a state-feedback controller on the abstraction allows us, through Cor. 3.4, to construct a corresponding output-feedback controller for the original system. For this we assume that the original system has an infinite state space—e.g., defined by a continuous-state dynamical system—and our goal is to compute a finite-state abstraction on which algorithmic techniques for state-based controller synthesis (e.g., [15, 28]) can be applied.

We first recall two well-known approaches to compute such finite-state abstractions which were developed for the setting where the original system has a finite state space, and show that they may not terminate for infinite-state systems, even if a finite-state realization of the topological closure of its external behavior exists. Based on this insight, we provide (Sec. 4.4) an algorithm for abstracting infinite-state systems which overcomes this problem.

4.1 Knowledge-Based Abstraction

A standard way to solve control-strategy synthesis problems over finite-state systems with partial observation [3, 20, 31] is to use a knowledge-based subset construction. Starting from the subsets of initial states generating the same output, the knowledge-based subset construction algorithm, given in Alg. 1, explores all inputs to the system and successively generates subsets of states that are indistinguishable given the full history of applied inputs and observed outputs. Such subsets $\tilde{x}$ of states of the original system $S$ become the states of the knowledge-based abstraction $\tilde{S}^k := \text{KA}(S)$. Note that every reachable state $\tilde{x}$ of $\tilde{S}^k$ computed via Alg. 1 has the property that all $x \in \tilde{x}$ have the same output; thus, we can define $H(\tilde{x})$ as the (unique) output $H(x)$ of some $x \in \tilde{x}$.

**Remark 2.** We restrict our attention to systems with strict transition function in this paper to simplify the discussion of the KA algorithm in Alg. 1 and KAM in Alg. 2. If not all inputs are enabled in every state, KA would need to distinguish state sets further based on the set of available inputs. This would require the controller to

HSCC ’20, April 22–24, 2020, Sydney, NSW, Australia

Rupak Majumdar, Necmiye Ozay, and Anne-Kathrin Schmuck
The next proposition formalizes the intuition that $\mathcal{S}^\mathcal{K}$ is a useful abstraction for a given output-feedback control problem over $S$. With Prop. 4.1 in place, it immediately follows from Cor. 3.4 that one can compute an output-feedback controller $C := \hat{C} \circ \text{LastX}_{\mathcal{S}^\mathcal{K}} \in \mathcal{W}(S, \hat{\psi})$ from an abstract state-feedback controller $\hat{C} \in \mathcal{W}(\mathcal{S}^\mathcal{K}, \hat{\psi})$, if it exists.

**Proposition 4.1.** Let $S$ be a system, $S^*$ its external trace system, and $\mathcal{S}^\mathcal{K} = \text{KA}(S)$. Then, $S^* \equiv_\alpha \mathcal{S}^\mathcal{K}$ with $\alpha = \text{LastX}_{\mathcal{S}^\mathcal{K}}$.

**Proof.** To simplify notation we define $\mathcal{S} := \mathcal{S}^\mathcal{K}$.

- We first prove that $\text{LastX}_{\mathcal{S}^\mathcal{K}}(\text{EHist}^\mathcal{S}(\bar{x})) = \{\bar{x}\}$ for all $\bar{x} \in \bar{X}$ by picking $\bar{x} = \bar{x}_0u_1x_1 \ldots \bar{x}_n$ and $\bar{z} = \bar{z}'_0u_0x_0 \ldots \bar{z}'_n$ s.t. $\bar{H}(\bar{x}_k) = \bar{H}(\bar{z}'_k)$ for all $k \in [0, n]$ and showing $\bar{x} = \bar{x}'_n$ by induction. For $k = 0$ we have $\bar{x}_0, \bar{z}'_0 \in \bar{x}_0$. As $\bar{H}(\bar{x}_0) = \bar{H}(\bar{z}'_0)$, we have $\bar{x}_0 = \bar{x}'_0$. Now let $k \in [1, n]$ and assume $\bar{x}_k = \bar{z}'_{k-1}$. Then it follows that there exists $y, y'$ s.t. $\bar{x}_k = F(\bar{x}_{k-1}, u_{k-1}) \cap H^{-1}(y)$ and $\bar{z}'_k = F(\bar{z}'_{k-1}, u_{k-1}) \cap H^{-1}(y')$. Again, $\bar{H}(\bar{x}_k) = \bar{H}(\bar{z}'_k)$ implies $y = y'$. Then it is easy to see that $\bar{x} = \bar{z}'_n$.

- We now show that equality holds for (A1)-(A3) from Def. 3.1: *(A1): By definition, $X^* = H(X_0)$; and by line 1 in Alg. 1, we have $\text{LastX}_{\mathcal{S}}(H(X_0)) = \bar{x}_0$. *(A2): Let $\bar{x} = \text{LastX}_{\mathcal{S}}(v)$ and $\bar{u} \in U$. Further, let $\bar{y}' = F(\bar{x}, u) \cap H^{-1}(y)$ and define $Y = \{y \in Y \mid \bar{y}' \neq \emptyset\}$. Now recall that $F(v, u) = \{v_y \mid F(\text{LastX}_{\mathcal{S}}(v), u) \cap H^{-1}(y) \neq \emptyset\}$. This implies $\bar{y}' \in \text{LastX}_{\mathcal{S}}(F^*(v, u))$ if $y \in Y$. *(A3): Let $\bar{x} = \text{EHist}^\mathcal{S}(\bar{x})$; we have $\text{LastX}_{\mathcal{S}}(F^*(v, u)) = \bigcup_{y \in Y} \text{LastX}_{\mathcal{S}}(\bar{y}'_y)$. From the definition of $\bar{F}$, it further follows that $\bar{y}' \in \bar{F}(\bar{x}, u)$ if $y \in Y'$ and in particular $\bar{F}(\bar{x}, u) = \bigcup_{y \in Y'} \bar{y}'_y$. Recalling that $\bar{x} = \text{LastX}_{\mathcal{S}}(v)$ this shows that $\text{LastX}_{\mathcal{S}}(F^*(v, u)) = \bar{F}(\text{LastX}_{\mathcal{S}}(v), u)$.

