Experimental Design Tools for Ordinary and Algebraic Differential Equations

K. Schittkowski[1]

Dept. of Computer Science, University of Bayreuth, D - 95440 Bayreuth, Germany

The purpose of this paper is to present practical tools to facilitate the interpretation of parameter estimation results and to optimize experimental designs, where the underlying dynamical model consists of systems of ordinary or algebraic differential equations. We present a heuristic procedure to compute significance levels of model parameters, and to allow successive elimination of redundant ones. For computing optimal experimental designs, we choose the A-criterion to evaluate the performance of the system, i.e., the identifiability of model parameters to be computed after getting the experimental data. Pseudo-weights are introduced and treated as design variables, to reduce the number of experiments and to figure out those time values at which experiments should be taken. A couple of practically relevant case studies from chemical engineering are included, which have been investigated before by other authors.

Keywords: parameter estimation, data fitting, identifiability, experimental design, confidence intervals, Fisher information matrix, least squares optimization, ordinary differential equation, differential algebraic equations

1 Introduction

Today there exists a large variety of statistical methods to analyze data and models especially for nonlinear regression or parameter estimation, see, e.g., the books of Bard [4], Beck and Arnold [7], Draper and Smith [13], Gallant [16], Ratkowsky [25], Seber [34], Seber and Wild [35], or Ross [26]. We are interested in designing experimental studies, for example in chemical or biological engineering, where a mathematical model is available in form of a system of ordinary or algebraic differential equations. The key idea is to estimate an unknown parameter vector $p = (p_1, \ldots, p_n)^T$ of a mathematical model that describes a real life situation, by minimizing the distance of some known experimental data from theoretically predicted values of a model function at certain time values. Thus, also model parameters that cannot be measured directly, can be identified by a least squares fit and analyzed subsequently in a quantitative way.

In mathematical notation, we want to solve a least squares problem of the form

$$p \in \mathbb{R}^n : \quad \begin{array}{l} \min \sum_{i=1}^{l} (h(p, t_i) - y_i)^2 \\ p_l \le p \le p_u \quad , \end{array} \tag{1}$$

[1] klaus.schittkowski@uni-bayreuth.de, http://www.klaus-schittkowski.de/

where $h(p,t) := \overline{h}(p, y(p,t), t)$ is a fitting function depending on the unknown parameter vector p, the time t, and the solution $y(p,t)$ of an underlying dynamical system of the form

$$\begin{array}{rcll} \dot{y}_1 & = & F_1(p,y,t) & , \; y_1(0) = y_1^0(p) \;\; , \\ & & \cdots & \\ \dot{y}_m & = & F_m(p,y,t) & , \; y_m(0) = y_m^0(p) \;\; . \end{array} \tag{2}$$

Without loss of generality, we assume that the initial time is zero.

Alternatively, the fitting criterion $h(p,t) := \overline{h}(p, y(p,t), z(p,t), t)$ may depend on m_d differentiable $y(p,t)$ and m_a additional algebraic variables $z(p,t)$. The dynamical system is given in the form of a system of differential algebraic equations

$$\begin{array}{rcll} \dot{y}_1 & = & F_1(p,y,z,t) & , \; y_1(0) = y_1^0(p) \;\; , \\ & & \cdots & \\ \dot{y}_{m_d} & = & F_{m_d}(p,y,z,t) & , \; y_{m_d}(0) = y_{m_d}^0(p) \;\; , \\ 0 & = & G_1(p,y,z,t) & , \; z_1(0) = z_1^0(p) \;\; , \\ & & \cdots & \\ 0 & = & G_{m_a}(p,y,z,t) & , \; z_{m_a}(0) = z_{m_a}^0(p) \;\; . \end{array} \tag{3}$$

The right-hand side of the differential equations and the initial values may depend on one or more of the model parameters to be estimated. Now $y(p,t)$ and $z(p,t)$ are solution vectors of a joint system of $m_d + m_a$ differential and algebraic equations (DAE). The system is called an index-1-problem or an index-1-DAE, if the algebraic equations can be solved subject to z, i.e., if the matrix

$$\nabla_z G(p,y,z,t) \tag{4}$$

possesses full rank, $G(p,y,z,t) = (G_1(p,y,z,t), \ldots, G_{m_a}(p,y,z,t))$.

Whenever we try to identify parameters of a dynamical model, the key question is whether it is possible at all to identify them and whether the computed parameters are uniquely determined. Proceeding from an arbitrary distribution of experimental time values t_i, $i = 1, \ldots, l$, and a model function $h(p,t)$, we are interested in the question whether the parameter p is uniquely determined, i.e., whether $h(p,t_i) = h(\hat{p},t_i)$ implies $p = \hat{p}$. If this property holds for arbitrary distributions of time values and for almost all $p \in I\!R^n$, we say that the model is *structurally globally identifiable* or *s.g.i.* in abbreviated form. If the same property is true only for all $\hat{p}$ in a neighborhood of p, we call the model *structurally locally identifiable* or *s.l.i.*, respectively.

There exist many methods to prove whether a model is s.g.i or not, see Walter [41], Walter and Pronzato [42], or Godfrey and DiStefano [17]. In most cases, these approaches are restricted to special model classes, for example to linear time-invariant differential equations. Ben-Zvi, McLellan, and McAuley [8] present a generalized Markov parameter approach for differential algebraic systems, and Tayakout-Fayolle, Jolimaitre, and Jallut [36] identified parameters in chromatography models.

Besides successive linearization, a more general method is applicable at least in some simple situations, for instance with polynomial, rational, or exponential

terms at the right-hand side of a differential equation, where we are able to analyze high-order derivatives 'by hand'. To outline the basic idea, we proceed from an initial value problem of the form (2) or (3), respectively, and a scalar fitting criterion $h(p,t)$. Assuming that $h(p,y,t)$ and $F(p,y,t)$ or $F(p,y,z,t)$ and $G(p,y,z,t)$, respectively, are infinitely often continuously differentiable with respect to y and t, we try to find a power series for $h(p,t)$ at $t=0$ in the form

$$h(p,t) = \sum_{k=0}^{\infty} \frac{1}{k!} a_k(p) t^k \quad .$$

Here $a_k(p)$ denotes the k-th one-sided derivative of $h(p,t)$ at $t=0$, i.e., the Taylor coefficient

$$a_k(p) = \lim_{t \searrow 0} \frac{d^k}{dt^k} h(p,t)$$

for $k = 0, 1, \ldots$. Thus, a sufficient condition for s.g.i. is that $a_k(p) = a_k(\hat{p})$ for $p, \hat{p} \in I\!R^n$ implies $p = \hat{p}$, see also Pohjanpalo [24] or Walter and Pronzato [42]. The identification of chemical engineering models is also studied by Gorskii and Spivak [20].

If, on the other hand, it is known in advance that a model is neither s.g.i. nor s.l.i., then it is much more difficult to find out which parameters are redundant or correlated and cannot be estimated subject to a given standard deviation of the experimental errors. One possible approach is investigated in this paper in more detail. Proceeding from a statistical analysis, a so-called *top-down classification* is proposed by Schneider, Posten, and Munack [32], see also Majer [22]. The idea is to find certain clusters of parameters by considering the absolutely largest coefficient of the eigenvector subject to the smallest eigenvalue of the Fisher information matrix, and to eliminate successively corresponding parameters, see Majer [22]. However, the approach is heuristic and requires additional statistical information about error distributions. It is possible that more experimental data and measurement sets must be supplied to stop the elimination process earlier, i.e., to identify more parameters.

Numerous other procedures have been investigated in the literature, where many of them are pure heuristic ones applicable only in special situations. Otarod and Happe [23], for example, compare correlations between different data sets. The underlying question how to identify relevant parameters, is also discussed in a more general setting how to select an appropriate model. This is discussed by Bortz and Nelson [10] in the context of HIV infection dynamics based on maximum likelihood formulations. Maximum likelihood estimators are also applied by Gorfine et al. [19] for testing statistical hypotheses in B Lymphocyte models. Generalized sensitivity functions are proposed by Thomaseth and Cobelli [37] for identifying parameters of input-output experiments.

Besides the question whether individual parameters can be identified or not, a more general problem is to investigate the identifiability of unknown coefficient functions of model equations. Particularly in case of partial differential equations, one is also interested in computing certain functions depending on the time or spatial variable by formulating an inverse problem.

However, inverse problems are often ill-posed. When trying to identify a coefficient function by measured output values, the following three difficulties occur:

1. A solution does not exist for a given set of parameters.
2. A solution is not unique.
3. The solution does not depend continuously on the unknown coefficient functions.

Otherwise, we say that the problem is well-posed in the sense of Hadamard, see Baumeister [6], DuChateau [14], or Banks and Kunisch [3] for more details.

The first question is how to get suitable confidence intervals for the estimated parameters. This is one of the main investigations when analyzing the output of a data fitting run. A short outline is given in Section 2, where we provide the basic analysis for the subsequent sections and introduce in particular the Fisher information matrix. The key idea is to investigate the confidence region subject to a given significance level, an ellipsoid which is typically approximated by a surrounding box, the confidence intervals. Whereas small interval lengths can be interpreted as well-identifiable parameters, larger intervals could be due to degenerated ellipsoids. For these and related modeling and simulation techniques, see also the books of Walter and Pronzato [42] and of van den Bosch and van der Klauw [39].

