Data-driven and Model-based Verification via Bayesian Identification and Reachability Analysis

Sofie Haesaert [a], Paul M.J. Van den Hof [a], Alessandro Abate [b]

[a] *Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*

[b] *Department of Computer Science, University of Oxford, Oxford, United Kingdom*

Abstract

This work develops a measurement-driven and model-based formal verification approach, applicable to dynamical systems with partly unknown dynamics. We provide a new principled method, grounded on Bayesian inference and on reachability analysis respectively, to compute the confidence that a physical system driven by external inputs and accessed under noisy measurements verifies a given property expressed as a temporal logic formula. A case study discusses the bounded- and unbounded-time safety verification of a partly unknown system, encompassed within a class of linear, time-invariant dynamical models with inputs and output measurements.

Key words: Temporal logic properties, Bayesian inference, Linear time-invariant models, Model-based verification, Reachability analysis, Data-driven validation, Statistical model checking

1 Introduction

The design of complex, high-tech, safety-critical systems such as autonomous vehicles, intelligent robots, and cyber-physical infrastructures, demands guarantees on their correct and reliable behaviour. Correct functioning and reliability over models of systems can be attained by the use of formal methods. Within the computer sciences, the formal verification of software and hardware has successfully led to industrially relevant and impactful applications [14]. Carrying the promise of a decrease in design faults and implementation errors and of correct-by-design synthesis, the use of formal methods, such as model checking [14], has become a standard in the avionics, automotive, and railway industries [37]. Life sciences [6,15] and robotic applications [5,11] have also recently benefited by the application of these successful techniques from the computer sciences: this has required a shift from finite-state to physical and cyber-physical models, which are of practical use in nowadays science and technology [26,35].

The strength of formal techniques, such as model checking, is bound to the fundamental requirement of having access to a given model, obtained from the knowledge of the behaviour of the underlying system of interest. In practice, for most physical systems the dynamical behaviour is known only in part: this holds in particular with biological systems [1] or with classes of engineered systems where, as a consequence, the use of uncertain control models built from data is a common practice [25]. As an example consider a battery cell to be placed in a car, of which we have only a partial model but know the demand limits that will be raised while in operation. Before installing the battery we can probe and measure its dynamics, and wish to verify that the battery will never heat up excessively under the demanded operational limits.

Only limited work within the formal methods community deals with the verification of models with partly unknown dynamics. Classical results [4,22] consider verification problems for non-stochastic models described by differential equations with bounded parametric uncertainty. Similarly, but for continuous-time *probabilistic* models, [9,10] explore the parameter space with the objective of model verification (respectively statistical or probabilistic). Whenever full state measurements of the system are available, Statistical Model Checking (SMC) [34,27] replaces numerical model-based procedures with empirical testing of formalised properties. SMC is limited to fully observable stochastic systems with little or

Email addresses: `s.haesaert@tue.nl` (Sofie Haesaert), `P.M.J.Vandenhof@tue.nl` (Paul M.J. Van den Hof), `alessandro.abate@cs.ox.ac.uk` (Alessandro Abate).

no non-determinism, and may require the gathering a large set of measurements. Extensions towards the inclusion of non-determinism have been studied in [21,28], with preliminary steps towards Markov decision processes. Related to SMC techniques, but bound to finite state models, [13,30,33] assume that the system is encompassed by a finite-state Markov chain and efficiently use data to learn the corresponding model and to verify it. Similarly, [3,8] employ machine learning techniques to infer finite-state Markov models from data over given logical formulae.

An alternative approach, allowing both partly unknown dynamics over uncountable (continuous) variables and noisy output measurements, is the usage of a Bayesian framework relating the confidence in a formal property to the uncertainty of a model built from data. When applied on nonlinearly parameterised, linear time invariant (LTI) models this approach introduces heavy computational issues, which can only be mitigated via statistical methods [18]. Instead, in order to obtain reliable and numerical solutions, we propose the use of linearly parameterised model sets defined through orthonormal basis functions to represent these partially unknown systems. This is a broadly used framework in system identification [24,25]: while maintaining the beneficial computational aspects of linear parameterisations, the choice of orthonormal basis functions allows for the incorporation of prior knowledge on the system behaviour. Practically, this has been widely used for the modelling of physical systems, such as the thermal dynamics of buildings [38].

This work investigates the verification of temporal logic properties over partially unknown systems, using both prior modelling knowledge and data drawn from the system in a Bayesian setting. Building on [19,20], we provide a complete framework and newly extend the modelling class in [19] to multi-input multi-output models. The focus of this work is further set apart from [20], which explored the design of experiments to ameliorate the data-driven verification procedure.

2 General Framework and Problem Statement

In this section we overview a new methodology to assess the confidence in whether a system $\mathbf{S}$ satisfies a given specification ψ, formulated in a suitable temporal logic, by integrating the partial knowledge of the system dynamics with data obtained from a measurement setup around the system.

Let us further clarify this framework. Let us denote with $\mathbf{S}$ a physical system, or equivalently its associated dynamical behaviour. A signal input $u(t) \in \mathbb{U}, t \in \mathbb{N}$, captures how the environment acts on the system. Similarly, an output signal $y_0(t) \in \mathbb{Y}$ indicates how the system interacts with the environment, or alternatively how the system can be measured. Note that the input and output signals are assumed to take values over continuous domains. The system dynamics can be described via mathematical models, which quantify the behavioural relation between its inputs and outputs. The knowledge of the behaviour of the system is often limited or uncertain, making it impossible to analyse its dynamics by means of a "true" model. In this case, a-priori available knowledge allows to construct a model set $\mathcal{G}$ with elements $\mathbf{M} \in \mathcal{G}$: this model class encompasses the uncertainty on the underlying system by means of a parameterisation $\theta \in \Theta$, $\mathcal{G} = \{\mathbf{M}(\theta)|\theta \in \Theta\}$. The unknown "true" model $\mathbf{M}(\theta^0)$ representing $\mathbf{S}$, is assumed to be an element of $\mathcal{G}$, namely $\theta^0 \in \Theta$. Model sets $\mathcal{G}$ obtained through first principles and with unknown parameters adhere to this standard setup.

Samples can be drawn from the underlying physical system via a measurement setup, as depicted in Figure 1. An experiment consists of a finite number (N_s) of input-output samples drawn from the system, and is denoted by $Z^{N_s} = \{u(t)_{ex}, \tilde{y}(t)_{ex}\}_{t=1}^{N_s}$, where $u(t)_{ex} \in \mathbb{U}$ (in general a continuous domain) is the input for the experiment and $\tilde{y}(t)_{ex}$ is a (possibly noisy) measurement of $y_0(t)_{ex}$. In general, the measurement noise can enter non-additively and be a realisation of a stationary stochastic process. [1] We assume that at the beginning of the measurement procedure (say at $t = 0$), the initial condition of the system, encompassed by the initial state $x(0)_{ex}$ of models in $\mathcal{G}$, is either known, or, when not known, has a structured uncertainty distribution that is based on the knowledge of past inputs and/or outputs. As reasonable, we implicitly consider only well-defined problems, such that for any model $\mathbf{M}(\theta)$ representing the system, given an input signal $u(t)_{ex}$ and an (uncertainty distribution for) $x(0)_{ex}$, the probability density distribution of the measured signal can be fully characterised.

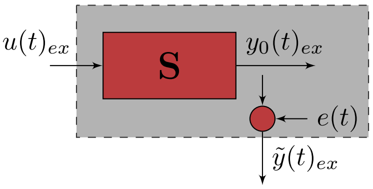


Fig. 1. System (smaller red box) and measurement setup (grey box). In the measurement setup the output $\tilde{y}(t)_{ex}$ includes the system output $y_0(t)_{ex}$ and the measurement noise $e(t)$. Data collected from experiments comprises the input $u(t)_{ex}$ and the measured output $\tilde{y}(t)_{ex}$ signals.

The end objective is to analyse the behaviour of system $\mathbf{S}$. We consider properties encoded as specifications ψ and expressed in a temporal logic of choice (to be detailed shortly). Let us remark that the behaviour of $\mathbf{S}$ to

[1] Notice that the operating conditions of the experiment, that is the input signal $u(t)_{ex}$, the initial state $x(0)_{ex}$, and the measurements $\tilde{y}(t)_{ex}$, have been indexed with "ex" to distinguish them from the conditions of interest for verification ("ver"), to be discussed shortly.

be analysed is bound to a set of operating conditions that are pertinent to the verification problem and that will be indexed by "*ver*": this comprises the set of possible input signals $u(t)_{ver}$ (e.g., a white or coloured noise signal, or a non-deterministic signal $u(t)_{ver} \in \mathbb{U}_{ver} \subseteq \mathbb{U}$), and of the set of initial states $x(0)_{ver} \in \mathbb{X}_{ver}$ for the mathematical models $\mathbf{M}(\theta)$ reflecting past inputs and/or outputs of the system. The system satisfies a property if the "true" model representing the system satisfies the property, namely $\mathbf{S} \vDash \psi$ if and only if $\mathbf{M}(\theta^0) \vDash \psi$.

In this work we consider the satisfaction of a property $\mathbf{M}(\theta) \vDash \psi$ as a *binary-valued mapping* from the parameter space Θ. More generally, when in addition to the measurements of the system also its internal transitions are disturbed by stochastic noise (known as process noise), then property satisfaction is a mapping from the parameter space Θ to the interval $[0, 1]$, and quantifies the probability that the model $\mathbf{M}(\theta)$ satisfies the property. This mapping generalises the definition of the satisfaction function discussed in [9], and is now stated as follows.

Definition 1 (Satisfaction Function) *Let $\mathcal{G}$ be a set of models $\mathbf{M}$ that is indexed by a parameter $\theta \in \Theta$, and let ψ be a formula in a suitable temporal logic. The satisfaction function $f_\psi : \Theta \to [0, 1]$ associated with ψ is*

$$f_\psi(\theta) = \mathbf{P}\left(\mathbf{M}(\theta) \vDash \psi\right). \qquad (1)$$

Let us assume that the satisfaction function f_ψ is measurable and entails a decidable verification problem (e.g., a model checking procedure) for all $\theta \in \Theta$ and properties ψ of interest. In this work we consider the verification of partly unknown physical systems with respect to a subset of linear time temporal logic properties. We are in a position to state the following.

Problem 1 *For a partly unknown physical system $\mathbf{S}$, under prior knowledge on the system given as a parameterised model class $\mathcal{G}$ supporting an uncertainty distribution over the parameterisation, gather possibly noisy data drawn from the measurement setup and verify properties on $\mathbf{S}$ expressed in a temporal logic of choice, with a formal quantification of the confidence of the assertion.*

2.1 A Bayesian Framework for Data-driven Modelling and Verification

Consider Problem 1. Denote with $\mathbf{P}(\cdot)$ and $p(\cdot)$ respectively a probability measure and a probability density function, both defined over a continuous domain. We employ Bayesian probability calculus [29] to express the confidence in a property as a measure of the uncertainty distribution defined over the set $\mathcal{G}$. By adopting the Bayesian framework, uncertainty distributions are handled as probability distributions of random variables. Therefore the confidence in a property is computed as a probability measure $\mathbf{P}(\cdot)$ via the densities $p(\cdot)$ over the uncertain variables.

Proposition 1 (Bayesian Confidence) *Given a specification ψ and a data set Z^{N_s}, the confidence that $\mathbf{S} \vDash \psi$ can be quantified via inference as*

$$\mathbf{P}\left(\mathbf{S} \vDash \psi \mid Z^{N_s}\right) = \int_\Theta f_\psi(\theta) p\left(\theta | Z^{N_s}\right) d\theta, \qquad (2)$$

where f_ψ is the satisfaction function given in (1). *The* a-posteriori *uncertainty distribution $p\left(\theta | Z^{N_s}\right)$, given the data set Z^{N_s}, is based on parametric inference over θ as*

$$p\left(\theta | Z^{N_s}\right) = \frac{p\left(Z^{N_s}|\theta\right)p(\theta)}{\int_\Theta p(Z^{N_s}|\theta)p(\theta)d\theta}\,, \qquad (3)$$

which assumes the knowledge of an uncertainty distribution $p(\theta)$ over the parameter set Θ, representing prior knowledge.

The statement can be formally derived based on standard Bayesian calculus, as in [29]. We have chosen to employ a Bayesian framework, as per (3), since it allows to reason explicitly over the uncertain knowledge on the system and to work with the data acquired from the measurement setup. This leads to the efficient incorporation of the available knowledge and to its combination with the data acquisition procedure, in order to compute the confidence on the validity of a given specification over the underlying system. As a special instance, this result can be employed for Bayesian hypothesis testing [39]. As long as the mapping f_ψ is measurable, the models in the model set (and hence the system represented by it) can be characterised by either probabilistic or non-probabilistic dynamics.

