

# Near-Optimal Smoothing Using a Maximum Entropy Criterion

Robert J. Larivee<sup>†</sup> and Steven D. Brown\*

Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716

Digital filters have become very common for smoothing and filtering of noisy chemical data. Many papers have appeared on the use of filters for optimizing signal-to-noise or for optimizing signal shape retention during the process of noise removal. In contrast, little work has been aimed at optimizing the filter to retain as much of the information present in the data while removing as much of the noise as possible. This paper reports a simple approach for optimizing digital filters of several types. The optimization criterion used here is based on the informational entropy of the smoothed signal. It is demonstrated here that, for many forms of data, a filter optimized by maximizing the informational entropy, but without knowledge of the true signal shape, performs as well as an equivalent, optimal filter optimized with the aid of prior knowledge of the signal shape. Filters optimized by maximizing the informational entropy have applications where accurate reconstruction must be accompanied by significant increases in signal-to-noise ratios of noisy, but unknown data. The maximum entropy optimization process is demonstrated on several different digital filters.

## INTRODUCTION

Digital filters have become common for smoothing and filtering of noisy data.<sup>1-3</sup> When a digital filter is applied to a noisy signal during data collection, real-time noise reduction, or filtering, results. When a digital "filter" is applied after data collection, smoothing of the noisy data is achieved. In either case, a well-designed digital filter or smoother acts as an ideal notch filter, allowing only the signal frequencies to pass, while attenuating all other frequencies in a predefined manner.<sup>4,5</sup> Because digital filtering and smoothing is usually done in software rather than in hard-wired circuits, ad hoc filtering schemes and modifications which selectively attenuate various frequencies are easily accomplished. This flexibility, along with the ease of implementation of these filters, has made digital filtering popular in analytical chemistry, as indicated by the increase of articles on this subject.<sup>6-10</sup>

There are a number of problems associated with digital filtering, however. The greatest difficulty with the use of digital filtering and smoothing methods lies in setting those adjustable parameters necessary to accomplish optimal noise reduction for a given data set. Often, these parameters are

associated with the convolution function which determines the frequencies that are attenuated. The range of frequencies passed by an optimal smoother is not fixed; it must be selected to match, to the extent possible, the noisy signal to be smoothed. Optimization of the set of adjustable parameters controlling the smoother passband is not easily accomplished, however, especially when the nature of the signal is unknown. If the wrong values are used, it is possible to distort the data inadvertently in attempts to reduce the noise content.<sup>1</sup> Previous work has considered this dilemma in terms of "hard" smoothing, where noise is significantly attenuated, but at the cost of signal shape, and "soft" smoothing, where noise is removed only to the extent that all of the possible details of the signal are preserved.<sup>11-13</sup> There is very little literature in chemistry describing which of the many smoother types is best suited for removal of noise superimposed on a particular signal shape, or on the optimization of a particular filter type. Comparisons of different filtering or smoothing methods of the same noisy data are also rare. Voightman and Winefordner compared several common analog filters—ones that can be realized in digital form as well—for use on spectroscopic data,<sup>14</sup> but they did not try to optimize the filtering process.

In this paper, an approach based on the Shannon information entropy is presented that addresses some of the problems associated with selecting the "best" parameters for digital filtering and smoothing. By maximization of the information entropy of a noisy response, it is possible to obtain a set of adjustable parameters necessary to ensure minimal loss of signal information in a smoothing step, even when the signal shape that is sought from the smoothing is unknown. When signals in Gaussian noise are examined, the smoothing obtained by application of this method is very close to that obtained with least-squares fitting of the noisy signal, using prior knowledge of the signal shape. The entropy-based optimization method is general, and it is demonstrated here in connection with three smoother algorithms.

