PARAMETER ESTIMATION OF COMPLEX SYSTEMS FROM SPARSE AND NOISY DATA

A Dissertation

by

YUNFEI CHU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2010

Major Subject: Chemical Engineering

Parameter Estimation of Complex Systems from Sparse and Noisy Data

PARAMETER ESTIMATION OF COMPLEX SYSTEMS FROM SPARSE AND NOISY DATA

A Dissertation

by

YUNFEI CHU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Juergen Hahn
Committee Members,	Aniruddha Datta
	Mahmoud El-Halwagi
	Arul Jayaraman
	Carl Laird
Head of Department,	Michael Pishko

December 2010

Major Subject: Chemical Engineering

ABSTRACT

Parameter Estimation of Complex Systems from Sparse and Noisy Data.

(December 2010)

Yunfei Chu, B.S., Tsinghua University;

M.S., Tsinghua University

Chair of Advisory Committee: Dr. Juergen Hahn

Mathematical modeling is a key component of various disciplines in science and engineering. A mathematical model which represents important behavior of a real system can be used as a substitute for the real process for many analysis and synthesis tasks. The performance of model based techniques, e.g. system analysis, computer simulation, controller design, sensor development, state filtering, product monitoring, and process optimization, is highly dependent on the quality of the model used. Therefore, it is very important to be able to develop an accurate model from available experimental data.

Parameter estimation is usually formulated as an optimization problem where the parameter estimate is computed by minimizing the discrepancy between the model prediction and the experimental data. If a simple model and a large amount of data are available then the estimation problem is frequently well-posed and a small error in data fitting automatically results in an accurate model. However, this is not always the case. If the model is complex and only sparse and noisy data are available, then the estimation problem is often ill-conditioned and good data fitting does not ensure accurate model predictions. Many challenges that can often be neglected for estimation involving simple models need to be carefully considered for estimation problems involving complex models.

To obtain a reliable and accurate estimate from sparse and noisy data, a set of techniques is developed by addressing the challenges encountered in estimation of complex models, including (1) model analysis and simplification which identifies the

important sources of uncertainty and reduces the model complexity; (2) experimental design for collecting information-rich data by setting optimal experimental conditions; (3) regularization of estimation problem which solves the ill-conditioned large-scale optimization problem by reducing the number of parameters; (4) nonlinear estimation and filtering which fits the data by various estimation and filtering algorithms; (5) model verification by applying statistical hypothesis test to the prediction error.

The developed methods are applied to different types of models ranging from models found in the process industries to biochemical networks, some of which are described by ordinary differential equations with dozens of state variables and more than a hundred parameters.

To my parents and my wife

ACKNOWLEDGEMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my family.

I owe my deepest appreciation to my advisor, Prof. Juergen Hahn for his excellent supervision, constant encouragement and persistent help in my research program. It is a great fortune for me to have the opportunity to work with Dr. Hahn and I really enjoyed the wonderful time in his group.

I am deeply grateful to my committee members, Prof. Mahmoud El-Halwagi, Prof. Arul Jayaraman, Prof. Carl Laird and Prof. Aniruddha Datta, for their guidance and support throughout the course of this research.

I would like to express my gratitude to my friends and colleagues, Zuyi Huang, Chunyan Qu, Yu Zhu, Abhay Singh, Chuili Sun, Obanifemi Aluko, Mitch Serpas, Loveleena Bansal, Jia Kang, and Shreya Maiti. They generously shared their knowledge and experience, and constantly provided valuable ideas and comments.

Finally, I would like to thank my parents, Guolin Chu and Yan Yu. They have always supported and encouraged me with their best wishes. I would also like to thank my wife, Liqing Zhang. She was always there to cheer me up and stand by me through good times and bad.

NOMENCLATURE

CRLB	Cramer-Rao Lower Bound
CSTR	Continuous Stirred-Tank Reactor
DOE	Design of Experiment
FAST	Fourier Amplitude Sensitivity Test
FIM	Fisher Information Matrix
GA	Genetic Algorithm
GSA	Global Sensitivity Analysis
KS	Kolmogorov-Smirnov
LSA	Local Sensitivity Analysis
MCMC	Markov Chain Monte Carlo
MINLP	Mixed Integer Nonlinear Programming
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
PS	Parameter Selection
SPSA	Simultaneous Perturbation Stochastic Approximation

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

Mathematical modeling is an important component of various disciplines in science and engineering. Though mathematical modeling, a real world problem can be translated into an equivalent mathematical problem, which facilitates the solution (Hangos and Cameron, 2001). In process engineering, models are widely used as replacements of the real system, in analysis, simulation, optimization, control, monitoring, and filtering (Fig. 1-1).

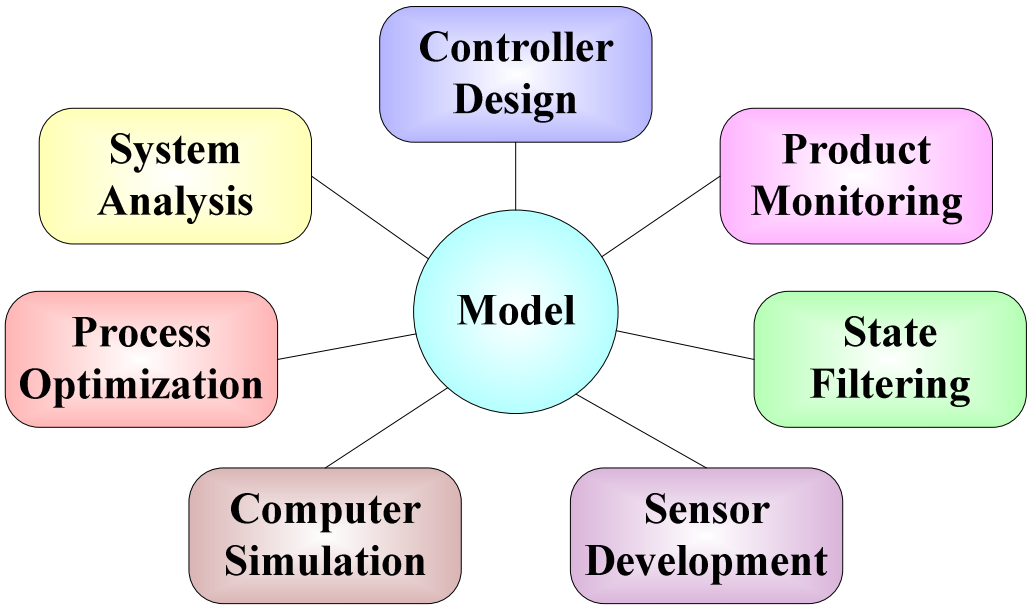


Fig. 1-1. Model-based techniques.

However, a model is just an approximation of a real system and there are differences between the model and the real system. Since a model-based technique is designed based on the model, yet the conclusions are often applied to the real system, the quality of any

---

This dissertation follows the style of Chemical Engineering Science.

technique that involves a model is closely dependent on the accuracy of the model. Therefore, model building is a core component of all model-based techniques. It should be noted that obtaining the model is the single most time consuming task in the application of model-based techniques and a majority (over 75%) of costs associated with a project can be attributed to modeling (Hjalmarsson, 2009). In fact, one of the chief barriers to the more widespread use of nonlinear models in advanced model-based techniques in the chemical/petroleum industry is the cost of model development and validation (Hussain, 1999). It is therefore important to understand what makes a modeling problem difficult and how to tackle the resulting difficulties.

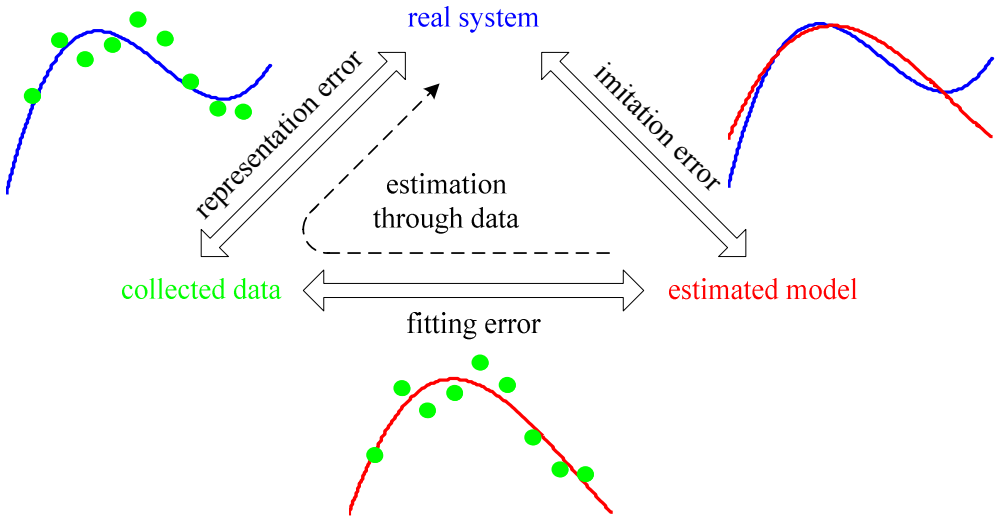


Fig. 1-2. Diagram of estimation procedure. (The real system is represented by the blue curve, the estimated model is represented by the red curve, and the collected data is represented by the greed points)

In mathematical modeling, the model structure is often determined by analysis of the underlying physical and chemical laws while the parameters are often updated from

experimental data. A diagram of an estimation procedure is shown in Fig. 1-2. Since the role of a model is to imitate the real system, it is the goal of parameter estimation is to reduce the discrepancy between the model and the real system. However, since a perfect model of any real system is never known, it is impossible to minimize the discrepancy directly. Instead, a set of data are sampled from the real system, as a representation of the real system, to evaluate the performance of predictions of the model. Then the value of model parameters is adjusted to reduce the fitting error which is a criterion to measure the discrepancy between the model and the data.

The fact that the parameter values are often determined by solving an optimization for data fitting often causes a misunderstanding that an estimation problem is just a special type of nonlinear programming problem, i.e., a least squares optimization problem. Admittedly solving the optimization problem is an important step in estimation and this problem is non-trivial for estimation of a nonlinear system. However, estimation is far more than simply an optimization problem for data fitting. It is clear from Fig. 1-2 that data fitting is just a way to achieve the actual goal to approximate the real system. A small value in the fitting error does not necessarily imply a small value in the discrepancy between the estimated model and the real system.

It should be noted that in the procedure of the data generation the representation error will occur unavoidably, which is the discrepancy between the data and the real system. Common representation errors include the measurement noise, discretization from sampling, and limited operating conditions for data collection. The data can never exactly describe the real system and the effect on the estimation results produced by the representation error in the data needs to be investigated carefully.

In the case of estimation of a simple model with a lot of data, the estimation problem is frequently well-posed and a small error in data fitting typically results in an accurate model. However, this is not always true. In the case of a complex model with sparse and noisy data, the estimation problem is often ill-conditioned and the phenomenon of over-fitting can occur where good data fitting can lead to poor model prediction capability.

Many issues are often neglected for estimation of a simple model, but need to be considered carefully for estimation of a complex model.

The first problem is the information-richness of the data. For estimation of a simple model, the amount of available data often far exceeds what is needed for estimating a few parameters. However, in the case of a complex model, the information content of each data point becomes critical: (1) a large number of parameters may need to be estimated, and (2) data generation is often difficult since a complex model implies a complex real system where conduction of an experiment can be expensive and time-consuming.

The second problem is related to uncertainty in the model. Since some discrepancy between the model and the real system is inevitable, uncertainty is an inherent characteristic of a model. In a complex model, there are many sources of the uncertainty, however, not all of them have equal influence on the behavior of interest of the model. A frequently asked question in this situation is "What uncertain sources really matter for a given property of the model?" Since it is very difficult or even impossible to reduce all uncertain sources in a complex system through estimation from data, it is helpful in practice to identify the important ones.

The third problem is the complexity of the model. Analysis of a complex model is difficult and there is usually no closed-form expression of the model predictions. The result of this is that simulation is the only way to investigate the model, however, simulation of a complex model is time consuming, especially if the model has to be solved repeatedly. This hinders the applications of a complex model, e.g. in the iterative optimization procedure for parameter estimation. Fortunately, complex models often contain a considerable degree of redundancy. It is possible and desirable to simplify a complex model and reduce it to a simple one, which can be more handily used for analysis and simulation.

The forth problem is that parameter estimation problems of complex models can be ill-conditioned. If highly correlated model parameters need to be estimated from noisy

data then not all parameters will be identifiable in practice. A regularization mechanism is required to as part of the estimation procedure to ensure reliable parameter estimates.

The fifth problem related to validation of the estimated model. For example, an ill-conditioned estimation problem may often result in the situation where a small fitting error does not necessarily lead to a small prediction error. In those cases it is insufficient to just check the fitting error and more sophisticated validation approaches are required.

A sixth problem results from the choice of estimation or filtering method. The question that needs to be answered is which methods are more appropriate for estimation of a complex model and how to integrate them with other procedures, e.g. regularization methods, in parameter estimation.

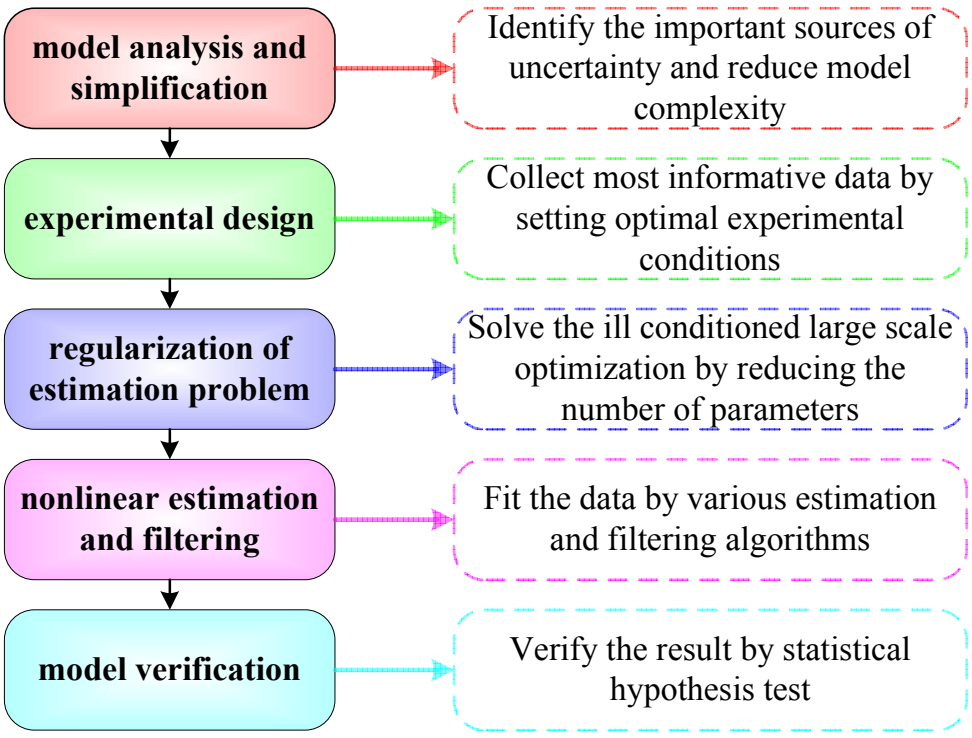


Fig. 1-3. Outline of research work in estimation of complex models.

To address the challenges encountered in estimation of complex models, a set of methodologies from model analysis and simplification to data fitting is developed in this dissertation. An overview over these techniques is shown in Fig. 1-3.

Stage 1: Model analysis and simplification

The purpose of this stage is to gain insight into a model by discovering the key factors which should be focused on. A large-scale model consists of a large number of parameters, however, the system behavior is often mainly determined by just a few of them. Sensitivity analysis is a powerful tool to identify these important components. In this work several new techniques for global sensitivity analysis were developed, which overcome some drawbacks of commonly used techniques. After identification of important components, the complex model can be reduced to a simple one facilitating the following analysis.

Stage 2: Optimal experimental design

This stage collects data for parameter estimation or model identification by adjusting the experimental conditions. Procedures include input design, sampling point selection, and sensor location. The main difficulty of experimental design is the inevitable parameter uncertainty since experimental design is always applied before parameter estimation can be performed. To deal with this problem, several robust strategies are introduced in this work.

Stage 3: Regularization of estimation problem

This stage focuses on solving the ill-conditioned problem of parameter estimation of complex systems. Parameter set selection is introduced as a technique to regularize the ill-conditioned problem and to reduce the effect of noise on the estimated parameter value. Additionally, parameter set selection serves as a simplification produce for the optimization problem resulting from parameter estimation. Procedures are presented in

this work to solve the resulting combinatorial selection problem under the effect of parameter uncertainty.

Stage 4: Nonlinear estimation and filtering

This stage fits the model parameters to experimental data. The common least squares estimation and maximum likelihood estimation techniques can be applied. The parameters can also be augmented as states and methods of nonlinear filtering can be applied.

Stage 5: Verification

It is determined at this stage if the experimental design and estimation results are sufficiently accurate. The fitting error is commonly used as a criterion, however, it may be insufficient for drawing definite conclusions in some cases. Statistical tests can be used to provide reasonable results.

The outline of this dissertation is as follows: Section 2 presents a comprehensive review of existing techniques involved in parameter estimation of complex systems. A comparative study of different sensitivity analysis techniques is presented in Section 3. Section 4 presents a robust parameter selection method for nonlinear dynamic systems and the integration of parameter selection with experimental design is presented in Section 5. An efficient algorithm via parameter clustering to solve the combinatorial selection problem is presented in Section 6 and in Section 7 a method to improve the prediction accuracy is discussed. A new robust method for experimental design is presented in Section 8. Conclusions are given in Section 9.

2. LITERATURE REVIEW

2.1 Sensitivity analysis techniques

Sensitivity analysis is a powerful tool to study how model parameter variations can qualitatively or quantitatively influence model behavior. The analysis can improve the understanding of the complex model as it can be used to rank the contribution of individual parts of the model to the feature of interest. A variety of approaches to sensitivity analysis have been developed (Borgonovo, 2006; Cacuci and Ionescu-Bujor , 2004; Frey and Patil, 2002; Hamby, 1994; Helton, 1993; Iman and Helton, 1988; Ionescu-Bujor and Cacuci , 2004; Klepper, 1997; Marino et al., 2008; Rabitz, 1983; Rubinstein, 1989; Saltelli et al., 2000, 2004, 2005, 2006, 2008; Turanyi, 1990; Wagner, 1995; Wallace, 2000). Four commonly used methods are investigated in this section: (i) differential analysis (Dickinson and Gelinas, 1976; Frank, 1978; Hwang et al., 1978), which approximates the model by the first-order Taylor series; (ii) the Morris method (Morris, 1991) which calculates the average sensitivity over an interval by computing sensitivity at several points in the parameter space; (iii) a sampling-based method (Hornberger and Spear, 1981; Iman et al., 1981), which computes a probabilistic-based mapping from the uncertain input to the output; and (iv) the variance-based method (Atherton et al., 1975; Cukier et al., 1973), which is based on the contributions of individual variables to the variance of the model output.

Differential analysis is a local method while the other three are global methods. Local sensitivity analysis perturbs one parameter at a time in a small range around the nominal values. The main drawback of local techniques is that the sensitivity value is generally dependent on the parameter value which is not precisely known prior to parameter estimation. Global sensitivity analysis simultaneously varies several parameters, often over a large range of the parameter values. As a result, global sensitivity analysis techniques are able to provide a more accurate description of the sensitivity than local analysis if the uncertainty of the parameter values is significant.

Sensitivity analysis has become a key step for mathematical modeling and analysis. This is also reflected by the wide range of its applications in chemical or biochemical engineering to identify the important parameters (Bentele et al., 2004; Cho et al., 2003; Daescu et al., 2003; Dunker et al., 1984; Hu and Yuan, 2006; Ingalls and Sauro, 2003; Perumal et al., 2009; Rabitz, 1981, 1987; Sandu et al., 2003; Yue et al., 2006; Zi et al., 2005). Sensitivity analysis also plays an important role in identifiability test (Brun, et al., 2001; Vajda and Rabitz, 1994; Vanrolleghem et al., 1995; Sun et al., 2001; Yeh, 1986), parameter selection (Brockmann et al., 2008; Brun et al. 2002; Machado et al., 2009; Yao et al., 2003; Weijers and Vanrolleghem, 1997), experimental design (Bardow, 2008; Buzzi-Ferraris and Forzattia, 1983; Franceschini and Macchietto, 2008; Hosten, 1974; Schittkowski, 2007), model reduction (Hay et al., 2009; Degenring et al., 2004; Ho, 2008; Liu et al., 2005; Sun and Hahn, 2006; Vajda et al., 1985), sensor network design (Cobb and Liebst, 1997; Stanimirovic et al., 2008; Zamprogna et al., 2005), state filtering (Huang et al., 2003; Jwo and Cho, 2007; Sorensen et al., 2006), controller design (Higham et al., 2004; Nikandrov and Swartz, 2009; Nagy and Braatz, 2003; Oniki, 1973; Sokolowski, 1987), and process optimization (Balsa-Canto et al., 2001; Castillo et al., 2006; Ozyurt and Barton, 2005).

A general form of the nonlinear dynamic system on which the analysis is performed on is assumed to be

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}) \text{ with } \mathbf{x}(0) = \mathbf{x}_0 \tag{2-1}$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}) \tag{2-2}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is the state vector, $\mathbf{u} \in \mathbb{R}^{n_u}$ is the input vector, $\mathbf{y} \in \mathbb{R}^{n_y}$ is the output vector and $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ is the parameter vector.

Differential analysis

Differential analysis is the most widely used method for sensitivity analysis. The technique approximates the output function by the truncated Taylor series

$$\mathbf{y}\left(t,\boldsymbol{\theta}+\Delta\boldsymbol{\theta}\right)\approx\mathbf{y}\left(t,\boldsymbol{\theta}\right)+\left.\frac{\partial\mathbf{y}}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\right|_{\boldsymbol{\theta}}\Delta\boldsymbol{\theta}+\frac{1}{2}\left(\mathbf{I}_{n_y}\otimes\Delta\boldsymbol{\theta}^{\mathrm{T}}\right)\left.\mathbf{H}_y\right|_{\boldsymbol{\theta}}\Delta\boldsymbol{\theta}\,, \tag{2-3}$$

where $\partial\mathbf{y}/\partial\boldsymbol{\theta}^{\mathrm{T}}\in\mathbb{R}^{n_y,n_\theta}$ is the Jacobian matrix of $\mathbf{y}$

$$\frac{\partial\mathbf{y}}{\partial\boldsymbol{\theta}^{\mathrm{T}}}=\begin{bmatrix}\dfrac{\partial y_1}{\partial\theta_1} & \cdots & \dfrac{\partial y_1}{\partial\theta_{n_\theta}}\\ \vdots & \ddots & \vdots\\ \dfrac{\partial y_{n_y}}{\partial\theta_1} & \cdots & \dfrac{\partial y_{n_y}}{\partial\theta_{n_\theta}}\end{bmatrix}, \tag{2-4}$$

and $\mathbf{H}_y\in\mathbb{R}^{n_y n_\theta,n_\theta}$ is the Hessian matrix

$$\mathbf{H}_y=\begin{bmatrix}\dfrac{\partial^2 y_1}{\partial\boldsymbol{\theta}^{\mathrm{T}}\partial\boldsymbol{\theta}}\\ \dfrac{\partial^2 y_2}{\partial\boldsymbol{\theta}^{\mathrm{T}}\partial\boldsymbol{\theta}}\\ \vdots\\ \dfrac{\partial^2 y_{n_y}}{\partial\boldsymbol{\theta}^{\mathrm{T}}\partial\boldsymbol{\theta}}\end{bmatrix}. \tag{2-5}$$

Both Jacobian matrix and Hessian matrix should be evaluated at the parameter point at which the Taylor series is expended and the approximation can be applied only when the parameter variation $\Delta\boldsymbol{\theta}$ is small. So this is a local technique. The Jacobian matrix and the Hessian matrix indicate the parameter effect on the output and they are defined as the first order local sensitivity and the second order local sensitivity respectively.

Besides the parameter sensitivity, the sensitivity of the output with respect to initial state value and the input value can also be calculated. The time function of the input $\mathbf{u}(t)$ is frequently parameterized by a finite set of parameters, denoted by $\mathbf{u}$. To simplify the expression, the initial states and the parameterized inputs are concatenated into an augmented parameter vector $\boldsymbol{\psi}\in\mathbb{R}^{n_\theta+n_x+n_u}$

$$\boldsymbol{\psi} = \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{x}_0 \\ \mathbf{u} \end{bmatrix}. \tag{2-6}$$

Various methods are developed to calculate the local sensitivity for systems described by Eq. (2-1) and Eq. (2-2). Among them, the method of direct differentiation is commonly used. The sensitivity values of the state variables $\mathbf{x}(t)$ with respect to a parameter ψ_i can be calculated by differentiating both sides of the state equations (2-1) to obtain

$$\frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \psi_i} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \psi_i} + \frac{\partial \mathbf{f}}{\partial \psi_i}, \quad i = 1 \cdots n_\theta + n_x + n_u . \tag{2-7}$$

If ψ_i is an initial value of a state, then the second term of the right hand side of Eq. (2-7) is zero. The initial conditions of the differential equations are given by

$$\left.\frac{\partial \mathbf{x}}{\partial \psi_i}\right|_{t=0} = \begin{cases} \mathbf{0}, & \text{if } \psi_i \in \boldsymbol{\theta} \text{ or } \psi_i \in \mathbf{u} \\ \mathbf{e}_j, & \text{if } \psi_i = \mathbf{x}_0(j) \end{cases}, \tag{2-8}$$

where $\mathbf{e}_j \in \mathbb{R}^{n_x}$ is a vector with entries of 1 on its j-th element and entries of 0 on all other elements. By solving the sensitivity equations (2-7) and the state equations (2-1) simultaneously the sensitivity values are calculated along the state/output trajectories.

The second order sensitivities can be calculated by differentiating both sides of Eq. (2-7)

$$\begin{aligned} \frac{d}{dt}\frac{\partial^2 \mathbf{x}}{\partial \psi_i \partial \psi_j} = {} & \frac{\partial \mathbf{f}}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial^2 \mathbf{x}}{\partial \psi_i \partial \psi_j} + \frac{\partial^2 \mathbf{f}}{\partial \psi_i \partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \psi_j} + \frac{\partial^2 \mathbf{f}}{\partial \mathbf{x}^{\mathrm{T}} \partial \psi_j}\frac{\partial \mathbf{x}}{\partial \psi_i} + \frac{\partial^2 \mathbf{f}}{\partial \psi_i \partial \psi_j} \\ & + (\mathbf{I}_N \otimes \frac{\partial \mathbf{x}^{\mathrm{T}}}{\partial \psi_j})\mathbf{H}_f \frac{\partial \mathbf{x}}{\partial \psi_i}, \quad i, j = 1 \cdots n_\theta + n_x + n_u \end{aligned} \tag{2-9}$$

where $\mathbf{H}_f \in \mathbb{R}^{n_x n_x, n_x}$ is the Hessian matrix of $\mathbf{f}$. To calculate the second order sensitivities, Eq. (2-9), Eq. (2-7) and Eq. (2-1) are solved simultaneously.

Based on the sensitivity of the state vector, the sensitivity of the output can be calculated as

$$\frac{\partial \mathbf{y}}{\partial \psi_i} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}^{\mathrm{T}}} \frac{\partial \mathbf{x}}{\partial \psi_i} + \frac{\partial \mathbf{g}}{\partial \psi_i}, \quad i = 1 \cdots n_\theta + n_x + n_u \tag{2-10}$$

and

$$\begin{aligned} \frac{\partial^2 \mathbf{y}}{\partial \psi_i \partial \psi_j} = & \frac{\partial \mathbf{g}}{\partial \mathbf{x}^{\mathrm{T}}} \frac{\partial^2 \mathbf{x}}{\partial \psi_i \partial \psi_j} + \frac{\partial^2 \mathbf{g}}{\partial \psi_i \partial \mathbf{x}^{\mathrm{T}}} \frac{\partial \mathbf{x}}{\partial \psi_j} + \frac{\partial^2 \mathbf{g}}{\partial \mathbf{x}^{\mathrm{T}} \partial \psi_j} \frac{\partial \mathbf{x}}{\partial \psi_i} + \frac{\partial^2 \mathbf{g}}{\partial \psi_i \partial \psi_j} \\ & + (\mathbf{I}_N \otimes \frac{\partial \mathbf{x}^{\mathrm{T}}}{\partial \psi_j}) \mathbf{H}_g \frac{\partial \mathbf{x}}{\partial \psi_i}, \quad i, j = 1 \cdots n_\theta + n_x + n_u \end{aligned} \tag{2-11}$$

where $\mathbf{H}_g \in \mathbb{R}^{n_x n_x, n_x}$ the Hessian matrix of $\mathbf{g}$.

Various algorithms and software packages exist for efficiently solving the differential equations for the sensitivity calculations (Byrnea and Hindmarsh, 1987), e.g., VODE (Brown et al., 1989), DASPK (Brown et al., 1994) or SUNDIALS (Hindmarsh et al., 2005). In this dissertation, the Matlab ODE solver is used as the sensitivity calculations can easily be incorporated with the other calculations.

Morris method

Unlike the differential analysis, the techniques for global sensitivity analysis do not employ the structure of the dynamic model described by Eq. (2-1) and Eq. (2-2). The model is only used to simulate the output value as a black-box model does. The Morris method (Morris, 1991) is a common screening method and it calculates the sensitivity measure by perturbing one parameter at a time

$$\mathbf{d}_i(t) = \frac{\mathbf{y}\left(t, \theta_1, \cdots, \theta_i \pm \Delta_i, \cdots, \theta_m\right) - \mathbf{y}\left(t, \theta_1, \cdots, \theta_i, \cdots, \theta_m\right)}{\Delta_i} \tag{2-12}$$

by a certain amount Δ_i , where $\mathbf{d}_i(t)$ is called the elementary effect of the *i*-th parameter at time *t*.

Similarly to the differential analysis techniques, the elementary effects are also dependent on the nominal value of the parameters. However, the elementary effects are computed as an average over a number of points in parameter space and will, therefore, reflect an average of the sensitivity over a region of the parameter space. The mean of the elementary effect is defined as the sensitivity measure

$$\frac{1}{r}\sum_{j}^{r}\mathbf{d}_{ij}(t) \tag{2-13}$$

where $\mathbf{d}_{ij}(t)$ is the elementary effect of the i-th parameter at the j-th sampling point and r is the number of sampling points used (Cropp and Braddock, 2002; Morris, 1991; Zador 2005). This measure of sensitivity is also commonly normalized to ensure that the use of different units does not affect sensitivity analysis results.

Sampling-based approach (Kolmogorov-Smirnov statistic)

Sampling-based approaches for sensitivity analysis are very popular because of their conceptual simplicity and ease of implementation. These methods characterize the uncertainty by assigning a probability distribution

$$p(\boldsymbol{\theta}) = p(\theta_1)\cdots p(\theta_{n_\theta}) \tag{2-14}$$

to every parameter θ_i of the parameter vector $\boldsymbol{\theta}$. The distribution function represents some of the knowledge about the uncertainty of the parameters. While a normal distribution can be a good representation for many systems where mean and variance are fairly well-known, a uniform distribution is often used if only the uncertainty range of the parameters is known (Saltelli et al., 2005, 2008).

The first step of sampling-based methods is to generate a sample set of parameter vectors from the distribution of the individual parameters. Three sampling procedures are widely used (Helton and Davis, 2002, 2003; McKay et al., 2000): random sampling, stratified sampling, and Latin hypercube sampling. The random sampling has the most obvious statistical meaning but the Latin hypercube sampling is more efficient for large number of parameters and is used in this work.

The next step is to evaluate the individual contribution of each parameter. Various methods can be used to calculate the sensitivity, such as regression analysis, correlation analysis, stepwise regression and rank transformation (Helton et al., 2005, 2006; Saltelli et al., 2000). In this section the Kolmogorov-Smirnov statistic is used as the sensitivity measure. The values of the output are recorded for simulations with varying parameter values. The difference between the values of the outputs for these different parameter

values and a nominal value of the output is computed and compared against a threshold. If the value is less than the threshold than the parameter value is classified as part of an "acceptable set", however, if the difference is larger than the threshold then the parameter value is assumed to belong to an "unacceptable set". The difference between the cumulative frequency distributions of the two sets is defined as the sensitivity measure of the output with respect to the parameter. The greater the difference between the two cumulative functions, the more sensitive is the output with respect to the parameter. The sensitivity measure of the sampling-based approach is

$$KS = \sup_x \left| F_{ai}(t,x) - F_{ui}(t,x) \right| \tag{2-15}$$

where F_{ai} and F_{ui} are the cumulative functions respectively corresponding to the 'acceptable set' and the 'unacceptable set' of the parameter θ_i. The greater the difference between the two cumulative functions, the more sensitive is the output with respect to the parameter.

Variance based method

Like the method based on the Kolmogorov-Smirnov statistic, the variance-based sensitivity characterizes the prior information of the parameter uncertainty by a probability density function (2.14). The conditional variance characterizes the individual contribution of a parameter to the total variance of the output, which is calculated by

$$\begin{aligned} &\mathrm{Var}\left[\mathrm{E}\left[y_j(t,\boldsymbol{\theta}) \mid \theta_i\right]\right] \\ &= \mathrm{E}\left[\left(\mathrm{E}\left[y_j(t,\boldsymbol{\theta}) \mid \theta_i\right] - \mathrm{E}\left[y_j(t,\boldsymbol{\theta})\right]\right)^2\right] \\ &= \int\left(\int\cdots\int y_j(t,\boldsymbol{\theta}) \prod_{k\neq i} p_k(\theta_k) \prod_{k\neq i} d\theta_k - \int\cdots\int y_j(t,\boldsymbol{\theta}) \prod_{k} p_k(\theta_k) \prod_{k} d\theta_k\right)^2 p_i(\theta_i)\, d\theta_i. \end{aligned} \tag{2-16}$$

There are two terms contained in the bracket in Eq. (2-16). The first term is the conditional mean of the output according to a particular parameter θ_i and the second one is the mean of the output over all parameters. The global sensitivity is often defined as the conditional variance divided by the total variance of the output (Saltelli et al., 2008)

$$\frac{\mathrm{Var}\left[\mathrm{E}\left[y_j(t,\boldsymbol{\theta})\,|\,\theta_i\right]\right]}{\mathrm{Var}\left[y_j(t,\boldsymbol{\theta})\right]} \tag{2-17}$$

or the normalized conditional variance (Chu et al., 2007)

$$\sqrt{\frac{\mathrm{Var}\left[\mathrm{E}\left[y_j(t,\boldsymbol{\theta})\,|\,\theta_i\right]\right]}{\mathrm{Var}\left[\theta_i\right]}}\,. \tag{2-18}$$

The advantage of the normalized conditional variance is that the global sensitivity is in some sense comparable to the local sensitivity as both have the same unit. Computation of the conditional variance is not trivial and various approaches for its computation have been presented, including the regression method (McKay, 1997; McKay et al., 1999), Sobol's method (Homma and Saltelli, 1996; Saltelli, 2002; Sobol, 2001), Bayesian approach (Oakley and O'Hagan, 2004; Zhang et al., 2009), high dimensional model representation (HDMR) (Li et al., 2002; Rabitz and Alis, 1999; Ziehn and Tomlin, 2009), state dependent parameter (SDP) (Ratto et al., 2007), polynomial chaos expansions (PCE) (Sudret, 2008), Fourier amplitude sensitivity test (FAST) (Cukier et al., 1973, 1975, 1978; McRae et al., 1982; Schaibly and Shuler, 1973), and extensions of FAST (Saltelli et al., 1999, 2010).

If the model is linear and parameters are independent

$$y_j(t,\boldsymbol{\theta}) = \sum_k a_{jk}(t)\,\theta_k \tag{2-19}$$

then the conditional mean of the parameter θ_i is given by

$$\mathrm{E}\left[y_j(t,\boldsymbol{\theta})\,|\,\theta_i\right] = a_{ji}(t)\,\theta_i + \sum_{k\neq i} a_{jk}(t)\,\mathrm{E}\left[\theta_k\right] \tag{2-20}$$

and the mean is

$$\mathrm{E}\left[y_j(t,\boldsymbol{\theta})\right] = \sum_k a_{jk}(t)\,\mathrm{E}\left[\theta_k\right]. \tag{2-21}$$

The conditional variance from Eq. (2-16) results in

$$\begin{aligned}&\mathrm{Var}\left[\mathrm{E}\left[y_j(t,\boldsymbol{\theta})\,|\,\theta_i\right]\right]\\&=\mathrm{E}\left[a_{ji}(t)^2\left(\theta_i-\mathrm{E}[\theta_i]\right)^2\right]\\&=a_{ji}(t)^2\,\mathrm{Var}[\theta_i].\end{aligned}\tag{2-22}$$

In this case the global sensitivity given by Eq. (2-18) matches the magnitude of the local sensitivity.

$$\sqrt{\frac{\mathrm{Var}\left[\mathrm{E}\left[y_j(t,\boldsymbol{\theta})\,|\,\theta_i\right]\right]}{\mathrm{Var}[\theta_i]}}=\left|a_{ji}(t)\right|=\left|\frac{\partial y_j(t,\boldsymbol{\theta})}{\partial\theta_i}\right|.\tag{2-23}$$

If the model is nonlinear and the parameter uncertainty is small, then the global sensitivity computed by these means returns results that approximate those computed by the absolute value of the local sensitivity analysis (Chu et al., 2007).

2.2. Optimal experimental design

Experimental design seeks to determine the experimental conditions to collect informative data that will improve the precision of estimated parameters. A large amount of literature exists on design of experiments, including several textbooks (e.g., Atkinson et al., 2007; Chaudhuri and Mykland, 1993; Emery and Nenarokomov, 1998; Forssell and Ljung, 2000; Franceschini and Macchietto, 2008; Hill, 1978; Kiefer and Wolfowitz, 1959; Ljung, 1999; Pronzato, 2008; Walter and Pronzato, 1990; Whittle, 1973).

For dynamic systems, the task of experimental design includes choosing input and output ports (Alonso et al., 2004; Harris et al., 1980; Keller and Bonvin, 1992; Singh and Hahn, 2005; 2006), designing the profile of the input signal (Levadi, 1966; Mehra, 1974; Hildebrand and Gevers, 2003), selection sampling points from output trajectory (D'Argenio, 1981; Knopman and Voss, 1987; Kutalik et al., 2004), and setting the initial value of the state variables (Saccomani et al., 2003). Each of these variables of a experiment has a significant bearing upon the information contained in the data for estimation.

Optimal experimental design depends upon the assumed model including the model structure and the assumptions about the error distribution. Assume a regression model for design is given by

$$\tilde{\mathbf{y}} = \mathbf{g}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{2-24}$$

where $\tilde{\mathbf{y}} = \left[\tilde{y}(t_1),\ \cdots,\ \tilde{y}(t_{n_t}) \right]^{\mathrm{T}}$ is the measured output, $\mathbf{g}(\boldsymbol{\theta}) = \left[g(t_1, \boldsymbol{\theta}), \cdots, g(t_{n_t}, \boldsymbol{\theta}) \right]^{\mathrm{T}}$ is the predicted value and $\boldsymbol{\varepsilon} = \left[\varepsilon(t_1),\ \cdots,\ \varepsilon(t_{n_t}) \right]^{\mathrm{T}}$ is the measurement noise. For dynamic systems the regression model is defined implicitly by a set of differential equations describing the system. If a system is nonlinear then an analytical expression of the regression model rarely exists.

Information about the noise is often required for experimental design, where it is a common assumption that noise is normally distributed with zero mean and a covariance matrix of $\boldsymbol{\Sigma}$. To simplify the notation in the following, a Cholesky decomposition of the inverse of the covariance matrix can be performed, i.e. $\boldsymbol{\Sigma}^{-1} = \mathbf{C}^{\mathrm{T}}\mathbf{C}$. A new regression model can then be obtained by multiplying both sides of the regression model shown in Eq. (2-24) by the matrix $\mathbf{C}$. The noise vector of this new model is $\mathbf{C}\boldsymbol{\varepsilon}$ which has a covariance matrix equal to the identity matrix. Due to this pre-processing, the covariance matrix of the noise can be assumed to be the identity matrix

$$\boldsymbol{\Sigma} = \mathbf{I} \tag{2-25}$$

without loss of generality.

If the covariance matrix of the noise is unknown it is possible to augment the parameter vector to include elements of the covariance matrix and estimate the covariance matrix simultaneously with other parameters. However, this approach further complicates the parameter estimation and experimental design and it is not uncommon to assume that one knows the covariance matrix of the noise in experimental design.

To measure the quality of a designed experiment a criterion is required. One criterion is the covariance matrix of the estimated parameters. It is possible to directly generate a distribution of estimated values of the parameters by using a Monte Carlo method. In this case, the experimental design can be performed using multiple sampling points of

the estimated parameter values (Asprey and Macchietto, 2000; Balsa-Canto et al., 2007; Hengl ea al., 2007; Joshi et al., 2006). However, these approaches can be computationally expensive since the covariance matrix can only be computed after the parameters have been estimated and it is also affected by the estimation algorithm. Alternatively, the Fisher information matrix (FIM) can be used as the inverse of the FIM provides the Cramer-Rao lower bound for the covariance matrix (Walter and Pronzato, 1990). It is desirable to minimize a criterion involving the inverse of the FIM or equivalently to maximize a criterion of the FIM in order to reduce a measure of the covariance matrix.

The outputs of a nonlinear dynamic system are affected by process and measurement noise and in general no closed-form solution of the FIM exists. However, for the purpose of simplicity only the measurement noise is commonly considered.

In the case of additive Gaussian noise the FIM is very closely related to the parameter sensitivity matrix. After the pre-processing procedure to whiten the noise, the measurement noise can be assumed to be normally distributed with the zero mean and the unit covariance matrix. As a result the measurements are also normally distributed given by

$$p\left(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}\right) = \left(2\pi\right)^{-n_y/2} \exp\left[-1/2\left(\tilde{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta})\right)^{\mathrm{T}}\left(\tilde{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta})\right)\right] \tag{2-26}$$

and the Fisher information matrix, **F**, is given by

$$\begin{aligned}\mathbf{F}\left(\boldsymbol{\theta}\right) &= \mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln p\left(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}\right) \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \ln p\left(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}\right)\right] \\ &= \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right)^{\mathrm{T}} \mathrm{E}\left[\left(\tilde{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta})\right)\left(\tilde{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta})\right)^{\mathrm{T}}\right] \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \\ &= \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right)^{\mathrm{T}} \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}.\end{aligned} \tag{2-27}$$

The FIM becomes the cross product of the local sensitivity matrix defined as

$$\mathbf{S} = \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}. \tag{2-28}$$

If the measurement noise is not normally distributed, calculation of FIM becomes more complicated (Das et al., 2010; Spall, 2005).

To reduce variations in the estimated parameter values, it is desired to maximize some measure of the Fisher information matrix to reduce its inverse. Such a measure can consist of a real function that operates on the Fisher information matrix, e.g., the experimental optimality criteria (Kiefer, 1959, 1974; Steinberg and Hunter, 1984). The most popular experimental optimality criterion is the *D*-optimality criterion (deAguiar et al., 1995; St. John and Draper, 1975; Wynn, 1972) which maximizes the logarithm of the determinant of the Fisher information matrix:

$$\varphi_D{}^* = \max \varphi_D(\mathbf{F}) = \max \det(\mathbf{F}). \tag{2-29}$$

This criterion minimizes the volume of the confidence ellipsoid with an arbitrary fixed confidence level for a least square estimator. Other common criteria include the *E*-optimality which maximizes the smallest eigenvalue of the Fisher information matrix

$$\varphi_E{}^* = \max \varphi_E(\mathbf{F}) = \max \lambda_{\min}(\mathbf{F}), \tag{2-30}$$

the modified *E*-optimality criterion which minimizes the condition number

$$\varphi_{ME}{}^* = \min \varphi_{ME}(\mathbf{F}) = \min \kappa(\mathbf{F}), \tag{2-31}$$

and the *A*-optimality criterion which minimizes the trace of the invese

$$\varphi_A{}^* = \min \varphi_A(\mathbf{F}) = \min \operatorname{tr}\left(\mathbf{F}^{-1}\right). \tag{2-32}$$

These criteria values will be far from the optimum if the Fisher information matrix is ill-conditioned. The criterion functions evaluate a design from different perspectives and to combine all aspects of interest a compound design criterion can be formulated with appropriate weights on each criterion (Atkinson and Bogacka, 1997, 2002; Cook and Wong, 1994).

While local sensitivity analysis can be applied to nonlinear models, there are several points that need to be carefully considered. One is that the results returned by local sensitivity analysis of a nonlinear system depend upon the values of the parameters that one wants to estimate. Obviously these values are not exactly known prior to estimation. The effect of the parameter values on the sensitivity values and, accordingly, on the

experimental design criterion represents one of the main problems associated with experimental design of nonlinear systems.

Several approaches have been developed to deal with this dependency. The most widely-used one is local design (Box and Lucas, 1959; Chernoff, 1953) which assumes that the true parameter values are close to the nominal values. If this is the case then the sensitivity vectors evaluated at the nominal values of the parameters can be used to design an experiment. However, this approach neglects the parameter uncertainty. Another approach is sequential design (Box and Hunter, 1962; Ford and Silvey, 1980; Wynn, 1970) which iterates between local design and parameter estimation. Using this technique, an experiment is designed based on the sensitivity evaluated at the previously estimated parameter values; the parameter values are then re-estimated based upon data generated from the designed experiment. The newly estimated parameter values are used for experimental design for the next iteration. The main drawback of this technique is that iterating between experimental design and parameter estimation may not result in a small number of experiments that need to be performed. This drawback is a significant one as reducing the experimental effort is one of the driving factors behind performing experimental design. Curvature based methods (Bates and Watts, 1980; Benabbas et al., 2005) which calculate the higher order sensitivity provide another direction to deal with the uncertainty in design of experiments.