In many practical situations, dynamical models contain too many parameters which are difficult to estimate simultaneously, i.e., are overdetermined. An important question is how to detect the relative significance of parameters and how to eliminate redundant ones based on a given experimental design. A heuristic approach going back to Schneider, Posten, and Munack [32] is presented in Section 3, which is computationally attractive and easy to implement. The idea is to analyze eigenvalues and eigenvectors of the covariance matrix. The absolutely largest coefficient of the eigenvector belonging to the largest eigenvalue is eliminated and marks a less significant parameter. The procedure is repeated until some criteria are satisfied, e.g., reaching a certain significance tolerance. The result is a serial order of parameters according to their relevance, and which helps to decide which parameters could be eliminated or whether additional experiments should be performed. A case study is presented in Section 4, a model for an isothermal batch reactor in form of a system of differential algebraic equations.

So far we proceeded from a given experimental design and tried to compute model parameters by a least squares fit. However, the initial design might not be the best one and the question is how to improve or even optimize it. Possible design parameters are time dependent input feeds, initial concentrations or temperatures. The goal is now to construct a suitable performance criterion depending on design parameters, additional constraints as far as necessary, and to solve the resulting nonlinear optimization problem to get better experimental designs.

Since the confidence intervals mentioned above are mainly determined by the diagonal elements of the covariance matrix, a possible objective function for computing optimal designs is the trace of this matrix, see Section 5. A couple of alternative criteria are available, by which the size and the structure of the ellipsoid are measured, for example based on the eigenvalues of the covariance matrix. These methods have been discussed in the literature before, see e.g. Winer, Brown and Michels [43], Ryan [28], Rudolph and Herrendörfer [27], Baltes et al. [1], or Lohmann et al. [21], often based on automatic differentiation for computing mixed partial derivatives of the right-hand side of the dynamical system. However, analytical derivatives are often not available. Instead, we apply simple forward differences and show that efficient computation of derivatives is possible if carefully implemented.

There remains the question how to locate experimental time values. Especially in case of time expensive experiments, it is highly desirable to minimize their number and to conduct experiments only within relevant time intervals. Thus, we apply the same strategy outlined before, but add artificial weight factors to the observations at a predefined, relatively dense grid specified in advance. These weights are considered then as design parameters. A particular advantage is that derivatives subject to weights are obtained without additional computational efforts.

To show that our approach is a stable and efficient procedure, two case studies are included in Section 6 and Section 7. The first one is a microbial growth model, see Banga et al. [2], which consists of a small system of only three differential equations, two model parameters and design parameters in form of initial concentrations and an input feed. However, one of the model parameters is extremely difficult to estimate and the authors decided to apply a stochastic search method. The other example is intensively investigated by Bauer et al. [5], the reaction of urethane. The model consists of three differential and three algebraic equations, and becomes more complex because of additional nonlinear equality and inequality constraints.

The techniques described so far do not depend on any special structure of the mathematical model. The only assumption is smoothness of the model functions, i.e., the functions should be twice continuously differentiable subject to the model and design parameters. All examples with practical background consist of ordinary differential or differential algebraic equations, since the imbedded solution process generates additional numerical noise making numerical results and conclusions more realistic.

The algorithms are implemented in Fortran as part of the interactive data fitting system EASY-FIT, see Schittkowski [29, 30], which can be downloaded from the home page of the author. All presented examples are part of its database and can be reevaluated or modified. Nonlinear optimization problems are solved by the SQP code NLPQLP, see Schittkowski [31].

2 Preliminaries: Fisher Information Matrix and Confidence Intervals

We proceed from a general nonlinear model in its simplest form

$$\eta = h(p,t) + \epsilon \ . \tag{5}$$

$h(p,t)$ is a scalar model function depending on a set of model parameters $p \in I\!R^n$. $t \in I\!R$ is the independent model variable, also called explanatory or regression variable, in most cases representing the time. The function $h(p,t)$ is supposed to be differentiable subject to p, and at least continuous with respect to t. It is assumed that there is a true parameter value $p^\star$, which is unknown and which is to be estimated by a least squares fit. The response function η is called the dependent model variable.

The above formulation proceeds for simplicity from a scalar variable t. Generalizations to multi-dimensional regression variables are possible without loss of generality. Also multi-response models where η possesses arbitrary dimension, can be considered, see Seber and Wild [35].

To estimate the true, but unknown parameter value $p^\star$ from given experimental data t_i and y_i, $i = 1, \ldots, l$, we minimize the least squares function

$$s(p) = \frac{1}{l-n} \sum_{i=1}^{l} (h(p,t_i) - y_i)^2 \tag{6}$$

over all $p \in I\!R^n$. Let $\hat{p}$ denote the solution of this data fitting problem. Then $\hat{p}$ is also called the ordinary least squares estimator (OLS) to distinguish it from alternative techniques, for example from the weighted or generalized least squares estimator. The question we are interested in is how far away $\hat{p}$ is from the true parameter $p^\star$.

It is assumed that the independent model values t_i are given a priori without errors, and that ϵ_i denotes the statistical error of the measurements or the response variable, respectively. Moreover, we suppose that the errors $\epsilon_i = y_i - h(p^\star, t_i)$ are independent and normally distributed with mean value zero and known constant variance σ^2, i.e., that $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, \ldots, l$.

The basic idea is to linearize the nonlinear model in a neighborhood of $p^\star$ and to apply linear regression analysis, since linear models are very well understood, see Seber [33]. By defining

$$f(p) = (h(p,t_1), \ldots, h(p,t_l)^T$$

and $\epsilon = y - f(p^\star)$, $q = p - p^\star$, $\epsilon = (\epsilon_1, \ldots, \epsilon_l)^T$, $y = (y_1, \ldots, y_l)^T$, we get from the first-order Taylor expansion

$$\begin{aligned} s(p) &= \|f(p) - y\|^2 \\ &\approx \|f(p^\star) + \nabla f(p^\star)^T (p - p^\star) - y\|^2 \\ &= \|\nabla f(p^\star)^T q - \epsilon\|^2 \ . \end{aligned}$$

Here $\|.\|$ denotes the Euclidian norm. We denote by $F^\star = \nabla f(p^\star)$ the Jacobian matrix of $f(p)$ at $p = p^\star$, and assume that $F^\star$ has full rank. A solution of the linear least squares problem is immediately obtained from the normal equations

$$\hat{q} = (F^\star {F^\star}^T)^{-1} F^\star \epsilon \ ,$$

from which we get a first-order approximation of the solution $\hat{p}$ by

$$\hat{p} = p^\star + (F^\star {F^\star}^T)^{-1} F^\star \epsilon \ .$$

From this approximation, some statistical properties known for linear models can be derived also for nonlinear ones. Under additional regularity assumptions, see Seber and Wild [35], $\hat{p}$ and $\hat{\sigma}^2 := s(\hat{p})$ are consistent estimates of $p^\star$ and σ^2, respectively. They converge with probability one to the true values, and are asymptotically normally distributed as l goes to infinity. Moreover, we know that due to the normal distribution of the errors, $\hat{p}$ is also a maximum likelihood estimator.

The error in parameters, $\hat{p} - p^\star$, is approximately normally distributed with mean value 0 and covariance matrix ${I_F^\star}^{-1}$, where $I_F^\star$ is defined by

$$I_F^\star := \frac{1}{\sigma^2} F^\star {F^\star}^T \ . \tag{7}$$

In addition, the expression

$$\frac{1}{n} (\hat{p} - p^\star)^T I_F^\star (\hat{p} - p^\star)$$

follows the F-distribution with $(n, l-n)$ degrees of freedom within the linearization error. Thus, an approximate $100(1-\alpha)\%$ confidence region for $p^\star$ is given by the set

$$\{\overline{p} : (\overline{p} - \hat{p})^T \hat{I}_F (\overline{p} - \hat{p}) \leq n F_{n,l-n}^\alpha\} \ , \tag{8}$$

where

$$\hat{I}_F := \frac{1}{\hat{\sigma}^2} \nabla f(\hat{p}) \nabla f(\hat{p})^T \tag{9}$$

estimates $I_F^\star$, a result very similar to the corresponding confidence region for linear models. $\hat{I}_F$ is called the Fisher information matrix. For a precise definition of this matrix and a proof see, e.g., Goodwin and Payne [18].

