On Useful Redundancy in Experiment Design For Nonlinear System Identification

Mazen Alamir, James S. Welsh and Graham C. Goodwin

***Abstract*— In the paper, a formulation is proposed for optimal experiment design dedicated to the identification of nonlinear systems. In particular, a recently mentioned redundancy property associated to dynamic systems related inverse problems is heavily exploited to guarantee global convergence. The paper considers general discrete-time nonlinear systems in which measurements are affected by bounded noise. An illustrative example is used to show the merits of the proposed approach.**

I. INTRODUCTION

Experiment design aims to compute an excitation signal that obtains maximal information from the system which can be used to identify the vector of parameters involved in the system model. Substantial research has been conducted in the area of experiment design appearing in both the statistics literature [4], [5], [16] and engineering literature [11], [8], [12], [17].

As experiments may be expensive, a *one shot* identification scheme is quite appealing. We utilize the term *one shot* here to indicate that an experiment once designed is applied to the plant where input/output data is captured and used in a system identification algorithm. Namely, there is no sequential [6], [15] nor iterative [7], [9] experiment design.

However, for such a scheme to be efficient, the excitation signal used in the experiment must be carefully computed. The computation of such signal for a one shot experiment in the absence of a priori knowledge on the parameter value is referred to as optimal robust experiment design [13].

When the system dynamics are linear, the characterization of *optimality* is possible through the use of the information matrix [8]. Robustness may then be obtained by using a min-max formalism in which maximization is done by the unknown parameter and minimization is performed by the control input [14]. Moreover, the use of Parseval's argument makes it possible to derive a minimization problem that is affine in the input spectrum.

This has been supported by the French (National Research Agency Grant) ANR-Capteurs Logiciels Plug & Play and has been done when the first author was visiting the Center Complex Dynamic Systems and Control at the University of Newcastle, Australia.

Mazen alamir is with Gipsa-lab, CNRS, University of Grenoble. BP 46, Domaine Universitaire, ST Martin d'Hères, France. mazen.alamir@gipsa-lab.inpg.fr

James S. Welsh and Graham C. Goodwin are with the School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan, NSW 2308, Australia James.Welsh@newcastle.edu.au, Graham.Goodwin@newcastle.edu.au

Extension of some ideas from the linear case to the nonlinear case can be considered as suggested in [10] provided that for a given excitation, the output prediction error is still affine in the parameter vector. Non convexity may appear for the computation of the excitation signal but the very definition of what might be considered as a *well defined* identification problem is still possible using the Hessian matrix of the output prediction error related cost.

When the parameter vector appears in a nonlinear form in the least squares related cost, the problem of non convexity makes impossible the very definition of "*well posdness* " due to the possible presence of local minimum. The latter may steer the identification process towards a faulty solution at which the properties of the Hessian matrix are irrelevant in characterizing the situation. Many existing works on nonlinear identification simply assume that such problems do not occur in the specific situation under study [2], [3]. In this case however, the first key property that characterizes a good excitation signal is the absence of local minima in the resulting cost function behavior. However, for highly nonlinear systems, asking for such a quasi-convexity property would be simply unrealistic.

The present paper suggests an alternative characterization through the use of the recently introduced concept of *safe redundancy* [1]. Roughly speaking, an inverse nonlinear problem is safe redundant if it can be put in a form in which the global minimum one is looking for is shared by a family of cost functions that do not share the same local minimums. Switching between a family of solvers, each using a different cost, enables one to safely cross the potential singularity of the original problem. Rephrasing the above discussion, one may say that the experiment design we are interested in amounts to find an excitation signal that makes the problem safely redundant in the sense mentioned above.

The paper is organized as follows: First, some definitions and notation are introduced in section II. The basic idea is then presented in section III. Some concepts of safe redundancy are recalled in section IV and its application in robust optimal experiment design for nonlinear systems is presented in section V. A simple example is used in section VI to illustrate the concepts and solutions proposed throughout the paper. Finally, section VII gives some concluding remarks and a road map for further

investigation.

II. Definitions and notation

In this paper, the following class of discrete-time nonlinear systems are considered:

$$x^+ = f(x,u,p) \quad ; \quad y = h(x,p) + w \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^{n_u}$, $p \in \mathbb{P} \subset \mathbb{R}^{n_p}$, $y \in \mathbb{R}^{n_y}$ and $w \in \mathbb{R}^{n_y}$ denote the state, the control input, the vector of unknown parameters, the measured output and the measurement error respectively. $\mathbb{P}$ is some compact set of admissible parameter vectors.

