A General Method of Computing the Derivative of Experimental Data

A. S. Lubansky and Y. Leong Yeow
Dept. of Chemical and Biomolecular Engineering, The University of Melbourne, Victoria 3010, Australia

Yee-Kwong Leong
School of Engineering, James Cook University, Townsville, Queensland 4812, Australia

S. Ranil Wickramasinghe and Binbing Han
Dept. of Chemical Engineering, Colorado State University, Fort Collins, CO 80523

DOI 10.1002/aic.10583
Published online September 6, 2005 in Wiley InterScience (www.interscience.wiley.com).

Many scientific and engineering investigations require the extraction of the first derivative from experimental data. Direct numerical differentiation is usually impractical because this amplifies the noise in the data, leading to unreliable results. This investigation shows that the problem of differentiating experimental data can be converted into one of solving an integral equation of the first kind. Tikhonov regularization is used to solve this integral equation, leading to a smooth first derivative. By using the built-in regularization parameter in the method noise amplification is kept under control. The performance of this method is demonstrated by applying it to data taken from the literature. © 2005 American Institute of Chemical Engineers *AIChE J,* 52: 323–332, 2006
Keywords: numerical differentiation, data interpolation, Tikhonov regularization, generalized cross-validation, inverse problem

Introduction

Performing numerical differentiation on a set of experimental data to obtain its rate of change or gradient is a routinely encountered operation. Practical examples include obtaining reaction rates from time–concentration data, converting measured spectra into derivative spectra, and calculating partial molar quantities from measured properties of mixtures. Despite its importance, there does not appear to be a generally accepted simple and reliable method for carrying out this differentiation operation. The present work aims to address this problem.

A major reason for the absence of a generally accepted method is that numerical differentiation is an ill-posed operation because it amplifies the unavoidable noise in the experimental data.[1] Thus direct differentiation by finite-difference approximation without taking precautions to control noise amplification can lead to unreliable or totally erroneous derivatives. Conversely, overfiltering the noise may suppress key features of the data.

Currently the most widely followed procedure for differentiating experimental data is to fit a set of polynomial or spline curves to the data and then to differentiate these curves analytically to obtain the required derivatives.[2,3] One such method is the popular Savitzky–Golay (SG) method, in which local least-squares regression is used to fit an nth order polynomial to each interior data point and m of its uniformly spaced neighboring points on either side.[4] Typically $n = 3$, 4, 5, or 6 and m varies from 2 to 12. Tables and general expressions have been developed that give the coefficients, and thus the derivatives, of the regression polynomials in terms of the $2m + 1$ data points.[4,5] This greatly simplifies the implementation of the SG method. One of the limitations of this method is that the data

Correspondence concerning this article should be addressed to A. S. Lubansky at a.lubansky@pgrad.unimelb.edu.au.

points must be uniformly spaced. Because derivatives are generated by differentiation of the fitted polynomials, noise amplification can become a serious issue. This is observed particularly in the second and higher derivatives. Where appropriate, the results generated by the SG method will be used to assess the performance of the method developed in the present investigation.

Instead of performing numerical operations on the experimental data to obtain the derivative directly the present investigation adopts an entirely different approach. The problem of data differentiation is converted into one of solving an integral equation of the first kind that gives the second derivative at a large number of closely and uniformly spaced points. This is then integrated to give the first derivative. One advantage of this approach is that it does not impose arbitrary assumptions about the functional form of the data. Another advantage is that there are many well-tested methods for solving this class of integral equations.[1] The Tikhonov regularization method, used in this investigation,[1] has a built-in regularization parameter that ensures a proper balance between retaining the essential features of the data and keeping noise amplification under control.

The mathematics of inverse problems and the performance of numerical methods developed to deal with them are often not physically obvious. Replacing the operation of differentiating a set of experimental data by one of solving an integral equation of the first kind may also appear to be counterintuitive. These are some of the difficulties confronting any would-be user of the technique described herein. For this reason, in the derivation of the equations of Tikhonov regularization, emphasis is placed on the physical meaning of the steps involved. Mathematical justification of these steps can be found in Engl et al.[1]

Governing Equations

The integral equation of the first kind

It is assumed that the experimental data are in the form of an x–y data set: $(x_1^M, y_1^M), (x_2^M, y_2^M), \ldots, (x_i^M, y_i^M), \ldots, (x_{N_D}^M, y_{N_D}^M)$ in ascending order of the independent variable x. y is the dependent variable for which the derivative $dy(x)/dx$ is required. N_D is the number of points in the data set. The superscript M signifies that these are experimentally measured quantities. For convenience, the first and second derivatives will be denoted by $r(x) = dy(x)/dx$ and $f(x) = d^2y(x)/dx^2$, respectively. $y(x)$ is related to these derivatives by the exact relationships

$$y^C(x) = \int_{x'=x_0}^{x} r(x')dx' + y_0 \tag{1a}$$

and

$$y^C(x) = \int_{x'=x_0}^{x} (x - x')f(x')dx' + y_0 + (x - x_0)r_0 \tag{1b}$$

where x_0 is an arbitrary reference point and, in this investigation, is taken to be x_1^M. Superscript C is used to distinguish the computed y from its experimentally measured counterpart. Equations 1a and 1b are, respectively, the one-term and two-term Taylor's series expansion about x_0 of $y(x)$ with the remainder term expressed in integral form.[6] They are to be solved for the unknown functions $r(x)$ and $f(x)$ together with the unknown constants $r(x_0) = r_0$ and $y(x_0) = y_0$. These are integral equations of the first kind and are well known to be ill posed. A numerical scheme, based on Tikhonov regularization, has been developed to solve Eq. 1b for $f(x)$, which is then integrated to yield $r(x)$. The same procedure can, with minor modifications, be applied to Eq. 1a to yield $r(x)$ directly.

Discretized equation

The experimental data can be separated into two known column vectors: $\mathbf{y}^M = (y_1^M, y_2^M, \ldots, y_i^M, \ldots, y_{N_D}^M)$ and $\mathbf{x}^M = (x_1^M, x_2^M, \ldots, x_i^M, \ldots, x_{N_D}^M)$. There is no requirement for $\mathbf{x}^M$ to be uniformly spaced. The span of the data $x_{N_D}^M - x_1^M$ is discretized into N_K uniformly spaced points $\mathbf{x}^C = (x_1^C = x_1^M, x_2^C, \ldots, x_j^C, \ldots, x_{N_K}^C = x_{N_D}^M)$ at distance $\Delta = (x_{N_D}^M - x_1^M)/(N_K - 1)$ apart.

