2016 European Control Conference
June 29 - July 1, 2016. Aalborg, Denmark

# Experiment Design for Formal Verification via Stochastic Optimal Control

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. Grey-box systems, which are specified as a model class, are subject to identification experiments that enable accepting or rejecting system properties expressed as formulae in a linear-time logic with a given confidence. We employ a Bayesian framework for the computation of the confidence level and for the design of experiments to increase the confidence. The experiment design is formulated as a stochastic optimal control problem, which solvable via dynamic programming. Applied to linear control systems, this work enables efficient data-driven verification of partly-known 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 approach.**

## I. Introduction

The area of system identification [1] investigates measurement-based model construction of physical systems, and deals with models characterised via noisy input-output measurements, with dynamics evolving over uncountable (continuous) state spaces. Within this area, a relevant research direction [2], [3], [4] has focused on the development of approaches towards identification for control, where the model building aspects are integrated around a cost function that needs to be optimised over.

In this work we are interested in the use of these techniques to verify or falsify system properties, such as safety or reachability requirements on the dynamics: these properties can be naturally formulated as specifications in a given temporal logic [5]. This problem has been first studied in [6]. Towards this goal, the input signals exciting the system should be chosen to maximise the amount of information gained. An optimal input typically depends on the knowledge of the true system, and the literature distinguishes three approaches to input design: 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. While the first approach is predominant [2], [4], some work has been done on the robust experiment design using the min-max approach [7]. On the other hand the third approach, well known from Bayesian statistics [8], is not yet widely employed.

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.

This work is supported by the European Commission IAPP project AMBI 324432, and by the John Fell OUP Research Fund. We acknowledge the support of NWO and of the Dutch Institute of Systems and Control.

In this work we embrace a Bayesian experiment design formulation for the verification of quantitative properties [6]. We will focus on linearly parameterised models classes, and linear-time properties that map into convex sets in the parameter space [9]. We show that the problem of experiment design can be reformulated as a stochastic optimal control problem over a Markov decision process (MDP) [10].

The work distinguishes itself from the standard experiment design problems for estimation, in that its main goal is not to estimate an optimal parameter. Instead, the overall goal of this work is to verify or falsify a property of the underlying system of interest. As such, there is only an indirect need to minimise the accuracy (and especially the variance) of the estimate. The idea of reformulating the experiment design problem as an MDP optimisation originates from [11], which is newly enhanced by embedding the posterior distribution of the model parameters into the state of the MDP. This extended MDP allows for an input design that depends on the collection of past data, since the state of the MDP encompasses the collected data via the updated posterior distribution. As such, in this work we extend previous results in [6] on Bayesian experiment design problem by synthesising a state-based policy, rather than a state-independent (open-loop) one as in [6]. The iterative design of experiments based on available knowledge has been explored in [12]: with focus on parameter estimation, the goal is expressed via the eigenvalues of the covariance matrix, and the input is designed via a receding horizon optimisation based on the nominal estimate of the system, which is updated in time during the experiment.

The article is structured as follows. Section II provides the background, based on the work in [6], [9], and introduces the goal of experiment design. Section III reframes this goal as a stochastic optimisation problem over an MDP, which is solved via dynamic programming. Section IV discusses a case study intended to elucidate the solution of the problem.

## II. Data-driven and model-based verification

The overall goal of this work can be stated as follows: *starting from available a-priori knowledge over system* $\mathbf{S}$*, iteratively and efficiently gather measurements until a specification* $\psi$ *defined over the system is verified or falsified with a given confidence* $1-\delta$.

We discuss the problem setup, first introduced in [6], [9], in the remainder of this section.

### A. System and model class

The system, denoted by $\mathbf{S}$ as in Figure 1, is measured in discrete time. An input signal $u(t), t \in \mathbb{N}$, captures how the

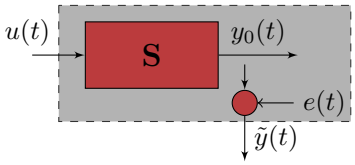


Fig. 1: System $\mathbf{S}$ has input $u(t)$ and output $y_0(t)$. In the measurement setup, the measured output $\tilde{y}(t)$ includes the system output $y_0(t)$ and the measurement noise $e(t)$.

environment acts on the system. Similarly, the output $y_0(t)$ 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 $e(t)$.

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 with 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 $\mathbf{S}$. The uncertainty about $\mathbf{M}(\theta^0)$ is structured as a distribution over the parameter set $\Theta$. It is then the (unknown) model denoted by $\mathbf{M}(\theta^0) = \mathbf{S}$ that we would ideally like to formally model-check.

It is possible to collect data of the system by exciting it with an input sequence $\mathbf{u}_{N_s} = \begin{bmatrix} u(0) & u(1) & \dots & u(N_s-1) \end{bmatrix}^T$, with $N_s$ the length of the input sequence. Noisy observations $\tilde{y}(t)$ of the output $y_0(t)$ are classically perturbed by 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 $\mathbf{u}_{N_s}$ as $\tilde{\mathbf{y}}_{N_s} = \begin{bmatrix} \tilde{y}(1) & \tilde{y}(2) & \dots & \tilde{y}(N_s) \end{bmatrix}^T$. The collected input-output data contains statistical information on the behaviour of the system, and allows to refine the uncertainty distribution over the parameter space, as discussed in the second part of this section.

