This document contains the **post-print pdf-version** of the refereed paper:

“Robustifying optimal experiment design for nonlinear, dynamic (bio)chemical systems”

by *Dries Telen, Dominique Vercammen, Filip Logist and Jan Van Impe*

which has been archived on the university repository Lirias (<https://lirias.kuleuven.be/>) of the KU Leuven.

The content is identical to the content of the published paper, but without the final typesetting by the publisher.

When referring to this work, please cite the full bibliographic info:

D. Telen, D. Vercammen, F. Logist, J. Van Impe (2014). Robustifying optimal experiment design for nonlinear, dynamic (bio)chemical systems. Computers and Chemical Engineering, 71, 415–425.

The journal and the original published paper can be found at:
<http://www.journals.elsevier.com/computers-and-chemical-engineering/>
<http://dx.doi.org/10.1016/j.compchemeng.2014.09.006>

The corresponding author can be contacted for additional info.

Conditions for open access are available at:
<http://www.sherpa.ac.uk/romeo/>

Robustifying optimal experiment design for nonlinear, dynamic (bio)chemical systems

D. Telen^a, D. Vercammen^a, F. Logist^{a,*}, J. Van Impe^a

^a*KU Leuven, Chemical Engineering Department, BioTeC & OPTEC,
W. de Croylaan 46, 3001 Leuven, Belgium*

Abstract

Dynamic experiments that yield as much information as possible are highly valuable for estimating parameters in nonlinear dynamic processes. Techniques for model-based optimal experiment design enable to systematically design such experiments. However, these experiments depend on the current best estimate of the parameters, which are not necessarily the true values. Consequently, in real experiments (i) the information content can be lower than predicted and (ii) state constraints can be violated. This paper presents a novel, computationally tractable formulation that enables the robustification of optimally designed experiments with respect to (i) information content and (ii) constraint satisfaction. To this end, the objective function is the expected value of a scalar function of the Fisher information matrix, which is efficiently computed using the sigma point method. This approach already has a robustifying effect. The sigma point method also enables the efficient computation of constraints' variance-covariance matrix, this can be exploited for further robustification.

Keywords: Optimal experiment design, Robust optimal control, Fisher information matrix, Sigma point, Parametric uncertainty

1. Introduction

In the last decades, optimal experiment design (OED) techniques have gained increasing attention in many (bio)chemical applications. This is due to the fact that it is often non trivial to identify parameters in dynamic systems which typically have a strong nonlinear nature (Cappuyns et al., 2007; Jaubertie et al., 2006; Schenkendorf et al., 2009; Telen et al., 2012, 2013, 2014). Especially, for cost intensive applications, it can be beneficial to design a control input in such a way that the experiment yields as much information as possible (see Franceschini and Macchietto (2008) for an overview).

An important challenge in OED is that the experiment has to be planned under incomplete information as the true parameter values are not known. This parametric uncertainty has some important consequences. First, a potential loss in information content is expected. So, experiments need to be *robust with respect to the information content* (Asprey and Macchietto, 2000). In Goodwin et al. (2007) and Körkel et al. (2004), it is suggested to formulate the robust experiment design formulation as a max-min optimization problem. Körkel and co-workers (Körkel et al., 2004) solved the inner optimization loop explicitly with a linear approximation. A

*Corresponding author

Email address: filip.logist@cit.kuleuven.be (J. Van Impe)

Bayesian approach can be found in Liepe et al. (2013). A different approach is to compute the expected value over the uncertain parameters of the scalar function of the Fisher information matrix. The idea of expected value optimization for OED objectives is already introduced in Pronzato and Walter (1985) and applied to a dynamic system for a Gaussian distribution of the parameters in Asprey and Macchietto (2002).

Second, an experiment can lead to violations of state constraints as the actual systems' behavior differs from the predicted one. So, there is the need to be *robust with respect to present state constraints*. In dynamic optimal control, such issues have been dealt with. Sometimes stochastic information about the uncertainties is available such that chance constraints can be formulated (Wendt et al., 2002; Srinivasan et al., 2003; Mitra, 2009; Galvanin et al., 2010; Recker et al., 2012). Here, the stochastic information can originate from a (previous) parameter identification procedure where typically confidence regions are generated along with estimates for the parameters (Franceschini and Macchietto (2008)).

In this paper, the expected value formulation for OED is exploited. Hence, the expected value of a scalar function of the Fisher information matrix needs to be computed. In practice this expected value is computationally challenging as it involves in theory the computation of an integral of the probability distribution of one of the scalar functions of the Fisher information matrix over the whole parameter space (Pronzato and Walter, 1985).

A first main contribution of this paper is that a computationally tractable formulation for expected value optimization of a nonlinear dynamic model is proposed. This formulation is based on the *sigma point method* (Julier and Uhlmann, 1996), which allows the approximation of the probability distribution (mean and variance-covariance matrix) through any nonlinear mapping. A second main contribution is to use this variance-covariance matrix for further robustifying the designed experiments in which the computed variance on the expected value can be taken directly into account.

The paper is structured as follows. In Section 2 the mathematical formulation of OED is introduced. The main contribution, i.e., the extension of the sigma point approach to the expected value computation and the robustification of state constraints is presented in Section 3. The difference between the traditional OED approach (Section 2) and the proposed robustified OED approach (Section 3) is also illustrated in Figure 1. Section 4 introduces the considered case studies, while the obtained results can be found in Section 5. Section 6 concludes the paper.

2. Optimal experiment design as a dynamic optimization problem

In this section, the formulation of the Fisher information matrix for nonlinear dynamic models is discussed first. Second, suitable choices for the objective function are described. In the third part the constraint formulation is discussed. In the fourth part, the formulation of OED as a dynamic optimization problem is explained. The section concludes with a discussion on how to solve dynamic optimization problems numerically.

2.1. Dynamic models

An adequate formulation for this paper of a dynamic, nonlinear model reads as follows:

$$\frac{dx(t)}{dt} = f(x(t), u(t), p, t) \quad \text{with} \quad x(0) = x_0, \quad (1)$$

$$r(t) = h(x(t)), \quad (2)$$

where $t \in [0, t_f]$ is the time, x are the state variables, u are the control variables and p are the parameters. The dynamic system equations are represented by f . The vector r indicates the system's output. The function $h(x(t))$ denotes the measurement function, which can be nonlinear.

