

# Comments on the Savitzky–Golay Convolution Method for Least-Squares Fit Smoothing and Differentiation of Digital Data

Sir: In 1964 a paper by Savitzky and Golay (1) describing a technique of smoothing and obtaining smoothed derivatives using convolution arrays derived from the coefficients of least-squares-fit formulas was published in this journal. That paper has become a classic judging by the number of citations (2)—well over 350 in more than 100 different journals—since its publication despite the fact that the paper contained numerous errors in the tables of convolution arrays. These errors were first pointed out and almost completely corrected by Steinier, Termonia, and Deltour (3) in 1972. The purpose of this letter is threefold: (1) to present two simple equations that may be used to check for errors in the arrays; (2) to note the simplicity involved in generating arrays wider than 25 points (the maximum width in the Savitzky–Golay tables), giving equations that may be used to easily expand the widths of all the convolution arrays discussed in the Savitzky–Golay paper; and (3) to call attention to examples that illustrate the beneficial use of smoothing techniques in extracting information other than simply peak position, height, and width from spectral data made up of peaks of known profile. More detailed discussions of least-squares-fit smoothing, than will be given here, may be found in the review papers of Nernst (4) and of Willson and Edwards (5).

The Savitzky–Golay convolution smoothing technique is based on fitting an array of  $n$  ( $= 2m + 1$ , with  $m$  a positive integer from 1 to 12) equally-spaced data points to a polynomial,

$$U(x) = c_0 + c_1x + c_2x^2 + \dots + c_jx^j \\ = \sum_{q=0}^j c_q x^q \quad (1)$$

These  $n$  points are normally a subgroup of a larger group of  $N$  points  $\{u_i(x_i)\}$  ( $x_{i+1} - x_i = \text{constant}$ ) and to smooth the entire set of points (except for  $m$  points at each end of the data array that must be treated separately) one takes a moving average. General formulas for smoothing by moving-average fits to a polynomial may be found in textbooks on the statistical handling of data (6). Savitzky and Golay pointed out that application of these smoothing formulas is equivalent in digital computers to convoluting a uniformly-spaced data array with a set of smoothing coefficients derived from the least-squares-fit formulas. They further pointed out that, while the coefficient  $c_0$  in Equation 1 gives the smoothed value of the data point at the middle of the  $n$ -point array being fitted to the polynomial ( $= \bar{u}_i$ ), the other coefficients  $c_1, c_2, \dots$ , and  $c_q$  in Equation 1 correspond to the smoothed values of the 1st, 2nd, ..., and  $q$ th derivatives at the midpoint of the array, divided by  $1!, 2!, \dots$ , and  $q!$ , respectively. They showed that these simple convolution arrays could be determined from the general equations for the  $c_q$ 's ( $q \neq 0$ ), and could be simply used to obtain smoothed derivatives in a single convolution operation.

The object of the digital least-squares-fit convolution procedure is thus the determination of a set of  $n$  coefficients,  $\{p_s^{(q)}\}$ , that may be used with a set of uniformly spaced digital data points centered around a data point  $u_i$  to obtain a smoothed value for the  $i$ th point,

$$\bar{u}_i = \sum_{s=-m}^m p_s^{(0)} u_{i+s} \quad (2)$$

or a different set of coefficients,  $\{p_s^{(q)}\}$ , that give the smoothed value of the  $q$ th derivative according to

$$\frac{d^q \bar{u}_i}{dx^q} = \sum_{s=-m}^m p_s^{(q)} u_{i+s} \quad (3)$$

$i = m + 1, \dots, N - m - 1$ .  $N$  = the number of points in the array to be smoothed.

Savitzky and Golay give tables of  $\{p_s^{(0)}\}$  and  $\{p_s^{(q)}\}$  for fitting to polynomials up to degree 6. Errors in those tables may be detected by two simple tests. First, the coefficients must be normalized such that

$$\sum_{s=-m}^m p_s^{(q)} s^q = (q!) \quad (4)$$

$q = 0, 1, 2, 3, 4$ , or  $5$ . Second, if  $q \neq 0$

$$\sum_{s=-m}^m p_s^{(s)} = 0 \quad (5)$$

The corrections given by Steinier et al. (3) have been checked and found to be correct except for a value given in the first row of their Table I. The correct value for the  $\pm 5$  point in column 25 of Savitzky–Golay Table I should be 342 instead of 343 as given by Steinier et al.

