ADVANCES AND SELECTED RECENT DEVELOPMENTS IN STATE AND PARAMETER ESTIMATION

Costas Kravaris*
Department of Chemical Engineering
University of Patras
Karatheodori 1, University Campus
GR 265 00 Patras
Greece

Juergen Hahn and Yunfei Chu
Artie McFerrin Department of Chemical Engineering
Texas A&M University
College Station, TX 77843-3122
U.S.A.

Abstract

This paper deals with two topics from state and parameter estimation. The first contribution of this work provides an overview of techniques used for determining which parameters of a model should be estimated. This is a question that commonly arises when fundamental models are used as these models often contain more parameters than can be reliably estimated from data. The decision of which parameters to estimate is independent of the observer/estimator design, however, it is directly affected by the structure of the model as well as the available data. The second contribution is an overview of recent developments regarding the design of nonlinear Luenberger observers, with special emphasis on exact error linearization techniques, but also discussing more general issues, including observer discretization, sampled data observers and the use of delayed measurements.

Keywords

State estimation, parameter estimation, observers, nonlinear system

Introduction

The use of fundamental models for process monitoring and control has become increasingly popular in recent years. However, the performance of a particular application does not only depend upon the algorithms used but also upon the quality of the model. This realization has led to several new research directions over the last few decades, two of which are reviewed in this work. One of these research areas focuses on the use of nonlinear models, and procedures required for dealing with these nonlinear models, to more appropriately describe a nonlinear system.

One type of approach for improving model accuracy, regardless if these are linear or nonlinear models, is to estimate model parameters from data. While there has been a significant interest in algorithms used for parameter estimation, the questions of how many and which parameters should be estimated have only been addressed more recently. The first part of this paper provides an overview of existing methods for selecting parameters for estimation.

The second part of this paper reviews theory and algorithms of nonlinear Luenberger observers for state and parameter estimation, focusing on recent methods and results from nonlinear systems theory. In a sense, this work complements a review paper presented at the previous CPC on particle filters and moving horizon estimators (Rawlings & Bakshi, 2006).

Regularization techniques for parameter estimation of complex dynamic models

This section focuses on parameter estimation which plays an important role in process monitoring as well as mathematical modeling. Despite significant advances over the last few decades, this topic is still an active area of research and several review articles dealing with parts of this problem have been published in the last decade (Ashyraliyev et al., 2009; Chou & Voit, 2009; Dochain, 2003; Esposito & Floudas, 2000; Jimenez-Hornero, Santos-Duenas & Garcia-Garcia, 2009; Maria, 2004; McLean & McAuley, 2011; Moles, Mendes & Banga, 2003). Oftentimes, parameter estimation deals with the algorithms used for performing the estimation; in fact the second contribution of this work focuses on methodologies for estimating states and parameters. However, it is equally important to decide which parameters of a model should be estimated, why a particular subset of the parameters should be estimated, and also how accurate the estimation results will be based upon available data. This section provides a review of existing techniques that can answer these questions. One of the motivating factors behind these techniques is that complex systems, e.g., chemical reaction networks, can contain dozens to hundreds of parameters (Schoeberl et al., 2002), however, it is often not possible to estimate more than a handful of these. In these cases, the accuracy of the estimates, and the model predictions resulting from these estimates, are strongly affected by the parameters chosen for estimation.

One challenge arising from estimation of complex systems is that the estimation problem is ill-conditioned. The reason for this is that a complex model contains a large number of parameters but not all of them are identifiable even if an unlimited amount of noise-free data would be available. Accordingly, the effects that changes in the parameters have on the outputs are correlated and the solution to the estimation problem is not unique. Furthermore, experimental data inevitably contain noise and the amount of available data is often limited. These limitations regarding the availability and quality of the data pose further challenges to the estimation problem since the optimal solution of the parameter values can be sensitive to variations in the data (Gutenkunst et al., 2007). Furthermore, similarly to what is widely-known in system identification, parameters that best fit the training data are not necessarily the best ones from a practical point of view (Slezak et al., 2010). Therefore, estimation of a complex system does not merely deal with determining the optimal solution to the data fitting problem but instead needs to focus on computing a solution which is robust to variations in the experimental data.

A second challenge that arises from estimation of a complex system is associated with the computational burden. Since a closed-form solution of the differential equations which describe a model is generally not available, it is only possible to evaluate the model via simulations. Since parameter estimation deals with solution of an optimization problem, the model needs to be evaluated repeatedly which can quickly result in estimation problems that are computationally prohibitive even for medium-scale problems.

A large number of techniques have been presented in the literature to address these problems in one way or another. A brief review of these techniques is provided in this section. The review is not meant to be comprehensive as this research area spans many different subtopics and is also an active area of research in many different fields of engineering. Instead, the work presented in this section focuses on techniques used for selecting a set of parameters for estimation. The reason for focusing on this area is that no review of existing techniques has previously been published in this field and that this approach aids regularizing ill-conditioned estimation problems.

This section is organized as follows. The formulation of the estimation problem for dynamic systems is presented next. After that, a general framework for regularization is proposed and three commonly used regularization techniques are compared. The following section focuses on the parameter selection procedure, which is one of the regularization approaches. As parameter selection encompasses a variety of methods, only the popular orthogonalization method is investigated further in the following section. This section concludes by presenting some suggestions for possible future research in this field.

Model formulation for estimation of dynamic systems

Parameter estimation aims to infer parameter values from available data so that the model predictions can accurately reflect the data (van den Bos, 2007). To estimate parameters of a dynamic system, a regression model is formulated that involves the differential equations. This formulation is presented in this subsection.

A time-invariant dynamic system is described by a set of ordinary differential equations as

$$\begin{cases} \dot{x}(t) = f(x(t), u(t), \theta) \\ y(t) = g(x(t), u(t), \theta) \end{cases} \tag{1}$$

where x is the state vector, u is the input vector, y is the output vector, and θ is the parameter vector.

The first step in parameter estimation is to derive an expression representing the parameter-output relationship. For the dynamic model shown in equation (1), the output **y** can be evaluated by model simulations, assuming that the initial state $x(0)$, the input profile $u(t)$, and the parameter vector **θ** are available. The resulting parameter-output relationship is denoted by $y(t,\theta)$ which is time-dependent, and generally lacks a closed-form solution expression.

The next step is to formulate the regression model by discretizing the output profiles and by including noise information. Given a set of time points $\{t_1, t_2, \cdots, t_m\}$, the output is sampled as

$$h(\theta) = \left[y_1(t_1,\theta), \cdots, y_1(t_m,\theta), \cdots, y_n(t_1,\theta), \cdots, y_n(t_m,\theta) \right]^T \tag{2}$$

where y_1, y_2, …, y_n are entries in the output vector y. After sampling, the continuous output profiles are discretized and the discretized output vector is only a function of the

parameters, denoted by $h(\theta)$. Since all measurements contain some level of noise, the data available for estimation are given by

$$\tilde{y}=h(\theta)+\varepsilon \quad (3)$$

where $\tilde{y}$ is the data vector, $h(\theta)$ represents the model prediction, and $\boldsymbol{\varepsilon}$ denotes the noise vector. Apart from the model structure, information about the noise distribution plays an important role in parameter estimation. This distribution determines both formulation of the optimization problem for parameter estimation and the statistics of the estimated parameter values. In practice, a description of the noise is usually not accurately known and it is often assumed to be Gaussian, denoted by $\varepsilon \sim N(0,\sigma^2 I)$ where the mean vector is 0 and the covariance matrix is $\sigma^2 I$.

The third step in parameter estimation is to formulate an optimization problem which computes the parameter estimates as the optimal solution to the problem. Maximum likelihood estimation is commonly used, which estimates parameters by maximizing the likelihood function. In the case of Gaussian noise, maximum likelihood estimation reduces to least squares estimation, which computes the parameter estimates by minimizing the difference between the model prediction and the measured data as

$$\hat{\theta}=\arg\min_{\theta}\left(\tilde{y}-h(\theta)\right)^{T}\left(\tilde{y}-h(\theta)\right) \quad (4)$$

where the difference is measured by the squared Euclidean norm.

It should be noted that estimation of a complex system is usually an ill-conditioned problem, i.e., the optimal least squares solution may be very sensitive to variations in the data. One approach to deal with this problem is to perform regularization to avoid ill-conditioning.

Regularization of ill-conditioned parameter estimation problems

When the estimation problem is ill-conditioned, a variety of regularization techniques can be applied to compute a robust solution. However, no detailed review of these regularization techniques exists in the literature and there is lack of a unified framework as part of which these regularization techniques can be viewed. Formulating such a general framework for regularization can provide insights into the commonalities but also into the differences that exist between the different techniques and, ultimately, help select an appropriate method. An attempt to present such a unified framework is made in this subsection where the framework combines three commonly used regularization techniques. Since regularization techniques for a nonlinear system are frequently extensions of those for a linear system, the linear case is investigated first.

The linear version of the regression model given by equation (3) is assumed be given by

$$\tilde{y}=H\theta+\varepsilon \quad (5)$$

The design matrix H is assumed to have a full column rank. If the design matrix is rank-deficient then the columns which are linearly dependent on others can be eliminated as well as the associated parameters. In this case, the reduced model will generate identical predictions to the original model and can be used for estimation. For a linear model, a closed form of the least squares solution can be computed and is given by

$$\hat{\theta}=\left(H^{T}H\right)^{-1}H^{T}\tilde{y} \quad (6)$$

Though widely used, least squares estimation will encounter difficulties when the design matrix is close to a rank-deficient matrix. In this case, the estimation problem is ill-conditioned and the optimal solution is extremely sensitive to the noise contained in the data. To illustrate this point, the covariance matrix of the parameter estimate is computed as

$$\mathrm{Var}\left[\hat{\theta}\right]=\sigma^2\left(H^{T}H\right)^{-1}=\sigma^2 V\Lambda^{-1}V^{T} \quad (7)$$

where the eigenvalue decomposition of the cross product matrix is applied:

$$H^{T}H=V\Lambda V^{T} \quad (8)$$

The matrix Λ is diagonal and diagonal entries are eigenvalues of the cross product matrix $H^{T}H$ or the squared singular values of the design matrix H. The fact that the design matrix H is close to rank deficiency implies that some diagonal entries in Λ are close to zero. Consequently some diagonal entries in Λ^{-1}, which are reciprocals of these small entries in Λ, are very large. A result of this is that a small amount of noise (denoted by a small σ^2) can cause a large variance in the parameter estimates.

