DESIGN OF EXPERIMENTS[1]

Design our experiments in such a way so that the information content of the data is the highest possible because the quality of the parameter estimates ultimately depends on "how good" the data are.

Two approaches to experimental design:

(1) *factorial design* (we design a set of preliminary experiments for a process that very little or essentially no information is available"

(2) *sequential design* (we attempt to satisfy one of the following two objectives:
   a. estimate the parameters as accurately as possible or
   b. discriminate among rival models and select the best one.

PRELIMINARY EXPERIMENTAL DESIGN (Factorial Design)

There are many books that address experimental design and present *factorial experimental design* in detail[2]. Two papers also summarize the subject[3]

Factorial design, and in particular 2^k designs, represent a generally sound strategy. The independent variables (also called *factors* in experimental design) are assigned two values or *factor levels* (a high and a low value) and experiments are conducted in all possible combinations of the independent variables (also called *treatments*). For example, if we have three *factors* (independent variables) to vary, the complete 2^3 factorial design corresponds to the following 8 experiments:

Run	*Independent Variables*
1	(x_1-Low, x_2-Low, x_3-Low)
2	(x_1-Low, x_2-Low, x_3-High)
3	(x_1-Low, x_2-High, x_3-Low)
4	(x_1-High, x_2-Low, x_3-Low)
5	(x_1-Low, x_2-High, x_3-High)
6	(x_1-High, x_2-Low, x_3-High)
7	(x_1-High, x_2-High, x_3-Low)
8	(x_1-High, x_2-High, x_3-High)

[1] Englezos, P. and N. Kalogerakis, "*Applied Parameter Estimation for Chemical Engineers*", Marcel-Dekker, New York, 2001.;

[2] Montgomery, D.C., "*Design and Analysis of Experiments*", 5th edition, Wiley,2001; Anderson, VL. and R.A. McLean, *Design of Experiments*, M. Dekker, NewYork, NY, 1974; Box G.E.P., Hunter, W.G., Hunter, J.S., "*Statistics for Experimenters*", Wiley, 1978.

[3] Hockman, K., and D. Berengut, "Design of Experiments', *Chemical Engineering*, November 1995, 142-145; Murphy, T.D., "Design and Analysis of Industrial Experiments", *Chem. Engineering*, June 1977, 168-182

SEQUENTIAL EXPERIMENTAL DESIGN FOR PRECISE PARAMETER ESTIMATION

Assume that N experiments have been conducted up to now and an estimate of the parameter vector is obtained

> *What are the best conditions to perform the next experiment so that the variance of the estimated parameters is minimized?*

Consider an algebraic equation model ($\mathbf{y} = \mathbf{f}(\mathbf{x},\mathbf{k})$).

The problem is restated as "f*ind the best experimental conditions* (i.e., $\mathbf{x}_{N+1}$) *where the next experiment should be performed so that the variance of the parameters is minimized*."

The covariance matrix of the parameters is $COV(\mathbf{k}) = \hat{\sigma}_\varepsilon^2 \mathbf{A}^{-1}$ (12.1)

where $$\mathbf{A} = \sum_{i=1}^{N} \mathbf{G}_i^T \mathbf{Q}_i \mathbf{G}_i \quad (12.2)$$

If for a moment we assume that the next experiment at conditions $\mathbf{x}_{N+1}$ has been performed, the *new* matrix **A** would be:

$$\mathbf{A}^{new} = \sum_{i=1}^{N} \mathbf{G}_i^T \mathbf{Q}_i \mathbf{G}_i + \mathbf{G}_{N+1}^T \mathbf{Q}_{N+1} \mathbf{G}_{N+1} = \mathbf{A}^{old} + \mathbf{G}_{N+1}^T \mathbf{Q}_{N+1} \mathbf{G}_{N+1} \quad (12.3)$$

and the resulting covariance matrix would be $COV(\mathbf{k}) = \hat{\sigma}_\varepsilon^2 \left[\mathbf{A}^{new}\right]^{-1}$ (12.4)

The implication here is that if the parameter values do not change significantly from their current estimates when the additional measurements are included in the estimation, we can *quantify* the effect of each additional experiment *before* it has been carried out!