## THEORY

**General.** The notation used in this paper is as follows: vectors are represented by bold capitalized letters, **R**, the individual scalar components of the vector are shown as non-capitalized letters with a subscript,  $r_i$ , and scalars not associated with a vector are distinguished as letters with no subscript,  $n$ . Functions are denoted by parentheses around the independent variable,  $f(t)$ , and the Fourier transform of these functions are written with a capitalized letter,  $F(\Omega)$ .

All real data obtained by physical measurements can be broken down into two components: the noise component and the signal component<sup>4,5</sup>

$$\mathbf{X} = \mathbf{S} + \mathbf{R} \quad (1)$$

where **X** is the raw data vector, **S** the signal vector, and **R** the noise vector. For the work reported here, the noise vector

\* Corresponding author.

<sup>†</sup> Present address: Department of Chemistry, Frostburg State University, Frostburg, MD 21532.

- (1) Willson, P. D.; Edwards, T. H. *Appl. Spec. Rev.* 1976, 12, 1.
- (2) Kaiser, J. F.; Reed, W. A. *Rev. Sci. Instrum.* 1977, 48, 1447.
- (3) Horlick, G. *Anal. Chem.* 1972, 44, 943.
- (4) Bialkowski, S. E. *Anal. Chem.* 1988, 60, 355A.
- (5) Bialkowski, S. E. *Anal. Chem.* 1988, 60, 403A.
- (6) Aubanel, E. E.; Myland, J. C.; Oldham, K. B.; Zoski, C. G. *J. Electroanal. Chem. Interfacial Electrochem.* 1985, 184, 239.
- (7) Jagannathan, S.; Patel, R. C. *Anal. Chem.* 1986, 58, 421.
- (8) Betty, K. R.; Horlick, G. *Appl. Spectrosc.* 1976, 30, 23.
- (9) Hayes, J. W.; Glover, D. W.; Smith, D. E.; Overton, M. W. *Anal. Chem.* 1973, 45, 277.
- (10) Myland, J. C.; Oldham, K. B.; Guoyi, Z. *Anal. Chem.* 1988, 60, 1610.

(11) Bromba, M. U. A.; Ziegler, H. *Anal. Chem.* 1983, 55, 1299.

(12) Bromba, M. U. A.; Ziegler, H. *Anal. Chem.* 1979, 51, 1760.

(13) Bromba, M. U. A.; Ziegler, H. *Anal. Chem.* 1984, 56, 2052.

(14) Voightman, E. G.; Winefordner, J. D. *Rev. Sci. Instrum.* 1986, 57, 957.

can be considered to be a sample of a random Gaussian distribution with a mean of zero. The assumption that the noise and signal components are additive and uncorrelated is reasonable for most chemical applications and is used in this paper.

Digital filtering and smoothing techniques can be described for a time series  $X$  as a convolution of the  $b$  discrete, time-domain data  $x(\tau)$  with a weighting function  $h(t - \tau)$  such that

$$\hat{s}(t) = \sum_{\tau=0}^b x(\tau)h(t - \tau) \quad (2)$$

where  $\hat{s}(t)$  is the smoothed signal, an estimate of the true signal  $s(t)$ ,<sup>15-17</sup> and where  $\tau$  is the lag associated with the convolution operation. Real data  $x(t)$  may be taken to be a discrete time series in the sampled independent variable  $t$

$$t = t_0 + n\Delta \quad (3)$$

where  $n$  is the cardinal number of the measurement,  $\Delta$  the interval between recorded samples, and where  $t_0$  is the initial value of the time series defined by eq 3. This time series can be set up in terms of time or any other suitable independent variable, so long as it meets the requirements above.