A regularization technique can be included in the least squares problem to reduce the effect of noise on the parameter estimates. Three commonly used regularization methods are investigated here, i.e., ridge regression (Box & Draper, 2007), principal component analysis (Jolliffe, 2002), and parameter selection (Miller, 2002).

In ridge regression, the term λI is added to the least squares solution shown in equation (6) and the resulting parameter estimate is given by

$$\hat{\theta}_{RR}=\left(H^{T}H+\lambda I\right)^{-1}H^{T}\tilde{y} \quad (9)$$

where λ is a nonnegative tuning parameter. The matrix $H^{T}H+\lambda I$ can be made to be far from being rank deficient if λ is set to be sufficiently large. If $\lambda = 0$ then the solution reduces to the ordinary least squares estimate. The well-known Levenberg–Marquardt algorithm commonly found in numerical analysis also uses a ridge regression technique (Marquardt, 1963).

Principal component analysis (or truncated singular value decomposition) and parameter selection deal with an ill-conditioned problem from a model-reduction perspective. When the design matrix H is close to being rank deficient, then this can be viewed as the problem where only some part of the procedure results in an ill-conditioned problem, while other parts may be easy to solve. The solution to this problem is to identify and, in a subsequent step, truncate the parts of the problem which result in an ill-conditioned parameter estimation problem. The remaining parameter estimation problem should be smaller in size and easier to solve.

One model reduction approach involves orthogonal projections. A projection matrix, denoted by P, has more rows than columns and satisfies

$$P^T P = I \tag{10}$$

The projected parameter vector is given by

$$\theta_P = P^T \theta \tag{11}$$

The original model given by equation (5) then reduces to

$$\tilde{y} = HP\theta_P + \varepsilon \tag{12}$$

where the least squares estimate of the reduced model is

$$\hat{\theta}_P = \left(P^T H^T H P\right)^{-1} P^T H^T \tilde{y} \tag{13}$$

The key to implement a regularization technique via such a model reduction approach is to determine an appropriate projection matrix.

Principal component analysis computes the projection matrix P by the eigenvalue decomposition shown in equation (8). The diagonal matrix Λ is partitioned into two sub-matrices by ordering the diagonal entries from the largest one to the smallest one

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix} \tag{14}$$

where Λ_1 contains the diagonal entries which are large and Λ_2 only has small diagonal entries. Accordingly, the matrix V in the eigenvalue decomposition can be partitioned as

$$V = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \tag{15}$$

Principal component analysis uses the projection matrix given by

$$P = V_1 \tag{16}$$

Parameter selection makes use of a different approach that can be described within this same framework. Parameter selection places an additional constraint on the projection matrix, i.e., the reduced parameters by the projection are a subset of the original set of parameters. Suppose a subset of parameters is selected and denoted as

$$\theta_S = \begin{bmatrix} \theta_{i_1}, \theta_{i_2}, \cdots, \theta_{i_s} \end{bmatrix}^T \tag{17}$$

where i_1, i_2, …, i_s are indices of the reduced parameters. Then the projection matrix is given by

$$P = \begin{bmatrix} e_{i_1}, e_{i_2}, \cdots, e_{i_s} \end{bmatrix} \tag{18}$$

where e_i denotes the i-th column of the identity matrix.

Even though the three regularization techniques have been separately derived, they can be viewed as different approaches within the same framework. They all share the same basic idea which is that constraints are added to reduce the feasible region of the estimation problem.

Formally, the regularized parameter estimation problem can be expressed as

$$\begin{aligned} &\hat{\theta}_R = \arg\min_{\theta} \left(\tilde{y} - H\theta\right)^T \left(\tilde{y} - H\theta\right) \\ &\text{s.t. } C(\theta) = 0 \end{aligned} \tag{19}$$

where $\hat{\theta}_R$ denotes the parameter estimate with regularization. The different technique vary in how the constraint function $C(\theta)$ is selected. To illustrate the constraints geometrically, a case of two parameters is considered where the parameter space is given by a plane. Fig. 1 displays the constraints as well as the corresponding feasible regions for the three regularization techniques.

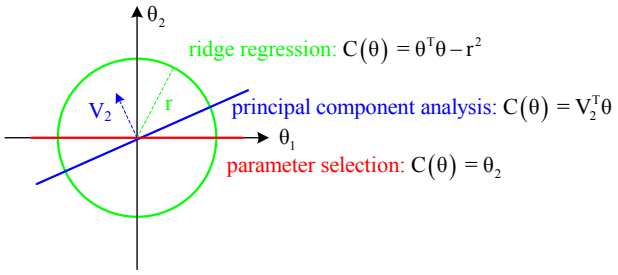

Figure 1. Feasible range and constraint for each regularization method.

The feasible range of ridge regression is a circle with radius equal to r. The value of r is dependent on the value of λ in the ridge regression solution shown in equation (9). For a general multi-dimensional parameter space, the feasible range of ridge regression is given by a hyper-sphere.

The feasible range of principal component analysis is a line, the direction of which is denoted by the first principal component. This line is perpendicular to the second principal component, here shown by the vector V_2. In a general case, the feasible range is a linear subspace spanned by the columns in V_1 in the expression shown in equation (15).

Similarly to principal component analysis, the feasible range of parameter selection is also given by a line. However, the linear subspace resulting from parameter selection is spanned by a subset of the original parameter axes while the subspace returned by principal component analysis can be arbitrary and combined by all original parameters. In this case, the feasible range of parameter selection is the axis of the selected parameter θ_1 which is equal to setting $\theta_2 = 0$.

Another regularization approach is to add a penalty term to the objective function, such as

$$\hat{\theta}_R = \arg\min_{\theta} \left(\tilde{y} - H\theta\right)^T \left(\tilde{y} - H\theta\right) + \gamma \left(L\theta\right)^T \left(L\theta\right) \tag{20}$$

where the coefficient $\gamma \geq 0$ and the matrix L are tuning parameters. If $\gamma = 0$ then the problem reduces to the ordinary least squares estimation. The expression shown in equation (20) can be regarded as the Lagrange form of the constrained optimization problem shown in equation (19). By properly selecting the tuning parameters shown in Table 1, the solution of the problem, including a penalty function from equation (20), is identical to the constrained problem shown in equation (19).

Table 1.
Selection of tuning parameters for the three regularization methods

Method	γ	L
Ridge regression	λ	I
Principal component analysis	∞	V_2^T
Parameter selection	∞	I_S^T

For ridge regression, the coefficient in the penalty term is chosen as $\gamma = \lambda$ and the matrix L is set to the

identity matrix. For principal component analysis, the matrix L is chosen as the transpose of V_2 of the eigenvalue decomposition shown in equation (15). $L\theta$ is the projection of the parameter vector θ onto the subspace spanned by V_2, which corresponds to the small eigenvalues of the cross product matrix from equation (8). The coefficient γ is set to a large number to move the parameter values away from the linear space spanned by V_2. A similar approach is applied for parameter selection where $\bar{S}$ denotes the indices of the unselected parameters and the matrix $I_{\bar{S}}$ consist of columns of the identity matrix corresponding to the unselected parameters. The coefficient γ is set to a large number so that the unselected parameters approach zero.

It should be noted that even though the general expression shown in equation (19) or (20) looks simple, it is non-trivial to determine the constraints or the tuning parameters for a regularization technique even for a linear model.

Many applications of ridge regression can be found in the literature; for some examples see (Ancheyta, Sanchez & Rodriguez, 2005; Ashyraliyev, Jaeger & Blom, 2008; Katare et al., 2004; Sandelinet al., 2006; Tobajas et al., 2007). Similarly, several applications of principal component analysis are found in (Bindlish, Rawlings & Young, 2003; Degenring et al., 2004). The focus of the following section is placed on parameter selection. Parameter selection has recently attracted increasing attention due to the following features:

- **Interpretability**. Parameter selection is not merely a mathematical solution to ill-conditioned estimation problems. It also provides a powerful tool for model analysis, e.g. the results indicate which parameters are important for a model's dynamic behavior and they also provide insight into the correlation of the effect that different parameters have on the outputs.
- **Simplification**. Parameter selection reduces the number of decision variables in the least squares problem since the unselected parameters are all fixed at a given value. This feature is very useful for estimation of complex models where the computational effort associated with the optimization might be a problem. This is specifically the case for online optimization-based control (e.g. MPC) or filtering (e.g. MHE).

Compared with parameter selection, ridge regression is a purely mathematical regularization algorithm which provides little information with regard to interpretability and does not reduce the optimization problem. Principal component analysis reduces the computational effort, however, the reduced parameters are linear combinations of all of the original parameters. Since the original parameters can even have different units, linear combinations of these parameters are not always trivial to interpret with regard to their physical meaning.

Regularization by parameter selection

Procedures for selecting a subset of parameters for estimation from all the parameters of a model fall into one of the following two categories: heuristic methods and optimization-based methods. A review of commonly used parameter selection methods is presented in this subsection.

Heuristics for parameter selection are derived based on the effects that variations in the parameters have on the model outputs. For example, if a variation in a parameter value has only a marginal effect on the model outputs then it is going to be difficult to estimate this parameter as even small amounts of noise in the measurements will have a significant impact on the estimated value. Similarly, if the effect that a variation of a parameter has on the outputs is correlated to the effect that variations of other parameters have, then it is not possible to uniquely determine the values of these parameters as the parameter estimation problem becomes ill-conditioned. However, in each of these two cases, it is possible to eliminate this parameter from consideration for estimation as this parameter does not result in a unique effect on the outputs that is required for a good fit. Since estimation of all parameters is neither desired nor necessary, estimation of a subset of parameters can both regularize an ill-conditioned estimation problem and simplify the associated optimization problem.

