Design of experiments for nonlinear system identification: A set membership approach

Design of Experiments for Nonlinear System Identification: a Set Membership Approach

Milad Karimshoushtari [a], Carlo Novara [a],

[a] *Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*

Abstract

Design of Experiments (DoE) is an important step in system identification. Regardless of the chosen model structure and identification method, the DoE quality determines an upper bound on the accuracy of the identified models. One of the greatest challenges in this context is to design an experiment which gives the maximum information about the dynamics of the system of interest. In this paper, a novel DoE algorithm for input-constrained MISO nonlinear systems, based on set membership identification, is proposed. The DoE algorithm is aimed to minimize the so-called radius of information, a quantity giving the worst-case model error. Two numerical examples are presented, showing the effectiveness of the approach and its potential in view of real-world applications.

Key words: DoE, DoDE, SM-DoE, Experiment Design, System Identification, Model Predictive Control, Adaptive Identification, Data Driven Control

1 Introduction

1.1 Background

In many technological areas, obtaining an accurate model of a dynamic system of interest is a fundamental step for any system analysis and/or design operation. However, building an accurate model using the physical laws governing the system may not be possible in several situations, due to the fact that these laws are not sufficiently well known or they are too complex, requiring a computationally expensive model that may be difficult to analyze or to use for design purposes. In this view, data-driven system identification approaches can be crucial in a wide range of applications.

Data-driven system identification can be seen as the science of building mathematical models of dynamic systems, using data and a "weak" prior physical knowledge. For example, the physical laws governing the system of interest may be not known but some less detailed information may be available, regarding its block structure, the type of involved dynamics, the type of involved nonlinearities, the system order, etc.

Typically, the identification process consists of the following main steps (not necessarily in the order reported here): 1) design of experiment (DoE); 2) selection of a suitable parametrized model structure; 3) identification of the model parameters (usually done through an optimization problem); 4) evaluation of the model quality through some validation analysis. In this process, the command input signal is the only means that can be used in the DoE phase to influence the information content of a dataset to be used for identification (this set is called identification dataset or training dataset). Regardless of the chosen model structure and identification method, the quality of the DoE determines the accuracy that can be achieved by any method.

1.2 State of the Art

One of the greatest challenges in this context is to design an experiment giving the maximum information about the system to be identified [11,17]. Most of the studies carried out so far have mainly focused on linear systems [14,39,36,9,35] and static systems [8,33,15]. On the other side, very few studies regarding nonlinear dynamic systems are available [28,6,10,12,27,20,19]. In fact, nonlinear systems are characterized by a significantly higher complexity than linear systems. While for linear systems the excitation properties of an input signal essentially depend on the signal frequencies, for nonlinear systems they also depend on the signal amplitudes [28]. For example, a white noise input signal is known in general to be appropriate for the identification of a linear system of any order. On the other side, it may not be suitable

⋆

Email addresses: milad.karimshoushtari@polito.it (Milad Karimshoushtari), carlo.novara@polito.it (Carlo Novara).

to allow an accurate exploration of the regressor domain of a nonlinear system, and this may lead to a low model accuracy. More specifically, because of the high complexity of the DoE problem for nonlinear systems, in most of the literature, the general DoE problem is either reduced to an input selection problem [28] or to a problem requiring quite restrictive assumptions. Typically, see e.g. [27,20,19], it is assumed that the exact model structure and some bounds of the parameters to identify are known. However, this assumption may not hold in many real-world situations. In this paper, we are interested in the more general case where the exact structure is not known. A more detailed discussion is given in Section 3.4.

Generally, the most popular DoE methods for nonlinear dynamic systems can be classified in two main categories: model-free and model-based methods. The idea in most of these methods, see e.g. [6,38,5,13], is to parameterize a pre-defined excitation signal and then optimize the signal parameters, called the design points, according to different criteria. In model-free DoE, no assumptions on the model structure are made. The typical approach is to distribute the design points in the input domain as much uniformly as possible. This DoE approach is also known as space-filling DoE.

In model-based DoE, after assuming a particular model structure (which can be a rough estimate of the model structure), the idea is to distribute the design points in the input domain, in such a way that the estimation of the model parameters is as much insensitive as possible to the measurement noise. In both model-free and model-based DoE, after designing the distribution of the design points in the input space, they are used as the parameter values of the pre-defined excitation signal. However, both methods provide no information about the optimal sequence of the design points. Although these methods are simple and adequate to capture the steady state behavior of a system of interest, they don't take into account the dynamics of the system. Therefore, by using these methods, capturing the nonlinear dynamic behavior of the system in the whole regressor domain turns out to be a heuristic/arbitrary process. In general, as far as the authors are aware, no DoE method for nonlinear dynamic systems can be found in the literature, which can ensure the exploration of the relevant regressor domain of a nonlinear system and, consequently, guarantee a desired model accuracy.

Due to imprecise prior knowledge, disturbances and measurement noise, in general, no identification process can result in a model that perfectly corresponds to the true system. Any identified model is always affected by some uncertainty. Understanding which are the regions of the regressor space where the model is most uncertain is a key element to build a proper DoE algorithm. However, knowing where the model is most uncertain is not sufficient. Since the unknown system is dynamic, the DoE algorithm has to be able to generate an input sequence such that the system moves toward those uncertain regions of the regressor space, in order to collect new measurements.

In the last three decades, there has been an increasing interest and research, formulating the identification problem in the Set Membership (SM) framework [22,37,21,26,25,29,30,24,4]. The main reason is the fact that SM identification allows us to properly quantify the uncertainty of the identified model in a deterministic manner. In SM nonlinear identification, no assumptions on the structure of the unknown system are required. Instead, two basic assumptions are made: An assumption on the regularity of the system, given by its Lipschitz constant or by bounds on its gradient, and another assumption on the noise boundedness. Then, an optimal estimate, with minimal guaranteed identification error and tight uncertainty bounds, is derived. This nonlinear SM approach does not require iterative minimization and thus avoids the issue of local minima. Since no optimization problems have to be solved, nonlinear SM identification is particularly suitable for adaptive identification, making the model more accurate over time by adding new measurements collected online. Because of these features, the applications of SM in robust control and experiment design is a promising research area [37].

1.3 Paper Contribution

In this paper, a novel online DoE algorithm for nonlinear dynamic MISO systems is proposed, that is able to reduce the worst-case model error, while considering input constraints of the system. The proposed DoE algorithm is able to guarantee any desired worst-case error larger than the measurement error in a finite-time experiment. The main contributions of this paper are the following. First, a so-called quasi-local nonlinear SM identification method is presented, that is characterized by less conservative bounds with respect to the global version of [22] and is simpler with respect to the local version of [22]. The second contribution of this paper is a novel adaptive Set Membership Predictive Control (SMPC) algorithm, that is able to drive the system toward the most uncertain regions of the regressor space. And finally, the third contribution of this paper is the online DoE algorithm itself. The effectiveness of the proposed DoE algorithm is illustrated in two simulation examples and compared to other DoE methods taken from the literature.

The paper is organized as follows. In section 2, the identification problem is formulated in the nonlinear SM framework and the quasi-local approach is introduced. In section 3, a static DoE algorithm, and a SM predictive controller are proposed, which are then used in the dynamic SM-DoE algorithm. In section 4, the proposed SM-DoE algorithm is tested in two simulation examples.

2 Quasi-Local Nonlinear Set Membership Identification

Consider a nonlinear discrete-time dynamic system in regression form:

$$\begin{aligned} y^{t+1} &= f_o(w^t), \\ w^t &= [y^t \dots y^{t-n_y+1}\ u^t \dots u^{t-n_u+1}]. \end{aligned} \tag{1}$$

where $y^t \in \mathbb{R}$, $u^t \in \mathbb{R}^m$, $f_o : \mathbb{R}^n \to \mathbb{R}$, $n = n_y + mn_u$ and the superscript is used to indicate the time index $t \in \mathbb{Z}$. Suppose that the function f_o is unknown but a set of noise corrupted data called *measurement dataset* generated by the system (1) is available.

$$\mathcal{D} \doteq \left\{\widetilde{y}^{t+1}, \widetilde{w}^t\right\}_{t=1}^{T-1}. \tag{2}$$

Then,

$$\widetilde{y}^{t+1} = f_o(\widetilde{w}^t) + d^t,\ t = 1, \dots, T-1. \tag{3}$$

where the term d^t accounts for the fact that y and w are not exactly known, due to possible disturbances and noises affecting the system.

The aim is to derive an estimate $\widehat{f}$ of f_o from the available measurements $\mathcal{D}$, using a suitable identification algorithm. An identification algorithm ϕ can be seen as a sequence of operations, providing some estimate $\widehat{f}$ of the unknown function f_o from the available measurements $\mathcal{D}$. Clearly, the algorithm ϕ should be chosen to give a small (possibly minimal) identification error $e(\widehat{f}) = \|f_o - \widehat{f}\|_p$, where $\|\cdot\|_p$ is the functional L_p norm, defined as

$$\|f\|_p \equiv \|f(\cdot)\|_p \doteq \begin{cases} \left[\int_{\mathcal{W}} |f(w)|^p\, dw\right]^{1/p}, p \in [1, \infty) \\ \operatorname{ess\,sup}_{w \in \mathcal{W}} |f(w)|\,, p = \infty \end{cases} \tag{4}$$

being $\mathcal{W}$ a compact and connected set in $\mathbb{R}^n$.

