

Bayesian Design of Experiments for Nonlinear Dynamic System Identification

Susanne Zaglauer
BMW Group
München, D-80788 Germany
e-mail: susanne.zaglauer@bmw.de

ABSTRACT

D-optimal and other model-based experimental designs, like the Query by Committee criterion, are often criticised because of their dependency to the statistical model and the lack of the explicit allocation of the identified irregularities of the assumed model. Furthermore, D-optimal experimental designs tend to weight the boundary area of the experimental space significantly. In extreme cases the boundary points of the experimental space are the experimental candidates. In order to defuse this critique a new online Bayesian design for the nonlinear dynamic system identification is introduced, which serves the flexibility and which is concurrently more resistant against the bias caused by the model. An advantage of this approach is also the small modification of the D-optimal design to increase the independency of the D-optimal design to the chosen model approach.

This contribution includes the presentation of different methods for the Design of Experiments for nonlinear dynamic system identification. Therefore, model-free and model-based based experimental designs are introduced and the new online Bayesian design for nonlinear dynamic system identification is represented. The presented design plans are appropriate to identify the dynamics of combustion engines, can be used to identify the parameters of both analytical and simulation models and are evaluated by simulations.

Categories and Subject Descriptors

G.3 [Probability and Statistics]: Experimental design
; I.6 [Simulation and Modeling]: Model Development

General Terms

Measurement, Performance, Design, Experimentation

Keywords

nonlinear dynamic system identification, Fischer Information, Design of Experiments, online Design of Experiments, D-optimal design, Bayesian design

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Simutools 2012, March 19-23, Desenzano del Garda, Italy
Copyright © 2012 ICST 978-1-936968-47-3
DOI 10.4108/icst.simutools.2012.247734

1. INTRODUCTION

The automotive industry is, based on the sales, the most important industry section in Germany and other countries ([1]). Because of many various boundary conditions, like CO_2 and fuel reduction, the calibration of the engine control unit is growing up for the development of the whole system automotive. Traditional methods fail because of the increasing complexity and the huge measurement effort.

Therefore, the statistic Design of Experiments (*DoE*) gains clearly in importance in engine development. The application area of the statistic *DoE* for modeling and systematic optimization in engine development goes way beyond the classic quality tasks and is applied to almost every steps of development. The goal is to achieve statistical verified results with low costs. In order to achieve this goal and to reduce the huge measurement effort the real dynamic systems are reproduced by parametric models, whose parameters are estimated on the basis of measured data. The *DoE*, which determines the distribution of the measuring points, is therefore of vital importance for the quality of the identified model. Only a data set, which contains enough information about the system, allows an accurate estimation of the parameters and the identification of a good model. So the target of the *DoE* is to cover the connections between target and influence factors systematically and to achieve a maximum of information of the system with as few experiments as possible. The design plans can also be used to identify the parameters of both analytical and simulation models and they are divided into model-free and model-based *DoE*. The model-based designs are dependent on the used model and the D-optimal designs ([8]) are the most popular ones. But these are often criticized because of their dependency to the statistical model and the lack of the explicit allocation of the identified irregularities. The main target of this contribution is the introduction and presentation of the new online Bayesian design, which is a modification of the D-optimal design and compensates the mentioned disadvantages.

This contribution is organized as follows. The second section contains the various design plans. First model-free designs and then model-based *DoE* are presented. Thus, a mathematical derivation of the Fischer Information Matrix is introduced, which plays an important role in the subsequent presented D-optimal design and in the new online Bayesian design. In the fourth section the previously introduced experimental designs are evaluated by simulations and a summary of all advantages of the Bayesian Design of Experiments for nonlinear dynamic system identification and an outlook in future projects conclude the contribution.

2. DESIGN OF EXPERIMENTS

An important question is how the measuring points should be distributed efficiently in the experimental space. The answer to this question is provided by the Design of Experiments. The goal is to identify the connections between target and influence factors systematically with as few experiments as possible. This means to achieve the maximum of information of the system under investigation with every measurement.

The kind of *DoE* determines the distribution of the points in the experimental space and various design methods can be found for example in [6], [7], [13] and [15]. At first model-free and then online model-based designs are presented, where the Bayesian design is new for the dynamic system identification and will be further examined.

2.1 Model-free Design of Experiments

In the practice it is not possible or it takes too many time to calculate any derivatives or covariance matrices from simulation models of technical design problems and so model-free Design of Experiments are very important in this context.

Model-free Design of Experiments do not care about the used model structure and try to cover the input space as equal as possible (space filling designs). The minimal distance is usually maximized, alternative strategies can be found in [11] for example. Thus, it can be guaranteed that informations about all areas of the input space are available. The most popular space filling designs are the random, the Latin Hypercube distribution ([5]) and the distance-based criteria, especially the S-optimal design ([9]).

