2015 American Control Conference
Palmer House Hilton
July 1-3, 2015. Chicago, IL, USA

Data-driven Property Verification of Grey-box Systems by Bayesian Experiment Design

S. Haesaert, P.M.J. Van den Hof, and A. Abate

Abstract— **A measurement-based statistical verification approach is developed for systems with partly unknown dynamics. These grey-box systems are subject to identification experiments which, new in this contribution, enable accepting or rejecting system properties expressed in a linear-time logic. We employ a Bayesian framework for the computation of a confidence level on the properties and for the design of optimal experiments. Applied to dynamical systems, this work enables data-driven verification of partly-known system dynamics with controllable non-determinism (inputs) and noisy output observations. A numerical case study concerning the safety of a dynamical system is used to elucidate this data-driven and model-based verification technique.**

I. 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. These guarantees can be attained by the use of formal methods [1]. Formal methods lead to the verification, over a model of the system, of specifications formulated in mathematically sound terms, for example by means of temporal logics. Temporal logic specifications include time-dependent properties such as reachability, obstacle avoidance, stability, and recurrence. Much recent research has dealt with the extension of formal methods from finite-state models, widely employed for software and hardware verification, to models of complex dynamical systems. This has led to applications in symbolic motion planning for robotics [2] and in the analysis of biological systems [3].

The strength of formal techniques is limited by its required access to a model built from the full knowledge of the behaviour of the underlying system. The goal of this line of research is data-driven and model-based verification for partly unknown physical systems that are accessible via noisy input-output measurements. With focus on properties of interest over the system, we investigate the computation of identification experiments that optimally excite the system with respect to specifications expressing such properties. We thus quite naturally assume that the system is available for experiments in an environment where we can change its input signals at will. Measurements are available as time series of input and noisy output signals: carrying information on the dynamics, they can be used to refine and decrease the uncertainty over the model and the properties of interest.

The area of system identification [4], [5] investigates measurement-based modelling of physical systems. Input signals excite the system behaviour observed via measurements, and can be chosen to maximise the amount of information gained. As the optimal input typically depends on the knowledge of the true system, the literature distinguishes three application oriented input-design approaches: an iterative approach, where an estimate of the nominal system is used to design the experiment at each stage; a min-max design that is robust to the worst-case scenario; and a Bayesian design that uses the prior uncertainty distribution over the model. The first approach predominates [4], [5], whereas some work has been done on the robust experiment design using the min-max approach [6]. On the other hand the third approach, well known from Bayesian statistics [7], is not yet widely employed.

Contribution: In this work we focus on the verification of systems modelled within a linearly-parameterised class of deterministic input-output models using Bayesian identification and experiment design. The contribution shows that for a subset of LTL specifications the confidence on the validity of a formal property can be computed using Bayesian inference over a finite sample set (cf. Section II). Since the performance of this data-driven method depends on the design of the experiment, we further define an optimality criterion that allows selecting an input using Bayesian experiment design. We display this approach on the safety verification of linearly-parameterised models (cf. Section III).

Related work: Recent work within the formal methods community also concerns the use of simulations and of measurements for verification. Statistical Model Checking (SMC) [8] employs finite executions generated by the model to find statistical evidence for the verification of bounded-time logical properties. SMC can be applied to *black-box systems* [9], which have a probability distribution that is not known. Beyond this, whilst SMC is well applicable to physical systems with unknown dynamics, it is in general limited to state-observable and fully-stochastic systems. Further, the presence of sets of inputs, disturbances, and unknown initial states or other forms of non-determinisms, are not easily incorporated into SMC [10], [11]. As an alternative, [12], [13] efficiently use data drawn from an input-output, finite state Markov system, to learn the corresponding model and to verify it. These results are bound to finite-state models, making them less applicable to more complex systems. Similarly, [14] use advanced machine learning techniques to infer finite-state Markov models from data over given logical formulae.

S. Haesaert and P.M.J. Van den Hof are with the Control Systems group in the Faculty of Electrical Engineering, Eindhoven University of Technology, The Netherlands. A. Abate is with the Department of Computer Science, Oxford University, UK.

978-1-4799-8686-6/$31.00 ©2015 AACC

II. Data-driven and Model-based verification

Let us recapitulate the overall goal of this work: *starting from available a-priori knowledge over system* $\mathbf{S}$*, iteratively and efficiently gather measurements until a specification* ψ *defined over the system is verified or falsified with a given confidence* δ.

System and Models: The system, denoted by $\mathbf{S}$ as in Figure 1, is measured in discrete time. An input signal $u(t) \in \mathbb{U}, t \in \mathbb{N}$, captures how the environment acts on the system. Similarly, the output $y_0(t) \in \mathbb{Y}$ indicates how the system interacts with the environment (namely, how it can be measured). The measurements $\tilde{y}(t)$ at $t \in \mathbb{N}$ of $y_0(t)$ are disturbed by the measurement noise $e(t)$.

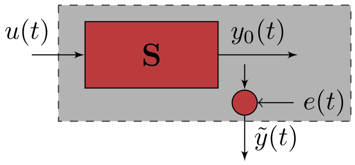

Fig. 1: System $\mathbf{S}$ has input $u(t)$ and output $y_0(t)$. In the measurement setup, the measured output $\tilde{y}(t)$ includes the system output $y_0(t)$ and the measurement noise $e(t)$.