In this paper *optimal experiment design for parameter estimation* is considered. In order to quantify the amount of information for parameter estimation in an experiment a suitable measure of the Fisher information matrix approach is employed (Walter and Pronzato, 1997), which is defined as:

$$F(t_f) = \int_0^{t_f} \frac{\partial x}{\partial p}(t)^\top C(t)^\top Q(t)^{-1} C(t) \frac{\partial x}{\partial p}(t) dt. \quad (3)$$

As the true values p are unknown, the Fisher information matrix is evaluated at the current best guess p^* . The Fisher information matrix combines information about the variance-covariance matrix of the output error measurements, $Q(t)^{-1}$ and the sensitivities of the states with respect to small variations in the model parameters, $\frac{\partial x}{\partial p}(t)$. The expression $C(t) = \frac{\partial h(x(t))}{\partial x}$ denotes the derivative of the measurement function $h(x(t))$ with respect to x . The measurement error, $\epsilon(t)$ is assumed to be additive to $h(x(t))$ and normally distributed with zero mean, and variance-covariance, $Q(t)$. The sensitivities are computed as the solution of the following equations:

$$\frac{d}{dt} \frac{\partial x}{\partial p}(t) = \frac{\partial f}{\partial x} \frac{\partial x}{\partial p}(t) + \frac{\partial f}{\partial p} \quad \text{with} \quad \frac{\partial x}{\partial p}(0) = \frac{\partial x_0}{\partial p}. \quad (4)$$

An interesting property of the Fisher information matrix is that under the assumption of unbiased estimators and uncorrelated Gaussian noise, the inverse of $F(t_f)$ approximates the lower bound of the parameter estimation variance-covariance matrix, i.e, the Cramér-Rao lower bound (Walter and Pronzato, 1997; Ljung, 1999). If a sampling decision is taken into account, the Fisher information matrix can be written as:

$$F(t_f) = \int_0^{t_f} w(t) \frac{\partial x}{\partial p}(t)^\top C(t)^\top Q(t)^{-1} C(t) \frac{\partial x}{\partial p}(t) dt, \quad (5)$$

where $w(t) \in \{0, 1\}$, is the decision variable which indicates whether or not a sample is taken. In this paper, this condition is taken into account in its relaxed form $w(t) \in [0, 1]$ (Sager, 2013). This avoids a mixed-integer optimization problem while allowing to get arbitrary close to the optimal solution due to the sum up rounding strategy (Sager, 2013).

2.2. Quantifying the information content

The Fisher information matrix $F(t_f)$ contains the information of a chosen experiment. To optimize this amount of information a scalar function of $F(t_f)$ has to be selected as the optimization problem's objective function. The objective function

is formulated as:

$$\Phi(F(t_f)) \quad (6)$$

where $\Phi(\cdot)$ is one of the criteria further described and $F(t_f)$ is the Fisher information matrix. Some widely used scalar functions (Walter and Pronzato (1997); Franceschini and Macchietto (2008)) are listed below:

- **A-criterion:** $\min[\text{trace}(F^{-1}(t_f))]$ A-optimal designs minimize the mean of the parameter estimation errors. The geometrical interpretation is the minimization of the enclosing frame around the joint confidence region.
- **D-criterion:** $\max[\det(F(t_f))]$ D-optimal designs minimize the geometric mean. Geometrically, this is minimizing the volume of the joint confidence region.
- **E-criterion:** $\max[\lambda_{\min}(F(t_f))]$ E-optimal designs aim at minimizing the largest parameter error. This corresponds to minimizing the length of the largest uncertainty axis of the joint confidence region.

A problem shared by all criteria is that the design of an experiment is dependent on the current best estimate of the parameters. A potential way to deal with this dependency is described in Körkel et al. (2004). It involves formulating a (conservative) max-min optimization problem, e.g., trying to maximize the worst determinant value of the Fisher information matrix. The inner optimization is approximated in a linear way leading to a challenging, but tractable optimization problem. A different approach for the robustification of experiment designs has been proposed in Pronzato and Walter (1985). It is assumed that the parameters p belong to a population with known/estimated statistics. As a measure of information content the *mathematical expectation* over this population is proposed:

$$E_p(\Phi(F(t_f))) \quad (7)$$

This means that the expected value of the scalar function of the Fisher information matrix over the parameter space needs to be computed. This can be done by solving a complex integral (Asprey and Macchietto, 2002) involving the Fisher information matrix and the parameter probability distribution. Note the difference with Equation (6) where only the current best estimates of the parameters are considered. The main difficulty of this formulation is how to compute the expected value in the context of dynamic process models. In the next Section, a computational tractable approach for the expected value OED formulation is presented.

2.3. Constraints

Processes in industry usually have to satisfy safety and operation constraints, e.g., temperatures must not exceed certain levels or feed rates can not exceed the pump limitations. Mathematically this is expressed as:

$$0 \geq c_p(x(t), u(t), t), \quad (8)$$

$$0 \geq c_t(x(t_f), t_f). \quad (9)$$

Here, the vector c_p indicates the path inequality constraints and c_t denotes the terminal inequality constraints on controls and states.

2.4. Dynamic optimization problem formulation

The complete *classic OED problem formulation* incorporating the required sensitivities and the Fisher information matrix is as follows:

$$\min_{u(\cdot), x(\cdot), \frac{\partial x}{\partial p}(\cdot), F(\cdot)} \Phi(F(t_f)) \quad (10)$$

subject to:

$$\frac{dx}{dt}(t) = f(x(t), u(t), p, t) \quad \text{with} \quad x(0) = x_0, \quad (11)$$

$$\frac{d}{dt} \frac{\partial x}{\partial p}(t) = \frac{\partial f}{\partial x} \frac{\partial x}{\partial p}(t) + \frac{\partial f}{\partial p} \quad \text{with} \quad \frac{\partial x}{\partial p}(0) = \frac{\partial x_0}{\partial p}, \quad (12)$$

$$\frac{d}{dt} F(t) = \frac{\partial x}{\partial p}(t)^\top C(t)^\top Q(t)^{-1} C(t) \frac{\partial x}{\partial p}(t) \quad \text{with} \quad F(0) = 0, \quad (13)$$

$$0 \geq c_p(x(t), u(t), t), \quad (14)$$

$$0 \geq c_t(x(t_f), t_f). \quad (15)$$

The first equation describes the actual system dynamics, the following two equations (12) and (13) are the required sensitivity equations and Fisher information matrix. Note that the symmetry in $F(t)$ can be exploited. Equations (14) and (15) denote the constraints. The left part of Figure 1 schematically summarizes the required ingredients for the traditional OED approach.