### B. Properties and confidence of satisfaction

Let us define $\Theta_\psi \subseteq \Theta$ to be the maximal feasible set of parameters, such that for every parameter in that set the property $\psi$ holds, i.e. $\forall \theta \in \Theta_\psi : \mathbf{M}(\theta) \models \psi$ and $\forall \theta \notin \Theta_\psi : \mathbf{M}(\theta) \not\models \psi$. The formula $\mathbf{M}(\theta) \models \psi$ reads as "the model satisfies property $\psi$". We are interested in temporal properties that, given a set of parameterised models $\mathbf{M}(\theta)$, translate to polytopic sets of feasible parameters (cf. Fig. 2): this is the case of specifications, for instance certain safety requirements, expressed in a fragment of linear temporal logic [9]. The work in [9] discusses how to synthesise set $\Theta_\psi$ for parameterised LTI models and specific sets of linear-time properties.

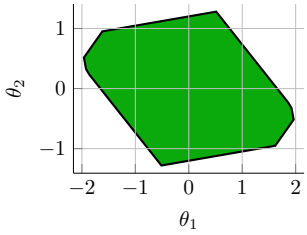


Fig. 2: Example of a feasible set $\Theta_\psi$ in a two-dimensional $\theta = (\theta_1, \theta_2)$ parameter space, obtained by synthesising parameters $\theta$ that are such that $\mathbf{M}(\theta) \models \psi$ [9].

The confidence (or credibility in Bayesian literature) in a specification $\psi$ defined over system $\mathbf{S}$ is computed based on the uncertainty distribution over $\Theta_\psi$: 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 $p(\theta | \tilde{\mathbf{y}}_{N_s}, \mathbf{u}_{N_s})$ can be used to compute the confidence as

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

According to the Bayesian probability calculus [8], the confidence of a property becomes the measure of the uncertainty distribution. Given a prior distribution $p(\theta)$ and a data set $\tilde{\mathbf{y}}_{N_s}$ obtained by taking $N_s$ measurements of $\tilde{y}(t)$, *the a-posteriori uncertainty distribution* $p(\theta | \tilde{\mathbf{y}}_t, \mathbf{u}_t)$ is based on parametric inference [8], [13] structured over the parameter set $\Theta$ as

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

### C. Experiment design for formal verification with confidence

As discussed earlier, we are interested in generating system data to refine the uncertainty distribution over the parameter space and to increase the confidence in the satisfaction (or lack thereof) of a property. We assume that the identification experiment is started at $t = 0$, and is stopped at a given time $t$ when we have a sufficient confidence over the statement $\mathbf{S} \models \psi$, namely $\mathbf{P}(\Theta_\psi \mid \tilde{\mathbf{y}}_t, \mathbf{u}_t) > 1 - \delta$, or over its complement $\mathbf{S} \not\models \psi$, namely $1 - \mathbf{P}(\Theta_\psi \mid \tilde{\mathbf{y}}_t, \mathbf{u}_t) > 1 - \delta$. We want to terminate the experiment as soon as possible: to this end we can choose the inputs $u(t)$ based on a model of the system, on $\tilde{\mathbf{y}}_t$ and $\mathbf{u}_t$, and within the set of allowed experiment inputs $u(t) \in \mathcal{E}$ (for example $\mathcal{E} := \{u \in [-u_{max}, u_{max}]\}$).

## III. Bayesian experiment design: MDP Formulation and Optimisation

Let us consider linearly-parameterised models, such as models parameterised with orthonormal basis functions. This model class is able to represent a wide set of physical systems [14, Chapter 4 and 7]. Consider models within a

linearly-parameterised model class $\mathcal{G}$ to have the following state-space realisation

$$\mathbf{M}(\theta): \quad \begin{cases} x(t+1) & = Ax(t) + Bu(t), \\ \hat{y}(t,\theta) & = \theta^T x(t), \end{cases} \tag{3}$$

which is linearly parameterised by $\theta = \begin{bmatrix}\theta_1 \dots \theta_n\end{bmatrix}^T \in \Theta \subset \mathbb{R}^n$. We assume that system $\mathbf{S}$ has a representation $\mathbf{M}(\theta^0)$ in this model set, with unknown parameter $\theta^0$, and has an output denoted as $y_0(t) = \hat{y}(t,\theta^0)$. Without loss of generality, 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. This is a common assumption for verification procedures and identification experiments, and can be relaxed to any known $x(0) \in \mathbb{R}^n$ or to any probability distribution for $x(0)$. As mentioned earlier, we assume that the measurements are perturbed by an additive zero-mean, white, Gaussian-distributed measurement noise with variance $\sigma_e^2$, i.e. $\mathcal{N}(0,\sigma_e^2)$, which is uncorrelated with the input.