The behaviour of a deterministic system can be described by mathematical models as a (causal) relation between the system input and output. In most cases the knowledge of the behaviour of a system is only partial, making it impossible to represent the system by a "true" model. In such cases, a-priori available knowledge allows to construct a model set $\mathcal{G}$, with elements $\mathbf{M} \in \mathcal{G}$ representing possible mathematical models of $\mathbf{S}$. Let us denote a parameterisation of the model set $\mathcal{G}$ as the mapping $\mathbf{M}(\cdot) : \Theta \to \mathcal{G}$, from the parameters $\theta \in \Theta$ in the parameter set, which is a subset of a Euclidean space $\Theta \subset \mathbb{R}^d$, to the models $\mathbf{M}$ in $\mathcal{G}$. This allows for a parametrised expression of the model set as $\mathcal{G} = \{\mathbf{M}(\theta) | \theta \in \Theta\}$. The chosen parameterised model set is assumed to contain the "true" model denoted as $\mathbf{M}(\theta^0)$, $\theta^0 \in \Theta$, which exactly represents the behaviour of the system $\mathbf{S}$. Hence $\mathcal{G}$ encompasses the part of the behaviour that is mechanistically known. The remaining uncertainty about $\mathbf{M}(\theta^0)$ is structured as a distribution over the parameter set Θ. It is then the (unknown) model denoted by $\mathbf{M}(\theta^0) = \mathbf{S}$ that we would ideally like to formally model-check.

Whenever the lack of knowledge on the system behaviour impedes a formal verification step, it is still possible to collect data of the system by exciting it with an input sequence $\mathbf{u} = \begin{bmatrix} u(0) & u(1) & \dots & u(N_s - 1) \end{bmatrix}^T$, with N_s the length of the input sequence. Via the measurement setup, as depicted in Figure 1, noisy observations $\tilde{y}(t)$ of the output $y_0(t)$ are measured. Classical noise characteristics deal with Gaussian white noise $e(t)$ that is additive to $y_0(t)$, i.e. $\tilde{y}(t) = y_0(t) + e(t)$. Let us denote the output samples obtained by exciting the system with the input $\mathbf{u}$ as $\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{y}(1) & \tilde{y}(2) & \dots & \tilde{y}(N_s) \end{bmatrix}^T$. Since the collected data contains statistical information on the behaviour of the system, it is possible to refine the uncertainty distribution over the parameter space, as discussed in the second part of this section.

Properties: Starting from a finite set of atomic propositions $p_i \in AP$, $i = 1, \dots, |AP|$, Linear-time Temporal Logic (LTL) [15] formulae are built 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 word composed of letters from the alphabet $\Sigma = 2^{AP}$, let $\pi_t = \pi(t), \pi(t+1), \pi(t+2), \ldots$ be a subsequence of π, then the satisfaction relation between π and ψ, namely $\pi \vDash \psi$ (or equivalently $\pi_0 \vDash \psi$) is defined recursively over π_t and the LTL syntax as $\pi_t \vDash \text{true} \Leftrightarrow \text{true}$, $\pi_t \vDash p \Leftrightarrow \ p \in \pi(t)$, $\pi_t \vDash \neg\psi \Leftrightarrow \ \pi_t \nvDash \psi$, $\pi_t \vDash \psi_1 \wedge \psi_2 \Leftrightarrow \ \pi_t \vDash \psi_1$ and $\pi_t \vDash \psi_2$, $\pi_t \vDash \bigcirc\psi \Leftrightarrow \ \pi_{t+1} \vDash \psi$, $\pi_t \vDash \psi_1 \, \mathsf{U} \, \psi_2 \Leftrightarrow \ \exists i \in \mathbb{N} : \pi_{t+i} \vDash \psi_2$, and $\forall j \in \mathbb{N} : 0 \leq j < i, \pi_{t+j} \vDash \psi_1$.

Of interest are formal properties encoded over the input-output behaviour of the system over the time horizon $t \geq 0$. Starting at an arbitrary time (say $t = 0$), the set of initial states of the system is given: we assume that this set encompasses the knowledge of past inputs and/or outputs of the system. The input signal is bounded and represents the possible external nondeterminism of the environment acting on the system. The output $y_0(t) \in \mathbb{Y}$ is labeled by a map $L : \mathbb{Y} \to \Sigma$, which assigns letters in the alphabet Σ to half spaces on the output, as $A_{p_i} y_0(t) \leq b_{p_i}$. In other words, sets of atomic propositions are associated to intervals over $\mathbb{Y} \subset \mathbb{R}$. A system, or equivalently the model that represents it, satisfies a property if all the words generated by the model verify that property. Since properties are encoded over the external (input-output) behaviour of the system, which is the behaviour of $\mathbf{M}(\theta^0)$, $\theta^0 \in \Theta$, we can 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 $\mathbf{M}(\theta^0) \vDash \psi$. Let us underline that properties are defined over the behaviour of the system, and not over the noisy measurements $\tilde{y}(t)$ of the system. Let us define Θ_ψ to be the maximal feasible set of parameters, such that for every parameter in that set the property ψ holds, i.e. $\forall \theta \in \Theta_\psi : \mathbf{M}(\theta) \vDash \psi$. This set has been alternatively described [16] as the level set of a satisfaction function, however since we are working with deterministic models the satisfaction function only takes binary values.

System Verification in a Bayesian Framework: We now argue that the characterisation of a distribution over the parameter set Θ can be used to compute a confidence in the satisfaction relation over the system $\mathbf{S} \vDash \psi$. This distribution encompasses the current uncertainty over $\mathbf{M}(\theta^0) = \mathbf{S}$, and can be characterised and refined using measurements of the system. Therefore, it is possible to accept or reject $\mathbf{S} \vDash \psi$ by drawing data from the measurement set-up until a certain *confidence level* is achieved. The necessary size of the data set to attain this confidence level depends on the chosen input data $\mathbf{u}$. This leads to an experiment design task: in order to optimise data efficiency, we structure the process of drawing and processing data as an iteration over 3 main stages, as represented schematically in Figure 2:

I) design (and perform) an experiment,
II) compute the corresponding parametric inference,
III) check if confidence in $\mathbf{S} \vDash \psi$ or in $\mathbf{S} \nvDash \psi$ is $> 1 - \delta$, else go to step **I**.