This error is not known, since from the available data, it is only known that $f_o \in \tilde{\mathcal{F}}$, where $\tilde{\mathcal{F}}$ is the set of all functions that could have generated the data. If no assumptions are made on f_o, this set, even in the case of exact measurements, is unbounded. Whatever algorithm ϕ is chosen, no information on the identification error can be derived, unless some assumptions are made on the function f_o and the noise d. The typical approach in the literature is to assume a finitely parameterized structure for f_o (linear, polynomial, neural network, etc.) and a statistical model for the noise, see [34]. In the SM approach, different and somewhat weaker assumptions are taken, not requiring the choice of the parametric structure for f_o, but related to its regularity. Moreover, the noise sequence $\{d^t\}_{t=1}^{T-1}$ is only supposed to be bounded.

Assumption 1. The noise d^t is unknown but bounded.

$$\left|d^t\right| \leqslant \varepsilon,\ t = 1, \dots, T-1. \tag{5}$$

Assumption 2. The function f_o is Lipschitz continuous on $\mathcal{W}$.

Based on Assumption 2, we can define the following quantity, called the *quasi-local Lipschitz parameter*:

$$\gamma(w) = \sup_{\widehat{w} \in \mathcal{W}, \widehat{w} \neq w} \frac{|f_o(w) - f_o(\widehat{w})|}{\|w - \widehat{w}\|}. \tag{6}$$

where $\|\cdot\|$ is the vector euclidean norm. Obviously, the global Lipschitz constant of f_o on $\mathcal{W}$ is given by

$$\Gamma = \sup_{w \in \mathcal{W}} \gamma(w). \tag{7}$$

Lemma 1. For any $w \in \mathcal{W}$, a $\gamma(w)$ exists, such that

$$|f_o(w) - f_o(\widehat{w})| \leqslant \gamma(w)\, \|w - \widehat{w}\|,\ \forall \widehat{w} \in \mathcal{W}.$$

Proof. The statement follows directly from (6). □

Let us now suppose that the quasi-local Lipschitz parameters $\gamma(\widetilde{w}^t),\ t = 1, \dots, T-1$, are known or can be estimated (a method for performing such an estimation is given in Section 2.5). On the basis of this information, we can define the following function set:

$$\begin{aligned} \mathcal{F} \doteq \{f : |f(w) - f(\widetilde{w}^t)| \leqslant \gamma(\widetilde{w}^t)\, \|w - \widetilde{w}^t\|\,, \\ \forall w \in \mathcal{W},\ t = 1, \dots, T-1\}. \end{aligned} \tag{8}$$

This allows us to introduce the *Feasible Function Set* (FFS), i.e. the set of all functions consistent with prior assumptions and measured data.

Definition 1. Feasible Function Set

$$\begin{aligned} FFS^T \doteq \{f \in \mathcal{F} : \left|\widetilde{y}^{t+1} - f(\widetilde{w}^t)\right| \leqslant \varepsilon, \\ t = 1, \dots, T-1\}. \end{aligned} \tag{9}$$

The Feasible Function Set summarizes all the information on the mechanism generating the data, that is available up to time T. If the prior assumptions hold, then $f_o \in FFS^T$, that is an important property for evaluating the accuracy of any estimate. Indeed, from the FFS definition, it follows that $f_o(w)$ is bounded as

$$\underline{f}(w) \leqslant f_o(w) \leqslant \overline{f}(w), \forall w \in \mathcal{W} \tag{10}$$

where

$$\begin{aligned} \overline{f}(w) &= \sup_{f \in FFS^T} f(w) \\ \underline{f}(w) &= \inf_{f \in FFS^T} f(w). \end{aligned} \tag{11}$$

Provided that the prior assumptions hold, $\overline{f}$ and $\underline{f}$ are tightest upper and lower bounds of f_o. For this reason, such functions are called *optimal bounds*.

In the set membership framework, validation of the prior assumptions is a fundamental step. It is usual to

introduce the concept of prior assumption validation as consistency with the available data; the prior assumptions are considered validated if at least one estimate consistent with these assumptions and the data exists, i.e. if FFS is not empty [21,4].

Definition 2. Validation of prior assumptions
Prior assumptions are considered validated if $FFS^T \neq \emptyset$

Note that the fact that prior assumptions are validated, i.e., that they are consistent with the present data, does not exclude that they may be invalidated by future data. In the following, the FFS^T is assumed to be non-empty. If not, values of the constants appearing in the assumptions on function f_o and on the noise d have to be suitably modified to give a non-empty FFS^T, as discussed in section 2.3.

Now, the notion of optimal algorithm is introduced. An identification algorithm ϕ is an operator mapping all the available information about the function f_o, the noise d, the measurement data $\mathcal{D}$ until time T, summarized by FFS^T, into an estimate $\widehat{f}$ of f_o:

$$\phi(FFS^T) = \widehat{f} \simeq f_o.$$

For a given estimate $\widehat{f}$, the related L_p error is:

$$e(\widehat{f}) = e(\phi(FFS^T)) = \|f_o - \widehat{f}\|_p.$$

This error cannot be exactly computed, since it is only known that $f_o \in FFS^T$. However, its tightest bound is given by

$$e(\widehat{f}) \leqslant \sup_{f \in FFS^T} \|f - \widehat{f}\|_p.$$

This motivates the following definition of the identification error, often indicated as the worst-case or guaranteed error.

Definition 3. Identification error.
The identification error of $\widehat{f} = \phi(FFS^T)$ is

$$E[\phi(FFS^T)] = E(\widehat{f}) \doteq \sup_{f \in FFS^T} \|f - \widehat{f}\|_p.$$

Looking for algorithms that minimize the identification error, leads to the following optimality concepts.

Definition 4. Optimal algorithm.
An algorithm ϕ^* is optimal if

$$\begin{aligned} E[\phi^*(FFS^T)] &= \inf_{\phi} E[\phi(FFS^T)] \\ &= \inf_{\widehat{f}} \sup_{f \in FFS^T} \|f - \widehat{f}\|_p \doteq \mathcal{R}_{\mathcal{I}}. \end{aligned}$$

$\mathcal{R}_{\mathcal{I}}$ is called the *radius of information* and is the minimum worst-case error that can be achieved on the basis of the prior and experimental information available up to time T. In other words, $\mathcal{R}_{\mathcal{I}}$ is a measure of the uncertainty associated with the identification process, for the given dataset and prior information. A reduction/minimization of $\mathcal{R}_{\mathcal{I}}$ can be obtained by a suitable experiment design procedure, as shown in Section 3.

After introducing the general framework of nonlinear SM identification, now we show that the optimal bounds and optimal algorithm formally introduced so far can be actually computed in closed form. We also provide necessary and sufficient conditions for assumption validations.

2.1 Optimal Bounds

The optimal bounds formally defined in (11) can be computed in closed form upon definition of the functions

$$\begin{aligned} f_u(w) &\doteq \min_{t=1,\dots,T-1} (\overline{h}^t + \gamma(\widetilde{w}^t)\|w - \widetilde{w}^t\|) \\ f_l(w) &\doteq \max_{t=1,\dots,T-1} (\underline{h}^t - \gamma(\widetilde{w}^t)\|w - \widetilde{w}^t\|) \end{aligned} \tag{12}$$

where $\overline{h}^t \doteq \widetilde{y}^{t+1} + \varepsilon^t$ and $\underline{h}^t \doteq \widetilde{y}^{t+1} - \varepsilon^t$.
The next result shows that f_u and f_l are optimal bounds, i.e., the tightest upper and lower bounds of f_o according to (11).

Theorem 1. *The functions f_u and f_l defined in (12) are optimal bounds, i.e.,*

$$\begin{aligned} \overline{f}(w) &= f_u(w) \\ \underline{f}(w) &= f_l(w). \end{aligned}$$

Proof. The complete proof is given in [16]. It can be obtained by minor modifications of the proof of Theorem 2 in [22]. □

2.2 Optimal Algorithm and Estimate

The functions $\overline{f}$ and $\underline{f}$ can be used to solve the problem of finding the optimal estimate of $f_o(w)$ for given $w \in \mathcal{W}$. Let the function f_c be defined as

$$f_c(w) \doteq \frac{1}{2}[\underline{f}(w) + \overline{f}(w)]. \tag{13}$$

where $\underline{f}(w)$ and $\overline{f}(w)$ are given in Theorem 1. The next result shows that the algorithm $\phi_c(FFS^T) = f_c$ is optimal for any L_p norm.

Theorem 2. *For any $L_p(\mathcal{W})$ norm, with $p \in [1, \infty]$:*
(i) The identification algorithm $\phi_c(FFS^T) = f_c$ is optimal.
(ii) The worst-case identification error of f_c is bounded as

$$E(f_c) = \inf_{\phi} E[\phi(FFS^T)] = \frac{1}{2}\|\overline{f} - \underline{f}\|_p = \mathcal{R}_{\mathcal{I}}.$$

Proof. The proof can be obtained by minor modifications of the proof of Theorem 7 in [22]. □

Remark 1. The point-wise bounds (10) provide an interval estimate of the unknown value $f_o(w)$. Interval estimates allow us to quantify the uncertainty associated with the identification process, and are thus important in system and control applications. Indeed, these estimates can be used e.g. for robust control design [31,40], fault detection [29] and experiment design [28].

2.3 Assumption Validation

Necessary and sufficient conditions for checking the assumptions validity are now given.

Theorem 3. *(i) A necessary condition for prior assumptions to be validated is:* $\overline{f}(\widetilde{w}^t) \geqslant \underline{h}^t$, $\underline{f}(\widetilde{w}^t) \leqslant \overline{h}^t$, $t = 1, \ldots, T-1$.
(ii) A sufficient condition for prior assumptions to be validated is: $\overline{f}(\widetilde{w}^t) > \underline{h}^t$, $\underline{f}(\widetilde{w}^t) < \overline{h}^t$, $t = 1, \ldots, T-1$.

