AIChE

Integrating Parameter Selection with Experimental Design Under Uncertainty for Nonlinear Dynamic Systems

Yunfei Chu and Juergen Hahn
Artie McFerrin Dept. of Chemical Engineering, Texas A&M University, College Station, TX 77843

DOI 10.1002/aic.11562
Published online July 9, 2008 in Wiley InterScience (www.interscience.wiley.com).

Models describing complex processes often contain a large number of parameters as part of the nonlinear system. It is usually not possible in practice to identify all parameters because of the number and quality of measurement data as well as interactions among the parameters. A common approach is to select a set of parameters for estimation whereas other parameters are fixed at their nominal values. Such a parameter selection procedure is often based on sensitivity analysis; however, the determined sensitivity values depend on assumed values of the parameters and initial states, as well as known trajectories of the input signals. In this work parameter selection and experiment design procedures are integrated into a unified framework, which optimizes a criterion of the Fisher information matrix and simultaneously takes the effect of uncertainty in the parameter values into account. A hybrid method combining a genetic algorithm and a simultaneous perturbation stochastic approximation is developed to solve the resulting mixed-integer nonlinear programming problem. The technique is illustrated on a model of a continuously-stirred tank reactor and of a signal transduction pathway.
 AIChE J, 54: 2310–2320, 2008
Keywords: parameter selection, experimental design, parameter estimation

Introduction

Parameter estimation involving large-scale dynamic models is an important but also a challenging task.[1–3] One problem for parameter estimation is that complex models often contain dozens or even hundreds of parameters whereas 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.[4] One solution to this is to select a subset of parameters to be estimated whereas all other parameters are fixed at a constant value. The question then becomes one of how to select the parameters to be estimated.

Several methods for parameter selection based on parameter sensitivity vectors have been proposed in the literature. These include, but are not limited to, the collinearity index method,[5] a column pivoting method,[6] an extension of the relative gain array,[7] a Gram-Schmidt orthogonalization method,[8] and a recursive approach based upon principal component analysis.[9]

A systematic approach for parameter selection is based on optimality criteria computed from the Fisher information matrix (FIM). The inverse of the FIM provides a lower bound for the covariance matrix of parameter estimators[4] and it can serve as a measure for the quality of a parameter set. A subset of identifiable parameters can be selected based on optimizing some criteria such as the *D*-optimality or the modified *E*-optimality criterion.[10,11] Parameter selection based on the

Correspondence concerning this article should be addressed to J. Hahn at hahn@tamu.edu.

sensitivity vectors can be regarded as some heuristic approaches to optimize certain optimality criterion implicitly. For example, the collinearity index is equivalent to the normalized E-optimality criterion,[12] and the orthogonalization method is a stepwise approach to maximize the D-optimality criterion.[13]

Parameter selection has been used in a variety of applications, ranging from ecological systems,[14] power systems,[15] production systems,[16] chemical reactions,[17] to biochemical networks.[18] However, one important drawback of common parameter selection schemes based on the FIM or sensitivity vectors is that these techniques depend on the chosen values of the parameters for nonlinear systems, even though the exact values of the parameters are not known before estimation. It has been demonstrated that parameter uncertainty will have a significant effect on the parameter selection.[13]

Another avenue for improving results obtained from parameter estimation is to collect a meaningful data set via experimental design.[19–23] The objective of experimental design is to determine initial conditions and to adjust time-varying inputs so as 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[24–26] 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 whereas 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 article 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 nontrivial 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.

Preliminaries

Optimality criteria of the Fisher information matrix

A criterion is required to measure the quality of a selected set of parameters or to design an experiment. One criterion is the covariance matrix of the estimated parameters, however, 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.[2,4] It is desirable to minimize a criterion involving the inverse of the FIM or equivalently to maximize a criterion of the FIM 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.

Assume that the measured output is a function of the parameters affected by measurement noise

$$\tilde{\mathbf{y}} = \mathbf{y}(\boldsymbol{\theta}) + \varepsilon, \tag{1}$$

where $\tilde{\mathbf{y}} = [\tilde{y}(t_1), \ldots, \tilde{y}(t_n)]^T$ is the observation of the output, $\mathbf{y}(\boldsymbol{\theta}) = [y(t_1, \boldsymbol{\theta}), \ldots, y(t_n, \boldsymbol{\theta})]^T$ is the true value and $\boldsymbol{\varepsilon} = [\varepsilon(t_1), \ldots, \varepsilon(t_n)]^T$ is the measurement noise. In practice, the measurement noise is often assumed to be normally distributed with zero mean and a covariance given by the matrix Σ. As a result, the measurements are also normally distributed and the Fisher information matrix, $\mathbf{F}$, is given by

$$\begin{aligned}\mathbf{F}(\boldsymbol{\theta}) &= E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\tilde{\mathbf{y}}|\boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \ln p(\tilde{\mathbf{y}}|\boldsymbol{\theta})\right] \\ &= \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right)^T \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^T}.\end{aligned} \tag{2}$$

If the measurement noise is uncorrelated and constant with time, the covariance matrix results in $\Sigma = \sigma^2\mathbf{I}$ and the FIM becomes the product of the transpose of the sensitivity matrix within itself. Without loss of generality for the procedure, it can be assumed that $\sigma^2 = 1$:

$$\mathbf{F}(\boldsymbol{\theta}) = \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^T}\right)^T \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^T}. \tag{3}$$