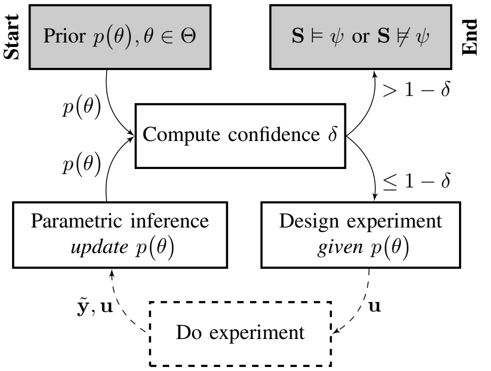

Fig. 2: The process of sequentially drawing data with the objective of verifying whether $\mathbf{S} \vDash \psi$ or not with a given confidence $1-\delta$. Starting from a prior parameters distribution $p(\theta)$, the diagram depict the key elements of the iteration with the white blocks, whereas the gray blocks represent the starting/ending parts.

With reference to Figure 2, the iteration is initiated with the construction of a distribution ('Prior'), which structures the initial limited knowledge by assigning a probability measure over the set of parameters. The first part in the iteration is to compute the confidence ('Compute confidence'), which allows leaving the iteration when the desired confidence level is achieved. On the contrary if the confidence level is not achieved, an identification experiment is designed to obtain more data from the system ('Design experiment'). The results of the experiment ('Do experiment') are then used to update the parameter distribution in 'Parametric inference'.

We employ conjugate priors [17] since they are closed under parametric inference and are in general quite expressive. Whenever the available knowledge is insufficient, that is whenever the confidence in accepting or rejecting a property is below the confidence threshold $1 - \delta$, the procedure will design and perform additional experiments.

Only in the pathologic instance where an infinitesimal deviation of the true parameter affects the satisfaction or rejection of the property, termination of the procedure cannot be attained. In practice, as long as the uncertainty distribution $p(\theta)$ has a level set δ that converges to a single point termination of the procedure follows almost surely.

In the following, we first explain how the confidence in a property is computed, then we discuss the parametric inference step via Bayesian identification for a given set of data $\mathbf{u}, \tilde{\mathbf{y}}$. Building on these two stages it becomes possible to compute optimal experiments, as in the corresponding stage.

*Confidence Computation (step **III**):* Denote the maximal feasible set $\Theta_\psi \subset \Theta$ such that $\forall \theta \in \Theta_\psi : \mathbf{M}(\theta) \vDash \psi$. The confidence in a specification ψ defined over the system $\mathbf{S}$ is computed based on the uncertainty distribution over Θ_ψ: given a prior uncertainty distribution $p(\theta)$, the confidence is computed as $\mathbf{P}(\Theta_\psi) = \int_{\Theta_\psi} p(\theta) d\theta$, whereas after an additional experiment and parametric inference (next paragraph), the a-posteriori uncertainty distribution $p(\theta|\tilde{\mathbf{y}}, \mathbf{u})$ can be used to compute the confidence as

$$\mathbf{P}(\Theta_\psi|\tilde{\mathbf{y}}, \mathbf{u}) = \int_{\Theta_\psi} p(\theta|\tilde{\mathbf{y}}, \mathbf{u}) d\theta \; . \tag{1}$$

Observe that according to Bayesian probability calculus for uncertainties [7], the confidence in a property becomes the measure of the uncertainty distribution.

*Parametric Inference (step **II**):* Given a prior distribution $p(\theta)$ and a data set $\tilde{\mathbf{y}}$ obtained by taking N_s measurements of $\tilde{y}(t)$, *the a-posteriori uncertainty distribution* $p(\theta|\tilde{\mathbf{y}}, \mathbf{u})$ is based on parametric inference [7], [17] structured over the parameter set Θ as

$$p(\theta \mid \tilde{\mathbf{y}}, \mathbf{u}) = \frac{p(\tilde{\mathbf{y}}|\theta, \mathbf{u}) p(\theta)}{\int_\Theta p(\tilde{\mathbf{y}}|\theta, \mathbf{u}) p(\theta) d\theta}. \tag{2}$$

Remark 1: Considerations of computational complexity limit the use of (1) and (2), since their solution is seldom analytical. In Section III we choose a model set that is linearly-parameterised, and has Gaussian distributions for both the measurement noise and the prior, which allows for closed-form solutions of (2). In this case, polyhedral expressions of the maximal feasible set Θ_ψ can be found for a subset of BLTL properties, which leads to easily obtainable computations of (1).

The computation of (1) for more general problems would demand the use of either *Monte-Carlo methods* to solve the relevant integrals directly without computing the feasible set first, or of *numerical approximation techniques* to obtain an (upper/lower-)approximation of the feasible set. The use of Monte-Carlo methods allows for an empirical computation of the confidence if, for each sample of the uncertainty distribution (a model in the model set), the property is decidable. Note that this transfers the bottleneck of computational complexity to the verification properties over the models, that is to $\mathbf{M}(\theta) \vDash \psi$. Numerical approximations can offer insight especially when enabled by the exploration of the parameter space in order to exploit the structure and properties of the models [18], [19].

