

Derivatives—a systematic approach to removing variability before applying chemometrics

David W. Hopkins

NIR Consultant, Battle Creek, Michigan, USA.

E-mail: dwh @ voyager.net

Introduction

Derivatives often present the data in spectra in a manner that brings out the information that we are interested in, such as the number and position of bands and their relative intensity and removes unwanted baseline variation. Two major methods of calculating derivative spectra are the Savitzky–Golay method of polynomial curve-fitting and the segment–gap method. Both of these methods can be presented and/or performed as convolution processes. The definition of segment–gap derivatives is clarified by presenting them as convolution functions. A new, normalised form of segment–gap derivative is derived that preserves the basic shapes of the calculated derivatives but brings the results into agreement with the Savitzky–Golay results over a range of parameters in both methods. The effectiveness of the convolution functions in removing the random high frequency noise in the spectra can be evaluated by calculating an index called RSSK/Norm, the square root of the sum of the squares of the convolution coefficients, divided by the normalisation constant. The agreement and utility of the two methods of calculating derivatives is demonstrated by evaluating the results of using derivatives up to the fourth order on the second overtone aromatic CH band of polystyrene, as an example, in the 1680 nm region of the NIR spectrum.

Basics

First, let us consider measurements of y (or Absorbance) measured at equal intervals of x (or wavelength). In reality, a derivative cannot be calculated in the mathematical sense from measurements taken at discrete intervals, because a derivative is defined for continuous functions as the limit of $\Delta y / \Delta x$ as Δx approaches 0. This analytical geometry approach is suggested by the dotted line “tangent” to the curve at 1672 nm in Figure 1. Therefore, the finite difference in y values for a selected difference in x is taken as an approximation to the desired first derivative. Various methods of calculating derivatives include the point-difference method, gap method, Savitzky–Golay¹ method and segment–gap method. (The segment–gap method was called the Norris Method in recent reviews.^{2,3} However, in accordance with Karl Norris’ wishes, the non-personal reference should be applied.) The first two methods are available in Grams software,⁴ and may be considered special cases of the segment–gap method. The Savitzky–Golay method is implemented in many software packages, Grams, Pirouette,⁵ Unscrambler⁶ and Vision/NSAS software,⁷ to name some I have used. There is not much detail written about the segment–gap method outside of the manuals from Foss/NIRSystems, so this method will be described in some detail here.

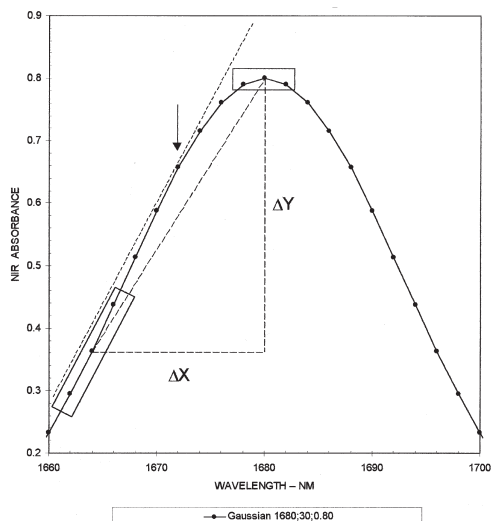


Figure 1. Calculation of a first derivative.

Savitzky–Golay method

Figure 1 illustrates the calculation of a first derivative. In this example, the slope $\Delta y / \Delta x$ is calculated for an interval of nine points. In the Savitzky–Golay method, the data points in the region around a central point (arrow in Figure 1) are fitted to a polynomial and the analytical value of the derivative at

Table 1. Savitzky–Golay method: first derivatives.