In recent model calculations to study the effects of random noise on an iterative deconvolution scheme (7), it was found that the 25-point smoothing arrays of Savitzky and Golay could be too narrow for the "best results" of smoothing where the buildup of the noise, remaining after smoothing, in the deconvolution process, must be balanced against the signal broadening that results from the smoothing. For single least-squares-fit smoothing, Edwards and Willson (8) have found, in a study of the effects of such smoothing on noise-free Gaussian lines, that the optimum width of the smoothing array is 0.7 times the full width at half maximum (FWHM) of the narrowest Gaussian line of their spectra. In a similar study considering Lorentzian- as well as Gaussian-shaped lines, Enke and Nieman (9) conclude that the best signal-to-noise enhancement from a single-pass (quadratic-cubic) smoothing occurs for a smoothing array that is twice as wide as the FWHM of the peak being smoothed. Thus the optimum width of a single-pass smoothing array will depend on the criteria set by the user. As found in our deconvolution studies, however, that optimum width can sometimes be greater than 25 points; and the choice is then between iterative, or repetitive, smoothings and generation of wider, single-pass smoothing arrays. Considering the Slutsky–Yule oscillatory distortions (10) that can build up with multiple smoothing, and the question of aesthetics—single-pass least-squares-fit smoothing involves exact fits of the original data to a known polynomial while multiple smoothing does not—use of a single least-squares-fit smoothing array of appropriate width may arguably be the more desirable method of smoothing data.

General formulas for the smoothing coefficients  $p_s^{(q)}$ , for fitting to polynomials of degree from 2 through 5, have already been published (4, 5, 6a). Similar formulas for calculating the  $p_s^{(q)}$ ,  $q = 1$  thru 5, have been derived and are presented in Table I together with the formulas for  $p_s^{(0)}$  using Roman numerals to label the eleven equations in accordance with the Savitzky–Golay tables of coefficients (1). With these eleven equations one can easily extend all of the tables of the Savitzky–Golay paper (1) for least-squares-fit convolution arrays with no limit to the number of points in an array.

The coefficients given in the Savitzky–Golay tables (1) are all integers. They correspond to the numerators of Equations I through XI in Table I (sometimes divided by a factor that

Table I. Equations for Calculating Savitzky-Golay Smoothing-Array Coefficients,  $p_s^{(a)}$ 