### A. MDP formulation

*Definition 1 (General Markov process):* A discrete-time MDP, denoted as $\Sigma = (\mathbb{X}, \mathbb{U}, T_x)$, is comprised of:

- a continuous (uncountable) state space $\mathbb{X} \subset \mathbb{R}^n$;
- an action space $\mathbb{U}$ consisting of a possibly uncountable number of actions, which we equate with $\mathcal{E}$;
- a Borel-measurable stochastic kernel $T_x$, which assigns to each state-action pair $x \in \mathbb{X}$ and $a \in \mathbb{U}$ a probability distribution $T_x(\cdot \mid x, a)$ over $\mathbb{X}$. □

For a given parameter $\theta$, model $\mathbf{M}(\theta)$ in (3) can be regarded as a discrete-time Markov process, expressed at time $k$ with a deterministic transition corresponding to a Dirac distribution (a point distribution), namely $T_x(dx' \mid x, u) = \delta_{Ax+Bu}(dx')$. Of interest to us is a Markov representation of the Bayesian inference procedure, where the posterior distributions in (2) can be computed recursively. More precisely, a prior $p(\theta) = \mathcal{N}(\mu, R)$ is updated via Bayesian inference from system-drawn data $\{\mathbf{u}_t, \tilde{\mathbf{y}}_t\}$ up to time $t$. Both the resulting posterior probability distribution $p(\theta|\mathbf{u}_t, \tilde{\mathbf{y}}_t) = \mathcal{N}(\mu^+, R^+)$, the data realisation $\tilde{\mathbf{y}}_t$, and the unknown true parameter $\theta$ can be described by random variables with Gaussian distributions given as

$$p(\tilde{\mathbf{y}}_t \mid \theta, \mathbf{u}_t) = \mathcal{N}\big(\Phi^T(\mathbf{u}_t)\theta, I\sigma_e^2\big), \tag{4a}$$

$$p(\tilde{\mathbf{y}}_t \mid \mathbf{u}_t) = \mathcal{N}\big(\Phi^T(\mathbf{u}_t)\mu, R_{\tilde{\mathbf{y}}_t}\big), \tag{4b}$$

$$R_{\tilde{\mathbf{y}}_t} = \left[\sigma_e^2 I + \Phi^T(\mathbf{u}_t) R \Phi(\mathbf{u}_t)\right],$$

$$p(\theta \mid \tilde{\mathbf{y}}_t, \mathbf{u}_t) = \mathcal{N}\big(\mu^+, R^+\big), \tag{4c}$$

$$R^+ = \left[R^{-1} + \sigma_e^{-2}\Phi(\mathbf{u}_t)\Phi^T(\mathbf{u}_t)\right]^{-1}, \tag{4d}$$

$$\mu^+ = R^+\left[R^{-1}\mu + \sigma_e^{-2}\Phi(\mathbf{u}_t)\tilde{\mathbf{y}}_t\right],$$

$$p(\mu^+|\theta, \mathbf{u}_t) = \mathcal{N}\big(\mu, R - R^+\big), \tag{4e}$$

with $\Phi(\mathbf{u}_t) = \begin{bmatrix}x(1) \ \dots \ x(t)\end{bmatrix} \in \mathbb{R}^{n\times t}$. In (4a), the distribution over the expected data $\tilde{\mathbf{y}}_t = \begin{bmatrix}\tilde{y}(1) \ \dots \ \tilde{y}(t))\end{bmatrix}^T$, conditioned on the parameter $\theta$ and the input sequence $\mathbf{u}_t$, can be computed from the distribution of the measurement noise. Its mean is a linear mapping of the input data to the matrix $\Phi(\mathbf{u}_t)$. Marginalised over the prior distribution, this is the data distribution conditioned on the input alone, as per (4b). The posterior distribution $p\big(\theta \mid \tilde{\mathbf{y}}_t, \mathbf{u}_t\big)$ in (4c) provides an expression for (2), and yields the prior distribution for the next iteration of the procedure.

Employing Gaussian distributions we can rewrite the iterative data collection above via data sets of length one as an MDP. More precisely, let the MDP be defined by the following stochastic transitions

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \\ \mu(t+1) &= \mu(t) + R(t)x(t+1)v(t), \\ R(t+1) &= R(t) - R(t)x(t+1)\Sigma(t)x(t+1)^T R(t), \\ v(t) &\sim \mathcal{N}\big(0, \Sigma(t)\big), \\ \Sigma(t) &= (\sigma_e^2 + x(t+1)^T R(t) x(t+1))^{-1}. \end{aligned}$$

We refer to this model as $\Sigma_\theta$ (as opposed to $\mathbf{M}(\theta)$ in (3)) for later use. Denote by $\mathbb{S}^n$ the set of real symmetric matrices $M = M^T \in \mathbb{R}^{n\times n}$. Since the covariance $R \in \mathbb{S}^n$, it can be uniquely defined by its upper triangular elements, denoted by $r \in \mathbb{R}^{(n+1)n/2}$. This means that there exists a one-to-one mapping from the variances in matrix $R \in \mathbb{S}^n$ to points $r \in \mathbb{R}^{(n+1)n/2}$, that is $f_R : \mathbb{R}^{(n+1)n/2} \to \mathbb{S}^n$ and $f_r : \mathbb{S}^n \to \mathbb{R}^{(n+1)n/2}$, where $f_r = f_R^{-1}$. For a given mapping $f_r$, the MDP $\Sigma_\theta$ has as state space $\mathbb{X}_\theta$ with elements $x_\theta = (x, \mu, r) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n(n+1)/2}$ and takes as input signals $u \in \mathbb{U}$.