For a numerical implementation, however, (8) is inconvenient. To get individual confidence intervals for the coefficients of $p^\star$, we consider an arbitrary linear combination $a^T p$. It is possible to show that approximately

$$\frac{a^T \hat{p} - a^T p^\star}{\sqrt{a^T {I_F^\star}^{-1} a}} \sim t_{l-n} \ , \tag{10}$$

where t_{l-n} is the t-distribution with $l-n$ degrees of freedom. A $100(1-\alpha)\%$ confidence interval is then given by

$$\left[\ a^T \hat{p} - t_{l-n}^{\alpha/2} \sqrt{a^T {I_F^\star}^{-1} a} \ , \ a^T \hat{p} + t_{l-n}^{\alpha/2} \sqrt{a^T {I_F^\star}^{-1} a} \ \right] \ . \tag{11}$$

When setting $a = e_i$ for $i = 1, \ldots, n$ successively, where e_i is the i-th unit vector, and when estimating $I_F^\star$ by $\hat{I}_F$, we get the approximate confidence intervals

$$\left[\hat{p_i} - t_{l-n}^{\alpha/2} \sqrt{\hat{d}_{ii}} \ , \ \hat{p_i} + t_{l-n}^{\alpha/2} \sqrt{\hat{d}_{ii}} \right] \tag{12}$$

for the i-th individual model parameter value $p_i^\star$, $i = 1, \ldots, n$. In this case, $\hat{p_i}$ is the i-th coefficient of $\hat{p}$ and $\hat{d}_{ii}$ the i-th diagonal element of $\hat{I}_F^{-1}$, see also Gallant [15] or Donaldson and Schnabel [12].

However, (12) is valid only approximately depending on the quality of the linearization or the curvature of $f(p)$, respectively. Donaldson and Schnabel [12] present some examples, where the confidence intervals are very poor. Thus, we have to be very careful when computing (12) without additional linearization checks.

3 Significance Levels by Eigenvalue/-vector Analysis of the Fisher Information Matrix

Proceeding from a parameter estimation model, corresponding data, and a successful least squares fit, significance levels of the estimated parameters are to be evaluated. If a model seems to be overdetermined, i.e., contains too many parameters compared to the number of equations, the levels give an impression of the significance of parameters and help to decide upon questions like

- which parameters can be identified,
- which parameters can be treated as constants,
- whether additional experimental should be added or not.

Moreover, overdetermined data fitting problems lead to unstable and slow convergence of Gauss-Newton-type least squares algorithms with a large number of iterations until termination tolerances are satisfied.

We have seen in the previous section that $\hat{\sigma}^2 \hat{I}_F^{-1}$ can be considered as an approximation of the covariance matrix $\sigma^2 {I_F^\star}^{-1}$, where $\hat{\sigma}^2 = s(\hat{p})$, see (6), and where $\hat{I}_F$ denotes the Fisher information matrix (9). $\hat{p}$ is a least squares estimate for the true, but unknown parameter $p^\star$. Assumptions are, as before, independent and normally distributed errors in the measurements with mean value 0 and variance σ^2.

A more rigorous analysis based on the maximum-likelihood function leads to the theorem of Cramér and Rao, which states that the inverse of the Fisher information matrix is a lower bound for the covariance matrix of the parameter errors.

Since all induced matrix norms are greater than the spectral radius of a matrix, we apply the L_2-norm, i.e.,

$$\|\hat{I}_F^{-1}\|_2^2 = |\lambda_{max}(\hat{I}_F^{-1})| = \frac{1}{|\lambda_{min}(\hat{I}_F)|} \ . \tag{13}$$

λ_{max} and λ_{min} denote the largest and smallest eigenvalues of a matrix, respectively. Since small eigenvalues of $\hat{I}_F$ enforce large entries of the covariance

matrix, we try to reduce them by successive elimination of parameters corresponding to large eigenvector coefficients. The order by which the variables are eliminated, can be considered as an indication about their relative significance, the highest level reflects the highest priority.

We proceed from a given significance tolerance $\gamma > 0$ for estimating the error in the parameter approximations, known experimental data, and an optimal solution $\hat{p}$ of the corresponding least squares data fitting problem. We try to satisfy

$$\|\hat{I}_F^{-1}\|_2 = \sqrt{\frac{1}{|\lambda_{min}(\hat{I}_F)|}} < \gamma \ . \tag{14}$$

Assuming a sufficiently accurate approximation of $p^\star$, the true parameter vector, we hope to get sufficiently small variances.

Note that very small or zero eigenvalues lead to the conclusion that some parameters cannot be estimated at all by the underlying model and the available data, or that there are combinations of highly correlated parameters, see Caracotsis ans Stewart [11]. To detect the significant parameters on the one hand and the redundant or dependent parameters on the other we apply the subsequent procedure, see also Schneider, Posten, and Munack [32] or Majer [22]. The idea is to successively eliminate parameters until (14) is satisfied. The cycle is terminated in one of the following situations:

1. The smallest eigenvalue of the Fisher information matrix is smaller than a threshhold value, see (14).
2. The parameter correlations are significantly reduced, e.g., by 25 %.
3. None of the above termination reasons is met and all parameters have been eliminated.

Algorithm 3.1 *Let $k = 1$, $\hat{J}_k = \emptyset$, $\hat{I}_F^k = \frac{1}{\hat{\sigma}^2}\nabla f(\hat{p})\nabla f(\hat{p})^T$, $\hat{p}$ minimizer of (6), and $\gamma > 0$ be given.*

1. *Compute the smallest eigenvalue λ_{min} of $\hat{I}_F^k$ and a corresponding eigenvector $v_{min} \in I\!R^n$, $v_{min} = (v_1^{min}, \ldots, v_n^{min})^T$.*
2. *If $\lambda_{min} > \frac{1}{\gamma^2}$, then stop. The required significance level is reached.*
3. *Determine j_0 with*
$$|v_{j_0}^{min}| = \max_{1 \le j \le n} |v_j^{min}| \ ,$$
eliminate the j_0-th row and column from $\hat{I}_F^k$, denote the resulting matrix by $\hat{I}_F^{k+1}$, and let $\hat{J}_{k+1} = \hat{J}_k \cup \{j_0\}$.
4. *If $k = n - 1$ then stop, a further reduction is not possible.*
5. *Replace $k + 1$ by k and repeat from Step 1.*

After termination, the indices in $\hat{J}_k$ represent the significance levels of the parameters. Level 1 corresponds to the first eliminated variable, level 2 to the second, etc. The final level can be assigned to several parameters indicating a group of identifiable parameters. Possible conclusions are to add more experimental data or to fix some parameters for subsequent evaluations. Thus, the determination of significance levels is part of the experimental design process to validate a parameter estimation model, see Majer [22] for more details.

Since we analyze parameter estimates in unstructured, complex and highly nonlinear dynamical models, is is recommended to scale the parameter vector $\hat{p}$ before starting the above procedure.

Example 3.1 *A linear ordinary differential equation describes a kinetic process in the form*

$$\begin{aligned} \dot{y}_1 &= -k_1 y_1 &&, y_1(0) = D \ , \\ \dot{y}_2 &= k_1 y_1 - k_2 y_2 &&, y_2(0) = 0 \ , \end{aligned} \tag{15}$$

see `LKIN_A1/A2/A3/A4` *in the EASY-FIT database. A 95 % confidence region as outlined in the previous section, is shown in Table 1, i.e.,* $\hat{c}_i = 2t_{l-n}^{\alpha/2}\sqrt{\hat{d}_{ii}}$, *see (12). The estimated error variance is* $0.41 \cdot 10^{-3}$, *the maximum correlation is* 0.57, *and the covariance values are sufficiently small.*
Now we introduce some additional parameters with severe internal dependencies,

$$\begin{aligned} \dot{y}_1 &= -\sqrt{k_{11}} k_{12} y_1 &&, y_1(0) = D_1 + 0.1 D_2 \ , \\ \dot{y}_2 &= k_{11} k_{12} y_1 - (k_{21} + 2k_{22}) y_2 &&, y_2(0) = 0 \ . \end{aligned} \tag{16}$$

The same statistical analysis as above leads to the significance intervals of Table 3. The correlation coefficients between k_{11} *and* k_{12}, *between* k_{21} *and* k_{22}, *and between* D_1 *and* D_2 *are exactly one. By successive elimination of parameters with highest coefficient of the eigenvector belonging to the lowest eigenvalue, see Table 2, priority levels are computed as shown in Table 3. They exactly reflect the artificially generated dependencies. The parameters* k_{11}, k_{22}, *and* D_1 *obtained the highest scores and are considered as the most significant ones. We even observe that the influence of* k_{22} *on the solution is greater than that of* k_{21} *as can be expected from their different coefficients in (16). An important side effect is that the maximum correlation is reduced from* 1.0 *to* 0.56. *The significance level for the successive parameter elimination is* $\gamma = 0.05$.
Besides of detecting dependencies among parameters, the proposed analysis helps to find redundant ones, as shown by a slight modification of (15). An additional redundant parameter r *is added to the first differential equation leading to a very small perturbation of the solution by choosing* $\epsilon = 10^{-14}$,

$$\begin{aligned} \dot{y}_1 &= -k_1 y_1 + \epsilon r &&, y_1(0) = D \ , \\ \dot{y}_2 &= k_1 y_1 - k_2 y_2 &&, y_2(0) = 0 \ . \end{aligned} \tag{17}$$

The priority analysis detects the redundant parameter r, *see Table 4. The starting value of the redundant parameter is not changed by the least squares algorithm.*