Hence, we can search all over the *operability region* (i.e., over all the potential values of $\mathbf{x}_{N+1}$ or the set of the feasible experimental conditions represented usually by a small number of grid points) to find the best conditions for the next experiment.

Optimality criteria to determine the conditions for the next experiment.

(Volume and Shape Design Criteria)

The Volume Design Criterion

We can minimize the variance of all parameters simultaneously by *minimizing the volume of the joint 95% confidence region*. *Minimization* of the volume of the ellipsoid

$$\left[\mathbf{k}-\mathbf{k}^*\right]^{\mathrm{T}}\mathbf{A}^{\mathrm{new}}\left[\mathbf{k}-\mathbf{k}^*\right]= p\,\hat{\sigma}_{\varepsilon}^2\,F^{\alpha}_{\mathrm{p},(\mathrm{N}+1)\mathrm{m}-\mathrm{p}} \qquad (12.5)$$

is equivalent to *maximization* of *det(*$\mathbf{A}^{\mathrm{new}}$) which in turn is equivalent to *maximization* of the product

$$\Pi = \lambda_1\lambda_2\ldots\lambda_{\mathrm{p}} \qquad (12.6)$$

where λ_i, i=1,…,p are the eigenvalues of matrix $\mathbf{A}^{\mathrm{new}}$.

The conditions ($\mathbf{x}_{\mathrm{N}+1}$) that yield the maximum value for *det*($\mathbf{A}^{\mathrm{new}}$) should be used to conduct the next experiment.

The Shape Design Criterion

When we have an ill-conditioned problem the volume criterion results in an elongated ellipsoid (like a cucumber) for the joint confidence region that has a small volume but the variance of the parameters can be very high.

We can determine the shape of the joint confidence region by examining the *cond*(**A**) which is equal to $\lambda_{\mathrm{max}}/\lambda_{\mathrm{min}}$ and represents *the ratio of the principal axes* of the ellipsoid. In this case we choose the experimental conditions which will yield the *minimum length for the largest principal axis of the ellipsoid*. This is equivalent to

maximization of λ_{min}

The conditions ($\mathbf{x}_{\mathrm{N}+1}$) that correspond to a maximum of λ_{min} are used to conduct he next experiment.

Implementation Steps

The steps to be followed to design the next experiment for the precise estimation of the model parameters is given below:

Step 1. Perform initial experiments (factorial design) to obtain initial estimates for the parameters and their covariance matrix.

Step 2. For each grid point of the operability region, compute the sensitivity coefficients and generate $\mathbf{A}^{new}$.

Step 3. Perform an eigenvalue decomposition of matrix $\mathbf{A}^{new}$ to determine its *condition number, determinant* and λ_{min}.

Step 4. Select the experimental conditions that correspond to a maximum $det(\mathbf{A}^{new})$ or maximum λ_{min}.

Step 5. Perform the experiment at the selected experimental conditions.

Step 6. Based on the additional measurement of the response variables, estimate the parameter vector and its covariance matrix.

Step 7. If the obtained accuracy is satisfactory, stop; else go back to Step 2 and select the conditions for an additional experiment.

In general, the search for the optimal $\mathbf{x}_{N+1}$ is made all over the operability region. Experience has shown however, that the best conditions are always found on the boundary of the operability region (Froment and Bischoff, 1990)[4].

[4] Froment, G.F., and K. B. Bischoff, *Chemical Reactor Analysis and Design*, 2nd ed., J. Wiley, New York, NY, 1990.

SEQUENTIAL EXPERIMENTAL DESIGN FOR MODEL DISCRIMINATION

Consider *r* rival models that can describe the system (if we perform a model adequacy test none of the models is rejected.)