The following characterization of the measurement error w is used:

Assumption 1: [Characterization of the meas. error] The measurement error w satisfies the following property: For all time k and all integer $k^* \geq 1$:

$$\frac{1}{k^*}\sum_{i=0}^{k^*-1} w(k+i) \in \phi_w(k^*-1)\cdot[-\bar{w},\bar{w}] \quad ; \quad \bar{w} \in \mathbb{R}^{n_y} \tag{2}$$

where $\phi_w(\cdot)$ is some function that takes values in $\mathbb{R}_+$ and that satisfies $\phi_w(0) = 1$. Note that the inclusion invoked in (2) is to be interpreted componentwise. ♡

Related to assumption 1, the following assumption states the a priori available knowledge on the parameters involved in the measurement error characterization:

Assumption 2: [Available knowledge on noise error] It is assumed that the vector $\bar{w}$ as well as an upper bound of the convergence-to-mean shaping function $\phi_w(\cdot)$ is known by some prior characterization step. ♡

The next assumption is related to the experimental conditions:

Assumption 3: [Experimental conditions]

1) Regardless the observability of the system, it is assumed that the state x_0 is well known at the beginning of the experiment. Therefore, reference to this initial state may be omitted. Note that slight error on the knowledge of the initial state can be implicitly account for through the measurement error vector w.

2) Measurements are supposed to be acquired with some sampling period τ.

3) The family of admissible excitation signals $u_{(\cdot)}$ is defined by a finite dimensional parametrization:

$$u_k = \mathcal{U}(k,q) \quad ; \quad q \in \mathbb{Q} \quad ; \quad k \in \{0,\ldots,k_{max}\} \tag{3}$$

where $\mathcal{U}$ is some given parametrization map. $\mathbb{Q} \subset \mathbb{R}^{n_q}$ is some compact set of admissible parameters while $k_{max} \in \mathbb{N}$ stands for the experiment duration in terms of sampling periods. ♡

Based on the above assumptions, the following notations are used throughout the paper:

- ✓ $x_k(q,p)$: the state that would be reached at instant k when starting at instant 0 using the control parameter q and under the system parameter p.
- ✓ $y_k(q,p,w_k) = h(x_k(q,p)) + w_k$: The measurement that would be obtained at instant k under the same conditions as above and assuming the measurement noise w_k at instant k.
- ✓ $y_k^0(q,p) = y_k(q,p,0)$
- ✓ y_k^m : the true measurement effectively obtained at instant k

In the following, the *true* value of the parameter p is denoted by p^r in order to make it readily distinguishable from any intermediate computation driven value p. Referring to these notations, the following function is extensively used in the remainder of the paper:

$$\Gamma_k(q,p^r,p) := \frac{1}{k_{max}-k+1}\sum_{i=0}^{k_{max}-k}\underbrace{[y_{k+i}^0(q,p^r) - y_{k+i}^0(q,p)]}_{\epsilon_{k+i}(q,p^r,p)} \tag{4}$$

which clearly satisfies the following recurrent equations:

$$\begin{aligned}\Gamma_0 &= \frac{1}{k_{max}+1}\sum_{i=0}^{k_{max}}\epsilon_i(q,p,p^r)\\ \Gamma_{k+1} &= \frac{1}{k_{max}-k}\Big[-\epsilon_k + (k_{max}-k+1)\cdot\Gamma_k\Big]\end{aligned}$$

In the sequel, the following family of time-weighting shifted discrete-Tchebychev polynomials are used:

$$\begin{aligned}&\bar{T}_i : \{1,\ldots,k_{max}\} \to \mathbb{R}_+\\ &\bar{T}_i(k) := \frac{1}{2}\Big[T_i(2k/k_{max}-1)+1\Big]\end{aligned}$$

where the T_i's are defined as follows:

$$T_0(k) = 1 \ ; \ T_1(k) = k \ ; \ T_{i+1}(k) = 2kT_i(k) - T_{i-1}(k)$$

Also, the following gate-like functions are used:

$$\Pi_i(k) = \begin{cases} 1 & \text{if } k \leq k_i^* \\ 0 & \text{otherwise}\end{cases} \tag{5}$$

where the k_i^* are some fixed integers $\in \{0,\ldots,k_{max}\}$.

Finally, the following truncated weighting function is defined on $\mathbb{R}^{n_y}\times\mathbb{R}^{n_y}$:

$$d^+(w,w^*) = \|max\{0,|w|-w^*\}\|_R \tag{6}$$

where again the internal expression is to be interpreted component-wise and where the positive definite weighting matrix $R \in \mathbb{R}^{n_y\times n_y}$ is chosen once for all throughout the paper.

