Available online at www.sciencedirect.com
SCIENCE DIRECT
Talanta 68 (2005) 156–164
Talanta
www.elsevier.com/locate/talanta

Evaluating the third and fourth derivatives of spectral data

Y. Leong Yeow [a,*], Safura Azali [a], S. Yen Ow [a], May C.L. Wong [a], Yee-Kwong Leong [b]

[a] *Department of Chemical and Biomolecular Engineering, The University of Melbourne, Vic. 3010, Australia*
[b] *School of Engineering, James Cook University, Townsville, Qld 4811, Australia*

Available online 5 July 2005

Abstract

The problem of differentiating spectral data to yield the third and fourth derivatives is converted into one of solving an integral equation of the first kind. This equation is solved by Tikhonov regularization. The method of General Cross Validation is used to guide the choice of the regularization parameter that keeps noise amplification under control. The performance of this route to third and fourth derivative spectra is demonstrated by applying it to a number of published spectra. A computational problem associated with General Cross Validation has been identified.

Keywords: Spectral conversion; Higher derivative spectra; Ill-posed problem; Integral equation; Tikhonov regularization

1. Introduction

The spectrum of a substance can take the form of a smooth featureless curve or one with a number of near overlapping peaks. In either case, it is difficult to relate the spectral data qualitatively or quantitatively to the chemical constituents of the substance and their concentrations. A variety of data processing techniques have been developed aimed at revealing the key features hidden in such spectra. Derivative spectroscopy is a popular technique used to enhance these hidden features [1]. Here the recorded spectral data is differentiated with respect to wavelength or wave number. The locations of the spectral peaks, including some of the hidden ones, will show up clearly as the zero-crossing points in the first (and third) derivative. Maxima and minima of the spectrum will appear as sharpened peaks and troughs in the even derivatives with the maxima and minima inverted in the second (and the sixth) derivative and re-inverted in the fourth (and the eighth) derivative. Features such as "shoulders" in the original spectral data will also show up as peaks in higher derivatives. The general applicability of derivative spectroscopy, particularly the higher derivatives, can be seen from the large number of applications, over five hundreds, reported by Talsky [1] who also provided a thorough discussions of the various aspects of derivative spectroscopy. More recently, Karpińska [2] reviewed the applications and developments in derivative spectroscopy. The computation and applications of higher derivative spectra, with special reference to the fourth derivative, have been reviewed by Antonov [3] and Lange and Balny [4]. Butler [5] provided an earlier review of fourth derivative spectroscopy.

A vast variety of methods have been developed to convert as measured spectral data into derivative spectra. Most methods are general in that they can be applied to different types of spectra, ranging from ultraviolet and visible spectra to FTIR spectra and emission spectra such as inductively coupled plasma atomic emission spectra. In most of the earlier investigations spectral differentiation is performed by specialized electronic hardware. Since the introduction of spectrometers interfaced with computers, these hardware have essentially been replaced by software that generates the derivatives by different numerical techniques. Irrespective of the method adopted or the nature of the spectra in question, computation of derivative spectra has a major difficulty. Differentiation is an inherently ill-posed operation in that it amplifies the unavoidable noise in the spectrum. Noise amplification becomes more serious as the order of

* Corresponding author. Fax: +61 3 8344 4153.
E-mail address: yly@unimelb.edu.au (Y.L. Yeow).

0039-9140/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2005.05.029

the derivative is increased. Naive methods of performing numerical differentiation without taking precaution to suppress noise will lead to first and second derivative spectra with greatly increased noise-to-signal ratio compared to that in the original spectrum. Further differentiation will often result in third and fourth derivative spectra that are dominated by noise.

A landmark development in derivative spectroscopy is the paper of Savitzky and Golay (SG) [6] in 1964 in which they described a computationally very efficient method that simultaneously smoothes the measured spectrum and generates the derivatives required. The SG method can now be found in most of the software that accompany the present generation of spectrometers. In this method, polynomials, typically fourth to eighth order, are fitted locally to each of the internal points of a spectral data set together with a selected number of its neighbouring points, typically 5–11 or more points, on either side. The fitted polynomials are then differentiated analytically to give the derivative at each of the internal point. In principle, this differentiation can be performed as many times as required to generate the desired order of derivative. Since locally fitted polynomials do not have continuous derivative from one point to the next, in practice noise amplification usually reaches an unacceptable level at the third and fourth derivatives. To overcome this problem, some investigators apply the SG method to smooth the lower order derivatives, usually the first or second, and then differentiate the smoothed derivatives to generate the higher derivatives. While this will suppress noise but it is also just as likely to filter out some of the essential features hidden in the original spectral data.

Recently, Yeow and Leong [7] adopted an entirely different approach to generating derivative spectra. Instead of differentiating the spectral data directly they converted the differentiation problem into one of solving an integral equation of the first kind for the second derivative spectrum and applied Tikhonov regularization [8] to solve the integral equation for the second derivative. They then integrated the second derivative to yield the first and zeroth (i.e. the original) order derivative spectra. In this method, noise amplification is kept under control by the user-specified parameter built into the Tikhonov regularization procedure. In their implementation of Tikhonov regularization, Yeow and Leong relied on Generalized Cross Validation (GCV) [9] to guide the selection of this regularization/noise suppression parameter.

The problem of computing the third and fourth derivatives of a set of spectral data can also be converted into one of solving an integral equation of the first kind for these higher derivatives. Tikhonov regularization can again be applied to solve this integral equation for the required higher derivatives. The main aim of this paper is to assess the performance of this way of obtaining fourth derivative spectra. This will be done by applying Tikhonov regularization to a number of spectral data taken from the literature. The fourth derivative is then integrated to yield the third and other lower order derivative spectra. The computed derivative spectra will be compared against that generated by the SG method and that by other methods reported in the literature.

Since the development of the integral equation for the fourth derivative spectrum and its solution by Tikhonov regularization follow closely the steps described by Yeow and Leong [7], only a brief description of these steps, highlighting the new developments, will be included in this paper.