Robust design (Asprey and Macchietto, 2002; Box and Draper, 1975; Dette et al., 2005) is an alternative to the aforementioned experimental design methods. Robust design evaluates the sensitivity not only at one point in the parameter space, but instead at many individual points. Approaches for robust design include, the min-max method (Hoel, 1965; Pronzato and Walter, 1988; Rojas et al., 2007; Goodwin et al. 2008) and the Bayesian method (Pronzato and Walter, 1985; Chaloner, 1993; Chaloner and Verdinelli, 1995). However, these robust methods are computationally expensive due to the evaluation of the sensitivity over a range of possible parameter values.

For a dynamic system the experimental design can be formulated as an optimal control problem to optimize a criterion function and various algorithms and software

have been developed (e.g., Banga et al., 2002; Bauer et al., 2000; Cook and Nachtsheim, 1980; Hamada et al., 2001; Korkel et al., 2004; Lohmann et al., 1992; Rasch et al. 2009; Schittkowski, 2009). However, most algorithms are applied to solve the local design problem. Due to the complexity of the optimization method heuristic methods are also helpful (Bohachevsky et al., 1986; Heredia-Langner et al., 2003; Lejeune, 2003).

2.3. Identifiability test and parameter selection

Successful parameter estimation depends, among other things, on parameter identifiability. Parameter identifiability can be determined either analytically or numerically (Grewal and Glover, 1976; Ljung, 1999; Walter 1987; Walter and Pronzato, 1990). Analytical identifiabilty investigates uniqueness of the solution derived from parameter estimation while numerical identifiability focuses on the robustness of the solution to the noise in the data. Additionally, analytical identifiabilty can be either global or local. While global identifiability includes local identifiability as a special case, it is significantly more difficult to determine global identifiability as approaches based upon differential algebra (Audoly et al., 1998, 2001; Ljung and Glad, 1994; Margaria et al., 2001), Taylor series approximations and similarity transformations (Chappell et al., 1990; Cobelli and DiStefano, 1980; Pohjanpalo, 1978; Vajda et al., 1989) are restricted to small systems. Local identifiability on the other hand is relatively straightforward to test by computing the rank of the parameter output sensitivity matrix. The state vector can be augmented with the parameters and observability of this augmented vector can be performed as it also includes parameter identifiability (Hermann and Krener, 1977). However the observability test is not trivial to perform and interpret for nonlinear systems. As the techniques introduced in this work are based upon these concepts, the definitions of identifiability (Jacquez and Perry, 1990; Rothenbe, 1971) are briefly reviewed next.

Definition 2.1: A parameter point $\boldsymbol{\theta}_0$ is said to be locally identifiable if there exists an open neighborhood of $\boldsymbol{\theta}_0$ containing no other $\boldsymbol{\theta}$ which produces the identical output $\mathbf{y}$.

Condition 2.1: Let $\boldsymbol{\theta}_0$ be a parameter point and the sensitivity matrix $\mathbf{S}(\boldsymbol{\theta}) = \partial \mathbf{y} / \partial \boldsymbol{\theta}^{\mathrm{T}}$ has constant rank in a neighborhood of $\boldsymbol{\theta}_0$. Then $\boldsymbol{\theta}_0$ is locally identifiable if and only if $\mathbf{S}(\boldsymbol{\theta}_0)$ has the full column rank.

It should be noted that it is a necessary condition that the rank of the sensitivity matrix does not change. If this condition is removed then the full column rank of the sensitivity matrix is just a sufficient condition for local identifiability, i.e., a rank-deficient sensitivity matrix does not imply that the parameters are not locally identifiable. The condition of constant rank has to be checked analytically and evaluating this condition for one nominal value of the parameters is usually not sufficient.

Analytical identifiability guarantees the existence of a unique solution in at least a small neighborhood of the nominal point. However, analytically identifiable of parameters does not guarantee accurate estimation in practice. If the sensitivity matrix has the full column rank but is ill-conditioned, then noise in the data will result in large variations of the estimated parameter values. While the parameters in this case are identifiable based upon the analytical conditions, it is questionable that accurate parameter estimates can be obtained in practice and it can be said that the system is not numerically identifiable.

Numerical identifiability (Jacquez and Greif, 1985) can be determined from the parameter covariance matrix. If the entries in the covariance are large then the parameters are not numerical identifiable. However, the covariance matrix can only be computed after the parameters have been estimated and it is affected by the choice of the estimation algorithm. As an alternative, the Fisher information matrix can be used as its inverse provides the Cramer-Rao lower bound for the covariance matrix (Ljung, 1999; Walter and Pronzato, 1990).

If some of the parameters are not numerically identifiable then a set of identifiable parameters are often selected for estimation. The Fisher information matrix of a subset of parameters becomes

$$\mathbf{F}_{\mathrm{L}} = \mathbf{L}^{\mathrm{T}}\mathbf{F}\mathbf{L} = \left(\mathbf{S}\mathbf{L}\right)^{\mathrm{T}}\left(\mathbf{S}\mathbf{L}\right) \tag{2-33}$$

where the selection matrix $\mathbf{L}$ is given by

$$\mathbf{L} = \begin{bmatrix} \mathbf{e}_{i_1} & \mathbf{e}_{i_2} & \cdots & \mathbf{e}_{i_{n_s}} \end{bmatrix}. \tag{2-34}$$

The set $\{i_1, i_2, \cdots, i_{n_s}\}$ denotes the index of the selected parameters and $\mathbf{e}_i$ is the i-th column of the identity matrix. The parameter selection problem then results in determining the matrix $\mathbf{L}$ which maximizes the value of a chosen measure of the Fisher information matrix from Eq. (2-33). This results in a combinatorial problem where n_s estimable parameters need to be selected from the set of n parameters. This type of problem is non-trivial to solve for a large number of parameters, especially if uncertainty in the parameters values is taken into account.

Various methods for parameter selection based on sensitivity analysis have been proposed in the literature. These include, but are not limited to, a collinearity index method (Brun et al., 2001), a column pivoting method (Velez-Reyes and Verghese, 1995), an extension of the relative gain array (Sandink et al., 2001), a Gram-Schmidt orthogonalization method (Lund and Foss, 2008; Yao et al. 2003), a recursive approach based upon principal component analysis (Li et al., 2004), and a combination of Hankel singular values and singular value decomposition (Sun and Hahn, 2006). A systematic approach for parameter selection is based on optimality criteria computed from the Fisher information matrix as the inverse of Fisher information matrix provides a lower bound for the covariance matrix of parameter estimators. A subset of identifiable parameters can be selected based upon optimizing some experimental criteria such as the *D*-optimality or the modified *E*-optimality criterion of the Fisher information matrix (Brun et al., 2001; Weijers and Vanrolleghem, 1997). Applications of parameter selection are wide-spread, ranging from ecological systems (Anh et al., 2006), power systems (Hiskens, 2001), production systems (Bastogne et al., 2007), chemical reactions (Kou et al., 2005), biochemical networks (Gadkar et al., 2005a, 2005b), to wastewater treatment processes (Sin and Vanrolleghem, 2007).

Selection of a subset of identifiable parameters for estimation results in a parameter set selection problem which is not trivial to solve. The orthogonalization method (Yao et al., 2003; Lund and Foss, 2008), which uses a modification of the Gram-Schmidt procedure, is one approach for solving this problem and involves the following steps:

Step 0 (Initiation). Set the number of parameters selected to zero, i.e., $n_s = 0$, and the projected sensitivity vectors to $\mathbf{s}_i^{(0)} = \mathbf{s}_i$, $i = 1, \cdots, n_\theta$.

Step 1 (Selection). Let $n_s = n_s + 1$ and select the parameter indexed by k which is determined by

$$k = \arg\max_i \left(\mathbf{s}_i^{(n_s)}\right)^{\mathrm{T}} \mathbf{s}_i^{(n_s)}. \tag{2-35}$$

Step 2 (Stopping test). If $\left(\mathbf{s}_k^{(n_s)}\right)^{\mathrm{T}} \mathbf{s}_k^{(n_s)} < \lambda$ (given threshold level) then stop.

Step 3 (Projection). Let $\mathbf{s}_i^{(n_s+1)} = \mathbf{s}_i^{(n_s)} - \dfrac{\left(\mathbf{s}_i^{(n_s)}\right)^{\mathrm{T}} \mathbf{s}_k^{(n_s)}}{\left(\mathbf{s}_k^{(n_s)}\right)^{\mathrm{T}} \mathbf{s}_k^{(n_s)}} \mathbf{s}_k^{(n_s)}$ and return to Step 1.

The key step is to project the sensitivity vectors of the unselected parameters on to the space orthogonal to that spanned by the sensitivity vectors of the previously selected parameters to remove the parameter's effect on the output covered by the previously selected parameters. The orthogonalization method has been widely used for analysis of biochemical reaction networks (Gadkar et al., 2005a, 2005b; Yue et al., 2006; Jaqaman and Danuser, 2006; Chu and Hahn, 2007; Jayasankar et al., 2009) since results returned by this technique have a clear interpretation and it is easy to implement.

A drawback of the orthogonalization method is that it is a heuristic approach to select identifiable parameters. A more systematic approach would be to optimize an experimental criterion of the Gram matrix of the sensitivity vectors, e.g. the *D*-optimality criterion (Brun et al., 2002)

$$\max_{\mathbf{z}} \det\left(\mathbf{S}(\mathbf{z})^{\mathrm{T}} \mathbf{S}(\mathbf{z})\right) \tag{2-36}$$

where $\mathbf{z} \in \{0,1\}^{n_\theta}$ denotes if a parameter is selected (z_i=1) or not (z_i=0) and $\mathbf{S}(\mathbf{z})$ is a submatrix of $\mathbf{S}$ consisting of the columns indexed by z_i=1. The determinant will have a small value close to zero if the sensitivity matrix is nearly rank-deficient. While it has been shown that the orthogonalization method is a forward selection method that maximizes the *D*-optimality criterion at each individual step (Chu and Hahn, 2007), it would be desirable to use a procedure which truly optimizes the *D*-optimality criterion.

2.4. Model reduction via balancing

Balancing of controllability gramians and observability gramians is a popular technique used in model reduction. Balancing of linear dynamic systems has been introduced by Moore (1981) and was later expanded to a certain class of nonlinear systems by Scherpen (1993). As the algorithms of balancing for a nonlinear system can present numerical difficulties, a class of balancing methods based upon empirical data has also been investigated (Hahn and Edgar, 2001, 2002a, 2002b, Hahn et al., 2003; Lall et al., 2002).

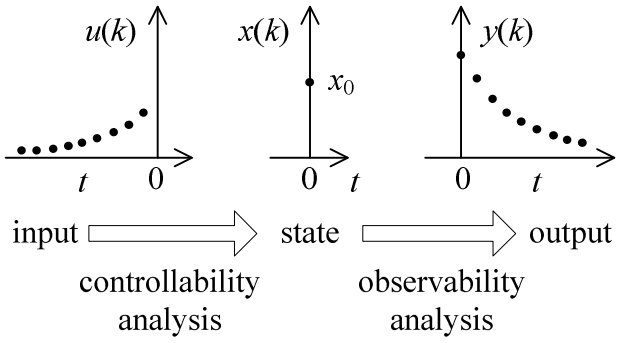


Fig. 2-1. Illustration of the role that states play in the input-output relationship. (The input $u(k)$ applies over the interval of $(-\infty, 0)$ while the output $y(k)$ is in $[0, \infty)$. Controllability analysis investigates the effect of inputs on the states while observability analysis analyzes the effect of initial perturbations of the states on the outputs.)

All balancing approaches are based upon the idea of retaining the states of a system that are most important when both controllability and observability are taken into account. Controllability analysis investigates the required energy of the inputs over the time interval $(-\infty, 0)$ to drive the system from the initial state $\mathbf{x}(-\infty) = 0$ to a current state $\mathbf{x}(0) = \mathbf{x}_0$. An illustration of this concept is shown in Fig. 2-1. Controllability analysis of a linear system can be performed by computing the controllability gramian $\mathbf{W}_C$. The eigenvectors of $\mathbf{W}_C$, corresponding to the largest eigenvalues, span the space which can most easily be reached using appropriate changes in the inputs.

Observability analysis investigates the effect that the current state $\mathbf{x}(0) = \mathbf{x}_0$ has on the outputs of the system over the time interval $(0, \infty)$. Conversely, if a state has a large effect on the outputs then the value of the state can be easily inferred from the output data. For linear systems, this state-to-output behavior can be measured by the observability gramian $\mathbf{W}_O$. The directions in state space which can most easily be inferred from the output data are given by the eigenvectors of $\mathbf{W}_O$ corresponding to the largest eigenvalues.

Once both gramians have been computed, a state transformation that balances the gramians, i.e., a transformation that turns both of them into diagonal matrices that are identical, can be applied:

$$\tilde{\mathbf{x}} = \mathbf{T}^{\mathrm{T}}\mathbf{x} \tag{2-37}$$

where $\mathbf{x}$ is the original state vector, $\tilde{\mathbf{x}}$ is the transformed state vector and $\mathbf{T}$ is the transformation matrix. One specific approach that computes $\mathbf{T}$ for the case where both gramians are full rank is given in Algorithm (2-1) below.

After balancing, each state of the transformed system is as observable as it is controllable and the importance of each state to the input-output behavior is given by the magnitude of the corresponding entries in $\mathbf{\Lambda}$.

Algorithm (2-1): Compute transformation matrix T for balancing

Step 1. Compute the gramians $\mathbf{W}_C$ and $\mathbf{W}_O$.

Step 2. Perform a Cholesky decomposition of $\mathbf{W}_C$, i.e., $\mathbf{W}_C = \mathbf{L}\mathbf{L}^{\mathrm{T}}$.

Step 3. Compute the eigenvalue decomposition of $\mathbf{L}^{\mathrm{T}}\mathbf{W}_O\mathbf{L}$, i.e., $\mathbf{L}^{\mathrm{T}}\mathbf{W}_O\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$,

where $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_{n_x}$ and $\mathbf{\Lambda} = \left[\lambda_1, \lambda_2, \cdots, \lambda_{n_x}\right]$.

Step 4. The matrix $\mathbf{T}$ is given by: $\mathbf{T} = \mathbf{L}\mathbf{U}\mathbf{\Lambda}^{-1/4}$.

2.5. Implementation of techniques using numerical algorithm

Simultaneous perturbation stochastic approximation

An optimization problem under uncertainty of some variables can be formulated by maximizing the expectation of a criterion function

$$\max_{\mathbf{w}} \mathrm{E}_{\mathbf{v}} \varphi(\mathbf{w}, \mathbf{v}) \tag{2-38}$$

where $\mathbf{w}$ represents a vector of the decision variables and $\mathbf{v}$ is a random variable following some distribution. The expectation can be computed by integrating the criterion function over the range of $\mathbf{v}$, which is then followed by maximizing the expectation as a function of $\mathbf{w}$. In this case, a gradient-based method can be used to update the value of $\mathbf{w}$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + a_k \mathbf{g}_k \tag{2-39}$$

where $\mathbf{g}_k$ is the gradient value in the k-th iteration

$$\mathbf{g}_k = \frac{\partial}{\partial \mathbf{w}} \mathrm{E}_{\mathbf{v}} \varphi(\mathbf{w}_k, \mathbf{v}). \tag{2-40}$$

However, for dynamic systems the state equations and the sensitivity equations need to be solved to compute one value of the criterion function resulting in a large computational burden. This is a point that needs to be addressed for solving this type of optimization problem.

The gradient of the expectation can be expressed by

$$\mathbf{g}_k = \mathrm{E}_{\mathbf{v}} \hat{\mathbf{g}}_k(\mathbf{v}) \tag{2-41}$$

where $\hat{\mathbf{g}}_k(\mathbf{v})$ is gradient of the criterion function at some value of $\mathbf{v}$

$$\hat{\mathbf{g}}_k(\mathbf{v}) = \frac{\partial}{\partial \mathbf{w}} \varphi(\mathbf{w}_k, \mathbf{v}) . \tag{2-42}$$

The gradient $\hat{\mathbf{g}}_k$ which is a stochastic variable is an unbiased estimate of $\mathbf{g}_k$ and can be used to update $\mathbf{w}$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + a_k \hat{\mathbf{g}}_k . \tag{2-43}$$

Use of the stochastic gradient $\hat{\mathbf{g}}_k$ instead of the gradient of the expectation $\mathbf{g}_k$ to solve the optimization problem is called a stochastic approximation (Robbins and Monro, 1951).

There are three procedures to compute the stochastic gradient: Calculate the partial derivative directly (Robbins and Monro, 1951), approximate the gradient by the ratio of the finite differences (Kiefer and Wolfowitz, 1952), or to approximate the gradient by a simultaneous perturbation (Spall, 1992):

$$\hat{\mathbf{g}}_k = \begin{bmatrix} \dfrac{\varphi(\mathbf{w}_k + c_k \boldsymbol{\Delta}_k, \mathbf{v}_k) - \varphi(\mathbf{w}_k - c_k \boldsymbol{\Delta}_k, \mathbf{v}_k)}{2 c_k \Delta_{k1}} \\ \vdots \\ \dfrac{\varphi(\mathbf{w}_k + c_k \boldsymbol{\Delta}_k, \mathbf{v}_k) - \varphi(\mathbf{w}_k - c_k \boldsymbol{\Delta}_k, \mathbf{v}_k)}{2 c_k \Delta_{kp}} \end{bmatrix} \tag{2-44}$$

where the perturbation is given by $\boldsymbol{\Delta}_k = [\Delta_{k1}, ..., \Delta_{kp}]^{\mathrm{T}}$. A sampling point of $\mathbf{v}_k$ is generated to evaluate $\hat{\mathbf{g}}_k$.

The parameters for SPSA can be selected as

$$a_k = a \Big/ (k + 1 + A)^{\alpha} \tag{2-45}$$

and

$$c_k = c \Big/ (k + 1)^{\gamma} . \tag{2-46}$$

Common values of α and γ are 1 and 1/6. Each component of the perturbation $\boldsymbol{\Delta}_k$ can use a Bernoulli ±1 distribution with probability of 1/2 for each ±1 outcome (Spall, 1998).

Evaluation of multi-dimensional integrals

Computation of a multi-dimensional integral is one of the main tasks for many sensitivity analysis methods, e.g., computation of the conditional variance in variance-based global sensitivity analysis and the quasi linearization method.

A multi-dimensional integral over the parameter space can be expressed as

$$I_f = \int_0^1 \cdots \int_0^1 f\left(\theta_1, \cdots, \theta_{n_\theta}\right) d\theta_1 \cdots d\theta_{n_\theta} \tag{2-47}$$

where the integration intervals are normalized with the lower bound set to zero and the upper bound set to unity.

Evaluation of such multi-dimensional integrals is not trivial. One general approach uses a Monte Carlo method (Robert and Casella, 2004). Monte Carlo methods generate a set of uniformly independent random points of the parameters, $\left\{\tilde{\boldsymbol{\theta}}_1, \cdots, \tilde{\boldsymbol{\theta}}_N\right\}$, and use the average value of the function over the samples to approximate the integral

$$I_N = S_R \frac{1}{N} \sum_{k=1}^{N} f\left(\tilde{\boldsymbol{\theta}}_k\right) \tag{2-48}$$

where S_R is the volume of the integration region. For the unit hyper-cube shown in Eq. (2-48), the value of S_R equals unity and the presence of this variable in the expression does not affect the numerical value, but does ensure that expression shown in Eq. (2-49) has the same unit as the one from Eq. (2-48). As given by the law of large numbers, I_N will approach I_f as the number of sampling points N approaches infinity

$$\lim_{N \to \infty} I_N = I_f \,. \tag{2-49}$$

Apart from the independently distributed random sequences, there are also deterministic sequences, called equi-distributed sequences, that are able to satisfy the condition given by Eq. (2-50). One method that uses deterministic equi-distributed sequences to evaluate the integral is the quasi Monte Carlo method (Niederreiter, 1978). An advantage of the quasi Monte Carlo method is that it can converge faster than standard Monte Carlo approaches.

One well-known equi-distributed sequence is generated from a set of rationally linear

independent numbers $\boldsymbol{\omega} = \left[\omega_1, \cdots, \omega_{n_\theta}\right]^{\mathrm{T}}$ ($n_\theta \geq 2$), i.e., for any integer $\lambda_1, \cdots, \lambda_{n_\theta}$

$$\sum_i \lambda_i \omega_i = 0 \text{ implies all } \lambda_i = 0 . \tag{2-50}$$

The sequence $\left\{\tilde{\boldsymbol{\theta}}_k = \left(k\boldsymbol{\omega}\right) \bmod 1\right\}$ is equi-distributed and the convergence condition given by Eq. (2-49) holds (Kuipers and Niederreiter, 1974). A continuous version of this sequence also exists (Kuipers and Niederreiter, 1974; Weyl, 1938) and is given by

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T g\left(\left(\boldsymbol{\omega}\tau\right) \bmod 1\right) d\tau = \int_0^1 \cdots \int_0^1 g\left(\boldsymbol{\theta}\right) d\theta_1 \cdots d\theta_{n_\theta} . \tag{2-51}$$

One important aspect of Eq. (2-52) is that the multi-dimensional integral can be transformed into a uni-dimensional integral, which is significantly easier to evaluate.

If the rationally linear independence condition is satisfied, then all elements of $\boldsymbol{\omega}$ are irrational numbers. As computers use a finite precision for representing numbers, the irrational $\boldsymbol{\omega}$ can not be recorded accurately and the rationally linear independence can not hold in practice. Instead a condition approximating the rationally linear independence has been presented in the literature (Cukier et al., 1978; McRae et al., 1982). Since only rational numbers can be recorded by a computer, the elements in $\boldsymbol{\omega}$ can be assumed to be integers without loss of generality. While it is not possible for the condition from Eq. (2-51) to hold, the equation can be satisfied by small integers $\lambda_1, \cdots, \lambda_{n_\theta}$ in the sense that

$$\sum_i \lambda_i \omega_i = 0 \text{ implies all } \lambda_i = 0 \text{ for any } \sum_i \left|\lambda_i\right| \leq M + 1 . \tag{2-52}$$

The number *M* is called the degree of independence which characterizes how close the condition given by Eq. (2-53) is to the one given by Eq. (2-51), which represents the general case for $\lim M \to \infty$. If $\boldsymbol{\omega}$ consists of only integers, then the function $g\left(\left(\boldsymbol{\omega}\tau\right) \bmod 1\right)$ is periodic with respect to τ and the integral can be evaluated over only one period of *T*. It has been shown that the error in the integration stems from the approximation involving the rationally linear dependence and that this error can be controlled by choosing a value for *M* (Cukier et al., 1978).

Markov chain Monte Carlo

Calculation of a random quantity according to a distribution can be performed by

$$\mathrm{E}\left[q(\boldsymbol{\theta})\right] = \int q(\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta} \tag{2-53}$$

where $q(\boldsymbol{\theta})$ is a function of the random vector $\boldsymbol{\theta}$ with the probability density function $p(\boldsymbol{\theta})$. One approach to compute the multidimensional integral is the Monte Carlo method. Various sampling points of $\boldsymbol{\theta}$ are generated according to $p(\boldsymbol{\theta})$ and then the integral can be approximated by the averaged value of $q(\boldsymbol{\theta})$ over all sampling points. However, in some situations the target density function is complex, e.g. a posterior density function, and it is non-trivial to generate sampling points using a direct method.

Markov chain provides a sophisticated sampling approach. A Markov chain is constructed so that its equilibrium distribution equals the desired one. The samples from the chain after transient steps are used to compute the expectation of the desired distribution. One class of methods to construct the Markov chain is given by the Metropolis-Hastings algorithm where an accept-reject procedure is used (Chib and Greenberg, 1995; Tierney, 1994). A specific algorithm is given by

Algorithm (2-2): Construct a random walk chain

Step 1. Initialize $\boldsymbol{\theta}(0) = \boldsymbol{\theta}_0$ and $k = 0$.

Step 2. Generate a sample $\boldsymbol{\psi}$ from the normal distribution $N\left(\mathbf{0}, \sigma_{\psi}^2 \mathbf{I}_{n_\theta}\right)$.

Step 3. Compute a potential value, $\boldsymbol{\theta}' = \boldsymbol{\theta}(k) + \boldsymbol{\psi}$.

Step 4. Generate a sample α from a uniform distribution in $[0,1]$.

Step 5. If $\alpha \le p(\boldsymbol{\theta}') / p(\boldsymbol{\theta}(k))$

$$\boldsymbol{\theta}(k+1) = \boldsymbol{\theta}'$$

else

$$\boldsymbol{\theta}(k+1) = \boldsymbol{\theta}(k).$$

Step 6. Set $k = k+1$ and return to Step 2.

This algorithm generates a potential value by adding a random value to the current value in Step 3. The potential value is accepted with probability $\min\left\{p(\boldsymbol{\theta}')/p(\boldsymbol{\theta}(k)),\ 1\right\}$ in Step 5. If the potential value is rejected then the current value is kept. The acceptance rate is a key factor for controlling the performance of a constructed Markov chain and it can be adjusted by choosing the variance σ_ψ^2 in Step 2. A recommended value for the equilibrium acceptance rate is 0.234 (Roberts and Rosenthal, 2001).

After the sampling points have been generated, the expectation (2) can be computed by

$$\mathrm{E}\left[q(\boldsymbol{\theta})\right] = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} q\left(\boldsymbol{\theta}(k)\right). \tag{2-54}$$

While it is not possible to use an infinite value for N in practice, N is generally chosen to be a large number. Similarly, the first few values of $q(\boldsymbol{\theta})$ are not included in the calculation as the system will not be near its equilibrium state.

Uniformly distributed random matrices

A set of n-by-n orthogonal matrices is given by $\mathcal{O}^n = \left\{\mathbf{U} : \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_n\right\}$. There is a unique distribution, denoted by μ, over the set $\mathcal{O}^n$ which is invariant under multiplication from either side by an orthogonal matrix, i.e. for the random matrix $\mathbf{X}$ over $\mathcal{O}^n$:

$$\mu(\mathbf{UX}) = \mu(\mathbf{XU}) = \mu(\mathbf{X}), \text{ for any } \mathbf{U} \in \mathcal{O}^n. \tag{2-55}$$

The distribution μ is called the uniform distribution of random orthogonal matrices (Anderson, 1984). A random orthogonal matrix $\mathbf{X}$ with the uniform distribution can be sampled using a QR factorization (Stewart, 1980).

Algorithm (2-3): Generation of uniformly distributed orthogonal matrix X

Step 1. Generate a random *n*-by-*n* matrix **Y** where each element of **Y** is independently sampled from the standard normal distribution.

Step 2. Compute the QR factorization of **Y**, i.e. **Y** = **QR** where the diagonal elements of **R** are all positive.

Step 3. A random orthogonal matrix **X** is given by $\mathbf{X} = \mathbf{Q}$.

3. COMPARATIVE SENSITIVITY ANALYSIS STUDY OF COMPLEX REACTION NETWORKS

3.1 Introduction

Mathematical models are increasingly being used as an important tool to investigate the underlying mechanism in a complex reaction network. However, a complex model, e.g. a large-scale reaction network, usually includes a great number of variables and parameters, resulting in a time consuming and ill-conditioned problem for analysis, optimization, and estimation. A powerful tool to tackle a complex reaction network is to apply sensitivity analysis including both local techniques and global techniques. Sensitivity analysis is able to identify a few important components which the following analysis can be focused on.

A great variety of techniques for sensitivity analysis exist. As no sensitivity analysis technique is known to work best for all situations, a comparison of the results returned by different techniques is required. In this section four techniques for sensitivity analysis are investigated by using a complex biochemical reaction network. These techniques are differential analysis, the Morris method, a sampling-based approach (Kolmogorov-Smirnov statistic), and the Fourier amplitude sensitivity test (FAST).

3.2 Model descriptions

The IL-6 signaling pathway model analyzed in this section was developed in a recent paper (Singh et al. 2006), which describes signal transduction in hepatocytes induced by IL-6. This model contains two pathways: Janus-associated kinases & signal transducers and transcription factors are activated in one pathway while the other pathway involves the activation of mitogen-activated protein kinases. This model consists of 68 nonlinear ordinary differential equations which include 118 parameters. The equations are derived according to the law of mass action or Michaelis-Menten kinetics and the parameters are the kinetic rate constants.

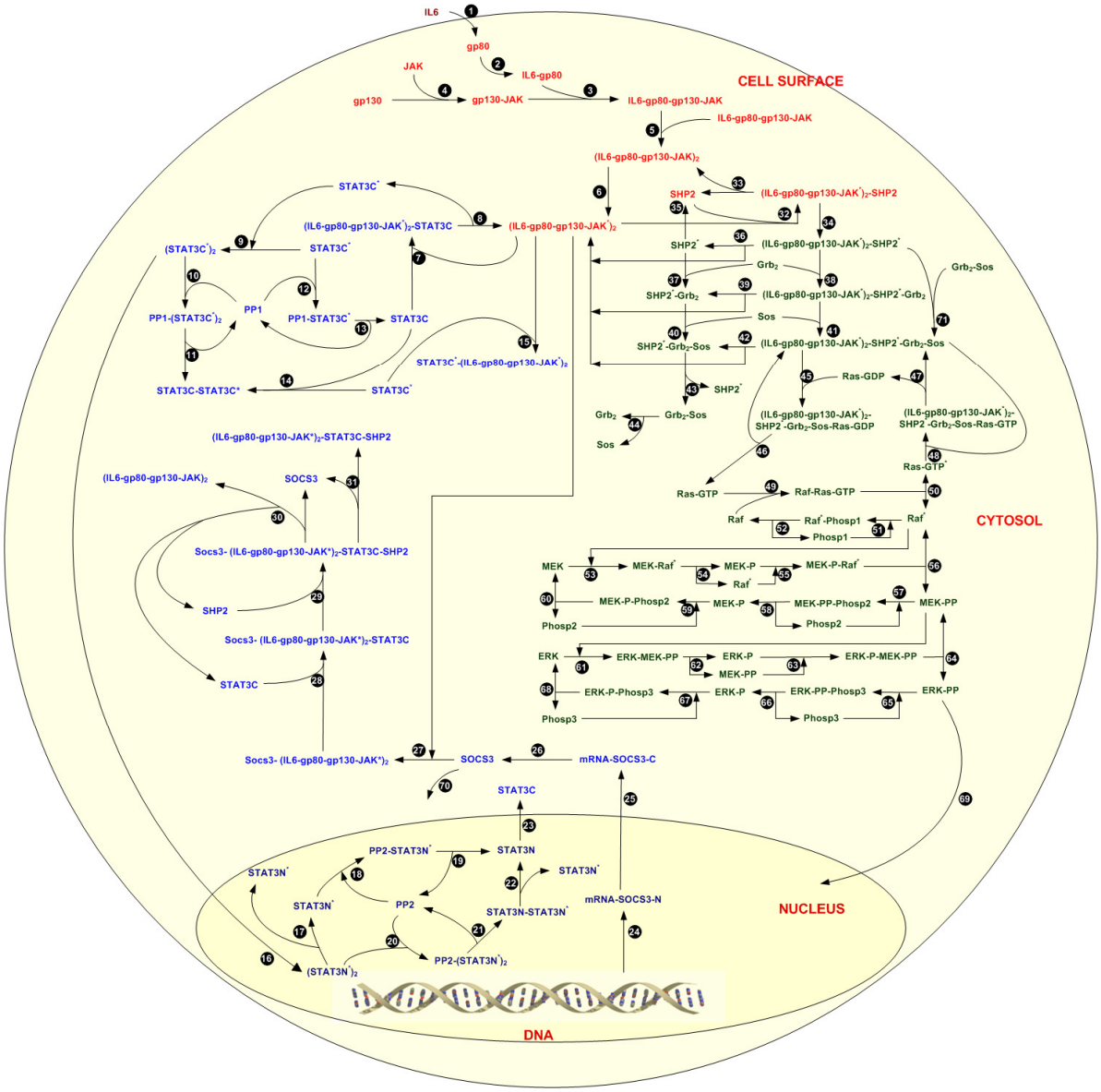


Fig. 3-1. IL-6 signaling pathway.

The state variables are the concentrations of the molecules in the pathway and the input variable is the concentration of IL-6 that stimulates the pathway. The output variable is the concentration of $(STAT3N^*)_2$ (dimer of activated STAT3 in the nucleus) as this transcription factor can be indirectly measured using a green fluorescent protein

(GFP) reporter system.

Due to the complexity of the system it is not possible to predict a priori which parts of the model are the main contributors to the dynamic behavior of the signaling pathway. The diagram of the IL-6 signaling pathway is shown in Fig. 3-1. The set of differential equations which describes the reaction network as well as the nominal values of the kinetic parameters are given in Chu et al., 2007.

3.3 Analysis of the signaling pathways and comparison of results

The four techniques for sensitivity analysis are applied to the described IL-6 signaling pathway model. The concentration of IL-6 serves as the input variable to the model for this analysis and it is changed from 0 to 0.5 nM at time 0. The simulations are carried out for a 24 hr time period as the dynamic response of the system is captured within this time interval. The time dependent sensitivity profile is sampled every minute to form the sensitivity vector.

Table 3-1

Summary of the sensitivity values calculated by the four methods. (The sensitivity values calculated by each method are given by the length of the sensitivity vector and are normalized by the largest sensitivity value.)

No.	Differential Analysis		Morris				Sampling-based				FAST			
			99-101%		10-1000%		99-101%		10-1000%		99-101%		10-1000%	
1	k_{f7}	1	k_{f7}	1	k_{f26}	1	k_{f7}	1	k_{f7}	1	k_{f7}	1	k_{f7}	1
2	k_{f32}	0.748	k_{f32}	0.761	k_{f7}	0.924	k_{f31}	0.969	k_{f21}	0.915	k_{f32}	0.753	k_{f21}	0.959
3	kf_{21}	0.713	kf_{21}	0.740	k_{f21}	0.861	V_{m24}	0.969	k_{f8}	0.748	k_{f21}	0.715	k_{f8}	0.729
4	k_{f8}	0.706	k_{f8}	0.732	k_{f27}	0.779	k_{f32}	0.891	k_{f26}	0.738	k_{f8}	0.707	k_{b7}	0.652
5	k_{b7}	0.667	k_{b7}	0.678	V_{m24}	0.683	k_{f27}	0.827	V_{m24}	0.731	k_{b7}	0.667	V_{m24}	0.624
6	k_{f20}	0.563	k_{f20}	0.573	k_{f8}	0.665	k_{f20}	0.766	k_{f29}	0.724	k_{f20}	0.563	k_{f42}	0.624
7	k_{b20}	0.549	k_{b20}	0.564	k_{b28}	0.635	k_{f8}	0.687	k_{f31}	0.703	k_{b20}	0.551	k_{f27}	0.609
8	k_{f42}	0.477	k_{f42}	0.489	k_{f31}	0.617	k_{b28}	0.684	k_{f28}	0.694	k_{f42}	0.478	k_{f26}	0.606
9	V_{m24}	0.450	k_{f26}	0.464	k_{f20}	0.585	k_{f70}	0.667	k_{f27}	0.673	V_{m24}	0.451	k_{f20}	0.558

Table 3-1 continued

No.	Differential Analysis		Morris				Sampling-based				FAST			
			99-101%		10-1000%		99-101%		10-1000%		99-101%		10-1000%	
10	k_{f26}	0.450	V_{m24}	0.463	k_{b7}	0.556	k_{b29}	0.641	k_{a26}	0.654	k_{f26}	0.451	k_{f48}	0.551
11	k_{f27}	0.447	k_{f27}	0.451	k_{b27}	0.517	k_{f21}	0.630	k_{b7}	0.639	k_{f27}	0.448	k_{f29}	0.529
12	k_{f45}	0.419	k_{f45}	0.427	k_{f29}	0.509	k_{b27}	0.611	k_{f48}	0.600	k_{b27}	0.421	k_{b27}	0.523
13	K_{m24}	0.413	k_{f31}	0.417	k_{a26}	0.471	k_{b20}	0.611	k_{b29}	0.587	k_{f45}	0.419	k_{f28}	0.523
14	k_{a26}	0.408	K_{m24}	0.417	K_{m24}	0.468	k_{b7}	0.609	K_{m24}	0.586	K_{m24}	0.415	k_{f31}	0.520
15	k_{f70}	0.407	k_{f70}	0.415	k_{f28}	0.466	k_{f29}	0.592	k_{f42}	0.580	k_{a26}	0.409	k_{a26}	0.517
16	k_{f31}	0.406	k_{a26}	0.415	k_{b29}	0.463	k_{f19}	0.588	k_{f70}	0.573	k_{f70}	0.408	k_{f32}	0.511
17	k_{b27}	0.392	k_{b27}	0.400	k_{f48}	0.428	k_{f28}	0.575	k_{b27}	0.572	k_{f31}	0.406	k_{f70}	0.508
18	k_{f28}	0.388	k_{f28}	0.395	k_{f70}	0.389	k_{a26}	0.571	k_{f20}	0.557	k_{f28}	0.389	k_{b28}	0.498
19	k_{b28}	0.387	k_{b28}	0.386	k_{f19}	0.368	K_{m24}	0.560	k_{b28}	0.540	k_{b28}	0.388	k_{b29}	0.490
20	k_{f29}	0.365	k_{f29}	0.376	k_{f42}	0.363	k_{f26}	0.540	k_{b20}	0.517	k_{f29}	0.367	K_{m24}	0.450
21	k_{b29}	0.359	k_{b29}	0.358	k_{f32}	0.338	k_{b39}	0.491	k_{b48}	0.506	k_{b29}	0.360	k_{b48}	0.445
22	k_{f71}	0.330	k_{f71}	0.338	k_{f18}	0.323	k_{f16}	0.463	k_{f32}	0.409	k_{f71}	0.331	k_{b20}	0.439
23	k_{b45}	0.303	k_{b45}	0.306	k_{f71}	0.320	k_{b10}	0.461	k_{f19}	0.401	k_{b45}	0.304	k_{f19}	0.406
24	k_{f19}	0.260	k_{f19}	0.265	k_{b48}	0.312	k_{f13}	0.435	k_{b18}	0.294	k_{f19}	0.269	k_{f45}	0.358
25	k_{f18}	0.220	k_{f18}	0.228	k_{b20}	0.283	k_{f18}	0.431	k_{f71}	0.275	k_{b38}	0.229	k_{f71}	0.324
26	k_{f36}	0.215	k_{b18}	0.218	k_{b18}	0.278	K_{m35}	0.419	k_{f18}	0.262	k_{f18}	0.221	k_{b18}	0.310
27	k_{b18}	0.214	k_{f36}	0.218	k_{f36}	0.221	k_{b17}	0.419	V_{m35}	0.233	k_{b48}	0.220	k_{f18}	0.307
28	k_{b48}	0.194	k_{f48}	0.194	V_{m35}	0.220	k_{f45}	0.399	k_{f45}	0.223	k_{f36}	0.219	k_{b45}	0.284
29	k_{f48}	0.189	k_{b48}	0.193	k_{f16}	0.216	k_{f42}	0.399	K_{m35}	0.217	k_{b18}	0.216	k_{f36}	0.219
30	k_{f6}	0.148	k_{f6}	0.146	k_{f17}	0.179	k_{f6}	0.388	k_{f17}	0.217	k_{f48}	0.197	k_{b17}	0.186
31	k_{f16}	0.135	k_{f16}	0.136	k_{f45}	0.170	V_{m35}	0.384	k_{b39}	0.180	k_{f6}	0.169	k_{f17}	0.172
32	k_{f38}	0.115	k_{f17}	0.118	K_{m35}	0.167	k_{f43}	0.361	k_{f36}	0.173	k_{f16}	0.155	V_{m35}	0.164
33	k_{b17}	0.115	k_{f38}	0.117	k_{b17}	0.160	k_{f71}	0.352	k_{f16}	0.165	k_{f17}	0.123	k_{f16}	0.158
34	k_{f17}	0.114	k_{b17}	0.116	k_{f6}	0.123	k_{f17}	0.341	k_{b45}	0.163	k_{b17}	0.118	k_{b39}	0.151
35	k_{f39}	0.099	k_{f39}	0.102	k_{f43}	0.117	k_{f48}	0.339	k_{b17}	0.134	k_{b10}	0.075	k_{f13}	0.147
36	k_{b38}	0.098	k_{b38}	0.099	k_{b45}	0.104	k_{f36}	0.331	k_{b38}	0.125	K_{m35}	0.066	K_{m35}	0.109
37	k_{f25}	0.092	k_{f25}	0.094	k_{b10}	0.101	k_{b45}	0.298	k_{f43}	0.121	k_{f13}	0.064	k_{f6}	0.094
38	k_{b71}	0.071	k_{b71}	0.072	k_{f13}	0.099	k_{b18}	0.297	k_{f6}	0.119	k_{b39}	0.047	k_{b10}	0.090
39	k_{f5}	0.066	k_{f5}	0.067	k_{b39}	0.091	k_{b48}	0.293	k_{f13}	0.072	V_{m35}	0.043	k_{f43}	0.073
40	k_{b5}	0.065	k_{f46}	0.067	k_{b38}	0.088	k_{b38}	0.277	k_{b10}	0.057	k_{f43}	0.035	k_{b38}	0.071

Local sensitivity analysis is performed on the model and additionally the three global sensitivity analysis techniques are used for a small uncertainty range (99%-101% nominal value of each parameter) and a large uncertainty range (10%-1000% nominal value) of the parameters.

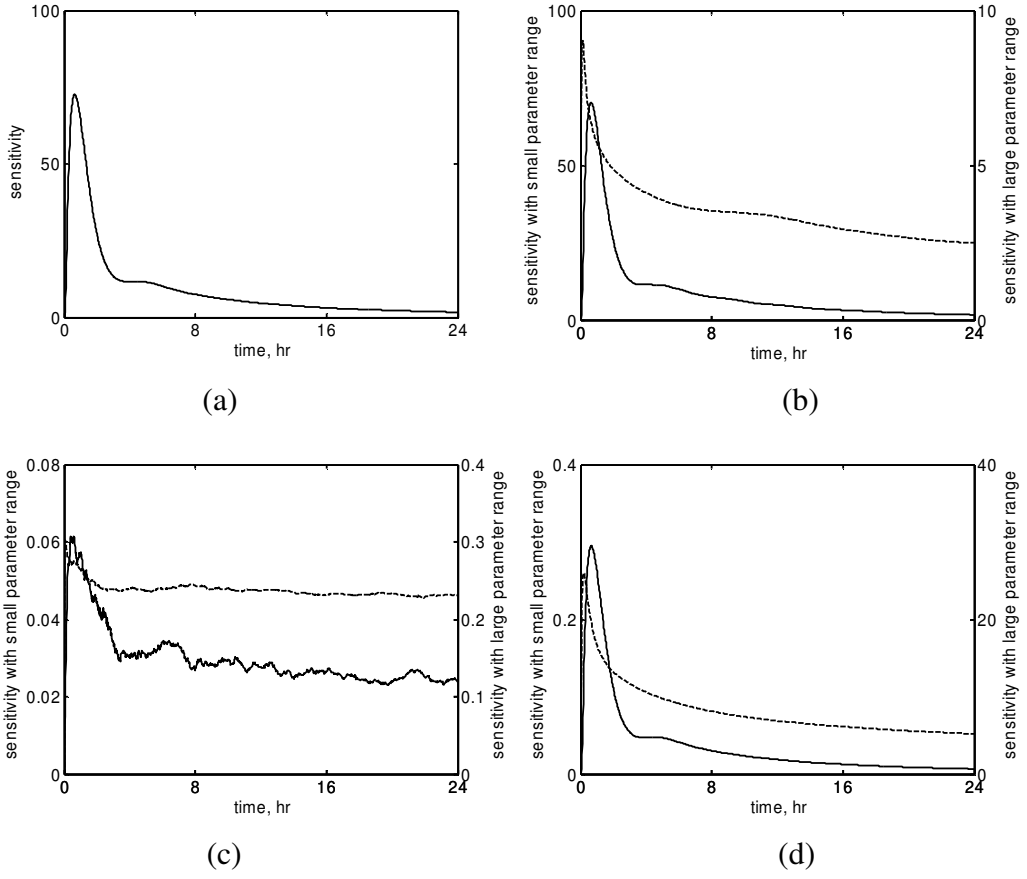


Fig. 3-2. Time dependent sensitivity profiles according to four techniques: (a) Differential analysis; (b) Morris method; (c) Sampling based method; (d) FAST method. (For the global sensitivity techniques, the sensitivity value is calculated for small parameter uncertainty (99%-100% of nominal value) represented by the solid line and for large parameter uncertainty (10%-1000% of nominal value) represented by the dashed line.)

As the used FAST algorithm is limited to the number of parameters that can be investigated and as the sampling-based approach tends to also be computationally demanding for systems with many parameters, the analysis for these two approaches are limited to the best 40 parameters identified by the Morris method. The reason for choosing the cutoff at the 40^{th} parameter is that the 40^{th} parameter has a sensitivity value that is less than 10% of the most important parameter identified by this technique. To compare the results by different techniques the lengths of the sensitivity vectors (normalized by the largest one) are listed and ordered in Table 3-1. The number of simulations for FAST is chosen to be 13001 as this number satisfies the Nyquist sampling theorem (Cukier et al., 1975). There are no restrictions on the number of simulations for the sampling-based approach and the same size with the FAST method is assigned. The time dependent sensitivity profiles of the activated transcription factor $(STAT3N^*)_2$ in the nucleus with respect to the parameter k_{f7} computed from the four techniques are shown in Fig. 3-2.