*Experiment Design (step **I**):* Every experiment contains statistical information on the behaviour of the system, which can be used to decide whether to accept or reject a specification over the system. The objective is to design an experiment $\mathbf{u}$ that optimally exploits the dynamic behaviour of the system and thus optimises the expected value of a criterion. In this subsection, the criterion for the Bayesian experiment design problem is the expected utility related to the acceptance or the rejection of a given specification of interest, based on the identification experiment. Let us define this criterion $J(\tilde{\mathbf{y}}, \mathbf{u})$ as a function of the measured output $\tilde{\mathbf{y}}$ and input $\mathbf{u}$ data, as $J : \mathbb{U}^{N_s} \times \mathbb{Y}^{N_s}$. The data realisation $\tilde{\mathbf{y}} \sim p(\tilde{\mathbf{y}}|\mathbf{u})$ has a probability density function which is conditioned on the input signal $\mathbf{u}$. Given a prior $p(\theta)$ the probability density distribution of the data can be expressed as

$$p(\tilde{\mathbf{y}}|\mathbf{u}) = \int_\Theta p(\tilde{\mathbf{y}}|\theta, \mathbf{u}) p(\theta) d\theta \; , \tag{3}$$

where $p(\tilde{\mathbf{y}}|\theta, \mathbf{u})$ is the data distribution conditioned on the input $\mathbf{u}$ and on the parameter θ. The Bayesian experiment design problem optimises the expected value of the criterion

J over the input signal $\mathbf{u}$ for a given prior $p\big(\theta\big)$, and is formulated as:

$$\max_{\mathbf{u}\in\mathcal{E}} \mathbf{E}\left[J(\tilde{\mathbf{y}},\mathbf{u}) \mid \tilde{\mathbf{y}} \sim p\big(\tilde{\mathbf{y}}|\mathbf{u}\big)\right], \tag{4}$$

where the set of allowed experiments $\mathcal{E}$ is defined as $\mathcal{E} := \{\mathbf{u} : u(t) \in \mathbb{U},\ \forall t = 0, \dots N_s - 1\}$, with $\mathbb{U}$ a bounded set, such as for instance $[-u_{\max}, u_{\max}],\ u_{\max} \in \mathbb{R}$.

In order to use the expected utility J related to the acceptance or rejection of a specification based on the identification experiment, consider that system $\mathbf{S}$ can be represented as $\mathbf{M}(\theta^0)$, with a nominal parameter θ^0. Although θ^0 is in general unknown, it can be perceived as a realisation of the uncertainty distribution over the parameters space, i.e. $\theta^0 \sim p\big(\theta\big)$, $\theta \in \Theta$. The acceptance or rejection of $\mathbf{S} \vDash \psi$ can be equivalently cast as the choice between hypothesis H_0: $\mathbf{M}(\theta^0) \vDash \psi$ and hypothesis H_1: $\mathbf{M}(\theta^0) \nvDash \psi$. This entails a decision which is valued with 1 when correct, and with 0 when incorrect. For a given choice of H_0 or H_1, and a nominal parameter θ^0, the utility is then a binary-valued function

$$\mathrm{ut}(H_i, \theta^0) = \begin{cases} 1 & \text{if } \begin{cases} H_0 & \text{and} \quad \mathbf{M}(\theta^0) \vDash \psi \\ H_1 & \text{and} \quad \mathbf{M}(\theta^0) \nvDash \psi, \end{cases} \\ 0 & \text{else.} \end{cases} \tag{5}$$

Note that ut has a 0 value when the chosen hypothesis is wrong, which is related in statistics to type I and type II error, respectively [20, page 514].

Conditional on a data set $\tilde{\mathbf{y}}$, the nominal parameter is distributed over the parameter space as $\theta^0 \sim p\big(\theta|\tilde{\mathbf{y}}, \mathbf{u}\big)$. The *expected utility* of a decision H_i conditional on the data set is thus $\mathbf{E}\left[\mathrm{ut}(H_i, \theta^0) \mid \theta^0 \sim p\big(\theta|\tilde{\mathbf{y}}, \mathbf{u}\big)\right]$. Note that the *expected utility* represents the confidence that $\mathbf{M}(\theta^0) \vDash \psi$ or $\mathbf{M}(\theta^0) \nvDash \psi$, and is a function of both the decision and the experiment $(\tilde{\mathbf{y}}, \mathbf{u})$. Thus when deciding on H_0 or H_1, the expected utility is either $\mathrm{ut}(H_0; (\tilde{\mathbf{y}}, \mathbf{u})) = \int_{\theta\in\Theta_\psi} p\big(\theta|\tilde{\mathbf{y}}, \mathbf{u}\big) d\theta = \mathbf{P}\big(\Theta_\psi|\tilde{\mathbf{y}}, \mathbf{u}\big)$, or $\mathrm{ut}(H_1; (\tilde{\mathbf{y}}, \mathbf{u})) = \int_{\theta\in\Theta\backslash\Theta_\psi} p\big(\theta|\tilde{\mathbf{y}}, \mathbf{u}\big) d\theta = \mathbf{P}\big(\Theta\backslash\Theta_\psi|\tilde{\mathbf{y}}, \mathbf{u}\big) = 1 - \mathbf{P}\big(\Theta_\psi|\tilde{\mathbf{y}}, \mathbf{u}\big)$. As a criterion, we then choose the *expected utility* maximised over the decision H_0 or H_1, namely

$$\begin{aligned} J(\tilde{\mathbf{y}}, \mathbf{u}) \doteq &\ \max_{H_i} \mathrm{ut}(H_i; (\tilde{\mathbf{y}}, \mathbf{u})) \\ = &\ \max\left\{\mathbf{P}\big(\Theta_\psi|\tilde{\mathbf{y}}, \mathbf{u}\big), \mathbf{P}\big(\Theta\backslash\Theta_\psi|\tilde{\mathbf{y}}, \mathbf{u}\big)\right\}. \end{aligned} \tag{6}$$

