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

Yunfei Chu and Juergen Hahn*
Artie McFerrin Department of Chemical Engineering
Texas A&M University
College Station, TX 77843-3122
hahn@tamu.edu

1. Introduction

Mathematical modeling has become an essential approach to understand the mechanisms and dynamics in a system. With advent of computing power and experimental techniques, modeling of large-scale complex systems is accessible and it has become a key component in various research areas. For example, in systems biology mathematical modeling has been a keystone (Kitano 2002). To study signal transduction pathways a number of large scale models have been developed (e.g. Schoeberl *et al*., 2002; Yamada *et al*., 2003; Singh *et al*., 2006) which provide a fresh insight into the biological systems. The structure of a model is often built by mechanism analysis while the parameters are often required to be updated from experimental data. The usefulness of a model is highly dependent on the quality of the model. However accurate estimation of the large scale systems is still a challenge.

A great number of parameters contrasted with the limited experimental data results in the over-parameterized model and not all the parameters are identifiable in practice. If parameters are not practically identifiable then a small amount of noise in the data will result in large variations of the estimated value of the parameters and the parameters can not be estimated accurately (Walter and Pronzato, 1990). Estimation of an over-parameterized model is an ill-posed inverse problem and some regularization methods are required to guarantee the uniqueness and stability of the solution (Aster *et al*., 2005). A widely used regularization is to select a subset of parameters to be estimated while all other parameters are fixed at a constant value. The question then becomes how to select the identifiable parameters to be estimated.

The identifiability of a set of parameters depends on the effect that changes in their values have on the output and the effect can be in turn characterized by the sensitivity vectors. Several methods for parameter selection based on sensitivity vectors have been proposed in the literature. These include, but are not limited to, a collinearity index method (Brun et al., 2001), a column pivoting method (Velez-Reyes and Verghese, 1995), an extension of the relative gain array (Sandink et al., 2001), a Gram-Schmidt orthogonalization method (Yao et al., 2003), a recursive approach based upon principal component analysis (Li et al., 2004) and a combination of Hankel singular value and singular value decomposition (Sun and Hahn, 2006). A systematic approach for parameter selection is based on optimality criteria computed from the Fisher information matrix. The inverse of Fisher information matrix provides a lower bound for the covariance matrix of parameter estimators (Walter and Pronzato, 1990) and it can serve as a measure for the quality of a parameter set. A subset of identifiable parameters can be selected based upon optimizing some criteria such as the *D*-optimality or the modified *E*-optimality criterion (Weijers et al., 1997; Brun et al., 2002). Parameter selection has been used in a variety of applications, ranging from ecological systems (Anh et al., 2006), power systems (Hiskens, 2001), production systems (Bastogne et al., 2007), chemical reactions (Kou et al., 2005), biochemical networks (Gadkar et al., 2005) to wastewater treatment processes (Sin and Vanrolleghem, 2007).

Generally by defining an objective function to measure the quality of a parameter set parameter selection can be formulated as a combinatorial optimization problem, however, it is

nontrivial to find the solution. For small scale system the exhaustive search (Weijers et al., 1997) can provide the exact solution. However, even for medial scale system the total number of possible combination of parameters is too large to be enumerated. Stochastic search such as the genetic algorithm (Chu and Hahn, 2007) provides a way to select parameters for medial and large scale systems. Another approach is the stepwise methods which select parameters one by one and in each step a parameter is selected to optimize the criterion function. The orthogonalization method including householder transformation (Velez-Reyes and Verghese, 1995) and Gram-Schmidt procedure (Yao et al., 2003) can be regarded as an approach to maximize *D*-optimality criterion stepwisely (Chu and Hahn, 2007).

However, the existing methods all concentrate on search in identifiable parameter sets. This work provides a different avenue for parameter selection. The pairwise indistinguishable parameter sets are determined first, in which any two parameters can not be distinguished. Since the parameters belonging to the same indistinguishable set can not be estimated simultaneously the search of all combinations of parameters can be reduced to search of parameters from different indistinguishable sets. The number of possible combinations will be reduced dramatically.

2. Background

2.1 Identifiability

Parameter identifiability is an important issue in estimation and basically it is cast into two category: analytical (or structural, priori) identifiabilty and numerical (or practical, posteriori) identifiability (Walter 1987; Walter and Pronzato, 1997; Ljung, 1999). The analytical identifiabilty studies the uniqueness of solution in estimation while the numerical identifiability investigates the stability of the solution. The analytical identifiabilty can be also divided into two kinds: the global identifiability which guarantees the unique solution in a large range of parameter space and the local identifiability which guarantees the unique solution in a neighborhood of a given parameter value. It is easy to see the global identifiability includes the local identifiability however it is much more difficult to test. The common methods including differential algebra (Ljung and Glad, 1994), Taylor series approach and similarity transformation (Chappell *et al*., 1990) are restricted to small scale systems.

