Computation of Two-Dimensional Polynomial Least-Squares Convolution Smoothing Integers

Kenneth L. Ratzlaff*

Instrumentation Design Laboratory, Department of Chemistry, The University of Kansas, Lawrence, Kansas 66045-0046

Jean T. Johnson

Department of Mathematics, Baker University, Baldwin City, Kansas 66006

The technique of smoothing by polynomial least-squares convolution (PLSC) techniques, commonly known to chemists as Savitzky–Golay smoothing, is arguably the most common digital data processing technique used by analytical chemists with the exception of the fit of a straight line to data to generate a calibration curve (*1*). This paper presents tools enabling a general extension of the method for smoothing two-dimensional data.

The PLSC smoothing technique, simple but effective in both concept and operation (*2*), can be used to remove noise that is high in frequency relative to the analytical information when data are acquired on an evenly spaced, linear, abscissa axis. The PLSC technique was first pointed out to the chemical community and smoothing coefficients were compiled in Savitzky and Golay's classic paper (*3*). Later papers made corrections to these tables (*1, 4*), and as digital data acquisition and signal processing became common, these integers saw frequent application.

However, analytical techniques increasingly operate in two dimensions: Excitation–emission fluorescence matrices (*5*), retention time/absorbance surfaces in liquid chromatography (*6*), and spatial/spectral maps of emission from atomic spectroscopy sources (*7*) are but three examples. In these cases the data will often benefit from smoothing in both dimensions. Consequently, when smoothing of two-dimensional data is necessary, techniques operating in two dimensions become imperative.

Together with an introduction to smoothing in two dimensions, a limited set of two-dimensional PLSC integers was published by Edwards (*8*). To compute the integers from the least-squares normal equations, the computer precision required for the computation rapidly overflowed the capability of even a CDC 6600 computer, and the tables were limited to smoothing areas up to a maximum size of 7 by 7. Because the objective was smoothing of photographic images whose bandwidth was the same in both dimensions, only symmetrical sets of integers were computed. Later, a somewhat more extensive but still limited set of two-dimensional PLSC integers was tabulated, computed by other techniques on a Macintosh computer (*2*).

When applying the original integer tables for linear smoothing, Madden (*1*) found them insufficient for his needs, limited to smoothing widths up to 25. Additionally, these tables are cumbersome and error-prone. Consequently, algebraic expressions were contributed by which these integers can be computed "on the fly". Although these formulas appear complex, evaluation on small computers can be easier than manipulating limited tables of data. Furthermore, there is no inherent limit to the smoothing area that can be employed.

Unfortunately, no similar set of formulas exists for two-dimensional smoothing. In this case the need is even more acute since there is no inherent reason why the smoothing width should be the same in both dimensions. Indeed, as will be shown later, the widths should typically be chosen independently.

Table I. Formulas for Use in Two-Dimensional PLSC Smoothing

Equation I. Quadratic/Cubic by Quadratic/Cubic

numerator:

$N(s,t) = -15(Qt^2 + Rs^2) + 9(3m^2 + 3m - 1)(3n^2 + 3n - 1) - 25m(m + 1)n(n + 1)$

denominator:

$D = (2m + 1)R(2n + 1)Q$

where:

$R = (2n - 1)(2n + 3)$

$Q = (2m - 1)(2m + 3)$

Equation II. Quartic/Quintic (n, t) by Quadratic/Cubic (m, s)

numerator:

$N(s,t) = -60Rs^2 + 945Qt^4 - 525Q(2n^2 + 2n - 3)t^2 + 20(61n^4 + 122n^3 - 161n^2 - 222n + 81)m^2 + 20(61n^4 + 122n^3 - 161n^2 - 222n + 81)m - 45(15n^4 + 30n^3 - 35n^2 - 50n + 12)$

denominator:

$D = 4QR(2m + 1)$

where:

$R = (2n - 3)(2n - 1)(2n + 3)(2n + 5)$

$Q = (2m - 1)(2m + 3)$

Equation III. Quartic/Quintic by Quartic/Quintic

numerator:

$N(s,t) = 945Rs^4 - 525R(2m^2 + 2m - 3)s^2 + 945Qt^4 - 525Q(2n^2 + 2n - 3)t^2 + (6176m^4 + 12352m^3 - 17416m^2 - 23592m + 10125)n^4 + 2(6176m^4 + 12352m^3 - 17416m^2 - 23592m + 10125)n^3 - 7(2488m^4 + 4976m^3 - 6608m^2 - 9096m + 3375)n^2 - 6(3932m^4 + 7864m^3 - 10612m^2 - 14544m + 5625)n + 675(15m^4 + 30m^3 - 35m^2 - 50m + 12)$

denominator:

$D = 4QR$

where:

$R = (2n - 3)(2n - 1)(2n + 3)(2n + 5)$

$Q = (2m - 3)(2m - 1)(2m + 3)(2m + 5)$

FORMULAS FOR TWO-DIMENSIONAL PLSC

Three formulas have been developed for use in two-dimensional PLSC smoothing and are presented in Table I. Equation I is derived from a model that is a quadratic/cubic polynomial in both dimensions:

$$y = b_0 + \sum_{k=1}^{3} a_k x^k + \sum_{k=1}^{3} c_k z^k$$

When the usual procedures for linear regression are followed (*9*), an equation for the sum of the squares of the residuals (SS) is written

$$\mathrm{SS} = \sum_{t=-n}^{n} \sum_{s=-m}^{m} (y - b_0 - \sum_{k=1}^{3} a_k x^k - \sum_{k=1}^{3} c_k z^k)^2$$

The normal equations are obtained by taking the derivatives of that equation and setting them to zero. The formulas for the integers are found by solving those equations algebraically. In one dimension, m defines the absolute value of the limit of the smooth, so that the smoothing width is $2m + 1$; s is the subscript of the smoothing integer in that dimension to be

* Author to whom correspondence should be addressed.

0003-2700/89/0361-1303$01.50/0

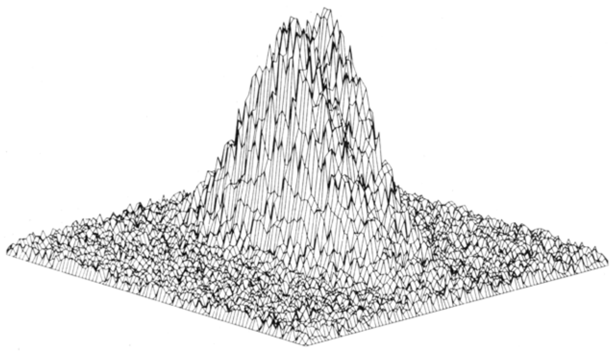

Figure 1. Model data surface. The parameters used to generate this surface are described in the text.

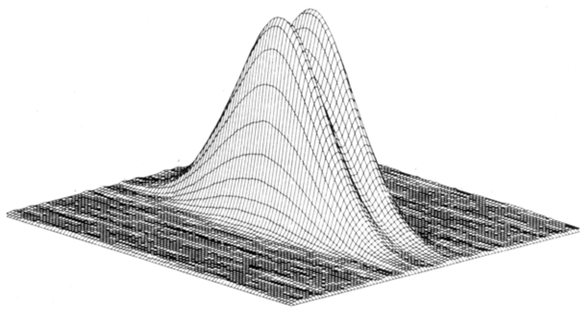

Figure 2. Model data surface after two-dimensional PLSC smoothing. The surface of Figure 1 was smoothed with a width of 5 in one dimension and 25 in the other. A quadratic/cubic model was used in both dimensions.

evaluated from $-m$ to m. Similarly in the other dimension, n and t define the width and subscript, respectively.

Equation II is derived from normal equations for a quadratic/cubic polynomial in one dimension and a quartic/quintic polynomial in the second:

$$y = b_0 + \sum_{k=1}^{3} a_k x^k + \sum_{k=1}^{5} c_k z^k$$

Equation III is derived from normal equations for quartic/quintic polynomials in both dimensions:

$$y = b_0 + \sum_{k=1}^{5} a_k x^k + \sum_{k=1}^{5} c_k z^k$$

A smoothed value can then be determined as

$$b_0(i,j) = (\sum_{t=-n}^{n} \sum_{s=-m}^{m} N(s,t)y(i+s,j+t))/D$$

where $N(s,t)$ and D are found in Table I.

Although eq 1 was tediously derived manually, the similar derivations of eq II and III would probably not have been possible within the careers of the authors. However, the algebra involved in the derivations was possible with MACSYMA (Symbolics, Cambridge, MA) software for algebraic manipulation operating on a VAX 8650 computer. The results were verified in two ways. First, when the smoothing width in one dimension is set to zero, the formulas that result are identical with those in ref 1. Second, when the smoothing operation is carried out on a surface generated from polynomials of the same order as the smooth operation, the residuals are zero.

PERFORMANCE ON TWO-DIMENSIONAL SURFACES

A synthetic surface, modeled as an excitation–emission fluorescence matrix, was used to demonstrate some considerations in applying the two-dimensional smoothing technique. Based on the noise profile observed in molecular fluorescence, a surface was derived consisting of a pair of closely spaced Gaussian-shaped peaks (Figure 1). The surface models acquisition of data when the background noise corresponds to about 60 e$^-$ and the peak maxima correspond to 36 000 e$^-$, the resultant noise is assumed to follow a Poisson distribution. Both peaks have a width (σ in the Gaussian expression) of 20 points in one dimension (denoted X) and 5 points in the other (denoted Z); the peak centers are separated in the Z dimension by 10 points.