The value of the unknown second derivative $f(x)$ at each point of $\mathbf{x}^C$ will be denoted by the column vector $\mathbf{f} = (f_1, f_2, f_3, \ldots, f_{N_K})$. To ensure that $\mathbf{f}$ is an accurate representation of $f(x)$, the number of discretization points N_K is usually set to be much larger than N_D. In the present investigation, typically $N_K = 151$ to 801. The discretized form of Eq. 1b is

$$y_i^C = \sum_{j=1}^{N_K} B_{ij}f_j + y_0 + (x_i^M - x_0)r_0 \qquad \text{for } i = 1, 2, \ldots, N_D \tag{2a}$$

or in matrix notation

$$\mathbf{y}^C = \mathbf{Bf} + \mathbf{1}y_0 + (\mathbf{x}^M - \mathbf{1}x_0)r_0 \tag{2b}$$

where $\mathbf{B}$ is an $N_D \times N_K$ matrix of known numerical coefficients arising from the approximation of the integral in Eq. 1b by numerical quadrature such as the trapezoidal or the Simpson's rule; $\mathbf{1}$ is a column vector with each element equal to 1. The unknowns $f_1, f_2, f_3, \ldots, f_{N_K}$, y_0, and r_0 are required to minimize:

(1) The sum of squares S_1 of the deviation of $y^C(x_i^M)$ from y_i^M for $i = 1$ to N_D, and

(2) The sum of squares S_2 of $d^2f(x)/dx^2$ at the interior discretization points x_j^C, $j = 2, 3, \ldots, N_K - 1$.

Condition (1) ensures that $y^C(x)$ approximates $y^M(x)$ closely and condition (2) ensures that $f(x)$ does not show spurious fluctuations.

Tikhonov regularization

In Tikhonov regularization, instead of satisfying conditions (1) and (2) separately, a linear combination $R = S_1 + \lambda S_2$ is minimized.[1] λ is the regularization parameter that balances these two requirements. A large λ favors the smoothness condition, whereas a small λ favors the accuracy condition. The appropriate choice of λ depends on N_D and N_K and also on the noise level in the data.[1] In this investigation generalized cross-validation (GCV) will be used to guide the selection of λ.[3] The

implementation of GCV will be described in the next subsection.

For any λ, the unknowns $f_1, f_2, f_3, \ldots, f_{N_K}$, y_0, and r_0 that minimize R are given by[1]

$$\mathbf{f}' = \left(\mathbf{B}'^T\mathbf{B}' + \frac{\lambda}{\Delta^4}\boldsymbol{\beta}^T\boldsymbol{\beta}\right)^{-1}\mathbf{B}'^T\mathbf{y}^M \tag{3}$$

For convenience $\mathbf{f}' = (f_1, f_2, f_3, \ldots, f_{N_K}, y_0, r_0)$ is used to denote all the unknowns and $\mathbf{B}'$ is the matrix $\mathbf{B}$ with the column vectors $\mathbf{1}$ and $\mathbf{x}^M - \mathbf{1}x_0$ added to it to reflect the incorporation of y_0 and r_0 into $\mathbf{f}$

$$\boldsymbol{\beta} = \begin{bmatrix} 1 & -2 & 1 & & & & 0 & 0 \\ & 1 & -2 & 1 & & & 0 & 0 \\ & & & \ddots & & & & \\ & & & 1 & -2 & 1 & 0 & 0 \end{bmatrix} \tag{4}$$

is the tridiagonal matrix arising from standard finite-difference approximation of the second derivatives $d^2f(x)/dx^2$. The two columns of 0 in Eq. 4 are again the consequence of the incorporation of y_0 and r_0, which do not feature in the smoothness condition, into $\mathbf{f}'$. In terms of $\mathbf{f}'$ and $\mathbf{B}'$

$$\mathbf{y}^C = \mathbf{B}'\mathbf{f}' \tag{5}$$

which is needed in the subsequent GCV computation.

Equation 3 is the linear algebraic equation that converts the experimental data $\mathbf{y}^M$ into the second derivative $f(x)$ described by $\mathbf{f}$. Because $f(x)$ is known at a large number of closely spaced points and is free of spurious fluctuations it can be integrated to give the first derivative $r(x)$ using any standard numerical integration procedure. In performing this integration the r_0 given by Eq. 3 is used as the boundary condition. Similarly, the resulting $r(x)$ is integrated to give the computed $y^C(x)$ with y_0 as the boundary condition. In this investigation the back-calculation of $r(x)$ and $y^C(x)$ are performed using commercial software independent of the computer codes developed to solve for $f(x)$. Thus the comparison of $y^C(x)$ with $y^M(x)$ serves as an independent check of the reliability of the $f(x)$ and $r(x)$ given by Eq. 3.

Generalized cross-validation

GCV, used to guide the selection of the regularization parameter λ, is based on the "leaving-out-one" principle.[3] The basic idea behind this can be summarized as follows. For a given λ the Tikhonov regularization computation is repeated N_D times, each time leaving out one data point. The sum of squares $V(\lambda)$ of the difference between the predicted value and the actual value for each of the left-out data points clearly depends on λ. In the "leaving-out-one" principle the optimum λ_{opt} is taken to be the minimizer of $V(\lambda)$. In the GCV implementation of the "leaving-out-one" principle $V(\lambda)$ is given by[3]

$$V(\lambda) = \frac{(\mathbf{y}^M - \mathbf{y}^C)^T(\mathbf{y}^M - \mathbf{y}^C)/N_D}{(1 - \mathrm{Tr}[\mathbf{A}]/N_D)^2} \tag{6}$$

Tr[$\mathbf{A}$] denotes the trace of the square matrix $\mathbf{A}$, known as the influence matrix, defined by[3]

$$\mathbf{A} = \mathbf{B}'\left(\mathbf{B}'^T\mathbf{B}' + \frac{\lambda}{\Delta^4}\boldsymbol{\beta}^T\boldsymbol{\beta}\right)^{-1}(\mathbf{B}')^T \tag{7}$$

Equations 3 and 5 together with the definition of $\mathbf{B}'$, $\boldsymbol{\beta}$, and $\mathbf{A}$ allow $V(\lambda)$ to be plotted against λ/Δ^4 and the λ_{opt} for a given data set to be determined.

Results

The reliability, simplicity, and generality of Eq. 3 will be demonstrated by applying it to a number of data sets. Apart from the first, all the data are from published literature. Because the experimental conditions of these data are documented in the original papers they will be omitted here. The derivatives given by Eq. 3 will be compared against those generated by the SG method or that were reported in the original papers. The back-calculated $y^C(x)$ will be checked against the original data or the smoothed function given by the SG method. The physical significance of the derivatives will not be discussed but the advantages of using $r(x)$ to analyze the experimental data will be briefly demonstrated.