III. Verification of Systems Representable by Linearly-Parameterised Models

In this section we provide a solution of the discussed new data-driven and model-based verification problem, according to the schematic process depicted in Figure 2, for single-input single-output systems in linearly-parameterised model sets, for a subset of properties expressed as bounded-horizon LTL formulae, and for a known Gaussian a priori uncertainty distribution and measurement noise. Under these restrictions, we obtain closed form solutions of (2) and convex sets for the feasible set Θ_ψ, which are then employed towards a relaxation of the criterion in (6) and a Monte Carlo solution to the Bayesian experiment design problem in (4). Linearly-parameterised model sets such as orthonormal basis function parameterisations are able to represent a wide set of systems [21, Chapter 4 and 7]. Models $\mathbf{M}$ within a linearly-parameterised model class $\mathcal{G}$ have the following state-space realisation:

$$\mathbf{M}(\theta) : \quad \begin{cases} x(t+1) & = Ax(t) + Bu(t), \\ \hat{y}(t, \theta) & = \theta^T x(t), \end{cases} \tag{7}$$

and are (linearly) parameterised by $\theta = \big[\theta_1 \dots \theta_n\big]^T \in \Theta \subset \mathbb{R}^n$. We assume that the system has a representation $\mathbf{M}(\theta^0)$ in this model set, with unknown parameter θ^0, and has an output denoted as $y_0(t) = \hat{y}(t, \theta^0)$. It is assumed that the initial state of the system and of the model representing it is $x(0) = 0$, both in the identification experiment and for the verification of the property. The noise disturbance, $e(t)$, on the measurement $\tilde{y}(t) = y_0(t) + e(t)$ is assumed to be an additive zero-mean, white, Gaussian-distributed measurement noise with variance σ_e^2 that is uncorrelated with the input. The following theorem can be derived for properties defined on the model output $y_0(t)$.

Theorem 1: Consider a linearly-parameterised model set, a bounded polyhedron for the set of initial states $x(0) \in \mathbb{X}_0$, and inputs $u(t) \in \mathbb{U}$ for $t \geq 0$. For every specification ψ expressed within the LTL fragment $\psi := \sigma|\bigcirc\psi|\psi_1 \wedge \psi_2$, with $\sigma \in \Sigma$, the feasible set of parameters $\Theta_\psi = \{\theta \in \Theta : \mathbf{M}(\theta) \vDash \psi\}$ is a polyhedron. ■

Several observations can be made. Firstly, the number of half planes characterising the set Θ_ψ may quickly 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 atomic propositions in the alphabet Σ. Secondly, the extension beyond the LTL fragment discussed above may lead to feasible sets that are in general not convex, and is therefore left for future work.

Recursive Parametric Inference: Let us denote the $(k+1)$-th iteration of the verification algorithm in Figure 2 as a combination of input design, experiment, and Bayesian identification starting from the prior knowledge gathered in the previous iterations. At the first iteration, the available knowledge is structured into a prior distribution $\mathcal{N}\big(\mu_0, R_0\big)$ over the parameter space, a multi-variate Gaussian with mean μ_0 and variance R_0. Employing Bayesian inference for the iterations of the identification procedure, the probability distributions in (2) and (3) can be computed recursively. At the $(k+1)$-th iteration the available knowledge is expressed as a prior $p\big(\theta\big) = \mathcal{N}(\mu_k, R_k)$ and in combination with data sets $\mathbf{u}$, $\tilde{\mathbf{y}}$ the distributions of interest are computed as

$$p\big(\tilde{\mathbf{y}} \mid \theta, \mathbf{u}\big) = \mathcal{N}\big(\Phi^T(\mathbf{u})\theta, I\sigma_e^2\big), \tag{8a}$$

$$\begin{aligned} p\big(\tilde{\mathbf{y}} \mid \mathbf{u}\big) &= \mathcal{N}\big(\Phi^T(\mathbf{u})\mu_k, R_{\tilde{\mathbf{y}}}\big), \\ R_{\tilde{\mathbf{y}}} &= \left[\sigma_e^2 + \Phi^T(\mathbf{u})R_k\Phi(\mathbf{u})\right], \end{aligned} \tag{8b}$$

$$\begin{aligned} p\big(\theta \mid \tilde{\mathbf{y}}, \mathbf{u}\big) &= \mathcal{N}\big(\mu_{k+1}, R_{k+1}\big), \\ R_{k+1} &= \left[R_k^{-1} + \sigma_e^{-2}\Phi(\mathbf{u})\Phi^T(\mathbf{u})\right]^{-1}, \\ \mu_{k+1} &= R_{k+1}\left[R_k^{-1}\mu_k + \sigma_e^{-2}\Phi(\mathbf{u})\tilde{\mathbf{y}}\right], \end{aligned} \tag{8c}$$

with $\Phi(\mathbf{u}) = \big[x(1) \ \dots \ x(N_s)\big] \in \mathbb{R}^{n\times N_s}$. In (8a), the distribution over the expected data $\tilde{\mathbf{y}} = \big[y(1) \ \dots \ y(N_s)\big]^T$, conditioned on the parameter θ and the input sequence $\mathbf{u}$, can be computed through the distribution of the measurement noise. Its mean is a linear mapping of the input data to the matrix $\Phi(\mathbf{u})$. Marginalised over the prior distribution, this is the data distribution conditioned on the input alone, as per (8b). The posterior distribution $p\big(\theta \mid \tilde{\mathbf{y}}, \mathbf{u}\big)$ (8c) provides an expression for (2), and corresponds to the prior distribution for the $(k+2)$-th iteration.

Bayesian Θ_ψ-Optimal Experiment: We solve approximatively the optimisation problem related to experiment design via an empirical approximation of the objective function and an input parameterisation.