2. Governing equations

2.1. Integral equation for fourth derivative spectra

Let $A(\lambda)$ represent a general spectrum and A_0 denote the spectral value at the arbitrary reference wavelength λ_0—usually taken to be the lowest wavelength in a set of recorded spectral data. $A(\lambda)$ can then be related to the fourth derivative spectrum and spectral properties at λ_0 by a four-term Taylor series expansion about λ_0:

$$A^{\mathrm{C}}(\lambda) = \frac{1}{6}\int_{\lambda'=\lambda_0}^{\lambda} (\lambda - \lambda')^3 h(\lambda')\mathrm{d}\lambda' + A_0 + (\lambda - \lambda_0)r_0 + \frac{(\lambda - \lambda_0)^2}{2} f_0 + \frac{(\lambda - \lambda_0)^3}{6} g_0 \quad (1)$$

The first term on the RHS is the remainder term, expressed as an integral, of the Taylor's series [10]. Eq. (1) is an exact expression for $A^{\mathrm{C}}(\lambda)$. Superscript C is used here to distinguish the computed spectrum from the experimentally measured spectral data which will carry the superscript M. The $h(\lambda) \equiv \mathrm{d}^4A(\lambda)/\mathrm{d}\lambda^4$ within the integral is the unknown fourth derivative to be computed. r_0, f_0 and g_0 in Eq. (1) are the unknown first, second and third derivative, respectively, of $A(\lambda)$ evaluated at λ_0. For the purpose of converting spectral data into its derivative, Eq. (1) is treated as an integral equation of the first kind to be solved, by Tikhonov regularization, simultaneously for the unknown function $h(\lambda)$ and the four unknown constants A_0, r_0, f_0 and g_0 [8].

2.2. Discretized equation

Following the general notation in [7], the spectral data will be represented by the vector $\mathbf{A}^{\mathrm{M}} = (A_1^{\mathrm{M}} = A_0, A_2^{\mathrm{M}}, \ldots, A_i^{\mathrm{M}}, \ldots, A_{N_D}^{\mathrm{M}})$ at the measurement wavelengths $\boldsymbol{\lambda}^{\mathrm{M}} = (\lambda_1^{\mathrm{M}} = \lambda_0, \lambda_2^{\mathrm{M}}, \ldots, \lambda_i^{\mathrm{M}}, \ldots, \lambda_{N_D}^{\mathrm{M}})$. The span of the wavelength $\lambda_{N_D}^{\mathrm{M}} - \lambda_1^{\mathrm{M}}$ and the number of spectral points N_D depend on the range and resolution of the spectrometer. Unlike the SG method, the λ_i^{M} do not have to be uniformly spaced. The span $\lambda_{N_D}^{\mathrm{M}} - \lambda_1^{\mathrm{M}}$ is discretized into N_K uniformly spaced points: $\boldsymbol{\lambda}^{\mathrm{C}} = (\lambda_1^{\mathrm{C}} = \lambda_1^{\mathrm{M}}, \lambda_2^{\mathrm{C}}, \ldots, \lambda_j^{\mathrm{C}}, \ldots, \lambda_{N_K}^{\mathrm{C}} = \lambda_{N_D}^{\mathrm{M}})$ at $\Delta = (\lambda_{N_D}^{\mathrm{M}} - \lambda_1^{\mathrm{M}})/(N_K - 1)$ distance apart. The values of the unknown fourth derivative spectra $h(\lambda)$ at these discretization points will be denoted by the vector $\mathbf{h} = (h_1, h_2, h_3, \ldots, h_{\mathrm{N_K}})$. To ensure accurate representation of

$h(\lambda)$, the number of discretization points N_K is usually set to be much larger than N_D, typically $N_K = 301$–801.

In terms of the discretized variables, Eq. (1) becomes:

$$A_i^{\mathrm{C}} = \sum_{j=1}^{N_K} B_{ij} f_j + A_0 + (\lambda_i^{\mathrm{M}} - \lambda_0) r_0 + \frac{(\lambda_i^{\mathrm{M}} - \lambda_0)^2}{2} f_0 + \frac{(\lambda_i^{\mathrm{M}} - \lambda_0)^3}{6} g_0, \quad \text{for } i = 1, 2, \ldots, N_D \tag{2a}$$

or in matrix notation:

$$\mathbf{A}^{\mathrm{C}} = \mathbf{B}\mathbf{h} + \mathbf{1}A_0 + (\boldsymbol{\lambda}^{\mathrm{M}} - \mathbf{1}\lambda_0) r_0 + \frac{(\boldsymbol{\lambda}^{\mathrm{M}} - \mathbf{1}\lambda_0)^2}{2} f_0 + \frac{(\boldsymbol{\lambda}^{\mathrm{M}} - \mathbf{1}\lambda_0)^3}{6} g_0 \tag{2b}$$

$\mathbf{B}$ is a $N_D \times N_K$ matrix of known numerical coefficients that arise from the approximation of the integral in Eq. (1) by numerical quadrature such as the trapezoidal or the Simpson's rule. $\mathbf{1}$ is a column vector of 1. The unknowns $h_1, h_2, h_3, \ldots, h_{N_K}$, A_0, r_0, f_0 and g_0 are required to minimize:

$$\text{(i) } S_1 = \sum_{i=1}^{N_D} (A_i^{\mathrm{M}} - A_i^{\mathrm{C}})^2 \tag{3a}$$

and

$$\text{(ii) } S_2 = \sum_{j=2}^{N_K - 1} \left(\frac{\mathrm{d}^2 h}{\mathrm{d}\lambda^2} \right)_j^2. \tag{3b}$$

Condition (i) ensures that the computed spectrum approximates the measured spectrum closely and condition (ii) ensures that the fourth derivative spectrum does not show spurious fluctuations.

2.3. Tikhonov regularization

In Tikhonov regularization, instead of minimizing (i) and (ii) separately, a linear combination $R = S_1 + \chi S_2$ is minimized [8]. χ is the weighting/regularization parameter that balances these two requirements. A large χ favours (ii) the smoothness condition while a small χ favours (i) the accuracy condition. The value of χ clearly depends on the noise level in the spectral data, the number of data points N_D and that of discretization points N_K, its numerical value, therefore, does not have any physical significance. Methods such as the popular Morozov Principle [8], the practical L-curve method [11] and the statistically based method of Generalized Cross Validation (GCV) [9] can be used to guide the choice of this parameter. In this investigation, the choice of χ will be based on GCV.