2.1.1 The Random Design of Experiments

As the name implies, random experimental designs involve randomly assigning experimental conditions. This type of experimental design is surprisingly powerful and often results in a high probability to create a good design. Random designs typically work well for large systems with many variables, 50 or more. There should be few interactions between the variables and very few variables that contribute significantly. Random design does not work very well with relatively smaller systems.

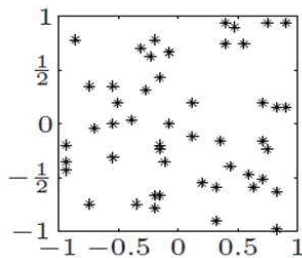


Figure 1: Representation of the random distribution of the design points.

The traditional random method gives only an ensemble of random numbers without any uniform distribution in the

input space (see Figure 1).

2.1.2 The Latin Hypercube Design

In this schema only one sampling point is in every column and row of a grid. For this purpose one sampling position is placed randomly in every cell along the grid diagonal. After this the rows of the grid are changed in that way, that a chosen criteria, like the maximizing of the minimal distance of the points, is fulfilled:

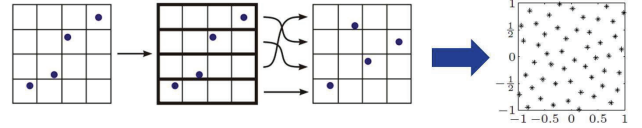


Figure 2: Representation of the maximin Latin-Hypercube-Distribution for the choice of the design points.

The Latin Hypercube design makes sure that the ensemble of random numbers represents the real oscillations and the points are distributed well in the input space.

2.1.3 Distance-based Criteria

The distance-based criteria are based on the distance $d(x, \mathcal{A})$ from a point x in the p -dimensional Euclidean space \mathbb{R}^p to a set $\mathcal{A} \subset \mathbb{R}^p$. This distance is defined as follows:

$$d(x, \mathcal{A}) = \min_{y \in \mathcal{A}} \|x - y\|,$$

where $\|x - y\|$ is the usual p -dimensional Euclidean distance:

$$\|x - y\| = \sqrt{(x_1 - y_1)^2 + \dots + (x_p - y_p)^2}.$$

S-optimality seeks to maximize the harmonic mean distance from each design point to all the other points in the design:

$$\max_{y \in \mathcal{C}} \frac{N_{\mathcal{D}}}{\sum_{y \in \mathcal{D}} \frac{1}{d(y, \mathcal{D} - y)}},$$

where \mathcal{C} is the set of candidate points and \mathcal{D} is the set of design points. For an S-optimal design, the distances $d(y, \mathcal{D} - y)$ are large, so the points are as spread out as possible. Since the S-optimality criterion depends only on the distances between design points, it is usually easier to compute and optimize than other distance-based criteria, like the U-optimal one, which depends on the distances between all pairs of candidate points.

An interesting approach is also the combination between the Latin Hypercube and the S-optimal design. Therefore, the rows of the Latin Hypercube grid are changed in that way, that the harmonic mean distance from each design point to all the other points in the design is maximized.

2.2 Model-based Design of Experiments

If the derivatives and the covariance matrix of the analytical or simulation model can be calculated, it is more useful to use model-based Design of Experiments.

Model-based Design of Experiments are aimed at the distribution of the design points in the input space that the model

parameter estimation is as insensitive as possible against the measurement noise. In contrast to the model-free *DoE* the model-based designs are more powerful and create a near optimal design. For this reason these designs are discussed more precisely than the model-free designs. In the following the Query by Committee ([12]) and the D-optimal design ([8]) are introduced and a new Bayesian design for the non-linear dynamic system identification is presented.

2.2.1 Query by Committee (QBC)

The Query by Committee (QBC) is an online Design of Experiments method and is based on a nonlinear model $g(u) = g(u, \theta)$, on which an optimized parameter vector $\hat{\theta}$ was estimated on the basis of measuring datas $(U, Y) = (u_i, y_i)_{i=1}^K$. Now you are looking for a procedure to determine K more measuring points in the input space on the basis of the candidate set. For the evaluation of the candidates a query criterion q_{qbc} is defined, which is a measure of the expected information content if one or more candidates are chosen. The next measuring points must be chosen in that way that the q_{qbc} achieves its maximum. Thus an optimization problem must be calculated again. An important query criterion is the variance of the model in one point:

$$q : U \rightarrow \mathbb{R}^+ \quad \text{with} \quad q_{qbc}(u) = \text{var}(g(u, \hat{\theta})).$$

It is not simple to calculate the variance of some models and so instead of only one model a model committee $\{g_i(u)\}$ of N models is used, then the output of the committee in one point $u \in U$ is usually the mean of the weighted outputs of the committee candidates.

$$g(u) = \frac{1}{N} \sum_{i=1}^N w_i g_i(u),$$

where w_i stands for the weighting of model i . The uncertainty of the committee can be quantified through the non-conformity or the dissense of the committee:

$$\text{var}(g(u)) = \frac{1}{N-1} \sum_{i=1}^N (g_i(u) - g(u))^2.$$

u_1 is calculated on the basis of q_{qbc} and every new measuring point u_j through the maximization of

$$q_{qbc}(u) = \text{var}(g(u)).$$

The variance of the committee reflects the model uncertainty. The expected information content by means of a measurement is more higher the higher the variance is. This criterion has many good properties, which can be advantageous:

- It is suitable for the consideration and integration in a query criterion of different network architectures or model types, which means different linear and nonlinear models.
- An evaluation requires less effort if the model committee is available.