Proof. The complete proof can be obtained by minor modifications of the proof of Theorem 1 in [3]. See also [16]. □

In the reminder of paper, it is assumed that the sufficient condition holds. If not, values of the constants appearing in the assumptions on function f_o and on the noise d have to be suitably modified. The above validation Theorem can be used for assessing the values of such constants so that sufficient conditions holds.

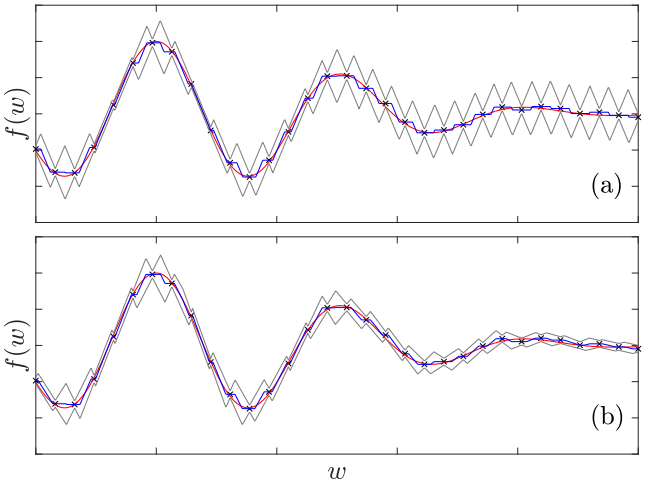

Fig. 1. (a) global bound, (b) quasi-local bound, $f_o(w)$ red line, Measurements black cross, $f_c(w)$ blue line, $\overline{f}(w), \underline{f}(w)$ grey line.

Figure 1 shows the global and quasi-local set membership bounds for a nonlinear function. In Figure 1(a), a constant global bound Γ was assumed to compute the optimal bounds [22]. In Figure 1(b), a quasi-local bound was assumed (Eq. (12)). The bounds are clearly tighter in the quasi-local model, especially when the function is "flat".

2.4 Radius of Information

The quantity $\mathcal{R}_{\mathcal{I}}$, called the *radius of information*, allows us to assess the accuracy achieved by the optimal estimate. In our approach, this quantity is used to quantify the informative content of a data set. Let us define the following error function:

$$f_e(w, \mathcal{D}) \equiv f_e(w) \doteq \frac{1}{2}[\overline{f}(w) - \underline{f}(w)]. \tag{14}$$

where $\mathcal{D}$ is the measurement dataset (2). Here, the dependence on $\mathcal{D}$ is explicited, in order to emphasize the fact that the function f_e is constructed from the measurement dataset. This function allows us to write the radius of information as

$$\mathcal{R}_{\mathcal{I}}^p = \|f_e(\cdot, \mathcal{D})\|_p. \tag{15}$$

The analytical computation of $\|f_e\|_p$ is not feasible, since f_e is a quite "complex" nonlinear function, defined over a multi-dimensional domain. Hence, following a standard approach, we compute numerically the norm, evaluating f_e on a finite set of points w^k and then approximating the norm as

$$\|f\|_p \simeq \widehat{\|f\|}_p = \begin{cases} \left[\sum_{k=1}^m a_k \left|f(w^k)\right|^p\right]^{1/p}, p \in [1, \infty) \\ \max\limits_{k=1,\ldots,m} \left|f(w^k)\right|, p = \infty \end{cases} \tag{16}$$

where a_k are suitably chosen coefficients. For $a_k = 1/m$ we have the widely used quasi-Monte Carlo algorithms [23]. The expression of $\mathcal{R}_{\mathcal{I}}^p$ given in (15) and computed according to (16) will be used in the next sections, in order to develop our DoE algorithms.

2.5 Parameter Estimation

Estimates of the noise bound ε, Lipschitz constant Γ, and the quasi-local Lipschitz parameter $\gamma(w)$ such that the assumptions are validated can be obtained by means of two algorithms given in [7] and reported in the following. The first algorithm is directly taken from [7], while the second one is a generalization of the corresponding one in [7].

The following theorems show that, under reasonable density conditions on the noise, the estimates given by these two algorithms converge to the corresponding true values.

Theorem 4. *(Theorem 2 of [7]) Let the set* $\{\widetilde{w}^t, d^t\}_{t=1}^{T-1}$ *appearing in (1) be dense on* $\mathcal{W} \times B_\varepsilon$ *as* $T \to \infty$. *Then,*

$$\lim_{T \to \infty} \widehat{\varepsilon} = \varepsilon. \qquad \square$$

Theorem 5. *(Theorem 3 of [7]) Let the set* $\{\widetilde{w}^t, d^t\}_{t=1}^{T-1}$

appearing in (1) be dense on $\mathcal{W} \times B_\varepsilon$ *as* $T \to \infty$*. Then,*

$$\lim_{T\to\infty} \widehat{\Gamma} = \Gamma. \qquad \square$$

Algorithm 1 Noise Bound Estimation ε

(1) Choose a "small" $\rho > 0$. for example: $\rho = 0.01 \max_{t,k=1,\dots,T-1} \|\widetilde{w}^t - \widetilde{w}^k\|$.
(2) Find the set of indexes: $I_t \doteq \{k : \|\widetilde{w}^t - \widetilde{w}^k\| \leqslant \rho\}$. if $I_t = \emptyset$ for all $t = 1, \dots, T-1$, go to step 1 and choose a larger ρ.
(3) For $t = 1, \dots, T-1$ compute $\delta\widetilde{y}^{t+1} = \max_{i \in I_t} \left|\widetilde{y}^{t+1} - \widetilde{y}^{i+1}\right|$. If $I_t = \emptyset$, set $\delta\widetilde{y}^{t+1} = \infty$.
(4) Obtain the estimate $\widehat{\varepsilon}$ of the noise bound ε as $\widehat{\varepsilon} = \frac{1}{2N} \sum_{t\in Q} \delta\widetilde{y}^{t+1}$ where $Q \doteq \{t \in \{1, \dots, T-1\} : \delta\widetilde{y}^{t+1} < \infty\}$ and $N \doteq \mathrm{card}(Q)$

Algorithm 2 Lipschitz Parameter Estimation γ, Γ

For $t = 1, \dots, T-1$ and $\widetilde{w}^t \neq \widetilde{w}^k$, compute

$$\widehat{\gamma}(\widetilde{w}^t) = \max_{k=1,\dots,T-1} \begin{cases} \frac{|\widetilde{y}^{t+1} - \widetilde{y}^{k+1}| + 2\widehat{\varepsilon}}{\|\widetilde{w}^t - \widetilde{w}^k\|} & \text{if } |\widetilde{y}^{t+1} - \widetilde{y}^{k+1}| > 2\widehat{\varepsilon} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Obtain the estimate $\widehat{\Gamma}$ of the Lipschitz constant Γ as

$$\widehat{\Gamma} = \max_{t,k=1,\dots,T-1} \begin{cases} \frac{|\widetilde{y}^{t+1} - \widetilde{y}^{k+1}| - 2\widehat{\varepsilon}}{\|\widetilde{w}^t - \widetilde{w}^k\|} & \text{if } |\widetilde{y}^{t+1} - \widetilde{y}^{k+1}| > 2\widehat{\varepsilon} \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

3 Set Membership Design of Experiments

The goal of DoE is to synthesize an input sequence to apply to the plant, in order to explore the regressor space $\mathcal{W}$ in the most effective way, allowing us to maximize the information extracted from the collected data, thus minimizing the uncertainty of the estimated model. In the set membership framework, the model uncertainty is measured by the radius of information (15), calculated in some selected norm.

In this section, we propose a novel set membership design of experiments (SM-DoE) approach for nonlinear dynamic systems, aimed at minimizing the radius of information.

3.1 Problem Formulation

Let us consider a multiple input single output (MISO) nonlinear system described by (1). The system is unknown and the output is corrupted by noise.

Let Assumptions 1 and 2 hold. And, let $U_t^T \doteq \{u^k\}_{k=t}^{T-1}$ be an input sequence from time t to time $T-1$. The problem considered in this section is the following.

Problem: Design an input sequence U_1^T that, applied to the nonlinear system (1), yields a minimal radius of information $\mathcal{R}_{\mathcal{I}}^p = \|f_e(w)\|_p$.

Ideally, a solution to this problem is given by:

$$\begin{aligned} U_1^{*T} = \arg\min_{U_1^T} \quad & \|f_e(\cdot, \mathcal{D})\|_p \\ \text{subject to} \quad & \widetilde{y}^{t+1} = f_o(\widetilde{w}^t) + d^t,\ t = 1, \dots, T-1 \\ & \mathcal{D} = \{\widetilde{y}^{t+1}, \widetilde{w}^t\}_{t=1}^{T-1} \end{aligned} \tag{19}$$

where $d^t, t = 1, \dots, T-1$ is the actual noise sequence.

However, for several reasons, this optimization problem cannot be used for DoE in real applications: 1) it requires to know f_o and the complete noise sequence $\{d^t\}_{t=1}^{T-1}$; 2) even assuming that f_o is known and the noise can be measured, the optimization problem can only be solved at time $T-1$, since at previous time instants $t < T-1$, the noise samples d^k are not known for $k > t$; 3) even in the case that f_o and the complete noise sequence are known, the optimization problem is highly nonlinear and non-convex, and thus hard to solve analytically. Nevertheless, in the simulations studies that will be presented in the paper, the problem (19) will be solved numerically (without guarantees of finding a global minimum) and the obtained estimate of the "ideal" optimal input sequence U_1^{*T} will be used as a term of comparison, to indicate the maximum performance that can be achieved by any DoE algorithm.

