Generalization of a Parameter Set Selection Procedure Based on Orthogonal Projections and the D-Optimality Criterion

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

DOI 10.1002/aic.12727
Published online July 22, 2011 in Wiley Online Library (wileyonlinelibrary.com).

Many models derived from first principles contain more parameters than can be reliably estimated from data. Selecting a subset of the parameters for estimation is one common approach to deal with this problem. One popular method sequentially selects parameters based on orthogonalization of the sensitivity vectors; however, it has the drawback that only one parameter is added at each step of the iteration and that no correlations of not yet chosen parameters can be taken into account. To address this drawback, a generalization of the parameter set selection procedure based on orthogonalization is presented in this work. The procedure can add any number of parameters at each iteration such that correlations among the parameters that will be added to the set of estimated parameters can be taken into account. It is shown that two existing parameter set selection techniques form special cases of the presented method. © 2011 American Institute of Chemical Engineers *AIChE J,* 58: 2085–2096, 2012

Keywords: parameter selection, parameter estimation, dynamic systems, optimal experimental design, D-optimality, biochemical pathways, orthogonal projection

Introduction

Mathematical modeling has become a common practice for monitoring, control, and optimization of many chemical processes. The performance of an application is not only dependent on the techniques used but also on the quality of the model. The most common approach to improve model accuracy is to estimate model parameters from available data. However, many first-principles-based models include more parameters than can be accurately estimated from data. This is due to correlations among the effects that parameters have on the outputs but also because measurements may be noisy and only available for certain operating conditions. One approach to deal with such a problem is to select a subset of parameters for estimation while the unselected parameters are fixed at some values, that is, usually their nominal values.

Procedures for selecting parameters for estimation require taking into account the magnitude of the effects that changes in the parameters have on the outputs, as well as the correlation among the effects. If the effect of a change in a parameter on the output is small, as measured by some norm, then it will not be possible to accurately estimate this parameter from data. Similarly, if variations in two or more parameters result in similar effects on the outputs then these effects are correlated and a change in the value of one parameter can be compensated for by changes in the values of the other parameters, that is, it is not possible to uniquely estimate the values of these parameters. In mathematical terms, these two conditions can be formulated as follows: (1) the norm of the sensitivity vectors of selected parameters should be sufficiently large; (2) no single sensitivity vector corresponding to a selected parameter can be (approximately) expressed as a linear combination of other sensitivity vectors of selected parameters. Various selection methods that are based on these two rules have been developed, including the orthogonalization method,[1,2] the collinearity index approach,[3,4] the principal component-based method,[5,6] the eigenvalue-based method,[7,8] the relative gain array-based method,[9] the hierarchical parameter clustering method,[10] and hybrid methods.[11,12] All methods simultaneously take the two selection criteria into account but differ in the way of how much weight is given to one criterion over the other as well how combinations of parameters are selected for estimation. All of these techniques have been shown to be computationally inexpensive and work well on the examples to which they have been applied; however, none of these techniques can ensure that an optimal set of parameters is chosen.

One approach to address this problem is to formulate parameter set selection as an optimization problem. This optimization problem tries to maximize an experimental criterion,[13] which measures the accuracy of the estimated parameter values, applied to the sensitivity matrix of the chosen parameters. The most commonly used experimental criterion for this approach is the *D*-optimality criterion,[14–17] which measures the volume of the confidence region for the estimated parameter values. The main advantage of optimization-based methods is that they explicitly link the estimation accuracy to the criterion value. However, the formulated optimization problem is combinatorial in nature and is nontrivial to solve for problems with a significant number of parameters to be estimated.

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

It is the goal of this article to present a new parameter set selection method which integrates several of the properties of the existing approaches. This technique combines the heuristic-based orthogonalization method with the *D*-criterion-based optimization technique. The presented method selects parameters sequentially by decomposing the complex combinatorial problem into a sequence of simple problems where it is often possible to even perform an exhaustive search of all viable alternatives. It is important to point out that each decomposed problem has the same structure as the full problem, as an orthogonal projection is applied on the sensitivity matrix to remove the parameter effect which is already covered by previously selected parameters. The main difference between the presented approach and the existing orthogonalization technique is that it is possible to add several parameters in the same iteration and thereby take correlations among the effects of the parameters to be added at this iteration into account. The number of parameters to be added at each iteration can be freely chosen. It will be shown that the orthogonalization method represents the special case where only one parameter is added at a time, while the solution of the entire optimization problem represents the other extreme where the entire parameter set to be estimated is chosen in one iteration. As the number of selected parameters is typically small compared to the total number of parameters, the computational effort of the procedure grows with the number of parameters to be added at each iteration and one should carefully consider the computational efforts as well as the estimation accuracy before making a decision on how many parameters should be added at each iteration. As the presented technique is flexible, it can be shown that it represents a generalization of the two commonly used approaches mentioned above.

The article is organized as follows. Background information about parameter selection, with special emphasis on the orthogonalization method, is provided in the next section. Then the developed method is presented and demonstrated by its application to two examples. Concluding remarks are given in the end.

Background Information Related to Parameter Set Selection

Local sensitivity analysis of dynamic systems is presented in this section as sensitivity analysis determines the effect that changes in the parameters have on the outputs, which forms the foundation for parameter selection. Then the commonly used orthogonalization method is described. The orthogonalization method sequentially selects parameters for estimation. Next, optimization-based selection methods are presented, with a specific emphasis on techniques using the *D*-optimality criterion.

Local sensitivity analysis

Local sensitivity is defined as the partial derivative of the output with respect to the parameter.[18,19] For a dynamic system, the direct differentiation method is often applied to compute the partial derivative.[20,21]

Suppose a dynamic system is described by a set of differential equations

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \theta), \text{with} \quad \mathbf{x}|_{t=0} = \mathbf{x}_0 \\ \mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t), \theta) \end{cases} \tag{1}$$

where $\mathbf{x} \in R^{n_x}$ is the state vector, $\mathbf{u} \in R^{n_u}$ is the input vector, $\mathbf{y} \in R^{n_y}$ is the output vector, and $\theta \in R^{n_\theta}$ is the parameter vector. Direct differentiation computes the local sensitivity of a parameter θ_i by solving the differential equations

$$\begin{aligned} \frac{d}{dt}\frac{\partial \mathbf{x}(t)}{\partial \theta_i} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \theta_i} + \frac{\partial \mathbf{f}}{\partial \theta_i}, \text{with} \quad \left.\frac{\partial \mathbf{x}}{\partial \theta_i}\right|_{t=0} = \mathbf{0} \\ \mathbf{s}_i(t) &= \frac{\partial \mathbf{g}}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \theta_i} + \frac{\partial \mathbf{g}}{\partial \theta_i} \end{aligned} \tag{2}$$

where the local sensitivity is

$$\mathbf{s}_i(t) = \frac{\partial \mathbf{y}(t)}{\partial \theta_i} \tag{3}$$

The local sensitivity is a function of time and is computed by solving the system equations (Eq. 1) and the sensitivity equations (Eq. 2) simultaneously. As the local sensitivity rarely has a closed-form expression, it is often discretized to facilitate its recording as well as subsequent analyses. Given a set of sampling time points $\{t_1, t_2, \cdots, t_n\}$, the sensitivity function is discretized as

$$\mathbf{s}_i = \left[\frac{\partial y_1(t_1)}{\partial \theta_i}, \cdots, \frac{\partial y_1(t_{n_t})}{\partial \theta_i}, \cdots, \frac{\partial y_{n_y}(t_1)}{\partial \theta_i}, \cdots, \frac{\partial y_{n_y}(t_{n_t})}{\partial \theta_i}\right]^{\mathrm{T}} \tag{4}$$

where $y_1, \cdots y_{n_y}$ are entries in the output vector $\mathbf{y}$, and $\mathbf{s}_i$ is the sensitivity vector of the parameter θ_θ. It should be noted that selecting sampling points will affect the values of the sensitivity vectors and, in turn, the parameter selection results. However, determining the sampling points is part of experimental design[22] which is beyond the scope of the work presented in this article. Instead, the sampling points are selected as uniformly distributed throughout an interval.

As the outputs and the parameters typically have different units, and there can be differences of orders of magnitude in their numerical values, normalization is often applied to the sensitivity value:

$$\frac{\bar{\theta}_i}{\sigma_j}\frac{\partial y_j}{\partial \theta_i} \tag{5}$$

where $\bar{\theta}_i$ is the nominal value of the parameter θ_i, and σ_j is the standard deviation of the measurement noise of the output y_j. It is commonly assumed that the measurement noise is uncorrelated and the covariance matrix is diagonal with diagonal entries σ_j. If the measurement noise is correlated, resulting in a nondiagonal covariance matrix Σ, then a Cholesky decomposition can be applied for the normalization. In this case, the covariance matrix can be decomposed as $\Sigma = \mathbf{R}\mathbf{R}^{\mathrm{T}}$, and the normalization results in multiplying the sensitivity vectors from the left with $\mathbf{R}^{-1}$. The sensitivity values are unitless after normalization, and the subsequent analysis will assume that the sensitivity values have been normalized.

