Effect of Data Length on the Region of Convergence in Parameter Estimation Using Quasilinearization

NICOLAS KALOGERAKIS
and
REIN LUUS

Department of Chemical Engineering and Applied Chemistry
University of Toronto
Toronto, Ontario M5S 1A4, Canada

Parameter estimation in dynamic models is important in process modeling and simulation. Recently, Wang and Luus (1978) proposed the use of shorter data length to substantially enlarge the region of convergence to the global optimum for the parameters. The purpose of this note is to demonstrate that quasilinearization algorithm, known for its rapid convergence rate, can also be improved in terms of expanded region of convergence, and can be successfully used even when accurate initial estimates of the unknown parameters are not available. Moreover, we wish to show the effect of using different sections of the data on the reliability of the estimated parameter values during the early iterations, when only portions of the data are used.

PYROLYSIS OF BENZENE

The example chosen here is that used by Seinfeld and Gavalas (1970) for parameter estimation. The following two-parameter model is used to describe the pyrolytic dehydrogenation of benzene to diphenyl and triphenyl:

$$\frac{dx_1}{dt} = -r_1 - r_2\,; \quad x_1(0) = 1 \tag{1}$$

$$\frac{dx_2}{dt} = \frac{r_1}{2} - r_2\,; \quad x_2(0) = 0 \tag{2}$$

where

$$r_1 = k_1[x_1^2 - x_2(2 - 2x_1 - x_2)/0.726] \tag{3}$$

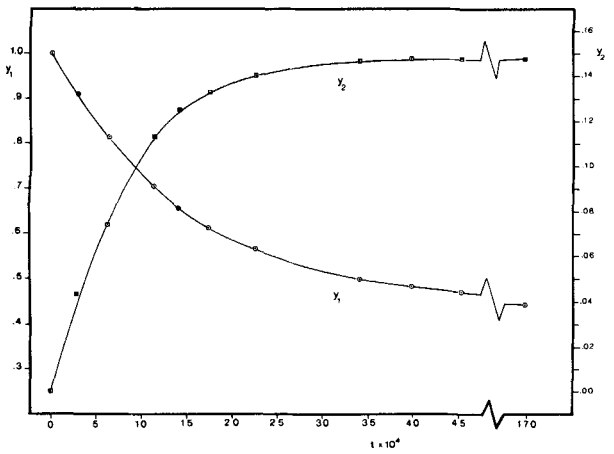

Figure 1. Data for pyrolysis of benzene:
○□ **actual data**
● ■ **interpolated data.**

$$r_2 = k_2[x_1x_2 - (1 - x_1 - 2x_2)(2 - 2x_1 - x_2)/3.852] \tag{4}$$

with x_1 and x_2 the gram-moles of benzene and diphenyl per gram-mole of pure benzene feed. The eight experimental measurements of y_1 and y_2 as a function of the residence time t, as given by Seinfeld and Gavalas (1970), are plotted in Figure 1. The measurements are related to the states x_1 and x_2 by

$$y_i(t_j) = x_i(t_j) + e_i(t_j)\,, \; i = 1,2\,; j = 1,2,\ldots,8 \tag{5}$$

where $e_i(t_j)$ indicates random experimental error.

The problem is to find the values of the parameters k_1 and k_2 which minimize the difference between the actual observations and the model predictions. We are thus interested in minimizing the expression

$$J = \sum_{j=1}^{8} \sum_{i=1}^{2} [y_i(t_j) - x_i(t_j)]^2 \tag{6}$$

EFFECT OF DATA LENGTH ON THE REGION OF CONVERGENCE

Quasilinearization algorithm is well known in parameter estimation for its fast speed of convergence. However, if the initial guess is not sufficiently close to the true parameter values, divergence may result. The difficulty of making a good initial guess can be overcome by the use of small portions of the data length during the early iterations, as proposed by Wang and Luus (1978, 1980).