Given a prior distribution $p\big(\theta\big) \sim \mathcal{N}\big(\mu(0), R(0)\big), \mu(0) = \mu, R(0) = R$, and given an initial state for $\mathbf{M}(\theta)$, the initial state of the MDP is given as $x_\theta(0) = (x(0), \mu(0), f_r(R(0)))$. At every time instant the input can be selected as a function of the state $x_\theta$ of the MDP.

*Definition 2 (Markov policy):* A Markov policy $\pi$ over the horizon $[0, N_s]$ is a sequence $\pi = (\pi_0, \pi_1, \dots, \pi_{N_s-1})$ of measurable maps, $\pi_k : \mathbb{X}_\theta \to \mathbb{U}, k = 0, 1, \dots, N_s - 1$, from the state space $\mathbb{X}_\theta$ to the action space $\mathbb{U}$. The set of Markov policies is denoted as $\Pi$. □

*Remark 1:* We refer the reader to [15] for a complete discussion on measurability issues related to Markov decision processes over continuous state and control spaces, and to corresponding optimal control problems. In this work we refer for simplicity to general measurability requirements, as in the previous statement. □

The goal of this work is the design of a Markov policy for $\Sigma_\theta$, to attain an efficient data collection for the verification (with a given confidence level) of a property defined over the unknown system $\mathbf{S}$.

### B. Experiment setup

Suppose we perform an experiment on $\mathbf{S}$ with a given policy $\pi \in \Pi$. At the start of the experiment, the MDP $\Sigma_\theta$ is initiated as $x_\theta(0) = (x(0), \mu(0), f_r(R_0))$, where $x(0)$ is the initial state of $\mathbf{S}$ and where the normal distribution $\mathcal{N}(\mu_0, R_0)$ represents the prior uncertainty distribution of $\theta_0$. At $t = 0$ the control input $u(0)$ is selected based on the policy

$u(0) = \pi_0(x_\theta(0))$. The subsequent transition from $x_\theta(0)$ to $x_\theta(1)$ is governed by deterministic transitions for $x(0)$ to $x(1)$ and for $r(0)$ to $r(1)$, and by a stochastic transition for $\mu(0)$ to $\mu(1)$ obtained from the measured output $\tilde{y}(1)$ as in (4c). At every subsequent time instant the input is chosen based on the current state and the stochastic part of the transition is obtained as a function of the measured output drawn from $\mathbf{S}$. For the latter, the MDP $\Sigma_\theta$ gives the uncertainty distribution of this transition based on the current state of the MDP $x_\theta(t) = (x(t), \mu(t), r(t))$. Remember that the current state of the system represents the collection of past measurements $\tilde{\mathbf{y}}_t$.

### C. Experiment design via MDP optimisation

The experiment is successfully completed at time $t \in \mathbb{N}$ when $\theta_0 \in \Theta_\psi$ or $\theta_0 \in \Theta \setminus \Theta_\psi$, with associated confidence of at least $1-\delta$, and where $\theta_0 \sim \mathcal{N}(\mu(t), R(t))$, with $R(t) = f_R(r(t))$. Denote set $K \subset \mathbb{X}_\theta$ as the set of states $(x, \mu, r)$ associated with the required confidence on $\theta_0 \sim \mathcal{N}(\mu, f_R(r))$. Hence a given state trajectory $\{x_\theta(t) \,|\, 0, 1, \ldots, N_s\}$ represents a successful experiment if it reaches the target set $K$. This property can be expressed as

$$\exists j \in [0, N_s] : x_\theta(j) \in K.$$

If in addition the state of the MDP is required to stay within a given safe set $A$, then the success of an experiment within the finite horizon becomes equivalent to a finite-horizon reach-avoid (or constrained reachability) property [5] over a safe set $A$ and target set $K$. This can be expressed as

$$\exists j \in [0, N_s] : x_\theta(j) \in K \wedge \forall i \in [0, j-1] : \, x_\theta(i) \in A \setminus K.$$

Specifications on the MDP that can be encoded as safety ones include all the requirements on the safe operating range of the system $\mathbf{S}$.

The probability associated to this event can be characterised as a boolean expression using indicator functions [15], [16], which leads to an expectation over the state trajectories as

$$r^\pi_{x_\theta(0)}(K, A) = \mathbb{E}^\pi_{x_\theta(0)}\left[\sum_{j \in [0, N_s]} \mathbf{1}_K(x_\theta(j)) \prod_{i=0}^{j-1} \mathbf{1}_{A \setminus K}(x_\theta(i))\right],$$

where $\mathbf{1}_B(x) = 1$ if $x \in B$, and otherwise it is equal to 0.