To quantitatively investigate parameter effects, local sensitivity analysis is often applied. Local sensitivity is defined as the partial derivative of the output with respect to the parameter:

$$s_i = \frac{\partial h(\theta)}{\partial \theta_i} \tag{21}$$

where s_i denotes the sensitivity vector of the parameter θ_i. If the model is linear as shown in equation (5) then the sensitivity vector of a parameter reduces to the column of the design matrix H corresponding to this parameter. For a nonlinear model of the form shown in equation (3), generated by the dynamic system from equation (1), the sensitivity profile of $\partial y/\partial \theta_i$ can be computed by solving the sensitivity equations

$$\begin{aligned} \frac{d}{dt}\frac{\partial x}{\partial \theta_i} &= \frac{\partial f}{\partial x^T}\frac{\partial x}{\partial \theta_i} + \frac{\partial f}{\partial \theta_i} \\ \frac{\partial y}{\partial \theta_i} &= \frac{\partial g}{\partial x^T}\frac{\partial x}{\partial \theta_i} + \frac{\partial g}{\partial \theta_i} \end{aligned} \tag{22}$$

together with the model equations. The sensitivity vector $\mathbf{s}_i$ is constructed by sampling $\partial y(t)/\partial \theta_i$ at the same sampling points which generate the output data from equation (2).

Based on the sensitivity analysis, the criteria for parameter selection are:

- **Magnitude of the effect**: The norm of the sensitivity vector of a selected parameter should be of significant value compared to the norms of the sensitivity vectors for other parameters.
- **Uncorrelated effect**: The sensitivity vectors of selected parameters cannot be highly correlated.

Some heuristic methods using these two rules for parameter selection are shown in Table 2. The orthogonalization method (Yao et al., 2003; Lund & Foss, 2008) forms a special case as it involves a simple procedure and the results are easy to interpret. This method

will be discussed separately as it is the most popular parameter selection technique.

Though simple to use, heuristic-based methods often use an ad hoc approach to balance the two criteria, i.e., magnitude of the effects and correlation among the effects, and the performance of the resulting estimation problem can often not be guaranteed. A more systematic way for parameter selection is to formulate an optimization problem. The sensitivity matrix which consists of the sensitivity vector as shown in equation (21) for each parameter has a direct relationship with the Fisher information matrix (Walter & Pronzato, 1997):

$$F = \frac{1}{\sigma^2} S^T S \tag{23}$$

where F denotes the Fisher information matrix and $S=\left[s_1, s_2, \cdots, s_p\right]$ is the sensitivity matrix. The scaling factor $1/\sigma^2$ has no effect on the following selection procedure and it can be assumed to be equal to unity without loss of generality. The inverse of the Fisher information matrix provides a lower bound for the covariance matrix of the parameter estimate. Therefore, the sensitivity matrix is directly related to the estimation accuracy and parameters can be selected to improve the estimation accuracy.

An experimental criterion function can be applied to the sensitivity matrix to quantitatively evaluate the estimation accuracy. The experimental criteria are named alphabetically and the most important criteria are:

$$\begin{aligned}
&\text{A-criterion: } \varphi_A\left(S^T S\right) = -\text{trace}\left(\left(S^T S\right)^{-1}\right) \\
&\qquad = -\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \cdots + \frac{1}{\lambda_p}\right) \\
&\text{D-criterion: } \varphi_D\left(S^T S\right) = -\det\left(S^T S\right) \\
&\qquad = \lambda_1 \lambda_2 \cdots \lambda_p \\
&\text{E-criterion: } \varphi_E\left(S^T S\right) = \lambda_{\min}\left(S^T S\right) \\
&\qquad = \min\left\{\lambda_1, \lambda_2, \cdots, \lambda_p\right\}
\end{aligned} \tag{24}$$

where the criterion φ is a matrix function and $\lambda_1, \lambda_2, \ldots, \lambda_p$ are eigenvalues of $S^T S$. Each criterion is a function of the eigenvalues. Since the eigenvalues of $S^T S$ are the squared singular values of S, the two criteria for parameter selection, i.e., magnitude of the effects and correlation among the effects, can be quantitatively evaluated and simultaneously taken into account by using an experimental criterion. For example, it is easy to check if a sensitivity vector in S has a small norm or if a submatrix of S has vectors that are linearly dependent, resulting in a small criterion value. A good subset of parameters for estimation can be selected by maximizing an experimental criterion. Among the criteria, the *D*-criterion is the most commonly used one, as it characterizes the volume of the confidence region of the parameter estimates (Walter & Pronzato, 1990).

Using any of the experimental criteria, the optimization problem for parameter selection can be formulated as

$$\begin{aligned}
&z^* = \arg\max_z \ \varphi\left(S_L^T S_L\right) \\
\text{s.t.} \quad &S_L = SL \\
&S = \left[s_1 \quad s_2 \quad \cdots \quad s_p\right] \\
&L = \left[e_{i_1} \quad e_{i_2} \quad \cdots \quad e_{i_s}\right], \text{ with } i_j \text{ that } z_{i_j} = 1 \\
&e_i \text{ is the i-th column of the identity matrix} \\
&z_i \in \{0,1\} \text{ and } \sum_{i=1}^{p} z_i = s
\end{aligned} \tag{25}$$

This is a mixed integer nonlinear programming problem. A binary decision vector z is introduced to denote which parameters are selected. If the entry $z_i = 1$ then the parameter θ_i is selected. Otherwise, the parameter is not selected. The number of selected parameters is set to s, which is equal to the numerical rank of the sensitivity matrix S. The matrix S_L is the sensitive matrix of the selected parameters and it is computed from the sensitivity matrix of all parameters S multiplied by the selection matrix L. The selection matrix consists of columns from the identity matrix corresponding to the binary vector z.

The main challenge of optimization-based methods is that the combinatorial problem is non-trivial to solve. The total number of combinations for selecting s parameters from a set of p parameters is $p!/s!(p-s)!$. The optimal solution may be difficult or even impossible to find if a significant number of parameters need to be selected from a large set of possible parameters. Instead, alternative formulations for equation (25) are often used to determine a sub-optimal solution with a reduced computational burden. These techniques include the sequential selection procedures, stochastic search techniques, and heuristic reduction approaches.

Sequential methods like forward selection are popular due to their simplicity (Blanchet, Legendre & Borcard, 2008). Forward selection decomposes the multi-dimensional search for the combinatorial problem to a sequence of one-dimensional searches. Only one parameter is selected at each step. The selected parameter is determined so that the objective function is maximized when it is added to the set of previously selected parameters.

Stochastic methods like genetic algorithm are another alternative for determining a sub-optimal solution for equation (25) with a manageable computational cost (Chu & Hahn, 2007). One property of a genetic algorithm is that it is based upon a population of potential solutions and might return several viable candidates of parameter sets to be selected. This property can be very useful as it is usually not possible to determine that one set of parameters is superior to all other ones for estimation. If a collection of potential parameters sets is computed then experience with the process can be used to make a final selection of which of these parameter sets to estimate.

Heuristic reductions such as a parameter clustering method (Chu & Hahn, 2009) can also be used for determining a sub-optimal set of parameters for estimation. Parameter selection via parameter clustering considerably simplifies the combinatorial problem by reducing the feasible range of the binary decision variables. Parameters

which have similar effects on the model outputs can be clustered into groups by a hierarchical clustering algorithm. Since parameters in a group are not distinguishable from each other, only one parameter needs to be estimated per group. The set of parameters for estimation is determined by selecting the representative parameters for each group. Since the number of groups is significantly smaller than the number of parameters, the optimization problem given by equation (25) can be solved for this subset of parameters. Alternatively, it is often possible to even perform an exhaustive search as the number of possible combinations is reasonably small. Numerical experiments have shown that clustering methods can find a good solution which in many cases is even identical to the optimal solution (Chu & Hahn, 2009).

Table 2.
Methods for parameter selection.

Method	Reference(s)	Feature
Orthogonalization	Lund & Foss, 2008 Yao et al., 2003	Orthogonal projection is applied to extract the effect of a parameter covered by previously selected parameters and the magnitude of the not yet covered effect determines which parameter is selected at each step.
Collinearity index	Brun, Reichert & Kunsch, 2001	Collinearity indices, based on the smallest eigenvalue or the condition number of the sensitivity matrix, are applied to select parameters.
Relative gain array	Sandink, McAuley & McLellan, 2001	Parameters are regarded as inputs and methods for determining proper input/output pairings in multivariate control are applied to select parameters.
Hankel singular value	Sun & Hahn, 2006	Parameters are regarded as inputs and balancing is used to rank parameters by their importance to the effect that they have on the outputs.
Hybrid	Li, Henson & Kurtz, 2004	PCA is used to determine the initial selection and subsequent parameters are selected based on the distance to the previously selected parameters.
D-optimal	Brun et al., 2002 Chu & Hahn, 2007	Parameters are selected by optimizing the *D*-optimality criterion.
Combined *D*- and modified *E*- crietion	Machado et al., 2009 Weijers & Vanrolleghem, 1997	Parameters are selected by optimizing the combined *D* and modified *E* criterion.

Parameter selection via orthogonal projection

Among the various methods for parameter selection, the orthogonalization method is the most widely used one. This technique is simple to implement since the procedure is directly related to the well-known Gram-Schmidt orthogonalization method. It is a sequential selection procedure and the criterion used to select a parameter at each step can provide insights into the relationship among parameters. The technique was originally presented as a heuristic method (Yao et al., 2003) and later identified as a forward selection which maximizes the *D*-criterion (Chu & Hahn, 2007). However, the idea to perform parameter selection via an orthogonalization procedure can be traced back much earlier for linear rank-degenerate least squares problems (Golub, Klema & Stewart, 1976; Golub & van Loan, 1996) where an alternative orthogonalization approach via a Householder transform was applied. Extensions of the orthogonalization procedure via Householder transform to nonlinear models have also been reported (Hiskens, 2001; Velez-Reyes & Verghese, 1995). Recently, the orthogonalization method has been extended to investigate prediction accuracy of a model and not just identifiability of the parameters (Chu, Huang, & Hahn, 2009), where the same orthogonalization procedure is applied, however, a different criterion is used to select the parameter at each step.