Consider the experiment design problem $\max_{\mathbf{u}\in\mathcal{E}} \mathbf{E}[J(\tilde{\mathbf{y}},\mathbf{u})|\tilde{\mathbf{y}}\sim p\big(\tilde{\mathbf{y}}|\mathbf{u}\big)]$ with $J(\tilde{\mathbf{y}},\mathbf{u})$ the expected utility as given in (6). Note that the posterior distribution $p\big(\theta|\tilde{\mathbf{y}},\mathbf{u}\big) = \mathcal{N}\big(\mu_{k+1}, R_{k+1}\big)$, hence $J(\tilde{\mathbf{y}},\mathbf{u})$ depends on the measurements $\tilde{\mathbf{y}}$ only through μ_{k+1}, as in (8c). It follows that the optimisation problem can be written as an expected value over μ_{k+1} (instead of $\tilde{\mathbf{y}}$), reducing the complexity of the problem from the horizon of the data to the dimensionality of the parameterisation,

$$\begin{aligned}\max_{\mathbf{u}\in\mathcal{E}} \quad & \int_{\Theta} \max\left\{\mathbf{P}\big(\Theta_\psi|\mu_{k+1},\mathbf{u}\big), \mathbf{P}\big(\bar{\Theta}_\psi|\mu_{k+1},\mathbf{u}\big)\right\} \\ & \qquad \times p\big(\mu_{k+1}|\mathbf{u}\big) d\mu_{k+1}, \qquad \text{with } \bar{\Theta}_\psi = \Theta \setminus \Theta_\psi \qquad (9) \\ \text{s.t.} \quad & p\big(\mu_{k+1}|\mathbf{u}\big) = \mathcal{N}\big(\mu_k, R_k - R_{k+1}\big).\end{aligned}$$

As an affine transformation of the measurements $\tilde{\mathbf{y}}$, the posterior mean μ_{k+1} is a random variable with a Gaussian distribution as $\mu_{k+1} = R_{k+1}\left[R_k^{-1}\mu_k + \sigma_e^{-2}\Phi(\mathbf{u})\tilde{\mathbf{y}}\right]$. Using this mean, a practical lower approximation of the maximisation inside the integral is found as $\mathbf{P}\big(\Theta_\psi|\tilde{\mathbf{y}},\mathbf{u}\big) = \int_{\Theta_\psi} p\big(\theta|\mu_{k+1},\mathbf{u}\big)d\theta$ for $\mu_{k+1}\in\Theta_\psi$, and $1-\mathbf{P}\big(\Theta_\psi|\tilde{\mathbf{y}},\mathbf{u}\big)$ else. This provides a relaxed version of (9), expressed as

$$\begin{aligned}\max_{\mathbf{u}\in\mathcal{E}} \quad & \int_{\Theta_\psi}\int_{\Theta_\psi} p\big(\theta|\mu_{k+1},\mathbf{u}\big)p\big(\mu_{k+1}|\mathbf{u}\big)d\theta d\mu_{k+1} \\ & \qquad + \int_{\bar{\Theta}_\psi}\int_{\bar{\Theta}_\psi} p\big(\theta|\mu_{k+1},\mathbf{u}\big)p\big(\mu_{k+1}|\mathbf{u}\big)d\theta d\mu_{k+1}. \qquad (10)\end{aligned}$$

The combined distribution of θ and μ_{k+1}, denoted by variable $\underline{\theta} = \begin{bmatrix}\theta^T & \mu_{k+1}^T\end{bmatrix}^T$, has a normal distribution $p\big(\underline{\theta} \mid \mathbf{u}\big) = \mathcal{N}\big(\mu_{\underline{\theta}}, R\big)$, with mean $\mu_{\underline{\theta}}^T = \begin{bmatrix}\mu_k & \mu_k\end{bmatrix}^T$ and covariance matrix

$$R = \begin{bmatrix} R_k & (R_k - R_{k+1}) \\ (R_k - R_{k+1}) & (R_k - R_{k+1})\end{bmatrix}.$$

Since the integral in (10) cannot in general be computed analytically, we can either compute it with an efficient numerical method or we can empirically approximate it.

Remark 2 (Numerical methods): Efficient numerical methods to compute the integral of multivariate densities over polytopes [22] are not an option. This is because these methods would approximate, depending on the mean of the prior, either the first term in (10) (the integral over $\Theta_\psi \times \Theta_\psi$) and neglect the second term, or the opposite. From a practical point of view this would make sense, since when the prior μ_k is in Θ_ψ we would expect the value of $\int_{\bar{\Theta}_\psi}\int_{\bar{\Theta}_\psi} p\big(\theta|\mu_{k+1},\mathbf{u}\big)p\big(\mu_{k+1}|\mathbf{u}\big)d\theta d\mu_{k+1}$ to be very small. But it can be observed that in the case of $\mu_k \in \Theta_\psi$ the second term is strictly increasing for a decrease in the variance of the posterior density $p\big(\theta \mid \tilde{\mathbf{y}}\big)$, whereas the first term is not. In conclusion, we opt for the alternative use of a Monte-Carlo approximation of the objective. ∎

Let ϵ be a dummy random variable with density distribution $\mathcal{N}\big(\underline{0}, I\big)$, which is independent of the decision variable $\mathbf{u}$. The value of the relaxed optimisation problem (10) can be approximated as

$$\hat{\mathbf{E}}J \approx \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}_{(\Theta_\psi\times\Theta_\psi)\cup(\bar{\Theta}_\psi\times\bar{\Theta}_\psi)}\big(\mu_{\underline{\theta}} + \Lambda\epsilon_i\big), \qquad (11)$$

with M realisations of $\epsilon_i \sim \mathcal{N}\big(\underline{0}, I\big)$ and $\Lambda\Lambda^T = R$. The realisations of $\mu_{\underline{\theta}} + \Lambda\epsilon_i$ have the same density distribution as $\underline{\theta} \sim \mathcal{N}\big(\mu_{\underline{\theta}}, R\big)$. Hence, for a given input $\mathbf{u}$, (11) is an unbiased estimate of (10) and it is also consistent, i.e., for $M \to \infty$ the estimated objective converges to the optimisation objective in (10) with probability 1. The N_s decision variables of $\mathbf{u}$ can be reduced by a parameterisation of the input signal $\mathbf{u}$ as $u(t) = \sum_{k=1}^{n_p}\beta_k \sin(\omega_k t + \alpha_k)$, with parameters $\alpha_k \in [0, 2\pi]$ and $\beta_k \in [0,\infty)$ for $k = 1,\ldots,n_p$ at predefined frequencies ω_k.