Remark 1 *In statistical model checking [27,34], the objective is to replace the computationally tolling verification of a system over bounded-time properties by the empirical (statistical) testing of the relevant specifications over finite executions drawn from the system. In contrast, our setup tackles the problem of efficiently incorporating data with prior knowledge, for the formal (deductive) verification of the behaviour of a system with partly unknown dynamics. As such our overall verification approach is, as claimed, both data-driven and model-based. Moreover, by separating the operational conditions of an experiment from those of importance for the verification procedure, the system can be verified over non-deterministic quantities, encompassing both controller and disturbance inputs, as well as modelling errors.*

2.2 Existing Computational Approaches

In the literature the satisfaction function is related to the exploration of a parameter set over the validity of

a formal property $f_\psi(\theta)$, and has been studied for autonomous models in continuous time in [4,16,22].
Bayesian inference is widely applicable to different types of properties and models, however its computational complexity might in practice limit its implementation. Analytical solutions to the inference equation (3) can be found if the prior is a conjugate distribution. For linear dynamical systems, closed-form solutions are given inter alia in [31].
In general (2)-(3) in Proposition 1 lack analytical solutions, and the assessment of the satisfaction function (1) may be computationally intensive. Statistical methods such as the one proposed in [18] on a similar Bayesian approach lead to involved computations and introduce additional uncertainty from Monte Carlo techniques.

In contrast with the reviewed literature, in the next section we propose a novel computational approach over discrete-time linear time-invariant systems. By exploiting linear parameterisations, analytical solutions of both the parameter inference and the satisfaction function are characterised, over properties expressed within a fragment of a temporal logic.

3 LTL Verification of LTI Systems

Consider a system $\mathbf{S}$ that can be represented by a class of finite-dimensional dynamical models that evolve in discrete-time, and are linear time-invariant (LTI). We focus the study to non probabilistic dynamics. These models depend on input and output signals ranging over $\mathbb{R}^m$ and $\mathbb{R}^p$, respectively, and on variables $x_{\mathbf{S}}(t)$ taking values in an Euclidean space, $x_{\mathbf{S}}(t) \in \mathbb{X} \subseteq \mathbb{R}^n$, where n, the state dimension, is the model order. The behaviour of such a system is encompassed by state-space models $(A_{\mathbf{S}}, B_{\mathbf{S}}, C_{\mathbf{S}}, D_{\mathbf{S}})$ as

$$\mathbf{S}: \quad \begin{cases} x_{\mathbf{S}}(t+1) &= A_{\mathbf{S}} x_{\mathbf{S}}(t) + B_{\mathbf{S}} u(t), \\ y_0(t) &= C_{\mathbf{S}} x_{\mathbf{S}}(t) + D_{\mathbf{S}} u(t), \end{cases} \tag{4}$$

where matrices $A_{\mathbf{S}}, B_{\mathbf{S}}, C_{\mathbf{S}}, D_{\mathbf{S}}$ are of appropriate dimensions. The experimental measurement setup, as depicted in Figure 1, consists of the signals $u(t)_{ex}$ and $\tilde{y}(t)_{ex} = y_0(t)_{ex} + e(t)$, representing the inputs and the measured outputs, respectively, and where $e(t)$ is an additive zero-mean, white, Gaussian-distributed measurement noise with covariance Σ_e that is uncorrelated from the inputs. N_s samples are collected within a data set $Z^{N_s} = \{u(t)_{ex}, \tilde{y}(t)_{ex}\}_{t=1}^{N_s}$.

3.1 Formalisation of Properties

System properties are expressed, over a finite set of atomic propositions $p_i \in AP$, $i = 1, \dots, |AP|$, in Linear Temporal Logic [2]. Any LTL formula ψ is built up recursively via the syntax

$$\psi ::= \text{true} \mid p \mid \neg\psi \mid \psi \wedge \psi \mid \bigcirc\psi \mid \psi \,\mathsf{U}\, \psi.$$

Let $\pi = \pi(0), \pi(1), \pi(2), \ldots \in \Sigma^{\mathbb{N}^+}$ be a string composed of letters from the alphabet $\Sigma = 2^{AP}$, and let $\pi_t = \pi(t), \pi(t+1), \pi(t+2), \ldots$ be a subsequence (postfix) of π. The satisfaction relation between π and ψ is denoted as $\pi \vDash \psi$ (or equivalently $\pi_0 \vDash \psi$). The semantics of the satisfaction relation are defined recursively over π_t and the syntax of the LTL formula ψ as follows:

$$\begin{array}{rll} \text{(true)} & \pi_t \vDash \text{true} & \Leftrightarrow \text{true} \\ \text{(atomic prop.)} & \pi_t \vDash p & \Leftrightarrow p \in \pi(t) \\ \text{(negation)} & \pi_t \vDash \neg\psi & \Leftrightarrow \pi_t \nvDash \psi \\ \text{(conjunction)} & \pi_t \vDash \psi_1 \wedge \psi_2 & \Leftrightarrow \pi_t \vDash \psi_1 \text{ and } \pi_t \vDash \psi_2 \\ \text{(next)} & \pi_t \vDash \bigcirc\psi & \Leftrightarrow \pi_{t+1} \vDash \psi \\ \text{(until)} & \pi_t \vDash \psi_1 \,\mathsf{U}\, \psi_2 & \Leftrightarrow \exists\, i \in \mathbb{N} : \pi_{t+i} \vDash \psi_2, \\ & & \qquad \text{and } \forall j \in \mathbb{N}: \\ & & \qquad 0 \le j < i, \pi_{t+j} \vDash \psi_1 \end{array}$$

This syntax allows to extend the study to more complex propositional formulae (such as disjunction or implication). Denote the k-bounded and unbounded invariance (or safety) operator as $\Box^k\psi = \bigwedge_{i=0}^{k} \bigcirc^i \psi$ and $\Box\psi = \neg(\texttt{true}\,\mathsf{U}\,\neg\psi)$, respectively.

It is of interest to refer formal properties expressed as LTL formulae to the input-output behaviour of a dynamical model, over a given time horizon $t \geq 0$. The output $y_0(t)_{ver} \in \mathbb{Y}$ is labeled by a map $L : \mathbb{Y} \rightarrow \Sigma$, which assigns symbols α in the alphabet Σ of the formulae discussed previously to half spaces on the output, as

$$L(y_0(t)_{ver}) = \alpha \in \Sigma \;\Leftrightarrow\; \textstyle\bigwedge_{p_i \in \alpha} A_{p_i} y_0(t)_{ver} \leq b_{p_i}, \tag{5}$$

for given $A_{p_i} \in \mathbb{R}^{1 \times p}$, $b_{p_i} \in \mathbb{R}$. In other words, sets of atomic propositions in AP are associated to polyhedra over $\mathbb{Y} \subset \mathbb{R}^p$. Let us underline that properties are defined over the behaviour $y_0(t)_{ver}$ of the model, and not over the noisy measurements $\tilde{y}(t)_{ex}$ of the model considered within the measurement setup. Additionally, for the verification problem the input signal is modelled as a bounded signal $u(t) \in \mathbb{U}_{ver}$, and represents external non-determinism from the environment acting on the system.

3.2 Model Set Selection

As a first step we need to embed the available a-priori knowledge on the underlying system within a parameterised model set. Note that although the goal of parameter exploration in formal verification has recently attracted quite some attention [4,16,22], there are as of yet no general scalable results for the computation of the satisfaction function for nonlinearly-parameterised, discrete-time LTI models. The use of linearly-parameterised model sets, especially those defined through orthonormal basis functions (as further elaborated next), has been widely used for the modelling of physical systems, such as the thermal dynamics of buildings [32,38].

Whilst in general the uncertainty about a model representing a linear time-invariant system does not map onto a linearly-parameterised model set, we argue that a linearly-parameterised model set can encompass a relevant class of models. For instance, any asymptotically stable LTI model can be represented uniquely by its (infinite) impulse response [23], and the coefficients of the impulse response define a linear parameterisation for this model. Further, for asymptotically stable systems, the coefficients of the impulse response converge to zero, so that a truncated set of impulse coefficients provides a good approximate LTI model set with a finite-dimensional, linear parameterisation. These impulse responses define a finite set of orthonormal basis functions [24, Chapters 4 and 7],[36] and construct a valid model set for a physical system solely based on knowledge of asymptotic stability. Alternative choices for an orthonormal basis such as Laguerre functions and Kautz functions [24], can incorporate additional and more extensive prior knowledge of the physical system.

We conclude that, as an alternative to the use of a nonlinearly parameterised set of models, structural information (even when not exact) can be used to select a set of orthonormal basis functions, whose finite truncation defines a finite-dimensional linearly-parameterised model set indexed over the coefficients of the basis functions. Thus, in the following we consider a linearly parameterised model set $\mathcal{G}$ that encapsulates system $\mathbf{S}$, and specifically $\mathcal{G} = \{(A, B, C(\theta), D(\theta)), \theta \in \Theta\}$.

A system satisfies a property if, assuming it can be equivalently represented by a mathematical model $\mathbf{M}(\theta^0)$, all the words generated by the model satisfy that property. Since properties are encoded over the external (input-output) behaviour of the system $\mathbf{S}$, which is the behaviour of $\mathbf{M}(\theta^0)$ (where in our case $\theta^0 \in \Theta$), we may equivalently assert that any property ψ is verified by the system, $\mathbf{S} \vDash \psi$, if and only if it is verified by the unknown model representing the system, namely $\mathbf{M}(\theta^0) \vDash \psi$. Within the modelling perspective offered in this work, let us introduce Θ_ψ to be the feasible set of parameters, such that for every parameter in that set the property ψ holds, i.e., $\forall \theta \in \Theta_\psi : \mathbf{M}(\theta) \vDash \psi$. More precisely, Θ_ψ is characterised as the level set of the satisfaction function f_ψ, $\Theta_\psi = \{\theta \in \Theta : f_\psi(\theta) = 1\}$. The quantification of Θ_ψ is of key importance in our work.

3.3 Safety Verification of Bounded-time Properties

Models $\mathbf{M}$ in the class $\mathcal{G}$ have the following representation $(A, B, C(\theta), 0)$:

$$\mathbf{M}(\theta): \quad \begin{cases} x(t+1) &= Ax(t) + Bu(t), \\ \hat{y}(t,\theta) &= C(\theta)x(t), \end{cases} \tag{6}$$

and are parameterised by $\theta \in \Theta \subset \mathbb{R}^{pn}$, $\theta = \text{vec}(C)$ and $C(\theta) \in \mathbb{R}^{p \times n}$. We assume a prior probability distribution $p(\theta)$, which structures the knowledge of the uncertainty in θ. In addition to this *strictly proper* model class, we will also allow for a *proper* model class $(A, B, C(\theta), D(\theta))$, where both the C and the D-matrices are parameterised, so that $\theta = \text{vec}([C\ D])$). For a given initial condition $x(0)$ and input sequence, the output of the "true" model $\hat{y}(t, \theta^0)$ is equal to the system output $y_0(t)$.

Consider a measurement setup as in Figure 1, related to an unknown parameter θ^0. Signals $u(t)_{ex}$ and $\tilde{y}(t)_{ex}$ represent the input and the measured output, respectively, and $e(t)$ is an additive zero-mean, white, Gaussian-distributed measurement noise with covariance Σ_e that is uncorrelated from the input. From this setup N_s samples are collected in a data set $Z^{N_s} = \{u(t)_{ex}, \tilde{y}(t)_{ex}\}_{t=1}^{N_s}$. Given the operating conditions of the experiment setup, the measured signal $\tilde{y}(t)_{ex}$ can be fully characterised: its probability density, conditional on the parameters θ, is

$$\begin{aligned} p\left(Z^{N_s}|\theta\right) &= \prod_{t=1}^{N_s} p\left(\tilde{y}(t)_{ex}|\theta\right) \\ &= \frac{1}{\sqrt{|\Sigma_e|^{N_s}(2\pi)^{pN_s}}} \exp\Bigg[\\ &- \frac{1}{2}\sum_{t=1}^{N_s} (\hat{y}(t,\theta) - \tilde{y}(t)_{ex})^T \Sigma_e^{-1} (\hat{y}(t,\theta) - \tilde{y}(t)_{ex})\Bigg], \end{aligned}$$

and can be directly used in Proposition 1. This conditional density $p\left(Z^{N_s}|\theta\right)$ depends implicitly on the given initial state $x(0)_{ex}$ and, in the case of a given uncertainty distribution over $x(0)_{ex}$, $p\left(Z^{N_s}|\theta\right)$ should be marginalised over $x(0)_{ex}$ [31]. The a-posteriori uncertainty distribution is obtained as the analytical solution of the parametric inference in (3) [31].

