

On the optimal worst-case experiment design for constrained linear systems [★]

Marko Tanaskovic ^a, Lorenzo Fagiano ^b, Manfred Morari ^a

^a*Automatic Control Laboratory, Swiss Federal Institute of Technology, Zurich, Switzerland.*

^b*ABB Switzerland Ltd., Corporate Research, Baden-Daettwil, Switzerland.*

Abstract

The problem of experiment design for constrained linear systems with multiple inputs is addressed. A parametric model of the system is considered. The presented theoretical results provide a guideline on how to design experiments that minimize the worst-case identification error, as measured by the radius of information of the set of feasible model parameters, calculated in any norm. In addition, it is shown that an alternative, simpler approach can be employed when input constraints are symmetric and the worst-case identification error is minimized in either 1- or ∞ -norm. For such cases, on the basis of the derived results, a computationally tractable algorithm for the experiment design is proposed. The presented results are valid for a general model representation, which admits the commonly used finite impulse response model as a special case. The features of the presented method are illustrated in a numerical example.

Key words: Design of experiments; System identification; Worst-case identification; Set membership identification; Parameter estimation; Reduced-complexity models; Impulse response.

1 Introduction

The goal of optimal experiment design in system identification is to choose the input sequence to be applied to the plant, in order to maximize the information contained in the collected data and thus minimize the uncertainty on the estimated model parameters. The literature on experiment design can be divided in two categories, i.e. “probabilistic” (see e.g. [11,15,27,29]) and “worst-case”, on the basis of the considered assumptions on the noise signal which affects the plant output.

In the worst-case framework, the information available on the noise is given just by its magnitude bound, thus moving away from any stochastic characterization. Under this assumption, Set Membership (SM) identification approaches (see e.g. [23], [30]) are used to derive the set, named the Feasible Parameter Set (*FPS*), containing all the parameter values such that the corresponding model is able to explain the collected measurements

within the assumed noise bound. In this framework, the size of the *FPS* can be considered as an indicator of the quality of the plant estimate, and it can be used as a criterion for optimal experiment design. In particular, it is of interest to design an experiment that minimizes the worst-case (i.e. for any possible noise realization) radius of the *FPS* (also called radius of information). In [1], the design of periodic input sequences with constrained magnitude that minimize the radius of information in 2-norm have been considered for single input, single output (SISO) systems parameterized by a finite impulse response (FIR) model. For the same class of models, in [18] it was shown that the input sequence that minimizes the worst-case radius of information in 1-norm with input magnitude constraints corresponds to a Galois sequence. In [2], in the same settings the shortest input sequences that minimize the worst-case radius of information in 1-, 2- and ∞ -norm were derived. In [5], a way to design input sequences that minimize the worst-case radius of information in ∞ -norm for conditional SM identification was presented. Moreover, the time complexity of the worst-case experiment design in 1-norm has been considered in [8] and [26]. More recently, experiment design aimed at minimizing the worst-case radius of information for nonlinear systems has been investigated in [17], [25] and for linear systems with quantized

[★] This paper was not presented at any IFAC meeting. Corresponding author M. Tanaskovic Tel. +41 44 632 0403. Fax +41 44 632 1211.

Email addresses: tmarko@control.ee.ethz.ch (Marko Tanaskovic), lorenzo.fagiano@ch.abb.com (Lorenzo Fagiano), morari@control.ee.ethz.ch (Manfred Morari).

measurements in [6], [7].

In summary, the existing studies on experiment design aimed at minimizing the worst-case radius of information address mainly the case of SISO systems parameterized by FIR models, and only magnitude constraints on the input are considered. However, many real world systems have multiple inputs subject to other types of constraints, like input rate constraints, which often arise due to physical limitations of the actuators, as well as constraints that couple different plant inputs. In addition, in order to reduce the model complexity (and the number of parameters that need to be identified), it is often beneficial to use different parameterizations, based on basis functions, instead of FIR models. The most commonly used basis functions are the Laguerre (see [31]) and Kautz (see [32]) ones as well as the generalized orthonormal basis functions (see [9], [24]). For the mentioned cases that are not covered by the literature, we provide here new results on the optimal worst-case experiment design. In particular, we consider systems with multiple inputs subject to general convex constraints, as well as general basis function parameterizations of the system's model. For these settings, our first result provides the answer to the question of how to design input sequences that can minimize the worst-case radius of information, computed in any norm. Although the result is very general, the computational complexity of the related input design procedure becomes quickly prohibitive. Then, we demonstrate how stronger results can be obtained when the input constraints are symmetric and if either the 1- or ∞ -norm radius of information is considered, leading to a computationally tractable algorithm to design the input sequence that actually minimizes the worst-case radius of information that can be achieved in the experiment. In the specific case of SISO, FIR models with only input magnitude constraints, our results correspond to the mentioned previous findings, hence providing a generalization of the existing theory.

The paper is organized as follows. In section 2 we introduce the notation and problem formulation. The main results are derived in section 3, while section 4 presents a numerical example. Finally, conclusions are given in section 5.

2 Problem Statement

We consider a multiple input, single output (MISO), strictly proper, discrete time, linear time invariant (LTI) system, with n_u inputs. At a generic time step t the measured output of the system is given by:

$$y(t) = \varphi(t)^T \theta^0 + e(t), \quad (1)$$

where T stands for the standard matrix transpose operator, $\theta^0 \in \mathbb{R}^m$ is a vector of m model parameters that describe the system, $e(t) \in \mathbb{R}$ is a term that accounts

for any measurement noise, process noise or plant-model mismatch and $\varphi(t) \in \mathbb{R}^m$ is a regressor vector that depends on the applied control inputs $u(t) \in \mathbb{R}^{n_u}$ as:

$$\varphi(t+1) = A\varphi(t) + Bu(t), \quad (2)$$

where $A \in \mathbb{R}^{m \times m}$ and $B \in \mathbb{R}^{m \times n_u}$ depend on the selected model parametrization.

Remark 2.1 *The results that we derive for MISO systems are also relevant for the experiment design of multiple input, multiple output (MIMO) systems, since any MIMO system can be decomposed into several MISO systems (one MISO system for each output).*

Remark 2.2 *Note that the system parametrization in (1) and (2) is slightly different from the standard regression form model assumed in the system identification literature (see e.g. [11,15,27,29]). We use such a model structure in order to state the results on the worst case experiment design for a broad range of basis function parameterizations. In fact the chosen model structure covers as special cases several commonly used model parameterizations. For example, for the case when $n_u = 1$ and an FIR plant model is used, A and B would have the following structure:*

$$A = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (3)$$

For the case when $n_u > 1$, A and B can be obtained by block diagonalizing the matrices in (3). Moreover, suitable A and B matrices can be derived for the case when the Laguerre (see [10]), Kautz (see [32]) or generalized basis functions (see [24]) are used.