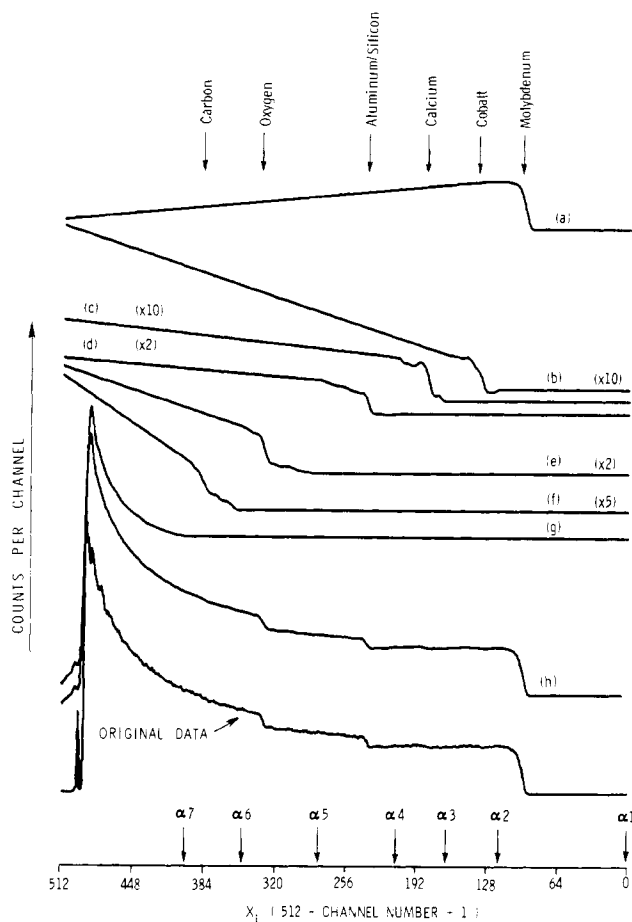
equation no.	order of smoothed derivative $(a)$	order of polynomial $(l)$	equation for $p_s^{(a)}$
I	0	2 or 3	$p_s^{(0)} = \frac{3(3m^2 + 3m - 1 - 5s^2)}{(2m + 3)(2m + 1)(2m - 1)}$
II	0	4 or 5	$p_s^{(0)} = \left(\frac{15}{4}\right) \frac{(15m^4 + 30m^3 - 35m^2 - 50m + 12) - 35(2m^2 + 2m - 3)s^2 + 63s^4}{(2m + 5)(2m + 3)(2m + 1)(2m - 1)(2m - 3)}$
III	1	2	$p_s^{(1)} = \frac{3s}{(2m + 1)(m + 1)(m)}$
IV	1	3 or 4	$p_s^{(1)} = \frac{5\{5(3m^4 + 6m^3 - 3m + 1)s - 7(3m^2 + 3m - 1)s^3\}}{(2m + 3)(2m + 1)(2m - 1)(m + 2)(m + 1)(m)(m - 1)}$
V	1	5 or 6	$p_s^{(1)} = \left(\frac{21}{4}\right) \left\{ 7(25m^8 + 100m^7 - 50m^6 - 500m^5 - 95m^4 + 760m^3 + 180m^2 - 300m + 72)s \right. \\ \left. - 105(6m^6 + 18m^5 - 15m^4 - 60m^3 + 17m^2 + 50m - 12)s^3 \right. \\ \left. + 33(15m^4 + 30m^3 - 35m^2 - 50m + 12)s^5 \right\} \\ \div \left\{ (2m + 5)(2m + 3)(2m + 1)(2m - 1)(2m - 3)(m + 3)(m + 2)(m + 1)(m)(m - 1)(m - 2) \right\}$
VI	2	2 or 3	$p_s^{(2)} = \frac{30\{3s^2 - m(m + 1)\}}{(2m + 3)(2m + 1)(2m - 1)(m + 1)(m)}$
VII	2	4 or 5	$p_s^{(2)} = \left(\frac{-105}{2}\right) \frac{15(6m^2 + 6m - 5)s^4 - 21(4m^4 + 8m^3 - 4m^2 - 8m + 5)s^2 + 5m(2m^5 + 6m^4 - m^3 - 12m^2 - m + 6)}{(2m + 5)(2m + 3)(2m + 1)(2m - 1)(2m - 3)(m + 2)(m + 1)(m)(m - 1)}$
VIII	3	3 or 4	$p_s^{(3)} = \frac{210\{5s^3 - (3m^2 + 3m - 1)s\}}{(2m + 3)(2m + 1)(2m - 1)(m + 2)(m + 1)(m)(m - 1)}$
IX	3	5 or 6	$p_s^{(3)} = \left(\frac{-945}{2}\right) \left\{ 7(6m^6 + 18m^5 - 15m^4 - 60m^3 + 17m^2 - 50m - 12)s \right. \\ \left. - 15(12m^4 + 24m^3 - 28m^2 - 40m + 39)s^3 + 77(2m^2 + 2m - 3)s^5 \right\} \\ \div \left\{ (2m + 5)(2m + 3)(2m + 1)(2m - 1)(2m - 3)(m + 3)(m + 2)(m + 1)(m)(m - 1)(m - 2) \right\}$
X	4	4 or 5	$p_s^{(4)} = \frac{(1890)\{3m(m^3 + 2m^2 - m - 2) - 5(6m^2 + 6m - 5)s^2 + 35s^4\}}{(2m + 5)(2m + 3)(2m + 1)(2m - 1)(2m - 3)(m + 2)(m + 1)(m)(m - 1)}$
XI	5	5 or 6	$p_s^{(5)} = \frac{(20790)\{15m^4 + 30m^3 - 35m^2 - 50m + 12\}s - 35(2m^2 + 2m - 3)s^3 + 63s^5}{(2m + 5)(2m + 3)(2m + 1)(2m - 1)(2m - 3)(m + 3)(m + 2)(m + 1)(m)(m - 1)(m - 2)}$

