Robust Optimal Experiment Design: A Multi-Objective Approach

Dries Telen ∗ Filip Logist ∗ Eva Van Derlinden ∗ Jan F. Van Impe ∗

Abstract: Optimal Experiment Design (OED) is an indispensable tool in order to reduce the amount of labour and cost intensive experiments in the modelling phase. The unknown parameters are often non-linearly present in the dynamic process models. This means that the Fisher Information Matrix also depends on the current guess for the parameters. In the early stage of the modelling phase these estimates are often highly uncertain. So designing an optimal experiment without taking this uncertainty into account is troublesome. In order to obtain an informative experiment, a robust optimisation approach is necessary. In recent work a formulation using an implicit weighted sum approach is proposed where the objective function is split in a nominal optimal experiment design part and a robust counterpart. This weighted sum has well known drawbacks in a Multi-Objective Optimisation approach. In this work these objectives are studied using advanced methods like the Normal Boundary Intersection and the Normalised Normal Constraint. In this way, the experimenter gets an overview of the different experiments possible. Furthermore, in past work the necessary third order derivatives are approximated using a finite difference approach. The results in this work are obtained using exact third order and fourth order derivatives by exploiting the symbolic and automatic derivation methods implemented in the ACADO-toolkit.

Keywords: Optimal Experiment Design, Robustness, Multi-objective Optimisation, Parameter Estimation, (Bio)Chemical Processes

1. INTRODUCTION

Nonlinear differential equation models are valuable tools for the analysis, design, optimisation and control of nonlinear dynamic processes. Before these models can be used in practice, they often need to calibrated (Walter and Pronzato, 1997). More precisely, model parameters have to be estimated such that the model predicts experimental data as accurately as possible. However, as experiments are often cost and/or labour intensive, Optimal Experiment Design techniques can help to systematically develop experiments which contain the most information and from which the parameters can be estimated with the highest accuracy (Franceschini and Macchietto, 2008). Applications in (bio)chemical engineering have been presented in, e.g., Balsa-Canto et al. (2008, 2010); Van Derlinden et al. (2010).

When a parameter appears non-linearly in the model, it is well-known that an optimally designed experiment depends on the current estimate value. In addition, in real experiments measurement noise can not be avoided. Both aspects introduce uncertainty in the optimisation problem and, hence, require a robust optimisation approach (Körkel et al., 2004). In robust optimisation, a so-called robust counterpart formulation has to be solved, i.e., a worst-case scenario needs to be optimised. These kinds of formulations typically yield a hard to solve min-max problem. In the case of robust Optimal Experiment Design a solution to these problems is shown in Körkel et al. (2004).

Finally, it has to be noted that the amount of uncertainty assumed in the process or the emphasis put on obtaining robust solutions induces a trade-off with respect to the original Optimal Experiment Design objective. In the approach of Körkel et al. (2004) an implicit weighted sum approach is used. In order to provide the experimenter with an overview of the trade-off between the two parts multi-objective optimisation techniques (Logist et al., 2010, 2011) are exploited. Furthermore, instead of approximating the third order derivatives with finite differences, these are calculated exactly by exploiting the symbolic and automatic differentiation feature available in the ACADO-toolkit (Houska et al., 2011).

This paper is structured as follows: first Section 2 introduces the formulation of a dynamic optimisation problem. Section 3 reviews the Optimal Experiment Design framework together with the robust counterpart, while Section 4 introduces the techniques of multi-objective optimisation. Section 5 presents the case study and the obtained results. The main conclusion is formulated in Section 6.