Remark 2.3 *According to the model in (1) and (2), the regressor vector depends on the past plant inputs only. The extension of the presented results to a model in which the regressor would depend on the past values of the plant output (e.g. ARX models) would not be straightforward. Such models are treated in the category of errors-in-variables problems in the SM literature (see e.g. [21]).*

We consider the following assumption on the matrices A and B .

Assumption 1 *The pair A, B is controllable and A has all the eigenvalues inside the unit circle.*

Note that Assumption 1 holds for all standard basis function parameterizations. The only knowledge that is assumed on the signal $e(t)$ is that its magnitude is bounded point-wise in time.

Assumption 2

$$|e(t)| \leq \epsilon, \forall t, \quad (4)$$

where $\epsilon \geq 0$ is a known constant.

Note that in practice it might be easier to estimate the bound on the magnitude of the signal $e(t)$ than to estimate its statistical properties. The magnitude bound of the measurement noise could for example be obtained from the sensor technical specifications.

We assume that the plant inputs are subject to constraints of the form:

$$\begin{aligned} C^T u(t) &\leq g, \quad \forall t \\ L^T \Delta u(t) &\leq f, \quad \forall t, \end{aligned} \quad (5)$$

where $\Delta u(t) = u(t) - u(t-1)$ is the rate of change of the input. The element-wise inequalities in (5) define convex sets through the matrices $C \in \mathbb{R}^{n_u \times n_g}$ and $L \in \mathbb{R}^{n_u \times n_f}$ and the vectors $g \in \mathbb{R}^{n_g}$ and $f \in \mathbb{R}^{n_f}$, where n_g and n_f are the number of linear constraints on the input magnitudes and on their rates, respectively. In addition, we consider the following assumption on the input constraints.

Assumption 3 *The set described by the input constraints (5) is compact and contains the origin.*

Note that just magnitude and rate constraints have been assumed in order to simplify the notation, however all the results hold also for additional polytopic input constraints satisfying Assumption 3.

Without loss of generality, we take the time step at which the experiment starts to be $t = 1$ and make the following assumption about the control inputs applied before this time step.

Assumption 4 $u(t) = 0, \forall t < 1$.

The actual value of the parameter vector θ^0 is unknown and needs to be estimated on the basis of the assumptions on the system model and on the magnitude bound of the signal $e(t)$, together with the information provided by the sequence of measurement data collected in the identification experiment. In particular, if we denote a sequence of M measured plant outputs and its corresponding sequence of applied regressor vectors as:

$$\mathbf{y}_M = \{y(t)\}_{t=1}^M, \quad \boldsymbol{\varphi}_M = \{\varphi(t)\}_{t=1}^M, \quad (6)$$

then the set of all the models that are consistent with the initial assumptions and collected measurements (6), named the Feasible Parameter Set, is defined as:

$$FPS \doteq \{\theta \in \mathbb{R}^m : |y(t) - \varphi(t)^T \theta| \leq \epsilon, t = 1, \dots, M\}. \quad (7)$$

An optimal estimate of the parameter vector θ^0 is the one that minimizes the worst-case estimation error. This estimate is given by the Chebyshev center of the FPS :

$$\theta_p^* \doteq \arg \inf_v \sup_{\theta \in FPS} \|v - \theta\|_p, \quad (8)$$

where $p \in [1, \infty)$. Clearly the resulting estimate will depend on the used norm. The worst-case error obtained by the estimate (8) is called the (p -norm) radius of information and it is given by:

$$r_p(FPS) \doteq \sup_{\theta \in FPS} \|\theta_p^* - \theta\|_p. \quad (9)$$

Often the estimate (8) is hard to compute; in these cases interpolatory algorithms represent a valid tractable alternative. These algorithms provide an estimate that is guaranteed to be contained in the FPS , and its worst-case error is bounded by twice the radius of information (8) (see e.g. [30]). Hence, the radius of information $r_p(FPS)$ is a reasonable measure of the quality of an estimate and therefore the experiment design should be aimed at minimizing this quantity. However, the radius of information achieved after a specific experiment depends on the applied control inputs, on the actual system's dynamics, and on the particular realization of the disturbance signal that occurred. Since the last two aspects are not part of the available prior information, it is of interest to consider the *worst-case* radius of information with respect to all possible values of the true system's parameters and all possible disturbance realizations:

$$\bar{r}_p(\boldsymbol{\varphi}_M) \doteq \sup_{\theta^0, |e(t)| \leq \epsilon, t=1, \dots, M} r_p(FPS). \quad (10)$$

It has been shown in [20] (Theorem 3), [22] (Theorem 2) and [19] (Theorem 1) that the worst-case radius of information can be computed as:

$$\begin{aligned} \bar{r}_p(\boldsymbol{\varphi}_M) &= \sup_{\theta} \|\theta\|_p \\ \text{Subject to:} & \\ |\varphi(t)^T \theta| &\leq \epsilon, \forall t \in [0, M]. \end{aligned} \quad (11)$$

The key to prove that (11) is equivalent to (10) is to show that the radius of information (9) does not depend on θ^0 and that it is maximized by a realization of $e(t)$ corresponding to a constant signal. Indeed changing θ^0 , for the same sequence of regressor vectors $\boldsymbol{\varphi}_M$ and the same realization of the signal $e(t)$ results in a translation of the FPS in space and hence does not affect the radius of information (9). The second point is more subtle to show and its proof can be found in the cited literature.

The worst-case radius of information $\bar{r}_p(\boldsymbol{\varphi}_M)$ represents an upper bound on the value of $r_p(FPS)$ that can be obtained in the identification experiment. In addition, from

(11) it follows that, for given prior information (model structure and disturbance/noise bounds) the worst-case radius of information depends only on the applied regressor vectors, and thus, according to (2), on the input sequence employed in the experiment. Therefore, our aim is to design an input sequence that minimizes the value of $\bar{r}_p(\varphi_M)$, while at the same time satisfying the input constraints (5). If such a sequence were found and used in the identification experiment, the corresponding worst-case radius of information would provide a guaranteed upper bound on the quality of the identified model set, no matter what the true system parameters are and what disturbance and noise realizations occur. In the next section, we address the problem of constructing input sequences that minimize the worst-case radius of information (10).

