

“© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

# D-Optimal Design for Information Driven Identification of Static Nonlinear Elements

Nalika Ulapane  
Swinburne University of Technology  
Hawthorn, Victoria, 3122, Australia  
aulapane@swin.edu.au  
nalika.ulapane@gmail.com

Sarath Kodagoda  
UTS Robotics Institute  
University of Technology Sydney  
Ultimo, Sydney, New South Wales, 2007, Australia  
Sarath.Kodagoda@uts.edu.au

Karthick Thiyagarajan  
UTS Robotics Institute  
University of Technology Sydney  
Ultimo, Sydney, New South Wales, 2007, Australia  
karthick.thiyagarajan@uts.edu.au

Linh Nguyen  
School of Engineering, IT and Physical Sciences  
Federation University Australia  
Churchill, Victoria, 3842, Australia  
l.nguyen@federation.edu.au

**Abstract**—Identification of static nonlinear elements (i.e., nonlinear elements whose outputs depend only on the present value of inputs) is crucial for the success of system identification tasks. Identification of static nonlinear elements though can pose several challenges. Two of the main challenges are: (1) mathematical models describing the elements being unknown and thus requiring black-box identification; and (2) collection of sufficiently informative measurements. With the aim of addressing the two challenges, we propose in this paper a method of predetermining informative measurement points offline (i.e., prior to conducting experiments or seeing any measured data), and using those measurements for online model calibration. Since we deal with an unknown model structure scenario, a high order polynomial model is assumed. Over fit and under fit avoidance are achieved via checking model convergence via an iterative means. Model dependent information maximization is done via a D-optimal design of experiments strategy. Due to experiments being designed offline and being designed prior to conducting measurements, this method eases off the computation burden at the point of conducting measurements. The need for in-the-loop information maximization while conducting measurements is avoided. We conclude by comparing the proposed D-optimal design method with a method of in-the-loop information maximization and point out the pros and cons. The method is demonstrated for the single-input-single-output (SISO) static nonlinear element case. The method can be extended to MISO systems as well.

**Index Terms**—Black box, design of experiments, DoE, D-Optimal, estimation, nonlinear systems, optimal design, optimization, over-fitting, system identification

## I. INTRODUCTION

System identification of static nonlinear elements (i.e., nonlinear elements whose outputs depend only on the present value of inputs) is crucial for identifying accurate models for control applications of dynamical systems [1]. A commonly encountered type of such systems include linear dynamics and static nonlinear elements [2]. Identification of some static elements is known to be time consuming, especially in automobile engine calibration [3]. This is typically down to having to take measurements when system responses have reached a

steady state. Furthermore, some identification tasks of static nonlinear elements warrant black-box identification as explicit model structures may be unknown. Due to such challenges it is desirable to reduce the time spent on measurement-taking. A possible way to accomplish this is by enabling the collection of a limited amount of data that is maximally informative in some sense for the calibration of a particular model. Our paper contributes to this space. This paper combines D-Optimal design of experiments to calculate informative measurements and efficient in-the-loop polynomial model calibration.

Previous work attempting information maximization focused efficient measurement taking have often relied on Fisher Information [4], [5] driven design of experiments [1], [4], [6], and so-called Active Learning strategies that have adopted techniques like ‘query by committee’ [7] and Gaussian Process (GP) [8] based uncertainty reduction [9], [10]. From such approaches, some quite powerful ones are the in-the-loop information maximization methods [9]. What such methods often do, is for any given prior set of measurements, they calculate the ‘next best measurement’ via information maximization in-the-loop with simultaneous measurement-taking and model calibration [9]. Although such methods have shown to accomplish the task of optimizing measurement-taking while identification of adequate models, they can add significant computation and real-time data interfacing challenges for the measurement-taking process [9]. The reason behind computation cost of such methods attribute to factors like the use of techniques like GP, or more generally, taking leads from Bayesian Optimization.

Drawing niche from there, our paper is attempting to relax the computation burden such in-the-loop methods pose. To accomplish this, we try to avoid in-the-loop information maximization. Instead, we try to design informative input sequences offline, prior to taking measurements and seeing any data. As such, we will have predetermined informative input sequences at hand for measurement taking, and when

taking measurements, we only try to calibrate an easy-to-calibrate polynomial model, and check the convergence of model parameter estimates.