Index	Quadratic polynomial fitting				Cubic polynomial fitting			
	5-point	7-point	9-point	11-point	5-point	7-point	9-point	11-point
-5	—	—	—	-5	—	—	—	300
-4	—	—	-4	-4	—	—	86	-294
-3	—	-3	-3	-3	—	22	-142	-532
-2	-2	-2	-2	-2	1	-67	-193	-503
-1	-1	-1	-1	-1	-8	-58	-126	-296
0	0	0	0	0	0	0	0	0
1	1	1	1	1	8	58	126	296
2	2	2	2	2	-1	67	193	503
3	—	3	3	3	—	-22	142	532
4	—	—	4	4	—	—	-86	294
5	—	—	—	5	—	—	—	-300
Norm	10	28	60	110	12	252	1188	5148
RSSK/Norm	0.316	0.189	0.129	0.095	0.950	0.512	0.338	0.246

the mid-point of the interval is taken as the value of the derivative at the wavelength of the central point. Savitzky and Golay¹ showed that the least-squares curve fitting and differentiation could be done in a single convolution operation, simply and elegantly. Convolution involves aligning the values of the convolution function with the values of the spectrum, multiplying the pairs of values, summing the products and dividing the sum by a normalisation constant. The result is entered as the value of the derivative at the central point. This process is then repeated by advancing to the points surrounding the next spectral point, so that a derivative spectrum is determined for the whole spectrum, point by point. They published convolution functions that enable the direct calculation of a smoothed spectrum, or derivatives up to an order of fifth derivative. Their tables contained a number of errors, which were corrected by Steinier *et al.*⁸ In Figure 1, a 9-point convolution interval is used to calculate the value at 1672 nm. Values for points in the spectrum are calculated by sliding the convolution interval along the data and calculating the derivative, point by point. The loss of undefined values at the ends of the spectrum is generally not a concern.

Table 1 also contains a parameter for each convolution function called RSSK/Norm, the square root of the sum of the squares of the convolution coefficients, divided by the normalisation constant. Howard Mark⁹ derived the relationship (Equation 79, p. 56) for an index of random error, the square root of the sum of squares of the calibration coefficients. This index expresses the sensitivity of calibration equations to the random errors in the spectral data. For results calculated by convolution, I have

Table 2. Segment-gap first derivatives.

Index	Segment, gap values (nanometers)								
	2, 2	2, 6	2, 10	6, 2	6, 6	6, 10	10, 2	10, 6	10, 10
-7	—	—	—	—	—	—	—	—	-1
-6	—	—	—	—	—	—	—	-1	-1
-5	—	—	—	—	—	-1	-1	-1	-1
-4	—	—	—	—	-1	-1	-1	-1	-1
-3	—	—	-1	-1	-1	-1	-1	-1	-1
-2	—	-1	0	-1	-1	0	-1	-1	0
-1	-1	0	0	-1	0	0	-1	0	0
0	0	0	0	0	0	0	0	0	0
1	1	0	0	1	0	0	1	0	0
2	—	1	0	1	1	0	1	1	0
3	—	—	1	1	1	1	1	1	1
4	—	—	—	—	1	1	1	1	1
5	—	—	—	—	—	1	1	1	1
6	—	—	—	—	—	—	—	1	1
7	—	—	—	—	—	—	—	—	1
Norm	1	1	1	3	3	3	5	5	5
RSSK/Norm	1.414	1.414	1.414	0.816	0.816	0.816	0.632	0.632	0.632

called the corresponding term RSSK/Norm. Simplifying the error analysis in this manner assumes that all measurements in the convolution interval have the same error. Thus, the RSSK/Norm represents a figure of merit for the noise reduction factor for the application of the convolution function, because the signals or magnitude of the resulting bands in the convoluted spectra are largely constant, over a range of convolution interval s andes.

Segment-gap method

For slowly changing functions, the derivative can often be approximated by taking the difference in y values for x locations separated by more than 1, up to as many as 15 points. One can see when too large a separation is selected, by observing that the basic shape of the derivative is degraded or changed, or minor or superimposed bands cease to be resolved. For such slowly changing functions, derivative curves with less noise can be obtained by taking the difference of two averages, formed from points surrounding the selected x locations. As a further simplification, the division of the difference in y values, or the difference in y -averages, by the x difference, Δx , is omitted. The term segment indicates the length of the x intervals over which y values are averaged, to obtain the two values that are subtracted to form the estimate of the derivative. The gap is the length of the x interval that separates the two segments that are averaged. This is shown in Figure 1, where a segment of three points and gap of five points is illustrated. The absorbance values in each of the boxes are averaged and the difference is taken as Δy , which is the value of the “derivative”, because the division by the Δx term is taken as

Table 3. Normalised segment-gap derivatives: first derivatives.