This is shown for this particular example in Figure 2, where the regions of convergence are plotted in the $k_1 - k_2$ plane for different portions of the data length. Although negative initial guesses for the rate constants k_1 and k_2 have no physical meaning, they are included in the plots to show the entire region of convergence.

Since there are two parameters to be estimated, two data points would be sufficient to yield the exact parameter values when the data are noise free. In most cases, owing to noise in the data, the estimated values of the parameters will not be the best estimates when only a small portion of the data length is used. However, by gradually increasing the data length to the point where the parameters do not change substantially, the whole data length can then be used to obtain the optimum parameter values.

Thus, in this example, by using the first two given data points only we obtain initially (356.7, 399.1), and then using all the data we obtain the best estimates (354.6, 400.2). The use of only the first two data points during the early iterations resulted in enlargement of the region of convergence as compared to the case where all the data are used. The corresponding regions of convergence are represented by curves b and a, respectively, in Figure 2.

Further decrease of the data length requires generation of artificial data. This is accomplished by least-squares polynomial

0001-1541-80-3835-0671-$00.75.

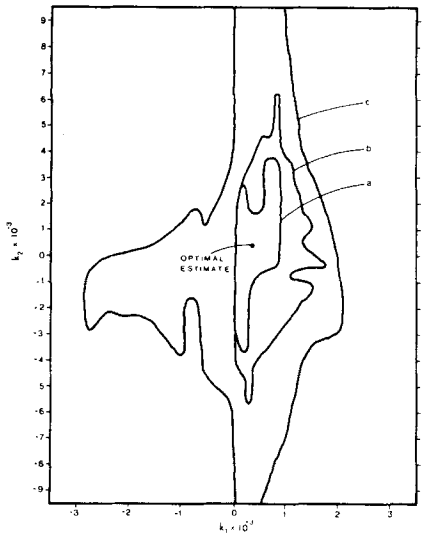

Figure 2. Effect of data length τ on the region of convergence:
a: $\tau = 169.7 \times 10^{-4}$
b: $\tau = 11.32 \times 10^{-4}$
c: $\tau = 5.63 \times 10^{-4}$.

curve fitting which can also be used to smooth the data when the noise level is high. In the present example, as shown in Figure 1, data points are not given for t less than 5.63×10^{-4}. By generating an artificial data point at $t = 2.815 \times 10^{-4}$ and applying quasilinearization algorithm using this generated data point and the first given data point, we obtain a substantial increase in the region of convergence. This is shown by curve c in Figure 2.

The rapid convergence associated with quasilinearization close to optimum is obtained even when the initial guess is far from the optimum values. Consider the case where as initial guess (1000, 9000) is chosen. As shown in Figure 3, the convergence to (352.6, 398.6) results in four iterations. Using subsequently all the data, omitting the artificial one, we achieve the optimum values (354.6, 400.2) in one further iteration.

The very fast convergence rate of quasilinerarization algorithm, requiring usually fewer than five iterations, combined with the fact that the main limitation of the method, the small region of convergence, is overcome, establish quasilinearization a very powerful method in parameter estimation. Moreover, the number of equations required for integration is $n(p+2)$, where n is the number of states and p the number of parameters; that is, only n more equations are required than Gauss-Newton method. Since quasilinearization requires fewer iterations to converge, it becomes computationally much more efficient than gradient methods.

EFFECT OF USING DIFFERENT SECTIONS OF THE DATA

Instead of restricting the use of data close to the initial state, one can just as well use different sections of the data during the first few iterations. For example, we may consider the data at $t = 11.32 \times 10^{-4}$ as constituting the initial state and the data at $t = 16.97 \times 10^{-4}$ as constituting the first data point. In other words, we may shift the data to allow any of the early data points to be the intial state.

By taking the data at $t = 11.32 \times 10^{-4}$ as initial state, the data at $t = 16.97 \times 10^{-4}$ as the first data point and by interpolation obtain an artificial data point at $t = 14.145 \times 10^{-4}$, the region of convergence is shown in Figure 4. As can be seen, the region of convergence remains enlarged, but the shape has changed somewhat.