Recall now that for a given specification ψ, we seek to determine a feasible set of parameters Θ_ψ, which is such that the corresponding models admit property ψ, namely $\mathbf{M}(\theta) \vDash \psi$, $\forall \theta \in \Theta_\psi$. Since models $\mathbf{M}(\theta)$ have a linearly-parameterised state space realisation as per (6), it follows that when the set of initial states $\mathbb{X}_{ver}$ and of inputs $\mathbb{U}_{ver}$ are bounded polyhedra, the verification of a class of safety properties expressed by formulae with labels as in (5) leads to a set of feasible parameters

Θ_ψ that is a polyhedron, which can be easily computed. More precisely, the following result can be derived.

Theorem 2 *Consider properties ψ composed within the LTL fragment $\psi ::= \alpha|\bigcirc\psi|\psi_1 \wedge \psi_2$, with $\alpha \in \Sigma$. Given a bounded polyhedral set (a polytope) of initial states $x(0) \in \mathbb{X}_{ver}$ and of inputs $u(t) \in \mathbb{U}_{ver}$ for $0 \leq t < \infty$, and considering a labelling map as in (5), then the feasible set Θ_ψ of the parameterised model set (6) is a polyhedron.*

Proof[of Theorem 2] Let $\otimes$ denote the Kronecker product. Consider the input set $\mathbb{U}_{ver}$ to be the convex hull of U, i.e. $\text{conv}(U) = \mathbb{U}_{ver}$. Similarly let the set of initial states be $\text{conv}(X_{ver}) = \mathbb{X}_{ver}$. Let the model set be given as $\mathbf{M}(\theta) = (A, B, C(\theta), D)$. We will temporarily assume that D is set to be equal to zero, and afterwards (cf. Point 3) we will show how to work with a parameterised D. As can be deduced from the *and* operations in (5), note that for simplicity the syntax fragment $\psi ::= \alpha|\bigcirc\psi|\psi_1 \wedge \psi_2$ with $\alpha \in \Sigma = 2^{AP}$ is equivalent to $\psi ::= p|\bigcirc\psi|\psi_1 \wedge \psi_2$ with $p \in AP$. We structure the proof in three parts.

1. We claim that for every specification ψ composed from the syntax fragment $\psi ::= p|\bigcirc\psi|\psi_1 \wedge \psi_2$ and $\theta \in \Theta$, the words generated by a model $\mathbf{M}(\theta) = (A, B, C(\theta), 0)$ with state $x(t)$ satisfy the specification ψ, denoted $< \mathbf{M}(\theta), x(t) > \vDash \psi$, if and only if

$$\left(\left(I_{n_\psi} \otimes x(t)\right)^T N_\psi + K_\psi \right) \theta \leq B_\psi. \tag{7}$$

The matrices $N_\psi \in \mathbb{R}^{nn_\psi \times np}, K_\psi \in \mathbb{R}^{n_\psi \times np}, B_\psi \in \mathbb{R}^{n_\psi}$ in the above satisfaction relation have dimensions that are functions of the parametrisation and of a property-dependent "dimension" n_ψ, which will be obtained inductively over the syntax of the specification. Next we focus on the study of fragments of the LTL syntax.

For any *atomic proposition* the model starting from state $x(t)$ satisfies a property p_i, i.e., $< \mathbf{M}(\theta), x(t) > \vDash p_i \Leftrightarrow A_{p_i} y \leq b_{p_i}$, with $A_{p_i} \in \mathbb{R}^{1\times p}$ and $b_{p_i} \in \mathbb{R}$. We construct the matrices N_{p_i}, K_{p_i} and B_{p_i} as follows. Consider $y(t)$ for a given $x(t)$ then

$$A_{p_i} y(t) = A_{p_i} C(\theta) x(t) = x(t)^T (I_n \otimes A_{p_i})\theta.$$

This yields $N_{p_i} = (I_n \otimes A_{p_i}) \in \mathbb{R}^{n\times np}$, $K_{p_i} = O_{1\times np} \in \mathbb{R}^{1\times np}$, and $B_{p_i} = b_{p_i} \in \mathbb{R}^{1\times 1}$.

The *next* operation $\bigcirc\psi_1$ with matrices $(N_{\psi_1}, K_{\psi_1}, B_{\psi_1})$ yields matrices

$$\begin{aligned} N_{\bigcirc\psi_1} &= \mathbf{1}_{|U|} \otimes \left(I_{n_{\psi_1}} \otimes A^T\right) N_{\psi_1}, \\ K_{\bigcirc\psi_1} &= \mathcal{U} \left(I_{n_{\psi_1}} \otimes B\right)^T N_{\psi_1} + \mathbf{1}_{|U|} \otimes K_{\psi_1}, \\ B_{\bigcirc\psi_1} &= \mathbf{1}_{|U|} \otimes B_{\psi_1}, \end{aligned}$$

where the i-th set of n_{ψ_1} rows of $\mathcal{U} \in \mathbb{R}^{|U|n_{\psi_1}\times m}$ is defined as

$$\left(I_{n_{\psi_1}} \otimes u_i^T\right) \text{ with } u_i \in U$$

and where $n_{\bigcirc\psi_1} = |U| n_{\psi_1}$. This can be derived as

$$\begin{aligned} &< \mathbf{M}(\theta), x(t) > \vDash \bigcirc\psi \Leftrightarrow \forall u(t) \in \mathbb{U}_{ver} : \\ &\quad \left(\left(I_{n_{\psi_1}} \otimes x(t+1)\right)^T N_{\psi_1} + K_{\psi_1} \right) \theta \leq B_{\psi_1}, \\ &\Leftrightarrow \forall u(t) \in \mathbb{U}_{ver} : \\ &\left(\left(I_{n_{\psi_1}} \otimes Ax(t)\right)^T N_{\psi_1} \right. \\ &\quad \left. + \left(I_{n_{\psi_1}} \otimes Bu(t)\right)^T N_{\psi_1} + K_{\psi_1} \right) \theta \leq B_{\psi_1}. \end{aligned}$$

Since the above is an affine function in $u(t)$, the image of every $u(t) \in \text{conv}(U) = \mathbb{U}_{ver}$ can be expressed as a convex combination of the values at the vertices $u_i \in U$, c.f. [6]. Thus an equivalent expression is

$$\begin{aligned} &\Leftrightarrow \forall u_i \in U : \left(\left(I_{n_{\psi_1}} \otimes Ax(t)\right)^T N_{\psi_1} \right. \\ &\left. + \left(I_{n_{\psi_1}} \otimes u_i\right)^T \left(I_{n_{\psi_1}} \otimes B\right)^T N_{\psi_1} + K_{\psi_1} \right) \theta \leq B_{\psi_1}, \end{aligned}$$

which can be rewritten as

$$\begin{aligned} &\Leftrightarrow \left(\mathbf{1}_{|U|} \otimes \left(I_{n_{\psi_1}} \otimes Ax(t)\right)^T N_{\psi_1} + \mathcal{U} \left(I_{n_{\psi_1}} \otimes B\right)^T N_{\psi_1} \right. \\ &\qquad \left. + \mathbf{1}_{|U|} \otimes K_{\psi_1} \right) \theta \leq \mathbf{1}_{|U|} \otimes B_{\psi_1}. \end{aligned}$$

Matrices $K_{\bigcirc\psi}$, and $B_{\bigcirc\psi}$ can be obtained directly. To obtain $N_{\bigcirc\psi}$ now rewrite the first term:

$$\begin{aligned} &\mathbf{1}_{|U|} \otimes \left(I_{n_{\psi_1}} \otimes x^T(t)\right) \left(I_{n_{\psi_1}} \otimes A^T\right) N_{\psi_1} \\ &\quad = \left(I_{|U|} \mathbf{1}_{|U|}\right) \otimes \left(I_{n_{\psi_1}} \otimes x^T(t)\right) \left(I_{n_{\psi_1}} \otimes A^T\right) N_{\psi_1} \\ &\quad = \left(I_{|U|n_{\psi_1}} \otimes x^T(t)\right) \left(\mathbf{1}_{|U|} \otimes \left(I_{n_{\psi_1}} \otimes A^T\right) N_{\psi_1}\right). \end{aligned}$$

The *and* operation $\psi_1 \wedge \psi_2$ for $(N_{\psi_1}, K_{\psi_1}, B_{\psi_1})$ and $(N_{\psi_2}, K_{\psi_2}, B_{\psi_2})$ with $n_{\psi_1\wedge\psi_2} = (n_{\psi_1} + n_{\psi_2})$ gives

$$N_{\psi_1\wedge\psi_2} = \begin{bmatrix} N_{\psi_1} \\ N_{\psi_2} \end{bmatrix}, K_{\psi_1\wedge\psi_2} = \begin{bmatrix} K_{\psi_1} \\ K_{\psi_2} \end{bmatrix}, \; B_{\psi_1\wedge\psi_2} = \begin{bmatrix} B_{\psi_1} \\ B_{\psi_2} \end{bmatrix}.$$

This can be derived from

$$\begin{aligned} &< \mathbf{M}(\theta), x(t) > \vDash \psi_1 \wedge \psi_2 \\ &\Leftrightarrow \bigwedge_{i\in\{1,2\}} \left(\left(I_{n_{\psi_i}} \otimes x(t)\right)^T N_{\psi_i} + K_{\psi_i} \right) \theta \leq B_{\psi_i} \\ &\Leftrightarrow \left(\left(I_{n_{\psi_1\wedge\psi_2}} \otimes x(t)\right)^T \begin{bmatrix} N_{\psi_1} \\ N_{\psi_2} \end{bmatrix} + \begin{bmatrix} K_{\psi_1} \\ K_{\psi_2} \end{bmatrix} \right) \theta \leq \begin{bmatrix} B_{\psi_1} \\ B_{\psi_2} \end{bmatrix}. \end{aligned}$$

2. The matrix-valued function

$$\left(\left(I_{n_\psi} \otimes x(0)\right)^T N_\psi + K_\psi\right)\theta$$

is affine in $x(0)$ (for a fixed θ), therefore its value at the initial condition $x(0) \in \mathbb{X}_{ver}$ is a convex combination of the function values at the vertices X_{ver} of $\mathbb{X}_{ver}$. Thus the satisfaction relation $< \mathbf{M}(\theta), x(0) > \models \psi$ represented by the multi-affine inequality holds uniformly over $x(0) \in \mathbb{X}_{ver}$ if and only if it holds for the vertices of $\mathbb{X}_{ver}$.
This gives a set of affine inequalities in θ, thus the feasible set Θ_ψ is a polyhedron and is given as

$$\left\{\theta \in \Theta : \bigwedge_{x_i \in X_{ver}} \left(\left(I_{n_\psi} \otimes x_i\right)^T N_\psi + K_\psi\right)\theta \leq B_\psi\right\}.$$

Let us remark that the set Θ_ψ is a polyhedron because it is formed by a finite set of half spaces.

3. To complete the proof of Theorem 2, we need to extend the results to models with parameterised D. The dynamics of model (A, B, C, D) with both C and D fully parameterised can be reformulated as

$$\begin{bmatrix} x(t+1) \\ u(t+1) \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} u(t+1)$$

$$y(t) = \begin{bmatrix} C & D \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}.$$

Using the new matrices $(\tilde{A}, \tilde{B}, \tilde{C}(\theta), 0)$ the obtained results still hold. For part **2.** set of vertices X_{ver} needs to be extended with the vertices of U as $X_{ver} \times U$. □

In the computation of the feasible set, the faces of the polyhedron Θ_ψ are shown to be a function of the vertices (recall that a polytope can be written as the convex hull of a *finite* set of vertices) of the bounded set of initial states $\mathbb{X}_{ver}$ and of the set of inputs $\mathbb{U}_{ver}$, and are also expected to grow in number as a function of the time horizon of the property.
The result in Theorem 2 is valid for any finite composition of the LTL fragment $\psi ::= \alpha | \bigcirc\psi | \psi_1 \wedge \psi_2$, as such it only holds for finite horizon properties. Properties defined over the infinite horizon will be the objective of Section 3.5.

Remark 2 *The feasible set Θ_ψ obtained in Theorem 2 is a Borel-measurable set as it defines (if not empty) a closed set in the parameter space.*

3.4 Case Study: Bounded-Time Safety Verification

3.4.1 Single-Input Single-Output System

Consider a system $\mathbf{S}$ and verify whether the output $y_0(t)_{ver}$ remains within the interval $\mathcal{I} = [-0.5,\ 0.5]$, labeled as ι, for the next 5 time steps, under $u(t)_{ver} \in \mathbb{U}_{ver} = [-0.2,\ 0.2]$ and $x(0)_{ver} \in \{0_2\} = \mathbb{X}_{ver}$. Introduce accordingly the alphabet $\Sigma = \{\iota, \tau\}$ and the labelling map $L : L(y) = \iota, \forall y \in \mathcal{I}$, $L(y) = \tau, \forall y \in \mathbb{Y} \backslash \mathcal{I}$. Now check whether the following LTL property holds: $\mathbf{S} \models \bigwedge_{i=1}^{5} (\bigcirc)^i \iota$.

