

Derivatives in Spectroscopy

Part I – The Behavior of the Derivative

Howard Mark and Jerome Workman Jr.

Derivatives of spectra ($dT/d\lambda$ or $dA/d\lambda$, and their wavenumber equivalents in Fourier transform–infrared spectroscopy) have been known and used in spectroscopy for a long time. First derivatives and second derivatives ($d^2T/d\lambda^2$ or $d^2A/d\lambda^2$) are both commonly used in modern spectroscopy, particularly in near-infrared (NIR) spectroscopy. They also enjoy widespread use in some non-optical spectroscopic techniques, such as nuclear magnetic resonance and electron spin resonance spectroscopies. The mathematics and behavior of the derivative is independent of the particular spectroscopic technique to which it is applied, however. But because our own backgrounds are in optical spectroscopy, we will discuss it, where pertinent, in terms of the spectroscopy we know best.

Studies of the application of derivatives to spectroscopy go back at least as far as 1953 (1–3). A more recent paper

contains a good bibliography of the work before its appearance (4). Since the advent of NIR spectroscopy as a popular analytical technique, the routine use of derivative spectra has burgeoned, along with the application to this method of spectroscopic analysis. Along with the increased applicability, interest has grown in the background and behavior of derivatives. Dave Hopkins especially has led the way in understanding the behavior of first and second derivatives, particularly their computation using Savitzky–Golay convolution functions (5, 6). We do not plan to deal with that aspect too extensively at this time, however.

The application of derivatives is not without problems, however, especially when the concern is to accurately represent the derivative of a given data spectrum. Therefore, understanding the nature of the problems encountered so that the proper decisions can be made regarding how the derivative should be calculated is crucial to obtaining optimum results.

Figure 1 illustrates some of the problems of derivatives. This figure depicts some of the basic behaviors underlying the use of the derivatives for spectroscopic analysis. The top curve in Figure 1 represents a synthetic spectrum with two Gaussian (normal) bands, one of 20-nm bandwidth and one of 60-nm bandwidth. Spectroscopic band shapes are conventionally considered to be either Gaussian or Lorentzian; in this column we will concentrate on Gaussian band shapes; therefore all our figures are based on Gaussian-shaped bands. We will, however, treat Lorentzian bands at appropriate points. In Figure 1 we present normal bands with spacing between wavelength points of 1 nm, a number that will become important later on. The middle curve represents the first “derivative” and the bottom curve the second “derivative” of the absorbance band. We are putting the term “derivative” in quotes, because they are, in fact, not true derivatives. The definition of a derivative includes the step of taking a limit as differences approach zero. In the real world, with real data, we can never calculate a true derivative because we must compute the differences between finite data points, and these must be taken over finite intervals, so that computed derivatives are approximations of the actual derivative.



Howard Mark

serves on the Editorial Advisory Board of *Spectroscopy* and runs a consulting service, Mark Electronics, 21 Terrace Avenue, Suffern, NY 10901, that provides assistance, training, and consultation in NIR spectroscopy as well as custom hardware and software design and development. He can be reached via e-mail at hlmark@prodigy.net.

Jerome Workman Jr. serves on the Editorial Advisory Board of *Spectroscopy* and is vice-president of research for Argose (Waltham, MA). He may be reached by e-mail at jworkman@argose.com.



The absorbance spectrum in Figure 1 is made from synthetic data, but mimics the behavior of real data in that both are represented by data points collected at discrete and (usually) uniform intervals. Therefore the calculation of a “derivative” from actual data is really the computation of finite differences, usually between adjacent data points. We will now remove the quotation marks from around the term, and simply call all the finite-difference approximations a derivative. As we shall see, however, often data points that are more widely spread are used. If the data points are sufficiently close together, then the approximation to the true derivative can be quite good. Nevertheless, a true derivative can never be measured when real data is involved.

Figure 1, however, still shows a number of characteristics that reveal the behavior of derivatives. First of all, we note that the first derivative crosses the x -axis at the wavelength where the absorbance peak has a maximum, and has maximum values (both positive and negative) at the point of maximum slope of the absorbance bands. These characteristics, of course, reflect the definition of the derivative as a measure of the slope of the underlying curve. For Gaussian bands, the maxima of the first derivatives also correspond to the standard deviation of the underlying spectral curve.