After the sensitivity vectors have been computed, the sensitivity matrix is constructed as

$$\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_{n_\theta}] \tag{6}$$

The effect of parameter changes on the output can be locally approximated by

$$\Delta \mathbf{y} \approx \mathbf{S}\Delta\theta = \mathbf{s}_1\Delta\theta_1 + \mathbf{s}_2\Delta\theta_2 + \cdots + \mathbf{s}_{n_\theta}\Delta\theta_{n_\theta} \tag{7}$$

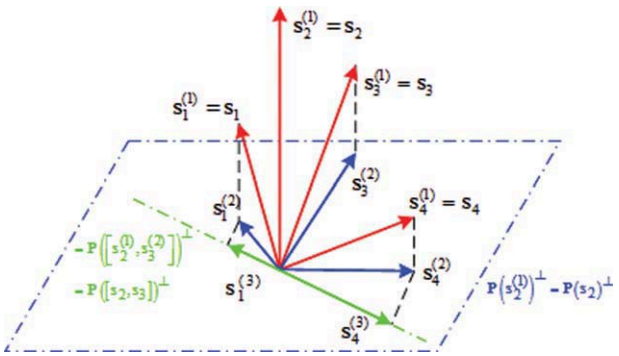

Figure 1. Illustration of the orthogonalization method via Gram–Schmidt algorithm.

Sensitivity vectors used in each iteration are marked in the same color (red for the first iteration, blue for the second iteration, and green for the third iteration). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

In this approximation, the sensitivity matrix can be viewed as the design matrix of the linear model for parameter set selection.

Orthogonalization method for parameter set selection

The orthogonalization method has been one of the most popular approaches for parameter set selection.[23–31] This technique sequentially applies orthogonalization projections of sensitivity vectors to select one parameter at a time.

Two different algorithms have been presented for performing the orthogonalization: the Gram–Schmidt algorithm[1,2] and the Householder transform.[32,33] Even though the algorithms use different approaches for the orthogonalization procedure, both orthogonalization algorithms use the same criterion to select the parameter at each iteration, and the final set of selected parameters will be identical. The algorithm based on the Householder transform is numerically more robust than the Gram–Schmidt algorithm[34]; however, the latter is easier to interpret for parameter set selection and will be used in this work. Parameter set selection based on the Gram–Schmidt algorithm involves the following steps:

Step 1: (Initiation). Calculate the sensitivity vectors $\mathbf{s}_1$,$\mathbf{s}_1$, $\cdots$, $\mathbf{s}_{n_\theta}$. Set the iteration index $k = 1$ and the projected sensitivity vectors as $\mathbf{s}_j^{(k)} = \mathbf{s}_j,\ j = 1, \cdots, n_\theta$.

Step 2: (Selection). Select the parameter, indexed by i_k, which has the largest norm of the projected sensitivity vector as

$$i_k = \arg\max_j \mathbf{s}_j^{(k)\mathrm{T}} \mathbf{s}_j^{(k)} \tag{8}$$

Step 3: (Stopping test). If $\mathbf{s}_{i_k}^{(k)\mathrm{T}} \mathbf{s}_{i_k}^{(k)} < \lambda$ (given threshold level) then stop.

Step 4: (Projection). Compute the orthogonal projection matrix $\mathbf{P}\left(\mathbf{s}_{i_k}^{(k)}\right)^{\perp}$ as

$$\mathbf{P}\left(\mathbf{s}_{i_k}^{(k)}\right)^{\perp} = \mathbf{I} - \frac{\mathbf{s}_{i_k}^{(k)} \mathbf{s}_{i_k}^{(k)\mathrm{T}}}{\mathbf{s}_{i_k}^{(k)\mathrm{T}} \mathbf{s}_{i_k}^{(k)}} \tag{9}$$

Let $\mathbf{s}_j^{(k+1)} = \mathbf{P}\left(\mathbf{s}_{i_k}^{(k)}\right)^{\perp} \mathbf{s}_j^{(k)}$, and return to Step 2 with $k = k + 1$.

The orthogonalization projection used in Step 4 is a key component of this procedure. The matrix $\mathbf{P}\left(\mathbf{s}_{i_k}^{(k)}\right)^{\perp}$ is a projection matrix and the operation $\mathbf{P}\left(\mathbf{s}_{i_k}^{(k)}\right)^{\perp} \mathbf{s}_j^{(k)}$ projects the sensitivity vector $\mathbf{s}_j^{(k)}$ onto the linear space perpendicular to the sensitivity vector $\mathbf{s}_{ik}^{(k)}$. Refer to Appendix 1 for more information about the projection matrix.

The orthogonal projection is performed at each iteration, that is, at the kth iteration, there are $k - 1$ projection matrices multiplied by a sensitivity vector:

$$\begin{aligned}\mathbf{s}_j^{(k)} &= \left(\mathbf{I} - \frac{\mathbf{s}_{i_1}^{(1)} \mathbf{s}_{i_1}^{(1)\mathrm{T}}}{\mathbf{s}_{i_1}^{(1)\mathrm{T}} \mathbf{s}_{i_1}^{(1)}}\right)\left(\mathbf{I} - \frac{\mathbf{s}_{i_2}^{(2)} \mathbf{s}_{i_2}^{(2)\mathrm{T}}}{\mathbf{s}_{i_2}^{(2)\mathrm{T}} \mathbf{s}_{i_2}^{(2)}}\right) \cdots \left(\mathbf{I} - \frac{\mathbf{s}_{i_{k-1}}^{(k-1)} \mathbf{s}_{i_{k-1}}^{(k-1)\mathrm{T}}}{\mathbf{s}_{i_{k-1}}^{(k-1)\mathrm{T}} \mathbf{s}_{i_{k-1}}^{(k-1)}}\right) \mathbf{s}_j \\ &= \mathbf{P}\left(\left[\mathbf{s}_{i_1}^{(1)}, \mathbf{s}_{i_2}^{(2)}, \cdots, \mathbf{s}_{i_{k-1}}^{(k-1)}\right]\right)^{\perp} \mathbf{s}_j \\ &= \mathbf{P}([\mathbf{s}_{i_1}, \mathbf{s}_{i_2}, \cdots, \mathbf{s}_{i_{k-1}}])^{\perp} \mathbf{s}_j\end{aligned} \tag{10}$$

The equality in the last line of Eq. 9 holds as the subspace spanned by the projected sensitivity vectors $\mathbf{s}_{i_1}^{(1)}, \mathbf{s}_{i_2}^{(2)}, \cdots, \mathbf{s}_{i_{k-1}}^{(k-1)}$ is the same as the one spanned by the original sensitivity vectors $\mathbf{s}_{i_1}$,$\mathbf{s}_{i_2}$,$\cdots$ $\mathbf{s}_{i_{k-2}}$ for Gram–Schmidt orthogonalization. This relationship is also true for the orthogonal complement subspace. Therefore, the criterion used in Step 2 is the norm of the projected sensitivity vector, which denotes the effect of parameters not covered by all previously selected parameters.

The number of selected parameters is also an important factor in the algorithm, which is determined by the threshold value λ in Step 3. Because of the sequential selection mechanism, the parameters are added one by one, that is, it is relatively simple to determine the number of selected parameters via a trial-and-error procedure.

Figure 1 illustrates the orthogonalization method. For illustration purposes, the sensitivity vectors are assumed to be in a Euclidean space, that is, $\mathbf{s}_i \in R^3$. Four sensitivity vectors, $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$, and $\mathbf{s}_4$, are displayed, corresponding to the parameters, θ_1, θ_2, θ_3, and θ_4, respectively. The first selected parameter is the one with the longest sensitivity vector, that is, 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. Among the projected sensitivity vectors, the one with the largest norm is found, which is $\mathbf{s}_3^{(2)}$. The parameter selected at the second step is θ_3. The projection is applied again and the remaining sensitivity vectors are projected onto the line perpendicular to the sensitivity vector $\mathbf{s}_3^{(2)}$. It can be seen from the figure that the line which is perpendicular to the projected sensitivity vectors $\mathbf{s}_2^{(1)}$ and $\mathbf{s}_3^{(2)}$ is also orthogonal to the original sensitivity vectors $\mathbf{s}_2$ and $\mathbf{s}_3$. This demonstrates the relationship shown in Eq. 10. At the third step, the parameter θ_4 is selected as its projected sensitivity vector is longer than that of the parameter θ_1.

The orthogonalization method is based on heuristics. While the selection criterion for each iteration is intuitive, the overall performance of the selected parameter set remains unclear due to the sequential nature of adding parameters.

Selecting sets of parameters via solution of an optimization problem

An alternative to heuristic-based methods is to formulate the parameter set selection problem as an optimization problem. This optimization problem selects m parameters out of a group of n_θ total parameters such that a criterion involving

the sensitivity matrix is maximized. The advantage of optimization-based methods is that they clearly state the criterion used for selecting parameters and that it is often possible to make predictions about the estimation accuracy of parameters based on the used criterion. It is common to use one of the experimental design criteria as part of the objective function and apply this criterion to the cross product of the sensitivity matrix.[35–37] The most commonly used criterion is the *D*-criterion resulting in the following optimization problem for parameter set selection:

$$\begin{aligned} I &= \arg\max_J \det\left(\mathbf{S}_J^{\mathrm{T}}\mathbf{S}_J\right) \\ \text{s.t.} \quad J &= \{j_1, j_2, \cdots, j_m\} \\ \mathbf{S}_J &= \mathbf{S}(:, J) \end{aligned} \tag{11}$$

After normalization of the sensitivity values with to the covariance of the measurement noise, the cross product matrix $\mathbf{S}_J^T\mathbf{S}_J$ is equal to the Fisher information matrix. The parameter set selection problem is equivalent to selecting m columns from the sensitivity matrix $\mathbf{S}$ of all parameters such that the objective function is maximized. The sensitivity matrix of selected parameters, denoted by $\mathbf{S}_J$, is constructed so that the determinant of the cross product matrix $\mathbf{S}_J^T\mathbf{S}_J$ is maximized. The set of indices of the selected columns is denoted by J and the submatrix of $\mathbf{S}(:, J)$ is defined as

$$\mathbf{S}(:, J) = \left[\mathbf{s}_{j_1}, \mathbf{s}_{j_2}, \cdots, \mathbf{s}_{j_m}\right] \tag{12}$$

It is easy to verify that a large value of the objective function can only be realized if the norm of the chosen individual parameter vectors is of significant value and also only if the sensitivity vectors are linearly independent.