Comparison of results by the four techniques for sensitivity analysis

It can be concluded from Table 3-1 that the results returned by the Morris method and the FAST method for a small parameter range are nearly identical to the ones computed by local analysis. This is not surprising as results from both the Morris method and FAST will reduce to results from a local method if the parameter-output relationship is sufficiently smooth and the parameter vary only in a small uncertainty range.

The ranking of the parameters for small changes for the sampling-based approach are also similar to the ones computed from differential analysis. The main differences between the results for these two methods arise from the different sensitivity measures used by the two techniques.

When the uncertainty range of the parameters is large, the nonlinear properties of the system become dominant and the parameter interactions will have a significant effect on the results. This effect can also be seen in the sensitivity analysis results. For example, as the uncertainty range increases, the importance of the parameter k_{f32}, as noted by its

position on the list, decreases from 2 to 21 by the Morris method, from 4 to 22 by the sampling based approach and from 2 to 16 by the FAST method. This change can serve as an indicator that local analysis may not always be appropriate when dealing with systems where parameter values are within a large uncertainty interval.

The results obtained by the sampling-based method and the FAST method are similar for a large uncertainty range of the parameters. When comparing the results generated from these two methods to those computed by the Morris method, then it is found that while the set of important parameters is similar, that there is nevertheless a difference in the ranking of the parameters. This difference is due to the fact that the Morris method has a more limited capability of capturing nonlinear of the parameter-output behavior than a sampling-based approach or FAST.

Due to the fact that the FAST method reduces to local sensitivity analysis, that it automatically generates time-dependent sensitivity profiles, and that it is computationally more efficient than the sampling-based approach if the contribution of the individual parameters to the uncertainty is calculated, the following discussion will focus on results returned by the FAST method.

Different dynamic effects of the parameters

It has been recognized that distinct temporal activation profiles of the same signaling proteins result in different gene-expression patterns and diverse physiological responses (Detre et al., 2006; Kholodenko, 2006. Hoffmann et al., 2002; Marshall, 1995) and, therefore, discriminating the temporal effects of the parameters is of great importance.

The sensitivity values listed in Table 3-1 denote the total effect that a parameter has on the output. However, parameters can have the same cumulative effect while at the same time have distinct dynamic behavior. The time dependent sensitivity profiles are required to analyze time-dependent effect that parameters have on the output.

According to the sensitivity profiles by FAST, the parameters can be classified roughly into three groups: (1) parameters, such as k_{f7}, whose sensitivity plot initially increases rapidly and then decreases slowly; (2) parameters, such as k_{f32}, whose

sensitivity plot increases sharply to a high peak and then decreases quickly to zero; and (3) parameters, such as k_{f29}, whose sensitivity plot rises gradually to a significant level.

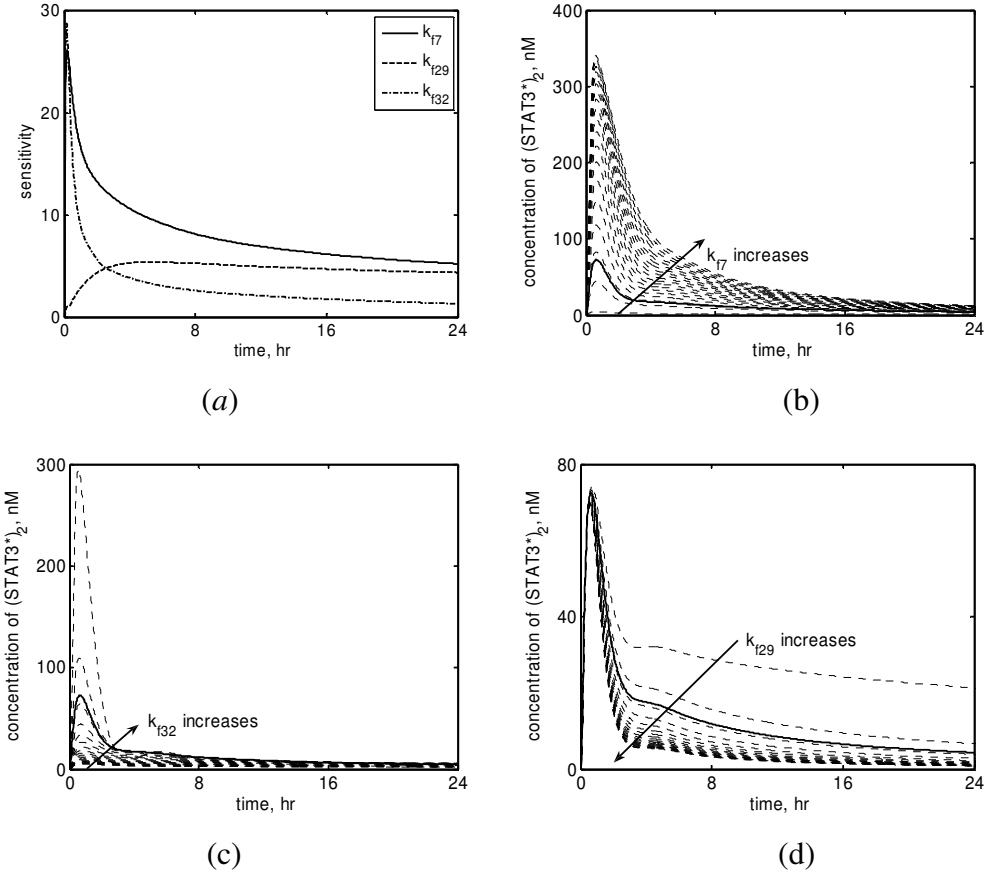


Fig. 3-3. Different dynamic effects of the parameters. (a) Time dependent sensitivity profiles by FAST; (b) Effect by variations of k_{f7}; (c) Effect by variations of k_{f32}; (d) Effect by variations of k_{f29}. (The parameter uncertainty range is from 10% to 1000% of the nominal value. The solid line is the concentration at the nominal value while the dashed lines are the concentrations at different values of the varied parameter.)

The sensitivity profiles of the three parameters are shown in Fig. 3-3(a). To illustrate the different effects of different groups of parameters, the concentration of $(STAT3N^*)_2$

at different values of each parameter (while the other parameters are held constant at their nominal values) are shown in Fig. 3-3(b-d). When k_{f7} is perturbed, it has a considerable impact on the amplitude and on the duration of the output signal. The effect of k_{f32} is mainly observed over a certain time interval and it has a large effect on the amplitude but little long-term effect, whereas k_{f29} has a significant influence over the entire time span of the simulation.

3.4 Conclusion

Mathematical modeling and simulation of complex signaling pathways has received increasing attention in the area of quantitative cell biology over the least few years. As many of the underlying biological mechanisms are not fully understood, it is important to study the effect of uncertainties on a system and determine which parameters should be estimated from data to account for these uncertainties. Towards this end, sensitivity analysis is a powerful tool to analyze mathematical models containing uncertain parameters. Four sensitivity analysis techniques were applied to the analysis of an IL-6 signaling pathway in this section and the results were discussed.

It can be concluded from the sensitivity analysis results that binding of the transcription factor STAT3 to the dimer of the phosphorylated receptor complex (IL6-gp80-gp130-JAK*)$_2$ is the most important reaction governing these pathways. Among the regulatory mechanisms in the pathway, reactions involving PP2 were determined to be the most important ones for the JAK/STAT pathway. Parameters associated with reactions involving SHP2 have a large effect on the initial response while parameters associated with reactions involving SOCS3 mainly affect the long-term behavior of the output. Parameters associated with reactions related to PP1 had the least effect of the ones mentioned here. On the Ras/MAPK side of the signaling pathway it was determined that that the receptor dimer binding to the exchange factor Sos through SHP2 and Grb2 is the most important intermediate that is affecting the STAT3 dimer in the JAK/STAT pathway.

Of these findings, the effect of MAPK on JAK/STAT is the most interesting as it

indicates a secondary level of control/regulation that is not obvious in current descriptions of IL-6 signaling or published data. Silencer RNA-mediated gene knockouts interfering with the formation of the (IL6-gp80-gp130-JAK*)2-SHP2*-Grb2-Sos complex can be used in future experiments to validate the effect of this secondary level of control on IL-6 signal transduction.

4. PARAMETER SET SELECTION FOR ESTIMATION FOR NONLINEAR DYNAMIC SYSTEMS

4.1 Introduction

Mathematical modeling plays an important role in study of complex dynamic systems and parameter estimation forms an essential component of deriving mathematical models. However, accurate estimation of parameters can be challenging as models can contain hundreds or even thousands of parameters while at the same time experimental data gathered for parameter estimation may be sparse and noisy. It is usually not possible to estimate the values of all the parameters accurately from the experimental data. It is the purpose of this section to develop a new approach for determining sets of parameters that should be estimated.

Parameter sensitivity analysis and experimental design are closely related techniques. The Fisher information matrix (FIM) serves as a measure of how much information about the parameters can be extracted from an experiment (Atkinson et al., 2007; Pazman, 1986; Silvery, 1980). If the Fisher information matrix is far from being singular in some sense then parameters are practically identifiable (Walter and Pronzato, 1990). A subset of parameters which can be estimated accurately is selected based upon optimizing certain criteria (Kiefer, 1959) as it is usually not possible to estimate the values of all parameters. A combination of the *D*-optimality and the modified *E*-optimality criteria has been used to determine identifiable parameters (Brun et al., 2002; Weijers and Vanrolleghem, 1997). If the Fisher information matrix is not close to being singular, then the norm of the sensitivity vectors is likely to be reasonably large and the angles between the sensitivity vectors are not small, either. Following these two rules, several parameter-selection techniques have been developed based on the sensitivity vectors, such as an orthogonalization method (Yao et al., 2003) and a recursive approach based upon principal component analysis (Li et al., 2004).

However these parameter selection approaches are local methods since parameter sensitivities will vary depending upon the choice of nominal values of parameters. The

inherent uncertainty in the parameter values poses a challenge on parameter selection. Sequential design is the most common approach to handle the described challenge (Issanchou et al., 2005): a set of initial values for the parameters is used for experimental design and to estimate parameters. The newly estimated parameter values are then used for another round of experimental design where values of the parameters are re-estimated. While such a procedure can be useful for systems where it is possible to perform a relatively large number of experiments, it can pose problems for systems such as intra-cellular signal pathways, as experiments can take weeks of preparation and can be expensive. Other procedures such as Bayesian methods (Chaloner and Verdinelli, 1995; Han and Chaloner, 2004) and maximin methods (Dette, 1997; Muller, 1995) require intensive computation and may prohibit applicability to systems with a large number of parameters.

Another challenge that arises for dynamic systems is that sensitivities need to be calculated along state trajectories which result in the Fisher information matrix being dependent not only on the parameter values but also on the initial states and inputs. It is the aim of this section to present a parameter set selection technique for dynamic systems described by nonlinear autonomous differential equations which will take the effect of uncertainties of the parameter values and initial states as well as changes of the inputs into account. Analysis of possible parameter sets to determine their likelihood to be the optimal set for parameter estimation as well as the magnitude of the region in parameter space under which a set will remain optimal form important components of this section.

A collection of (sub-)optimal parameter sets is investigated rather than just focusing on the “optimal” set due to the following reasons: (i) the differences in the values of the optimality criteria between the “optimal” set and a suboptimal set may be negligible and it may not be possible to distinguish between them in practice; (ii) the “optimal” set may only be the best set at the nominal point and it may be worse than a suboptimal set if the nominal values of the parameters are slightly different than was originally thought; (iii) further analysis can concentrate on these important sets rather than considering all

possible subsets of parameters; (iv) some experimental limitations may not have been taken into account when deriving the "optimal" set of parameters and determining several sets of potential candidates for parameter estimation can allow more flexibility for conducting experiments. A collection of suboptimal sets is determined by a genetic algorithm and is subsequently analyzed to determine the key factors influencing the sensitivity and to compute which parameter sets work best when uncertainty in the nominal values of the parameters is taken into account.

4.2 Presentation of a new parameter subset selection procedure

This section presents a new procedure for parameter set selection for parameter estimation of nonlinear dynamic systems. The contribution of this technique is that it combines a method for selecting parameter sets with uncertainty analysis to determine when a parameter set that is suboptimal for the nominal values of the parameters may become optimal due to changes of the nominal values. A flow diagram of the procedure that is used in this section is shown in Fig. 4-1.

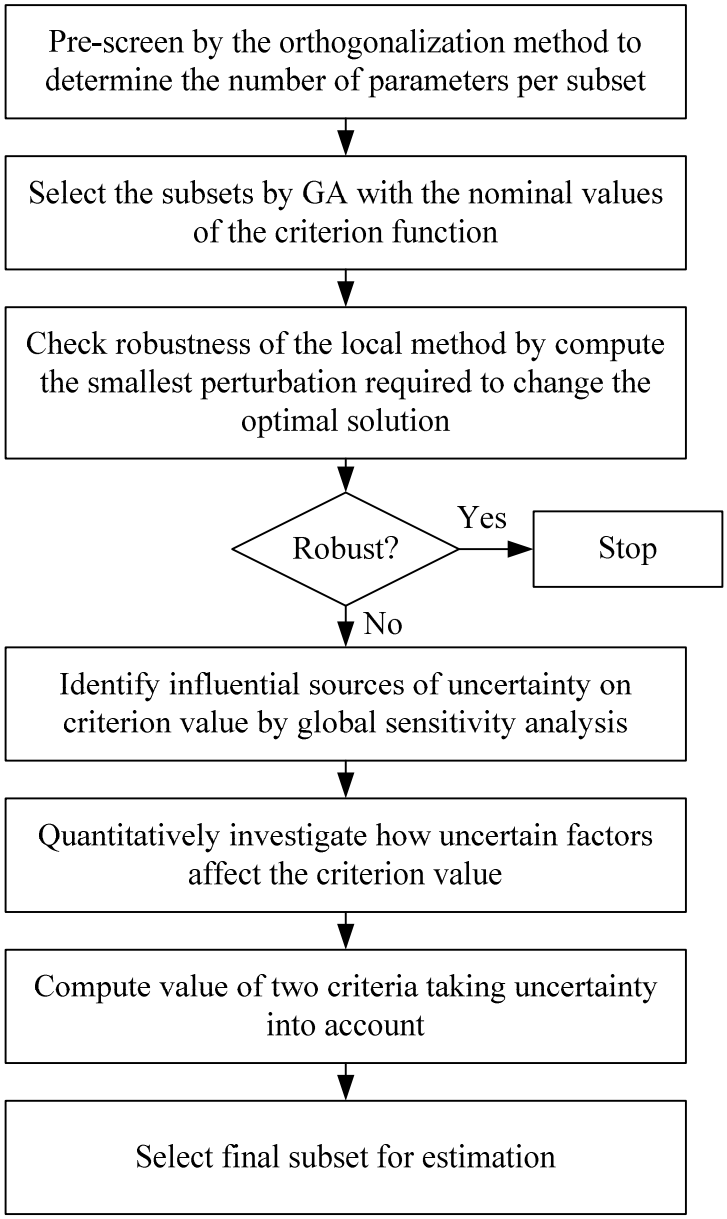


Fig. 4-1. Flow diagram of procedure for parameter subset selection.

Parameter subset selection by GA

Parameter selection procedures search for a subset of parameters which maximizes an optimality criterion. One specific form of such an optimization problem is given by

$$
\begin{aligned}
&\mathbf{z}^{*} = \arg\max_{\mathbf{z}} \ \phi_D(\mathbf{F}(\mathbf{z})) \\
&\text{s.t. } \mathbf{F}(\mathbf{z}) = \mathbf{FIM}^{(i_1,\cdots i_{n_s})}_{(i_1,\cdots i_{n_s})} \text{ with } i_j \text{ that } z_{i_j} = 1,\ j = 1\cdots n_s \\
&\quad z_1 + z_2 + \cdots + z_{n_\theta} = n_s \\
&\quad z_i \in \{0,1\},\ i = 1\cdots n_\theta .
\end{aligned}
\tag{4-1}
$$

The decision vector $\mathbf{z} \in \{0,1\}^{n_\theta}$ denotes whether a parameter is included in the selected parameter subset. If $z_i = 1$ then θ_i belongs to the selected subset with the size of n_s. The value of n_s can be determined through prescreening by the orthogonalization method. **FIM** is the Fisher information matrix of all parameters. **F**(**z**) is the Fisher information matrix of the parameters included in the selected subset and it is equal to the principal submatrix of **FIM** with the indices of the non-zero decision variables (the entries of column i_j and row i_k, $j, k = 1 \ldots n_s$).

This optimization problem results in a nonlinear integer programming problem. While an exhaustive search is a simple approach to find the optimal solution, this is not a practical approach for any problem of reasonable size. Sequential methods which add parameters to the subset one at a timer are able to significantly reduce the computational burden. It will be shown that the orthogonalization method is a sequential approach which maximizes the *D*-criterion at each step. To elaborate on this point the QR decomposition is used to express the orthogonalization

$$\mathbf{S} = \mathbf{QR}, \tag{4-2}$$

where **S** is the normalized sensitivity matrix of the selected parameters, **Q** is an orthogonal matrix and **R** is an upper triangular matrix. The columns of **Q** form the unit orthogonal bases of the space spanned by the sensitivity vectors (the columns of **S**) and the columns of **R** are the coordinates of the sensitivity vectors on the orthogonal bases. When a new parameter is selected, its sensitivity vector is added to **S**, a new base is added to **Q** and the coordinates of the sensitivity vector on the bases are added to **R**. The new diagonal entry of **R** denotes the projected value of the last sensitivity vector on the space normal to the sensitivity vectors of the previously selected parameter. The orthogonal method maximizes the square of the new diagonal entry of **R** at each step

when a new parameter is selected. The determinant of the information matrix is related to the determinant of **R** by

$$\det(\mathbf{S}^T\mathbf{S}) = \det(\mathbf{R}^T\mathbf{R}) = \det(\mathbf{R})^2 . \tag{4-3}$$

Because **R** is upper triangular the determinant of the information matrix is equal to the product of the squared diagonal entries of **R**. Accordingly, the orthogonalization method which maximizes the squared diagonal entry of **R** at each step can be regarded as a sequential method that maximizes the *D*-criterion at each step. However, due to the sequential nature of the orthogonalization method, it is possible that parameter sets with even larger criterion values may be missed as they can only be found by a simultaneous approach. That being said, this procedure can still be implemented as a pre-screening tool as it is straightforward to implement and does not require extensive computations.

It is important to select a set of estimable parameters for parameter estimation, however, the parameter set corresponding to the optimal criterion value at the nominal point may not always be the best choice due to the optimality criterion changing with the nominal values of the parameters. Accordingly, a procedure is required to not only to compute the optimal set of parameters but also to determine a collection of suboptimal parameter sets. This can be achieved by using a genetic algorithm (GA) (Goldberg, 1989; Michalewicz, 1994) to solve the optimization problems shown in Eq. (4-1). One distinct property of a GA is that it involves a population of potential solutions to the problem. Multiple candidate solutions are considered simultaneously and according to the evolution law good population member has a larger chance to be preserved in the new generation than unfit members. After many generations, the population will usually contain many members with high fitness values. This property makes GA very suitable to solve the problem of subset selection. A collection of (sub-)optimal solutions can be formed by choosing good candidates from each generation with a value of the optimality criterion larger than a threshold level α. This procedure will return a collection of parameter sets with near optimal value of the optimization problem shown in Eq. (4-1).

Determine the region in parameter space for which local results remain valid

Due to continuity of the optimality criterion, the optimal subset selected at the nominal value will still be the best set in a neighborhood around the nominal point. However, if the nominal values of the parameters can vary significantly, then the results computed by local sensitivity analysis may not be accurate over the entire range. A technique is presented in this subsection which determines the smallest magnitude of parameter changes that is required such that the parameter set with the optimal value at the nominal point will lose its "top positions" to another set of parameters. The magnitude of the variation under which the chosen parameter set does not change is an indicator of the robustness of the results computed by the local method.

Since an analytical expression describing the relationship between the criterion function and the nominal values of to the parameters is usually not known in practice, a linear approximation of the sensitivity vectors is used:

$$\left.\frac{\partial \mathbf{y}}{\partial \theta_i}\right|_{\boldsymbol{\psi}+\Delta\boldsymbol{\psi}} = \left.\frac{\partial \mathbf{y}}{\partial \theta_i}\right|_{\boldsymbol{\psi}} + \left.\frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_i}\right|_{\boldsymbol{\psi}} \Delta\boldsymbol{\psi} \,. \tag{4-4}$$

The sensitivity matrix contains the sensitivity vectors of a subset of parameters $\theta_{i_1}, \theta_{i_2}, \cdots, \theta_{i_{n_s}}$ and can be expressed by

$$\left[\frac{\partial \mathbf{y}}{\partial \theta_{i_1}} \quad \cdots \quad \frac{\partial \mathbf{y}}{\partial \theta_{i_P}}\right]_{\boldsymbol{\psi}+\Delta\boldsymbol{\psi}} = \left[\frac{\partial \mathbf{y}}{\partial \theta_{i_1}} \quad \cdots \quad \frac{\partial \mathbf{y}}{\partial \theta_{i_{n_s}}}\right]_{\boldsymbol{\psi}} + \left[\frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_1}} \quad \cdots \quad \frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_{n_s}}}\right]_{\boldsymbol{\psi}} \left(\mathbf{I}_{n_s} \otimes \Delta\boldsymbol{\psi}\right) \tag{4-5}$$

where $\boldsymbol{\Psi}$ is the vector which characterizes the operating conditions. To simplify the notation,

$$\mathbf{S}_I = \bar{\mathbf{S}}_I + \mathbf{W}_I \left(\mathbf{I}_{n_s} \otimes \mathbf{d}\right) \tag{4-6}$$

will be used, where the matrices are

$$\mathbf{S}_I = \left[\frac{\partial \mathbf{y}}{\partial \theta_{i_1}} \quad \cdots \quad \frac{\partial \mathbf{y}}{\partial \theta_{i_{n_s}}} \right]_{\boldsymbol{\psi}+\Delta\boldsymbol{\psi}}$$

$$\bar{\mathbf{S}}_I = \left[\frac{\partial \mathbf{y}}{\partial \theta_{i_1}} \quad \cdots \quad \frac{\partial \mathbf{y}}{\partial \theta_{i_{n_s}}} \right]_{\boldsymbol{\psi}}$$

$$\mathbf{W}_I = \left[\frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_1}} \quad \cdots \quad \frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_{n_s}}} \right]_{\boldsymbol{\psi}}$$

$$\mathbf{d} = \Delta\boldsymbol{\psi}$$

$$I = \left\{ i_1 \quad i_2 \quad \cdots \quad i_{n_s} \right\}.$$

One should note that the linear approximation of the sensitivity vectors is used rather than the linear approximation of the optimality criterion itself as linearization of the sensitivity vector offers a more accurate approximation.

Suppose that a parameter set at the nominal point (indicated by indices *J*) has a larger criterion value than another parameter set (indicated by indices *I*). The smallest perturbation required to change the order of two parameter sets can be calculated by the following optimization problem:

$$\begin{aligned} & \min \ \left\| \mathbf{d} \right\|_2 \\ \text{s.t.} \ & \phi_D \left(\mathbf{S}_I{}^{\mathrm{T}} \mathbf{S}_I \right) > \phi_D \left(\mathbf{S}_J{}^{\mathrm{T}} \mathbf{S}_J \right) \\ & \mathbf{S}_I = \bar{\mathbf{S}}_I + \mathbf{W}_I \left(\mathbf{I}_P \otimes \mathbf{d} \right) \\ & \mathbf{S}_J = \bar{\mathbf{S}}_J + \mathbf{W}_J \left(\mathbf{I}_P \otimes \mathbf{d} \right) \\ & \mathbf{d}_{\mathrm{L}} \le \mathbf{d} \le \mathbf{d}_{\mathrm{U}}. \end{aligned} \tag{4-7}$$

The last inequality constraint provides an upper and a lower bound for variation of the parameter vector such that constraints on parameters by physics can be taken into account. For example, all the kinetic parameters in a model referring to rate constants should always be positive. It should be noted that due to the linear approximation of the sensitivity matrix it may be possible that the variation calculated may not change the order of the two subsets. In this case the sensitivity values can be re-evaluated at the perturbed parameter value calculated by the first solution of optimization problem and

the optimization problem is solved again. This is an iterative procedure that is performed until a perturbation is found that will change the order of the criterion values of the two sets.

Sampling-based method to identify sources of uncertainty that affect the value of the optimality criterion

The technique presented in this subsection uses global sensitivity analysis to determine how sensitive the optimality criterion is to sources of uncertainty. For the most part, these sources of uncertainty are due to changes in the values of the parameters, however, changes in initial conditions can also be considered.

A sampling-based method with Latin hypercube sampling is used in this section since it is the most efficient sampling way for large systems. The optimality criterion is evaluated at each sampling point by simulating the model. The Kolmogorov-Smirnov (KS) statistic of the criterion value with respect to a parameter is calculated to serve as the global sensitivity measures following the procedure described below:

Step 1. Determine the uncertainty range of each parameter.

Step 2. Generate uniformly distributed samples of the parameters by Latin hypercube sampling.

Step 3. Calculate the first order sensitivities by solving the state equations and the sensitivity equations simultaneously for each sample value and compute the value of the optimality criterion.

Step 4. Calculate the objective function for each sample

$$f_{\phi}(k) = \left(\phi_D(k) - \bar{\phi}_D\right)^2,$$

where $\phi_D(k)$ is the criterion value calculated at the k-th sample, $\bar{\phi}_D$ is the criterion value calculated at the nominal value.

Step 5. Calculate the mean value of $f_{\phi}(k)$ and group the sample values of each parameter into two sets. If $f_{\phi}(k)$ is larger than the mean value, then the k-th

sample value of the parameter is placed into the ‘unacceptable set’; otherwise it is put into the ‘acceptable set’.

Step 6. Compute the two cumulative distribution functions of the sample values contained in the two sets for each parameter and calculate the *KS* statistic.

From the sampling points the criterion functions which are not subject to the parameter uncertainty can be calculated. Due to the uncertainty a subset of parameters can be estimated more accurate than another subset at one point but less accurate at another point. The mean criterion value of a subset indicates the overall performance of a subset. A good estimator of the mean criterion is the average criterion value on the sampling points. However it is the case that a subset can have a large mean criterion value because it has a very large criterion value in a small range but has low criterion value over most of the parameter space. In practice the situation where the subset has large criterion value is unlikely and it is more likely that the subset is worse than others. One may prefer to select the subset which has the largest probability to have the largest criterion value in the uncertain range. The probability of each subset to be the top one can be calculated from the sampling points as well. From the explanation above the two criteria may not be completely consistent and an example is in the case study in the next section. One is often at loss to choose the criterion before selection. This is another motivation to select a collect of subsets. After calculation of the value of the two criteria of the subsets one is easy to make a balance among different criteria.

Quantitative investigation of the effect of uncertain factors on the optimality criterion

Even though global sensitivity analysis is able to identify the important uncertain factors affecting the value of the optimality criterion, it is unable to determine quantitatively how changes in the nominal parameter values affect estimation accuracy. The gradient of the criterion function can provide such information and it can be used directly to determine the optimal setting of adjustable variables. The mathematical procedure for this technique is provided in the following.

Assume a selected subset is $\left\{\theta_{i_1}, \theta_{i_2}, \cdots, \theta_{i_{n_s}}\right\}$ and the sensitivity matrix is

$$\mathbf{S}(\boldsymbol{\psi}) = \left[\begin{array}{cccc} \dfrac{\partial \mathbf{y}}{\partial \theta_{i_1}} & \dfrac{\partial \mathbf{y}}{\partial \theta_{i_2}} & \cdots & \dfrac{\partial \mathbf{y}}{\partial \theta_{i_{n_s}}} \end{array}\right]\Bigg|_{\boldsymbol{\psi}=[\boldsymbol{\theta}^{\mathrm{T}}, \mathbf{x}_0{}^{\mathrm{T}}, \mathbf{u}^{\mathrm{T}}]^{\mathrm{T}}} \tag{4-8}$$

where $\mathbf{S}$ is evaluated at some value of the parameter $\boldsymbol{\psi}$ and the *D*-criterion is a function of $\boldsymbol{\psi}$

$$\phi_D(\boldsymbol{\psi}) = \log\det\left(\mathbf{S}(\boldsymbol{\psi})^{\mathrm{T}}\mathbf{S}(\boldsymbol{\psi})\right). \tag{4-9}$$

Differentiation of the criterion function results in

$$d\phi_D(\boldsymbol{\psi}) = 2\mathrm{trace}\left\{\left(\mathbf{S}(\boldsymbol{\psi})^{\mathrm{T}}\mathbf{S}(\boldsymbol{\psi})\right)^{-1}\mathbf{S}(\boldsymbol{\psi})^{\mathrm{T}}\, d\mathbf{S}(\boldsymbol{\psi})\right\}. \tag{4-10}$$

Since the differentiation of each element in the sensitivity matrix is

$$d\left(\frac{\partial \mathbf{y}}{\partial \theta_{i_j}}\right) = \frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_j}} d\boldsymbol{\psi}\,, \tag{4-11}$$

the differential of the sensitivity matrix is

$$d\mathbf{S}(\boldsymbol{\psi}) = \left[\begin{array}{cccc} \dfrac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_1}} d\boldsymbol{\psi} & \dfrac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_2}} d\boldsymbol{\psi} & \cdots & \dfrac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_{n_s}}} d\boldsymbol{\psi} \end{array}\right]. \tag{4-12}$$

Substituting Eq. (4-12) into the optimality criterion results in

$$d\phi_D(\boldsymbol{\psi}) = 2\mathrm{trace}\left\{\left[\begin{array}{cccc} \mathbf{A}_1 d\boldsymbol{\psi} & \mathbf{A}_2 d\boldsymbol{\psi} & \cdots & \mathbf{A}_{n_s} d\boldsymbol{\psi} \end{array}\right]\right\}, \tag{4-13}$$

where

$$\mathbf{A}_{i_1} = \left(\mathbf{S}^{\mathrm{T}}\mathbf{S}\right)^{-1}\mathbf{S}^{\mathrm{T}}\frac{\partial^2 \mathbf{y}}{\partial \boldsymbol{\psi}^{\mathrm{T}} \partial \theta_{i_1}}. \tag{4-14}$$

Finally,

$$d\phi_D(\boldsymbol{\psi}) = 2\left(\mathbf{a}_1{}^{\mathrm{T}} + \mathbf{a}_2{}^{\mathrm{T}} + \cdots + \mathbf{a}_{n_s}{}^{\mathrm{T}}\right) d\boldsymbol{\psi}\,. \tag{4-15}$$

where $\mathbf{a}_i{}^{\mathrm{T}}$ is the *i*-th row of $\mathbf{A}_i$ and the partial derivative of ϕ_D with respect to $\boldsymbol{\psi}$ is

$$\frac{\partial \phi_D}{\partial \boldsymbol{\psi}} = 2\sum_{j=1}^{n_s} \mathbf{a}_j\,. \tag{4-16}$$

The magnitude of the gradient is an indicator of the effect that changes in a parameter have on the criterion function value. The sign of the gradient indicates whether a change of the value of a parameter increases or decreases the optimality criterion. The gradient shown in Eq. (4-16) is in fact the local sensitivity of the criterion function. However, this is not to be confused with the sensitivity of the output. The sensitivity of the output is used to compute the value of the criterion for parameter selection while the sensitivity of the criterion function is used to study the effects that parameter uncertainty has on the criterion value.

4.3 Case studies

Two examples are used to illustrate the developed techniques. The first case study deals with an exothermic continuously-stirred tank reactor while the second one analyzes a detailed model describing an IL-6 transduction network in liver cells.

Parameter set selection for a CSTR

This model describes an exothermic CSTR in which a first-order reaction A→B is taking place (Muske and Georgakis, 2003):

$$A \rightarrow B, \quad R_A = k \exp(-E/RT) c_A . \tag{4-17}$$

The reactor is described by the following differential equations

$$\begin{aligned} \dot{c}_A &= \frac{F}{V}(c_A^f - c_A) - R_A \\ \dot{T} &= \frac{F}{V}(T^f - T) + \frac{\Delta H}{\rho C_P} R_A - \frac{hA}{\rho C_P V}(T - T_c) . \\ \dot{T}_c &= \frac{F_c}{V_c}(T_c^f - T_c) + \frac{hA}{\rho_c C_{P_c} V_c} \end{aligned} \tag{4-18}$$

The three states of the system are the concentration of component A, the temperature of the reactor and the temperature of the coolant jacket. The reactor temperature is chosen as the only output of the system.

All parameters in Eq. (4-18) are assumed to be constant. It can be seen that ρ and C_P never appear by themselves and only in the form of their product in Eq. (4-18). Due to

this only the product of the two parameters can be estimated. The same situation arises for the product of ρ_c and C_{Pc}. To take this observation into account the parameters C_P, and C_{Pc} are set to their nominal value and are not considered for parameter set selection. This leaves nine parameters (No.1-No.9 in Table 4-1) as candidates considered for estimation. The feed flow rate and the coolant flow rate are the two input variables. These 11 variables plus the 3 initial conditions of the states make up the augmented parameter vector for sensitivity analysis. The reactor volume, the cooling jacket volume and the heat transfer area are design parameters whose values are exactly known. Thus there is no need to consider them for parameter estimation.

Table 4-1

Nominal value of CSTR parameters

No.	Parameter	Variable	Value
1	Feed temperature	T^f	20 ^oC
2	Feed composition	c^f_A	2500 mol/m^3
3	Fluid density	ρ	1025 kg/m^3
4	Heat of reaction	ΔH	160 kJ/mol
5	Activation energy	E/R	255 K
6	Preexponential factor	k	2.5 h^{-1}
7	Coolant inlet temperature	T^f_c	10 ^oC
8	Coolant density	ρ_c	1000 kg/m^3
9	Heat transfer coefficient	h	1000 W/m^2•^oC
10	Feed flow rate	F	0.1 m^3/h
11	Coolant flow rate	F_c	0.15 m^3/h
12	Initial state of composition	c_{A0}	1000 mol/m^3
13	Initial state of reactor temperature	T_0	20 ^oC
14	Initial state of coolant temperature	T_{c0}	20 ^oC
	Reactor volume	V	0.2 m^3
	Cooling jacket volume	V_c	0.055 m^3
	Heat transfer area	A	4.5 m^2
	Coolant heat capacity	C_{Pc}	1.2 kJ/kg•^oC
	Fluid heat capacity	C_P	1.55 kJ/kg•^oC

The sensitivities of the reactor temperature with respect to the parameters are calculated by the direct method and normalized. In a next step the orthogonalization method is applied. The results are shown in Table 4-2 where the overall sensitivity and the rank value are shown for each parameter. It can be seen that while the output may be sensitive to some parameters that these parameters may nevertheless have a small rank value as they are highly correlated to parameters already chosen for the set. The method indicates that the coolant density ρ_c, the pre-exponential factor k and the fluid density ρ form a set of three parameters that has the largest effect on the reactor temperature. The rank value of the 4th parameter is less than 0.7% of sum of the first three, and therefore the size of the parameter set is chosen to be three (n_s=3). The set $\{\rho_c, k, \rho\}$ is a suboptimal selection under the *D*-optimality. In fact the set is the optimal in this case but this is not always true (the next case is an example).

Table 4-2

Parameters of the CSTR model ordered by the orthogonalization method

Parameter	ρ_c	k	ρ	c^f_A	h	ΔH	T^f_c	E/R	T^f
Rank Value	9.29	0.79	0.13	0.07	0.008	0.001	0	0	0
Sensitivity	9.29	1.30	0.56	2.09	7.11	2.66	3.72	1.07	0.29

The total number of the possible subsets of parameters is C^3_9=84 and, it is therefore possible to perform an exhaustive search evaluating each set of parameters. The ten sets with the highest criterion value are shown in Table 4-3. 2000 simulations with the augmented parameters varying from 0.5-2 (normalized value) have been performed to investigate the change of the criterion value with the uncertainty. The mean value of the criterion for each set for these 2000 simulations is listed in Table 4-3. It can be seen that there are significant differences between the criterion values at the nominal point and the mean values of the criteria for changes in the nominal value of the parameters. For example, the 8th set of parameters results in a higher mean value of the criterion under

the influence of uncertainty in the parameter values than the best set for the nominal point. It can be concluded that determining a set of parameters to be estimated from data at a nominal point may not lead to an optimal conclusion. The last row in Table 4-3 denotes the probability for each subset to be the optimal set for the simulations that were run for the uncertain parameters. The probability is computed by the number of simulations where a subset has the largest criterion value divided by the total number of the simulations that were performed. The 7th parameter set from Table 4-3 has the largest probability to be the optimal set.

Table 4-3

Collection of suboptimal subsets for CSTR model

No.	1	2	3	4	5	6	7	8	9	10
Parameter subset	ρ k ρ_c	c^f_A k ρ_c	c^f_A ΔH ρ_c	c^f_A E/R ρ_c	ρ k h	ρ E/R ρ_c	ρ ΔH ρ_c	k ρ_c h	c^f_A ΔH h	c^f_A k h
Criterion value	-0.15	-0.23	-0.27	-0.53	-0.54	-0.61	-0.68	-0.85	-0.94	-0.98
Mean criterion value	-0.13	-1.35	-1.35	-1.38	-0.89	-0.27	-0.34	0.21	-0.77	-0.70
Probability to be the optimum	0.150	0.061	0.029	0.053	0.037	0.129	0.170	0.153	0.083	0.137

It can be seen from this analysis that there is not one set of parameters that will be the best one for both criteria if uncertainty is taken into account. Instead it is more useful to provide a collection of parameter sets as well as criteria to evaluate them and to have a user chose certain set based upon experience with a process. For example, even though the 8th subset has a slightly lower probability to be the best set compared to the 7th set, the mean criterion value of the 8th parameter set is larger than the one for the 7th set. Therefore, the 8th parameter set is the best choice for parameter estimation for this example. However, other criteria, e.g., experience that one has with a process, may also play a factor when choosing one parameter set over another one.

Table 4-4

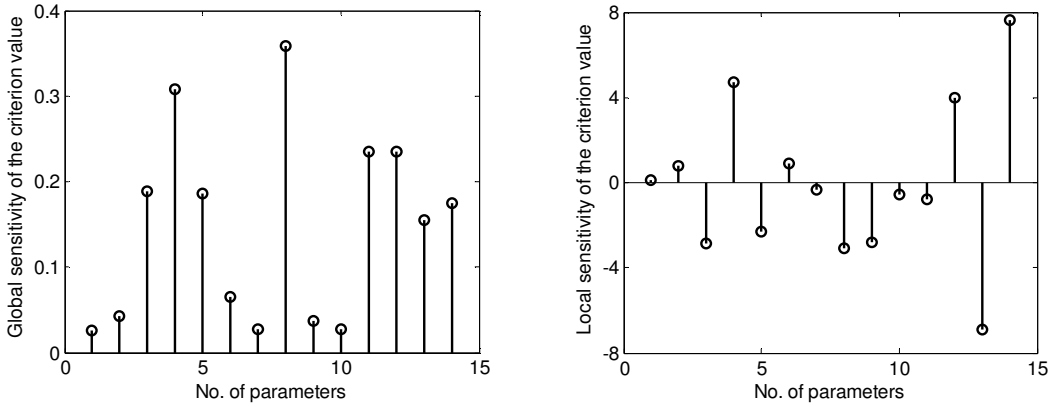
Smallest variation required to change order of a subset with the 1st one for the CSTR model

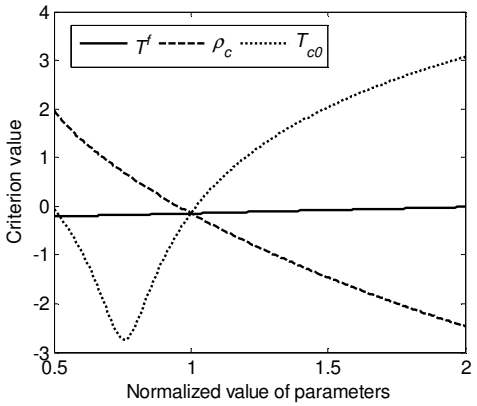
No. of subset	2	3	4	5	6	7	8	9	10
Variation magnitude	0.007	0.011	0.032	0.200	0.268	0.075	0.163	0.050	0.052

Table 4-4 lists the smallest variation of the augmented parameters required to change the order of a parameter set with the 1st set. From Table 4-4 it can be concluded that a small change in the nominal value of the parameters (0.7%) can change the selection of an optimal parameter set. Since the optimal set at the nominal point is extremely sensitive to the nominal values and since these nominal values are by definition imprecise, which is the reason why they need to be estimated, it is questionable if choosing an optimal parameter set simply based upon local sensitivity analysis returns meaningful results. Another important conclusion that can be drawn from the results shown in Table 4-4 is that the magnitude of the smallest perturbation required to change the order of two subsets is not proportional to the difference of the criterion value between the two sets. The difference of criterion value between the 6th subset and the 1st subset is less than that between the 10th subset and the 1st subset. However, the variation required to change the order of the 6th subset with the 1st subset is much larger than the one required to make the 10th set more important than the one currently ranked 1st.

The global sensitivities of the criterion values with respect to the parameters are calculated to identify the influential uncertain sources. The *KS* statistic of the 1st parameter set, as one representative of a global sensitivity measure, is computed from the sampled points (Fig. 4-2(a)). To study how variations of the parameters affect the criterion values the gradient of the criterion function are also computed and shown in Fig. 4-2(b). The gradient is in fact the local sensitivity of the criterion function. It can be

concluded that the local and global sensitivity results in different information. The initial value of the coolant temperature T_{c0} (No. 14) has the largest magnitude of the local sensitivity while the coolant density ρ_c (No. 8) has the largest global sensitivity.





(c)

Fig. 4-2. Sensitivity analysis of criterion value. (a) Global sensitivity of criterion value; (b) Local sensitivity of criterion value; (c) Change of criterion value with variation of T^f, ρ_c and T_0.

To investigate the reasons behind these different observations for local and global analysis, the criterion values have been plotted for variations of some specific parameters in Fig. 4-2(c). It can be seen that varying T_{c0} strongly changes the criterion value at the nominal value but has a diminishing effect for large values of T_{c0}. Also the criterion value does not decrease monotonically as T_{c0} decreases. The criterion value changes monotonically with changes in ρ_c in the whole range. On the other hand, the feed temperature T^f (No. 1) has only marginal effects by changes in its nominal value and it has small value of both global sensitivity and local sensitivity.

Parameter subset selection of an IL-6 signaling pathway

Modeling and analysis of intracellular signaling networks is an important area in systems biology. Signaling pathways are the cellular information routes by which cells sense their surroundings and adjust to environmental changes or hormonal stimuli. The signaling network includes various components which detect, amplify, and integrate diverse external signals to generate responses such as changes in enzyme activity or gene expression.

The IL (interleukin)-6-type cytokines are an important family of mediators involved in the regulation of the acute-phase response to injury and infection (Heinrich et al., 2003). Several models of the IL-6 signaling pathway have been proposed and a recently developed model is presented in the paper by Singh et al. 2006, which describes signal transduction in hepatocytes induced by IL-6 (Fig. 3-1). This model contains two signaling mechanisms: Janus-associated kinases (JAK) & signal transducers and transcription factors 3 (STAT3) are activated in one pathway while the other pathway involves the activation of mitogen-activated protein kinases (MAPK). The model is described by 68 nonlinear ordinary differential equations including 118 parameters. The equations are derived according to the law of mass action or Michaelis-Menten kinetics and the parameters are the kinetic rate constants. The states are the concentrations of the molecules involved in the pathway. The input is the concentration of IL-6 that stimulates the pathway and the output is the concentration of the transcription factor $(STAT3N^*)_2$

(dimer of activated STAT3 in the nucleus). For the detailed model of the differential equations and the nominal values of parameters one can see Singh et al. 2006, and Chu et al. 2007.

The investigated model contains a total of 118 parameters. Since the analysis procedure could be computationally prohibitive for such a large number of parameters, only the 50 most important parameters, as identified by the sensitivity value, will be investigated here. Also, 16 of the 68 states have initial conditions different from zero and variations in these initial conditions are also considered in this section.

The order of parameters selected by the orthogonalization method is shown in Table 4-5. (The parameter k_{fi} is the rate constant of the forward reaction in the i-th pathway and k_{bi} is the rate constant of the backward reaction of in the i-th pathway.) It can be seen that having more than 6 parameters does not provide much of a benefit as the additional contribution of the 7^{th} parameter is less than 1% to what can already be achieved by choosing 6 parameters. Accordingly the size of the subset is determined to be 6 (n_s=6).