The second derivative, in contrast, has its maximum value at the same wavelength as the underlying peak, although in the negative direction. The second derivative crosses the x -axis at the point of maximum slope of the underlying (first derivative) curve, and because of that, presents a much sharper band than the underlying absorbance band does. The problem arises, however, that this sharpening effect is accompanied by the creation of two artifact peaks, the two positive-going peaks that flank the negative-going portion of the second derivative. In complicated spectra, therefore, it can sometimes be difficult to distinguish true spectral features from the artifacts created by the second derivative calculation.

Finally, we note that the magnitudes of both the first and second derivatives of the narrow absorbance band are considerably greater than corresponding magnitudes for the wider absorbance band. This characteristic is a reflection of the fact that the slope of the narrower band truly is greater than that of the broader band of the same height, as can be seen in the expanded views of the two absorbance bands in Figure 1. For the same ΔX , the narrow absorbance band has a much larger value of ΔY than the broad absorbance band does, therefore $\Delta Y/\Delta X$ (the derivative) is larger for that band. A similar situation is true for the second derivative as well. There is a further consideration as well; the mathematical definition of a normal curve includes a premultiplying factor of $1/(\sigma(2\pi)^{1/2})$, which makes the area under the normal curve equal to unity. Therefore, the wider the bandwidth, the smaller the maximum value of the curve will be, further reducing the slope as compared to a narrower band.

It is interesting and useful to consider this quantitatively. The expression for the normal distribution is (7)

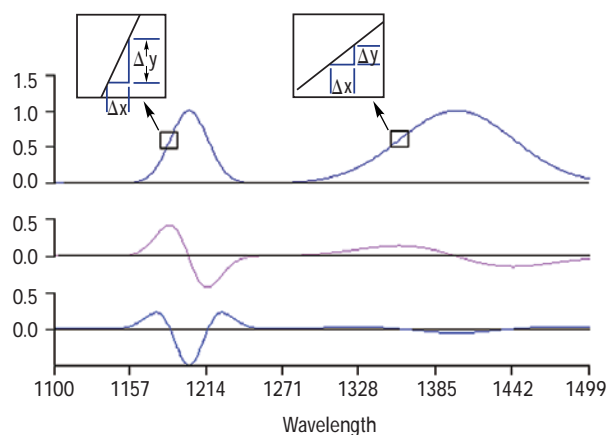


Figure 1. Two Gaussian absorbance bands and their respective first and second “derivatives” (finite differences). The top spectrum represents a synthetic Gaussian absorbance spectrum, the middle a first “derivative” and the bottom a second “derivative.” Note that the ordinate of the first “derivative” has been expanded by a factor of 10 and the second “derivative” by another factor of 10. The wavelength spacing between data points is 1 nm. The narrow band has a bandwidth (full width at half height) of 20 nm, the broad one is 60 nm.

$$Y = \frac{1}{\sigma(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad [1a]$$

The corresponding expression for the Lorentzian distribution is (8, see p. 211)

$$Y = \frac{2}{\pi\sigma} \times \frac{1}{1 + \left[\frac{2(\mu-X)}{\sigma}\right]^2} \quad [1b]$$

where σ is the measure of bandwidth (and equals the standard deviation for the normal curve) and μ is the wavelength corresponding to the peak center.

We note parenthetically here that equation 1a includes the premultiplying factor for constant area. The expression for a normal curve of constant maximum height (of unity) will be simply

$$Y = e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad [2]$$

The first derivative of the normal distribution, from the expression in equation 1a, then is

$$\frac{dY}{dX} = \frac{1}{\sigma(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \frac{d}{dX} \left(-\frac{1}{2} \left(\frac{X-\mu}{\sigma} \right)^2 \right) \quad [3]$$

$$\frac{dY}{dX} = \frac{1}{\sigma(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \left(-\frac{1}{2\sigma} \frac{d}{dX} (X-\mu)^2 \right) \quad [4]$$

$$\frac{dY}{dX} = \frac{1}{\sigma(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \left(-\frac{1}{2\sigma} 2(X-\mu) \right) \quad [5]$$

$$\frac{dY}{dX} = \frac{-(X-\mu)}{\sigma^2(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad [6a]$$

Equation 6a is derived from the constant-area expression for the normal curve; from the constant-height expression we obtain

$$\frac{dY}{dX} = \frac{-(X-\mu)}{\sigma} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad [6b]$$

The origin of the features seen qualitatively in Figure 1 can be observed in either of equations 6a or 6b. When $X = \mu$, then the derivative is zero, and the sign of the derivative changes from positive when $X < \mu$ to negative when $X > \mu$. The presence of the negative exponential term ensures that the derivative will asymptotically approach zero as X approaches infinity in both directions.