3 Worst-case experiment design

3.1 Preliminaries

We first define the set of admissible regressors as the set containing all the regressor vectors that can be reached from the origin (see Assumption 3) in N steps, given the input constraints (5):

$$\Phi_N \doteq \{\varphi \in \mathbb{R}^m : \varphi = [A^{N-1}B, A^{N-2}B, \dots, B]U, U \in \mathcal{U}\}, \quad (12)$$

where $\mathcal{U} \subset \mathbb{R}^{Nn_u}$ is the set of all the admissible control input sequences of length N , defined by

$$\mathcal{U} \doteq \{U \in \mathbb{R}^{Nn_u} : WU \leq Z\}, \quad (13)$$

where

$$W = \begin{bmatrix} C & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & L & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C & \dots & \mathbf{0} & \mathbf{0} & -L & L & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & C & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & L & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & C & \mathbf{0} & \mathbf{0} & \dots & -L & L \end{bmatrix}^T \in \mathbb{R}^{N(n_g+n_f) \times Nn_u}$$

$$Z = \begin{bmatrix} \underbrace{g^T \dots g^T}_N & \underbrace{f^T \dots f^T}_N \end{bmatrix}^T \in \mathbb{R}^{N(n_g+n_f)}, \quad (14)$$

with $\mathbf{0}$ denoting zero matrices of appropriate dimension. Note that, from (13) and (14), it follows that \mathcal{U} is a polytope and therefore Φ_N is also a polytope, as it is a linear projection of a polytope. The fact that Φ_N is a polytope is a consequence of the assumption that the regressor vector depends on the past control inputs only. In the case of an ARX model, Φ_N would not be a polytope,

which prevents the direct extension of the presented results to a more general ARX model structure (see also Remark 2.3). Theoretically, the set of admissible regressors is given by $\Phi_\infty = \lim_{N \rightarrow \infty} \Phi_N$. However, considering Φ_∞ would imply allowing infinite experimentation time, which is not realistic. Moreover, since by Assumption 1 the matrix A has all the eigenvalues inside the unit circle, it holds that $\lim_{N \rightarrow \infty} A^{N-1}B = \mathbf{0}$, i.e. Φ_N is a good approximation of Φ_∞ for N sufficiently large. Therefore, in our analysis we will make the following assumption about the set of admissible regressor vectors.

Assumption 5 *The set of admissible regressor vectors is given by Φ_N , where N is a finite, but possibly very large number.*

In practice, a sufficiently large N should be selected by the experiment designer. We present a suitable procedure to select N in section 3.3.

Remark 3.1 *Note that if the matrix A is nilpotent, for example when an FIR model parametrization is used (see e.g. (3)), then $\exists N' : \forall N \geq N', \Phi_N = \Phi_{N'}$ and therefore $\Phi_\infty = \Phi_{N'}$, i.e. the set of admissible regressors can be calculated exactly for a finite value of N .*

We next define the minimal worst-case radius of information \bar{r}_p^* as:

$$\bar{r}_p^* \doteq \inf_{M, \varphi(t) \in \Phi_N, t=1, \dots, M} \bar{r}_p(\varphi_M), \quad (15)$$

i.e. \bar{r}_p^* is the minimal worst-case radius of information achievable by considering all the admissible regressors and all possible experiment's lengths M . From (11) and (15) it follows that the minimal worst-case radius of information can be achieved by an infinitely long input sequence that gives rise to all the vectors in Φ_N :

$$\bar{r}_p^* = \sup_{\theta \in S} \|\theta\|_p, \quad (16)$$

where

$$S = \{\theta \in \mathbb{R}^m : |\varphi^T \theta| \leq \epsilon, \forall \varphi \in \Phi_N\}. \quad (17)$$

The first question that we address is whether there exists a *finite* sequence of inputs that achieves \bar{r}_p^* in any norm. To this end, we first define the set $-\Phi_N$ as:

$$-\Phi_N \doteq \{\varphi \in \mathbb{R}^m : -\varphi \in \Phi_N\}, \quad (18)$$

and then the set $\hat{\Phi}_N$ as the convex hull of Φ_N and $-\Phi_N$:

$$\hat{\Phi}_N \doteq \text{conv}(\Phi_N \cup -\Phi_N). \quad (19)$$

The set $\hat{\Phi}_N$ is also a polytope, and as such it is uniquely defined by its vertices. We denote the set of all the

vertices of $\hat{\Phi}_N$ by $V(\hat{\Phi}_N)$. Note that, by construction, $V(\hat{\Phi}_N)$ can be split into two sets:

$$V(\hat{\Phi}_N) = \left\{ V^+(\hat{\Phi}_N) \cup V^-(\hat{\Phi}_N) \right\}, \quad (20)$$

where $\forall v \in V^+(\hat{\Phi}_N)$, $-v \in V^-(\hat{\Phi}_N)$ and $V^+(\hat{\Phi}_N) \cap V^-(\hat{\Phi}_N) = \emptyset$, with the set $V^+(\hat{\Phi}_N)$ satisfying $V^+(\hat{\Phi}_N) \subset V(\Phi_N)$, where $V(\Phi_N)$ is the set of all the vertices of Φ_N . Note that, depending on the shape of the set Φ_N , there might be more than one way to split the set $V(\hat{\Phi}_N)$ into $V^+(\hat{\Phi}_N)$ and $V^-(\hat{\Phi}_N)$.

We also recall the following definition of polytope's polar from [12].

Definition 3.1 A polar of a polytope $P \subset \mathbb{R}^m$ containing the origin is defined as:

$$P^* \doteq \{x \in \mathbb{R}^m : y^T x \leq 1, \forall y \in P\} \quad (21)$$

We further recall two theorems related to the polarity of polytopes that are instrumental for our results.

Theorem 3.1 (Theorem 9.1 in [4])

$$P^* = \{x \in \mathbb{R}^m : v^T x \leq 1, \forall v \in V(P)\}, \quad (22)$$

where $V(P)$ denotes the set of all the vertices of the polytope P .

Theorem 3.2 (Theorem 6.4 in [4]) Let $bd(P)$ denote the set of all the points on the boundary of a polytope P . Then $\forall v \in bd(P)$, it holds that $\max_{x \in P^*} v^T x = 1$, i. e.

the hyperplane $F = \{x \in \mathbb{R}^m : v^T x = 1\}$ is a supporting hyperplane of P^* .

