Bondi, A., "Physical Properties of Molecular Crystals, Liquids, and Glasses," John Wiley & Sons, Inc., New York (1968).
Casellato, F., C. Vecchi, and A. Girelli, "Differential Calorimetric Study of Polycyclic Aromatic Hydrocarbons," *Thermochim Acta*, **6**, 361 (1973).
Finke, H. L., J. F. Messerly, S. H. Lee, A. G. Osborn, and D. R. Douslin, "Comprehensive Thermodynamic Studies of Seven Aromatic Hydrocarbons," *J. Chem. Thermodyn.*, **9**, 937 (1977).
Hildebrand, J. H., J. M. Prausnitz, and R. L. Scott, "Regular and Related Solutions," Van Nostrand Reinhold Co., New York (1970).
McLaughlin, E., and H. A. Zainal, "The Solubility Behavior of Aromatic Hydrocarbons in Benzene," *J. Chem. Soc.*, 863 (March, 1959).
Prausnitz, J. M., "Molecular Thermodynamics of Fluid-Phase Equilibria," Prentice-Hall, Inc. (1969).
Timmermans, J., "Physico-Chemical Constants of Pure Organic Compounds Vol. I and II," Elsevier Publishing Co., New York (1950/1965).
Weimer, R. F., and J. M. Prausnitz, "Complex Formation Between Carbon Tetrachloride and Aromatic Hydrocarbons," *J. Chem. Phys.*, **42**, 3643 (1965).

Manuscript received October 27, 1981; revision received March 29, and accepted April 19, 1982.

Avoiding Unwarranted Inflection Points in Fitting of Data

This paper presents a method for correlating univariable experimental data, which avoids unwarranted inflection points by means of mathematical constraints fitted to sections of the data. The method was applied successfully to correlate thermodynamic data which are otherwise difficult to fit in without inflection points.

YEHUDA TAITEL
School of Engineering
Tel-Aviv University
Tel-Aviv 69978, Israel

and

ABRAHAM TAMIR
Department of Chemical Engineering
Ben-Gurion University of the Negev
Beer-Sheva 84120, Israel

SCOPE

In the correlation of thermodynamic or other kinds of data, one usually strives to achieve simultaneously the following objectives: high accuracy and a minimal number of parameters in the series expansion. The latter is particularly important because, by increasing the number of parameters, inflection points tend to appear in unexpected locations due to mathematical reasons rather than physical. There are many physical cases, especially those of thermodynamic properties as a function of concentration, where functional dependence is known to be smooth and free of inflection points, either in the entire range or at several known sections. In such a case, correlation of empirical data that introduces false inflection points, caused by improper mathematical correlation or inaccurate noise bearing data, may cause serious distortion of the implied thermodynamic behavior. For example, when a search is made for azeotropic behavior of vapor pressure as a function of concentration, improper correlation may alter the implied azeotropic properties of such a system.

The available correlation methods (Landis and Nilson, 1962; Klaus and Van Ness 1967; Tamir, 1981) provide simplicity, interpolation ability, smoothness and low-ordered polynomials. However, no mathematical constraints were imposed in the process of fitting of the data so as to avoid the appearance of unwarranted inflection points. The object of this work is to establish a method which can avoid unwarranted inflection points. This is achieved by incorporating the method of "sectionwise fitting" while imposing the constraint that avoids inflection points. "Sectionwise fitting" is the term used for the method in which the entire range of data points is divided into sections which are correlated separately.

CONCLUSIONS AND SIGNIFICANCE

The principal significance of this work is the development of a method for correlating data by means of cubic polynomials with one independent variable, which incorporates mathematical constraints that would prevent the appearance of undesirable inflection points. In order to improve the goodness of fit, the method of "sectionwise fitting" is used.

The method has been applied successfully to vapor-liquid equilibrium data, which otherwise is difficult to correlate without inflection points. Correlation is obtained by minimizing the sum of the squares (S.O.S.) of deviations between the experimental points and the analytical representation, subject to constraint of continuity of the first derivative. The S.O.S. defines the loss function for inaccuracy; however, other loss functions may also be chosen, though the S.O.S. is the most common one. An additional constraint of smoothness of the second derivative can also be imposed. In spite of that, less accuracy is achieved by increasing the number of constraints, namely, the S.O.S. is increased.