OED for a dynamic system with n_x state variables and n_p parameters requires in addition to the n_x states $n_x \cdot n_p$ sensitivities. Furthermore $\frac{n_p}{2} \cdot (n_p + 1)$ Fisher elements are needed. The total number of states for a single objective function evaluation is:

$$n_x + n_p \cdot n_x + n_p/2 \cdot (n_p + 1). \quad (16)$$

2.5. Dynamic optimization problem solution

OED is a special class of dynamic optimization problems. These optimization problems are infinite dimensional, since for every point t in the interval an optimal value has to be found. They can be solved by converting them to a finite Nonlinear Programming (NLP) problem by means of discretization. Two different approaches can be distinguished. A *sequential direct method* as *Single Shooting* Vassiliadis et al. (1994) discretizes the controls while the *simultaneous direct approaches* discretize both states and controls. Within these simultaneous approaches there exist: *Multiple Shooting* Bock and Plitt (1984) and *Orthogonal Collocation* Biegler (2007). In Single Shooting only the controls are discretized which results in general in a smaller NLP compared with the simultaneous approaches. Violations of state constraints can only be checked after integration and are not always straightforward to take into account. In Multiple Shooting also the states are discretized such that state constraints can be checked more easily in these discretization points. However, in between these points violations can still occur. In collocation schemes the states' discretization is even finer than in Multiple Shooting, meaning that state constraint violations are more unlikely to occur. The price to pay is, however, that a much larger NLP has to be solved and sparsity has to be taken into account. Nevertheless dedicated numerical integrators are not required as in the shooting approaches.

3. A computationally tractable formulation for robustified OED

In this section, the application of the sigma point approach to the expected value of OED-design criteria is discussed in view of making the experiment design

more robust. Subsequently, the approach on how to robustify even more state constraints using the sigma point method is presented. This section concludes with a description of the numerical implementation and the employed software.

3.1. The sigma point approach

The sigma point method is a method for the approximation of nonlinear transformations of probability distributions. It has been developed in Julier and Uhlmann (1996) and it is based on the intuition that it is easier to approximate a distribution with a fixed number of parameters, i.e., the *sigma points*, than to approximate an arbitrary nonlinear function. Assume the following nonlinear function:

$$z = g(y) , \quad (17)$$

in which the n -dimensional variable y has a specific distribution with a given expectation value $\bar{y}$ and variance-covariance matrix P_{yy} . The sigma points are chosen as:

$$\gamma_0 = \bar{y} , \quad (18)$$

$$\gamma_i = \bar{y} + \sqrt{(n + \kappa)P_{yy}}_i \text{ with } i = 1, \dots, n , \quad (19)$$

$$\gamma_i = \bar{y} - \sqrt{(n + \kappa)P_{yy}}_{i-n} \text{ with } i = n + 1, \dots, 2n , \quad (20)$$

which results in total in $2n+1$ sigma points. Here, $\sqrt{P_{yy}}_i$, denotes the i -th column of the matrix square root, which can be computed by, e.g., a Cholesky decomposition. These sigma points are subsequently evaluated by the nonlinear function:

$$\zeta_i = g(\gamma_i) \text{ with } i = 0, \dots, 2n . \quad (21)$$

The predicted mean can be computed as:

$$\bar{z} = \frac{1}{n + \kappa} \left(\kappa \zeta_0 + \frac{1}{2} \sum_{i=1}^{2n} \zeta_i \right) , \quad (22)$$

while the predicted variance-covariance matrix is approximated by:

$$P_{zz} = \frac{1}{n + \kappa} (\kappa (\zeta_0 - \bar{z})(\zeta_0 - \bar{z})^\top) + \frac{1}{n + \kappa} \left(\frac{1}{2} \sum_{i=1}^{2n} (\zeta_i - \bar{z})(\zeta_i - \bar{z})^\top \right) . \quad (23)$$

An answer to the question how to choose the parameter κ and a thorough theoretical analysis of the approximation errors can be found in Julier and Uhlmann (1996) ($\kappa=3-n$). Note also that how to choose these sigma points is not unique. Several methods, each with their advantages have been proposed, e.g., Julier (2002). In the next section, the extension to expected value optimization for OED is described.

3.2. Expected value computation for OED

The expected value of the first moment or mean of the distribution after a nonlinear transformation can be approximated through the sigma point method Julier and Uhlmann (1996). Assume the current best guess for the parameters p with a variance-covariance matrix V . This can be the result of a literature study or the result of a previous parameter estimation procedure. The nonlinear function $g(y)$ is in the optimal experiment design context the evaluation of the objective function $\Phi(F)$ for given initial conditions and controls. The parameters p are the n -dimensional variable y in Equation (17). If n_p is the number of uncertain parameters, then the

objective function has to be evaluated $2n_p + 1$ times. The evaluations differ only in the values for the parameters p :

$$\pi_0 = p, \quad (24)$$

$$\pi_i = p + \sqrt{(n_p + \kappa)V_i} \text{ with } i = 1, \dots, n_p, \quad (25)$$

$$\pi_i = p - \sqrt{(n_p + \kappa)V_{i-n_p}} \text{ with } i = n_p + 1, \dots, 2n_p. \quad (26)$$

Here, $\sqrt{V_i}$, denotes the i -th column of the matrix square root which can be computed by, e.g., a Cholesky decomposition. The expected value for the objective function $E(\Phi(F(t_f))) = \bar{\Phi}(F(t_f))$ can be computed as:

$$\bar{\Phi}(F(t_f)) = \frac{1}{n_p + \kappa} \left(\kappa \Phi(F(\pi_0)) + \frac{1}{2} \sum_{i=1}^{2n_p} \Phi(F(\pi_i)) \right), \quad (27)$$

in which $\Phi(F(\pi_i))$ is the scalar function of the Fisher information matrix $F(t_f)$ evaluated in the i -th sigma point, i.e., where the parameter values are $p = \pi_i$, while F_i denotes the corresponding Fisher information matrix. Note that the sigma point approach is not used to quantify the information content as has been proposed in, e.g., Heine et al. (2008); Schenkendorf et al. (2009), but is used to quantify the uncertainty in the Fisher information matrix or in the constraints.