Similarly, from equation 6a we can derive the expression for the second derivative of the normal distribution

$$\begin{aligned} \frac{d^2Y}{dX^2} &= \frac{-(X-\mu)}{\sigma^2(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \frac{d}{dX} \left(-\frac{1}{2} \left(\frac{X-\mu}{\sigma} \right)^2 \right) + \\ &e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \frac{d}{dX} \left(\frac{-(X-\mu)}{\sigma^2(2\pi)^{1/2}} \right) \end{aligned} \quad [7]$$

$$\begin{aligned} \frac{d^2Y}{dX^2} &= \frac{-(X-\mu)}{\sigma^2(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \left(-\frac{1}{2\sigma} 2(X-\mu) \right) + \\ &\left(\frac{-1}{\sigma^2(2\pi)^{1/2}} \right) e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \end{aligned} \quad [8]$$

$$\frac{d^2Y}{dX^2} = \left(\frac{(X-\mu)^2}{\sigma^3(2\pi)^{1/2}} - \frac{1}{\sigma^2(2\pi)^{1/2}} \right) e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad [9a]$$

And from equation 6b we obtain

$$\frac{d^2Y}{dX^2} = \left(\frac{(X-\mu)^2}{\sigma^2} - \frac{1}{\sigma} \right) e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad [9b]$$

For the Lorentzian distribution, from equation 1b the first derivative is

$$\frac{dY}{dX} = \frac{2}{\pi\sigma} \times \frac{-1}{\left(1 + \left[\frac{2(\mu-X)}{\sigma}\right]^2\right)^2} \frac{d}{dX} \left(1 + \left[\frac{2(\mu-X)}{\sigma}\right]^2 \right) \quad [10]$$

$$\frac{dY}{dX} = \frac{2}{\pi\sigma} \times \frac{-8(\mu-X)}{(\sigma^2 + 4(\mu-X)^2)^2} \quad [11]$$

And then the second derivative of the Lorentzian distribution is

$$\begin{aligned} \frac{d^2Y}{dX^2} &= \frac{2}{\pi\sigma} \times \left[\frac{(\sigma^2 + 4(\mu-X)^2)^2 \frac{d}{dX} (-8(\mu-X))}{(\sigma^2 + 4(\mu-X)^2)^4} - \right. \\ &\left. \frac{(-8(\mu-X)) \frac{d}{dX} ((\sigma^2 + 4(\mu-X)^2)^2)}{(\sigma^2 + 4(\mu-X)^2)^4} \right] \end{aligned} \quad [12]$$

$$\begin{aligned} \frac{d^2Y}{dX^2} &= \frac{2}{\pi\sigma} \times \left[\frac{-8(\sigma^2 + 4(\mu-X)^2)^2}{(\sigma^2 + 4(\mu-X)^2)^4} + \right. \\ &\left. \frac{8(\mu-X) \times 2(\sigma^2 + 4(\mu-X)^2) \frac{d}{dX} (\sigma^2 + 4(\mu-X)^2)}{(\sigma^2 + 4(\mu-X)^2)^4} \right] \end{aligned} \quad [13]$$

$$\begin{aligned} \frac{d^2Y}{dX^2} &= \frac{2}{\pi\sigma} \times \left[\frac{-8(\sigma^2 + 4(\mu-X)^2)^2}{(\sigma^2 + 4(\mu-X)^2)^4} + \right. \\ &\left. \frac{16(\sigma^2(\mu-X) + 4(\mu-X)^3) \frac{d}{dX} (\sigma^2 + 4(\mu-X)^2)}{(\sigma^2 + 4(\mu-X)^2)^4} \right] \end{aligned} \quad [14]$$

$$\begin{aligned} \frac{d^2Y}{dX^2} &= \frac{2}{\pi\sigma} \times \left[\frac{-8(\sigma^2 + 4(\mu-X)^2)^2}{(\sigma^2 + 4(\mu-X)^2)^4} + \right. \\ &\left. \frac{16(\sigma^2(\mu-X) + 4(\mu-X)^3)(8(\mu-X))}{(\sigma^2 + 4(\mu-X)^2)^4} \right] \end{aligned} \quad [15]$$

$$\frac{d^2Y}{dX^2} = \frac{2}{\pi\sigma} \times \left[\frac{-8(\sigma^4 + 8(\mu - X)^2 + 64(\mu - X)^4)}{(\sigma^2 + 4(\mu - X)^2)^4} + \frac{128(\sigma^2(\mu - X)^2 + 4(\mu - X)^4)}{(\sigma^2 + 4(\mu - X)^2)^4} \right] \quad [16]$$