* (A3): Observe that $y = H(\bar{x})$ for $\alpha = \text{LastX}_{\mathcal{S}}$. Then $H(y(\bar{x})) = H(\text{EHist}^\mathcal{S}(\bar{x})) = H(\hat{\mathcal{S}}(\bar{x})), \text{hence } H(\hat{\mathcal{S}}(\bar{x})) = \{H(\bar{x})\}$. □

Alg. 1 incrementally constructs $\mathcal{S}^\mathcal{K}$ from $S$ by forward exploration from the initial states. As the abstract state space $\bar{X} \subseteq 2^X$ contains subsets of $X$ it terminates if $X$ is finite. This case is the one most prominently discussed in existing literature, e.g., in [3, 31]. However, Alg. 1 might also terminate if $\bar{X}$ is infinite (see, e.g., the example in Sec. 4.3), given that the necessary operations (in particular "Post" and "Intersect") can be implemented if state subsets are infinite. If $X$ is infinite, Alg. 1 might however also not terminate even if there exists a finite-state realization of $S$. This is shown in Ex. 4.2. It is interesting to note that this might still be the case even if $X = \{0\}$. This can be verified by checking that Alg. 1 does also not terminate if all states in the system $\mathcal{S}$ depicted in Fig. 2 are initial.

**Example 4.2.** Consider the infinite state system $\mathcal{S}$ in Fig. 2, with $U = \{u\}, Y = \{A, B\}$. By omitting the trivial input, the external language $\text{Ext}(S)$ of this system is $\mathcal{A}^T(A)^* \cup \mathcal{A}^T(B)^*$, for which one can construct a finite trace equivalent system, for instance, using one of the methods discussed in the following sections. Yet, Alg. 1 will separate every state labeled with $B$, leading to an infinite chain of states with observation $B$, and will therefore not terminate.

### 4.2 Bisimulation Minimization

The knowledge-based abstraction algorithm $\text{KA}$ computes reachable subsets going forward, but it may fail to terminate by trying to distinguish states that are language equivalent to already computed ones, that is, states that generate the same future sequence of outputs under the same input sequence. Thus, one could first compute a bisimulation quotient $[2, 9, 17]$ of the system $S$ and only then compute the knowledge-based abstraction. It is possible that an infinite-state system has a finite bisimulation quotient; in that case, constructing the quotient first will allow the knowledge-based abstraction to terminate (see Fig. 2 (bottom) for an example).

For a system $S = (X, X_0, U, F, Y, H)$, a partition of the set $X$ is a set of non-empty sets of $X$, called blocks, that are pairwise disjoint and whose union is $X$. A partition is *stable* if the following properties hold. First, for each block $\bar{x}$ of the partition, every state in the block has the same output: for all $x, x' \in \bar{x}$, we have $H(x) = H(x')$. Second, for each block $\bar{x}$ of the partition, every state in the block has the same possible future outputs: for all $x, x' \in \bar{x}$ and $u \in U$, we have $Post_u(x) = Post_u(x')$, where $Post_u(x) := \{x' \in F(x, u) \cap \bar{x}' \neq \emptyset\}$.

Using the notion of a stable partition of $X$ we can define the *bisimulation abstraction* $\mathcal{S}^\beta = (\bar{X}, \bar{X}_0, \bar{F}, \bar{Y}, \bar{H})$ of $S$ as follows. The set of abstract states $\bar{X}$ is the minimal stable partition of $X$. The initial abstract states $\bar{X}_0$ are those blocks that contain some initial states from $X$. The abstract transition function is defined as $\bar{F}(\bar{x}, u) = \{\bar{x}' \in \bar{X} \mid \exists x \in \bar{x}. F(x, u) \subseteq \bar{x}'\}$. Moreover, since every state in each block of the partition has the same output, we can uniquely define $\bar{H}(\bar{x})$ to be the output of some state in $\bar{x}$.

A partition refinement algorithm [9, 19] can be used to compute $\mathcal{S}^\beta$ from $S$. Unlike Alg. 1, this algorithm proceeds backwards by splitting blocks based on their predecessors, starting with the partition defined by the outputs, i.e., $\{q \in 2^X \setminus \emptyset\} \ni \exists y \in Y \mid y = H^{-1}(q)$.
This algorithm may terminate if $X$ is infinite and the necessary operations are implementable over infinite state subsets. Going back to the system described in Ex. 4.2 we see that the bisimulation quotient $S_{bi}$ (depicted in Fig. 2 (bottom left)) is finite, while the original system $S$ (depicted in Fig. 2 (top left)) and its knowledge-based abstraction $S_{K}$ (depicted in Fig. 2 (top right)), are infinite. Applying the KA algorithm on $S_{bi}$ returns the desired finite state abstraction (depicted in Fig. 2 (bottom right)) which allows for output feedback control. However, if $S$ is infinite-state, the partition refinement algorithm is not guaranteed to terminate even if the knowledge-based abstraction of the original system is finite. This is further illustrated by the example discussed in the next section, which shows that knowledge-based abstraction and bisimulation minimization are incomparable and the suggested procedure to compute $S_{bi}$ first, before utilizing KA, may not terminate.

### 4.3 Illustrative Example

Before explaining KAM, we introduce an illustrative example. Consider the infinite state system $S$ depicted in Fig. 3 (top left) with $U = \{u\}$ and $Y = \{A, B, C, D, E, F\}$. It consists of one initial state $a_1$ which outputs $A$, an infinite chain of states $b_i, i \in \mathbb{N}$, all of which output $B$, and four different modules $A_i^f$, $A_{i}^{D}$ (light blue, dashed), $A_{i}^{l}$ (dark blue, dashed), $A_{i}^{L}$ (light orange, dotted) and $A_{i}^{h}$ (dark orange, dotted), attached to one $b$-state each. System $S$ is constructed s.t. modules of type $D$ (resp. of type $E$) are reachable after output $B$ has occurred an odd (resp. even) number of times, i.e., from all states $X_{B}^{odd} = \{b_{2i+1}\}_{i \in \mathbb{N}}$ (resp. from all states $X_{B}^{even} = \{b_{2i}\}_{i \in \mathbb{N}}$). However, the sequence of class $I$ and $II$ modules of the same type $i \in \{E, D\}$ is irregular, i.e., there is no $\omega$-regular expression to describe how $A_i^f$ and $A_i^{D}$ modules repeat.

By closely investigating the modules of the same $i$-type it can be observed that modules $A_i^f$ and $A_{i}^{D}$ for the same $i \in \{D, E\}$ are external language equivalent. Therefore, the regularity of alternating between type $D$ and type $E$ modules is enough to obtain a sound finite-state realization $S$ of $S$ depicted in Fig. 3 (top right).

**KA-algorithm (Sec. 4.1).** The KA algorithm computes the abstract state space by combining all states with the same observable past while going forward. For the system $S$ in Fig. 3 (top left) it constructs state subsets as depicted in Fig. 3 (bottom left). We see that the KA algorithm discovers that class I modules are a sound realization of class II modules, i.e., $S_{K}$ only consists of class I modules s.t. type $D$ and type $E$ modules are reachable from states in $X_{B}^{odd}$ and $X_{B}^{even}$ respectively. However, the KA algorithm still does not terminate on this example as it explores language equivalent states unnecessarily. I.e., by computing state subsets only going forward, it computes a new, not yet explored subset of $b$-states in every iteration. The KA-algorithm is not able to generalize and thereby merge all states corresponding to $X_{odd}^{f}$ or $X_{even}^{f}$ due to their unique future.