Effects of parameters on the outputs are often correlated for systems with many parameters, resulting in a FIM that is nearly rank-deficient. One way to reduce the variance of the estimate is to reparameterize the model by parameter selection

$$\boldsymbol{\psi} = \mathbf{L}^T\boldsymbol{\theta}, \tag{4}$$

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

$$\mathbf{L} = [\,\mathbf{e}_{i_1} \quad \mathbf{e}_{i_2} \quad \cdots \quad \mathbf{e}_{i_m}\,]. \tag{5}$$

The set $\{i_1, i_2, \ldots, i_m\}$ denotes the index of the selected parameters and $\mathbf{e}_i$ is the i-th column of the identity matrix. The parameters in $\boldsymbol{\psi}$ are a subset of $\boldsymbol{\theta}$ selected for estimation. The FIM of the newly selected parameters is given by

$$\mathbf{F}(\boldsymbol{\psi}) = \mathbf{L}^T\mathbf{F}(\boldsymbol{\theta})\mathbf{L}, \tag{6}$$

To quantify the information content of the FIM, a set of real functions of the FIM are used, for example, the experimental optimality criteria which are named alphabetically.[19] The most popular experimental optimality criterion is the D-optimality criterion, which maximizes the logarithm of the determinant of the FIM:

$$\varphi_D^* = \max\, \varphi_D(\mathbf{F}) = \max\, \log\, \det(\mathbf{F}). \tag{7}$$

This criterion minimizes the volume of the confidence ellipsoid with an arbitrary fixed confidence level for a least

square estimator. This criterion is used in this work, however, the presented techniques can be easily generalized to other criteria found in the literature.[19–23]

Simultaneous perturbation stochastic approximation

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

$$\max_{\mathbf{w}} E_{\mathbf{v}}\, \varphi(\mathbf{w}, \mathbf{v}), \tag{8}$$

where $\mathbf{w}$ represents a vector of the decision variables and $\mathbf{v}$ is a random variable following some distribution. One possible solution for this is to evaluate the expectation by numerical integration over the range of $\mathbf{v}$, resulting in a nonlinear programming involving $\mathbf{w}$. Any of the common techniques for solving nonlinear programming problems can be used for this problem. However, in some cases, the expectation is difficult to be calculated because of the complexity of the criterion function φ. In these cases, a method of stochastic approximation can be applied as it can solve the problem without explicit evaluation of the integration.

The basic version of the stochastic approximation updates the decision variables by[27]

$$\mathbf{w}_{k+1} = \mathbf{w}_k + a_k \hat{\mathbf{g}}_k, \tag{9}$$

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

$$\hat{\mathbf{g}}_k(\mathbf{v}_k) = \frac{\partial}{\partial \mathbf{w}} \varphi(\mathbf{w}_k, \mathbf{v}_k). \tag{10}$$

It should be noted the gradient $\hat{\mathbf{g}}_k$ is not the gradient of the objective function, but it is an unbiased estimator for the gradient. The averaged $\hat{\mathbf{g}}_k$ is the gradient of the objective function; however, the averaging is done implicitly in the iterations (Eq. 9). It has been pointed out that the averaging across the different iterations can be of good use for solving such kinds of problems[27] and the intensive calculation of the expectation by integration can also be avoided.

In some cases, the derivative of the criterion function is difficult to compute directly and gradient-free versions of stochastic approximation have been developed. These include finite perturbation[28] and simultaneous perturbation.[29] The SPSA method only requires two evaluations of the criterion function to approximate the gradient as

$$\hat{\mathbf{g}}_k = \begin{bmatrix} \frac{\varphi(\mathbf{w}_k + c_k\Delta_k, \mathbf{v}_k) - \varphi(\mathbf{w}_k - c_k\Delta_k, \mathbf{v}_k)}{2c_k\Delta_{k1}} \\ \vdots \\ \frac{\varphi(\mathbf{w}_k + c_k\Delta_k, \mathbf{v}_k) - \varphi(\mathbf{w}_k - c_k\Delta_k, \mathbf{v}_k)}{2c_k\Delta_{kp}} \end{bmatrix}, \tag{11}$$

where the perturbation is given by $\Delta_k = [\Delta_{k1}, \ldots, \Delta_{kp}]^T$. A sampling point of $\mathbf{v}_k$ is generated to evaluate $\hat{\mathbf{g}}_k$. The SPSA has been shown to have comparable performance with the finite perturbation at lower computational cost.[29]

The parameters for SPSA can be selected as

$$a_k = a/(k + 1 + A)^{\alpha}, \tag{12}$$

and

$$c_k = c/(k + 1)^{\gamma}. \tag{13}$$

Common values of α and γ are 1 and 1/6, respectively. Each component of the perturbation Δ_k can use a Bernoulli ± 1 distribution with probability of 1/2 for each ± 1 outcome. The underlying theories of SPSA and the implementation can refer to the literature.[30,31]

Genetic algorithm

Genetic algorithm (GA) belong to a class of nongradient-based global optimization techniques. One distinct feature of a GA is that it returns a number of potential solutions in the form of coded binary strings for the decision variables, called chromosomes.[32,33]