The structure of the paper is as follows: Section II presents the relevant problem formulation mathematically; Section III describes step-by-step the algorithmic workflow that has to be followed to accomplish informative input design, measurement-taking, and model convergence; Section IV discusses the implication of the results, and Section V concludes this paper identifying the merits and demerits of the proposed method.

## II. PROBLEM FORMULATION

Suppose there is a static nonlinear SISO element in the form of  $y = f(x) + \epsilon$  where  $x \in \mathbb{R}$  is the input,  $y \in \mathbb{R}$  is the output and  $\epsilon \in \mathbb{R}$  is noise. For example, suppose this element is a component of a real-world dynamical system from which certain inputs and outputs can be measured. The specific assumptions and constraints regarding the element and noise are given in the subsequent sections.

Now, although the input  $x$  and the output  $y$  are accessible, the structure of the mathematical function  $f$  performing the operation of the element remains unknown. Thus, to facilitate modeling of such an element, the problem at hand becomes to find a model of the form  $\hat{y} = f_\theta(x)$ , such that the model adequately describes the element in question. Here,  $\theta \in \mathbb{R}^m, m \in \mathbb{Z}^+$  is the set of model parameters. Since the structure of  $f$  is unknown, determining a model  $f_\theta$  becomes an empirical task, and thus it essentially becomes a black-box identification task.

The strategy we adopt in this paper to solve this problem can be summarized as follows. First, decide on a model architecture  $f_\theta$ . Second, determine a sequence of inputs  $x_1, x_2, \dots, x_N$  such that the vector  $\tilde{x}_N$  is given by

$$\tilde{x}_N = [x_1, x_2, \dots, x_N]^T \quad (1)$$

for  $\tilde{x}_N \in \mathbb{R}^N$ , that is maximally informative in some sense with respect to the selected model (D-optimal design-based information maximization [1], [4] is adopted in this paper). One constraint that has to be respected, is that once constructed, the order of elements in  $\tilde{x}_N$  should not be changed since the elements of  $\tilde{x}_N$  are arranged in an order of informativeness. Third, conduct an adequate amount of measurements  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  such that

$$\tilde{x}_n = [x_1, x_2, \dots, x_n]^T \quad (2)$$

for  $\tilde{x}_n \in \mathbb{R}^n, \tilde{x}_n \subset \tilde{x}_N, [\dots]^T$  denoting matrix transpose, and

$$\tilde{y}_n = [y_1, y_2, \dots, y_n]^T \quad (3)$$

for  $\tilde{y}_n \in \mathbb{R}^n$  that results in the convergence of  $\theta$  estimates calculated as follows.

$$\theta^* = \arg \min_{\theta} \frac{1}{n} (\tilde{y}_n - f_\theta(\tilde{x}_n))^T (\tilde{y}_n - f_\theta(\tilde{x}_n)) \quad (4)$$

The solution  $\theta^*$  obtained in that manner would yield the model  $f_\theta$  that represents the system  $f$ , completing the system

identification objective that was set out initially. Section III describes the methodology proposed to achieve the aforesaid goal.

## III. METHODOLOGY

As set out in the Introduction, the methodology presented in this section covers a SISO static nonlinear element case. The methodology is composed of the steps presented as subsections within this section. These steps should be followed sequentially.

### A. Step 1: Know the element variables

The first step is to know what the inputs and outputs are to the element and whether they are measurable. With respect to the SISO case focused in this paper, we would know that the element is in the form

$$y = f(x) + \epsilon \quad (5)$$

where  $x \in \mathbb{R}$  is the input,  $y \in \mathbb{R}$  is the output and  $\epsilon \in \mathbb{R}$  is noise. The caveat is that the structure of the function  $f$  will be unknown at this stage. Thus, for modeling purposes, black-box identification becomes warranted.

The element in question will have to satisfy the following constraints in order to enable model identification as proposed in this paper.