2. OPTIMAL CONTROL FORMULATION

Optimal Experiment Design for nonlinear dynamic systems is a particular class of optimal control problems. The
general formulation of an optimal control problem is as follows:

$$\min_{\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}, t} J$$ (1)

subject to:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}, t) \quad t \in [0, t_f]$$ (2)

$$0 = \mathbf{b}_c(\mathbf{x}(0), \mathbf{p})$$ (3)

$$0 \geq \mathbf{c}_p(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}, t)$$ (4)

where \( \mathbf{x} \) are the state variables, \( \mathbf{u} \) the time-varying control inputs and \( \mathbf{p} \) the model parameters. The vector \( \mathbf{f} \) represents the dynamic system equations (on the interval \( t \in [0, t_f] \)) with initial conditions given by the vector \( \mathbf{b}_c \). The vector \( \mathbf{c}_p \) indicates path inequality constraints on the states and controls. Furthermore, the vector \( \mathbf{y} \) is introduced, which contains the measured outputs. These are usually a subset of the state variables \( \mathbf{x} \). The decision variables can be grouped together as \( \mathbf{z} = [\mathbf{x}(t), \mathbf{u}(t), t_f] \).

Note that \( \mathbf{p} \) is no longer a degree of freedom, but considered known. All vectors \( \mathbf{z} \) that satisfy Equations (2) to (4) form the set of feasible solutions \( S \).

3. OPTIMAL EXPERIMENT DESIGN FOR NONLINEAR DYNAMIC SYSTEMS

In this section the Optimal Experiment Design framework is explained. The first part focusses on the Optimal Experiment Design formulation, the second part explains the robust formulation as presented in Körkel et al. (2004).

3.1 Optimal Experiment Design

In Optimal Experiment Design for parameter estimation, some scalar function of the Fisher Information Matrix is used as the objective function. This matrix is defined as:

$$\mathbf{F} = \int_0^{t_f} \left( \left[ \frac{\partial \mathbf{y}(p,t)}{\partial \mathbf{p}} \right]^T Q \left( \frac{\partial \mathbf{y}(p,t)}{\partial \mathbf{p}} \right) \right) d\mathbf{p}$$ (5)

The true values \( \mathbf{p}^* \) are unknown so the Fisher depends on the current best estimate. Two parts constitute the Fisher information matrix: the inverse of the measurement error variance-covariance matrix, \( Q \) and the sensitivities of the output model to small variations in the model parameters, \( \frac{\partial \mathbf{y}(p,t)}{\partial \mathbf{p}} \). The latter can be found as the solution to:

$$\frac{d}{dt} \frac{\partial \mathbf{y}(p,t)}{\partial \mathbf{p}} = \frac{\partial \mathbf{f}(p,t)}{\partial \mathbf{y}} \frac{\partial \mathbf{y}(p,t)}{\partial \mathbf{p}} + \frac{\partial \mathbf{f}(p,t)}{\partial \mathbf{p}}$$ (6)

An interesting property of the Fisher Information Matrix is that under the assumption of unbiased estimators and uncorrelated Gaussian noise, the inverse of \( \mathbf{F} \) approximates the lower bound of the parameter estimation variance covariance matrix, which is the Cramér-Rao lower bound. The most widely used scalar functions are (Pukelsheim, 1993; Walter and Pronzato, 1997; Franceschini and Macchietto, 2008):

**A-criterion:** \( \min [\text{trace}(\mathbf{F}^{-1})] \) A-optimal designs minimise the mean of the parameter estimation errors. The geometrical interpretation is the minimisation of the enclosing frame around the joint confidence region. In order to decrease the computational effort, the problem is reformulated as maximising the trace of \( \mathbf{F} \).

**D-criterion:** \( \max[\det(\mathbf{F})] \) D-optimal designs minimise the geometric mean. Geometrically, this is minimising the volume of the joint confidence region.

**E-criterion:** \( \max[\lambda_{\text{min}}(\mathbf{F})] \) E-optimal designs aim at minimising the largest parameter error. Uncertainty regarding other parameters is thus neglected. This corresponds to minimising the length of the largest uncertainty axis of the joint confidence region.

**Modified E-criterion:** \( \min \left( \frac{\det(\mathbf{F})}{\lambda_{\text{min}}(\mathbf{F})} \right) \) The modified E-criterion (ME-criterion) minimises the condition number of the Fisher information matrix. A priori the optimal value of one is known. This corresponds to circular joint confidence regions.