Synthetic data

The discrete points in Figure 1a are data described by $y(x) = 3[1 - \exp(-x/2)]\cos x$. To simulate experimental measurements random noise, uniformly distributed between +20 and −20%, has been added to the data. In this data set $N_D = 51$ and the points are uniformly distributed at the intervals of $\Delta = 0.1$. This example allows the derivatives given by Eq. 3 to be compared against the exact derivatives $d^2y(x)/dx^2 = -0.75\exp(-x/2)\cos x - 3[1 - \exp(-x/2)]\cos x - 3\exp(-x/2)\sin x$ and $dy(x)/dx = 1.5\exp(-x/2)\cos x - 3[1 - \exp(-x/2)]\sin x$. The back-calculated $y^C(x)$ is checked against the exact original function. The added noise of ±20% is probably beyond that normally encountered in analytical measurements. The purpose of imposing such a large noise is to test the limits of the Tikhonov regularization procedure.

The second derivative given by Eq. 3 is shown as a continuous curve in Figure 1b. In this calculation $N_K = 201$. This large N_K, compared to $N_D = 51$, allows the $f(x)$ given by Eq. 3 to be treated as essentially a continuous curve and $r(x)$ to be back-calculated by simple numerical integration. The resulting $r(x)$ is shown as a continuous curve in Figure 1c. Similarly the $y^C(x)$ back-calculated from $r(x)$ is shown as a continuous curve in Figure 1a.

The GCV plot for the data in Figure 1a is shown in Figure 1d, where it can be seen that $\lambda_{opt}/\Delta^4 = 0.48$. Numerical experimentation indicates that, as long as λ is close to λ_{opt}, small changes to λ do not have significant effects on the derivatives given by Eq. 3. Approximate minimization of the GCV function $V(\lambda)$ is adequate as a guide in the choice of λ for this and also for all subsequent examples. There is no direct physical significance associated with this particular λ_{opt} because it depends on N_D, N_K, the noise level, and the distribution of the data points. It is therefore not helpful and may even be misleading to tabulate the λ_{opt} for the various data sets. The

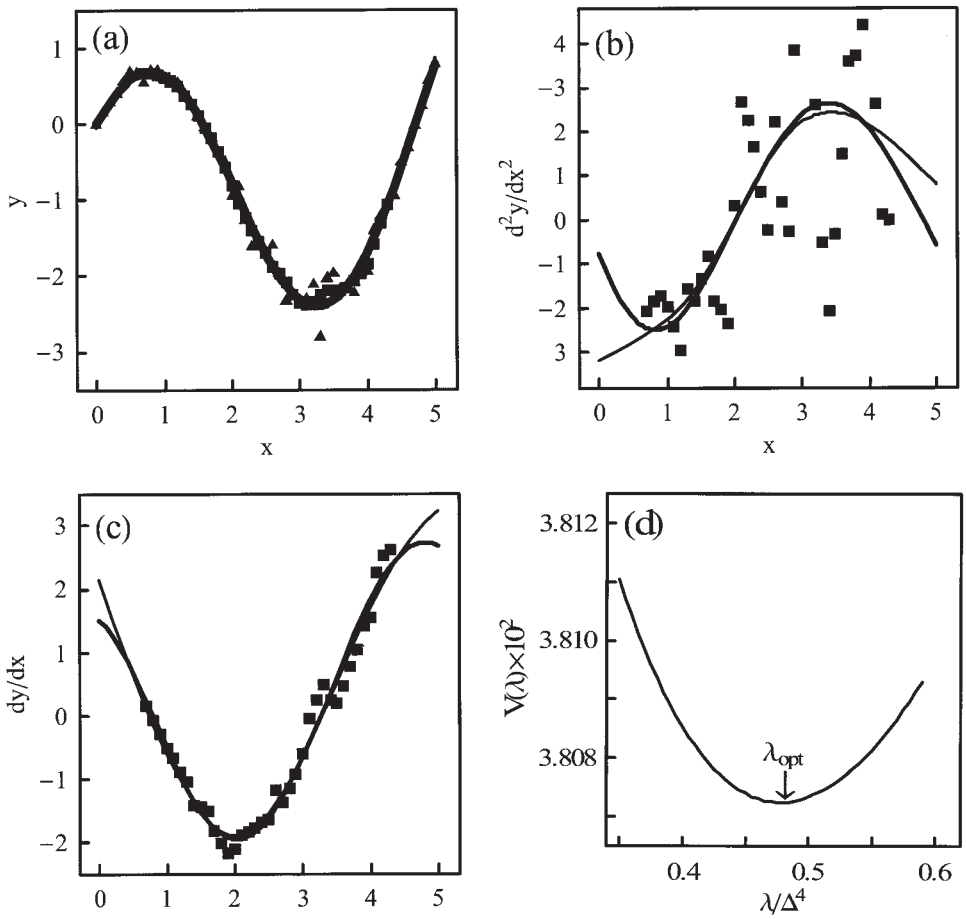

Figure 1. Data with added random noise.

(a) The lighter curve is the exact $y(x)$ and the darker curve is the $y(x)$ given by Eq. 3. ▲: data with 20% random noise; ■: smoothed data by the Savitzky–Golay (SG) method. (b) The lighter curve is the exact second derivative and the darker curve is that given by Eq. 3. ■: second derivative by the SG method. (c) The lighter curve is the exact first derivative and the darker curve is that given by Eq. 3. ■: first derivative by the SG method. (d) Plot of GCV function showing $\lambda_{opt}/\Delta^4 = 0.48$.

shape of the GCV plots for all the data sets investigated below is similar to that in Figure 1d and will therefore be omitted.

For comparison, the exact $y(x)$ and the exact derivatives are shown in Figures 1a–1c, respectively as lighter curves. Given that the exact $y(x)$ coincides very closely with the back-calculated $y^C(x)$, the lighter curve has to be thickened for it to show up clearly. It can be seen from Figures 1b and 1c that, apart from the two ends, Eq. 3 has reliably succeeded in extracting the derivatives from the "noisy" data. The cause of the significant error at the two ends is not clear. A possible contributory factor is the particular pattern of the imposed random noise. Another possible reason is the way the two end points were treated. Instead of excluding the two end points from the minimization of $d^2f(x)/dx^2$ alternatives, imposing $d^2f(x)/dx^2 = 0$ at the end points should be considered. Justification and performance of such alternatives depend on the physical nature of the data. These have not been investigated.