III. THE BASIC IDEA

Based on the above notations, the identification problem one would be faced to at the end of the expriment is to find $p \in \mathbb{P}$ such that the following inclusions are satisfied for all $k \in \{1, \dots, k_{max}\}$:

$$\frac{1}{k_{max}-k+1} \sum_{i=0}^{k_{max}-k} \left[y^0_{k+i}(q,p) - y^m_{k+i} \right] \in \\ \in \phi_w(k_{max}-k) \cdot [-\bar{w}, \bar{w}] \quad (7)$$

in which, $y^m_{(\cdot)}$ is the sequence of measurements acquired during the experiment. Recall that the set of all possible sequences of measurements is parameterized by all possible values of $p^r \in \mathbb{P}$ and all possible sequences of measurement noise $w_{(\cdot)}$, namely:

$$y^m_i = y^0_i(q, p^r) + w_i \quad (8)$$

Injecting (8) into (7) and using the definition (4) enables to claim that all possible identification problems that one may be faced to are given by the following set of inclusions:

$$\Gamma_k(q,p^r,p) + \frac{1}{k_{max}-k+1} \sum_{i=0}^{k_{max}-k} w_{k+i} \in \\ \in \phi_w(k_{max}-k) \cdot [-\bar{w}, \bar{w}] \quad (9)$$

Note that making robust experiment design based on the above inclusions requirement amounts to look for excitation signals such that the corresponding problem is well conditioned for all possible p^r and all possible sequence of measurement errors $w_{(\cdot)}$. This would be quite cumbersome to compute. This is why in the present work, the set of inclusions (9) is replaced by the following set of noise-free set of inclusions (using assumption 1):

$$\Gamma_k(q,p^r,p) \in 2\phi_w(k_{max}-k) \cdot [-\bar{w}, \bar{w}] \quad (10)$$

which clearly enlarges the set of admissible parameters but hugely simplifies the robust identification problem[1].

Using the notation (6), the inclusions (10) can be transformed into the following conditions that must be satisfied for all $k \in \{0, \dots, k_{max}\}$:

$$\Psi_k(q,p^r,p) \quad := \quad d^+\big(\Gamma_k(q,p^r,p), 2\phi_w(k_{max}-k)\bar{w}\big) = 0$$

and since the Ψ_k's must vanish for all k, the following cost function must vanish whatever is the weighting function $\Phi_i(\cdot)$ being used:

$$J_i(q,p^r,p) = \sum_{k=0}^{k_{max}} \Phi_i(k) \cdot \|\Psi_k(q,p^r,p)\| \quad (11)$$

[1]Note that such a step consisting in *removing* the disturbances from the equations is unavoidable even in handling linear systems related robust experiment design problems [14]. Nevertheless, the time characterization (2) of the measurement error is explicitly used here in the very definition of the cost function (11) despite of the simplification that lead to the set of inlcusion (10)

The above discussion can be summarized by the following fact:

Fact 3.1:

For each excitation parameter q, each true unknown parameter p^r and each family of weighting functions $\{\Phi_i(\cdot)\}_i$, the solution of the *modified* identification problem is a global minimum of ALL the cost functions $J_i(q,p^r,\cdot)$ (in the decision variable p) that are obtained by changing the value of $i \in \mathbb{N}$. ♡

It is worth noting that it is the use of the dead-zone map d^+ that enables the above fact to hold despite the presence of measurement noise and errors.

Fact 3.1 strengthens the very particular nature of the optimization problems that arise when solving dynamic systems-related inverse problems. This fact has been first underlined in [1] when studying the nonlinear state estimation problem. Moreover, this feature has been exploited in order to built up an iterative algorithm that is robust against local minima. In this paper, some ideas from [1] are adapted to the parameter estimation problem. In particular, the sufficient condition for global convergence is used in order to define a cost function on the excitation parameter q leading to optimal experiment design. In order to do so, some results from [1] are first recalled in section IV and adapted to the identification context in section V.

IV. RECALLS ON AVOIDING SINGULARITIES IN NONLINEAR REDUNDANT INVERSE PROBLEMS

Unlike convex optimization problems where *well posdness* can be directly linked to the Hessian matrix properties, for non convex problems, this notion is not easy to define. The aim of this recall is to suggest a new measure of well posdness for a particular class of non-convex optimization problems.

Consider a constrained optimization problem defined on $\mathbb{R}^{n_p}$ by:

$$\mathcal{P}_0 \quad : \quad \min_{p \in \mathbb{P}} J_0(p) \quad ; \quad \mathbb{P} \subset \mathbb{R}^{n_p}$$

where J_0 is a positive cost function that admits the global minimum p^r and probably other local minima. This problem may result from identification concern for a given excitation vector q and a given true parameter value p^r.