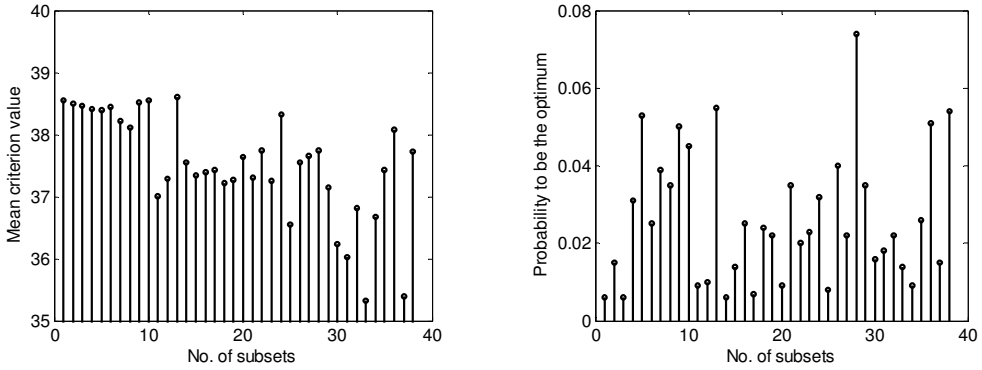
Table 4-5

First 10 parameters in the IL-6 model ordered by the orthogonal method

No.	1	2	3	4	5	6	7	8	9	10
Parameter	k_{f7}	k_{f31}	k_{f21}	k_{f70}	k_{f6}	k_{f48}	k_{f45}	k_{f26}	k_{f18}	k_{f8}
Rank value	637.3	239.6	187.5	63.2	46.3	23.0	11.7	5.3	3.2	1.0

The total number of possible sets with six parameters is C^6_{50}, which is roughly $1.6 \cdot 10^7$. It is not possible to perform an exhaustive search in this case due to the large number of possible sets, however, a selection procedure based upon a GA can still be applied. A certain threshold for determining a cutoff of sets to be considered has been found by trial and error. This cutoff has been set to a value of 55 (α=55) resulting in 38 parameter sets which are considered for further analysis. The ten parameter sets with the largest criterion values are shown in Table 4-6. It can be seen that the differences in the

criterion values between these ten sets is small, which can serve as an indicator that further analysis may be warranted rather than simply using the "optimal" set for parameter estimation.



(a) (b)

Fig. 4-3. Results calculated by sampling-based method for IL-6 model. (a) Mean criterion value of a subset; (b) Probability for a subset to be the optimal one.

The mean criterion value is calculated for 2000 simulations where the nominal values of the augmented parameters can vary from 30% to 300% and the results are shown in Fig. 4-3(a). The mean criterion value of the 1st subset is about 38.5 while the 10th and the 13th subsets have larger mean value. The probability of a subset to be the optimal one due to uncertainty in the nominal values is shown in Fig. 4-3(b). The 1st subset which is optimal at the nominal point has small probability to preserve its top position for small changes in the (estimated) nominal value. Its probability of being the top choice is only 8% of the probability of the 28th set, and there are 35 parameter sets that have larger probability of being the optimal choice. It can be concluded that if the uncertainty effects are not considered then the selected parameter set may be far from being the optimal one.

Table 4-6

10 sets of parameters with largest performance indices for the IL-6 model

No.	1	2	3	4	5	6	7	8	9	10
Parameter subset	k_{f31}	k_{f31}	k_{f31}	k_{f31}	k_{f7}	k_{f7}	k_{f31}	k_{f31}	k_{f31}	k_{f31}
	k_{f21}	k_{f21}	k_{f21}	k_{f21}	k_{f31}	k_{f31}	k_{f21}	k_{f21}	k_{f21}	k_{f21}
	k_{f70}	k_{f6}	k_{f70}	k_{f6}	k_{f21}	k_{f21}	k_{f70}	k_{f70}	k_{f70}	k_{f6}
	k_{f6}	k_{f48}	k_{f6}	k_{f32}	k_{f6}	k_{f70}	k_{f48}	k_{f16}	k_{f6}	k_{f48}
	k_{f48}	k_{f32}	k_{f32}	k_{f26}	k_{f48}	k_{f6}	k_{f16}	k_{f32}	k_{f32}	k_{f32}
	k_{f32}	k_{f26}	k_{b48}	k_{b48}	k_{f26}	k_{f48}	k_{f32}	k_{b48}	k_{f42}	k_{f27}
Criterion value	56.83	56.82	56.63	56.62	56.58	56.57	56.51	56.32	55.92	55.89

The smallest distance from the nominal point to the point at which another parameter set has larger criterion value than the 1st parameter set is shown in Table 4-7. It can be seen that a small perturbation of 0.2% of the nominal values in parameter space is able to change the optimal solution. This observation also indicates that the 1st subset which was determined by local analysis is likely to lose its top position due to uncertainties in the nominal values of the parameters.

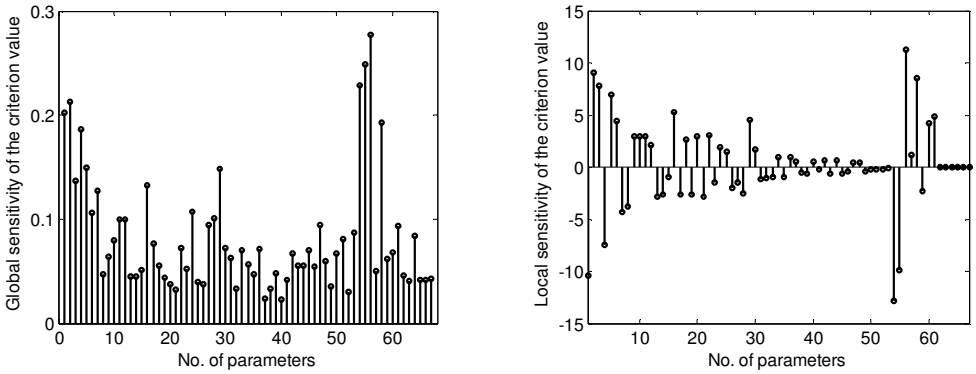
Table 4-7

Smallest variation required to change the order of a subset with the 1st one for the IL-6 model

No. of subset	2	3	4	5	6	7	8	9	10
Smallest variation	0.002	0.07	0.05	0.06	0.08	0.08	0.10	0.13	0.15

To study how individual parameters change the criterion value, the global sensitivities and the local sensitivities of the criterion with respect to the 1st parameter set are shown in Fig. 4-4. It can be seen that the initial concentration of JAK (No. 54) which has the largest magnitude of the local sensitivity also has the 3rd largest contribution

when global sensitivity analysis is applied. Similarly the initial concentration of SHP2 (No. 55) which is determined as being most important by global sensitivity also has the 2nd largest magnitude for local sensitivity.



(a) (b)

Fig. 4-4. Sensitivity analysis of criterion value of the best subset in the IL-6 model. (a) Global sensitivities of criterion value; (b) Local sensitivities of criterion value.

From the local sensitivities it can be seen how some biological mechanism affect the estimation accuracy. The initial concentration of STAT3C (No. 55) has the largest local sensitivity and an increase of the initial value raises the optimality criterion. STAT3C is one of the main proteins in the JAK/STAT signaling pathway. The initial concentration of JAK (No. 54) also has large positive sensitivity. JAK is an essential component for forming the receptor complex which is in turn required to initiated signal transduction. The initial concentration of SHP2 (No. 56) has the largest magnitude among the negative sensitivities. Increase of the initial value of SHP2 will decrease the value of the optimality criterion for this parameter set. SHP2 is an important protein for signaling through the MAPK pathway. The initial conditions of the two inhibitors PP1 (No. 57) which deactivates STAT3C in the cytoplasm and PP2 (No. 58) which deactivates STAT3N in the nucleus also have a negative effect on the value of the objective function.

Similar argument can be made for important parameters of the signal transduction pathway model. The parameter k_{f7} (No.1) has the larger positive sensitivity than any other parameters. k_{f7} is involved in the reaction where STAT3 is activated by the receptor complex and large values of k_{f7} increase the rate of activation. The parameter k_{f32} (No.2) has the largest negative sensitivity. k_{f32} is involved in the reaction where SHP2 enables signal transduction through the MAPK pathway which limits the transduction through the JAK/STAT pathway.

It is important to point out that the concentration of the cytokine IL-6 (No. 51) has only a mildly positive effect as is determined by local sensitivity analysis. The reason for this is that the nominal value of the input is so large that the cells are saturated with IL-6 and a change in the value of IL-6 will only have a minor effect on the output. However, it should be pointed out that this behavior will be very different if the IL-6 concentration were lower by an order or magnitude or more.

4.4 Conclusion

Selection of parameters which can be estimated accurately from data is a prerequisite for successful estimation. While it is straightforward to perform parameter sensitivity analysis to determine a set of parameters to be estimated, it may happen that the determined set is not the best one for estimation. The reason for this is that results from local parameter sensitivity analysis depend upon the nominal values of the parameter, which are by definition not precisely known, and on values of the initial conditions and inputs. This section investigated these points as a family of parameter sets can be selected by the *D*-optimality criterion in combination with the orthogonalization method, where the optimization was performed based upon a genetic algorithm.

In a second step, the smallest perturbation required to change the optimal solution is determined to check if the results returned by the local method are acceptable. It has been illustrated in the case studies that the optimal solution can be extremely sensitive to parameter uncertainty and a more detailed analysis may be required. This analysis should start by determining which sources of uncertainty are affecting the value of an

optimality criterion. A method based upon global sensitivity analysis and another techniques based upon local sensitivity analysis of the criterion value are presented in this section. Furthermore, the mean criterion value and the probability for a subset to be the optimal one for a specified region of the parameter space are used to evaluate the chosen sets of parameters.

The result of the presented technique is a collection of candidate sets of parameters for estimation with detailed information about the effect of uncertainty in the parameter values, initial conditions, and inputs on the optimality criterion. The provided information is also helpful for evaluating data used for parameter estimation or designing future experiments.

5. INTEGRATING PARAMETER SELECTION WITH EXPERIMENTAL DESIGN UNDER UNCERTAINTY FOR NONLINEAR DYNAMIC SYSTEMS

5.1 Introduction

Parameter estimation involving large-scale dynamic models is an important but also a challenging task (Ljung, 1999; Nelles, 2001. Walter and Pronzato, 1997). One problem for parameter estimation is that complex models often contain dozens or even hundreds of parameters while only a limited amount of data is available as conducting experiments can be time consuming and costly. Therefore, many models of complex systems are often over-parameterized and not all the parameters are estimable in practice. If parameters are not practically estimable then a small amount of noise in the data will result in large variations of the estimated values of the parameters and the parameters cannot be estimated accurately (Walter and Pronzato, 1990). One solution to this is to select a subset of parameters to be estimated while all other parameters are fixed at a constant value.

Parameter selection has been used in a variety of applications. However, one important drawback of common parameter selection schemes based upon the FIM or sensitivity vectors is that these techniques depend upon the chosen values of the parameters for nonlinear systems, even though the exact values of the parameters are not known prior to estimation. It has been demonstrated that parameter uncertainty will have a significant effect on the parameter selection.

Another avenue for improving results obtained from parameter estimation is to collect a meaningful data set via experimental design. The objective of experimental design is to determine initial conditions and to adjust time-varying inputs to generate a data set with an optimal amount of information. The effect that uncertainty in the parameter values has on experimental design needs to be taken into account for nonlinear systems and robust strategies should be applied.

Parameter selection and experimental design are often considered separately, however, results from the two procedures affect each other for nonlinear systems:

Parameter set selection is highly dependent on the experiment condition while the experimental design is also dependent on the parameters selected for estimation. For example, it can happen that the best experiment design for a specific set of parameters may be a bad choice for another parameter set.

This section presents an integrated approach to parameter set selection and experimental design which also takes parameter uncertainty into account. This is achieved by formulating an optimization problem which is a mixed-integer nonlinear programming problem (MINLP) that optimizes a criterion of the FIM. As this is a non-trivial problem, a hybrid method combining a genetic algorithm (GA) and a simultaneous perturbation stochastic approximation (SPSA) is developed. The technique computes a collection of (sub-)optimal parameter sets, rather than a single optimal set, as well as the optimal experimental settings to estimate the sets.

5.2 Integrating selection of parameters with experimental design

This section first presents an example that illustrates the effect that parameter set selection, experimental design, and uncertainty in the model parameter values have on one another. This is followed by the formulation of the optimization problem whose result represents the solution of the integrated experimental design and parameter set selection procedure under uncertainty. The last subsection describes solution techniques used for solving this optimization problem.

Motivating example

Consider a system with one input, three parameters, and two output variables:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \theta_1 + \theta_2 + 5/12\theta_3 + 3u^2\theta_3 + 1/2\theta_3^2 \\ 4\theta_2 + 5/4\theta_3 - u^2\theta_3 + 3/2\theta_3^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \tag{5-1}$$

where $\mathbf{y}=[y_1,\ y_2]^T$ are output variables, $\boldsymbol{\theta}=[\theta_1,\ \theta_2,\ \theta_3]^T$ are parameters, u is an input variable determining the experimental condition, and $\boldsymbol{\varepsilon}=[\varepsilon_1,\ \varepsilon_2]$ represent noise with a Gaussian distribution with

$$\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0},\ \mathrm{Var}(\boldsymbol{\varepsilon}) = \mathbf{I}.$$

Not all three parameters can be determined uniquely since the sensitivity matrix

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \begin{bmatrix} 1 & 1 & 5/12 + 3u^2 + \theta_3 \\ 0 & 4 & 5/4 - u^2 + 3\theta_3 \end{bmatrix} \tag{5-2}$$

is column-rank deficient.

As the sensitivity matrix has a rank of two, two parameters are selected for estimation. There are three possible combinations of parameters to be estimated and the *D*-criterion for each possible set of parameters is given by

$$\begin{aligned} \varphi_{1,2} &= 4\ln(2), \\ \varphi_{1,3} &= 2\ln\left(3\theta_3 - u^2 + 5/4\right), \\ \varphi_{2,3} &= \ln\left(\theta_3^{\,2} + (26u^2 + 5/6)\theta_3 + 169u^4 + 65/6u^2 + 25/144\right), \end{aligned} \tag{5-3}$$

where $\varphi_{i,j}$ denotes the criterion value of the set consisting of parameters i and j. Using Bayesian statistics, the unknown parameters can be regarded as random variables and some distribution function can be used to characterize the parameter uncertainty. In this example the parameters are assumed to be uniformly distributed from 0 to 2 and the nominal value is assumed to be the mean value ($\overline{\theta}_i = 1$). The input is assumed to be in the range from -1 to 1 with a nominal value of zero, $\overline{u} = 0$.

Since the criterion value is a function of the parameters and the input variables, there are several possibilities for computing a function value. One approach is to determine the criterion with all parameters and inputs set to their nominal value:

$$\overline{\varphi} = \varphi\left(\overline{u}, \overline{\boldsymbol{\theta}}\right). \tag{5-4}$$

The most commonly used methods for parameter selection make use of Eq. (5-4), which assumes that the nominal values of the parameters are close to the true values and that the input variables have only a minor effect on the criterion value. However, this assumption may not be accurate and a good set of parameters evaluated at their nominal values may become suboptimal for other values. Instead, it is better to use the mean criterion value over the uncertain range of the parameters:

$$\mathrm{E}_{\boldsymbol{\theta}}[\varphi] = \mathrm{E}_{\boldsymbol{\theta}}\left[\varphi(\bar{u}, \boldsymbol{\theta})\right]. \tag{5-5}$$

A parameter set which has a large mean criterion value has a good average performance over the uncertainty range of the parameters. The criteria from Eq. (5-4) and Eq. (5-5) have so far not taken into account that the input variables can be changed. When integrating parameter set selection and experimental design, the mean criterion values have to be evaluated at their optimal input trajectory, which may be different for each parameter set:

$$\mathrm{E}_{\boldsymbol{\theta}}[\varphi]^{*} = \mathrm{E}_{\boldsymbol{\theta}}\left[\varphi(u^{*}, \boldsymbol{\theta})\right] \text{ with } u^{*} = \arg\max_{u} \mathrm{E}_{\boldsymbol{\theta}}\left[\varphi(u, \boldsymbol{\theta})\right]. \tag{5-6}$$

Table 5-1

Evaluation of each subset of parameters using different criteria

Conditions	$\varphi_{1,2}$	$\varphi_{1,3}$	$\varphi_{2,3}$
$u = \bar{u},\ \boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$; Eq. (5-4)	2.77	2.89	0.70
$u = \bar{u}$, average over $\bar{\boldsymbol{\theta}}$; Eq. (5-5)	2.77	2.69	0.50
$u = u^{*}_{i,j}$, average over $\bar{\boldsymbol{\theta}}$; Eq. (5-6)	2.77	2.69	5.34

Table 5-1 lists the criterion values for Eq. (5-4, 5-5, 5-6) of each parameter set given by Eq. (5-3). It can be concluded from Table 5-1 that the nominal criterion indicates that the parameter set $\{\theta_1, \theta_3\}$ is the optimal choice for estimation. However if parameter uncertainty is taken into account then the parameter set $\{\theta_1, \theta_2\}$ is the best choice. If experimental design is considered in addition to uncertainty in the parameter values then the parameter set $\{\theta_2, \theta_3\}$ is the optimal choice for parameter estimation. The fact that different evaluations of the criteria result in selecting different parameters demonstrates that parameter selection is highly dependent on the experimental condition. Additionally,

uncertainty in the parameter values can not be neglected as it can also have a significant impact on the results.

Problem formulation

For simple models, like the illustrative example, it is possible to determine the optimal parameter set and the optimal experimental design analytically. However, this is almost never the case in practice where more complex nonlinear dynamic systems are found. This section describes the problem formulation whose solution will result in the criterion given by Eq. (5-6).

Variables of a model that affect parameter estimation can be classified as belonging to one of the following four categories: (i) time-varying input variables, $\mathbf{u}(t)$, that can be manipulated; (ii) time-invariant inputs, $\mathbf{v}$, that can only be adjusted at the beginning of an experiment and will remain constant thereafter; (iii) parameters, $\boldsymbol{\theta}$, whose values are not known and need to be estimated; and (iv) unknown factors, $\boldsymbol{\delta}$, whose values are not known and will not be estimated. The Fisher information matrix, $\mathbf{F}$, is a function of these four types of variables

$$\mathbf{F} = \mathbf{F}(\mathbf{u}(t), \mathbf{v}, \boldsymbol{\theta}, \boldsymbol{\delta}) . \tag{5-7}$$

To evaluate the FIM some knowledge about all four kinds of variables is required. Although the value of the parameters and the unknown factors can not be obtained accurately, some *a priori* information about their uncertainty such as the range or distribution of their values is often available. These two types of variables can then be described by random variables according to some distributions based on the knowledge of their uncertainty. The criterion function should always be evaluated over the uncertainty range of the parameters and unknown factors instead of at their nominal values.

The values of the inputs determine the experimental conditions for generating the data set to be used for parameter estimation. Since the inputs can be manipulated they should be varied such that an information rich data set is obtained. It has been shown that the selection of parameters is dependent on the experimental design while at the

same time the optimal values of the input variables is also dependent on the parameters selected for estimation. Therefore, parameter set selection and experimental design need to be performed simultaneously.

A new formulation of the parameter set selection and experimental design problem is required to take the effect that these four types of variables have on the FIM into account. Eq. (5-8) describes the resulting optimization problem: The objective function is the expectation of the criterion value based upon the FIM over a range of values for $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$. The first two constraints are the system equations while the third and the fourth constraints are the sensitivity equations. The sensitivity matrix is formed by combining the sensitivity values at different time points. Some columns of the sensitivity matrix are selected according to the decision variable $\mathbf{z}$ to compute the FIM only for the parameters to be selected. The number of parameters per set, n_z, can be determined by singular value decomposition of the sensitivity matrix. The input variables $\mathbf{u}(t)$ and $\mathbf{v}$ determine the experimental conditions.

$$
\begin{aligned}
&\mathbf{z}^*, \mathbf{u}(t)^*, \mathbf{v}^* = \arg\max_{\mathbf{z},\mathbf{u}(t),\mathbf{v}} \ \mathop{\mathrm{E}}_{\boldsymbol{\theta},\boldsymbol{\delta}}\left[\varphi\left(\mathbf{F}(\mathbf{z},\mathbf{u}(t),\mathbf{v},\boldsymbol{\theta},\boldsymbol{\delta})\right)\right] \\
\text{s.t.} \quad & \frac{d\mathbf{x}}{dt} = \mathbf{f}\left(\mathbf{x},\mathbf{u}(t),\mathbf{v},\boldsymbol{\theta},\boldsymbol{\delta}\right), \quad \mathbf{x}(0) = \mathbf{x}_0\left(\mathbf{v},\boldsymbol{\delta}\right) \\
& y = h\left(\mathbf{x},\mathbf{u}(t),\mathbf{v},\boldsymbol{\theta},\boldsymbol{\delta}\right) \\
& \frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}, \quad \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}(0) = \mathbf{0} \\
& \frac{\partial y}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \frac{\partial h}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + \frac{\partial h}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \\
& \mathbf{S} = \left[\begin{array}{ccc} \dfrac{\partial y(t_1)}{\partial \boldsymbol{\theta}} & \cdots & \dfrac{\partial y(t_n)}{\partial \boldsymbol{\theta}} \end{array}\right]^{\mathrm{T}} \\
& \mathbf{F}=\mathbf{L}^{\mathrm{T}}\left(\mathbf{S}^{\mathrm{T}}\mathbf{S}\right)\mathbf{L} \\
& \mathbf{L} = \left[\begin{array}{cccc} \mathbf{e}_{i_1} & \mathbf{e}_{i_2} & \cdots & \mathbf{e}_{i_m} \end{array}\right], \text{ with } i_j \text{ that } z_{i_j} = 1 \\
& \mathbf{e}_i \text{ is the } i\text{-th column of the identity matrix} \\
& \mathrm{z}_i \in \{0,1\} \text{ and } \sum_i z_i = n_z \\
& \mathbf{L}_{\mathbf{u}} \le \mathbf{u}(t) \le \mathbf{U}_{\mathbf{u}} \\
& \mathbf{L}_{\mathbf{v}} \le \mathbf{v} \le \mathbf{U}_{\mathbf{v}} \\
& \boldsymbol{\theta} \text{ is random vector with density function } p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \\
& \boldsymbol{\delta} \text{ is random vector with density function } p_{\boldsymbol{\delta}}(\boldsymbol{\delta}).
\end{aligned}
\tag{5-8}
$$

As the manipulated variables $\mathbf{u}(t)$ are a function of time belonging to an infinite-dimensional function space, it is required to convert this infinite-dimensional problem into a finite-dimensional one by parameterizing the input variables.[35] Various expressions can be used and a common one is to describe each $u_i(t)$ by a polynomial with parameters $a_{i,j,k}$:

$$
u_i(t) = a_{i,j,n}t^n + a_{i,j,n-1}t^{n-1} + \cdots + a_{i,j,1}t + a_{i,j,0}, \quad t \in T_j, \tag{5-9}
$$

where $u_i(t)$ is the i-th input variable and T_j is the j-th time interval. For simplicity parameterization by the zero order polynomial is often used in practice. The vector $\mathbf{u}$

$$
\mathbf{u} = \left[\begin{array}{cccc} a_{1,1,1} & \cdots & a_{i,j,k} & \cdots \end{array}\right]^{\mathrm{T}} \tag{5-10}
$$

is used to denote the coefficients parameterizing the input variables and will replace $\mathbf{u}(t)$ in the optimization problem given by Eq. (5-8).

After parameterization of the input variables the optimization problem results in a mixed-integer nonlinear programming problem (MINLP). These types of optimization problem are generally not trivial to solve. Furthermore, two additional aspects have to be taken into account that increase the complexity of the problem: (1) The objective function includes an expectation and it may not be possible to evaluate this expectation exactly if the number of uncertain factors, inputs, and parameters is large; and (2) One is generally less interested in determining a single optimal set of parameters to be estimated but rather in obtaining a collection of parameter sets that have a high criterion value. If the values of the criterion have similar magnitudes for several sets of parameters, then the choice of which set to use for parameter estimation can be made based upon insight into the system. Since it is not possible to accurately describe the uncertain factors, it is a reasonable assumption to choose any of the parameter sets and its corresponding experimental conditions that results in a large criterion value.

One approach to evaluate the expectation is to numerically integrate the value over all uncertain factors. In this case, the determination of the continuous decision variables becomes a nonlinear programming problem and existing software such as LOQO (Vanderbei and Shanno, 1999), UOBYQA (Powell, 2002), SNOPT (Gill et al., 2002), or IPOPT (Wachter and Biegler, 2006) can be used. The solution of the nonlinear programming problem can then be coupled with solution of the binary programming problem that selects the parameters to be estimated.

However, numerical integration over the uncertain factors and parameters is computationally demanding. The first order sensitivity values of the parameters are required to calculate the criterion value of the FIM. For dynamic systems the sensitivities are calculated by solving a set of differential equations. These sensitivity equations need to be solved for a number of values in the uncertain range of the parameters and uncertain factors to compute a value of the expectation. Computation of the expectation at each iteration of the optimization is a task that becomes too computationally intensive for large number of uncertain factors and parameters. One alternative to this is to use a method of the stochastic approximation. While stochastic

methods also have their own set of drawbacks, they can be applied for determining approximate solutions of optimization problems of a significant scale. Since the goal for selecting parameter sets is not to come up with one optimal set, but rather to return a collection of sets that are good candidates for parameter estimation, there is no significant drawback to finding an approximate solution from using a stochastic technique. SPSA is computationally inexpensive as it is a derivative-free method that only requires two criterion values to approximate the gradient in each iteration step (Chin, 1997).

Since a stochastic optimization method is used for determining the continuous variables in this section, it is sensible to also use a stochastic technique for determining the discrete variables. Genetic algorithm will be used for the discrete variables as they return a population of possible solutions as a result of the algorithm. This property is consistent with the aim to determine multiple sets for parameter selection. Also, since **z** is a vector of binary variables, no reformulation is required for determining **z** via a GA.

A hybrid method combining GA and SPSA

A hybrid heuristic method which integrates GA and SPSA is developed in this subsection. A GA is used to update the discrete decision variable **z** while SPSA updates the continuous variables **u** and **v**. Since the focus is on how to combine the two algorithms, a basic implementation of each algorithm is used. In this hybrid algorithm the GA schedules which parameter sets will have their input variables updated by the SPSA. The fitness function is computed after the input variables have been updated in order to generate a new generation for the GA. If a parameter set is removed from the current generation, the information about the number of iteration steps which have been completed by the SPSA and the determined input trajectory are recorded. If a previously removed parameter set reappears in a later generation of the GA, then the last recorded input trajectory is used as the starting point for SPSA and not a nominal trajectory. A diagram of the algorithm is shown in Fig. 5-1.

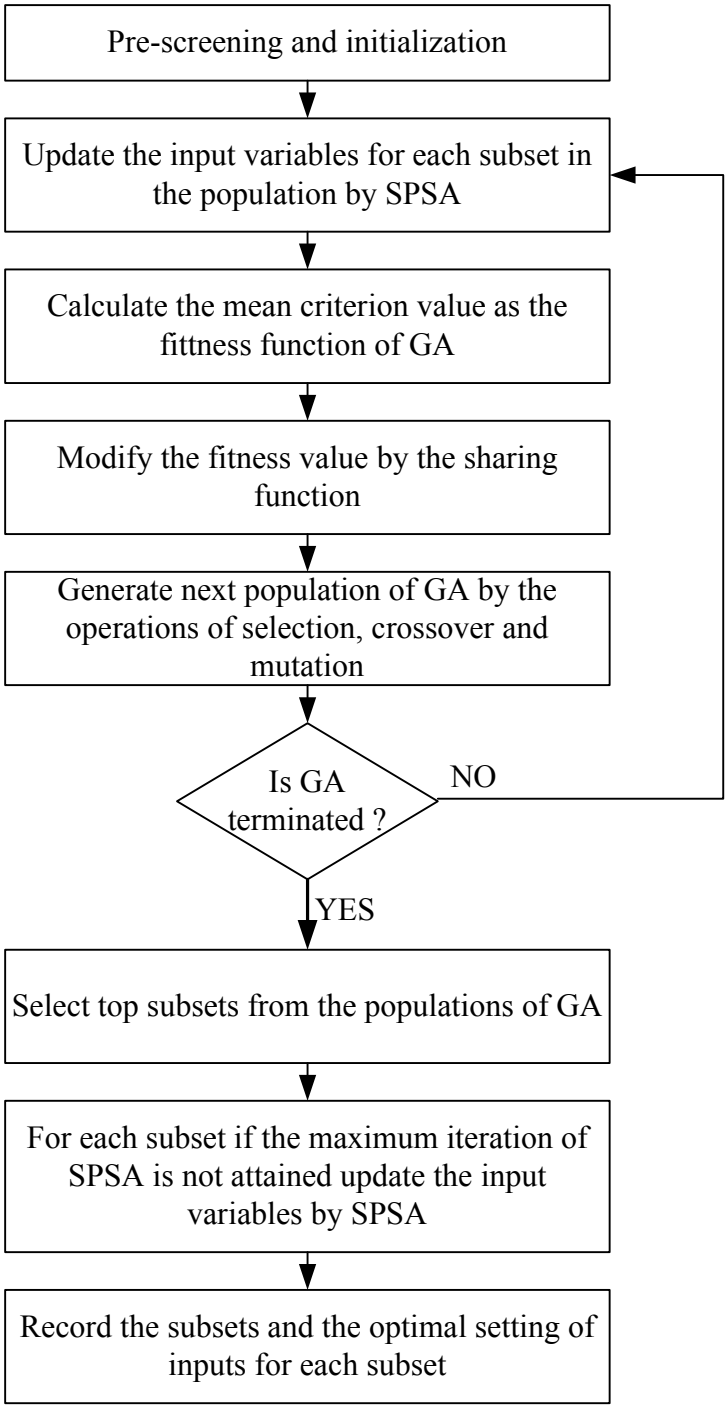


Fig. 5-1. The procedure for integrating parameter set selection with experimental design.

The input variables for each parameter set in the current GA population are updated by SPSA. Since parameter sets with large fitness values are more likely to remain in the population, they have a larger chance to have their input variable profiles updated by SPSA. When the input variables of a parameter set are updated, the mean criterion value over the uncertain variables should be evaluated. At each step of SPSA, the criterion values at 2 different sampling points of parameters are. The criterion value is averaged across different iterations as an approximation of the mean criterion value. As the input variables converge to an optimal value, the differences among the input values between two successive iterations are reduced. Accordingly, the averaged criterion value computed from different iterations will approach the averaged criterion value at the optimal input.

Each parameter set in the current population has its input variables updated by the SPSA. However, the SPSA does not to determine the optimal input trajectory for each parameter set in the current population. Instead, the SPSA performs several iterations to improve the input trajectories for each parameter set in the current population. Due to this is it ensured that a parameter set with a high fitness value which has large probability to appear in the population of GA will have its input trajectory updated frequently resulting in a good approximation of its optimal input trajectory. On the other hand, not too much computation time is wasted on parameter sets that are likely to be removed from a population due to their low fitness value. The number of iterations performed by SPSA to update the continuous variables during each generation depends upon the update history of the parameter set and on the generation number. The reason for this is that the values of the input variables are likely far from their optimal values during the first few generations and can change significantly, whereas the input variable profiles will only require minor modifications for parameter sets that have remained in the population for several generations.

To clarify the procedure of iterating between the GA and the SPSA, an artificial case is presented in Fig. 5-2. Three parameter sets determined by the GA are chosen for update by SPSA. Each parameter set is associated with a value of the vector variable z,

mean criterion value φ, input variable u and the current iteration number n for SPSA. The number n denotes how many iterations of SPSA are used to update the input variables for each parameter set in the current population of GA. SPSA updates the profile of the input variable u, i.e., in this case a scalar value, and the iteration number of SPSA is changed from 0 to 10. Next, the updated input variables are used to calculate the mean criterion value for each parameter set and the mean criterion values represent the fitness functions used by the GA to generate the second generation. Since parameter set 3 has the smallest criterion value, it will be removed from the second generation. However, its information is retained. The input variable u for each parameter set in the population is again updated by SPSA and the mean criterion value for each parameter set is recalculated at the new value of the input variable. GA uses the criterion value to generate the next generation and the parameter set 3 is reintroduced into the population. Since the information of parameter set 3 is preserved, the input variable can be updated from the already recorded information by SPSA.

When the solution is near the optimal one, fewer iterations of SPSA are required to update the decision variables. For example, 10 iterations of SPSA are performed for each parameter set in the 1^{st} generation. However, as the input variables approach their optimal values the number of iterations can be reduced to decrease the computational effort. Only 5 iterations of SPSA are performed for parameter sets 1 and 2 in the 2^{nd} generation of the GA since the input trajectories of these parameter sets have already been updated once. Since parameter set 4 is first introduced in the 2^{nd} iteration, its input trajectory is updated by 10 iterations of SPSA. When the input variables of parameter sets 1 and 2 are updated in the third generation then the number of iterations by SPSA can be further reduced to 3. This procedure is repeated until the input trajectories have converged to desired values or until the maximum number of generations of the GA has been reached. In this example the number of iterations for SPSA is determined however the number can be adapted by the improvement of the objective function.

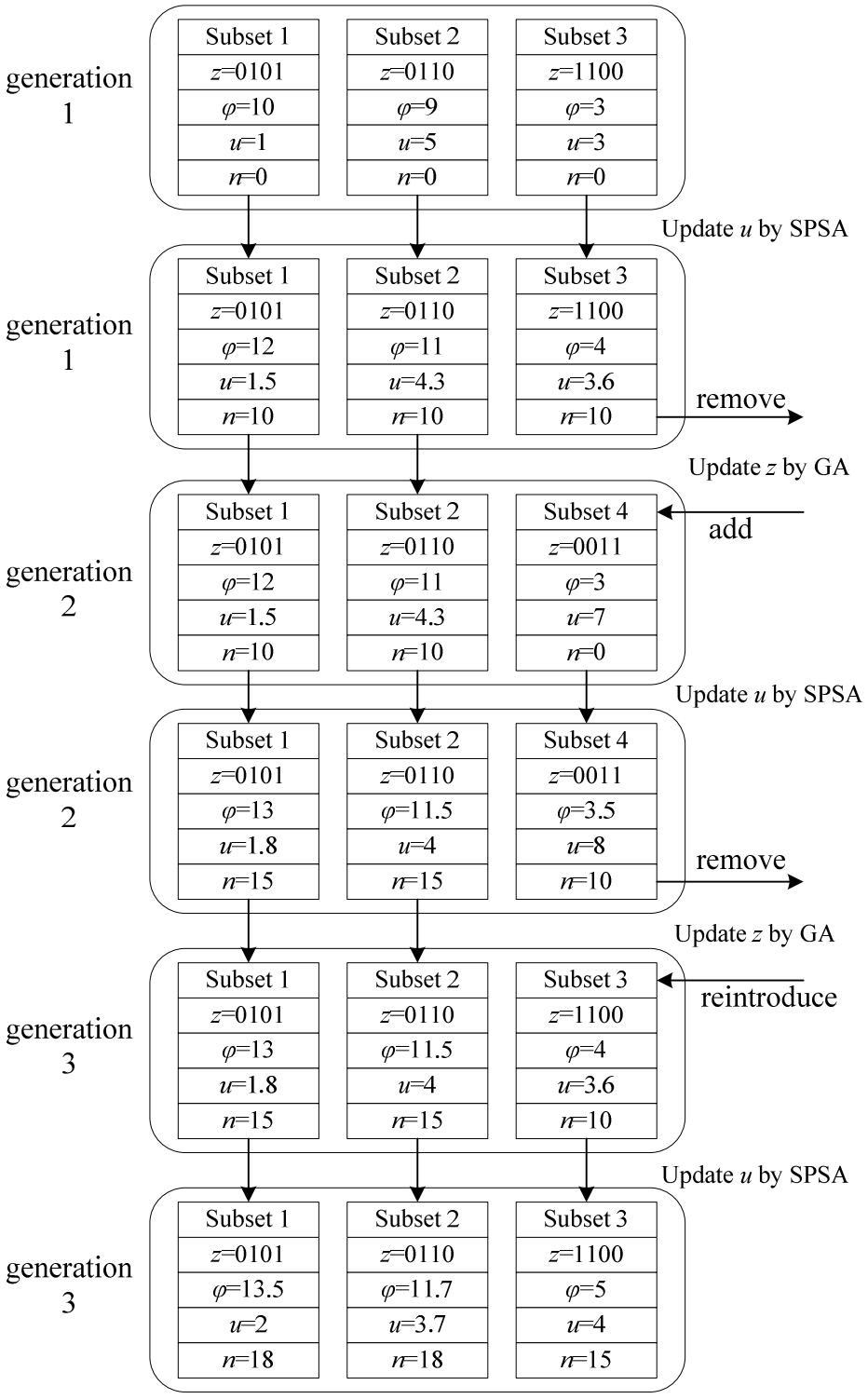


Fig. 5-2. Illustration of steps of algorithm by using an example problem.

5.3 Case studies

This section presents two detailed case studies illustrating the presented procedure. The first case study deals with a continuously-stirred tank reactor (CSTR) while the second case study involves a model of a signal transduction pathway.

Application of the procedure to a CSTR

This model describes an exothermic CSTR in which a first-order reaction A→B is taking place (Muske and Georgakis, 2003)

$$A \rightarrow B, \quad R_A = k \exp(-E/RT) c_A . \tag{5-11}$$

The reactor is described by the following differential equations

$$\begin{aligned} \dot{c}_A &= \frac{F}{V}(c_A^f - c_A) - R_A \\ \dot{T} &= \frac{F}{V}(T^f - T) + \frac{\Delta H}{C_P} R_A - \frac{hA}{C_P V}(T - T_c) . \\ \dot{T}_c &= \frac{F_c}{V_c}(T_c^f - T_c) + \frac{hA}{C_{P_c} V_c} \end{aligned} \tag{5-12}$$

The three states of the system are the concentration of component A, the temperature of the reactor, and the temperature of the coolant jacket. The reactor temperature is chosen as the only output of the system.

The variables in the system belong to one of 5 categories listed in Table 5-2. The first 9 variables shown in Table 5-2 are the parameters considered for estimation. The initial concentration is not measured and belongs to the category of unknown factors. The two inlet flows can be manipulated and they are the time-varying input variables. The coolant temperature can be manipulated as well but its value is constant as it belongs to the category of time-invariant input variables. The last three variables are known parameters which will not be considered in the following analysis.

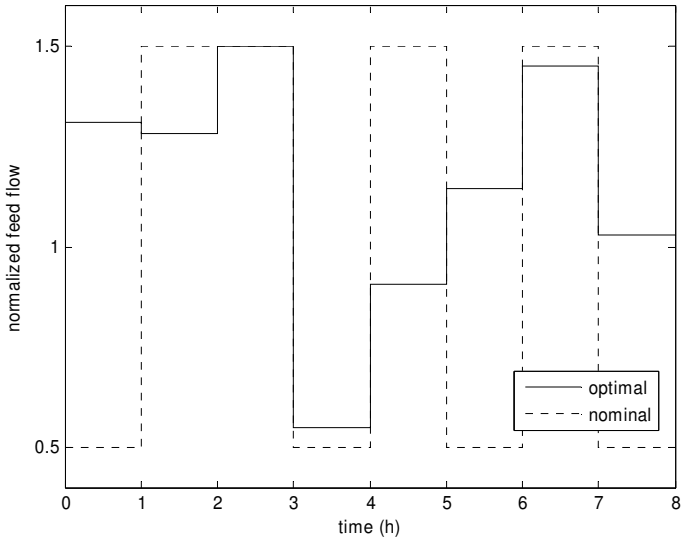
Table 5-2

Nominal values of variables in the CSTR model

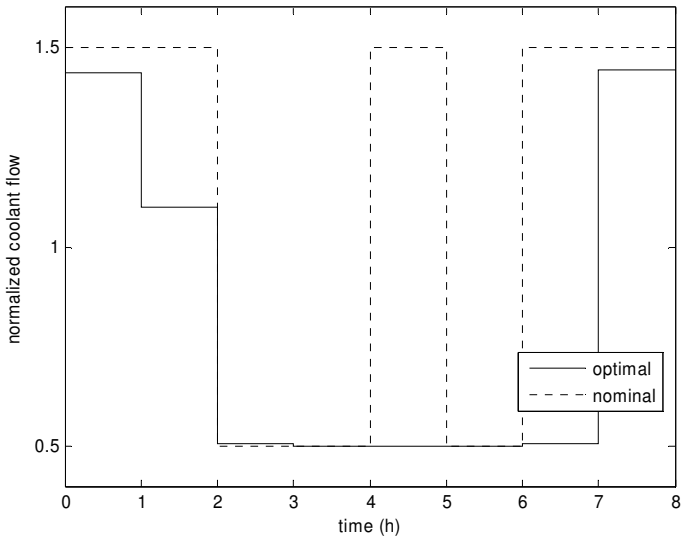
Parameter	Variable	Nominal Value	Type	symbol
Feed temperature	T^f	20 ^oC	Parameters	$\boldsymbol{\theta}$
Feed concentration	c^f_A	2500 mol/m^3		
Fluid heat capacity	C_P	1600 kJ/ $m^3 \bullet {}^oC$		
Heat of reaction	ΔH	160 kJ/mol		
Activation energy	E/R	255 K		
Preexponential factor	k	2.5 h^{-1}		
Coolant inlet temperature	T^f_c	10 ^oC		
Coolant heat capacity	C_{Pc}	1200 kJ/ $m^3 \bullet {}^oC$		
Heat transfer coefficient	h	1100 W/$m^2 \bullet {}^oC$		
Initial concentration	c_{A0}	1000 mol/m^3	Unknown factor	$\boldsymbol{\delta}$
Feed flow rate	F	0.1 m^3/h	Time variant input variables	$\mathbf{u}(t)$
Coolant flow rate	F_c	0.15 m^3/h		
Initial coolant temperature	T_{c0}	30 ^oC	Time invariant input variables	$\mathbf{v}$
Reactor volume	V	0.2 m^3	Known parameters	
Cooling jacket volume	V_c	0.055 m^3		
Heat transfer area	A	4.5 m^2		

All the variables are normalized by their nominal values to remove the possibility that scaling affects the procedure. The uncertain parameters and the unknown factor are assumed to be uniformly distributed in the range from 25% to 175% of their nominal values. It is assumed that the input variables can be changed from 50% to 150% of their nominal values. It is possible to use distributions other than uniform distributions for describing the uncertainty without modifying the procedure. The zeroth order polynomial is used to parameterize the input variables. The time horizon for collecting data is 8 hrs and it is assumed that the manipulated input variables can be changed every hour.

A singular value decomposition of the FIM is computed for the nominal values of the parameters and a pre-determined input profile to determine the number of parameters to be estimated. The first singular value is 0.8, followed by 0.1 and all other singular values are smaller then 0.05. Accordingly, it is appropriate to set the number of parameters to be estimated to two (n_z=2).



(a)



(b)

Fig. 5-3. Input trajectories. (a) Feed flow rate F; (b) Coolant flow rate F_c.

An exhaustive search over all parameter sets can be performed and a genetic algorithm is not required for this example as the total number of parameter sets containing two parameters is only 36. The optimal input trajectories can be computed by SPSA for each subset of parameters. The values of the parameters for SPSA were chosen to be: α=1, γ=1/6, a=1, c=0.2, A=100. The maximal number of iteration is set to 500. The algorithm is implemented in Matlab and the computation time for determining the optimal input trajectories for a parameter set is approximately 3 minutes on a computer with a P-IV CPU and 2 GB of memory. The time dependent profiles of the two input variables, the feed flow (F) and the coolant flow (F_c), are shown in Fig. 5-3. Additionally, the time invariant input variable, T_{c0}, is set to 60% of its nominal value.

The values of the three criteria are listed for all parameter sets in Table 5-3. Column 3 shows the optimal mean criterion values calculated according to Eq. (5-6). Column 4 contains the mean criterion values calculated according to Eq. (5-5) and Column 5 shows the nominal criterion values calculated according to Eq. (5-4). The nominal inputs for all results in columns 4 and 5 are those shown in Fig. 5-3. The procedure has been repeated several times to ensure that the results are reproducible as using stochastic optimization techniques and computing an expectation of a criterion over a set of uncertain parameters introduces stochastic elements into the procedure. The results of these repeated numerical experiments was that the parameter set to be estimated remains unchanged and only minor differences can be found in the criterion values.

Table 5-3

Criterion values for all parameter sets consisting of 2 parameters

No.	Subsets	Optimal mean *D*-criterion	Mean *D*-criterion	Norminal *D*-criterion
1	C_{Pc}, h	6.70	6.06	5.20
2	T^f_c, h	6.46	6.11	5.78
3	ΔH, h	6.26	5.29	4.98
4	T^f_c, C_{Pc}	6.25	5.75	5.67
5	ΔH, C_{Pc}	6.20	5.09	4.92
6	c^f_A, C_{Pc}	5.99	4.51	4.61
7	c^f_A, h	5.94	4.78	4.62
8	C_P, C_{Pc}	5.86	4.57	3.79
9	C_P, h	5.57	4.32	3.43
10	k, C_{Pc}	5.42	4.27	4.18
11	k, h	5.26	4.16	4.03
12	E/R, C_{Pc}	4.74	3.68	3.83
13	E/R, h	4.45	3.56	3.68
14	C_P, T^f_c	2.91	2.18	1.57
15	T^f, C_{Pc}	2.40	1.45	1.34
16	T^f, h	2.34	1.51	1.24
17	ΔH, T^f_c	2.12	1.34	1.05
18	c^f_A, T^f_c	1.94	1.24	1.37
19	C_P, ΔH	1.84	1.14	0.52
20	k, T^f_c	1.82	1.25	1.37
21	c^f_A, C_P	1.74	0.74	0.28
22	E/R, T^f_c	1.12	0.62	1.00
23	C_P, k	0.55	-0.53	-1.16
24	c^f_A, k	0.27	0.01	0.22
25	c^f_A, ΔH	0.22	0.02	0.34
26	ΔH, k	-0.29	-0.44	-0.32
27	C_P, E/R	-0.37	-1.16	-1.54
28	c^f_A, E/R	-0.40	-0.58	-0.14
29	ΔH, E/R	-0.92	-0.98	-0.63
30	T^f, T^f_c	-1.32	-1.70	-1.66
31	T^f, C_P	-2.16	-3.04	-3.62
32	T^f, ΔH	-2.64	-2.74	-2.86
33	T^f, c^f_A	-2.86	-3.21	-3.00
34	T^f, k	-3.80	-3.96	-4.12
35	T^f, E/R	-4.46	-4.61	-4.51
36	E/R, k	-5.54	-6.72	-6.55

Application of the procedure to a signal transduction network

A model of the JAK/STAT signaling pathway (Yamada et al., 2003) is used in this subsection to illustrate the techniques for parameter selection and experimental design for complex dynamic systems. Fig. 5-4 shows the structure of the signaling pathway under investigation. The model includes 32 state variables and 53 parameters. The input is the concentration of IFN-γ while the output is the concentration of STAT1n*-STAT1n* which is a transcription factor and can be indirectly measured using a green fluorescent protein reporter system. The reactions in the pathway are numbered and the parameter names are derived from the reaction number.