We assume that system $\mathbf{S}$ can be represented as an element of a model set $\mathcal{G}$, with models expressed via transfer functions characterised by second-order Laguerre-basis functions [23] (a special case of orthonormal basis functions). This translates to the following parameterised state-space representation:

$$\begin{aligned} x(t+1) &= \begin{bmatrix} a & 0 \\ 1-a^2 & a \end{bmatrix} x(t) + \begin{bmatrix} \sqrt{1-a^2} \\ (-a)\sqrt{1-a^2} \end{bmatrix} u(t), \\ \hat{y}(t,\theta) &= \theta^T x(t)\ . \end{aligned} \quad (8)$$

The parameter set is chosen as $\theta \in \Theta = [-10, 10]^2$, whereas the coefficient a is chosen to be equal to 0.4. We select, as prior available knowledge on the system, a uniform distribution $p(\theta)$ on the model class, and pick a known variance $\sigma_e^2 = 0.5$ for the white additive noise on the measurement. The set of feasible parameters $\Theta_\psi \subset \Theta$ is represented in Figure 2 and is computed according to Theorem 2. Based on the prior available knowledge, the confidence associated to $\theta^0 \in \Theta_\psi$ amounts to 0.0165: this quantity is obtained by numerical computation of (2) with probability distribution [2] $p(\theta)$. Thereafter, we have set up an experiment on the system with "true parameter" $\theta^0 = [1\ 0]^T$ (Figure 2) and with input signal $u(t)_{ex}$, a realisation of a white noise with a uniform distribution over $[-0.2, 0.2]$, and measured $\tilde{y}(t)_{ex}$ for 200 consecutive time instances. In comparison to the confidence obtained with the prior $p(\theta)$, the uncertainty distribution is refined as $p\left(\theta | Z^{N_s}\right)$, and the resulting confidence in the property is increased to 0.779, as per (2). Along this line of experiments, we have repeated the test 100 times, for several instances of the parameter θ^0 characterising the underlying system $\mathbf{S}$. In all instances, after obtaining 200 measurements the a-posteriori probability is used to assess the confidence in the safety of the system, as displayed in Table 1 via mean and variance terms. Observe that θ^0 is just outside of Θ_ψ for values $[-1, -1]^T$ and $[1, 1]^T$. For $\theta^0 = [-1, 1]^T$ and $\theta^0 = [1, -1]^T$, the true parameter lies in the feasible set but very close to the edges: this is reflected in the results in Table 1. For

[2] Integrals are solved via the numerical integration tool in `Matlab`.

Table 1
Mean (μ) and variance (σ^2) of the confidence obtained from 100 experiments with 200 measurements each.

θ^0	μ	σ^2	θ^0	μ	σ^2
$\begin{bmatrix}-1 & -1\end{bmatrix}^T$	0.348	0.073	$\begin{bmatrix}1 & -1\end{bmatrix}^T$	0.491	0.085
$\begin{bmatrix}-1 & 0\end{bmatrix}^T$	0.705	0.060	$\begin{bmatrix}1 & 0\end{bmatrix}^T$	0.730	0.056
$\begin{bmatrix}-1 & 1\end{bmatrix}^T$	0.492	0.086	$\begin{bmatrix}1 & 1\end{bmatrix}^T$	0.339	0.065

the points clearly inside the feasible set the confidence generally becomes high with a low variance, whereas for the points closest to the edges ($\theta^0 = [-1, 1]^T$ and $\theta^0 = [1, -1]^T$) the variance is higher and the confidence has only increased up to around 0.49. In comparison, the points just outside the feasible edge give a lower confidence then the former two. Consider as an example the $[1, 1]^T$ case: for this the observed initial increase from 0.0165 (i.e., the a-priori confidence) to around 0.34 is expected and reflects the closeness of the parameter to the feasible set; additional measurements will steer the confidence down towards zero (parameter outside of the feasible set). In conclusion, the experiments show that the measurements can be used to quantify the confidence level.

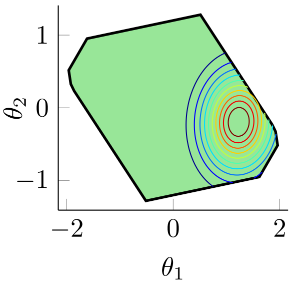


Fig. 2: Feasible set of parameters $\Theta_\psi \subset \Theta$, and contour lines of the posterior $p\left(\theta | Z^{N_s}\right)$, obtained for $\theta^0 = [1 \;\; 0]^T$ after 200 measurements.

3.4.2 *Multiple-Input/Output system: Feasible Set*

In order to showcase the workings of the feasible set computations for multiple-input multiple-output (MIMO) models, we consider a straightforward extension of (8), that is model

$$x(t+1) = \begin{bmatrix} a & 0 \\ 1-a^2 & a \end{bmatrix} x(t) + \begin{bmatrix} \sqrt{1-a^2} & 0 \\ 0 & -a\sqrt{1-a^2} \end{bmatrix} u(t),$$
$$\hat{y}(t,\theta) = \begin{bmatrix} \theta_1 & \theta_3 \\ \theta_2 & \theta_4 \end{bmatrix} x(t) \; . \qquad (9)$$

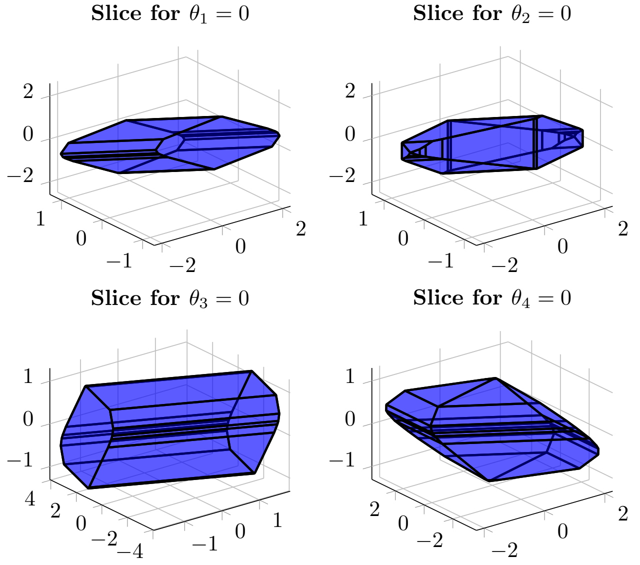
We verify whether the output $y_0(t)$ remains for 5 time steps within the polytope described by

$$\begin{bmatrix} -1 & 0 \\ -1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} y_0 \leq \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix},$$

with $x(0)_{\text{ver}} \in \{0_2\}$ and under inputs constrained within

$$\mathbb{U}_{\text{ver}} = \text{conv}\left(\left[\begin{smallmatrix} -0.2 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0.2 \\ 0.1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0.1 \\ -0.1 \end{smallmatrix}\right]\right).$$

Using the results in the previous subsection we compute the feasible set: in Figure 3 we display slices of the feasible set, where a slice is obtained by fixing one parameter θ_i to a value within the feasible set.


Slice for $\theta_1 = 0$ **Slice for $\theta_2 = 0$**

Slice for $\theta_3 = 0$ **Slice for $\theta_4 = 0$**

Fig. 3. Three-dimensional plots of the four-dim feasible set.

3.5 *Verifying Unbounded-Time Properties Using Invariant Sets*

In this section we extend the approach of Section 3.3, to hold on the LTL fragment $\psi ::= \alpha | \bigcirc \psi | \psi_1 \wedge \psi_2$ with additionally the *unbounded* invariance (safety) operator. The subsection is built up as follows:

- first we connect the notion of positive invariance with that of feasible set;
- then we discuss how to practically compute a feasible set for invariance properties, with the set of initial states limited to be the origin;
- this is then extended to computing feasible sets under initial states in a polytope that includes the origin;
- finally, we interpret these results and complete the section with results on the verification of unbounded-time properties.

Recall the form of the k-bounded and of the unbounded invariance operators, namely $\square^k \psi = \bigwedge_{i=0}^{k} \bigcirc^i \psi$ and $\square \psi = \neg(\texttt{true} \, \mathsf{U} \, \neg\psi)$, respectively. The extension from a k-bounded operator, covered by the result in Theorem 2, to the unbounded invariance one, is based on the concept of robust positive invariance [7, Def. 4.3], recalled next.

Definition 2 *For the system $x(t+1) = Ax(t)+Bu(t)$, the set $\mathcal{S} \subseteq \mathbb{X}$ is said to be robustly positively invariant if, for all $x(0) \in \mathcal{S}$ and $u(t) \in \mathbb{U}_{ver}$, the condition $x(t) \in \mathcal{S}$ holds for all $t \geq 0$.*

Recall that the feasible set Θ_ψ is defined as the set of parameters for which property ψ holds, namely $\forall\theta \in \Theta_\psi$: $\mathbf{M}(\theta) \vDash \psi$. The satisfaction relation $\mathbf{M}(\theta) \vDash \psi$ depends implicitly on the set of initial states $x(0) \in \mathbb{X}_{ver}$ and on the set of inputs $\mathbb{U}_{ver}$. Let us extend the definition of the feasible set to explicitly account for its dependence on the set of initial conditions: given a bounded and convex set $\mathcal{S} \subset \mathbb{X}$, let $\Theta_\psi(\mathcal{S})$ be defined as the set of parameters in Θ for which the parameterised models $\mathbf{M}(\theta)$ initialised with $x(0) \in \mathcal{S}$ satisfy ψ over input signals $u(t) \in \mathbb{U}_{ver}$ $t \geq 0$. Hence the feasible set Θ_ψ can be written as a function of the set of initial states $\mathbb{X}_{ver}$, that is $\Theta_\psi(\mathbb{X}_{ver})$. Thus the extended map $\Theta_\psi(\cdot)$ takes subsets of the state space into subsets of the parameter space. Note that if $\mathcal{S}$ is a robustly positively invariant set that includes the set of initial states $\mathbb{X}_{ver} \subseteq \mathcal{S}$, then for all $\theta \in \Theta_\psi(\mathcal{S})$ the models $\mathbf{M}(\theta)$ satisfy ψ over all infinite-time model traces $x(t)$: this allows to state that $\mathbf{M}(\theta) \vDash \Box\psi$. We can show that the following holds.

Lemma 3 *The function $\Theta_\psi(\cdot) : 2^{\mathbb{X}} \to 2^{\Theta}$, for specifications obtained as $\psi ::= \alpha \mid \bigcirc\psi \mid \psi_1 \wedge \psi_2$, is monotonically decreasing: that is if $\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \mathbb{X}$, then $\Theta_\psi(\mathcal{S}_2) \subseteq \Theta_\psi(\mathcal{S}_1)$.*

Proof[of Lemma 3] We leverage the notation used in the proof of Theorem 2. Provided that the parameterised model is given as $(A, B, C(\theta), 0)$, we show that any $\theta \in \Theta_\psi(\mathcal{S}_2)$ is also an element of $\theta \in \Theta_\psi(\mathcal{S}_1)$. Suppose $\mathcal{S}_2$ has a finite number of vertices $x_i \in \mathcal{V}(\mathcal{S}_2)$, then for any $\theta \in \Theta_\psi(\mathcal{S}_2)$:

$$\bigwedge\nolimits_{x_i\in\mathcal{V}(\mathcal{S}_2)} \left((I_{n_\psi} \otimes x_i)^T N_\psi + K_\psi\right)\theta \leq B_\psi,$$

and for every $x \in \mathcal{S}_2$

$$\left((I_{n_\psi} \otimes x)^T N_\psi + K_\psi\right)\theta \leq B_\psi.$$

Since the vertices $x_j \in \mathcal{V}(\mathcal{S}_1)$ are also elements of $\mathcal{S}_2$, then

$$\bigwedge\nolimits_{x_j\in\mathcal{V}(\mathcal{S}_1)} \left((I_{n_\psi} \otimes x_j)^T N_\psi + K_\psi\right)\theta \leq B_\psi$$

and $\theta \in \Theta_\psi(\mathcal{S}_1)$. This reasoning can be trivially extended to include models with parameterised D matrices. Increasing the number of vertices of $\mathcal{S}_1$ and $\mathcal{S}_2$, does not change the result, hence the same holds if $\mathcal{S}_1$ and $\mathcal{S}_2$ are convex sets. □

Based on the result in Lemma 3, we conclude that the maximal feasible set $\Theta_{\Box\psi}$ is obtained as a mapping from the minimal robustly positively invariant set $\mathcal{S}$ that includes $\mathbb{X}_{ver}$: $\Theta_{\Box\psi} = \Theta_\psi(\mathcal{S})$. This leads next to consider under which conditions such minimal robustly positively invariant set $\mathcal{S}$ can be exactly computed or approximated.