Consider also an optimizer that uses some iteration $\mathcal{S}$:

$$p^{(i+1)} = \mathcal{S}(p^{(i)}, J_0(\cdot))$$

in order to find the solution p^r of $\mathcal{P}_0$. r successive iterations of $\mathcal{S}$ lead to the multi-steps updating map that is denoted hereafter by $\mathcal{S}^{(r)}$, namely:

$$p^{(i+r)} = \mathcal{S}^{(r)}(p^{(i)}, J_0)$$

The sequence of values that is *induced* by the successive iterations starting from some initial guess $p^{(0)}$, namely

$$\left\{\mathcal{S}^{(r)}(p^{(0)}, J_0)\right\}_{r\in\mathbb{N}}$$

is called hereafter, the $(p^{(0)}, J_0)$-solver path as it depends on both the initial guess and the cost function J_0 being used. Note that under convexity condition one may expect the following asymptotic property to hold

$$\text{(Under convexity)} \quad \lim_{r\to\infty} \mathcal{S}^{(r)}(p^{(0)}, J_0) = p^r$$

while in the general case, the solver path may be trapped by some local minima of J_0. The redundancy invoked in fact 3.1 may be used to avoid such situation. For this, consider the following definition:

Definition 4.1: [N-Safely Redundant Problems]

The optimization problem $\mathcal{P}_0$ is called *N-safely redundant* if and only if the following conditions hold

1) There exists a finite sequence of N cost functions J_i defined on $\mathbb{R}^{n_y}$ sharing the same global minimum $p^r \in \mathbb{P}$.
2) There exists a solver $\mathcal{S}$ and a finite integer $r^* \in \mathbb{N}$ such that the following inequality

$$\Delta_N^{\gamma}(p) := \min_{i\in\{0,\dots,N\}} \Big[J_0(\mathcal{S}^{(r^*)}(p, J_i)) - \gamma J_0(p)\Big] \leq 0 \tag{12}$$

holds for some $\gamma \in [0, 1[$ and all $p \in \mathbb{P}$. Moreover:

$$\mathcal{S}^{(r^*)}(p, J_{\imath^*}) \in \mathbb{P} \tag{13}$$

where $\imath^*$ is the optimal argument of the minimization invoked in (12) ♡

Note that conditions (12) and (13) simply state that regardless the current iterate p, there exists at least one of the (p, J_i)-solver pathes that leads to a decrease in the value of the original cost J_0 and this after at most r^* iterations. Figure 1 shows a schematic view of a 2-safely redundant optimization problem in which J_0 admits 4 local minima (including the global minimum).

Definition 4.1 suggests the algorithm depicted on figure 2. This algorithm can be better understood through the following comments:

Step 0
The initial guess is set to $p^{(0)}$ and the iteration index m is set to 0.
Step 1
This loop is the main loop that ends as soon as the current iterate $p^{(m)}$ corresponds to the stopping condition on $J_0(p^{(m)}) \leq \varepsilon$.
Step 1.1
When the logical variable $again$ is true, a new solver path has to be *tried* based on the next element in the sequence of cost functions $\{J_i\}_{i=1}^N$ as the preceding one failed to decrease J_0. This is checked in **Step 1.2.2** where $again$ is updated.
Step 1.2
As long as $again$ is true AND there is remaining solver pathes to be explored ($i \leq N$), the next solver path is explored by performing r^* steps of the optimizer iterations (**Step 1.2.1**) and $again$ is updated accordingly (**Step 1.2.2**). If significant decrease is achieved then $again$ becomes false and **Step 1.2.4** is fired leading to the update of $p^{(m+1)}$ and the iteration index m and a new iteration is started. Otherwise, the next cost function corresponding solver path is visited by incrementing i (**Step 1.2.3**).

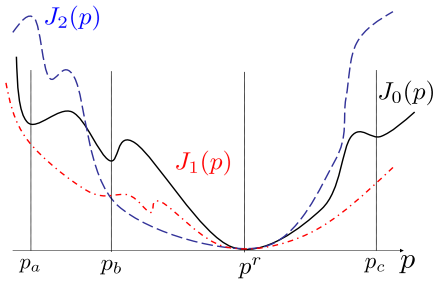

Fig. 1. Typical scheme for a 2-safely redundant optimization problem defined by the cost function J_0. J_0 (in black-solid line) admits one global minimum p^r and 3 local minima p_a, p_b and p_c. However, the singularities p_a and p_c can be crossed following the (p_a, J_1) and the (p_b, J_1)-solver pathes respectively while the singularity p_b can be crossed following the (p_b, J_2)-solver path.

Algorithm A_1

0. Initialization $p^{(0)}$ initial guess, $m \leftarrow 0$
1. while $(J_0(p^{(m)}) > \varepsilon)$ **do**
  1.1 $i \leftarrow 1$; again$\leftarrow$ *true*
  1.2 while $(again \quad \& \quad i \leq N)$ **do**
    1.2.1 $\xi^{(m,i)} \leftarrow \mathcal{S}^{r^*}(p^{(m)}, J_i)$
    1.2.2 $again \leftarrow \Big(J_0(\xi^{(m,i)}) > \gamma J_0(p^{(m)})\Big)$
    1.2.3 If $again$ **then** $i \leftarrow i + 1$
    1.2.4 Else $m \leftarrow m + 1$, $p^{(m)} \leftarrow \xi^{(m,i)}$
  End while
End while

Fig. 2. The algorithm that enables redundancy to be exploited for N-safe redundant optimization problems.