2.2.2 D-optimal design

Beside the Query by Committee the D-optimal design is also a powerful design. To estimate the variance of a parameter estimation of any parametric model the Fischer Information Matrix can be used. The D-optimal and the following Bayesian design are based on the Fischer Information (FI) and therefore it is explained in this section (see also [14]).

The Fischer Information

The Fischer Information describes the information content of a random variable x at the parameter θ from which the Likelihood-function $L(\theta) = p(x, \theta)$ is dependent

$$FI(\theta) = E \left[\frac{\partial \ln p(x, \theta)^2}{\partial \theta} \right].$$

For an unbiased estimator $\hat{\theta}(x)$ it applies:

$$E[\hat{\theta}(x) - \theta] = \int [\hat{\theta}(x) - \theta] p(x, \theta) dx = 0. \quad (1)$$

The Likelihood-function $p(x, \theta)$ describes the probability that for a given θ the random variable x is observed. If $p(x, \theta)$ has a "sharp peak" it is easy to estimate the "correct" parameter θ , i.e. the data include a lot of information about θ .

If the Likelihood-function is flat and sweeping many data are needed to estimate the parameter θ (the data include few information about θ).

By differentiation of equation (1) we obtain:

$$\frac{\partial}{\partial \theta} \int [\hat{\theta}(x) - \theta] p(x, \theta) dx = \int (\hat{\theta}(x) - \theta) \frac{\partial}{\partial \theta} p(x, \theta) dx - \int p(x, \theta) dx.$$

As the Likelihood-function is a probability distribution, it applies:

$$\int p(x, \theta) dx = 1$$

and through the relation $(\ln(x))' = \frac{1}{x}$

$$\frac{\partial}{\partial \theta} p(x, \theta) = p(x, \theta) \frac{\partial}{\partial \theta} \ln p(x, \theta).$$

This results in the correlation:

$$\int (\hat{\theta}(x) - \theta) p(x, \theta) \frac{\partial}{\partial \theta} \ln p(x, \theta) dx = 1.$$

Through factorization of the integrand one obtains:

$$\int \left((\hat{\theta}(x) - \theta) \sqrt{p(x, \theta)} \right) \left(\sqrt{p(x, \theta)} \frac{\partial}{\partial \theta} \ln p(x, \theta) \right) dx = 1.$$

By squaring the equation and usage of the Cauchy-Schwarz-inequality it is implied:

$$\left[\int (\hat{\theta}(x) - \theta)^2 p(x, \theta) dx \right] \cdot \left[\int \left(\frac{\partial}{\partial \theta} \ln p(x, \theta) \right)^2 p(x, \theta) dx \right] \geq 1. \quad (2)$$

The second factor of (2) is the Fischer Information

$$FI(\theta) = \int \left(\frac{\partial \ln p(x, \theta)}{\partial \theta} \right)^2 p(x, \theta) dx.$$

The first factor is the expected square error of the estimator θ , because

$$E[(\hat{\theta}(x) - \theta)^2] = \int (\hat{\theta} - \theta)^2 p(x, \theta) dx.$$

Thus, the uncertainty in the parameters is characterized by the inverse of the Fischer Information.

This result is also presented by the Cramer-Rao-inequality which says that the inverse of the Fischer Information is a lower bound at the variance of an unbiased estimator of θ and the precision with who θ can be estimated is limited by (for the proof see [18]):

$$\text{var}(\hat{\theta}(x)) \geq FI(\theta)^{-1}. \quad (3)$$

With N parameters with $\theta = [\theta_1, \theta_2, \dots, \theta_N]$ the Fischer Information is a $N \times N$ matrix FIM

$$FIM(\theta) = E \left[\frac{\partial}{\partial \theta} \ln p(x, \theta) \frac{\partial}{\partial \theta} \ln p(x, \theta)^\top \right]$$

and in equation (3) the variance must be replaced by the covariance matrix.