**Bisimulation-Quotient (Sec. 4.2).** A partition refinement algorithm computing the bisimulation quotient of $S$ merges states with the same observable future going backward. For the system $S$ in Fig. 3 (top left) it immediately discovers that all states in $X_{F} = \{f_i\}_{i \in \mathbb{N}}$ as well as $X_{G} = \{g_i\}_{i \in \mathbb{N}}$ have the same observable future (namely $F^{o}$ and $G^{o}$ respectively). It further merges all states contained in the same $A_i^f$ module into one equivalence class (see Fig. 3 (bottom right) indicated by the four color/line patterns). However, as it proceeds backwards, it does not take into account the reachable portion of all state subsets and thereby considers states within class I and II modules of the same type as different. This differentiates $b$ states depending on the class of modules they are connected to (indicated by the coloring of the $b$-states in Fig. 3 (bottom right)). As the partition refinement algorithm constructs equivalence classes going backward, it generates a distinct equivalence class for the left and right "color pattern" $a$ $b$ state "sees". As we assume that class I and II modules are irregularly sequenced, there exist infinitely many such equivalence classes and the algorithm therefore never terminates.

**Combining both algorithms.** For this example, running the KA algorithm first and the partition refinement algorithm second, results in the finite state abstraction $S$ depicted in Fig. 3 (top right). This is, however, not practically implementable, as the KA algorithm never terminates. Further, we have shown that for Ex. 4.2 one needs to execute the partition refinement algorithm first, followed by the KA algorithm. One can therefore construct an example where one reachable part of the state space requires executing the KA algorithm first, while the other part requires the partition refinement algorithm to be executed first. In this case, no order would lead to the desired result.

### 4.4 Knowledge Abstraction with Minimization

We now present the Knowledge-based Abstraction algorithm with Minimization (KAM), given in Alg. 2, which interlaces the forward Knowledge-based Abstraction (KA) with backward refinement-based Minimization (M). We also illustrate the algorithm using the example from Sec. 4.3.

**Algorithm Description.** KAM generates a rooted, labeled tree and a cover set $\text{Cover} \subseteq 2^X$. The nodes of the tree are kept in $\text{EXP}_X$ and the edges in $\text{EXF}_X$. The edges are labeled with inputs from $U$. The nodes are labeled with a three-tuple $\langle v, q, c \rangle \in \text{EXP}_X$, consisting of a sequence $v$ of external events seen when reaching the current node from the root of the tree, a block $q \subseteq X$ in the current Cover, and a subset of states $c \subseteq X$ (called a cell). Intuitively, a tuple $\langle v, q, c \rangle \in \text{EXP}_X$ remembers the observed input/output sequence from the initial states (in $v$), the available knowledge about the current state (in $c$), and the current "guesses" on states which are future observation-equivalent to $c$ (in $q$). The cells $c$ and blocks $q$ correspond to the data structures manipulated by the KA and the Minimization algorithm, respectively, and are initialized similarly: Cover is initialized with the partition induced by $H$ on $X$ (line 1, see Sec. 4.2), cells are initialized with all initial cover blocks containing an initial state (line 3). Note that the initialization of cells simplifies as we have assumed that $H$ respects the initial state set $X_0$.

**Example 4.3.** For the example in Sec. 4.3, we see that the partition induced by $H$ on $X$ results in the initial cover set $\text{Cover} = \{X_y \mid y \in Y\}$ s.t. $X_y$ collects all states of $S$ that generate the output $y$, e.g., $X_A := \{a_1\}$ and $X_C := \{c_i\}_{i \in \mathbb{N}}$. On the other hand, there is only one initial cell, namely $\{a_1\}$ with $H(\{a_1\}) = A$. This results in the initialization of $\text{EXP}_X$ with $\langle A, X_A, \{a_1\} \rangle$ (see Fig. 4 (left)).

The main loop of KAM (lines 5–21) grows the tree by iterating between a forward exploration (as in KA) and backward refinement...
Figure 3: Infinite-state system $S$ (top left) discussed in Sec. 4.3, its sound finite-state abstraction $\hat{S}$ (top right), part of its infinite-state knowledge abstraction $\hat{S}^k$ (bottom left) and its infinite bisimulation quotient $\hat{S}^k_B$ (bottom right). The single input $U = \{u\}$ is omitted and outputs $Y = \{A, \ldots, F\}$ are indicated next to the respective state. A state subset $\{a_i\}$ denotes the set $\{a_i \mid i \in \mathbb{N}\}$.

Figure 4: Exploration tree EXP of $S$ in Fig. 3 computed by Alg. 2 (left) and the abstract system $\hat{S}^k$ extracted after its 5th iteration (right). Nodes are labeled by $t_k$ (blue) for easier reference and the single input $u$ is omitted to avoid clutter. Diamond-enclosed numbers indicate the iteration in which this transition is explored. Dotted red arcs indicate cover block refinements in the iteration of the main while loop depicted by the red circled number and caused by the line of $\text{Refine}$ indicated on its top right. E.g., $X_B$ of $t_{21}$ is refined by re-calling $\text{Refine}$ in line 35 after $X_C$ of $t_{21}$ was refined in line 23 (as $t_1$ is a predecessor of $t_{21}$). The notation $35/31$ in $t_{22}$ indicates that its cover block $X_B$ is refined by line 31 after re-calling $\text{Refine}$ via line 35 on node $t_{22}$.

Having thus created all the children for a node $(v, q, c)$, if $c$ is a proper subset of $q$, the next step in KAM is to check if $q$, the current guess for the observation equivalence class for $c$, needs to be refined. Refinement is performed by the function $\text{Refine}$ (Alg. 2, line 15) and works similarly to the bisimulation algorithm.

In contrast to the usual bisimulation algorithm, $\text{Refine}(\langle v, q, c \rangle)$ only splits a block $q$ based on its possible successors in the tree if this split respects $c$, thereby avoiding the splitting of indistinguishable states, which caused the non-termination issue discussed in Sec. 4.2. One can intuitively think of $s \subseteq X$ computed in line 27 of Alg. 2 as the set of all states which are equivalent to $c$ in terms of their one-step observable future. However, in contrast to the bisimulation algorithm, KAM only adds $s$ to Cover but does not add its complement $q \setminus s$ (see line 29). This is due to the fact that this operation might not respect the currently available cells and again split indistinguishable states. If $q \setminus s$ is indeed needed, it will be discovered by another call to $\text{Refine}$.