To attain short experiments, let us *penalise* the time it takes to achieve the required confidence. Whilst above we have attached a value of 1 to each trajectory reaching the target set, let us instead consider a discount factor $\gamma \in (0, 1)$ attached to each time step that $x_\theta$ stays in $A \setminus K$ before reaching $K$. Denoting $\mathbf{1}^\gamma_B(x) = \gamma \mathbf{1}_B(x)$, the expression for the experiment design objective becomes

$$s^\pi_{x_\theta(0)}(K, A) = \mathbb{E}^\pi_{x_\theta(0)}\left[\sum_{j \in [0, N_s]} \mathbf{1}_K(x_\theta(j)) \prod_{i=0}^{j-1} \mathbf{1}^\gamma_{A \setminus K}(x_\theta(i))\right].$$

In the sequel we refer to $s^\pi_{x_\theta(0)}(K, A)$ as the *discounted reach-avoid property*. Notice that unlike the reach-avoid probability $r^\pi_{x_\theta(0)}(K, A)$, the discounted equivalent $s^\pi_{x_\theta(0)}(K, A)$ is not induced by a probability over the trajectories of $\Sigma_\theta$. Still the discounted quantity $s^\pi_{x_\theta(0)}(K, A)$ over the traces of $\Sigma_\theta$ can be written as a reach-avoid probability $r^\pi_{\bar{x}_\theta(0)}(K, A)$ for an extended MDP $\bar{\Sigma}_\theta$, which includes a transition probability of $(1-\gamma)$ to model the possibility that the experiment is terminated preemptively. $\bar{\Sigma}_\theta$ extends the state space of $\Sigma_\theta$ with two discrete modes: $q_1$ for the experiment being *active* and $q_2$ for the experiment being *interrupted preemptively*.

For a given policy $\pi$, the time-dependent value function $\mathbf{W}^\pi_k : \mathbb{X}_\theta \to [0, 1]$, defined as

$$\mathbf{W}^\pi_k(x_\theta) = \mathbb{E}^\pi\left[\sum_{j \in [k+1, N_s]} \mathbf{1}_K(x_\theta(j)) \prod_{i=k+1}^{j-1} \mathbf{1}^\gamma_{A \setminus K}(x_\theta(i)) \,\middle|\, x_\theta(k) = x_\theta\right],$$

is the $\gamma$-*discounted* probability that the state trajectory $\{x_\theta(k+1), \ldots, x_\theta(N_s)\}$, starting from $x_\theta(k)$, will reach the target set $K$ within the time horizon $[k, N_s]$, while staying within the safe set $A$. This function allows expressing the discounted reach-avoid probability backward recursively, as follows.

*Proposition 3:* Given a policy $\pi = (\pi_0, \pi_1, \ldots, \pi_{N_s-1})$, define function $\mathbf{W}^\pi_k : \mathbb{X}_\theta \to [0, 1]$ by backward recursion

$$\mathbf{W}^\pi_k(x_\theta) = \mathbb{E}^{\pi_k}_{x_\theta}\left[\mathbf{1}_K(x^{t+1}_\theta) + \mathbf{1}^\gamma_{A \setminus K}(x^{t+1}_\theta) \mathbf{W}^\pi_{k+1}(x^{t+1}_\theta)\right],$$

with the compact notation $x^{t+1}_\theta \sim T_x(\cdot \,|\, x_\theta, \pi_k(x))$ for $k = N_s - 1, N_s - 2, \ldots, 0$, and initialised with $\mathbf{W}^\pi_{N_s}(x_\theta) = 0$. Then for any initial state $x_\theta(0) \in \mathbb{X}_\theta$, the discounted probabilistic reach-avoid property $s^\pi_{x_\theta(0)}(K, A)$ is

$$s^\pi_{x_\theta(0)}(K, A) = \mathbf{1}_K(x_\theta(0)) + \mathbf{1}^\gamma_{A \setminus K}(x_\theta(0)) \mathbf{W}^\pi_0(x_\theta(0)). \quad \square$$

*Proof:* The proof follows [16, Lemma 4], where the above statement is proven for a value function $V^\pi_k(x) = \mathbf{1}_K(x) + \mathbf{1}_{A \setminus K}(x) W^\pi_k(x)$. To allow for the discounting, extend the state space with two discrete modes as described above. Consider an extended safe set $\bar{A} := \{q_1\} \times A$ and a target set $\bar{K} := \{q_1, q_2\} \times K$. Let the probability of going from $q_1$ to $q_2$ be $1 - \gamma$ for any continuous state in $A \setminus K$. The proof follows [16]. ■

Rather than selecting and fixing a policy $\pi$, as done above, we now focus on the optimal control problem, which seeks the Markov policy $\pi^*$ that maximises the discounted probabilistic reach-avoid property, and which is such that $s^*_{x_\theta(0)}(K, A) = \sup_{\pi \in \Pi} s^\pi_{x_\theta(0)}(K, A)$. This optimal policy can be characterised as follows.