For any χ, the unknowns $h_1, h_2, h_3, \ldots, h_{N_K}$, A_0, r_0, f_0 and g_0 that minimize R is given by [8]:

$$\mathbf{h}' = \left(\mathbf{B}'^{\mathrm{T}} \mathbf{B}' + \frac{\chi}{\Delta^4} \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\beta} \right)^{-1} \mathbf{B}'^{\mathrm{T}} \mathbf{A}^{\mathrm{M}}. \tag{4}$$

where $\mathbf{h}' = (h_1, h_2, h_3, \ldots, h_{N_K}, A_0, r_0, f_0, g_0)$ is used to denote all the unknowns and $\mathbf{B}'$ is the matrix $\mathbf{B}$ with the column vectors $\mathbf{1}$, $\boldsymbol{\lambda}^{\mathrm{M}} - \mathbf{1}\lambda_0$, $(\boldsymbol{\lambda}^{\mathrm{M}} - \mathbf{1}\lambda_0)^2/2$ and $(\boldsymbol{\lambda}^{\mathrm{M}} - \mathbf{1}\lambda_0)^3/6$ added to it to reflect the incorporation of A_0, r_0, f_0 and g_0 into $\mathbf{h}'$. $\boldsymbol{\beta}$ is the modified tri-diagonal matrix arising from standard finite difference approximation of $\mathrm{d}^2 h(\lambda)/\mathrm{d}\lambda^2$ at the uniformly distributed discretization points,

$$\boldsymbol{\beta} = \begin{bmatrix} 1 & -2 & 1 & & & 0 & 0 & 0 & 0 \\ & 1 & -2 & 1 & & 0 & 0 & 0 & 0 \\ & & & \ddots & & . & . & . & . \\ & & & 1 & -2 \quad 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{5}$$

The four extra columns of 0 in Eq. (5) are again the consequence of the incorporation of A_0, r_0, f_0 and g_0, which do not feature in the smoothness condition, into $\mathbf{h}'$. In terms of $\mathbf{h}'$and $\mathbf{B}'$:

$$\mathbf{A}^C = \mathbf{B}'\,\mathbf{h}'. \tag{6}$$

Eq. (6) is needed in the computation of the GCV function below.

Eq. (4) is the linear algebraic equations that convert spectral data $\mathbf{A}^{\mathrm{M}}$ into the fourth derivative spectrum $h(\lambda)$ described by $\mathbf{h}$. As $h(\lambda)$ is known at a large number of closely and uniformly spaced wavelengths it can be integrated successively with relative ease to give the third, second and first derivative spectra and a back-calculated version $A^{\mathrm{C}}(\lambda)$ of the original spectrum. In the integration operations the g_0, f_0, r_0 and A_0 given by Eq. (4) are used as the boundary conditions. All the integration will be performed using commercial software independent of the computer code developed to solve for $h(\lambda)$. Thus the comparison of the back-calculated $A^{\mathrm{C}}(\lambda)$ with $A^{\mathrm{M}}(\lambda)$ serves as an independent check of the reliability of the $h(\lambda)$ given by Eq. (4). Such comparison will be performed in all the examples described below.

2.4. The leaving-out-one principle and Generalized Cross Validation

GCV that is used to guide the choice of χ is based on the "leaving-out-one" principle [9]. In principle, the computation described by Eq. (4) can be repeated N_D times each time leaving out one data point. The optimum χ is taken to be the value that minimizes the sum of squares $V(\chi)$ of the difference between the predicted value and the actual value for each of the left out data point. It can be shown that, in the GCV implementation of the "leaving-out-one" principle, $V(\chi)$ is given by [9]:

$$V(\chi) = \frac{(\mathbf{A}^{\mathrm{M}} - \mathbf{A}^{\mathrm{C}})^{\mathrm{T}} (\mathbf{A}^{\mathrm{M}} - \mathbf{A}^{\mathrm{C}})/N_D}{(1 - \mathrm{Tr}[\mathbf{E}]/N_D)^2}. \tag{7}$$

Tr[**E**] denotes the trace of the square matrix **E**, known as the influence matrix, defined by [9]:

$$\mathbf{E} = \mathbf{B}' \left(\mathbf{B}'^{\mathrm{T}} \mathbf{B}' + \frac{\chi}{\Delta^4} \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\beta} \right)^{-1} (\mathbf{B}')^{\mathrm{T}}. \quad (8)$$

Eqs. (4) and (7) together with the definition of $\mathbf{B}'$, $\boldsymbol{\beta}$ and $\mathbf{E}$ allow $V(\chi)$ to be evaluated and plotted against χ/Δ^4. Minimization of $V(\chi)$ is used to locate the optimum χ.

3. Data and results

Eq. (4) will now be used to compute the fourth derivative of a number of spectral data sets taken from published literature. The experimental conditions of these data can be found in the original papers and will be omitted here.

3.1. A synthesized spectrum of two Gaussian bands

The discrete points in Fig. 1(a) represent a simulated UV absorption spectrum over the wavelength 265 nm $\leq \lambda \leq$ 295 nm with $\Delta\lambda = 0.5$ nm. This is generated by the sum of the following two Gaussian bands:

$$A(\lambda) = 0.293 \exp\left[-\left(\frac{\lambda - 277}{8.0065} \right)^2 \ln 2 \right] \quad \text{and}$$

$$A(\lambda) = 0.03 \exp\left[-\left(\frac{\lambda - 289}{3.8851} \right)^2 \ln 2 \right] \quad (9)$$

where λ is in nm. These two Gaussian bands are shown as lighter curves in Fig. 1(a). This is the spectrum used by Lange et al. [12] to demonstrate their method for obtaining fourth derivative spectrum. The disparity in amplitude and close proximity of the two Gaussian peaks mean that the weaker band located at 289 nm is completely masked by the stronger band at 277 nm. One of the aims of derivative spectroscopy is to expose such a hidden peak.