The approach to DoE that we propose can actually be applied in real situations, without knowing a-priori the true function and noise sequence. A key feature is that the approach is sequential: at each time step, on the basis of the current and past measured data, the approach individuates what is the next point of the regressor domain that the system has to visit, in order to maximally reduce the radius of information. In the case where the system is a static function of the input - i.e., where $w^t = u^t$ - the optimal input is obviously chosen equal to the individuated next point. In the general case of dynamic systems, it may be not possible to visit the desired point, since the system future regressor depends not only on the current input but also on the past input and output values. Hence, once the next point to visit has been found, a model predictive control (MPC) strategy is used to drive the system toward that point. The proposed MPC strategy is based on a nonlinear set membership model identified from the past data and updated at each time step on the basis of the new measurement.

The DoE algorithms for static and dynamic systems are treated separately in the next two subsections. Indeed, although the idea behind both the algorithms is the same and the algorithm for static systems can be seen as a subcase of that for dynamic ones, the algorithm for static systems is much simpler. It thus deserves a separate presentation, in order to allow the user an easy algorithm implementation and usage in the static case.

3.2 Static Set Membership DoE

Consider a static nonlinear system of the form

$$z^t = f_o(w^t),\ w^t = u^t \in \mathcal{W}. \tag{20}$$

In this case, solving the experiment design problem is easier with respect to the general dynamic case, since the system trajectory depends only on the current input and not on the past input and output values. Hence, it is possible to obtain a measurement of the function f_o at any desired point of the regressor domain $\mathcal{W}$. The static Set-Membership DoE algorithm that we propose is the following.

Algorithm 3 Static Set-Membership DoE

(1) Choose the initial regressor w^1 (e.g., the center of the regressor domain $\mathcal{W}$);
Measure $\widetilde{z}^1 = f_o(w^1)$;
Define the measurement dataset $\mathcal{D} = \left\{\widetilde{z}^1, w^1\right\}$.

(2) While $t < T$, solve the optimization problem

$$\begin{aligned} &\mathcal{W}_M^t = \arg\max_{w\in\mathcal{W}} f_e^t(w, \mathcal{D});\\ &w^t \in \mathcal{W}_M^t \end{aligned} \tag{21}$$

Measure $\widetilde{z}^t = f_o(w^t)$;
Add $\widetilde{z}^t$ and w^t to the dataset $\mathcal{D} := \mathcal{D} \cup \{\widetilde{z}^t, w^t\}$;
$t := t + 1$.

The vector w^t is any point in $\mathcal{W}_M^t$ and f_e^t is the error function (14) computed at time instant t.

The algorithm is iterative. At each iteration, a point in the regressor domain where the uncertainty is maximum is considered. The optimization problem (21) is nonlinear and non-convex. However, as discussed in section 2.4, we evaluate f_e on a finite set of points in the domain $\mathcal{W}$, making the computation easy (thus obtaining in general a sub-optimal solution). The following result holds for Algorithm 3.

Theorem 6. *Let T be the number of steps in Algorithm 3 and $\mathcal{R}_\mathcal{I}(t)$ be the radius of information computed at time t. Then, there exists a T such that $\mathcal{R}_\mathcal{I}(t) \leqslant \varepsilon$, for all $t \geqslant T$.*

Proof. Let us define the following set:

$$\mathcal{W}_e^t = \left\{w \in \mathcal{W} : \Gamma\|w - w^t\| < f_e(w) - \varepsilon\right\}. \tag{22}$$

The following inequalities hold at each iteration of the algorithm:

$$f_e^{t+1}(w) < f_e^t(w) \quad \forall w \in \mathcal{W}_e^t, \tag{23}$$

$$f_e^{t+1}(w) = f_e^t(w) \quad \forall w \in \mathcal{W} \setminus \mathcal{W}_e^t. \tag{24}$$

If $\mathcal{W}_M^t \subseteq \mathcal{W}_e^t$, we have $f_e^t(w) < f_e^t(w^t)$ for all $w \in \mathcal{W}\setminus\mathcal{W}_e^t$ and from (24) we can write $f_e^{t+1}(w) < f_e^t(w^t) = \mathcal{R}_\mathcal{I}(t)$ for all $w \in \mathcal{W}\setminus\mathcal{W}_e^t$. From (23) we have $f_e^{t+1}(w) < \mathcal{R}_\mathcal{I}(t)$ for all $w \in \mathcal{W}_e^t$,. Therefore, $f_e^{t+1}(w) < \mathcal{R}_\mathcal{I}(t)$ for all $w \in \mathcal{W}$. which also means

$$\mathcal{R}_\mathcal{I}(t+1) < \mathcal{R}_\mathcal{I}(t). \tag{25}$$

If $\mathcal{W}_M^t \nsubseteq \mathcal{W}_e^t$, from (23), at each iteration we have

$$\mathcal{W}_M^{t+1} = \mathcal{W}_M^t \setminus \mathcal{W}_e^t. \tag{26}$$

Since $w^t \in \mathcal{W}_e^t, w^t \in \mathcal{W}_M^t$ it is evident that $\mathcal{W}_M^t \cap \mathcal{W}_e^t \neq \emptyset$. Therefore, $\mathcal{W}_M^{t+1} \subset \mathcal{W}_M^t$ which means $\mathcal{W}_M$ is shrinking at each time step therefore in a finite n_i steps we have $\mathcal{W}_M^{t+n_i} \subseteq \mathcal{W}_e^{t+n_i}$. Thus from (25) we have,

$$\mathcal{R}_\mathcal{I}(t+n_i+1) < \mathcal{R}_\mathcal{I}(t). \tag{27}$$

$\mathcal{R}_\mathcal{I}$ is a positive definite function and $\mathcal{R}_\mathcal{I} \leqslant \varepsilon$ if $\mathcal{D} = \mathcal{W}$. Also from (25),(27) we can say $\mathcal{R}_\mathcal{I}$ is a decreasing function. Therefore, as $t \to \infty$, Then $\mathcal{R}_\mathcal{I}(t) \leqslant \varepsilon$.

□

3.3 Dynamic Set Membership DoE

Suppose that the DoE has to be carried out for a nonlinear dynamic system, written in the general regression form (1). Unlike the static case (20), it is not possible to evaluate the regression function at any desired point w, since the system regressor depends not only on the current input but also on the past input and output values. The idea that we propose is to use an algorithm similar to Algorithm 3 to generate desired reference points w^r in combination with an MPC controller making the dynamic system visit the desired point w^r. The MPC approach that we propose is novel and is called set membership model predictive control (SMPC).

3.3.1 Set Membership Model Predictive Control

In recent years, there has been an increasing interest in set membership predictive control laws, designed from experimental data [1,2,32,37]. However, such approaches, implicitly or explicitly, assume that a sufficiently informative set of data is available and do not consider the problem of experiment design. In this section, we propose a novel MPC approach, called SMPC, able to perform together experiment design and controller design.

To formulate the SMPC approach, a state-space-like representation of the plant (1) and related models is needed. To this aim, we introduce the following pseudo-state:

$$\begin{aligned} x^t &= [y^t \dots\ y^{t-n_y+1} \quad u^{t-1} \dots\ u^{t-n_u+1}]\\ &= [x_{(1)}^t \dots\ x_{(n_y)}^t \quad x_{(n_y+1)}^t \dots\ x_{(n_y+n_u)}^t]. \end{aligned} \tag{28}$$

where $x^t \in \mathcal{X}$ is equal to $w^t \in \mathcal{W}$ without the input

sample at time t:

$$w^t = [x^t_{(1)} \dots\ x^t_{(n_y)} \quad u^t \quad x^t_{(n_y+1)} \dots\ x^t_{(n_y+n_u)}].$$

$\mathcal{W}$ and $\mathcal{X}$ are bounded sets in $\mathbb{R}^n$ and $\mathbb{R}^{n-m}$, respectively, with $\mathcal{W} \subseteq (\mathcal{X} \times \mathbb{R}^m)$.

The state space representation of the plant and the one-step prediction of the model are given by

$$\begin{aligned} &x^{t+1} = f_o(x^t, u^t) \\ &f_o(x^t, u^t) \doteq [f_o(w^t)\ x^t_{(1)} \dots\ x^t_{(n_y-1)}\ u^t \dots\ x^t_{(n_y+n_u-1)}] \end{aligned} \tag{29}$$

$$\begin{aligned} &\widehat{x}^{t+1}_c = f_c(x^t, u^t) \\ &f_c(x^t, u^t) \doteq [f_c(w^t)\ x^t_{(1)} \dots\ x^t_{(n_y-1)}\ u^t \dots\ x^t_{(n_y+n_u-1)}] \end{aligned} \tag{30}$$

In the notation above, $f(x^t, u^t)$ returns a vector of pseudo-states, while $f(w^t)$ returns a scalar. f_c is the central estimate of set membership model (Eq. (13)) and $\widehat{\cdot}$ represents the estimate given by the model.