Feasible set for invariance properties with $\mathbb{X}_{ver} = \{0_n\}$

For $\mathbb{X}_{ver} = \{0_n\}$, assuming a bounded interval $\mathbb{U}_{ver}$ with the origin in its interior, and under some basic assumptions on the dynamics (to be shortly discussed), the minimal robustly positively invariant set can be shown to be a bounded and convex set that includes the origin [7]. Maintaining the condition of $\mathbb{U}_{ver}$ being bounded and having the origin in its interior, we first consider the case that $\mathbb{X}_{ver} = \{0_n\}$ and characterise $\mathcal{S}$ via tools available from set theory in systems and control; thereafter we look at extensions to more general sets of initial states $\mathbb{X}_{ver}$.

Assume that $\mathbb{U}_{ver}$ includes the origin, and denote the forward reachability mappings initialised with $\mathcal{R}^{(0)} := \{0_n\} \subset \mathbb{X}$ as

$$\mathcal{R}^{(i)} := \mathrm{Post}(\mathcal{R}^{(i-1)}), \tag{10}$$

with set operation $\mathrm{Post}(X) := \{x' = Ax + Bu, x \in X, u \in \mathbb{U}_{ver}\}$. Denote the limit reachable set as $\mathcal{R}^\infty = \lim_{i\to\infty} \mathcal{R}^{(i)}$. From literature we recall that properties of these i-step reachable sets, as given in [7] include the following: for a reachable pair (A, B) and an asymptotically stable matrix A, the ∞-reachable set $\mathcal{R}^\infty$ is bounded and convex [7, Proposition 6.9]. Specifically, the k-step reachable set converges to the ∞-reachable set via (10), since it is monotonically increasing $\mathcal{R}^{(i)} \subseteq \mathcal{R}^{(i+1)}$. Moreover, $\mathcal{R}^\infty$ is the minimal robustly positively invariant set for the system, so that any positively invariant set includes $\mathcal{R}^\infty$ [7, Proposition 6.13]. Thus, starting from $x(0) = 0_n$, all $x(t) \in \mathcal{R}^\infty$, and furthermore of interest to this work we conclude that $\Theta_{\Box^k\psi} = \Theta_\psi(\mathcal{R}^{(k)})$ and that $\Theta_{\Box\psi} = \Theta_\psi(\mathcal{R}^\infty)$.

Feasible set for invariance properties under polytopic sets of initial states

More generally, if $\mathbb{X}_{ver} \subseteq \mathcal{R}^\infty$ and under the same assumptions on matrices A, B and $0 \in \mathbb{U}_{ver}$, then $\mathcal{R}^\infty$ is the minimal robustly positively invariant set that includes $\mathbb{X}_{ver}$, and $\Theta_\psi(\mathcal{R}^\infty) = \Theta_{\Box\psi}$. For finite iterations the reachable sets $\mathcal{R}^{(i)}$ are polytopes, and if $\mathcal{R}^{(i)} = \mathcal{R}^{(i+1)}$, then $\mathcal{R}^{(i)} = \mathcal{R}^\infty$. Though the iterations can stop in finite time, in general the number of iterations to obtain $\mathcal{R}^\infty$ can be infinite. Whilst the minimal robustly positively invariant set is not necessarily closed or a polytope, there exist methods to approximate $\mathcal{R}^\infty$ as detailed in [7]. For instance, for stable systems, $\mathcal{R}^{(k)}$ is shown to converge to $\mathcal{R}^\infty$, in the sense that for all $\epsilon > 0$ there exists $\bar{k}$ such that for $k \geq \bar{k}$, $\mathcal{R}^{(k)} \subseteq \mathcal{R}^\infty \subseteq (1+\epsilon)\mathcal{R}^{(k)}$ [7, Proposition 6.9].

Recall that the maximal feasible set $\Theta_{\Box\psi}$ is obtained as a mapping from the minimal robustly positively invariant set $\mathcal{S}$ including $\mathbb{X}_{ver}$, so that $\Theta_{\Box\psi} = \Theta_\psi(\mathcal{S})$. Let us extend the study to the case where the conditions $\mathbb{X}_{ver} = \{0_n\}$ or its extension $\mathbb{X}_{ver} \subseteq \mathcal{R}^\infty$ do not apply, while the condition on the bounded set $\mathbb{U}_{ver}$ is maintained, that is $0 \in \mathbb{U}_{ver}$. Consider the more general case where the set of initial states is a polytope but not necessarily a subset of $\mathcal{R}^\infty$. Denote the union of the forward reachability mappings initialised with $\mathcal{R}^{(0)}_{\mathbb{X}_{ver}} := \mathbb{X}_{ver} \subseteq \mathbb{X}$ as

$$\mathcal{R}^{(i)}_{\mathbb{X}_{ver}} := \mathcal{R}^{(i-1)}_{\mathbb{X}_{ver}} \cup \mathrm{Post}(\mathcal{R}^{(i-1)}_{\mathbb{X}_{ver}}) \, . \qquad (11)$$

This set is also known in the literature as the *reach tube*. The corresponding set for infinite time is denoted as $\mathcal{R}^{\infty}_{\mathbb{X}_{ver}} = \lim_{i\to\infty} \mathcal{R}^{(i)}_{\mathbb{X}_{ver}}$. Notice that in the earlier case when $\mathbb{X}_{ver} \subseteq \mathcal{R}^\infty$, then $\mathcal{R}^\infty = \mathcal{R}^\infty_{\mathbb{X}_{ver}}$. The iteration is monotonically increasing $\mathcal{R}^{(i)}_{\mathbb{X}_{ver}} \subseteq \mathcal{R}^{(i+1)}_{\mathbb{X}_{ver}}$, and whenever $\mathcal{R}^{(i)}_{\mathbb{X}_{ver}} = \mathcal{R}^{(i+1)}_{\mathbb{X}_{ver}}$ it stops after a finite number of iterations with $\mathcal{R}^{\infty}_{\mathbb{X}_{ver}} = \mathcal{R}^{(i)}_{\mathbb{X}_{ver}}$. Of course, also in this more general case, the number of iterations can be unbounded, however the convergence properties of $\mathcal{R}^{(i)}$ extend directly to the case of sets $\mathcal{R}^{(i)}_{\mathbb{X}_{ver}}$. Since $\mathcal{R}^{(i)}_{\mathbb{X}_{ver}}$ is a union of polytopes, it is not guaranteed to be a convex set. Still, it can be shown via arguments as in the proof of Theorem 2 that the computation of the feasible set $\Theta_\psi(\mathcal{S})$ boils down to that of $\Theta_\psi\big(\mathrm{conv}(\mathcal{S})\big)$.

Remark 3 *Let us illustrate the convergence property for sets $\mathcal{R}^{(i)}_{\mathbb{X}_{ver}}$ as follows. For every vertex $x^i(0) \in \mathbb{X}_{ver}$, select a decomposition $x^i_r + x^i_s$ with $x^i_r \in \mathcal{R}^\infty$, which minimises $\|x^i_s\|$ for a chosen vector norm $\|\cdot\|$. Since every element $x(0) \in \mathbb{X}_{ver}$ is a convex combination of the vertices $x^i(0)$, it follows that for all $x(0) \in \mathbb{X}_{ver}$:*

$$x(0) = \sum_i a_i x^i(0) = \sum_i a_i x^i_r(0) + \sum_i a_i x^i_s(0)$$
$$\in \mathrm{conv}(x^i_r(0)) + \mathrm{conv}(x^i_s(0)) \subseteq \mathcal{R}^\infty + \bar{\mathbb{X}}_{ver},$$

with $\sum_i a_i = 1$ for $a_i \geq 0$, where $\bar{\mathbb{X}}_{ver} = \mathrm{conv}(x^i_s(0))$, and where we have employed the standard operation of set addition. We obtain that $\mathbb{X}_{ver} \subseteq \mathcal{R}^\infty + \bar{\mathbb{X}}_{ver}$, and that the minimal positively invariant set $\mathcal{R}^\infty_{\mathbb{X}_{ver}}$ can be bounded by $\mathcal{R}^\infty + \lim_{k\to\infty} \bigcup_{i=0}^k A^i \bar{\mathbb{X}}_{ver}$. Under the discussed conditions on $\mathbb{U}_{ver}$ and (A, B), previously necessary for $\mathcal{R}^\infty$ to be a bounded and convex polytope, $A^i\bar{\mathbb{X}}_{ver}$ will converge to $\{0_n\}$. Thus, the iteration $\mathcal{R}^{(k)}_{\mathbb{X}_{ver}}$ is monotonically increasing and bounded, hence it converges. If $\bar{\mathbb{X}}_{ver}$ includes the origin in its interior, then there exists a finite iteration step k, such that $\bigcup_{i=0}^k A^i\bar{\mathbb{X}}_{ver} = \bigcup_{i=0}^{k+1} A^i\bar{\mathbb{X}}_{ver}$. Moreover, for any reachable pair (A, B) and asymptotically stable A, the closure of the minimal robustly positively invariant set $\mathcal{R}^\infty_{\mathbb{X}_{ver}}$ includes the origin.

Robust approximations of the feasible set via $\Theta_\psi(\cdot)$

In order to exploit the convergence in the computation of the feasible set for invariance properties, we need to bound the error incurred in the use of approximations of the sets $\mathcal{R}^\infty_{\mathbb{X}_{ver}}$ or $\mathcal{R}^\infty$. Let $\mathcal{B}$ denote a unit ball centred at the origin and let the Hausdorff distance between sets $\mathcal{R}_1$ and $\mathcal{R}_2$ be defined as

$$\delta_H(\mathcal{R}_1, \mathcal{R}_2) = \inf\{\epsilon \geq 0 | \mathcal{R}_1 \subseteq \mathcal{R}_2 + \epsilon\mathcal{B}, \mathcal{R}_2 \subseteq \mathcal{R}_1 + \epsilon\mathcal{B}\}.$$

We can show that the following holds.

Lemma 4 *Let us consider a model set under a reachable pair (A, B), an asymptotically stable A, and let the input set $\mathbb{U}_{ver}$ include the origin. Consider a polytope $\mathcal{R}$, and a property ψ comprised of $\psi ::= \alpha | \bigcirc\psi | \psi_1 \wedge \psi_2$, with $\alpha \in \Sigma$, for which $\Theta_\psi(\mathcal{R})$ is a non-empty polytope with vertices v_i and the origin in its interior. Let A be bounded as $\|A\|_2 \leq 1$. Then for any $\epsilon_x \geq 0$,*

$$\Theta_\psi(\mathcal{R} + \epsilon_x\mathcal{B}) \subseteq \Theta_\psi(\mathcal{R}) \subseteq \Theta_\psi(\mathcal{R} + \epsilon_x\mathcal{B}) + \epsilon_\theta\mathcal{B} \qquad (12)$$

if $\epsilon_\theta \geq \dfrac{\epsilon_x\epsilon_p \max_i(\|v_i\|)^2}{1 + \epsilon_x\epsilon_p \max_i(\|v_i\|)}$, *for* $\epsilon_p := \max\limits_{p\in AP} \dfrac{\|A_p\|_2}{|b_p|}$.

Proof[of Lemma 4] **1.** $\Theta_\psi(\mathcal{R} + \epsilon_x\mathcal{B}) \subseteq \Theta_\psi(\mathcal{R})$
Based on the definition of this set (c.f. the proof of Theorem 2), the set operation $\Theta_\psi(\cdot)$ is monotonically decreasing as in Lemma 3. Therefore $\Theta_\psi(\mathcal{R} + \epsilon_x\mathcal{B}) \subseteq \Theta_\psi(\mathcal{R})$ holds.