3.2 Robust Optimal Experiment Design

In nonlinear dynamic models the elements of the Fisher matrix depend on the current estimate of the parameters. It is however possible that there is a large uncertainty with respect to these parameter values. This explains the need for robust optimal experiment design. This means that the information content is much less dependent on the current estimate of the parameters. Parameters are assumed to be normally distributed:

$$\mathbf{p} \sim N(\mathbf{p}_0, \Sigma)$$ (7)

where \( \mathbf{p}_0 \) is the mean value and \( \Sigma \) the variance. The formulation used as robust optimal experiment design in Körkel et al. (2004) is a min-max optimisation problem of the following form:

$$\min_{\mathbf{x}(t), \mathbf{u}(t), t_f} \max_{||p-p_0||_2, \gamma, \lambda \leq \gamma} \Phi(\mathbf{F}(p))$$ (8)

where \( \gamma \) is a confidence quantile. In Körkel et al. (2004), this problem is solved by applying a Taylor expansion with respect to the parameters of the objective function, which leads to the following robust experimental design optimisation problem:

$$\min_{\mathbf{x}(t), \mathbf{u}(t), t_f} \Phi(\mathbf{F}(p)) + \gamma \left\| \frac{d}{dp} \Phi(\mathbf{F}(p)) \right\|_{2, \Sigma}$$ (9)

It is clear that there are two parts in this objective function: the nominal Optimal Experiment Design objective, \( \Phi(\mathbf{F}(p)) \) and the robust objective, \( \left\| \frac{d}{dp} \Phi(\mathbf{F}(p)) \right\|_{2, \Sigma} \). The latter objective aims at having a cost surface which is flat with respect to the unknown parameters. An interesting thing about the formulation above is that this can be viewed as a multi-objective optimisation problem consisting of two different, possibly conflicting, objectives. Using some value for the quantile boils down to weighing the different objectives. In order to study the trade-off between the two objectives in a systematic way a multi-objective approach can be used.

A numerically challenging aspect of the robust part is that there are second-order derivatives with respect to the original model formulation in the objective function. In order to calculate the second-order derivatives the approach described in Balsa-Canto et al. (2001, 2004) is used. The problem, however, remains that in order to optimise...
the resulting optimal control problem, third and in case of an exact Hessian calculation fourth order derivatives are necessary. To this extent the ACADO toolkit (Houska et al., 2011) is used. One of the advantages of the ACADO toolkit with respect to this robust optimal experiment design problem is that models are expressed symbolically. This allows an easy derivation of the necessary equation to formulate the robust problem. Using its symbolic and automatic differentiation features, the necessary third and even fourth order derivatives can be obtained.

4. MULTI-OBJECTIVE OPTIMISATION

The general formulation of a dynamic optimisation problem involving multiple objectives can be described as:

\[
\min_{x(t), u(t), p, t} \{ J_1(z), \ldots, J_m(z) \} \tag{10}
\]

subject to Equations (2) - (4). As concept for optimality in multi-objective optimisation, Pareto optimality is used (see, e.g., Miettinen (1999)). A solution is called Pareto optimal if there exists no other feasible solution that improves at least one objective function without worsening another. The decision variables \( z \) belong to the feasible set \( S \), defined previously, and the vector of all individual cost functions is defined as \( J = [J_1(z), \ldots, J_m(z)]^T \).

Methods for generating Pareto sets can be classified into two classes: (i) methods converting the MOO problem into a series of parametric single objective optimisation (SOO) problems (e.g., weighted sum, \ldots), and (ii) methods tackling directly the MOO problem (e.g., stochastic evolutionary algorithms (Deb, 2001)). However, since methods of the latter class require a repeated evaluation of the objectives (and, thus, also of the underlying models), they can become time consuming for the class of systems under study. Therefore, only deterministic techniques from the first class will be studied, because they use fast, gradient-based techniques.