*Proposition 4:* Define functions $\mathbf{W}^*_k : \mathbb{X}_\theta \to [0, 1]$, by the backward recursions

$$\mathbf{W}^*_k(x_\theta) = \sup_{u \in \mathbb{U}} \mathbb{E}^u_{x_\theta}\Big[\mathbf{1}_K(x^{t+1}_\theta) + \mathbf{1}^\gamma_{A \setminus K}(x^{t+1}_\theta) \mathbf{W}^*_{k+1}(x^{t+1}_\theta)\Big],$$

with $x^{t+1}_\theta \sim T_x(\cdot \,|\, x_\theta, u)$ for $k = N_s - 1, N_s - 2, \ldots, 0$, and initialized by $W^*_{N_s}(x_\theta) = 0$. Then for any initial state $x_0 \in \mathbb{X}$ the optimal probabilistic reach-avoid property $s^*_{x_\theta(0)}(K, A)$ can be expressed as

$$s^*_{x_\theta(0)}(K, A) = \mathbf{1}_K(x_\theta(0)) + \mathbf{1}^\gamma_{A \setminus K}(x_\theta(0)) \mathbf{W}^*_0(x_\theta(0)).$$

Furthermore, $\pi_k^* : \mathbb{X}_\theta \to \mathbb{U}$ for $k = N_s - 1, N_s - 2, \ldots, 0$, is such that $\forall x_\theta \in \mathbb{X}_\theta$:

$$\pi_k^*(x_\theta) = \arg\max_{u \in \mathbb{U}} \mathbb{E}^u_{x_\theta}\Big[\mathbf{1}_K(x_\theta^{t+1}) + \mathbf{1}^\gamma_{A\setminus K}(x_\theta^{t+1})\mathbf{W}^*_{k+1}(x_\theta^{t+1})\Big]$$

and $\pi^* = (\pi_0^*, \pi_1^*, \ldots, \pi_{N_s-1}^*)$ is the optimal Markov policy for the discounted probabilistic reach-avoid. □

*Proof:* The proof follows from that for probabilistic reach-avoid properties given in [15] and [16, Theorem 6]. ■

The above proposition shows that there exists an optimal policy for the discounted reach-avoid problem that is deterministic. This implies that for every state $x_\theta$ of $\Sigma_\theta$ the optimal policy delivers a single control input, instead of a probability distribution over the set of possible control actions. This means that the set of deterministic Markov policies $\Pi$ is sufficiently exciting for the experiment design problem.

The computation of $s^*_{x_0}(K, A)$ is based on $N_s$ recursions given in Proposition 4, which can be denoted by a dynamic programming operator $\mathbf{T}$ such that $\mathbf{W}^*_k = \mathbf{T}\mathbf{W}^*_{k+1}$. Therefore the value of the optimal $\gamma$-discounted probabilistic reach-avoid property can be written as the composition of $N_s$ mappings as $s^*_{x_\theta(0)}(K, A) = \mathbf{1}_K(x_\theta(0)) + \mathbf{1}^\gamma_{A\setminus K}(x_\theta(0))\left(\mathbf{T}^{N_s}\mathbf{W}^*_{N_s}\right)(x_\theta(0))$. Let us qualitatively comment on the behaviour of the backwards recursions from $\mathbf{W}^*_{k+1}$ to $\mathbf{W}^*_k$. Employing a $\gamma$-discounting, $\mathbf{T}$ is a contractive mapping, which allows tapping on typical results in stochastic optimal control [10]. As a result of this contractivity property, the mapping $\mathbf{T}^{N_s}\mathbf{W}^*_{N_s}$ will converge, for increasing values of $N_s$, to a unique value function associated to a corresponding infinite horizon problem. Hence, for problems over a long time horizon $N_s$, we expect to obtain a stationary policy [10], leading to the practical use of a time-independent and deterministic policy for the experiment design problem. Computed offline, this policy can then be implemented online during the identification experiment.

### D. Computational aspects

Although we have attained a formal characterisation of the experiment design problem by a stochastic optimal control formulation, the computation of the exact solution is seldom analytical. This is not only because the backwards recursions $\mathbf{T}$ cannot be in general expressed explicitly, but also because the target set $K$ will often not have an analytical expression. Instead, the associated stochastic dynamic programming problem ought to be solved approximately.

This might lead to high computational costs of finding an accurate optimal policy. But these computations can be done offline, before the identification experiment, and approximation errors do not influence the validity of the confidence in a property of the system, they only change the optimality and hence the time that the experiment will take to complete. Approximate solutions can be obtained via numerical procedures or via sample-based algorithms. For a reduction of high computational costs related to the state-space dimensionality, one could consider a policy iteration scheme, which computes the solution to the associated infinite horizon problem [17], or a policy search scheme. Especially methods that rely on local dimensionality reduction to solve the problem approximately represent an area of interest for future work.

## IV. Analysis of the experiment design problem

In this section we will look at a case study with a one-dimensional parameterised model. This allows us to analyse and clarify in depth the stochastic optimal control reformulation of the experiment design problem. This one-dimensional problem leads to an optimisation problem over a three-dimensional MDP.