The fourth derivative of the spectral data in Fig. 1(a) given by Eq. (4) is shown as a dark curve in Fig. 1(b). For comparison, the exact fourth derivative obtained by analytical differentiation of Eq. (9) is shown as a lighter curve on the same plot. The two derivative spectra are in very close agreement so much so that the analytical curve has to be thickened

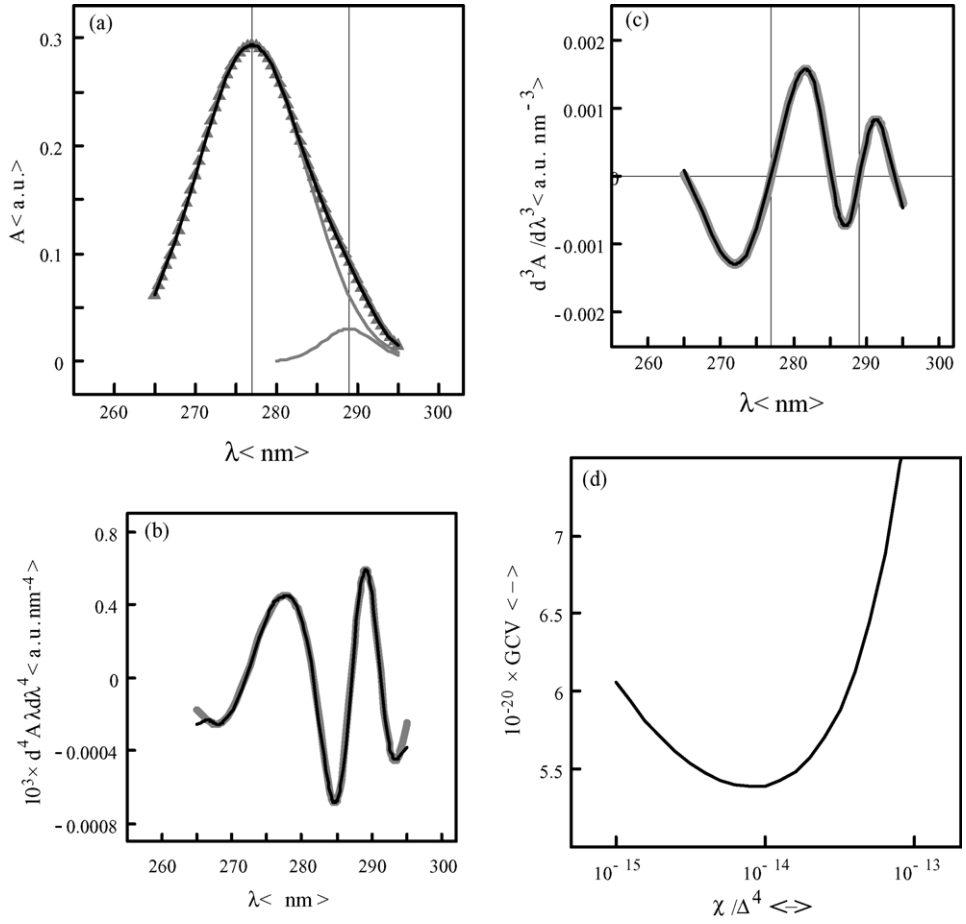

Fig. 1. Synthetic UV spectra. (a) Absorption spectra. ▲: Exact spectral data based on Eq. (9), lighter curves: constituent Gaussian bands, continuous curves: back-calculated from Eq. (4). (b) Fourth derivative spectra. Continuous curve: from Eq. (4), thick lighter curve: exact fourth derivative. (c) Third derivative spectra. Continuous curve: back-calculated from Eq. (4), thick lighter curve: exact third derivative. (d) The GCV function used to guide the selection of χ showing the optimum value at around $\chi/\Delta^4 \approx 7.5 \times 10^{-15}$.

for it to show up. It can also be seen that the fourth derivative spectrum has succeeded in revealing the location of the two Gaussian peaks. The spectrum back-calculated from the fourth derivative given by Eq. (4) is shown as a continuous curve in Fig. 1(a) and is in very good agreement with the original spectral data. The exact location of the two Gaussian peaks can be determined by the zero-crossing points of the back-calculated third derivative spectrum shown in Fig. 1(c) as a dark curve. For comparison, the analytical third derivative is shown as a lighter curve in the same plot.

The GCV function used to determine the optimum χ for the synthetic spectrum is shown in Fig. 1(d). $\chi_{opt}/\Delta^4 \approx 7.5 \times 10^{-15}$. Spectral data are converted into a dimensionless form prior to applying Eq. (4), hence the χ/Δ^4 shown here is in dimensionless form. As mentioned above, no physical significance should be read into this particular value. It is noticed that, in most of the examples considered in this paper, the optimum χ/Δ^4 varies from 3.5×10^{-12} to 3.5×10^{-15}. In order to cope with χ/Δ^4 of this order of magnitude it was necessary to perform the computation in extended precision, typically most numbers are kept to 16 decimal points. This greatly increased the computation cost of obtaining the fourth derivative compared to that for obtaining the second derivative as reported by Yeow and Leong [7].

3.2. UV absorption spectrum of papaverine hydrochloride

The discrete points in Fig. 2(a) are the tabulated absorption spectral points for a 10^{-5} mol l^{-1} papaverine hydrochloride solution [13]. These data span the wavelength range 206.5 nm $\leq \lambda \leq$ 264 nm at the interval of $\Delta\lambda = 1$ nm except for the first data point where $\Delta\lambda = 0.5$ nm. Yeow and Leong [7] used this spectrum to demonstrate their Tikhonov-based procedure for obtaining second derivative spectrum.

