

# Input design using cylindrical algebraic decomposition

Håkan Hjalmarsson and Freja Egebrand

**Abstract**—Experiment design for system identification has seen significant progress in the last decade. One contribution has been to derive convex relaxations of such problems. Consider that only a scalar function of the system parameters is of interest. A standard step in such a case is to first linearize this function with respect to the estimated parameters. The objective of this contribution is twofold: firstly, to examine if there are cases where the linearized approximation is inadequate, and secondly to explore how to improve upon this approximation. By way of examples we show that it is not difficult to construct examples where linearization is insufficient. Furthermore, we introduce the use of higher order approximations and we formally show that this leads to polynomial optimization problems under Gaussian assumptions. We propose the use of cylindrical algebraic decomposition as a method to obtain exact solutions for this type of problems. Numerical examples are provided.

## I. INTRODUCTION

As data quality limits the modeling accuracy, experiment design is a critical issue in data-driven modeling. Poor data can never be remedied by post-processing. The topic cuts across a wide range of disciplines, including machine learning, statistics, statistical learning and system identification. The overview [1] provides an outline of the general issues and state-of-the art. In regards to the system identification area, numerous contributions have advanced the area significantly during the last decade [2, 3, 4, 5]. In particular we point to contributions regarding robust and adaptive designs [6, 7] and the introduction of the concept of least-costly design [8].

A key issue has been to find problem formulations that allow the corresponding optimization problems to be re-cast as convex programs. There exists a quite general framework for linear models [9] and there are methods emerging for nonlinear models [10, 11, 12].

Spearheaded by [13] already 25 years ago, there has been a clear shift from general purpose design criteria such as D-optimal and E-optimal design to applications oriented design criteria [4]; there are for example methods emerging for as diverse applications as deconvolution [14], fusion reactors [15], magnetic resonance imaging, and MPC [16].

In most applications, the objective function is nonlinear in the estimated parameters. In order to obtain a convex problem it is common to linearize this function. The rationale for this is that accurate models are required in high performance

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agreement No. 267381

The authors are with the Automatic Control Laboratory, ACCESS Linnaeus Center, Electrical Engineering, KTH – Royal Institute of Technology, S-100 44 Stockholm, Sweden. Emails: {hakan.hjalmarsson|freja.egebrand}@ee.kth.se

applications and therefore parameter deviations will have to be small, this in turn implying that errors in the objective function are well captured by a first order Taylor approximation. However, the error in such an approximation can be significant for low to medium performance applications. One such illustrative example is given in [17] where pole and zero estimates are considered. One possibility to handle this problem is to a posteriori re-scale the experimental effort so that the design objective is reached. However, this leads to a suboptimal solution. Our contribution is to extend the Taylor approximation to higher orders. This leads to polynomial optimization problems, or polynomial matrix inequality problems. There exist several approaches to handle such optimization problems. One line of research is based on sum-of-squares relaxations [18] which can be shown to converge [19]. In this contribution, we will follow another path and use cylindrical algebraic decomposition (CAD) [20]. This technique provides exact solution for constrained polynomial optimization problems. In the control literature, CAD has been applied to MPC [21], but it does not seem to have been used in an experiment design context.

The outline of the paper is as follows. In Section II we outline a prototype experimental problem. It is deliberately kept simple so as to not confound the key issues. It can easily be extended to the frameworks in [9, 4]. In Section III we introduce the idea of extending the Taylor approximation and derive the corresponding optimization problem. Section VII contains a numerical illustration and Section VII concludes the paper.

## II. OPTIMAL INPUT DESIGN

Let  $\theta \in \mathbb{R}^n$  denote the model parameters, with  $\theta^o$  denoting the true ones. We assume that  $N$  input-output samples is used in the estimation and the resulting parameter estimate is denoted  $\hat{\theta}_N$ . We will assume that the estimate is normal distributed with mean  $\theta^o$  and covariance matrix  $P/N$ :

$$\delta\hat{\theta}_N := \hat{\theta}_N - \theta^o \in N(0, P/N) \quad (1)$$

This is usually an approximation as in general this is valid only asymptotically in the sample size  $N$  [22]. However, our discussion carries over to the asymptotic case.

We denote the design variables for the experiment design by  $r = [r_0 \ \dots \ r_m]^T \in \mathbb{R}^m$ . We will assume that the inverse of the normalized covariance matrix  $P$  depends affinely on the design variables, i.e.

$$P^{-1} = \sum_{k=0}^m r_k R_k \quad (2)$$

for some known matrices  $\{R_k\}_{k=0}^m$ . The reason for this assumption is that for linear models (e.g. ARX, ARMAX, Box-Jenkins, linear state space models with tailor made parametrizations) in the asymptotic regime ( $N$  large), the parametrization (2) can be achieved if, e.g.,  $r$  is taken to be the autocorrelation coefficients of the input, i.e.

$$r_k = E[u_t u_{t-k}], \quad (3)$$

see [9] for details. In this case, the actual input is then generated by filtering white noise through the filter corresponding to the stable spectral factor of the spectrum corresponding to  $r$ . It is also possible to use (2) for certain nonlinear models, see [12] for details. In the following we will assume that the design variables  $r_k$  have the interpretation (3).