3.2 Main results

We now have all the ingredients to derive the result on how to design a finite input sequence that yields the minimal possible worst-case radius of information in any norm.

Theorem 3.3 Let the Assumptions 1, 2, 3 and 5 hold. The minimal worst-case radius of information \bar{r}_p^* , $\forall p \in [1, \infty)$, is achieved by a finite input sequence that generates all the regressor vectors in $V^+(\hat{\Phi}_N)$.

Proof 1 We first note that the set \mathcal{S} in (17) can be written in a slightly different form:

$$\mathcal{S} = \{\theta \in \mathbb{R}^m : \varphi^T \theta \leq \epsilon, \forall \varphi \in \Phi_N \cup -\Phi_N\}. \quad (23)$$

In addition, we consider another set, defined by the elements of $\hat{\Phi}_N$:

$$\mathcal{S}_c = \left\{ \theta \in \mathbb{R}^m : \varphi^T \theta \leq \epsilon, \forall \varphi \in \hat{\Phi}_N \right\}. \quad (24)$$

Since, by construction, $\Phi_N \cup -\Phi_N \subseteq \hat{\Phi}_N$, it holds that $\mathcal{S}_c \subseteq \mathcal{S}$. In addition, we note that, by construction, $\hat{\Phi}_N$ contains the origin and therefore, by Definition 3.1, the polytope \mathcal{S}_c is a polar of the polytope $\frac{1}{\epsilon} \hat{\Phi}_N$, where $\frac{1}{\epsilon} \hat{\Phi}_N = \left\{ \varphi \in \mathbb{R}^m : \varphi = \frac{1}{\epsilon} v, \forall v \in \hat{\Phi}_N \right\}$. Therefore, from Theorem 3.1 it holds that:

$$\mathcal{S}_c = \left\{ \theta \in \mathbb{R}^m : v^T \theta \leq \epsilon, \forall v \in V(\hat{\Phi}_N) \right\}. \quad (25)$$

By using the fact that $V(\hat{\Phi}_N)$ can be split into $V^+(\hat{\Phi}_N)$ and $V^-(\hat{\Phi}_N)$ as in (20), (25) can be rewritten as:

$$\mathcal{S}_c = \left\{ \theta \in \mathbb{R}^m : |v^T \theta| \leq \epsilon, \forall v \in V^+(\hat{\Phi}_N) \right\}. \quad (26)$$

Since, by construction, $V^+(\hat{\Phi}_N) \subset \Phi_N$, it has to hold that $\mathcal{S} \subseteq \mathcal{S}_c$, and as also $\mathcal{S}_c \subseteq \mathcal{S}$, it follows that $\mathcal{S}_c = \mathcal{S}$. Then, from (16) and (26), the claim of the Theorem follows directly. \blacksquare

The result of Theorem 3.3 gives a guideline on how to design identification experiments to minimize the worst-case radius of information \bar{r}_p^* in any norm ($\forall p \in [1, \infty)$). We note that the input sequence given by Theorem 3.3 also minimizes the volume of FPS (see [28]). Theorem 3.3 implies that in order to design the input sequence that minimizes the worst-case radius of information in any norm, the vertices of Φ_N need to be computed. However, finding the vertices of a polytope described by a set of linear inequalities is in general a computationally hard problem. Dedicated software tools for finding polytope vertices exist [14], but can be used only for relatively low dimensional spaces and for rather simple polytopes. Moreover, the number of vertices of Φ_N would, in general, grow more than exponentially with the increase in the number of the parameters to be identified, implying impractical experimentation time even for a small number of unknown coefficients.

However, it is reasonable to expect that if one would aim at minimizing the worst-case radius of information for a specific norm, there could be a different and less complex way of designing the experiment. In the remainder of this section, we show that this is indeed the case and that polytopic geometry can be exploited in order to provide a tractable way to design experiments if the goal is to minimize the worst-case radius of information in 1- or ∞ -norm, and if the set Φ_N is centrally symmetric, such that $\Phi_N = -\Phi_N = \hat{\Phi}_N$. Note that this condition is met if the constraints (5) are symmetric, which is often the

case in practical problems. The use of 1- and ∞ -norm is quite common in the context of SM identification (see e.g. [21]).

In order to state the result for 1-norm, we define the following set of vectors:

$$V^1 = \left\{ \begin{array}{l} v \in \mathbb{R}^m : v = [1, x_1, \dots, x_{m-1}]^T, \\ x_i \in \{-1, 1\}, i = 1, \dots, m-1 \end{array} \right\}. \quad (27)$$

Note that, by construction, the set V^1 has 2^{m-1} elements. The vectors in V^1 are selected such that they form a 1-norm ball:

$$D^1 = \{x \in \mathbb{R}^m : |v^T x| \leq 1, \forall v \in V^1\}. \quad (28)$$

Note that there is more than one possible way to define the set V^1 such that the set D^1 in (28) corresponds to the 1-norm ball. In fact, the definition of the set V^1 in (27) is taken as a notational convention and the result that we derive holds for any other combination of vectors such that D^1 is a 1-norm ball.

To each of the vectors $v_i \in V^1, i = 1, \dots, 2^{m-1}$, we associate a scalar $\alpha_i > 0, i = 1, \dots, 2^{m-1}$ that is computed as the maximal value that guarantees the satisfaction of $\alpha_i v_i \in \Phi_N$:

$$\begin{aligned} \alpha_i &= \max k \\ \text{Subject to:} & \\ kv_i &\in \Phi_N. \end{aligned} \quad (29)$$

Note that by using (12) and (13), the optimization problem (29) can be rewritten as:

$$\begin{aligned} \alpha_i &= \max_{k,U} k \\ \text{Subject to:} & \\ kv_i &= [A^{N-1}B, A^{N-2}B, \dots, B]U \\ WU &\leq Z, \end{aligned} \quad (30)$$

which is a linear program (LP) that can be solved efficiently with available software tools. We further denote the smallest scalar among all $\alpha_i, i = 1, \dots, 2^{m-1}$, with α^1 :

$$\alpha^1 \doteq \min\{\alpha_1, \dots, \alpha_{2^{m-1}}\}. \quad (31)$$

Note that finding α^1 amounts to solving an LP of the form (30) for each vector in V^1 and then finding the smallest number among the obtained solutions.

We are now in position to give the result on how to design the identification experiment in order to achieve the minimal worst-case radius of information \bar{r}_1^* .