So, the complete problem formulation for *expected value optimization in OED* using the sigma point approach is:

$$\min_{u(\cdot), x_i(\cdot), \frac{\partial x_i}{\partial p}(\cdot), F_i(\cdot)} E(\Phi(F(t_f))) \quad \text{with } i = 0, \dots, 2n_p \quad (28)$$

subject to:

$$\frac{dx_i}{dt}(t) = f(x_i(t), u(t), \pi_i, t) \quad \text{with } x_i(0) = x_0 \quad (29)$$

$$\frac{d}{dt} \frac{\partial x_i}{\partial p}(t) = \frac{\partial f}{\partial x} \frac{\partial x_i}{\partial p}(t) + \frac{\partial f}{\partial p} \quad \text{with } \frac{\partial x_i}{\partial p}(0) = \frac{\partial x_0}{\partial p}, \quad (30)$$

$$\frac{d}{dt} F_i(t) = \frac{\partial x_i}{\partial p}^\top(t) C(t)^\top Q(t)^{-1} C(t) \frac{\partial x_i}{\partial p}(t) \quad \text{with } F_i(0) = 0, \quad (31)$$

$$0 \geq c_p(x_i(t), u(t), t), \quad (32)$$

$$0 \geq c_t(x_i(t_f), t_f). \quad (33)$$

The states resulting from the sigma point approach are implemented in parallel with the original system such that in total for a system with n_x original states and n_p parameters:

$$\left(n_x + n_p \cdot n_x + \frac{n_p}{2} \cdot (n_p + 1) \right) \cdot (2n_p + 1), \quad (34)$$

states need to be computed in the dynamic optimization, but leads to an experiment in which the uncertainty of the parameters is taken into account in the OED (see also Figure 1). In the approach proposed in this work, the expected value can be computed without knowledge of the actual distribution (Julier and Uhlmann, 1996). Furthermore, in the current approach the computation of a complex integral over the parameter space is avoided (Asprey and Macchietto, 2002).

3.3. Further robustifying constraints

An additional advantage of the proposed technique is that in these sigma points the constraints are also required to be satisfied. This means that in a single optimization procedure, state constraints for $2n_p + 1$ parameter sets are required to be satisfied, yielding an OED which is more robust as uncertainty information for the different parameter values are taken into account. Furthermore, besides the computation of the expected Fisher information matrix, it is also possible to compute the expected constraint evolution, $E(c_p(t)) = \bar{c}_p(t)$, and variance-covariance matrix $P_{c_p c_p}$ with negligible additional computational effort:

$$\bar{c}_p(t) = \frac{1}{n_p + \kappa} \left(\kappa c_p(t, \pi_0) + \frac{1}{2} \sum_{i=1}^{2n_p} c_p(t, \pi_i) \right), \quad (35)$$

$$P_{c_p c_p}(t) = \frac{\kappa}{n_p + \kappa} ((c_p(t, \pi_0) - \bar{c}_p(t))(c_p(t, \pi_0) - \bar{c}_p(t))^T) + \left(\frac{1}{2(n_p + \kappa)} \sum_{i=1}^{2n_p} (c_p(t, \pi_i) - \bar{c}_p(t))(c_p(t, \pi_i) - \bar{c}_p(t))^T \right), \quad (36)$$

Here, $c_p(t, \pi_i) = c_{p,i}$ denotes the constraint profile resulting from the evaluation of the dynamic system in the i -th sigma point, i.e., $p = \pi_i$. The computed uncertainty of the variance-covariance matrix of the expected value can be taken into account to further robustify the optimization with respect to the constraints by computing a so-called backoff value (Srinivasan et al., 2003; Galvanin et al., 2010). The expected value is already robust as the state constraints have to be satisfied by all the sigma points. The uncertainty as measured by the variance-covariance matrix on the expected value can subsequently be taken into account by (Srinivasan et al., 2003; Galvanin et al., 2010):

$$0 \geq \bar{c}_{p,j}(t) + \alpha \sqrt{P_{c_p c_p, jj}(t)} \quad \text{with } j = 1, \dots, n_{c_p}, \quad (37)$$

with α a chosen parameter. Here, $\bar{c}_{p,j}(t)$ denotes the j -th element of the vector $\bar{c}_p(t)$ with dimension n_{c_p} and $P_{c_p c_p, jj}(t)$ is the corresponding diagonal element of the variance-covariance matrix. This constraint formulation ensures that the expected value in a certain probability region is satisfied. If the expected value is assumed to be normally distributed, α can be chosen to specific % significance levels. A similar approach for robust dynamic optimization can be found in (Srinivasan et al., 2003; Galvanin et al., 2010; Logist et al., 2011; Houska et al., 2012; Recker et al., 2012; Ostrovsky et al., 2013) in which, the variance-covariance matrix is approximated using a first order approximation. So, the *robustified expected value optimization problem* is:

$$\min_{u(\cdot), x_i(\cdot), \frac{\partial x_i}{\partial p}(\cdot), F_i(\cdot)} E(\Phi(F)) \quad \text{with } i = 0, \dots, 2n_p \quad (38)$$

subject to:

$$\frac{dx_i}{dt}(t) = f(x_i(t), u(t), \pi_i, t) \quad \text{with} \quad x_i(0) = x_0, \quad (39)$$

$$\frac{d}{dt} \frac{\partial x_i}{\partial p}(t) = \frac{\partial f}{\partial x} \frac{\partial x_i}{\partial p}(t) + \frac{\partial f}{\partial p} \quad \text{with} \quad \frac{\partial x_i}{\partial p}(0) = \frac{\partial x_0}{\partial p}, \quad (40)$$

$$\frac{d}{dt} F_i(t) = \frac{\partial x_i}{\partial p}^\top(t) C(t)^\top Q(t)^{-1} C(t) \frac{\partial x_i}{\partial p}(t) \quad \text{with} \quad F_i(0) = 0, \quad (41)$$

$$0 \geq c_p(x_i(t), u(t), t), \quad (42)$$

$$0 \geq c_t(x_i(t_f), t_f), \quad (43)$$

$$0 \geq \bar{c}_{p,j}(t) + \alpha \sqrt{P_{c_p c_p, jj}(t)} \quad \text{with} \quad j = 1, \dots, n_{c_p}. \quad (44)$$