We must perform additional experiments to determine which one of the rival models is the correct one *by rejecting all the rest*. Obviously, it is very important that the next experiment has *maximum discriminating power*. We illustrate this point with a very simple example. Assume that we have the following two rival single-response models,

Model 1: $y^{(1)} = k_1^{(1)} x + k_2^{(1)}$

Model 2: $y^{(2)} = k_2^{(2)} e^{k_1^{(2)} x}$

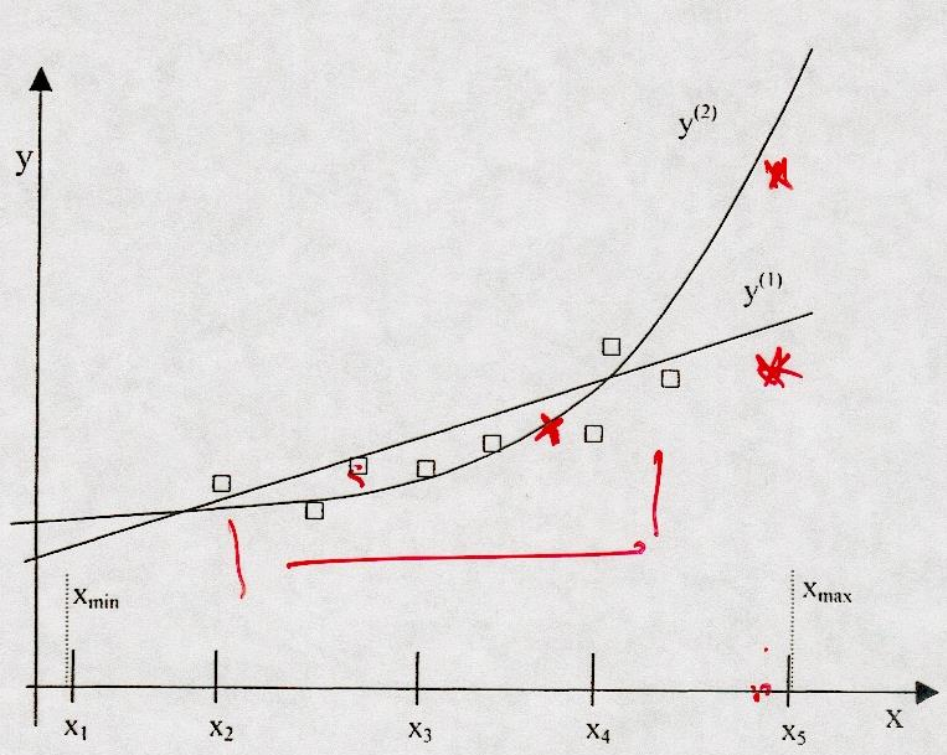

Figure 12.1 The two rival models can describe equally well data taken near x_2 and x_4, whereas data taken near x_1 or x_3 or particularly x_5 will have a high discriminating power.

As seen the vicinity of x_5 is the area of the operability region $[x_{min}, x_{max}]$ where the divergence between the two models is maximized.

The Divergence Design Criterion[5]

Hunter and Reimer (1965) proposed the *simple Divergence* criterion that can readily be extended to multi-response situations.

Hunter and Reimer (1965) proposed the *simple Divergence* criterion that can readily be extended to multi-response situations. In general, if we have performed N experiments; the experimental conditions $\mathbf{x}_{N+1}$ for the next one are obtained by maximizing the *weighted Divergence* between the rival models, defined as

$$D(\mathbf{x}_{N+1}) = \sum_{i=1}^{r} \sum_{j=i+1}^{r} \left[\mathbf{y}^{(i)} - \mathbf{y}^{(j)}\right]^{T} \mathbf{Q} \left[\mathbf{y}^{(i)} - \mathbf{y}^{(j)}\right] \quad (12.12)$$

where r is the number of the rival models and $\mathbf{Q}$ is a user-supplied weighting matrix. The role of $\mathbf{Q}$ is to normalize the contribution of the response variables that may be of a different order of magnitude (output vector, $\mathbf{y}$, could be temperature in degrees K, pressure in Pa and concentration as mole fraction).