It goes without saying that by the very definition of N-safe redundancy, the above algorithm leads to a globally convergent iterations despite possible local minima of J_0. This is formally stated in the following proposition:

Proposition 1: [Global convergence]
If the optimization problem is N-safely redundant then the algorithm A_1 stops with the iterate $p^{(m)}$ that satisfies the

following property:

$$J_0(p^{(m)}) \leq \varepsilon$$

Consequently, if $\varepsilon = 0$ is used, the sequence $p^{(m)}$ globally asymptotically converges to the unique global minimum p^r.
Proof By definition of N-safe redundancy expressed in condition (12), each time the loop of **step 1.2** is visited, there is an index $\imath$ that makes the logical variable *again* false since

$$J_0(\xi^{(m,\imath)}) \leq \gamma J_0(p^{(m)}) \tag{14}$$

leading to the updating **step 1.2.4**:

$$p^{(m+1)} = \xi^{(m,\imath)} \tag{15}$$

Based on (14)-(15), one obtains the following discrete dynamic

$$J_0(p^{(m+1)}) \leq \gamma J_0(p^{(m)}) \tag{16}$$

which clearly ends the proof. □

V. Optimal Experiment Design

The preceding section showed that for a N-safely redundant optimization problem in the decision variable p to be globally solvable by the cross-singularity algorithm depicted above, the inequality [see (12)]

$$\Delta_N^\gamma(p) \leq 0 \tag{17}$$

must be satisfied.

When applying this scheme to the identification driven optimization problem, it is worth recalling that all the expressions used in section IV are defined for a given actual model parameter $p^r \in \mathbb{P}$ and a given excitation parameter $q \in \mathbb{Q}$. Therefore, the optimal experiment design problem amounts to find an excitation vector $q \in \mathbb{Q}$ that makes the identification problem efficiently and globally solvable whatever is the value of the unknown actual parameter p^r and this for the least possible value of the contraction factor γ appearing in (12). One way to do this is to solve the following min-max optimization problem in the unknown q:

$$\min_{q\in\mathbb{Q}} \bar{\Delta}_N^\gamma(q) := \max_{(p^r,p)\in\mathbb{P}\times\mathbb{P}} \Delta_N^\gamma(q, p^r, p) \tag{18}$$

for decreasing values of $\gamma \in [0, 1[$ until no solution exists for which the optimal value is ≤ 0.

It is worth mentioning some properties of the optimization problem (18):

1) The optimal cost of (18) is equal to 0 since

$$\bar{\Delta}_N^\gamma(q) \geq \Delta_N^\gamma(q, p^r, p^r) = 0$$

as all the cost functions J_i vanish by definition.

2) In general, there is no unique optimal solution to (18) since any value of the excitation parameter q that makes $J_0(q, p^r, p)$ uniformly (for all p^r) N-safely redundant vanishes the cost function and is therefore a global minimum.

3) The computational complexity of $\Delta_N^\gamma(\cdot)$ is almost exclusively linked to the system integration task (enabling the Ψ_k's in (11) to be computed] and not much on the value of N since each computation of a new J_i costs a scalar product [see (11)].

VI. Illustrative examples

Let us consider the modified Van-der-Pol oscillator to which an excitation control input has been added as well as an unknown parameter p:

$$\begin{aligned} \dot{x}_1 &= px_2 + u \\ \dot{x}_2 &= -9x_1 + (1 - x_1^2)x_2 \\ y &= x_1 + w \end{aligned}$$

The parameter p is supposed to belong to the following compact set $\mathbb{P} = [0.5, 2.0]$. w is assumed to be a white noise of variance $W = 0.001$. It is also assumed that the experiment starts with the initial state $x_0 = \big(0.1, 0\big)$. The basic period $\tau = 0.0625\ sec$ is used for the experiment design. Finally, the integer k_{max} that defines the experiment duration is taken equal to $k_{max} = 60$.

The parametrization of the excitation signal is given by $\mathcal{U}(t, q) = B(t) \cdot q$ where:

$$B(t) = (\sin(\frac{2\pi t}{T_e}), \sin(\frac{4\pi t}{T_e}), \sin(\frac{6\pi t}{T_e}), \sin(\frac{8\pi t}{T_e}))$$

The basic period $T_e = 1.2\ sec$ has been used. The admissible set $\mathbb{Q}$ for the excitation parameter vector q is taken equal to $\mathbb{Q} = [-0.15, +0.15]^4$ which corresponds to somehow a limited energy excitation signal. The initial guess $q_0 = 1 \in \mathbb{R}^4$ has been taken to initialize the iterations. For each value of q the computation of $\bar{\Delta}_N^\gamma(q)$ is obtained by taking the maximum value of $\Delta_N(q, p^r, p)$ over a uniform grid covering the admissible domaine $\mathbb{P}^2 = [0.5, 2.0]^2$ and containing 25 nodes.