Theorem 3.4 *Let the Assumptions 1, 2, 3 and 5 hold. If Φ_N is centrally symmetric ($\Phi_N = -\Phi_N = \hat{\Phi}_N$), the minimal worst-case radius of information \bar{r}_1^* is achieved*

with the input sequence that generates the regressor vectors $\alpha^1 v_i, \forall v_i \in V^1$, and its value is $\bar{r}_1^ = \frac{\epsilon}{\alpha^1}$.*

Proof 2 *Let us denote by S^1 the set formed by the regressor vectors defined by $\alpha^1 v_i, \forall v_i \in V^1$, as:*

$$S^1 = \{\theta \in \mathbb{R}^m : |\alpha^1 v_i^T \theta| \leq \epsilon, \forall v_i \in V^1\}. \quad (32)$$

From the way V^1 is defined, it follows that $S^1 = \frac{\epsilon}{\alpha^1} D^1$, i.e. S^1 is the 1-norm ball with radius $\frac{\epsilon}{\alpha^1}$. Since, by construction, $\alpha^1 v_i \in \Phi_N, \forall v_i \in V^1$, it has to hold that $S \subseteq S^1$. Moreover, from the assumption that Φ_N is centrally symmetric, it holds that Φ_N contains the origin and therefore, from the definition 3.1, S is a polar of $\frac{1}{\epsilon} \Phi_N$, where $\frac{1}{\epsilon} \Phi_N = \{\varphi \in \mathbb{R}^m : \varphi = \frac{1}{\epsilon} v, \forall v \in \Phi_N\}$. In addition, from the way α^1 is defined in (31), it follows that $\exists v \in V^1 : \alpha^1 v \in \text{bd}(\frac{1}{\epsilon} \Phi_N)$. Moreover, from the fact that Φ_N is symmetric ($\Phi_N = -\Phi_N = \hat{\Phi}_N$), it follows that $-\alpha^1 v \in \text{bd}(\frac{1}{\epsilon} \Phi_N)$. Then, from Theorem 3.2, it holds that the two parallel faces of S^1 defined by $\alpha^1 v$ and $-\alpha^1 v$: $F_+ = \{x \in \mathbb{R}^m : \alpha^1 v^T x = \epsilon\}$ and $F_- = \{x \in \mathbb{R}^m : -\alpha^1 v^T x = \epsilon\}$ are the supporting hyperplanes of S . Since $S \subseteq S^1$ and S^1 has at least two parallel faces that are the supporting hyperplanes of S , S^1 is the tightest possible 1-norm ball containing S . This can be shown by contradiction. Let us assume (for the sake of contradiction) that S^1 is not the tightest possible 1-norm ball containing S and that there exists some 1-norm ball $S' = qD^1, q < \frac{\epsilon}{\alpha^1}$ such that $S \subseteq S' \subset S^1$. Let us denote the faces of S' that are parallel to F_+ and F_- by F'_+ and F'_- . Then it holds that: $F'_+ \in \{x \in \mathbb{R}^m : \alpha^1 v^T x < \epsilon\}$ and $F'_- \in \{x \in \mathbb{R}^m : -\alpha^1 v^T x < \epsilon\}$. Since F_+ and F_- are the supporting hyperplanes of S (i.e. $\max_{x \in S} \alpha^1 v^T x = \epsilon$ and $\max_{x \in S} -\alpha^1 v^T x = \epsilon$), it means that F'_+ and F'_- are contained in the supporting half-spaces of S and therefore intersect S , which is a contradiction. Therefore, S^1 is indeed the tightest possible 1-norm ball containing S and hence it has to hold that:

$$\sup_{\theta \in S} \|\theta\|_1 = \sup_{\theta \in S^1} \|\theta\|_1 = \frac{\epsilon}{\alpha^1}, \quad (33)$$

Which completes the proof. \blacksquare

Theorem 3.4 provides both a way to design an optimal experiment and the value of the related minimal worst-case radius of information \bar{r}_1^* . The latter can be calculated in advance and is given by $\bar{r}_1^* = \frac{\epsilon}{\alpha^1}$. This value gives a good intuition of the effects of input constraints and noise on the experiment. In fact, the value of α^1 can be regarded to as a measure of the size of the set of admissible regressors, Φ_N , and the ratio $\frac{\epsilon}{\alpha^1}$ as a noise-to-signal ratio: the smaller this value, the better the worst-case outcome of the experiment.

Remark 3.2 *The fact that the minimal possible worst-*

case radius of information can be calculated a priori can be exploited to evaluate the quality of an experiment realization a posteriori. In particular, if the achieved radius of information is much smaller than the minimal worst-case value, it would indicate a favorable realization of measurement error and hence good quality of the parameter estimate. Vice versa, if the obtained radius of information is close to the worst-case, it would mean that an unfortunate realization of the signal $e(t)$ has occurred and the experiment could be repeated in hope that a more fortunate realization might occur.

Note that in the case of ∞ -norm, we can define the following set of vectors:

$$V^\infty = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}, \quad (34)$$

where the set V^∞ has m elements. In addition, if we define α^∞ analogously to α^1 in (31), we can state the following corollary of Theorem 3.4.

Corollary 3.1 *If Assumptions 1, 2, 3 and 5 hold and Φ_N is centrally symmetric, the minimal worst-case radius of information \bar{r}_∞^* is achieved with the input sequence that generates the regressor vectors $\alpha^\infty v_i$, $\forall v_i \in V^\infty$ and its value is $\bar{r}_\infty^* = \frac{\epsilon}{\alpha^\infty}$.*

It can be easily shown that in the case of SISO, FIR plant models with only input magnitude constraints, the result of Theorem 3.4 and Corollary 3.1 become equivalent to the existing results on the worst-case experiment design (see e.g. [2], [18]). In these settings, for ∞ -norm, the optimal input sequence is given by an input pulse of maximal magnitude and for 1-norm, the optimal input sequence is a Galois sequence. Therefore, Theorem 3.4 and Corollary 3.1 can be considered as a generalization of the existing results to the case of input constrained MISO systems, when more general convex constraints and model parameterizations are used.