Equation 2 can also be expressed as

$$\hat{s}(t) = x(t) * h(t) \quad (4)$$

where  $*$  indicates the convolution operation. Equation 4 can be brought from the time domain into the frequency domain, using the Fourier transform to obtain

$$\hat{S}(\Omega) = X(\Omega)H(\Omega) \quad (5)$$

where the weighting function  $H(\Omega)$  can be represented as its Fourier transform

$$H(\Omega) = \sum_{n=0}^b h_n e^{in\Delta\tau} \quad (6)$$

where  $b$  is the number of points in the system being examined. The weighting function in eq 6 describes the weighting of each point  $x_i(\Omega)$  in the Fourier domain representation of  $\hat{S}$ . Ideally, those frequencies which represent pure noise will receive a zero weight value, and those which represent only signal will receive a weight value of one. Determination of this weighting function is the critical step in filtering and smoothing.

The relationship between eqs 4, 5, and 6 is important in understanding how filtering and smoothing routines operate. Generally the systematic changes attributed to the signal are confined to a few low frequencies, while many common forms of noise cover a relatively wide range of frequencies. If this relationship of noise and signal holds, it is easily seen that the number of frequencies required to describe a simple, broadband signal to an acceptable degree of accuracy are also relatively few, usually consisting of frequency components from the lower frequency region. The number of frequencies needed to adequately define the signal by the Fourier series defined in eqs 4-6 is determined by the Nyquist sampling theorem.<sup>15</sup> Note that this determination of the frequencies necessary to describe a signal adequately can be considered equivalent to the sampling of the signal in that either undersampling or under representation (with fewer frequencies represented than those required by the Nyquist theorem) leads to aliasing and distortion of the true signal shape.

If the signal shape is known and the noise is zero-mean and white (or can be made white by application of a whitening

filter), determining the best weighting function is straightforward. The optimal weighting function—in the least-squares sense—for a known signal is the matched smoother, where, for real data, the weighting function is identical to the frequency-domain representation of the signal.<sup>4,15,18</sup> When the signal shape is unknown, however, a matched smoother cannot be used. In these cases, determining the best weighting function for smoothing or filtering is far from straightforward, and nearly-optimal smoothers are attractive, if they can be found. Any smoother other than the matched smoother, even when properly sized, cannot provide better signal-to-noise ratio enhancement in the least-squares sense, however.

**Polynomial Smoothing.** Probably the most familiar digital smoother used when the signal shape is not known prior to filtering or smoothing is the well-known polynomial smoother first described in the chemical literature by Savitzky and Golay.<sup>19-22</sup> In this method, a portion of the signal containing  $2N + 1$  points is fitted to a Gram polynomial of the  $k$ th order,  $P_k^N(n)$ , where

$$P_k^N(n) = \sum_{j=0}^k \left[ (-1)^{(k+j)} \frac{(k+j)^{(2j)} (N+n)^j}{(j!)^2 (2N)^{(j)}} \right] \quad (7)$$

defines the terms in the polynomial and where the weighting coefficient for each term in the polynomial is given by

$$c_k^N = \frac{1}{2k+1} + \frac{(2N+k+1)^{(k+1)}}{(2N)^{(k)}} \quad (8)$$

The  $(x)^{(j)}$  notation used in these equations is defined as

$$(x)^{(j)} = x(x-1)(x-2)\dots(x-j+1) \quad (9)$$

These polynomials form a set of orthogonal functions on the finite number of points in the  $2N + 1$  point "window" of data under examination. Generally, the center point of the window is then set to the estimated value obtained from the polynomial model obtained from eq 7. Other points in the window can be used, but there are some compelling reasons to avoid them.<sup>23</sup> The window is moved to the next point and the process is repeated. This procedure is repeated until all points in the data vector are smoothed. In effect, this method approximates a complex signal shape with a simpler, polynomial function over the points included in the smoother window. The width of the window used for the local approximation and the order of the polynomial allow some latitude in matching the frequency characteristics of the smoother and the signal. In eq 6, the  $n$ th element of the weight for polynomial smoothing,  $h_n$ , is

$$h_n = \sum_{k=0}^m (1/c_k^N) P_k^N(0) P_k^N(n) \quad (10)$$

where  $m$  is the degree of the polynomial being fitted to the noisy data. The weight defined by eq 10 can be modified by changing either the window size or the degree of the polynomial.