In many applications of GA, one chromosome will dominate the population after a few generations. To ensure diversity of the population, a sharing function can be used to modify the fitness of each chromosome.[34] One form of sharing function is

$$h(d_{ij}) = \begin{cases} 1 - \frac{d_{ij}}{D}, & \text{if } d_{ij} \le D \\ 0, & \text{otherwise} \end{cases}, \tag{14}$$

where d_{ij} is the distance between chromosome i and chromosome j, and D is the parameter controlling the diversity. The fitness function of a chromosome is modified by

$$f_i' = \frac{f_i}{\sum_j h(d_{ij})}. \tag{15}$$

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:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \theta_1 + \theta_2 + 5/12\theta_3 + 3u^2\theta_3 + 1/2\theta_3^2 \\ 4\theta_2 + 5/4\theta_3 - u^2\theta_3 + 3/2\theta_3^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \tag{16}$$

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 $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2]^{\mathrm{T}}$ represent noise with a Gaussian distribution with

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

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

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \begin{bmatrix} 1 & 1 & 5/12 + 3u^2 + \theta_3 \\ 0 & 4 & 5/4 - u^2 + 3\theta_3 \end{bmatrix}. \tag{17}$$

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

$$\begin{aligned} \varphi_{1,2} &= 4\ln(2), \\ \varphi_{1,3} &= 2\ln\left(3\theta_3 - u^2 + 5/4\right), \\ \varphi_{2,3} &= \ln\left(\theta_3^2 + \left(26u^2 + 5/6\right)\theta_3 + 169u^4 + 65/6u^2 + 25/144\right), \end{aligned} \tag{18}$$

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 ($\bar{\theta}_i = 1$). The input variable is assumed to be in the range from -1 to 1 with a nominal value of zero, $\bar{u} = 0$.

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:

$$\bar{\varphi} = \varphi\left(\bar{u}, \bar{\boldsymbol{\theta}}\right). \tag{19}$$

The commonly used methods for parameter selection make use of Eq. 19, 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:

$$E_{\boldsymbol{\theta}}[\varphi] = E_{\boldsymbol{\theta}}[\varphi(\bar{u}, \boldsymbol{\theta})]. \tag{20}$$

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. 19 and Eq. 20 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:

$$E_{\boldsymbol{\theta}}[\varphi]^* = E_{\boldsymbol{\theta}}[\varphi(u^*, \boldsymbol{\theta})] \text{ with } u^* = \arg\max_u E_{\boldsymbol{\theta}}[\varphi(u, \boldsymbol{\theta})]. \tag{21}$$

Table 1 lists the criterion values for Eqs. 19–21 of each parameter set given by Eq. 18. It can be concluded from Table 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 cannot be neglected as it can also have a significant impact on the results.

Table 1. Evaluation of each Subset of Parameters Using Different Criteria

Conditions	$\varphi_{1,2}$	$\varphi_{1,3}$	$\varphi_{2,3}$
$u = \bar{u}$, $\theta = \bar{\theta}$; Eq. 19	2.77	2.89	0.70
$u = \bar{u}$, average over $\bar{\theta}$; Eq. 20	2.77	2.69	0.50
$u = u^*_{i,j}$, average over $\bar{\theta}$; Eq. 21	2.77	2.69	5.34

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. 21.

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

$$\mathbf{F} = \mathbf{F}(\mathbf{u}(\mathrm{t}), \mathbf{v}, \boldsymbol{\theta}, \boldsymbol{\delta}). \tag{22}$$

To evaluate the FIM some knowledge about all four kinds of variables is required. Although, the value of the parameters and the unknown factors cannot 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. Equation 23 describes the resulting optimization problem:

$$
\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}(\mathbf{x}, \mathbf{u}(t), \mathbf{v}, \boldsymbol{\theta}, \boldsymbol{\delta}), \quad \mathbf{x}(0) = \mathbf{x}_0(\mathbf{v}, \boldsymbol{\delta}) \\
&\quad y = h(\mathbf{x}, \mathbf{u}(t), \mathbf{v}, \boldsymbol{\theta}, \boldsymbol{\delta}) \\
&\quad \frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^T} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}^T}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^T} + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}^T}, \ \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^T}(0) = 0 \\
&\quad \frac{\partial y}{\partial \boldsymbol{\theta}^T} = \frac{\partial h}{\partial \mathbf{x}^T}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^T} + \frac{\partial h}{\partial \boldsymbol{\theta}^T} \\
&\quad \mathbf{S} = \left[\frac{\partial y(t_1)}{\partial \boldsymbol{\theta}} \quad \cdots \quad \frac{\partial y(t_n)}{\partial \boldsymbol{\theta}}\right]^T \\
&\quad \mathbf{F} = \mathbf{L}^T\left(\mathbf{S}^T\mathbf{S}\right)\mathbf{L} \\
&\quad \mathbf{L} = [\mathbf{e}_{i_1} \quad \mathbf{e}_{i_2} \quad \cdots \quad \mathbf{e}_{i_m}], \text{ with } i_j \text{ that } z_{i_j} = 1 \\
&\quad \mathbf{e}_i \text{ is the } i\text{-th column of the identity matrix} \\
&\quad \mathbf{z}_i \in \{0, 1\} \text{ and } \sum_i z_i = n_z \\
&\quad \mathbf{L}_\mathbf{u} \le \mathbf{u}(t) \le \mathbf{U}_\mathbf{u} \\
&\quad \mathbf{L}_\mathbf{v} \le \mathbf{v} \le \mathbf{U}_\mathbf{v} \\
&\quad \boldsymbol{\theta} \text{ is random vector with density function } p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \\
&\quad \boldsymbol{\delta} \text{ is random vector with density function } p_{\boldsymbol{\delta}}(\boldsymbol{\delta})
\end{aligned}
\tag{23}
$$

The objective function is the expectation of the criterion value based on the FIM over a range of values for $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$. The first two constraints are the system equations whereas 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.