We will let  $J(\theta)$  denote some property of the application when model parameter  $\theta$  is used in the design. We illustrate this with an example

*Example 2.1 (Model based control):* Let  $\theta \in \mathbb{R}^n$  be the model parameters for a discrete-time linear time-invariant model

$$y_t = G(q, \theta)u_t(\theta) + H(q)e_t \quad (4)$$

where  $q$  denotes the shift operator  $qu_t = u_{t+1}$ ,  $u_t$ ,  $y_t$  and  $e_t$ , are the input, the output and a zero mean white noise disturbance, respectively. Suppose that the true system is in the model class, i.e. there is a  $\theta = \theta^o$  such that (4) corresponds to the true system. Suppose now that  $\theta$  is used to design a feedback controller  $C(q, \theta)$  which is applied to the true system,  $C$  could for example be a model reference controller, the  $H_2$ -optimal or an  $H_\infty$  controller. This gives the equations

$$\begin{aligned} y_t(\theta) &= G(q, \theta^o)u_t(\theta) + H(q, \theta^o)e_t \\ u_t(\theta) &= -C(q, \theta)y_t(\theta) \end{aligned}$$

where  $s_t(\theta)$  denotes the true signal  $s_t$  when the controller  $C(q, \theta)$  is used in the feedback loop.

In this control application, we could, for example, take  $J(\theta)$  to be the achieved phase-margin when  $C(q, \theta)$  is used. ■

One possible measure of the performance degradation in the application resulting from using model parameter  $\theta$  instead of the true parameter  $\theta^o$  (which for all reasonable application designs should be the best) is

$$V_{deg}(\theta) = (J(\theta) - J(\theta^o))^2 \quad (5)$$

Now consider that the parameter that is used in the design of the application (the controller  $C(q, \theta)$  in Example 2.1) is obtained by parameter estimation, i.e.  $\theta = \hat{\theta}_N$ . Since  $\hat{\theta}_N$  is stochastic it is necessary to study the average of the performance degradation

$$V_{deg} = E \left[ (J(\hat{\theta}_N) - J(\theta^o))^2 \right] \quad (6)$$

where the expectation is over the random components of  $\hat{\theta}_N$ .

The experiment design problem that we will consider here is the following least-costly experiment design problem

$$\begin{aligned} \min_r & Nr_0 \\ \text{s.t.} & V_{deg} < \frac{1}{\gamma} \\ & T_m > 0 \end{aligned} \quad (7)$$

where  $T_m$  denotes the symmetric Toeplitz matrix with  $r_0, \dots, r_m$  in the first row. The last constraint ensures that  $r_0, r_1, r_m$  can be extended to a valid auto-correlation sequence [23]. The interpretation of (7) is that we would like to use a minimal amount of input energy in our identification experiment in order to achieve a certain accuracy (represented by  $\gamma \in \mathbb{R}$ ) of the property  $J(\hat{\theta}_N, \theta^o)$ .

### III. FIRST ORDER APPROXIMATION OF THE DESIGN CRITERION

Typically (7) is non-convex. However, a semi-definite program (SDP) can be obtained by linearizing  $J$ . The procedure is as follows: Write

$$J(\theta) \approx J(\theta^o) + L^T(\theta - \theta^o) \quad (8)$$

where  $L = \left. \frac{dJ(\theta)}{d\theta} \right|_{\theta=\theta^o}$  so that

$$V_{deg} \approx E \left[ L^T(\hat{\theta}_N - \theta^o)(\hat{\theta}_N - \theta^o)^T L \right] = L^T P L / N$$

where we have used (1) in the second equality. The problem (7) can thus be approximated by

$$\begin{aligned} \min_r & Nr_0 \\ \text{s.t.} & L^T P L / N < \frac{1}{\gamma} \\ & P > 0 \\ & T_m > 0 \end{aligned} \quad (9)$$

where we have added the non-singularity condition  $P > 0$  which is equivalent to  $P^{-1} > 0$ . Applying Schur-complements twice to the first two inequalities in (9) gives the equivalent problem

$$\begin{aligned} \min_r & Nr_0 \\ \text{s.t.} & P^{-1} > \frac{\gamma}{N} L L^T \\ & T_m > 0 \end{aligned} \quad (10)$$

Finally, inserting (2) gives

$$\begin{aligned} \min_{r_0, \dots, r_m} & Nr_0 \\ \text{s.t.} & \sum_{k=0}^m r_k R_k > \frac{\gamma}{N} L L^T \\ & T_m > 0 \end{aligned} \quad (11)$$

which is an SDP in  $r_0, \dots, r_m$ . This is the type of optimization problem that is obtained for a wide range of problems [9, 4].