Summarizing the above description, we see that $\text{Refine}$ refines the Cover set based on the one-step future of the computed cell.
Algorithm 2 KAM: Knowledge Abstraction and Minimization

Require: \( S = (X, X_0, U, F, Y, H) \)

1. \( \text{Cover} \leftarrow \{ q \in 2^X \setminus \{0\} \mid \exists y \in Y . q = H^{-1}(y) \} \);
2. \( \text{EXP}_T \leftarrow \emptyset \);
3. \( \text{EXP}_X \leftarrow \{(H(c), q, c) \mid q \in \text{Cover} \land c = q \land X_0 \neq \emptyset \} \);
4. \( \text{EXP}_T \leftarrow \emptyset \);
5. while \( \text{EXP}_X \neq \{(q, c) \mid \exists \nu . (v, q, c) \in \text{EXP}_X \} \) do
6. \( \text{EXP}_T \leftarrow \{(q, c) \mid \exists \nu . (v, q, c) \in \text{EXP}_X \} \);
7. for \( (v, q, c) \in \text{EXP}_X \) s.t. \(|\nu|\) is maximal do
8. for \( u \in U, y \in Y \) do
9. \( v' = vu y \);
10. \( c' = F(c, u) \cap H^{-1}(y) \neq \emptyset \);
11. \( Q' = \{q' \in \text{Cover} \mid c' \subseteq q' \land q' \text{ is minimal}\} \);
12. \( \text{EXP}_X \leftarrow \text{EXP}_X \cup \{(v', q', c') \mid q' \in Q'\} \);
13. \( \text{EXP}_T \leftarrow \text{EXP}_T \cup \{(v, q, c, u, (v', q', c')) \mid q' \in Q'\} \);
end for
15. if \( c \subseteq q \) then \( \text{Refine}((v, q, c)) \);
16. end if
17. end for
18. \( \tilde{S} \leftarrow \text{EXTRACT}(\text{EXP}_X, \text{EXP}_T) \);
19. if \( \text{TermCond()} == \text{true} \) then return \( \tilde{S} \);
20. end if
21. end while
22. return \( \tilde{S} \);
23. function \( \text{Refine}((v, q, c)) \)
24. for \( u \in U \) do
25. \( \text{Post}_Q u \leftarrow \{q' \in \text{Cover} \mid \{(v, q, c, u, (\cdot, \cdot, \cdot)) \in \text{EXP}_T \} \};
26. end for
27. \( s \leftarrow \{x \in q \mid \forall u \in U . F(x, u) \subseteq \text{Post}_Q u \};
28. if \( s \subseteq q \) then
29. \( \text{Cover} \leftarrow \text{Cover} \cup \{s\} \);
30. for all \( (v', q', c') \in \text{EXP}_X \) s.t. \( q = \emptyset \) do
31. if \( c \subseteq q \) then change \( (\tilde{v}, q, c) \) to \( (\tilde{v}, s, \tilde{c}) \) in \( \text{EXP}_T, X, F \);
32. end if
33. end for
34. for all \( (\tilde{v'}, q', \tilde{c'}) \), \( (\tilde{v}, q, c) \in \text{EXP}_T \) s.t. \( \tilde{q} = s \land \tilde{c}' \subset q' \) do
35. \( \text{Refine}((\tilde{v'}, q', \tilde{c'}) \);\)
36. end for
37. end if
38. end function
39. function \( \text{EXTRACT}(\text{EXP}_X, \text{EXP}_T) \)
40. \( \tilde{X} \leftarrow \{q \in 2^X \mid (\cdot, \cdot, \cdot) \in \text{EXP}_X \} \);
41. \( \tilde{X}_0 \leftarrow \{x_0 \cap H^{-1}(y) \in 2^X \setminus \{0\} \mid y \in Y \} \);
42. \( \tilde{F} \leftarrow \{(x, u, q, c) \mid (\cdot, \cdot, \cdot) \in \text{EXP}_T \} \);
43. \( \tilde{H}(\tilde{x}) = y \) if \( y \in \tilde{H}(\tilde{x}) \);
44. return \( \tilde{S} = (\tilde{X}, \tilde{X}_0, U, \tilde{F}, \tilde{Y}, \tilde{H}) \);
45. end function

Given this refinement, all previously obtained relations between cells and blocks need to be re-evaluated as \( s \subset \mathcal{Q} \) implies that \( s \) is now the minimal cover of \( c \), if \( c \) was previously related to \( q \) in \( \text{EXP}_X \) (see line 31). Thus, KAM updates its guess on the set of states possibly external language equivalent to a state in \( c \). This, however, might imply new block splits in cell/block pairs reaching \( c \), which have been checked for refinement in previous iterations of the algorithm. This is taken care of by the recursive call to \( \text{Refine} \) in line 35. Note that the recursion always moves up to the parent in the tree, and thus it eventually terminates. One can show that after the recursive call to \( \text{Refine} \) terminates, we always have a single minimal cover box \( q \) for every cell \( c \) computed so far. That is, given the relation \( \bar{a}(c) = \{q \in \text{Cover} \mid (c, q) \in \text{EXP}_X \} \) for \( \text{EXP}_X := \{(q, c) \mid \exists \nu . (v, q, c) \in \text{EXP}_X \} \), we have \( |\bar{a}(c)| = 1 \) (see [14, Lem. A.2] for a formal proof).