The SG method has been applied to smooth the "noisy" data in Figure 1a and to yield the first and second derivatives. The results of this method are shown in Figures 1a–1c as filled squares. In the present simple implementation of the SG method fourth-order polynomials with 15 regression points (that is, $n = 4$ and $m = 7$) at distance $\Delta = 0.1$ apart were fitted locally to the noisy data. From Figure 1a it is clear that both the SG method and Eq. 3 have successfully smoothed the noisy data. For the second derivative, Figure 1b shows that the SG method has not performed well, which is a consequence of noise amplification by successive differentiation. Numerical experimentation indicates that for the SG method to yield a reliable second derivative the number of regression points for each of the local fourth-order polynomials has to be increased to >19. This has the undesirable effect of reducing the range of x covered by the derivative by 10 intervals, a significant reduction out of a total of 50 intervals, from each end of the data.

From a comparison of the first derivatives in Figure 1c it is clear that the fourth-order 15-point SG method has succeeded in capturing the general trend of the first derivative. Apart from the reduction in the range of x covered, the first derivative of the SG method shows significant noise compared to the relatively smooth curve given by Eq. 3. It is these local fluctuations that caused the failure of the SG method at the second derivative level. It should be noted that the relatively poor performance of the SG method in this example is atypical. This is caused by the unrealistically high noise level in the data.

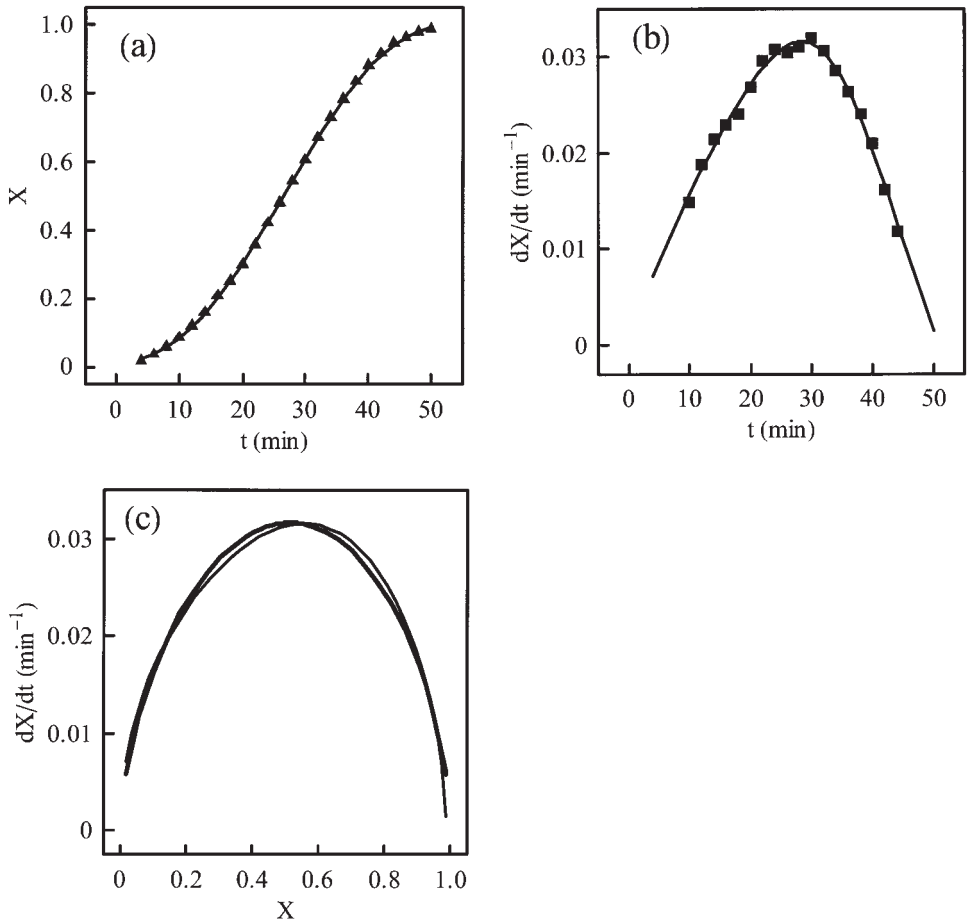

Figure 2. Reduction of copper oxide particles.

(a) Variation of fractional completion of reduction with time. Points are the data of Hamada et al.[7] and the curve is back-calculated from the best-fit rate Eq. 8. (b) Reduction rate vs. time. The curve is generated from Eq. 3 and the discrete points (■) are obtained by the SG method. (c) The continuous dark curve is the model-independent reduction rate plotted against fractional conversion and the lighter curve is the best-fit Eq. 8.

Reduction kinetics of copper oxide particles

Hamada et al.[7] investigated the reduction of copper oxide particles to copper by hydrogen. A set of their data showing the fractional reduction X against time is reproduced as discrete points in Figure 2a. These are for 0.46-μm oxide particles at 81°C under hydrogen partial pressure of 101 kPa and flow rate of 400 cm^3 min^{-1}. The reduction rate dX/dt given by Eq. 3 is plotted against t in Figure 2b and against the back-calculated X in Figure 2c. For comparison, the SG method has also been applied to obtain the reduction rate dX/dt as a function of time and the results are shown as filled squares in Figure 2b. These rate data were obtained using fourth-order polynomials with seven regression points. There is good general agreement between the SG rate and that given by Eq. 3, with the former again exhibiting significant local fluctuations.

Following Hamada et al., a rate model of the form

$$\frac{dX}{dt} = k(1 - X)^{\alpha}X^{\beta} \tag{8}$$

is used to describe the kinetics of the reduction.[7] This model is also suggested by the shape of the computed X vs. dX/dt curve in Figure 2c. The parameters of this semiempirical model are the rate constant k and the indices α and β. Least-squares regression between Eq. 8 and the computed X vs. dX/dt curve leads to $k = 0.074661$ min^{-1}, $\alpha = 0.60196$, and $\beta = 0.63412$. The best-fit Eq. 8 is shown as a lighter curve in Figure 2c. Because the computed rate in Figure 2c is a relatively simple curve the regression computation is straightforward and trouble free. This simple regression computation should be compared against the regression steps reported by Hamada et al.[7] They first fitted a fifth-order polynomial to the measured t vs. X^M data and then differentiated this polynomial to give dX/dt. This is followed by regression computation between dX/dt, X, and $(1 - X)$ to determine the kinetic parameters. They reported: $k = 0.069$ min^{-1}, $\alpha = 0.52$, and $\beta = 0.60$. These are in acceptable agreement with the results of the present investigation.