2. $\Theta_\psi(\mathcal{R}) \subseteq \Theta_\psi(\mathcal{R} + \epsilon_x\mathcal{B}) + \epsilon_\theta\mathcal{B}$
Consider the case where the model is parameterised as $(A, B, C(\theta), 0)$. To prove (12), we first find an ϵ_θ as a function of ϵ_x such that

$$\Theta_\psi(\mathcal{R}) \subseteq \Theta_\psi(\mathcal{R} + \epsilon_x\mathcal{B}) + \epsilon_\theta\mathcal{B}. \qquad (13)$$

Let v_i be the vertices of the polytope $\Theta_\psi(\mathcal{R})$, i.e., $v_i \in \mathcal{V}(\Theta_\psi(\mathcal{R}))$ (as used in Lemma 3), then (13) holds if and only if $v_i \in \Theta_\psi(\mathcal{R} + \epsilon_x\mathcal{B}) + \epsilon_\theta\mathcal{B}$. Equivalently, this means that there exists an $r_\theta \in \epsilon_\theta\mathcal{B}$ such that $v_i - r_\theta \in \Theta_\psi(\mathcal{R} + \epsilon_x\mathcal{B})$. This is equivalent to demanding that for every $x_j \in \mathcal{V}(\mathcal{R})$, $v_i \in \mathcal{V}(\Theta_\psi(\mathcal{R}))$ and $r_x \in \epsilon_x\mathcal{B}$, there exists a vector $r_\theta \in \epsilon_\theta\mathcal{B}$:

$$\begin{aligned} &\big((I_{n_\psi} \otimes (x_j^T + r_x^T))N_\psi + K_\psi\big)(v_i - r_\theta) \leq B_\psi \\ \Leftrightarrow \quad &\big((I_{n_\psi} \otimes x_j^T)N_\psi + K_\psi\big)(v_i - r_\theta) \\ &\quad + \big((I_{n_\psi} \otimes r_x^T)N_\psi\big)(v_i - r_\theta) \leq B_\psi. \end{aligned}$$

Take $(v_i - r_\theta) = (1 - \alpha_i)v_i$ with $\alpha_i \in [0, 1)$, then

$$\begin{aligned} &\big((I_{n_\psi} \otimes x_j^T)N_\psi + K_\psi\big)(1 - \alpha_i)v_i \\ &\quad + \big((I_{n_\psi} \otimes r_x^T)N_\psi\big)(1 - \alpha_i)v_i \leq B_\psi \end{aligned}$$

holds if

$$(1-\alpha_i)(I_{n_\psi}\otimes r_x^T)N_\psi v_i \le \alpha_i B_\psi. \tag{14}$$

Separate the matrix N_ψ and B_ψ into its block matrices $N_\psi^j = [N_\psi]_{\{1+(j-1)n:nj\}\times\{1:n\}}$ and $B^j = [B_\psi]_j$, such that inequality (14) is equivalent to the set of inequalities

$$\begin{aligned}(1-\alpha_i)r_x^T N_\psi^j v_i' &\le \alpha_i B^j, \text{ for } j=1,\ldots,n_\psi\\ \Leftrightarrow \quad r_x^T N_\psi^j v_i' &\le \frac{\alpha_i}{(1-\alpha_i)}B^j \quad .\end{aligned}$$

Given that 0 is in the interior of $\Theta_\psi(\mathcal{R})$, it follows that $B_j > 0$ for $j=1,\ldots,n_\psi$

$$\max_j \left(r_x^T N_\psi^j v_i'\right)(B^j)^{-1} \le \frac{\alpha_i}{(1-\alpha_i)} \quad .$$

The term on the left can be upper bounded based on the Cauchy-Schwarz inequality

$$\begin{aligned}\max_j \left(r_x^T N_\psi^j v_i'\right)(B^j)^{-1} &\le \max_j \|(N_\psi^j)^T r_x\|_2 \|v_i'\|_2 (B^j)^{-1}\\ &\le \max_j \|(N_\psi^j)^T\|_2 \|r_x\|_2 \|v_i'\|_2 (B^j)^{-1} \text{ and } \|r_x\|_2 \le \epsilon_x\\ &\le \epsilon_x \epsilon_p \|v_i'\|_2.\end{aligned}$$

The last inequality follows from the introduction of the precision of the labelling, denoted as ϵ_p, and defined as

$$\epsilon_p = \max_{p\in AP} \frac{\|A_p\|_2}{|b_p|}. \tag{15}$$

Remember that $\|L\otimes K\|_2 = \|L\|_2\|K\|_2$. Then based on Theorem 2 and on the condition $\|A\|_2 \le 1$, it can be shown that

$$\max_j \|(N_\psi^j)^T\|_2 |B^j|^{-1} \le \max_{p\in AP}\frac{\|A_p\|_2}{|b_p|}.$$

Note that $\frac{\alpha_i}{(1-\alpha_i)}$ monotonically increases with α_i for $\alpha_i \in [0,1)$. Therefore a bound on α_i can be found as

$$\alpha_i = (\epsilon_x\epsilon_p\|v_i\|)/(1+\epsilon_x\epsilon_p\|v_i\|) \text{ for } j=1,\ldots,n_\psi. \tag{16}$$

It follows that (13) holds if

$$\epsilon_\theta \ge \max(\|v_i\|_2)\frac{\epsilon_x\epsilon_p\max(\|v_i\|_2)}{1+\epsilon_x\epsilon_p\max(\|v_i\|_2)}. \tag{17}$$

For the case that the model is parameterised in both C and D, i.e., $(A,B,C(\theta),D(\theta))$ the derivation is a bit more cumbersome (cf. proof of Theorem 2), but can be repeated with no change to the end result. □

Let us briefly discuss the conditions under which Lemma 4 is applicable. The requirement that $\Theta_\psi(\mathcal{R})$ is not empty is raised to avoid the trivial case where $\Theta_\psi(\mathcal{R}) = \emptyset$ in (12) holds for all ϵ_θ. The condition that $\Theta_\psi(\mathcal{R})$ is a polytope (and hence bounded) is necessary to obtain a bounded Hausdorff distance. This distance quantifies the difference between two sets, and is a necessary step to bound the approximation error. The requirement that $\Theta_\psi(\mathcal{R})$ includes the origin is a sufficient condition and relates to well-posedness for bounded input sets including the origin. When considering invariance properties defined for $0\in\mathbb{U}_{ver}$ and for any polytope $\mathbb{X}_{ver}$, the requirement that $0_n \in \Theta_\psi(\cdot)$ is necessary for $\Theta_{\square\psi}$ to be non-empty: this can be intuitively illustrated by noting that under an assumption of asymptotic stability for A, for any θ and for $u(\cdot)=0$, the output $\hat{y}(t,\theta)$ of the model in (6) converges to 0. Hence for a property to be satisfied under these conditions it should at least hold for the zero output, which is equivalent to demanding that it holds for the parameter $\theta = 0_n$. For any atomic proposition $p_i \in AP$ (see Equation (5)) it can be shown that there is an invertible mapping between the row vectors, proportional to the normals of the faces of the polyhedral set $\Theta_{p_i}(x(0))$, and the initial state $x(0)$. Therefore, if $\mathcal{R}^{(k)}$ has the origin in its interior, then $\Theta_{p_i}(\mathcal{R}^{(k)})$ has to be bounded, and as a consequence so does any feasible set comprising this atomic proposition. This holds for $k\ge n$ if (A,B) is a reachable pair and if $\mathbb{U}_{ver}$ has 0 in its interior. Under the same conditions there exists a k such that $\mathcal{R}^{(k)}_{\mathbb{X}_{ver}}$ has 0_n in its interior. The generalisation to the case dealing with an Hausdorff distance of the feasible set for invariance properties with a set of inputs $0\notin\mathbb{U}_{ver}$ is outside of the scope of this work.

Convergence properties of robust approximations

We can employ Lemma 4 to bound the Hausdorff distance between $\Theta_\psi(\mathcal{R}^{(k)}_{\mathbb{X}_{ver}})$ and $\Theta_{\square\psi}$. If $\mathbb{X}_{ver} = \{0_n\}$ and the spectral radius of A is strictly less than 1 (that is $\rho(A) < 1$ or equivalently A is asymptotically stable), then the Hausdorff distance can be bounded as

$$\delta_H(\mathcal{R}^{(k)},\mathcal{R}^\infty) \le \epsilon(k) := \|A^k\|_2 \max_{u\in\mathbb{U}}(|u|)c_1, \tag{18}$$

with c_1 a bound on $\sum_{i=0}^\infty \|A^iB\|$, which is the peak-to-peak performance of the dynamical system formed by (A,B). The derivation of the inequality above, and of the subsequent results, can be found in the Appendix. Stronger results can be obtained via dedicated software for these computations [17]. In the case that $\mathbb{X}_{ver} \not\subseteq \mathcal{R}^\infty$ then the forward reachable iteration can be rewritten as

$$\mathcal{R}^{(k)}_{\mathbb{X}_{ver}} = \big(\bigcup_{i=0}^{k} A^i\mathbb{X}_{ver}\big) + \mathcal{R}^{(k)}.$$

The Hausdorff norm can be bounded as

$$\delta_H(\mathcal{R}^{(k)}_{\mathbb{X}_{ver}}, \mathcal{R}^{\infty}_{\mathbb{X}_{ver}}) \leq \epsilon(k) + \|A^{k+1}\|_2 \delta_H \left(\mathbb{X}_{ver}, \{0_n\}\right).$$

Note that for $\rho(A) < 1$ the norm $\|A^k\|_2 \to 0$ for $k \to \infty$. In case the conditions of Lemma 4 on $\mathcal{R}^{(k)}_{\mathbb{X}_{ver}} \subseteq \mathbb{X}$ and $\Theta_\psi\big(\mathcal{R}^{(k)}_{\mathbb{X}_{ver}}\big)$ hold, the Hausdorff distance $\delta_H(\Theta_{\Box^k\psi}, \Theta_{\Box\psi})$ can be bounded by

$$\|A^k\|_2 \max_i(\|v_i\|)^2 \epsilon_p\big(\max_{u\in\mathbb{U}}(|u|)c_1 + \|A\|\delta_H(\mathbb{X}_{ver}, \{0_n\})\big). \tag{19}$$

Verification of unbounded-time properties

Based on the convergence properties of the feasible set, the asymptotic behaviour of the confidence computed in Proposition 1 can be stated as follows.

Corollary 5 (Convergence) *Under the conditions of Lemma 4, the feasible sets $\Theta_{\Box^k\psi}$ and $\Theta_{\Box\psi}$ are measurable; further, for a Gaussian distribution $p(\theta) \sim \mathcal{N}(\mu_\theta, R_\theta)$ with a covariance $R_\theta \succ 0$, $\mathbf{P}\left(\theta \in \Theta_{\Box^k\psi}\right) \to \mathbf{P}\left(\theta \in \Theta_{\Box\psi}\right)$ for $k \to \infty$.*

Proof[of Corollary 5] For a strictly positive R_θ, the Gaussian density distribution takes finite values over the parameter space, therefore the convergence of a monotonically-decreasing polytope over the parameter space induces the convergence of the associated probability measure. □

Theorem 2 can now be generalised to include unbounded-time invariance properties as follows.

Theorem 6 *Consider a polytopic set of initial states $x(0) \in \mathbb{X}_{ver}$, inputs $u(t) \in \mathbb{U}_{ver}$ for $t \geq 0$, and a labelling map as in (5). Let $\hat{\mathcal{R}}^{\infty}_{\mathbb{X}_{ver}}$ be a polytopic superset of the minimal robustly positively invariant set that includes $\mathbb{X}_{ver}$, denoted as $\mathcal{R}^{\infty}_{\mathbb{X}_{ver}}$. Then the feasible set admits a polyhedral subset $\hat{\Theta}_\psi \subset \Theta_\psi$ for every specification ψ expressed within the LTL fragment $\psi ::= \alpha|\bigcirc\psi|\psi_1 \wedge \psi_2|\Box\psi$, and if $\hat{\mathcal{R}}^{\infty}_{\mathbb{X}_{ver}} = \mathcal{R}^{\infty}_{\mathbb{X}_{ver}}$ then $\hat{\Theta}_\psi = \Theta_\psi$.*

Proof[of Theorem 6] Every property $\phi ::= p|\bigcirc\psi|\psi_1 \wedge \psi_2|\Box\psi$ with $p \in AP$ can be rewritten as $\psi_1 \wedge \Box\psi_2$ where ψ_1 and ψ_2 have syntax $\psi ::= p|\bigcirc\psi|\psi_1 \wedge \psi_2$. Consider a property $\psi_* = \psi_1 \wedge \Box\psi_2$, and let us leverage equivalences among LTL formulae [2]. For $\bar{\psi}_1$ and $\bar{\psi}_2$ in $\psi ::= p|\bigcirc\psi|\psi_1 \wedge \psi_2$ the properties $\psi_{1*} := \psi_1 \wedge \bar{\psi}_1$ and $\psi_{2*} := \psi_2 \wedge \bar{\psi}_2$ are such that $\psi_* \wedge (\bar{\psi}_1 \wedge \Box\bar{\psi}_2) \equiv \psi_{1*} \wedge \Box\psi_{2*}$. Now consider $\bigcirc\psi_* \equiv \bigcirc(\psi_1 \wedge \Box\psi_2) \equiv \bigcirc(\psi_1 \wedge \Box\psi_2) \equiv (\bigcirc\psi_1) \wedge (\bigcirc\Box\psi_2)$, from the distributive law of $\bigcirc$. Using the semantics of $\bigcirc$ and $\Box$, it follows that $\bigcirc\Box\psi_2$ is equivalent to $\Box\bigcirc\psi_2$. Thus for $\psi_{1*} := \bigcirc\psi_1$ and $\psi_{2*} := \bigcirc\psi_2$ it holds that $\bigcirc\psi_* \equiv \psi_{1*} \wedge \Box\psi_{2*}$. Take $\Box\psi_* \equiv \Box(\psi_1 \wedge \Box\psi_2) \equiv (\Box\psi_1) \wedge (\Box\Box\psi_2)$ based on the distributive law (c.f. [2, p.248]), which is subsequently equal to $(\Box\psi_1) \wedge (\Box\psi_2) \equiv \Box(\psi_1 \wedge \psi_2)$ by applying the idempotency law and the distributive law. Hence $\Box\psi_* \equiv \Box(\psi_1 \wedge \psi_2)$.
In conclusion, every property $\phi ::= p|\bigcirc\psi|\psi_1 \wedge \psi_2|\Box\psi$ can be written as $\psi_1 \wedge \Box\psi_2$ where ψ_1 and ψ_2 have syntax $\psi ::= p|\bigcirc\psi|\psi_1 \wedge \psi_2$: this is because we have shown that every operation ($\bigcirc$, $\wedge$, $\Box$) preserves this rewriting.