The main challenge of optimization-based parameter set selection is the combinatorial nature of the optimization problem as the total number of possible parameter combinations is $n_\theta!/[m!\ (n_\theta - m)!]$. One result of this challenge is that sequential selection techniques are often used in practice as they are computationally much less expensive. This is despite the fact that optimization-based approaches can return the optimal parameter set whereas sequential selection approaches usually only result in suboptimal parameter sets.

Generalization of a Parameter Set Selection Procedure based on Orthogonal Projections and the *D*-Optimality Criterion

This section presents a procedure which forms an extension of both the parameter set selection via orthogonalization technique and parameter set selection via solution of an optimization problem involving the *D*-criterion. It will be shown that the orthogonalization method is a sequential approach to (suboptimally) solve the combinatorial problem given by Eq. 11. A general decomposition of the optimization problem involving the *D*-criterion is presented, which can add any number of parameters at each iteration.

Relationship between parameter set selection involving the orthogonalization method and via D-optimality-based approaches

Parameter set selection is often performed either via a heuristic-based approach such as the orthogonalization method or by formulating an optimization problem involving an experimental design criterion. While these seem to be entirely separate approaches, they can be viewed as two extreme cases of the technique presented in this work. Furthermore, there are properties of both types of approaches which should be preserved in an integrative framework. For example, the *D*-criterion provides a formal measure for evaluating the parameter set selected by the orthogonalization method and can also be used to quantify by how much estimation performance changes with each parameter added to the set. Similarly, the orthogonalization method can be used to quickly compute a suboptimal solution of the combinatorial optimization problem involving the *D*-criterion to provide a good starting point for a more sophisticated search algorithm.

In fact the orthogonalization-based method is a sequential method which solves the *D*-criterion-based approach. The key to the similarity between the two procedures is the update of the determinant when a new parameter is selected.

At the first iteration of the forward selection procedure, the determinant reduces to $\det(\mathbf{s}_j^{\mathrm{T}}\mathbf{s}_j) = \mathbf{s}_j^{\mathrm{T}}\mathbf{s}_j$ which is the squared norm of the sensitivity vector. Accordingly, the sensitivity vector which maximizes the determinant in the forward selection procedure is the one with the largest norm used by the orthogonalization method, that is, both methods select the same parameter at the first step.

At an intermediate iteration ($2 \le k \le m$) of the forward selection procedure, the sensitivity matrix of the previously selected parameters is assumed to be $\mathbf{S}_{k-1} = [\mathbf{s}_{i_1}, \mathbf{s}_{i_2}, \cdots, \mathbf{s}_{i_{k-1}}]$ and a new sensitivity vector, assumed to be $\mathbf{s}_j$, is added. Then the new sensitivity matrix is $\mathbf{S}_k = [\mathbf{S}_{k-1}\ \mathbf{s}_j]$ and the *D*-criterion of the augmented matrix is

$$\begin{aligned} &\det\left(\mathbf{S}_k^{\mathrm{T}}\mathbf{S}_k\right) \\ &= \det\left(\begin{bmatrix} \mathbf{S}_{k-1}^{\mathrm{T}}\mathbf{S}_{k-1} & \mathbf{S}_{k-1}^{\mathrm{T}}\mathbf{s}_j \\ \mathbf{s}_j^{\mathrm{T}}\mathbf{S}_{k-1} & \mathbf{s}_j^{\mathrm{T}}\mathbf{s}_j \end{bmatrix}\right) \\ &= \det\left(\mathbf{S}_{k-1}^{\mathrm{T}}\mathbf{S}_{k-1}\right)\left(\mathbf{s}_j^{\mathrm{T}}\mathbf{s}_j - \mathbf{s}_j^{\mathrm{T}}\mathbf{S}_{k-1}\left(\mathbf{S}_{k-1}^{\mathrm{T}}\mathbf{S}_{k-1}\right)^{-1}\mathbf{S}_{k-1}^{\mathrm{T}}\mathbf{s}_j\right) \end{aligned} \tag{13}$$

where the matrix of $\mathbf{s}_j^{\mathrm{T}}\mathbf{s}_j - \mathbf{s}_j^{\mathrm{T}}\mathbf{S}_{k-1}\left(\mathbf{S}_{k-1}^{\mathrm{T}}\mathbf{S}_{k-1}\right)^{-1}\mathbf{S}_{k-1}^{\mathrm{T}}\mathbf{s}_j$ is the Schur complement of $\mathbf{s}_j^{\mathrm{T}}\mathbf{s}_j$. Please refer to Appendix 2 for more information about the Schur complement as it is used in this procedure.

The determinant is decomposed into the product of two terms by computing the Schur complement. The first term is the determinant of the previous sensitivity matrix which is unaffected by adding a new sensitivity vector; however, the second term is affected by the selection of the new sensitivity vector. Forward selection determines a parameter to be added such that this additional parameter will maximize the determinant, that is, the parameter selected by forward selection is the one with the largest value of the Schur complement:

$$\begin{aligned} &\mathbf{s}_j^{\mathrm{T}}\mathbf{s}_j - \mathbf{s}_j^{\mathrm{T}}\mathbf{S}_{k-1}\left(\mathbf{S}_{k-1}^{\mathrm{T}}\mathbf{S}_{k-1}\right)^{-1}\mathbf{S}_{k-1}^{\mathrm{T}}\mathbf{s}_j \\ &= \mathbf{s}_j^{\mathrm{T}}\left(\mathbf{I} - \mathbf{S}_{k-1}\left(\mathbf{S}_{k-1}^{\mathrm{T}}\mathbf{S}_{k-1}\right)^{-1}\mathbf{S}_{k-1}^{\mathrm{T}}\right)\mathbf{s}_j \\ &= \mathbf{s}_j^{\mathrm{T}}\mathbf{P}(\mathbf{S}_{k-1})^{\perp}\mathbf{s}_j \\ &= \left\|\mathbf{P}(\mathbf{S}_{k-1})^{\perp}\mathbf{s}_j\right\|_2^2 \end{aligned} \tag{14}$$

where the projection matrix $\mathbf{P}(\mathbf{S}_{k-1})^{\perp}$ projects the sensitivity vector $\mathbf{s}_j$ onto the subspace spanned by the columns of $\mathbf{S}_{k-1}$, which is equal to

$$\mathbf{P}(\mathbf{S}_{k-1})^{\perp} = \mathbf{P}([\mathbf{s}_{i_1}, \mathbf{s}_{i_2}, \cdots, \mathbf{s}_{i_{k-1}}])^{\perp} \tag{15}$$

Combining this expression with Eq. 10 results in

$$\mathbf{P}(\mathbf{S}_{k-1})^{\perp}\mathbf{s}_j = \mathbf{s}_j^{(k)} \quad \text{and} \quad \left\|\mathbf{P}(\mathbf{S}_{k-1})^{\perp}\mathbf{s}_j\right\|_2^2 = \mathbf{s}_j^{(k)^{\mathrm{T}}}\mathbf{s}_j^{(k)} \tag{16}$$

It can be seen that the squared norm as shown in Eq. 15 is the parameter selection criterion shown in Eq. 7, that is, in Step 2 of the orthogonalization method. Therefore, the criterion used for forward selection and the one used for the orthogonalization method compute the same parameters to be added at each step.

Generalized orthogonalization method by selecting multiple parameters at a time

The orthogonalization-based method is a computationally inexpensive approach for selecting parameters for estimation. However, parameter selection based on orthogonalization cannot ensure that the selected parameter set is the optimal one as only one parameter is selected at a time. On the other hand, determining the exact solution of the parameter set selection problem by solving the optimization problem is computationally prohibitive for medium to large-scale problems. Therefore, it is desirable to develop a method which can return better results than those produced by the orthogonalization-based approach, yet whose computational requirements are less than if the optimization problem is solved. This section develops a generalized orthogonalization-based method which has the described properties.

One of the main drawbacks of parameter selection involving the orthogonalization method is that only one parameter is selected at each iteration, that is, the technique may miss that a combination of two parameters may result in more accurate estimates than can be achieved if two parameters are chosen sequentially. The generalized orthogonalization method presented in this section addresses this point as it allows to add any number of parameters at each iteration. In the extreme case where only one parameter is added at each iteration, the technique reduces to the existing orthogonalization method. Similarly, in the extreme case where the entire parameter set is selected in one iteration, the method reduces to the existing optimization-based technique. The advantage of the presented technique is that it inherits many of the properties of the existing approaches, while it allows to choose the computational requirement based on the number of parameters selected at each iteration.