This method is also useful for data exhibiting more than one extremum, for example, a binary mixture with two azeotropes or heat of mixing data with negative and positive values.

0001-1541-83-6606-0153-$2.00.

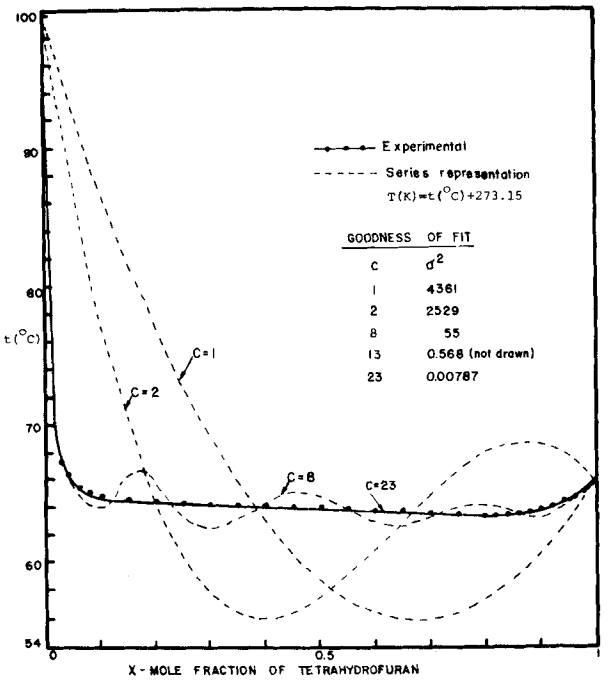

Figure 1. Experimental and series representation of the boiling temperature vs. concentration for tetrahydrofuran-water at 760 mm Hg (101,325 Pa).

A common method for correlating experimental data is by using a polynomial series expansion. In order to increase the accuracy of such a correlation, by minimizing the sum of squares (S.O.S.), a higher order polynomial with a large number of adjustable parameters is used. In such a case, however, unwarranted inflection points may appear in unexpected locations due to mathematical rather than physical reasons. This is clearly reflected in Figure 1, where boiling temperature at atmospheric pressure of mixtures of tetrahydrofuran-water (Hirata 1975) are plotted vs. composition. As can be seen from data points, the mixture exhibits azeotropic behavior. The dotted lines are calculated minimizing the S.O.S. which contain 1, 2 and 8 adjustable parameters in the polynomial series expansion. As observed, the fitting of the entire data by a polynomial expansion with one parameter, $c = 1$, is very poor. Increasing the number of parameters, c, improves the fit, but for $c = 8$, the series representation indicates that the system contains seven azeotropes, physically unrealistic. The use of $c = 13$ and $c = 23$ substantially reduces the S.O.S., but the function obtained exhibits an oscillatory behavior and the appearance of $c - 1$ azeotropes. A possible way to handle this problem is to use the spline-fit method, first introduced by Landis and Nilson (1962). This method uses local cubic interpolating polynomials between consecutive values of the interpolated data. By requiring continuity in the first and second derivatives at junction points, an overall fit of the data is achieved. However, since the curve must pass exactly through each data point, the spline-fit method accomplishes no smoothing of data. In order to achieve smoothing, Klaus and Van Ness (1967) extended the spline-fit method as follows: instead of defining the interval boundaries to pass through every data point, interval boundaries are determined arbitrarily, so that each interval may contain a number of data points. Moreover, constraints are placed on cubic polynomials, so that at every interval boundary the adjacent cubics give the same values for the dependent variable and its first two derivatives. Finally, the Lagrange's popular method of undetermined multipliers is applied to minimize the objective function yielding parameters of the cubic polynomial for each interval. This method undoubtedly proved useful in correlating thermodynamic data, otherwise difficult to correlate; however, the said method does not prevent the possible appearance of inflection points.

Recently, Tamir (1981) proposed a simple method which is based on "sectionwise-fitting" to explore azeotropic behavior of mixtures. By ignoring the need to maintain continuity between sections, the extension of "sectionwise-fitting" to ternary and multicomponent systems is quite straightforward. On the other hand, extension of the spline-fit to ternary mixtures is much more complicated, and the degree of complexity increases with the number of components. It was found that "sectionwise-fitting" reduces the occurrence of undesirable inflection points (undesirable azeotropes) but, as with the spline-fit, elimination is incomplete.