Going back to equations 6 and 11, how do the magnitudes of the derivatives change with σ ? Since the maximum first derivative occurs when $X - \mu = \sigma$, let us substitute σ for $X - \mu$ in equation 6a and for the normal distribution we get

$$\frac{dY}{dX_{MAX}} = \frac{-\sigma}{\sigma^2(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{\sigma}{\sigma}\right)^2} = \frac{-e^{-\frac{1}{2}}}{\sigma^2(2\pi)^{1/2}} \quad [17]$$

and in equation 11 for the Lorentzian distribution

$$\frac{dY}{dX} = \frac{2}{\pi\sigma} \times \frac{-8(\sigma)}{(\sigma^2 + 4(\sigma)^2)^2} = \frac{-16\sigma}{25\pi\sigma^5} = \frac{-16}{25\pi\sigma^4} \quad [18]$$

For the normal distribution, the exponential term has become a constant, and we see that the maximum magnitude of the derivative is inversely proportional to σ (regardless of whether we start with the constant-area expression or the constant-height expression). This confirms our observation from Figure 1. For the Lorentzian distribution, we see that the derivative decreases with the fourth power of the bandwidth.

Similarly, the maximum second derivative occurs when $X = \mu$, so inserting this equality into equation 9a for the normal distribution gives us

$$\frac{d^2Y}{dX^2_{MAX}} = \left(0 - \frac{e^{-\frac{1}{2}}}{\sigma^2(2\pi)^{1/2}} \right) = \frac{-e^{-\frac{1}{2}}}{\sigma^2(2\pi)^{1/2}} \quad [19]$$

And substituting $X - \mu = 0$ into equation 16 gives us the corresponding value for the Lorentzian distribution

$$\begin{aligned} \frac{d^2Y}{dX^2} &= \frac{2}{\pi\sigma} \times \frac{-8(\sigma^4 + 8(0)^2 + 64(0)^4) + 128(\sigma^2(0)^2 + 4(0)^4)}{(\sigma^2 + 4(0)^2)^4} \\ &= \frac{2 \times (-8\sigma^4)}{\pi\sigma \times \sigma^8} = \frac{-16}{\pi\sigma^5} \end{aligned} \quad [20]$$

The negative sign in equations 19 and 20 tells us that the maximum second derivative is a negative value, which agrees with Figure 1, and it also tells us that the magnitude of the second derivative decreases inversely as the square of σ (for the normal band shape) and inversely as the fifth power of σ (for the Lorentzian band shape), that is, as the bandwidth of the absorbance band, increases. This explains why the deriva-

tives of the broad absorbance band decrease with respect to the narrow absorbance band as we see in Figure 1, and more so as the derivative order increases.

The Behavior of Computed Derivatives

Now, equations 6 and 9 are mathematically exact. But as we observed when discussing Figure 1, a representation of a derivative based on finite differences is only an approximation. How good is this approximation, and how quickly does it get bad? That depends somewhat on how the derivative is calculated. We made a point of noting that the derivative in Figure 1 was calculated from synthetic data, with abscissa (wavelength) spacing of 1 nm. This value of spacing was chosen so that the two methods of calculation would default to the same result. We note above that the definition of a derivative includes the operation of division by ΔX (or by dX , in the mathematically exact case). Some computer programs that purport to calculate derivatives do not include the step of performing that division, while others do. The results will vary considerably in the two cases. We will begin our discussion by considering the simpler case, where we do not divide by ΔX . This provides the numerator term for the derivative definition, and also for the approximation; this allows us to examine the behavior of that term in isolation. In some cases, this is all that is used or needed: it provides a qualitative observation of the overall shape of the spectrum that is of interest, for example. Sometimes it is done this way when the data are used for quantitative or qualitative analysis, and the spectral data from the unknown samples, the samples which are to be analyzed on a routine basis, are treated the same way as the calibration data. Indeed, since the numerator term differs from the correct derivative approximation only by a scaling factor, it can be difficult to tell just from looking at the derivative curve whether it is a correctly calculated derivative or not, especially if the scale is not present.