#### IV. POLYNOMIAL OPTIMAL EXPERIMENT DESIGN

##### A. A polynomial inequality for the performance degradation

We will now derive a general expression for the first constraint in (7) without invoking the approximation (8). We start with assuming that  $J$  is a multivariate polynomial. Let  $\delta\theta = \theta - \theta^\circ$  with elements  $\{\delta\theta_i\}$ . Thus we assume that

$$J(\theta) = J(\theta^\circ) + \sum_{q=1}^Q \frac{1}{q!} (\delta\theta \circ \nabla)^q J(\theta^\circ) \quad (12)$$

where we use the operator notation

$$\delta\theta \circ \nabla = \sum_{i=1}^n \delta\theta_i \frac{\partial}{\partial \theta_i} \quad (13)$$

*Lemma 4.1:* When  $J$  is a multivariate polynomial, the constraint

$$V_{deg} < \frac{1}{\gamma} \quad (14)$$

can be written as a multivariate polynomial inequality under the constraint  $P > 0$ .

*Proof:* Inserting (12) in (5) gives

$$V_{deg}(\theta) = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \frac{1}{q_1!q_2!} \{(\delta\theta \circ \nabla)^{q_1} J(\theta^\circ)\} \{(\delta\theta \circ \nabla)^{q_2} J(\theta^\circ)\} \quad (15)$$

which in (6) gives

$$V_{deg} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \frac{1}{q_1!q_2!} \mathbb{E} \left[ \left\{ (\delta\hat{\theta}_N \circ \nabla)^{q_1} J(\theta^\circ) \right\} \left\{ (\delta\hat{\theta}_N \circ \nabla)^{q_2} J(\theta^\circ) \right\} \right] \quad (16)$$

Now invoking (1), and using that all higher order moments of  $\delta\hat{\theta}_N$  are polynomials in the elements of  $P$  (they can be obtained from the moment generating function) gives the expression

$$V_{deg} = \sum_k g_k(\theta^\circ) p_k(P) \quad (17)$$

where  $p_k$  are multivariate polynomials in the elements of  $P$  and where the coefficients  $g_k(\theta^\circ)$  are functions of the partial derivatives of  $J$  at  $\theta = \theta^\circ$ .

Now recall the matrix inversion formula

$$M^{-1} = \frac{\text{adj}(M)}{\det(M)} \quad (18)$$

where  $\text{adj}(M)$  is the adjoint matrix of  $M$ . Applying this formula to  $P$  and using the parametrization (2) we observe that the elements of  $P$  are rational functions of  $r$  with the common denominator  $d(r) = \det(\sum_{k=0}^m r_k R_k)$ . Next we observe that  $P > 0$  implies that  $d(r) > 0$ .

This means that for each  $k$ ,  $p_k(P)$  in (17) is a rational function in  $r$  where the denominator is some power of the polynomial  $d(r)$ . Therefore, multiplying (14) with the highest power of  $d(r)$  that occurs in the  $p_k(P)$  results in a polynomial inequality. As  $d(r)$ , on  $P > 0$ , is strictly positive, the new condition is equivalent to (17) on the set of  $r$  such that  $P > 0$ . This concludes the proof. ■

##### B. Topelitz positivity as a polynomial inequality

The term  $L^T P L / N$  in the first constraint in (9) corresponds to one of the terms in (17). In order to improve the approximation more terms can be added giving the rational inequality constraint

$$\sum_{k=1}^{\alpha} g_k(\theta^\circ) w_k(r) < \frac{1}{\gamma} \quad (19)$$

for some  $\alpha$ . By multiplying with the greatest common denominator a polynomial inequality is obtained.

The constraint  $T_m > 0$  can also be written as a set of polynomial inequalities.

*Lemma 4.2:* The inequality

$$T_m > 0 \quad (20)$$

is equivalent to the  $m + 1$  polynomial inequalities

$$\det(T_{q-1})r_0 - \tilde{r}_q^T \det(T_{q-1})T_{q-1}^{-1}\tilde{r}_q > 0, \quad q = 0, \dots, m \quad (21)$$

where  $\tilde{r}_q = [r_q \quad r_{q-1} \quad \dots \quad r_1]^T$ , and where  $T_{-1} = I$  and  $\tilde{r}_0 = 0$ .

*Proof:* We prove this by induction. For  $k = 0$  the condition is

$$T_0 = r_0 > 0 \quad (22)$$

which clearly is a polynomial inequality constraint.