3.3 Optimal experiment design

The minimal worst-case radii of information \bar{r}_1^* and \bar{r}_∞^* depend on the disturbance and noise bound ϵ and on the values of α^1 and α^∞ , which themselves depend on the assumed model structure, input constraints and chosen value of N . From the way Φ_N is defined in (12), it follows that the values of α^1 and α^∞ increase (\bar{r}_1^* and \bar{r}_∞^* decrease) with the increase of N . However, as already mentioned, since by Assumption 1 the matrix A is stable, the values of α^1 and α^∞ converge asymptotically as N increases. This fact can be exploited to determine the value of N to be used in the experiment design. This can

be done by setting a tolerance bound on the relative decrease of the minimal worst-case radius of information $\bar{\Delta}r$ and then increasing N until the relative change in the resulting minimal worst-case radius of information becomes smaller than the defined bound $\bar{\Delta}r$. Note that the calculated value of N will influence the resulting experiment length M , since a larger value of N would likely lead to higher M . Algorithm 1 summarizes the proposed procedure for determining N in the case of 1-norm (for the ∞ -norm the procedure is analogous).

Algorithm 1 Algorithm to determine the value of N

```

 $N = 1;$ 
Calculate  $\alpha^1$  according to (31), based on  $\Phi_N$ ;
 $\bar{r}_{1,old}^* = \frac{\epsilon}{\alpha^1};$ 
 $\Delta r = 1;$ 
while  $\Delta r \geq \bar{\Delta}r$  do
     $N = N + 1;$ 
    Calculate  $\alpha^1$  according to (31), based on  $\Phi_N$ ;
     $\bar{r}_1^* = \frac{\epsilon}{\alpha^1};$ 
     $\Delta r = \frac{|\bar{r}_1^* - \bar{r}_{1,old}^*|}{\bar{r}_{1,old}^*};$ 
     $\bar{r}_{1,old}^* = \bar{r}_1^*;$ 
end while

```

For a fixed value of N , Theorems 3.3 and 3.4 and Corollary 3.1 also provide the guidelines on how to calculate the sets of regressor vectors that should be explored in order to minimize the worst-case radius of information. In order to design an identification experiment, an input sequence that visits all such regressor vectors has to be generated. Usually one would seek for the shortest possible input sequence that gives rise to all the regressor vectors of interest, in order to reduce the required experimentation time. However, finding the shortest sequence would, in general, be a very difficult combinatorial problem. Therefore, we propose a suboptimal, but computationally tractable algorithm for finding a sequence of control inputs starting from the set of regressor vectors that should be generated. The approach, summarized in Algorithm 2 below, is such that it guarantees that all the regressor vectors of interest are visited, but does not give the guarantee that the computed input sequence is the shortest possible.

The algorithm is iterative and it employs a greedy approach. At each iteration, an element from the set of the regressor vectors that have not been visited and that can be reached from the current regressor vector φ_0 in the smallest number of steps is found, together with the sequence of inputs that should be applied in order to make the transition. This vector is then removed from the set of the regressor vectors that have not been visited and it is taken as the current regressor vector. The calculated input sequence is concatenated with the previously calculated input sequences. Initially, the current regressor vector φ_0 is selected as a zero vector (see Assumption 4)

and the set of the regressor vectors that have not been visited is populated by all those that have to be visited.

The proposed algorithm relies on the fact that if a given regressor vector, φ_T , can be reached from the current one, φ_0 , in k steps, given the input constraints (5), then the following system of linear equalities and inequalities must be feasible:

$$\begin{aligned} & \exists \varphi(t) \in \mathbb{R}^m, t \in [0, k], u(t), \Delta u(t) \in \mathbb{R}^{n_u}, t \in [1, k] \\ & \text{Such that:} \\ & \varphi(0) = \varphi_0 \\ & \varphi(k) = \varphi_T \\ & \left. \begin{aligned} \varphi(t) &= A\varphi(t-1) + Bu(t-1) \\ \Delta u(t) &= u(t) - u(t-1) \\ C^T u(t) &\leq g \\ L^T \Delta u(t) &\leq f \end{aligned} \right\} \forall t = 1, \dots, k \end{aligned} \quad (35)$$

where $u(0)$ is the last element of the input sequence that was applied in order to achieve the regressor vector φ_0 . A solution of (35), if it exists, represents the input sequence that steers the regressor vector from φ_0 to φ_T in k steps. Checking whether a system of linear equalities and inequalities that form a convex set has a solution, and finding that solution, is a problem that can be efficiently solved by using linear programming techniques [3]. Note that, since $\varphi_T \in \Phi_N$ and the pair A, B is controllable, the problem (35) will always have a solution for k large enough. Finding the regressor vector that can be reached in the smallest number of steps is then done by consecutively solving problem (35) for all the terminal regressor vectors that have not been visited, with k initially set to 1 and then incremented until a feasible solution is found.

Note that the computational complexity of the Algorithm 2 depends on the number of regressor vectors that have to be visited. This number scales linearly for the ∞ - and exponentially for the 1-norm with the number of unknown parameters m . Note however that the Algorithm 2 relies on linear programs only, which can be solved very efficiently. Therefore, the Algorithm can be executed in reasonable time even for large number of regressor vectors that need to be visited.

4 Numerical example

In order to illustrate the derived results and the features of the proposed algorithm, we consider a numerical example in which we compare the identification performance for the case when an input sequence that minimizes the worst-case radius of information is used, with the case when an input signal based on using random binary sequences is used. We consider a SISO system model, parameterized by Laguerre basis functions. The

Algorithm 2 Algorithm to compute the input sequence

```

 $\varphi_0 = 0$ ;
Initialize not_visited by the regressor vectors to be visited;
Initialize  $U$  as an empty sequence
while not_visited is not empty do
   $k = 1$ ;
  found = 0;
  while not found do
    for  $i = 1 : \text{size}(\text{not\_visited})$  do
       $\varphi_T = \text{not\_visited}[i]$ ;
      if (35) has a solution for  $k, \varphi_0$  and  $\varphi_T$  then
         $U^*$  is the solution of (35);
        found = 1;
        break;
      end if
    end for
     $k = k + 1$ ;
  end while
   $\varphi_0 = \text{not\_visited}[i]$ ;
  remove not_visited[ $i$ ] from not_visited;
   $U = [U, U^*]$ ;
end while