If the random vector y assumes a multivariate Gaussian distribution $y \sim N(\mu(\theta), \Sigma(\theta))$ with mean value $\mu(\theta)$ as known function of the unknown parameter vector θ and covariance matrix Σ the probability distribution of y is then

$$p(y, \theta) = \frac{1}{(2\pi)^{n/2} (\det(\Sigma))^{1/2}} \exp \left(-\frac{(y - \mu)^\top \Sigma^{-1} (y - \mu)}{2} \right)$$

$$\Rightarrow FIM(\theta) = \frac{\partial}{\partial \theta} \mu^\top(\theta) \Sigma^{-1} \frac{\partial}{\partial \theta} \mu(\theta) + \frac{1}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta} \right). \quad (4)$$

With uncorrelated Gaussian Noise $\Sigma(\theta) = \sigma^2 I$ it follows from equation (4) and [17]

$$FIM(\theta) = \frac{1}{\sigma^2} \frac{\partial \mu^\top(\theta)}{\partial \theta} \frac{\partial \mu(\theta)}{\partial \theta}. \quad (5)$$

After the introduction of the Fischer Information Matrix now the D-optimal design is presented. Therefore, a nonlinear dynamic system

$$\hat{y}_k = f(u_{k-1}, \dots, u_{k-id}, \hat{y}_{k-1}, \dots, \hat{y}_{k-od}) + \epsilon_k. \quad (6)$$

is used, whereby the next output \hat{y}_k of the model is dependent to the last inputs $u_k \in \mathbb{R}^n$ and outputs $\hat{y}_k \in \mathbb{R}$. The parameters id and od stand for the maximal input and output delay. The noise ϵ should be Gaussian distributed with mean 0 and uncorrelated in k :

$$E(\epsilon_k) = 0, \quad E(\epsilon; \epsilon_k^\top) = \delta_{i,k} C$$

with $C \in \mathbb{R}^{m \times m}$ as known positive definite matrix ($C = \sigma^2 I_m$), σ^2 : variance and $\delta_{i,k}$: Kronecker Delta.

The D-optimal design, which minimizes the volume of the confidence ellipsoid, which approximates the used parameters, is defined by the maximization of the determinant of the inverse of the Fischer Information Matrix. It is now differentiated between the offline and the online D-optimality.

Offline D-optimality

The offline D-optimal design can be calculated before starting the measurements and thus are called offline *DoE*.

Therefore, the *FIM* can only be calculated independently of the estimated parameters if models are linear in their parameters. Thus in practice the complexity of the approximation problem is often estimated and an experimental design based on an accordingly complex polynomial model is calculated. For a model $\hat{y} = \mathcal{X}\theta$, which is linear in its parameters, the D-optimal criterion is given by

$$\max \det(\mathcal{X}^\top \mathcal{X})^{-1}.$$

This is independent of the parameter θ and so it is possible to determine the *DoE* completely without any measurement.

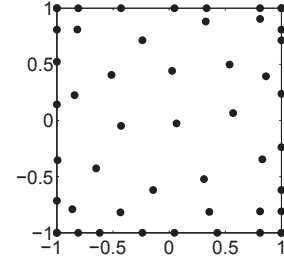


Figure 3: Distribution of the offline D-optimal design.

Figure 3 shows a typically D-optimal distribution with 50 design points based on a polynomial model of third order. In contrast to other distributions mainly the boundary of the input space is covered well. Thus, no extrapolation appears in the boundary area.

Online D-optimality

In contrast to the offline D-optimality, model-based online methods allow to adapt the *DoE* under consideration of the gained information. With the use of adaptive *DoE* it is possible to improve models at certain areas of the input space which are relevant to the identification problem. Now to find online an optimal distribution of the points in the experimental space the *FIM* of the discrete models must be maximized over a considered time horizon:

$$\max \det \left(\sum_{j=k}^{k+N_s} FIM(\theta_k^{(j)}) + FIM_0^{(i)}(\theta_k^{(i)}) \right)$$

where

$$FIM_0^{(i)}(\theta_k^{(i)}) = \sum_{j=1}^{k-1} FIM(\theta_k^{(j)})$$

is the *FIM*, which depends on the previous training data and the actual parameter estimation $\theta_k^{(i)}$.

The D-optimal criterion has also many good properties, which can be advantageous:

- D-optimal Design of Experiments support in contrast to other classical design plans also polynomials of degree > 2 .
- If experimental data are given or certain areas of the variation domain should especially be considered, these points can be included into the design plan.

- The experimental domain can be arbitrary pretended, uninteresting areas can be excluded for example.

But the D-optimal design has also some disadvantages (see [16]):

- D-optimal Design of Experiments are very dependent to the statistical model.
- They have the lack of the explicit allocation of the identified irregularities of the assumed model.
- D-optimal Design of Experiments tend to weight the boundary area of the experimental space significantly. In extreme cases the boundary points of the experimental space are the experimental candidates.

2.2.3 Bayesian design

To compensate the disadvantages of the D-optimal design now the new Bayesian design is presented. In the following a nonlinear dynamic model described in (6) is used and we differentiate again between offline and online.