The fourth derivative of these spectral data given directly by Eq. (4) is shown as a continuous curve in Fig. 2(b). For comparison the fourth derivative spectrum given by the SG method is shown as filled squares on the same plot. In this implementation of the SG method sixth order polynomials with 15 regression points (i.e. 7 points on either side) at $\Delta\lambda = 1$ nm apart were used. The large number of regression points means the span of wavelength covered by the SG-based spectrum is reduced by 7 points on each end. And

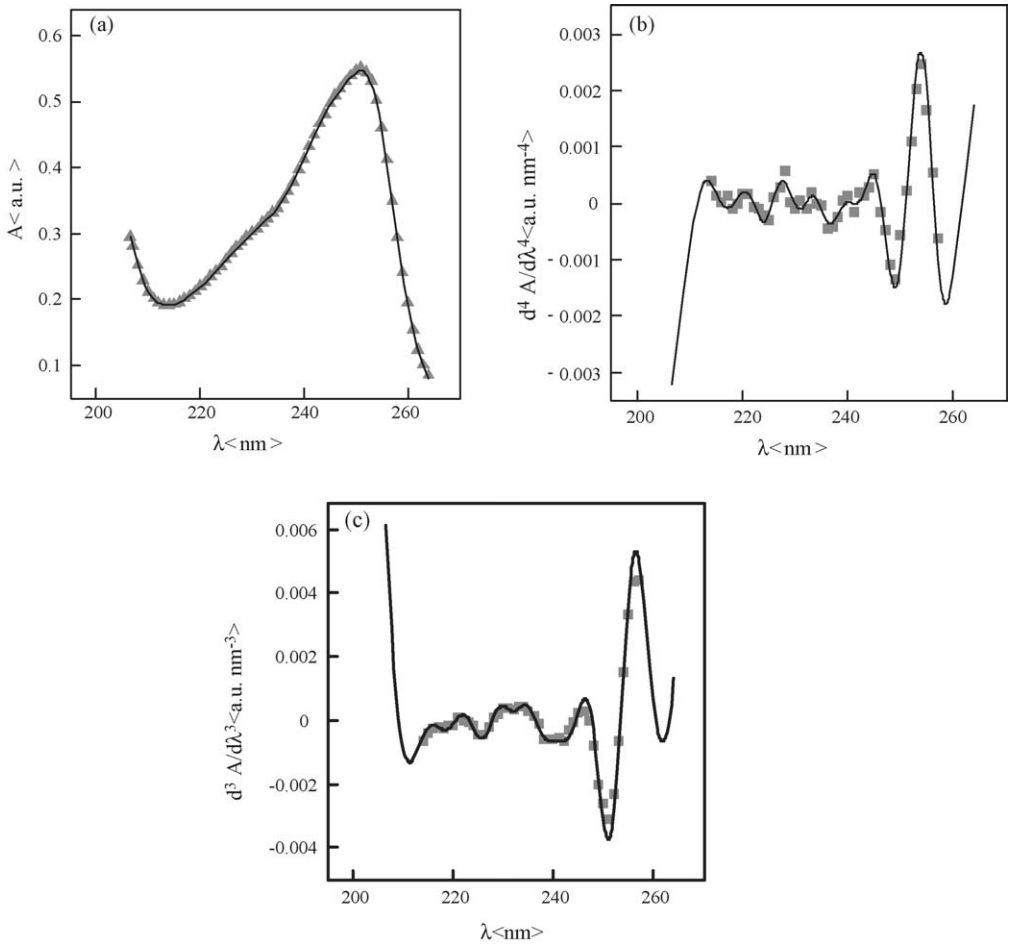

Fig. 2. UV spectra of papaverine hydrochloride. (a) Absorption spectra. ▲: Spectral data from Lang [13], continuous curve: back-calculated from Eq. (4). (b) Fourth derivative spectra. ■: Sixth order SG method, continuous curve: from Eq. (4). (c) Third derivative spectra. ■: Sixth order SG method, continuous curve: from Eq. (4).

as expected, the SG spectrum is significantly more noisy than that given by Eq. (4). They are, however, in satisfactory agreement in terms of the location and amplitude of all the main turning points. The order of polynomials and the number of regression points used in the SG method were determined by a trial-and-error process. For example if 13 instead of 15 regression points were used with the sixth order polynomials it was found that the resulting fourth derivative is considerably more noisy compared to that given by Eq. (4). Conversely if 17 regression points were used the derivative curve is over smoothed and this is reflected by a reduction in the height of the amplitude of the SG peak in the neighbourhood of 255 nm. Similar numerical experimentation were also performed using lower and higher order polynomials.

The back-integrated third derivative spectrum for the papaverine hydrochloride solution is shown in Fig. 2(c) as a continuous curve. The corresponding derivative given by sixth order SG with 15 regression points is shown as discrete points. The two third order derivative spectra are in closer agreement than the fourth derivatives in Fig. 2(b). The agreement is furthered improved when the corresponding second and first derivative spectra and the back-calculated original spectra are compared. For example, the back-calculated spectrum based on Tikhonov regularization is shown as a continuous curve in Fig. 2(a). The corresponding curve based on the SG method is essentially indistinguishable from this curve and from the original spectral data. For clarity, the SG spectrum is not shown.

3.3. UV absorption cross-section of bromofluorobutene

The discrete points in Fig. 3(a) are part of the UV absorption cross-section data of 2-bromo-3-3-4-4-4-pentafluorobutene-1 reported by Orkin et al. [14]. These data points are at a uniform interval of $\Delta\lambda = 0.5$ nm apart. They are again the data used by Yeow and Leong [7] to demonstrate their method for computing second derivative.

The fourth derivative spectra generated by Eq. (4) and by the SG method are shown in Fig. 3(b). In this case, the SG method is based on eighth order polynomials with 19 regression points at $\Delta\lambda = 0.5$ nm apart. Following the procedure

