

Derivatives in Spectroscopy

Part II – The True Derivative

Howard Mark and Jerome Workman Jr.

We continue where we left off in our last column (1), and we start with some discussion regarding the observations we made concerning the change in the magnitude of the computed values of the derivatives (first and second) as the wavelength spacing over which they are computed is changed. As we normally do when continuing a subseries, we resume the equation numbering and figure numbering from where we left off in the previous column. To recap, we noted that at small spacings, the numerator of the computed approximation to the derivative was a close approximation to the shape of the true derivative, and the magnitude increased as the spacing increased, linearly for the first derivative and faster than linearly for the second derivative. In fairly short order, however, in both cases the rate of increase of the derivative magnitude started falling off as

the spacing continued to increase. The falloff in the rate of increase was accompanied by some secondary effects: wavelength shifts of the peak derivative value, and various kinds of distortion of the shape of the derivative. At very large spacings (larger than the bandwidth of the peak), the derivative was replaced by what was essentially a tracing of the shape of the underlying peak: a double tracing for the first derivative and triple for the second derivative.

At small spacing values, however, it now becomes clear why increasing the spacing is desirable. Because in real data the noise of the measured spectrum is constant (because the underlying spectrum from which various derivative approximations are calculated is the same spectrum each time), increasing the spacing of the derivative computation increases the signal part of the signal-to-noise ratio (S/N), thereby improving S/N. As we saw, however, too-large spacings were deleterious, both for distorting the shape of the peak and for producing inaccurate approximations to the derivative numerator.

So now the question arises: what is the nature of the way the magnitude increases with spacing? In Figure 6a we show an expanded view of the region of the first derivative of the normal curve, around the region of the maximum of the underlying (normal) absorbance band. Similarly, Figure 6b shows the corresponding view for the second derivative. The first derivative is well-approximated by a straight line in this region. The second derivative is seen to be approximated by a parabola; a not-unexpected result when considering that this represents the result obtained from a truncated Taylor series approximation of the curve.



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 h1mark@nearinfrared.com.

Jerome Workman Jr.

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



Therefore, for a first derivative, as the x spacing increases, the magnitude of the derivative calculation increases proportionately. In the case of the second derivative, increasing magnitude of the x spacing causes the magnitude of the calculated derivative to increase as the square of the spacing; this is the source of the initial upward curvature we noted in Figure 5 of our previous column (1) for the second derivative.

Better Derivative Approximations

Now we will examine the behavior of the derivative approximation when both the numerator and denominator terms are used. In Figure 7 we present the curves of this computation of the derivative corresponding to the numerator-only computation presented in Figure 2 of our previous column (1). Here we note several differences between Figure 7 and Figure 2. In Figure 7a we see that there is virtually no difference between any of the five curves; they are all producing essentially the same values, in contrast to Figure 2, in which the differences were increasing with spacing. The reason is that for the range of spacings used, all the derivative approximations calculated are reasonably good approximations of the true derivative. Therefore, because they all estimate the same true value, they are all essentially equal to each other.

In Figure 7b we notice even more differences from the corresponding part of Figure 2. The first thing we notice is one characteristic that is the same: the maximum values of the derivative curves shift as the spacing increases. But another difference that is at least as prominent is that the maximum value decreases as the spacing increases. This is exactly opposite to the behavior we noticed in the numerator term, where the maximum increased with the spacing. A third difference we notice from the corresponding part of Figure 2 is that at the point where the first derivative crosses the x -axis, the slope of the derivative also decreases with increasing spacing, while in Figure 2b the slope increased with the spacing, except for the largest values of spacing included in that plot.

Similarly, in Figure 7c both the maximum value of the derivative and the slope at the zero-crossing decrease, whereas in Figure 2c the maximum of the calculated derivative remained constant, although the slope at the zero-crossing decreased.

In the three parts of Figure 8 we see that the second derivative behaves similarly, except that it starts out smaller than the first derivative does, by almost an order of magnitude.

Figure 9 confirms this: the second derivative is smaller than the first (remember, all this is for the normal distribution; other distributions may behave differently). Figure 9 also shows how the correct computation of the derivative differs from the computation of the numerator only, which we saw in reference 1. The so-called derivative computed from the numerator term only increased and then leveled off as the spacing increased, whereas Figure 9 shows that the correct computation starts out with an (almost) constant value of the derivative, which then decreases, with an asymptotic approach to zero.