The orthogonalization method has been widely used for analysis of complex models, e.g. biochemical pathways (Gadkar, Gunawan & Doyle, 2005; Jaqaman & Danuser, 2006; Yue et al., 2006), chemical reaction networks (Kou et al., 2005; Lin, Biegler & Jacobson, 2010; Puskas et al., 2005), and industrial process systems (Burth, Verghese & Velez-Reyes, 1999; Hiskens, 2004; Ma et al., 2008). Since the orthogonalization method is important in both theory and practice, a detailed description of the technique is provided next.

The orthogonalization method aims to determine a set of orthogonal bases from the sensitivity vectors. The norm and linear dependence of sensitivity vectors are taken into account simultaneously in the orthogonalization procedure. The orthogonalization method can be implemented by either the Gram-Schmidt algorithm (Yao et al., 2003; Lund and Foss, 2008) or the Householder transform (Hiskens, 2001; Velez-Reyes & Verghese, 1995). The two orthogonalization procedures return identical result. The procedure involving Householder transforms is more numerically stable than the Gram-Schmidt algorithm (Golub & van Loan, 1996), however, the latter is simpler to interpret.

Parameter selection via the Gram-Schmidt orthogonalization procedure involves the following steps:

Step 0 (Initiation). Set the number of selected parameters to zero, i.e., $i = 0$, and the projected sensitivity vectors to the original sensitivity vectors as $s_i^{(0)} = s_i$, $i = 1, \cdots, p$

Step 1 (Selection). Select the parameter indexed by k which is determined by

$$k = \arg\max_j \left(s_j^{(i)}\right)^T s_j^{(i)} \qquad (26)$$

Step 2 (Stopping test). If $\left(s_k^{(i)}\right)^T s_k^{(i)} < \lambda$ (given threshold level) then stop.

Step 3 (Projection). Let $s_j^{(i+1)} = s_j^{(i)} - \frac{\left(s_j^{(i)}\right)^T s_k^{(i)}}{\left(s_k^{(i)}\right)^T s_k^{(i)}} s_k^{(i)}$ and return to Step 1 with $i = i+1$.

The key step in the algorithm is to project the sensitivity vectors of the unselected parameters onto the space orthogonal to that spanned by the sensitivity vectors of the previously selected parameters. The projection aims to remove the parameter's effect on the output covered by the previously selected parameters. The parameter which has the largest not yet covered effect is selected at each step.

Fig. 2. illustrates parameter selection via the Gram-Schmidt orthogonalization procedure. Four parameters are investigated and their sensitivity vectors are drawn in the figure. The first selected parameter is the one with the longest sensitivity vector, i.e., θ_2 is selected. Then the sensitivity vectors of the unselected parameters, θ_1, θ_3, θ_4, are projected onto the plane perpendicular to the sensitivity vector of θ_2. The projected sensitivity vector with the largest length is determined to be $s_3^{(1)}$, which results in θ_3 being selected as the parameter for this step. The projection is applied again and the remaining sensitivity vectors are projected on to the line perpendicular to the sensitivity vector $s_3^{(1)}$. The parameter θ_4 is selected at this step since its projected sensitivity vector is longer than that of the parameter θ_1. The sequence of selected parameters is θ_2, θ_3, and θ_4. From the figure it can be seen that the length of the sensitivity vectors is reduced after each projection. The remaining length of the sensitivity vectors can be used as a stopping criterion for the algorithm.

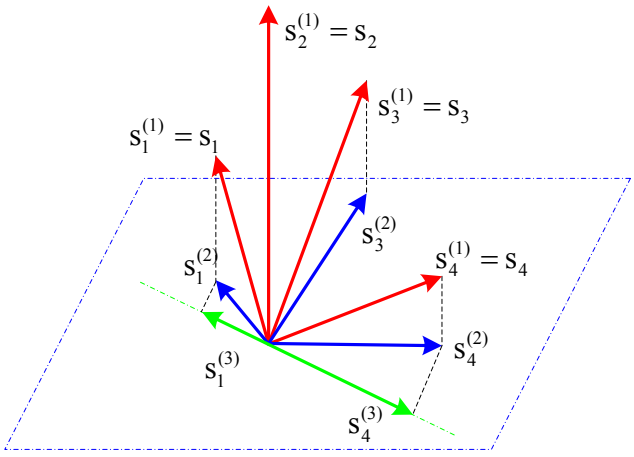

Figure 2. Illustration of the orthogonalization method via the Gram-Schmidt algorithm. Sensitivity vectors for each step have the same color (red for the initial step, blue for the first step, and green for the second step).

Even though the orthogonalization method is simple and easy to interpret, it is not clear how selection improves the estimation performance. The link between the orthogonalization method and selection based upon the *D*-criterion was derived in Chu & Hahn (2007), where it was shown that the orthogonalization method is a forward selection which maximizes the *D*-criterion. The proof in Chu & Hahn (2007) was made based on QR factorization. In this work, a more direct proof is based on the Schur complement:

At the initial step the longest sensitivity vector is chosen for initialization of the orthogonalization method. Given that the cross product of a vector is a scalar which is equal to the squared norm of the vector and the determinant of a scalar is itself, the longest sensitivity vector is also the one which maximizes the *D*-criterion if one parameter is selected. After the initial step ($i \geq 1$) of the procedure, the sensitivity matrix of the previously selected parameters is assumed to be $S^{(i-1)}$ and a new sensitivity vector, assumed to be $s^{(i-1)}$, is added. Then the new sensitivity matrix is given by $S^{(i)} = \left[S^{(i-1)} \quad s^{(i-1)}\right]$ and the *D*-criterion is

$$\begin{aligned} & \varphi_D\left(S^{(i)T}S^{(i)}\right) \\ & \det\left(S^{(i)T}S^{(i)}\right) \\ & \det\left(\begin{bmatrix} S^{(i-1)T}S^{(i-1)} & S^{(i-1)T}s^{(i-1)} \\ s^{(i-1)T}S^{(i-1)} & s^{(i-1)T}s^{(i-1)} \end{bmatrix}\right) \\ & \det\left(S^{(i-1)T}S^{(i-1)}\right)\left(s^{(i-1)T}s^{(i-1)} - s^{(i-1)T}S^{(i-1)}\left(S^{(i-1)T}S^{(i-1)}\right)^{-1}S^{(i-1)T}s^{(i-1)}\right) \end{aligned} \tag{27}$$

where the matrix given by $s^{(i-1)T}s^{(i-1)} - s^{(i-1)T}S^{(i-1)}\left(S^{(i-1)T}S^{(i-1)}\right)^{-1}S^{(i-1)T}s^{(i-1)}$ is the Schur complement of $S^{(i-1)T}S^{(i-1)}$.

By calculating the Schur complement, the determinant is decomposed into the product of two terms. The first term is the determinant of the previous sensitivity matrix which does not change by adding the new sensitivity vector. Adding the new sensitivity vector only changes the second term. Forward selection determines which parameter should be added such that the determinant will be maximized. The result is that the parameter selected by forward selection is the one with the largest value of the Schur complement, which can be further expressed as

$$\begin{aligned} & s^{(i-1)T}s^{(i-1)} - s^{(i-1)T}S^{(i-1)}\left(S^{(i-1)T}S^{(i-1)}\right)^{-1}S^{(i-1)T}s^{(i-1)} \\ & s^{(i-1)T}\left(I - S^{(i-1)}\left(S^{(i-1)T}S^{(i-1)}\right)^{-1}S^{(i-1)T}\right)s^{(i-1)} \\ & s^{(i-1)T}P^{(i-1)}s^{(i-1)} \\ & \left(P^{(i-1)}s^{(i-1)}\right)^T\left(P^{(i-1)}s^{(i-1)}\right) \\ & s^{(i)T}s^{(i)} \end{aligned} \tag{28}$$

where $P^{(i-1)} = I - S^{(i-1)}\left(S^{(i-1)T}S^{(i-1)}\right)^{-1}S^{(i-1)T}$ is a projection matrix. This matrix projects $s^{(i-1)}$ to $s^{(i)}$ in the space orthogonal to the columns in $S^{(i-1)}$ and $s^{(i)T}s^{(i)}$ is the

squared norm of the projected vector. Therefore, the procedure for adding a new parameter via forward selection with the *D*-criterion is identical to the orthogonalization method.

The connection between the orthogonalization method and forward selection provides insights into the orthogonalization procedure. Due to this connection, the orthogonalization method can be compared and integrated with other methods used for parameter set selection. The approach cannot guarantee that an optimal solution is found, just like any other forward selection procedure, however, it is simple to implement and can be applied to large-scale systems. Furthermore, it can provide a good initial guess of which parameters to estimate and it can also be used as a guideline to determine the number of parameters that should be selected for estimation.

Nonlinear Luenberger Observers

The previous section reviewed techniques for dealing with models which are overparameterized. While it has been discussed in detail how one can determine which parameters of a model should be estimated from data, the actual task of parameter estimation has not been dealt with so far. In order to address this point, this section will present an overview of one class of observer, i.e., nonlinear Luenberger observers, which can be used for state and parameter estimation.

State estimation plays a key role in process monitoring and process control applications. When the process dynamics is approximately linear, the classic Kalman filter or Luenberger observer provides an effective method for on-line state estimation. However, in the presence of process nonlinearities, it is well-known that linear state estimators or observers can be inadequate (Soroush, 1998). For this reason, there have been many research efforts over the past decades to develop nonlinear state estimation methods for chemical processes, which use the nonlinear process model and account for the nonlinear dynamic behavior (Bastin & Dochain, 1990; Dochain, 2003; Kantor, 1989; Kazantzis & Kravaris, 2000; Kurtz & Henson, 1998; Rajaraman et al., 2004; Rajaraman et al., 2006; Singh & Hahn, 2005; Soroush, 1997; Soroush, 1998; Valluri and Soroush, 1996). It should be noted that the focus of the techniques discussed in this section will be on observer-based approaches and not on moving-horizon estimators or particle filters as advances involving these techniques have been discussed at a paper presented at the previous CPC (Rawlings & Bakshi, 2006).