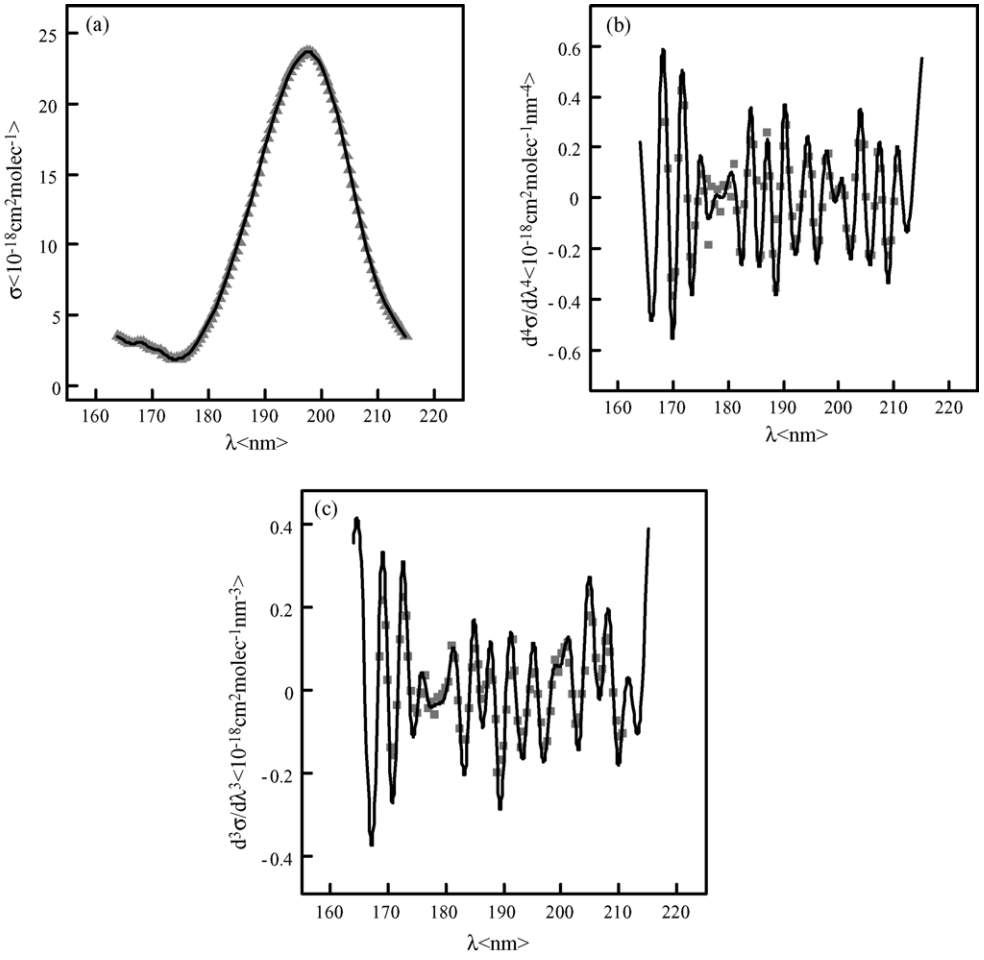

Fig. 3. UV absorption of 2-bromo-3-3-4-4-4-pentafluoro-butene-1. (a) Absorption cross-section. ▲: Measured spectral data from Orkin et al. [14], continuous curve: from Eq. (4). (b) Fourth derivative spectra. ■: Eighth order SG method, continuous curve: from Eq. (4). (c) Third order derivative spectra. ■: Eighth order SG method, continuous curve: from Eq. (4).

in the previous example, the corresponding third derivative spectra by these two different methods are shown in Fig. 3(c). The back-calculated spectra based on Tikhonov regularization is shown as a continuous curve in Fig. 3(a). All these plots indicate that the results generated by Eq. (4) and by the SG method are in acceptable agreement. Without knowledge of the expected spectral behaviour of the fluorobromoalkene in question, it is not possible to interpret the peaks in the computed derivative spectra.

3.4. Absorption spectra of molecular chlorine

Seery and Britton [15] and Hubinger and Nee [16] reported the absorption spectrum of molecular chlorine at 298 °C. Their data, plotted as absorption cross-sections, are shown as discrete points in Fig. 4(a) [17]. These two set of spectral data, covering slightly different spans of wavelength, are in good agreement. $\Delta\lambda = 10$ nm for both data sets. Eq. (4) is used to convert these two data sets into fourth derivative separately. The aim here is to test whether Eq. (4) is capable of converting two nearby spectra into nearby fourth derivatives. This constitutes a critical test of the Tikhonov regularization procedure.

The two fourth derivative spectra based on Eq. (4) are shown in Fig. 4(b). The relatively simple shape of the Cl_2 spectrum allows Hubinger and Nee [16] to use the following expression:

$$\sigma(\lambda) = 2.55 \times 10^{-19} \exp\left\{-86.6\left[\ln\left(\frac{339.5}{\lambda}\right)\right]^2\right\} + 8.72 \times 10^{-21} \exp\left\{-80.0\left[\ln\left(\frac{406.5}{\lambda}\right)\right]^2\right\} \quad (10)$$

to describe their measured absorption cross-section. In this expression, σ is in cm^2 $molecule^{-1}$ and λ is in nm. The analytical fourth derivative spectrum obtained by differentiating this expression is shown as a lighter curve in Fig. 4(b). Apart from the two ends, all the fourth derivative spectra in this figure are in acceptable agreement. In particular the two fourth derivative spectra given by Eq. (4) are in closer agreement than with the analytical result. The small number of spectral