4.1 Weighted Sum (WS)

The (convex) weighted sum is in practice the most widely used technique for combining different objectives. The resulting parametric SOO problem is the following:

\[
\min_{x \in S} J_{ws} = \sum_{i=1}^{m} w_i J_i(z) \tag{11}
\]

with \( w_i \geq 0 \) and \( \sum_{i=1}^{m} w_i = 1 \). By consistently varying the weights \( w_i \) an approximation of the Pareto set is obtained. However, despite its simplicity, the weighted sum approach has several intrinsic drawbacks (Das and Dennis, 1997). A uniform distribution of the weights does not necessarily results in an even spread on the Pareto front and points in non-convex parts of the Pareto set cannot be obtained.

4.2 Normal Boundary Intersection (NBI)

This method has been proposed by Das and Dennis (1998) to mitigate the above mentioned drawbacks of the WS. NBI tackles the MOO problem from a geometrically intuitive viewpoint. It first builds a plane in the cost space \( J_p \) which contains all convex combinations of the individual minima, i.e., the convex hull of individual minima (CHIM), and then constructs (quasi-)normal lines to this plane. The rationale behind the method is that the intersection between the (quasi-)normal from any point \( J_p \) on the CHIM, and the boundary of the feasible cost space closest to the origin is expected to be Pareto optimal. Hereto, the MOO objective problem is reformulated as to maximise the distance \( \lambda \) from a point \( J_p \) on the CHIM along the quasi-normal through this point, without violating the original constraints. Technically, this requirement of lying on the quasi-normal introduces additional equality constraints, resulting in the following formulation:

\[
\max_{\lambda} \lambda \quad \Phi - \lambda \Phi e = J(z) - J^* \tag{12}
\]

subject to:

\[
\Phi w - \lambda \Phi e = J(z) - J^* \tag{13}
\]

where \( \Phi \) is the \( m \times m \) pay-off matrix in which the \( i \)-th column is \( J(z_i^*) - J^* \), with \( z_i^* \) being the minimiser of the \( i \)-th objective \( J_i \) and \( J^* \) being the utopia point, which contains the minima of the individual objectives \( J_i(z_j^*) \). \( w \) is a vector of weights such that \( \sum_{i=1}^{m} w_i = 1 \) with \( w_i \geq 0 \), and \( e \) is a vector containing all ones. Now, \( \Phi w J \) describes a point in the CHIM and \( -\Phi e \) defines the (quasi-)normal to the CHIM pointing towards the origin. When the points on the CHIM are selected with an equal spread (via a uniform distribution of \( w \)), also an equal spread on the Pareto frontier in the objective space is obtained.

4.3 Normalised Normal Constraint (NNC)

NNC as introduced by Messac et al. (2003), employs similar ideas as NBI, but combines them with features of the \( \varepsilon \)-constraint method (Haines et al., 1971). This \( \varepsilon \)-constraint method minimises the single most important objective function \( J_k \), while the \( m - 1 \) other objective functions are added as inequality constraints \( J_i \leq \varepsilon_i \), which are interpreted as hyperplanes reducing the feasible criterion space. After normalisation of the objectives, NNC also first constructs a plane through all individual minima (called here, the utopia hyperplane). Then NNC minimises a selected (normalised) objective \( J_k \), given the original constraints, and while additionally reducing the feasible space by adding \( m - 1 \) hyperplanes through a selected point \( J_p \) in the utopia plane. These hyperplanes are chosen perpendicular to each of the \( m - 1 \) utopia plane vectors, which join the individual minimum \( J(z_i^*) \) corresponding to the selected objective \( J_k \), with all other individual minima \( J(z_i^*) \). Hence, this approach leads to an additional set of inequality constraints:

\[
\min_{z \in S} J_k \tag{14}
\]

subject to:

\[
(J(z_k^*) - J(z_i^*))^T (J(z) - J_p) \leq 0 \quad i = 1 \ldots m, i \neq k. \tag{15}
\]

As in NBI, evenly distributed points on the utopia plane \( J_p \) can be selected by a uniform variation of a vector \( w \), which also ensures an even spread on the Pareto set.