The surface, after smoothing, is depicted in Figure 2. In this example, a quadratic/cubic smooth was applied in both

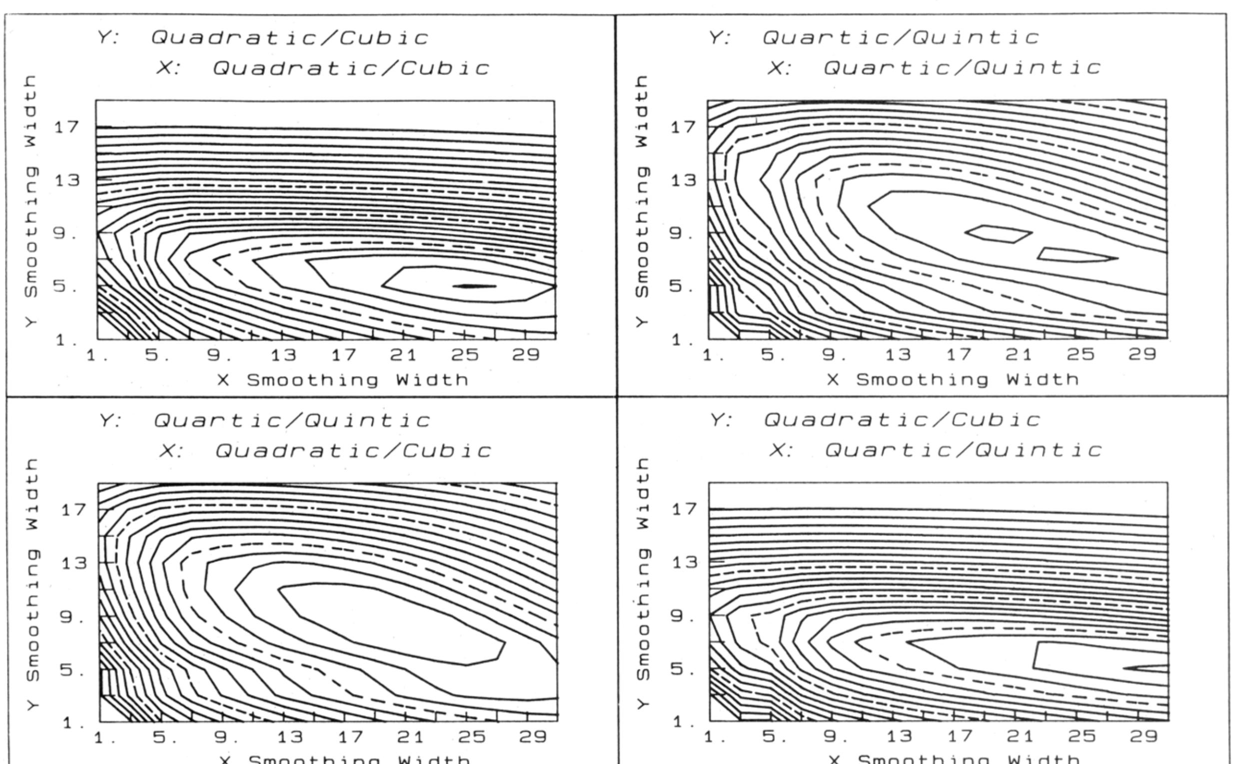

Figure 3. Contour plots depicting the root mean error after two-dimensional smoothing of the data of Figure 1. Identical line patterns used in the four plots represent the same contour levels, allowing quantitative comparisons. The contour lines are spaced logarithmically.

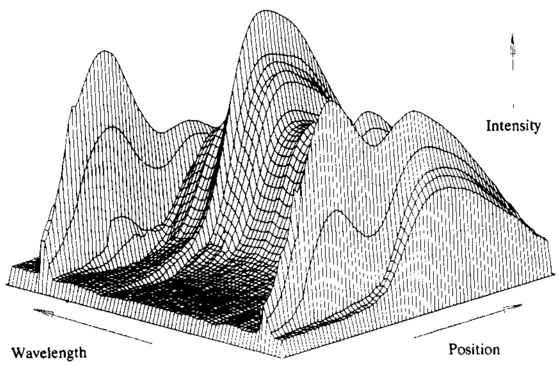

Figure 4. Smoothed response surface for Mg in a dc plasma. One dimension represents wavelength, and the other dimension represents vertical height in the center of the dc plasma.

dimensions, 25 points wide in the X dimension and 5 points wide in the Z dimension.