The key to the generalized orthogonalization method is the Schur complement as it is directly involved in decomposition of the determinant resulting from the optimization problem but also plays an important role for the orthogonal projection. The general case is considered next where any number of parameters can be added to a set of previously selected parameters. Suppose the sensitivity matrix of already selected parameters is $\mathbf{S}_1$, and the sensitivity matrix of parameters added at this iteration is $\mathbf{S}_2$. Then the augmented sensitivity matrix after adding the parameters is $[\mathbf{S}_1, \mathbf{S}_2]$ and the update of the determinant of the cross product matrix is

$$\begin{aligned}
&\det\left([\mathbf{S}_1,\mathbf{S}_2]^{\mathrm{T}}[\mathbf{S}_1,\mathbf{S}_2]\right)\\
&= \det\left(\begin{bmatrix}\mathbf{S}_1^{\mathrm{T}}\mathbf{S}_1 & \mathbf{S}_1^{\mathrm{T}}\mathbf{S}_2\\ \mathbf{S}_2^{\mathrm{T}}\mathbf{S}_1 & \mathbf{S}_2^{\mathrm{T}}\mathbf{S}_2\end{bmatrix}\right)\\
&= \det\left(\mathbf{S}_1^{\mathrm{T}}\mathbf{S}_1\right)\det\left(\mathbf{S}_2^{\mathrm{T}}\mathbf{S}_2 - \mathbf{S}_2^{\mathrm{T}}\mathbf{S}_1\left(\mathbf{S}_1^{\mathrm{T}}\mathbf{S}_1\right)^{-1}\mathbf{S}_1^{\mathrm{T}}\mathbf{S}_2\right)\\
&= \det\left(\mathbf{S}_1^{\mathrm{T}}\mathbf{S}_1\right)\det\left(\left(\mathbf{P}(\mathbf{S}_1)^{\perp}\mathbf{S}_2\right)^{\mathrm{T}}\left(\mathbf{P}(\mathbf{S}_1)^{\perp}\mathbf{S}_2\right)\right)
\end{aligned} \tag{17}$$

where the projection matrix is given by

$$\mathbf{P}(\mathbf{S}_1)^{\perp} = \mathbf{I} - \mathbf{S}_1\left(\mathbf{S}_1^{\mathrm{T}}\mathbf{S}_1\right)^{-1}\mathbf{S}_1^{\mathrm{T}} \tag{18}$$

The change in the determinant resulting from the addition of more parameters is given by multiplication of the original determinant value with a new term. This term is the determinant of the cross product of the projected $\mathbf{S}_2$ onto the space which is orthogonal to the one spanned by the columns in $\mathbf{S}_1$. In other words, the selection of the new parameters relies on the part of their effect which is not covered by previously selected parameters.

As several parameters can be selected at each iteration, the Schur complement is no longer a scalar and the derivation shown in Eq. 14 does not hold anymore. Instead, the cross product of the projected sensitivity matrix has to be used.

According to the determinant decomposition shown in Eq. 17, the problem of adding several parameters to a set of existing parameters is transformed into solving a *D*-optimality problem based on the projected sensitivity matrix

$$\begin{aligned}
&I = \arg\max_J \det\left([\mathbf{S}_1,\mathbf{S}_2]^{\mathrm{T}}[\mathbf{S}_1,\mathbf{S}_2]\right)\\
&\text{s.t.}\quad \mathbf{S}_2 = \mathbf{S}^{(k)}(:,J)\\
\Leftrightarrow\ &I = \arg\max_J \det\left(\left(\mathbf{P}(\mathbf{S}_1)^{\perp}\mathbf{S}_2\right)^{\mathrm{T}}\left(\mathbf{P}(\mathbf{S}_1)^{\perp}\mathbf{S}_2\right)\right)\\
&\text{s.t.}\quad \mathbf{S}_2 = \mathbf{S}^{(k)}(:,J)
\end{aligned} \tag{19}$$

where J denotes the indices of the added parameters which determines the matrix of $\mathbf{S}_2$, while $\mathbf{S}_1$ remains unchanged. A summary of the described parameter set selection procedure is given below:

Algorithm: Generalized orthogonalization method by selecting any number of parameters at a time

Assume a sequence of numbers of selected parameters at each iteration is given by $M = \{m_1, m_2, m_3, \cdots\}$

Step 1: (Initiation). Compute the sensitivity matrix $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_{n_\theta}]$. Set the iteration index $k = 1$ and the projected sensitivity matrix as $\mathbf{S}^{(k)} = \mathbf{S}$.

Step 2: (Selection). Select the set of parameters, indexed by $I^{(k)} = \left\{i_1^{(k)}, i_2^{(k)}, \cdots, i_{m_k}^{(k)}\right\}$, which has the largest determinant computed from the projected sensitivity matrix as

$$\begin{aligned}
&I^{(k)} = \arg\max_J \det\left(\mathbf{S}_J^{\mathrm{T}}\mathbf{S}_J\right)\\
&\text{s.t.}\quad \mathbf{S}_J = \mathbf{S}^{(k)}(:,J)\\
&\qquad J = \{j_1, j_2, , j_{m_k}\}
\end{aligned} \tag{20}$$

Step 3: (Stopping test). If $\det\left(\mathbf{S}_{I^{(k)}}^{\mathrm{T}}\mathbf{S}_{I^{(k)}}\right) < \lambda$ where $\mathbf{S}_{I^{(k)}} = \mathbf{S}^{(k)}\left(:, I^{(k)}\right)$ and λ is a given threshold level, then stop.

Step 4: (Projection). Compute the orthogonal projection matrix $\mathbf{P}(\mathbf{S}_{I^{(k)}})^{\perp}$ as

$$\mathbf{P}(\mathbf{S}_{I^{(k)}})^{\perp} = \mathbf{I} - \mathbf{S}_{I^{(k)}}\left(\mathbf{S}_{I^{(k)}}^{\mathrm{T}}\mathbf{S}_{I^{(k)}}\right)^{-1}\mathbf{S}_{I^{(k)}}^{\mathrm{T}} \tag{21}$$

Let $\mathbf{S}^{(k+1)} = \mathbf{P}(\mathbf{S}_{I^{(k)}})^{\perp}\mathbf{S}^{(k)}$, and return to Step 2 with $k = k + 1$.

The optimization problem shown in Eq. 20 determines the selected parameters, indexed by $I^{(k)}$, at each iteration. While

Step 2 uses the same formulation is the original D-optimality problem, the scale of this problem can be significantly smaller than the original one, as the problem given by Eq. 20 only needs to determine one or a few parameters at each iteration, whereas the original optimization problem determined the entire parameter set at once.

The selected parameters are the union of the sets of parameters selected at each iteration as

$$I^{(1)} \cup \cdots \cup I^{(k)} \cup \cdots = \left\{ i_1^{(1)}, i_2^{(1)}, \cdots, i_{m_1}^{(1)}, \cdots, i_1^{(k)}, i_2^{(k)}, \cdots, i_{m_k}^{(k)}, \cdots \right\} \quad (22)$$

The number of parameters selected at each iteration can be chosen based on the computational cost associated with performing the procedure.

If $m_1 = m$ and $m_{k\neq 1} = 0$, then all parameters are selected in one iteration, and the method reduces to the one where the optimization problem is solved. This selection ensures that an optimal solution is found, however, this approach is computational expensive or may even be infeasible for large problems. If all $m_k = 1$, then the method reduces to the regular orthogonalization method. In this case, the algorithm is not computationally demanding; however, the found solution may be far from the optimal one.

Selecting several parameters at each iteration is one approach to improve the solution. If only one parameter is selected at each iteration then only the relationship of this parameter to the already selected ones is considered. However, if several parameters are selected at each iteration then interactions among these parameters are taken into account as well as the relationship of this parameter set to the already selected parameters. This comes at the cost of having to search through all potential parameter combinations that could be added at each iteration, though. At the same time, selecting several parameters at each iteration has the potential to improve the quality of the selected parameter sets.

The optimization problem given by Eq. 20 is solved at each iteration of the procedure. Apart from selecting several parameters at each iteration, another procedure which can improve the quality of the solution is to conduct the search at each iteration from different starting sets. Specifically, the algorithm records not only the optimal solution at each iteration but also several potential solutions, that is, n_k solutions will be recorded at the kth iteration. Each parameter set is then the starting set to which new parameters are added in the next iteration. During the next iteration, the algorithm will solve the optimization problem given by Eq. 20 repeatedly, once for each of the already recorded parameter sets. Selecting multiple parameter sets extends the search region of the sequential algorithm and increases the likelihood of determining the optimal solution.

A diagram showing the procedural steps of the generalized orthogonalization method by selecting multiple parameter sets with any number of parameters at each iteration is shown in Figure 2. The first step in this procedure is to compute the sensitivity matrix of all parameters. At the kth iteration, n_k parameter sets are determined and each parameter set includes m_k parameters. The sequences of $N = \{n_1, n_2, \cdots, n_k, \cdots\}$ and $M = \{m_1, m_2, \cdots, m_k, \cdots\}$ determine the specific structure of the search tree. The values chosen for all n_i and m_j are based on the computational burden appropriate for a particular problem. A more detailed discussion of the choice of n_i and m_j is provided in the next section.

Another advantage of providing several good solutions rather than just the optimal solution is that it is challenging to take all factors into account for parameter set selection and that sometimes the final decision of which parameters to estimate needs to be made based on experience with the system. That being said, it is much easier to make this decision based on a handful of good parameter sets that can be chosen rather than determining the entire set of parameters to be estimated from experience.

Case Studies

In this section, the presented method is applied to two examples. The first one is an artificial example where 10,000 sensitivity matrices are randomly generated. This example illustrates the performance of the method for different choices of $N = \{n_1, n_2, \cdots, n_k, \cdots\}$ and $M = \{m_1, m_2, \cdots, m_k, \cdots\}$. The second example deals with a complex biochemical reaction network described by a set of ordinary differential equations.