[5] Hunter, W.G., A.M. Reimer, *Technometrics*, 7, 307-323, 1965

Model Adequacy Tests for Model Discrimination

Once the new experiment has been conducted, all models are tested for adequacy. Depending on our knowledge of σ_ε^2 we may perform a χ^2-test, an F-test or Bartlett's χ^2-test (when σ_ε^2 is completely unknown). The first two tests have been discussed previously as cases I and II.

Case III: Bartlett's χ^2-test

Let us assume that $\sigma_{\varepsilon,i}^2$ is the estimate of σ_ε^2 from the i^{th} model. The associated degrees of freedom for the i^{th} model are denoted with $(d.f)_i$. Bartlett's χ^2-test is based on testing the hypothesis

$$H_0 : \sigma_{\varepsilon,1}^2 = \sigma_{\varepsilon,2}^2 = \sigma_{\varepsilon,3}^2 = \ldots = \sigma_{\varepsilon,r}^2$$

H_1 : not all of the $\sigma_{\varepsilon,i}^2$ (i=1,…,r) are equal.

Using a χ^2-test and a level of significance (usually α=0.01 or α=0.05).

$$\text{if } \chi^2_{data} > \chi^2_{\nu=r-1,1-\alpha} \Rightarrow \text{Reject } H_O$$

where $\chi^2_{\nu=r-1,1-\alpha}$ is obtained from the tables and χ^2_{data} is computed from the data as follows:

First, we generate the *pooled variance*, $\hat{\sigma}_p^2$,

$$\hat{\sigma}_p^2 = \frac{\sum_{i=1}^{r}[(d.f.)_i - 1]\hat{\sigma}_{\varepsilon,i}^2}{\sum_{i=1}^{r}(d.f.)_i} \qquad (12.14)$$

which contains lack of fit, and then we compute χ^2_{data} from

$$\chi^2_{data} = \frac{ln[\hat{\sigma}_p^2]\sum_{i=1}^{r}(d.f.)_i - \sum_{i=1}^{r}(d.f.)_i\, ln[\hat{\sigma}_{\varepsilon,i}^2]}{1+\frac{1}{3(r-1)}\left(\sum_{i=1}^{r}\frac{1}{(d.f.)_i} - \frac{1}{\sum_{i=1}^{r}(d.f.)_i}\right)} \quad (12.15)$$

When the hypothesis H_O is rejected, we drop the model with the highest $\hat{\sigma}_{\varepsilon,i}^2$ and we repeat the test with one less model.
We keep on removing models until H_O cannot be rejected any more. These models are now used in the determination of the overall divergence for the determination of the experimental conditions for the next experiment.

NOTE: It should be emphasized that only the χ^2-test and the F-test are true model adequacy tests. Consequently, they may eliminate all rival models if none of them is truly adequate. On the other hand, Bartlett's χ^2-test does not guarantee that the retained model is truly adequate. It simply suggests that it is the best one among a set of inadequate models!

Implementation Steps for Model Discrimination

Based on the material presented above, the implementation steps to design the next experiment for the discrimination among *r* rival model are:

Step 1. Perform a series of initial experiments (*factorial design*) to obtain initial estimates for the parameters and their covariance matrix for each of the *r* rival models.

Step 2. For each grid point of the operability region, compute the Weighted Divergence, *D*. In the computations we consider only the models which are adequate at the present time (not all the rival models).

Step 3. Select the experimental conditions that correspond to a maximum *D*.

Step 4. Perform the experiment at the selected experimental conditions.