```

plant is given as a linear combination of 5 filters:

$$y(z) = \sum_{k=1}^5 \theta_k^0 \frac{\sqrt{1-0.4^2}}{z-0.4} \left[\frac{1-0.4z}{z-0.4} \right]^{k-1} u(z) + e(z), \quad (36)$$

where z is the time shift operator ($zu(t) = u(t+1)$) and $y(z)$, $u(z)$ and $e(z)$ are the \mathcal{Z} transforms of the plant input, output and disturbance signals. The coefficients θ_k^0 , $k = 1, \dots, 5$ are considered to be unknown and need to be estimated from an identification experiment. The plant model with the structure in (36) can be written in the regressor form (1)-(2) by taking the vector of the unknown parameters to be $\theta^0 = [\theta_1^0, \dots, \theta_5^0]^T$ and the matrices in (2) as (see [10]):

$$A = \begin{bmatrix} 0.4000 & 0 & 0 & 0 & 0 \\ 0.8400 & 0.4000 & 0 & 0 & 0 \\ -0.3360 & 0.8400 & 0.4000 & 0 & 0 \\ 0.1344 & -0.3360 & 0.8400 & 0.4000 & 0 \\ -0.0538 & 0.1344 & -0.3360 & 0.8400 & 0.4000 \end{bmatrix} \quad (37)$$

$$B = [0.9165 \quad -0.3666 \quad 0.1466 \quad -0.0587 \quad 0.0235]^T$$

It is assumed that the control input of the system is constrained such that $|u(t)| \leq 2$ and $|\Delta u(t)| \leq 1$. In addition, we assume the bound on the disturbance $e(t)$ to be $\epsilon = 0.4$ and in the simulations we generated $e(t)$ by taking a random value from a uniform distribution in the interval $[-\epsilon, \epsilon]$ at each time step. For the given identification problem, we would like to design an experiment that minimizes the worst-case radius of information in 1-norm. To this end, by using Algorithm 1, we selected N

that defines the set of admissible regressors (Φ_N) to be $N = 30$, corresponding to the tolerance of $\overline{\Delta r} = 10^{-4}$. We created the set of regressor vectors that should be visited, by solving an LP as in (35) for each element of the set V^1 and computing the value of the parameter α^1 as in (31). The corresponding sequence of inputs is then obtained by using the Algorithm 2, it is 108 samples long and is shown in Fig. 1. For this example it took 0.54 seconds to calculate N and 0.86 seconds to execute the Algorithm 2 on a laptop with Intel i7-3667U processor, when using Gurobi [13] for solving the optimization problems.

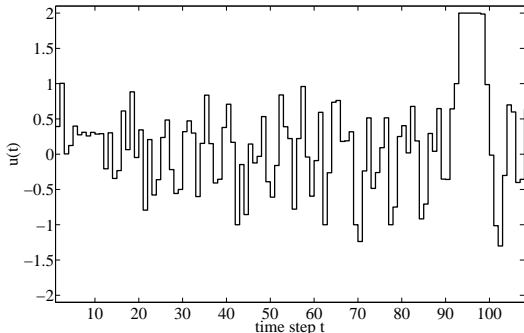


Fig. 1. Input sequence minimizing the worst-case radius of information in 1-norm.

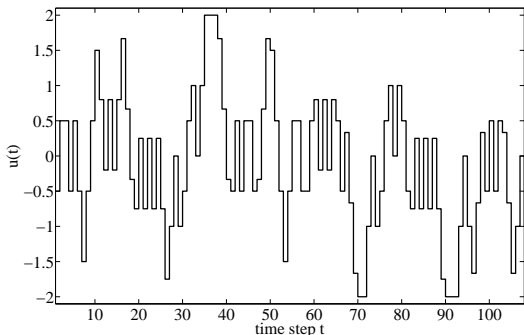


Fig. 2. Typical example of an input sequence generated by using a random binary signal.

We performed a series of identification experiments with two different input sequences and the same measurement error realization. In one case we use the input sequence given in Fig.1, which guarantees to achieve the minimal worst-case radius of information, while in the other case we use an input sequence of the same length whose generation is based on a random binary signal.

This sequence is formed by first generating a random binary sequence with the probability of the sign change taken as a random variable in the interval $[0.2, 0.8]$ and then by finding a sequence that is the most similar to it (it has the smallest 2-norm difference) and that satisfies the input rate constraints. A typical excitation signal

obtained in this way is shown in Fig. 2. Random binary sequences are a recommended heuristic for generating input signals in identification experiments, see e.g. [16]. During the experiments, the values of the model parameters were set to: $\theta^0 = [2, 1.6, 1, 0.6, 0.2]^T$. We repeated the identification experiment 4000 times. The obtained values for the radius of information for the first 150 realizations are compared in Fig. 3, while in Fig. 4 we compare the empirical probability density functions for the obtained values of the radius of information over all the realizations. Taking the average over all the experiments, the obtained radius of information is very similar for both cases. However, when the control input sequence of Fig. 1 is used, the radius of information is never larger than 0.78, which is the theoretical value for the minimal worst-case radius of information. On the other hand, when the input sequence based on the random binary signal is used, for some unfavorable realizations of the error signal, the radius of information can be larger than the theoretically derived value for the minimal worst-case radius of information. In fact, in 13.7% of the experiment realizations using random input, the obtained radius of information was greater than the maximal value guaranteed by our approach. The same kind of numerical analysis was done for the case of ∞ -norm radius of information. In this case in 16.2% of the experiment realizations the radius of information obtained when using the random input was larger than the worst case bound guaranteed by our approach.

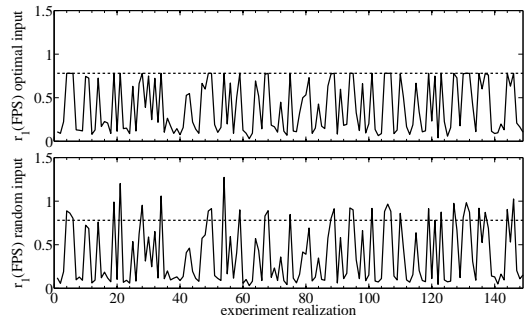


Fig. 3. The upper plot shows the obtained radius of information when the optimal input sequence is used. The lower plot shows the obtained radius of information when the input based on a random binary signal is used. The dashed line indicates the value of the minimal worst-case radius of information according to Theorem 3.4.

Hence, as expected from our results, when using the input sequence derived with Theorem 3.4 and Algorithm 2, there is an upper bound (that can be calculated beforehand) to the radius of information that can be obtained for any measurement error realization. This is not the case with a more standard input sequence often used in system identification experiments, which can result in quite large radius of information in case the measurement error is unfavorable.