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 $a_{i,j,k}$:

$$
u_i(t) = a_{i,j,n}t^n + a_{i,j,n-1}t^{n-1} + \cdots + a_{i,j,1}t + a_{i,j,0}, \quad t \in T_j, \tag{24}
$$

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 $\mathbf{u}$

$$
\mathbf{u} = [a_{1,1,1} \quad \cdots \quad a_{i,j,k} \quad \cdots]^T. \tag{25}
$$

is used to denote the coefficients parameterizing the input variables and will replace $\mathbf{u}(t)$ in the optimization problem given by Eq. 23.

After parameterization of the input variables, the optimization problem results in a MINLP. These types of optimization problem are generally not trivial to solve. Furthermore, two additional aspects have to be taken into account that increases 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 on 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,[36] UOBYQA,[37] SNOPT,[38] or IPOPT[39] 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 find 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.[40]

Since a stochastic optimization method is used for determining the continuous variables in this work, it is sensible to also use a stochastic technique for determining the discrete variables. GA will be used 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 $\mathbf{z}$ is a vector of binary variables, no reformulation is required for determining $\mathbf{z}$ via a GA.

A hybrid method combining genetic algorithm and simultaneous perturbation stochastic approximation

A hybrid heuristic method, which integrates GA and SPSA is developed in this subsection. A GA is used to update the

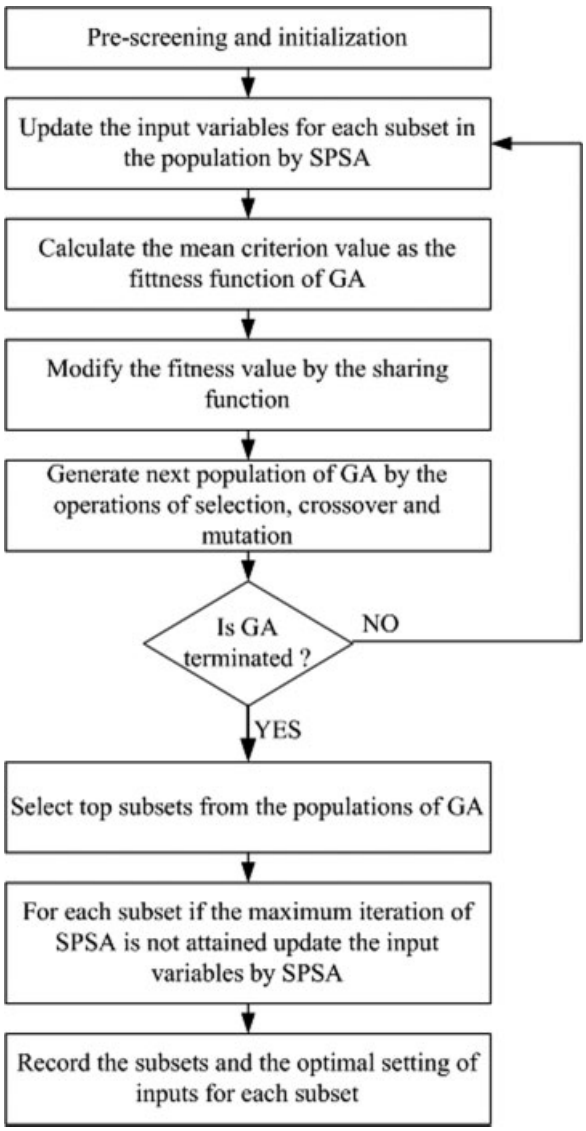

Figure 1. The procedure for integrating parameter set selection with experimental design.

discrete decision variable **z** whereas 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 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 Figure 1.

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. The criterion values are evaluated at two different sampling points of the parameters for each step of SPSA. 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. Because of 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. However, 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 on 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 Figure 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 para-