Now assume that the statement holds for  $T_{m-1}$ . We have that

$$T_m = \begin{bmatrix} T_{m-1} & \tilde{r}_m \\ \tilde{r}_m^T & r_0 \end{bmatrix} \quad (23)$$

Using Schur complements gives

$$T_m > 0 \Leftrightarrow r_0 - \tilde{r}_m^T T_{m-1}^{-1} \tilde{r}_m > 0, \quad T_{m-1} > 0 \quad (24)$$

In regards to the first inequality on the right of  $\Leftrightarrow$  in (24), by the formula (18),  $T_{m-1}^{-1}$  is a rational function in  $r_0, r_1, \dots, r_{m-1}$ , where, due to the right-most inequality in (24), i.e.,  $T_{m-1} > 0$ , the common denominator is positive. Thus multiplying with  $\det(T_{m-1})$ , the first inequality becomes a polynomial inequality. Thus we have shown that  $T_m > 0$  is equivalent to the polynomial inequalities that are assumed equivalent to  $T_{m-1} > 0$  and the polynomial inequality

$$\det(T_{m-1})r_0 - \tilde{r}_m^T \det(T_{m-1})T_{m-1}^{-1}\tilde{r}_m > 0$$

Induction now proves the lemma. ■

##### C. Polynomial optimal experiment design

Using Lemma 4.1 and Lemma 4.2 gives that (7) can be written as

$$\min_r N r_0$$

$$\text{s.t. } Q(r, \theta^\circ) \geq 0$$

$$\det(T_{q-1})r_0 - \tilde{r}_q^T \det(T_{q-1})T_{q-1}^{-1}\tilde{r}_q > 0, \quad q = 0, \dots, m \quad (25)$$

where  $Q$  is a polynomial in  $r$ .

## V. CYLINDRICAL ALGEBRAIC DECOMPOSITION AND EXPERIMENT DESIGN

For a set of polynomial inequalities in  $m$  variables, CAD is a partition of  $\mathbb{R}^m$  into what is called cells where all the inequalities are either true or false in one cell. Information regarding the cells are contained in a set of polynomials where variables systematically have been eliminated. In our case by examining the CAD corresponding to the inequalities in (25) and finding the cell where  $r_0$  is minimum, the optimum can be found. The computational complexity is very high (doubly exponential). There are readily available implementations, e.g. QEPcad [24] and Mathematica. We refer to [20] for a full account of the method. In the control community, CAD has been used for model predictive control [21]. The last reference also contains a lucid description of the principles of CAD, see also [25] for an introduction.

An interesting feature is that the computational complexity reduces drastically if only the minimum value is computed and not the optimizing point. The Mathematica implementation of CAD makes use of this feature.

In terms of experiment design, this opens up for the possibility to assess the loss when lower order approximations are used.

An intrinsic problem in optimal experiment design is that the optimization problem depends on the unknown system (in (25) represented by the dependence of  $Q$  on the true parameter vector  $\theta^o$ ). This problem can be addressed by either robust designs [26, 6] or adaptive (or sequential) designs [27, 7]. However, another interesting property of CAD is that the problem is allowed to be parametric, i.e. the solutions can be computed explicitly as a function of certain parameters of the problem. This may be used to obtain explicit information of how  $\theta^o$  influences the optimal solution.

## VI. NUMERICAL EXAMPLE

In this section a numerical example is presented to illustrate the method. The problem formulation used is the least-costly design in (7) and the model of the system is the FIR model

$$y_t(\theta) = \theta_1 u_{t-1} + \theta_2 u_{t-2} + e_t \quad (26)$$

The disturbance  $e_t$  is a zero mean white Gaussian process with variance 1.  $J(\theta)$  is chosen as  $J(\theta) = J(\theta_1, \theta_2) = \theta_1 \theta_2 + \theta_1^2 \theta_2^2$  and the true parameter vector  $\theta^o$  is given by  $\theta^o = [\frac{1}{10} \frac{1}{7}]^T$ . The covariance matrix  $P$  is in this case given by [22]

$$P = T_1^{-1} = \begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix}^{-1}$$

where  $r_0$  and  $r_1$  corresponds to the autocorrelation coefficients of the input as in (3). Thus  $r_0$  and  $r_1$  are the design variables in the input design.

Inserting  $J(\theta)$  in (6) gives

$$V_{deg} = \text{Var}(J(\theta)) = \text{Var}(\theta_1 \theta_2 + \theta_1^2 \theta_2^2) \quad (27)$$

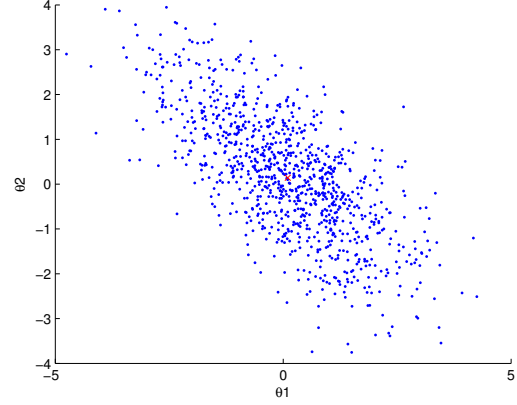


Fig. 1. The obtained parameter estimates when the objective function is linearized,  $n = 1$ . The true parameter  $\theta^o$  is marked as a red cross.