The local identifiability is relatively easy to check. The condition is based on the rank of the sensitivity matrix. The local identifiability is related with this work and to make it clear the definition and the condition from Rothenbe (1971) are used.

Definition 1: A parameter point $\boldsymbol{\theta}_0$ is said to be locally identifiable if there exists an open neighborhood of $\boldsymbol{\theta}_0$ containing no other $\boldsymbol{\theta}$ which produces the identical observations $\mathbf{y}$.

Condition 1: Let $\boldsymbol{\theta}_0$ be a parameter point and the sensitivity matrix $\mathbf{S}(\boldsymbol{\theta}) = \partial \mathbf{y} / \partial \boldsymbol{\theta}^{\mathrm{T}}$ has constant rank in a neighborhood of $\boldsymbol{\theta}_0$. Then $\boldsymbol{\theta}_0$ is locally identifiable if and only if $\mathbf{S}(\boldsymbol{\theta}_0)$ is nonsingular.

It should be noted the condition that the sensitivity matrix has constant rank is necessary. If this condition is removed then the nonsingularity of the sensitivity matrix is just a sufficient condition for local identifiability, or, in another word, a rank deficient sensitivity matrix does not imply that the parameters are not locally identifiable. The condition of constant rank has to be checked analytically and only the value at a parameter point is not enough.

The analytical identifiability guarantees the existence of the unique solution at least in a small range. However the analytically identifiable parameters may not be estimated accurately in practice. If sensitivity matrix is not singular but is close to singularity, the noise in data will be amplified in the procedure of estimation and result in large variations in estimated parameter value. To obtain an accurate estimation it is also required that the sensitivity matrix should be far

from singularity. However the term of 'far from singularity' has no unified definition and various functions are used to measure the distance to singularity of the sensitivity matrix. A set of them are the experimental optimality criteria presented in the next subsection.

2.2 Optimality criteria of the Fisher Information Matrix

The numerical identifiability can be measured by the parameter covariance matrix. If the covariance matrix is large then the parameters are not numerical identifiable. However, the covariance matrix can only be computed after the parameters have been estimated and it is also affected by the estimation algorithm. Alternatively, the Fisher information matrix can be used as its inverse provides the Cramer-Rao lower bound for the covariance matrix (Ljung, 1999).

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

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

where $\tilde{\mathbf{y}} = [\tilde{y}(t_1), \cdots, \tilde{y}(t_n)]^{\mathrm{T}}$ is the observation of the output, $\mathbf{y}(\boldsymbol{\theta}) = [y(t_1,\boldsymbol{\theta}), \ldots, y(t_n, \boldsymbol{\theta})]^{\mathrm{T}}$ is the true value and $\boldsymbol{\varepsilon} = [\varepsilon(t_1), \ldots, \varepsilon(t_n)]^{\mathrm{T}}$ 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 $\boldsymbol{\Sigma}=\sigma^2\mathbf{I}$. The Fisher information matrix, $\mathbf{F}$, is given by

$$\mathbf{F}(\boldsymbol{\theta}) = \left(\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right)^{\mathrm{T}} \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}. \tag{2}$$

Without loss of generality for the procedure, it can be assumed that $\sigma^2 = 1$.

To reduce the variations in estimated parameters it is desired to maximize the Fisher information matrix in some sense in order to reduce its inverse. To measure the size of the Fisher information matrix precisely a set of real function is defined. Such functions are called the experimental optimality criteria and named alphabetically (Kiefer, 1959). The most popular experimental optimality criterion is the *D*-optimality criterion which maximizes the logarithm of the determinant of the Fisher information matrix:

$$\varphi_D^{\ *} = \max \varphi_D(\mathbf{F}) = \max \log\det(\mathbf{F}). \tag{3}$$

This criterion minimizes the volume of the confidence ellipsoid with an arbitrary fixed confidence level for a least square estimator. Other common criteria (Kiefer, 1959; Atkinson et al., 2007) includes the *E*-optimality which maximizes the least eigenvalue of Fisher information matrxi

$$\varphi_E^{\ *} = \max \varphi_E(\mathbf{F}) = \max \lambda_n(\mathbf{F}). \tag{4}$$

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

$$\varphi_{ME}^{\ *} = \min \varphi_{ME}(\mathbf{F}) = \min \kappa(\mathbf{F}). \tag{5}$$

To optimize these criteria, the Fisher information matrix should be far from singularity. And from relation between the Fisher information matrix and the sensitivity matrix (Eq.2) it is equivalent to that the sensitivity matrix is far from singularity. The optimality criteria provide a set of functions to measure the distance to singularity of the sensitivity matrix for numerical identifiability test.