Randomly generated sensitivity matrix

In this example, the sensitivity matrix is generated randomly. The number of data points sampled from the output is assumed to be 4 and the number of all parameters is denoted by n_θ, that is, the sensitivity matrix $\mathbf{S}$ has the dimension 4-by-n_θ. Each entry in the sensitivity matrix is randomly generated according to a standard uniform distribution in the interval [0, 1].

At most four parameters can be estimated from the 4 data points, that is, $m = 4$ for this example. Estimating fewer parameters has also been investigated as it is common that more data are available than parameters have to be estimated. However, these results are not shown here as the estimation problem becomes easier, the more data are available. Instead, the generalized orthogonalization method is applied with different numbers of parameter sets selected at each iteration, denoted by the sequence $N = \{n_1, n_2, \cdots, n_k, \cdots\}$, and different numbers of selected parameters in each parameter set, denoted by the sequence $M = \{m_1, m_2, \cdots, m_k, \cdots\}$.

First, the sequence N is set to $N = \{1,1, \cdots, 1, \cdots\}$. To select four parameters, there are eight possible sequences for M: {1, 1, 1, 1}, {1, 1, 2}, {1, 2, 1}, {1, 3}, {2, 1, 1}, {2, 2}, {3, 1}, and {4}. Parameter set selection according to each sequence is applied, and 10,000 sensitivity matrices are generated randomly to test the performance of each sequence. The mean D-criterion value returned by each sequence is plotted in Figure 3a. To investigate the influence of the total number of parameters, the selection is performed for three different numbers of parameters: $n_\theta = 10, 15, 20$.

The regular orthogonalization method, which selects one parameter at a time, corresponds to the sequence $M = \{1,1,1,1\}$ while solving the optimization problem, which selects four parameters simultaneously at a time, corresponds to the sequence $M = \{4\}$. Solution of the optimization problem returns the best parameter set while other selection sequences only return a suboptimal solution. To compare the performance, the returned D-criterion value of each sequence is normalized by dividing by the optimal value.

The sequence corresponding to $M = \{1,1,1,1\}$ returns the smallest D-criterion value which is only 67% of the optimal value for the case of $n_\theta = 20$. This shows that the regular orthogonalization method may return a solution far from the optimal one and that optimality at each iteration cannot guarantee an optimal solution for selecting the entire parameter

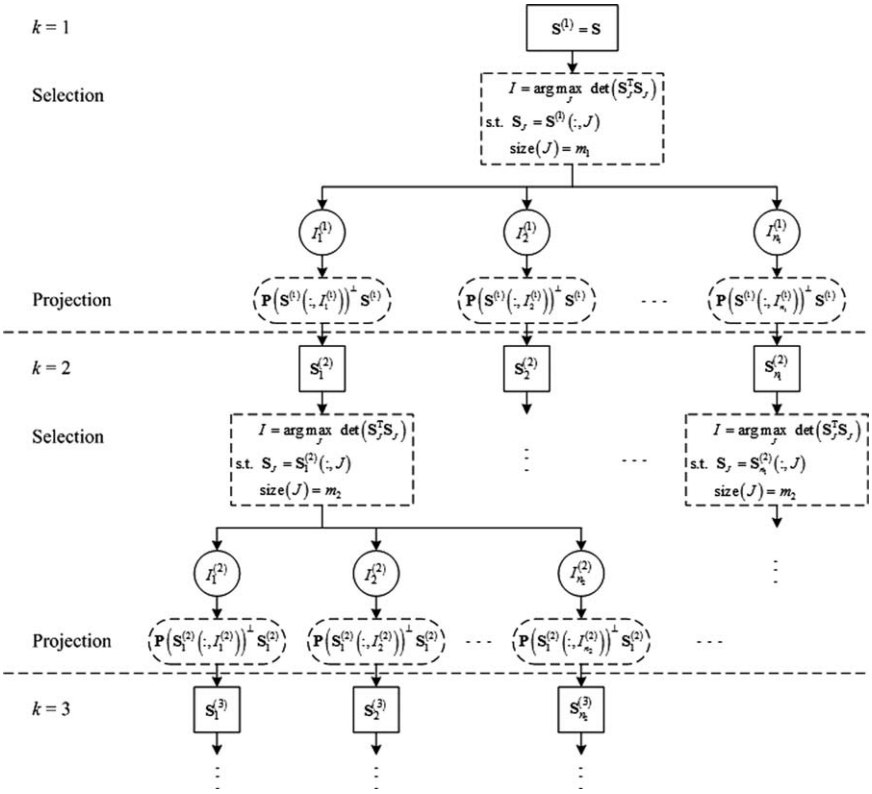

Figure 2. Flow diagram of the generalized orthogonalization method which selects several parameters at each step.

Solid circles denote the indices of selected parameters at each iteration and the solid squares denote the projected sensitivity matrices according to the selection. The dashed rectangles denote that an optimization problem needs to be solved and the dashed rectangles with rounded sides denote procedures where orthogonal projections are performed. The search tree is branched at the selection nodes (dashed rectangles) where n_k best parameter sets are selected by the exhaustive search and each set contains m_k parameters.

set. When the number of parameters increases, the performance of the regular orthogonalization method will further deteriorate, for example, the value of the D-criterion drops below 50% of the optimal value for the case of $n_\theta = 50$.

An improvement is to select several parameters at a time rather than a single one. In fact, any of the cases where more than one parameter was selected at least once return better results than the case where $M = \{1,1,1,1\}$ is shown in Figure 3a. Another observation is that the sequences which select several parameters at the first iteration perform better than those which just select a single parameter at this iteration. This observation is not surprising as selected parameters will be retained at all subsequent iterations, that is, it is especially important to determine appropriate choices for the first few parameters as they will affect the procedure at every iteration.

While selecting several parameters at a time returns larger values of the D-criterion, this comes at the price of increased computational effort. To investigate the computational burden, the number of possible parameter combinations for each selection sequence is computed and normalized by dividing by the number of possible combinations if all parameters of the set were chosen at once. Figure 3b shows a comparison of the computational effort associated with the different choices of M. It can be clearly seen that the computational effort grows with the number of parameters chosen at each iteration. Furthermore, if three parameters are chosen in one iteration, then this is computationally more demanding for this example than if two parameters are chosen at two separate iterations. If the entire parameter set is chosen in one iteration, then the number of parameter combinations increases dramatically with the number of all parameters n_θ. There are 210 possible combinations for the case of $n_\theta = 10$; however, this number increases to 4845 for the case of $n_\theta = 20$. The required computational effort should be kept in mind when the sequence $M = \{m_1, m_2, \cdots, m_k, \cdots\}$ is selected.

For this example, a good balance between the quality of the chosen parameter set and the computational effort is given by the sequences $M = \{2,1,1\}$ (or $M = \{2,2\}$). In the case of $n_\theta = 20$, the criterion computed for $M = \{2,1,1\}$ has 81% of the optimal value which is much larger than 67% computed for adding one parameter per iteration while the computational burden associated with $M = \{2,1,1\}$ is only 4.64% of determining the entire parameter set in one iteration. It should be noted that larger computational costs do not automatically imply better performance of the determined parameter sets. For example, the sequence $M = \{1,3\}$ requires more computations to be determined than the sequence corresponding to $M = \{2,1,1\}$ according to Figure 3b; however, the performance of the latter is better according to Figure 3a. This serves as another indicator that it is desirable to select several parameters at the first iteration.

Another approach to improve performance is to select several parameter sets at each iteration and then perform the next iteration independently for each of these parameter sets, that is, the influence of the sequence N will be investigated next. To limit the number of possible outcomes and allow

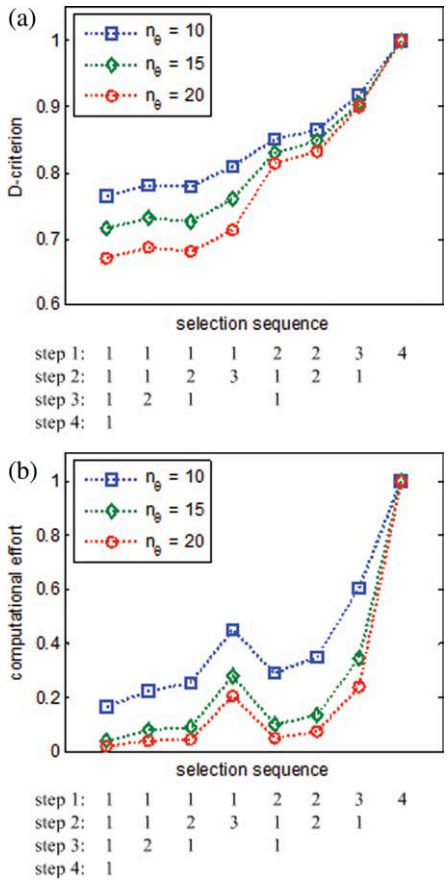

Figure 3. Results returned by the generalized orthogonalization method with different selection sequences.

(a) Normalized mean *D*-criterion value over 10,000 randomly selected sensitivity matrices. (b) Computational effort denoted by the normalized number of possible parameter combinations. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

for a comparison to be made, only the sequence $M = \{2,1,1\}$ with $n_\theta = 15$ is investigated here. In this situation, the sequence N is specified as $N = \{n_1, n_2, n_3\}$. It has been found that only the number of selected parameter sets at the first iteration, n_1, has a major influence and that setting $n_2 = n_3 = 1$ will only have a negligible impact on the discussion.