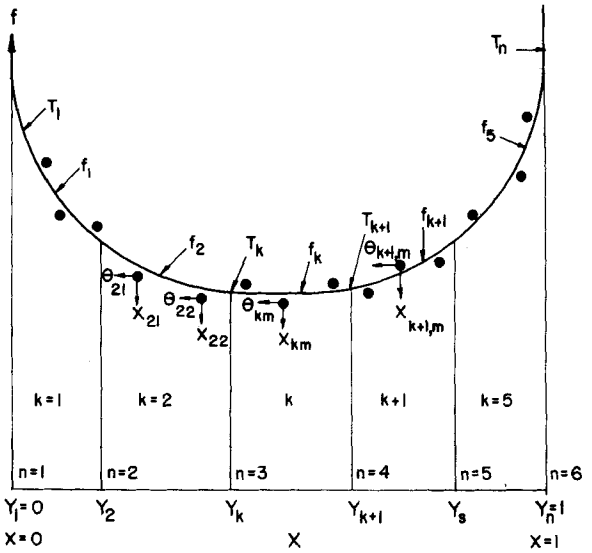

Figure 2. Nomenclature of the method.

ANALYSIS

It is desirable to find an analytical function $f(X)$ where X (usually concentration) ranges from 0 to 1 in such a way so as to fit well the experimental data, avoiding undesirable inflection points. The X-axis is divided into $(n-1)$ sections of somewhat arbitrary length, designated by the boundaries $Y_1, Y_2, \ldots Y_k \ldots Y_n$, shown in Figure 2. Note that $Y_1 = 0$ and $Y_n = 1.0$. Each section is fitted with a third order polynomial $f_k(X)$, valid for $Y_k < X < Y_{k+1}$ of the form:

$$f_k = \left[1 - 3\left(\frac{X-Y_k}{L_k}\right)^2 + 2\left(\frac{X-Y_k}{L_k}\right)^3\right]T_k + \left[3\left(\frac{X-Y_k}{L_k}\right)^2 - 2\left(\frac{X-Y_k}{L_k}\right)^3\right]T_{k+1} + \left[(X-Y_k) - 2\frac{(X-Y_k)^2}{L_k} + \frac{(X-Y_k)^3}{L_k^2}\right]C_k + \left[-\frac{(X-Y_k)^2}{L_k} + \frac{(X-Y_k)^3}{L_k^2}\right]C_{k+1} \quad (1)$$

This form is somewhat similar to the one used by Klaus and Van Ness (1967).

T_k and C_k are the adjustable parameters that correspond respectively to the values of the function f and its first derivatives at the boundaries Y_k. The fact that $f_1(0) = T_1$ and $f_{n-1}(1) = T_n$ are presumed to be known since they express the value of f (boiling temperature or pressure) of the pure substances.

Using Eq. 1, the first derivative is given by:

$$\frac{df_k}{dX} = \left[-6\frac{(X-Y_k)}{L_k^2} + 6\frac{(X-Y_k)^2}{L_k^3}\right]T_k + \left[6\frac{(X-Y_k)}{L_k^2} - 6\frac{(X-Y_k)^2}{L_k^3}\right]T_{k+1} + \left[1 - 4\left(\frac{X-Y_k}{L_k}\right) + 3\left(\frac{X-Y_k}{L_k}\right)^2\right]C_k + \left[-2\left(\frac{X-Y_k}{L_k}\right) + 3\left(\frac{X-Y_k}{L_k}\right)^2\right]C_{k+1} \quad (2)$$

while the second derivative is

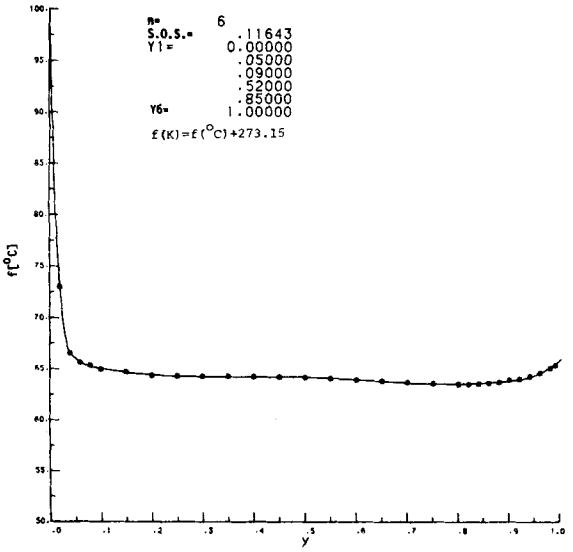

Figure 3. Correlation of data for the system tetrahydrofuran-water: function and its first derivative are smooth.