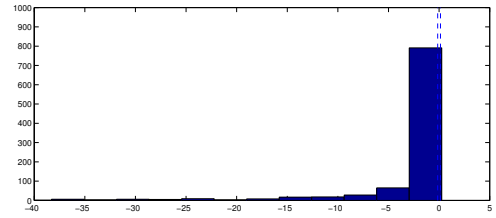


Fig. 2. A histogram of  $J(\theta^o) - J(\theta)$  when  $n = 1$ . The dashed lines are placed at  $\frac{-1}{\sqrt{\gamma}}$  and  $\frac{1}{\sqrt{\gamma}}$ .

The last constraint in (7),  $T_1 > 0$ , is in this case equivalent to  $r_0 \geq |r_1|$ . The problem (7) can thus in this example be formulated as

$$\begin{aligned} \min_{r_0, r_1} N r_0 \\ \text{s.t. } \text{Var}(\theta_1 \theta_2 + \theta_1^2 \theta_2^2) < \frac{1}{\gamma} \\ r_0 \geq |r_1| \end{aligned} \quad (28)$$

In the initial simulations we choose  $N = 1000$  and  $\gamma = 50$ . As a start we try to solve the problem by the approach in Section III. We linearize the function  $J(\theta)$  in order to obtain an SDP in  $r_0$  and  $r_1$ . The SDP is then solved with standard methods in Mathematica and a solution,  $r_0 = 0.0107955$  and  $r_1 = 0.00755685$ , is obtained. Spectral factorization of the corresponding spectrum is used to obtain a realization of the input signal  $u_t$  that corresponds to the obtained autocorrelations coefficients. From this  $u_t$  it is possible to simulate values of  $y_t(\theta^o)$ . Using  $u_t$  and  $y_t$  we estimate  $\theta$  using least-squares and from this estimate,  $\hat{\theta}$ , the estimate of  $J$ ,  $J(\hat{\theta})$ , is obtained. The results of 100 Monte-Carlo simulations are presented in Figure 1 and Figure 2. In Figure 1 the true value  $\theta^o$  is plotted together with the estimated values of  $\theta$ . Figure 2 is a histogram of the values of  $J(\theta^o) - J(\theta)$ . The dashed lines (hardly indistinguishable in the figure) correspond to the allowed region of the standard deviation  $[\frac{-1}{\sqrt{\gamma}}, \frac{1}{\sqrt{\gamma}}]$ .

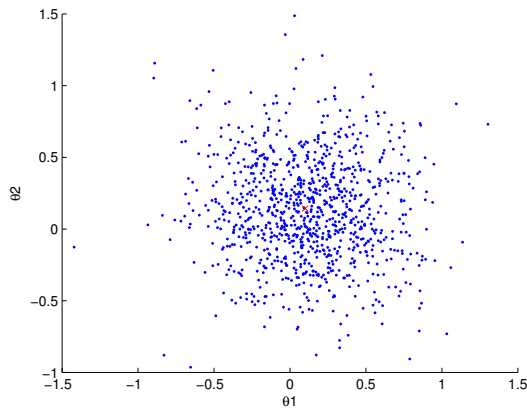


Fig. 3. The obtained parameter estimates when the objective function is approximated by a Taylor expansion of order 2,  $n = 2$ . The true parameter  $\theta^0$  is marked as a red cross.

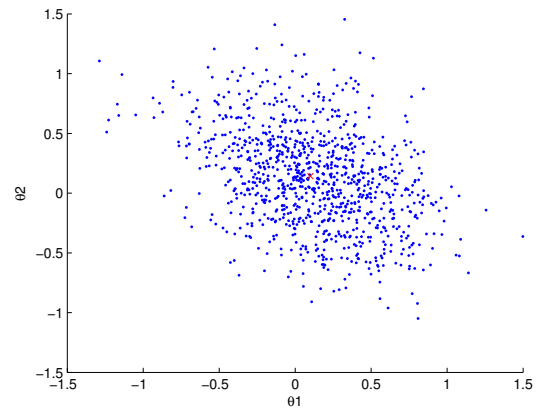


Fig. 5. The obtained parameter estimates when the objective function is approximated by a Taylor expansion of order 4,  $n = 4$ . The true parameter  $\theta^0$  is marked as a red cross.