From (29) and (30), it follows that

$$x^{t+1} = f_o(x^t, u^t) = f_c(x^t, u^t) + [e^t \quad 0 \ \dots\ 0]. \tag{31}$$

where the model uncertainty is described in terms of additive perturbation e^t, which is known to be bounded as

$$\left|e^t\right| \leqslant f_e(w^t) \leqslant \mathcal{R}^\infty_I \quad \forall w \in \mathcal{W}. \tag{32}$$

The sequence of inputs $\left\{u^i\right\}_{i=t}^{t+k-1}$, starting from a generic time instant t, up to a time instant $t+k-1$, is indicated with U^k_t. The state of the plant at time $t+k$ obtained starting from a generic "initial" state x^t and applying the input sequence U^k_t is defined as

$$\begin{aligned} &\mathcal{S}_o(x^t, U^k_t) \doteq x^{t+k} : \\ &\quad x^{t+n+1} = f_o(x^{t+n}, u^{t+n}) \quad \forall n \in [0, k-1]. \end{aligned} \tag{33}$$

The set of all possible plant state values at time $t+k$ that originate from a generic "initial" state x^t by applying the input sequence U^k_t to the system (31) is defined as

$$\begin{aligned} &\mathcal{S}(x^t, U^k_t) = \{\, \widehat{x}^{t+k} : \\ &\quad \widehat{x}^{t+n+1} = f_c(\widehat{x}^{t+n}, u^{t+n}) + [e^{t+n} \quad 0 \ \dots\ 0], \\ &\quad \left|e^{t+n}\right| \leqslant f_e(\widehat{w}^{t+n}), \forall n \in [0, k-1] \,\}. \end{aligned} \tag{34}$$

$$\widehat{w}^{t+n} = [\widehat{x}^{t+n}_{(1)} \dots\ \widehat{x}^{t+n}_{(n_y)} \quad u^{t+n} \quad \widehat{x}^{t+n}_{(n_y+1)} \dots\ \widehat{x}^{t+n}_{(n_y+n_u)}].$$

Note that this set is generated by all possible sequences $\{e^{t+n}\}_{n=0}^{k-1}$ such that $|e^{t+n}| \leqslant f_e(\widehat{w}^{t+n})$, for all $n \in [0, k-1]$. Clearly, it holds that $\mathcal{S}_o(x^t, U^k_t) \in \mathcal{S}(x^t, U^k_t)$. It is also true that if $\mathcal{R}^\infty_I = 0$ then $\mathcal{S}_o(x^t, U^k_t) = \mathcal{S}(x^t, U^k_t)$. The size of the set $\mathcal{S}(x^t, U^k_t)$ can be interpreted as the uncertainty of the state at time $t+k$. In other words, the uncertainty of the trajectory points, when a certain input sequence U^k_t is applied to the system.

Recalling the idea behind our approach from section 3.1, the aim is to reduce the radius of information by collecting measurements where the uncertainty is maximum. Thus, suppose that we want to take a measurement at a point w^r, or its equivalent (x^r, u^r), where the uncertainty amplitude is $f_e(w^r)$. Our approach consists in using a SMPC controller (to be defined later) to drive the plant state x^t to a neighborhood of x^r, called the reference set $\mathcal{X}_r \subset \mathbb{R}^{n-m}$, defined as

$$\mathcal{X}_r \doteq \{x : \Gamma \left\|x - x^r\right\|_2 + \varepsilon < \lambda f_e(w^r), \lambda \in (0, 1]\} \tag{35}$$

This set is a ball of radius $\frac{\lambda f_e(w^r) - \varepsilon}{\Gamma}$, centered at x^r; λ is a design parameter, allowing us to change the size of the reference set. When the state of the system x^t is inside $\mathcal{X}_r$ i.e. $x^t \in \mathcal{X}_r$, by applying u^r as input to the system and adding the new measurement to the dataset $\mathcal{D}$, the uncertainty $f_e(w^r)$ will be reduced by at least a factor of λ.

Assumption 3. For any "initial" state x^t and reference state x^r, there exists a control sequence U^k_t that moves the state from x^t to x^r:

$$\begin{aligned} \forall x^t, x^r \in \mathcal{X},\ &\exists K < \infty,\ \exists U^k_t \in \mathcal{U} : \\ &\mathcal{S}_o(x^t, U^k_t) = x^r \text{ for } k < K \end{aligned} \tag{36}$$

where $\mathcal{U}$ is the set of all possible input sequences compatible with the input constraints.

Assumption 4. For any input sequence U^k_t, the state of the system (29) remains inside the compact set $\mathcal{X}$:

$$\forall U^k_t \in \mathcal{U}, \forall t \geqslant 0, k \geqslant 1 : \quad x^t \in \mathcal{X}. \tag{37}$$

Assumption 3 is a standard controllability assumption. Assumption 4 is a mild boundedness assumption, just requiring that the system trajectory does not tend to infinity. Note that many identification and DoE approaches require stability, which is a stronger assumption than boundedness.

Let us define the set of potential trajectory horizons from x^t to x^r as follows:

$$\begin{aligned} \mathcal{I}(x^t, x^r) \doteq \{&i \in \mathbb{N} : i < K, \\ &\exists U^i_t \text{ such that } x^r \in \mathcal{S}(x^t, U^i_t)\}. \end{aligned} \tag{38}$$

For each element of the set $\mathcal{I}(x^t, x^r)$, there exists an input sequence such that $x^r \in \mathcal{S}(x^t, U^i_t)$. Assumption 3 ensures that this set is non-empty and finite for any initial and reference states. The optimization problem solved in the SMPC approach is:

$$
\begin{aligned}
J^*(x^t, x^r, i) &= \max_{U_t^i} J(x^t, x^r, U_t^i) \\
&\text{subject to } U_t^i \in \mathcal{U} \\
&\quad x^r \in \mathcal{S}(x^t, U_t^i) \\
J(x^t, x^r, U_t^i) &= \sum_{n=1}^{i} \operatorname{diam}(\mathcal{S}(x^t, U_t^n))
\end{aligned} \tag{39}
$$

where $i = \min\{\mathcal{I}(x^t, x^r)\}$ and $\operatorname{diam}(\mathcal{S})$ is the diameter of the set $\mathcal{S}$.

The controller is implemented according to a receding horizon strategy. The control law, indicated as $u^t = \mathcal{K}(x^t, x^r)$, means solving (39) and applying the first element of the maximizer U_t^{i*} as the control action u^{t*} to the system, and adding the new measurement to the set $\mathcal{D}$ ($\mathcal{D} := \mathcal{D} \cup \{\widetilde{y}^{t+1}, \widetilde{w}^t\}$). Then, these operations are repeated at each time t. The resulting control sequence, starting from a generic time instant t up to another instant k is denoted as $\mathcal{K}_t^k$.

Theorem 7. *Let Assumptions 3 and 4 hold. Starting from any initial state $x^t \in \mathcal{X}$, the state of the system controlled by the feedback law $\mathcal{S}_o(x^t, \mathcal{K}_t^k)$, will visit a point inside reference set $\mathcal{X}_r$ in finite time. That is,*

$$
\begin{aligned}
&\forall x^t, x^r \in \mathcal{X},\ \exists K < \infty : \\
&\qquad \mathcal{S}_o(x^t, \mathcal{K}_t^k) \in \mathcal{X}_r \ \ \textit{for some} \ \ k < K.
\end{aligned}
$$

Proof. Assumption 3 ensures that the set $\mathcal{I}$ is not empty and the optimization problem (39) is always feasible for all $i \in \mathcal{I}$. From the definition of the cost function, the following inequality holds for all feasible solutions.

$$
\begin{aligned}
&\forall U_t^i \quad \text{such that} \quad U_t^i \in \mathcal{U}, x^r \in \mathcal{S}(x^t, U_t^i) : \\
&\operatorname{diam}(\mathcal{S}(x^t, U_t^i)) \leqslant J^*(x^t, x^r, i).
\end{aligned} \tag{40}
$$

From (40) and the fact that $\mathcal{X}_r$ is a ball centered at x^r we can conclude that if $J^*(x^t, x^r, i) \leqslant \operatorname{diam}(\mathcal{X}_r)/2$, then the set $\mathcal{S}(x^t, U_t^i)$ is inside $\mathcal{X}_r$. i.e. $\mathcal{S}_o(x^t, U_t^i) \in \mathcal{S}(x^t, U_t^i) \subset \mathcal{X}_r$. This holds for all feasible solutions which means the state of the real system will be inside $\mathcal{X}_r$ in i steps. Therefore, in order to prove the theorem, we have to prove that

$$
\forall \varepsilon > 0, \exists K \quad \text{such that} \quad J_k^* < \varepsilon \quad \text{for} \quad k < K. \tag{41}
$$

where J_k^* is the cost computed at time instant k. At each time step we solve (39) and apply the first element of the maximizer U_t^{i*} as control action u^{t*} and add a new measurement to the dataset $\mathcal{D}$. The following inequalities hold when a new measurement is added

$$
\operatorname{diam}(\mathcal{S}_{k+1}(x^t, u^{t*})) = 2 f_e(w^t) < 2\varepsilon. \tag{42}
$$

$$
\mathcal{S}_{k+1}(x^t, u^{t*}) \subset \mathcal{S}_k(x^t, u^{t*}). \tag{43}
$$

$$
\operatorname{diam}(\mathcal{S}_{k+1}(x^t, u^{t*})) < \operatorname{diam}(\mathcal{S}_k(x^t, u^{t*})). \tag{44}
$$

where the subscript $k+1$ indicates a measurement is added to the dataset $\mathcal{D}$, which happens at each time step.