Table 1: Confidence Intervals for (15)

p	$\hat{p}_i$	$\hat{c}_i$
k_1	0.1126	0.0034
k_2	0.0571	0.0022
D	102.4778	1.79

Table 2: Elimination of Parameters for (16)

k	$\lambda_{min}(\hat{I}_F^k)$	$\lvert v_{j_0}\rvert$	j_0
1	$-0.37 \cdot 10^{-11}$	0.89	3
2	$-0.12 \cdot 10^{-11}$	0.99	6
3	0.46.7	0.97	1
4	2578.3	1.0	5

Table 3: Confidence Intervals and Priority Levels for Overdetermined System (16)

p	$\hat{p}_i$	$\hat{c}_i$	$\hat{J}_k$
k_{11}	0.1328	0.0041	3
k_{12}	0.8476	0.00065	4
k_{21}	0.0314	0.00047	1
k_{22}	0.0128	0.00094	4
D_1	51.2396	0.96	4
D_2	51.2396	0.96	2

Table 4: Priority Levels for Redundant System (17)

p	$\hat{p}_i$	$\hat{J}_k$
k_1	0.1126	2
k_2	0.0571	2
D	102.4778	2
r	1.000	1

4 Case Study: Isothermal Batch Reactor

By the first case study, we illustrate the computation of significance levels as outlined in Section 3. The corresponding identifiers in the EASY-FIT database are `BATCH_F1/F2/F3/F4`. The kinetic model describes a chemical reaction system of an isothermal batch reactor, see Biegler, Damiano and Blau [9] or Majer [22],

$$\begin{aligned}
\dot{x}_1 &= -k_2 x_2 x_8 \ , \\
\dot{x}_2 &= -k_1 x_2 x_6 + k_{-1} x_{10} - k_2 x_2 x_8 \ , \\
\dot{x}_3 &= \ \ k_2 x_2 x_8 + k_1 x_4 x_6 - 0.5 k_{-1} x_9 \ , \\
\dot{x}_4 &= -k_1 x_4 x_6 + 0.5 k_{-1} x_9 \ , \\
\dot{x}_5 &= \ \ k_1 x_2 x_6 + k_{-1} x_{10} \ , \\
\dot{x}_6 &= -k_1 x_2 x_6 - k_1 x_4 x_6 + k_{-1} x_{10} + 0.5 k_{-1} x_9 \ , \\
0 &= -x_7 + x_6 + x_8 + x_9 + x_{10} - Q^+ \ , \\
0 &= -x_8 (K_2 + x_7) + K_2 x_1 \ , \\
0 &= -x_9 (K_3 + x_7) + K_3 x_3 \ , \\
0 &= -x_{10} (K_1 + x_7) + K_1 x_5 \ .
\end{aligned} \tag{18}$$

We have

$$\begin{aligned}
Q^+ &= 0.0131 \ , \\
T_b &= 342.15 \ , \\
k_1 &= \exp\left(p_1 - \left(\tfrac{1}{T} - \tfrac{1}{T_b}\right) \exp(p_7)\right) \ , \\
k_2 &= \exp\left(p_2 - \left(\tfrac{1}{T} - \tfrac{1}{T_b}\right) \exp(p_8))\right) \ , \\
k_{-1} &= \exp\left(p_3 - \left(\tfrac{1}{T} - \tfrac{1}{T_b}\right) \exp(p_9)\right) \ , \\
K_1 &= \exp(-p_4) \ , \\
K_2 &= \exp(-p_5) \ , \\
K_3 &= \exp(-p_6) \ ,
\end{aligned}$$

nine parameters to be estimated, and three experimental data sets obtained under different temperatures and initial concentrations,

T	$x_1(0)$	$x_2(0)$	$x_3(0)$	$x_4(0)$	$x_5(0)$	$x_6(0)$
313.15	1.7066	8.3200	0.0100	0.0000	0.0	0.0131
340.15	1.6497	8.2200	0.0104	0.0017	0.0	0.0131
373.00	1.5608	8.3546	0.0082	0.0086	0.0	0.0131

In case of estimating only one data set, it is obvious that there are strong internal dependencies between p_1 and p_7, p_2 and p_8, and p_3 and p_9. This is reflected by the priorities listed in Table 5, where $\gamma = 0.1$. At least one of the two corresponding priorities, i.e., of $p_1(= 3)$, $p_8(= 1)$, and $p_9(= 2)$, obtained

	$T = 340.15$		$T = 313.15$, $T = 340.15$		$T = 313.15$ $T = 340.15$, $T = 373$	
p	$\hat{p}$	$\hat{J}_k$	$\hat{p}$	$\hat{J}_k$	$\hat{p}$	$\hat{J}_k$
p_1	$2.8 \cdot 10^{-5}$	3	0.0	3	0.0	3
p_2	1.6	6	1.1	4	1.1	3
p_3	8.7	5	2.8	4	5.0	3
p_4	34.2	6	21.6	4	24.9	3
p_5	26.5	4	17.2	2	18.0	2
p_6	32.5	6	20.1	4	23.2	3
p_7	10.6	6	9.3	4	9.3	3
p_8	8.6	1	8.9	4	8.9	3
p_9	0.0002	2	0.04	1	0.04	1

Table 5: Parameter Values and Priorities for Isothermal Batch Reactor

the lowest possible value. The remaining parameters p_2, p_3, and p_7 seem to be identifiable.

To improve the number of parameters which can be identified, we add one and then two additional data sets, see again Table 5. For three different data sets, seven of nine parameters are considered as identifiable within one group similar to the results obtained by Majer [22]. We also observe that the parameter values get more and more stabilized, and some of them for three data sets are quite far away from the parameter values obtained for one data set.

5 Experimental Design

Optimal experimental design is extensively discussed in the literature, see the subsequent case studies and the references cited there. We present an approach based on approximations of partial second derivatives by forward differences. It is shown that a relatively large perturbation tolerances of the difference formula lead to nevertheless sufficiently accurate solutions of the nonlinear optimization problem to be solved.

Mathematical models describe the dynamical behavior of a system with the goal to allow numerical estimation of model parameters a user is interested in. These parameters identify the system under consideration, and are to be verified by experiments.

However, the experimental design often depends on parameters which must be set in advance to be able to measure certain output data of an experiment. Examples are initial concentration of substrates, input feeds of a chemical reactor, temperature distributions, etc. In addition, our model may depend on universal physical parameters like gas constant, absolute temperature, or gravitational constant.

To determine the experimental design parameters in an optimal way, we first

have to find a suitable guess for the model parameters either from the literature or some preliminary experiments. We have seen in the previous sections that the covariance matrix determines the confidence region of the model parameters, see (8). Since we have now additional freedom to design an experiment, we can use the design parameters to minimize the volume of the corresponding ellipsoid based on a suitable criterion.

To formalize the situation, we denote the model parameters by $p \in \mathbb{R}^{n_p}$ and the design parameters by $q \in \mathbb{R}^{n_q}$. In case of a dynamic, i.e., time dependent parameter, for example a control function, we assume that the control function is approximated by finitely many parameters.

Now we extend our model function $h(p,t)$ by the design parameters, $h(p,q,t)$, and assume that we have a set of experimental time values t_k, $k = 1, \ldots, l$. Moreover, we let

$$f(p,q) = (h(p,q,t_1), \ldots, h(p,q,t_l))^T$$

and denote by $F(p,q) = \nabla_p f(p,q)$ the Jacobian matrix of $f(p,q)$ subject to p for a given $q \in \mathbb{R}^{n_q}$. For simplicity, we assume that $F(p,q)$ has full rank for all p and q.

A formal performance measure is available based on the covariance matrix $C(p,q) := I(p,q)^{-1}$, where $I(p,q) := F(p,q)F(p,q)^T$ denotes an approximation of the Fisher information matrix, and where we omit a guess for the error variances of measurements to simplify the notation. In other words, we assume that all experimental data are measured with constant error. The volume of a confidence region for a given model parameter p is given by

$$\{\overline{p} : (\overline{p} - p)^T I(p,q)(\overline{p} - p) \leq \alpha_{n_p}\} \tag{19}$$

with a statistical parameter α_{n_p}, see (8).

Formula (19) describes an ellipsoid, and the goal is to minimize its volume on the one hand, but on the other to prevent also degenerate situations where the maximal and minimal eigenvalue drift away. This is to be achieved by adapting the design parameter q for a given model parameter p, which is obtained either from a preliminary experiment, literature, or a reasonable guess. Possible criteria are available either for $C(p,q)$ or $I(p,q)$, respectively, depending on the procedure how to measure or estimate the volume and the structure of the ellipsoid. The most popular ones are

D	:	$\det(C(p,q))$
A	:	$\text{trace}(C(p,q))$
$A^\star$	:	-trace$(I(p,q))$
E	:	$\lambda_{min}(I(p,q))$
$E^\star$ or C	:	$\lambda_{min}(I(p,q))/\lambda_{max}(I(p,q))$

Here $\lambda_{min}(I(p,q))$ and $\lambda_{max}(I(p,q))$ denote the minimal and maximal eigenvalues of $I(p,q)$. For a more detailed discussion, see, e.g., Winer, Brown and Michels [43] or Ryan [28].