One approach that has been tried by many researchers and industrial practitioners in an attempt to handle process nonlinearities has been the extended Kalman filter (Gelb, 1974) and, more recently, along the same lines, the extended Luenberger observer (Zeitz, 1987). In this approach, the design is based on an approximate local linearization of the system around a reference trajectory. Even though the extended Kalman filter has found many industrial applications, there have been many studies that established its serious difficulties in the presence of strong process nonlinearities (e.g. Kantor, 1989; Valluri and Soroush, 1996).

This section will try to give an overview of some alternative approaches for handling process nonlinearities in observer design, originating from nonlinear systems theory, and aiming at overcoming the difficulties of the extended Kalman filter / extended Luenberger observer. The mathematical formulation will be deterministic and the class of applications that will be targeted will be primarily related to process monitoring. It could also include parameter estimation problems, when these are appropriately reformulated as state estimation problems for an extended system that accounts for the nature of parameter variability (e.g. random steps, random ramps, etc.).

In particular, in the present work, we consider unforced nonlinear systems

$$\begin{aligned} \frac{dx}{dt} &= f(x) \\ y &= h(x) \end{aligned} \tag{29}$$

with $x \in \mathbb{R}^n$ being the state vector, $y \in \mathbb{R}$ the measured output, $f : \mathbb{R}^n \to \mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}$ nonlinear functions. The objective is to construct a state observer to generate an estimate of the state vector x, driven by the output measurement y.

The understanding here is that system (29) represents a chemical process, including its control loops.

It is possible to generalize the well-known Luenberger observer (Luenberger, 1966; Luenberger, 1971; Chen, 1984) to nonlinear systems as follows:

$$\frac{d\hat{x}}{dt} = f(\hat{x}) + L(y - h(\hat{x})) \tag{30}$$

where $\hat{x} \in \mathbb{R}^n$ is the estimate of the state vector and L is a constant gain vector. The choice of L will determine the stability properties of the error dynamics, and the question is if it would be possible to select the observer gain L to guarantee rapid decay of the error over a large enough operating region. In general, analysis beyond a small neighborhood of an equilibrium point can be very difficult.

More generally, it is possible to consider a Luenberger observer with state-dependent gain:

$$\frac{d\hat{x}}{dt} = f(\hat{x}) + L(\hat{x})(y - h(\hat{x})) \tag{31}$$

and this opens up many more possibilities. In fact, under certain conditions, it is possible to select the gain function $L(x)$ to "shape" the resulting error dynamics. In this direction, there has been a rich systems theory literature, involving a variety of mathematical approaches, including high-gain observers (Ciccarela et al., 1993; Ciccarela et al., 1995; Gauthier et al., 1992; Gauthier & Kupka, 1994; Gauthier & Kupka, 2001), observer error Lyapunov function methods (Tsinias, 1989; Tsinias, 1990), exact linearization methods (Andrieu & Praly, 2006; Astolfi & Praly, 2006; Guay, 2002; Kazantzis & Kravaris, 1998; Kreisselmeier & Engel, 2003; Krener & Isidori, 1983; Krener & Respondek, 1985; Krener & Xiao, 2002; Krener & Xiao, 2005; Xia & Gao, 1989; Xiao, 2006a) and

optimization-based approaches (Kang, 2006; Moraal & Grizzle, 1995; Rao et al., 2003; Zimmer, 1994).

It would not be possible to review the entire spectrum of theoretical ideas and results on nonlinear observers in a single review paper. Instead, we will focus on those results that are applicable to general nonlinear systems and, at the same time, are directly amenable to engineering calculations. There have a number of new developments in the past decade, but there has not been a review paper to "put the pieces together" in a unified framework, from an applications perspective. This will be attempted in the present section.

Exact Linearization of the error dynamics, with prescribed eigenvalues

In exact linearization design methods, the objective is to build an observer so that the resulting error dynamics is linear in curvilinear coordinates, and with pre-specified rate of decay of the error. This leads to the following selection of the state-dependent observer gain (Kazantzis & Kravaris, 1998):

$$L(x) = \left[\frac{\partial T}{\partial x}(x)\right]^{-1} b \tag{32}$$

where $T(x)$ is an invertible function from $\mathbb{R}^n$ to $\mathbb{R}^n$ that satisfies the system of linear partial differential equations:

$$\frac{\partial T}{\partial x}(x) f(x) = AT(x) + bh(x) \tag{33}$$

and the matrices A and b are design parameters.

Under the above choice of gain, the observer (31) can be expressed in transformed coordinates $z = T(\hat{x})$ as:

$$\begin{aligned} \frac{dz}{dt} &= Az + by \\ \hat{x} &= T^{-1}(z) \end{aligned} \tag{34}$$

which means that the observer has linear dynamics in transformed coordinates. Moreover, the error dynamics follows the differential equations:

$$\frac{d}{dt}\left(T(\hat{x}) - T(x)\right) = A\left(T(\hat{x}) - T(x)\right) \tag{35}$$

hence the rate of decay of the error in curvilinear coordinates is governed by the eigenvalues of the matrix A.

The key to the application of the exact linearization method is the linear partial differential equation (33). Existence, uniqueness and invertibility of the solution can be guaranteed under appropriate non-resonance conditions between A and the Jacobian of f(x), observability of (29) and controllability of (A,b) (see Kazantzis & Kravaris (1998) and Krener & Xiao (2002) for details on local results, Andrieu & Praly (2006) and Astolfi & Praly (2006) for global results).

Computationally, equation (5) can be handled via a power series solution (Kazantzis & Kravaris, 1998). The computational effort increases with the size of the state vector x and the truncation order of the power series.

Comparing the observer under consideration to the constant-gain observer (30), one can immediately see the advantage: the observer has guaranteed error convergence properties that can hold far beyond a neighborhood of the equilibrium point. Of course, there is a cost involved: it is the computational effort in solving the associated partial differential equations.

Observer for state and disturbance estimation

Problems of simultaneous state and disturbance estimation can be handled as state estimation problems for an extended system. Consider

$$\begin{aligned} \frac{dx}{dt} &= \mathcal{F}(x, w) \\ y &= \mathcal{H}(x, w) \end{aligned} \tag{36}$$

where x is the state vector and w is the vector of disturbances, which are assumed to be of known nature (e.g. step, ramp, …) but unknown size. In the case of constant or step disturbances, they will satisfy $\frac{dw}{dt} = 0$; more generally,

$$\frac{dw}{dt} = \mathcal{S}(w) \tag{37}$$

Thus, the problem of state and disturbance estimation becomes a problem of state estimation for the extended system of (36) and (37), i.e. of

$$\begin{aligned} \frac{d}{dt}\begin{bmatrix} x \\ w \end{bmatrix} &= \begin{bmatrix} \mathcal{F}(x, w) \\ \mathcal{S}(w) \end{bmatrix} \\ y &= \mathcal{H}(x, w) \end{aligned} \tag{38}$$

with state $x_{ext} = \begin{bmatrix} x \\ w \end{bmatrix}$. Therefore, a state observer can be built for the extended system

$$\frac{d}{dt}\begin{bmatrix} \hat{x} \\ \hat{w} \end{bmatrix} = \begin{bmatrix} \mathcal{F}(\hat{x}, \hat{w}) \\ \mathcal{S}(\hat{w}) \end{bmatrix} + \mathcal{L}(\hat{x}, \hat{w})\left(y - \mathcal{H}(\hat{x}, \hat{w})\right) \tag{39}$$

so that the error dynamics is linear and with prescribed eigenvalues, following the method outlined in the previous subsection (Kravaris et al., 2007).

In some applications, it may be desirable to implement the state and disturbance observer in modular form (Friedland, 1986), with a basic observer built for the disturbance-free part of the process dynamics and, on top of it, a disturbance observer and a state-estimate corrector, as shown in the following figure:

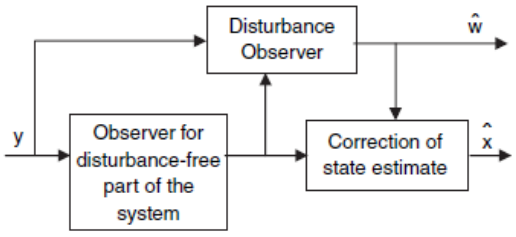

Figure 3. State & disturbance observer in modular form

Given an observer design for the extended system (38) via the exact linearization method of the previous subsection, it is possible to convert the observer equations

to the above modular structure (Kravaris and Savoglidis, 2008). The modular implementation necessitates some additional, relatively straightforward calculations, but has advantages from the point of view of fault detection and isolation applications.

It should be noted that disturbance estimation cannot just result in improved process monitoring, but can be used for improving feedback control as the controller can make use of this additional information about the disturbance to be rejected.