A. Identification of the map $\phi_w(\cdot)$ [see (2)]

Although in the case of Gaussian noise, analytical computations may be performed in order to compute the upper bound $\phi_w(\cdot)$ as well as the value of $\bar{w}$. Here, simulation-based computations are used in order to simulate sensor characterization oriented experiment that may be necessary for non gaussian measurement errors. For this to be done, 1000 runs obtained using the RANDN function of MATLAB and the maximum of the absolute value of $\phi_w(k)$ has been reported for $k = 1, \ldots, 4 \times k_{max}$. The results can be viewed on Figure 3 where an approximated upper bound is also plotted that is given by:

$$\bar{w} = 0.0031 \quad ; \quad \phi_w(k) = \frac{1}{k^{0.4}}$$

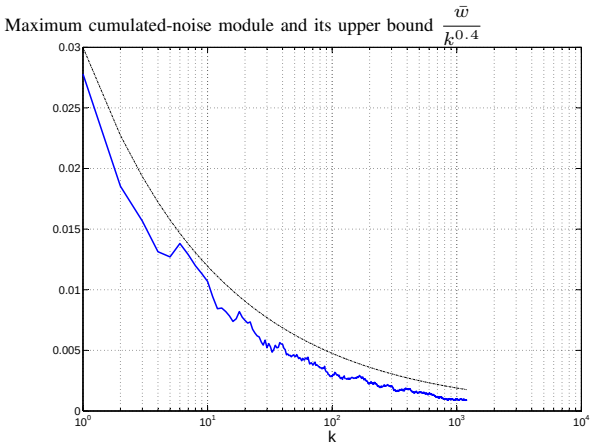

Fig. 3. Identification of the noise characterization parameters $\bar{w}$ and $\phi_w(\cdot)$. 1000 runs of the MATLAB function RANDN are used. Each run contains $20 \times k_{max} = 1200$ samples. For each $k \in \{1, \ldots, 20 \times k_{max}\}$, the maximum over these runs of the absolute value of the cumulated noise is plotted (blue continuous line) versus its suggested upper bound $\phi_w(k)\bar{w}$ (black dotted line) with $\bar{w} = 0.0031$ and $\phi_w(k) = 1/k^{0.4}$.

B. Computations & Discussion

The aim of this section is to illustrate the concepts and points raised throughout the paper. First of all, figure 4 shows how the concept of N-safe redundancy applies in the particular case of the example under interest.

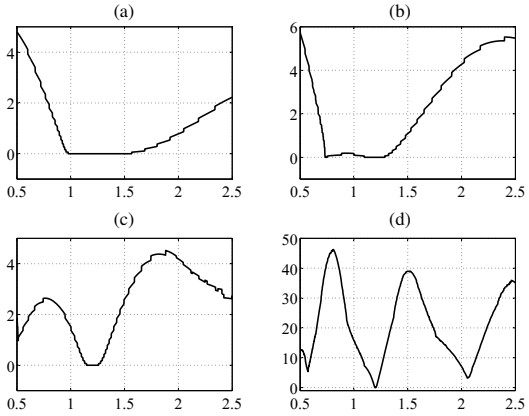

Fig. 4. Evolution of the cost functions $J_i(q, p^r, \cdot)$ vs the decision variable p for the initial excitation vector $q = (1, 1, 1, 1)$ and the true parameter $p^r = 1.2$. The cost functions are obtained using $\Phi_i = \Pi_i$ for $i = 1, 2, 3, 4$ where the weighting profiles are given by (5) in which $k_1^* = 30$ (a), $k_2^* = 35$ (b), $k_3^* = 40$ (c) and $k_4^* = 60$ (d). Note how the resulting optimization problem is clearly 4-safely redundant in the sense of definition 4.1. Note that in this case J_0 is equal to J_4 (figure (d)).

Indeed, Figure 4 shows the different allures of the cost functions $J_i(q, p^r, \cdot)$ defined by (11) (for a particular value of the pair (q, p^r)) when 4 different weighting profiles Φ_i are taken equal to the Π_i's defined by (5). It can be clearly noticed how the solver path corresponding to the first cost function [figure 5.(a)] can be used to approach the vicinity of the real value $p^r = 1.2$ avoiding the singularities shown on the fourth cost function [figure 5.(d)]. The latter can then be used to achieve final convergence.

The benefit from using redundancy when optimizing experiment design is shown on figures 5 and 6. Indeed, figure 5 shows the worst-case optimization problem over all possible values of (p^r, p) for different optimal experiment design parameters (the Φ_i's are here taken equal to the shifted Tchebychev polynomials $\bar{T}_i$ defined in section II)

- **[top-left]**: Cost function $J_0(q_0, 0.5, \cdot)$ for the initial excitation vector $q_0 = (1, 1, 1, 1)$ and the original cost function J_0. This plot shows clearly that if $p^r = 0.5$ then there are too many bad initial guesses that may lead to a bad identification (assuming a gradient based approach).