However, the reliability of estimates of the parameters can be influenced by the choice of the section of the data during the early iterations. In order to examine the reliability of the first estimates of k_1 and k_2, new measurements y_1 and y_2 were generated from

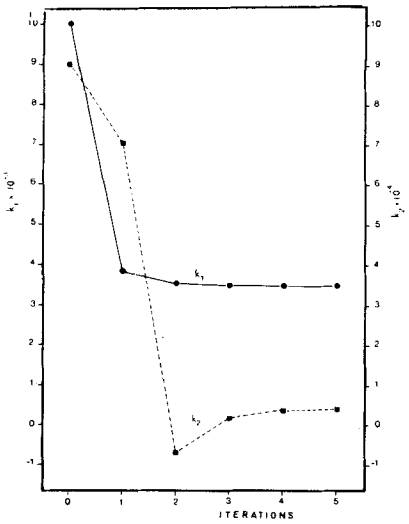

Figure 3. Convergence of parameter estimation.

$$y_i(t_j) = x_i(t_j)[1 + \sigma \cdot R_{ij}] \, , \; i = 1,2 \, , \; j = 1,2 \qquad (7)$$

where R_{ij} are random numbers uniformly distributed between -1 and 1, σ is the maximum noise level and $x_i(t_j)$ is the exact solution of Equations (1) and (2), obtained using 354.6 and 400.2 as the exact parameter values for k_1 and k_2, respectively. One hundred runs were performed for each different starting point, the results of which are shown in Table 1.

The standard deviation of k_1, a measure of the reliability in the estimate, increases gradually as the starting point is moved away

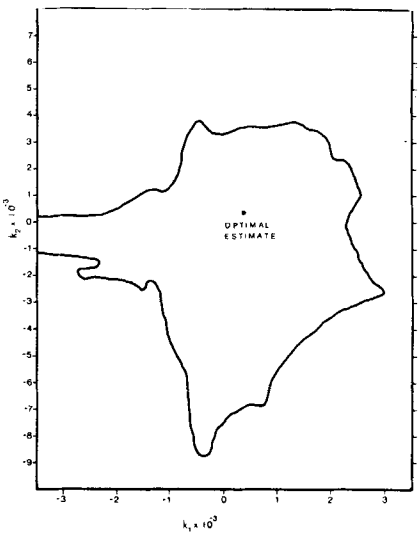

Figure 4. Region of convergence with initial state at $t = 11.32 \times 10^{-4}$; $\tau = 5.65 \times 10^{-4}$.

TABLE 1. RELIABILITY OF PARAMETER ESTIMATES AS DETERMINED BY SECTION OF DATA USED

$\sigma = 0.01$

Starting point	k_1	k_2
0	354.9 ± 7.4	403.1 ± 72.5
5.63×10^{-4}	354.9 ± 9.5	402.1 ± 33.1
11.32×10^{-4}	354.9 ± 13.1	402.2 ± 30.9
16.97×10^{-4}	355.7 ± 18.3	402.8 ± 34.7
34.00×10^{-4}	357.7 ± 55.1	407.3 ± 78.4

$\sigma = 0.02$

Starting point	k_1	k_2
0	355.3 ± 14.8	405.6 ± 145.1
5.63×10^{-4}	355.4 ± 19.1	404.2 ± 66.3
11.32×10^{-4}	355.6 ± 26.2	404.5 ± 61.9
16.97×10^{-4}	356.8 ± 36.7	405.7 ± 69.5
34.00×10^{-4}	363.8 ± 111.6	418.6 ± 159.1

from $t=0$. However, for k_2 the standard deviation at $t_o=0$ is as high as at $t_0=34 \times 10^{-4}$ and passes through a minimum at $t_0 = 11.32 \times 10^{-4}$. The selection of the initial state in this case should come as a trade off between gains in the reliability of k_2 and losses for k_1. It appears that a good compromise is obtained when the starting point $t = 11.32 \times 10^{-4}$ is used.