State observer in discrete time

Chemical processes are typically continuous-time systems, but the state observer operates in discrete time, and it is driven by discrete-time output measurements (sampled data). There is a general question of whether to perform discretization at the beginning or at the end.

i) Discretizing the dynamic process model from the beginning, with time step equal to the sampling period, can always be accomplished using an appropriate numerical method, subject to some numerical error. Thus, in discrete time, system (29) is approximately represented in the form:

$$\begin{aligned} x(k+1) &= \Phi(x(k)) \\ y(k) &= h(x(k)) \end{aligned} \tag{40}$$

Using this discrete-time representation, it is possible to formulate exact observer linearization approaches in a completely analogous manner as with continuous-time systems (Califano et al., 2003; Chung & Grizzle, 1990; Kazantzis & Kravaris, 2001; Kazantzis, 2009; Kravaris, 2009; Lee and K. Nam, 1991; Lee & Nam, 1991; Lin & Byrnes, 1995; Xiao et al., 2003; Xiao, 2006b; Xiao et al., 2008). In particular, it is possible to construct a discrete-time state observer with linear dynamics in curvilinear coordinates, and with prescribed eigenvalues (Kazantzis & Kravaris, 2001):

$$\begin{aligned} z(k+1) &= A_d z(k) + b_d y(k) \\ \hat{x}(k) &= T_d^{-1}(z(k)) \end{aligned} \tag{41}$$

where $T_d(x)$ is an invertible function that satisfies the system of linear functional equations:

$$T_d(\Phi(x)) = A_d T_d(x) + b_d h(x) \tag{42}$$

and the matrices A_d and b_d are design parameters. The resulting error dynamics follows the difference equations:

$$T_d(\hat{x}(k+1)) - T_d(x(k+1)) = A_d\left(T_d(\hat{x}(k)) - T_d(x(k))\right) \tag{43}$$

i.e. the rate of decay of the error in curvilinear coordinates is governed by the eigenvalues of the matrix A_d.

ii) Discretizing the continuous-time observer (34) is also a possibility, if sampling is relatively fast. Then, assuming that the output is approximately constant in between measurements (zero-order hold), system (34) can be discretized as follows:

$$\begin{aligned} z(k+1) &= A_d z(k) + b_d y(k) \\ \hat{x}(k) &= T^{-1}(z(k)) \end{aligned} \tag{44}$$

where $T(x)$ is the solution of the system of linear partial differential equations (33) and $A_d = e^{AP}$, $b_d = \int_0^P e^{A\tau} b \, d\tau$, with P being the sampling period. Notice that (44) is an exact discretization of (34) when y(t) is piecewise constant, because of the linearity of the dynamics of the continuous-time observer.

A comparison of discretization at the beginning versus discretization at the end should be made at this point. In the former case, there is always a numerical error in the representation (40), whose effect on the observer design and its propagation during observer implementation could be a significant issue. Apart from this difficulty, observer design in discrete time is very similar to observer design in continuous-time and there is a mathematical similarity between the linear partial differential equations (33) and the linear functional equations (42), both in terms of the existence-uniqueness-invertibility conditions for their solution, and in terms of the calculation method (power series in both cases) and computational effort (see Kazantzis & Kravaris (2001) and Xiao et al. (2003)).

On the other hand, discretization at the end directly utilizes the continuous-time design and it seems to be the most meaningful approach in the presence of relatively fast sampling. The reason is that, because of the linearity of the observer dynamics, exact discretization of the continuous-time observer is possible and therefore, stability of the error dynamics and rate of decay of the error will be preserved after discretization. This is a key advantage of the observer linearization method.

In either case, the discrete-time observer will not provide estimates of the inter-sample behavior of the process states.

Sampled-data observer

When sampling is not performed at a very fast rate, inter-sample behavior becomes important and needs to be accurately estimated by the observer. For this purpose, the model (29) could be used to predict the evolution of the output during the time period in between two consecutive measurements. In particular, the model can be used to estimate of the rate of change of the output, in order to be able to continuously apply a correction on the most recent measurement during this period.

Since the rate of change of the output is $\frac{dy}{dt} = L_f h(x)$, where $L_f h(x) = \sum_{j=1}^{n} f_j(x) \frac{\partial h}{\partial x_j}(x)$ denotes the Lie derivative, it is possible to estimate of the rate of change of the output from the Lie derivative, evaluated at the state estimate $\hat{x}$:

$$\frac{d\hat{y}}{dt} = L_f h(\hat{x}), \quad t \in [t_i, t_{i+1})$$

where t_i, t_{i+1} denote two consecutive sampling instants.

The above can be integrated on line, where the most recent measurement is used as the initial condition. This leads to the following inter-sample output predictor:

$$\frac{d\psi}{dt} = L_f h(\hat{x}), \quad t \in [t_i, t_{i+1})$$
$$\psi(t_i) = y(t_i) \tag{45}$$

with ψ representing the output prediction.

When the continuous-time observer (31) is driven by the output predictor (45), this results in the following sampled-data observer:

$$\frac{d\hat{x}}{dt} = f(\hat{x}) + L(\hat{x})\big(\psi - h(\hat{x})\big), \quad t \in [t_i, t_{i+1})$$
$$\frac{d\psi}{dt} = L_f h(\hat{x}), \quad t \in [t_i, t_{i+1}) \tag{46}$$
$$\psi(t_i) = y(t_i)$$

Figure 4 depicts the construction of the sampled-data observer, as the coupling of the continuous-time observer with the inter-sample output predictor.

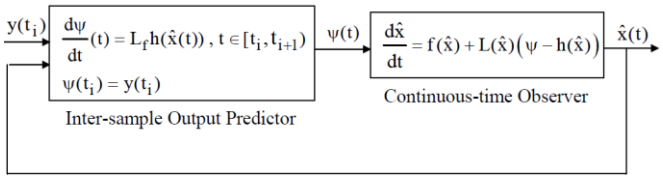

Figure 4. Structure of the sampled-data observer

It was shown in Karafyllis & Kravaris (2009) that, as long as the sampling period does not exceed a certain limit, stability of the error dynamics and robustness with respect to measurement error for the continuous-time observer (31) imply stability of the error dynamics and robustness with respect to measurement error for the sampled-data observer (46). In other words, the sampled-data implementation inherits the key properties of the continuous-time design and, in fact, these properties hold at all times, not just at the sampling instants.

Observer design in the presence of delayed measurements

Measurement delays are quite common in practice and, therefore, appropriate modifications will be introduced to the observer equations to be able to handle this situation. Given a nonlinear system with delayed output of the form

$$\frac{dx}{dt}(t) = f(x(t))$$
$$y(t) = h(x(t-\theta)) \tag{47}$$

where θ is the measurement delay, it is possible to construct an observer with delayed linear dynamics:

$$\frac{dz}{dt}(t-\theta) = Az(t-\theta) + by(t)$$
$$\hat{x}(t) = T^{-1}(z(t)) \tag{48}$$

As long as the transformation map $T(z)$ satisfies (33), this observer will generate an estimate of $x(t-\theta)$ from the output measurements up to time t, and the error dynamics for the estimate of $x(t-\theta)$ will follow (35).

Furthermore, the model (29) can be utilized to predict the state at time t, given the estimate at time $t-\theta$, in the same spirit as in Smith-predictor methods for dead time compensation. This leads to the following overall observer:

$$\frac{d\hat{z}_0}{dt}(t) = A\hat{z}_0(t) + by(t)$$
$$\frac{d\hat{x}}{d\tau}(\tau) = f(\hat{x}(\tau)), \ \tau \in [t-\theta, t], \quad \hat{x}(t-\theta) = T^{-1}(\hat{z}_0(t)) \tag{49}$$

where $\hat{z}_0(t)$ represents the delayed state estimate in transformed coordinates and $\hat{x}(t)$ the estimate of the state at current time, predicted from the delayed one through the model equations. The observer (49) is a special case of a chain observer for handling delayed output measurements (see Germani et al. (2002) and Kazantzis & Wright (2005) for generalization involving m-step prediction).

In discrete-time, the handling of delays is much more straightforward, since the mathematical form of the state-space description (40) remains unaffected. However, since chemical processes are continuous-time systems, it would make more sense to go by an observer of the form (49), appropriately discretized, and possibly coupled with an inter-sample predictor along the lines of the previous subsection.

Functional observers

In many applications, the entire state vector is not really needed to be estimated, but rather, some function(s) of the state vector. Then, it would make sense to try to reduce the order of the observer - if possible - in order to reduce computational effort in real-time implementation. This leads to the development of functional observers for nonlinear systems, in the same spirit as they have been developed for linear systems (Luenberger, 1966; Luenberger, 1971; Chen, 1984). In particular, considering the nonlinear system (29), suppose that the function

$$v = q(x) \tag{50}$$

needs to be estimated instead of the entire state vector, and an observer is sought, of dimensionality less than n.

The construction of a functional observer in the context of exact linearization proceeds with the same steps as before, the main difference being that now the mapping $T(x)$ is not from $\mathbb{R}^n$ to $\mathbb{R}^n$, but from $\mathbb{R}^n$ to $\mathbb{R}^\ell$, $\ell < n$ and the observer dynamics $\frac{dz}{dt} = Az + by$ is ℓ - dimensional. The mapping $T(x)$ must still satisfy the partial differential equation (33) and in addition, $q(x)$ must be expressible as a function of $T(x)$ and $h(x)$, say $q(x) = g(T(x), h(x))$. Then

$$\frac{dz}{dt} = Az + by$$
$$\hat{v} = g(z, y) \tag{51}$$

will be a functional observer for system (29) and the function to be estimated given by (50).

A reduction of observer order to $\ell = (n-1)$ is always feasible, but further reduction is possible only under special conditions on $q(x)$ (see Kravaris (2011) for the pertinent conditions, as well as applications to chemical and biochemical reactors).

Computational issues

Reducing computational requirements for the real-time implementation of the nonlinear observer is, of course, significant, but the critical issue in applying nonlinear observers is how to reduce computational effort and complexity in calculating state-dependent observer gains.

When solving the linear partial differential equations (33) or the linear functional equations (42) using standard symbolic calculation packages, computational complexity can be excessively large even for medium-scale problems. To be able to handle computational complexity, automatic differentiation algorithms seem to be the answer – see Röbenack (2004) and references therein, regarding the use of automatic differentiation algorithms for multivariate Taylor series expansions, Faà di Bruno formulas, Lie derivatives, etc.

Future Research

Even though a wide range of methods have been developed to deal with ill-conditioned estimation problems resulting from complex models, there are still many open problems waiting to be solved. Some suggestions for potential future research directions are made in the following paragraphs.

(1) **Development of New Regularization Techniques**. Regularization generally adds a constraint on the parameters values to the estimation problem to reduce variability in the parameter estimates. Three commonly used techniques have been discussed in this section. These are based upon the Euclidean norm, however, it should be possible to extend this work to the more general L_p norm. Some promising results involving the 1-norm have been reported (Tibshirani, 1996), however, more work remains to be done to come up with procedures used by a general norm.