Even though the fitted curve in Figure 2c shows perceptible deviation from the computed curve given by Eq. 3, the t vs. X^C relationship back-calculated from the best-fit Eq. 8, shown as a continuous curve in Figure 2a, is in very good agreement with the original data with a correlation coefficient > 0.999. This is an indication that regression computation based on the com-

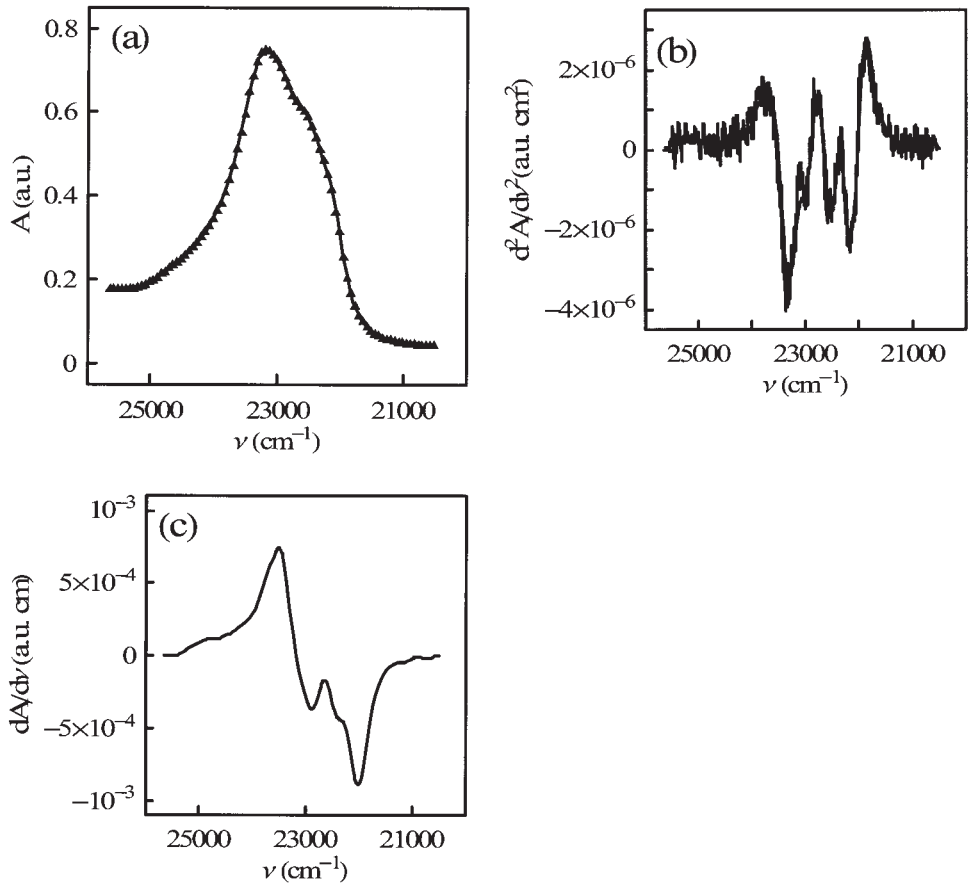

Figure 3. Spectra of cytochrome c oxidase.

(a) Points are the digitized spectrum of Horvath et al.[9] and the continuous curve is back-calculated from Eq. 3. (b) The lighter curve is the second-derivative spectrum reported by Horvath et al. and the darker curve is that given by Eq. 3. (c) First-derivative spectrum given by Eq. 3.

puted rate is not only a simple but also a more sensitive way of determining the kinetic parameters in the assumed rate model.[8]

Derivative spectrum of cytochrome c oxidase

It is a standard practice to convert measured spectra into first-, second-, and even higher-derivative spectra to reveal hidden information in the original spectra.[9] The SG method is the standard procedure for computing derivative spectra.[4] Because noise amplification is a serious problem for higher derivatives, elaborate and repeated filtering often has to be performed to suppress noise. In contrast, Eq. 3 gives the first- and second-derivative spectra directly.

The discrete points in Figure 3a are the absorption spectrum $A^M(\nu)$ of cytochrome c oxidase digitized from a spectrum reported by Horvath et al.,[10] $N_D = 332$ in this digitized version of the spectrum. The second- and first-derivative spectra given by Eq. 3 are shown as continuous curves in Figures 3b and 3c, respectively. The back-calculated absorption spectrum is shown as a continuous curve in Figure 3a. This is in excellent agreement with the digitized spectrum and the correlation coefficient is again >0.999.

To verify its reliability, the second-derivative spectrum given by Eq. 3 is compared against that obtained by Horvath et al. using the SG method. Their second-derivative spectrum is shown as a lighter curve in Figure 3b. Apart from the high-frequency fluctuations exhibited by the lighter curve, the two second-derivative spectra are in good agreement in terms of general trends, location of the major turning points, and their corresponding values. The high-frequency fluctuations exhibited by the derivative spectrum of Horvath et al. are likely to be the artifacts of their computation process rather than the genuine spectral response of the cytochrome c oxidase molecules.

Partial molar volumes of water–methanol system

Sandler[11] tabulated the density of a water–methanol mixture at 25°C as a function of the mole fraction of water x_w. His data are converted into specific molar volume v and plotted against x_w, as discrete points, in Figure 4a. These data can be converted into the partial molar volume of water and of methanol using the following relationships[11]

$$\bar{v}_w = v + (1 - x_w)\frac{\partial v}{\partial x_w} \qquad \text{and} \qquad \bar{v}_m = v - x_w\frac{\partial v}{\partial x_w} \tag{9}$$

Temperature and pressure are held fixed in the partial derivative $\partial v/\partial x_w$.