On the other hand, computing only the numerator term is not recommended when results are to be compared between different instruments or laboratories. It is also not recommended when performing theoretical studies is of interest, or when the results of experiments are to be compared to theoretical expectations, since it does not, in general, reflect the actual value of the true derivative. Given the minor computational burden, however, the proper computation of including the division should always be done. Here we start with the examination of the numerator term alone for its pedagogical value.

The question arises: Because the definition of the derivative specifies taking a limit as differences approach zero, wouldn't the best results be obtained from using the smallest possible differences? The answer is "Yes, but . . .". The "but" reflects the fact that while synthetic data are noise-free, real data contain noise. In this column we consider only the noise-free synthetic data we created, but it is clear that with real data, containing real and irreducible noise, computing smaller and smaller differences will eventually bring us to the point where the differences equal and then become less than

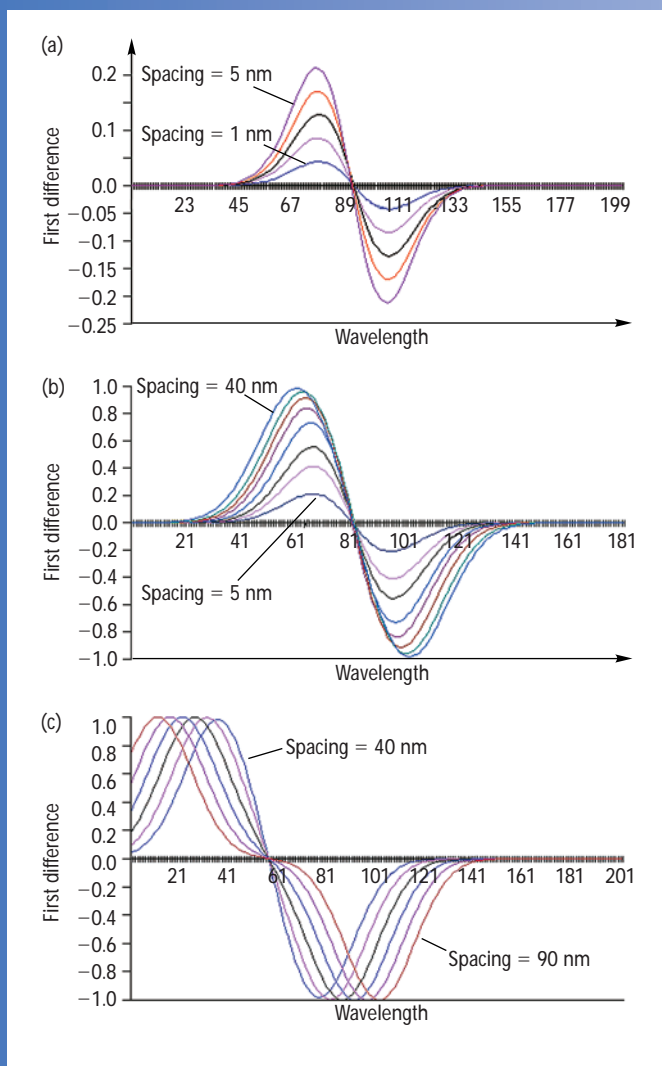


Figure 2. First differences calculated using different spacings between the data points used to calculate the finite difference for the numerator term only, as an approximation to the derivative. The underlying curve is the 20-nm bandwidth absorbance band in Figure 1, with data points every nm. (a) difference spacings = 1–5 nm, (b) spacings = 5–40 nm, (c) spacings = 40–90 nm.

the noise level. Derivative calculations are indeed known to be fraught with noise problems. In the interest of examining the behavior of the derivative, however, we are going to ignore the effect of the noise in this column, although we will eventually return to that question.