$$\frac{d^2f_k}{dX^2} = \left[-\frac{6}{L_k^2} + 12\frac{(X-Y_k)}{L_k^3}\right]T_k + \left[\frac{6}{L_k^2} - 12\frac{(X-Y_k)}{L_k^3}\right]T_{k+1} + \left[-\frac{4}{L_k} + 6\frac{(X-Y_k)}{L_k^2}\right]C_k + \left[-\frac{2}{L_k} + 6\frac{(X-Y_k)}{L_k^2}\right]C_{k+1} \quad (3)$$

where

$$L_k = Y_{k+1} - Y_k \quad (3a)$$

The sum of squares (S.O.S.) is defined as:

$$\sigma^2 = \sum_{k=1}^{n-1}\sum_{m=1}^{M_k}[\theta_{km} - f_{km}]^2 \quad (4)$$

where θ_{km} is the value of the m-th experimental point in the k interval ($Y_k < X < Y_{k+1}$), and f_{km} is the value of the function f at that X_{km} point (Figure 2). M_k is the number of data points in the k-th interval. Minimizing this function by adjusting T_k and C_k to satisfy

$$\frac{\partial\sigma^2}{\partial T_i} = 0 \qquad i = 2,3,\ldots n-1 \quad (5)$$

$$\frac{\partial\sigma}{\partial C_i} = 0 \qquad i = 1,2,\ldots n \quad (6)$$

yields a correlating fit that will satisfy smoothness of the function and its first derivative.

If desired, it is possible to impose too the smoothness of the second derivative. This smoothness requires that

$$\phi_k = \frac{d^2f_{k-1}(Y_k)}{dx^2} - \frac{d^2f_k(Y_k)}{dx^2} = 0 \quad k = 2,3,\ldots n-1 \quad (7)$$

Minimizing σ^2 subject to the constraint $\phi_k = 0$ is achieved by the method of Lagrange multipliers, namely solving:

$$\frac{\partial\sigma^2}{\partial T_i} + \sum_{j=2}^{n-1}\lambda_j\frac{\partial\phi_j}{\partial T_i} = 0 \qquad i = 2,3,\ldots n-1 \quad (8)$$

$$\frac{\partial\sigma^2}{\partial C_i} + \sum_{j=2}^{n-1}\lambda_j\frac{\partial\phi_j}{\partial C_i} = 0 \qquad i = 1,2,\ldots n \quad (9)$$

along with Eq. 7.

As previously mentioned, it is not always desirable to impose the continuity of the second derivative since the imposing of additional constraints will decrease the accuracy, namely increase σ^2.

So far no precautions were taken to avoid inflection points. If the solution obtained is free of inflection points, we consider the results to be satisfactory. If not, additional constraints should be imposed to avoid those unwarranted inflection points.

The procedure takes advantage of the fact that the function $f_k(X)$ is a polynomial of a third degree. Consequently, its second derivative d^2f_k/dX^2 is linear with respect to X. Avoiding inflection

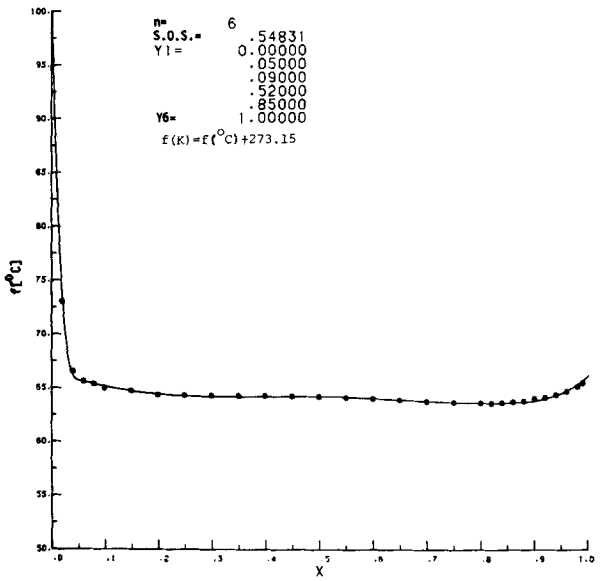

Figure 4. Correlation of data for the system tetrahydrofuran-water: function, its first and second derivatives are smooth.