- 1)  $x$  and  $y$  are measurable.
- 2)  $y$  is a continuous and smooth function of  $x$ , but the structure of the function  $f(\cdot)$  is unknown.
- 3)  $x$  has a range  $x_{min} < x < x_{max}$ , where  $x_{min}, x_{max} \in \mathbb{R}$ , and the values of  $x_{min}$  and  $x_{max}$  are known.
- 4) The output remains bounded  $\forall x$  in  $x_{min} < x < x_{max}$ ; implying that  $|y| < y_{max}, \forall y$ , where  $|*|$  denotes absolute value and  $y_{max} \in \mathbb{R}^+$ ; however the value of  $y_{max}$  may be unknown.
- 5) Noise  $\epsilon$  is zero mean white noise where  $|\epsilon| < \epsilon_{max}, \forall \epsilon$ , and  $\epsilon_{max} \in \mathbb{R}^+$ ; however, the value of  $\epsilon_{max}$  may be unknown.
- 6)  $\frac{\epsilon_{max}}{y_{max}} < \gamma$  where  $\gamma \in \mathbb{R}^+, \gamma \ll 1$ , however, the value of  $\gamma$  may be unknown.

### B. Step 2: Select a model structure

The problem in focus deals with an unknown model structure. An element providing input-output relationship of the form  $f, f: \mathbb{R} \rightarrow \mathbb{R}$  has to be modeled (see (5)).

The Weierstrass Approximation Theorem states that a continuous real-valued function defined on a real interval can be uniformly approximated as closely as desired by a polynomial function [11]. Following that theorem, we choose a polynomial  $f_\theta, f_\theta: \mathbb{R} \rightarrow \mathbb{R}$  of the following form

$$\hat{y} = f_\theta(x) \quad (6)$$

where

$$f_\theta(x) = \sum_{i=0}^m \theta_m x^m \quad (7)$$

to model the element in question. When the set of model parameters is denoted as

$$\theta = [\theta_0, \theta_1, \dots, \theta_m]^T \quad (8)$$

where  $\theta \in \mathbb{R}^{m+1}$ ,  $m \in \mathbb{Z}^+$ , and a set of output estimates is denoted as

$$\hat{y} = [\hat{y}_1, \hat{y}_2, \dots]^T \quad (9)$$

the aforesaid model can be expressed in matrix form in the following manner

$$\hat{y} = X\theta \quad (10)$$

where  $X$  is a matrix of which a generic  $i^{\text{th}}$  row denoted by  $\bar{X}_i$  is given as follows.

$$\bar{X}_i = [1, x_i, x_i^2, \dots, x_i^m] \quad (11)$$

When using the proposed method, the order of the model, i.e.,  $m$ ,  $m \in \mathbb{Z}^+$  has to be chosen by the user based on some a priori knowledge or some reasonable judgment about the expected complexity of the element. A safe approach to follow is to select a ‘high’ order model to start with. The parameter estimation steps described in the coming subsections will overcome any likely over-fitting. However, the authors are not able to prescribe a particular number that is reflective of ‘high’ order at this stage. As said before, this becomes a number that should be decided by the user based on any a priori knowledge and sensible judgment about the anticipated complexity of the element in question.

### C. Step 3: Compute informative measurement points

This computation of informative measurements is an iterative process done offline prior to taking any measurements and seeing any measured data. The model structure of adequate order (i.e.,  $f_\theta$ ) selected as per subsection III-B will be the sole guide for this computation. In addition to that, the constraints and the relevant knowledge about the element as mentioned in subsection III-A will have to be satisfied.

To start with, we define a value  $N$ ,  $N \in \mathbb{Z}^+$ , and a sequence given by  $\tilde{x}_N = [x_1, x_2, \dots, x_N]^T$  as in (1), that contains ALL possible measurement points  $x_i$  that can be measured, where  $x_i \in \mathbb{R}$ , and  $x_{\min} \leq x_i \leq x_{\max} \forall x_i \in \tilde{x}_N$ . This means, if we do not have time or any other constraints, an ideal and maximally informative measurement task would involve measuring the whole sequence  $\tilde{x}_N$ .