The number of required states is the same as in the previous subsection, the difference however is that additional constraints as expressed by Equation (44) are added (see Figure 1).

Remark. A Monte-Carlo simulation can be performed after the experiment design to investigate the robust character of the constraints (Schenkendorf et al., 2009) by sampling different parameter sets. If the a-priori desired confidence level related with the chosen α , is not reached, this α can be increased. The sigma point approach serves as a predictor step while the Monte-Carlo simulations are a corrector step. This allows an iterative procedure where the use of computationally expensive Monte-Carlo simulations are avoided in the optimization procedure.

4. Case studies

This section presents the considered case studies. A first biochemical case study is a well-mixed fed-batch bioreactor. The second case study is the chemical Williams-Otto reactor. Note that for the two case studies two parameters are considered. This is for visualization purposes, the proposed technique is not restricted to a two parameter case.

4.1. A fed-batch bioreactor

This case study considers a well-mixed fed-batch bioreactor (Cappuyns et al., 2007). The dynamic model equations are given by:

$$\frac{dC_s}{dt} = -\sigma C_x + \frac{u}{\nu} C_{s,\text{in}} - \frac{u}{\nu} C_s, \quad (45)$$

$$\frac{dC_x}{dt} = \mu C_x - \frac{u}{\nu} C_x, \quad (46)$$

$$\frac{d\nu}{dt} = u, \quad (47)$$

with initial conditions: $x(0) = (3.0, 0.25, 7.0)^\top$, and constraints: $0 \leq u(t) \leq 1$ L/h, in which C_s [g/L] is the concentration of limiting substrate, C_x [g/L] the biomass concentration and ν [L] the bioreactor volume. u [L/h] is the volumetric feed rate, containing a substrate concentration $C_{s,\text{in}}$. The specific growth rate studied in this case is the monotonic *Monod* type. The mathematical expression for the growth rate is: $\mu = \mu_{\max} \frac{C_s}{K_s + C_s}$. The substrate consumption rate is given by the linear law: $\sigma = \mu/Y_{X|S} + m$, where $Y_{X|S}$ is the yield and m the maintenance factor. The end time t_f is set to 15 h. Both the substrate and biomass concentration are assumed to be measurable, $h(t) = (C_s, C_x)$. The used measurement error variances are $\sigma_{C_s}^2 = 0.10 \cdot 10^{-2}$ g²/L² and $\sigma_{C_x}^2 = 6.25 \cdot 10^{-4}$ g²/L². The number of control

intervals is set to 20. The number of states to be calculated is for this case study 45 as $n_x = 2$ and $n_p = 2$. Note that state ν is not influenced by the uncertain parameters and thus omitted from the uncertainty propagation.

The first uncertain parameter is $\mu_{\max}$, the maximum specific growth rate. The second parameter under consideration is K_s , the half-saturation constant. The parameter values used in the OED procedure are 0.421 and 0.439, respectively. The initially assumed variance-covariance matrix is:

$$V = \begin{pmatrix} 0.10^2 & 0 \\ 0 & 0.10^2 \end{pmatrix}. \quad (48)$$

For the remaining constants and parameter values, the reader is referred to Capuyns et al. (2007).

4.2. The Williams-Otto reactor

The case study described in this section is the Williams-Otto fed-batch reactor (Hannemann and Marquardt, 2010; Logist et al., 2011). The following reactions take place in the reactor: $A + B \rightarrow C$, $B + C \rightarrow P + E$ and $C + P \rightarrow G$. Reactant A is initially present while reactant B is added continuously. During the exothermic reactions products P and E as well as G are formed. The heat produced in the reactions is removed by a cooling jacket. The dynamic model equations are:

$$\frac{dx_A}{dt} = \frac{x_A u_1}{1000R} - k_1 \eta_1 x_A x_B, \quad (49)$$

$$\frac{dx_B}{dt} = \frac{(1 - x_B)u_1}{1000R} + k_1 \eta_1 x_A x_B - k_2 \eta_2 x_B x_C, \quad (50)$$

$$\frac{dx_C}{dt} = \frac{-x_C u_1}{1000R} + k_7 \eta_1 x_A x_B - k_3 \eta_2 x_B x_C - k_6 \eta_3 x_C x_P, \quad (51)$$

$$\frac{dx_P}{dt} = \frac{-x_P u_1}{1000R} + k_2 \eta_2 x_B x_C - k_4 \eta_3 x_C x_P, \quad (52)$$

$$\frac{dx_E}{dt} = \frac{-x_E u_1}{1000R} + k_3 \eta_2 x_B x_C, \quad (53)$$

$$\frac{dx_G}{dt} = \frac{-x_G u_1}{1000R} + k_5 \eta_3 x_C x_P, \quad (54)$$

$$\frac{dT}{dt} = \frac{(T_F - T)u_1}{1000R} + k_8 \eta_1 x_A x_B + k_9 \eta_2 x_B x_C + k_{10} \eta_3 x_C x_P - l_1(T - 1000u_2), \quad (55)$$

$$\frac{dR}{dt} = \frac{u_1}{1000}, \quad (56)$$

with initial conditions:

$x(0) = (1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 65, 2.0)^\top$, and constraints:

$$\begin{aligned} 60 \leq T(t) \leq 90 \quad , \quad 0 \leq u_1(t) \leq 5.784 \quad , \\ 0.02 \leq u_2(t) \leq 0.1 \quad , \quad R(t_f) \leq 5 \quad , \end{aligned}$$

in which x_i with $i \in \{A, B, C, P, E, G\}$ are dimensionless concentrations, T is the reactor temperature and R the volume. The time t [h] is the independent variable. The Arrhenius dependencies are $\eta_1 = \exp(\frac{-6666.7}{T+273.15})$, $\eta_2 = \exp(\frac{-8333.3}{T+273.15})$ and $\eta_3 = \exp(\frac{-11111.0}{T+273.15})$. The controls of the system are u_1 , i.e., the feed rate of B and u_2 , the scaled jacket fluid temperature. The end time is fixed to 1000 h. Only the temperature profile is assumed to be measurable. The number of control intervals

is set to 50. The number of states to be calculated, is for this case study 121, see Equation (34) as $n_x = 7$ and $n_p = 2$ with one state to describe the reactor volume which does not depend on the uncertain parameters.

Two uncertain parameters are considered as well for the OED of the Williams-Otto reactor. The first is k_1 , which is a kinetic coefficient of the reaction, while l_1 is the heat transfer coefficient. The employed parameter values are $1.6599 \cdot 10^6$ and $2.435 \cdot 10^{-4}$, respectively. The initially assumed variance-covariance matrix is:

$$V = \begin{pmatrix} (1.6599 \cdot 10^5)^2 & 0 \\ 0 & (2.435 \cdot 10^{-5})^2 \end{pmatrix}. \quad (57)$$

For the remaining constants and parameter values, the reader is referred to Hanne-mann and Marquardt (2010); Logist et al. (2011).

4.3. Numerical implementation

OED for nonlinear dynamic models is a subclass of dynamic optimization which quickly leads to a high number of states. These problems are solved in this work by discretizing the differential states and the controls via a direct collocation scheme (Biegler, 1984, 2010). For the states, fourth order Lagrange polynomials are chosen while for the controls, a piecewise constant approach is selected. The resulting nonlinear program is solved with CasADi, a software tool for automatic differentiation and dynamic optimization (Andersson et al., 2012). The required gradients, jacobians and Hessians are automatically generated in CasADi due to the symbolic implementation. These are then used to solve the NLP using the interior point optimization routine IPOPT (Wächter and Biegler, 2006). The tolerance of the optimization routine is set to 10^{-6} .

5. Results

In this section, the obtained numerical results are discussed. The first part discusses the results for the fed-batch bioreactor. The second part presents the results of the Williams-Otto reactor.

5.1. A fed-batch bioreactor

For the fed-batch bioreactor two different designs are considered: the E-design and the expected E-design (EE-design). Given the collocation scheme, this case study results in 3680 discretized state and 20 control variables to be determined by the NLP-solver subject to 3999 equality and 301 inequality constraints. In addition, the number of continuous sampling hours available for sampling is restricted in these designs to 3 hours. Remark that it is assumed that measurements are taken at the same time for all states. As the maintenance factor m is assumed to be zero, different realizations for the parameters do not result in invalid experiments and no state constraints have to be satisfied. So, the focus on the *robustness* in this case study is *with respect to the information content*. The obtained state and control profiles for a total of 3 h sampling are depicted in Figure 2. In addition, the start of the sampling for a period of 0.75 h is each time denoted by the squares or diamonds. Remarkably, the obtained profiles are quite similar. The main difference between the control action of the two designs is a shift of a similar feeding phase towards the end of the experiment in the EE-design. The moment when measurements are to be taken coincides for both designs with this feeding phase. Note that the expected state confidence region is almost reduced to zero in Figures 2(c) and 2(d). This is possible if the state at that time point is the same for all the sigma point realizations, see Figures 2(a) and 2(b). From these Figures 2(a) and 2(b), it is observed

Table 1: First, second and third quartile of the parameter quality measure for the E- and EE-design.

	E-design	EE-design
Q_1	0.85	0.77
Q_2	2.65	1.59
Q_3	21.8	5.58

that the largest difference with the nominal profile originates from sigma points π_1 and π_3 , which is the influence of $\mu_{\max}$. If the objective function is evaluated using the current best estimates of the parameters with the control input originating from the E-design and the EE-design, a decrease of the information content (measured as the predicted minimal eigenvalue of the Fisher information matrix) of 7.9 % for the 4 period sampling case is observed. This illustrates that the robustification of the design comes at a cost, i.e., a reduction of information content.

To investigate how the objective function changes when the actual system has parameters which are different from the ones for which the experiment has been designed, 200 different parameter sets for each of the designs are sampled from the assumed parameter distribution. The previously found E-design and EE-design are applied with 4 sampling periods accounting for 3 sampling hours. For each of the resulting 200 profiles the predicted objective function value is computed first. Second, for each of the 200 state profiles a noise realization based on the assumed measurement error variance is sampled. This results in 200 noisy datasets for both designs. For these sets, a parameter identification procedure is performed and the information content is quantified by computing the empirically observed smallest eigenvalue of the resulting Fisher information matrix. The resulting objective function scores are depicted in Figure 3. A strong dependency of the objective function is observed in relationship with $\mu_{\max}$, which is also hinted by the dependency of the variance estimate of the previous paragraph. For the E-design, a decrease of the objective function is observed when $\mu_{\max}$ differs from the nominal value where the sampled $\mu_{\max}$'s are larger than the mean value. In the region where $\mu_{\max}$ is smaller than the mean value, a sharp decrease in information content is observed compared to the EE-designs. Note that the objective function values obtained after the parameter identification procedure are in agreement with the predicted values. Remark, however, that the decrease in information content when $\mu_{\max}$ is larger than its nominal value is more abrupt than predicted. The EE-design leads to a more flat objective surface and outperforms the E-design in regions far from the nominal parameter values. The quality is also evaluated at the level of the resulting estimates. The parameter estimation quality measure employed is the mean relative parameter error (MPE) [%]: $MPE = \frac{100}{n_p} \sum_i^{n_p} \frac{|p_{\text{identified},i} - p_{\text{sampled},i}|}{p_{\text{sampled},i}}$ [%]. The distribution of the parameter quality index is illustrated by the 3 quartiles in Table 1. It is observed that the two first quartiles are more or less close for the two designs, note however that the EE-design scores consistently better than the E-design. The difference is more evident in the third quartile, where it is 21.8% versus 5.58% illustrating that the EE-design is indeed more robust for the parameter identification. To examine the influence of a single parameter on the parameter estimation accuracy expressed in MPE, Figure 4 is displayed. It depicts the cross section of the 2D surface fitted through the different realizations for both the E- and EE-design. One parameter is kept at its initially assumed value while the other parameter is varied up to 50%. From Figure 4 it is observed that $\mu_{\max}$ influences the MPE more than K_s . Furthermore, in the neighborhood of the initially assumed value the E-design is better than the robust case, while further from the initially assumed value, the