If the parameters are not all numerical identifiable a set of identifiable parameters are often selected for estimation. When a set of parameters selected the Fisher information matrix become

$$\mathbf{F}_{\mathbf{L}} = \mathbf{L}^{\mathrm{T}}\mathbf{F}\mathbf{L} = (\mathbf{S}\mathbf{L})^{\mathrm{T}}(\mathbf{S}\mathbf{L}). \tag{6}$$

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

$$\mathbf{L} = \begin{bmatrix} \mathbf{e}_{i_1} & \mathbf{e}_{i_2} & \cdots & \mathbf{e}_{i_m} \end{bmatrix}. \tag{7}$$

The set $\{i_1, i_2, \cdots, i_m\}$ denotes the selected parameters and $\mathbf{e}_i$ is the i-th column of the identity matrix. Then the problem of parameter selection becomes search of the selection matrix to let the Fisher information matrix far from singularity by optimizing some criterion function.

However, this is a combinatorial problem and not trivial to solve. A way to obtain an adequately good solution is the forward selection, which adds a new parameter into the selected set each step and the parameter added is determined by optimizing the criterion. One kind of forward selection is the Gram-Schmidt orthogonalization method (Yao et al., 2003). Using this technique, the first parameter selected has the largest norm of the sensitivity vector. In a second step the sensitivity vectors of other parameters are projected on to the space orthogonal to the first sensitivity vector. From the projected sensitivity vectors, the longest one is chosen and the corresponding parameter is selected as the second parameter in the set. The same procedure is repeated to select the remaining parameters until the number of parameters to be estimated is reached or until the length of the projected sensitivity vectors decreases below a certain threshold. Another version of orthogonalization method uses the Householder transformation for orthogonalization (Velez-Reyes and Verghese, 1995) which extends the linear version (Golub and van Load, 1989) to the nonlinear systems. The basic procedure is the same as the Gram-Schmidt procedure except there is a Householder mapping in each step. It has been shown that the orthogonalization method is in fact a stepwise method to maximize *D*-optimality criterion (Chu and Hahn, 2007). The forward selection can not guarantee to find the optimal solution however it can quickly find a suboptimal solution close to the optimal one. Because of it simplicity, the orthogonalization method has attracted a wide attention (e.g., Burth, 1999; Gadkar et al. 2005; Kou et al., 2005; Yue et al., 2006; Jaqaman and Danuser, 2006).

3 Investigation of pairwise indistinguishable parameter sets

If the sensitivity vectors of two parameters are not parallel then the parameters can not be reduced to one and they are distinguishable. However, if the sensitivity vectors are close to paralleling the two parameters will have very similar effects and in practice the difference in effects to distinguish them may be so tiny that they are overwhelmed by the noise in the data. In this case the parameters are said to be numerically pairwise distinguishable.

Before search for the numerically distinguishable sets a value to measure how close a sensitivity vector parallel to one another is required. Naturally the angle between two sensitivity vectors provides a way to measure the closeness of the vectors. A similarity measure of two parameters can be defined by the angle as

$$\cos\phi_{ik} = \frac{\left|\mathbf{s}_i{}^{\mathrm{T}}\mathbf{s}_k\right|}{\left\|\mathbf{s}_i\right\|_2 \left\|\mathbf{s}_k\right\|_2}, \tag{8}$$

where $\phi_{ik} \in [0, \pi/2]$ is the angle between sensitivity vector $\mathbf{s}_i$ and $\mathbf{s}_k$. The range of the similarity measure is from 0 to 1. When it equals 1, the two parameters have equivalent effect and the case returns to the analytical indistinguishability. When it equals 0, the sensitivity vectors of the parameters are orthogonal and the parameters have distinct effects. It is should be noted that the similarity measure is the absolute value of Pearson's correlation coefficient. The absolute value is because the parameters will have equivalent effects when their sensitivity vectors parallel to each other no matter if they are in the identical direction or in the opposite direction.

Based on the similarity measure, the parameters can be grouped by the clustering algorithms. Clustering is an unsupervised classification which partition a data set into subsets (groups), so that the data in each subset share some common trait. In this work the parameter set is partitioned by the parameter effects and the parameters in the same group have similar effects. The groups clustered are indeed the numerical indistinguishable required. There are a great number of clustering algorithms (Duda et al., 2001; Theodoridis and Koutroumbas, 2006) and the agglomerative hierarchical clustering is used.

This method builds the hierarchy from the individual parameter to the whole set by progressively merging groups. In the initial stage, each parameter is clustered into a one parameter group. At each middle stage two groups are merged into a new group and the two

groups selected have the largest similarity measure. There are several approaches to define the similarity between groups based on the similarity between individual parameters. The approach of complete linkage is used which uses the least similarity value between the parameters in two groups as the similarity measure of the two groups. By this approach the least similarity value of parameters in a group can be controlled by choosing the number of groups. The fewer groups partitioned the smaller the least similarity value in a group.