Example 4.5. For the example in Sec. 4.3, we see that for the tuple \( t_0 \) we have \( c = q = X_A = \{a_1\} \), hence, \( \text{Refine} \) is not called in the first iteration of KAM. In its second iteration, it computes the leaves \( t_{21} \) and \( t_{22} \) in the main while loop and then checks the parent node \( t_1 \) for refinement. For this, it computes all cover cells reachable by \( b_1 \) (which is Post\( Q = \bigcup \{X_B, X_C\} \)) and then computes all states in \( q = X_B \) with the same reachable cover blocks (which is \( s = X_B \)). As \( q = s \), no split occurs and a new iteration of the main while loop starts. After the computation of the leaves \( t_{31} \rightarrow t_{35} \) KAM checks the parent node \( t_2 \) for refinement. Here we obtain Post\( Q = X_B \) and \( s = X_C \). As \( s \subseteq q = X_C \), the cell \( X_C \) \( X_B \) of \( X_C \) is added to Cover. As there is no other node in the tree with a cell component contained in \( X_C \), we only update the block component of \( t_{21} \) (indicated by the red dotted arrow pointing to it in Fig. 4) and schedule all its predecessors for refinement. Therefore, node \( t_1 \) is checked for refinement again. Given the new cover cell \( X_C \), we now obtain \( \text{Post}_Q = \bigcup \{X_B, X_C\} \) and \( s = X_C \). This updates the cover element of \( t_1 \) and \( t_{33} \). This schedules only \( t_{22} \) for refinement, as \( t_0 \) does not fulfill the condition that \( c \subseteq q \). Now checking \( t_{22} \) for refinement still gives \( \text{Post}_Q = \bigcup \{X_B, X_C\} \) as we have not yet added the cover element \( X_C \). This is due to the fact that we do not know whether this element is indeed needed and respects the constructed state subsets. We therefore leave node \( t_{22} \) unchanged and proceed to the forth iteration of the main while loop. This computes the leaves \( t_{41} \rightarrow t_{45} \). During this computation we now have the new cover cell \( X_C \) available and KAM uses this smaller cover cell to correctly tack the equivalence class for \( t_{44} \) (indicated in green in Fig. 4). Now the only interesting refinement check is on \( t_{32} \) which discovers the new cover cell \( X_{C_{even}} \) and induces the further refinement of node \( t_{22} \) introducing the cover cell \( X_{B_{even}} \). This updates \( t_{22} \) and \( t_{45} \). Due to space constraints, we do not depict the constructed tree further. It should be noted that \( t_{43} \) clusters \( d_1 \) and \( d_2 \), as these states are not distinguishable based on the past observations. Therefore, calling \( \text{Refine} \) on \( t_{43} \) in the next iteration of KAM will not refine the equivalence class \( X_B \). Post\( Q = \bigcup \{X_F, X_G\} \) and we therefore obtain \( s = X_B \). The same happens for nodes \( d_1 \) and \( d_2 \). This prevents the non-termination issue of the bisimulation algorithm for this example.

After exploration and refinement, KAM extracts an abstraction \( \tilde{S} \) via the function \( \text{EXTRACT} \) in line 19. Intuitively, \( \text{EXTRACT} \) projects the tree in \( \text{EXP}_T \) to the blocks in the current Cover set which are reachable. It thereby "forgets" the forward-computed cells. For the example in Sec. 4.3 the abstraction extracted after the fifth iteration of KAM is depicted in Fig. 4 (right). It can be observed that Fig. 4 (right) coincides with the abstraction \( \tilde{S} \) in Fig. 3 (top right) up to a renaming of states.
Termination. Intuitively, KAM should terminate if Cover stabilizes. Then, all distinguishable subsets which are observation-equivalent have been discovered, and hence, imply Ext(S) = Ext( ̃S). That is, we would ideally like to have TermCond() == true in line 19 iff Cover has stabilized. Unfortunately, even if we observe that Cover has not changed in the current iteration, we do not know if it will never change again. This is because KAM bases its search for cover splits on the already constructed state-subsets. There might be a very long input/output event sequence which only causes a subset split after a long exploration phase. As the state space of S is infinite, we cannot check if this will ever happen. Interestingly, this is also true for fully initialized systems (i.e., where $X = X_0$). Thus, this termination check is undecidable.

One interesting special case where termination is decidable occurs if the KA algorithm (Alg. 1) terminates (which is for example always the case if $X$ is finite). In this case, we can show that $\text{EXP}_1 = \text{EXP}_{x,1}$ holds in the $l$-th iteration of Alg. 2 iff $\Gamma = ̃X$ holds in the $l$-th iteration of Alg. 1 (see [14, Lem. A.3] for a formal proof of this statement). While Cover might have stabilized earlier, we know it has surely stabilized by then.

Finite-State Abstractions. The termination condition discussed above aims at computing a sound finite-state realization of the external behavioral closure of $S$ which might not exist. Indeed, for arbitrary non-linear dynamical systems there rarely ever exists an exact finite-state realization in this sense, even if their input and output sets are finite. Therefore, as the name suggests, abstraction-based controller synthesis is usually only aiming at computing a finite-state abstraction which is accurate enough to synthesize an abstract controller for the given specification.

In this context, it is interesting to investigate whether the system $\tilde{S}^\text{S}$ computed in line 18 of Alg. 2 after running the while loop in line 5–21 finitely often, is indeed a sound abstraction of $S$ in the sense of Def. 3.1 and therefore allows for abstraction based control in the sense of Cor. 3.4. Interestingly, this is only true if KAM has already explored all possible output events which are reachable in $S$ at least once when terminated. This is for example trivially satisfied if $X_0 = X$. Additionally, whenever Cover stabilizes after a finite number of iterations, KAM indeed computes a sound realization of $S$. This is formalized in the following theorem.

**Theorem 4.6.** Let $S$ be a system, $S^*$ its external trace system and $\tilde{S}^\text{S}$ an abstract system extracted in line 18 of KAM(S) in some iteration. Further, let $Y^\prime = \{y \in Y \mid \exists q, c \in \text{EXP}_1 \cdot ̃H(q) = y\}$ and Reach($Y^\prime$) = $\{y \in Y \mid \exists p \in \text{EPref}(S), y = \text{Last}(p)\}$. If $Y^\prime = \text{Reach}(Y^\prime)$ it holds that $S^* \leq_{a} \tilde{S}^\text{S}$ with $a = \text{LastX}_{S^*}$. Further, if Cover has stabilized, we additionally have $S^* \leq_{a} \tilde{S}^\text{S}$.

In order to prove Thm. 4.6, we first prove Prop. 4.7 below which formalizes the intuition that, under the given premises, the cell/block pairs $(q, c) \in \text{EXP}_1$ available when extracting $\tilde{S}^\text{S}$ in line 18 of Alg. 2 actually induce a sound abstraction relation between $\tilde{S}^\text{S}$ and $S^*$. I.e., we always have $S^\text{S} \leq_{a} \tilde{S}^\text{S}$ for

$$\tilde{a}(c) := \{q \in \tilde{X}^\prime \mid \langle q, c \rangle \in \text{EXP}_1\}.\tag{1}$$

Further, Prop. 4.7 shows that $S^\text{S} \leq_{a} \tilde{S}^\text{S}$ if $\tilde{S}^\text{S}$ is finite-state (and thereby Cover has stabilized). With this result Thm. 4.6 becomes a simple corollary of Prop. 4.7 and Prop. 4.1 by utilizing the compositionality of sound abstractions (see [14, Prop. A.1]).