Consider the parameterised model:

$$\mathbf{M}(\theta) : x(t+1) = \frac{1}{2}x(t) + u(t), \quad y_0(t) = \theta x(t),$$

with measurements taken as

$$\tilde{y}(t) = y_0(t) + e(t), \quad e(t) \sim \mathcal{N}(0, 0.5).$$

Assume that the feasible set of parameters is given as $\Theta_\psi = [-1,\ 1]$ (the details of property $\psi$ not being of interest here). The objective is to design an experiment such that the confidence in the decision whether $\mathbf{S} \vDash \psi$ or $\mathbf{S} \nvDash \psi$ based on the posterior probability distribution is at least $1 - \delta$, with $\delta = 0.01$. Consider a given set of allowed inputs for the experiment, $\mathcal{E} = \{-1, 0, 1\}$, and a maximal experiment time of $N_s = 10$. Since the state transitions of $\mathbf{M}(\theta)$ are strictly stable, no additional requirements are raised on the allowed range of $x(t)$. Hence the safe set is chosen as $A = \mathbb{X}_\theta$. At the start of the experiment the prior uncertainty is given as $\mathcal{N}(\mu_0, R_0)$, i.e. $\mu(0) := \mu_0$ and $R(0) := R_0$, and the system is initialised at $x(0) = x_0$ which, with no loss in generality, is assumed to be $x_0 = 0$.

### A. Standard iterative experiment design

The standard approach to experiment design from the literature [11] would be to synthesise inputs over the whole time horizon $\mathbf{u}_{N_s}$ before performing the experiment, based on an approximation of the posterior $p\big(\theta \mid \tilde{\mathbf{y}}_{N_s}, \mathbf{u}_{N_s}\big)$ chosen as $\mathcal{N}(\mu(0), R(N_s)))$. Note that $R(N_s)$ is the variance obtained after applying $\mathbf{u}_{N_s}$ over the time horizon. The objective to design an experiment such that $\max\{\mathbf{P}\big(\Theta_\psi|\tilde{\mathbf{y}}_{N_s}, \mathbf{u}_{N_s}\big), 1 - \mathbf{P}\big(\Theta_\psi|\tilde{\mathbf{y}}_{N_s}, \mathbf{u}_{N_s}\big)\} \geq 1 - \delta$ would then be approximated as

$$\max\{\mathbf{P}\big(\hat{\theta} \in \Theta_\psi\big), 1 - \mathbf{P}\big(\hat{\theta} \in \Theta_\psi\big)\} \geq 1 - \delta, \quad \text{with } \hat{\theta} \sim \mathcal{N}(\mu(0), R(N_s))).$$

At certain values of $\mu_0$ the used approximation $\max\{\mathbf{P}\big(\hat{\theta} \in \Theta_\psi\big), 1 - \mathbf{P}\big(\hat{\theta} \in \Theta_\psi\big)\}$ decreases monotonically for decreasing values of $R(N_s)$. This behaviour can especially be observed at the edges of the feasible set $\Theta_\psi$, that is for $\mu_0 = 1$ or $\mu_0 = -1$, but it also affects $\mu_0$ close to these edges. The decrease in maximal confidence for $\mu_0 \in \{-1, 1\}$ can be explained as follows. When starting the experiment the confidence in $\mathbf{S} \nvDash \psi$ is strictly larger than $0.5$. Since $\mu_0$ is located on the edge of $\Theta_\psi$, decreasing the variance gives a confidence in the feasible set that tends to $0.5$. Hence for a decreasing variance, the maximal confidence tends to $0.5$.

Considering (4d), the design that optimises this confidence is the one that keeps $R(N_s) \approx R(0)$, which translates to a design that minimises $x(t)$, and that for $x(0) = 0$ applies $u(t) = 0$ for all $t$. A required confidence of $1-\delta > 0.5$ would never be attained with this standard approach. We contrast this outcome with the new approach, elucidated next.

### B. Stochastic optimal control formulation

Consider a reformulation of the experiment design problem for a discounted probabilistic reach-avoid problem over an MDP. We again consider the case where we stop the experiment when a confidence of at least $1 - \delta$ is attained. We select a rather high discount factor $\gamma = 0.6$, since we expect the average time that the experiment takes to be short.

Based on the proposed MDP reformulation of the experiment design problem we obtain a three-dimensional MDP with state $(x, \mu, R)$ for the discounted reach-avoid problem. The stochastic optimal control problem is solved approximately via fitted value iteration [18]. In this algorithm the integration action in the backwards mapping is replaced by samples and the value functions are fitted with a neural network. We have chosen a network with 2 layers, with respectively 15 and 10 neurons, and `tansig` functions for all neurons. Within the 10 backwards mappings applied to obtain the outcomes plotted in Fig. 3 and 4, the approximate value function iteration has already converged. Notice that the approximate function is denoted by $\hat{\mathbf{W}}_0^*$, in contrast to the exact value function $\mathbf{W}_0^*$. Extending the time horizon beyond $N_s = 10$ will not change the value function $\hat{\mathbf{W}}_0^*$, hence the value functions $\hat{\mathbf{W}}_0^*$ in Figures 3 and 4 additionally display the approximate solution to an infinite horizon problem.