Step 5. Based on the new measurement of the response variables, estimate the parameter vector and its covariance matrix for *all* rival models.

Step 6. Perform the appropriate model adequacy test (χ^2-test, F-test or Bartlett's χ^2-test) for *all* rival models. If more than one model remains adequate, go back to Step 2 to select the conditions for another experiment.

We compute the Divergence *only* among adequate models in Step 2 in order to base the decision for the next experiment on the models that are still competing.

However, once the new data point becomes available it is good practice to update the parameter estimates for all models (adequate or inadequate ones). Practice has shown that under conditions of high experimental errors, as additional information becomes available some models may become adequate again!

SEQUENTIAL EXPERIMENTAL DESIGN FOR ODE SYSTEMS

Recall that the mathematical model is of the form,

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}, \mathbf{k}) \quad ; \quad \mathbf{x}(t_0) = \mathbf{x}_0$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t)$$

where

$\mathbf{k}=[k_1,k_2,\ldots,k_p]^T$ is a *p-dimensional* vector of unknown parameters;
$\mathbf{x}=[x_1,x_2,\ldots,x_n]^T$ is an *n-dimensional* vector of state variables;
$\mathbf{x}_0=[x_{10},x_{20},\ldots,x_{n0}]^T$ are initial conditions for the state variables
$\mathbf{u}=[u_1,u_2,\ldots,u_r]^T$ is an *r-dimensional* vector of manipulated variables which are normally set by the operator/experimentalist.

Design of the next experiment in this case requires the following:

(i) Selection of the initial conditions, $\mathbf{x}_0$.
(ii) Selection of the sampling rate and the final time.
(iii) Selection of the manipulated variables, $\mathbf{u}$, which could be kept either constant or even be optimally manipulated over time.

The sequential experimental design can be made either for precise parameter estimation or for model discrimination purposes.

Selection of *Optimal Sampling Interval* and *Initial State* for Precise Parameter Estimation

The optimality criteria based on which the conditions for the next experiment are determined are the same for dynamic and algebraic systems.

For precise parameter estimation we examine the determinant and the eigenvalues of matrix $\mathbf{A}^{new}$

$$\mathbf{A}^{new} = \mathbf{A}^{old} + \mathbf{K}\left[\sum_{i=1}^{N_P} \mathbf{G}^T(t_i)\mathbf{C}^T\mathbf{Q}\mathbf{C}\mathbf{G}(t_i)\right]\mathbf{K}$$

where N_P is the number of data points to be collected, $\mathbf{K}=diag(k_1,k_2,\ldots,k_p)$, $\mathbf{G}(t)$ is the parameter sensitivity matrix, $(\partial\mathbf{x}^T/\partial\mathbf{k})^T$ and matrix $\mathbf{A}^{old}$ is matrix $\mathbf{A}$ of the normal equations based on the experimental runs performed up to now.

If the number of runs is N_R and the number of data points gathered in each run is N_P, $\mathbf{A}^{old}$ is of the form,

$$\mathbf{A}^{old} = \mathbf{K}\sum_{j=1}^{N_R}\left[\sum_{i=1}^{N_P} \mathbf{G}^T(t_i)\mathbf{C}^T\mathbf{Q}\mathbf{C}\mathbf{G}(t_i)\right]_j\mathbf{K}$$

Our selection of the initial state, $\mathbf{x}_0$, and the value of the manipulated variables (kept constant during the experiment) vector, $\mathbf{u}(t)$ determine a particular experiment. Thus, the *operability region* is defined as a closed region in the $[x_{0,1},x_{0,2,\ldots},x_{0,n},u_1,u_2,\ldots,u_r]^T$ –space.

In addition to the selection of the experimental conditions for the next experiment (i.e., selection of the *best grid point* in the operability region according to the volume or the shape criterion), the *sampling rate* and the *final time or optimal time interval* must also be specified.