SUMMARY

Quasilinearization algorithm was applied to a typical chemical engineering parameter estimation problem. Using only portions of the data length during the early iterations, as proposed by Wang and Luus (1978), the size of the region of convergence was significantly enlarged. Without the problem of having to have a good initial guess, quasilinearization becomes a very powerful tool for estimating parameters, since it possesses the property of very fast convergence rate and is easy to program.

Furthermore, the effect of using different sections of the data during the early iterations was demonstrated. Although a method to obtain the optimum section of the data is not proposed, the effect of using different starting points on the reliability of the estimated parameter values, during the early iterations, was illustrated.

NOTATION

e_i = random experimental error associated with state i
i = index relating to state variable
j = index relating to point in time
J = sum of squares of deviations performance index given by Equation (6)
k_1 = rate constant relating to first reaction (parameter to be determined)
k_2 = rate constant relating to second reaction (parameter to be determined)
n = number of state variables
p = number of parameters to be determined
r_1 = rate of first reaction given by Equation (3)
r_2 = rate of second reaction given by Equation (4)
R_{ij} = random numbers uniformly distributed between -1 and 1
t = residence time
x_1 = moles of benzene per mole of benzene feed
x_2 = moles of diphenyl per mole of benzene feed
y_i = recorded measurement of x_i
σ = maximum noise level
τ = data length used for parameter estimation

LITERATURE CITED

Seinfeld, J. H., and G. R. Gavalas, "Analysis of Kinetic Parameters from Batch and Integral Reactor Experiments," *AIChE J.*, **16**, 644 (1970).

Seinfeld, J. H., and L. Lapidus, *Mathematical Methods in Chemical Engineering*, Vol. 3, Prentice Hall, Englewood Cliffs, N. J. (1974).

Wang, B. C., and R. Luus, "Modification of Gauss-Newton Method for Parameter Estimation to Expand the Region of Convergence," AIChE 71st Annual Meeting, Miami, Paper No. 78d, (Nov. 15, 1978).

———, "Increasing the Size of Region of Convergence for Parameter Estimation Through the Use of Shorter Data-Length," *Int. J. Control* (accepted for publication, 1980).

Manuscript received October 16, 1979; revision received December 21 and accepted January 21, 1980.

Photopolymerization in a Continuous Stirred-Tank Reactor: Experiment

H. T. CHEN
C. N. KUAN
S. SETTHACHAYANON
and
P. A. CHARTIER

Department of Chemical Engineering
New Jersey Institute of Technology
Newark, New Jersey 07102

Addition polymerizations are usually initiated through the use of catalysts. Alternate means of initiation are available, including initiation by absorption of radiation (ionizing or ultraviolet). Both catalyst decomposition and radiation absorption lead to the production of free radicals which initiate the polymerization reaction. However, there are differences between the initiation modes which can profoundly affect the ease with which the reaction is carried out and the character of the product. Thus, in the case of radiation initiation, the rate of initiation is essentially independent of temperature, and because of the physical separation of radiation source and the reaction medium, the initiation rate may be changed very rapidly. By contrast, the rate of initiation with chemical initiators is usually strongly temperature dependent, and because the chemical initiator is in solution in the reaction medium, the initiation rate cannot be changed very quickly. Because of these facts, radiation initiation may lead to greater reactor stability and greater ease of reactor control. An important consequence of greater stability and ease of reactor control should be less frequent instances of runaway reactions and explosive decompositions, or the elimination of these events altogether. This consequence would result in greater safety of operation and loss of production.

Polymerization processes initiated by the thermal decomposition of a catalyst are well established. But very little work has been done on the engineering aspect of photopolymerizations (Yemin and Hill, 1969; Jain et al., 1970; Chen and Hill, 1971; Hill and Chen, 1972; Sandru and Smith, 1973; Ibarra and Smith, 1974; Mendiratta et al., 1975; Chen and Steenrod, 1975), in part because of incomplete exploration of their characteristics and advantages. Notice that a number of studies of the design and

Correspondence concerning this note should be addressed to H. T. Chen.

0001-1541-80-3833-0672-$00.75.