When a new measurement is added to the dataset $\mathcal{D}$, the uncertainty of the successive predicted states might also be reduced

$$
\mathcal{S}_{k+1}(x^t, U_t^n) \subseteq \mathcal{S}_k(x^t, U_t^n) \quad \forall n \in [2, i]. \tag{45}
$$

From (45), since the size of the uncertainty of the predicted states might be reduced, two things could happen. If $x^r \in \mathcal{S}_{k+1}(x^t, U_t^{i*})$, from (44) we have

$$
J_{k+1}^*(x^t, x^r, i_{k+1}) < J_k^*(x^t, x^r, i_k) \quad , \quad i_{k+1} = i_k. \tag{46}
$$

If $x^r \notin \mathcal{S}_{k+1}(x^t, U_t^{i*})$, which means U_t^{i*} is no longer a feasible solution. In such conditions, from the definition of the optimization problem, one of the following inequalities hold

$$
\operatorname{card}(\mathcal{I}_{k+1}) \leqslant \operatorname{card}(\mathcal{I}_k), \quad i_{k+1} > i_k. \tag{47}
$$

or

$$
J_{k+1}^*(x^t, x^r, i_{k+1}) < J_k^*(x^t, x^r, i_k), \quad i_{k+1} = i_k. \tag{48}
$$

Now consider the following function

$$
v(x^t, x^r) = \sum_{i \in \mathcal{I}(x^t, x^r)} J^*(x^t, x^r, i). \tag{49}
$$

From (46), (47), (48) and (49) we have

$$
\begin{aligned}
v_{k+1}(x^t, x^r) &< v_k(x^t, x^r) \\
v_{k+2}(x^{t+1}, x^r) &< v_{k+1}(x^{t+1}, x^r) \\
v_{k+3}(x^{t+2}, x^r) &< v_{k+2}(x^{t+2}, x^r) \\
&\dots
\end{aligned} \tag{50}
$$

Finally, consider the integral of the function v over the compact set $\mathcal{X}$

$$
V(x^r) = \int_{x \in \mathcal{X}} v(x, x^r) dx. \tag{51}
$$

From (49), (51) we can say $V(x^r)$ is a positive definite function $V(x^r) \geqslant 0$ since $J^* \geqslant 0$ and $V(x^r) = 0$ if and only if $\mathcal{R}_{\mathcal{I}}^{\infty} = 0$. From (50), (51) it holds that

$$
V_{k+1}(x^r) - V_k(x^r) < 0, \quad \forall k > 0. \tag{52}
$$

Therefore $\lim_{k \to \infty} V_k(x^r) = 0$ which is true if and only if $\lim_{k \to \infty} J_k^* = 0$. □

The dynamic Set Membership DoE is implemented in Algorithm 4. The algorithm is iterative. At each iteration, a reference regressor w^r is computed to be visited. Ideally this reference should be where f_e is maximum (similarly to Algorithm 3). However, if the reference is close to the estimated state, it can be visited more

Algorithm 4 Dynamic Set Membership DoE

(1) Select a reference regressor w^r to be visited which has a high uncertainty and its equivalent pseudo-state x^r is close to the estimated state:

$$w^r, x^r = \arg \min_{w^r \in \mathcal{W}, x^r \in \mathcal{X}} (\left\| \widehat{x}^{t+1} - x^r \right\|_2 + \frac{\delta}{f_e(w^r)}). \tag{53}$$

(2) Compute $\mathcal{X}_r$ according to (35) with a suitable λ.
(3) Apply the following criterion:

$$\begin{array}{ll} \text{if} & \widehat{x}^{t+1} \in \mathcal{X}_r \\ \text{then} & u^t = u^r \in w^r \\ \text{else} & u^t = \mathcal{K}(x^r, x^t). \end{array}$$

(4) Evaluate $\widetilde{y}^{t+1} = f_o(\widetilde{w}^t) = f_o(x^t, u^t)$.
(5) Add $\widetilde{y}^{t+1}$ and $\widetilde{w}^t$ to the dataset $\mathcal{D} := \mathcal{D} \cup \{\widetilde{y}^{t+1}, \widetilde{w}^t\}$.
(6) Update γ and Γ according to Algorithm 2.
(7) Set $t := t + 1$ and go to step (1).

quickly. Equation (53) combines these two objectives. In step 2, the reference set $\mathcal{X}_r$ is computed. In step 3, if the central estimate is inside the reference set, the input is generated according to the corresponding u^t of the vector w^r. Otherwise, the input is generated by the SMPC controller $\mathcal{K}$. Finally, at each iteration, a new measurement is taken and added to the measurement dataset $\mathcal{D}$ and the Lipschitz bounds γ, Γ are updated. To compute f_e in step 1 and 2, a global bound Γ is used and a quasi-local bound γ is used in step 3.

Corollary 1. For any desired radius of information $\mathcal{R}_d \geqslant \varepsilon$, there exists a finite number of steps T of Algorithm 4 such that $\mathcal{R}_{\mathcal{I}}(t) \leqslant \mathcal{R}_d$, for all $t \geqslant T$.

According to Theorem 6, if the system is static (which means it is possible to take a measurement anywhere in the regressor domain $\mathcal{W}$) then, Algorithm 3 can reach any desired radius of information. Also, according to Theorem 7, it is shown that, for a dynamic system, the SMPC controller can visit any desired reference point in the regressor domain. Algorithm 4 is the combination of Algorithm 3 and the SMPC controller assuming a large value of δ in equation (53). Thus, we can conclude that Algorithm 4 can reach any desired radius of information i.e. any desired worst-case error larger than the measurement error in a finite-time experiment.

3.4 Discussion and comparison with other DoE approaches

As mentioned in the introduction, not much research has been carried out in DoE for nonlinear systems in the set membership context, and the existing literature shows the following limitations: 1)The experiment design problem is reduced to an input selection problem [28]; 2)The method only works for a specific class of nonlinear systems [27,19]. 3)The exact model structure (given for example by the system physical equations) is assumed to be known [27,20,19]. 3)The parameter bounds are assumed to be known [27,20,19]. 3)If the assumed parameter bounds are conservative the algorithm does not work [27,18].

On the other side, due to its non-parametric nature, our DoE approach is not affected by these limitations. Indeed, the approach does not require to know the structure of the equations describing the system to identify and, as a consequence, no parameter bounds are required. We can say that our approach is more general than the other set membership DoE approaches for nonlinear systems available in the literature. Indeed, besides the use of less restrictive assumptions, our approach is independent on the particular identification method used. It can work with methods relying on physical models and methods involving black-box parameterizations (this latter case is considered in the examples presented below). On the contrary, the other existing approaches are usually based on a specific identification method (e.g., least squares or minimization of the worst-case parametric error).

It is worth mentioning that the philosophy behind our DoE approach is different from the one followed in linear DoE (in the time-domain). Indeed, our approach is non-parametric, and the goal is to explore the regressor domain in order to minimize the worst-case function error. On the contrary, in linear DoE, a parametric model structure is typically assumed, and the goal is to have convergence of the estimates (statistical setting), [17,35], or minimal worst-case parametric errors (set membership setting), [36]. For this reason, in the case of linear system, our approach is not equivalent to any other linear approach that we found in the literature. Note that, in principle, our approach can be effective also when the system to identify is linear. However, its utilization in the linear case may be not convenient due to a higher complexity with respect to the methods tailored specifically for linear systems.

4 Simulation Results

In this section, we present two simulation studies to illustrate the SM-DoE algorithm. The first example is concerned with a simulated nonlinear system, previously studied in [28]. The input signals obtained by the SM-DoE approach are compared with other input signals and with the optimal one, discussed in section 3.1. The second example, also studied in [6], is a nonlinear dynamic system with a static nonlinearity, where we compare the SM-DoE approach with three other DoE methods, taken from the literature.

4.1 Example 1

This section is concerned with the DoE for the following nonlinear dynamic system:

$$y^{t+1} = 0.88\, y^t - 0.12 \tanh(15\, y^t) + 0.06\, u^t. \tag{54}$$

Assuming the initial condition $y^1 = 0$. Three input signals have been considered:

$$\begin{aligned} U_{(1)} &= \{3\sin(0.2t), t = 1, 2, \ldots, T\}, \\ U_{(2)} &= \left\{3\sin(0.0009t^2), t = 1, 2, \ldots, T\right\}, \\ U_{(3)} &= \{WN(0, 4, t), t = 1, 2, \ldots, T\}. \end{aligned} \tag{55}$$

where $WN(0, 4, t)$ is a white gaussian noise of mean 0 and variance 4. For each of these signals, a simulation of the system (54) with length $T = 300$ was performed. The output signal was corrupted by a uniform random noise with amplitude $\leqslant 0.01$. The corresponding radius of information $\mathcal{R}_{\mathcal{I}}$ was computed. The involved regressor is

$$w^t = [y^t \; u^t].$$

The regressor domain of interest $\mathcal{W}$ is the rectangular region indicated in Figure 2 and defined by

$$\begin{aligned} \mathcal{W} \doteq \{w : w_1 \leq 0.35,\, w_1 \geq -0.35, \\ w_2 \leq 3.5,\; w_2 \geq -3.5\}. \end{aligned} \tag{56}$$

The values $\varepsilon = 0.01$ and $\gamma(\widetilde{w})$ were computed according to Algorithm 1,2. The ideal optimal input sequence $U_{Optimal}$ was computed according to (19), using Matlab Global Optimization Toolbox, providing the minimum possible radius of information. Finally, a fifth input signal U_{SM-DoE} was obtained, using the proposed SM-DoE algorithm. For each of the five input signals, a dataset was obtained.

In order to assess the quality of each dataset, a quasi-local set membership model was identified. Then, the prediction accuracy was validated on a 100×100 grid in the rectangular region defined in (56). The following accuracy indexes were considered to evaluate the model accuracy:

$$\begin{aligned} &\text{RMSE} = \|\widetilde{y} - \hat{y}\|_2 / \sqrt{N}. \\ &\text{FIT} = 100\left(1 - \frac{\|\widetilde{y} - \hat{y}\|_2}{\|\widetilde{y} - \text{mean}(\widetilde{y})\|_2}\right). \end{aligned} \tag{57}$$

where $\widetilde{y}$ indicates the measured output vector and $\hat{y}$ is the predicted output vector and N is the length of these two vectors.

Table 1 shows the radius of information and the accuracy on the validation set of the identified set membership models for each input signal.