Index	Point-diff	Segment, gap values (nanometers)								
		2, 2	2, 6	2, 10	6, 2	6, 6	6, 10	10, 2	10, 6	10, 10
-7	—	—	—	—	—	—	—	—	—	-1
-6	—	—	—	—	—	—	—	—	-1	-1
-5	—	—	—	—	—	—	-1	-1	-1	-1
-4	—	—	—	—	—	-1	-1	-1	-1	-1
-3	—	—	—	-1	-1	-1	-1	-1	-1	-1
-2	—	—	-1	0	-1	-1	0	-1	-1	0
-1	—	-1	0	0	-1	0	0	-1	0	0
0	-1	0	0	0	0	0	0	0	0	0
1	1	1	0	0	1	0	0	1	0	0
2	—	—	1	0	1	1	0	1	1	0
3	—	—	—	1	1	1	1	1	1	1
4	—	—	—	—	—	1	1	1	1	1
5	—	—	—	—	—	—	1	1	1	1
6	—	—	—	—	—	—	—	—	1	1
7	—	—	—	—	—	—	—	—	—	1
Norm	1	2	4	6	12	18	24	30	40	50
RSSK/Norm	1.414	0.707	0.354	0.236	0.204	0.136	0.102	0.105	0.079	0.063

constant scale factor that is ignored. Ignoring the division by the Δx term also has the result that derivatives calculated with increasing segments and gaps have greater amplitudes. This feature is a major distinguishing characteristic between segment–gap derivatives and other procedures. The segment–gap derivatives are a very flexible and useful technique for analysing spectra, particularly when the wavelength is sampled at intervals less than the spectral bandwidth of the instrument. Because the segment and gap method is available in the Foss/NIRSystems software that I have used, all further discussion will describe the features of that implementation.

In practice, the length of the segment and gap intervals is specified in actual x units (i.e. nanometers). Furthermore, to maintain wavelength accuracy, the segment and gap intervals are constrained to be an odd number of points, so that averages or results correspond to the midpoints of the in-

Table 4. Normalised segment–gap derivatives: second derivatives.

Index	Segment, gap values (nanometers)								
	2, 2	2, 6	2, 10	2, 14	2, 18	6, 2	6, 6	6, 10	10, 2
-10	—	—	—	—	1	—	—		—
-9	—	—	—	—	0	—	—	1	—
-8	—	—	—	1	0	—	—	1	1
-7	—	—	—	0	0	—	1	1	1
-6	—	—	1	0	0	—	1	0	1
-5	—	—	0	0	0	1	1	0	1
-4	—	1	0	0	0	1	0	0	1
-3	—	0	0	0	0	1	0	0	0
-2	1	0	0	0	0	0	0	0	-2
-1	0	0	0	0	0	-2	-2	-2	-2
0	-2	-2	-2	-2	-2	-2	-2	-2	-2
1	0	0	0	0	0	-2	-2	-2	-2
2	1	0	0	0	0	0	0	0	-2
3	—	0	0	0	0	1	0	0	0
4	—	1	0	0	0	1	0	0	1
5	—	—	0	0	0	1	1	0	1
6	—	—	1	0	0	—	1	0	1
7	—	—	—	0	0	—	1	1	1
8	—	—	—	1	0	—	—	1	1
9	—	—	—	—	0	—	—	1	—
10	—	—	—	—	1	—	—		—
Norm	4	16	36	64	100	48	108	192	180
RSSK/Norm	0.612	0.153	0.068	0.038	0.024	0.088	0.039	0.022	0.030

tervals (rather than allow an error of a half-interval that occurs if even intervals are allowed). In the case where a 2 nm sampling interval is used for spectra, specifying a gap of either 0, 1 or 2 nm (for example) results in the same computation, with a gap of one point. The next higher permitted gap values are 6, 10, 14 nm, etc., due to the requirement of specifying intervals with an odd number of points. Similarly, the permitted segment values are 2, 6, 10, 14 nm, etc., for the same reason. This ambiguity (that deep rounding of the input occurs, with the effect that the identical results are obtained with specified intervals of 3, 4, 5 or 6 nm, for example) is generally invisible to the user, but it is good to keep it in mind when trials for optimal conditions are made.