Equation 3 is used to compute $\partial v/\partial x_w$ from the data in Figure

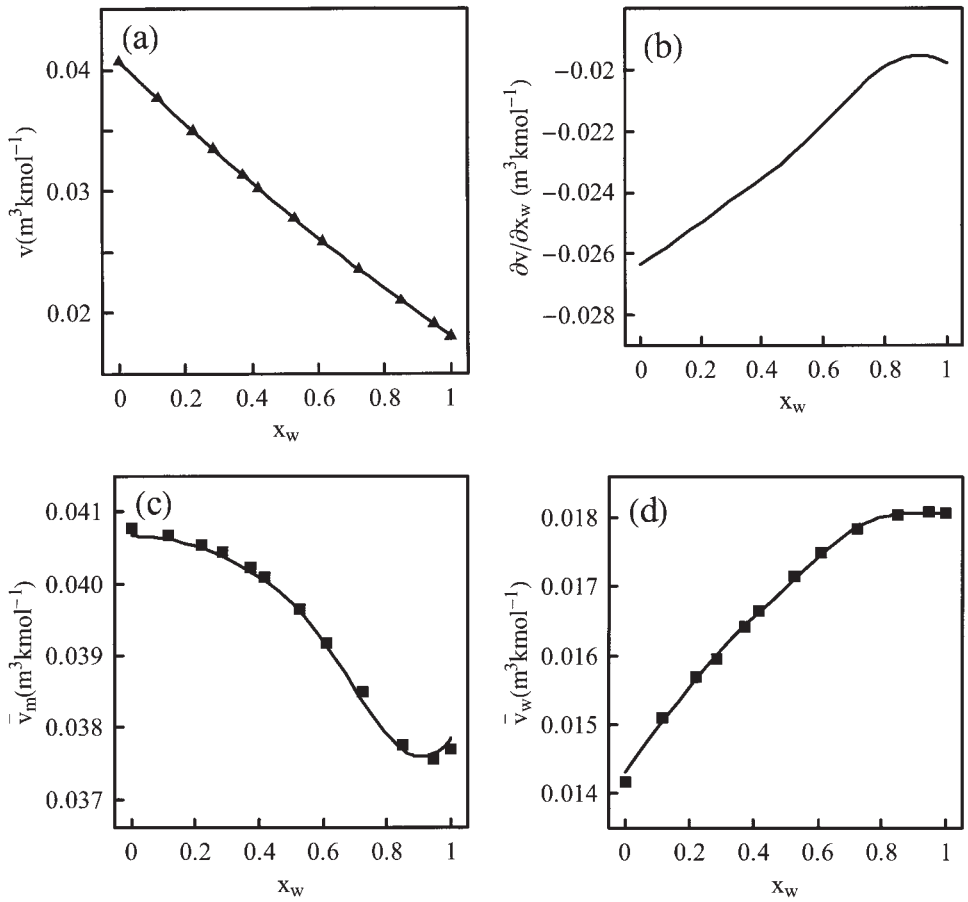

Figure 4. Molar volumes of water–methanol system.

(a) Discrete points are molar volume based on the tabulated data of Sandler[12] and the continuous curve is back-calculated from Eq. 3. (b) First derivative of molar volume with respect to mole fraction of water given by Eq. 3. (c) and (d) The continuous curves are the partial molar volumes of water and of methanol computed from Eq. 10 and the discrete points (■) are the corresponding results from Sandler.[12]

4a and the outcome is shown as a continuous curve in Figure 4b. As in the previous examples, the back-calculated molar volume is shown as a continuous curve in Figure 4a. The partial molar volumes $\bar{v}_w$ and $\bar{v}_m$ given by Eq. 9 are shown as continuous curves in Figures 4c and 4d. For comparison the corresponding partial molar volumes reported by Sandler[11] are shown as discrete points on the same plots. Sandler obtained these partial volumes by fitting a fifth-order polynomial to the excess molar volumes of the mixtures and then differentiating the fitted curve. The close agreement between the two sets of partial molar volume curves, obtained by different methods, confirms the reliability of Eq. 3 as a method for differentiating experimental data.

In the construction of the partial molar volume curve by excess molar volumes, the density of the mixture at any mole fraction is usually measured to $\pm 10^{-4}\%$ accuracy.[11] Such stringent accuracy is required to ensure that the error bars associated with the excess volumes, obtained as the difference between two density readings that are almost equal to one another, do not blow up to an unacceptable level. The route to partial molar volume by Eq. 3 does not involve such differences and therefore the data accuracy requirements are far less stringent. This should allow significant simplification in the density measurement techniques. These observations also apply to the determination of other partial molar quantities such as partial molar enthalpy.

Isothermal Joule–Thomson coefficient of carbon dioxide

The isothermal Joule–Thomson coefficient ϕ of a fluid can be computed from experimentally measured isobaric specific volume data v using the thermodynamic relationship[11]

$$\phi = -T^2\left[\frac{\partial(v/T)}{\partial T}\right]_P \tag{10}$$

after using Eq. 3 to evaluate the derivative required. The discrete points in Figure 5a show the measured density of carbon dioxide by Ihmels and Gmehling[12] at approximately constant pressure of 15 MPa (varying from 14.978 to 14.984 MPa) replotted as v/T vs. T. This is then converted into second and first derivatives using Eq. 3 and shown as continuous curves in Figures 5b and 5c.

The first derivative in Figure 5c is substituted into Eq. 10 to give the isothermal Joule–Thomson coefficient and this is shown as a continuous curve in Figure 5d. Following the same procedure, the data of Ihmels and Gmehling,[12] at $P = 10$ MPa

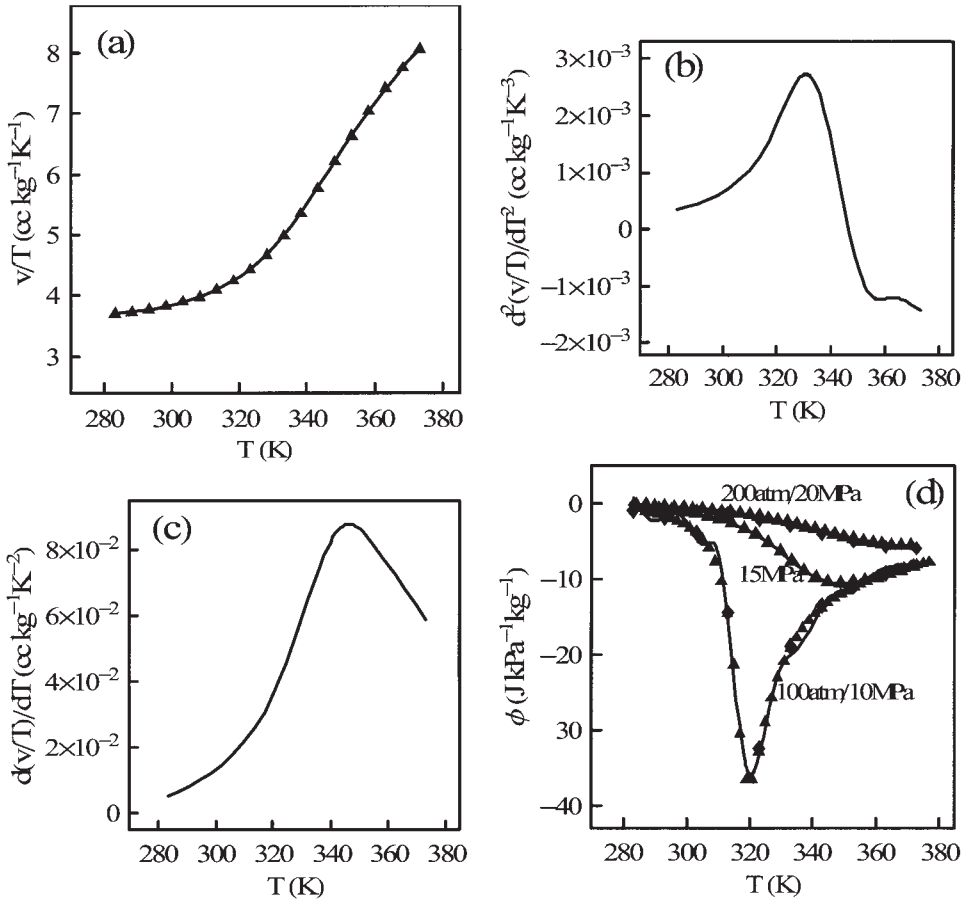

Figure 5. Isobaric specific volume data of carbon dioxide.