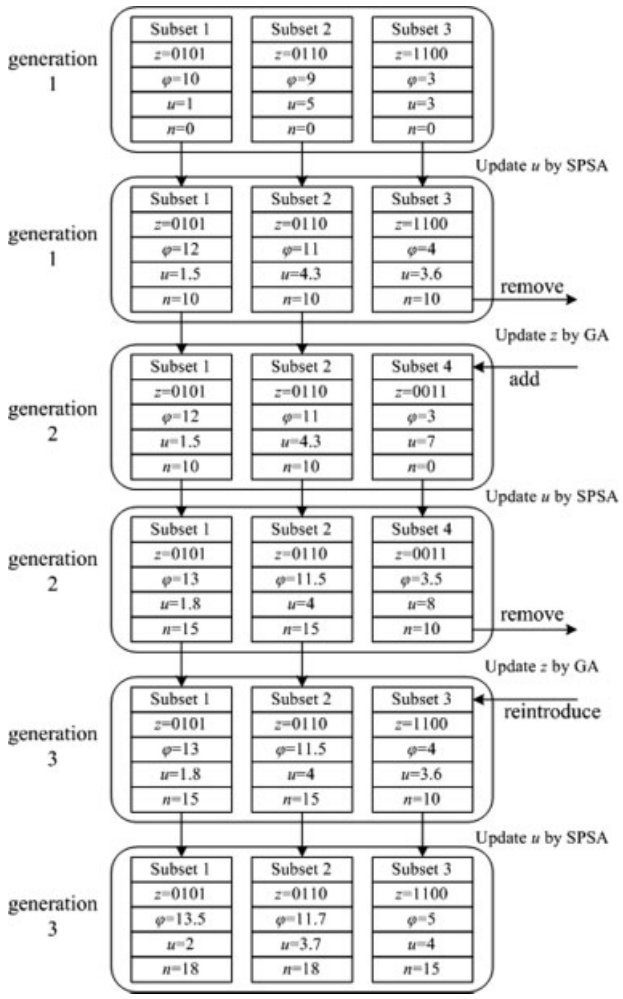

Figure 2. Illustration of steps of algorithm by using an example problem.

meter set in the 1st generation. However, as the input variables approach their optimal values the number of iterations can be reduced to decrease the computational effort. Only five iterations of SPSA are performed for parameter sets 1 and 2 in the 2nd generation of the GA, as the input trajectories of these parameter sets have already been updated once. Since parameter set 4 is first introduced in the 2nd 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 three. 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. The number of iterations for SPSA is predetermined in this illustrative example, as this is the basic implementation of the algorithm. However, the number can be adapted depending upon changes in the objective function value. For example, if there is a large improvement in the value of the objective function in previous iterations then the number of iterations for SPSA for the parameter sets in current population can remain the same. Similarly, the number of iterations can be reduced if the value of the objective function has changed only by minor amounts during the last few iterations. The total number of SPSA for a parameter set can also be determined by the convergence properties of the value of the objective function.

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) whereas 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[41]

$$\mathrm{A} \rightarrow \mathrm{B}, R_{\mathrm{A}} = k \exp(-E/RT) c_{\mathrm{A}}. \tag{26}$$

The reactor is described by the following differential equations

$$\begin{aligned} \dot{c}_{\mathrm{A}} &= \frac{F}{V}\left(c_{\mathrm{A}}^{f} - c_{\mathrm{A}}\right) - R_{\mathrm{A}} \\ \dot{T} &= \frac{F}{V}\left(T^{f} - T\right) + \frac{\Delta H}{C_{\mathrm{P}}} R_{\mathrm{A}} - \frac{hA}{C_{\mathrm{P}} V}(T - T_{\mathrm{c}}). \\ \dot{T}_{\mathrm{c}} &= \frac{F_{\mathrm{c}}}{V_{\mathrm{c}}}\left(T_{\mathrm{c}}^{f} - T_{\mathrm{c}}\right) + \frac{hA}{C_{\mathrm{P_c}} V_{\mathrm{c}}} \end{aligned} \tag{27}$$

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 five categories listed in Table 2. The first nine variables shown in Table 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.

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 h and it is assumed that the manipulated input variables can be changed every hour.

Determine the Number of Parameters to be Estimated

A singular value decomposition of the FIM is computed for the nominal values of the parameters and a predetermined input profile to determine the number of parameters to be

Table 2. Nominal Values of Variables in the CSTR Model

Parameter	Variable	Nominal Value	Type	Symbol
Feed temperature	T^f	20°C	Parameters	θ
Feed concentration	c_A^f	2500 mol/m^3		
Fluid heat capacity	C_P	1600 kJ/m^3 °C		
Heat of reaction	ΔH	160 kJ/mol		
Activation energy	E/R	255 K		
Preexponential factor	k	2.5 h^{-1}		
Coolant inlet temperature	T_c^f	10°C		
Coolant heat capacity	C_{P_c}	1200 kJ/m^3 °C		
Heat transfer coefficient	h	1100 W/m^2 °C		
Initial concentration	c_{A0}	1000 mol/m^3	Unknown factor	δ
Feed flow rate	F	0.1 m^3/h	Time variant input variables	u(t)
Coolant flow rate	F_c	0.15 m^3/h		
Initial coolant temperature	T_{c0}	30°C	Time invariant input variables	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		

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 ($n_z = 2$).

Determine Parameter Sets for Estimation

An exhaustive search over all parameter sets can be performed, and a GA 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: $\alpha = 1$, $\gamma = 1/6$, $a = 1$, $c = 0.2$, $A = 100$. 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 min 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 Figure 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 3. Column 3 shows the optimal mean criterion values calculated according to Eq. 21. Column 4 contains the mean criterion values calculated according to Eq. 20 and Column 5 shows the nominal criterion values calculated according to Eq. 19. The nominal inputs for all results in Columns 4 and 5 are those shown in Figure 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.