The calculation of the segment–gap derivatives can readily be expressed as application of convolution functions to the spectral data. The averages in the segments and the gap derivatives can be calculated in a single process of multiplying the spectral values in the convolution interval by factors of 1, summing and dividing by a normalisation factor that performs the averaging within the segments. This normalisation only accounts for the averaging in the segment and the derivative is still not normalised for the length of the convolution interval. The convolution functions for calculating first derivatives with selected combinations of segment and gap are presented in Table 2.

A normalised form of the segment–gap derivatives can be calculated by the application of the same formula that can be deduced from the Savitzky–Golay first derivatives, Equation (1).

$$\text{Norm} \sum_{i = -(n-1) / 2}^{(n-1) / 2} i * C_i \quad (1)$$

where i is the index of the set of convolution function integers, C_i are the convolution function integers and n is the number of points in the convolution interval.

The convolution functions for calculating first derivatives with selected combinations of segment and gap for the normalised segment–gap derivatives are presented in Table 3.

Second derivatives are frequently used in NIR spectral analysis. Fourth derivatives have been used for UV and Visible spectral analysis,^{10,11} but not widely used in NIR analysis. Selected second derivative convolution functions are presented for the normalised segment–gap method in Table 4. The usual segment–gap method convolution functions for the second and fourth derivatives are similar to those given in Table 4, except that the Norm values are 1, 3, 5, etc, depending on the number of points in the segment chosen, as seen already for the first derivatives in Table 2. The normalisation factors for the second, third and fourth derivatives are calculated from equations (2), (3) and (4), which can be deduced from the corresponding Savitzky–Golay tables. For these equations, the index i takes the same range as in Equation (1).

$$\text{Norm} = 1/2 \sum i^2 C_i \quad (2)$$

$$\text{Norm} = 1/6 \sum i^3 C_i \quad (3)$$

$$\text{Norm} = 1/24 \sum i^4 C_i \quad (4)$$

Applications and discussion

The application of various methods for calculating derivatives to a scan of a piece of polystyrene 0.94 mm thick, backed by Spectralon (Labsphere, Inc, North Sutton, NH, USA) and measured in reflectance mode in a Foss/NIRSystems Model 6500 scanning spectrophotometer. The polystyrene was cut from the bottom of a disposable Petri dish (Baxter Healthcare Corp., McGaw Park, IL, USA).

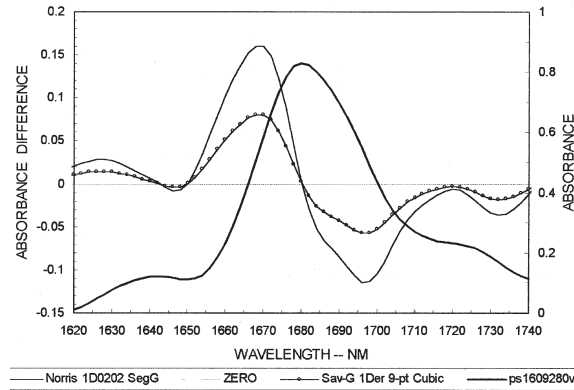


Figure 2. First derivatives of polystyrene, 1620 to 1740 nm region. Bold curve is the original spectrum; the fine solid curve is the segment-gap derivative, calculated with segment and gap of 2 nm; the curve marked with circles is the calculated with the Savitzky-Golay method, 9-point cubic fit.

The results of using these convolution functions in an Excel spreadsheet agree perfectly to the fifth decimal place, with the normal calculations carried out in the Vision and NSAS software (segment-gap derivatives) and Pirouette, Unscrambler and Grams32 (Savitzky-Golay derivatives). The new convolution functions were generated and evaluated in Excel. The results will be presented for the second overtone aromatic CH band of polystyrene as an example. Although the 1680 band appears as a nice, single absorption band, the results demonstrate the ability of the derivatives to identify bands that are highly overlapping. The first derivative results (Figure 2) suggest that this band is actually at least a doublet of very closely situated absorptions, from the two different slopes in the 1670–1695 nm region.