(a) Discrete points are v/T based on the data of Ihmels and Gmehling[12] and the continuous curve is back-calculated from Eq. 3. Nominal P = 15 MPa. (b) and (c) Second and first derivatives at 15 MPa of v/T given by Eq. 3. (d) Isothermal Joule–Thomson plots for different P values. Continuous curve: Eq. 3, ◆: Forsythe[13]; ▲: NIST.[14]

(varying between 9.984 and 10.025 MPa) and 20 MPa (varying from 19.975 to 19.985 MPa), are converted into ϕ curves and shown on the same figure. For comparison the adiabatic Joule–Thomson coefficient μ data for CO_2 at 100 and 200 atm reported in Forsythe[13] are converted into ϕ and shown as discrete points in Figure 5d. In converting μ to ϕ the following simple relationship was used: $\phi = -\mu C_p$. For this purpose the C_p was taken from the NIST database.[14] The tabulated μ values for CO_2 at 10, 15, and 20 MPa from the NIST database[14] have also been converted into ϕ and are shown as labeled curves in Figure 5d.

It can be seen that at 15 and 20 MPa the ϕ from different sources are in good agreement, thus verifying the reliability of Eq. 3. At 10 MPa, the agreement is less satisfactory. It is likely that this is caused by the small variations in pressure of the data of Ihmels and Gmehling.[12] For example, when Eq. 3 is applied to the exact isobaric specific volume data of Span and Wagner,[15] the resulting ϕ curves are in excellent agreement with that computed from the adiabatic Joule–Thomson coefficient μ tabulated in Forsythe[13] and the NIST database.[14] The small variations in pressure, from 10 MPa, in the data of Ihmels and Gmehling can be regarded as experimental noise and the resulting ϕ curve shows the residual noise after the smoothing effects by the regularization parameter λ, determined by GCV. It is with the aim of demonstrating the effects of data noise that Eq. 3 was applied to these approximately isobaric data instead of the exactly isobaric interpolated data of NIST or the state-equation–based data of Span and Wagner.

Discussion

From the applications described above it is clear that Eq. 3 is the key equation for converting experimental data into derivative curves. The derivation of this equation may appear lengthy, but Eq. 3 itself is relatively straightforward to apply and all the numerical operations in setting up the matrices and their subsequent manipulations can be performed using standard commercial computer software.

In deriving Eq. 3 no assumptions were made regarding the functional form of the relationship between the variables x and y. Thus—irrespective of the nature of the experimental data—the same equation can be used. This has been borne out by the examples described above. Back-calculations and comparison with derivatives obtained by other methods confirm the reliability of this equation. Equation 3 does not require the data to be uniformly spaced. This feature is particularly important in that very often experimental measurements are not available at uniformly spaced intervals.

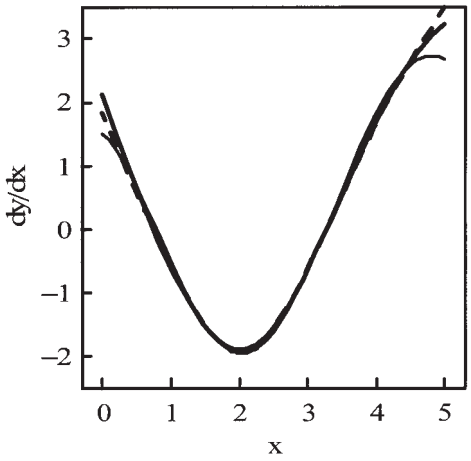

Figure 6. First derivatives of the data in Figure 1.

Lighter curve: exact first derivative; darker curve: first derivative by the second derivative given by Eq. 3; broken curve: first derivative given directly by the solution of Eq. 1a.

Apart from the applications reported herein, Eq. 3 or equations similar to it have been applied with some success to a number of practical situations where the derivatives of noisy experimental data are required for different purposes.[16,17] In some of these applications the Tikhonov regularization–GCV procedure is applied to Eq. 1a to give the first derivative directly instead of by the second derivative as was done in this study. In other applications, such as in pendant droplet tensiometry, where both the first and second derivatives of the droplet profile are required in the inverse problem of deducing the surface tension from measured droplet radius vs. height data, Eq. 1b then becomes the relevant equation to consider. In derivative spectrometry where the third- or fourth-derivative spectra are required,[9] the Tikhonov regularization–GCV procedure is applied to the integral equation obtained by the three- or four-term Taylor's series expansion with the remainder term in integral form. As the order of the derivative is increased so does the amount of the computation involved and the reliability of the derivatives can be expected to decrease. It is therefore interesting to compare the first derivative given directly by Eq. 1a with that back-calculated from the second derivative given by Eq. 1b. The dark and light continuous curves in Figure 6 are the same as those shown in Figure 1c, that is, the first derivative back-calculated from the second derivative and the exact first derivative of the simulated data in Figure 1a. The broken curve is the first derivative given directly by applying Tikhonov regularization–GCV to Eq. 1a. It can be seen that the two computed first derivatives approximate the exact derivative equally well. They differ from one another and from the exact derivative only at the two ends of the data range. This example indicates that the indirect route to the first derivative adopted in this investigation has not caused significant degradation in quality.

Recently Eilers described a method for smoothing experimental data.[18] His method has many features in common with the method described herein. Essentially for data smoothing the $\mathbf{B}'$ in Eq. 3 is replaced by an identity matrix $\mathbf{I}$ of the appropriate size. Eilers also applied the "leaving-out-one" principle to guide the choice of the regularization parameter. As has been demonstrated in the applications considered above, Eq. 3 followed by the necessary back-calculation can also be used to smooth noisy experimental data. The converse process of differentiating the smoothed functions of Eilers to obtain the derivative is generally not recommended because of noise amplification associated with differentiation.

In all the applications described above λ_{opt} was determined from GCV function plots similar to that in Figure 1d. Construction of the GCV plot is the most time consuming step in the entire process of converting a set of experimental data into a derivative curve. Even though the typical computing time required is not excessive, no more than 10 to 20 min on a 1-GHz PC, it is the price that one has to pay for completely automating the choice of λ, although this can be avoided. With some practical experience in Tikhonov regularization and, more importantly, guided by physical understanding of the nature of the data set under investigation it is quite possible to arrive at the appropriate λ without having to construct the GCV curve.