points means that the function $f_k(X)$ should always be concave, namely,

$$\frac{d^2f_k}{dX^2} > 0 \quad (10)$$

or convex, namely,

$$\frac{d^2f_k}{dX^2} < 0 \quad (11)$$

Since d^2f_k/dX^2 is a linear function at $Y_k < X < Y_{k+1}$, it is possible to ensure satisfying conditions 10 or 11 throughout the entire interval by imposing:

$$\frac{d^2f_k}{dX^2} = 0 \quad (12)$$

only at one boundary point, Y_k or Y_{k+1}.

Thus, the procedure of the solution is as follows:

1) σ^2 is minimized by solving (Eqs. 5,6) or Eqs. (8,9), subject to the desirability of the smoothness of the second derivative. If the solution obtained is free of inflection points, the solution is satisfactory and the process of "sectionwise fitting" is terminated.

2) If inflection points do appear, when d^2f_k/dX^2 changes its sign, we impose the additional constraint (Eq. 12) at each boundary point, where the value of the second derivative is "wrong" (namely, negative for the concave case and positive for the convex case) and solve Eqs. 8 and 9 adding the additional constraints given by Eq. 12 to those given by Eq. 7. Although it is possible to eliminate all inflection points simultaneously, it is advisable to do so one by one, namely, by eliminating one inflection point at a time. This demands frequent solution of Eqs. 8 and 9 but the latter procedure, in this case, may sometimes provide a somewhat more accurate solution, for it is possible that the elimination of an inflection point may cause the elimination of a neighboring inflection point as well.

The above-mentioned procedure is somewhat lengthy and requires, of course, the use of a computer. Details are omitted here for the sake of brevity, but authors who may find difficulties in the computations should, therefore, contact the authors directly.

RESULTS AND DISCUSSION

The method was tested by correlating the vapor pressure at 30°C for ethanol-heptane (Klaus and Van Ness, 1967) and the equilibrium temperature of tetrahydrofuran-water at 760 mm Hg (Hirata, 1975, p. 490). Both systems exhibit azeotropic behavior and are believed to have a single azeotrope.

Figure 1 demonstrates, for the tetrahydrofuran-water system,

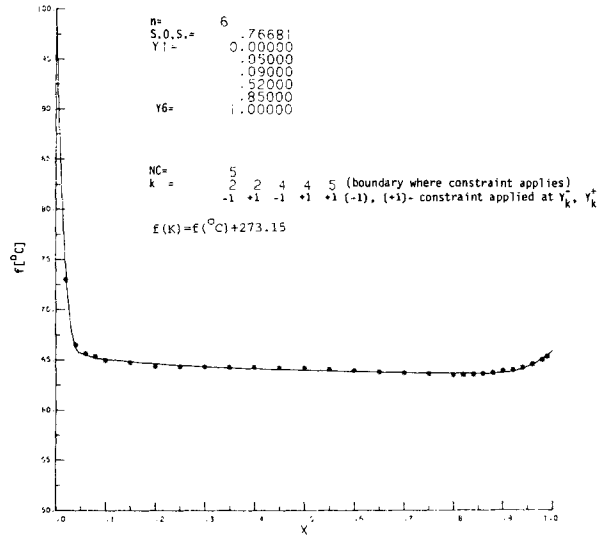

Figure 5. Correlation of data for the system tetrahydrofuran-water: function and its first derivative are smooth; inflection points are avoided.

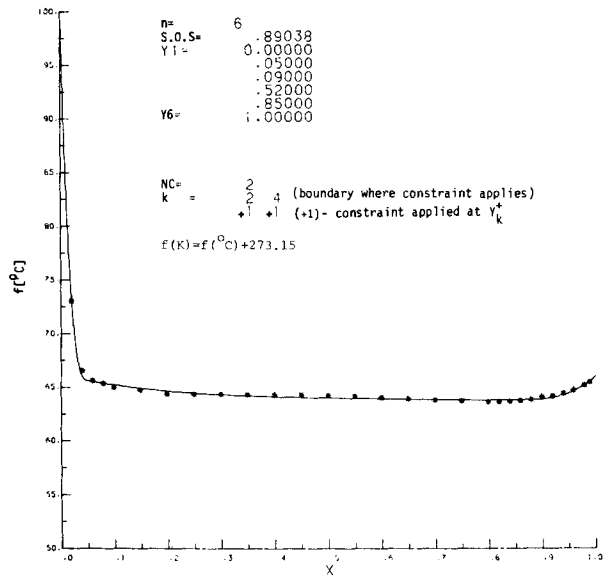

Figure 6. Correlation of data for the system tetrahydrofuran-water: function, its first and second derivatives are smooth; inflection points are avoided.