Investigation of the pairwise indistinguishable sets reveals a relationship among parameters. The information of which parameters have the similar effects is very helpful to study the mechanism of a system. An important application is to select the identifiable parameter set. Parameter selection can be formulated as a combinatorial optimization problem to maximize some criterion function which measures the degree of identifiability of a set of parameters. A commonly used criterion function is the *D*-optimality criterion which is formulated as

$$\begin{aligned} &\mathbf{z}^* = \arg\max_{\mathbf{z}} \ \log\det\left(\mathbf{F}(\mathbf{z})\right) \\ &\text{s.t. } \mathbf{F}(\mathbf{z}) = \mathbf{FIM}_{(i_1,\cdots i_{n_s})}^{(i_1,\cdots i_{n_s})} \text{ with } i_j \text{ that } z_{i_j} = 1,\ j = 1\cdots n_s \\ &\quad z_1 + z_2 + \cdots + z_{n_\theta} = n_s \\ &\quad z_i \in \{0,1\},\ i = 1\cdots n_\theta \end{aligned} \tag{9}$$

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

However, the solution to the combinatorial problem is not trivial. Even for medial scale system (e.g., 10 selected from 50) the total number of possible combination of parameters is too large to be searched exhaustively. Clustering parameters provides a new approach to solve the combinatorial problem for parameter selection. Search of which parameters can not be estimated together is performed before search of which parameters can be estimated simultaneously. Since the parameters in a pairwise indistinguishable set can not be estimated together there is no need to try their combinations in the search. Identification of indistinguishable sets inserts the new constraints to the optimization problem to reduce the possible number of combination

$$\sum_i z_{n_i} \le 1, \tag{10}$$

where z_{n_i} represents if the *i*-th parameter in the *n*-th indistinguishable set is selected or not and in an indistinguishable set at most one parameter can be selected for estimation. Further, a representative parameter can be selected from each indistinguishable set and the parameters for estimation can be searched from the combinations of the representative parameters. Clustering parameters can be regarded as a 'dimensional reduction' method. It can reduce the number of possible combination dramatically and make it possible for an exhaustive search.

Algorithm of parameter selection base on parameter clustering

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

Step 2. Determine n_s, the number of parameters per set, by singular value decomposition of the sensitivity matrix or the methods of forward selection.

Step 3. Wipe out the insignificant parameters whose sensitivity vectors have small length (e.g., less than 5% of the largest one).

Step 4. Cluster the parameters into n_g ($n_g \ge n_s$) groups with the similarity measure (Eq. 8) by hierarchical clustering.

Step 5. Select the parameter which has the largest sensitivity vector in a group as the representative of the group.

Step 6. Select n_s parameters from n_g representatives to optimize the criterion function by exhaustive search.

There are different approaches to calculate the parameter sensitivity in Step 1 and one commonly used is to calculate the sensitivity value $\partial \mathbf{y}(t)/\partial \boldsymbol{\theta}^{\mathrm{T}}$ by solving the system equations

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \mathbf{f}\left(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}\right) \\ \mathbf{y} &= \mathbf{h}\left(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}\right) \end{aligned}, \tag{11}$$

and the sensitivity equations simultaneously

$$\begin{aligned} \frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \\ \frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} &= \frac{\partial \mathbf{h}}{\partial \mathbf{x}^{\mathrm{T}}}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \end{aligned}, \tag{12}$$

where **x**, **u**, **y**, **θ** are states, inputs, outputs and parameters respectively. The sensitivity vector is consisted of the sensitivity value at different time points

$$\mathbf{s}_i = \left[\partial \mathbf{y}^{\mathrm{T}}(t_1)/\partial \theta_i \quad \partial \mathbf{y}^{\mathrm{T}}(t_2)/\partial \theta_i \quad \cdots \quad \partial \mathbf{y}^{\mathrm{T}}(t_n)/\partial \theta_i\right]^{\mathrm{T}}, \tag{13}$$

where $\mathbf{s}_i$ is the sensitivity vector with respect to parameter θ_i.

The number of parameters per set can be determined by the numerical rank of the sensitivity matrix. Each column of the sensitivity matrix is a sensitivity vector of a parameter. The number of columns is equal to the number of parameters. However, due to the parameter correlation sensitivity matrix will be rank-deficient or close to singularity. In this case some singular values of the sensitivity matrix are zero or close to zero. The numerical rank of the sensitivity matrix is the number of the largest singular values greater than a threshold. The forward selection as the orthogonalization method can be also used to determine the number of parameters per set and it will have the similar results with singular value decomposition of the sensitivity matrix.