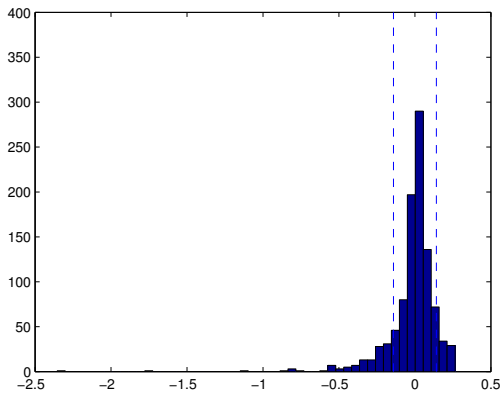


Fig. 4. A histogram of  $J(\theta^0) - J(\theta)$  when  $n = 2$ . The dashed lines are placed at  $\frac{-1}{\sqrt{\gamma}}$  and  $\frac{1}{\sqrt{\gamma}}$ .

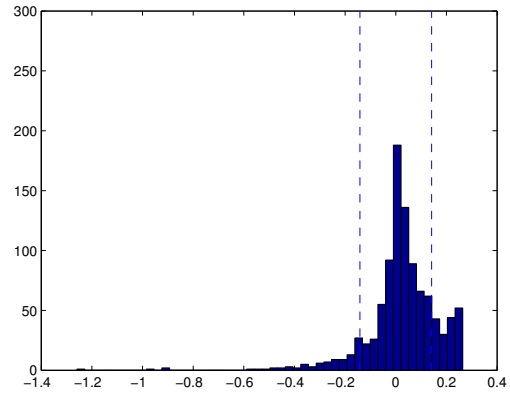


Fig. 6. A histogram of  $J(\theta^0) - J(\theta)$  when  $n = 4$ . The dashed lines are placed at  $\frac{-1}{\sqrt{\gamma}}$  and  $\frac{1}{\sqrt{\gamma}}$ .

Figure 1 shows that the parameter estimate  $\hat{\theta}$  could differ quite significantly from the true parameter  $\theta^0$ . The histogram in Figure 2 emphasize that the input design based on linearizing  $J(\hat{\theta})$  gives very poor accuracy. Hence the first attempt to solve the problem by linearization is inadequate. In order to improve the accuracy of estimate of  $J$  the Taylor approximation is extended to higher order. In this example simulations of order  $n = 2$  and  $n = 4$  are performed. In the case of  $n = 2$  the solution is  $r_0 = 0.0829441$  and  $r_1 = 0.00388727$  and when  $n = 4$  the obtained autocorrelations coefficients are  $r_0 = 0.0751584$  and  $r_1 = 0.0288413$ . The resulting estimates are shown in Figure 3, Figure 4, Figure 5 and Figure 6.

The first thing to notice is that the parameter estimates  $\hat{\theta}$  are much more centered around  $\theta^0$  when the Taylor approximation is of higher order, than in the linearized case. This is of course due to that more input power is used. However, notice also that the “parameter clouds” have different shapes. This is due to that the color of the input spectrum is different. In the linearization approach,  $r_1$  is

close to  $r_0$  giving an input with low frequency behaviour which in turn gives a good estimate of the static gain, i.e. the sum of  $\theta_1$  and  $\theta_2$ . For the higher order solutions,  $r_1$  is much smaller than  $r_0$ , corresponding to an almost white input. This in turn means that the parameter estimates are almost uncorrelated which in Figures 3 and 5 shows up as disc shaped “parameter clouds”.

The histograms of  $J(\theta^0) - J(\theta)$  in Figure 4 and Figure 6 show that the estimates from the higher order solutions give much more accurate estimates of the quantity of interest  $J(\theta^0)$  as compared with the linearization approach (see Figure 4). Another interesting observation is that the estimate of order  $n = 4$  seems to give lower variance than when the order of the Taylor approximation is  $n = 2$ . However the improvement is not as significant as the improvement from the linear case.

In the example above  $\gamma = 50$ . As discussed, this leads to a major difference in accuracy between the case when the function is linearized and when the function is approximated by a Taylor approximation of higher order. As mentioned

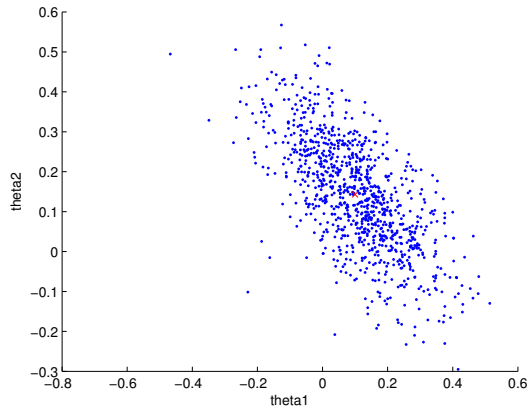


Fig. 7. The obtained parameter estimates when the objective function is linearized,  $n = 1$ , and when  $\gamma$  is increased to  $\gamma = 5000$ . The true parameter  $\theta^0$  is marked as a red cross.