**Proposition 4.7.** *Given the premises of Thm. 4.6, it holds that $S^\text{S} \leq_{a} \tilde{S}^\text{S}$ with $a$ as in (1). Further, if Cover has stabilized, we additionally have $S^\text{S} \leq_{a} \tilde{S}^\text{S}$.*

**Proof.** To simplify notation we use $\tilde{S} := S^\text{S}$ and $\tilde{S}$ := $\tilde{S}^\text{S}$.

- **We first show that equality holds for (A1) and (A3) from Def. 3.1.**
  - (A1): Observe that line 1 in Alg. 1 and line 41 in Alg. 2 literally match. Further, for all $x \in X_0$ we have that $(c, x, x)$ is in the initial cover set (line 1 in Alg. 2) and thereby $(\tilde{x}, \tilde{x}) \in \text{EXP}_1$, as we have assumed $X_0$ to respect $H$. As Alg. 2 always maintains $x \subseteq \tilde{x}$ for any $(\tilde{x}, \tilde{x}) \in \text{EXP}_1$ and all elements in Cover only get refined, we see that there is no other $x^\prime \in \tilde{x}$ related to $x \in \tilde{x}$. We therefore have $\tilde{a}(X_0) = \tilde{X}_0$. (A3): It is easy to see that for all $x \in \tilde{X}$ holds that $x, x^\prime \in \tilde{x}$ implies $H(x) = H(x^\prime) = \tilde{H}(x)$. As $x \subseteq \tilde{x}$ for all $(\tilde{x}, \tilde{x}) \in \text{EXP}_1$, we have $\tilde{H}(x) = \tilde{H}(\tilde{x})$ for all related states.

- **Now we show that (A2) holds with equality for all $(\tilde{x}, x) \in \text{EXP}_1$ (possibly a subset of $\text{EXP}_1$).** For this, observe that $S$ is extracted in the last iteration of the while loop in line 5–21 of Alg. 2 and therefore the recursive function Refine was applied to all $(\tilde{x}, x) \in \text{EXP}_1$ with $\tilde{x} \subset x$ and has terminated. We can therefore utilize [14, Lem. A.2] implying $|\tilde{a}(x)| = 1$ for all $x$ present in $\text{EXP}_1$.

  - (A2) for $\text{EXP}_1$: Pick $\tilde{x} \in X$, $u \in U$ and $x_0 = F(\tilde{x}, u) \cap H^{-1}(y)$. Further, define $Y^\prime = \{y \in Y \mid \tilde{x} \neq y \}$ and let $Q^\prime$ contain all $\tilde{x}^\prime \in x$ s.t. $(\tilde{x}^\prime, y)_{\tilde{x}^\prime} \in \text{EXP}_1$ and $y \in Y^\prime$. Using the same argument as in the proof of Prop. 4.1 we have $F(\tilde{x}, u) = \bigcup_{y \in Y} \{x_y\}_{\tilde{x}^\prime}$. Further, we extract $\tilde{S}$ after all covers have been refined. With this we know that $F(\tilde{x}, u) = \text{Post}(Q^\prime(\tilde{x}, x))$, as otherwise there would exists a refinement $s \subset \tilde{x}$ in the sense of line 27 in Alg. 2. This further implies that for all $(\tilde{x}, x), (\tilde{x}, x)_{\tilde{x}} \in \text{EXP}_1$ we have that $\text{Post}(Q^\prime((\tilde{x}, x)), (\tilde{x}, x)) = \text{Post}(Q^\prime((\tilde{x}, x)_{\tilde{x}}), (\tilde{x}, x))$. With this it follows that $Q' = F(\tilde{x}, u)$. This implies $\tilde{a}(F(\tilde{x}, u)) = \tilde{a}(F(\tilde{x}, u)) = \tilde{a}(\tilde{x}, u)$.

- **It remains to show that (A2) holds with equality for a stable cover and with inclusion for an unstable one** for tuples $(\tilde{x}, x) \in \text{EXP}_1 \setminus \text{EXP}_1$. First, one can verify that $(\tilde{x}, x) \in \text{EXP}_1 \setminus \text{EXP}_1$ if a tuple $(\tilde{x}, x, x) \in \text{EXP}_1$ in the last iteration of the while loop before extracting $\tilde{S}^\text{S}$, and (b) if there exists no tuple $(\tilde{x}, x) \in \text{EXP}_1$ for an arbitrary $x$. While (a) is obvious, we show that (b) also holds. It follows from [14, Lem. A.2], that after completing every iteration of the while-loop in line 21 it holds for every $\tilde{x}$ already constructed, that there exists a unique $x^\prime$ s.t. $(\tilde{x}^\prime, x) \in \text{EXP}_1$. Thus, assume that $(\tilde{x}, x, x) \in \text{EXP}_1 \setminus \text{EXP}_1$. Further, a stable cover implies that there already exists another tuple $(\tilde{x}, x, x) \in \text{EXP}_1$ for which all outgoing transitions are contained in $\text{EXP}_1$. With this, we use the same reasoning as for $\text{EXP}_1$ to construct $Q'$ and to show that (A2) holds with equality.

- (A2) for $\text{EXP}_1 \setminus \text{EXP}_1$ with unstable Cover: If the Cover is not
stable, we cannot ensure that \( \bar{x} \) is stable for any \( (\bar{x}, \bar{x}') \in \text{EXP}^1 \setminus \text{EXP}_1 \), i.e., would not be refined in the next iteration of the whole loop. Further, we have to make sure that there exists another tuple \( (\bar{x}, \bar{x}') \in \text{EXP}_1 \). Now recall that we initialize Cover with the largest subsets \( \mathcal{X}_0 \subseteq X \) that generate the same output \( y \). As \( y^* = \text{Reach}(Y) \), we know that all initial cover cells \( \mathcal{X}_0 \) with \( y \in \text{Reach}(Y) \) will be explored (and possibly refined) at least once in Alg. 2. As \( \bar{x} \in \text{Cover} \) and by construction \( \bar{x} \subseteq \mathcal{X}_0 \), for \( y = \text{Reach}(Y) \) we know that \( (\bar{x}, \bar{x}') \in \text{EXP}_1 \). With this we can use the same reasoning as in the proof of (A2) for \( \text{EXP}_1 \) to construct \( Q' \). If it is stable, the argument reduces to the previous one. If it is not, we have \( \bar{f}(\bar{x}, u) \in \text{Post}Q_y((\bar{x}, \bar{x}')) \). With this, the same arguments as in the proof of (A2) for \( \text{EXP}_1 \) show that (A2) holds with inclusion, i.e., \( \bar{a}(\bar{f}(\bar{x}, u)) \subseteq \bar{F}(\bar{a}(\bar{x}), u) \) where \( \bar{a}(\bar{x}) \) contains all minimal \( \bar{x} \)'s covering \( \bar{x} \).