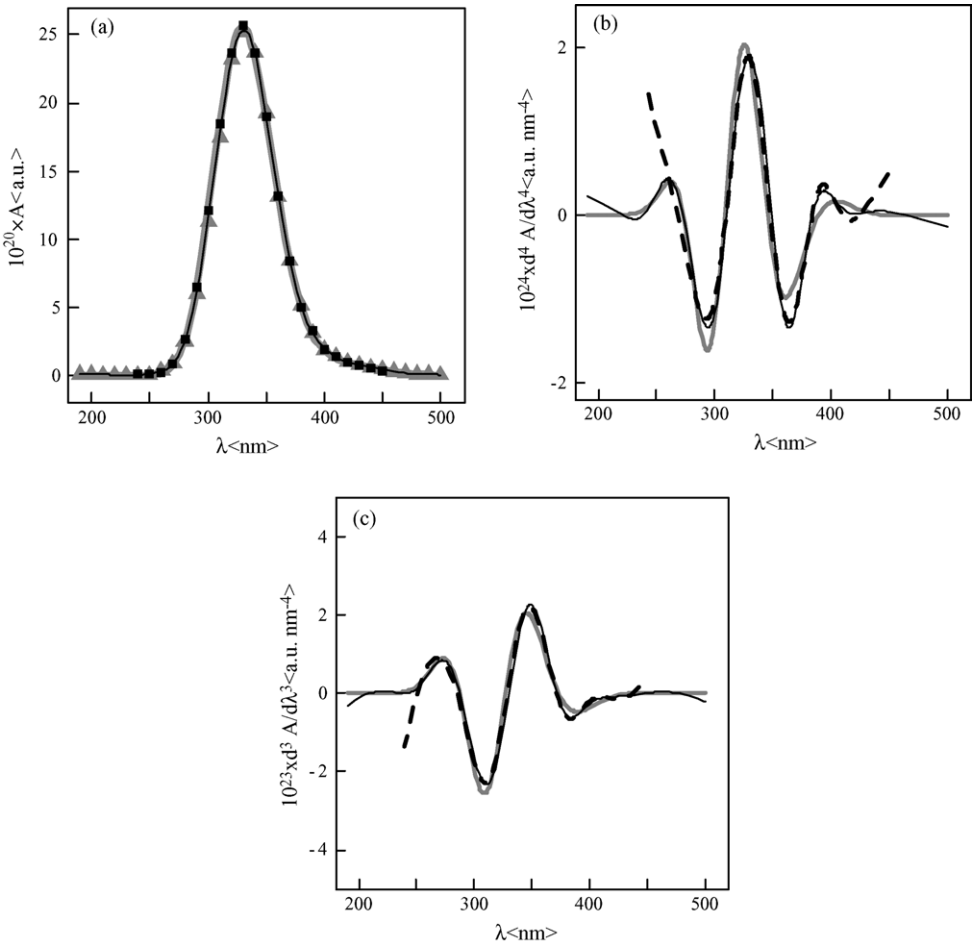

Fig. 4. UV spectrum of Cl_2. (a) Absorption spectra. ■: Seery and Britton [15], ▲: Hubinger and Nee [16], light thick curve: Eq. (10), dark continuous curve: back-calculated from Eq. (4) based on data from [16]. (b) Fourth derivative spectra. Light thick curve: analytical derivative of Eq. (10), continuous dark curve: from Eq. (4) based on data from [16], dark broken curve: Eq. (4) based on data from [15]. (c) Light thick curve: analytical derivative of Eq. (10), continuous dark curve: from Eq. (4) based on data from [16], dark broken curve: Eq. (4) based on data from [15].

points and the relatively large $\Delta\lambda$ of the data means that it is not practical to use the SG method to evaluate the fourth derivative spectrum. This paucity of data points does not pose a problem for Eq. (4). As expected, there is even closer agreement between the two third derivative spectra given by Eq. (4) and that given by analytical differentiation of Eq. (10) (see Fig. 4(c)). The two back-calculated spectra from Eq. (4) and that based on Eq. (10) are essentially indistinguishable from one another and only the back-calculated curve based on the data of Hubinger and Nee [16] is plotted in Fig. 4(a).

4. Discussion

All the examples considered above indicate that the general principle of converting the problem of differentiating spectral data into one of solving an integral equation of the first kind can be extended to fourth order derivative spectra. The method based on Tikhonov regularization coupled with GCV can again be applied to solve this integral equation. Most of the advantages of this approach observed in second order derivative spectroscopy [7] are retained in the fourth derivative. These include the general applicability of the method and its ability to keep noise amplification under control. Unlike the SG method, the method is able to cope with non-uniformly spaced spectral data and again unlike the SG method there is no reduction in the span of wavelength covered by the derivative spectra.

In their earlier investigation, Yeow and Leong [7] applied Tikhonov regularization to solve an integral equation of the first kind that gives the second derivative spectra directly. The second derivative for the papaverine hydrochloride solution they obtained by this method is shown as a lighter curve in Fig. 5. For comparison the second derivative spectrum back-calculated from the fourth derivative in Fig. 2(b) is shown as a dark curve on the same plot. The two second derivative spectra are in good agreement. Similar comparison of second derivative spectra has also been performed with the bromofluoroalkene data and good agreement is again observed. Since the two second derivative spectra are obtained by solving different integral equations each with its own regularization parameter, the agreement achieved can be taken as an indication that the general computational procedure, particularly the choice of χ based on GCV, is working reliably.

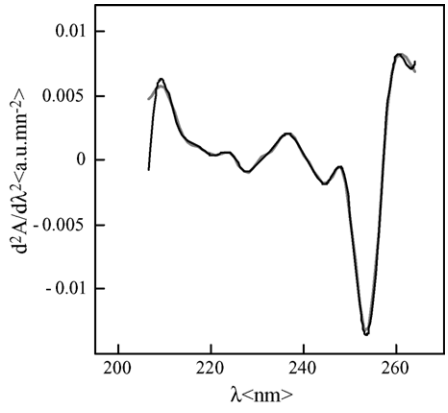

Fig. 5. Second derivative of papaverine hydrochloride. Continuous dark curve: back-calculated from the fourth derivative in Fig. 2b light thick curve: second derivative from [7].

In order to ensure that the SG method will lead to a smooth fourth derivative spectrum, it was necessary to use sixth or higher order polynomials with a large number of regression points. This has the undesirable effect of greatly reducing the span of the wavelength covered by the resulting derivative spectra. As already mentioned the alternative approach is to apply the SG method with lower order polynomials twice, once to obtain the second derivative and then to apply the SG method a second time to smooth the raw second derivative spectrum and to give the third and fourth derivatives with reduced noise. Repeated application of the SG method to smooth the derivative obtained in a previous step is a widely adopted practice even at second derivative level [18,19]. A similar approach can be adopted in Tikhonov regularization as an alternative route to fourth derivative spectra. Here Tikhonov regularization, such as that reported in [7], is first used to convert a set of spectral data into its second derivative. The second derivatives at the original measurement wavelengths are then evaluated and Tikhonov regularization is applied again to compute the second derivative of these second derivative data to give the desired fourth derivative. The performance of such an approach has been investigated. The discrete points in Fig. 6(a) are the second derivative, reported in [7], of papaverine hydrochloride at the original measurement points [13]. Applying their method to these second derivative spectral points gives the fourth derivative spectrum shown as a light thick curve in Fig. 6(b). For comparison, the fourth derivative spectrum given directly by Eq. (4) is shown as a darker curve on the same plot. This is the same spectrum as that in Fig. 2(b). There is very good agreement between the two fourth derivative spectra. This indirect route to fourth derivative has slightly reduced, but not eliminated, the computational problem associated with the evaluation of the GCV function. Comparison of the fourth derivatives in Fig. 6(b), together with the comparison in Fig. 5, may have added further to the confidence in the results given by Eq. (4), but numerical experimentation on this indirect route to fourth derivative has not exhibited significant advantage over the direct route based on Eq. (4).