Offline Bayesian Design of Experiments

As seen before the D-optimal design maximizes the determinant of the Fischer Information Matrix [2]

$$FIM(\theta) = \frac{1}{\sigma^2} \frac{\partial \mu^\top(\theta)}{\partial \theta} \frac{\partial \mu(\theta)}{\partial \theta} = \frac{1}{\sigma^2} \sum_{k=1}^K R_k R_k^\top$$

with $R_k = \left(\frac{\partial f(p, u_k)}{\partial p} \right)_{p=\hat{p}}^\top$.

R_k should be chosen such that the uncertainty of θ is small and the increase of the determinant of $(R_k R_k^\top)$ reduces the error variances of the coefficients which are proportional to $(R_k R_k^\top)^{-1}$.

In this definition the design matrix has one column for every term in the assumed model. This model is an approximation to the behavior of the real system. It seems quite natural to ask for the quality of the approximation. But to maximize $\det(R_k R_k^\top)$ the whole experimental effort is spent by the precise estimation of the assumed model parameters. Therefore the D-optimal design does not supply the detected irregularities in the model. To avoid this you may try with the Bayesian design to add central points or other control points both to the factorial and to the optimal design to decide whether the model has uncertainties.

Control points in a design are useful mainly for the protection against higher order effects which are included in the model. These terms are called potential terms. The assumed model contains only primary terms. Typically the sample size is not large enough to estimate all primary and potential terms simultaneously. The issue here is to develop an approach which estimates the primary terms precisely during a general traceability and an estimation is provided for the potential terms.

Suppose there are q potential terms that are just possibly important in addition to the p primary terms that you really want to fit. It is $R = (R_{k_{pri}} | R_{k_{pot}})$, whereby X has $p+q$ columns. The sample size n is often not large enough to estimate all $p+q$ parameter precisely, typically $p < n < p+q$ holds.

Now a introduction of scaling and centering is necessary to

allow the use of standardized prior distributions of the coefficients to work well at all diverse contexts. Without loss of generality it would be supposed that the residual error of y is equal to 1 such that every non constant term varies between -1 and 1 and every potential term is approximately uncorrelated with all primary terms. Therefore

$$\max(R_{k_{pri}}) = 1 \text{ and } \min(R_{k_{pri}}) = -1$$

hold for every non constant primary term and

$$\max(R_{k_{pot}}) - \min(R_{k_{pot}}) = 1$$

holds for every potential term with $\sigma^2 = 1$. For every pair of primary and potential terms holds

$$\sum_{\text{candidates}} R_{k_{pot}} R_{k_{pri}} = 0$$

whereby \max, \min and \sum are calculated for the design over the whole set of points.

The scaling can be reached by execution of a regression of the potential terms against the primary terms with use of the potential points to evaluate α and Z whereby

$$\alpha = (R_{k_{pri}}^\top R_{k_{pri}})^{-1} R_{k_{pri}}^\top R_{k_{pot}}$$

$$r = R_{k_{pot}} - R_{k_{pri}} \alpha$$

and

$$Z = \frac{r}{\max(r) - \min(r)}$$

whereas \max, \min and the definition of α use the set of the potential points.

Therefore the definition of R_k leads to $(R_{k_{pri}} | Z)$ instead of $(R_{k_{pri}} | R_{k_{pot}})$. In addition α is the lowest quadratic regression coefficient of $R_{k_{pot}}$ on $R_{k_{pri}}$ and r is the residual of the regression.

Now K is a $(p+q) \times (p+q)$ diagonal matrix whose first p diagonal elements are equal to 0 and whose remaining q diagonal elements are equal to 1. Under this conditions it follows that $\sigma^2 = 1$ and $Y | \theta \sim \mathcal{N}(R_k \theta, I)$. With the Bayesian rule for conjugate Gaussian distributions it leads to $\theta | Y \sim \mathcal{N}(b, [R_k R_k^\top + K/\tau^2]^{-1})$ with $b = [R_k R_k^\top + K/\tau^2]^{-1} R_k^\top Y$. In the case of the availability of the prior distribution the Bayesian design is that design which maximizes $(R_k R_k^\top + K/\tau^2)$ with respect to R_k . A great advantage of this approach is that D-optimal designs only need little modifications to solve the problem. In our experience no other approaches are known to solve singular design problems with the help of a suitable prior distribution to make the D-optimal design less model dependent.

=====

Example

This example from [4] represents a simple situation, in which the ordinary D-optimal approach works poorly but the proposed Bayesian modification works well.

Suppose that A and B are quantitative factors. An interaction model is looked for an experiment with $n = 5$ whereby the design points are $(a_i, b_i), i = 1, \dots, 5$ with a discretized 5×5 set of possible points: $a_i, b_i \in \{-1, -0.5, 0, 0.5, 1\}$ as shown in Figure 4.

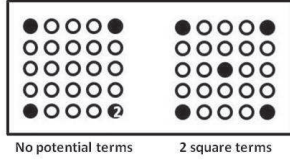


Figure 4: D-optimal design for Interaction model (left) and Bayesian version (right) if $n = 5$ and the Potential terms are (A^2, B^2) .