When designing a two-dimensional smoothing procedure, one must choose a smoothing width for each dimension. There have been a number of discussions of how to optimize the smoothing width in linear applications (*10, 11*). The choice of widths involves the trade-off of two factors: First, any smoothing procedure will generate some distortion of the peak, which increases with smoothing width; however, the noise is increasingly attenuated as the width increases. The position of the intersection of these two factors depends on peak shape and noise level in addition to the order of the smoothing integers.

For each of the smoothing equations, the surface was smoothed for each combination of smoothing widths from 1 to 31 in the X dimension and from 1 to 19 in the Z dimension. The mean square error was computed for each case, summing the squares of the difference between the smoothed surface and the noise-free surface; the noise was accumulated only over the portion of the surface in which the magnitude of the noise-free surface exceeds 0.1% of the peak maximum.

The smoothing results are shown as a series of contour plots in Figure 3. While this demonstration does not purport to provide a rule for optimum smoothing, it does show that a nonsymmetrical smoothing procedure is required. The most obvious and significant point is that the minima occur at parameters that are far from symmetric.

As a final "real-world" example, the smoothing operation was applied to a surface acquired with a two-dimensional charge-coupled device array. The image, shown in Figure 4, results from the excitation of Mg in a dc plasma, with spatial resolution in one dimension and spectral resolution in the other. In the spectral dimension the emission peaks are very narrow while in the spatial dimension, there exist broad regions of both atomic and continuum emission. Consequently, the smoothing width is small in the wavelength dimension and large in the spatial dimension. The resultant surface clearly and rapidly conveys an optimal position for the measurement of Mg emission.

CONCLUSIONS

Two-dimensional smoothing integers for PLSC are readily computed from the formulas provided. These formulas make possible the selection of any width in either dimension according to the requirements of the data set. With these tools, smoothing procedures for surfaces can be optimized separately in each dimension.

ACKNOWLEDGMENT

We express our appreciation to Rafi Jalkian, Jeff Kolczynski, Skip Pomeroy, and Bonner Denton of the University of Arizona for providing images acquired by charge-coupled device imaging technology, including the data on which Figure 4 is based.

LITERATURE CITED

(1) Madden, H. H. *Anal. Chem.* **1978**, *50*, 1383.
(2) Ratzlaff, K. L. *Introduction to Computer-Assisted Experimentation*; Wiley-Interscience: New York, 1987.
(3) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627.
(4) Steinier, J.; Termonia, Y.; Deltour, J. *Anal. Chem.* **1972**, *44*, 1906.
(5) Rossi, T. M.; Warner, I. M. *Appl. Spectrosc.* **1984**, *38*, 422.
(6) Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. *Chemometrics: A Textbook*; Elsevier: New York, 1988.
(7) Kolczynski, J. D.; Pomeroy, R. S.; Jalkian, R. D.; Denton, M. B., to be submitted for publication.
(8) Edwards, T. R. *Anal. Chem.* **1982**, *54*, 1519.
(9) Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: New York, 1969.
(10) Enke, C. G.; Nieman, T. A. *Anal. Chem.* **1976**, *48*, 705A.
(11) Edwards, T. H.; Willson, P. D. *Appl. Spectrosc.* **1974**, *28*, 541.

RECEIVED for review December 20, 1988. Accepted March 1, 1989.

Experimental Determination of the Coefficient in the Steady State Current Equation for Spherical Segment Microelectrodes

Zbigniew Stojek

Department of Chemistry, Warsaw University, ul. Pasteura 1, 02-093 Warsaw, Poland

Janet Osteryoung*

Department of Chemistry, SUNY University of Buffalo, Buffalo, New York 14214

A variety of shapes of microelectrodes (i.e., electrode with at least one dimension in the micrometer range) have been examined and characterized as reported in the literature so far. This includes disks, rings, cylinders, lines, and hemispheres (*1*). Usually microelectrodes are made from metals or graphitic materials. However, mercury microelectrodes are also needed. Several such microelectrodes have been prepared and applied (*2–13*).

Application of mercury microelectrodes for voltammetric purposes requires that the mercury surface is coherent and does not consist of isolated droplets. It has been shown that hemispherical mercury electrodes which work well can be prepared by electrodeposition of the predetermined amount of mercury on a platinum microdisk in a separate plating step (*4, 8, 10, 13*). Iridium has also been proposed as a substrate (*9*). However, iridium wires thinner than 127 μm are not available commercially. Recently a dropping mercury microelectrode has been constructed. The operation of the device is based on heating a mercury reservoir which drives the droplets through a capillary (*14*).

In many of these cases the geometry is not well-defined. For example, Wehmeyer and Wightman (*4*) assume that