Step 3 is a screening of parameters based on their sensitivity values only. Science the parameters have small sensitivity value are difficult to be estimated no matter what combinations they are in, these parameters can be discarded in further analysis. The number of groups is a key variable in the algorithm which controls the discrepancy value between the original function and the reduced one and it also have a significant effect on the optimal solution searched. Using the hierarchical clustering the hierarchical tree is obtained. From the tree it is easy to see the relation between the least similarity value and the number of groups. So the number of groups can be determined by the least similarity value required.

A parameter in an indistinguishable set can compensate the effects of other parameters in the set. However, if the sensitivity value of the parameter is small a large change of the value is required for the compensation and if the sensitivity value is large a small change is required. A parameter with a large sensitivity value is easy to be estimated from the noise data than a parameter with a small sensitivity value. So the representative of a set is selected as the one has the largest sensitivity value in Step 5. The number of combination of representatives is much smaller than the number of all parameters and an exhaustive search is feasible.

4 Case study

Modeling of biochemical reactions in a cell has been a keystone in systems biology. However a difficulty in the modeling is that the models often contain even hundreds of parameters while the experimental data gathered is still scarce. So not all the parameters are identifiable and a set of identifiable parameters is often selected for estimation. Parameter

selection, such as the Gram-Schmidt orthogonalization method, has attracted a wide attention in modeling of large scale biochemical networks (Gadkar et al. 2005; Yue et al., 2006; Jaqaman and Danuser, 2006).

To illustrate the technique a model of a signal transduction network (Huang *et al*., 2007) is used. The model, shown in Figure 1, contains two pathways: Janus-associated kinases & signal transducers and transcription factors are activated in one pathway while the other pathway involves the activation of mitogen-activated protein kinases and it is updated from the model of Singh *et al*. 2006. This model consists of 72 nonlinear ordinary differential equations which include 124 parameters. The details description of the model and value of the parameters can be seen in the references (Singh *et al*., 2006; Chu *et al*., 2007; Huang *et al*., 2007)

From the 124 parameters 7 is selected for estimation. The number of parameters selected is determined by singular value decomposition of the sensitivity matrix. The singular values beyond the largest 7 are close to zero. Use of the orthogonalization method will have the same result. The 74 insignificant parameters the sensitivity value of which are less than 5% of the largest one are removed before selection. The problem then becomes a selection of 7 parameters from 50 to optimize the *D*-optimality criterion. The total number of possible combinations is near 10^7 and there is an intensive computation burden for exhaustive search. To compare, the orthogonal method, a genetic algorithm and the clustering method are all applied.