Table 1
Radius of Information and set membership model accuracy corresponding to the input sequences.

Inputs	$U_{(1)}$	$U_{(2)}$	$U_{(3)}$	U_{Optimal}	$U_{\text{SM-DoE}}$
$\mathcal{R}_{\mathcal{I}}^{\infty}$	0.568	0.536	0.209	0.053	0.055
$\mathcal{R}_{\mathcal{I}}^{2}$	0.210	0.177	0.055	0.030	0.033
RMSE	0.0391	0.0350	0.0214	0.0061	0.0062
FIT	0.69	0.72	0.85	0.96	0.96

It can be noted from Table 1 that the optimal input sequence and the SM-DoE sequence provide much lower radius of information compared to the sinusoidal and random inputs. The fact that the data generated from $U_{Optimal}$ and U_{SM-DoE} provide lower radius of information, and consequently a higher identification accuracy, is related to the a more effective exploration of the regressor domain $\mathcal{W}$. This can be observed in Figure 2, where the "measured" regressors are shown for the five simulations.

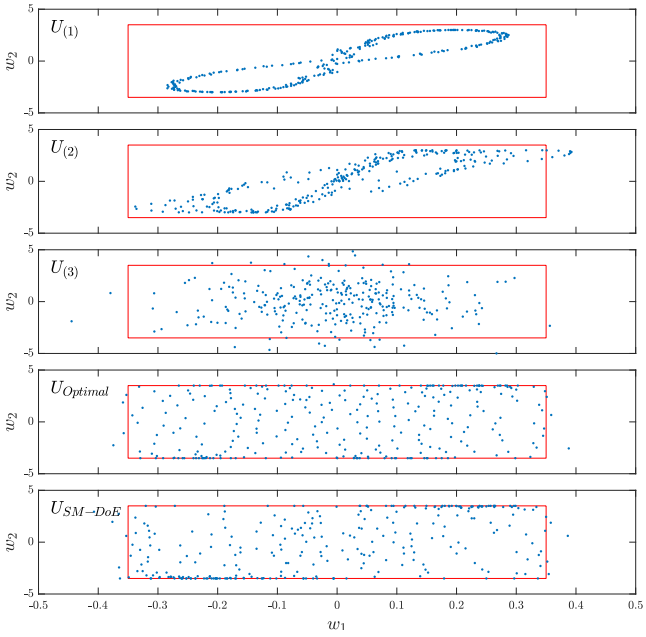

Fig. 2. Measured regressor $\{\widetilde{y}^t, \widetilde{u}^t\}_{t=1}^{300}$ for different input sequences.

Figure 3 shows the radius of information and model accuracy during the SM-DoE process. It can be seen that only half of the experiment was sufficient to derive an accurate model. It is also evident that reducing the radius of information directly leads to increasing the model accuracy.

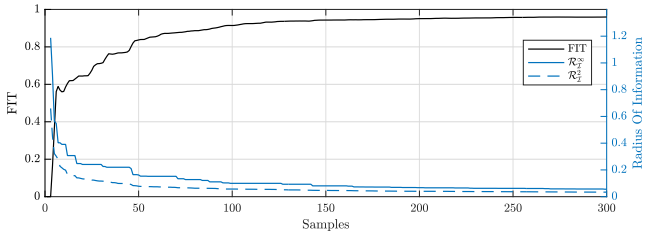

Fig. 3. Radius of Information and model accuracy during SM-DoE experiment.

4.2 Example 2

To evaluate the performance of proposed SM-DoE compared to other DoE methods, a simulation study was

performed considering a nonlinear dynamic system previously investigated by [6]. In this example, three different DoE methods, as well as the SM-DoE algorithm, were tested. The system under investigation is the following:

$$\dot{y} = g(y, x) = 2x/(2.4\cos(10x + 4) - 0.5y + 3.3). \quad (58)$$

where

$$x = f(u_1, u_2) = \cos(9\sqrt{u_1^2 + u_2^2} + 2) + 0.5\cos(11u_1 + 2) + 15((u_1 - 0.4)^2 + (u_2 - 0.4)^2)^2. \quad (59)$$

This system was discretized using the forward Euler method with a sampling time of 0.5 s. The system has two inputs, and the single output is corrupted by a uniform bounded noise of amplitude $\leqslant 0.025$. An illustration of the scaled functions of the system is shown in Figure 4. the function g of the discretized system is different from that of the continuous-time system.

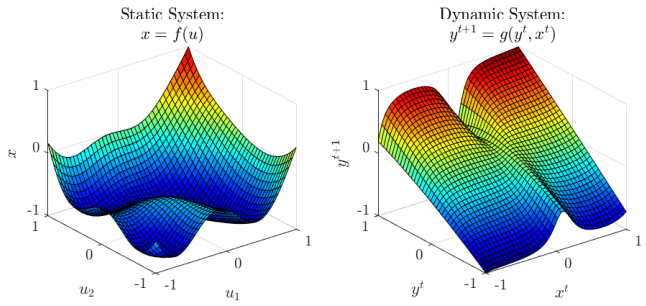

Fig. 4. Nonlinear Dynamic System.

In the SM-DoE algorithm, the regressor has been defined as $w^t = [y^t\, u_1^t\, u_2^t]$, and values $\varepsilon = 0.05$ and $\Gamma = 10, \gamma(\widetilde{w})$ were computed according to Algorithms 1 and 2. Algorithm 4 has been applied to the system according to Figure 5, with $\delta = 0.25$, $\lambda = 0.5$ and duration $T = 560s$.

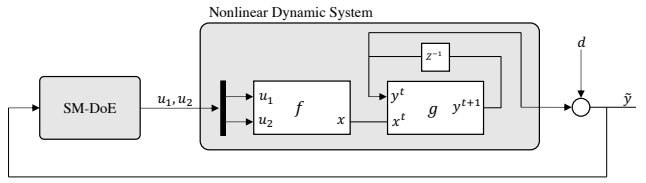

Fig. 5. Set Membership DoE scheme.

For comparison, three other DoE methods were also considered. As mentioned in the introduction, the most popular DoE methods for nonlinear dynamic systems are classified into two main categories: model-free and model-based methods [6,38,5,13]. The idea of these methods is to parameterize a pre-defined excitation signal, and then optimize the signal parameters called the design points, according to different criteria. For example, a widely applied excitation signal in industrial identification tasks is the amplitude modulated pseudo-random binary signal (APRBS). The APRBS signal is a sequence of N fixed steps with associated hold times T_{h_i} and amplitudes $a_i \in [u_{min}\, u_{max}]$ $i = 1, \ldots, N$. Since the values of the amplitudes are free design parameters in the following they are called design points. In Figure 6, a schematic APRBS signal in the time domain is plotted. Given the length T of the signal, the hold times T_{h_i} determine the number of steps and thus influences the frequency characteristics of the signal. It's important to choose an appropriate minimum hold time $T_{h_{min}}$ to assure that the system has a reasonable time to settle.

Besides the minimum hold time, the distribution of the design points a_i $i = 1, \ldots, N$ is essential for the quality of the excitation signal.

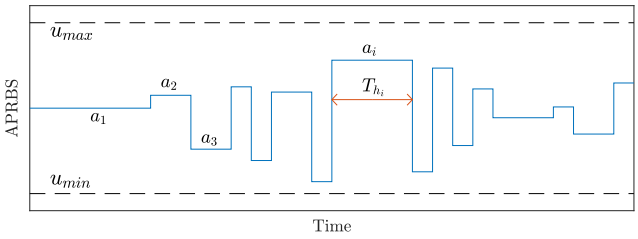

Fig. 6. APRBS signal in time domain.

In model-free DoE, no assumptions on the model structure are made. The typical approach is to distribute design points in the input domain as much uniformly as possible. This DoE approach is also known as space-filling DoE. The most popular space-filling DoE technique is based on the Latin Hypercube (LHC) distribution. To calculate an LHC distribution, the input space is divided into N intervals. In every column and row, only one design point is placed. Figure 7(b) shows an example of LHC distribution of 50 design points for a two-dimensional input space. In model-based DoE, after assuming a particular model structure, the idea is to distribute the design points in the input domain, in such a way that the estimation of the model parameters is as much insensitive as possible to the measurement noise. The most popular model-based DoE is the D-optimal distribution technique. Figure 7(c) shows an example of the D-optimal distribution technique with 50 design points based on a polynomial model of the third order.

After designing the distribution of the points in the input space, they are used for a parameterized excitation signal, like the above mentioned APRBS signal. Both model-free and model-based DoEs do not take into account the dynamics of the system and do not provide any indication about the optimal sequence of the design points.

In this example, three distributions (Random, LHC, D-optimal) have been constructed with 50 design points

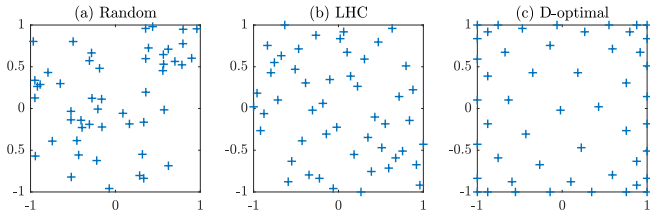

Fig. 7. distribution of 50 design points for two-dimensional input space.

that are shown in Figure 7. For each distribution, 10 different APRBS signals have been constructed with a random sequence of the design points with $T_{h_{min}} = 6s$ and duration of $T = 560s$. (A total of 30 input sequences). $T_{h_{min}}$ was chosen by trial and error which gave the best results.