For example, instead of GCV, λ can be chosen so that the average and maximum difference between $y^C(x)$ and $y^M(x)$ are of the same order of magnitude as the error bars of the data. This is often referred to as the Morozov principle.[1] This principle combined with physically based expectation of the behavior of the derivative curves can serve as an alternative guide to the optimum λ. Figure 7 shows the plots of d^2X/dt^2 for the data of Hamada in Figure 2a for $\lambda/\Delta^4 = 0.017$ (light wavy curve), 0.34 (dark continuous curve), and 6.8 (thick broken curve). Based on the physical expectation that d^2X/dt^2 should vary smoothly with t, it can be concluded that $\lambda/\Delta^4 = 0.017$ is too small. On the other hand $\lambda/\Delta^4 = 6.8$ may be slightly too large because a substantial proportion of the resulting d^2X/dt^2 (and dX/dt not shown) varied almost linearly with t. The GCV function is at a minimum for $\lambda/\Delta^4 = 0.34$ and this is the parameter used in Figure 2. For GCV to perform satisfactorily N_D should be larger than about 25 to 30.[3] In many practical situations this condition is not satisfied so alternative ways of choosing λ, such as the Morozov principle, should be explored.

Finally, it is relevant to reiterate that the derivative generated by Eq. 3 is not very sensitive to small changes of the regularization parameter around λ_{opt}, and thus consequently fine tun-

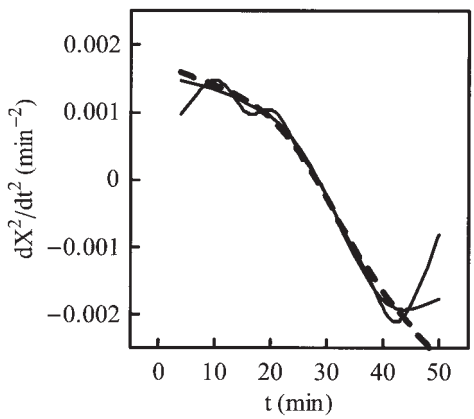

Figure 7. Second derivatives of Hamada et al. data.[7]

$\lambda/\Delta^4 = 0.017$ (light wavy curve, too small), 6.80 (thick broken curve, too large), 0.34 (dark continuous curve, GCV optimum).

ing of λ is generally not required. For example, for the copper oxide reaction, least-squares fitting of the computed rates for the three λ values in Figure 7 resulted in the following rate equations

$$\lambda/\Delta^4 = 0.017 \qquad \frac{dX}{dt} = 0.0740013(1 - X)^{0.596079}X^{0.629143} \tag{11}$$

$$\lambda/\Delta^4 = 0.34 \qquad \frac{dX}{dt} = 0.0746683(1 - X)^{0.602038}X^{0.634162} \tag{12}$$

$$\lambda/\Delta^4 = 6.80 \qquad \frac{dX}{dt} = 0.0754342(1 - X)^{0.609496}X^{0.639866} \tag{13}$$

For the 400-fold change in λ/Δ^4, the three sets of kinetic parameters are not significantly different from one another and the $X(t)$ curves back-calculated from them are in fact indistinguishable from one another. They differ from one another only in the first derivative and more so in the second derivative, as seen in Figure 7. This is consistent with the assertion that computed derivative data provide a more sensitive route to the kinetic parameters.

Conclusions

Tikhonov regularization provides a simple and reliable method for converting experimental data into derivative curves. The procedure is independent of the nature of the data and can cope with nonuniformly distributed data points. The GCV function or the Morozov principle can be used to guide the choice of the regularization parameter. The method can be extended to higher derivatives and can also be used to smooth experimental data.

Literature Cited

1. Engl HW, Hanke M, Neubauer A. *Regularization of Inverse Problems*. Dordrecht, The Netherlands: Kluwer; 2000.
2. Dierckx P. *Curve and Surface Fitting with Splines.* Oxford, UK: Oxford Univ. Press; 1993.
3. Wahba G. *Spline Models for Observational Data*. Philadelphia, PA: SIAM; 1990.
4. Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem.* 1964;36:1627-1639.
5. Madden HH. Comments on Savitzky–Golay convolution method for least-squares fit smoothing and differentiation of digital data. *Anal Chem.* 1978;50:1383-1386.
6. Burkill JC. *A First Course in Mathematical Analysis*. Cambridge, UK: Cambridge Univ. Press; 1978.
7. Hamada S, Kudo Y, Tojo T. Preparation and reduction kinetics of uniform copper particles from copper(I) oxides with hydrogen. *Colloids Surf.* 1992;67:45-51.
8. Yeow YL, Wickramasinghe SR, Han B, Leong Y-K. A new method of processing the time–concentration data of reaction kinetics. *Chem Eng Sci.* 2003;58:3601-3610.
9. Talsky G. *Derivative Spectrophotometry: Low and High Orders*. Weinheim, Germany: VCH; 1994.
10. Horvath MP, Copeland RA, Makinen MW. The second derivative electronic absorption spectrum of cytochrome *c* oxidase in the Soret region. *Biophys J.* 1999;77:1694-1711.
11. Sandler SI. *Chemical and Engineering Thermodynamics*. 3rd Edition. New York, NY: Wiley; 1999.
12. Ihmels EC, Gmehling J. Densities of toluene, carbon dioxide, carbonyl sulfide, and hydrogen sulfide over a wide temperature and pressure range in the sub- and supercritical state. *Ind Eng Chem Res.* 2001;40:4470-4477.
13. Forsythe WE. *Smithsonian Physical Tables*. 9th rev. Edition. Norwich, NY: Knovel; 2003.
14. NIST Chemistry WebBook. http://webbook.nist.gov/chemistry/. Boulder, CO: National Institute of Standards and Technology; 2003.
15. Span R, Wagner W. A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 K at pressures up to 800 MPa. *J Phys Chem Ref Data.* 1996;25:1509-1596.
16. Nguyen YT, Vu TD, Wong HK, Yeow YL. Solving the inverse problem of capillary viscometry by Tikhonov regularization. *J Non-Newtonian Fluid Mech.* 1999;87:103-116.
17. Yeow YL, Taylor JW. Obtaining the shear rate profile of steady laminar tube flow of Newtonian and non-Newtonian fluids from nuclear magnetic resonance imaging and laser Doppler velocimetry data. *J Rheol.* 2002;46:351-365.
18. Eilers PHC. A perfect smoother. *Anal Chem.* 2003;75:3631-3636.

Manuscript received Dec. 22, 2004, and revision received Apr. 27, 2005.