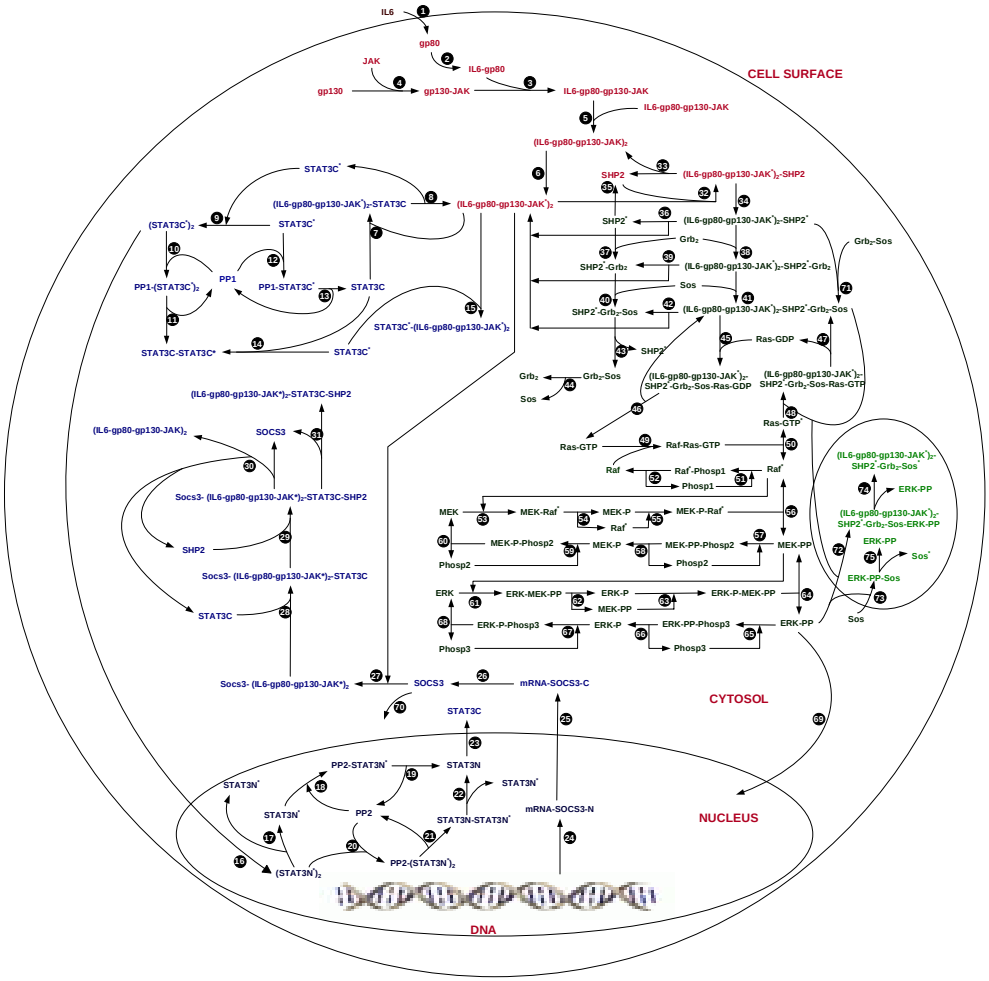

Figure 1. The model of IL-6 signaling pathway

Figure 2 shows the dendrogram of hierarchical clustering of parameters. It is obvious to see that some parameters have very high similarity value. The sensitivity vectors of those

parameters nearly parallel and it is difficult to distinguish the effects of any two parameters. The least similarity value can be also read from the figure. For example, the dash line shows the least similarity when the parameters are clustered into 12 groups. As increase of the number of groups the dash line will move down and the least similarity value will increase. Table 1 lists the least similarity value changing with the number of groups.

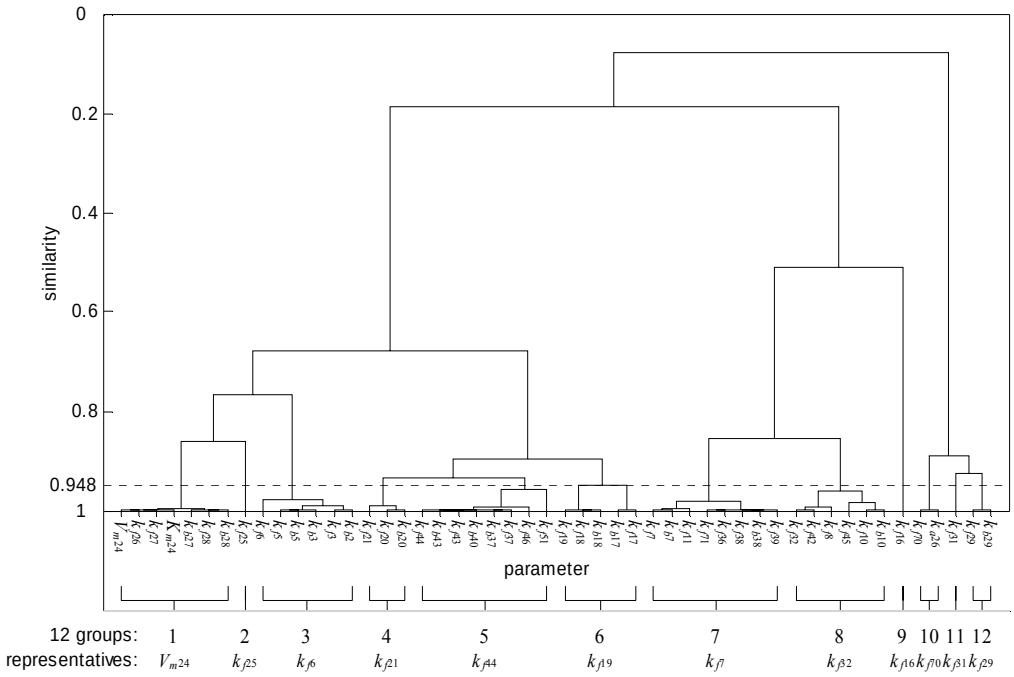

Figure 2. The dendrogram of hierarchical clustering of parameters

When the parameters are grouped, the parameter which has the longest sensitivity vector in a group is chosen as the representative of the group. Then the parameters for estimation are selected from the representatives by an exhaustive search to maximize the *D*-optimality criterion. The largest criterion value searched for different number of groups is shown in Table 1. The criterion value is small when the parameters are clustered into 7 groups. This is because the similarity of parameters in a group is not large enough that the representative is able to cover the total effects of a group. There is also a large discrepancy between the original model and the reduced one. However, as increase of the number of groups the largest criterion value rises. When the parameters are clustered into 12 groups the criterion value increases to the highest one and it stays at the value as increase of the number of groups. It gives a sign that an adequately good solution is found. The parameters in the 12 groups and the representatives are shown in Figure 2.