All possible solutions of the combinatorial problem at the first iteration are evaluated and ordered by their criterion values. The top n_1 parameter sets are recorded, and the selection performed at the next iteration is independently performed for each parameter set identified as one of the top n_1 sets from the first iteration. In a final step, a collection of n_1 parameter sets is returned, and the best parameter set in the collection is identified.

The criterion values returned by this procedure are shown in Figure 4a, and the corresponding computational efforts can be found in Figure 4b. Similar to Figure 3, both *D*-criterion values and the computational efforts are normalized by the values for computing the entire parameter set in one iteration. It can be seen that when two parameter sets are selected in the first iteration that the criterion value increases significantly from 82.4%, for the case of only further investigating the best parameter set identified in the first iteration, to 90.9%. The computational effort after the first iteration is doubled since twice the number of cases has to be evaluated for the remaining iterations. It can easily be shown that the computational effort for the remaining steps scales linearly with the number of parameter sets retained in the first iteration; however, the improvements made in the criterion value become less as more sets are investigated. A good compromise for this example is to choose $n_1 = 3$ which returns an adequately good solution (94.4%) with a reasonable computational effort (28.6%).

This example demonstrates the performance of the generalized orthogonalization method for different settings of $M = \{m_1, m_2, \cdots, m_k, \cdots\}$ and $N = \{n_1, n_2, \cdots, n_k, \cdots\}$. Selecting several parameters in each iteration and retaining several of the best solutions, instead of just the best one, are two approaches to balance the computational effort and the quality of the solution. Because of the mechanisms involved in sequential selection, the settings at the first iteration have a more significant impact on the outcome than changes made at later iterations. It should be noted that these observations were made for this specific example and that different conclusions could be drawn for different applications.

Complex biochemical reaction network

This example deals with a model of a signal transduction pathway. Parameter estimation is a key component for modeling these types of systems[38]; however, a biochemical reaction network usually contains a large number of strongly correlated parameters and experimental data are often scarce and noisy as concentration measurements inside living cells are nontrivial to obtain. It has been previously reported that it is difficult to fit the majority of parameters accurately from experimental data,[39] indicating a need for parameter set selection before parameter estimation is performed.

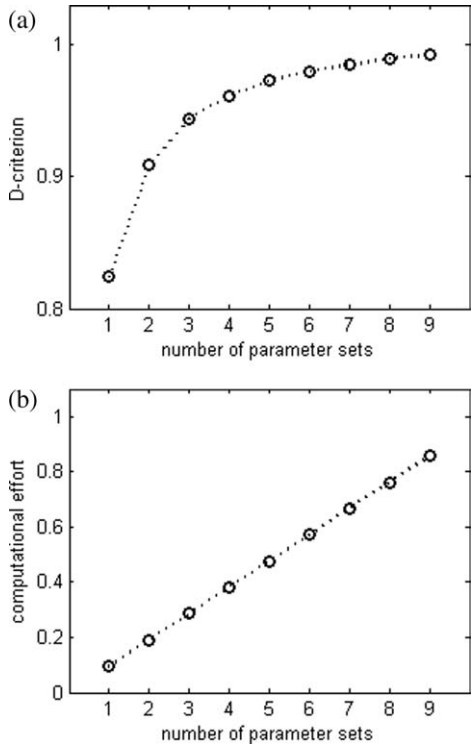

Figure 4. Results for selecting different parameter sets at the first step.

(a) normalized *D*-criterion and (b) computational effort denoted by the normalized number of possible parameter combinations.

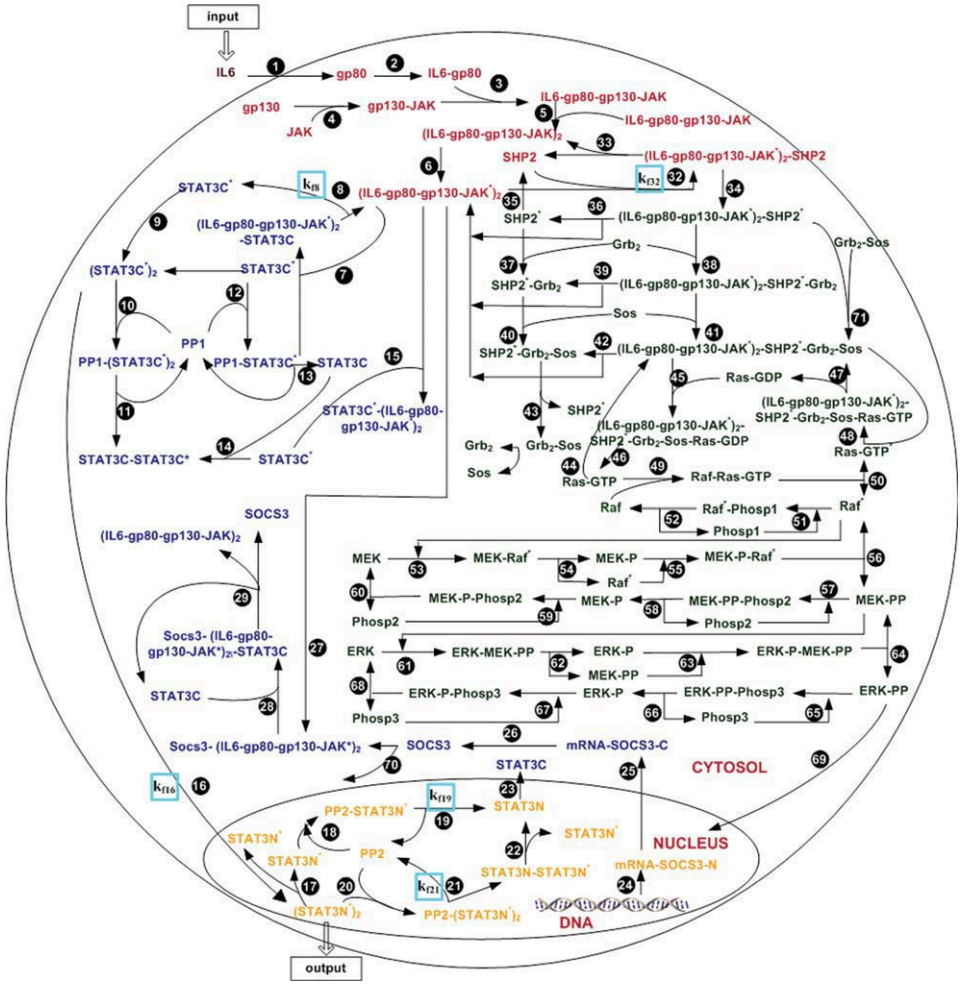

Figure 5. Structure of IL-6 signal transduction pathway.

Each reaction is numbered and the kinetic parameters are named according to the number of the involved reaction. The selected parameters in the best set {k_{f32}, k_{f8}, k_{f21}, k_{f16}, k_{f19}} are marked by a cyan rectangle. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

To illustrate the technique presented in this work, a model of a signal transduction network for hepatocytes stimulated by Interleukin-6 (IL-6) is used. This model, shown in Figure 5, describes signal transduction from the input, that is, the concentration of IL-6 outside of the cell, to the output, the concentration of the activated transcriptions factor STAT3 in the nucleus. The model is described by 66 nonlinear ordinary differential equations and includes 115 parameters. The concentrations of the proteins are the state variables and the parameters mainly describe the reaction rate constants. A detailed description of the model and the nominal values of the parameters and initial values of the states can be found in the literature.[40,41]

The model is simulated from 0 to 3 h to capture the main dynamic properties of the system. The local sensitivity is computed by the direct differentiation method. The sensitivity profiles are normalized by dividing by the nominal values of the parameters, as the parameters have different units and their nominal values vary by orders of magnitude. The calculated sensitivity profiles are sampled every 5 min, which is also the sampling time of the output.

The measurement noise is assumed to be Gaussian with zero mean and variance equal to 4 ($\sigma^2 = 4$). As the measurements are all sampled from the same output variable and the measurement noise is assumed to be uncorrelated, the normalization of the sensitivity by dividing by the standard derivation has no effect on the subsequent parameter selection problem.

The sensitivity matrix is made up of the sensitivity vectors. The sensitivity matrix is ill-conditioned and a majority of the singular values of the sensitivity matrix are close to zero, that is, most parameters are not estimable. The number of selected parameters is determined from the numerical rank of the sensitivity matrix. Five parameters will be estimated, $m = 5$, as the sum of the first five singular values is more than 99% of the sum of all singular values of the sensitivity matrix.