For the set of initial states $\mathbb{X}_{ver}$, a property ψ is invariant

$$\langle \mathbf{M}(\theta), x(0)\rangle \vDash \Box\psi, \ \forall x(0) \in \mathbb{X}_{ver}$$

if and only if $\forall x \in \mathcal{R}^{\infty}_{\mathbb{X}_{ver}} : \langle \mathbf{M}(\theta), x\rangle \vDash \psi$. Let $\hat{\mathcal{R}}^{\infty}_{\mathbb{X}_{ver}}$ be a polytopic superset of $\mathcal{R}^{\infty}_{\mathbb{X}_{ver}}$ with a finite set of vertices $v_{\mathcal{R}} \in V_{\mathcal{R}}$. Then the subset approximation of the feasible set $\Theta_{\Box\psi}$ follows as $\Theta_{\Box\psi} \supseteq \hat{\Theta}_{\Box\psi} =$

$$\left\{\theta \in \Theta : \bigwedge_{v_{\mathcal{R}} \in V_{\mathcal{R}}} \left((I_{n_\psi} \otimes v^T_{\mathcal{R}})N_\psi + K_\psi\right)\theta \leq B_\psi\right\},$$

where $\hat{\Theta}_{\Box\psi} \subseteq \Theta_{\Box\psi}$. Note that if $\hat{\mathcal{R}}^{\infty}_{\mathbb{X}_{ver}} = \mathcal{R}^{\infty}_{\mathbb{X}_{ver}}$ then $\hat{\Theta}_{\Box\psi} = \Theta_{\Box\psi}$. The feasible set of $\psi_1 \wedge \Box\psi_2$ is equal to $\Theta_{\psi_1 \wedge \Box\psi_2} = \Theta_{\psi_1} \cap \Theta_{\Box\psi_2}$. And $\Theta_{\psi_1 \wedge \Box\psi_2}$ can be upper and lower bounded as $\Theta_{\psi_1} \cap \hat{\Theta}_{\Box\psi_2} \subseteq \Theta_{\psi_1 \wedge \Box\psi_2} \subseteq \Theta_{\psi_1} \cap \Theta_{\Box^k\psi_2}$ with $k \in \mathbb{N}$. This proves Theorem 6 for the case where the model is $(A, B, C(\theta), 0)$. The proof for the model class with additional parameterisation of D can be derived similarly. □

The extension beyond the LTL fragment discussed above may lead to feasible sets that are in general not convex, and is therefore beyond the scope of this work.

3.6 Case Study (continuation): Unbounded-Time Safety Verification

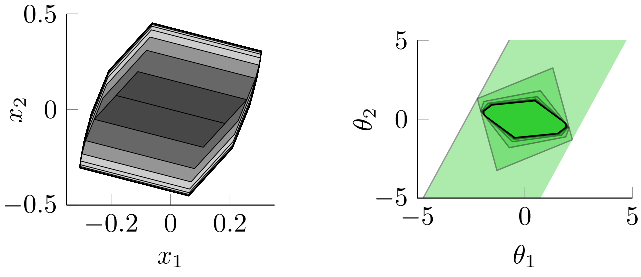
We study convergence properties for the safety specification ι considered in the case study in Section 3.4, maintaining the same operating conditions as before for the verification step and for the experiments. In Figure 4a the forward reachability sets $\mathcal{R}^{(k)}$ with $k = 1, \ldots, 20$ are obtained for the model dynamics in (8). Figure 5 (upper plot) displays bounds $\epsilon(k)$ on the Hausdorff distances $\delta_H(\mathcal{R}^{(k)}, \mathcal{R}^{\infty})$ computed with (18): starting from a slanted line segment for $\mathcal{R}^{(1)}$ as in Figure 4a, it can be observed that the forward reachable sets $\mathcal{R}^{(k)}$ converge rapidly, as confirmed with the error bound displayed in Figure 5 (upper plot).

Based on $\mathcal{R}^{(k)}$, the feasible set for the k-bounded invariance $\Box^k\iota$ can be computed as $\Theta_{\Box^k\iota} = \Theta_\iota\big(\mathcal{R}^{(k)}\big)$. The feasible sets $\Theta_{\Box^k\iota}$ with $k = 1, \ldots, 20$ are plotted in Figure

4b. Observe that the feasible set $\Theta_{\square^1 \iota}$ is not bounded, but for $k \geq 2$ the feasible sets are bounded and, as expected, decrease in size with time. In Figure 5 (middle plot) bounds on the Hausdorff distances $\delta_H(\Theta_{\square \iota}, \Theta_{\square^k \iota})$ are given for $k = 2, \ldots, 20$ (no finite bound is computed for the index $k = 1$, since for that instance the feasible set is not bounded). Let us conclude this case study looking at confidence quantification, as a function of the time horizon. Figure 5 (lower plot) represents the confidence over the property $\mathbf{P}\left(\theta \in \Theta_{\square^k \iota} \mid Z^{N_s}\right)$, for indices $k = 1, \ldots, 20$. Unlike the case discussed in Section 3.4, which focused on looking at statistics of the confidence via mean and variance drawn over multiple experiments, we zoom in on asymptotic properties by considering a data set Z^{N_s} comprising a single trace made up of 200 measurements, simulated under the same conditions as in Section 3.4, and with $\theta_0 = [1\ 0]^T$. From the resulting probability density distribution $p\left(\theta \mid Z^{N_s}\right)$, it may be observed that the confidence converges rapidly to a nonzero value.

4 Discussion on the Generalisation of the Results

The discussed approach based on polytopes allows for analytical expressions of the feasible set, however the implementation may not scale to models with very large dimension: in particular, the number of half-planes characterising the feasible set may increase with the time bound of the LTL formula ψ (that is, with the repeated application of the $\bigcirc$ operator), and with the cardinality of the set of atomic propositions in the alphabet Σ. Still, these computations are essentially equivalent to those of known reachability algorithms, therefore the method is extensible well beyond the 2-dimensional case study, especially when applying sophisticated reachability analysis tools in the literature [17,12]. Therefore the discussed limitations related to the current implementation of the approach, ought to be dealt with in the future by the use of tailored and less naïve computational approaches.



(a) The first 20 iterations of the forward reachable set $\mathcal{R}^{(k)}$, $k = 1, \ldots, 20$ for the case study. The reachable sets grow in size from dark grey ($k = 1$) to light grey ($k = 20$), so that $\mathcal{R}^{(k-1)} \subseteq \mathcal{R}^{(k)}$.

(b) The feasible sets for the k-bounded invariance property $\square^k \iota$, with $k = 1, \ldots, 20$, obtained for the case study.

Fig. 4. Reachable and feasible sets for the unbounded-time verification problem.

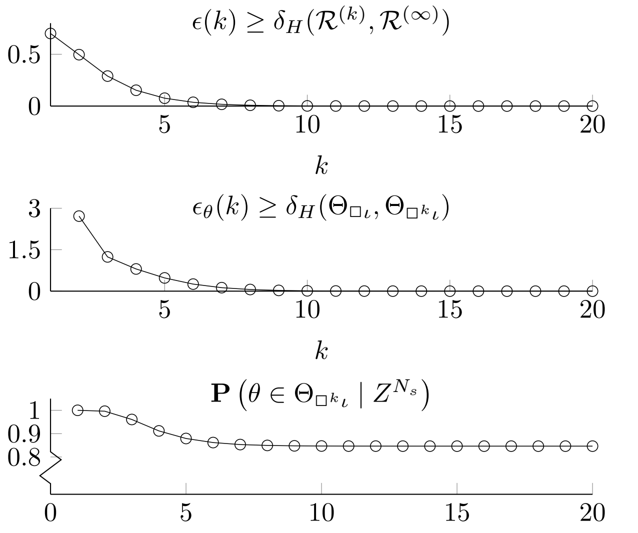


Fig. 5. (Upper plot) Error bound on the approximation level of the k-th forward reachable sets, which is such that $\mathcal{R}^{(\infty)} \subseteq \mathcal{R}^{(k)} + \epsilon(k)$ for $k = 1, \ldots, 20$. (Middle plot) The Hausdorff distance $\epsilon_\theta(k)$ between $\Theta_{\square^k \psi}$ and $\Theta_{\square \psi}$ with $k = 2, \ldots, 20$, obtained for the case study.(Lower plot) Confidence that $\mathbf{S} \vDash \square^k \iota$ for $k = 1, \ldots, 20$ for the case in Section 3.4, with a new experiment consisting of 200 samples collected as Z^{N_s}.

In the discussion of model selection, we elaborated possible generalisations beyond linearly-parameterised model sets. Future extension will in particular deal with hybrid models, since when systems are not linear, their (local) behaviour is often well approximated with piecewise-linear dynamical models.

We are presently working to extensions of the considered set of logic formulae of interest, and plan to employ experiment design to optimise the input-output signal interaction for efficient data usage over general classes of models, as initially attempted in [20]. Additionally, the design of control policies that optimise properties of interest over partly unknown systems is topic of current work.

Finally, current work targets the applicability of tractable solutions to model-based and data-driven verification over complex physical systems.

5 Conclusions

This paper has introduced a new framework for the integrated formal verification and modelling of physical systems with partly unknown dynamics. A Bayesian framework allowing for the efficient incorporation of measurement data and prior information has been combined with a verification procedure based on safety analysis. The new approach allows for the computation of the confidence level over the validity of a property of interest on the unknown system. The method has been applied to the verification of LTI models of systems over bounded and unbounded safety properties (a fragment of LTL logics), and its computational overhead has been focus of discussion.

Acknowledgements

The research of Sofie Haesaert is supported by DISC through a personal grant from the NWO graduate program.

References

[1] A. Abate, R. C. Hillen, and S. A. Wahl. Piecewise affine approximation of fluxes and enzyme kinetics from in-vivo ^{13}C labeling experiments. *International Journal of Robust and Nonlinear Control*, pages 1120–1139, 2012. Special Issue on System Identification for Biological Systems.

[2] C. Baier and J.-P. Katoen. Principles of model checking. *MIT Press*, 2008.

[3] E. Bartocci, L. Bortolussi, and G. Sanguinetti. Learning temporal logical properties discriminating ECG models of cardiac arrhytmias. *CoRR*, abs/1312.7523, 2013.

[4] G. Batt, C. Belta, and R. Weiss. Model checking genetic regulatory networks with parameter uncertainty. In *HSCC*, pages 61–75. Springer, 2007.

[5] C. Belta, A. Bicchi, M. Egerstedt, E. Frazzoli, E. Klavins, and G. J. Pappas. Symbolic planning and control of robot motion [grand challenges of robotics]. *Robotics & Automation Magazine, IEEE*, 14(1):61–70, 2007.

[6] C. Belta, L. C. G. J. M. Habets, and V. Kumar. Control of multi-affine systems on rectangles with applications to hybrid biomolecular networks. In *Proceedings of the Conference on Decision and Control*, pages 534–539, 2002.

[7] F. Blanchini and S. Miani. *Set-Theoretic Methods in Control.* Birkhäuser Basel, 1st edition, 2007.

[8] L. Bortolussi and G. Sanguinetti. Learning and designing stochastic processes from logical constraints. In *QEST*, pages 89–105. Springer, 2013.

[9] L. Bortolussi and G. Sanguinetti. Smoothed model checking for uncertain continuous time Markov chains. *CoRR*, abs/1402.1450, 2014.

[10] L. Brim, M. Češka, S. Dražan, and D. Šafránek. Exploring parameter space of stochastic biochemical systems using quantitative model checking. In N. Sharygina and H. Veith, editors, *CAV*, volume 8044 of *LNCS*, pages 1–17. Springer, 2013.