Table 1. The results by clustering method for different number of groups

No. of groups	7	8	9	10	11	12	13	14	15
Least similarity	0.862	0.889	0.894	0.924	0.935	0.948	0.959	0.962	0.977
Criterion	-4.139	-0.346	2.471	3.680	3.680	4.113	4.113	4.113	4.113
Discrepancy	0.430	0.305	0.272	0.074	0.031	0.029	0.014	0.009	0.008

To have an idea of how good the subset searched by clustering method, the optimal subsets searched by the orthogonalization method and the genetic algorithm are listed in Table 2. Rather

than the optimal one the genetic algorithm can also find some suboptimal sets listed in Table 2 as well. The optimal set searched by clustering method is identical with that searched by genetic algorithm. The set has the criterion value larger than the one searched by the orthogonalization method. The orthogonalization method is a forward selection approach to maximize the criterion stepwise. If the number of parameters per set is one then the *D*-optimality criterion is reduced to the length of the sensitivity vector. So in the first step the method of forward selection will select the parameter which has the longest sensitivity value. The parameter in this case is k_{f7}. However the optimal set searched by clustering method and genetic algorithm does not include this parameter. This indicates that the forward selection may fail to find the optimal solution.

Ability to search a collection of (sub-)optimal solutions rather than the only optimal one is an advantage for the genetic algorithm. From the solutions some relationship among the parameters can be found. For example, the second set searched by genetic algorithm has only one parameter, k_{a26}, different from the one, k_{f70}, in the first set and the difference in the criterion value between the two sets is minor. This means that in practice the two parameters are interchangeable. However, the clustering method shows it more clearly. The two parameters are clustered in the same group (the 10th group) shown in Figure 2 so one can be used to replace another. The same situation happens between the 3rd set and the 1st one. The different parameters k_{f18} and k_{f19} are in the same group (the 6th group). Rather than providing the optimal set for parameter selection clustering method can also reveal the relationship among parameters.

Table 2. The results by the three methods

	Parameters selected	Criterion
Clustering	k_{f32}, k_{f70}, k_{f16}, k_{f21}, k_{f31}, k_{f44}, k_{f19}	4.113
Forward selection	k_{f7}, k_{f6}, k_{f21}, k_{f70}, k_{f31}, k_{f19}, k_{f44}	3.850
Genetic algorithm	k_{f32}, k_{f70}, k_{f16}, k_{f21}, k_{f31}, k_{f44}, k_{f19}	4.113
	k_{f32}, k_{a26}, k_{f16}, k_{f21}, k_{f31}, k_{f44}, k_{f19}	4.106
	k_{f32}, k_{f70}, k_{f16}, k_{f21}, k_{f31}, k_{f44}, k_{f18}	4.088
	k_{f32}, k_{a26}, k_{f16}, k_{f21}, k_{f31}, k_{f44}, k_{f18}	4.071
	k_{f32}, k_{f70}, k_{f6}, k_{f21}, k_{f31}, k_{f44}, k_{f18}	4.064

5 Conclusion

A great number of parameters are involved in a large scale system. However, the effects of some parameters are every similar and these parameters can not be estimated simultaneously. Investigation of pairwise indistinguishable set is to uncover when the effect on the output produced by variations of parameters in a set can be compensated by change of one parameter in the set. Identification of the pairwise indistinguishable sets is very useful in parameter selection. Parameter selection is a combinatorial problem to optimize a criterion function of the set. Identification of the indistinguishable sets provides a 'dimensional reduction' technique to simplify the optimization problem. The search of all possible combinations of parameters is reduced to search of combinations of the representatives and the number of sets for enumeration is reduced so dramatically that an exhaustive search is feasible to find the solution.

References

1. 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. Ecological Modelling 193 (3-4): 182-204 2006.
2. Aster RC, Borchers B, Thurber CH. Parameter estimation and inverse problems. Boston: Elsevier Academic Press, 2005.
3. Atkinson A, Donev A, Tobias R. Optimum experimental designs, with SAS. Oxford: Oxford University Press, 2007.