The way to minimize noise effects is to exaggerate the differences, by computing finite differences at larger and larger wavelength intervals, and this is often done in practice. In Figure 2 we present the results of computing finite difference approximations to a derivative (for the normal case), using different spacings (that is, the wavelength difference between the data points we compute the finite difference between; we will sometimes call this Δx and freely intermix the two terms). For the derivatives in Figure 2, the underlying ab-

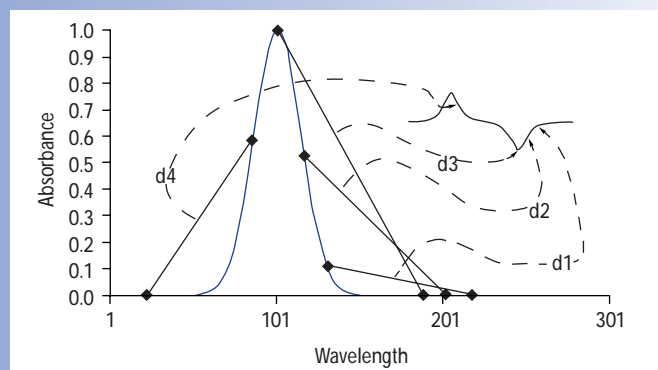


Figure 3. Derivative computed over a very large spacing. With a large spacing, one point used for the difference is on the baseline and the other traces over the shape of the curve.

sorbance curve is the narrower one from Figure 1, having a 20-nm bandwidth.

We see from Figure 2 that, in contrast to the mathematically ideal behavior of a true derivative, the behavior of a finite difference depends on how it is calculated. As Figure 2a shows, at small spacings, the shape of the computed difference curve closely mimics the true derivative, and has a magnitude that is proportional to the spacing. Figure 2b shows that as the spacing increases, several changes occur:

- The relationship between the difference spacing and the magnitude of the derivative departs from the degree of proportionality we observe at smaller spacings. As the spacing increases, the maximum value of the computed difference asymptotically approaches the value of unity.
- There is a shift in the wavelength corresponding to the maximum value of the derivative.
- Close examination of Figure 2b will reveal a decrease in the slope of the difference curve at the point it crosses the x -axis, even though we are not using the denominator term of the derivative calculation.

Figure 2c shows that at sufficiently large spacing values, the concept of this being a derivative breaks down entirely. The derivative curve has separated into two features, each of them appearing to be a normal curve, although one of them is negative. As the spacing continues to increase, the two features move farther and farther apart.

Figure 3 shows how this occurs. When the spacing is very wide — that is, wider than the breadth of the absorbance band near the baseline — one of the points used to compute the difference is always on the baseline, while the other point “rides” over the peak and traces its shape. As the point of the derivative slides along the x -axis, eventually the two points exchange roles, and the other feature is traced out, but with the opposite sign.

Now we look at the second derivative similarly. Some of this has been presented previously in the literature (9), although in less detail than we do here. Figures 4a–c present second derivatives calculated using the same spacings as for the differences in Figure 2. In Figure 4 we see that the second