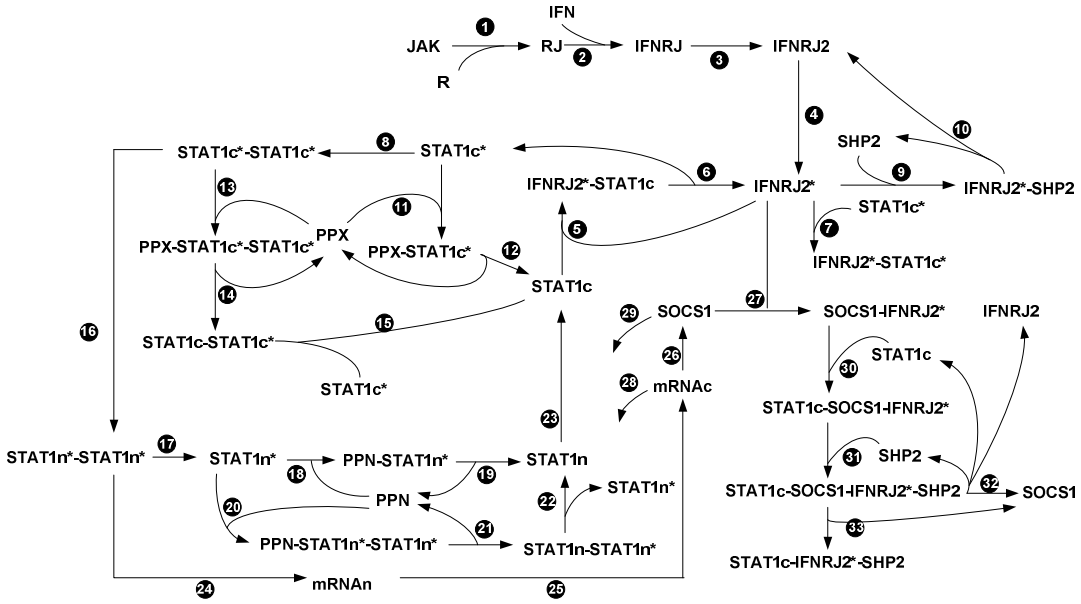


Fig. 5-4. Structure of the JAK/STAT signaling pathway.

All the variables are normalized by their nominal values for this case study. The uncertain parameters and the unknown factors are assumed to be uniformly distributed in the range from 50% to 150% of their nominal values. The input is assumed range from

50% to 150% of its nominal value. Little is known about the uncertainty distribution of these parameters and a uniform distribution is no more or less likely to accurately describe the parameter values than any other distribution. The experiment is performed over a period of 8 hrs and the input can be changed every 30 minutes and remains constant in between the changes.

A singular value decomposition of the FIM is used to determine the number of parameters to be estimated. Selecting four parameters for estimation is sufficient as the condition number of the FIM is large and the magnitude of the singular values drops significantly after the fourth singular value (n_z=4).

The presented algorithm is implemented in Matlab. The size of the population of the GA is set to 30, 3 elites are used, and the maximum generation number is set to 100. The parameter, *D*, of the sharing function is set to one. Roulette selection, scattered crossover, and uniform mutation are used. The following parameter values are chosen for SPSA: α=1, γ=1/6, a=0.7, c=0.2, A=100.

The input sequence is assumed to consist of 16 values as the variable can be changed every 30 minutes over a time horizon of 8 hrs. Determining an input sequence with 16 changes for a problem where a stochastic optimization method is used, will likely produce slightly different results each time the optimization is performed. Accordingly, a term that penalizes deviations of the input from its nominal value has been added to the objective function from Eq. (5-8):

$$\mathbf{z}^*, \mathbf{u} = \arg\max_{\mathbf{z},\mathbf{u}} \; \mathop{\mathrm{E}}_{\boldsymbol{\theta},\boldsymbol{\delta}}\left[\varphi\left(\mathbf{F}(\mathbf{z},\mathbf{u},\boldsymbol{\theta},\boldsymbol{\delta})\right)\right] - \lambda_u \left\|\mathbf{u}-\bar{\mathbf{u}}\right\|^2, \tag{5-13}$$

where $\bar{\mathbf{u}}$ is the nominal value of $\mathbf{u}$ and λ_u is a penalty coefficient for changes in the manipulated variable. Including this penalty term ensures that the input is only changed from the nominal value if such a change has a significant positive effect on the criterion value. Fig. 5-5 shows the mean input values and their standard deviations for one parameter set where the problem was solved ten times.

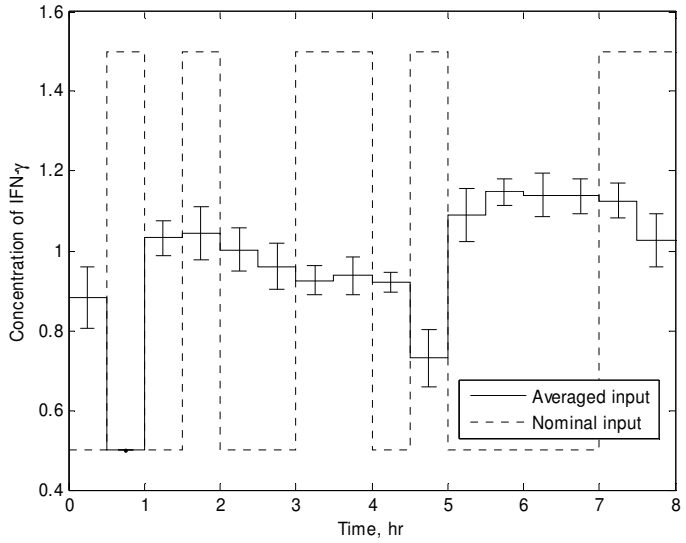


Fig. 5-5. Averaged input signal and error bars for ten solutions of the algorithm.

It required approximately 28 hrs of computation time on a computer with a P-IV CPU and 2 GB of memory to obtain the collection of 30 (sub-)optimal parameter sets shown in Table 5-4 and the corresponding optimal experimental conditions for the sets.

Table 5-4 lists the 30 parameter sets with the optimal mean criterion value for their respective experimental conditions (Eq. 5-6) shown in the third column. For comparison purposes, the mean criterion value (Eq. 5-5) at the nominal operating conditions (column 4) and the nominal criterion value (Eq. 5-4) are also shown in the table (column 5).

It can be seen there is a significant difference in the ranking of the individual parameter sets. The set consisting of $\{k_{f5}, k_{f6}, k_{f21}, k_{f29}\}$ has the largest nominal criterion value, however, it is only ranked at the 21st position by the mean criterion value. More importantly, 3 of the 4 parameters selected for this set by the nominal *D*-criterion are different from the parameters chosen by the optimal mean *D*-criterion.

Table 5-4

Criterion values for 30 selected parameter sets

No.	Subsets	Optimal mean *D*-criterion	Mean *D*-criterion	Norminal *D*-criterion
1	k_{f6}, k_{f19}, k_{f33}, k_{b30}	4.941	3.187	4.028
2	k_{f6}, k_{f19}, k_{f28}, k_{b30}	4.792	3.248	3.907
3	k_{f6}, k_{f21}, k_{f33}, k_{b30}	4.791	3.838	4.232
4	k_{f6}, k_{f19}, k_{f29}, k_{b30}	4.763	3.384	4.007
5	k_{f6}, k_{f21}, k_{f29}, k_{b30}	4.756	3.592	4.169
6	k_{f6}, k_{f19}, k_{f26}, k_{f33}	4.692	3.646	4.423
7	k_{f6}, k_{f19}, V_{m24}, k_{f33}	4.689	3.644	4.423
8	k_{f6}, k_{f19}, V_{m24}, k_{f29}	4.682	3.482	4.562
9	k_{f6}, k_{f19}, k_{f21}, k_{f29}	4.630	3.450	4.748
10	k_{f6}, k_{f19}, k_{f21}, k_{f28}	4.628	3.465	4.747
11	k_{f6}, k_{f19}, V_{m24}, k_{f28}	4.547	3.359	4.480
12	k_{f6}, k_{f19}, k_{f26}, k_{f29}	4.523	3.484	4.562
13	k_{f6}, k_{f21}, k_{f29}, k_{b18}	4.438	3.262	4.555
14	k_{f6}, k_{f21}, k_{f28}, k_{b18}	4.416	3.279	4.556
15	k_{f6}, k_{f18}, k_{f21}, k_{f28}	4.404	3.349	4.606
16	k_{f6}, k_{f19}, k_{f21}, k_{b30}	4.398	3.619	4.528
17	k_{f6}, k_{f21}, k_{f26}, k_{f28}	4.394	3.234	4.387
18	k_{f6}, k_{f18}, k_{f21}, k_{f33}	4.363	3.373	4.353
19	k_{f5}, k_{f6}, k_{f21}, k_{f29}	4.327	3.269	4.898
20	k_{f6}, k_{f21}, k_{f28}, k_{b30}	4.324	3.474	4.111
21	k_{f5}, k_{f6}, k_{f19}, k_{f28}	4.311	3.267	4.675
22	k_{f6}, k_{f18}, k_{f33}, k_{b30}	4.297	2.595	3.049
23	k_{f6}, k_{f21}, k_{f33}, k_{b18}	4.251	3.303	4.304
24	k_{f6}, k_{f19}, k_{f25}, k_{f33}	4.242	3.151	3.886
25	k_{f6}, k_{f21}, k_{b18}, k_{b30}	4.222	3.430	4.351
26	k_{f6}, k_{f19}, k_{f31}, k_{b30}	4.215	2.635	3.048
27	k_{f6}, k_{f18}, k_{f21}, k_{b30}	4.204	3.500	4.400
28	k_{f6}, k_{f21}, k_{f29}, K_{m24}	4.185	2.963	3.977
29	k_{f6}, k_{f16}, k_{f21}, k_{f29}	4.172	3.428	2.901
30	k_{f6}, k_{f21}, k_{f26}, k_{f33}	4.162	3.604	4.284

Another important observation is that the criterion value changes significantly, when the experimental conditions are optimized for a chosen parameter set. This is especially

important insofar as the nominal setting of the input values were chosen in a matter that ensures a reasonable level of excitation as the input was varied from its smallest to its largest values in pulses of varying duration. However, improving experimental design does not only affect the values of the criteria, but also the ranking of different parameter sets. The set $\{k_{f6}, k_{f19}, k_{f33}, k_{b30}\}$ has the largest optimal mean criterion value, however, when the input is fixed at the nominal point this parameter set is ranked the 26^{th} by the mean criterion value. It is also noted that this parameter set only has the 23^{rd} largest nominal criterion value. This exemplifies that some potentially good parameter sets may be missed if parameter uncertainty is neglected and if the effect of experimental design is ignored.

While it may seem trivial to determine if a set is the best or the 26^{th} best among hundreds of thousands of possible sets, it is important to point out that there are significant differences in the criterion values even among the best 30 sets shown in Table 5-4. This becomes even more important once it is recognized that the criterion value involves computation of the logarithm of the determinant of the FIM.

This example illustrates the complex nature of the optimization problem given by Eq. (5-8). Future work will focus on decomposition of the optimization problem to reduce the computational burden and enable application of the presented procedure to even larger models.

5.4 Conclusions

This section presented an integrated approach for selecting parameters for estimation and experimental design while taking uncertainty in the parameter values into account. Integrating these two approaches is important insofar as experimental design and selection of parameters to be estimated influence one another for nonlinear systems. Additionally, the nominal values of parameters that have yet to be estimated also have an effect on both experimental design and parameter set selection. The integrated approach formulates an optimization problem where the expectation of a criterion involving the Fisher information matrix is maximized by varying the parameters to be estimated and

the experimental conditions. This optimization problem is a MINLP which is non-trivial to solve. A hybrid method combining a genetic algorithm and a simultaneous perturbation stochastic approximation is developed to determine an approximate solution. The presented solution technique uses an iterative approach where the GA determines the discrete variables representing the set of parameters to be estimated, and the SPSA computes the values of the continuous variables, i.e., the experimental conditions.

One other aspect of the presented work in this section is that a collection of parameter sets, each with its own optimal experimental design, is determined, rather than one optimal result. The reason for this is that one may have a specific preference for estimating certain parameters or using specific experimental conditions, even though this may restrict the results and not be optimal. However, by providing a collection of solutions and a measure for the quality of the determined parameter set/experimental design, it is possible to make an informed decision about which result to use.

6. PARAMETER SET SELECTION VIA CLUSTERING OF PARAMETERS INTO PAIR-WISE INDISTINGUISHABLE GROUPS OF PARAMETERS

6.1 Introduction

Mathematical modeling continues to play a key role in various branches of engineering and science. The structure of these models is generally determined from insight into the system while the parameter values are taken from the literature or in some cases estimated from experimental data. While it would be preferable to estimate as many parameters from data as possible, the number of parameters in these models usually exceeds the number of those which can be reliably estimated from available data. If parameters are not identifiable then even a low level of noise in the data will result in large variations of the estimated value of the parameters and the parameters can not be estimated accurately. A commonly used regularization is to select a subset of parameters to be estimated while all other parameters are fixed at a constant value.

Most of the parameter selection approaches can be formulated as a combinatorial optimization problem, however, solving these optimization problems is nontrivial. For systems with only few parameters an exhaustive search can be used. However, the total number of possible combination of parameters is too large to be enumerated even for systems with only a few dozen parameters. Stochastic search techniques such as genetic algorithms can provide a solution for larger systems, however, convergence of these algorithms is not guaranteed. Another approach is to use a sequential approach where one parameter is selected at a time. The main disadvantage of these types of approaches is that some combinations of parameters may give better results, but could be excluded because of parameters selected at earlier steps.

The methods mentioned above focus on searching the space of possible parameter sets, a procedure that is strongly affected by the number of parameters of the model. This section presents a different approach for parameter selection. The number of parameters to be considered is first reduced by determining several groups of parameters where the parameters within a group are pair-wise indistinguishable, i.e., they cannot be estimated

together. It is then possible to only consider one parameter per group for the parameter set selection procedure. This technique significantly reduces the combinatorial problem resulting from a large number of parameters and enables solution of the parameter set selection problem using existing approaches.

This section develops the analytic and numerical methods for sorting parameters into groups. It is shown that parameters in an analytical pair-wise indistinguishable set can be re-parameterized by a new parameter. A procedure for carrying out this step numerically is also shown. While the numerical procedure cannot guarantee pair-wise indistinguishability, it does have the advantage that it can be used to lump parameters with very similar effects into the same group.

Clustering of parameters into groups can be viewed as some form of model reduction as the number of parameters to be considered is reduced in the process. The difference of the behavior of the original model with all parameters and the model with a reduced parameter set can be described by a measure. The magnitude of this measure can be controlled by the number of groups of parameters to be considered as is investigated in this section. It is illustrated in an example that the presented procedure can not only find an adequately good solution compared to the forward selection or an approach based upon a genetic algorithm but it can also give important insight into the effect that parameters have on the model output.

6.2 Parameter set selection via clustering of parameters into pair-wise indistinguishable groups of parameters

As the number of parameters in many fundamental models far exceeds the number of parameters that can be accurately estimated from available data, it is necessary to determine a subset of parameters which can be estimated. Parameter selection can be viewed as a special case of model reduction as only the values for some parameters are determined from parameter estimation while all other parameters are assumed to remain at their constant value.

This section presents two techniques for determining subsets of parameters to be estimated. The first technique is an analytical approach which derives the condition for which the output of a system with fewer parameters is identical to the one for the entire parameter set. The second method is a numerical approach which does not require that the outputs are identical but instead investigates the error bound that results from including fewer parameters in a model.

Analytical approach

Determining which parameters can be lumped in a model is a problem that is related to parameter identifiability. Each time a parameter is not locally identifiable, it is possible to reduce the parameter by setting it to a constant value. If the effects that two parameters have on the output are identical then each parameter may be individually identifiable, however, only one of the two parameters needs to be considered and the other can be set to a constant value. The following definition and proposition provide the mathematical description of this situation.

Definition 6-1: A parameter set is said to be a pair-wise indistinguishable set when any two parameters in the set are not locally identifiable.

It can be seen that if the two parameters are not locally identifiable then their sensitivity matrix is rank deficient. This implies that the sensitivity vectors of two parameters are parallel. The effect that variations of any parameter in a pair-wise indistinguishable set have on the output can be compensated by changing any other parameter of the set while the others were set to the nominal value. The statement is backed up by Proposition 6-1.

Proposition 6-1.

Assume the output $\mathbf{y} \in \mathbb{R}^m$ is an analytical function of the parameter $\boldsymbol{\theta} \in \mathbb{R}^n$

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\theta}), \tag{6-1}$$

and $\bar{\boldsymbol{\theta}}$ is a nominal value of the parameter. If the sensitivity value of the output with respect to θ_i (i=1,…,r) is nonzero

$$\frac{\partial f_s}{\partial \theta_i} \neq 0, \quad i = 1, \cdots, r \quad \text{and} \quad s = 1, \cdots, m. \tag{6-2}$$

then for any $\boldsymbol{\theta}$ in a neighborhood of $\bar{\boldsymbol{\theta}}$, there exist a function $\psi(\boldsymbol{\theta})$ such that

$$\mathbf{f}(\theta_1, \cdots, \theta_{r-1}, \theta_r, \theta_{r+1}, \cdots, \theta_n) = \mathbf{f}(\bar{\theta}_1, \cdots, \bar{\theta}_{r-1}, \psi(\theta_1, \cdots, \theta_n), \theta_{r+1}, \cdots, \theta_n) \tag{6-3}$$

if and only if the sensitivity vectors of the output with respect to θ_i (i=1,…,r) are parallel to each other

$$\frac{\partial \mathbf{f}}{\partial \theta_i} + \alpha_i \frac{\partial \mathbf{f}}{\partial \theta_r} = 0, \qquad i = 1, \cdots, r-1 \tag{6-4}$$

where α_i is a function of $\boldsymbol{\theta}$. (The proof of Proposition 6-1 can be found in Chu and Hahn, 2009.)

While two parameters in a pair-wise indistinguishable set may have zero sensitivity value, i.e., each parameter is not locally identifiable, this case would violate the condition of nonzero sensitivity values from Proposition 6-1. As this is a trivial case, it can be easily excluded by checking the sensitivity values for each parameter.

Proposition 6-1 states that the effects that changes of the values of parameters in a pair-wise indistinguishable group have on the outputs can be lumped. As a result of this, only one parameter is needed to represent the group and all other parameters can be fixed at their nominal values. Accordingly, the output function can be re-parameterized to have fewer parameters, where one possible re-parameterization is given by Proposition 6-1.

Illustrative example

A simple nonlinear regression model is used to illustrate the presented analytical procedure. Let

$$\mathbf{f}(\theta_1,\theta_2,\theta_3)=\begin{bmatrix}\theta_1\theta_2+\theta_3\\ \theta_1\theta_2\theta_3\end{bmatrix}=\mathbf{g}(h,\theta_3), \tag{6-5}$$

where it can easily be seen that a substitution $h=\theta_1\theta_2$ can be made. However, this result is derived here using the procedure presented above. The sensitivity vectors for θ_1 and θ_2 are computed to be

$$\frac{\partial\mathbf{f}}{\partial\theta_1}=\begin{bmatrix}\theta_2\\ \theta_2\theta_3\end{bmatrix} \text{ and } \frac{\partial\mathbf{f}}{\partial\theta_2}=\begin{bmatrix}\theta_1\\ \theta_1\theta_3\end{bmatrix}. \tag{6-6}$$

These two sensitivity vectors are parallel and are related by the following differential equation:

$$\frac{\partial\mathbf{f}}{\partial\theta_1}-\frac{\theta_2}{\theta_1}\frac{\partial\mathbf{f}}{\partial\theta_2}=0. \tag{6-7}$$

Eq. (6-7) can be used to compute the re-parameterization of h. The characteristic ordinary differential equation is given by

$$\frac{d\theta_1}{1}=\frac{d\theta_2}{-\theta_2/\theta_1}. \tag{6-8}$$

which can be solved by separation of variables and the solution is

$$\theta_1\theta_2=C, \tag{6-9}$$

where C is a constant. A first integral is $h=\theta_1\theta_2$ and it is chosen as the variable for re-parameterizing the model. By fixing one parameter, like $\theta_1=\bar{\theta}_1$, it can be obtained

$$\mathbf{f}(\theta_1,\theta_2,\theta_3)=\mathbf{f}(\bar{\theta}_1,\theta_1\theta_2/\bar{\theta}_1,\theta_3). \tag{6-10}$$

Change of the two parameters, θ_1 and θ_2, can be replaced by changing only θ_2 while θ_1 is fixed at the nominal value.

Numerical approach

The procedure presented in the last section results in a set of characteristic ordinary differential equations which need to be solved. As it is rarely possible to analytically solve this expression, a numerical approach to clustering parameters into groups is presented here. This numerical approach does not require the sensitivity vectors to be

parallel, however, the angle between the sensitivity vectors should be small. In this case the parameters can be viewed as being pair-wise indistinguishable with a certain numerical precision.

A similarity measure of the effect of two parameters on the output can be defined by

$$\cos\phi_{ik} = \frac{\left|\mathbf{s}_i^{\mathrm{T}}\mathbf{s}_k\right|}{\left\|\mathbf{s}_i\right\|_2 \left\|\mathbf{s}_k\right\|_2}, \tag{6-11}$$

where $\phi_{ik} \in [0, \pi/2]$ is the angle between the sensitivity vectors $\mathbf{s}_i$ and $\mathbf{s}_k$. The value of the similarity measure ranges from zero to one where a value of unity indicates that the two vectors are parallel to one another and that the two parameters cannot be distinguished. A value of zero refers to the sensitivity vectors of the parameters being orthogonal, i.e., the parameters have a distinct effect on the outputs. It should be noted that the absolute value is used for the similarity measure as it is of little importance which orientation the sensitivity vectors have.

The parameters can be clustered into groups based upon the similarity measure. Agglomerative hierarchical clustering is used here since it is easy to determine from the hierarchical tree how many groups the parameters should be clustered into. However, other clustering algorithms (Duda et al., 2006; Theodoridis and Koutroumbas, 2006) could be used with only a minor difference in the outcome.

Agglomerative hierarchical clustering forms groups by repeatedly merging different groups of parameters. Initially, each parameter is in a group by itself. In a second step, the two groups with the largest similarity measure are merged into a new group. The similarity within a group can be controlled by the number of groups that one chooses to have.

Since the parameters in a numerically pair-wise indistinguishable set have similar effects on the output, a parameter in the set can be selected as the representative for the group. Parameter set selection is simplified by this procedure as the number of parameters can be reduced to the number of groups. However, since the sensitivity vectors of the parameters in a group are not perfectly parallel, it has to be taken into

account that there will be a discrepancy between the parameter-output effect of the original system and the one with a reduced number of parameters.

This discrepancy can be measured by the prediction gap between the two functions

$$d(\boldsymbol{\theta}) = \min_{\boldsymbol{\psi}} \left\| \mathbf{f}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\psi}) \right\|_2, \tag{6-12}$$

where **f** is the original output function and **g** is derived from **f** when only one parameter per group is considered and all other ones are fixed at their nominal values. The individual parameters of $\boldsymbol{\psi}$ are the representative parameters for each group and are a subset of $\boldsymbol{\theta}$. The prediction gap indicates how well a model with a reduced parameter set can approximate the behavior of the original model.

It is non-trivial to compute this discrepancy for general nonlinear functions. Due to this an approximation of *d* based upon linearization is used in this section. The truncated Taylor series approximation of the original function, **f**, with respect to the parameters is given by

$$\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\bar{\boldsymbol{\theta}}) + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}), \tag{6-13}$$

and the approximated Taylor expression of the function **g** is

$$\begin{aligned} \mathbf{g}(\boldsymbol{\psi}) &\approx \mathbf{g}(\bar{\boldsymbol{\psi}}) + \frac{\partial \mathbf{g}}{\partial \boldsymbol{\psi}^{\mathrm{T}}} (\boldsymbol{\psi} - \bar{\boldsymbol{\psi}}) \\ &= \mathbf{f}(\bar{\boldsymbol{\theta}}) + \sum_i \frac{\partial \mathbf{f}}{\partial \theta_{s_i}} (\psi_i - \bar{\theta}_{s_i}) \end{aligned} \tag{6-14}$$

where s_i is the index for the remaining parameters. The approximation of the discrepancy becomes

$$\mathbf{f}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\psi}) = \mathbf{S}\mathbf{x} - \mathbf{T}\mathbf{y}, \tag{6-15}$$

where $\mathbf{S} = \left. \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right|_{\bar{\boldsymbol{\theta}}}$, $\mathbf{T} = \left. \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}_{\mathbf{s}}^{\mathrm{T}}} \right|_{\bar{\boldsymbol{\theta}}}$, $\mathbf{x} = \boldsymbol{\theta} - \bar{\boldsymbol{\theta}}$ and $\mathbf{y} = \boldsymbol{\psi} - \bar{\boldsymbol{\theta}}_{\mathbf{s}}$. The discrepancy is dependent on the value of parameters and the worst case can be considered

$$d = \max_{\|\mathbf{x}\|_2 = 1} \min_{\mathbf{y}} \left\| \mathbf{S}\mathbf{x} - \mathbf{T}\mathbf{y} \right\|_2. \tag{6-16}$$

Since the discrepancy may increase unbounded with an increased in the length of **x,** a constraint is placed on the length of **x**.

The similarity of parameters within a group can be controlled by determining the number of groups for the parameter set. In the extreme case where each group only contains one parameter, the discrepancy between the original function and the one with a reduced parameter set is zero. However, the discrepancy will increase as fewer groups are used and the similarity within groups decreases. It will be shown in the following that the discrepancy can be bounded by a decreasing function of the least similarity value found in a group.

Proposition 6-2. Let $\mathbf{s}_k$ be the k-th column vector of the matrix **S**. Then the discrepancy

$$d = \max_{\|x\|_2=1} \min_{y} \left\| \mathbf{S}\mathbf{x} - \mathbf{s}_k y \right\|_2 \tag{6-17}$$

can be bounded by

$$d \le \sqrt{1-\cos^2 \phi_s} \sqrt{\sum_{i \ne k} \left\| \mathbf{s}_i \right\|_2^2}\,, \tag{6-18}$$

where $\cos\phi_s$ is the smallest similarity value in the group. (The proof of Proposition 6-2 is in Chu and Hahn, 2009.)

Proposition 6-2 provides a bound for one group. Summation of the upper bounds for each group results in an upper bound for the entire problem.

Proposition 6-3. Let $\mathbf{s}_{l_i}$ be the l_i-th column of the matrix **S** where l is the index of the group and i is the index of the sensitivity vector in a group. The discrepancy (Eq. 23) can be bounded by

$$d \le \sqrt{\sum_l d_l^2}\,, \tag{6-19}$$

where d_l is the discrepancy of the l-th group. (The proof of Proposition 6-3 is Chu and Hahn, 2009.)

Application of determining pair-wise indistinguishable groups of parameters for parameter estimation

One important step for parameter estimation is to select the set of parameters to be estimated. It is possible to formulate the parameter set selection procedure as an optimization problem, such as

$$\begin{aligned} &\mathbf{z}^* = \arg\max_{\mathbf{z}} \ \log\det\left(\mathbf{F}(\mathbf{z})\right) \\ \text{s.t. } &\mathbf{F}(\mathbf{z}) = \mathbf{F}^{(i_1,\cdots i_{n_s})}_{(i_1,\cdots i_{n_s})} \text{ with } i_j \text{ that } z_{i_j} = 1,\ j = 1\cdots n_s \\ &z_1 + z_2 + \cdots + z_n = n_s \\ &z_i \in \{0,1\},\ i = 1\cdots n. \end{aligned} \tag{6-20}$$

where the decision vector $\mathbf{z} \in \{0,1\}^n$ denotes whether a parameter is included in the selected parameter subset. If z_i=1 then θ_i belongs to the selected subset with the size of n_s. The matrix $\mathbf{F}$ is the Fisher information matrix of all parameters. $\mathbf{F}(\mathbf{z})$ is the Fisher information matrix of the parameters included in the selected subset and it is equal to the principal submatrix of $\mathbf{F}$ with the indices of the non-zero decision variables (the entries of column i_j and row i_k, $j, k = 1 \ldots n_s$).

The optimization problem given by Eq. (6-20) is nontrivial to solve as the number of possible combinations of parameters grows drastically with the number of parameters in the problem. Reducing the number of parameters to be considered can significantly reduce the computational burden. The parameter clustering algorithm can be used as described in the previous subsection. Since only one parameter per set can be reliably estimated from data, only one parameter per group needs to be considered. Even though it is possible to select any parameter in a group as the representative parameter, the parameter with the greatest length of the sensitivity vector is selected by the algorithm. The reason for doing so is that the value of a parameter with a large sensitivity vector would need to be changed by a smaller amount during the parameter estimation procedure than the value of a parameter with a small sensitivity vector, even though both parameters would be estimated from the same data set. For example, if the sensitivity vector of one parameter is an order of magnitude larger than the one for another parameter in the same set, then the estimated value for the change in the parameters

would be an order of magnitude larger for the parameter corresponding to the smaller sensitivity vector. Since it can be assumed that the initial values of the parameters are reasonably close to their true values, it follows that estimating the parameters which require a smaller adjustment of their values is an acceptable approach.

The number of binary variables from Eq. (6-20), i.e., the number of parameters in the problem is then reduced from n, the number of all parameters to n_g, the number of groups.

Algorithm of parameter selection base on parameter clustering

Step 1. Calculate the sensitivity vectors of the outputs with respect to the parameters.

Step 2. Determine n_s, number of parameters in the subset that will be estimated, by singular value decomposition of the sensitivity matrix.

Step 3. Set parameters whose sensitivity vectors have small length (e.g., less than 5% of the largest one) to their nominal values.

Step 4. Cluster the parameters into n_g ($n_g \geq n_s$) groups by hierarchical clustering based upon the similarity measure from Eq. (6-11).

Step 5. For each group select the parameter which has the largest sensitivity vector as the representative of the group.

Step 6. Select n_s parameters from n_g representatives to optimize the criterion function by solving the optimization problem given by Eq. (6-20).

The number of parameters per set from Step 2 can be determined by the rank of the sensitivity matrix. Each column of the sensitivity matrix is a sensitivity vector of a parameter. The number of columns is equal to the number of parameters. However, due to correlation between parameters, the sensitivity matrix may be ill-conditioned. The rank of the sensitivity matrix can be determined by the number of singular values greater than a certain threshold. The value of the threshold should be problem-specific, however, it is possible to use some rule of thumb. If there is a gap of an order of magnitude or

more between the singular values then it is appropriate to choose the number of parameters to be estimated equal to the number of singular values that are larger than the cut-off determined by this gap. Step 3 represents a simple methodology for reducing the parameter set as no parameter with a small length of the sensitivity vector needs to be considered. Step 4 performs clustering of the remaining parameters into groups. The decrease of the number of groups clustered will reduce the computation resources for the optimization problem while increasing the discrepancy between the reduced model and the original model. The number of groups should be determined as the smallest one which can let the discrepancy less than a threshold value. The parameter with the largest sensitivity vector for each group is chosen as the representative of this group in Step 5. Step 6 selects the parameters to be considered for solution of the optimization problem from Eq. (6-20) by taking one parameter per cluster as described in Step 5.

The presented technique can significantly reduce the computational burden for solving the optimization problem given by Eq. (6-20) as the computational effort for solution of this problem grows drastically with the number of parameters to be considered.

6.3 Case study

To illustrate the technique presented in this section, a model of a signal transduction pathway for hepatocytes stimulated by Interleukin-6 is used which is updated from Singh et al., 2006. The model, shown in Fig. 6-1, contains two pathways: Janus-associated kinases & signal transducers and transcription factors are activated in one pathway while the other pathway involves the activation of mitogen-activated protein kinases. This model consists of 66 nonlinear ordinary differential equations and includes 115 parameters. The state variables are the concentrations of the proteins in the pathway and the input variable is the concentration of Interleukin-6 outside of the cell that initiates signal transduction. The output variable is the concentration of $(STAT3N^*)_2$ (dimer of activated STAT3 in the nucleus) as this transcription factor can be indirectly measured using a green fluorescent protein (GFP) reporter system. A detailed description

of the original version of the model and the nominal values of the parameters can be found in the literature (Singh et al., 2006, Chu and Hahn, 2009), however, the model has been updated to describe the mechanism that SOCS3 and SHP2 compete for the same binding site on the receptor.

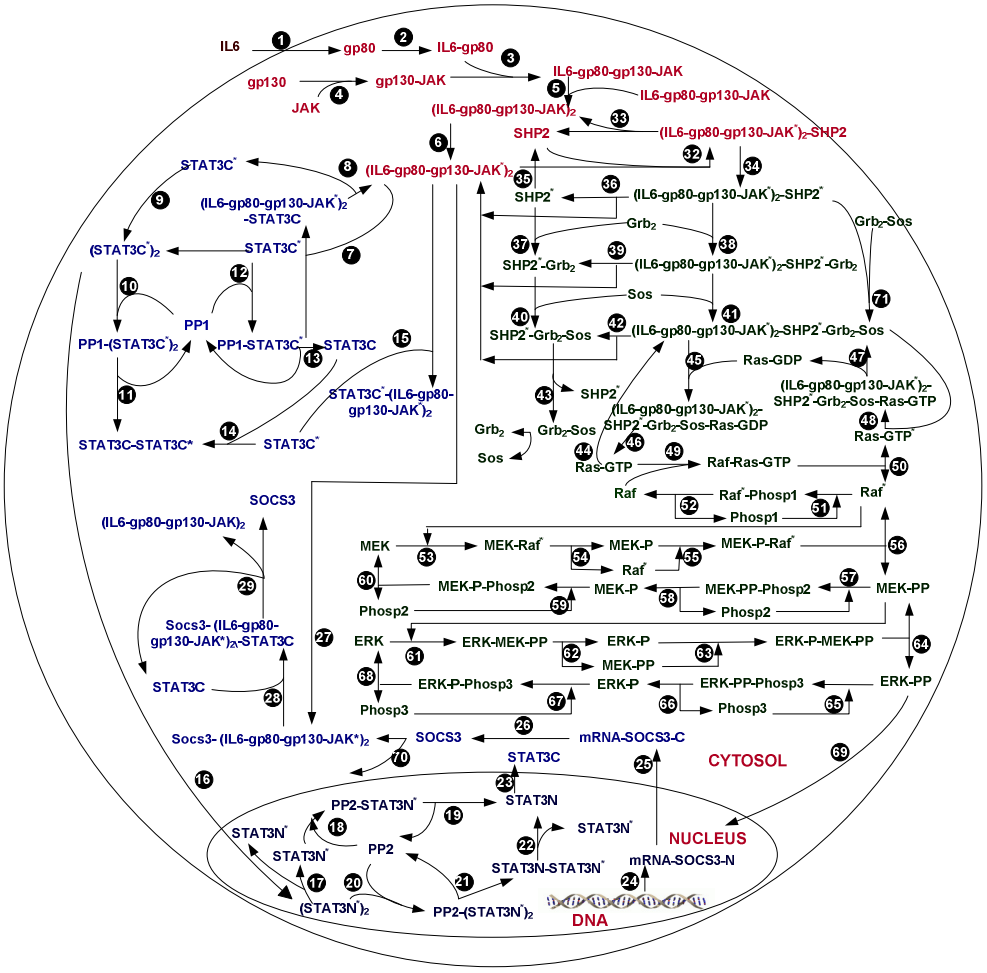


Fig. 6-1. Model of the Interleukin-6 signaling pathway.

The Fisher information matrix is computed in a first step. The sensitivity value is sampled every minute during the time interval from 0 to 12 hr to form the sensitivity vector. Singular value decomposition of the sensitivity matrix determines that the 9^{th} through 115^{th} singular values are close to zero. Accordingly, a parameter set consisting of 8 parameters will be selected. In a second step, the lengths of the sensitivity vectors are analyzed. 70 of the parameters have sensitivity vectors with a length that is less than 5% of the length of the largest sensitivity vector. These 70 parameters will be set to their nominal values and not considered further. The problem to be solved turns into a problem where a combination of 8 parameters needs to be chosen from a set of 45 parameters such that the *D*-optimality criterion is maximized. If an exhaustive search were to be performed then the number of possible parameter sets that would have to be evaluated would be $\sim 2\times 10^{8}$. For the purpose of comparison, the forward selection (the orthogonalization method), a solution of the optimization problem via genetic algorithm, and the clustering method introduced in this section are applied and discussed in the following.

Fig. 6-2 shows the dendrogram of hierarchical clustering of parameters. It can be concluded that the similarity values between some of the parameters is very high as their sensitivity vectors are almost parallel. The diagram also illustrates how the selection of the similarity value influences the number of group. For example, if 11 groups are used then the smallest similarity value is equal to 0.941 which is illustrated by the dashed line. An increase in the number of groups leads to an increase of the lowest similarity value of the system.

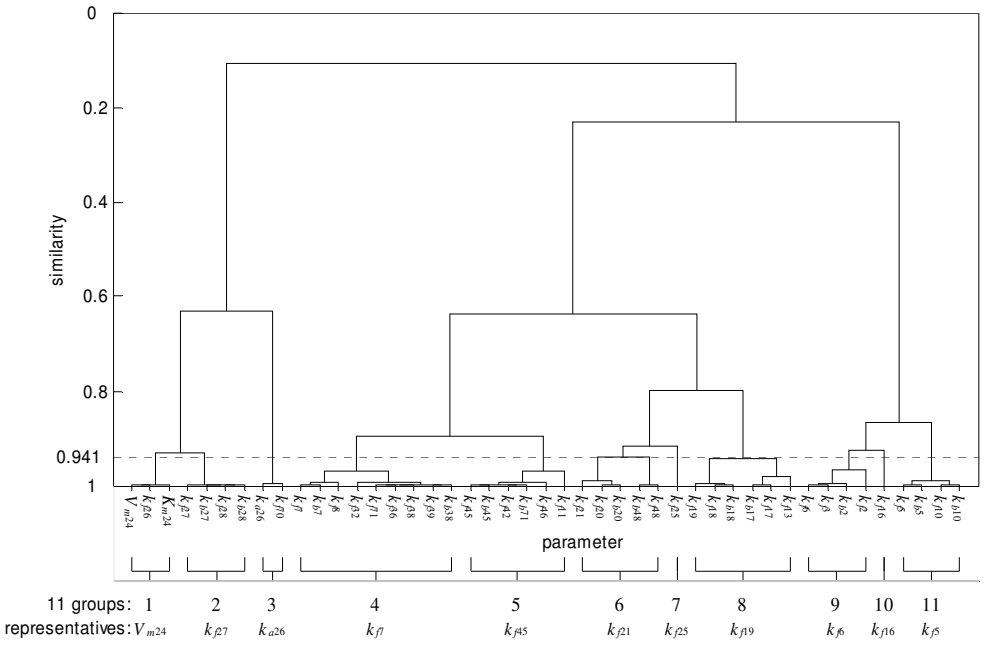


Fig. 6-2. Dendrogram of hierarchical clustering of parameters.

Reducing the parameter space via parameter clustering can be viewed as one type of model reduction. The discrepancy between the original model and the reduced model is important as it indicated how many groups need to be selected to appropriately represent the original model. As discussed in Proposition 6-2 and Proposition 6-3, the model discrepancy can be bounded by the least similarity measure. At the same time, the smallest similarity measure can be determined by choosing the number of groups from the dendrogram in Fig. 6-2. Therefore the number of groups can be determined by assuming the discrepancy to be less than a certain threshold value. Table 6-1 lists the least similarity measure and the discrepancy value for different number of groups. If the parameters are clustered into more groups, then the least similarity measure is increased and the discrepancy value is decreased. In this case the number of groups is determined to be equal to 11 as the discrepancy value drops below 0.05.

Table 6-1

Results of the clustering method for different number of groups

No. of groups	8	9	10	11	12	13	14	15
Least similarity	0.917	0.925	0.930	0.941	0.943	0.966	0.968	0.969
Discrepancy	0.240	0.149	0.095	0.046	0.045	0.021	0.02	0.02

The parameter with the largest sensitivity vector in each group is chosen as the representative parameters for the group. The parameters selected for estimation are now chosen from the set of 11 representative parameters instead of the original 45 parameters. The optimization problem has reduced to determining a set of 8 parameters out of 11 possible parameters to maximize the *D*-optimality criterion of the Fisher information matrix, as compared to the original problem that involved choosing a set of 8 parameters out of 45 parameters. The computational effort decreases significantly, from $\sim 2\times10^8$ possible combinations to 165, due to this reduction in the number of parameters that need to be considered.

Table 6-2

Results by the three methods

	Parameters selected	Criterion
Clustering	k_{f7}, k_{a26}, k_{f21}, k_{f19}, k_{f6}, k_{f45}, k_{f25}, V_{m24}	4.391
Forward selection	k_{f7}, k_{f21}, k_{f70}, k_{f16}, k_{f27}, k_{f19}, k_{f42}, k_{f25}	3.918
Genetic algorithm	k_{f7}, k_{a26}, k_{f21}, k_{f19}, k_{f6}, k_{f45}, k_{f25}, V_{m24}	4.391

The results returned by the presented algorithm need to be put in a proper context of what would be achieved using other method for determining parameter sets. In order to do this, a comparison between the presented technique, the forward selection, and solution of the *D*-optimality criterion optimization problem via a genetic algorithm has been performed. The summary of these results is shown in Table 6-2. It can be seen that the presented technique outperforms the forward selection as the forward selection adds

one parameter at a time to the parameter set and does not take into account that combinations of parameters may give better results than choosing one parameter at a time. As a result of this, the forward selection procedure does not determine a set of parameters that maximizes the *D*-optimality criterion.

The genetic algorithm returned the same results as the technique based upon parameter clustering as the optimal solution, even though the GA determined the best 8 parameters from the set of 45 parameters that were still considered after pre-screening. While both the GA and the clustering technique returned the same results, it should be pointed out that a GA requires more computational effort and it is not guaranteed that a GA will converge to an optimal solution. Also, it is possible to combine the clustering technique with a GA that only considers one parameter per cluster. The main reason this was not performed in this example was because determining 8 parameters from a set of 11 is a small enough problem that an exhaustive search can easily be implemented.

While parameter selection for parameter estimation is a main application area for determining pair-wise indistinguishable parameters, there are other areas where the presented technique has value. For example, it is well known that robustness is a common property of biological networks, e.g., resilience to perturbations in current conditions. The kinetic parameters in a biochemical network can change due to alteration of enzyme activity caused by a mutation or a disease. However, biological networks have a certain tolerance regarding variations in kinetic parameters while still being able to maintain their functions. One important reason for robustness is redundancy built into a network. As some components of a network may fail, there are other components that have a similar effect which allows the network to function properly. Determining indistinguishable sets can reveal some of these redundancies inherent to a system.

Investigation of pair-wise indistinguishable sets can also identify key parts of a network which can be further investigated. While it is straightforward to determine important parameters by comparing the lengths of sensitivity vectors, it is non-trivial to compare the different effects that changes in sensitivity over time have on a system. It has been shown that parameters which have similar cumulative effects can have distinct

dynamic behaviors. It has also been recognized that distinct temporal activation profiles of the same signaling proteins can result in different gene-expression patterns and diverse physiological responses. Clustering of parameters involves investigation of similarity of time-dependent sensitivity profiles. It should be noted that some groups only contain one parameter, such as k_{f16}, k_{f25} and k_{f31} in the shown example. These parameters have distinct sensitivity profiles and their effects cannot be compensated for by other parameters.

6.4 Conclusions

This section presented a technique for determining which parameters should be estimated in a model. While it is possible to solve this problem by maximizing the *D*-optimality criterion of the Fisher information matrix, it has to be recognized that the computational effort required for solving such a problem grows drastically with the number of parameters under investigation. The technique presented in this section addresses this problem as parameters which have a similar effect on the outputs are clustered and only one parameter per group needs to be considered for solution of the optimization problem.

This section presented the underlying theory for determining pair-wise indistinguishable parameter sets and also developed an algorithm that can be used for determining the parameters to be estimated. The approach has been illustrated in a case study involving a signal transduction network. The technique was able to reduce the optimization problem from determining a set of 8 parameters out of a total of 115 parameters to finding a set of 8 parameters out of 11 parameters. A comparison with the results returned by forward selection and by a genetic algorithm has been made and it was found that the technique computed a better set of parameters than the one determined by forward selection. The results computed by the genetic algorithm were identical to the ones computed by the presented technique, even though the genetic algorithm investigated a much larger parameter space than the 11 parameters selected by

clustering. This example serves as an illustration of the technique as the optimal solution was found with very little computational effort.