5. APPLICATION TO A PREDICTIVE MICROBIAL GROWTH MODEL

This section introduces first the case study from the field of predictive microbiology. In the second part the optimisation results are shown and discussed.
5.1 Predictive microbial model

In this case study an optimal experiment for parameters of the Cardinal Temperature Model with Inflection (Rosso et al., 1993) are designed. This is a secondary model to the model of Baranyi and Roberts (Baranyi and Roberts, 1994). This latter model describes the cell density as a function of time whereas the former incorporates the dependency of the specific growth rate on temperature. The model equations of the Baranyi and Roberts model are:

\[ \frac{dn}{dt} = \frac{Q}{Q + 1} \mu_{\text{max}}(T(t))[1 - \exp(n - n_{\text{max}})] \]  
\[ \frac{dQ}{dt} = \mu_{\text{max}}(T(t))Q \]

in which \( n \) [ln(CFU/ml)] is the natural logarithm of the cell density, \( n_{\text{max}} \) the maximum value for \( n \), \( Q \) [-] the physiological state of the cells. In the current Optimal Experiment Design the state \( Q \) is omitted (Van Derlinden et al., 2010). This because the duration of the microbial lag phase determined by the prior and actual experimental conditions, which is modelled through \( Q \), cannot be predicted. The model equations thus reduce to a logistic growth model for the design of the experiments. The temperature dependency described by the Cardinal Temperature Model with Inflection is given by:

\[ \mu_{\text{max}} = \mu_{\text{opt}} \gamma(T) \]  
\[ \gamma(T) = \frac{(T - T_{\text{min}})^2(T - T_{\text{max}})}{(T_{\text{opt}} - T_{\text{min}})(T_{\text{opt}} - T_{\text{max}})(T_{\text{opt}} + T_{\text{min}} - 2T)\} \]

with \( \gamma(T) \) as expressed in Table 1. In this model four parameters have to be estimated: \( T_{\text{min}}, T_{\text{opt}}, T_{\text{max}} \) and \( \mu_{\text{opt}} \). The Cardinal Temperature Model with Inflection is illustrated in Figure 1. It depicts the practical interpretation of the four parameters. The values of the parameters for are shown in Table 2. In this case study the parameters for *Escherichia coli K12* are considered. The initial condition is \( n(0) = 7 \text{ln(CFU/ml)} \). The end time is fixed to 38 h. For model validity reasons the dynamic temperature profiles are constrained to:

\[ 15^\circ C \leq T(t) \leq 45^\circ C \]  
\[ -5^\circ C/h \leq \Delta T(t) / \Delta t \leq 5^\circ C/h \]

Temperatures below \( T_{\text{min}} \) or above \( T_{\text{max}} \) result in a maximum specific growth rate, \( \mu_{\text{max}} \) set to zero. The simplest approach is to estimate the four parameters from one single experiment. In practice however the four parameters are divided in six two-parameter combinations. More details on the case study can be found in Van Derlinden et al. (2010).

5.2 Resulting Optimal Experiments

In this case study the F-criterion with respect to the two parameter problem: \( T_{\text{min}} \) and \( T_{\text{max}} \), is considered as an example to illustrate the approach. In the Optimal Experiment Design the following variance-covariance is assumed regarding the current estimate of the parameter values:

\[ \Sigma = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix} \]