is common to the denominator of those equations). It should be noted, however, that the  $p_s^{(q)}$  coefficients are all less than unity. Thus, before using the integer coefficients from the Savitzky-Golay tables, those integers must be divided by a normalizing factor, also given as an integer, in order to obtain the correct  $p_s^{(q)}$  that satisfy the normalization condition, Equation 4. The denominators of Equations I through XI give the integer normalizing factors of the Savitzky-Golay tables. In calculating the integer entries in corrected Savitzky-Golay tables (11) with a PDP-10 computer, double-precision techniques were needed. When it was realized that the precision needed in calculating the integers of Savitzky and Golay was much greater than needed for most applications of least-square-fit smoothing, the  $p_s^{(q)}$  were computed directly with less precision leading to considerable simplification in the computer program (11).

There are at least two data-processing procedures in which least-squares-fit smoothing/differentiation techniques are found to be potentially beneficial. In these cases, a buildup of random noise is inherent in the procedure and smoothing is found useful in ameliorating the signal-to-noise ratio in the final result. In both cases the experimental spectra which are smoothed do *not* consist simply of lines of known lineshape (e.g., Voigt) but rather of step- or Heaviside-like signals on a large, slowly-varying background. The object of the processing in the first case, Auger electron spectroscopy on solid surfaces, is to deconvolute the broadening and loss features from the background-corrected signal to obtain the "true" Auger line shape. In the second case, Rutherford backscattering spectroscopy from thick solid targets, it is the height, shape, and location in energy of the threshold signals that are to be determined by the data-processing.

Auger transitions that involve the valence electrons of a solid result in signals whose line shapes contain information on the valence-band density-of-states of the solid. These signals are, in general, distorted in shape because of the inelastic interactions that some of the Auger electrons undergo before exiting the solid; and deconvolution is necessary to extract the undistorted Auger line shape from the measured signal. In the iterative (van Cittert) deconvolution technique that we use, random noise is found to build up rapidly with the number of iterations (7). This buildup can eventually offset any improvement in the shape of the convolution-distorted signal brought about by the deconvolution. Judicious use of smoothing techniques is found to significantly shift the "break-even" number of iterations at which the noise buildup begins to offset the benefits of deconvolution-shape improvement on the signal. Results of model calculations investigating the buildup of noise in deconvolution have been published (7).

The application of nuclear scattering techniques to elemental analysis of solids was first discussed by Rubin, Passell, and Bailey in an article in this journal (12). Since that pioneering work, extensive use has been made of Rutherford backscattering spectroscopy for analysis of the near-surface regions of solids. The threshold energy for backscattering is directly relatable to the mass of the scattering nucleus, and the height of the threshold signal is related to the surface concentration of the scattering element in the sample. Using least-squares-fit smoothing coefficients, Peisach (13) has investigated the use of differentiation to obtain energy-location and step-height information from backscattering data and has found that such techniques of data analysis can lead to "improved definition of spectral energies and improved precision in determining elemental surface concentrations". Potential-modulation techniques that are commonly used in many other spectroscopies to obtain directly measured derivative spectra are not practical with the experimental



**Figure 1.** Spectral decomposition of Rutherford backscattering data. Curves a through f—resolved thresholds for molybdenum, cobalt, calcium, aluminum (and/or silicon), oxygen and carbon, respectively. Curve g—residual signal after subtracting curves a through f from the complete spectrum. Curve h—sum of curves a through g, all with unit scaling. The lowest curve gives the backscattering spectrum before decomposition

systems currently used for backscattering spectroscopy, thus making it necessary to mathematically differentiate the measured spectra using a computer. Unfortunately, for some unexplained reason, Peisach used two different arrays for smoothing and differentiation, thus somewhat blurring the utility of using least-squares-fit arrays to obtain smoothed derivatives in a single convolution.