- **[top-right]**: Cost function $J_0(\hat{q}_0^{0.3}, 0.5, \cdot)$ after experiment optimization using no redundancy ($N = 0$). The resulting excitation vector is denoted by $\hat{q}_0^{0.3}$. Note that the local minimum problem still arise. Since the contraction with $\gamma = 0.3$ is not achievable, the optimization tried to reduce the value of the cost function by reducing the overall excitation.

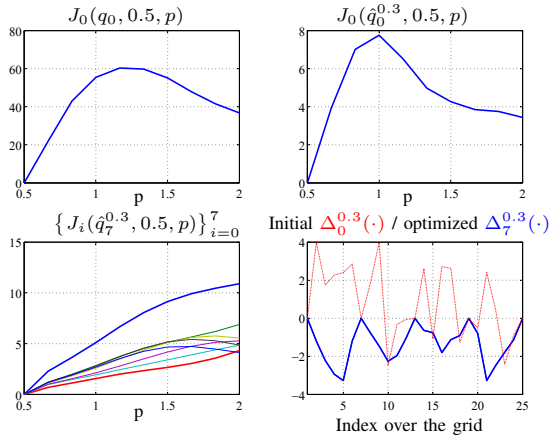

Fig. 5. (a) Evolution of the worst case cost function $J_0(q_0, p^r, \cdot)$ for the initial excitation signal q_0. (b) Evolution of the optimized worst-case cost function J_0 if no redundancy is used ($N = 0$). (c) Evolution of the worst case cost functions J_i, $i = 0, \ldots, 7$ when $N = 7$ is used to exploit redundancy. (d) Evolution over the grid in the (p^r, p)-plane of the cost functions $\Delta_0^{0.3}(q_0, \cdot, \cdot)$ [red dotted] and $\Delta_7^{0.3}(\hat{q}, \cdot, \cdot)$ [blue solid].

- **[bottom-left]**: Cost functions $J_i(\hat{q}_7^{0.3}, 0.5, \cdot)$ for $i = 1, \ldots, 7$ obtained after redundancy based experiment optimization with $N = 7$. Note that even in the worst case, there is no more local minimum and global convergence can be achieved whatever is the initial guess.

- **[bottom-right]**: This plot shows the values of the cost

functions $\Delta_0^{0.3}(q_0,(p^r,p)_j)$ and $\Delta_7^{0.3}(\hat{q}_7^{0.3},(p^r,p)_j)$ at the grid nodes $\{(p^r,p)_j\}_{j=0}^{25}$ respectively before and after the excitation vector optimization process. Positive values mean that the contraction condition is not satisfied at the corresponding grid nodes while negative value means that the contraction is achieved.

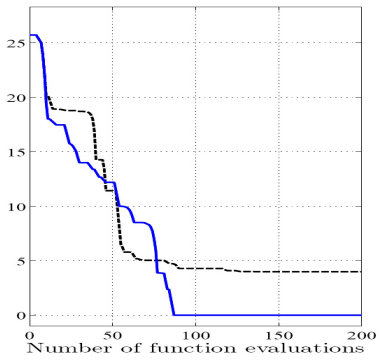

Fig. 6. Evolution of the cost functions $\bar{\Delta}_0^{0.3}(q)$ [black dotted] and $\bar{\Delta}_7^{0.3}(q)$ [blue solid] during the optimization of the excitation vector q. Note that in the absence of redundancy use, no solution can be found that achieves the contraction factor $\gamma = 0.3$

Figure 6 gives the evolution of the cost functions $\bar{\Delta}_0^{0.3}(q)$ (black dotted) and $\bar{\Delta}_7^{0.3}(q)$ during the experiment optimization using respectively $N = 0$ or $N = 7$. Note that when no redundancy is used, the optimizer fails in founding an excitation vector q that guarantees the contraction over the grid nodes. The resulting evolution of the excitation vector's components during the optimization is given on figure 7.

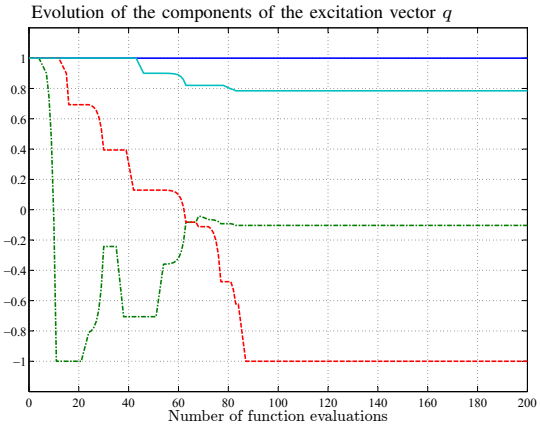

Fig. 7. Evolution of the components of the excitation vector q during experiment optimization with the redundancy $N = 7$ and the contraction factor $\gamma = 0.3$.