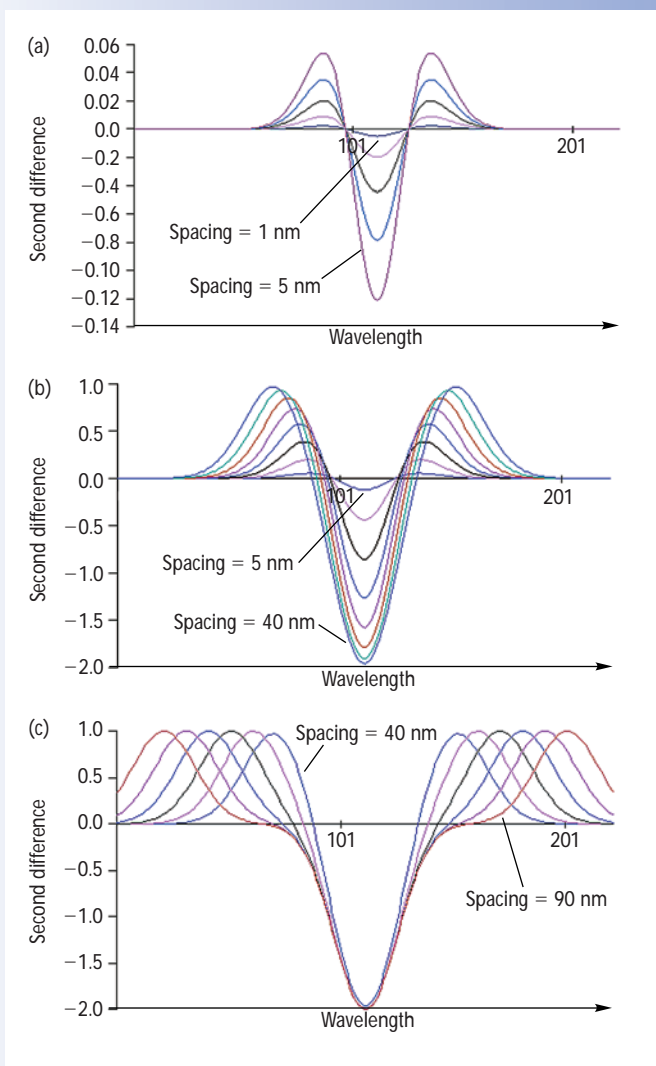


Figure 4. Second differences calculated using different spacings between the data points used to calculate the finite difference for the numerator term only, as an approximation to the derivative. The underlying curve is the 20-nm bandwidth absorbance band in Figure 1, with data points every nm. (a) difference spacings = 1–5 nm, (b) spacings = 5–40 nm, (c) spacings = 40–90 nm.

derivative is subject to some of the same effects as the first derivative:

- Linear (proportional) change in amplitude at small spacings
- Nonlinear change in amplitude at large spacings.

On the other hand, there is no shift in the wavelength of the central maximum, although Figures 4b and 4c show that the artifact peaks do change their wavelength. Replacing the shift in wavelength, however, is a broadening of the central peak. We noted above that one characteristic of the second derivative is the narrowing of this peak compared with the underlying absorbance band. As the spacing over which the derivative is computed increases, however, this resolution enhancement effect decreases and eventually disappears. The

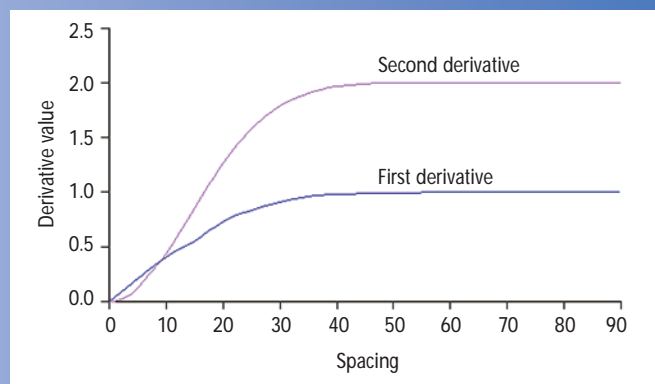


Figure 5. Maximum computed derivative magnitude determined by the spacing of the points used in the computation. Note that the sign of the second derivative has been reversed to simplify comparison with the first derivative behavior.

reason is similar to that for the first derivative, as shown in Figure 3: at very large spacings the points used to compute the derivative eventually wind up simply tracing over the underlying absorbance band, with the result that, since second derivatives are essentially computed from three points, three copies of the underlying absorbance band are produced, albeit with different signs.

In Figure 5 we show the variation of the computed derivatives as determined by the spacing used in the computation. Another feature that can be seen in Figure 5, which is also observable in Figure 4 albeit with some difficulty, is that at small spacings the maximum derivative value is not simply proportional to the spacing but changes faster than proportionately to the spacing; the overall curve of calculated derivative value versus spacing is sigmoidal.

We continue in our next column by examining the behavior of the derivative calculation when the division of the ΔY term is divided by the ΔX term, to form an approximation to the true derivative.

References

1. F. Singleton and G.L. Collier, British patent no. 760,729, applied for December 16, 1953.
2. F. Singleton and G.L. Collier, *Chem. Ind.* (London), 1519 (1955).
3. A.T. Giese and C.S. French, *Appl. Spectrosc.* **9**, 78 (1955).
4. M.J.D. Low and H. Mark, *Appl. Spectrosc.* **24**(1), 129–130 (1970).
5. D. Hopkins, *NIR News* **12**(3), 3–5 (2001).
6. D. Hopkins, *Near Infrared Analysis* **2**, 1–13 (2001).
7. H. Mark and J. Workman, *Spectroscopy* **2**(9), 37–43 (1987).
8. J.D. Ingle and S.R. Crouch, *Spectrochemical Analysis* (Prentice-Hall, Upper Saddle River, NJ, 1988).
9. G.E. Ritchie and H. Mark, *NIR News* **13**(1), 4–6 (2002). ■