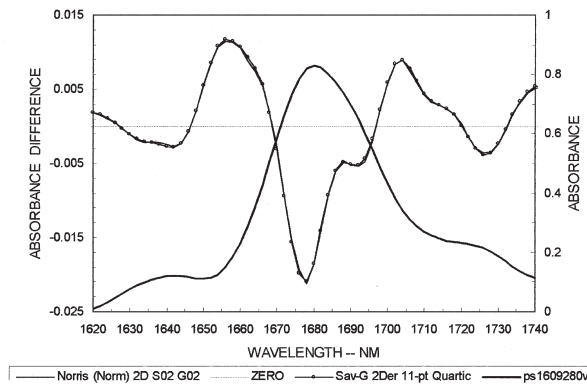


Figure 3. Second derivatives start to resolve two bands in the 1680 nm region. Bold curve is the original spectrum; the fine solid curve is the segment-gap derivative (Normalised), calculated with segment and gap of 2 nm; the curve marked with circles is the calculated with the Savitzky-Golay method, 11-point quartic polynomial fit.

Figure 2 demonstrates a key difference between the segment-gap and Savitzky-Golay results. The segment-gap derivative is of greater magnitude, because no normalisation for the segment and gap length is made. This is usually of no importance when the method is used in a calibration or chemometric analysis; it shows up clearly when results are compared, as in Figure 2. Figure 2 demonstrates that the curve shapes are comparable, differing only by a scale factor.

Normalisation of the segment-gap derivatives as described above, using Equations (1) through (4), brings the segment-gap derivatives into the same scale as the Savitzky-Golay convolution methods. This works because the normalisation constant calculated by Equations (1) through (4) properly determines the effective Δx for the method of calculating Δy determined by the convolution factors, to properly determine the derivative estimator $\Delta y / \Delta x$. Figure 4 compares the results of the segment-gap (normalised) first derivative with segment and gap of 2 nm (only three points in the convolution interval, but RSSK/Norm is 0.707) with the Savitzky-Golay method, 9-point cubic polynomial fit, RSSK/Norm of 0.338); the agreement is so close that the two curves are practically superimposed.

The second derivative suggests the occurrence of bands at 1678 and 1690 nm (Figure 3). This demonstrates the band sharpening and better resolution of complicated spectra that has been noted.^{10,11} Looking critically, the asymmetry of the 1680 band can be noted in the raw absorbance scan, suggesting that the band may be more complex than a single absorber. The segment-gap (normalised) second derivative (5-point convolution interval, RSSK/Norm of 0.612) certainly agrees well with the Savitzky-Golay method, 11-point quartic polynomial fit, RSSK/Norm of 0.256).

Third derivative results also indicate the presence of two bands (results not shown). Further, the fourth derivative results show the clearest discrimination of multiple bands in the 1680 nm region of polystyrene; band positions are indicated by the positive peaks in the derivatives (arrows at 1677 and 1694 nm in Figure 4). Clearly the segment-gap (Normalised) method (nine points, RSSK/Norm of 0.523) and the Savitzky-Golay 9-point quartic fit method agree very well (RSSK/Norm of 0.313).

The fourth derivatives (see Figure 4) also suggest the resolution of other bands in the neighborhood of 1680, possibly due to second overtones of $-\text{CH}-$ and $-\text{CH}_2-$. There are suggestions of bands at 1632 and 1645 nm that contribute to the low-intensity band observed at about 1642 in the polystyrene

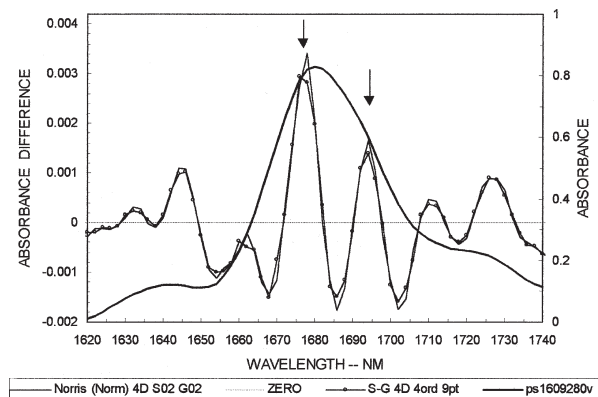


Figure 4. Fourth derivatives of polystyrene clearly indicates two bands in the 1680 region; band positions are indicated by the positive peaks in the derivatives (arrows at 1677 and 1694 nm). Bold curve is the original spectrum; the fine solid curve is the segment-gap Derivative (Normalised), calculated with segment and gap of 2 nm; the curve marked with circles is the calculated with the Savitzky-Golay method, 9-point 4th order polynomial (quartic) fit.