A. Time Interval Determination

We choose the time interval over which the output vector (measured variables) is most sensitive to the parameters[6]. For that purpose we use the *Information Index*[7] which for parameter k_j is defined as follows

$$I_j(t) = k_j \left(\frac{\partial \mathbf{y}^T}{\partial k_j} \right) \mathbf{Q} \left(\frac{\partial \mathbf{y}}{\partial k_j} \right) k_j \quad ; \quad j=1,\ldots,p$$

where **Q** is a suitably chosen weighting matrix.

The scalar $I_j(t)$ is simply a weighted sum of squares of the sensitivity coefficients of the output vector with respect to parameter k_j at time t.

Hence, $I_j(t)$ can be viewed as an index measuring the overall sensitivity of the output vector with respect to parameter k_j at time t.

Using Equation 12.17 and the definition of the parameter sensitivity matrix, the above equation becomes

$$I_j(t) = k_j \boldsymbol{\delta}_j^T \mathbf{G}^T(t) \mathbf{C}^T \mathbf{Q} \mathbf{C} \mathbf{G}(t) \boldsymbol{\delta}_j k_j \quad ; \quad j=1,\ldots,p \qquad (12.21)$$

where $\boldsymbol{\delta}_j$ is a *p-dimensional* vector with 1 in the j^{th} element and zeros elsewhere.

> The procedure for the selection of the most appropriate time interval requires the integration of the state and sensitivity equations and the computation of $I_j(t)$, j=1,…,p at each grid point of the operability region. Next by plotting $I_j(t)$, j=1,…,p versus time (preferably on a log scale) the time interval $[t_1, t_{Np}]$ where the information indices are excited and become large in magnitude is determined. This is the time period over which measurements of the output vector should be obtained.

---

[6] Kalogerakis, N., and R. Luus, "Sequential Experimental Design of Dynamic Systems Through the Use of Information Index", *Can. J. Chem. Eng.*, 62, 730-737 (1984).

[7] Kalogerakis, N., and R. Luus, "Improvement of Gauss-Newton Method for Parameter Estimation through the Use of Information Index", *Ind. Eng. Chem. Fundam.*, 22, 436-445 (1983b).

B. Sampling Rate Determination

Once the best time interval has been obtained, the sampling times within this interval should be obtained. If the number of data points, Np, is relatively large, in order to cover the entire time interval of interest it is practical to choose $t_2, t_3,\ldots,t_{Np-1}$ by a log-linear interpolation between t_1 ($t_1>0$) and t_{Np}, namely[8],

$$t_{i+1} = t_i\left(\frac{t_{N_P}}{t_1}\right)^{\frac{1}{N_P-1}} \quad ; \quad i=1,2,\ldots,N_P-2$$

The steps that should be followed to determine the best grid point in the operability region for precise parameter estimation of a dynamic system are:

Step 1. Compute the Information Indices, $I_j(t)$, j=1,…,p at each grid point of the operability region..
Step 2. By plotting $I_j(t)$, j=1,…,p versus time determine the optimal time interval $[t_1,t_{Np}]$ for each grid point.
Step 3. Determine the sampling times $t_2, t_3,\ldots$ using Equation 12.22 for each grid point.
Step 4. Integrate the state and sensitivity equations and compute matrix $\mathbf{A}^{new}$.
Step 5. Using the desired optimality criterion (volume or shape), determine the best experimental conditions for the next experiment.

NOTE: In Steps 2 to 4 we may consider only to the grid points on the boundary because as is the case with algebraic systems, the optimum conditions are expected to lie on the boundary of the operability region.

[8] Kalogerakis, N., and R. Luus, "Sequential Experimental Design of Dynamic Systems Through the Use of Information Index", *Can. J. Chem. Eng.*, 62, 730-737 (1984).