Case Study – Bounded-time Safety Verification : Consider a system $\mathbf{S}$ with input signals with support $u(t) \in \mathcal{U} = [-0.2,\ 0.2]$. For simplicity let us select a fixed initial state $x_0 = \begin{bmatrix}0 & 0\end{bmatrix}^T$. Verify whether the output $y_0(t)$ remains within the interval $\mathcal{I} = \big[-0.5,\ 0.5\big]$, labeled as ι, for the next 4 time steps. Introduce accordingly the alphabet $\Sigma = \{\iota, \tau\}$ and the labelling map $L : L(y) = \iota, \forall y \in \mathcal{I}$, $L(y) = \tau, \forall y \in \mathbb{Y}\setminus\mathcal{I}$. Now check whether the following finite-horizon LTL property holds: $\mathbf{S} \vDash \bigwedge_{i=1}^{4}(\bigcirc)^i\iota$. We assume that system $\mathbf{S}$ can be represented as an element of a model set $\mathcal{G}$ with transfer functions characterised by second-order Laguerre-basis functions [21] (a special case of orthonormal basis functions), which translates to the following parameterised state-space representation:

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

The coefficient a is chosen to be $a = 0.4$. We further consider, as prior available knowledge on the system, a distribution $p(\theta) = \mathcal{N}\big(\mu_k, R_k\big)$ on the model class, and a known variance $\sigma_e^2 = 0.5$ for the white additive measurement noise. The output of the computation of the feasible set is in Figure 3. We are interested in an experiment of length $N_s = 100$ with the input parameterised as a multi-sine with frequencies $(\omega_0, 2\omega_0, \ldots, 5\omega_0)$ and fundamental frequency $\omega_0 = 2\pi/10$.

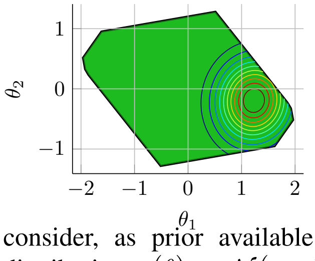

Fig. 3: The green region in the parameter space $[\theta_1\ \theta_2]^T$ is the feasible set of the case study. The contour lines give the density function of a possible a-posteriori distribution over the parameter space (for confidence quantification).

The results of the experiment design problem are given in

Figure 4 and compared to the results for a classical D-optimal experiment design [23] on a single iteration of the verification algorithm in Figure 2. For D-optimality we minimise the determinant of R_{k+1}. Both the Θ_ψ- and D-optimal experiment designs have been performed for priors with several different mean values μ_k, and with a fixed variance of $R_k = 0.2 I_{2\times 2}$. Note that the D-optimal experiment is independent of the prior mean.

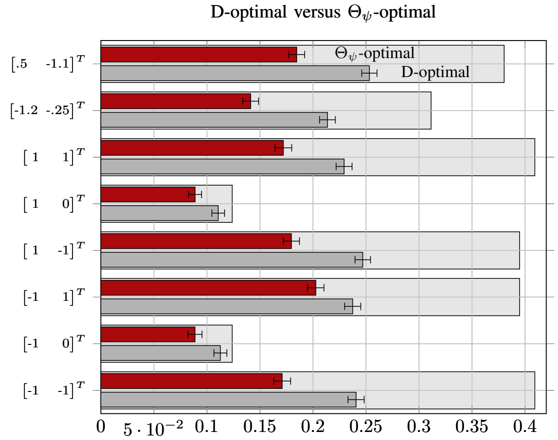

Fig. 4: Empirical evaluation of $1-\mathbf{E}[J]$ on the horizontal axis for the safety verification case study for both the Θ_ψ-optimal (red bar, top) and the D-optimal experiment design (grey bar, below). The wider light grey bar gives $1-\max\{\mathbf{P}(\Theta_\psi), \mathbf{P}(\bar{\Theta}_\psi)\}$. The means of the priors in the experiment design are given on the vertical axis. On the bars for the empirical evaluation of $1-\mathbf{E}[J]$, their standard deviation is also drawn on the bars by the symbol $\vdash\!\dashv$.

After designing an experiment $\mathbf{u}$, the optimisation objective $\mathbf{E}\left[J(\tilde{\mathbf{y}}, \mathbf{u}) \mid \tilde{\mathbf{y}} \sim p(\tilde{\mathbf{y}}|\mathbf{u})\right]$ is evaluated empirically. For this 400 data samples $\tilde{\mathbf{y}}$ are drawn from the distribution $p(\tilde{\mathbf{y}}|\mathbf{u})$, first drawing a parameterisation from the prior distribution $\theta \sim p(\theta)$, and subsequently performing an identification experiment. In Figure 4 the empirical evaluation of $\mathbf{E}\left[J(\tilde{\mathbf{y}}, \mathbf{u})\right]$ for both the Θ_ψ- and D- optimal experiment designs are plotted together with the attainable result without performing additional experiments $\max\{\mathbf{P}(\Theta_\psi), \mathbf{P}(\bar{\Theta}_\psi)\}$. Note that the figure displays in fact the values $1-\mathbf{E}J$ and $1-\max\{\mathbf{P}(\Theta_\psi), \mathbf{P}(\bar{\Theta}_\psi)\}$ for convenience, and also give the standard deviation of the empirical evaluations.