robust design outperforms the E-design. From Figure 4 it can be inferred that if $\mu_{\max}$ is accurate, the value of K_s hardly influences the parameter accuracy for both designs.

5.2. The Williams-Otto reactor

For the Williams-Otto reactor, the D-design, the expected D-design (ED-design) and the robustified expected D-design (robust ED-design) are considered. Given the employed collocation scheme, this case study results in at most 24200 discretized state and 50 control variables to be determined by the NLP subject to 24200 equality constraints and 300 inequality constraints. Note the presence of a strict state constraint with respect to reactor temperature. So, to ensure the approximate 95 % confidence region of the expected temperature evolution satisfies the state constraint, the following equations are added:

$$E(T(t)) + 2\sqrt{P_{TT}(t)} \leq 90, \quad (58)$$

$$E(T(t)) - 2\sqrt{P_{TT}(t)} \geq 60. \quad (59)$$

The constituting temperature profiles for the expected temperature profile are illustrated in Figure 5(a) for the ED-design result. The obtained temperature and expected temperature profiles are depicted in Figures 5(b) and 5(c). From Figure 5(a), it is clear that in one of sigma point temperature profiles the present state constraint is active but the predicted 95 % confidence bound almost satisfies the state constraint for the expected D-design while the robust ED-design is within the constraints. Furthermore, in the nominal D-design the state constraint is active at several different times. In Figures 5(e) and 5(f) the differences in control actions between the three designs is displayed. The initial feeding stops earlier in the ED-design/robust ED-design compared with the D-design. In addition, the heating of the system stops earlier in the ED-design/robust ED-design compared with the D-design. The objective function value, evaluated with the current best guess of the parameter (as one would do if the classic FIM approach is used) is 17.5 % lower in the robust ED-design than in the D-design while in the robust ED-design, this value is 18.6 % lower. So, a price for the robustification is observed, i.e., the information content as measured by the FIM is lower in the best guess of the parameters in the case of the expected design formulations compared with the classic OED approach.

As the nominal D-design results in several time instances in an active state constraint, it is to be expected that changes in the parameters will result in a violation of this constraint. To investigate the *robustness with respect to this state constraint*, 500 different parameter sets are drawn given the distribution in the previous section. These simulations result for the ED-design in 473 designs which remain within the state constraints and for the robust ED-design 477 designs (≥ 95 %) which are feasible while the D-design results in only 211 feasible experiments. Parameter sets leading to valid experiments are depicted in Figure 6. For values of the heat coefficient below the nominal value almost no feasible experiment can be found for the D-design experiment. The ED-design and robust ED-design, however result in this region in valid experiments (up to 95.4 %). The resulting mean reactor temperature profiles for the three designs and their empirical 95 % confidence bounds are illustrated in Figure 5(d). The latter figure indicates that changes in the heat coefficient can indeed lead to higher temperatures than allowed in the case of the D-design control action. If, however, the experiment is designed using the expected value approach, inherently robustifying as each of the sigma point profiles are required to satisfy the present constraints. Also note that the predicted confidence bounds of the ED-design are in agreement with the empirical observations, illustrating the potential of the proposed methodology. In addition, the robustified

ED-design gives rise to an experiment with empirical confidence bounds, satisfying the state constraints, as depicted in Figure 5(d).

6. Conclusion

The problem addressed in this paper is that in OED for nonlinear dynamic models, experiments are planned under incomplete information due to the presence of uncertain parameters. This can lead to a loss in the predicted information content and to experiments that violate present state constraints. In this paper, a computationally tractable methodology is proposed to use the expected value OED design criteria formulation, based on the sigma point method. The expected value is computationally challenging as it involves the computation of an integral over the parameter domain. So, the sigma point method is employed such that a tractable computation of the expected value of the Fisher information matrix is obtained and the computation of a complex integral is avoided. Moreover, the computation of the expected value has an inherent robust character, as for each of the sigma points, all present state constraints have to be satisfied. Furthermore, the sigma point method also allows the computation of the variance-covariance matrix of the expected value. This variance-covariance matrix can be used for a further robustification of the designed experiments in which the computed variance on the expected value can be taken directly into account. The presented methodology is illustrated on two different case studies with different OED objective functions.

ACKNOWLEDGMENTS

DT and DV each hold an IWT-PhD grant. JVI holds the chair Safety Engineering sponsored by essenscia. The research was supported by: PFV/10/002 (OPTEC), OT/10/035, FWO KAN2013 1.5.189.13, FWO-G.0930.13 and BelSPO: IAP VII/19 (DYSCO).

References

- Andersson, J., Akesson, J., Diehl, M., 2012. CasADi - a symbolic package for automatic differentiation and optimal control. In: Proceedings of the 6th International Conference on Automatic Differentiation.
- Asprey, S., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. Computers and Chemical Engineering 24, 1261 – 1267.
- Asprey, S., Macchietto, S., 2002. Designing robust optimal dynamic experiments. Journal of Process Control 12 (4), 545 – 556.
- Biegler, L., 1984. Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. Computers and Chemical Engineering 8, 243–248.
- Biegler, L., 2007. An overview of simultaneous strategies for dynamic optimization. Chemical Engineering and Processing: Process Intensification 46, 1043–1053.
- Biegler, L., 2010. Nonlinear Programming: Concepts, Algorithms, and Applications to Chemical Processes. MOS-SIAM.
- Bock, H., Plitt, K., 1984. A multiple shooting algorithm for direct solution of optimal control problems. In: Proceedings of the 9th IFAC world congress, Budapest. Pergamon Press, pp. 243–247.