7. IMPROVING PREDICTION CAPABILITIES OF MODELS OF COMPLEX BIOCHEMICAL REACTION NETWORKS VIA PARAMETER SELECTION AND ESTIMATION

7.1 Introduction

Parameter estimation is generally conducted by minimizing an objective function describing the difference between predicted model outputs and experimental measurements. A significant effort has been placed on developing computationally efficient parameter estimation techniques (Mendes and Kell, 1998; Rodriguez-Fernandez et al., 2006; Gennemark and Wedelin, 2007; Balsa-Canto et al., 2008). These techniques mainly focus on computational approaches for determining the solution that minimizes the objective function and do not address how sensitive the estimated values are to noise in the data. The importance of this point needs to be emphasized as models of biochemical reaction networks often contain a large number of correlated parameters while the experimental data is often scarce and noisy, resulting in not all parameters being identifiable in practice.

Estimation of an unidentifiable model is an ill-posed problem and the optimal solution is not unique or may not be stable (Aster et al., 2005). Noise in the data will lead to a large variation of the estimated parameter values. Since not all parameters can be estimated accurately, the focus of estimation of complex biochemical models should be placed on improvement of the accuracy of the model prediction (Gutenkunst et al., 2007; Piazza et al., 2008; Wilkinson et al., 2008; Dimelow and Wilkinson, 2009; Hlavacek, 2009), rather than on obtaining the best fit of the model to the data. In essence, a procedure that ensures that the model does not over-fit the data is required, similar to techniques found in system identification (Aster et al., 2005; Walter and Pronzato, 1997). The main difference between this approach and techniques used for system identification is that the model structure here is determined from insight into the system and it is known that the model will have more parameters than can be estimated.

A common regularization approach is to reduce the number of free parameters. This is achieved by selecting a subset of parameters that will be estimated while all other parameters are fixed at their nominal values. Determining which parameters should be estimated is called the parameter selection problem and plays a key role for the ability of a model to fit data. However, most parameter selection techniques concentrate on parameter identifiability with a lesser emphasis on accuracy of the model predictions. Even though the concepts are related, improving the prediction accuracy of a model is not equivalent to improving identifiability.

The aim of this paper is to develop a technique for determining parameter values from limited amount of potentially noisy data with the goal of enhancing the accuracy of model predictions. A forward selection method is presented which minimizes the mean squared error of the prediction. A comparison with the well-cited orthogonalization method (Yao et al., 2003; Lund and Foss, 2008) is made. Both techniques belong to a class of forward selection procedures, however, each has a different objective function which is minimized and different criteria are employed at each step for selecting parameters. The advantage of the presented method is that it produces a more accurate prediction than the orthogonalization method. An additional advantage is that the presented technique provides a criterion that determines how many parameters should be estimated and explicitly takes parameter uncertainty into account. A numerical experiment for estimation of parameters of an NF-κB signal transduction network (Lipniacki et al., 2004) is conducted to illustrate the presented method. It is shown that estimation of an appropriately selected subset of parameters is sufficient to fit the data and results in more accurate model predictions than estimating all parameters.

7.2 Motivating example

This section illustrates the differences of two parameter set selection procedures, one focusing on parameter identifiability and one dealing with prediction accuracy. It will also be highlighted that it may be appropriate to estimate only a subset of parameters of a model if the effects that the parameters have on the outputs are highly correlated.

Assume the following linear regression model for the illustrative example:

$$\tilde{\mathbf{y}} = \mathbf{S}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{7-1}$$

where the noise vector $\boldsymbol{\varepsilon}$ is Gaussian distributed with zero mean and a variance of $\sigma^2\mathbf{I}$ and the matrix $\mathbf{S}$ is given by

$$\mathbf{S}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 1 & a & 0 & \cdots & 0 & \cdots & 1 & a & 0 & \cdots & 0 \\ 1 & 0 & a & \ddots & \vdots & \cdots & 1 & 0 & a & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 & \cdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 1 & 0 & \cdots & 0 & a & \cdots & 1 & 0 & \cdots & 0 & a \end{bmatrix}_{m\times nm} . \tag{7-2}$$

The model has m free parameters and it is assumed that $n \cdot m$ data points are given and that the value of a is positive ($a>0$). When the value of a is small, then the columns of $\mathbf{S}$ are almost linearly dependent and many combinations of parameter values have similar output values resulting in an ill-conditioned model for estimation.

When all parameters are estimated, the mean squared error for the estimated parameters is

$$\mathrm{MSE}\left(\hat{\boldsymbol{\theta}}\right) = \mathrm{E}\left[\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)^{\mathrm{T}}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right)\right] = \frac{a^2 + 2(m-1)}{na^2}\sigma^2 \tag{7-3}$$

where $\hat{\boldsymbol{\theta}}$ is the estimated parameter value by the least squares method and $\boldsymbol{\theta}^*$ is the true value. The mean squared error of the prediction is

$$\mathrm{MSE}\left(\hat{\mathbf{y}}\right) = \mathrm{E}\left[\left(\hat{\mathbf{y}} - \mathbf{y}^*\right)^{\mathrm{T}}\left(\hat{\mathbf{y}} - \mathbf{y}^*\right)\right] = m\sigma^2 \tag{7-4}$$

where $\hat{\mathbf{y}}$ is the prediction under the parameter estimated and $\mathbf{y}^*$ is the true value. It can be seen that the two mean squared errors are distinctively different. When the value of a decreases the mean squared error of the parameters will increase, however, the mean squared error of the prediction is not dependent on the value of a. An experimental design which increases the value of a improves parameter identifiabilty but has no effect on the accuracy of output predictions. Furthermore, accurate prediction of the output can be more easily achieved than accurate estimation of the parameter values for over-parameterized models as the MSE shown in Eq. (7-4) is independent of a. This result

supports a recent observation that updating biochemical reaction networks should focus on the prediction rather than obtaining exact parameter values (Gutenkunst et al., 2007; Wilkinson et al., 2008).

Since the effects of parameters on the outputs are correlated, it is sufficient to only estimate a small number of parameters. In this case a small value of a indicates that it is sufficient to estimate only one parameter. For illustration purposes, the true parameter value is assumed to be $\theta_i^* = 1$, $i = 1 \cdots m$. If only the parameter θ_1 is estimated and the values of the other parameters are fixed at 0, which are different from the true values, then the mean squared error of the prediction is

$$\mathrm{MSE}(\hat{\mathbf{y}}_1) = \mathrm{tr}\left[\mathrm{E}\left(\hat{\mathbf{y}}_1 - \mathbf{y}^*\right)\left(\hat{\mathbf{y}}_1 - \mathbf{y}^*\right)^{\mathrm{T}} \right] = \sigma^2 + (m-1)na^2 \tag{7-5}$$

where $\hat{\mathbf{y}}_1$ is the prediction when only θ_1 is estimated.

Comparing the two mean squared errors from Eq. (7-4) and Eq. (7-5) it can be concluded that estimation of only one parameter is preferred to estimation of all parameters in the case of

- Highly correlated parameters, i.e., a small value of a
- Considerable noise levels, i.e., a large value of σ^2
- Limited amount of experimental data, i.e., a small value of n
- A large number of parameters, i.e., a large value of m

In fact, all of these conditions are commonly encountered for estimation problems involving complex models of biochemical reaction networks. While the chosen example was a generic one to illustrate a point, it is likely that benefits can be achieved from estimating only a subset of the parameters of a model if one or more of these conditions hold.

7.3 Parameter set selection procedure

This section introduces a parameter set selection approach that minimizes the mean squared error of the prediction. This approach bears some similarity with conventional parameter set selection methods as it is based on parameter sensitivity analysis. However,

unlike other methods, the presented technique aims to improve the accuracy of model output predictions rather than to improved parameter identifiability. Using the sensitivity matrix, the regression model can be linearized at the true value of the parameters

$$\mathbf{y}(\boldsymbol{\theta}) = \mathbf{y}(\boldsymbol{\theta}^*) + \mathbf{S}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) \tag{7-6}$$

where $\mathbf{S} = \partial \mathbf{g}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^{\mathrm{T}} \big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$. Without loss of generality, the parameter vector can be decomposed into two parts

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_s \\ \boldsymbol{\theta}_u \end{bmatrix} \tag{7-7}$$

where $\boldsymbol{\theta}_s$ is the vector of parameters selected for estimation and $\boldsymbol{\theta}_u$ is the vector of the remaining unselected parameters. Correspondingly the sensitivity matrix can be decomposed into two parts

$$\mathbf{S} = [\mathbf{T} \quad \mathbf{W}]. \tag{7-8}$$

The linearized model can be written as

$$\mathbf{y}(\boldsymbol{\theta}_s, \boldsymbol{\theta}_u) = \mathbf{y}(\boldsymbol{\theta}^*) + \mathbf{T}(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*) + \mathbf{W}(\boldsymbol{\theta}_u - \boldsymbol{\theta}_u^*). \tag{7-9}$$

If only the selected parameters are estimated and the unselected parameters are fixed at the nominal value $\overline{\boldsymbol{\theta}}_u$, then the model for estimation of the selected parameters becomes

$$\begin{aligned} \mathbf{y}_s(\boldsymbol{\theta}_s) &= \mathbf{y}(\boldsymbol{\theta}_s, \overline{\boldsymbol{\theta}}_u) \\ &= \mathbf{y}(\boldsymbol{\theta}^*) + \mathbf{T}(\boldsymbol{\theta}_s - \boldsymbol{\theta}_s^*) + \mathbf{W}(\overline{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u^*). \end{aligned} \tag{7-10}$$

The least squares estimate of the parameter is given by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s^* &= (\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\left[\tilde{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta}^*) - \mathbf{W}(\overline{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u^*)\right] \\ &= (\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\boldsymbol{\varepsilon} - (\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}(\overline{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u^*). \end{aligned} \tag{7-11}$$

The error between the prediction and the true value is obtained

$$\hat{\mathbf{y}}_s - \mathbf{y}(\boldsymbol{\theta}^*) = \mathbf{T}(\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\boldsymbol{\varepsilon} + \left[\mathbf{I} - \mathbf{T}(\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\right]\mathbf{W}(\overline{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u^*) \tag{7-12}$$

where $\hat{\mathbf{y}}_s = \mathbf{y}_s(\hat{\boldsymbol{\theta}}_s)$. The mean squared error is given by

$$\begin{aligned}\mathrm{MSE}\left(\hat{\mathbf{y}}_s\right) &= \mathrm{E}\left[\left(\hat{\mathbf{y}}_s - \mathbf{y}\left(\boldsymbol{\theta}^*\right)\right)^{\mathrm{T}}\left(\hat{\mathbf{y}}_s - \mathbf{y}\left(\boldsymbol{\theta}^*\right)\right)\right] \\ &= n_s\sigma^2 + \left(\overline{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u^*\right)^{\mathrm{T}}\left[\mathbf{W}^{\mathrm{T}}\mathbf{W} - \mathbf{W}^{\mathrm{T}}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}\right]\left(\overline{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u^*\right)\end{aligned} \tag{7-13}$$

where n_s is the number of selected parameters. It can be seen that the mean squared error consists of two terms. The first part is the variance term which results from the noise in the data and it is proportional to the number of free parameters used and the variance of the noise. The second one is the bias term which stems from the difference between the fixed values of the unselected parameters and their true values.

The parameters selected for estimation can be determined by minimizing the mean squared error. To illustrate the relationship between the results and the choice of the selection vector **z**, let

$$\mathbf{T} = \mathbf{S}\left(\mathbf{z}\right) \text{ and } \mathbf{W} = \mathbf{S}\left(\mathbf{1} - \mathbf{z}\right) \tag{7-14}$$

where **1** is a vector where each element has a value of unity. Then the optimization problem for parameter selection is given by

$$\min_{\mathbf{z}} \sigma^2 \sum_{i=1}^{n_\theta} z_i + \delta\left(\mathbf{z}, \boldsymbol{\theta}^*\right) \tag{7-15}$$

where the number of selected parameters is equal to the sum of elements in **z** and $\delta\left(\mathbf{z}, \boldsymbol{\theta}^*\right)$ is the bias term in Eq. (7-13) after substitution of Eq. (7-14).

However, this problem can not generally be solved since the objective function is dependent on the true value of parameters which is never known prior to estimation. This dependence results from the evaluation of the sensitivity matrix as well as the difference between the fixed values of the unselected parameters and their true values. However, there are several options for dealing with this type of problem, each one involving an approximation. The simplest one replaces the true parameter values by their nominal values, however, this is not a very realistic assumption and can return questionable results.

A second option is to formulate the problem as a minimax problem. In this case, the largest value of the bias term for all feasible parameter values is calculated and the

parameters are selected such that they minimize the largest bias value. However, the minimax procedure is difficult to implement since the explicit expression of the sensitivity as a function of the parameter values can usually not be explicitly obtained for realistic examples and numerical evaluation of the expression is computationally expensive. Furthermore, optimizing for the worst possible parameter values may lead to a conservative result since this case may rarely occur.

A third option is to uses available information about the parameters distribution to characterize the parameter uncertainty. The mean value of the bias term over the range of the uncertain parameters is calculated and then the parameters are selected by minimizing the mean value. In this case, the problem is given by

$$\min_{\mathbf{z}} \sigma^2 \sum_{i=1}^{n_\theta} z_i + \mathrm{E}_{\boldsymbol{\theta}^*}\left[\delta\left(\mathbf{z},\boldsymbol{\theta}^*\right)\right]. \tag{7-16}$$

This option is more mathematically tractable than the minimax procedure and is adopted in this paper. The mean value can be calculated by a Monte-Carlo method: a set of points of the parameter values are sampled according to the parameter distribution and the mean value is calculated from the sampled values.

While it has been discussed how the issue of parameter uncertainty can be handled for formulation of the optimization problem, the remaining combinatorial problem is still not trivial to solve, especially for biochemical reaction networks with a large number of parameters. Even though stochastic methods like genetic algorithm can be applied, it can generally not be guaranteed that the found solution is indeed optimal or reproducible. Another approach is to use a forward selection procedure, which is a sequential method that selects one parameter at each step. Since the variance term in the mean squared error shown in Eq. (7-16) is independent of the specific parameters selected, the objective function for the forward selection can be given by the bias:

$$\min_{\mathbf{z}} \mathrm{E}_{\boldsymbol{\theta}^*}\left[\delta\left(\mathbf{z},\boldsymbol{\theta}^*\right)\right]. \tag{7-17}$$

The forward selection procedure minimizes the objective function (7-17) by decomposing the multi-dimensional search into a sequential uni-dimensional search. All elements of $\mathbf{z}$ are initially set to zero. In a first step, the objective function is evaluated

for all situations where only one element of **z** is equal to one. The smallest value of the objective function is found and the corresponding element in **z** is set to one. In a second step, the objective function is evaluated for all situations where only two elements in **z** are equal to one, but where the element determined in the first step has to be one of these two. The elements in **z** which have the smallest value of the objective function for all possible cases are set to one. This procedure is repeated where in each step one additional element of **z** is given a value of one.

An advantage of the forward selection procedure is that a stopping rule naturally arises: the variance term in the mean squared error increases with the number of parameters selected while the bias term decreases. Initially, the bias term is dominant and selection of more parameters reduces the mean squared error. However, after a certain number of parameters have been selected, the variance term dominates and selecting more parameters will increase the mean squared error. Since the increase in the variance term is proportional to the number of selected parameters by a factor of σ^2, the forward selection procedure should be stopped when the reduction in the bias term by adding a new parameter is less than σ^2.

Relationship to the orthogonalization method

One important aspect of any theoretical contribution is to put the work into the proper context of existing approaches. A comparison is made here between the presented method and the orthogonalization method since both methods belongs to a class of forward selection procedures. This comparison will be made based upon two properties: comparison of the objective functions of the two methods and comparison of the criterion for selecting the next parameter at each step.

Some assumptions need to be made for carrying out this comparison. Since the orthogonalization method is a local technique and does not take dependence of the sensitivity matrix on the parameter values into account, the sensitivity matrix is assumed to be a constant matrix in the following comparison. For notational purposes, assume the decomposition of the sensitivity matrix is $\mathbf{S} = \begin{bmatrix} \mathbf{T} & \mathbf{W} \end{bmatrix}$ corresponding to the selected

parameters and unselected parameters. In the presented method the parameter uncertainty is assumed to be

$$\mathrm{E}\left(\boldsymbol{\theta}^*\right)=\overline{\boldsymbol{\theta}} \text{ and } \mathrm{Var}\left(\boldsymbol{\theta}^*\right)=\mathbf{I}. \tag{7-18}$$

The objective function of the orthogonalization method is the determinant of the Gram matrix of **T**. The determinant of the Gram matrix of **S** is

$$\det\left(\mathbf{S}^{\mathrm{T}}\mathbf{S}\right)=\det\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)\det\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}-\mathbf{W}^{\mathrm{T}}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}\right). \tag{7-19}$$

The matrix $\mathbf{W}^{\mathrm{T}}\mathbf{W}-\mathbf{W}^{\mathrm{T}}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}$ appears in the bias term in Eq. (7-13) and represents the Schur complement of $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ in $\mathbf{S}^{\mathrm{T}}\mathbf{S}$. Since the determinant of $\mathbf{S}^{\mathrm{T}}\mathbf{S}$ is a constant which is not dependent on the selected parameters, maximizing $\det\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)$ is equivalent to minimizing $\det\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}-\mathbf{W}^{\mathrm{T}}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}\right)$. Under the assumption mentioned above, the objective function from Eq. (7-17) of the presented method is given by

$$\begin{aligned}\mathrm{E}_{\boldsymbol{\theta}^*}\left[\delta\left(\mathbf{z},\boldsymbol{\theta}^*\right)\right]&=\mathrm{tr}\left[\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}-\mathbf{W}^{\mathrm{T}}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}\right)\mathrm{Var}\left(\boldsymbol{\theta}_u^*\right)\right]\\&=\mathrm{tr}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W}-\mathbf{W}^{\mathrm{T}}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}\right).\end{aligned} \tag{7-20}$$

It is apparent that both methods aim to minimize some measure of the matrix $\mathbf{W}^{\mathrm{T}}\mathbf{W}-\mathbf{W}^{\mathrm{T}}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}$. However, each method uses different measures as one makes use of the product of the eigenvalues of the matrix and the other of the sum of the eigenvalues, as represented by the determinant and the trace of the matrix, respectively.

It is should be noted that minimizing the trace of $\mathbf{W}^{\mathrm{T}}\mathbf{W}-\mathbf{W}^{\mathrm{T}}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}$ is not equivalent to maximizing the trace of $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ as there is no relationship for the trace that is equivalent to the one for the determinant used in equation in Eq. (7-19).

The criterion for the orthogonalization method to add a new parameter at each step is the squared norm of the projected sensitivity vector $\mathbf{s}_i^{(n_s)}$. The projected sensitivity

vector $\mathbf{s}_i^{(n_s)}$ is in fact the residual vector $\mathbf{r}_i$ of the sensitivity vector $\mathbf{s}_i$ regressed on the sensitivity vectors of the selected parameters:

$$\mathbf{r}_i = \mathbf{w}_i - \mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{w}_i \tag{7-21}$$

where $\mathbf{w}_i$ is i-th column in $\mathbf{W}$. The residual characterizes the effect of an unselected parameter on the output which is not covered by the effects of the selected parameters. The forward selection selects the parameters which has the largest residual as given by the squared norm of the residual vector. However, the presented method sequentially minimizes the trace as seen by Eq. (7-20) and it is equal to the sum of squared residuals

$$\mathrm{tr}\left(\mathbf{W}^{\mathrm{T}}\mathbf{W} - \mathbf{W}^{\mathrm{T}}\mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{W}\right) = \sum_i \mathbf{r}_i^{\mathrm{T}}\mathbf{r}_i \,. \tag{7-22}$$

Thus the criterion for the presented method to select a parameter at each step is that the selection leads to the smallest sum of the residuals of the unselected parameters. In other words, the orthogonalization method selects the parameter which has the largest effect not covered by the selected parameters at each step while the presented method selects the parameter which results in the smallest value of the total effects produced by the remaining parameters.

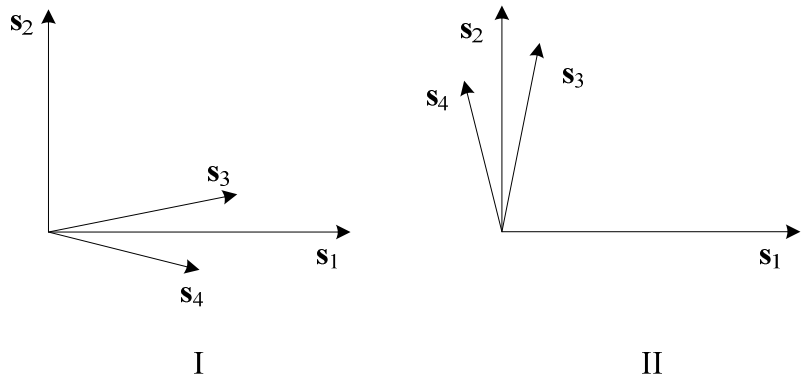


Fig. 7-1. Illustration of the different selections made by the two forward selection methods.

An illustration of the different results returned by the two forward selection procedures is given in Fig. 7-1 and the following explanations. Assume that there are the four sensitivity vectors $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4\}$. In both cases $\mathbf{s}_1$ and $\mathbf{s}_2$ are fixed at the same position but $\mathbf{s}_3$ and $\mathbf{s}_4$ are different for Case I and Case II. The orthogonalization method selects $\mathbf{s}_1$ first in both cases while the presented method select $\mathbf{s}_1$ first in case I but selects $\mathbf{s}_2$ first in case II as the effects of $\mathbf{s}_3$ and $\mathbf{s}_4$ are similar to the effect of $\mathbf{s}_2$. In fact if the length of the four sensitivity vectors is fixed but the directions may change, then the orthogonalization method will select $\mathbf{s}_1$ first in all situations while the presented method will make the first selection according to the lengths of the sensitivity vectors as well as the directions. The later selection can produce a more accurate prediction. Additionally, it is a disadvantage of the orthogonalization technique that there is no clear rule that indicates when to stop adding parameters to the set to be estimated. However, it should be noted that Eq. (7-20) cannot be used if the sensitivity matrix depends on the parameter values, while the forward selection technique that maximizes the criterion shown in Eq. (7-17) is still applicable. Not being able to take parameter uncertainty into account is another disadvantage of the orthogonalization method compared with the presented method.

7.4 Detailed case study

Model description

NF-κB (nuclear factor kappa B) is a transcription factor that plays a key role in regulating numerous genes involved in pathogen or cytokine inflammation, immune response, cell proliferation and survival. A model of the NF-κB signaling pathway (Lipniacki et al., 2004) is used in this work to illustrate the presented technique. This model consists of 15 nonlinear ordinary differential equations and includes 26 parameters with nonzero nominal value. The state variables are the concentrations of the proteins in the pathway and the concentrations of free nuclear NF-κB is selected as the

output. The nominal value of the parameters, the initial value of the states, and the Matlab codes for the model are available in the reference of Lipniacki et al. (2004).

Simulation parameters used for the case study

The algorithms used for model simulation, sensitivity calculation, parameter selection and estimation are all implemented in Matlab. The simulations cover a time span from 0 to 8 hrs. The output is sampled every 15 minutes over this time interval and measurement noise is assumed to have zero mean and a covariance matrix of $0.05^2\mathbf{I}$ ($\sigma^2 = 0.0025$). The true parameter values are assumed to be uniformly distributed in the range from 25% to 175% of the nominal values of the parameters. Though the value of the kinetic parameters may sometimes change by orders of magnitude, it is not completely unrealistic to have the estimated parameter value within a bound closer to the nominal values (Wilkinson et al., 2008).

Parameter selection

The sensitivity of the output with respect to the parameters is calculated by the direct differential method (Rabitz et al., 1983; Turanyi, 1990) which solves the sensitivity equations simultaneously with the system equations. The sensitivity matrix is formed by sampling the sensitivity value at the given time points. To calculate the mean value of the criterion from Eq. (7-17), 10^4 parameter values are randomly sampled over their uncertainty range. For each parameter value the sensitivity matrix is evaluated and recorded. Table 7-1 shows the results of the forward selection procedure where the parameters are sequentially selected. The value of the objective function (7-17) at each step is listed in the table as well for the selected parameter sets. It can be seen that when the number of selected parameters increases, that the value of the bias decreases. Initially adding a new parameter will generate a decrease in the value of the bias which is larger than the variance, however, after selection of the 7^{th} parameter the decrease resulting from selecting the 8^{th} parameter is $(0.0011 - 0.0002) = 0.0009$, which is less than the variance of the noise ($\sigma^2 = 0.0025$). This is a good indicator to stop the procedure after

the first 7 parameters $\{i_{1a}, c_{1a}, a_3, k_v, c_5, c_4, c_{3a}\}$ have been selected for estimation. The orthogonalization method has also been applied to this problem for comparison purposes. For this method parameters are selected based on the sensitivity matrix evaluated at the nominal value of the parameters. Since there is no clear criterion for how many parameters should be selected, the procedure was also stopped after the 7th parameter has been selected. The following parameter set was obtained for estimation: $\{k_v, c_{4a}, a_3, c_5, i_1, i_{1a}, a_1\}$. It should be noted that three out of the selected seven parameters are different for the results returned by the two techniques. The effect that the choice of these different parameter sets has on the prediction accuracy of the output will be investigated in the following subsections.

Table 7-1

Forward selection procedure to minimize the mean squared error

step	1	2	3	4	5	6	7	8	9
parameter	i_{1a}	c_{1a}	a_3	k_v	c_5	c_4	c_{3a}	i_1	k_{prod}
bias	0.3018	0.1132	0.0385	0.0189	0.0076	0.0039	0.0011	0.0002	0.0001

Parameter estimation and evaluation of prediction accuracy

Several data sets are created for different true values of the parameters of the system in order to provide a realistic representation of the model uncertainty. 10 parameter values are randomly sampled where the parameters are assumed to have a uniform distribution over their uncertainty range. Furthermore, 50 data sets are generated for each set of parameter values by adding noise to the data. Three different sets of parameters are estimated from the created data sets and comparisons are made: the presented method that selects parameters to minimize the mean squared error of prediction (referred to as the 'presented method'), the method that selects parameters by the orthogonalization procedure (referred to as the 'orthogonalization method') and one method that estimates all of the parameters.

Table 7-2 shows the averaged fitting errors over 50 data sets for each of the 10 parameter values returned by the three methods. The fitting error is the sum of squared errors between the predicted output and the measured data divided by the variance of the noise:

$$e\left(\hat{\boldsymbol{\theta}}\right)=\frac{1}{\sigma^2}\left(\mathbf{y}\left(\hat{\boldsymbol{\theta}}\right)-\tilde{\mathbf{y}}\right)^{\mathrm{T}}\left(\mathbf{y}\left(\hat{\boldsymbol{\theta}}\right)-\tilde{\mathbf{y}}\right) \tag{7-23}$$

where $\hat{\boldsymbol{\theta}}$ is the nonlinear least squares estimate of parameters. It can be seen that estimation of all parameters returns a smaller fitting error in all cases than can be achieved by either of the two parameter set selection techniques. However, even though less than one quarter of the parameters are fitted by the techniques that use parameter set selection, the fitting error is close to the one returned by estimation of all parameters. It can be concluded for this biochemical reaction network that the effects that the parameters have on the output are highly correlated and that not all parameters can be reliably estimated in practice.

Table 7-2

Comparison of average fitting error for the different techniques applied to ten different data sets

parameter value		1	2	3	4	5	6	7	8	9	10
averaged fitting error	presented method	21.61	24.50	19.83	13.26	20.12	22.09	24.05	17.14	25.47	13.72
	estimation of all parameters	19.26	22.30	17.71	11.20	18.83	19.87	21.08	15.51	23.36	11.84
	orthogonalization method	21.53	24.49	19.01	15.80	19.96	21.72	25.01	17.08	26.82	13.61

When the noise level is significant a small fitting error does not necessarily imply a small prediction error, e.g., as given by the sum of the squared errors between the predicted output and the true value of the output:

$$d\left(\hat{\boldsymbol{\theta}}\right)=\frac{1}{\sigma^2}\left(\mathbf{y}\left(\hat{\boldsymbol{\theta}}\right)-\mathbf{y}^*\right)^{\mathrm{T}}\left(\mathbf{y}\left(\hat{\boldsymbol{\theta}}\right)-\mathbf{y}^*\right). \tag{7-24}$$

The averaged prediction errors for the 50 data sets generated for each of the 10 different cases are shown in Table 7-3 for the three procedures.

Table 7-3

Comparison of mean squared error for the different techniques applied to ten different data sets

parameter value		1	2	3	4	5	6	7	8	9	10
averaged fitting error	presented method	3.915	3.592	4.033	6.371	3.580	3.671	4.048	4.337	3.850	4.529
	estimation of all parameters	6.252	6.478	5.816	8.762	5.987	6.199	5.807	6.569	5.811	6.747
	orthogonalization method	4.618	3.967	4.558	11.291	3.988	4.326	5.997	5.150	6.208	5.130

It can be seen for all the 10 different parameter values that the averaged prediction error returned by the presented method is less than that returned by estimation of all parameters as well as that returned by the orthogonalization method. As the data sets used in this example include 50 entries, a statistical test is required to confirm if the difference in the averaged values returned by the three methods in Table 7-3 is significant. A hypothesis test is formulated as follows:

$$H_0:\bar{m}_A-\bar{m}_B=\bar{m} \text{ against } H_1:\bar{m}_A-\bar{m}_B>\bar{m} \tag{7-25}$$

where $\bar{m}_A$, $\bar{m}_B$ and $\bar{m}$ denote the mean value returned by method A, method B and the gap, respectively. The subscript A denotes the orthogonal method or estimation of all parameters while the subscript B denotes the presented method in the comparison. The test checks if the null hypothesis can be rejected for a given significant level α and a difference $\bar{m}$. If this is the case then the difference between the averaged values is statistically significant. To further explain this procedure, $\bar{m}$ is first set to zero and the

P-value of the test is calculated, i.e., the smallest α such that the null hypothesis can be rejected. Next α is set to 0.05 and $\bar{m}_c$, the largest $\bar{m}$, is calculated such that the null hypothesis can be rejected. From the value of $\bar{m}_c$ it can be concluded with 95% confidence to what degree a method compares favorably to another technique. $\bar{m}_c$ serves as a good indicator for describing the gap between the performance of two methods rather than the differences in the averaged values.

Table 7-4

Statistical significance test of the performance of the presented method vs. estimation of all parameters

parameter value	1	2	3	4	5	6	7	8	9	10
P-value	3.85E-9	5.68E-8	2.33E-9	4.67E-8	4.60E-8	2.05E-9	3.18E-9	8.38E-7	3.92E-5	2.25E-7
value of $\bar{m}_c$	1.77	2.1	1.36	1.81	1.76	1.93	1.33	1.54	1.57	1.59
$\bar{m}_c$ %	0.452	0.585	0.337	0.284	0.492	0.526	0.329	0.355	0.408	0.351

The *P*-value and the value of $\bar{m}_c$ are listed in Table 7-4 for a comparison of the presented method and estimation of all parameters. All *P*-values are close to zero, i.e., it can be concluded with a probability close to one that the presented method returns better results than estimation of all parameters and that the differences in the mean squared errors is not the result of randomness in the data set. All values of $\bar{m}_c$ in the table are positive which also confirms this interpretation. The table also lists the relative significant gap $\bar{m}_c$% which is the value of $\bar{m}_c$ divided by the averaged squared error by the presented method in Table 7-3. The conclusion from the table is that the presented method returns a smaller prediction error than estimation of all parameters by at least 28.4%, by at most 58.5% and on average by 41.2%.

Table 7-5

Statistical significance test of the performance of the presented method vs. the orthogonalization method

parameter value	1	2	3	4	5	6	7	8	9	10
P-value	9.50E-3	3.43E-4	8.12E-4	7.38E-7	1.39E-3	1.70E-5	6.4E-15	1.57E-8	5.3E-10	2.54E-8
value of $\bar{m}_c$	0.21	0.2	0.26	3.41	0.19	0.41	1.72	0.64	2.05	0.44
$\bar{m}_c$ %	0.054	0.056	0.064	0.535	0.053	0.112	0.425	0.148	0.532	0.097

Similarly, the *P*-value and the value of $\bar{m}_c$ are listed in Table 7-5 for a comparison of the presented method and the orthogonalization method. From the values it can be concluded that the presented method outperforms the orthogonalization method. However, the gap between the two techniques is smaller than what is shown in Table 7-4, indicating that the orthogonalization method is likely to return a better result than estimation of all parameters.

7.5 Conclusions

An observed universal property of complex biochemical networks is the "sloppiness" of parameter sensitivities: The eigenvalues of the Gram matrix of the sensitivity vectors can vary by many orders of magnitude (Gutenkunst et al., 2007). This indicates that parameters in a model are strongly correlated and effects on the output produced by changes of a parameter can be compensated by changes of another one. These "sloppy" parameter sensitivities result in an unidentifiable model, i.e., many combinations of parameter values can result in similar model predictions. This observation is also confirmed by other results involving parameter estimation (Piazza et al., 2008; Dimelow and Wilkinson., 2009). This "sloppiness" property indicates that focusing on exact parameter values may not be the best route for deriving models and instead the focus should be shifted to the accuracy of model predictions (Gutenkunst et al., 2007; Wilkinson et al., 2008; Hlavacek, 2009).

Based upon these observations, a method that improves accuracy of model predictions is developed in this work. The technique selects a subset of the parameters for parameter estimation and fixes the remaining parameters at their nominal values. Correlations among the effects that changes of the parameters have on the outputs are considered when selecting a subset of parameters. It should be emphasized that selecting a subset of parameters for estimation not only simplifies the estimation procedure itself, but can also lead to better prediction accuracy as over-fitting is avoided.

The technique presented in this work is different from other parameter set selection methods because it aims at improving the accuracy of the model prediction rather than to improve parameter identifiability. The method belongs to a class of forward selection techniques which sequentially minimizes the mean squared error of the prediction. The effect of parameter uncertainty on the objective function is explicitly taken into account in this parameter selection technique. In addition to this, a criterion that determines how many parameters should be selected for estimation is derived for this procedure based upon a tradeoff between the variance and the bias in the mean squared error.

The presented method for parameter selection is evaluated in a detailed case study involving a model of the NF-κB signal transduction pathway. The results returned by the presented method are compared with results by estimation of all parameters as well as those by the orthogonalization method. Even though only a subset of parameters is estimated by this technique, the fitting errors are almost as small as the ones for estimating all parameters. This supports the conclusion that many combinations of parameter values are able to generate equally good fitting (Gutenkunst et al., 2007; Piazza et al., 2008; Wilkinson et al., 2008; Dimelow and Wilkinson, 2009; Hlavacek, 2009). However, it has also been shown that the presented technique can result in better prediction accuracy than estimation of all parameters as well as better prediction accuracy than those resulting from the orthogonalization method.

8. QUANTITATIVE OPTIMAL EXPERIMENTAL DESIGN USING GLOBAL SENSITIVITY ANALYSIS VIA QUASI LINEARIZATION

8.1 Introduction

Experimental design has received a significant amount of attention in statistics and system identification (Ljung, 1999; Steinberg and Hunter, 1984; Walter and Pronzato, 1990). Qualitative design is one aspect of experimental design and consists of selecting input/output variables and identifiable parameters. Quantitative design, on the other hand, deals with determining input shapes and sampling schedules based on optimization of a suitable criterion. Local parametric sensitivities, i.e., partial derivatives of the output with respect to parameters, play an important role in both qualitative and quantitative experimental design. Various criteria for experimental design have been developed based on local sensitivity analysis.

Global sensitivity analysis has more recently received a lot of attention as an alternative to local sensitivity analysis. Global sensitivity analysis characterizes the effect of a parameter on the output while explicitly taking information about parameter uncertainty into account. A significant amount of work has been done using global sensitivity analysis instead of local sensitivity analysis for experimental design (Brockmann et al., 2008; Chhatre et al., 2008; Cho et al., 2003; Degerman et al., 2009; King et al., 2007; Kontoravdi et al., 2005; Martinez et al., 2009; Sidoli et al., 2005; Varella et al., 2010; Yue et al., 2008; Zi et al., 2005). However, these efforts mainly focused on qualitative experimental design, i.e., determining important parameters that should be estimated. While it has been recognized that global sensitivity analysis outperforms local sensitivity for determining important parameters, reports of quantitative optimal experimental design by global sensitivity analysis, e.g. selection of sampling points and determination of input profiles by optimizing a experimental criterion, are rare (for an exception, see Martinez et al, 2009).

The main obstacle to using global sensitivity analysis techniques for quantitative optimal experimental design is that there is a lack of optimality criteria that can be

applied to the global sensitivity values. The common experimental criteria are derived for linear systems where the results returned by each criterion characterize a specific attribute related to the precision of the estimated parameter values. These criteria are real functions of the design matrix of the linear model. In the case of a nonlinear model the local sensitivity matrix can be used as the design matrix; while it is possible to use the experimental criteria on the local sensitivity matrix, it has to be taken into account that the nonlinear behavior of the model is not taken into account in this case. However, it should be pointed out that the use of experimental criteria can not be easily extended to the global sensitivity matrix. If criteria developed for the local sensitivity matrix are applied to the global sensitivity matrix, such as was done by Martinez et al., 2009, then the experimental designs may be inconsistent with the traditional designs as the global sensitivity measures are not guaranteed to reduce to the local sensitivity when the parameter uncertainty is negligible, or even when the model is linear. A consequence of this is that it is difficult to interpret the results returned by an experimental criterion applied to the global sensitivity matrix.

It is the goal of this section to develop a new global sensitivity analysis measure to be used for optimal experimental design. This global sensitivity analysis is performed via quasi linearization and the computed global sensitivity matrix is shown to be an extension of the design matrix of the linearized model. Due to this property, the existing experimental criteria can be applied to the global sensitivity matrix. The technique is consistent with traditional experimental design as results from the global sensitivity analysis reduce to the ones derived using local sensitivity analysis if the model is linear or if the parameter uncertainty is approaching zero. However, the presented approach is a global technique as the parameter uncertainty is explicitly taken into account during the computation. Due to this, quantitative experimental design based on the global sensitivity analysis can be performed, which may result in an improvement compared to a design based upon local sensitivity analysis. The technique is illustrated in three case studies, one where the parameter identifiability is tested, one where the optimal sampling points are determined, and one where the optimal input profile is computed.

8.2 Motivation behind derivation of a new global sensitivity analysis technique

The main drawback of local sensitivity analysis applied to nonlinear systems is that the sensitivity values are affected by the parameter values. To overcome this drawback a wide variety of global sensitivity methods have been developed. It is generally accepted that global sensitivity analysis is superior to local sensitivity analysis for identification of influential parameters as is also evidenced by a large number of applications of global sensitivity analysis. However, the use of global sensitivity analysis for designing inputs and outputs has been much more limited and no generally acceptable criterion for quantitative experimental design involving global sensitivity analysis has been proposed in the literature. The reason for this is that quantitative experimental design generally uses a criterion of the sensitivity matrix for determining experimental conditions; however, it is unclear how existing experimental design criteria can be applied to the sensitivity values computed from global sensitivity analysis techniques.

Even though it is straightforward to construct the global sensitivity matrix, similar to the local sensitivity matrix, and it has been suggested to apply existing experimental criteria to the global sensitivity matrix, the results of such a design can be problematic. The reason for this statement is that such a design involving the global sensitivity matrix is inconsistent with the traditional designs, e.g., if the global sensitivity matrix fails to reduce to the design matrix when the model is linear. One resulting problem is that it that interpretation of the results returned by such a method are unclear. A simple example is presented here to illustrate this point. Consider the following two linear regression models

$$\text{Model I: } \begin{matrix} g_1(\theta_1,\theta_2)=\theta_1+\theta_2 \\ g_2(\theta_1,\theta_2)=\theta_1+\theta_2 \end{matrix}, \qquad \text{Model II: } \begin{matrix} g_1(\theta_1,\theta_2)=\theta_1+\theta_2 \\ g_2(\theta_1,\theta_2)=\theta_1-\theta_2 \end{matrix}. \tag{8-1}$$

These models do not contain noise terms as it is the purpose of this illustrative example to assess structural identifiability. The local sensitivity matrix is the design matrix as given by

$$\mathbf{S}_{\text{I}}=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} \text{ and } \mathbf{S}_{\text{II}}=\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}. \tag{8-2}$$

The identifiability of the parameters can be determined directly from the value of the experimental criterion. The *D*-optimality criterion value is zero for Model I since the sensitivity matrix is rank deficient while the criterion value is nonzero for Model II due to the full rank of the sensitivity matrix. As a result, the parameters in Model I are unidentifiable while the ones for Model II are identifiable. However, the *D*-optimality criterion values of the global sensitivity matrix calculated from the conditional variance are both zero since the global sensitivity matrices are identical and equal to $\mathbf{S}_I$, which would falsely suggest that both models are unidentifiable. The reason for the incorrect results returned by this method based upon global sensitivity is that the sign information is lost in the computation of the conditional variance and that this global sensitivity matrix does not reduce to the design matrix.

It should be pointed out that the presented example just used one global sensitivity analysis method to illustrate a point. Since a wide variety of methods for global sensitivity analysis exist, it is beyond the scope of this section to compare all of them. Instead the focus is on variance-based methods as they have been frequently applied in qualitative experimental design to indentify influential parameters. Other global sensitivity indices, e.g., the Kolmogorov-Smirnov statistic also fail to reduce to the local sensitivity due to several reasons. In contrast to these approaches, the technique introduced in this section can be used for global sensitivity analysis, but also reduces to existing techniques for small uncertainty in the parameters. It can be shown that this method includes the sensitivity defined by the Pearson correlation coefficient as a special case.

8.3 Optimal experimental design using global sensitivity analysis via quasi linearization

Development of a new global sensitivity measure for optimal experimental design

The development of a new global sensitivity analysis technique that can be used for quantitative optimal experimental design, instead of existing local sensitivity analysis

methods, is the main contribution of this section. This technique has the advantage that parametric uncertainty can be explicitly taken into account by applying existing experimental design criteria to the global sensitivity matrix developed in this section. This extension of local methods to global sensitivity analysis is achieved via quasi linearization.

While the exact values of parameters are never known before estimation is performed, it is common that some prior information about the parameter uncertainty is available. The region of possible parameter values is often characterized by a hyper-rectangle and each parameter is distributed within an interval. A reasonable choice of the nominal value of a parameter is the mean parameter value $\boldsymbol{\theta}_{\text{nominal}} = \mathrm{E}[\boldsymbol{\theta}]$. To simplify the notation, deviation variables are introduced by subtracting the nominal value from the original one, i.e., $\boldsymbol{\theta} = \boldsymbol{\theta}_{\text{original}} - \boldsymbol{\theta}_{\text{nominal}}$, and then the nominal value of the deviation variable is

$$\bar{\boldsymbol{\theta}} = \mathbf{0} \ . \tag{8-3}$$

The interval of an uncertain parameter is assumed as

$$\boldsymbol{\theta} \in \left[\boldsymbol{\theta}^{\mathrm{L}}, \boldsymbol{\theta}^{\mathrm{U}} \right] \tag{8-4}$$

where $\boldsymbol{\theta}^{\mathrm{L}}$ is the lower bound and $\boldsymbol{\theta}^{\mathrm{U}}$ is the upper bound. Similarly, the output can be transformed such that

$$\mathbf{g}\left(\bar{\boldsymbol{\theta}}\right) = \mathbf{0} \ . \tag{8-5}$$

It should be noted that introducing deviation variables only represents a change of notation and has no effect on the parameter sensitivity analysis itself and/or the experimental design.

Since the goal is to perform experimental design for nonlinear systems, a linear approximation of the original model can be useful. Using the notation introduced in Eq. (8-3)-(8-5) a regression model can be written in deviation variables as

$$\mathbf{g}\left(\boldsymbol{\theta}\right) \approx \sum_{i} \mathbf{s}_i \theta_i \ . \tag{8-6}$$

This linear approximation also provides a straightforward technique for evaluating

sensitivity: According to this expression, the coefficient vector $\mathbf{s}_i$ is the sensitivity vector of the parameter θ_i.

The most common approximation is the local linearization resulting in the local sensitivity value. However, several alternatives to the described local linearization exist, one of which will be used in this section. One alternative is to regard the regression model as a nonlinear system mapping of the inputs of $\boldsymbol{\theta}$ to the outputs $\mathbf{g}(\boldsymbol{\theta})$. To study the behavior of the system and investigate the effect of the inputs, the system is stimulated by an input

$$\theta_i = \alpha_i v_i(\psi_i) \tag{8-7}$$

where $\alpha_i = \theta_i^{\mathrm{U}} - \theta_i^{\mathrm{L}}$ is the magnitude of the uncertainty and the input function is chosen such that

$$\psi_i \in [0,1] \text{ and } v_i(\psi_i) \in \left[\frac{\theta_i^{\mathrm{L}}}{\theta_i^{\mathrm{U}} - \theta_i^{\mathrm{L}}}, \frac{\theta_i^{\mathrm{U}}}{\theta_i^{\mathrm{U}} - \theta_i^{\mathrm{L}}} \right]. \tag{8-8}$$

The best linear approximation to the nonlinear model for this specific input can be calculated by minimizing the squared errors of the approximation

$$\begin{aligned}
&\min_{\mathbf{s}_0,\mathbf{s}_1,\cdots,\mathbf{s}_{n_\theta}} \int_0^1 \cdots \int_0^1 \left\| \mathbf{g}\left(\alpha_1 v_1(\psi_1),\cdots,\alpha_{n_\theta} v_{n_\theta}(\psi_{n_\theta})\right) - \sum_k \mathbf{s}_k \alpha_k v_k(\psi_k) \right\|^2 \prod_k d\psi_k \\
&= \min_{\mathbf{s}_0,\mathbf{s}_1,\cdots,\mathbf{s}_{n_\theta}} \sum_j \int_0^1 \cdots \int_0^1 \left(g\left(t_j, \alpha_1 v_1(\psi_1),\cdots,\alpha_{n_\theta} v_{n_\theta}(\psi_{n_\theta})\right) - \sum_k s_k(t_j) \alpha_k v_k(\psi_k) \right)^2 \prod_k d\psi_k \\
&= \sum_j \min_{s_0(t_j),s_1(t_j),\cdots,s_{n_\theta}(t_j)} \int_0^1 \cdots \int_0^1 \left(g\left(t_j, \alpha_1 v_1(\psi_1),\cdots,\alpha_{n_\theta} v_{n_\theta}(\psi_{n_\theta})\right) - \sum_k s_k(t_j) \alpha_k v_k(\psi_k) \right)^2 \prod_k d\psi_k .
\end{aligned} \tag{8-9}$$

where $\mathbf{s}_k = \left[s_k(t_1),\cdots,s_k(t_{n_t}) \right]^{\mathrm{T}}$.