In a similar study of differentiated backscattering data in our laboratory (11), single convolution arrays were used to obtain smoothed derivatives. Our conclusions as to the advantages of using smoothed derivatives in analyzing Rutherford backscattering data were similar to those reached by Peisach (13). In an extension of these differentiation techniques, it was found possible to easily decompose the backscattering spectrum into separate threshold components. The data used in this study were obtained with singly-charged, 2-MeV helium ions incident on the surface of an alumina-supported, silica-promoted, cobalt-molybdate catalyst (14), and are shown as the lowest curve in Figure 1. Plotted in this curve are the numbers of helium ions backscattered into a fixed detector as a function of the energy (which is proportional to channel number) of the backscattered ions. Because the threshold signals in these data turn on as one goes from high to low energy, the data are plotted here against a new abscissa parameter  $x_i (= 512 - \text{channel number} + 1)$ .

Spectral decomposition of these data was carried out using a variation of a background correction scheme devised by Houston (15). One can calculate a background-corrected

spectrum  $R_2(x_i, \alpha_j)$ , using Houston's notation, where the subscript 2 indicates that the second derivative of the original data is first taken, and then this second derivative is integrated twice to recover the original spectrum minus the constant and linear components lost during differentiation. Using a least-squares-fit smoothed-derivative array, the smoothing operation occurs automatically in the process of determining the second derivative.  $x_i$  is the independent variable used in the integration, and  $\alpha_j$  is the lower limit of integration. Then calculating  $R_2(x_i, \alpha_j)$  and  $R_2(x_i, \alpha_{j+1})$ , where  $\alpha_j$  and  $\alpha_{j+1}$  are two  $x_i$  values in the spectrum below and above the threshold feature to be extracted, respectively, the extracted threshold signal is given by  $R_2(x_i, \alpha_j) - R_2(x_i, \alpha_{j+1})$ . The locations of the threshold signals are indicated by arrows at the top of Figure 1. The values of  $x_i$  that were chosen for the  $\alpha_j$ 's in the decomposition of the original spectrum were  $\alpha_1 = 1$ ;  $\alpha_2 = 118$ ;  $\alpha_3 = 165$ ;  $\alpha_4 = 210$ ;  $\alpha_5 = 280$ ;  $\alpha_6 = 349$ ; and  $\alpha_7 = 400$ . These  $\alpha_j$ 's were chosen to lie in regions of the spectrum assumed to contain only background, and no threshold-signal, information. There were 512 data points in the spectrum and a 19-point convolution array (Equation VII of Table I) was used to determine the second derivative.

The seven spectral components resulting from the decomposition are plotted as curves (a) through (g) in Figure 1. In the original data, three thresholds are clearly discernible while the other three are less so. In the component spectra (curves a through g), all six thresholds are clearly seen. As this example calculation was performed for illustrative purposes only, linear extrapolation of the signals below the threshold energy (in the direction of increasing  $x_i$  values) was made instead of using the more correct reciprocal-energy dependence (14). Note that the decrease in counts-per-channel below threshold for the molybdenum signal indicates an enhancement of the molybdenum concentration at the solid surface. This result was corroborated by backscattering data taken with 1.4-MeV protons incident on the sample (14). Curve h gives the sum of curves a through g, all with unit scaling, and is seen to be essentially a smoothed version of the original data. Thus, the differentiation techniques demonstrated by Peisach (13) as useful in analyzing backscattering spectra can be simply extended to give component threshold signals. Least-squares-fit smoothing is beneficial in this operation in ameliorating the noise buildup in the differentiation. It provides a simple one-operation method

of obtaining the smoothed derivative.

The Savitzky-Golay convolution method of least-squares-fit smoothing and differentiation of digital data is versatile and extremely simple to use. Using only the coefficients given in the Savitzky-Golay tables (1), one is limited to a 25-point smooth, however. Wider smoothing arrays are sometimes desirable and may be simply generated using the eleven equations given in Table I of this letter. Relaxing the condition that integer numerator and denominator values for the  $p_s^{(q)}$  be calculated separately leads to further simplification in the computer program. Smoothing has been shown in data processing involving deconvolution (7) and/or differentiation (11, 12) to be beneficial. Peak position, height, and width—of lines of known profile—are not the only information that may be sought in some spectroscopies. The smoothing techniques of Savitzky and Golay offer an extremely simple aid in extracting additional lineshape information.