For the benefit of subsequent computations, we impose each element in  $\tilde{x}_N$  to be unique (i.e.,  $x_i \neq x_j$  for  $i \neq j \forall x_i, x_j \in \tilde{x}_N$ ). Furthermore, we impose  $\tilde{x}_N$  to be in ascending order, such that  $x_1 = x_{\min}$ ,  $x_N = x_{\max}$ , and  $x_i < x_{i+1} \forall x_i, x_{i+1} \in \tilde{x}_N$ . Also, we impose all adjacent elements to be equally spaced (i.e.,  $|x_{i-1} - x_i| = |x_i - x_{i+1}| \forall x_{i-1}, x_i, x_{i+1} \in \tilde{x}_N$ ).

Now suppose, limitation of time or another constraint is prohibiting us from measuring the full sequence  $\tilde{x}_N$ . Therefore, we would like to measure a subset of  $\tilde{x}_N$ , that is maximally informative. Constructing such a maximally informative subset is the next immediate task at hand. The following approach is proposed to construct such an informative subset.

Select  $k$ ,  $k \in \mathbb{Z}^+$ ,  $k \ll N$  arbitrary measurement points from  $\tilde{x}_N$  as prior measurements, and construct an arbitrary sequence  $\tilde{x}_k$  given by

$$\tilde{x}_k = [x_1, x_2, \dots, x_k]^T \quad (12)$$

where  $\tilde{x}_k \subset \tilde{x}_N$ . The value for  $k$  here can be decided upon the user’s discretion and the number can be as low as  $k = 1$ .

The matrix  $X_k$  should then be constructed using the set  $\tilde{x}_k$ .  $X_k$  come as

$$X_k = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & x_2^3 & \dots & x_2^m \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_k & x_k^2 & x_k^3 & \dots & x_k^m \end{bmatrix} \quad (13)$$

from which the set of model output estimates  $\hat{y}_k$  corresponding to the set of input points  $\tilde{x}_k$  can be written as

$$\hat{y}_k = X_k \theta \quad (14)$$

for a set of model parameters  $\theta$ ,  $\theta \in \mathbb{R}^{m+1}$ ,  $m \in \mathbb{Z}^+$  (as given in (8)) that has to be estimated.

For a given  $\tilde{x}_k$  and a corresponding matrix  $X_k$ , we would now like to find the next most informative measurement point  $x_{k+1}$  where  $x_{k+1} \in \tilde{x}_N \setminus \tilde{x}_k$ , such that

$$x_{k+1} := [\tilde{x}_k^T, x_{k+1}]^T \quad (15)$$

becomes maximally informative in some sense for calibrating the selected  $f_\theta$  model. Since a polynomial model structure is dealt with, D-optimal design [1], [4] is adopted as the strategy for maximizing information.

For an available set  $\tilde{x}_k$  and the corresponding matrix  $X_k$ , the next most informative measurement point  $x_{k+1}^*$  is calculated as

$$x_{k+1}^* = \arg \max_{x_{k+1}} \frac{1}{k+1} |X_{k+1}^T X_{k+1}| \quad (16)$$

where  $|*|$  in (16) denotes the matrix determinant and the matrix  $X_{k+1}$  is given by

$$X_{k+1} = \begin{bmatrix} & & X_k & & & \\ 1 & x_{k+1} & x_{k+1}^2 & x_{k+1}^3 & \dots & x_{k+1}^m \end{bmatrix} \quad (17)$$

for any  $x_{k+1} \in \tilde{x}_N \setminus \tilde{x}_k$ . The solution for (16) becomes the solution for the D-optimal design problem [1], [4].

By studying equations (12) to (16) it can be seen that by starting at any available sequence  $\tilde{x}_k$ , the process can be incremented to find a corresponding  $\tilde{x}_{k+1}$ . Thus, the process is iterative, and can be continued until  $k$  reaches its maximum allowable value  $N$  as was specified before. Thus the integer  $N$  would act as a stopping criteria for the iterative generation of informative measurement points. In a different perspective, if continued until  $k$  reaches  $N$ , this process can be understood as a rearrangement of the points in  $\tilde{x}_N$  in an order of informativeness with respect to the selected model  $f_\theta$ . The rearranged version of  $\tilde{x}_N$  constructed in that manner will henceforth be denoted as  $\tilde{x}_N^*$ .