Application of the procedure to a signal transduction network

A model of the JAK/STAT signaling pathway[42] is used in this subsection to illustrate the techniques for parameter selection and experimental design for complex dynamic systems. Figure 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 whereas the output is the concentration of STAT1n*-

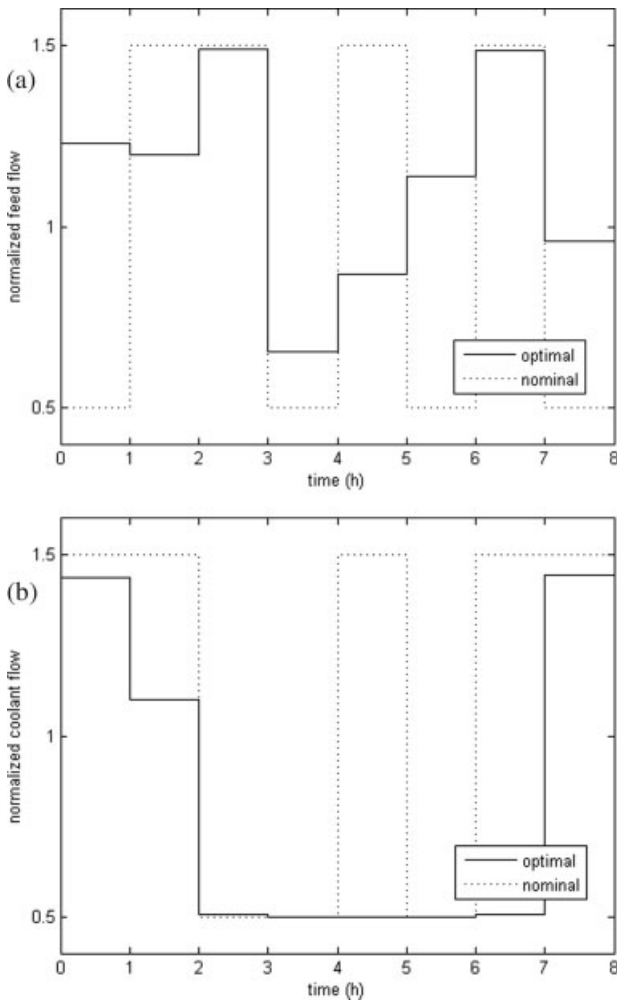

Figure 3. The input trajectories for the (a) feed flow rate F, and (b) coolant flow rate F_c.

Table 3. Criterion Values for All Parameter Sets Consisting of Two Parameters

No.	Subsets	Optimal Mean *D*-Criterion	Mean *D*-Criterion	Norminal *D*-Criterion
1	C_{P_c}, h	6.70	6.06	5.20
2	T_c^f, h	6.46	6.11	5.78
3	ΔH, h	6.26	5.29	4.98
4	T_c^f, C_{P_c}	6.25	5.75	5.67
5	ΔH, C_{P_c}	6.20	5.09	4.92
6	c_A^f, C_{P_c}	5.99	4.51	4.61
7	c_A^f, h	5.94	4.78	4.62
8	C_P, C_{P_c}	5.86	4.57	3.79
9	C_P, h	5.57	4.32	3.43
10	k, C_{P_c}	5.42	4.27	4.18
11	k, h	5.26	4.16	4.03
12	E/R, C_{P_c}	4.74	3.68	3.83
13	E/R, h	4.45	3.56	3.68
14	C_P, T_c^f	2.91	2.18	1.57
15	T^f, C_{P_c}	2.40	1.45	1.34
16	T^f, h	2.34	1.51	1.24
17	ΔH, T_c^f	2.12	1.34	1.05
18	c_A^f, T_c^f	1.94	1.24	1.37
19	C_P, ΔH	1.84	1.14	0.52
20	k, T_c^f	1.82	1.25	1.37
21	c_A^f, C_P	1.74	0.74	0.28
22	E/R, T_c^f	1.12	0.62	1.00
23	C_P, k	0.55	−0.53	−1.16
24	c_A^f, k	0.27	0.01	0.22
25	c_A^f, 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_A^f, E/R	−0.40	−0.58	−0.14
29	ΔH, E/R	−0.92	−0.98	−0.63
30	T^f, T_c^f	−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_A^f	−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

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.

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 h and the input can be changed every 30 min and remains constant in between the changes.

Determine the Number of Parameters to be Estimated

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 ($n_z = 4$).

Algorithm Parameters

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: $\alpha = 1$, $\gamma = 1/6$, $a = 0.7$, $c = 0.2$, $A = 100$.