- Cappuyns, A., Bernaerts, K., Smets, I., Ona, O., Prinsen, E., Vanderleyden, J., Van Impe, J., 2007. Optimal fed batch experiment design for estimation of monod kinetics of *Azospirillum brasilense*: From theory to practice. *Biotechnology Progress* 23 (5), 1074–1081.
- Franceschini, G., Macchietto, S., 2008. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science* 63, 4846–4872.
- Galvanin, F., Barolo, M., Bezzi, F., Macchietto, S., 2010. A backoff strategy for model-based experiment design under parametric uncertainty. *AIChE Journal* 56, 2088–2102.
- Goodwin, G., Rojas, C., Welsh, J., Feuer, A., 2007. Robust optimal experiment design for system identification. *Automatica* 43, 993–1008.
- Hannemann, R., Marquardt, W., 2010. Continuous and discrete adjoints for the hessian of the lagrangian in shooting algorithms for dynamic optimization. *SIAM Journal on Scientific Computing* 31, 4675–4695.
- Heine, T., Kawohl, M., King, R., 2008. Derivative-free optimal experimental design. *Chemical Engineering Science* 63, 4873–4880.
- Houska, B., Logist, F., Van Impe, J., Diehl, M., 2012. Robust optimization of nonlinear dynamic systems with application to a jacketed tubular reactor. *Journal of Process Control* 22, 1152–1160.
- Jaubertie, C., Denis-Vidal, L., Coton, P., Joly-Blanchard, G., 2006. An optimal input design procedure. *Automatica* 42, 881 – 884.
- Julier, S., 2002. The scaled unscented transformation. In: *American Control Conference, 2002. Proceedings of the 2002*. Vol. 6. IEEE, pp. 4555–4559.
- Julier, S., Uhlmann, J., 1996. A general method for approximating nonlinear transformations of probability distributions. Tech. rep., Department of Engineering Science, University of Oxford.
- Körkel, S., Kostina, E., Bock, H., Schlöder, J., 2004. Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optimization Methods and Software Journal* 19 (3-4), 327–338.
- Liepe, J., Filippi, S., Komorowski, M., Stumpf, M., 2013. Maximizing the information content of experiments in systems biology. *PLoS Computational Biology* 9, e1002888.
- Ljung, L., 1999. *System Identification: Theory for the User*. Prentice Hall.
- Logist, F., Houska, B., Diehl, M., Van Impe, J., 2011. Robust multi-objective optimal control of uncertain (bio)chemical processes. *Chemical Engineering Science* 66 (20), 4670 – 4682.
- Mitra, K., 2009. Multiobjective optimization of an industrial grinding operation under uncertainty. *Chemical Engineering Science* 64 (23), 5043–5056.
- Ostrovsky, G., Ziyatdinov, N., Lapteva, T., 2013. Optimal design of chemical processes with chance constraints. *Computers and Chemical Engineering* 59, 74 – 88.
- Pronzato, L., Walter, E., 1985. Robust experiment design via stochastic approximation. *Mathematical Biosciences* 75, 103–120.

- Recker, S., Kühn, P., Diehl, M., Bock, H. G., 2012. Sigmapoint approach for robust optimization of nonlinear dynamic systems. In: Proceeding of SIMULTECH 2012 . pp. 199–207.
- Sager, S., 2013. Sampling decisions in optimum experimental design in the light of Pontryagin's maximum principle. *SIAM Journal on Control and Optimization* 51, 3181–3207.
- Schenkendorf, R., Kremling, A., Mangold, M., 2009. Optimal experimental design with the sigma point method. *IET Systems Biology* 3, 10–23.
- Srinivasan, B., Bonvin, D., Visser, E., Palanki, S., 2003. Dynamic optimization of batch processes II. Role of measurements in handling uncertainty. *Computers and Chemical Engineering* 27, 27–44.
- Telen, D., Logist, F., Van Derlinden, E., Van Impe, J., 2012. Optimal experiment design for dynamic bioprocesses: a multi-objective approach. *Chemical Engineering Science* 78, 82–97.
- Telen, D., Houska, B., Logist, F., Van Derlinden, E., Diehl, M., and Van Impe, J., 2013. Optimal experiment design under process noise using Riccati differential equations. *Journal of Process Control* 23, 613–629.
- Telen, D., Logist, F., Quirynen, R., Houska, B., Diehl, M., and Van Impe, J., 2014. Optimal experiment design for nonlinear dynamic (bio)chemical systems using sequential semidefinite programming. *AIChE Journal* 60, 1728–1739.
- Vassiliadis, V., Sargent, R., Pantelides, C., 1994. Solution of a class of multistage dynamic optimization problems. 2. Problems with path constraints. *Industrial and Engineering Chemistry Research* 33, 2123–2133.
- Wächter, A., Biegler, L., 2006. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical Programming* 106 (1), 25–57.
- Walter, E., Pronzato, L., 1997. Identification of Parametric Models from Experimental Data. Springer, Paris.
- Wendt, M., Li, P., Wozny, G., 2002. Nonlinear chance constrained process optimization under uncertainty. *Industrial and Engineering Chemistry Research* 41, 3621–3629.

Figure 1: Illustration of the traditional OED formulation (left) and the proposed robustified OED formulation (right). The traditional approach is explained in Section 2, while the new approach based on the sigma points method (SP) is presented in Section 3.

Figure 2: State, expected state evolution with confidence bounds (CB) and control action of the E-design and EE-design for the bioreactor case with 4 sampling periods.

Figure 3: Illustration of the predicted (left) and empirically observed (right) objective function surface for the E- and EE-design.

Figure 4: The MPE in function of individual parameter deviations for the E- and EE-design.

Figure 5: Expected reactor temperature evolution and control actions of the D-design, ED-design and robust ED-design for the Williams-Otto reactor case.

Figure 6: Illustration of the projected resulting objective function surface for the three designs: D-design, ED-design and the robust ED-design. The black ellipsoids denote the 66 % and 95 % joint parameter confidence region.