As seen on the left side of Figure 4 the D-optimal design concentrates on the four corner points replicated in one of them. If now the two pure quadratic terms represent the potential terms in the Bayesian model (with $\tau = 1$) then the design give a Bayesian justification for the addition of the central point which can be seen on the right side of Figure 4. To check the evaluations the i -th row of $R_{(5 \times 6)}$ is $(1, a_i, b_i, a_i b_i, a_i^2 - 0.5, b_i^2 - 0.5)$. The first $p = 4$ columns are primary terms, the last $q = 2$ columns of X are potential terms such that $p = 4 < n = 5 < p + q = 6$. One may choose $(a_i, b_i), i = 1, \dots, 5$ to maximize the determinate of $(R^T R + K)$ whereby

$$K = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The choice of a central point appears by a default value of $\tau = 1$. If K is replaced by K/τ^2 the central point would be added for all $\tau > 0.61$. If $\tau < 0.61$ the design concentrates on the four corner points like the D-optimal one. The sudden change whether a central point is added or not takes place at $\tau = 0.612$. The choice of τ plays an important role for the design and is set to the default value $\tau = 1$ at the simulation results in this contribution because at this value an improvement of the normal D-optimal design is expected with high probability.

=====

Online Bayesian Design of Experiments

For an online optimization for the prediction horizon T_i the optimal design point s_i^* is defined with the Fischer Information matrix of (5):

$$\max_{s_i} \frac{T_s}{T_i} \sum_{j=1}^N \left(\det \left(\sum_{k=1}^{T_i/T_s} \omega_j FIM_k^{(j)}(\hat{p}, u_k(s_i, s_{i-1})) + K/\tau + F_0^{(j)} \right) \right)$$

under the constraints

$$h(u_t) \leq 0, \quad \forall t \geq k \quad (7)$$

$$g(y_t) \leq 0, \quad \forall t \geq k + 1 \quad (8)$$

whereby $U = \{U_i\}_{i=1}^K$ is the input sequence for all K design points and the total time $T = \sum T_i = \sum T_{h,i} + T_{r,i}$ and the inequalities are described by the functions $h : \mathbb{R}^n \rightarrow \mathbb{R}^q$ and $g : \mathbb{R}^m \rightarrow \mathbb{R}^p$. ω_j stands for the weighting of model j and N

for the number of the model in the committee. The input sequence for a design point s_i with holding time $T_{h,i}$ and ramp time $T_{r,i}$ is described by (see [2])

$$U_i(s_{i-1}, s_i, T_{h,i}) = \{u_k\}_{k=1}^{T_i/T_s}$$

$$u_k(s_i, s_{i-1}) = \begin{cases} s_i, & \text{für } \frac{T_{r,i}}{T_s} \leq k \leq \frac{T_i}{T_s} \\ g \cdot k + s_{i-1}, & \text{für } 1 \leq k \leq \frac{T_{r,i}}{T_s}, \end{cases}$$

whereby the slope g is limited by the maximal slope g_{max} and evaluated by

$$g = \frac{g_{max}}{T_{r,i}} \alpha$$

$$T_{r,i} = \left\lceil \frac{1}{T_s} \max_j (|\alpha_j|) \right\rceil T_s$$

$$\alpha = \frac{1}{g_{max}} (s_i - s_{i-1}).$$

$F_0^{(j)}$ is the *FIM*, which depends on the previous training data and the actual parameter estimation. So long as the holding times $T_{h,i}, i = 1, \dots, K$ are independent to the design points they should be determined at the beginning only and the input sequence U_i must be optimized for the unique parameter, the design point s_i .

3. EXPERIMENTS

In the following only the online model-based designs are compared to each other, because they are more powerful and the performance of the new online Bayesian design for the nonlinear dynamic system identification should be shown. Therefore, the Bayesian design is compared with the Query by Committee criterion and the D-optimal design by means of test examples with a committee of Neural Nets and then the Bayesian design is compared with the D-optimal design by means of test examples with a polynomial. With these experiments it is shown that the Bayesian design achieves better results both for the presented dynamic system identification and also for the static system identification. The following various test are performed by the test system:

- Comparison of the Bayesian design with the Query by Committee criterion and the D-optimal design on the basis of a committee of Neural Nets.
- Comparison of the Bayesian design with the D-optimal design on the basis of a polynomial.