#### D. Step 4: Collecting measurements while calibrating model

Having computed  $N$  informative measurement points as described in subsection III-C and being made available in the arrangement of  $\tilde{x}_N^*$ , the objective now should be to collect a subset of those measurements. This subset should lead to some convergence of parameter (i.e.,  $\theta$ ) estimates with respect to the selected model (i.e.,  $f_\theta$ ).

It should be noted that when collecting a subset of measurements, the order of elements in the arranged sequence  $\tilde{x}_N^*$  cannot be changed. Collecting a subset of measurements now implies that starting measuring from the first element of  $\tilde{x}_N^*$  and proceeding sequentially through to the last element of  $\tilde{x}_N^*$  until model convergence is achieved. The following strategy is proposed to perform this task.

It is proposed to select three constants  $p$ ,  $q$ , and  $r$  such that

$$\frac{N}{p} = q + \frac{r}{p} \quad (18)$$

where  $p, q, r \in \mathbb{Z}^+$  and  $r$  is the remainder when  $N$  is divided by  $p$ . Here,  $p$  is a batch size. This means, when conducting measurements,  $p$  sequential and distinct measurements will be collected at a stretch. A set of  $p$  measurements collected likewise is known as a batch of measurements.

To account for noise, authors propose each measurement in a batch of  $p$  being repeated  $s$ ,  $s \in \mathbb{Z}^+$  times, and the mean of the  $s$  repetitions being taken as a single measurement as done in [9].

The values of  $p$  and  $s$  have to be decided by the user. It makes sense to select  $p$  to be

$$p \ll N \quad (19)$$

as this will make number of possible batches (i.e.,  $q$ ) larger. The choice for the value of  $s$  will often depend on how much time can be afforded for collecting measurements.

The idea now is to iterate along collecting batches of measurements. According to (18) it will be possible to measure  $q$  batches at maximum. We will next define a stopping criteria that will enable stopping measurements at fewer batches while guaranteeing a degree of satisfactory model calibration.

#### E. Stopping criteria

Suppose  $q_0, q_0 \in \mathbb{Z}^+$ ,  $1 < q_0 < q$ , batches of measurements have been collected. In this subsection we will inquire what the stopping criteria will be if the measurement taking is to stop at  $q_0$  batches while guaranteeing a degree of model convergence. If in any way the defined stopping criteria cannot be reached, measurement-taking is set to proceed until all measurements in  $\tilde{x}_N^*$  are taken.

Consider the case of having measured  $q_0$ ,  $1 < q_0 < q$  batches. The total number of measurements available then will be  $u$ ,  $u \in \mathbb{Z}^+$  given as follows.

$$u = p \times q_0 \quad (20)$$

The corresponding measurement sets  $\tilde{x}_u$  and  $\tilde{y}_u$  will be given in the following form.

$$\tilde{x}_u = [x_1, x_2, \dots, x_u]^T \quad (21)$$

$$\tilde{y}_u = [y_1, y_2, \dots, y_u]^T \quad (22)$$

The first  $v$  rows,  $v \in \mathbb{Z}^+$ ,  $v < u$ , of the vectors  $\tilde{x}_u$  and  $\tilde{y}_u$ , or conversely, the first  $v$  informative measurements, are then isolated into the vectors  $\tilde{x}_v$  and  $\tilde{y}_v$  given as follows.

$$\tilde{x}_v = [x_1, x_2, \dots, x_v]^T \quad (23)$$

$$\tilde{y}_v = [y_1, y_2, \dots, y_v]^T \quad (24)$$

The values  $u$  and  $v$  will yield a fraction  $\delta$  given by

$$\delta = \frac{v}{u} \quad (25)$$

where  $\delta < 1$ . A value for  $\delta$  has to be chosen by the user and a rationale for the choice of value is discussed later in this subsection.