As the previous example has shown that it may be beneficial to add several parameters in the first step and to also retain several of the best parameter subsets, the sets M and N are chosen as follows: $M = \{3,2\}$ and $N = \{10, 5\}$. The first iteration will determine three parameters, that is, $m_1 = 3$. All combinations of three parameters are evaluated and the 10 best subsets, that is, $n_1 = 10$, are recorded. Then the second iteration will add the remaining two parameters, that is, $m_2 = 2$, to each of the 10 subsets selected in the first iteration. At this iteration, the five best subsets, that is, $n_2 = 5$, with two parameters are recorded. A total of $10 \times 5 = 50$ parameter sets are returned at the end of the procedure. The determination of the sets M and N follows the rules that a

Table 1. Selected Parameter Sets for the IL-6 Model

No.	Step 1			Step 2		Criterion (%)
1	k_{f32}	k_{f8}	k_{f21}	k_{f16}	k_{f19}	100
2	k_{f32}	k_{f8}	k_{f21}	k_{f16}	k_{f18}	91.9
3	k_{f32}	k_{f8}	k_{f21}	k_{f16}	k_{b18}	87.8
4	k_{f32}	k_{f8}	k_{f21}	k_{f6}	k_{f19}	70.6
5	k_{f32}	k_{f8}	k_{f21}	k_{f6}	k_{f18}	64.6
6	k_{f32}	k_{f8}	k_{f21}	k_{f6}	k_{b18}	61.8
7	k_{f7}	k_{f8}	k_{f21}	k_{f16}	k_{f19}	55.7
8	k_{f32}	k_{f16}	k_{f20}	k_{f8}	k_{f19}	54.1
9	k_{f7}	k_{f8}	k_{f21}	k_{f16}	k_{f18}	52.0
10	k_{f32}	k_{f16}	k_{b20}	k_{f8}	k_{f19}	51.9
11	k_{f32}	k_{f6}	k_{f21}	k_{f19}	V_{m24}	51.5
12	k_{f7}	k_{f8}	k_{f21}	k_{f16}	k_{b18}	49.6
13	k_{f32}	k_{f16}	k_{f20}	k_{f8}	k_{f18}	49.2
14	k_{f7}	k_{f6}	k_{f21}	k_{f19}	V_{m24}	48.7

large number at the initial step will be more important than a large number at the subsequent steps. Other choices of M and N are also possible; however, the results returned are very similar. For comparison purposes, the regular orthogonalization method is also applied, which is equivalent to choosing $M = \{1,1,1,1,1\}$ and $N = \{1,1,1,1,1\}$.

The parameter set returned by the regular orthogonalization method is $\{k_{f7}, k_{f6}, k_{f21}, k_{f19}, V_{m24}\}$, and this set is also included in the collection returned by the presented method. However, the presented method determines 13 parameter sets which have a criterion value larger than that of the regular orthogonalization method. Theses 14 parameter sets are listed in Table 1. The best determined parameter set is $\{k_{f32}, k_{f8}, k_{f21}, k_{f16}, k_{f19}\}$, which has only two common parameters with the set $\{k_{f7}, k_{f6}, k_{f21}, k_{f19}, V_{m24}\}$ computed by the regular orthogonalization method. From the criterion values of the different parameter sets, it can be concluded that the set computed by the regular orthogonalization method has a criterion value that is only 48.7% of the value for the best determined set. This serves as another indicator that the regular orthogonalization technique may return a parameter set which is significantly worse than the optimal set.

The sensitivity profiles of the parameters selected by the orthogonalization method are shown in Figure 6a, and the sensitivity profiles of the parameters in the best set selected by the presented method are plotted in Figure 6b. The parameter k_{f7} has the sensitivity vector with the largest norm and it is the first one selected by the regular orthogonalization method. However, selection of k_{f7} excludes the parameter combination of k_{f32} and k_{f8} even though these two parameters are key for constructing better parameter sets. For example, the best six parameter sets listed in Table 1 all include k_{f32} and k_{f8} while none of them includes k_{f7}.

To evaluate the performance of the selected parameter sets, parameter estimation is conducted. Artificial data are generated by adding noise, with zero mean and variance of $\sigma^2 = 4$, to the output value. 200 data sets have been generated and for each data set the parameter sets $\{k_{f7}, k_{f6}, k_{f21}, k_{f19}, V_{m24}\}$ and $\{k_{f32}, k_{f8}, k_{f21}, k_{f16}, k_{f19}\}$ were estimated. The estimation error is defined as the square root of the mean square error of the parameter estimate. The estimation error for estimating $\{k_{f7}, k_{f6}, k_{f21}, k_{f19}, V_{m24}\}$ is 29.0% and the error for estimating $\{k_{f32}, k_{f8}, k_{f21}, k_{f16}, k_{f19}\}$ is only 22.1%. Here, the estimation error is expressed as the percentage since it is normalized by dividing by the nominal values of the parameters. The absolute difference between the estimation errors of the two parameter sets is 6.9%, and the relatively difference is 23.8%, which indicates a considerable improvement in the estimation accuracy if the set $\{k_{f32}, k_{f8}, k_{f21}, k_{f16}, k_{f19}\}$ is estimated instead of the set $\{k_{f7}, k_{f6}, k_{f21}, k_{f19}, V_{m24}\}$. Apart from the averaged error over the 200 data sets, the distribution of the estimation error is shown in Figure 7. The histogram illustrates that estimating $\{k_{f7}, k_{f6}, k_{f21}, k_{f19}, V_{m24}\}$ returns a larger estimation error than estimating $\{k_{f32}, k_{f8}, k_{f21}, k_{f16}, k_{f19}\}$.

Conclusions

Parameter set selection is an important technique for determining which parameters should be estimated from data. This article presented a new method for parameter set selection, which includes the regular orthogonalization method and optimization of the D-criterion applied to the sensitivity matrix as special cases. The difference between the regular orthogonalization procedure and the presented technique is that it is possible to select several parameters at each iteration and not just a single parameter. The choice of which set of parameters should be added at each iteration is made based on the D-criterion. As such, if only one parameter is added at each iteration, then the technique reduces to the existing orthogonalization procedure. Similarly, if the entire parameter set is to be determined in one iteration, then the method reduces to the solution of the D-criterion-based optimization problem for parameter selection. This more generalized procedure provides some flexibility such that a good trade-off between estimation accuracy and computational effort can be found. Furthermore, the procedure allows to retain several of the best solutions at each iteration and not just the optimal one. The reason for this is that it is often not possible to determine one best set of parameters to be estimated and other criteria, such as experience with a particular system, need to be taken into account as well. However, this is only possible if several good

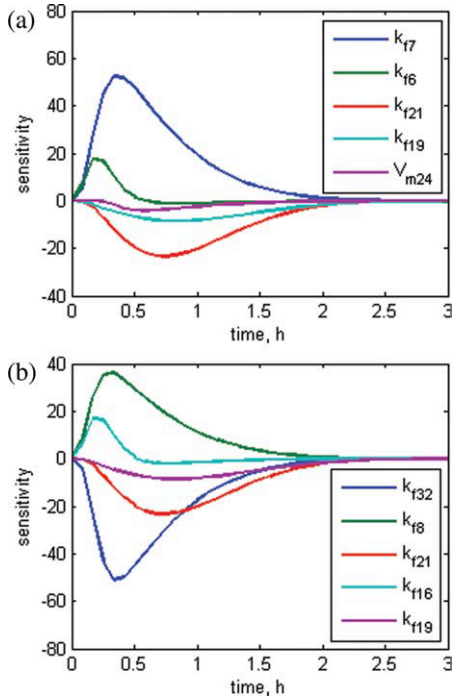

Figure 6. Sensitivity profiles of parameters selected by (a) the regular orthogonalization method and (b) the presented method.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

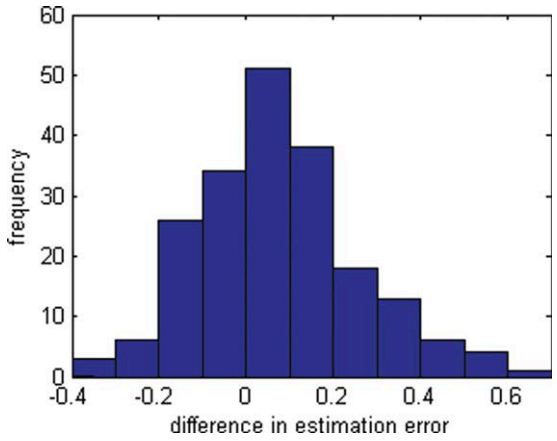

Figure 7. Distribution of the differences in the estimation errors.

The difference is the estimation error of estimating $\{k_{f7}, k_{f6}, k_{f21}, k_{f19}, V_{m24}\}$ minus that of estimating $\{k_{f32}, k_{f8}, k_{f21}, k_{f16}, k_{f19}\}$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

solutions are provided and ranked according to a criterion and not just one solution.

The method has been applied in two case studies. While the optimal solution cannot be guaranteed unless the extreme case of selecting all parameters in one iteration is considered, it was found that there is a significant advantage of the parameter sets determined by this method over the one computed from the regular orthogonalization procedure.