3.1 Comparison of the design on the basis of dynamic system

The results are presented through Box plots. At first the Bayesian design is compared to the Query by Committee criterion and the D-optimal design on the basis of a committee of Neural Nets. The comparison is performed by 100 different training data sets. Every of these data sets include 1024 data points. The measured outputs are afflicted with noise whereby the signal to noise ratio (*SNR*) is 20. Seven Neural networks are in the committee weighted with a Bayesian method and an adaptive choice of the neurons. As design criteria the D-optimal, the Query by Committee

and the Bayesian design are taken. Altogether 30 percent of the training data are trained online and 70 percent offline. The prediction horizon has the value 1. The neural nets are excited by an APRBS signal (see [3]) and as an amount of error for the evaluation the normalized root mean squared error (*NRMSE*) is used:

$$NRMSE = \frac{\sqrt{e^T e}}{K(\text{var}(y))},$$

where e is the difference between system output and model output and $\text{var}(y)$ is the variance of the measurement signal. Many measured variables, which are relevant for the behavior of a combustion engine (e.g. the exhaust gas temperature), feature a *PT1* behavior with a time constant depending on the operating point. The proposed modifications to the *DoE* are also targeted on an improved steady state behavior or a high model quality next to the static optima. To achieve this a static nonlinear system followed by a *PT1* element with a time constant dependent nonlinear on the output of the static system. This allows to check easily the static accuracy of the dynamic model globally and close to the assumed optima. The *RadCos* function is used as static nonlinearity with input level $u \in [0, 1]^2$ [10]:

$$x = \cos\left(9\sqrt{u_1^2 + u_2^2} + 2\right) + 0.5 \cos(11u_1 + 2) \\ + 15((u_1 - 0.4)^2 + (u_2 - 0.4)^2).$$

The used *PT1* element is described by the following differential equation:

$$T\dot{y} + y - Kx = (2.4 \cos(10x + 4) - 0.5y + 4)\dot{y} + y - x$$

for which an anti derivative exists and which is discretised by $h = 0.5s$. An illustration of the scaled transfer function of the system is shown in Figure 5.

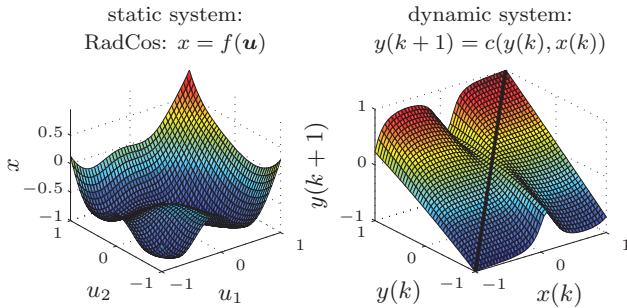


Figure 5: Composed dynamic test system

As seen on Figure 6 the Bayesian design (Bayesian) has the lowest mean value and the lowest scattering of the errors. The D-optimal design (Dopt) is the second best design and the Query by Committee (QBC) is the worst one. As validation data set the data sets which are not used for the training of the model committee are used for the transient model behavior.

3.2 Comparison of the designs on the basis of a static system

Ater the prove that the Bayesian design performs better than the other designs at the identification of dynamic systems the Bayesian design is now compared to the D-optimal

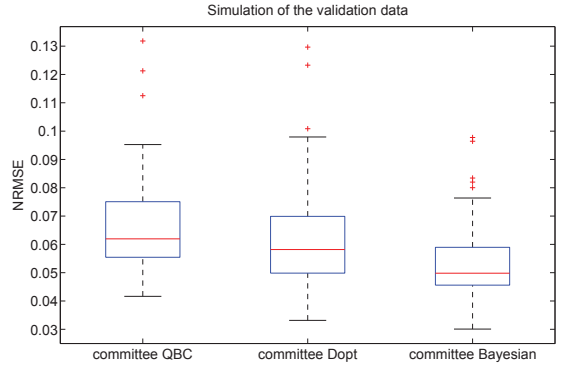


Figure 6: Comparison of the design criteria by the simulation errors.

one on the basis of a polynomial. Therefore, it is shown that the Bayesian design can not only be used with dynamic system but also with static systems. A polynomial of degree 6 (see Figure 7), which corresponds to the *RadCos* function before, is given and the order of the polynomial on which the D-optimal design is calculated is 7. 60 design points are used and the validation data are calculated on the basis of a Latin Hypercube distribution.

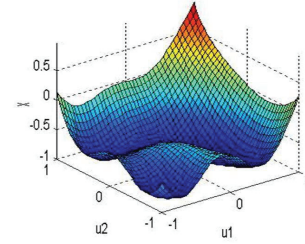


Figure 7: Polynomial of degree 6.

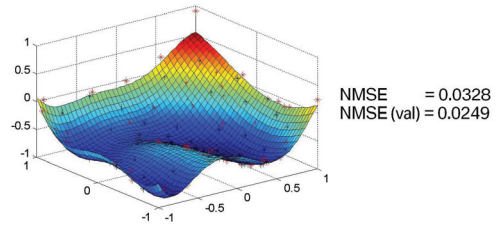


Figure 8: Approximation of the polynomial of degree 6 on the basis of a D-optimal design based on a polynomial of degree 7.