Following (13), the matrices  $X_u$  and  $X_v$  should then be constructed from the sets  $\tilde{x}_u$  and  $\tilde{x}_v$ . The corresponding sets of parameters  $\theta_u$  and  $\theta_v$  are then estimated as follows (recall the shape of  $\theta$  from (8)).

$$\theta_u = \arg \min_{\theta} \frac{1}{u} (\tilde{y}_u - X_u \theta)^T (\tilde{y}_u - X_u \theta) \quad (26)$$

$$\theta_v = \arg \min_{\theta} \frac{1}{v} (\tilde{y}_v - X_v \theta)^T (\tilde{y}_v - X_v \theta) \quad (27)$$

$\theta_u$  and  $\theta_v$  are denoted as follows.

$$\theta_u = [\theta_u^{(0)}, \theta_u^{(1)}, \dots, \theta_u^{(m)}]^T \quad (28)$$

$$\theta_v = [\theta_v^{(0)}, \theta_v^{(1)}, \dots, \theta_v^{(m)}]^T \quad (29)$$

Now, we define a cost functions  $Diff(u)$  such that

$$Diff(u) = \frac{1}{m+1} \sum_{j=0}^m \frac{|\theta_u^{(j)} - \theta_v^{(j)}|}{|\theta_u^{(j)}|} \quad (30)$$

which is defined for  $\theta_u^{(j)} \neq 0$ .  $|\ast|$  in (30) denotes the absolute value.

The stopping criteria is defined for measurement-taking to stop when the condition in (31)

$$Diff(u) \leq d \quad (31)$$

becomes satisfied for a threshold value  $d$ ,  $d \in \mathbb{R}^+$ ,  $d \ll 1$  specified by the user, subject to

$$u \geq m + 1 \quad (32)$$

given that the order of the assumed model is  $m$  (recall model order from (7)).

The intuition of this stopping criteria is that the parameter set  $\theta_v$  estimated from the first  $v$  informative measurements, is on average within  $100d\%$  of the parameter set  $\theta_u$  estimated from the first  $u$  informative measurements (recall that  $u > v$  from (25)). Thus, the stopping criteria indicates the convergence of the estimated model parameter values. Smaller the value of  $d$ , greater the convergence will be, and the harder to achieve.

The value of  $\delta$  (in (25)) that has to be set by the user, indicates what fraction  $v$  will be with respect to  $u$ . A popular

choice can be  $\delta = 0.7$ , following the common 70:30 (which is sometimes exploited as 70:100) training and testing data split regularly practiced in machine learning exercises [12]. Generally, a range of  $0.5 < \delta < 1$  should be adequate. But, one would want to avoid getting too close to 0.5 and too close to 1. Thus, around the  $\sim 0.7$  mark can be considered a sensible choice.

When the stopping criteria is reached either by way of  $Diff(u) \leq d$  or the measurement count reaching  $N$ , the resulting identified model can be expressed as a set of parameters in the following manner.

$$\theta = \theta_u \quad (33)$$

This parameter set  $\theta$  would yield a model  $f_\theta$  (recall its shape from (7)) that represents the nonlinear element  $f$  in question, accomplishing the system identification task.

#### IV. RESULTS

##### A. The static nonlinear element

The methodology was experimented by trying to identify a model for the static nonlinear element given in (34). The element's properties are given in Table I.

$$y = \exp(-2x) \sin(7\pi x) + \exp(-3x) + 0.5 + \epsilon. \quad (34)$$

TABLE I  
THE IMPORTANT PROPERTIES OF THE ELEMENT IN (34).

Parameter	Symbol	Value
Minimum of $x$	$x_{min}$	0
Maximum of $x$	$x_{max}$	1
Noise Threshold	$\gamma$	0.05

##### B. Computing informative measurements

The parameters set for computing informative measurements are given in Table II.

TABLE II  
THE PARAMETERS SET FOR COMPUTING INFORMATIVE MEASUREMENTS

Parameter	Symbol	Value
Model order	$m$	15
Maximum number of measurements	$N$	1,001

##### C. Taking measurements

Table III provides the parameters set for measurement-taking and termination. Once the informative input sequence was computed, the termination of the corresponding measurement-taking process was investigated with noisy measurements. The number of measurements taken to reach the termination criteria across 50 repetitions of this measurement-taking process are plotted as a histogram in Figure 1.