Notation

$I = \{i_1, i_2, \ldots, i_m\}$ = indices for selected parameters
$I^{(k)} = \{i_1^{(k)}, i_2^{(k)}, \cdots, i_{mk}^{(k)}\}$ = indices for selected parameters at the k-th iteration
$J = \{j_1, j_2, \cdots, j_m\}$ = indices for parameter combinations
m = number of selected parameters
$M = \{m_1, m_2, \cdots, m_k, \cdots\}$ = sequence of the numbers of parameters selected at each iteration in the sequential selection procedure
n_t = number of sampling points
n_u = number of input variables
n_x = number of state variables
n_y = number of output variables
n_θ = number of all parameters in the model
$N = \{n_1, n_2, \cdots, n_k, \cdots\}$ = sequence of the numbers of sets of selected parameters at each iteration in the sequential selection procedure
$\mathbf{P}([\mathbf{s}_1, \mathbf{s}_2, \cdots \mathbf{s}_k])$ = projection matrix which projects a vector onto the space orthogonal to the one spanned by the vectors of $\mathbf{s}_1, \mathbf{s}_2, \cdots \mathbf{s}_k$
$\mathbf{P}(\mathbf{S})^\perp$ = projection matrix which projects a vector onto the space orthogonal to the one spanned by the columns in $\mathbf{S}$
$\mathbf{R}$ = Cholesky factor of $\boldsymbol{\Sigma}$
$\mathbf{s}_i$ = sensitivity vector of the parameter θ_θ
$\mathbf{s}_i^{(k)}$ = projected vector of $\mathbf{s}_i$ at the k-th iteration
$\mathbf{s}_i(t)$ = time dependent sensitivity value of the parameter θ_θ
$\mathbf{S}$ = sensitivity matrix of all parameters
$\mathbf{S}_k$ = sensitivity matrix when k parameters are selected
$\mathbf{S}_J$ = sensitivity matrix of the parameters indexed by J
t_i = i-th sampling point
$\mathbf{u}$ = input vector
$\mathbf{x}$ = state vector
$\mathbf{x}_0$ = initial value of the state variables
$\mathbf{y}$ = input vector
y_i = i-th output variable
θ = parameter vector
θ_i = i-th parameter
$\bar{\theta}_i$ = nominal value of the parameter θ_i
σ_i = standard deviation of the measurements from y_i
$\boldsymbol{\Sigma}$ = variance-covariance matrix of all measurements
$\|\cdot\|_2^2$ = squared Euclidean norm

Literature Cited

1. Yao KZ, Shaw BM, Kou B, McAuley KB, Bacon DW. Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimability and experimental design. *Polym React Eng.* 2003;11:563–588.
2. Lund BF, Foss BA. Parameter ranking by orthogonalization—applied to nonlinear mechanistic models. *Automatica.* 2008;44:278–281.
3. Brun R, Reichert P, Kunsch HR. Practical identifiability analysis of large environmental simulation models. *Water Resour Res.* 2001;37:1015–1030.
4. Omlin M, Brun R, Reichert P. Biogeochemical model of Lake Zurich: sensitivity, identifiability and uncertainty analysis. *Ecol Model.* 2001;141:105–123.
5. Turanyi T. Reduction of large reaction mechanisms. *New J Chem.* 1990;14:795–803.
6. Degenring D, Froemel C, Dikta G, Takors R. Sensitivity analysis for the reduction of complex metabolism models. *J Process Control.* 2004;14:729–745.
7. Schittkowski K. Experimental design tools for ordinary and algebraic differential equations. *Ind Eng Chem Res.* 2007;46:9137–9147.
8. Quaiser T, Monningmann M. Systematic identifiability testing for unambiguous mechanistic modeling—application to JAK-STAT, MAP kinase, and NF-kappa B signaling pathway models. *BMC Syst Biol.* 2009;3:50.
9. Sandink CA, McAuley KB, McLellan PJ. Selection of parameters for updating in on-line models. *Ind Eng Chem Res.* 2001;40:3936–3950.
10. Chu YF, Hahn J. Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters. *Ind Eng Chem Res.* 2009;48:6000–6009.
11. Li RJ, Henson MA, Kurtz MJ. Selection of model parameters for off-line parameter estimation. *IEEE Trans Control Syst Technol.* 2004;12:402–412.
12. Sun CL, Hahn J. Parameter reduction for stable dynamical systems based on Hankel singular values and sensitivity analysis. *Chem Eng Sci.* 2006;61:5393–5403.
13. Walter E, Pronzato L. Qualitative and quantitative experiment design for phenomenological models—a survey. *Automatica.* 1990;26:195–213.
14. 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 Technol.* 1997;36:69–79.
15. 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.
16. Chu YF, Hahn J. Parameter set selection for estimation of nonlinear dynamic systems. *AICHE J.* 2007;53:2858–2870.
17. Machado VC, Tapia G, Gabriel D, Lafuente J, Baeza JA. Systematic identifiability study based on the Fisher Information Matrix for reducing the number of parameters calibration of an activated sludge model. *Environ Modell Softw.* 2009;11:1274–1284.
18. Rabitz H, Kramer M, Dacol D. Sensitivity analysis in chemical kinetics. *Annu Rev Phys Chem.* 1983;34:419–461.
19. Turanyi T. Sensitivity analysis of complex kinetic systems. Tools and applications. *J Math Chem.* 1990;5:203–248.
20. Caracotsios M, Stewart WE. Sensitivity analysis of initial value problems with mixed odes and algebraic equations. *Comput Chem Eng.* 1985;9:359–365.
21. Leis JR, Kramer MA. The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. *ACM Trans Math Softw.* 1988;14:45–60.
22. Chu YF, Hahn J. Integrating parameter selection with experimental design under uncertainty for nonlinear dynamic. *AICHE J.* 2008;54:2310–2320.
23. Burth M, Verghese GC, Velez-Reyes M. Subset selection for improved parameter estimation in on-line identification of a synchronous generator. *IEEE Trans Power Syst.* 1999;14:218–225.
24. Hiskens IA. Power system modeling for inverse problems. *IEEE Trans Circuits Syst I-Regul Pap.* 2004;51:539–551.
25. Gadkar KG, Gunawan R, Doyle FJ. Iterative approach to model identification of biological networks. *BMC Bioinformatics.* 2005;6:155.

26. 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.
27. Puskas JE, Chan SWP, McAuley KB, Shaikh S, Kaszas G. Kinetics and mechanisms in carbocationic polymerization: the quest for true rate constants. *J Polym Sci Pol Chem*. 2005;43:5394–5413.
28. Jaqaman K, Danuser G. Linking data to models: data regression. *Nat Rev Mol Cell Biol*. 2006;7:813–819.
29. Yue H, Brown M, Knowles J, Wang H, Broomhead DS, Kell DB. Insights into the behaviour of systems biology models from dynamic sensitivty and identifiability analysis: a case study of an NF-kappa B signalling pathway. *Mol Biosyst*. 2006;2:640–649.
30. Lin WJ, Biegler LT, Jacobson AM. Modeling and optimization of a seeded suspension polymerization process. *Chem Eng Sci*. 2010;65:4350–4362.
31. Raue A, Kreutz C, Maiwald T, Klingmuller U, Timmer J. Addressing parameter identifiability by model-based experimentation. *IET Syst Biol*. 2011;5:120–130.
32. Velez-Reyes M, Verghese GC. Subset selection in identification, and application to speed and parameter estimation for induction machines. In: *Proceedings of the 4th IEEE Conference on Control Applications*, Albany, 1995:991–997.
33. Hiskens IA. Nonlinear dynamic model evaluation from disturbance measurements. *IEEE Trans Power Syst*. 2001;16:702–710.
34. Golub GH, van Loan CF. *Matrix Computations*, 3rd ed. Baltimore: Johns Hopkins University Press, 1996.
35. Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: State of the art. *Chem Eng Sci*. 2008;63:4846–4872.
36. Kreutz C, Timmer J. Systems biology: experimental design. *FEBS J*. 2009;276:923–942.
37. Miao HY, Xia XH, Perelson AS, Wu HL. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev*. 2011;53:3–39.
38. Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG. Systems biology: parameter estimation for biochemical models. *FEBS J*. 2009;276:886–902.
39. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*. 2007;3:1871–1878.
40. Singh A, Jayaraman A, Hahn J. Modeling regulatory mechanisms in IL-6 signal transduction in hepatocytes. *Biotechnol Bioeng*. 2006;95:850–862.
41. Chu Y, Jayaraman A, Hahn J. Parameter sensitivity analysis of IL-6 signalling pathways. *IET Syst Biol*. 2007;1:342–352.
42. Seber GAF. *A Matrix Handbook for Statisticians*. New York: Wiley-Interscience, 2007.

Appendix 1. Orthogonal Projection

Given a matrix $\mathbf{S} = [\mathbf{s}_1,\mathbf{s}_2,\cdots,\mathbf{s}_n]$ of full column rank, let $\mathbf{P}(\mathbf{S})$ denote the orthogonal projection matrix which projects a vector onto the linear space spanned by the columns in $\mathbf{S}$. The projection matrix is given by[42]

$$\mathbf{P}(\mathbf{S}) = \mathbf{S}\left(\mathbf{S}^{\mathrm{T}}\mathbf{S}\right)^{-1}\mathbf{S}^{\mathrm{T}} \tag{A1}$$

This projection matrix is symmetric and idempotent, that is,

$$\mathbf{P}(\mathbf{S})^{\mathrm{T}} = \mathbf{P}(\mathbf{S}) \quad \text{and} \quad \mathbf{P}(\mathbf{S})^2 = \mathbf{P}(\mathbf{S}) \tag{A2}$$

Let $\mathbf{P}(\mathbf{S})^{\perp}$ denote the projection matrix which projects a vector onto the linear space perpendicular to the columns in $\mathbf{S}$. This matrix can be computed as follows:

$$\mathbf{P}(\mathbf{S})^{\perp} = \mathbf{I} - \mathbf{P}(\mathbf{S}) \tag{A3}$$

Appendix 2. Schur Complement

Assume a given a square matrix of $\mathbf{A}$ can be partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \tag{B1}$$

If the submatrix of $\mathbf{A}_{11}$ is nonsingular, then the Schur complement of $\mathbf{A}_{22}$ in $\mathbf{A}$ is given by[42]

$$\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \tag{B2}$$

An important application of the Schur complement is computation of the determinant, which can make use of the following expression:

$$\det(\mathbf{A}) = \det(\mathbf{A}_{11})\det\left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right) \tag{B3}$$

Manuscript received May 3, 2011, and revision received Jun. 16, 2011.