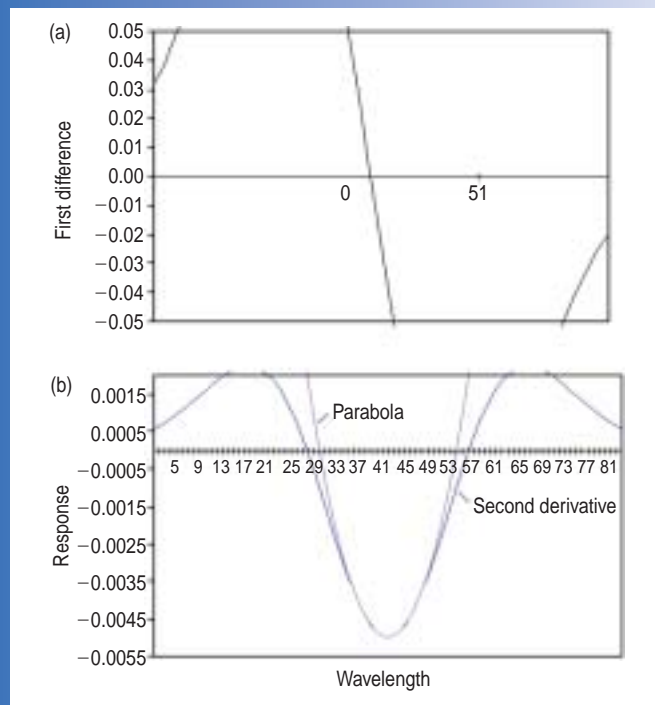


Figure 6. Expansions of the first and second derivative curves. (a) The region around the zero-crossing of the first derivative can be approximated with a straight line. (b) The region around the peak of the second derivative can be approximated with a parabola.

Can we explain all these effects? Of course we can, and in fact, the explanation is almost obvious. When spacings are small, the computed derivative is a good approximation to the true derivative. As long as this is the case, the exact value of ΔX used to compute the derivative is unimportant, because as we saw in Figure 5, the first difference ΔY increases almost linearly with ΔX . Therefore, all values of ΔX give the same result for the computation, because $\Delta Y/\Delta X$ is constant regardless of spacing.

As we observed from Figure 5, however, as ΔX continues to increase, ΔY ceases to increase proportionately. Strictly speaking, this happens immediately when ΔX becomes finite, and the question of whether the amount is noticeable is a matter of degree, how much difference it makes in a particular application. Nevertheless, whatever point that is, the initial increase in ΔX carries a corresponding increase in ΔY , and beyond that point it is no longer proportional. At that point, the computed value of the estimate of the true derivative starts to decrease.

Furthermore, as we noted last time, at sufficiently large spacings (ΔX) the numerator term ceased to increase. As we also noted before, at this point the various data points used for the computation are each individually tracing out the shape of the underlying curve. Therefore, because this constant maximum value is divided by a continually increasing value, as ΔX in the denominator continues to increase, we can expect that the quotient $\Delta Y/\Delta X$ will decrease, and this is the behavior we observe.