The last line in Eq. (8-9) exemplifies that the optimization can be performed separately for different time points t_j. To simplify the notation, the index of t_j is omitted

$$\min_{s_0,s_1,\cdots,s_{n\theta}} J = \int_0^1 \cdots \int_0^1 \left(g\left(\alpha_1 v_1(\psi_1),\cdots,\alpha_{n_\theta} v_{n_\theta}(\psi_{n_\theta})\right) - \sum_k s_k \alpha_k v_k(\psi_k) \right)^2 \prod_k d\psi_k \ . \tag{8-10}$$

This expression is a least squares optimization and the solution can be calculated from

$$\frac{\partial J}{\partial s_i} = 0,\ i = 1 \cdots n_\theta \tag{8-11}$$

which directly leads to

$$\int_0^1 \cdots \int_0^1 g v_i \prod_k d\psi_k - \sum_j \left(\alpha_j \int_0^1 \cdots \int_0^1 v_i v_j \prod_k d\psi_k \right) s_j = 0,\ i = 1 \cdots n_\theta \tag{8-12}$$

where the solution using matrix notation is given by

$$\begin{bmatrix} s_1 \\ \vdots \\ s_{n_\theta} \end{bmatrix} = \begin{bmatrix} \alpha_1 \int_0^1 \cdots \int_0^1 v_1 v_1 \prod_k d\psi_k & \cdots & \alpha_{n_\theta} \int_0^1 \cdots \int_0^1 v_1 v_{n_\theta} \prod_k d\psi_k \\ \vdots & \ddots & \vdots \\ \alpha_1 \int_0^1 \cdots \int_0^1 v_{n_\theta} v_1 \prod_k d\psi_k & \cdots & \alpha_{n_\theta} \int_0^1 \cdots \int_0^1 v_{n_\theta} v_{n_\theta} \prod_k d\psi_k \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 \cdots \int_0^1 g v_1 \prod_k d\psi_k \\ \vdots \\ \int_0^1 \cdots \int_0^1 g v_{n_\theta} \prod_k d\psi_k \end{bmatrix}. \tag{8-13}$$

A multi-dimensional integral needs to be evaluated for each element of the matrix and the vector on the right side of Eq. (8-13). To limit the computational effort, it is assumed that the input functions are orthogonal:

$$\int_0^1 \cdots \int_0^1 v_i(\psi_i) v_j(\psi_j) \prod_k d\psi_k = 0 \quad \text{for any } i \neq j. \tag{8-14}$$

Then the sensitivity value can be computed from

$$s_i = \frac{\int_0^1 \cdots \int_0^1 g\left(\alpha_1 v_1(\psi_1), \cdots, \alpha_{n_\theta} v_{n_\theta}(\psi_{n_\theta})\right) v_i(\psi_i) \prod_k d\psi_k}{\alpha_i \int_0^1 \cdots \int_0^1 v_i(\psi_i)^2 \prod_k d\psi_k}. \tag{8-15}$$

Another reason to choose orthogonal input functions is that the defined sensitivity value will reduce to the local sensitivity value when the range of parameter uncertainty tends to zero. For an illustration of this statement, suppose that the uncertainty range of each parameter decreases simultaneously with the same α; then the limit of Eq. (8-15) is

$$
\begin{aligned}
\lim_{\alpha\to 0} s_i &= \lim_{\alpha\to 0} \frac{\int_0^1 \cdots \int_0^1 g\left(\alpha v_1(\psi_1), \cdots, \alpha v_{n_\theta}\left(\psi_{n_\theta}\right)\right) v_i(\psi_i) \prod_k d\psi_k}{\alpha \int_0^1 \cdots \int_0^1 v_i(\psi_i)^2 \prod_k d\psi_k} \\
&= \frac{\int_0^1 \cdots \int_0^1 \left[\lim_{\alpha\to 0} \frac{1}{\alpha} g\left(\alpha v_1(\psi_1), \cdots, \alpha v_{n_\theta}\left(\psi_{n_\theta}\right)\right)\right] v_i \prod_k d\psi_k}{\int_0^1 \cdots \int_0^1 v_i^2 \prod_k d\psi_k} \\
&= \frac{1}{\int_0^1 \cdots \int_0^1 v_i^2 \prod_k d\psi_k} \int_0^1 \cdots \int_0^1 \left.\frac{\partial g\left(\alpha v_1(\psi_1), \cdots, \alpha v_{n_\theta}\left(\psi_{n_\theta}\right)\right)}{\partial \alpha}\right|_{\alpha=0} v_i \prod_k d\psi_k \qquad \text{(8-16)} \\
&= \frac{1}{\int_0^1 \cdots \int_0^1 v_i^2 \prod_k d\psi_k} \int_0^1 \cdots \int_0^1 \left(\left.\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_1}\right|_{\boldsymbol{\theta}=0} v_1 + \cdots + \left.\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_{n_\theta}}\right|_{\boldsymbol{\theta}=0} v_{n_\theta} \right) v_i \prod_k d\psi_k \\
&= \frac{1}{\int_0^1 \cdots \int_0^1 v_i^2 \prod_k d\psi_k} \int_0^1 \cdots \int_0^1 \left.\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}=0} v_i v_i \prod_k d\psi_k \\
&= \left.\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta}=0}.
\end{aligned}
$$

Selecting appropriate inputs for Eq. (8-7) is a critical step in this quasi-linearization procedure. The range condition given by Eq. (8-8) and the orthogonality condition from Eq. (8-14) should be satisfied. Additionally, the inputs should sufficiently stimulate the system to create a rich data set for the global sensitivity computed by Eq. (8-15).

Several candidates for input functions are commonly used in various types of nonlinear systems analysis: piecewise constant functions, ramp functions, and sinusoidal functions. In particular, sinusoidal functions are commonly used as frequency response characteristics can be determined and the sensitivity given by equation (8-15) is related to the Fourier coefficient of the output.

However, one is not restricted to these input types and can instead determine the input function according to the prior distribution of the parameter, if this distribution is known. Using this approach, the independent variable ψ_i is regarded as a random variable with a uniform distribution over the unit interval. According to the distribution

of the parameter θ_i, $F(\theta_i)$, the input function can be selected as

$$v_i(\psi_i) = \frac{1}{\alpha_i} F^{-1}(\psi_i) \ . \tag{8-17}$$

A multi-dimensional integral needs to be evaluated to compute the global sensitivity from Eq. (8-15). In most cases, an analytical solution does not exist and a standard Monte Carlo method can be applied instead: A set of values of $\boldsymbol{\psi}$ are sampled according to the uniform distribution and recorded as $\{\boldsymbol{\psi}^k\}$ where k is an index for the run of simulation. The parameter value is calculated using the input function and the output is evaluated at each parameter point to generate a set of $\{g^k\}$. In the special case where the input function is a linear function

$$\theta_i = \alpha_i \psi_i \tag{8-18}$$

the sensitivity results in

$$s_i = \frac{\int_0^1 \cdots \int_0^1 g(\alpha_1 \psi_1, \cdots, \alpha_{n_\theta} \psi_{n_\theta}) \psi_i \prod_i d\psi_i}{\alpha_i \int_0^1 \cdots \int_0^1 \psi_i^2 \prod_i d\psi_i} \ . \tag{8-19}$$

The calculation by the Monte Carlo method is then given by

$$s_i = \frac{\sum_k g^k \psi_i^k}{\alpha_i \sum_k \psi_i^k \psi_i^k} = \frac{\sum_k g^k \theta_i^k}{\sum_k \theta_i^k \theta_i^k} = \rho(g, \theta_i) \frac{\sigma_g}{\sigma_{\theta_i}} \tag{8-20}$$

where ψ_i^k and θ_i^k are the i-th elements of the sample vectors $\boldsymbol{\psi}^k$ and $\boldsymbol{\theta}^k$, respectively. For this special case, the sensitivity is the uncentered Pearson correlation coefficient $\rho(g, \theta_i)$ normalized by the ratio of the standard deviations of the output and the parameter.

A more efficient approach to evaluate the multi dimensional integral is the quasi Monte Carlo method. A set of rationally linear independent numbers $\{\omega_n\}$ is selected. The multi-dimensional integral is then transferred to a uni-dimensional integral

$$\begin{aligned}&\int_0^1 \cdots \int_0^1 g\left(\alpha_1 v_1(\psi_1), \cdots, \alpha_{n_\theta} v_{n_\theta}\left(\psi_{n_\theta}\right)\right) v_i(\psi_i) \prod_k d\psi_k \\ &= \int_0^T g\left(\alpha_1 v_1\left((\omega_1 \tau) \bmod 1\right), \cdots, \alpha_{n_\theta} v_{n_\theta}\left(\left(\omega_{n_\theta} \tau\right) \bmod 1\right)\right) v_i\left((\omega_i \tau) \bmod 1\right) d\tau\end{aligned} \tag{8-21}$$

where the upper bound of the integral, T, equals the least common multiple of $\{1/\omega_n\}$ since the function $v_n\left((\omega_n \tau) \bmod 1\right)$ is periodic with a period of $1/\omega_n$.

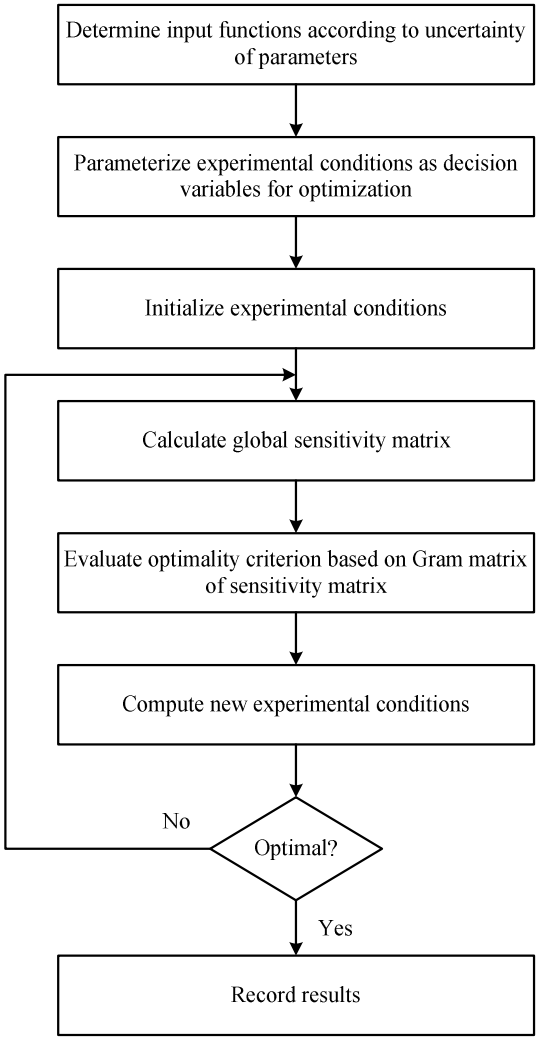


Fig. 8-1. Flowchart of experimental design based on global sensitivity analysis.

Optimal experimental design involving global sensitivity

The global sensitivity vector $\mathbf{s}_i$ is formed by computing the global sensitivity value at different sampling points in time. Since Eq. (8-6) is a linear approximation, the experimental design optimality criteria derived for linear models can also be used in this case. The only modification is that the sensitivity matrix consists of the global sensitivity values, as computed from Eq. (8-15), instead of the local sensitivity values. Since the global sensitivity is able to reduce to the local sensitivity, the design by global sensitivity analysis reduces to the one by local sensitivity analysis when the parameter uncertainty is small. At the same time, the effect of parameter uncertainty is taken into account in the presented procedure and, as a result, the technique can be applied to models with a significant degree of uncertainty.

The flowchart for the experimental design procedure based on global sensitivity analysis is shown in Fig. 8-1. The first step is to determine the parameter bounds using available information. This information can be obtained from the literature, preliminary experiments, or by modeling and analyzing the mechanisms. The next step is to parameterize the experimental conditions. For example, the input profile is often represented by some form that involves only a few parameters, such as a series of piecewise constant functions, to reduce the resulting optimization problem to a finite-dimensional problem. Other experimental conditions that can be parameterized are selection of measurements, sampling points, or initial conditions. All of these variables can be included in the decision vector. The optimal design is then determined by solving an optimization problem. The objective function of this optimization problem is an experimental criterion based on the global sensitivity matrix calculated from Eq. (8-15). The most popular criterion is the determinant of the Gram matrix of the sensitivity matrix or the trace of the inverse of the Gram matrix, however, other criteria can also be applied.

8.4 Three illustrative examples

This section presents three examples that illustrate different aspects of the presented

experimental design procedure. The first example is a generic one while the second and third examples describe chemical reactors.

Identifiability test of a simple model

This test of structural parameter identifiability aims to check whether the parameter values can be determined uniquely from noise free data. If multiple solutions exist for parameter estimation, then the parameters are not identifiable and the estimation problem is ill-posed. Identifiability of a linear regression model is directly related to the rank of the design matrix. If the design matrix has full column rank then the parameters are identifiable and if the matrix is rank deficient, then the parameters are not identifiable. For a nonlinear model the identifiability can be locally evaluated by the rank of the local sensitivity matrix. If the sensitivity matrix is full rank in a neighborhood of a given point, then the parameters are identifiable in a neighborhood of this point. It should be noted that the sensitivity value at only one point may be insufficient for determining identifiability as the rank of the sensitivity matrix may change in the neighborhood of this point.

Consider the model

$$\begin{aligned} g_1(\theta_1,\theta_2) &= \theta_1 + \theta_2^3 \\ g_2(\theta_1,\theta_2) &= \theta_1 - \theta_2^3 \end{aligned} \tag{8-22}$$

where $\theta_1,\theta_2 \in [-\alpha,\alpha]$ and the nominal value vector is $\boldsymbol{\theta} = \mathbf{0}$. As this is a relatively simple example, it is possible to compute an analytical solution for the sensitivity analysis and to conclude that the model is identifiable over this region.

In a first step, the local sensitivity matrix is computed for the nominal values

$$\mathbf{S} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}. \tag{8-23}$$

This sensitivity matrix has a rank of one, which contradicts the observations made about the system above. The reason for this is that the local sensitivity matrix changes rank in a neighborhood containing the nominal value.

As a second method, the global sensitivity matrix is computed using a variance-

based method, where the parameter uncertainty is characterized by a uniform distribution over the region

$$\mathbf{S} = \begin{bmatrix} 1 & \sqrt{\frac{3}{7}\alpha^2} \\ 1 & \sqrt{\frac{3}{7}\alpha^2} \end{bmatrix}. \tag{8-24}$$

This sensitivity matrix also has a rank of one. The reason for this result is that the information about the sign is lost while computing the conditional variance.

To compare these results, the global sensitivity matrix is computed via quasi linearization from Eq. (8-15) for the same parameter uncertainty as the one used for the variance-based method

$$\mathbf{S} = \begin{bmatrix} 1 & \frac{3}{5}\alpha^2 \\ 1 & -\frac{3}{5}\alpha^2 \end{bmatrix}. \tag{8-25}$$

The rank of the sensitivity matrix is two, unless α approaches zero in which case Eq. (8-25) reduces to Eq. (8-23). The results are consistent with what is known about the system.

Apart from identifiability, it is also important to compare other results returned by these three methods. For example, the local sensitivity identifies the parameter θ_1 as the influential parameter regardless of the range of parameter uncertainty. In contrast to this, both global sensitivity methods determine that the uncertainty range has an effect on which of the two parameters is most influential. If the range is small then the parameter θ_1 is influential, however, if the range is large then the parameter θ_2 becomes more important. This is due to the structure of the system where the parameter θ_1 appears linearly while the parameter θ_2 is taken to the third power in Eq. (8-22). This ability to take the parameter uncertainty into account is one of the advantages of global sensitivity analysis.

Batch reactor with two reactions in series

Suppose two consecutive reactions are taking place in a batch reactor (Fogler, 2005)

$$\mathrm{A} \xrightarrow{k_1} \mathrm{B} \xrightarrow{k_2} \mathrm{C}$$

in which species B is the desired product. The reactions are irreversible and first order with regard to species A and B, respectively. For the initial concentrations, $C_A(0)=1$ mol/l and $C_B(0)=0$, the concentration of B is

$$C_B = \frac{k_1}{k_2 - k_1}\left(e^{-k_1 t} - e^{-k_2 t}\right). \tag{8-26}$$

Even though this is a linear dynamic system, the output C_B is nonlinearly dependent on the parameters k_1 and k_2. The ranges of the kinetic parameters are chosen as

$$k_1 \in \left[1-\alpha_1 \quad 1+\alpha_1\right] \text{ and } k_2 \in \left[1-\alpha_2 \quad 1+\alpha_2\right] \tag{8-27}$$

and the nominal values are $\bar{k}_1 = 1\ \mathrm{min}^{-1}$ and $\bar{k}_2 = 1\ \mathrm{min}^{-1}$.

Three sensitivity measures are calculated for the two parameters: the local sensitivity, the global sensitivity via the conditional variance computed by FAST and the global sensitivity via the quasi linearization. The set of rationally independent numbers are selected as $\omega_1 = 3$ and $\omega_2 = 7$. To demonstrate the effect of parameter uncertainty on the experimental design, two sets of uncertain ranges are used: a small uncertainty with $\alpha_1 = 0.1\ \mathrm{min}^{-1}$ and $\alpha_2 = 0.1\ \mathrm{min}^{-1}$ and a large uncertainty with $\alpha_1 = 0.9\ \mathrm{min}^{-1}$ and $\alpha_2 = 0.9\ \mathrm{min}^{-1}$. In both cases the parameters are assumed to be uniformly distributed over these intervals.

In the case of $\alpha_1 = 0.1\ \mathrm{min}^{-1}$ and $\alpha_2 = 0.1\ \mathrm{min}^{-1}$, the sensitivity profiles are shown in Fig. 8-2. The global sensitivity via quasi linearization reduces to the local sensitivity. For the global sensitivity via the variance-based method, only the magnitude of the sensitivity value reduces to the local sensitivity since the global sensitivity values are always non-negative.

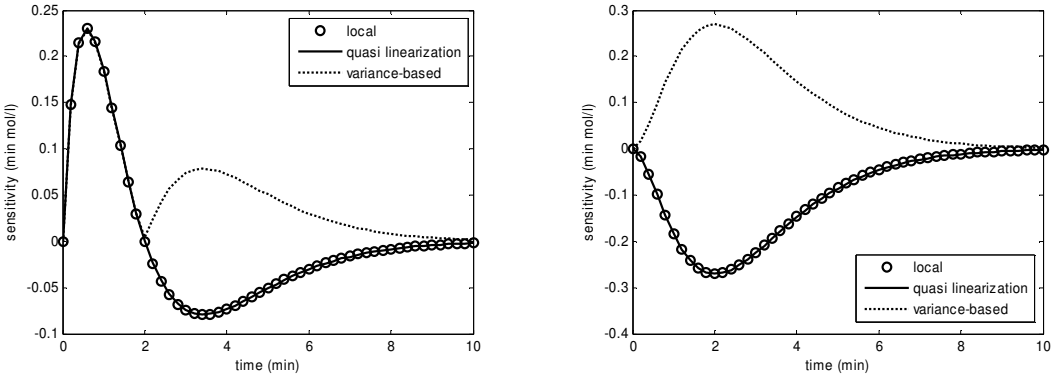


Fig. 8-2. Sensitivity of the concentration of the species B with respect to (a) k_1 and (b) k_2 in the case of small uncertainty. (The uncertainty ranges are $\alpha_1 = 0.1 \text{ min}^{-1}$ and $\alpha_2 = 0.1 \text{ min}^{-1}$.)

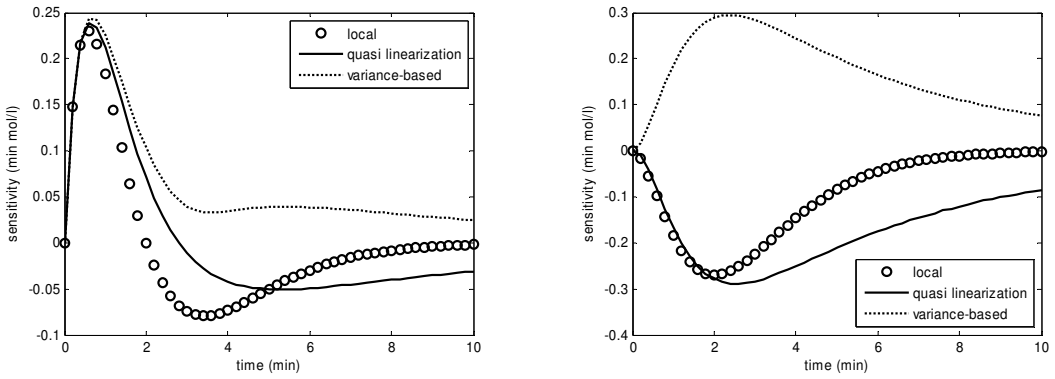


Fig. 8-3. Sensitivity of the concentration of the species B with respect to (a) k_1 and (b) k_2 in the case of large uncertainty. (The uncertainty ranges are $\alpha_1 = 0.9 \text{ min}^{-1}$ and $\alpha_2 = 0.9 \text{ min}^{-1}$.)

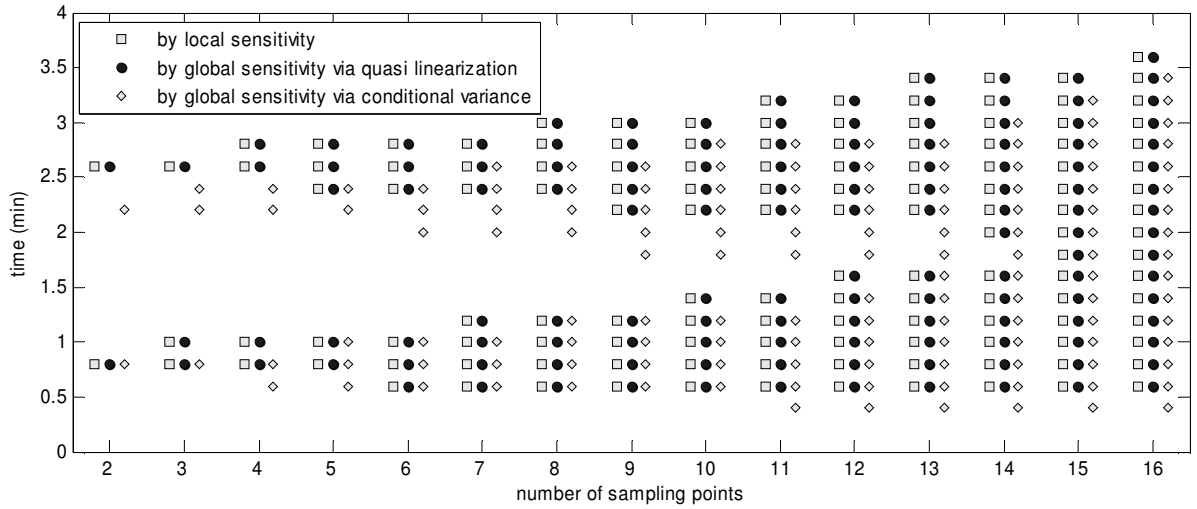
In the case of $\alpha_1 = 0.9\ \text{min}^{-1}$ and $\alpha_2 = 0.9\ \text{min}^{-1}$, the three methods return different sensitivity profiles for both parameters as is shown in Fig. 8-3. The local sensitivity profile is the same as the one for the small uncertainty case since the sensitivity value is unaffected by the uncertainty. However, the values of each global sensitivity measure are different for the two cases since the information about the parameter uncertainty is taken into account for the calculation of the sensitivity value.

A comparison of the experimental designs returned by the three sensitivities is performed by selecting the optimal sampling points based on the *D*-optimality criterion of the sensitivity matrix. The candidate sampling points were chosen every 0.2 min for a time span from 0 to 10 min. At least two and at most 50 sampling points were required to estimate the two parameters. The optimal sampling points were computed using the three sensitivity measures for each number of sampling points. The sets of the sampling points for small uncertainties are shown in Fig. 8-4(a) and those for large uncertainties are shown in Fig. 8-5(a). For some number of sampling points the results returned by the different methods are identical and those results are not shown.

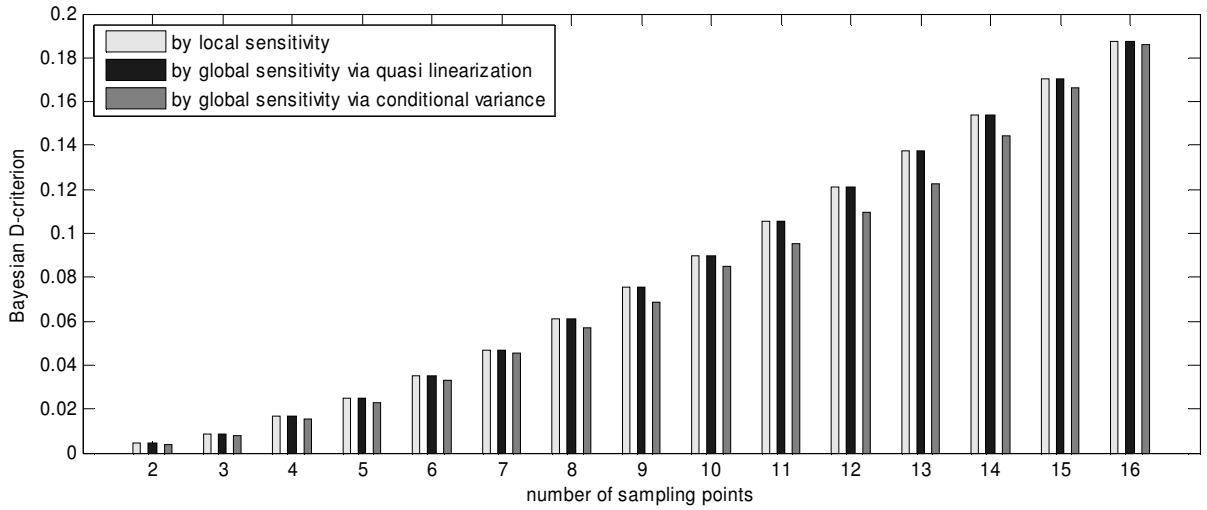
To evaluate the performance of each method over the entire uncertainty region the Bayesian *D*-optimality criterion is calculated for each design. The Bayesian *D*-optimality criterion is the mean value of the *D*-optimality evaluated according to the parameter uncertainty

$$\varphi_{BD}(\xi) = \int \cdots \int \varphi_D(\xi, \boldsymbol{\theta})\, p(\boldsymbol{\theta}) \prod_i d\theta_i \tag{8-28}$$

where ξ denotes a experimental design, $\varphi_D(\xi, \boldsymbol{\theta})$ is the *D*-criterion of the local sensitivity matrix evaluated at a parameter point $\boldsymbol{\theta}$ for the given design, and $p(\boldsymbol{\theta})$ is the density function of the parameters. The value of $\varphi_D(\xi, \boldsymbol{\theta})$ assesses the design in a neighborhood of the parameter value $\boldsymbol{\theta}$ and the mean value describes the overall performance of a design over the entire uncertainty region. The Bayesian criterion is a widely used approach to evaluate a design under uncertainty and is generally acknowledged to be superior to the criterion value at only one given point.

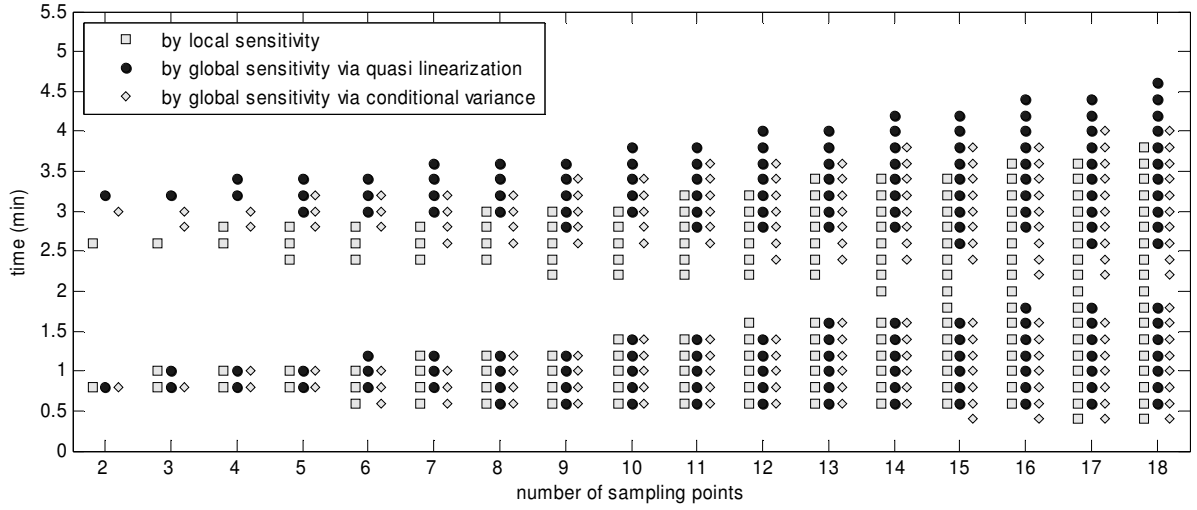


(a)

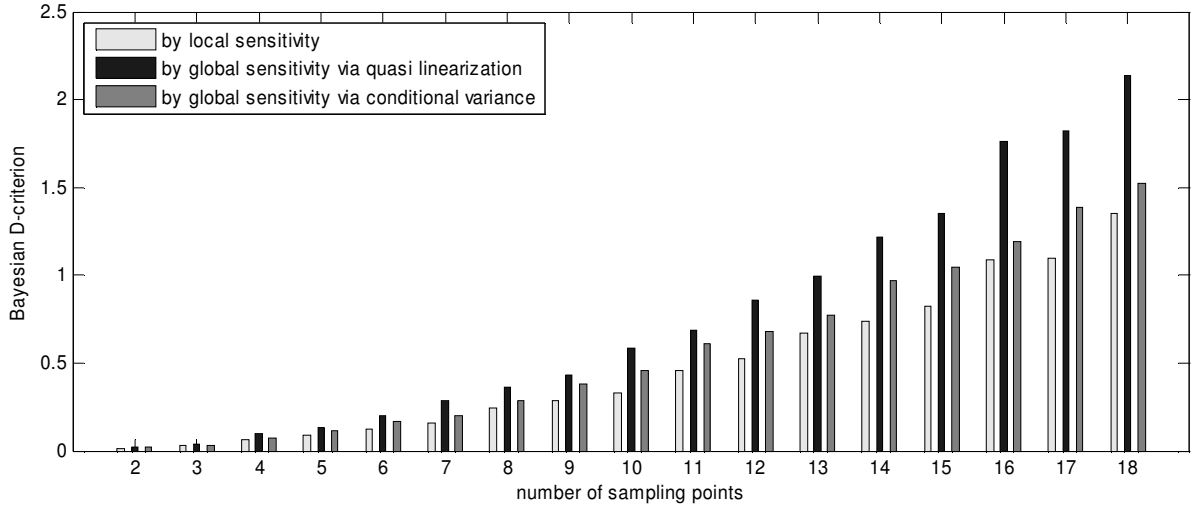


(b)

Fig. 8-4. Experimental designs in the case of small uncertainty. (a) Selected time points; (b) Bayesian *D*-criterion. (The uncertainty ranges are $\alpha_1 = 0.1\ \text{min}^{-1}$ and $\alpha_2 = 0.1\ \text{min}^{-1}$.)



(a)



(b)

Fig. 8-5. Experimental designs in the case of large uncertainty. (a) Selected time points; (b) Bayesian *D*-criterion. (The uncertainty ranges are $\alpha_1 = 0.9\ \text{min}^{-1}$ and $\alpha_2 = 0.9\ \text{min}^{-1}$.)

The Bayesian criteria of the designs computed by the three sensitivity analysis techniques for small uncertainties are shown in Fig. 8-4(b). Since the parameter uncertainty is negligible, the Bayesian criterion is close to the local *D*-criterion at the nominal point, i.e., the design based on the local sensitivity matrix is near optimal. The design based upon global sensitivity analysis via quasi linearization achieves approximately the same performance as the local design since the global sensitivity reduces to the local sensitivity. However, the design by global sensitivity analysis using conditional variances returns a smaller value of the Bayesian criterion.

The results of the Bayesian criterion for large uncertainties are shown in Fig. 8-5(b). The design based upon global sensitivity analysis via quasi linearization returns the best performance while the design based upon local sensitivity analysis returns the smallest criterion value. The design by local sensitivity analysis achieves the best performance when the true parameters are close to the nominal parameter values. However, if the parameter uncertainty is significant then the best design at one point can be the worst at another point and on average the local design is sub-optimal. The designs by global sensitivity analysis return better results than the local design since the parameter uncertainty is taken into account.

To verify the significance of the difference in the mean criterion values shown in Fig. 8-4(b) and Fig. 8-5(b) a hypothesis test is performed

$$H_0 : m_A - m_B = 0 \text{ against } H_1 : m_A - m_B > 0 \tag{8-29}$$

where m_A and m_B are the mean values of method *A* and method *B*, respectively. In this case the subscript *A* denotes the design by the quasi linearization method while the subscript *B* denotes the design by the local method or the design by the variance-based method. The *P*-value of the test for every case is close to zero which indicates that the difference between the mean values is significant.

Reactor with van de Vusse reaction kinetics

The second case study deals with an isothermal CSTR in which a van de Vusse reaction is taking place (van de Vusse, 1964)

$$\mathrm{A} \underset{k2}{\overset{k_1}{\rightleftarrows}} \mathrm{B} \xrightarrow{k_3} \mathrm{C}$$
$$\mathrm{A} \xrightarrow{k_4} D$$

The model consisting of the component balances for species A and B is given by

$$\begin{aligned} \dot{C}_A &= -k_1 C_A + k_2 C_B - k_4 C_A^2 + u\left(C_{Af} - C_A\right) \\ \dot{C}_B &= k_1 C_A - \left(k_2 + k_3\right) C_B - u C_B \\ y &= C_A \end{aligned} \quad .$$

The objective is to design a profile of the input u and the initial conditions C_{A0} and C_{B0} to generate an output y so that the kinetic parameters k_1, k_2, k_3, k_4 can be accurately estimated. The nominal values of the kinetic parameters were taken from the literature (Doyle et al., 1995) and are listed in Table 8-1.

The kinetic parameters for estimation were assumed to be log-uniformly distributed from 50% of the nominal value to 200% of the nominal value where the nominal value is the mean value. The *A*-optimality criterion is used to find the optimal experimental condition. This criterion minimizes the sum of the variances of the estimated parameters. Since the parameters have different units, they are normalized by dividing them by their nominal values $\theta_i = k_i / \overline{k}_i$, $i = 1 \cdots 4$. After normalization all parameters have no unit and are distributed from 0.5 to 2 with the mean equal to 1.

Table 8-1

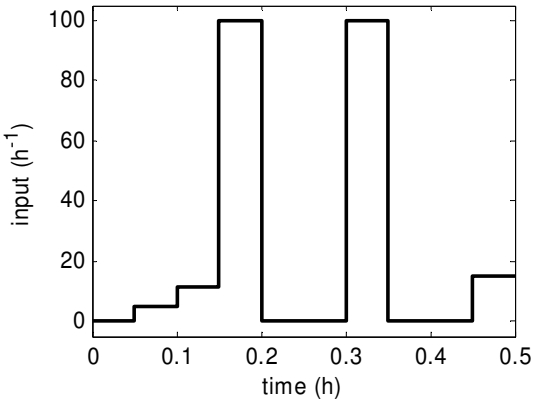
Values of the parameters

type	variable	nominal value	range	unit
parameter for estimation	k_1	50	25~100	h^{-1}
	k_2	100	50~200	h^{-1}
	k_3	100	50~200	h^{-1}
	k_4	10	5~20	$l\ mol^{-1}\ h^{-1}$
design variable	u	-	0~100	h^{-1}
	C_{A0}	-	0~5	$mol\ l^{-1}$
	C_{B0}	-	0~5	$mol\ l^{-1}$
constant	C_{Af}	10	-	$mol\ l^{-1}$

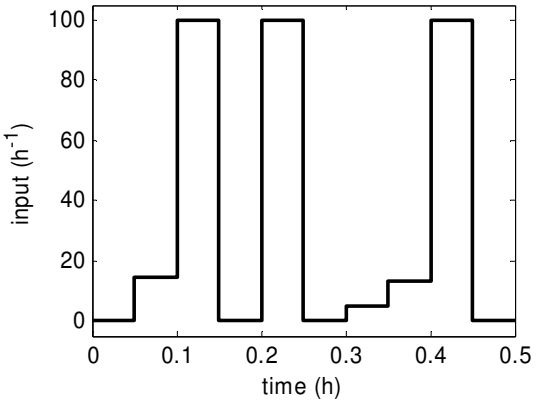
The data for estimation were generated by adding Gaussian distributed noise with zero mean and variance $\sigma^2 = 0.01$ to the output. The output was sampled every 0.01 hr in the range from 0 hr to 0.5 hr. The input was assumed to be piecewise constant over the time interval and during each 0.05 hr the input was fixed at some level. The range of the initial values was chosen to be from 0 to 5 mol l^{-1}.

Fig. 8-6(a) shows the optimal profiles of the input according to the *A*-optimality criterion computed by the local sensitivity analysis and Fig. 8-6(b) shows the profile computed by the global sensitivity analysis via quasi linearization. There are distinct differences between the two input profiles. The initial values returned by the two designs are identical for C_{A0}=0 and C_{B0}=5 mol l^{-1}.

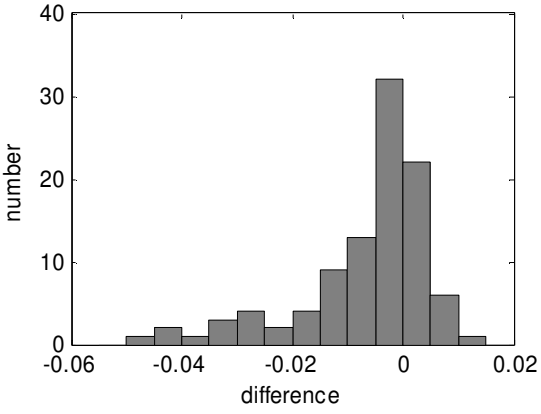
The *A*-optimal design minimizes the variance of the estimated parameters. If the true parameter values are identical to the nominal values then the local design is optimal. To calculate the variance of the estimated parameter values, 100 data sets were generated by adding different noise signals to the output and estimate the parameters for each data set. The variance of the parameters is computed from the estimated parameter values. The variance of the parameters for the design using local sensitivity analysis is 0.0769 while the variance of the parameters for the design involving global sensitivity analysis is 0.0795. However, if the true parameter values are not close to the nominal values, then the design by local sensitivity analysis may return poorer results than the design by global sensitivity analysis. To illustrate the effect of parameter uncertainty on the design, 100 parameter values were sampled over the uncertainty range. The averaged variances returned by the local sensitivity analysis experimental design is 0.1108 and the averaged variance returned by the global sensitivity analysis experimental design is 0.1039. Fig. 8-6(c) shows the distribution of the differences of the variances of estimated parameters between the two designs. It can be seen that the design using global sensitivity analysis returns on average smaller variances than the design based upon local sensitivity analysis.



(a)



(b)



(c)

Fig. 8-6. Optimal input profile by (a) Local design and (b) Global design. (c) Distribution of differences in the variance of estimated parameters by the two designs.

8.5 Conclusions

Local sensitivity analysis is a widely used technique in experimental design, however, the dependence of the sensitivity results on the parameter values makes the design only valid in a neighborhood of the nominal parameter values. Global sensitivity analysis does not have this drawback and provides a promising alternative for experimental design. However, most existing global sensitivity analysis techniques do not reduce to local sensitivity analysis procedures, even if the model under investigation is linear. As a result, most applications of global sensitivity analysis deal with qualitative experimental design, i.e., determination of important parameters.

This section presented a global sensitivity analysis technique that can under appropriate conditions reduce to a local sensitivity analysis method. The technique is derived via quasi linearization of the nonlinear model and the parameter uncertainty is explicitly taken into account in the calculation of the global sensitivity. This technique is then incorporated into a quantitative experimental design procedure as it represents an extension of an existing local sensitivity analysis procedure. Existing optimal design criteria, such as the *D*-optimality criterion or the *A*-optimality criterion, can be applied to the global sensitivity matrix to select optimal sampling points and determine the optimal input profile. It was shown in case studies that the design based on global sensitivity analysis outperforms the design based on local sensitivity analysis if the entire uncertainty space of the parameters is considered.

9. CONCLUSIONS

This dissertation presents several new techniques that deal with problems related to parameter estimation of complex models. Reducing model complexity is one of the keys to reliable parameter estimation and sensitivity analysis plays a key role in this task. The work compares several widely-used local and global methods for sensitivity analysis and also develops new techniques to overcome the drawbacks of existing methods. These techniques can be applied to identify the important sources of uncertainty, which, in turn, contribute most to variation in the model behavior. Further analysis can then focus on the identified important components while other unimportant components can be safely eliminated.

Sensitivity analysis is a very useful screening tool, however, an approach that regularizes the ill-conditioned estimation problem is still required. This work presents several methods for parameter selection to determine a subset of estimable parameters. The subsequent estimation algorithm only adjusts the value of these selected parameters to fit the data while the unselected parameters are fixed at a constant value. The parameter selection methods not only reduce the effect of noise in the data and return reliable estimation results but also reduce the computational load of the parameter estimation problem.

Another key to improve the estimation accuracy is to increase the information content in the experimental data. This goal can be achieved by optimal experimental design including determination of input profiles, choice of the outputs, and selection of sampling points. Some new techniques for robust experimental design are developed in this work.

The developed methods are applied to different types of models ranging from models found in the process industries to biochemical network models, some of which are described by ordinary differential equations with dozens of state variables and more than a hundred parameters.

9.1 Contributions

Specifically, the contributions of this dissertation are listed in chronological order:

(1) This work compared commonly used sensitivity analysis techniques and applied them to a complex model of the IL-6 signaling transduction network. New insights into the sensitivity results and the underlying mechanism of the network were discovered. (Chu et al., 2007)

(2) This work developed a robust parameter selection procedure for estimation of complex nonlinear dynamic systems. The effect of the parameter uncertainty on the selection was taken into account and the returned result was a collection of parameter subsets rather than only one subset, which was a desirable property in practice. The relationship of the frequently-used orthogonal selection approach with the forward selection framework was investigated. (Chu and Hahn, 2007)

(3) This work created an approach to incorporate parameter selection and experimental design. The two approaches were often performed separately, however, they were highly correlated. The optimality of each individual procedure might not guarantee the optimality for the whole so it was more reasonable to consider them simultaneously. To solve the resulting mix integer nonlinear programming under uncertainty, an efficient method combining stochastic approximation and genetic algorithm had been presented. (Chu and Hahn, 2008)

(4) This work presented a novel algorithm to solve the combinatorial problem of parameter selection. The indistinguishability of parameters was investigated first and the indistinguishable parameters were then grouped by a hierarchical clustering algorithm. The grouping significantly reduced the search region and simplified the solution. The method was as efficient as other forward selection methods however it was able to return a better result. Parameter clustering also provided a useful tool to investigate the underlying mechanism of the analyzed model. (Chu and Hahn, 2009)

(5) This work developed a method to increase the prediction accuracy of a model from parameter selection and estimation. The relation and difference between output predictability and the parameter identifiability were investigated. A new

orthogonalization method was presented which could solve the resulting optimization problem efficiently while returning a more accurate prediction than other methods. (Chu et al., 2009)

(6) This work presented a robust technique for experimental design based on global sensitivity analysis. A new global sensitivity analysis method was developed. The property, which distinguished it from other global sensitivity analysis methods, was that the results were consistent with the local sensitivity analysis. The technique could take the parameter uncertainty into account while avoiding calculation of the partial derivatives which made it less computationally demanding than other robust design strategies. (Chu and Hahn, 2010)

9.2 Future work

Several extensions of the presented work are possible.

Solution to the combinatorial design problem

Some experimental designs, e.g. sampling time selection and sensor location can be formulated as a combinatorial problem of selecting rows from the sensitivity matrix to maximize an experimental criterion. For some criteria the continuous relaxation of the combinatorial problem is a convex problem, the global optimal solution of which can be found efficiently. Taking advantage of the continuous relaxation problem can provide an upper bound for the discrete optimization problem. Future work can focus on branch-and-bound algorithms to solve the combinatorial problem for sensor location. Parameter selection is another combinatorial problem that select columns from a sensitivity matrix, however, it is more difficult to formulate as no such advantage can be employed.

Regularization of state estimation problem

State estimation infers the values of unmeasured state variables from measured state variables. State estimation is an important research area in process engineering and a large variety of techniques have been developed. Similar to parameter estimation, state

estimation of a complex system can also be ill-conditioned. Future work can extend the regularization procedure for the parameter estimation problem to state estimation. Further, the procedure can be extended to the simultaneous estimation of both parameters and states to provide answers to questions such as which parameters should be estimated to infer the state variables of interest, or can a reduced order model provide a more accurate inference of the given states than the full order model?