For our numerical implementation, we use the A-criterion, since the computationally attractive confidence intervals by which the size of the ellipsoid is estimated, take only the diagonal elements of the covariance matrix into account, see (12). This leads for each $p \in \mathbb{R}^{n_p}$ to the optimization problem

$$q \in \mathbb{R}^{n_q} : \quad \begin{array}{l} \min \ \mathrm{trace}(C(p,q)) \\ q_l \le q \le q_u \ , \end{array} \tag{20}$$

where we add additional bounds for the variables q.

There remains the question how to compute the derivatives of the objective function

$$\phi(q) := \mathrm{trace}(C(p,q)) \tag{21}$$

subject to q in an efficient way. Numerical differentiation of $\phi(q)$ subject to q by a difference formula based on a previous numerical differentiation of $h(p,q,t)$ subject to p by another difference formula is unstable because of accumulation of truncation errors. It is assumed that second order analytical mixed-partial derivatives are not available. Thus, we try to find a reasonable compromise which nevertheless leads to sufficiently stable procedure.

Partial differentiation formulae of the objective function of (20) subject to q_r, $1 \le r \le n_q$, are well-known from the literature, see e.g., Bauer et al. [5], and are repeated for the matter of completeness,

$$\begin{aligned} \frac{\partial}{\partial q_r}\phi(q) &= \frac{\partial}{\partial q_r}\left(\mathrm{trace}(C(p,q))\right) \\ &= \mathrm{trace}\left(\frac{\partial}{\partial q_r} I(p,q)^{-1}\right) \\ &= \mathrm{trace}\left(-I(p,q)^{-1}\ \frac{\partial}{\partial q_r} I(p,q)\ I(p,q)^{-1}\right) \\ &= -\mathrm{trace}\left(I(p,q)^{-1}\ \frac{\partial}{\partial q_r}\left(F(p,q)F(p,q)^T\right)\ I(p,q)^{-1}\right) \\ &= -\mathrm{trace}\left(I(p,q)^{-1}\ \left(\frac{\partial}{\partial q_r}F(p,q)F(p,q)^T \right.\right. \\ &\qquad \left.\left. + F(p,q)\frac{\partial}{\partial q_r}F(p,q)^T\right)\ I(p,q)^{-1}\right) \ . \end{aligned} \tag{22}$$

There remains differentiation of the $l \times n_p$ matrix

$$\begin{aligned} \frac{\partial}{\partial q_r}F(p,q) &= \frac{\partial}{\partial q_r}\nabla_p f(p,q) \\ &= \left(\frac{\partial^2}{\partial q_r \partial p_i} h(p,q,t_k)\right)_{i=1,n_p;k=1,l} \ . \end{aligned} \tag{23}$$

Typically, the remaining mixed partial derivatives derivatives are provided by automatic differentiation of the underlying ODE or DAE solver, respectively, or by internal numerical differentiation. However, we proceed from more general models and allow higher-index DAEs, switching points with variable transition conditions, initial values depending on parameters, and especially from dynamical systems where analytical derivatives of the right-hand side are not available. Thus, the mixed partial derivatives of the model function $h(p,q,t)$ subject to p and q are approximated by forward differences

$$\begin{aligned}\frac{\partial^2}{\partial q_r \partial p_i} h(p,q,t_k) \quad &\approx \quad \frac{1}{\epsilon_r \epsilon_i} \left((h(p+\epsilon_i e_i, q+\epsilon_r e_r, t_k) + h(p,q,t_k)) \right. \\ &\qquad \left. - (h(p,q+\epsilon_r e_r,t_k) + h(p+\epsilon_i e_i, q, t_k)) \right)\end{aligned} \tag{24}$$

for $k = 1, \ldots, l$ and $i = 1, \ldots, n_p$. Here, $e_i \in I\!R^{n_p}$ and $e_r \in I\!R^{n_q}$ are the i-th and r-th unit vectors, respectively, and ϵ_i, ϵ_r are suitable perturbation tolerances, e.g., chosen by $\epsilon_i := \max(\nu, |p_i|)\epsilon$ and $\epsilon_r := \max(\nu, |q_r|)\epsilon$. The tolerances $\epsilon > 0$ and $\nu > 0$ must be selected very carefully.

Equation (24) is written in a form to show that cancelation appears only once. Since the evaluation of the objective function $\phi(q)$, i.e., of $F(p,q) = \nabla_p f(p,q)$, requires also an approximation of first derivatives of the form

$$\frac{\partial}{\partial p_i} h(p,q,t_k) \quad \approx \quad \frac{1}{\epsilon_i} \left(h(p+\epsilon_i e_i, q, t_k) - h(p,q,t_k) \right) \quad , \tag{25}$$

only two additional evaluation of h are required to get the mixed second-order derivatives (24).

The perturbation tolerance ϵ should not be too small. Depending on the condition number of the information matrix, large values like $\epsilon = 0.01$ or even $\epsilon = 0.1$ are applicable and can lead to stable solution processes subject to a surprisingly small optimality criterion, as shown by the subsequent case studies.

The experimental design approach discussed above assumes that the time values t_i are known. However, there are very many situations where one would like to know in advance their approximate number and also their optimal locations, to improve the confidence intervals of the parameters to be estimated, and to reduce the number of time-consuming or expensive experiments.

Our idea is to proceed from a given set of time values which could be large and dense, and to formulate an experimental design optimization problem as before by introducing additional weights w_k, $k = 1, \ldots, l$. Thus, we replace the k-th model function $h(p,q,t_k)$ by $w_k h(p,q,t_k)$ with an additional factor w_k, $k = 1, \ldots, l$, which is treated as an optimization variable of the optimum design problem (20),

$$\begin{aligned} q \in I\!R^{n_q}, w \in I\!R^l : \quad & \min \ \text{trace}(C(w,p,q)) \\ & \textstyle\sum_{k=1}^{l} w_k = 1 \quad , \\ & q_l \le q \le q_u \quad , \\ & \tau \le w_k \le 1 \quad , \qquad\qquad k = 1, \ldots, l \quad . \end{aligned} \tag{26}$$

The covariance matrix $C(w,p,q) := I(w,p,q)^{-1}$ depends now on additional weights, $I(w,p,q) := F(w,p,q)F(w,p,q)^T$, $F(w,p,q) := \nabla_p f(w,p,q)$, and finally

$$f(w,p,q) := (w_1 h(p,q,t_1), \ldots, w_l h(p,q,t_l))^T \quad .$$

Note that for stability reasons, a small lower bound $\tau > 0$ is introduced. We have to prevent the possibility that weights become zero at an intermediate iterate leading to an indefinite information matrix.

Corresponding partial derivatives of the objective function

$$\phi(w,q) = \text{trace}(C(w,p,q))$$

subject to a weight w_k are obtained from

$$\begin{aligned} \frac{\partial}{\partial w_k}\phi(w,q) &= -\text{trace}\Big(I(w,p,q)^{-1} \Big(\frac{\partial}{\partial w_k} F(w,p,q) F(w,p,q)^T \\ &\qquad + F(w,p,q) \frac{\partial}{\partial w_k} F(w,p,q)^T \Big) \; I(w,p,q)^{-1} \Big) \quad , \end{aligned} \tag{27}$$

see (22), and from

$$\begin{aligned} \frac{\partial}{\partial w_k} F(w,p,q) &= \frac{\partial}{\partial w_k} \nabla_p f(w,p,q) \\ &= \left(\frac{\partial^2}{\partial w_k \partial p_i} (w_k h(p,q,t_k)) \right)_{i=1,n_p;k=1,l} \\ &= \left(\frac{\partial}{\partial p_i} h(p,q,t_k) \right)_{i=1,n_p;k=1,l} \quad , \end{aligned} \tag{28}$$

confer also (23). Thus, we get the weight derivatives more or less for free, since the partial derivatives subject to the model parameters are known from the computation of the objective function.

6 Case Study: Unstructured Microbial Growth Model

Banga et al. [2] consider a design problem based on an unstructured microbial growth model to determine feed rate profiles in fed-batch bio-reactors. They mention that numerical instabilities prevent application of gradient-based optimization procedures. Instead, the use a stochastic search algorithm.