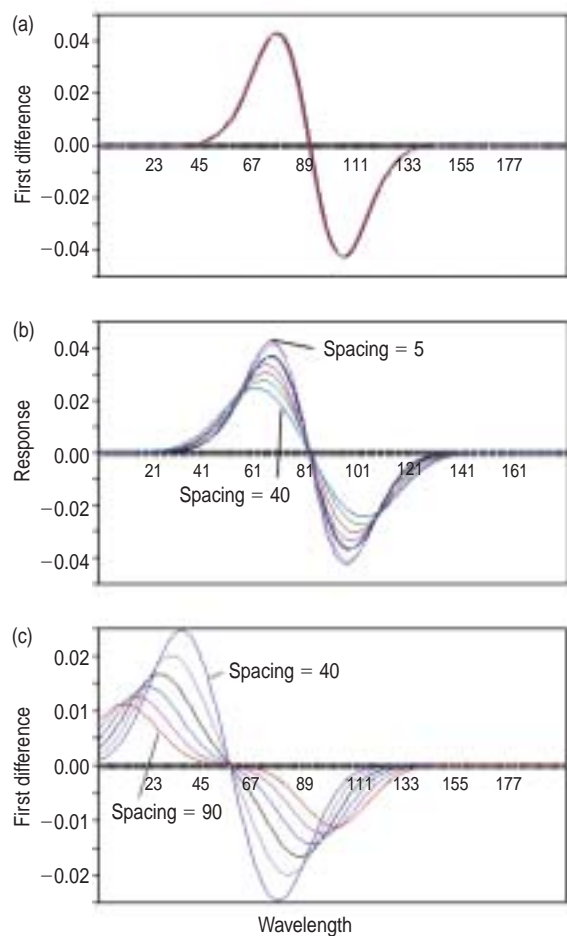


Figure 7. First derivatives calculated using spacings of (a) 1–5 nm, (b) 5–40 nm, and (c) 40–90 nm for finite difference approximation to the true derivative. The underlying curve is the 20-nm bandwidth absorbance band in Figure 1, with data points every nanometer.

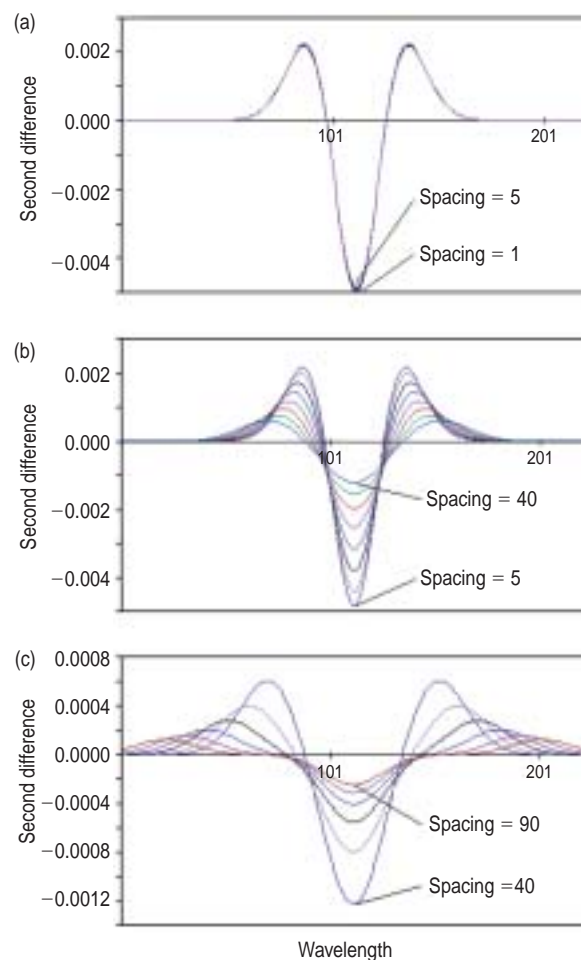


Figure 8. Second derivatives calculated using spacings of (a) 1–5 nm, (b) 5–40 nm, and (c) 40–90 nm for the finite difference approximation to the true derivative. The underlying curve is the 20-nm bandwidth absorbance band in Figure 1, with data points every nanometer.

One final point: we see from Figure 9 that, as we noted before, the true value of the second derivative of a normal curve (at its maximum) is roughly an order of magnitude smaller than the first derivative (or at least, the largest value of the first derivative). In the presence of noise, therefore, S/N will be degraded by this factor, from this cause alone.

We also have noted before that adding or subtracting noisy data causes the variance to increase as the number of data points added together (2). The noise of the first derivative, therefore, will be larger than that of the underlying absorbance band by a factor of the square root of two.

We also showed previously that if a random variable (for example, a measurement contaminated with noise) is multiplied by a constant (c , say), then the variance of the product is increased by a factor of c^2 (3). A second derivative calculation is equivalent to using coefficients 1, -2 , 1 as multipliers for three data points spaced at the desired x -spacing. The variance of the spectrum, then, is multiplied by $1^2 + 2^2 + 1^2 = 6$. Therefore, the standard deviation of the noise contribu-

tion to a first derivative is $2^{1/2}$ greater than the noise of the spectrum, while the noise contribution to the second derivative is $6^{1/2}$ times the noise of the spectrum. Therefore, the noise of the second derivative is $3^{1/2}$, or roughly 1.7, times that of the first derivative. So from both aspects, the signal-to-noise ratio of the second derivative is worse than that of the first. The increase of the noise is clearly the lesser of the contributions, compared with the full order of magnitude reduction of the signal part of S/N. Second derivatives have become *de rigueur* as a data treatment of choice for spectral data, and there are reasons for that, which we have discussed. But they also carry with them the burden of a severely reduced S/N compared to first derivatives. When selecting a data treatment, therefore, one should know the disadvantages as well as the benefits of each one. Although derivative treatments have long been in use for analysis of spectroscopic data, the quantitative study of the derivative transform has not previously been widely disseminated, but is worth having. There may be times when a second derivative