### LITERATURE CITED

- (1) A. Savitzky and M. J. E. Golay, *Anal. Chem.*, **36**, 1627 (1964).
- (2) "Citations Index" (1965-1976).
- (3) J. Steinier, Y. Termonia, and J. Deltour, *Anal. Chem.*, **44**, 1906 (1972).
- (4) R. R. Nernst, *Adv. Magn. Reson.*, **2**, 39 (1966).
- (5) P. D. Willson and T. H. Edwards, *Appl. Spectrosc. Rev.*, **12**, 1 (1976).
- (6) (a) E. Whittaker and G. Robinson, "The Calculations of Observations", 4th ed., Blackie and Son, Ltd., London, 1944, pp 291-296. (b) P. G. Guest, "Numerical Methods of Curve Fitting", Cambridge University Press, 1961, pp 349-353. (c) M. G. Kendall and A. Stuart, "The Advanced Theory of Statistics", Hafner Publishing, Inc., New York, N.Y., 1966, Vol. 3, "Design and Analysis, and Time-Series", pp 366-375.
- (7) H. H. Madden and J. E. Houston, *Adv. X-Ray Anal.*, **19**, 657 (1976); *J. Appl. Phys.*, **47**, 3071 (1976).
- (8) T. H. Edwards and P. D. Willson, *Appl. Spectrosc.*, **28**, 541 (1974).
- (9) C. G. Enke and T. A. Nieman, *Anal. Chem.*, **48**, 705A (1976).
- (10) M. G. Kendall and A. Stuart, ref 6c, pp 375-379.
- (11) H. H. Madden and D. G. Schreiner, "Convolution Method for Least-Squares-Fit Smoothing and Differentiation of Digital Data", Report No. SAND76-0283, Sandia Labs., Albuquerque, N.M., 1976, unpublished work.
- (12) S. Rubin, T. O. Passell, and L. E. Bailey, *Anal. Chem.*, **29**, 736 (1957).
- (13) M. Peisach, *Thin Solid Films*, **19**, 297 (1973).
- (14) J. A. Borders and G. D. Peterson, "Ion Backscattering Studies of Synthol Catalyst Materials", Report No. Sand77-0479, Sandia Labs., Albuquerque, N.M., 1977, unpublished work.
- (15) J. E. Houston, *Rev. Sci. Instrum.*, **45**, 897 (1974).

Hannibal H. Madden

Sandia Laboratories,  
Albuquerque, New Mexico 87185

RECEIVED for review April 18, 1977. Accepted May 22, 1978.  
This work was supported by the U.S. Department of Energy under Contract AT(29-1)789.

## Extension of the Rasberry-Heinrich Equation for X-ray Fluorescence Analysis

Sir: For an "infinitely" thick sample, the x-ray fluorescence intensity is shown to be given by the following expression:

$$I_j^Q = K_j C_j \int_{\lambda_{\min}}^{\lambda_{Kab}} \frac{I_0(\lambda) \mu_j(\lambda)}{\mu_Q(\lambda) + \mu_Q(\lambda_j)} d\lambda \quad (1)$$

where  $K_j$  is a factor to take care of the geometrical arrangement of the sample-detector and of fundamental parameters of element  $j$ .  $C_j$  is weight fraction of element  $j$ .  $\mu_Q(\lambda)$  is mass absorption coefficient of the sample (denoted by  $Q$ ) at the incident wavelength,  $\lambda$ .  $\mu_Q(\lambda_j)$  is mass absorption coefficient of the sample for the primary fluorescence radiation

of element  $j$ .  $\mu_j(\lambda)$  is mass absorption coefficient of element  $j$  for the incident wavelength,  $\lambda$ .  $\lambda_{\min}$  is minimum wavelength of the generator tube as given by the excitation voltage.  $\lambda_{Kab}$  is wavelength of the K-absorption edge of element  $j$ .

Equation 1 does not include higher order fluorescence effects (secondary, tertiary, etc.) that occur when the sample contains more than one chemical element. Furthermore, we consider monochromatic incident radiation in order to make the physical processes therein easier to understand. This would lead to a better understanding of what happens when the incident spectra is polychromatic, as it is in most of the practical cases.