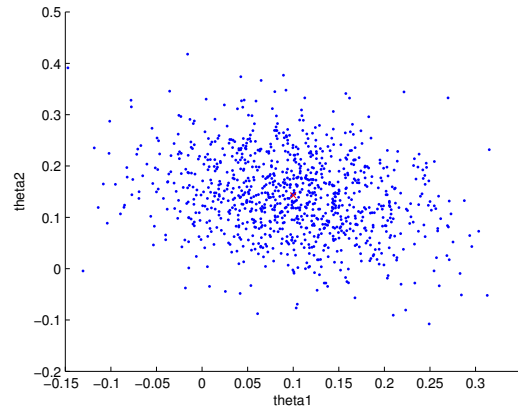


Fig. 9. The obtained parameter estimates when the objective function is approximated by a Taylor expansion of order 2,  $n = 2$  and when  $\gamma$  is increased to  $\gamma = 5000$ . The true parameter  $\theta^0$  is marked as a red cross.

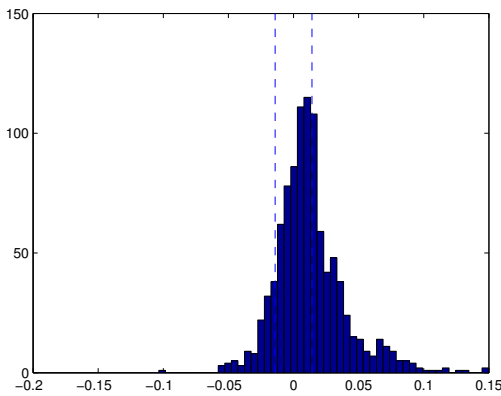


Fig. 8. A histogram of  $J(\theta^0) - J(\theta)$  when  $n = 1$  and  $\gamma = 5000$ . The dashed lines are placed at  $\frac{-1}{\sqrt{\gamma}}$  and  $\frac{1}{\sqrt{\gamma}}$ .

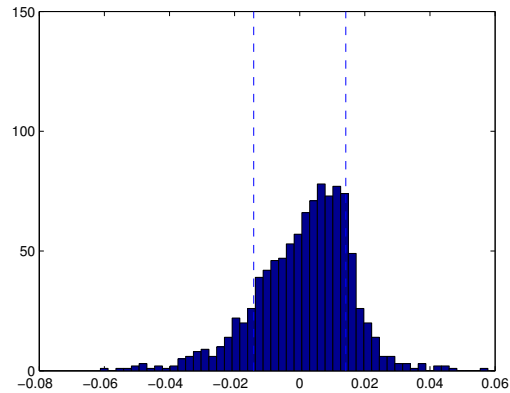


Fig. 10. A histogram of  $J(\theta^0) - J(\theta)$  when  $n = 2$  and  $\gamma = 5000$ . The dashed lines are placed at  $\frac{-1}{\sqrt{\gamma}}$  and  $\frac{1}{\sqrt{\gamma}}$ .

in Section I this is typical for low to medium performance applications. An interesting continuation of this example is thus to increase  $\gamma$  in order to simulate a model in high performance applications. In the following we set  $\gamma = 5000$ . On the basis of the discussions in Section I we expect the difference between when the objective function is linearized and when a Taylor approximation of higher order is used to be smaller. When the problem is solved in Mathematica we get the solution  $r_0 = 1.07955$  and  $r_1 = 0.755685$  for the linearized case and  $r_0 = 1.71264$ ,  $r_1 = 0.473605$  when  $n = 2$ . In Figure 7 and Figure 8 the results for the linearized case is shown. These should be compared to Figure 9 and Figure 10 when the objective function is a second order Taylor approximation.

By studying Figure 7 and Figure 9 we note that both the case with a linearized objective function and the case when the objective function is approximated by a Taylor approximation of order 2 leads to good estimates of  $\theta^0$ . The histograms in Figure 8 and Figure 10 show the same. If the figures are studied a bit more carefully, a slight

difference between the two cases can be detected. The approximation of order 2 seems to give a slightly more accurate approximation of  $\theta^0$  and  $J(\theta^0)$ . However it is quite obvious that the difference between the cases is much smaller than in the initial examples with a lower  $\gamma$ , as expected from the discussion in Section I. These last results stress that a linearization of the objective function is sufficient in high performance applications. There is little to gain by higher order approximations in such cases.

## VII. CONCLUSIONS

In this contribution we have illustrated that the standard approach of linearizing the quantity of interest may give poor results in experiment design. This holds in particular when the demand on accuracy is low. As a remedy we proposed to increase the order of the approximation. We formally showed that this leads to optimization problems that are polynomial in the decision variables and we have proposed the use of cylindrical algebraic decomposition to solve such problems. The computational complexity grows very quickly with the size of the problem but can be reduced by just computing

the cost function. This value can be used to benchmark the linearized solution.