(2) **Incorporating Parameter Uncertainty**. Most regularization methods are derived based upon linear models. When extensions to nonlinear systems are made then these extensions generally involve linearization in one form or another. For example, the local sensitivity matrix, which is the result of the linearization, plays the same role as the design matrix in the linear model. It is possible to extend existing techniques for parameter set selection from linear to nonlinear models by using this local sensitivity matrix. However, one problem that arises is that the local sensitivity matrix of a nonlinear systems is dependent upon the, not-yet accurately known, values of the parameters. It has been reported that even a small change in the parameter values can significantly change the results calculated based on the local sensitivity matrix (Chu & Hahn, 2007). Developing methods which can incorporate more information about the parameter uncertainty into the parameter set selection procedure is one promising area of research.

(3) **Making Use of Global Sensitivity Analysis**. One approach to deal with the challenge mentioned under (2) is to use global sensitivity analysis as it can incorporate information about parameter uncertainty into the sensitivity analysis. However, this has the drawback that the experimental design criteria cannot be applied to the commonly used global sensitivity analysis techniques. For example, most global sensitivity analysis techniques return results that are inconsistent with the sensitivity matrix derived from local sensitivity analysis even if the model is linear (Chu & Hahn, 2010). In order to address this point, it would be required to either derive a new set of experimental design criteria (unlikely) or to modify existing global sensitivity analysis techniques such that the procedures can be viewed as an extension of the linear methods, which would then allow to incorporate the experimental design criteria into the procedure.

(4) **Integrating Parameter Set Selection and Experimental Design.** Many different variables such as the input profiles, measured outputs, initial states, and sampling points can vary from one experiment to another. While it is possible to incorporate information about all of these experimental design variables into the sensitivity analysis procedure, it is usually not taken into account that additional experiments could be designed to further improve the results. Sensitivity analysis and experimental design are two tasks that are usually performed separately from each other, even though they strongly affect each other (Chu & Hahn, 2008). The long-term goal in this area would be to integrate the two procedures.

(5) **Investigation of New Solution Techniques for the Parameter Set Selection Optimization Problem**. The majority of this section dealt with different approaches for determining a sub-optimal solution, including regularization methods, to the parameter set selection optimization problem. However, it would be ideal if this problem could be optimally solved while also taking uncertainty in the measurements and the model structure into account. This would require new formulations/decompositions of the mixed-integer nonlinear programming problem resulting from the parameter set selection problem.

One important area of future research for the design and implementation of nonlinear Luenberger observers is given by the following:

(6) **Development of Software Packages for Designing Nonlinear Observers**. Reduction of the computational effort when computing state-dependent observer gains is a critical issue for the application of nonlinear observers. An important future research direction will therefore be the development of efficient and effective

software packages to assist engineers in design calculations for nonlinear observers.

Conclusions

This paper provided an overview of two areas dealing with improving accuracy of models used for process monitoring and control. The first part of the paper reviewed background information on and techniques used for selection of model parameters for estimation, while the second part dealt with nonlinear Luenberger observer design for state and parameter estimation.

Acknowledgments

The authors (JH, YC) gratefully acknowledge partial financial support from the American Chemical Society (Grants ACS-PRF#48144-AC9 and ACS-PRF#50978-ND9) and the National Science Foundation (Grant NSF# 0941313).

References

Ancheyta, J., Sanchez, S., & Rodriguez, M. A. (2005). Kinetic modeling of hydrocracking of heavy oil fractions: A review. *Catalysis Today*, 109, 76-92.

Andrieu, V., & Praly, L. (2006). On the existence of a Kazantzis-Kravaris/Luenberger observer. *SIAM Journal on Control and Optimization*, 45, 432-456.

Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J. A., & Blom, J. G. (2009). Systems biology: parameter estimation for biochemical models. *The FEBS journal*, 276, 886-902.

Ashyraliyev, M., Jaeger, J.,& Blom, J. G. (2008). Parameter estimation and determinability analysis applied to Drosophila gap gene circuits. *BMC Systems Biology*, 2, 83.

Astolfi, A., & Praly, L. (2006). Global complete observability and output-to-state stability imply the existence of a globally convergent observer. *Mathematics of Control, Signals, and Systems*, 18, 32-65.

Bastin, G. & Dochain, D. (1990). *On-Line Estimation and Adaptive Control of Bioreactors*. Amsterdam: Elsevier.

Bindlish, R., Rawlings, J. B., & Young, R. E. (2003). Parameter estimation for industrial polymerization processes. *AIChE Journal*, 49, 2071-2078.

Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Forward selection of explanatory variables. *Ecology*, 89, 2623-2632.

Box, G. E. P., & Draper, N. R. (2007). *Response surfaces, mixtures, and ridge analyses*. Wiley-Interscience.

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.

Brun, R., Reichert, P.,& Kunsch, H. R. (2001). Practical identifiability analysis of large environmental simulation models. *Water Resources Research*, 37, 1015-1030.

Burth, M., Verghese, G. C., &Velez-Reyes, M. (1999). Subset selection for improved parameter estimation in on-line identification of a synchronous generator. *IEEE Transactions on Power Systems*, 14, 218-225.

Califano, C., Monaco, S., & Normand-Cyrot, D. (2003). On the observer design in discrete-time. *Systems & Control Letters, 49*, 255-265.

Chen, C. T. (1984). *Linear System Theory and Design*. New York: Holt, Rinehart and Winston.

Chou, I. C.,& Voit E. O. (2009). Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Mathematical Biosciences*, 219, 57-83.

Chu, Y. F., & Hahn, J. (2007). Parameter set selection for estimation of nonlinear dynamic systems. *AIChE Journal*, 53, 2858-2870.

Chu, Y. F., & Hahn, J. (2008). Integrating parameter selection with experimental design under uncertainty for nonlinear dynamic systems. *AIChE Journal*, 54, 2310-2320.

Chu, Y. F., & 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. F., & Hahn, J. (2010). Quantitative optimal experimental design using global sensitivity analysis via quasi-linearization. *Industrial & Engineering Chemistry Research*, 49, 7782-7794.

Chu, Y. F., Huang, Z. Y., Hahn J. (2009). Improving prediction capabilities of complex dynamic models via parameter selection and estimation. *Chemical Engineering Science*, 64, 4178-4185.

Chung, S. T. & Grizzle, J. W. (1990). Sampled-data observer error linearization. *Automatica*, 26, 997-1007.

Ciccarela, G., Dalla Mora, M., & Germani, A. (1993). A Luenberger-like observer for nonlinear systems. *International Journal of Control*, 57, 537-556.

Ciccarela, G., Dalla Mora, M. & Germani, A. (1995). A robust observer for discrete-time nonlinear systems. *Systems & Control Letters*, 24, 291-300.

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.

Dochain, D. (2003). State and parameter estimation in chemical and biochemical processes: a tutorial. *Journal of Process Control*, 13, 801-818.

Draper, N.R., & Guttma, I. (1995). Confidence intervals versus regions. *Journal of the Royal Statistical Society. Series D*., 44, 399-403.

Esposito, W. R.,& Floudas, C. A. (2000). Global optimization for the parameter estimation of differential-algebraic systems. *Industrial & Engineering Chemistry Research*, 39, 1291-1310.

Fontoura, J. M. R., Santos, A. F., Silva, F. M., Lenzi, M. K., Lima, E. L. &Pinto, J. C. (2003). Monitoring and control of styrene solution polymerization using NIR spectroscopy. *Journal of Applied Polymer Science*, 90, 1273-1289.

Friedland, B. (1986). *Control System Design. An Introduction to State-Space Methods*. New York: McGraw-Hill.

Gadkar, K. G., Gunawan, R., & Doyle, F. J. (2005). Iterative approach to model identification of biological networks. *BMC bioinformatics*, 6, 155.

Gauthier, J. P., Hammouri, H., & Othman, S. (1992). A simple observer for nonlinear systems: Application to bioreactors. *IEEE Transactions on Automatic Control*, 37, 875-880.

Gauthier, J. P. & Kupka, I. A. K. (1994). Observability and observers for nonlinear systems. *SIAM journal on control and optimization*, 32, 975-994.

Gauthier, J. P. & Kupka, I. A. K. (2001). *Deterministic observation theory and applications*. New York: Cambridge University Press.

Gelb, A. (1974). *Applied optimal estimation*. Cambridge, MA: MIT Press.

Germani, A. Manes, C., & Pepe, P. (2002). A new approach to state observation of nonlinear systems with delayed output. *IEEE Transactions on Automatic Control*, 47, 96-101.

Golub, G., Klema, V., & Stewart, G. W. (1976). Rank degeneracy and least squares problems. *Technical Report*

STAN-CS-76-559, Computer Science Department, Stanford University.

Golub, G. H., & van Loan, C. F. (1996). *Matrix computations. 3rd ed*. Johns Hopkins University Press.

Guay, M. (2002). Observer linearization by output-dependent time-scale transformations. *IEEE Transactions on Automatic Control*, 47, 1730-1735.

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.

Hiskens, I. A. (2001). Nonlinear dynamic model evaluation from disturbance measurements. *IEEE Transactions on Power Systems*, 16, 702-710.

Hiskens, I. A. (2004). Power system modeling for inverse problems. IEEE transactions on circuits and systems. I, Fundamental theory and applications. *IEEE transactions on circuits and systems. I, Fundamental theory and applications*, 51, 539-551.

Jaqaman, K., & Danuser, G. (2006). Linking data to models: data regression. *Nature Reviews. Molecular Cell Biology*, 7, 813-819.

Jimenez-Hornero, J. E., Santos-Duenas, I. M., & Garcia-Garcia, I. (2009). Optimization of biotechnological processes. The acetic acid fermentation. Part II: Practical identifiability analysis and parameter estimation. *Biochemical Engineering Journal*, 45, 7-21.

Jolliffe, I. T. (2002). *Principal component analysis (2nd ed)*. Springer.