the inability of a single polynomial expansion to yield satisfactory results for this system, by showing that the number of inflection points is $(c - 1)$. Note that, although for $c = 23$ the correlation appears to be perfect, it is unacceptable owing to its 22 inflection points that do exist, though not noticeable in the figure. Figures 3 and 4 show the results of the first step of the procedure, namely, obtaining a correlation that minimizes σ^2 but which does not avoid inflection points. Figure 3 shows the results for a smooth function and its first derivative, while in Figure 4 the second derivative is also smooth. As seen in the latter case, σ^2 is larger owing to imposing additional constraints.

The problem of deciding the number of subdivisions and the location of the boundaries Y_k, although somewhat arbitrary and a matter of experience of the user of the method, deserves the following considerations. Usually one strives to use a minimum number of subdivisions so that the solution will be presented with a minimum number of parameters. On the other hand, by increasing the number of subdivisions, the representation of each section is more accurate, namely, the S.O.S. is reduced.

Once the number of sub-divisions has been chosen, the location of the boundaries, Y_k, may be determined by solving an optimization problem where the minimal value among all minimal values of σ^2 (according to Eq. 4) is searched for. This problem is a very time consuming one and our determination of the location of Y_k's was rather intuitive. One can consult Fisher's method (1958) for the above purpose.

A general rule to observe, however, is that small intervals should be used in the region where the function is steep, and large ones where the function is flat. In the example demonstrated in Figures 3 and 4, five intervals were used ($n = 6$) and the boundary locations were chosen at $Y_k = 0$, 0.05, 0.09, 0.052, 0.85 and 1. The results presented in Figures 3 and 4 appear to correlate the data quite well. Nevertheless, one can clearly see that both exhibit inflection points which is not permitted physically, and they appear here because of inappropriate experimental scatter.

Figures 5 and 6 are the results of the second step where unwarranted inflection points were suppressed and both figures exhibit concave curves that correlate the data quite well, though with somewhat higher S.O.S. (σ^2). Figure 5 shows the results for a case where continuity of the second derivative is not imposed. Elimination of inflection points was achieved by imposing $d^2f/dX^2 = 0$, at Y_2^-, Y_2^+, Y_4^-, Y_4^+ and Y_5^+. Since d^2f/dX^2 is discontinuous at the boundaries, it has different values both on the left and right sides of each boundary and one can impose a zero value of d^2f/dX^2 at Y_k^- (left side of the boundary) and Y_k^+ (right side of the boundary) separately. In the case of a continuous second derivative (Figure 6), the elimination of inflection points was achieved by requiring d^2f/dX^2 to be zero at Y_2 and Y_4. The results obtained are summarized in Table 1. The results for the ethanol-heptane system are also included in the Table though, for the sake of brevity, the figures for the latter system are omitted since it does not give any new insight.

In the two systems considered in the Table, inflection points do appear when the "normal" spline-fit method is used (cases 1 and 2). Imposing additional constraints to avoid the inflection points increases, as expected, the sum of squares, and thus decreases the apparent goodness of fit of the data. However, since the absence of inflection points is required by the physical phenomena, the final results of $f(X)$ are a better correlation of the data than those obtained in case 1 or 2.

Inflection points may appear due to mathematical reasons, namely, the best fit is one that incorporates inflection points, but it also depends largely on the particular data points used and the

TABLE 1. COMPARISON OF RESULTS

(Case 1: f and f' are smooth; Case 2: f,f' and f'' are smooth; Case 3: f,f' are smooth and inflection points are eliminated; Case 4: f,f',f'' are smooth and inflection points are eliminated.)