For each set of data generated by the considered input sequences, a Neural-Network model has been identified with 18 sigmoid neurons. Table 2 shows the identified model accuracy for each DoE method. This table shows that the accuracy of the model derived from SM-DoE data is significantly higher.

Table 2
model accuracy mean and standard deviation.

Inputs	FIT	RMSE
Random	0.69 ± 0.03	0.149 ± 0.017
LHC	0.68 ± 0.04	0.155 ± 0.022
D-optimal	0.76 ± 0.01	0.115 ± 0.008
SM-DoE	0.91	0.043

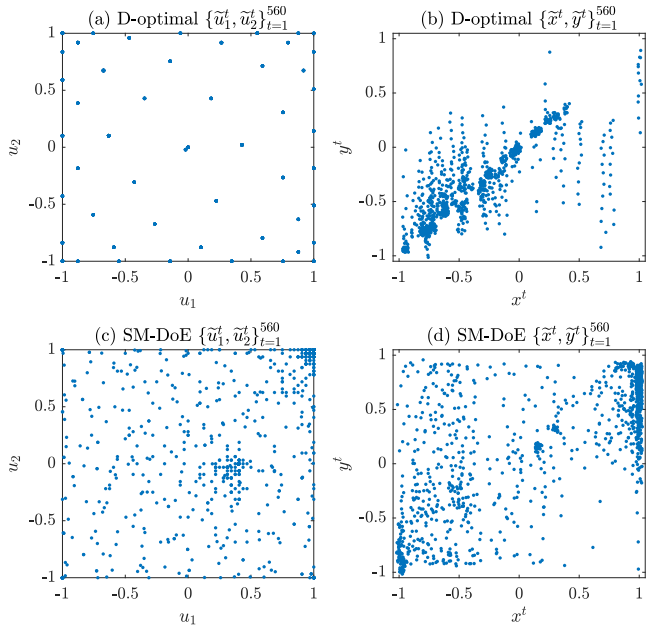

Fig. 8. Measured Regressor. (a),(b) D-optimal design, (c),(d) SM-DoE.

The measured regressors of the D-optimal design and the SM-DoE approach are shown in Figure 8. The domain is the same as the one in Figure 4. Both the experiments have the same duration. In the D-optimal case, although the input space is covered very well (In Fig. 8a, the design points are on top of each other due to the nature of the APRBS signal), we can see that the most nonlinear regions of the dynamic system are not explored (Fig. 8b), and the measurements are more concentrated around the diagonal which represents the steady state behavior of this system. Thus, the dynamic system nonlinearities are not captured by the data. On the other hand, the SM-DoE was able to better explore the whole regressor domain (Fig.8c, 8d).

5 Conclusion

The aim of this paper was to develop a systematic DoE method for nonlinear dynamic systems. We formulated the problem in a set membership framework and proposed a quasi-local nonlinear set membership approach that results in less conservative uncertainty bounds compared to the global approach. Then, we proposed a SM-DoE algorithm for input-constrained MISO nonlinear dynamic systems. The algorithm uses a novel SMPC controller to move the system toward the most uncertain regions of the regressor space and take new informative measurements. The proposed SM-DoE algorithm minimizes the worst-case model error. Thus, it is able to guarantee any desired worst-case error larger than the measurement error in a finite-time experiment. Applications of the proposed method are clearly most useful in areas where experiments are expensive and/or a very accurate model is desired. The DoE approach presented in this paper may also be of interest for future studies on adaptive data-driven nonlinear control design.

References

[1] M. Canale, L. Fagiano, and M. C. Signorile. Design of robust predictive control laws using set membership identified models. *Asian Journal of Control*, 15(6):1714–1722, 2013.

[2] M. Canale, L. Fagiano, and M. C. Signorile. Nonlinear model predictive control from data: a set membership approach. *International Journal of Robust and Nonlinear Control*, 24(1):123–139, 2014.

[3] M. Canale, C. Novara, and M. Milanese. NSM constrained approximation of lipschitz functions from data. *Systems & control letters*, 59(7):396–403, 2010.

[4] J. Chen and G. Gu. *Control-Oriented System Identification: An H_∞ Approach*. John Wiley & Sons, New York, 2000.

[5] M. Deflorian, F. Klöpper, and J. Rückert. Online dynamic black box modelling and adaptive experiment design in combustion engine calibration. *IFAC Proceedings Volumes*, 43(7):703–708, 2010.

[6] M. Deflorian and S. Zaglauer. Design of experiments for nonlinear dynamic system identification. *IFAC Proceedings Volumes*, 44(1):13179–13184, 2011.

[7] L. Fagiano and C. Novara. Learning a nonlinear controller from data: theory, computation and experimental results. *IEEE Transactions on Automatic Control*, 61(7):1854–1868, 2016.

[8] V. V. Fedorov. *Theory of optimal experiments*. Elsevier, 2013.

[9] U. Forssell and L. Ljung. Some results on optimal experiment design. *Automatica*, 36(5):749–756, 2000.

[10] S. T. Glad and L. Ljung. Model structure identifiability and persistence of excitation. In *29th IEEE Conference on Decision and Control*, pages 3236–3240. IEEE, 1990.

[11] G. C. Goodwin and R. L. Payne. *Dynamic system identification: experiment design and data analysis*, volume 136. Academic press New York, 1977.

[12] D. Gorinevsky. On the persistency of excitation in radial basis function network identification of nonlinear systems. *IEEE Transactions on Neural Networks*, 6(5):1237–1244, 1995.

[13] M. Gringard and A. Kroll. On the parametrization of APRBS and multisine test signals for the identification of nonlinear dynamic TS-models. *IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8, 2016.

[14] H. Hjalmarsson. From experiment design to closed-loop control. *Automatica*, 41(3):393–438, 2005.

[15] K. Hsu, C. Novara, T. Vincent, M. Milanese, and K. Poolla. Parametric and nonparametric curve fitting. *Automatica*, 42(11):1869–1873, 2006.

[16] M. Karimshoushtari. *Design of Experiments for Nonlinear System Identification*. PhD thesis, Politecnico di Torino, Torino, Italy, 2019.

[17] L. Ljung. System identification: theory for the user. *PTR Prentice Hall, Upper Saddle River, NJ*, pages 1–14, 1999.

[18] L. Lu and B. Yao. Set-membership identification based adaptive robust control of systems with unknown parameter bounds. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 847–852. IEEE, 2009.

[19] L. Lu and B. Yao. Experimental design for identification of nonlinear systems with bounded uncertainties. In *Proceedings of the 2010 American Control Conference*, pages 4504–4509. IEEE, 2010.

[20] S. W. Marvel and C. M. Williams. Set membership experimental design for biological systems. *BMC systems biology*, 6(1):21, 2012.

[21] M. Milanese, J. Norton, H. Piet-Lahanier, and É. Walter. *Bounding Approaches to System Identification*. Plenum Press, 1996.

[22] M. Milanese and C. Novara. Set membership identification of nonlinear systems. *Automatica*, 40(6):957–975, 2004.

[23] M. Milanese and C. Novara. Computation of local radius of information in SM-IBC identification of nonlinear systems. *Journal of Complexity*, 23:937–951, 2007.

[24] M. Milanese and C. Novara. Unified set membership theory for identification, prediction and filtering of nonlinear systems. *Automatica*, 47(10):2141–2151, 2011.

[25] M. Milanese and M. Taragna. H_∞ set membership identification: A survey. *Automatica*, 41:2019–2032, 2005.

[26] M. Milanese and A. Vicino. Optimal algorithms estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, 27:997–1009, 1991.

[27] A. R. G. Mukkula and R. Paulen. Model-based design of optimal experiments for nonlinear systems in the context of guaranteed parameter estimation. *Computers & Chemical Engineering*, 99:198–213, 2017.

[28] C. Novara. Experiment design in nonlinear set membership identification. In *American Control Conference*, pages 1566–1571. IEEE, 2007.

[29] C. Novara. Sparse set membership identification of nonlinear functions and application to fault detection. *International Journal of Adaptive Control and Signal Processing*, 30(2):206–223, 2016.

[30] C. Novara, M. Canale, M. Milanese, and M. C. Signorile. Set Membership inversion and robust control from data of nonlinear systems. *International Journal of Robust and Nonlinear Control*, 24(18):3170–3195, 2014.

[31] F. Randy and K. P. V. *Robust Nonlinear Control Design*. Birkhäuser Basel, Boston, 1996.

[32] M. C. Signorile. *Robust control of nonlinear systems from data*. PhD thesis, Politecnico di Torino, 2012.

[33] T. W. Simpson, D. K. Lin, and W. Chen. Sampling strategies for computer experiments: design and analysis. *International Journal of Reliability and Applications*, 2(3):209–240, 2001.

[34] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31:1691–1723, 1995.

[35] P. Stoica and T. Soderstrom. A useful input parameterization for optimal experiment design. *IEEE Transactions on Automatic Control*, 27(4):986–989, 1982.

[36] M. Tanaskovic, L. Fagiano, and M. Morari. On the optimal worst-case experiment design for constrained linear systems. *Automatica*, 50(12):3291–3298, 2014.

[37] M. V. Tanaskovic. *Application of Set Membership Identification to Controller Design*. PhD thesis, ETH Zurich, 2015.

[38] S. Zaglauer. Bayesian design of experiments for nonlinear dynamic system identification. In *Proceedings of the 5th International ICST Conference on Simulation Tools and Techniques*, pages 85–92. ICST (Institute for Computer Sciences, Social-Informatics and ..., 2012.

[39] M. B. Zarrop. *Optimal experiment design for dynamic system identification*, volume 21. Springer, 1979.

[40] Q. Zhihua. *Robust Control of Nonlinear Uncertain Systems*. Wiley series in nonlinear science, 1998.