The process is described by two differential equations and the integration of

an input function $F_{in}(t)$,

$$\begin{aligned} \dot{C}_S &= -\sigma C_X + F_{in}(t)\frac{C_{Sin} - C_S}{V} \ , & C_S(0) &= C_S^0 \ , \\ \dot{C}_X &= \mu C_X - F_{in}(t)\frac{C_X}{V} \ , & C_X(0) &= \frac{C_X^0}{V_0} \ , \\ \dot{V} &= F_{in}(t) \ , & V(0) &= V_0 \ , \end{aligned} \tag{29}$$

where

$$\begin{aligned} \mu &= \mu_m \frac{C_S}{K_p + C_S + C_S^2/K_i} \ , \\ \sigma &= \frac{\mu}{Y_{XS} + m} \ , \\ V_0 &= V^\star \frac{C_{Sin}}{C_{Sin} - C_S^0} \end{aligned}$$

and $m = 0.29$, $C_{Sin} = 500$, $Y_{XS} = 0.47$, $V^\star = 7$, and $\mu_m = 2.1$. F_{in} is a piecewise linear input control function chosen very close to the optimal solution found by Versyck [40], K_p, K_i are model parameters to be estimated, and initial values C_S^0, C_X^0 are design parameters. It turns out that K_p is very difficult to estimate. Starting from some reasonable initial guesses $K_p = 10$, $K_i = 0.1$, $C_S^0 = 40$, $C_X^0 = 10$, artificial measurements are generated and perturbed by a uniform error of 5 %. Then, confidence intervals are computed for the design parameters K_p and K_i. The corresponding data fitting problems are found in the EASY-FIT database under the identifiers `MICGROWX/Y/Z`.

In the next step, we consider the feed controls at 19 grid points as additional design parameters, and 20 constraints are added to prevent that $C_S(t)$ falls below zero. The perturbation tolerance for gradient approximations by forward differences is set to $\epsilon = 0.01$. NLPQLP needs 18 iterations to reduce the performance criterion from $1.3 \cdot 10^5$ to 0.009 under termination accuracy 10^{-8}. Optimal design parameters are the initial concentrations $C_S^0 = 38.9$ and $C_X^0 = 14.3$, and the optimal feed curve is shown in Figure 1. In Figures 2 and 3 the corresponding state functions $C_S(t)$ and $C_X(t)$ are plotted.

After getting the optimal design parameters, the confidence intervals are computed in the same way as for the starting values. Standard deviations are reduced from 239.8 to 0.093 for K_p and from 0.0096 to 0.0022 for K_i. Moreover, the correlation coefficient is reduced from 0.99 to 0.19.

In addition to the model and design parameters, also weights are to be computed at an equidistant grid of 43 time values. Thus, the optimization problem (26) gets 86 additional variables. NLPQLP computes a solution in seven iterations with termination accuracy 10^{-6}. The total number of experiments is reduced to 5, see Figure 4, and would have to be taken into account only for C_S. Model parameter $K_p = 10$ is estimated subject to a confidence level 0.16 and $K_i = 0.1$ subject to 0.00011. The location of the time values seems to be at the critical points which determine the structure of the dynamical system.

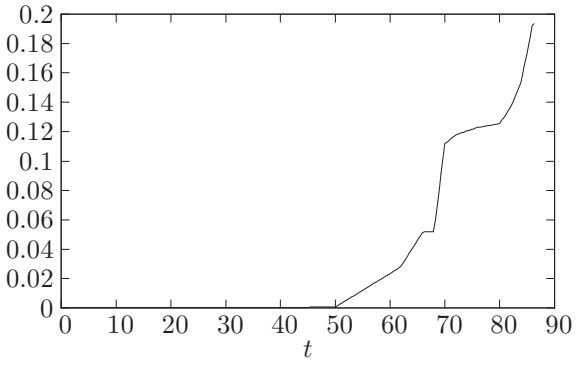

Figure 1: Control Function $F_{in}(t)$

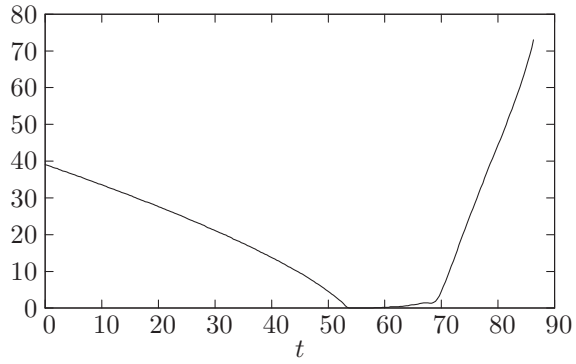

Figure 2: State Function $C_S(t)$

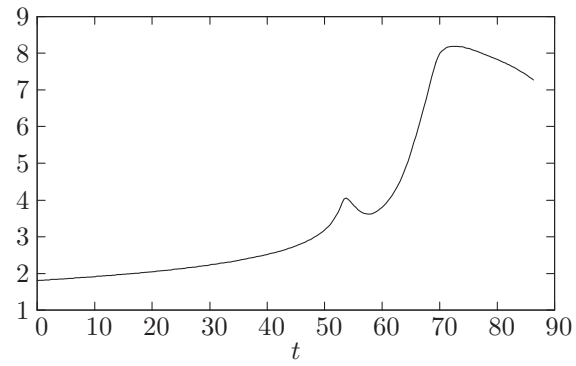

Figure 3: State Function $C_X(t)$

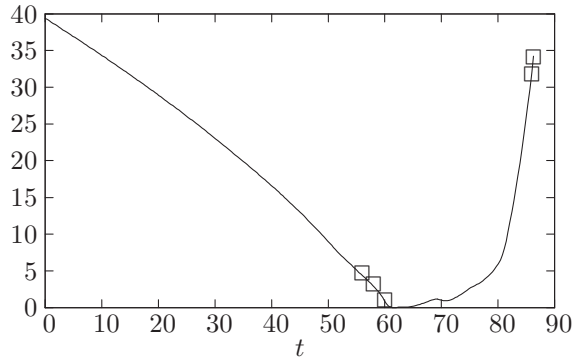

Figure 4: State Function $C_S(t)$ with Experimental Data

7 Case Study: Chemical Reaction of Urethane

A practically relevant example is studied by Bauer et al. [5], the reaction of urethane, see `URETHANX/W` in the EASY-FIT database. The corresponding DAE describing the reaction of phenylisocyanate (n_1), butanol (n_2), urethane (n_3), allophante (n_4), and isocyanurate (n_5) consists of three differential and three algebraic equations of index 1,

$$\begin{array}{lcll} \dot{n}_3 &=& V(r_1 - r_2 + r_3) \ , & n_3(0) = 0 \\ \dot{n}_4 &=& V(r_2 - r_3) \ , & n_4(0) = 0 \\ \dot{n}_5 &=& V r_4 \ , & n_5(0) = 0 \\ 0 &=& n_1 + n_3 + 2n_4 + 3n_5 - n_{a1} - n_{1ea}(t) \ , & \\ 0 &=& n_2 + n_3 + n_4 - n_{a2} - n_{2eb}(t) \ , & \\ 0 &=& n_6 - n_{a6} - n_{6ea}(t) - n_{6eb}(t) \ , & \end{array} \tag{30}$$

where n_6 denotes the solvent and

$$\begin{array}{lcllcl} V &=& \sum_{i=1}^{6} \frac{M_i n_i}{\rho_i} \ , & k_1 &=& k_{ref1} \exp\left(-E_{a1}(1/T(t) - 1/T_{ref1})/R\right) \ , \\ r_1 &=& k_1 \dfrac{n_1 n_2}{V^2} \ , & k_2 &=& k_{ref2} \exp\left(-E_{a2}(1/T(t) - 1/T_{ref2})/R\right) \ , \\ r_2 &=& k_2 \dfrac{n_1 n_3}{V^2} \ , & k_3 &=& k_2/k_c \ , \\ r_3 &=& k_3 \dfrac{n_4}{V} \ , & k_4 &=& k_{ref4} \exp\left(-E_{a4}(1/T(t) - 1/T_{ref4})/R\right) \ , \\ r_4 &=& k_4 \dfrac{n_1^2}{V^2} \ , & k_c &=& k_{c2} \exp\left(-d_{h2}(1/T(t) - 1/T_{g2})/R\right) \ . \end{array}$$

Two input feeds are given in form of non-decreasing functions $feed_a(t)$ and $feed_b(t)$, $t \in [0, 80]$, and define $n_{1ea}(t) = n_{a1ea} feed_a(t)$, $n_{2eb}(t) = n_{a2eb} feed_b(t)$,

$n_{6ea}(t) = n_{a6ea} feed_a(t)$, and $n_{6eb}(t) = n_{a6eb} feed_b(t)$. Mol ratios, active ingredients, and the initial volume have to satisfy certain bound constraints,

$$\begin{aligned}
0.1 &\leq MV_1 \leq 10 \ , \\
0 &\leq MV_2 \leq 1000 \ , \\
0 &\leq MV_3 \leq 10 \ , \\
0 &\leq g_a \leq 0.8 \ , \\
0 &\leq g_{aea} \leq 0.9 \ , \\
0 &\leq g_{aeb} \leq 1 \ , \\
0 &\leq V_a \leq 0.00075 \ ,
\end{aligned}$$

and are connected to the remaining parameters by analytical equations

$$\begin{aligned}
MV_1(n_{a1} + n_{a1ea}) &= n_{a2} + n_{a2eb} \ , \\
MV_2 n_{a1} &= n_{a1ea} \ , \\
MV_3 n_{a1} &= n_{a2eb} \ , \\
g_a(n_{a1}M_1 + n_{a2}M_2 + n_{a6}M_6) &= n_{a1}M_1 + n_{a2}M_2; \ , \\
g_{aea}(n_{a1ea}M_1 + n_{a6ea}M_6) &= n_{a1ea}M_1 \ , \\
g_{aeb}(n_{a2eb}M_2 + n_{a6eb}M_6) &= n_{a2eb}M_2 \ , \\
V_a &= n_{a1}M_1/\rho_1 + n_{a2}M_2/\rho_2 + n_{a6}M_6/\rho_6 \ .
\end{aligned}$$

which play the role of nonlinear equality constraints. Constant data are given for

$$\begin{aligned}
M_1 &= 0.11911 \ , & \rho_1 &= 1095 \ , & T_{ref1} &= 363.16 \ , \\
M_2 &= 0.07412 \ , & \rho_2 &= 809 \ , & T_{ref2} &= 363.16 \ , \\
M_3 &= 0.19323 \ , & \rho_3 &= 1415 \ , & T_{ref4} &= 363.16 \ , \\
M_4 &= 0.31234 \ , & \rho_4 &= 1528 \ , & T_{g2} &= 363.16 \ , \\
M_5 &= 0.35733 \ , & \rho_5 &= 1451 \ , & R &= 8.314 \ , \\
M_6 &= 0.07806 \ , & \rho_6 &= 1101 \ . &&
\end{aligned}$$