In order to solve the dynamic multi-objective optimisation problem, the ACADO-Multi-Objective is used (Loth et al., 2010). The underlying dynamic optimisation problem is solved using a Single Shooting approach (Sargent and Sullivan, 1978). The integrator used is Runge-Kutta45. The integrator tolerance and the KKT-tolerance of the SQP-optimiser are both set to \( 10^{-6} \). The Pareto front is obtained in 77.53 seconds and the total number of SQP iterations is 111. The decision variables is a dynamic piecewise linear temperature profile split in 19 control intervals. From Figure 2 it is clear that if one would use different weights and thus different values for \( \gamma \), these would lead to a single point in the Pareto front. So even changing the weights results in the same experiment. In order to obtain a complete overview of the different to robustify the optimal experiment design the two advanced methods are more insightful. This also validates the idea of viewing the robust counterpart as an additional objective. In the Pareto fronts of the NBI/NNC a sharp trade-off between the two objectives is observed. One of the risks of this approach however is that the optimal experiment design is over-robustified. To explain this notion Figure 3 is displayed. These depict the nominal objective function value with respect to changes in the parameter values. The upper surface is the surface corresponding to the nominal optimal experiment. It illustrates a high dependency on the current values of the parameters. The lower surface is obtained when the robust objective is forced to its lowest value. This results in an almost flat surface, however the amount of information is lower than the lowest value in the cost surface of the nominal
optimal experiment design. When the trade-off experiment (indicated in the Pareto front) however is used, a more flat surface than in the nominal case is obtained. The lowest possible value of this surface is higher than the one of the nominal design. The surface is also less curved than the nominal design, the price to pay is the fact that the value for the objective function is slightly lower than in the nominal case.

When the result of the WS approach is studied in detail, it is clear that the trade-off design suffers from being too robust. It has already less curvature than the nominal case, the nominal objective value is on the other hand lower. The maximal value in the objective surface is lower than the minimal value in the nominal optimal experiment case. This illustrates that using the WS approach yields a too robust design in which too much information is lost.

The resulting evolution of the state and the applied temperature profiles for the different experiments obtained using the NBI/NNC approach are depicted in Figure 4. Figure 4 shows that the most robust design is non-informative. Already after eight hours the population reaches its maximum cell population numbers. As such, no information on the microbial dynamics as a function of temperature can be obtained beyond this point. In the other two experiments, growth is possible up to 38 hours

6. CONCLUSION

In the current paper a robust formulation of the optimal experiment design is studied. Instead of viewing the formulation as a single objective, a multi-objective optimisation approach is adopted. This is due to the fact that the previous formulation is in essence a weighted sum approach. This multi-objective optimisation approach is tackled using the efficient, deterministic optimisation procedures: Normal Boundary Intersection and Normalised Normal Constraint.

In the case study it is illustrated that using these advanced methods a robust optimal experiment design can be obtained whereas the weighted sum resulted in a single trade-off solution which has a lower value for its robustness measure. This experiment results in a local objective surface which is lower than the nominal optimal experiment design. The NBI/NNC approach results in several possible trade-off experiments which have a less curved objective surface but at the expense of a lower nominal information value. If the focus is put too much on the robust formulation, this results in experiments which suffer from the same problem as the WS approach demonstrates, i.e., a flatter objective surface but lying below the nominal objective surface.
Fig. 4. Corresponding state evolution (upper) and temperature profile (lower) for the selected points of the Pareto front.

An interesting extension to the current formulation is to take second order information of the objective cost surface into account. This results however in a problem formulation which requires third order derivative information in the objective and requires fifth order information if the problem is solved in an exact Hessian approach. In future work the current approach will also be compared to the robust dynamic optimisation approach discussed in Logist et al. (2011) or Houksa (2011).

ACKNOWLEDGEMENTS

Work supported in parts by Projects PFV/10/002, OPTEC (Centre-of-Excellence Optimisation in Engineering) and KP/09/005 SCORES-4CHEM knowledge platform of the Katholieke Universiteit Leuven and by the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian Federal Science Policy Office. Dries Telen has a Ph.D grant of the institute for the Promotion of Innovation through science and Technology in Flanders (IWT-Vlaanderen). Jan Van Impe holds the chair Safety Engineering sponsored by the Belgian Chemistry and life sciences federation essencia.

REFERENCES