Computing the Optimal Input Trajectories via Genetic Algorithm/Simultaneous Perturbation Stochastic Approximation

The input sequence is assumed to consist of 16 values as the variable can be changed every 30 min over a time horizon of 8 h. Determining an input sequence with 16 changes for a problem where a stochastic optimization method is

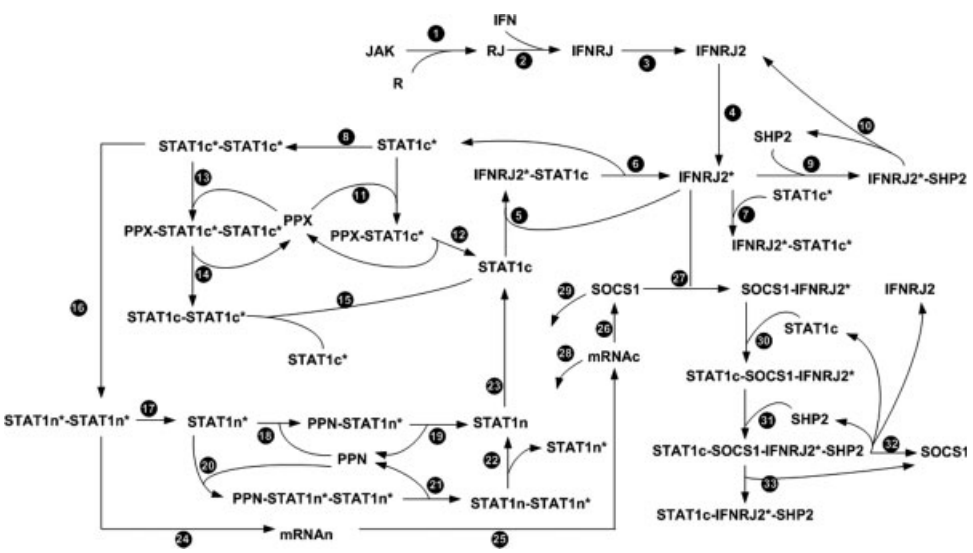

Figure 4. Structure of the JAK/STAT signaling pathway.

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. 23:

$$\mathbf{z}^*, \mathbf{u}^* = \arg \max_{\mathbf{z},\mathbf{u}} \mathop{\mathrm{E}}_{\boldsymbol{\theta},\delta}[\varphi(\mathbf{F}(\mathbf{z}, \mathbf{u}, \boldsymbol{\theta}, \delta))] - \lambda_u \|\mathbf{u} - \bar{\mathbf{u}}\|^2, \quad (28)$$

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. Figure 5 shows the mean input values and their standard deviations for one parameter set where the problem was solved 10 times.

It required approximately 28 h 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 4 and the corresponding optimal experimental conditions for the sets.

Discussion of the Results

Table 4 lists the 30 parameter sets with the optimal mean criterion value for their respective experimental conditions (Eq. 21) shown in the third column. For comparison purposes, the mean criterion value (Eq. 20) at the nominal operating conditions (Column 4) and the nominal criterion value (Eq. 19) 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, three of the parameters selected for this set by the nominal *D*-criterion are different from the parameters chosen by the optimal mean *D*-criterion.

Another important observation is that the criterion value changes significantly, when the experimental conditions are optimized for a chosen parameter set. This is particularly important, so far 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 26th by the mean criterion value. It is also noted that this parameter set only has the 23rd 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.

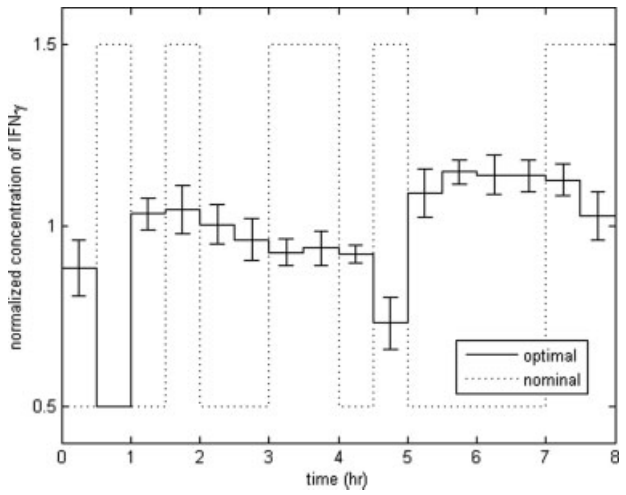

Figure 5. Averaged input signal and error bars for 10 solutions of the algorithm.

Table 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

While it may seem trivial to determine if a set is the best or the 26th 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 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. 23. 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.

Conclusions

This article 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 so far 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 FIM is maximized by varying the parameters to be estimated and the experimental conditions. This optimization problem is a MINLP, which is nontrivial to solve. A hybrid method combining a GA and a SPSA 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 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.

Acknowledgment

The authors gratefully acknowledge financial support from the ACS Petroleum Fund (Grant PRF 48144-A(9)).