**Proof of Thm. 4.6.** As sound abstractions compose in the expected way (see [14, Prop. A.1]), we obtain a chain of sound abstractions \( S^* \preceq_{\text{LastX}_K} S^2 \preceq_{\alpha} S^4 \) from Prop. 4.7 and Prop. 4.1, implying \( S^* \preceq_{\alpha} S \) with \( a \alpha \) \( \text{LastX}_K \). It can be further observed from the tree-structure generated by KAM that every external prefix \( v \) of \( S \) corresponds to a unique tuple \( (q, c) \in \text{EXP}^1 \). Further, the same external prefix \( v \) reaches the state \( c \) of \( S^2 \) and the state \( q \) of \( S^4 \). As Prop. 4.7 shows that these states \( c \) and \( q \) are related via \( \alpha \), we have \( \text{LastX}_K = a \alpha \text{LastX}_K \). With this, the first claim of Thm. 4.6 follows. The second claim follows similarly. □

**Iterative ABCD with KAM.** By combining Cor. 3.4 and Thm. 4.6 we can compute an output-feedback controller \( C \vdash \bar{C} \circ \text{LastX}_K \in \mathcal{W}(S, \psi) \) from an abstract state-feedback controller \( \bar{C} \in \mathcal{W}(S^2, \psi) \) whenever the latter synthesis problem allows for such a solution, i.e., \( \mathcal{W}(S^2, \psi) \neq \emptyset \). Hence, ABCD with output feedback is sound in this case. Given that \( S^2 \) is in general only known to abstract \( S \), we are however losing completeness. That is, if \( \mathcal{W}(S^2, \psi) = \emptyset \), it does not imply that there is no solution to the original synthesis problem \( (S, \psi) \).

We can however take an eager abstraction-refinement approach instead to retain relative completeness. That is, whenever \( \mathcal{W}(S^2, \psi) = \emptyset \), we run KAM for some more steps, extract a new abstract \( S^2 \), and again try to synthesize a controller. We give up, once an upper bound \( L \) on the iterations of KAM is reached. This eager approach relies on the insight that abstractions extracted after more iterations of KAM refine earlier abstractions (see Thm. 4.8). Further, this abstraction-refinement procedure is relative complete.

That is, if there is a topologically closed finite-state abstraction \( S \) for which \( \mathcal{W}(S, \psi) \neq \emptyset \), there always exists a large enough \( L \) s.t. the abstraction \( S^2 \) extracted from KAM in the \( L \)'s iteration allows to solve the controller synthesis problem, i.e., \( \mathcal{W}(S^2, \psi) \neq \emptyset \).

**Theorem 4.8.** Given the premises of Thm. 4.6, let \( S^1 \) be the system computed in line 18 of Alg. 2 after one more iteration of Alg. 2 after \( S^2 \) was extracted. Then \( S^2 \preceq S^1 \).

**Proof.** Let \( \text{EXP}_1, \text{EXP}^1 \) and \( \text{EXP}^2, \text{EXP}^2_1 \) be the sets computed when extracting \( S^2 \) and \( S^1 \), respectively. Further let us define an abstraction map candidate \( a_2 \) using three cases. I.e., \( q \in a_1(p) \) if there exists \( c \) s.t. either (a) \( (q, c) \in \text{EXP}_1 \) and \( p \neq q \), or (b) \( (q, c) \in \text{EXP}^1 \setminus \text{EXP}^1_1 \), and there exists \( c' \) s.t. \( q \) is related to \( p \) in (a) or (b).

This definition induces the following three cases for the proof.

- (a) holds for \( (q, p) \): This implies \( (q, c) \in \text{EXP}^1_1 \). It follows from the same arguments as used in the proof of Prop. 4.7 that equality holds for (A1)-(A4) in Def. 3.1 w.r.t. \( S^1 \) both for \( S^1 \) and \( S^2 \). As \( a_1 \) reduces to the identity map in this case, the claim trivially follows.

- (b) holds for \( (q, p) \): Then it follows again that equality holds for (A1)-(A4) in Def. 3.1 w.r.t. \( S^1 \) but it follows from Thm. 4.6 that only inclusion holds for (A3) w.r.t. \( S^3 \). Formally, we fix \( c \) existentially quantified in the definition of case (b) before. Then we have \( a_1(\bar{f}(c, u)) = \bar{f}(a_1(c), u) \) where \( a_1(c) \) contains the unique minimal \( p \) covering \( c \) and \( a_1(\bar{f}(c, u)) \subseteq \bar{F}(a_1(c), u) \) where \( a_1(c) \) contains all minimal \( q \)’s covering \( c \). We have \( p \subseteq q \) for all \( q \in a_1(c) \) due to the additional refinement step run before extracting \( S^1 \). In particular, we have \( a(c) = a_1(p) \). Hence, \( a_1(\bar{f}(c, u)) = \bar{f}(a_1(c), u) \) and \( a_1(\bar{f}(c, u)) \subseteq \bar{F}(a_1(c), u) \). Now define \( C = \bar{F}(c, u) \). If for all \( c \in C' \) case (a) or (b) holds, we have that \( a_1(c) \) maps to a unique \( p' \). In this case it holds that \( a_1(\bar{F}(c, u)) = \bar{F}(a_1(c), u) \) and therefore \( a_1(\bar{F}(c, u)) \subseteq \bar{F}(a_1(p), u) \), what proves the statement. Now for any \( c' \) for which case (c) applies there exists a \( c'' \) s.t. case (a) or (b) applies while \( a(c'') = a(c'') \) and \( a_1(c') = a_1(c'') \). With this, the previous argument applies and the claim follows.

- (c) holds for \( (q, p) \): Fix \( c \) existentially quantified in the definition of (c) and recall that there exists \( c' \) s.t. \( a(c) = a(c') \) and \( a_1(c) = a_1(c') \) and case (a) or (b) (or both) applies for \( c' \). Hence, without loss of generality we can replace \( c \) by \( c' \) and the claim follows. □

**Remark 3.** The idea of abstraction-refinement for controller synthesis is also often applied in the context of l-complete abstractions [18, 21, 24, 30]. Similar to KAM, l-complete abstractions are constructed forward and generalize from initial observations to equivalence classes. Here, the equivalence classes collect states which share the same l-long external history. l-complete abstractions are typically constructed from the external behavior of \( S \) and do not assume the state dynamics of \( S \) to be known. They thereby do not utilize the memory structure implicitly given by the state dynamics of \( S \) in their generalization step. Therefore, KAM generates tighter abstractions whenever the underlying state transition system is known, but l-complete abstractions are to be preferred if this is not the case.

**Symbolic Implementations.** KAM differs from the simultaneous reachability and bisimulation minimization algorithm of Lee and Yannakakis [12] as it constructs an external language- (not bisimulation-) equivalent system. Hence, it only applies post computations and intersection with outputs, but does not take set differences. This is in fact crucial in implementations. For example, for affine systems with polyhedral initial sets and output sets, one can implement the algorithm exactly using a convex polyhedral abstract domain, as both post computations and intersections maintain convexity while set differences do not.