Kang, W. (2006). Moving horizon numerical observers of nonlinear control systems. *IEEE Transactions on Automatic Control*, 51, 344-350.

Kantor, J. C. (1989). A finite dimensional nonlinear observer for an exothermic stirred-tank reactor. *Chemical Engineering Science*, 44, 1503-1510.

Karafyllis, I., & Kravaris, C. (2009). From continuous-time design to sampled-data design of observers. *IEEE Transactions on Automatic Control*, 54, 2169-2174.

Katare, S., Bhan, A., Caruthers, J. M., Delgass, W. N., & Venkatasubramanian, V. (2004). A hybrid genetic algorithm for efficient parameter estimation of large kinetic models. *Computers & Chemical Engineering*, 28, 2569-2581.

Kazantzis, N. & Kravaris, C. (1998). Nonlinear observer design using Lyapunov's auxiliary theorem. *Systems & Control Letters*, 34, 241-247.

Kazantzis, N. & Kravaris, C. (2000). Nonlinear observer design for process monitoring. *Industrial & engineering chemistry research*, 39, 408-419.

Kazantzis, N. & Kravaris, C. (2001). Discrete-time nonlinear observer design using functional equations. *Systems & Control Letters*, 42, 81-94.

Kazantzis, N. & Wright, R. A. (2005). Nonlinear observer design in the presence of delayed output measurements. *Systems & Control Letters*, 54, 877-886.

Kazantzis, N. (2009). Map invariance and the state reconstruction problem for nonlinear discrete-time systems. *European Journal of Control*, 15, 105-119.

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 Research*, 44, 2428-2442.

Kravaris, C., Sotiropoulos, V., Georgiou, C., Kazantzis, N., Xiao, M., & Krener, A. J. (2007). Nonlinear observer design for state and disturbance estimation. *Systems & Control Letters*, 56, 730-735.

Kravaris, C. & G. Savoglidis. (2008). Modular design of nonlinear observers for state and disturbance estimation. *Systems & Control Letters*, 57, 946-957.

Kravaris, C. (2009). Modular design of discrete-time nonlinear observers for state and disturbance estimation. *European Journal of Control*, 15, 120-130.

Kravaris, C. (2011). Functional observers for nonlinear systems. *In Proceedings of the 9th IEEE International Conference on Control & Automation*. Santiago, Chile, in press.

Kreisselmeier, G., & Engel, R. (2003). Nonlinear observers for autonomous Lipschitz continuous systems. *IEEE Transactions on Automatic Control* , 48, 451-464.

Krener, A. J., & Isidori, A. (1983). Linearization by output injection and nonlinear observers. *Systems & Control Letters*, 3, 47-52.

Krener, A. J., & Respondek, W. (1985). Nonlinear observers with linearizable error dynamics. *SIAM Journal on Control and Optimization*, 23, 197-216.

Krener, A. J. & Xiao, M. (2002). Nonlinear observer design in the Siegel domain. *SIAM Journal on Control and Optimization*, 41, 932-953.

Krener, A. J. & Xiao, M. (2005). Nonlinear observer design for smooth systems. In Perruquetti, W., & Barbot, J. P. (Eds.) *Chaos in Automatic Control*, 411-422. Marcel Dekker.

Kurtz, M. J., & Henson, M. A. (1998). State and disturbance estimation for nonlinear systems affine in the unmeasured variables. *Computers & Chemical Engineering*, 22, 1441-1459.

Lee, W., & Nam, K. (1991). Observer design for autonomous discrete-time nonlinear systems. *Systems & Control Letters*, 17, 49-58.

Luenberger, D. G. (1966). Observers for multivariable systems. *IEEE Transactions on Automatic Control*, 11, 190-197.

Luenberger, D. G. (1971). An introduction to observers. *IEEE Transactions on Automatic Control*, 16, 596-602.

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.

Lin, W., & Byrnes, C. I. (1995). Remarks on linearization of discrete-time autonomous systems and nonlinear observer design. *Systems & Control Letters*, 25, 31-40.

Lin, W. J., Biegler, L. T., Jacobson, A. M. (2010). Modeling and optimization of a seeded suspension polymerization process. *Chemical Engineering Science*, 65, 4350-4362.

Lund, B. F., & Foss, B. A. (2008). Parameter ranking by orthogonalization - Applied to nonlinear mechanistic models. Automatica, 44, 278-281.

Ma, J., Han, D., He, R. M., Dong, Z. Y. Hill, D. J. (2008). Reducing identified parameters of measurement-based composite load model. *IEEE Transactions on Power Systems*, 23, 76-83.

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.

Maria, G. (2004). A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. *Chemical and Biochemical Engineering Quarterly*, 18, 195-222.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 1963 11, 431-441.

McLean, K. A. P., & McAuley, K. B. (2011) Mathematical modeling of chemical processes - obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *Canadian Journal of Chemical Engineering*, In Press.

Miller, A. J. (2002). *Subset selection in regression. 2nd ed*. CRC.

Moles, C. G., Mendes, P., & Banga, J, R. (2003). Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Research*, 13, 2467-2474.

Moraal, P. E. & Grizzle, J. W. (1995). Observer design for nonlinear systems with discrete-time measurements. *IEEE Transactions on Automatic Control*, 40, 395-404.

Puskas, J. E., Chan, S. W. P., McAuley, K. B., Shaikh, S., & Kaszas G. (2005). Kinetics and mechanisms in carbocationic polymerization: The quest for true rate constants. *Journal of Polymer Science. Part A, Polymer Chemistry*, 43, 5394-5413.

Rajaraman, S. & Hahn, J., & Mannan, M. S. (2004). A methodology for fault detection, isolation, and identification for nonlinear processes with parametric uncertainties. *Industrial & Engineering Chemistry Research*, 43, 6774-6786.

Rajaraman, S. & Hahn, J., & Mannan, M. S. (2006). Sensor fault diagnosis for nonlinear processes with parametric uncertainties. *Journal of Hazardous Materials*, 130, 1-8.

Rawlings, J. B. & Bakshi, B.R. (2006). Particle filtering and moving horizon estimation. *Computers & Chemical Engingeering*, 30, 1529-1541.

Rao, C. V., & Rawlings, J. B., & Mayne, D. Q. (2003). Constrained state estimation for nonlinear discrete-time systems: stability and moving-horizon approximations. *IEEE Transactions on Automatic Control*, 48, 246-258.

Robenack, K. (2004). Computation of the observer gain for extended Luenberger observers using automatic differentiation. *IMA Journal of Mathematical Control and Information*, 21, 33-47.

Sandelin, F., Oinas, P., Salmi, T., Paloniemi, J., & Haario, H. (2006). Dynamic modelling of catalytic liquid-phase reactions in fixed beds- Kinetics and catalyst deactivation in the recovery of anthraquinones. *Chemical Engineering Science*, 61, 4528-4539.

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.

Santos, L. O., Afonso, P. A. F. N. A., Castro, J. A. A. M., Oliveira, N. M. C., & Biegler, L. T. (2001). On-line implementation of nonlinear MPC: an experimental case study. *Control Engineering Practice*, 9, 847-857.

Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D., Muller, G. (2002). Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Biotechnology*, 20, 370-375.

Singh, A. K., & Hahn, J. (2005). State estimation for high-dimensional chemical processes. *Computers & Chemical Engineering*, 29, 2326-2334.

Slezak, D. F., Suarez, C., Cecchi, G. A., Marshall, G., Stolovitzky, G. (2010). When the optimal is not the best: Parameter estimation in complex biological models. *PLoS ONE*, 5, 10.

Soroush, M. (1997). Nonlinear state-observer design with application to reactors. *Chemical Engineering Science*, 52, 387-404.

Soroush, M. (1998). State and parameter estimations and their applications in process control. *Computers & Chemical Engineering*, 23, 229-245.

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.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58, 267-288.

Tobajas, M., Mohedano, A. F., Casas, J. A., & Rodriguez, J. J. (2007). A kinetic study of reuterin production by Lactobacillus reuteri PRO 137 in resting cells. *Biochemical Engineering Journal*, 35, 218-225.

Tran, M., Varvarezos, D. K., & Nasir, M. (2005). The importance of first-principles, model-based steady-state gain calculations in model predictive control - a refinery case study. *Control Engineering Practice*, 13, 1369-1382.

Tsinias, J. (1989). Observer design for nonlinear systems. *Systems & Control Letters*, 13, 135-142.

Tsinias, J. (1990). Further results on the observer design problem. *Systems & Control Letters*, 14, 411-418.

Valluri, S., & Soroush, M. (1996). Nonlinear state estimation in the presence of multiple steady-states. *Industrial & Engineering Chemistry Research*, 35, 2645-2659.

van den Bos, A. (2007). *Parameter estimation for scientists and engineers*. Wiley-Interscience.

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*, 991-997.

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.

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.

Xia, X. H., & Gao, W. B. (1989). Nonlinear observer design by observer error linearization. *SIAM Journal on Control and Optimization*, 27, 199-216.

Xiao, M. Q., Kazantzis, N., Kravaris, C., & Krener, A. J. (2003). Nonlinear discrete-time observer design with linearizable error dynamics. *IEEE Transactions on Automatic Control*, 48, 622-626.

Xiao, M. Q. (2006a). The global existence of nonlinear observers with linear error dynamics: a topological point of view. *Systems & Control Letters*, 55, 849-858.

Xiao, M. Q. (2006b). A direct method for the construction of nonlinear discrete-time observer with linearizable error dynamics. *IEEE Transactions on Automatic Control*, 51, 128-135.

Xiao, M. Q., Kazantzis, N., Kravaris, C. (2008). On the non-linear discrete-time observer design problem. *International Journal of Modelling, Identification and Control*, 4, 3-11.

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.

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 sensitivity and identifiability analysis: a case study of an NF-kappa B signalling pathway. *Molecular BioSystems*, 2, 640-649.

Zeitz, M. (1987). The extended Luenberger observer for nonlinear systems. *Systems & Control Letters*, 9, 149-156.

Zimmer, G. (1994). State observation by on-line minimization. *International Journal of Control*, 60, 595-606.