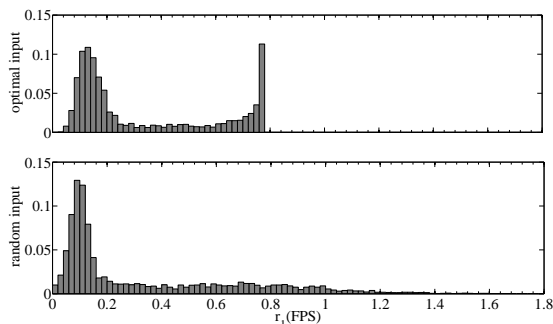


Fig. 4. Empirical probability density functions of the radius of information in different identification experiment realizations for the optimal input (upper plot) and for the input based on the random binary sequence (lower plot).

5 Conclusion

We have presented a way to compute the set of regressor vectors that should be visited in order to minimize the worst-case radius of information in any norm for the case of input constrained MISO linear systems with a general parametrization. In addition, we addressed the particular cases when the radius of information is calculated in 1- and ∞ -norm and showed that there exists a computationally tractable way to design an optimal experiment in such settings, provided that the input constraints are symmetric. The presented results provide a generalization of previous findings, related to the specific case of SISO systems with FIR models. Finally, we showed with a numerical example that the bound on the quality of the resulting model guaranteed by our approach can be violated with a non-negligible probability by more common random input sequences, which are suboptimal from the point of view of minimizing the worst-case radius of information. At the same time, the average performance achieved with our approach is similar to the one obtained with random input sequences.

References

- [1] E.-W. Bai, R. Tempo, and H. Cho. Membership set estimators: size, optimal inputs, complexity and relations with least squares. *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, 42:266–277, 1995.
- [2] G. Belforte and T. T. Tay. Optimal input design for worst-case system identification in $l_1/l_2/l_\infty$. *System & Control Letters*, 20:273–278, 1993.
- [3] D. Bertsimas and J. Tsitsiklis. *Introduction to linear optimization*. Athena Scientific, 1997.
- [4] A. Brøndsted. *An introduction to convex polytopes*. Springer-Verlag, New York, 1982.
- [5] M. Casini, A. Garulli, and A. Vicino. On input design in l_∞ conditional set membership identification. *Automatica*, 42:815–823, 2006.
- [6] M. Casini, A. Garulli, and A. Vicino. Input design in worst-case system identification using binary sensors. *IEEE Transactions on Automatic Control*, 56:1186–1191, 2011.
- [7] M. Casini, A. Garulli, and A. Vicino. Input design in worst-case system identification with quantized measurements. *Automatica*, 48:1997–3007, 2012.
- [8] M. A. Dahleh, T. V. Theodosopoulos, and J. N. Tsitsiklis. The sample complexity of worst-case identification of FIR linear systems. *Systems & Control Letters*, 20:157–166, 1993.
- [9] P. M. J. Van den Hof, P. S. C. Heuberger, and J. Bokor. System identification with generalized orthonormal basis functions. *Automatica*, 31:1821–1834, 1995.
- [10] G. A. Dumont and Y. Fu. Non-linear adaptive control via Laguerre expansion of Volterra kernels. *International Journal of Adaptive Control and Signal Processing*, 7:367–382, 1993.
- [11] M. Gevers. Identification for control: From the early achievements to the revival of experimental design. *European Journal of Control*, 11:335–352, 2005.
- [12] B. Grünbaum. *Convex Polytopes*. Interscience, New York, 1967.
- [13] Inc. Gurobi Optimization. Gurobi optimizer reference manual, 2013. <http://www.gurobi.com>.
- [14] M. Herceg, M. Kvasnica, C.N. Jones, and M. Morari. Multi-Parametric Toolbox 3.0. In *Proc. of the European Control Conference*, pages 502–510, Zürich, Switzerland, July 17–19 2013. <http://control.ee.ethz.ch/~mpt>.
- [15] H. Hjalmarsson. From experiment design to closed-loop control. *Automatica*, 41:393–438, 2005.
- [16] L. Ljung. *System identification: theory for the user*. Prentice Hall, New Jersey, 1999.
- [17] L. Lu and B. Yao. Experimental design for identification of nonlinear systems with bounded uncertainties. In *Proc. of the American Control Conference*, pages 4504–4509, Baltimore, Maryland, USA, June 30–July 2 2010.
- [18] P. M. Mäkilä. Robust identification and Galois sequences. *International Journal of Control*, 54:1189–1200, 1991.
- [19] A. G. Marchuk and K. Yu. Oshipenko. Best approximation of functions specified with an error at a finite number of points. *Mathematical Notes*, 17:207–212, 1975.
- [20] C. A. Micchelli and T. J. Rivlin. A survey of optimal recovery. *Optimal estimation in Approximation Theory*, editors: C. A. Micchelli and T. J. Rivlin, pages 1–54, 1977.
- [21] M. Milanese, J. Norton, H. Piet-Lahanier, and E. Walter. *Bounding approaches to system identification*. Plenum press, New York, 1996.
- [22] M. Milanese and R. Tempo. Optimal algorithms theory for robust estimation and prediction. *IEEE Transactions on Automatic Control*, 30:730–738, 1985.
- [23] M. Milanese, R. Tempo, and A. Vicino. *Robustness in identification and control*. Plenum, 1989.
- [24] B. M. Ninness, H. Hjalmarsson, and F. Gustafsson. The fundamental role of generalized orthonormal bases in system identification. *IEEE Transactions on Automatic Control*, 33:1133–1139, 1999.
- [25] C. Novara. Experiment design in nonlinear set membership identification. In *Proc. of the American Control Conference*, pages 1566–1571, New York, USA, July 11–13 2007.
- [26] K. Poolla and A. Tikku. On the time complexity of worst-case system identification. *IEEE Transactions on Automatic Control*, 39:944–950, 1994.
- [27] L. Pronzato. Optimal experimental design and some related control problems. *Automatica*, 44:303–325, 2008.
- [28] L. Pronzato and E. Walter. Experiment design in a bounded-error context: comparison with d-optimality. *Automatica*, 25:383–391, 1989.

- [29] F. Pukelsheim. *Optimal experimental design*. Wiley, New York, 1993.
- [30] J.F. Traub and H. Woźniakowski. *A General Theory of Optimal Algorithms*. Academic Press, New York, 1980.
- [31] B. Wahlberg. System identification using Laguerre models. *IEEE Transactions on Automatic Control*, 36:551–562, 1991.
- [32] B. Wahlberg. System identification using Kautz models. *IEEE Transactions on Automatic Control*, 39:1276–1282, 1994.