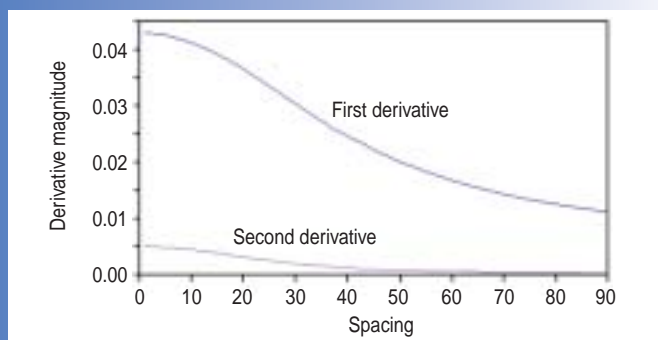


Figure 9. Maximum magnitudes of first and second derivative approximations as the spacing is varied.

transform is not giving adequate results, and in some of those cases, using a first derivative transform may be preferable.

REFERENCES

1. H. Mark and J. Workman, *Spectroscopy* 18(4), 32–37 (2003).
2. H. Mark and J. Workman, *Spectroscopy* 3(3), 40–42 (1988).
3. H. Mark and J. Workman, *Spectroscopy* 3(8), 13–15 (1988). ■

Corrections to Part I of This Series

A sharp-eyed reader found an error in our derivation of the derivative of the normal distribution (1). In going from Equation 3 to Equation 4, when we factored the constants from the derivative, we should have taken out $1/\sigma^2$, whereas we accidentally took out $1/\sigma$. Therefore, several equations, from Equation 4 on, are off by a factor of σ . The correct equations follow:

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

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

$$\frac{dY}{dX} = \frac{-(X-\mu)}{\sigma^3(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad [6a]$$

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

Similarly, the correct equations for the second derivative of the Normal distribution are as follows:

$$\frac{d^2Y}{dX^2} = \frac{-(X-\mu)}{\sigma^3(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \frac{d}{dX} \left(-\frac{1}{2} \left(\frac{X-\mu}{\sigma} \right)^2 \right) + e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \frac{d}{dX} \left(\frac{-(X-\mu)}{\sigma^3(2\pi)^{1/2}} \right) \quad [7]$$

$$\frac{d^2Y}{dX^2} = \frac{-(X-\mu)}{\sigma^3(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \left(-\frac{1}{2\sigma^2} 2(X-\mu) \right) + \left(\frac{-1}{\sigma^3(2\pi)^{1/2}} \right) e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad [8]$$

$$\frac{d^2Y}{dX^2} = \left(\frac{(X-\mu)^2}{\sigma^5(2\pi)^{1/2}} - \frac{1}{\sigma^3(2\pi)^{1/2}} \right) e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad [9a]$$

$$\frac{d^2Y}{dX^2} = \left(\frac{(X-\mu)^2}{\sigma^4} - \frac{1}{\sigma^2} \right) e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad [9b]$$

$$\frac{dY}{dX_{MAX}} = \frac{-\sigma}{\sigma^3(2\pi)^{1/2}} e^{-\frac{1}{2}\left(\frac{\sigma}{\sigma}\right)^2} = \frac{-e^{-\frac{1}{2}}}{\sigma^2(2\pi)^{1/2}} \quad [17]$$

$$\frac{d^2Y}{dX^2_{MAX}} = \left(0 - \frac{e^{-\frac{1}{2}}}{\sigma^3(2\pi)^{1/2}} \right) = \frac{-e^{-\frac{1}{2}}}{\sigma^3(2\pi)^{1/2}} \quad [19]$$

The formulas given for the Lorentzian distribution are correct. Also, because the differences are constant scaling factors, the relative magnitudes and the discussions of the derivatives are unaffected, except that the second derivative of the normal distribution changes inversely as the cube of σ rather than its square.

Our thanks to Colin Christy for pointing out this error.