		Tetrahydrofuran-Water (Equilibrium Temperature $f(X = 0) = 100°C$, $f(X = 1) = 65.97°C$)[a]			Ethanol-Heptane (Equilibrium Pressure $f(X = 0) = 58.7$ mm Hg, $f(X = 1) = 78.8$ mm Hg)[b]		
Case	n	No. of Data Points	NC*	σ^2	No. of Data Points	NC*	σ^2
1	6	29	—	0.116	18	—	1.70
2	6	29	—	0.548	18	—	7.02
3	6	29	5	0.767	18	3	3.05
4	6	29	2	0.890	18	1	10.82

* NC = number of constraints needed to avoid inflection points (cases 3 and 4)

[a] $T(K) = t(°C) + 273.15$

[b] $P(Pa) = 133.32\ P\ (mm\ Hg)$

nature of their scatter.

Figures 3 and 4 show that, for the tetrahydrofuran-water system, the raw data indeed require the appearance of inflection points and, as a result, the correction of this anomaly by our method is necessitated. It is quite possible that for more accurate data inflection points will not appear, in which case no correction is needed.

The application of the present method to binary systems with two azeotropes—which are very rare—is quite straightforward. The curve of the boiling temperature versus composition exhibits, in this case, a maximum and a minimum and the application of this "sectionwise-fitting" is a natural extension. Similarly, heat of mixing data for systems which exhibit both negative and positive values may be correlated by "sectionwise-fitting" using this method.

ACKNOWLEDGMENT

Thanks to Professor Alexander Apelblat for his contribution to this work, and to Elisha Elijah for his splendid cooperation in editing this paper.

NOTATION

c = number of adjustable parameters
C_k = the derivatives at $X = Y_k$
f_k = analytical representation in the kth interval ($Y_k < X < Y_{k+1}$)
L_k = $Y_{k+1} - Y_k$
M_k = number of data points at interval k
n = number of boundaries
NC = number of constraints to avoid inflection points
$S.O.S.$ = σ^2 = sum of squares, Eq. 4
T_k = value of the dependent variable f at Y_k
X = independent variable
X_{km} = X of the mth experimental point between Y_k and Y_{k+1}
Y_k = the boundaries of intervals

Greek Letters

λ_j, λ_j' = Legrange multiplier
ϕ_j = constraint
θ_{km} = experimental temperature at X_{km}

Subscript

k = the k interval

LITERATURE CITED

Fisher, W. D., "On Grouping for Maximum Homogeneity," *Amer. Statistical Assoc. J.*, **53**, 789 (1958).

Hirata, M., S. Ohe, and K. Nagahama, *Computer-Aided Data Book of Vapor-Liquid Equilibria*, Kodansha Ltd., Elsevier Scientific Publishing Co. (1975).

Klaus, R. L., and H. C. Van Ness, "An Extension of the Spline-Fit Technique and Application to Thermodynamic Data," *AIChE J.*, 13(6), 1132 (1967).

Landis, F., and E. N. Nilson, "The Determination of Thermodynamic Properties by Direct Differentiation Technique," *Progress in International Research on Thermodynamics and Transport Properties*, 218, Academic Press, New York (1962).

Tamir, A., "Azeotropes Prediction by Sectionwise-Fitting," *Chem. Eng. Sci.*, **36**, 37 (1981).

Manuscript received June 8, 1981; revision received March 8, and accepted March 23, 1982.

R & D NOTES

Kinetic Analysis of a Non-Isothermal Reactor: Comments on a Note by Cooper and Alley

R. S. UPADHYE and V. G. FOX

Air Products and Chemicals, Inc.
Allentown, PA 18105

Cooper and Alley (1981) have recently published a technique for analyzing non-isothermal kinetic data for a plug flow reactor (PFR). The data needed for this analysis are the concentration and the temperature profiles along the length of the reactor. Since the temperature varies along the length of the reactor, a plot of $ln(f_x/f_o)$ vs. x where f_x and f_o represent the mole fractions of the reactant at length x and at the entrance of the reactor, respectively, against x will not be linear. Cooper and Alley present a transformation of the independent variable, x, such that the plot of $ln(f_x/f_o)$ versus the transformed x yields a straight line. The slope of this line represents k_o, the rate coefficient evaluated at the inlet temperature. Cooper and Alley have assumed previous knowledge of the value of the activation energy, and used this value in defining the transformation factors. They pointed out that knowledge of the activation energy, E, is necessary to use this technique. Further, they have stated that if the value of E is not available, a second set of measurements, starting at a different initial temperature may