Sensitivity analysis of black-box models

Empirical black-box models like neural networks are an important tool to model the input-output relationship of given data. A main problem in applications of such type of models is to determine the model structure. A complex structure can result in more accurate models, however, it increases the risk of over-fitting the data. Parameter selection methods can be generalized to black-box models to deal with over-fitting.

Process design under uncertainty

Uncertainty is an inherent characteristic of any process system. The potential effect of uncertainty on model-based optimization results for process design is often not negligible. The techniques of uncertainty and sensitivity analysis studied in this work can be applied to the solution of the optimization problem. First the uncertainty of the parameter values in the design problem can be summarized. Next sensitivity analysis can be applied to identify the important sources of the uncertainty which can significantly influence the optimal solution. The focus can then be placed on these important uncertain parameters.

Analysis including subsequent applications

The uncertainty and sensitivity analysis in this work is applied to the output of the model. However a model is often built for subsequent applications, e.g. control or monitoring. It is desired to perform the analysis on the final results of these applications. For example, a model is often used to design a controller. Some controllers are sensitive

to model uncertainty while others are more robust. Such application-oriented uncertainty and sensitivity analysis can be very helpful in practice.

Structural uncertainty and sensitivity analysis

The uncertainty and sensitivity analysis can be applied to identify the important uncertain parameters. However, in the initial stage of mathematical modeling even the structure of a model is often not exactly known and it is helpful to determine which parts of the model are important. This kind of questions can be answered by extending the sensitivity analysis to some on-off switching parameters. These parameters indicate if the corresponding part of the model is included (with the value of 1) or not (with the value of 0). The analysis of these switching parameters can also be performed simultaneously with analysis of other parameters, however, since the value of some parameters is binary, a new techniques for sensitivity analysis and interpretation for the results are needed.

Measure of nonlinearity

A nonlinear model generally results in more problems than a linear one and more sophisticated techniques are required. However, these sophisticated techniques are more difficult to implement. It is natural to ask the question when a simple method will fail and a sophisticated alternative has to be applied. For example it is well known that local sensitivity analysis is valid for linear or mildly nonlinear models and global sensitivity analysis is more preferable even if the computation is more expensive. However, there is no clear answer to the question of at which degree of nonlinearity of a model the local technique will fail. Similarly, it is unclear if it is required for a given model to apply a global technique. However, these types of questions can be answered with an extension of the techniques presented in this work.

REFERENCES

Alonso, A.A., Frouzakis, C.E., Kevrekidis, I.G., 2004. Optimal sensor placement for state reconstruction of distributed process systems. AIChE Journal 50, 1438-1452.

Anderson, T.W., 1984. An introduction to multivariate statistical analysis 2nd ed. Wiley, New York.

Anh, D.T., Bonnet, M.P., Vachaud, G., Minh, C.V., Prieur, N., Duc, L.V., Anh, L.L., 2006. Biochemical modeling of the Nhue River (Hanoi, Vietnam): Practical identifiability analysis and parameters estimation. Ecological Modelling 193, 182 - 204.

Asprey, S.P., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. Computers & Chemical Engineering 24, 1261-1267.

Asprey, S.P., Macchietto, S., 2002. Designing robust optimal dynamic experiments. Journal of Process Control 12, 545-556.

Aster, R.C., Borchers, B., Thurber C.H., 2005. Parameter Estimation and Inverse Problems. Elsevier Academic Press, Amsterdam.

Atherton, R.W., Schainker, R.B., Ducot, E.R., 1975. On the statistical sensitivity analysis of models for chemical kinetics. AIChE Journal 21, 441-448.

Atkinson, A.C., Bogacka, B., 1997. Compound D- and D-s-optimum designs for determining the order of a chemical reaction. Technometrics 39, 347-356.

Atkinson, A.C., Bogacka, B., 2002. Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. Chemometrics and Intelligent Laboratory Systems 61, 17-33.

Atkinson, A.C., Donev, A.N., Tobias, R.D., 2007. Optimum experimental designs, with SAS. Oxford University Press, Oxford.

Audoly, S., Bellu, G., D'Angio, L., Saccomani, M.P., Cobelli, C., 2001. Global identifiability of nonlinear models of biological systems. IEEE Transactions on Bio-medical Engineering 48, 55-65.

Audoly, S., D'Angio, L., Saccomani, M.P., Cobelli, C., 1998. Global identifiability of linear compartmental models - A computer algebra algorithm. IEEE Transactions on Bio-medical Engineering 45, 36-47.

Balsa-Canto, E., Banga, J.R., Alonso, A.A., Vassiliadis, V.S., 2001. Dynamic optimization of chemical and biochemical processes using restricted second-order information. Computers & Chemical Engineering 25, 539 -546.

Balsa-Canto, E., Peifer, M., Banga, J.R., Timmer, J., Fleck, C., 2008. Hybrid optimization method with general switching strategy for parameter estimation. BMC Systems Biology 2, 26.

Balsa-Canto, E., Rodriguez-Fernandez, M., Banga, J.R., 2007. Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. Journal of Food Engineering 82, 178-188.

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 approach. Industrial & Engineering Chemistry Research 41, 2425-2430.

Bardow, A., 2008. Optimal experimental design of ill-posed problems: The METER approach. Computers & Chemical Engineering 32, 115-124.

Bastogne, T., Thomassin, M., Masse, J., 2007. Selection and identification of physical parameters from passive observation. Application to a winding process. Control Engineering Practice 15, 1051-1061.

Bates, D.M., Watts, D.G., 1980. Relative Curvature Measures of Nonlinearity. Journal of the Royal Statistical Society. Series B 42, 1 -25.

Bauer, I., Bock, H.G., Korkel, S., Schloder, J.P., 2000. Numerical methods for optimum experimental design in DAE systems. Journal of Computational and Applied Mathematics 120, 1-25.

Benabbas, L., Asprey, S.P., Macchietto, S., 2005. Curvature-based methods for designing optimally informative experiments in multiresponse nonlinear dynamic situations. Industrial & Engineering Chemistry Research 44, 7120-7131.

Bentele, M., Lavrik, I., Ulrich, M., Stosser, S., Heermann, D.W., Kalthoff, H., Krammer, P.H., Eils, R., 2004. Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis. Journal of Cell Biology 166, 839-851.

Bohachevsky, I.O., Johnson, M.E., Stein, M.L., 1986. Generalized simulated annealing for function optimization. Technometrics 28, 209-217.

Borgonovo, E., 2006. Measuring uncertainty importance: Investigation and comparison of alternative approaches. Risk Analysis 26, 1349-1361.

Box, G.E.P., Lucas, H.L., 1959. Design of experiments in non-linear situations. Biometrika 46, 77-90.

Box, G.E.P., Hunter, W.G., 1962. Useful method for model-building. Technometrics 4, 301-318.

Box, G.E.P., Draper, N.R., 1975. Robust designs. Biometrika 62, 347-352.

Brockmann, D., Rosenwinkel, K.H., Morgenroth, E., 2008. Practical identifiability of biokinetic parameters of a model describing two-step nitrification in biofilms. Biotechnology and Bioengineering 101, 497-514.

Brown, P.N., Byrne, G.D., Hindmarsh, A.C., 1989. VODE—A variable-coefficient ode solver. SIAM Journal of Scientific and Statistical Computing 10, 1038-1051.

Brown, P.N., Hindmarsh, A.C., Petzold, L.D., 1994. Using Krylov methods in the solution of large-scale differential-algebraic systems. SIAM Journal of Scientific and Statistical Computing 15, 1467-1488.

Brun, R., Reichert, P., Kunsch, H.R., 2001. Practical identifiability analysis of large environmental simulation models. Water Resources Research 37, 1015-1030.

Brun, R., Kuhni, M., Siegrist, H., Gujer, W., Reichert, P., 2002. Practical identifiability of ASM2d parameters - systematic selection and tuning of parameter subsets. Water Research 36, 4113-4127.

Buzzi-Ferraris, G., Forzattia, P., 1983. A new sequential experimental design procedure for discriminating among rival models. Chemical Engineering Science 38, 225 -232.

Byrnea, G.D., Hindmarsh, A.C., 1987. Stiff ODE solvers: A review of current and coming attractions. Journal of Computational Physics 70, 1-62.

Cacuci, D.G., Ionescu-Bujor, M., 2004. A comparative review of sensitivity and uncertainty analysis of large-scale systems - II: Statistical methods. Nuclear Science and Engineering 147, 204-217.

Castillo, E., Conejo, A.J., Minguez, R., Castillo, C., 2006. A closed formula for local sensitivity analysis in mathematical programming. Engineering Optimization 38, 93-112.

Chaloner, K., 1993. A note on optimal Bayesian design for nonlinear problems. Journal of Statistical Planning and Inference 37, 229 -235.

Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: A review. Statistical Science 10, 273-304.

Chappell, M.J., Godfrey, K.R., Vajda, S., 1990. Global identifiability of the Parameters of Nonlinear-systems with specified inputs-A comparison of methods. Mathematical Biosciences 102, 41-73.

Chaudhuri, P., Mykland, P.A., 1993. Nonlinear experiments: optimal design and inference based on likelihood. Journal of the American Statistical Association 88, 538 -546.

Chernoff, H., 1953. Locally optimal designs for estimating parameters. Annals of Mathematical Statistics 24, 586-602.

Chhatre, S., Francis, R., Newcombe, A.R., Zhou, Y.H., Titchener-Hooker, N., King, J., Keshavarz-Moore, E., 2008. Global sensitivity analysis for the determination of parameter importance in bio-manufacturing processes. Journal of Chemical Technology and Biotechnology 51, 79-90.

Chib, S., Greenberg, E., 1995; Understanding the Metropolis-Hastings algorithm. American Statistician 49, 327 -335.

Chin, D.C., 1997. Comparative study of stochastic algorithms for system optimization based on gradient approximations. IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics 27 2, 244-249.

Cho, K.H., Shin, S.Y., Kolch, W., Wolkenhauer, O., 2003. Experimental design in systems biology, based on parameter sensitivity analysis using a Monte Carlo

method: A case study for the TNFα-mediated NF-κB signal transduction pathway. Simulation 79, 726-739.

Chu, Y., Hahn, J., 2007. Parameter set selection for estimation of nonlinear dynamic systems. AICHE Journal 53, 2858-2870.

Chu, Y., Hahn, J., 2008. Integrating parameter selection with experimental design under uncertainty for nonlinear dynamic systems. AIChE Journal 54, 2310-2320.

Chu, Y., Hahn, J., 2009. Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters. Industrial & Engineering Chemistry Research 48, 6000-6009.

Chu, Y., Hahn, J., 2010. Quantitative optimal experimental design using global sensitivity analysis via quasi-linearization. Industrial & Engineering Chemistry Research 49, 7782-7794.

Chu, Y., Huang, Z., Hahn, J., 2009. Improving prediction capabilities of complex dynamic models via parameter selection and estimation. Chemical Engineering Science 64, 4178-4185.

Chu, Y., Jayaraman, A., Hahn, J., 2007. Parameter sensitivity analysis of IL-6 signalling pathways. IET Systems Biology 1, 342-352.

Cobb, R.G., Liebst, B.S., 1997. Sensor placement and structural damage identification from minimal sensor information. AIAA Journal 35, 369-374.

Cobelli, C., DiStefano, J.J., 1980. Parameter and structural identifiability concepts and ambiguities: A critical review and analysis. American Journal of Physiology 239, 7-24.

Cook, R.D., Nachtsheim, C.J., 1980. A Comparison of Algorithms for Constructing Exact D-Optimal Designs. Technometrics 22, 315-324.

Cook, R.D., Wong, W.K., 1994. On the equivalence of constrained and compound optimal designs. Journal of the American Statistical Association 89, 687-692.

Cropp, R.A., Braddock, R.D., 2002. The new Morris method: An efficient second-order screening method. Reliability Engineering & System Safety 78, 77-83.

Cukier, R.I., Fortuin, C.M., Shuler, K.E., Pestschek, A.G., Schaibly, J.H., 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. Journal of Chemical Physics 59, 3873-3878.

Cukier, R.I., Levine, H.B., Shuler, K.E., 1978. Nonlinear sensitivity analysis of multiparameter model systems. Journal of Computational Physics 26, 1-42.

Cukier, R.I., Schaibly, J.H., Shuler, K.E., 1975. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. III. Analysis of the approximations. Journal of Chemical Physics 63, 1140-1149.

Daescu, D.N., Sandu, A., Carmichael, G.R., 2003. Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: II - Numerical validation and applications. Atmospheric Environment 37, 5097-5114.

D'Argenio, D.Z., 1981. Optimal sampling times for pharmacokinetic experiments. Journal of Pharmacokinetics and Pharmacodynamics 9, 739 -756.

Das, S., Spall, J.C., Ghanem, R., 2010. Efficient Monte Carlo computation of Fisher information matrix using prior information. Computational Statistics & Data Analysis 54, 272-289.

deAguiar, P.F., Bourguignon, B., Khots, M.S., Massart, D.L., PhanThanLuu, R., 1995. D-optimal designs. Chemometrics and Intelligent Laboratory Systems 30, 199-210.

Degenring, D., Froemel, C., Dikta, G., Takors, R., 2004. Sensitivity analysis for the reduction of complex metabolism models. Journal of Process Control 14, 729 -745.

Degerman, M., Westerberg, K., Nilsson, B., 2009. Determining critical process parameters and process robustness in preparative chromatography - A model-based approach. Chemical Engineering & Technology 32, 903-911.

Detre, C., Kiss, E., Varga, Z., Ludanyi, K., Paszty, K., Enyedi, A., Kovesdi, D., Panyi, G., Rajnavolgyi, E., Matko, J., 2006. Death or survival: Membrane ceramide controls the fate and activation of antigen-specific T-cells depending on signal strength and duration. Cellular Signalling 18, 294-306.

Dette, H., 1997. Designing experiments with respect to 'standardized' optimality criteria. Journal of the Royal Statistical Society Series B – Methodological 59, 97-110.

Dette, H., Melas, V.B., Pepelyshev, A. Strigul, N., 2005. Robust and efficient design of experiments for the Monod model. Journal of Theoretical Biology 234, 537-550.

Dickinson, R.P., Gelinas, R.J., 1976. Sensitivity analysis of ordinary differential equation systems—A direct method. Journal of Computational Physics 21, 123-143.

Dimelow, R.J., Wilkinson, S.J., 2009. Control of translation initiation: A model-based analysis from limited experimental data. Journal of the Royal Society Interface 6, 51-61.

Doyle, F.J., Ogunnaike, B.A., Person, R.K., 1995. Nonlinear model-based control using 2nd-order volterra models. Automatica, 31, 697-714.

Duda, R.O., Hart, P.E., Stork, D.G., 2006. Pattern Classification, 2nd edition. Wiley, New York.

Dunker, A.M., 1984. The decoupled direct method for calculating sensitivity coefficients in chemical kinetics. Journal of Chemical Physics 81 5, 2385-2393.

Emery, A.F., Nenarokomov, A.V., 1998. Optimal experiment design. Measurement Science & Technology 9, 864-876.

Fogler, H.S., 2005. Elements of Chemical Reaction Engineering, 4th edition. Prentice Hall PTR, Upper Saddle River, NJ.

Ford, I., Silvey, S.D., 1980. A sequentially constructed design for estimating a nonlinear parametric function. Biometrika 67, 381 -388.

Forssell, U., Ljung, L., 2000. Some results on optimal experiment design. Automatica 36, 749-756.

Franceschini, G., Macchietto, S., 2008. Model-based design of experiments for parameter precision: State of the art. Chemical Engineering Science 63, 4846 -4872.

Frank, P.M., 1978. Introduction to System Sensitivity Theory. Academic Press, New York.

Frey, H.C., Patil, S.R., 2002. Identification and review of sensitivity analysis methods. Risk Analysis 22, 553-578.

Gadkar, K.G., Gunawan, R., Doyle, F.J., 2005a. Iterative approach to model identification of biological networks. BMC Bioinformatics 6, 155.

Gadkar, K.G., Varner, J., Doyle, F.J., 2005b. Model identification of signal transduction networks from data using a state regulator problem. Systems Biology 2, 17-30.

Gennemark, P., Wedelin, D., 2007. Efficient algorithms for ordinary differential equation model identification of biological systems. IET Systems Biology 1, 120-129.

Gill, P.E., Murray, W., Saunders, M.A., 2002. SNOPT: An SQP algorithm for large-scale constrained optimization. SIAM Journal on Optimization 12, 979-1006.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, MA.

Goodwin, G.C., Aguero, J.C., Welsh, J.S., Yuz, J.I., Adams, G.J., Rojas, C.R., 2008. Robust identification of process models from plant data. Journal of Process Control 18, 810-820.

Grewal, M., Glover, K., 1976. Identifiability of linear and nonlinear dynamical systems. IEEE Transactions on Automatic Control 21, 833 -836.

Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., Sethna, J.P., 2007. Universally sloppy parameter sensitivities in systems biology models. PLOS Computational Biology 3, 1871-1878.

Hahn, J., Edgar, T.F., 2001. A gramian based approach to nonlinearity quantification and model classification. Industrial & Engineering Chemistry Research 40, 5724 -5731.

Hahn, J., Edgar, T.F., 2002a. An improved method for nonlinear model reduction using balancing of empirical gramians. Computers & Chemical Engineering 26, 1379 -1397.

Hahn, J., Edgar, T.F., 2002b. Balancing approach to minimal realization and model reduction of stable nonlinear systems. Industrial & Engineering Chemistry Research 41, 2204 -2212.

Hahn, J., Edgar, T.F., Marquardt, W., 2003. Controllability and observability covariance matrices for the analysis and order reduction of stable nonlinear systems. Journal of Process Control 13, 115 -127.

Hamada, M, Martz, H.F., Reese, C.S., Wilson, A.G., 2001. Finding near-optimal Bayesian experimental designs via genetic algorithms. American Statistician 55, 175-181.

Hamby, D.M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. Environmental Monitoring and Assessment 32, 135 -154.

Han, C., Chaloner, K., 2004. Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics. Biometrics 60, 25-33.

Hangos, I.T., Cameron, K.M., 2001. Process modelling and model analysis. Academic Press, San Diego.

Harris, T.J., Macgregor, J.F., Wright, J.D., 1980. Optimal sensor location with an application to a packed bed tubular reactor. AIChE Journal 26, 910 -916.

Hay, A., Borggaard, J.T., Pelletier, D., 2009. Local improvements to reduced-order models using sensitivity analysis of the proper orthogonal decomposition. Journal of Fluid Mechanics 629, 41-72.

Heinrich, P.C., Behrmann, I., Haan, S., Hermanns, H.M., Muller-Newen, G., Schaper, F., 2003. Principles of interleukin (IL)-6-type cytokine signalling and its regulation. Biochemical Journal 374, 1-20.

Helton, J.C., 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliability Engineering & System Safety 42, 327 -367.

Helton, J.C., Davis, F.J., 2002. Illustration of sampling-based methods for uncertainty and sensitivity analysis. Risk Analysis 22, 591-622.

Helton, J.C., Davis, F.J., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety 81, 23-69.

Helton, J.C., Davis, F.J., Johnson J.D., 2005. A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. Reliability Engineering & System Safety 89, 305-330.

Helton, J.C., Johnson, J.D., Sallaberry, C.J., Storlie, C.B., 2006. Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliability Engineering & System Safety 91, 1175-1209.

Hengl, S., Kreutz, C., Timmer, J., Maiwald, T., 2007. Data-based identifiability analysis of non-linear dynamical models. Bioinformatics 23, 2612-2618.

Heredia-Langner, A., Carlyle, W.M., Montgomery, D.C., Borror, C.M., Runger, G.C., 2003. Genetic algorithms for the construction of D-optimal designs. Journal of Quality Technology 35, 28-46.

Hermann, R., Krener, A., 1977. Nonlinear controllability and observability. IEEE Transactions on Automatic Control 22, 728-740.

Higham, N.J., Konstantinov, M., Mehrmann, V., Petkov, P., 2004. The sensitivity of computational control problem. IEEE Control Systems Magazine 24, 28-43.

Hildebrand, R., Gevers, M., 2003. Identification for control: Optimal input design with respect to a worst-case nu-gap cost function. SIAM Journal on Control and Optimization 41, 1586-1608.

Hill, P.D., 1978. A review of experimental design procedures for regression model discrimination. Technometrics 20, 15 -21.

Hindmarsh, A.C., Brown, P.N., Grant, K.E., Lee, S.L., Serban, R., Shumaker, D.E., Woodward, C.S., 2005. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. ACM Transactions on Mathematical Software 31, 363-396.

Hiskens, I.A., 2001. Nonlinear dynamic model evaluation from disturbance measurements. IEEE Transactions on Power Systems 16, 702 -710.

Hjalmarsson, H., 2009. System identification of complex and structured systems. European Journal of Control 15, 275-310.

Hlavacek, W.S., 2009. How to deal with large models? Molecular Systems Biology 5, 240.

Ho, T.C., 2008. Kinetic modeling of large-scale reaction systems. Catalysis Reviews 50, 287-378.

Hoel, P.G., 1965. Minimax designs in two dimensional regression. Annals of Mathematical Statistics 36, 1097-1106.

Hoffmann, A., Levchenko, A., Scott, M.L., Baltimore, D., 2002. The I kappa B-NF-kappa B signaling module: Temporal control and selective gene activation. Science 298, 1241-1245.

Homma, T., Saltelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering & System Safety 52, 1-17.

Hornberger, G.M., Spear, R.C., 1981. An approach to the preliminary-analysis of environmental systems. Journal of Environmental Management 12, 7-18.

Hosten, L.H., 1974. A sequential experimental design procedure for precise parameter estimation based upon the shape of the joint confidence region. Chemical Engineering Science 29, 2247 -2252.

Hu, D.W., Yuan, J.M., 2006. Time-dependent sensitivity analysis of biological networks: Coupled MAPK and PI3K signal transduction pathways. Journal of Physical Chemistry 110, 5361-5370.

Huang, Y.J., Reklaitis, G.V., Venkatasubramanian, V., 2003. A heuristic extended Kalman filter based estimator for fault identification in a fluid catalytic cracking unit. Industrial & Engineering Chemistry Research 42, 3361-3371.

Hussain, M.A., 1999. Review of the applications of neural networks in chemical process control - simulation and online implementation. Artificial Intelligence in Engineering 13, 55-68.

Hwang, J.T., Dougherty, E.P., Rabitz, S., Rabitz, H., 1978. Greens function method of sensitivity analysis in chemical-kinetics. Journal of Chemical Physics 69, 5180-5191.

Iman, R.L., Helton, J.C., 1988. An investigation of uncertainty and sensitivity analysis techniques for computer models. Risk Analysis 8: 71-90.

Iman, R.L., Helton, J.C., Campbell, J.E., 1981. An approach to sensitivity analysis of computer models. Part 1. Introduction, input variable selection and preliminary variable assessment. Journal of Quality Technology 13, 174-183.

Ingalls, B.P., Sauro, H.M., 2003. Sensitivity analysis of stoichiometric networks: An extension of metabolic control analysis to non-steady state trajectories. Journal of Theoretical Biology, 222, 23-36.

Ionescu-Bujor, M., Cacuci, D.G., 2004. A comparative review of sensitivity and uncertainty analysis of large-scale systems - I: Deterministic methods. Nuclear Science and Engineering 147, 189-203.

Issanchou, S., Cognet, P., Cabassud, M., 2005. Sequential experimental design strategy for rapid kinetic modeling of chemical synthesis. AIChE Journal 51, 1773-1781.

Jacquez, J.A., Greif, P., 1985. Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. Mathematical Biosciences 77, 201 -227.

Jacquez, J.A., Perry, T., 1990. Parameter estimation: Local identifiability of parameters. American Journal of Physiology 258, 727-736.

Jaqaman, K., Danuser, G., 2006. Linking data to models: Data regression. Nature Reviews Modecular Cell Biology 7, 813-819.

Jayasankar, B.R., Ben-Zvi, A., Huang B., 2009. Identifiability and estimability study for a dynamic solid oxide fuel cell model. Computers & Chemical Engineering 33, 484-492.

Joshi, M., Seidel-Morgenstern, A., Kremling, A., 2006. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. Metabolic Engineering 8, 447-455.

Jwo, D.J., Cho, T.S., 2007. A practical note on evaluating Kalman filter performance optimality and degradation. Applied Mathematics and Computation 193, 482-505.

Keller, J.P., Bonvin, D., 1992. Selection of input and output variables as a model reduction problem. Automatica 28, 171 -177.

Kholodenko, B.N., 2006. Cell-signalling dynamics in time and space. Nature Reviews. Molecular Cell Biology 7, 165-176.

Kiefer, J., Wolfowitz, J., 1952. Stochastic estimation of the maximum of a regression function. Annals of Mathematical Statistics 23, 462-466.

Kiefer, J., 1959. Optimum experimental designs. Journal of the Royal Statistical Society. Series B, Statistical Methodology 21, 272-319.

Kiefer, J., Wolfowitz, J., 1959. Optimum designs in regression problems. Annals of Mathematical Statistics 30, 271-294.

Kiefer, J., 1974. General equivalence theory for optimum designs (approximate theory). Annals of Statistics 2, 849-879.

King, J.M.P., Titchener-Hooker, N.J., Zhou, Y., 2007. Ranking bioprocess variables using global sensitivity analysis: A case study in centrifugation. Bioprocess and Biosystems Engineering 30, 123-134.

Klepper, O., 1997. Multivariate aspects of model uncertainty analysis: Tools for sensitivity analysis and calibration. Ecological modelling 101, 1-13.

Knopman, D.S., Voss, C.I., 1987. Behavior of sensitivities in the one-dimensional advection-dispersion equation: Implications for parameter estimation and sampling design. Water Resources Research 23, 253-272.

Kontoravdi, C., Asprey, S.P., Pistikopoulos, E.N., Mantalaris, A., 2005. Application of global sensitivity analysis to determine goals for design of experiments: An example study on antibody-producing cell cultures. Biotechnology Progress 21, 1128-1135.

Korkel, S., Kostina, E., Bock, H.G., Schloder, J.P., 2004. Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. Optimization Methods & Software 19, 327-338.

Kou, B., McAuley K.B., Hsu, C.C., Bacon, D.W., Yao, K.Z., 2005. Mathematical model and parameter estimation for gas-phase ethylene homopolymerization with supported metallocene catalyst. Industrial & Engineering Chemistry Researc 44, 2428-2442.

Kuipers, L., Niederreiter, H., 1974. Uniform Distribution of Sequences. Wiley, New York.

Kutalik, Z., Cho, K.H., Wolkenhauer, O., 2004. Optimal sampling time selection for parameter estimation in dynamic pathway modeling. Biosystems 75, 43-55.

Lall, S., Marsden, J.E., Glavaski, S., 2002. A subspace approach to balanced truncation for model reduction of nonlinear control systems. International Journal of Robust and Nonlinear Control 12, 519 -535.

Lejeune, M.A., 2003. Heuristic optimization of experimental designs. European Journal of Operational Research 147, 484-498.

Levadi, V., 1966. Design of input signals for parameter estimation. IEEE Transactions on Automatic Control 11, 205-211.

Lohmann, T., Bock, H.G., Schloeder, J.P., 1992. Numerical methods for parameter estimation and optimal experiment design in chemical reaction systems. Industrial & Engineering Chemistry Research 31, 54-57.

Li, G.Y., Wang, S.W., Rabitz, H., Wang, S.Y., Jaffe, P., 2002. Global uncertainty assessments by high dimensional model representations (HDMR). Chemical Engineering Science, 57, 4445-4460.

Li, R.J., Henson, M.A., Kurtz, M.J., 2004. Selection of model parameters for off-line parameter estimation. IEEE Transactions on Control Systems Technology 12, 402-412.

Lipniacki, T., Paszek, P., Brasier, A.R., Luxon, B., Kimmel, M., 2004. Mathematical model of NF-kappa B regulatory. Journal of Theoretical Biology 228, 195-215.

Liu, G., Swihart, M.T., Neelamegham, S., 2005. Sensitivity, principal component and flux analysis applied to signal transduction: The case of epidermal growth factor mediated signaling. Bioinformatics 21, 1194-1202.

Ljung, L., Glad, T., 1994. On global identifiability for arbitrary model parametrizations. Automatica 30, 265 -276.

Ljung, L., 1999. System Identification: Theory for the User, 2nd edition. Prentice Hall PTR, Upper Saddle River, NJ.

Lund, B.F., Foss, B.A., 2008. Parameter ranking by orthogonalization - Applied to nonlinear mechanistic models. Automatica 44, 278-281.

Machado, V.C., Tapia, G., Gabriel, D., Lafuente, J., Baeza, J.A., 2009. Systematic identifiability study based on the Fisher Information Matrix for reducing the number

of parameters calibration of an activated sludge model. Environmental Modelling & Software 24, 1274-1284.

Marshall, C.J., 1995. Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation. Cell 80, 179-185.

Margaria, G., Riccomagno, E., Chappell, M.J., Wynn, H.P., 2001. Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences. Mathematical Biosciences 174, 1-26.

Marino, S., Hogue, I.B, Ray, C.J. Kirschner, D.E., 2008. A methodology for performing global uncertainty and sensitivity analysis in systems biology. Journal of Theoretical Biology 254, 178-196.

Martinez, E.C., Cristaldi, M.D., Grau, R.J., 2009. Design of dynamic experiments in modeling for optimization of batch processes. Industrial & Engineering Chemistry Research 48, 3453-3465.

Mehra, R., 1974. Optimal input signals for parameter estimation in dynamic systems--Survey and new results. IEEE Transactions on Automatic Control 19, 753-768.

McKay, M.D., 1997. Nonparametric variance-based methods of assessing uncertainty importance. Reliability Engineering & System Safety 57, 267-279.

McKay, M.D., Morrison, J.D., Upton S.C., 1999. Evaluating prediction uncertainty in simulation models. Computer Physics Communications 117, 44-51.

McKay, M.D., Beckman, R.J., Conover, W.J., 2000. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 42, 55-61.

McRae, G.J., Tilden, J.W., Seinfeld, J.H., 1982. Global sensitivity analysis-A computational implementation of the Fourier amplitude sensitivity test (FAST). Computers & Chemical Engineering 6, 15-25.

Mendes, P., Kell, D.B., 1998. Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation. Bioinformatics 14, 869-883.

Michalewicz, Z., 1994. Genetic algorithms + data structures = evolution programs. Springer, Berlin.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33, 161-174.

Moore, B., 1981. Principal component analysis in linear systems: Controllability, observability, and model reduction. IEEE Transactions on Automatic Control 26, 17-32.

Muller, C.H., 1995. Maximin efficient designs for estimating nonlinear aspects in linear-models. Journal of Statistical Planning and Inference 44, 117-132.

Muske, K.R., Georgakis, C., 2003. Optimal measurement system design for chemical processes. AIChE Journal 49, 1488-1494.

Nagy, Z.K., Braatz, R.D., 2003. Robust nonlinear model predictive control of batch processes. AIChE Journal 49, 1776-1786.

Nelles, O., 2001. Nonlinear system identification: from classical approaches to neural networks and fuzzy models. Springer, Berlin.

Niederreiter, H., 1978. Quasi-Monte. Carlo methods and pseudo-random numbers. Bulletin of the American Mathematical Society 84, 957 -1041.

Nikandrov, A., Swartz, C.L.E., 2009. Sensitivity analysis of LP-MPC cascade control systems. Journal of Process Control 19, 16-24.

Oakley, J.E., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: A Bayesian approach. Journal of the Royal Statistical Society. Series B, Statistical Methodology 66, 751-769.

Oniki, H., 1973. Comparative dynamics (sensitivity analysis) in optimal control theory. Journal of Economic Theory 6, 265-283.

Ozyurt, D.B., Barton, P.I., 2005. Large-scale dynamic optimization using the directional second-order adjoint method. Industrial & Engineering Chemistry Research 44, 1804 -1811.

Pazman, A., 1986. Foundations of Optimum Experimental Design. D. Reidel Publishing Company, Dordrecht, Netherlands.

Perumal, T.M., Wu, Y., Gunawan, R., 2009. Dynamical analysis of cellular networks based on the Green's function matrix. Journal of Theoretical Biology, 261, 248-259.

Piazza, M., Feng, X.J., Rabinowitz, J.D., Rabitz, H., 2008. Diverse metabolic model parameters generate similar methionine cycle dynamics. Journal of Theoretical Biology 251, 628-639.

Pohjanpalo, H., 1978. System identifiability based on the power series expansion of the solution. Mathematical Biosciences 41, 21 -33.

Powell, M.J.D., 2002. UOBYQA: Unconstrained optimization by quadratic approximation. Mathematical Programming 92, 555-582.

Pronzato, L., Walter, E., 1985. Robust experiment design via stochastic-approximation. Mathematical Biosciences 75, 103-120.

Pronzato, L., Walter, E., 1988. Robust experiment design via maximin optimization. Mathematical Biosciences 89, 161 -17.

Pronzato, L., 2008. Optimal experimental design and some related control problems. Automatica 44, 303-325.

Rabitz, H., 1981. Chemical sensitivity analysis theory with applications to molecular dynamics and kinetics. Computers & Chemistry 5, 167-180.

Rabitz, H., 1987. Chemical dynamics and kinetics phenomena as revealed by sensitivity analysis techniques. Chemical Reviews 87, 101-112.

Rabitz, H., Alis, O.F., 1999. General foundations of high-dimensional model representations. Journal of Mathematical Chemistry 25, 197-233.

Rabitz, H., Kramer, M., Dacol, D., 1983. Sensitivity Analysis in Chemical Kinetics. Annual Review of Physical Chemistry 34, 419-461.

Rasch, A., Bucker, H.M., Bardow, A., 2009. Software supporting optimal experimental design: A case study of binary diffusion using EFCOSS. Computers & Chemical Engineering 33, 838-849.

Ratto, M., Pagano, A., Young, P., 2007. State dependent parameter metamodelling and sensitivity analysis. Computer Physics Communications 177, 863-876.

Robbins, H., Monro, S., 1951. A stochastic approximation method. Annals of Mathematical statistics 22, 400-407.

Robert, C.P., Casella, G., 2004. Monte Carlo Statistical Methods, 2nd edition. Springer, New York.

Roberts, G.O., Rosenthal, J.S., 2001. Optimal scaling for various Metropolis-Hastings algorithms. Statistical Science 16, 351 -367.

Rodriguez-Fernandez, M., Mendes, P., Banga, J.R., 2006. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. Biosystems 83, 248-265.

Rojas, C.R., Welsh, J.S., Goodwin, G.C., Feuer, A., 2007. Robust optimal experiment design for system identification. Automatica 43, 993-1008.

Rothenbe, T.J., 1971. Identification in parametric models, Econometrica 39, 577-591.

Rubinstein, R.Y., 1989. Sensitivity analysis and performance extrapolation for computer simulation models. Operations Research 37, 72 -81.

Saccomani, M.P., Audoly, S., D'Angio, L., 2003. Parameter identifiability of nonlinear systems: The role of initial conditions. Automatica 39, 619-632.

Saltelli, A., 2002. Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications 145, 280-297.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Physics Communications 181, 259-270.

Saltelli, A., Chan, K., Scott, E.M., 2000. Sensitivity Analysis. Wiley, New York.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global sensitivity analysis: The primer. Wiley, New York.

Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2005. Sensitivity analysis for chemical models. Chemical Reviews 105, 2811-2827.

Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2006. Sensitivity analysis practices: Strategies for model-based inference. Reliability Engineering & System Safety 91, 1109-1125.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice: A guide to Assessing Scientific Models. Wiley, New York.

Saltelli, A., Tarantola, S., Chan, K.P.S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41, 39-56.

Sandink, C.A., McAuley, K.B., McLellan, P.J., 2001. Selection of parameters for updating in on-line models. Industrial & Engineering Chemistry Research 40, 3936 - 3950.

Sandu A., Daescu D.N., Carmichael, G.R., 2003. Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: Part I - theory and software tools. Atmospheric Environment 37, 5083-5096.

Schaibly, J.H., Shuler, K.E., 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II Applications. Journal of Chemical Physics 59, 3879-3888.

Scherpen, J.M.A., 1993. Balancing for nonlinear systems. Systems & Control Letters 21, 143-153.

Schittkowski, K., 2007. Experimental design tools for ordinary and algebraic differential equations. Industrial & Engineering Chemistry Research 46, 9137-9147.

Sidoli, F.R., Mantalaris, A., Asprey, S.P., 2005. Toward global parametric estimability of a large-scale kinetic single-cell model for mammalian cell cultures. Industrial & Engineering Chemistry Research 44, 868-878.

Silvey, S.D., 1980. Optimal Design: An Introduction to the Theory for Parameter Estimation. Chapman and Hall, London.

Sin, G., Vanrolleghem, P.A., 2007. Extensions to modeling aerobic carbon degradation using combined respirometric-titrimetric measurements in view of activated sludge model calibration. Water Research 41, 3345-3358.

Singh, A.K., Hahn, J., 2005. Determining optimal sensor locations for state and parameter estimation for stable nonlinear systems. Industrial & Engineering Chemistry Research 44, 5645-5659.

Singh, A.K., Hahn J., 2006. Sensor location for stable nonlinear dynamic systems: Multiple sensor case. Industrial & Engineering Chemistry Research 45, 3615-3623.

Singh, A., Jayaraman, A., Hahn J., 2006. Modeling regulatory mechanisms in IL-6 signal transduction in hepatocytes. Biotechnology and Bioengineering, 95, 850-862.

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation 55, 271-280.

Sokolowski J., 1987. Sensitivity analysis of control constrained optimal control problems for distributed parameter systems. SIAM Journal on Control and Optimization 25, 1542 -1556.

Sorensen, J.V.T., Madsen, H., Madsen, H., 2006. Parameter sensitivity of three Kalman filter schemes for assimilation of water levels in shelf sea models. Ocean modelling 11, 441-463.

Soroush, M., Kravaris, C., 1992. Discrete-time nonlinear controller synthesis by input/output linearization. AIChE Journal 38, 1923-1945.

Spall, J.C., 1992. Multivariate stochastic-approximation using a simultaneous perturbation gradient approximation. IEEE Transactions on Automatic Control 37, 332-341.

Spall, J.C., 1998. Implementation of the simultaneous perturbation algorithm for stochastic optimization. IEEE Transactions on Aerospace and Electronic Systems 34, 817-823.

Spall, J.C., 2005. Monte Carlo computation of the Fisher information matrix in nonstandard settings. Journal of Computational and Graphical Statistics 14, 889-909.

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation 55, 271-280.

Stanimirovic, O., Hoefsloot, H.C.J., De Bokx, P.K., Smilde, A.K., 2008. Variable selection methods as a tool to find sensor locations for distributed parameter systems. Industrial & Engineering Chemistry Research 47, 1184-1191.

Steinberg, D.M., Hunter, W.G., 1984. Experimental design: Review and comment. Technometrics 26, 71 -97.

Stewart, G.W., 1980. The efficient generation of random orthogonal matrices with an application to condition estimators. SIAM Journal on Numerical Analysis 17, 403 - 409.

St. John, R.C., Draper, N.R., 1975. D-optimality for regression designs: A review. Technometrics 17, 15 -23.

Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering & System Safety 93, 964-979.

Sun, N., Sun, N.Z., Elimelech, M., Ryan, J.N., 2001. Sensitivity analysis and parameter identifiability for colloid transport in geochemically heterogeneous porous media. Water Resources Research 37, 209-222.

Sun C.L., Hahn, J., 2006. Parameter reduction for stable dynamical systems based on Hankel singular values and sensitivity analysis. Chemical Engineering Science, 61, 5393-5403.

Theodoridis, S., Koutroumbas, K., 2006. Pattern Recognition, 3rd edition. Academic Press, Boston.

Tierney, L., 1994. Markov Chains for Exploring Posterior Distributions. Annals of Statistics 22, 1701 -1728.

Turanyi, T., 1990. Sensitivity analysis of complex kinetic systems—Tools and applications. Journal of Mathematical Chemistry 5, 203-248.

Vajda, S., Godfrey, K.R., Rabitz, H., 1989. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. Mathematical Biosciences 93, 217 -248.

Vajda, S., Rabitz, H., 1994. Identifiability and distinguishability of general reaction systems. Journal of Physical Chemistry 98, 5265-5271.

Vajda, S., Valko, P., Turanyi, T., 1985. Principal component analysis of kinetic models. International Journal of Chemical Kinetics 17, 55 -81.

Vanderbei, R.J., Shanno, D.F., 1999. An interior-point algorithm for nonconvex nonlinear programming. Computational Optimization and Applications 13, 231-252.

van de Vusse, J.G., 1964. Plug-flow type reactor versus tank reactor. Chemical Engineering Science 19, 994-997.

Vanrolleghem, P.A., Van Daele, M., Dochain, D., 1995. Practical identifiability of a biokinetic model of activated sludge respiration. Water Research 29, 2561-2570.

Varella, H., Guerif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. Environmental Modelling & Software 25, 310-319.

Velez-Reyes, M., Verghese, G.C., 1995. Subset selection in identification, and application to speed and parameter estimation for induction machines. Proceedings of the 4th IEEE Conference on Control Applications, Albany, NY, pp. 991-997.

Wachter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming 106 1, 25-57.

Wagner, H.M., 1995. Global sensitivity analysis. Operations Research 43, 948-969.

Wallace, S.W., 2000. Decision making under uncertainty: Is sensitivity analysis of any use? Operations Research 48, 20-25.

Walter, E., 1987. Identifiability of Parametric Models. Pergamon Press, New York.

Walter, E., Pronzato, L., 1990. Qualitative and quantitative experiment design for phenomenological models - A survey. Automatica 26, 195-213.

Walter, E., Pronzato, L., 1997. Identification of Parametric Models from Experimental Data. Springer, Berlin.

Weijers, S.R., Vanrolleghem, P.A., 1997. A procedure for selecting best identifiable parameters in calibrating activated sludge model no.1 to full-scale plant data. Water Science and Technology 36, 69 -79.

Weyl, H., 1938. Mean motion. American Journal of Mathematics 60, 889 -896.

Whittle, P., 1973. Some general points in the theory of optimal experimental design. Journal of the Royal Statistical Society. Series B, Statistical Methodology 35, 123 - 130.

Wilkinson, S.J., Benson, N., Kell, D.B., 2008. Proximate parameter tuning for biochemical networks with uncertain kinetic parameters. Molecular Biosystems 4, 74-97.

Wynn, H.P., 1970. The sequential generation of D-optimum experimental designs. Annals of Mathematical Statistics 41, 1655-1664.

Wynn, H.P., 1972. Results in the theory and construction of D-optimum experimental designs. Journal of the Royal Statistical Society. Series B, Statistical Methodology 34, 133-147.

Yamada, S., Shiono, S., Joo, A., Yoshimura, A., 2003. Control mechanism of JAK/STAT signal transduction pathway. FEBS letters 534, 190-196.

Yao, K.Z., Shaw, B.M., Kou, B., McAuley, K.B., Bacon, D.W., 2003. Modeling ethylene/butene copolymerization with multi-site catalysts: Parameter estimability and experimental design, Polymer Reaction Engineering 11, 563-588.

Yeh, W.W.G., 1986. Review of parameter identification procedures in groundwater hydrology : The inverse problem. Water Resources Research 22, 95 -108.

Yue, H., Brown, M., He, F., Jia, J.F., Kell, D.B., 2008. Sensitivity Analysis and Robust Experimental Design of a Signal Transduction Pathway System. International Journal of Chemical Kinetics 40, 730-741.

Yue, H., Brown, M., Knowles, J., Wang, H., Broomhead D.S., Kell D.B., 2006. Insights into the behaviour of systems biology models from dynamic sensitivty and identifiability analysis: A case study of an NF-kappa B signalling pathway. Molecular Biosystems 2, 640-649.

Zador, J., Zsely, I.G., Turanyi, T., Ratto, M., Tarantola, S., Saltelli, A., 2005. Local and global uncertainty analyses of a methane flame model. Journal of Physical Chemistry 109, 9795-9807.

Zamprogna, E., Barolo, M., Seborg, D.E., 2005. Optimal selection of soft sensor inputs for batch distillation columns using principal component analysis. Journal of Process Control 15, 39 -52.

Zhang, H.X., Dempsey, W.P., Goutsias, J., 2009. Probabilistic sensitivity analysis of biochemical reaction systems. Journal of Chemical Physics 131, no. 094101.

Zi, Z.K., Cho, K.H., Sung, M.H., Xia, X.F., Zheng, J.S., Sun, Z.R., 2005. In silico identification of the key components and steps in INF-γ induced JAK-STAT signaling pathway. FEBS Letters 579, 1101-1108.

Ziehn, T., Tomlin, A.S., 2009. GUI-HDMR - A software tool for global sensitivity analysis of complex models. Environmental Modelling & Software 24, 775-785.

VITA

Yunfei Chu received his Bachelor of Science degree in automation from Tsinghua University, Beijing, China, in 2002. He received his Master of Science degree in control theory and control engineering from the same university in 2005. He entered the Chemical Engineering program at Texas A&M University in 2005 and received his Doctor of Philosophy degree in 2010. His research interests include parameter estimation, sensitivity analysis, experimental design, state filtering, model reduction, reaction networks, process optimization and control. He published 10+ papers in leading journals, e.g. AIChE Journal, Chemical Engineering Science, Industrial & Engineering Chemistry Research.

Dr. Chu may be reached c/o Dr. Juergen Hahn at Artie McFerrin Department of Chemical Engineering, Texas A&M University, 3122 TAMU, College Station, TX 77843. His email is YFoChu@gmail.com.