Model parameters to be estimated and for which some initial guesses are available, are

$$\begin{aligned}
k_{ref1} &= 5 \cdot 10^{-4} \ , & E_{a1} &= 3.52 \cdot 10^4 \ , \\
k_{ref2} &= 8 \cdot 10^{-8} \ , & E_{a2} &= 8.5 \cdot 10^4 \ , \\
k_{ref4} &= 1 \cdot 10^{-8} \ , & E_{a4} &= 3.5 \cdot 10^4 \ , \\
k_{c2} &= 1.7 \cdot 10^{-1} \ , & d_{h2} &= 1.08 \ .
\end{aligned}$$

Table 6: Design Parameters of the Urethane Problem

p	initial	final
MV_1	1.0	0.24299
MV_2	0.3	1.13686
MV_3	0.3	0.25613
g_a	0.75	0.80000
g_{aea}	0.5	0.67401
g_{aeb}	0.4	0.32820
V_a	2.75	11.29776
n_{a1}	0.106	0.68736
n_{a2}	0.106	0.18089
n_{a6}	0.0876	0.30515
n_{a1ea}	0.0319	0.78145
n_{a2eb}	0.0319	0.17605
n_{a6ea}	0.0486	0.57685
n_{a6eb}	0.0454	0.34219

The two input feed controls and the time-dependent temperature are piecewise linear functions defined at 10 grid points between $t = 8$ and $t = 80$. The corresponding 30 support values are experimental design parameters together with the bounded parameters MV_1, MV_2, MV_3, g_a, g_{aea}, g_{aeb}, V_a and the parameters n_{a1}, n_{a2}, n_{a6}, n_{a1ea}, n_{a2eb}, n_{a6ea}, and n_{a6eb}, which are coupled by a set of seven equations mentioned above. To sum up, the whole optimization problem consists of 8 model parameters, 47 design parameters, 8 nonlinear equality constraints, and 20 linear inequality constraints to satisfy monotonicity of the input feeds. In addition, there are 10 time values between 0 and 80, four measurement functions n_1, n_3, n_4, and n_5, and model variables are scaled to one.

The initial design is based on the data of Table 6. First, we suppose that the parameters given above, are the result of 'real' data fitting run. Experimental data are generated at the 10 time values and random errors based on a uniform distribution with relative deviation of 1 %. Subsequently, confidence intervals subject to model parameters are computed as described in Section 1, see (12). The results are listed in Table 7. The maximum standard deviation is more than 800 %, i.e., it is practically impossible to estimate the model parameters based on the given design data.

Our code is executed with $\epsilon = 10^{-2}$ for the approximation of partial derivatives. The optimization routine NLPQLP of Schittkowski [31] terminates after 72 iterations reducing the performance criterion from 2.7×10^9 to $6.3 \cdot 10^4$, with termination tolerance 10^{-8}. Optimal design values are listed in Table 6. Corresponding state and control functions are shown in Figures 5 to 8. As for the initial design, confidence intervals are computed for the design parameters, see Table 7. Now all deviations are below 15 %.

Table 7: Confidence Intervals for Urethane Problem

p	initial (%)	final (%)	weights (%)
k_{ref1}	1.14	14.70	7.44
k_{ref2}	56.45	11.53	2.73
k_{ref4}	1.62	1.30	0.20
k_{c2}	859.13	4.71	0.00058
E_{a1}	0.77	0.18	2.34
E_{a2}	1.24	1.95	0.95
E_{a4}	1.38	0.84	0.14
d_{h2}	676.79	6.21	0.00026

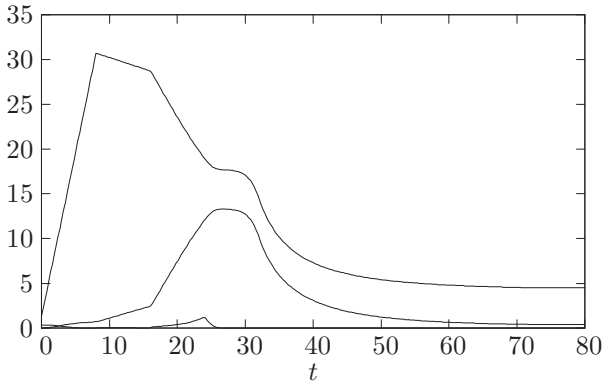

Figure 5: State Functions $n_1(t)$, $n_2(t)$, $n_3(t)$

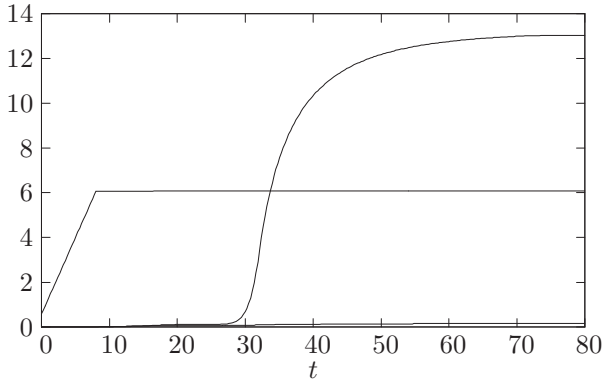

Figure 6: State Functions $n_4(t)$, $n_5(t)$, $n_6(t)$

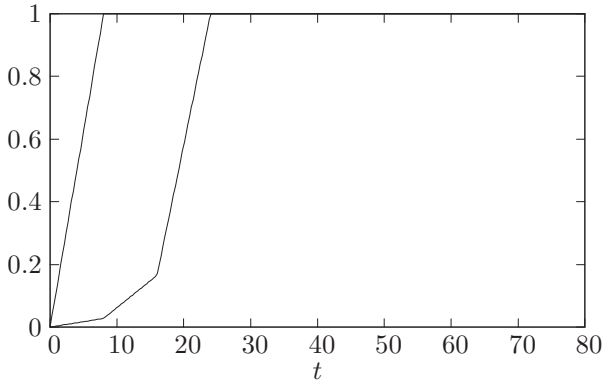

Figure 7: Control Functions $feed_a(t)$ and $feed_b(t)$

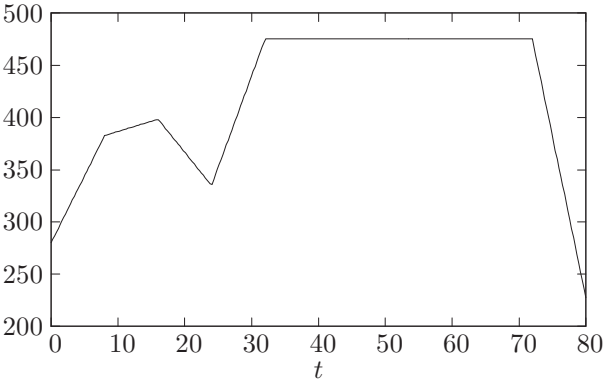

Figure 8: Temperature $T(t)$

Table 8: Optimal Weights for Urethane Problem

i	t_i	$n_1(t_i)$	$n_3(t_i)$	$n_4(t_i)$	$n_5(t_i)$
1	12	0.183	0.267		
2	16	0.0604			
3	18	0.0073			
4	20			0.0022	0.0018
5	26			0.0929	
6	28	0.03	0.131		0.0111
7	30				0.171
8	32	0.074			0.036
9	34				0.0035
10	36				
11	38	0.0123		0.0934	0.0025
12	40				0.0031
13	80	0.0068			

We consider again the urethane problem because of its practical relevance, see also (30). It is pointed out in Bauer et al. [5] that the experiments are expensive and that it is highly desirable to reduce their number as much as possible. We proceed from 40 equidistant time values between 0 and 80 for the four measurable output functions n_1, n_3, n_4, and n_5, and try to reduce their number to only the significant ones without loosing the desired identification option as computed in the previous section. Note that all substrates are measured independently of each other, i.e., we have a total of $l = 160$ experimental data from where relevant ones are to be extracted.