As seen on Figure 8 and 9 the Bayesian design leads to better results than the D-optimal one. As mentioned before you can see that the D-optimal design tends to weight the boundary area of the experimental space significantly and that the design is very dependent to the statistical model and the lack of the explicit allocation of the identified irregularities of the assumed model. The normalized mean square error (*NMSE*), which is the mean quadratic error between

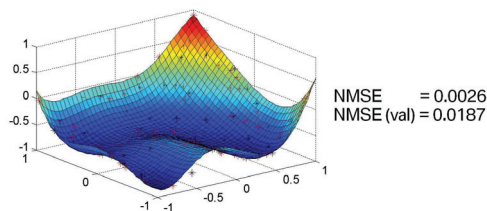


Figure 9: Approximation of the polynomial of degree 6 on the basis of a Bayesian design based on a polynomial of degree 7.

the measured output and the model output weighted with the variance of the measured output, is significantly lower with the Bayesian design both with the training data and with the validation data. The same result is shown if the order of the polynomial on which the D-optimal design is calculated is underestimated.

4. CONCLUSIONS

This contribution presents various methods for the Design of Experiments for nonlinear dynamic system identification. At first model-free designs, like the random design, the Latin Hypercube and the distance-based criteria, are introduced. Then model-based designs are discussed. The Query by Committee criterion is mentioned and the Fischer Information matrix is calculated, which plays an important role in the D-optimal design, which is then presented. The Fischer Information matrix is also used especially for the new Bayesian design for the nonlinear dynamic system identification, where central points or other control points are added to the D-optimal design to decide whether the design has uncertainties. Furthermore, it is demonstrated that the D-optimal design needs a few modifications to make the design less independent to the chosen model approach. The presented new Bayesian design allows to identify more accurate dynamic both analytical and simulation models under consideration of their later application. This helps to consider the transient engine behavior in model based optimization tasks. As shown in the subsequently simulation with dynamic systems like Neural Networks and static systems like polynomials as model approaches the Bayesian design provides better results as the D-optimal and the Query by Committee on both the validation data and the steady state error. For the future the design will be used in practice to achieve better results and therefore a reduction of emissions and consumption.

References

[1] The german automotive business brief, 2011.

[2] M. Deflorian, F. Klöpper, and J. Rückert. Online dynamic black box modelling and adaptive experiment design in combustion engine calibration. In *IFAC Symposium Advances in Automotive Control*, 2010.

[3] M. Deflorian and S. Zaglauer. Design of experiments for nonlinear dynamic system identification. In *IFAC 18th World Congress, Milano 2011*, 2011.

[4] W. DuMouchel and B. Jones. A simple bayesian modification of d-optimal designs to reduce dependence on an assumed model. *Technometrics*, 36, NO. 1:11, February 1994.

[5] M. E. Johnson, L. M. Moore, and D. Ylvisaker. Minimax and maximin distance designs. *Journal of statistical planning and inference*, 26(2):131–148, 1990.

[6] J. P. C. Kleijnen, S. Sanchez, T. W. Lucas, and T. M. Cioppa. A users guide to the brave new world of designing simulation experiments. *Inform Journal on Computing*, 17:263–289, 2005.

[7] J. R. Koehler and A. B. Owen. Computer experiments. *Handbook of Statistics*, 13:261–308, 1996.

[8] T. J. Mitchell. An algorithm for the construction of d-optimal experimental designs. *Technometrics*, 16(2):203–210, 1974.

[9] M. D. Morris and T. J. Mitchell. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43:381–402, 1995.

[10] J. Poland. *Modellgestützte und Evolutionäre Optimierungsverfahren für die Motorentwicklung*. PhD thesis, Fakultät für Informatik der Eberhard-Karls-Universität Tübingen, 2002.

[11] A. Schreiber and R. Isermann. Integrated methodology for stationary and dynamic measurement and modeling of combustion engines. In *Challenges in the field of net Powertrain Concepts*, 2009.

[12] H. Seung, M. Opper, and H. Sompolinsky. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 287–294. ACM New York, NY, USA, 1992.

[13] T. W. Simpson, D. K. J. Lin, and W. Chen. Sampling strategies for computer experiments: design and analysis. *International Journal of Reliability and Application*, 2:209–240, 2001.

[14] M. Stehlik and W. G. Müller. Fisher information in the design of computer simulation experiments. In *6th International Conference on Inverse Problems in Engineering: Theory and Practice Journal of Physics: Conference Series 135*, 2008.

[15] R. Vitali. *Response Surface Methods for High Dimensional Structural Design Problems*. PhD thesis, University of Florida, 2000.

[16] E. Walter and L. Pronzato. *Identification of Parametric Models from Experimental Data*. Springer-Verlag, London, 1997.

[17] M. Witczak. Toward the training of feed-forward neural networks with the D-optimum input sequence. *IEEE Transactions on neural networks*, 17(2):357–373, 2006.

[18] S. Zaglauer. *Motoroptimierung durch Modellkomitees*. Master's thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2010.