**5 HYBRID SYSTEM EXAMPLES**

We now present two continuous-state discrete-time hybrid system examples and show how our approach can be used to design abstractions useful for output-feedback control. Along the way, we also compare our approach with several alternatives and show...
Figure 5: Graphical representation of $\Sigma_1$ (far left) and $\Sigma_2$ (far right), showing the state space $X$ with the partition induced by the output maps $H_1$ and $H_2$, respectively. For $\Sigma_1$, $X_j = F(X_i, u_i)$ (dashed blue) indicates the reachable set of $X_i = H^{-1}_1(y_{00})$ (solid blue). Intersecting $X_j$ with the partition generates transitions (blue) originating in $y_{00}$ in the finite-state abstraction (middle). Similarly, $X_j = F(X_i, u_2)$ (dashed red) is reached from $X_i = H^{-1}_2(y_{02})$ (solid red) generating transitions (red) originating in $y_{02}$ in the abstraction.

In order to apply this process, we need to select a grid size $\eta$ when constructing $\Sigma_1$. We denote the resulting abstraction with $\Sigma_1^{(\eta)}$. We can start with $\eta = 1$ as discussed before. This, however, induces non-determinism and it can be easily seen by inspecting Fig. 5 (middle), that there does not exist a controller in the abstraction that allows us to surely transition from $y_{00}$ to $y_{22}$ and back infinitely often—in the abstraction, applying the necessary input sequence might lead to visiting $y_{02}$ instead of $y_{22}$. One can try a finer grid size, e.g., $\eta = 0.03$, but the problem still does not admit a solution for $\Sigma_1^{(0.03)}$. By inspection, the problem only has a solution if $\eta$ is chosen such that 0.2 is an integer multiple of $\eta$. Here, 0.2 is the greatest common divisor of 0.4 (the increments the dynamics make) and 1 (the "fidelity" of the outputs). So, the set of grid sizes that gives a solution is a measure-zero set in $R_{>0}$ and, in general, the "right" grid size is dictated by the dynamics and output map. Further, even if we use an automatic refinement tool like Mascot, the step size of the refinement of $\eta$ is a design parameter and thus, the tool may not ever explore an integer multiple of 0.2.

We now turn to solving the output-feedback control problem $(\Sigma_1, \psi_1)$ by directly applying the algorithms discussed in Sec. 4 to $\Sigma_1$ without constructing $\Sigma_1$ first. For this example, all three algorithms (i.e., KA, KA with bisimulation quotient, and KAM) will produce the same abstraction. This is due to the fact that the dynamics of the system are such that the post and the pre operations over $F$ cancel out. Therefore the forward and backward algorithms are essentially performing the same operations. Further, all of them terminate and generate a sound realization. Thus, these algorithms automatically figure out that the largest cover of $X$ which merges states with the same future under any applied input sequence has size $\eta = 0.2$.

Example 5.2. We consider another switched system $\Sigma_2$ with the same dynamics as $\Sigma_1$ but with changed output space $Y_2 = \{y_{00}, \ldots, y_{22}, y_{22a}, y_{22b}\}$ s.t. $H_2$ maps the upper left and lower right triangle of $y_{22}$ to $y_{22a}$ and $y_{22b}$, respectively (see Fig. 5 (right) for an illustration). The specification $\psi_2$ requires to repeatedly visit $y_{00}$ and either $y_{22a}$ or $y_{22b}$ infinitely often after starting in $y_{00}$.

Consider running KAM on $\Sigma_2$. First observe that we are now initializing KAM with the triangle shape domains of $H(y_{22a})$ and $H(y_{22b})$ in addition to the the boxed domains for all remaining outputs. This will result in little triangles right above and right below the diagonal of $y_{23}$, which collect reachable state subsets with the same output. However, in the remaining part of the state space, KAM will converge to the same rectangular grid as it does for $\Sigma_1$. The intuitive reason for this is that the post of any set $H^{-1}(y)$ with $y \notin \{y_{22a}, y_{22b}\}$ remains a box. Therefore, we can never distinguish whether we observe $y_{22a}$ or $y_{22b}$ if we transition to a box on the diagonal of $y_{23}$, no matter how fine we grid. Further, the post of any such box will be either $y_{22a} \cup y_{22b}$ again, $\{y_{00}\}$ (for $u = u_1$) or $y_{22b}$ (for $u = u_2$). With this it is easy to see that boxes of size $\eta = 0.2$ are again the largest partition of $X$ that form equivalence classes respecting observable subsets. KAM will therefore compute the same sound realization for $\Sigma_2$ as for $\Sigma_1$. If we however run KA (with or without the bisimulation quotient) one would additionally chop every box of size $\eta = 0.2$ into an upper left and lower right...
triangle. This unnecessary doubles the state space of the abstraction, but still resulting in a sound realization.

Let us now consider computing an abstraction $\Sigma^2_2$ by forward simulation of $\Sigma_2$ first, using SCOTS. Then we immediately get into trouble, because we cannot find a rectangular grid that respects the output map, as needed to fulfill (A3) in Def. 3.1. This approach would therefore directly fail in this example.

Finally, consider a system $\Sigma_3$ which has an unbounded state space $X_3 = \mathbb{R}^2$ with transition function defined by $F$ of $\Sigma_1$ but without the wrapping of its input argument. The output set $Y_3$ and the output function $H_3$ of $\Sigma_3$ are given by tiling the entire $\mathbb{R}^2$ space irregularly with the 3x3 blocks of observations $Y_1$ and $Y_2$ along with their respective output maps $H_1$ and $H_2$. We still have a finite set of inputs and outputs. By recalling that KAM produces the same sound realization for $\Sigma_1$ and $\Sigma_2$, we can use the same arguments as in the example of Sec. 4.3 to see that KAM will generate the same sound realization for $\Sigma_3$ as for $\Sigma_1$ and $\Sigma_2$, while all other algorithms will produce infinite-state abstractions. Admittedly, while the example distinguishes KAM from the other algorithms, it is not clear how to symbolically represent the algorithm in this case.

ACKNOWLEDGMENTS

We thank Manuel Mazo Jr. for useful comments on a draft of this paper. Majumdar and Schmuck were funded in part by the DFG project 389792660-TRR 248 and by the ERC under the Grant Agreement 610150. Ozay was supported in part by ONR grant N00014-18-1-2501, NSF grant ECCS-1553873, and an Early Career Faculty grant from NASA’s Space Technology Research Grants Program.

REFERENCES