The optimization routine needs 132 iterations to reduce the performance criterion from $1.65 \cdot 10^{19}$ to $8.0 \cdot 10^{14}$ under stopping tolerance 10^{-8}. Nineteen weights are above the lower bounds as shown in Table 8, and the corresponding confidence levels are found in Table 7. They are significantly smaller than in case of the 40 measurements taken in the previous section, and also the reduction of experimental expenses is significant.

8 Conclusions

A couple of practical tools are proposed to compute optimal designs, to analyze results of a data fitting run based on confidence intervals, and to get significance levels for the estimated parameters. Of particular interest is the reduction of the number of experimental measurements to be retrieved, by weight optimization. The feasibility of this approach and the robustness of the numerical implementation are evaluated by some practical case studies.

The successive elimination of redundant parameters is very simple and does not depend on the mathematical structure of the underlying model equations,

i.e., the procedure is also applicable in situations different from those discussed in this paper. The idea is only based on eigenvalue and eigenvector computations which are easily performed by available standard routines. However, the Fisher information matrix might become singular and highly ill-conditions requiring a careful implementation of the code.

Experimental design is discussed to show that the generally accepted paradigm of analytical derivatives is not always mandatory. At least for the two non-trivial test cases we show that simple forward differences may be applied after some safeguards. The stability of the solution depends on a careful choice of a tolerance for approximating derivatives by forward differences. More numerical tests and analytical investigations are necessary to find out guidelines how to choose the tolerance in advance in a proper way.

References

[1] Baltes M., Schneider R., Sturm C., Reuss M. (1994): *Optimal experimental design for parameter estimation in unstructured growth models*, Biotechnologivcal Progress, Vol. 10, 480-488

[2] Banga J.R., Versyck K.J., Van Impe J.F. (2002): *Computation of optimal identification experiments for nonlinear dynamic process models: a stochastic global optimization apporoach*, Industrial and Engineering Chemistry Research, Vol. 41, 2425-2430

[3] Banks H.T, Kunisch K. (1989): *Estimation Techniques for Distributed Parameter Systems,* Birkhäuser, Boston, Basel, Berlin

[4] Bard Y. (1974): *Nonlinear Parameter Estimation*, Academic Press, New York, London

[5] Bauer I., Bock H.G., Körkel S., Schlöder J. (2000): *Numerical methods for optimum experimental design in DAE systems*, Journal of Computational and Applied Mathematics, Vol. 120, 1-25

[6] Baumeister J. (1987): *Stable Solution of Inverse Problems,* Vieweg, Braunschweig

[7] Beck J.V., Arnold K.J. (1977): *Parameter Estimation in Engineering and Science*, John Wiley, New York

[8] Ben-Zvi A., McLellan P.J., McAuley K.B. (2003): *Identifiability of linear time-invariant differential-algebraic systems. I. The generalized Markov parameter approach*, Industrial and Engineering Chemistry Research, Vol. 42, 6607-6618

[9] Biegler L.T., Damiano J.J., Blau G.E. (1986): *Nonlinear parameter estimation: a case study comparison*, AIChE Journal, Vol. 32, No. 1, 29-45

[10] Bortz D.M., Nelson P.W. (2006): *Model selection and mixed-effects modeling of HIV infection dynamics*, Bulletin of Mathematical Biology, Vol. 68, 2005-2025

[11] Caracotsios M., Stewart W.E. (1985): *Sensitivity analysis of initial value problems with mixed ODE's and algebraic equations*, Computers and Chemical Engineering, Vol. 9, 359-365

[12] Donaldson J.R., Schnabel R.B. (1987): *Computational experience with confidence regions and confidence intervals for nonlinear least squares*, Technometrics, Vol. 29, 67-82

[13] Draper N.R., Smith H. (1981): *Applied Regression Analysis*, John Wiley, New York

[14] DuChateau P. (1995): *An introduction to inverse problems in partial differential equations for engineers, physicists, and mathematicians, a tutorial,* in: Proceedings of the Workshop on Parameter Identification and Inverse Problems in Hydrology, Geology, and Ecology, J. Gottlieb, P. DuChateau eds., Kluwer Academic Publishers, Dordrecht, Boston, London 3 - 50

[15] Gallant A.R. (1975): *Nonlinear Regression*, American Statistics, Vol. 29, No. 2, 73-81

[16] Gallant A.R. (1987): *Nonlinear Statistical Models*, John Wiley, New York

[17] Godfrey K.R., DiStefano J.J. (1985): *Identifiability of model parameters*, in: IFAC Identification and System Parameter Estimation, P. Joung ed., Pergamon Press, Oxford, 89-114

[18] Goodwin G.C., Payne R.L. (1977): *Dynamic System Identification: Experiment Design and Data Analysis*, Academic Press, New York

[19] Gorfine M., Freedman L., Shahaf G., Mehr R. (2003): *Maximum likelihood ratio test in complex models: An application to B lymphocyte development*, Bulletin of Mathematical Biology, Vol. 65, 1131-1139

[20] Gorskii V.G., Spivak S.I. (1989): *Identifiability of parameters One of the most important steps in constructing mathematical models in chemistry*, Journal of Structural Chemistry, Vol. 29, 924-930

[21] Lohmann T.W., Bock H.G., Schlöder J.P. (1992): *Numerical methods for parameter estimation and optimal experiment design in chemical reaction systems*, Industrial and Engineering Chemistry Research, Vol. 31, 54-57

[22] Majer C. (1998): *Parameterschätzung, Versuchsplanung und Trajektorienoptimierung für verfahrenstechnische Prozesse*, Fortschrittberichte VDI, Reihe 3, Nr. 538, VDI, Düsseldorf

[23] Otarod M., Happel J. (2004): *Identifiability of kinetic parameters of methanol synthesis in plug-flow tracing*, AIChE Journal, Vol. 44, 1897-1902

[24] Pohjanpalo H. (1978): *System identifiability based on power series expansion of the solution*, Mathematical Bioscience, Vol. 41, 21-33

[25] Ratkowsky D.A. (1988): *Nonlinear Regression Modeling*, Marcel Dekker, New York

[26] Ross G.J.S. (1990): *Nonlinear Estimation*, Springer, Berlin

[27] Rudolph P.E., Herrendörfer G. (1995): *Optimal experimental design and accuracy of parameter estimation for nonlinear regression models used in long-term selection*, Biomedical Journal, Vol. 37, 183-190

[28] Ryan T.P. (2007): *Modern Experimental Design*, Wiley Series in Probability and Statistics

[29] Schittkowski K. (2002): *EASY-FIT: A software system for data fitting in dynamic systems*, Structural and Multidisciplinary Optimization, Vol. 23, No. 2, 153-169

[30] Schittkowski K. (2002): *Numerical Data Fitting in Dynamical Systems-A Practical Introduction with Applications and Software*, Kluwer Academic Publishers, Dordrecht, Boston, London

[31] Schittkowski K. (2006): *NLPQLP: A Fortran implementation of a sequential quadratic programming algorithm with distributed and non-monotone line search - user's guide, version 2.2*, Report, Department of Computer Science, University of Bayreuth

[32] Schneider R., Posten C., Munack A. (1992): *Application of linear balance equations in an online observation system for fermentation processes*, Proceedings of the IFAC *Modelling and Control of Biotechnical Processes*, Boulder, Colorado, 319-322

[33] Seber G.A.F. (1977): *Linear Regression Analysis*, John Wiley, New York

[34] Seber G.A.F. (1984): *Multivariate Observations*, John Wiley, New York

[35] Seber G.A.F., Wild C.J. (1989): *Nonlinear Regression*, John Wiley, New York

[36] Tayakout-Fayolle M., Jolimaitre E., Jallut C. (2000): *Consequence of structural identifiability properties on state-model formulation for linear inverse chromatography*, Chemical Engineering Science, Vol. 55, 2945-2956

[37] Thomaseth K., Cobelli C. (1999): *Generalized sensitivity functions in physiological system identification*, Annals of Biomedical Engineering, Vol. 27, 606-616

[38] van den Bosch B.A.J. (1978): *Identification of parameters in distributed chemical reactors*, in: Distributed Parameter Systems, W.H. Ray, D.G. Lainiotis eds., Marcel Dekker, New York, Basel, 47-134

[39] van den Bosch P.P.J., van der Klauw A.C. (1994): *Modeling, Identification and Simulation of Dynamical Systems*, CRC Press, Boca Raton, Ann Arbor, London, Tokyo

[40] Versyck K.J., Claes J., Van Impe J. (1997): *Practical identification of unstructured growth kinetics by application of optimal experimental design*, Biotechnical Progress, Vol. 13, 524-531

[41] Walter E. (1982): *Identifiability of State Space Models*, Lecture Notes in Biomathematics, Vol. 46, Springer, Berlin

[42] Walter E., Pronzato L. (1997): *Identification of Parametric Models*, Springer, Paris, Milan, Barcelone

[43] Winer B.J., Brown D.R., Michels K.M. (1971): *Statistical Principles In Experimental Design*, McGraw-Hill