#### REFERENCES

- [1] L. Pronzato, "Optimal experimental design and some related control problems," *Automatica*, vol. 44, no. 2, pp. 303–325, February 2008.
- [2] M. Gevers, "Identification for control: From the early achievements to the revival of experiment design," *European journal of control*, vol. 11, no. 4-5, pp. 335–352, 2005, semi-plenary lecture at IEEE Conference on Decision and Control - European Control Conference.
- [3] H. Hjalmarsson, "From experiment design to closed loop control," *Automatica*, vol. 41, no. 3, pp. 393–438, March 2005.
- [4] —, "System identification of complex and structured systems," *European Journal of Control*, vol. 15, no. 4, pp. 275–310, 2009, plenary address. European Control Conference.
- [5] L. Ljung, "Perspectives in system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.
- [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, June 2007.
- [7] L. Gerencsér, H. Hjalmarsson, and J. Mårtensson, "Identification of ARX systems with non-stationary inputs - asymptotic analysis with application to adaptive input design," *Automatica*, vol. 45, no. 3, pp. 623–633, March 2009.
- [8] X. Bombois, G. Scorletti, M. Gevers, P. M. J. Van den Hof, and R. Hildebrand, "Least costly identification experiment for control," *Automatica*, vol. 42, no. 10, pp. 1651–1662, 2006.
- [9] H. Jansson and H. Hjalmarsson, "Input design via LMIs admitting frequency-wise model specifications in confidence regions," *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1534–1549, 2005.
- [10] H. Hjalmarsson and J. Mårtensson, "Optimal input design for identification of non-linear systems: Learning from the linear case," in *American Control Conference*, New York City, USA, July 11-13 2007.
- [11] T. L. Vincent, C. Novara, K. Hsu, and K. Poolla, "Input design for structured nonlinear system identification," *Automatica*, vol. 46, no. 6, pp. 990–998, Jul 2010.
- [12] C. A. Larsson, H. Hjalmarsson, and C. R. Rojas, "On optimal input design for nonlinear FIR-type systems," in *Proceedings 49th IEEE Conference on Decision and Control*, Atlanta, GA, USA, 2010.
- [13] M. Gevers and L. Ljung, "Optimal experiment designs with respect to the intended model application," *Automatica*, vol. 22, no. 5, pp. 543–554, 1986.
- [14] X. Bombois, H. Hjalmarsson, and G. Scorletti, "Identification for robust  $H_2$  deconvolution filtering," *Automatica*, vol. 46, no. 3, pp. 577–584, Mar. 2010.
- [15] E. Olofsson, H. Hjalmarsson, C. R. Rojas, P. Brunzell, and J. Drake, "Vector dither experiment design and direct parametric identification of reversed-field pinch normal modes," in *Proceedings 48th IEEE Conference on Decision and Control*, Shanghai, China, 2009.
- [16] C. Larsson, C. Rojas, and H. Hjalmarsson, "MPC oriented experiment design," in *18th IFAC World Congress*, Milano, Italy, 2011.
- [17] R. Vuerinckx, R. Pintelon, J. Schoukens, and Y. Rolain, "Obtaining accurate confidence regions for the estimated zeros and poles in system identification problems," *IEEE Trans. on Automatic Control*, vol. 46, no. 4, pp. 656–659, April 2001.
- [18] P. Parillo, "Semidefinite programming relaxations for semialgebraic problems," *Mathematical Programming Ser. B*, vol. 96, no. 2, pp. 293–320, 2003.
- [19] D. Henrion and J. Lasserre, "Convergent relaxations of polynomial matrix inequalities and static output feedback," *IEEE Trans. on Automatic Control*, vol. 51, no. 2, pp. 192–202, 2006.
- [20] B. Caviness and J. Johnson, *Quantifier elimination and cylindrical algebraic decomposition*. Vienna: Springer Verlag, 1998.
- [21] I. Fotiou, P. Parrilo, and M. Morari, "Nonlinear parametric optimization using cylindrical algebraic decomposition," in *Proceedings 42nd IEEE Conference on Decision and Control, and the European Control Conference 2005*, Seville, Spain, December 2005, pp. 3735–3740.
- [22] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.
- [23] C. Byrnes, S. Gusev, and A. Lindquist, "From finite covariance windows to modeling filters: A convex optimization approach," *SIAM Review*, vol. 43, pp. 645–675, 2001.
- [24] C. Brown, "QEPCAD B: a program for computing with semialgebraic sets using CADs," *ACM SIGSAM Bulletin*, vol. 37, pp. 97–108, 2003.
- [25] M. Jirstrand, "Cylindrical algebraic decomposition - an introduction," Dept of EE, Linköping University, S-581 83 Linköping, Sweden, Tech. Rep. LiTH-ISY-R-1807, Dec. 1995.
- [26] L. Pronzato and E. Walter, "Robust experiment design via maximin optimization," *Mathematical Biosciences*, vol. 89, no. 2, pp. 161 – 176, 1988.
- [27] I. Ford and S. Silvey, "A sequentially constructed design for estimating a nonlinear parametric function," *Biometrika*, vol. 67, pp. 381–388, 1980.