TABLE III  
PARAMETERS SET FOR MEASUREMENT-TAKING AND TERMINATION

Parameter	Symbol	Value
Model order	$m$	15
Maximum number of measurements	$N$	1,000
Batch size	$p$	5
Number of batches	$q$	200
Model convergence data ratio	$\delta$	0.7
Termination threshold	$d$	0.01

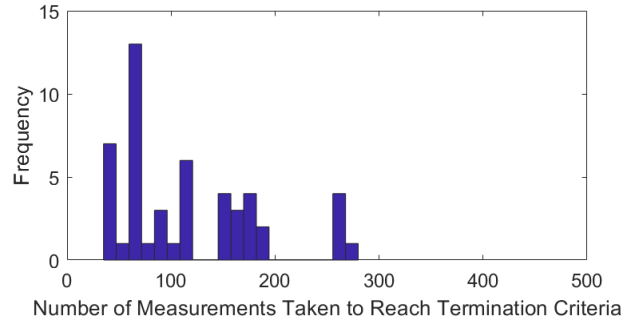


Fig. 1. Histogram over 50 trials from the work of this paper showing the number of measurements taken for the model to converge, mean = 114 measurements, std = 70 measurements.

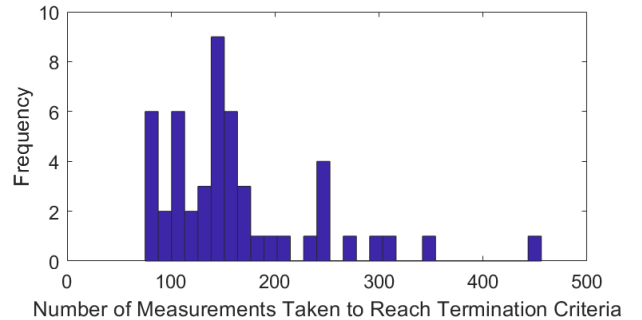


Fig. 2. Histogram over 50 trials from [9], showing the number of measurements taken for the model to converge, mean = 164 measurements, std = 76 measurements.

Figure 2 (taken from [9]) provides a histogram similar to that of Figure 1. The histogram of Figure 2 shows the number of measurements taken to reach the termination criteria across 50 repetitions of identifying a model for the same element in (34), via an in-the-loop information maximization method that uses computationally intensive Gaussian Process (GP) regression [8], [13]. The work of [9] and the results of Figure 2 also stem from a termination criteria of 0.01 indicative of model convergence (similar to the value of  $d$  in this work, see Table III).

A downside of the in-the-loop information maximization method of [9], is the repetitive use of GP while taking measurements. This poses a computation and data interfacing burden to measurement-taking, as on one side GP is computationally intensive, and on the other side access to data

is required in real-time to calculate next best informative measurement points. Unlike that method in [9], the method proposed in our present paper does the informative input design offline before conducting any measurements or seeing any data. Therefore, an advantage of the present method is its ability to reduce the computation and interfacing burden that prevail with in-the-loop methods like in [9].

Although the advantage of reducing computation and interfacing burden remains with the present method, it has a weakness in not being able to produce models as ideal as what an in-the-loop method can produce. This is evident from Figure 3 (showing a model produced from in-the-loop information maximization in [9]) and 4 (showing a model produced from the present method). On closely observing Figure 4, it can be seen how the identified model subtly deviates from the general trend of data while the model in Figure 3 almost never does so. Authors suspect the reason behind this could be the limitation of the present method doing information maximization offline using only  $x$  data, while in-the-loop methods (like in [9]) might be doing superior information maximization as they can exploit both  $x$  and  $y$  data streams.