absorbance spectrum. Similarly, the shoulder at the 1720 nm region appears to be due to two underlying bands indicated at 1711 and 1727 nm. The band at 1662 nm is possibly false, an artifact of superposition of side-bands of the adjacent bands. This problem has been noted previously,^{10,11} with the suggestion that other methods of spectral characterisation, for example, low temperature spectroscopy or fractionation and purification of various components, should be employed to demonstrate the separation of physically distinct absorbers. Morrey¹² has described the problems that can occur in the application of derivatives to band structure elucidation; he showed that bands can be created in a third-band illusion, or obliterated by the adjacent bands. Such problems need to be recognised in the interpretation of higher order derivatives, but they should not preclude their use.

The selection of which derivative method to use depends significantly on evaluating whether a result is too noisy, or so highly smoothed that details are distorted or lost. It may be necessary to accept a larger amount of noise to obtain good spectral resolution. The trade-off may depend on the object of the analysis. The RSSK/Norm values can offer some guidance in selection.

Conclusions

The segment–gap derivatives have been presented in detail, interpreting the method in terms of convolution. This comparison with the Savitzky–Golay procedures suggests an elegant method to normalise the segment–gap derivatives to obtain consistent signal magnitudes even at various segment and gap selections. The RSSK/Norm values show that even derivatives that do not explicitly involve smoothing achieve signal-to-noise improvement, by the utilisation of gaps of one or more points. The higher order derivatives are useful in interpreting complex spectra and it may be expected that derivatives as high as the fourth order will be useful in chemometric analysis of qualitative and quantitative applications.

The RSSK/Norm values provide a basis for comparing the effectiveness of convolution methods in reducing the propagation of spectral noise. Longer convolution intervals reduce the noise in the results and, typically, choosing a higher order polynomial results in higher noise levels, which can be offset by the ability to utilise longer convolution intervals. However, observation of the results of convolution by any method is necessary, to avoid over-smoothing and loss of resolution of underlying bands. From the results of analysing the polystyrene spectra, it appears that derivative convolution functions with mid-range RSSK/Norm values between approximately 0.3 to 0.7 are most effective for analysing the underlying band structure in the 1680 nm region. Values greater than 1 indicate poor noise rejection, while values less than 0.1 indicate significant loss of band structure at the expense of the noise reduction.

Note added in proof

Upon further presentation and discussion of these calculations, it seems useful to simplify the nomenclature and call the RSSK/Norm values RSSC, root sum of squares of coefficients.

Acknowledgements

This paper reflects my long fascination with the elegance of the convolution process and my hope to bring a little more science to the art of spectral evaluation. I would like to thank Karl Norris, Howard Mark, Richard Kramer and Fred McClure for long discussions on the calculation and interpretation of derivatives.

References

1. A. Savitzky and M.J.E. Golay, *Anal. Chem.* **36**(8), 1627 (1964).
2. D.W. Hopkins, *NIR news* **12**(3), 3 (2001).
3. D.W. Hopkins, *Near Infrared Analysis* **2**(1), 1 (2001).

4. Grams/32 Software is part of a chemometrics package from Thermo Galactic, Salem, NH, USA.
5. Pirouette Software is a chemometrics package from Infometrix, Inc, Woodinville, WA, USA.
6. Unscrambler Software is a chemometrics package from CAMO, Inc, Oslo, Norway.
7. Vision and NSAS are software programs for the collection and analysis of NIR data, available from FOSS/NIRSystems, Silver Spring, MD, USA.
8. J. Steinier, Y. Termonia and J. Deltour, *Anal. Chem.* **44(11)**, 1906 (1972).
9. H. Mark, *Principles and Practice of Spectroscopic Calibration*. John Wiley & Sons, New York, USA, p. 170 (1991).
10. W.L. Butler and D.W. Hopkins, *Photochem. Photobiol.* **12**, 439 (1970).
11. W.L. Butler and D.W. Hopkins, *Photochem. Photobiol.* **12**, 451 (1970).
12. J.R. Morrey, *Anal. Chem.* **40(6)**, 905 (1968).