Literature Cited

1. Walter E, Pronzato L. *Identification of Parametric Models from Experimental Data*. Berlin: Springer, 1997.
2. Ljung L. *System Identification: Theory for the User*, 2nd ed. Upper Saddle River: Prentice Hall PTR, 1999.
3. Nelles O. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Berlin: Springer, 2001.
4. Walter E, Pronzato L. Qualitative and quantitative experiment design for phenomenological models — a survey. *Automatica*. 1990;26:195–213.
5. Brun R, Reichert P, Kunsch HR. Practical identifiability analysis of large environmental simulation models. *Water Resour Res*. 2001;37: 1015–1030.
6. Velez-Reyes M, Verghese GC. 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. 1995:991–997.
7. Sandink CA, McAuley KB, McLellan PJ. Selection of parameters for updating in on-line models. *Ind Eng Chem Res*. 2001;40:3936–3950.
8. Yao KZ, Shaw BM, Kou B, McAuley KB, Bacon DW. Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimability and experimental design. *Polymer React Eng*. 2003;11:563–588.
9. Li RJ, Henson MA, Kurtz MJ. Selection of model parameters for off-line parameter estimation. *IEEE Trans Contr Syst Tech*. 2004;12:402–412.
10. Weijers SR, Vanrolleghem PA. A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data. *Water Sci Tech*. 1997;36:69–79.
11. Brun R, Kuhni M, Siegrist H, Gujer W, Reichert P. Practical identifiability of ASM2d parameters — systematic selection and tuning of parameter subsets. *Water Res*. 2002;36:4113–4127.
12. Reichert P, Vanrolleghem P. Identifiability and uncertainty analysis of the River Water Quality Model No. 1 (RWQM1). *Water Sci Tech*. 2001;43:329–338.
13. Chu Y, Hahn J. Parameter set selection for estimation for nonlinear dynamic systems. *AIChE J*. 2007;53:2858–2870.
14. Anh DT, Bonnet MP, Vachaud G, Minh CV, Prieur N, Duc LV, Anh LL. Biochemical modeling of the Nhue river (Hanoi, Vietnam): practical identifiability analysis and parameters estimation. *Ecol Model*. 2006;193:182–204.
15. Hiskens IA. Nonlinear dynamic model evaluation from disturbance measurements. *IEEE Trans Power Syst*. 2001;16:702–710.
16. Bastogne T, Thomassin M, Masse J. Selection and identification of physical parameters from passive observation: application to a winding process. *Contr Eng Pract*. 2007;15:1051–1061.
17. Kou B, McAuley KB, Hsu CC, Bacon DW, Yao KZ. Mathematical model and parameter estimation for gas-phase ethylene homopolymerization with supported metallocene catalyst. *Ind Eng Chem Res*. 2005;44:2428–2442.
18. Gadkar KG, Varner J, Doyle FJ. Model identification of signal transduction networks from data using a state regulator problem. *Syst Biol*. 2005;2:17–30.
19. Kiefer J. Optimum experimental designs. *J Roy Stat Soc B*. 1959;21: 272–319.
20. Goodwin GC, Payne RL. *Dynamic System Identification: Experiment Design and Data Analysis*. New York: Academic Press, 1977.
21. Silvey SD. *Optimal Design: An Introduction to the Theory for Parameter Estimation*. London: Chapman and Hall, 1980.
22. Pukelsheim F. *Optimal Design of Experiments*. New York: Wiley, 1993.
23. Atkinson A, Donev A, Tobias R. *Optimum Experimental Designs, with SAS*. Oxford: Oxford University Press, 2007.
24. Asprey SP, Macchietto S. Designing robust optimal dynamic experiments. *J Process Contr*. 2002;12:545–556.
25. Korkel S, Kostina E, Bock HG, Schloder JP. Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optim Meth Software*. 2004;19:327–338.
26. Rojas CR, Welsh JS, Goodwin GC, Feuer A. Robust optimal experiment design for system identification. *Automatica*. 2007;43:993–1008.
27. Robbins H, Monro S. A stochastic approximation method. *Ann Math Stat*. 1951;22:400–407.
28. Kiefer J, Wolfowitz J. Stochastic estimation of the maximum of a regression function. *Ann Math Stat*. 1952;23:462–466.
29. Spall JC. Multivariate stochastic-approximation using a simultaneous perturbation gradient approximation. *IEEE Trans Automat Contr*. 1992;37:332–341.
30. Spall JC. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Trans Aero Electron Syst*. 1998;34:817–823.
31. Spall JC. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Hoboken: Wiley-Interscience, 2003.
32. Goldberg DE. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading: Addison-Wesley, 1989.
33. Michalewicz Z. *Genetic Algorithms + Data Structures = Evolution Programs*. Berlin: Springer-Verlag, 1994.
34. Goldberg DE, Richardson J. Genetic algorithms with sharing for multimodal function optimization. In: Grefenstette JJ, editor. *Proceedings of the 2nd International Conference on Genetic Algorithms*, Cambridge, MA. 1987:41–49.
35. Kraft D. On converting optimal control problems into nonlinear programming problems. *Comput Math Model*. 1985;15:261–280.
36. Vanderbei RJ, Shanno DF. An interior-point algorithm for nonconvex nonlinear programming more options. *Comput Optim Appl*. 1999;13:231–252.
37. Powell MJD. UOBYQA: unconstrained optimization by quadratic approximation. *Math Program*. 2002;92:555–582.
38. Gill PE, Murray W, Saunders MA. SNOPT: an SQP algorithm for large-scale constrained optimization. *SIAM J Optim*. 2002;12:979–1006.
39. Wachter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program*. 2006;106:25–57.
40. Chin DC. Comparative study of stochastic algorithms for system optimization based on gradient approximations. *IEEE Trans Syst Man Cybern B Cybern*. 1997;27:244–249.
41. Muske KR, Georgakis C. Optimal measurement system design for chemical processes. *AIChE J*. 2003;49:1488–1494.
42. Yamada S, Shiono S, Joo A, Yoshimura A. Control mechanism of JAK/STAT signal transduction pathway. *FEBS Lett*. 2003;534:190–196.

Manuscript received Jan. 21, 2008, revision received Mar. 24, 2008, and final revision received May 14, 2008.