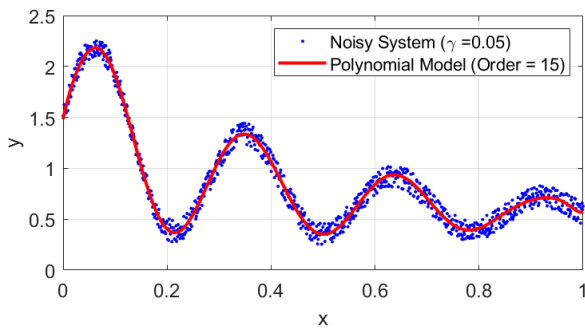


Fig. 3. Behavior of an estimated model in [9] alongside the noisy system.

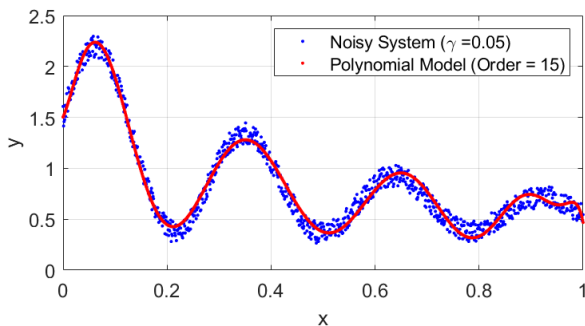


Fig. 4. Behavior of an estimated model in the present paper alongside the noisy system.

## V. CONCLUSIONS

An information driven measurement taking and system identification method was presented for identifying static nonlinear

elements. The method involves computing informative measurement points offline via D-optimal design information maximization, and taking measurements on those calculated points while performing fast in-the-loop calibration of a polynomial model. D-optimal design criteria for computing informative measurement points, and progressing (and termination) criteria for measurement-taking were proposed. The key benefit of this method is its ability to reduce the burden of in-the-loop computation when taking measurements, as information maximization and designing of input points is done offline prior to taking any measurements, or seeing any data. This method showed ability to converge faster (i.e., in fewer measurements) than a previously presented in-the-loop information maximization method [9] that operated on comparable model convergence constraints. However, this method showed a disadvantage in the identified model being less ideal than what could be identified via in-the-loop information maximization as done in [9]. Therefore, it will be useful to further improve offline information maximization methods to become able to perform equivalent to in-the-loop information maximization methods, while preserving the advantage of computational simplicity.

## REFERENCES

- [1] L. Ljung, *System Identification: Theory for the User*, ser. Prentice Hall information and system sciences series. Prentice Hall PTR, 1999. [Online]. Available: <https://books.google.com.au/books?id=nHFoQgAACAAJ>
- [2] S. A. Billings and S. Fakhouri, "Identification of systems containing linear dynamic and static nonlinear elements," *Automatica*, vol. 18, no. 1, pp. 15–26, 1982.
- [3] L. Eriksson and L. Nielsen, *Modeling and control of engines and drivelines*. John Wiley & Sons, 2014.
- [4] M. Deflorian and S. Zaglauer, "Design of experiments for nonlinear dynamic system identification," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 13 179–13 184, 2011.
- [5] J. J. Rissanen, "Fisher information and stochastic complexity," *IEEE transactions on information theory*, vol. 42, no. 1, pp. 40–47, 1996.
- [6] C. R. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer, "Robust optimal experiment design for system identification," *Automatica*, vol. 43, no. 6, pp. 993–1008, 2007.
- [7] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992, pp. 287–294.
- [8] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.
- [9] N. Ulapane, K. Thiyagarajan, and S. Kodagoda, "System identification of static nonlinear elements: A unified approach of active learning, overfit avoidance, and model structure determination," in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2020, pp. 1001–1006.
- [10] R. Chin, A. I. Maass, N. Ulapane, C. Manzie, I. Shames, D. Nešić, J. E. Rowe, and H. Nakada, "Active learning for linear parameter-varying system identification," *arXiv preprint arXiv:2005.00711*, 2020.
- [11] M. H. Stone, "The generalized weierstrass approximation theorem," *Mathematics Magazine*, vol. 21, no. 5, pp. 237–254, 1948.
- [12] N. Ulapane, K. Thiyagarajan, and S. Kodagoda, "Binary spectrum feature for improved classifier performance," in *2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 2020, pp. 1117–1122.
- [13] —, "Hyper-parameter initialization for squared exponential kernel-based gaussian process regression," in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2020, pp. 1154–1159.