Notice in Fig. 4 that for decreasing variance an increasing amount of the state space has a value equal to one. These are the states for which $\mathbf{S} \models \psi$ can be accepted or rejected with confidence at least $1 - \delta = 0.99$.

The optimal policy can now be computed as in Proposition 3 by selecting at each state $x_\theta = (x, \mu, R) \in A \backslash K$ the action that maximises the expected value function. Notice in Fig. 3 and 4 that when starting the experiment at $x = 0$ and with $\mu$ initialised at the edges of the feasible set, i.e. $\mu = 1$ and $\mu = -1$, it can be observed that the function is *locally* convex hence the optimal control action is either equal to 1 or to $-1$. More precisely, not applying an input (i.e. $u(0) = 0$) would mean that $\mu$ and $R$ would not change over time. But any input different than zero will excite the system, therefore

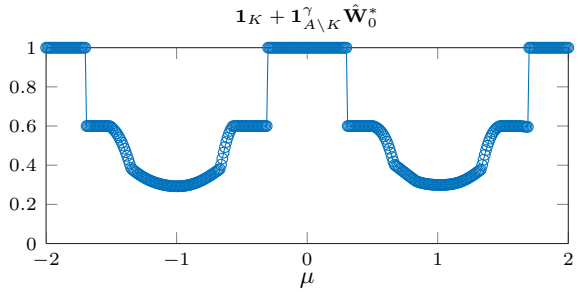


Fig. 3: Slice of the discounted optimal reach-avoid probability over the values of $\mu$ for the given parameters $R = 0.3$, $x = 0$.

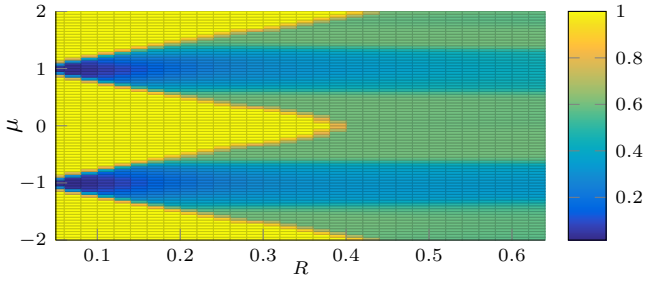


Fig. 4: Surface plot of the $\gamma$-discounted reach-avoid probability $\mathbf{1}_K + \mathbf{1}^{\gamma}_{A\backslash K}\hat{\mathbf{W}}_0^*$ for $x(0) = 0$ evaluated over varying values of the posterior mean $\mu(0)$ and covariance $R(0)$. The light yellow area in the plot is in the target set $K$.

decrease the variance and create a stochastic transition kernel with a non-singular variance for $\mu$. Hence with probability 1 the value at the next time step will improve.

## REFERENCES


[1] L. Ljung, *System identification - Theory for the user*, 2nd ed. PTR Prentice Hall, 1999.

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

[3] X. Bombois, G. Scorletti, M. Gevers, P. Van den Hof, and R. Hildebrand, "Least costly identification experiment for control," *Automatica*, vol. 42, no. 10, pp. 1651–1662, 2006.

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

[5] C. Baier and J.-p. Katoen, "Principles of model checking," *MIT Press*, vol. 950, 2008.

[6] 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*, 2015, pp. 1800–1805.

[7] 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.

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

[9] S. Haesaert, A. Abate, and P. M. J. V. den Hof, "Data-driven and model-based verification: A bayesian identification approach," in *54th IEEE Conference on Decision and Control*, Dec 2015, pp. 6830–6835.

[10] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal control : The discrete time case*. Athena Scientific, 1996.

[11] C. A. Larsson, "Application-oriented experiment design for industrial model predictive control," Ph.D. dissertation, KTH, 2014.

[12] J. D. Stigter, D. Vries, and K. J. Keesman, "On adaptive optimal input design: a bioreactor case study," *AIChE J.*, vol. 52, no. 9, pp. 3290–3296, 2006.

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

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

[15] A. Abate, M. Prandini, J. Lygeros, and S. Sastry, "Probabilistic Reachability and Safety for Controlled Discrete Time Stochastic Hybrid Systems," *Automatica*, vol. 44, no. 11, pp. 2724–2734, 2008.

[16] S. Summers and J. Lygeros, "Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem," *Automatica*, vol. 46, pp. 1951–1961, 2010.

[17] D. Bertsekas, "Approximate policy iteration: a survey and some new methods," *J. Control Theory Appl.*, vol. 9, no. 3, pp. 310–335, 2011.

[18] S. Haesaert, R. Babuska, and A. Abate, "Sampling-based Approximations with Quantitative Performance for the Probabilistic Reach-Avoid Problem over General Markov Processes," sep 2014. [Online]. Available: http://arxiv.org/abs/1409.0553