[11] J. W. Burdick, N. du Toit, A. Howard, C. Looman, J. Ma, R. M. Murray, and T. Wongpiromsarn. Sensing, navigation and reasoning technologies for the DARPA urban challenge. Technical report, DTIC Document, 2007.

[12] D. Cattaruzza, A. Abate, P. Schrammel, and D. Kroening. Unbounded-time analysis of guarded lti systems with inputs by abstract acceleration. In *International On Static Analysis*, pages 312–331. Springer, 2015.

[13] Y. Chen and T. D. Nielsen. Active learning of Markov decision processes for system verification. In *Conference on Machine Learning and Applications*, pages 289–294, 2012.

[14] E. M. Clarke. The birth of model checking. In *25 Years of Model Checking*, pages 1–26. Springer, 2008.

[15] D. Del Vecchio and E. D. Sontag. Engineering principles in bio-molecular systems: From retroactivity to modularity. *European Journal of Control*, pages 389 – 397, 2009.

[16] G. Frehse, S. K. Jha, and B. H. Krogh. A counterexample-guided approach to parameter synthesis for linear hybrid automata. In *HSCC*, pages 187–200. Springer Berlin Heidelberg, 2008.

[17] G. Frehse, C. Le Guernic, A. Donzé, S. Cotton, R. Ray, O. Lebeltel, R. Ripado, A. Girard, T. Dang, and O. Maler. Spaceex: Scalable verification of hybrid systems. In S. Q. Ganesh Gopalakrishnan, editor, *Proc. 23rd International Conference on Computer Aided Verification (CAV)*, LNCS. Springer, 2011.

[18] B. M. Gyori, D. Paulin, and S. K. Palaniappan. Probabilistic verification of partially observable dynamical systems. *CoRR*, abs/1411.0976, 2014.

[19] S. Haesaert, P. M. J. Van den Hof, and A. Abate. Data-driven and model-based verification: A bayesian identification approach. In *Proceedings of the 54th IEEE Conference on Decision and Control (CDC)*, pages 6830–6835. IEEE, 2015.

[20] S. Haesaert, P. M. J. Van den Hof, and A. Abate. Data-driven property verification of grey-box systems by Bayesian experiment design. In *American Control Conference*, pages 1800–1805, 2015.

[21] D. Henriques, J. G. Martins, P. Zuliani, A. Platzer, and E. M. Clarke. Statistical model checking for Markov decision processes. In *QEST*, pages 84–93, 2012.

[22] T. Henzinger and H. Wong-Toi. Using hytech to synthesize control parameters for a steam boiler. In *Formal Methods for Industrial Applications*, pages 265–282. Springer Berlin Heidelberg, 1996.

[23] P. S. C. Heuberger, P. M. J. Van den Hof, and O. H. Bosgra. A generalized orthonormal basis for linear dynamical systems. *Automatic Control, IEEE Transactions on*, 40(3):451–465, 1995.

[24] P. S. C. Heuberger, P. M. J. Van den Hof, and B. Wahlberg. *Modelling and identification with rational orthogonal basis functions.* Springer London, 2005.

[25] H. Hjalmarsson. From experiment design to closed-loop control. *Automatica*, pages 393–438, 2005.

[26] E. A. Lee. Cyber physical systems: Design challenges. In *Proc. of Object Oriented Real-Time Distributed Computing*, pages 363–369. IEEE Computer Society, 2008.

[27] A. Legay, B. Delahaye, and S. Bensalem. Statistical model checking: An overview. In H. Barringer, Y. Falcone, B. Finkbeiner, K. Havelund, I. Lee, G. Pace, G. Roşu, O. Sokolsky, and N. Tillmann, editors, *Runtime Verification*, volume 6418 of *LNCS*, pages 122–135. Springer Berlin Heidelberg, 2010.

[28] A. Legay and S. Sedwards. Lightweight Monte Carlo algorithm for Markov decision processes. *CoRR*, abs/1310.3609, 2013.

[29] D. V. Lindley. The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, pages 293–337, 2000.

[30] H. Mao and M. Jaeger. Learning and model-checking networks of I/O automata. In *Proc. of Asian Conference on Machine Learning*, 2012.

[31] V. Peterka. Bayesian Approach to System Identification. *Trends Prog. Syst. Identif.*, 1981.

[32] B. C. Reginato, R. Z. Freire, G. H. D. C. Oliveira, N. Mendes, and O. Abadie, Marc. Predicting the temperature profile of indoor buildings by using orthonormal basis functions. In *Conf. on Building Performance Simulation Association*, United Kingdom, 2009.

[33] K. Sen, M. Viswanathan, and G. Agha. Learning continuous time Markov chains from sample executions. In *QEST*, pages 146–155, 2004.

[34] K. Sen, M. Viswanathan, and G. Agha. Statistical model checking of black-box probabilistic systems. In R. Alur and D. Peled, editors, *CAV*, volume 3114 of *LNCS*, pages 202–215. Springer, 2004.

[35] P. Tabuada. *Verification and Control of Hybrid Systems: a Symbolic Approach.* Springer, 2009.

[36] P. M. J. Van den Hof, P. S. C. Heuberger, and J. Bokor. System identification with generalized orthonormal basis functions. *Automatica*, pages 1821–1834, 1995.

[37] M. Y. Vardi. From philosophical to industrial logics. In *Logic and Its Applications*, pages 89–115, Berlin, Heidelberg, 2009. Springer-Verlag.

[38] G. S. Virk and D. L. Loveday. Model-based control for HVAC applications. In *Conf. on Control Applications*, pages 1861–1866. IEEE, 1994.

[39] P. Zuliani, A. Platzer, and E. M. Clarke. Bayesian statistical model checking with application to Stateflow/Simulink verification. *Formal Methods in System Design*, pages 338–367, 2013.

Derivation of the Bounds in Section 3.5

We derive the Hausdorff distance used in the subsection "Convergence properties of robust approximations".

1. Hausdorff distance of forward reachable mappings. We sketch the method to bound the Hausdorff distance, whereas a more formal derivation can be found in the literature on robustly positively invariant sets [7].

The k-step forward reachable set equals to

$$\mathcal{R}^{(k)} = \bigcup_{i=1}^{k} \Big\{ \sum_{j=1}^{i} A^{j-1} B u(i-j), \quad \text{for } u(j) \in \mathbb{U}_{ver} \Big\}.$$

For $0 \in \mathbb{U}_{ver}$, the minimal invariant set $\mathcal{R}^{\infty}$ can be written as

$$\mathcal{R}^{(\infty)} = \Big\{ \sum_{j=0}^{i-1} A^j B u(j) + A^i \sum_{k=0}^{\infty} A^k B u(k), \\ \text{for } u(\cdot) \in \mathbb{U}_{ver} \Big\}.$$

If the spectral radius of a A is strictly smaller than 1, $\rho(A) < 1$, then

$$\mathcal{R}^{(\infty)} \subseteq \mathcal{R}^{(k)} + \epsilon(k)\mathcal{B},$$

with $A^k \sum_{i=0}^{\infty} A^i B u(k) \subseteq \epsilon(k)\mathcal{B}$, for $u(\cdot) \in \mathbb{U}_{ver}$. Note that $\epsilon(k)$ is bounded for $\rho(A) < 1$. For a matrix A without defective eigenvalues, i.e. where the eigenvectors form a complete basis, this L_1 norm (the peak-to-peak performance) can be easily bounded using the spectral radius of A, by selecting

$$\begin{aligned} \epsilon(k) &= \frac{|\rho(A)|^k}{1 - |\rho(A)|} \|B\|_2 \max_{u \in \mathbb{U}_{ver}} (|u|) \\ &\geq \|A^k\|_2 \sum_{i=0}^{\infty} \|A^i B\|_2 |u(k)|. \end{aligned}$$

In case that the matrix A is defective, we opt to bound the L_1-norm by exploiting the absolute sum of the L_2 induced norm for A^i, $i \to \infty$: $\sum_{i=0}^{\infty} \|A^i\|_2$. Note that $\|A^i\|_2$ converges to 0 for $i \to \infty$ since $\rho(A) < 1$, therefore there exists a finite l such that $\|A^l\|_2 < 1$ and we can upper bound the absolute sum as

$$\begin{aligned} \sum_{i=0}^{\infty} \|A^i\|_2 &\leq \Big(\sum_{i_1=0}^{l-1} \|A^{i_1}\|_2 \Big) \Big(\sum_{i_2=0}^{\infty} \|A^l\|_2^{i_2} \Big) \\ &= \Big(\sum_{i_1=0}^{l-1} \|A^{i_1}\|_2 \Big) \frac{1}{1 - \|A^l\|_2}. \end{aligned}$$

Thus in general, the Hausdorff distance can be bounded as

$$\delta_H(\mathcal{R}^{(k)}, \mathcal{R}^{(\infty)}) \leq \epsilon(k) = \|A^k\|_2 \max_{u \in \mathbb{U}_{ver}} (|u|) c_1,$$

with $c_1 = \frac{\left(\sum_{i_1=0}^{l} \|A^{i_1}\|_2 \right)}{1 - \|A^l\|_2} \|B\|_2$ for l such that $\|A^l\|_2 < 1$. Note that c_1 can be replaced by any bound on the L_1 norm (the peak-to-peak performance) of the dynamical system formed by (A, B).

In case that $\mathbb{X}_{ver} \not\subseteq \mathcal{R}^{\infty}$ then the forward reachable iteration can be rewritten as

$$\mathcal{R}^{(k)}_{\mathbb{X}_{ver}} = \Big(\bigcup_{i=0}^{k} A^i \mathbb{X}_{ver} \Big) + \mathcal{R}^{(k)},$$

for which we know that

$$\mathcal{R}^{(\infty)}_{\mathbb{X}_{ver}} \subseteq \mathcal{R}^{(k)}_{\mathbb{X}_{ver}} + \epsilon(k) + \|A\|^{k+1} \delta_H(\mathbb{X}_{ver}, \{0\}).$$

Thus the Hausdorff norm is upper bounded as $\delta_H(\mathcal{R}^{(k)}_{\mathbb{X}_{ver}}, \mathcal{R}^{(\infty)}_{\mathbb{X}_{ver}}) \leq \epsilon(k) + \|A^{k+1}\| \delta_H(\mathbb{X}_{ver}, \{0\})$.

2. Hausdorff distance on feasible sets. Suppose that the conditions in Lemma 4 hold for $\mathcal{R}^{(k)}_{\mathbb{X}_{ver}}$, then we can compute a value for ϵ_θ such that $\Theta_\psi(\mathcal{R}^{(k)}_{\mathbb{X}_{ver}}) \subseteq \Theta_\psi(\mathcal{R}^{(k)}_{\mathbb{X}_{ver}} + \epsilon_x \mathcal{B}) + \epsilon_\theta \mathcal{B}$, where ϵ_x is a bound on the Hausdorff distance $\delta_H(\mathcal{R}^{(k)}_{\mathbb{X}_{ver}}, \mathcal{R}^{(\infty)}_{\mathbb{X}_{ver}})$.

The set operation $\Theta_\psi(\cdot)$ is monotonically decreasing, therefore $\Theta_\psi(\mathcal{R}^{(k)}_{\mathbb{X}_{ver}} + \epsilon(k)\mathcal{B}) \subseteq \Theta_{\Box\psi} = \Theta_\psi\left(\mathcal{R}^{\infty}_{\mathbb{X}_{ver}}\right) \subseteq \Theta_\psi\left(\mathcal{R}^{(k)}_{\mathbb{X}_{ver}}\right) = \Theta_{\Box^k\psi}$, and $\Theta_{\Box^k\psi} \subseteq \Theta_\psi(\mathcal{R}^{(k)}_{\mathbb{X}_{ver}} + \epsilon(k)\mathcal{B}) + \epsilon_\theta\mathcal{B} \subseteq \Theta_{\Box\psi} + \epsilon_\theta\mathcal{B}$, and

$$\Theta_{\Box\psi} \subseteq \Theta_{\Box^k\psi} \subseteq \Theta_{\Box\psi} + \epsilon_\theta\mathcal{B}.$$

Based on Lemma 4, with $\epsilon_p = \max_{p_i} \frac{|A_{p_i}|}{|b_{p_i}|}$, we obtain

$$\epsilon_\theta = \frac{\epsilon_x\epsilon_p \max_i(\|v_i\|)^2}{1 + \epsilon_x\epsilon_p \max_i(\|v_i\|)} \leq \epsilon_x\epsilon_p \max_i(\|v_i\|)^2.$$

Note that since $\|A^k\|_2$ converges to 0 for $k \to \infty$ for $\rho(A) < 1$, and since $\max_i(\|v_i\|)^2$ is not increasing, the error ϵ_θ also converges to 0.