As mentioned above, all the GCV functions for fourth derivative involve numbers that are of the order of 10^{-12} or smaller. In order to prevent loss of accuracy, to cope with number of this magnitude, it is necessary to carry out the computation to 14–16 decimal points. This slowed down the computation significantly. This can be partially circumvented by rescaling the numbers in the intermediate steps of Tikhonov regularization with the aim avoiding numbers that are vastly different in magnitude. This has been partially successful in some of the cases involved. Systematic investigation of the numerical and computation problems involved

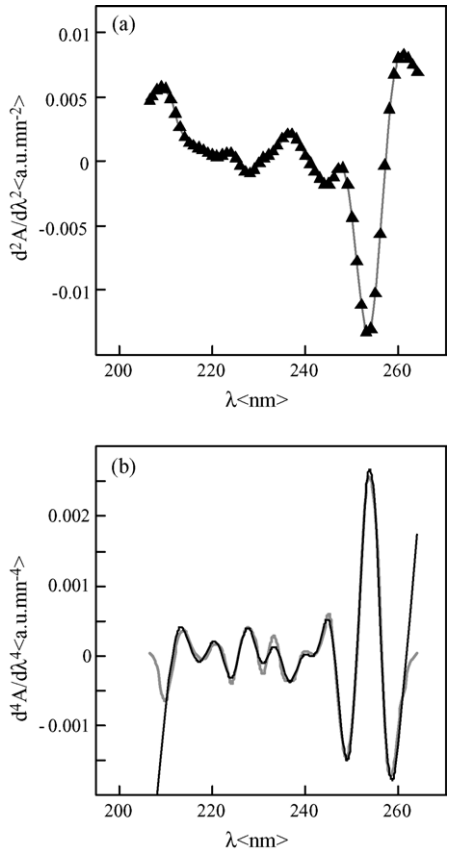

Fig. 6. Fourth derivative of papaverine hydrochloride. (a) ▲: Computed second derivative from [7], light thick curve: back-calculated from the light thick fourth derivative curve in (b). (b) Light thick curve: fourth derivative computed from discrete points in (a), dark continuous curve: fourth derivative from Fig. 2(b).

have not been attempted. These numerical and computational problem will need to be solved if the procedure is to be extended to yet higher derivatives, such as the sixth and the eighth derivatives.

The time consuming computation of the GCV function can be avoided if one is prepared to apply the Morozov principle [8] to guide the choice of the regularization parameter. Here, the parameter is adjusted so that the average difference between A^{C} and A^{M} is of the same order as the estimated average error bar of the spectral data. This normally only requires solving Eq. (4) at a small number of selected χ. Apart for the average difference between A^{C} and A^{M}, the choice of χ is also guided by one's knowledge of the expected derivative spectrum, particularly by the location of the expected peaks and the ability to interpret any unexpected peaks in terms of the spectral behaviour of the constituents in the substance under investigation. It should also be mentioned that since the derivative spectra are not very sensitive to small changes in χ, and therefore, fine tuning of χ is normally not required.

5. Conclusions

Taylor's series with the remainder term in an integral form provides an integral equation of the first kind for the fourth derivative of a spectral data set. Tikhonov regularization can be applied to solve this equation. The method can be applied to different types of spectral data. GCV provides a means of locating the appropriate χ. The numerical and computation problems associated with the GCV function for the fourth derivative will need to be solved if the method is to be extended to yet higher derivatives. Physical knowledge of the spectral behaviour of the system under investigation can be used to guide the choice of this parameter.

References

[1] G. Talsky, Derivative Spectrophotometry: Low and High Orders, VCH, Weinheim, 1994.
[2] J. Karpińska, Talanta 64 (2004) 801–822.
[3] L. Antonov, Anal. Chim. Acta 349 (1997) 295–301.
[4] R. Lange, C. Balny, Biochim. Biophys. Acta 1595 (2002) 80–93.
[5] W.L. Butler, Methods Enzymol. 56 (1979) 501–515.
[6] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627–1639.
[7] Y.L. Yeow, Y.-K. Leong, Appl. Spectro. 59 (2005) 584–592.
[8] H.W. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 2000.
[9] G. Wahba, Spline Models for Observational Data, SIAM, Philadelphia, 1990.
[10] J.C. Burkill, A First Course in Mathematical Analysis, Cambridge University Press, Cambridge, 1978.
[11] P.C. Hansen, SIAM Rev. 34 (1992) 561–580.
[12] R. Lange, J. Frank, C.-L. Saldana, C. Balny, Eur. Biophys. J. 24 (1996) 277–283.
[13] L. Lang, Absorption Spectra in the Ultraviolet and Visible Region, vol. 1, Akadémiai Kiadó, Budapest, 1959.
[14] V.L. Orkin, F. Louis, R.E. Huie, M.J. Kurylo, J. Phys. Chem. A 106 (2002) 10195–10199.
[15] D.J. Seery, D. Britton, J. Phys. Chem. 68 (1964) 2263–2266.
[16] S. Hubinger, J.B. Nee, J. Photochem. Photobiol. A: Chem. 86 (1995) 1–7.
[17] A. Noelle, M. Allen, N.P. Ernsting, A.K. Richter, UV/Vis Spectra Data Base, third ed., Science-SoftCon CD-ROM, 2004.
[18] J. Yang, Z. Piao, X. Zeng, Spectrochim. Acta 46B (1991) 953–965.
[19] M.P. Horvath, R.A. Copeland, M.W. Makinen, Biophys. J. 77 (1999) 1694–1711.