Selection of *Optimal Sampling Interva*l and I*nitial State* for Model Discrimination

The best grid point of the operability region is chosen by maximizing the overall divergence, defined for dynamic systems as

$$D(\mathbf{u},\mathbf{x}_0) = \sum_{l=j+1}^{r} \sum_{j=1}^{r} \sum_{i=1}^{N_P} \left[\mathbf{y}^{(j)}(t_i) - \mathbf{y}^{(l)}(t_i)\right]^T \mathbf{Q} \left[\mathbf{y}^{(j)}(t_i) - \mathbf{y}^{(l)}(t_i)\right]$$

Again, we consider that **u** is kept constant throughout an experimental run. The design procedure is the same as for algebraic systems.

Determination of *Optimal Inputs f*or Precise Parameter Estimation and Model Discrimination

As mentioned previously, the independent variables which determine a particular experiment and are set by the experimentalist are the initial state, $\mathbf{x}_0$ and the vector of the manipulated variables (also known as *control input*), **u**.

In the previous section we considered the case where **u** is kept constant throughout an experimental run. The case where **u** is allowed to vary as a function of time constitutes the problem of *optimal inputs*. It should be noted that the use of constant inputs is often more attractive to the experimentalist because of the ease of experimentation.

In general, it is expected that by allowing some or all of the manipulated variables to change over time, a greater amount of information will be gathered from a run. A l problem can be formulated as follows,

> *Given the governing differential equations (the model) and the initial conditions,* $\mathbf{x}_0$*, determine the optimal inputs,* **u**(t)*, so that the determinant (for the volume criterion) or the smallest eigenvalue (for the shape criterion) of matrix* $\mathbf{A}^{new}$ *is maximized.*

This optimal control problem is not a typical "textbook problem"[9]. It is not even clear whether the optimal solution can be readily computed. As a result, one should consider suboptimal solutions.

[9] Luus, R., *Iterative Dynamic Programming*, Chapman & Hall, CRC, London, UK, 2000;

EXAMPLES

12.1.1 Consecutive Chemical Reactions

As an example, for precise parameter estimation of dynamic systems we consider the simple consecutive chemical reactions in a batch reactor used by Hosten and Emig (1975) and Kalogerakis and Luus (1984) for the evaluation of sequential experimental design procedures of dynamic systems. The reactions are carried out isothermally and are as follows

$$A \rightarrow B \rightarrow C$$

Both steps are irreversible and kinetically of first order. The governing differential equations are

$$\frac{dx_1}{dt} = -k_1\, exp\left(-\frac{k_2}{T}\right)x_1 \quad ; \quad x_1(0) = x_{0,1}$$

$$\frac{dx_2}{dt} = k_1\, exp\left(-\frac{k_2}{T}\right)x_1 - k_3\, exp\left(-\frac{k_4}{T}\right)x_2 \quad ; \quad x_2(0) = 0$$

where x_1 and x_2 are the concentrations of *A* and *B* (*g/L*) and T is the reaction temperature (*K*).

Both state variables x_1 and x_2 are assumed to be measured and the standard error in the measurement (σ_ε) in both variables is equal to 0.02 (*g/L*). Consequently, the statistically correct choice for the weighting matrix **Q** is the identity matrix.
The experimental conditions which can be set by the operator are the initial concentrations of *A* and *B* and the reaction temperature T. For simplicity the initial concentration of *B* is assumed to be zero, i.e., we always start with pure *A*.

NOTE See: Kalogerakis, N., and R. Luus, "Sequential Experimental Design of Dynamic Systems Through the Use of Information Index", *Can. J. Chem. Eng.*, 62, 730-737 (1984

Murray, L.E. and E.K. Reiff Jr., "Design of Transient Experiments for Identification of Fixed Bed Thermal Transport Properties", *Can. J. Chem. Eng.*, 62, 55-61 (1984);

Kalogerakis, N., and R. Luus, "Sequential Experimental Design of Dynamic Systems Through the Use of Information Index", *Can. J. Chem. Eng.*, 62, 730-737 (1984).