Recall that according to the above discussion, well posdness of the non convex optimization problem may be directly linked to the minimum value of the contraction factor γ for which an optimal cost function $\bar{\Delta}_N^{\gamma}(\cdot) = 0$ is still possible to achieve. Computations on the example under study lead to a minimal value of $\gamma = 0.1$. Figure 8 shows (for different values of the contraction factor $\gamma = 0.9, 0.3, 0.2$ and 0.1 and for each value of the pair (p^r,p) on the grid) the index of the first solver-path that achieve the required contraction. For instance, the bottom-left subplot of figure 8 shows that when trying to achieve the contraction factor of $\gamma = 0.2$ for the pair (p^r,p) indexed by 17, the first solver path that achieves this contraction is the last one (index=7). Note therefore that the smaller γ is, the more one needs redundancy to achieve the objective.

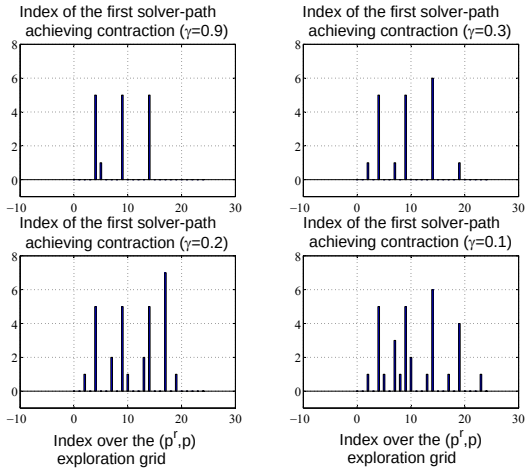

Fig. 8. Analysis of the optimal solutions for different values of the contraction factor: $\gamma = 0.9$ [top-left], $\gamma = 0.3$ [top-right], $\gamma = 0.2$ [bottom-left] and $\gamma = 0.1$ [bottom-right]. Horizontal coordinate represents the index over the (p^r,p) grid while the vertical abscissa represents the index of the first solver-path achieving the contraction γ for that particular value of the pair (p^r,p). The smallest is γ, the more redundancy is required.

VII. Conclusion

In this paper, a robust optimal design scheme is proposed for the parametric identification of nonlinear dynamic systems under bounded measurement noise. The scheme extensively uses the concept of N-redundant optimization problem recently introduced in [1]. As in any min-max optimization based formulation, heavy computations are involved and dedicated iterative algorithms have to be developed although the use of redundancy may reduce the overall needed functions evaluations.

References

[1] M. Alamir. On useful redundancy in dynamic inverse problems related optimization. In *Proceeding of the IFAC-World Congress*, South Korea, July 2008.

[2] M. W. Braun, R. Oritz-Mojica, and D. E. Rivera. Application of minimum crest factor multisinusoidal signals for "plant-friendly" identification of nonlinear process systems. *Control Engineering Practice*, 10:301–313, 2002.

[3] S. H. Chung, D. L. Ma, and R. D. Braatz. Optimal model-based experimental design in batc crystallization. *Chemometrics and Intelligent Laboratory Systems*, 50:83–90, 2000.

[4] D.R. Cox. *Planning of Experiments*. Wiley, New York, 1958.

[5] V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York and London, 1972.

[6] I. Ford and S.D. Silvey. A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika*, 67(2):381–388, 1980.

[7] M. Gevers. Identification for control: From the early achievements to the revival of experiment design. *European Journal of Control*, 11:1–18, 2005.

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

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

[10] H. Hjalmarsson and J. Martensson. Optimal input design for identification of nonlinear systems: Learning from the linear case. In *Proceeding of the 2007 American Control Conference*, New York, USA., 2007.

[11] V.S. Levadi. Design of input signals for parameter estimation. *IEEE Transactions on Automatic Control*, AC-11(2):205–211, 1966.

[12] R.K. Mehra. Optimal inputs for system identification. *IEEE Transactions on Automatic Control*, AC-19:192–200, 1974.

[13] L. Pronzato. Optimal experimental design and some related control problem. *Automatica*, 44:303–325, 2007.

[14] C. R. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer. Robust optimal experiment design for system identification. *Automatica*, 43:993–1008, 2007.

[15] E. Walter and L. Pronzato. *Identification of Parametric Models from Experimental Data*. Springer-Verlag, Berlin, Heidelberg, New York, 1997.

[16] P. Whittle. Some general points in the theory of optimal experimental design. *J.R. Stat. Soc.*, (1):123–130, 1973.

[17] M. Zarrop. *Optimal Experiment Design for Dynamic System Identification*, volume 21 of *Lecture Notes in Control and Information*. Springer, Berlin, New York, 1979.