In Figure 4, the result shows that the empirical value of $\mathbf{E}[J]$ is higher for the Θ_ψ-optimal experiment design than for the D-optimal experiment design for all the given mean values. It can be observed that this is especially significant when $\max\{\mathbf{P}(\Theta_\psi), \mathbf{P}(\bar{\Theta}_\psi)\}$ is smaller. The authors have observed that in this case the posterior variances of the Θ_ψ-optimal experiment design tend to align with the closest faces of the feasible region. For mean values that lie farther from the boundaries of the feasible region such as $[1 \;\; 0]^T$ and $[-1 \;\; 0]^T$, the $\max\{\mathbf{P}(\Theta_\psi), \mathbf{P}(\bar{\Theta}_\psi)\}$ is already quite big and the difference between the Θ_ψ- and D- optimal design is less significant. It can be concluded that the Θ_ψ-optimal experiment gives a significant improvement with respect to $\mathbf{E}[J]$ in comparison to the D-optimal solution.

IV. Future Work

The present work relies on the underlying knowledge of the exact model structure (limited to linearly parameterised models) and of the noise dynamics. It is of interest to extend it to more general model structures and to consider more complex linear-time properties. Future work will employ this theory as a practical tool for property optimisation via controller synthesis.

References

[1] E. M. Clarke, "The birth of model checking," in *25 Years of Model Checking*. Springer, 2008, pp. 1–26.

[2] 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*, vol. 14, no. 1, pp. 61–70, 2007.

[3] B. Yordanov, G. Batt, and C. Belta, "Model checking discrete-time piecewise affine systems: Application to gene networks," in *European Control Conference*, 2007.

[4] H. Hjalmarsson, "From experiment design to closed-loop control," *Automatica*, vol. 41, no. 3, pp. 393–438, 2005.

[5] ——, "System identification of complex and structured systems," *European journal of control*, vol. 15, no. 3, pp. 275–310, 2009.

[6] C. R. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer, "Robust optimal experiment design for system identification," *Automatica*, vol. 43, no. 6, pp. 993–1008, 2007.

[7] D. V. Lindley, "The philosophy of statistics," *Journal of the Royal Statistical Society: Series D*, vol. 49, no. 3, pp. 293–337, 2000.

[8] H. L. S. Younes and R. G. Simmons, "Probabilistic verification of discrete event systems using acceptance sampling," in *Computer Aided Verification*. Springer, 2002, pp. 223–235.

[9] K. Sen, M. Viswanathan, and G. Agha, "Statistical model checking of black-box probabilistic systems," in *Computer Aided Verification*, ser. LNCS, R. Alur and D. Peled, Eds. Springer, 2004, pp. 399–401.

[10] D. Henriques, J. G. Martins, P. Zuliani, A. Platzer, and E. M. Clarke, "Statistical model checking for Markov decision processes," in *Quantitative Evaluation of Systems*. IEEE, 2012, pp. 84–93.

[11] A. Legay, S. Sedwards, and L. Traonouez, "Lightweight verification of Markov decision processes with rewards," *CoRR*, 2014. [Online]. Available: http://arxiv.org/abs/1410.5782

[12] K. Sen, M. Viswanathan, and G. Agha, "Learning continuous time Markov chains from sample executions," in *Quantitative Evaluation of Systems*. IEEE, 2004, pp. 146–155.

[13] Y. Chen and T. D. Nielsen, "Active learning of Markov decision processes for system verification," *Conf. on Machine Learning and Applications*, vol. 2, pp. 289–294, 2012.

[14] L. Bortolussi and G. Sanguinetti, "Learning and designing stochastic processes from logical constraints," in *Quantitative Evaluation of Systems*, ser. LNCS. Springer, 2013, pp. 89–105.

[15] P. Tabuada and G. J. Pappas, "Model checking LTL over controllable linear systems is decidable," *Hybrid Systems: Computation and Control*, vol. 2623, pp. 498–513, 2003.

[16] L. Bortolussi and G. Sanguinetti, "Smoothed model checking for uncertain continuous time Markov chains," *arXiv preprint arXiv:1402.1450*, 2014.

[17] V. Peterka, "Bayesian approach to system identification," *Trends and Progress in System identification*, pp. 239–304, 1981.

[18] G. Frehse, S. K. Jha, and B. H. Krogh, "A counterexample-guided approach to parameter synthesis for linear hybrid automata," in *Hybrid Systems: Computation and Control*, 2008, vol. 4981, pp. 187–200.

[19] T. A. Henzinger and H. Wong-Toi, *Using HyTech to synthesize control parameters for a steam boiler*, ser. LNCS, J.-R. Abrial, E. Börger, and H. Langmaack, Eds. Springer Berlin Heidelberg, 1996.

[20] K. S. Shanmugan and A. M. Breipohl, *Random Signals: Detection, Estimation and Data Analysis*. Wiley, 1988.

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

[22] L. Blackmore and M. Ono, "Convex chance constrained predictive control without sampling," in *Proceedings of the AIAA Guidance, Navigation and Control Conference*, 2009, pp. 7–21.

[23] M. Gevers, X. Bombois, R. Hildebrand, and G. Solari, "Optimal experiment design for open and closed-loop system identification," *Communications in Information and Systems*, vol. 11, no. 3, p. 197, 2011.