4. Bastogne T, Thomassin M, Masse J. Selection and identification of physical parameters from passive observation. Application to a winding process. Control Engineering Practice 15 (9): 1051-1061 2007.
5. Brun R, Reichert P, Kunsch HR. Practical identifiability analysis of large environmental simulation models. Water Resources Research 37 (4): 1015-1030 2001.
6. Brun R, Kuhni M, Siegrist H, Gujer W, Reichert P. Practical identifiability of ASM2d parameters - systematic selection and tuning of parameter subsets. Water Research 36 (16): 4113-4127 2002.
7. Burth M, Verghese GC, Velez-Reyes M. Subset selection for improved parameter estimation in on-line identification of a synchronous generator. IEEE Transactions on Power Systems. 14 (1): 218-225 1999.
8. Chappell MJ, Godfrey KR, Vajda S. Global identifiability of the Parameters of Nonlinear-systems with specified inputs-A comparision of methods. Mathematical Biosciences 102 (1): 41-73 1990.
9. Chu Y, Jayaraman A, Hahn J. Parameter Sensitivity Analysis of IL-6 Signaling Pathways. IET Systems Biology 1 (6): 342-352 2007.
10. Chu Y, Hahn J. Parameter Set Selection for Estimation for Nonlinear Dynamic Systems. AIChE Journal 53 (11): 2858-2870 2007.
11. Duda RO, Hart PE, Stork DG. Pattern classification (2nd edition). New York: Wiley, 2001.
12. Gadkar KG, Varner J, Doyle FJ. Model identification of signal transduction networks from data using a state regulator problem. IEE Proceedings Systems Biology 2 (1): 17-30 2005.
13. Golub GH, van Loan CF. Matrix computations (2nd edition). Baltimore: Johns Hopkins University Press, 1989.
14. Hiskens IA. Nonlinear dynamic model evaluation from disturbance measurements. IEEE Transactions on Power Systems 16 (4): 702-710 2001.
15. Huang Z, Chu Y, Senocak F, Jayaraman A, Hahn J. Model Update of Signal Transduction Pathways in Hepatocytes based upon Sensitivity Analysis. Proceedings Foundations of Systems Biology. 45-50 Stuttgart, Germany 2007
16. Jaqaman K, Danuser G. Linking data to models: data regression. Nature Reviews Molecular Cell Biology 7 (11): 813-819 2006.
17. Kiefer J. Optimum experimental designs. Journal of the Royal Statistical Society Series B - Statistical Methodology 21 (2): 272-319 1959.
18. Kitano H. Systems biology: A brief overview. Science 295 (5560): 1662-1664, 2002
19. Kou B, McAuley KB, Hsu CC, Bacon DW, Yao KZ. Mathematical model and parameter estimation for gas-phase ethylene homopolymerization with supported metallocene catalyst. Industrial & Engineering Chemistry Research 44 (8): 2428-2442 2005.
20. Li RJ, Henson MA, Kurtz MJ. Selection of model parameters for off-line parameter estimation. IEEE Transactions on control systems technology 12 (3): 402-412 2004.
21. Ljung L, Glad T. On global identifiability for arbitrary model parametrizations. Automatica 30 (2): 265-276 1994.
22. Ljung L. System identification: theory for the user (2nd edition). Upper Saddle River: Prentice Hall PTR, 1999.
23. Rothenbe TJ. Identification in parametric models. Econometrica 39 (3): 577-591 1971.
24. Sandink CA, McAuley KB, McLellan PJ. Selection of parameters for updating in on-line models. Industrial & Engineering Chemistry Research 40 (18): 3936-3950 2001.
25. Singh A, Jayaraman A, Hahn J. Modeling regulatory mechanisms in IL-6 signal transduction in hepatocytes. Biotechnology and Bioengineering 95 (5): 850-862, 2006.
26. Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nature Biotechnology 20 (4): 370-375, 2002.

27. Sin G, Vanrolleghem PA. Extensions to modeling aerobic carbon degradation using combined respirometric-titrimetric measurements in view of activated sludge model calibration. Water Research 41 (15): 3345-3358, 2007.
28. Sun CL, Hahn J. Parameter reduction for stable dynamical systems based on Hankel singular values and sensitivity analysis. Chemical Egnineering Science 61 (16): 5393-5403, 2006.
29. Theodoridis S, Koutroumbas K. Pattern recognition (3rd edition). Boston: Elsevier/Academic Press, 2006.
30. 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. 991 – 997, 1995.
31. Walter E. Identifiability of parametric models. New York: Pergamon Press, 1987.
32. Walter E, Pronzato L. Qualitative and quantitative experiment design for phenomenological models - A survey. Automatica 26 (2): 195-213 1990.
33. Walter E, Pronzato L. Identification of parametric models from experimental data. Berlin: Springer, 1997.
34. Weijers SR, Vanrolleghem PA. A procedure for selecting best identifiable parameters in calibrating activated sludge model no.1 to full-scale plant data. Water Science and Technology 36 (5): 69-79 1997.
35. Yamada S, Shiono S, Joo A, Yoshimura A. Control mechanism of JAK/STAT signal transduction pathway. FEBS LETTERS 534 (1-3): 190-196, 2003.
36. Yao KZ, Shaw BM, Kou B, McAuley KB, Bacon DW. Modeling ethylene/butene copolymerization with multi-site catalysts: Parameter estimability and experimental design. Polymer Reaction Engineering 11 (3): 563-588 2003.
37. 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. Molecular Biosystems. 2 (12): 640-649, 2006.