To determine how window size and number of terms in the series defined by eq 10 affect the different frequency components present in the noisy data, it is most convenient to consider the smoother transfer function. By changing the variable in eq 6 such that  $\theta$  is equal to  $\pi\omega/\omega_s$ , with  $\omega_s$  the point where Nyquist folding occurs, a plot of  $H(\theta)$  versus  $\theta$  can be produced. This plot describes the transfer function for the

(18) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* 1964, 36, 1627; 1972, 44, 1906.

(19) Madden, H. H. *Anal. Chem.* 1978, 50, 1383.

(20) Enke, C. G.; Nieman, T. A. *Anal. Chem.* 1976, 48, 705A.

(21) Willson, P. D.; Polo, S. R. *J. Opt. Soc. Am.* 1981, 71, 599.

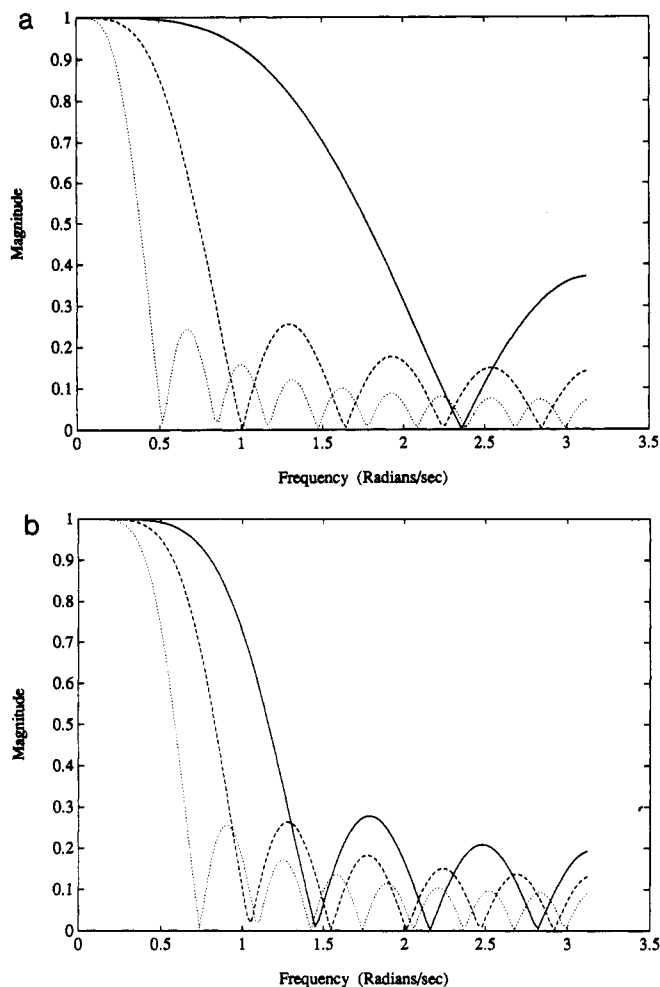
(22) Ernst, R. R. *Rev. Sci. Instrum.* 1965, 36, 1689.

(23) Dougherty, T. P.; Wentzell, P. D.; Crouch, S. R. *Anal. Chem.* 1987, 59, 367.

(15) Brigham, O. E. *The Fast Fourier Transform*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1974.

(16) Bracewell, R. N. *The Fourier Transform and Its Applications*; McGraw-Hill: New York, NY, 1965.

(17) Brereton, R. G. *Chemom. Intell. Lab. Sys.* 1986, 1, 17.



**Figure 1.** (a) The transfer functions for a third-order polynomial smoother with window sizes of 5 (—), 11 (---), and 21 (···) points. (b) The transfer functions for a fifth-order polynomial with windows of 11 (—), 15 (---), and 21 (···) points.

smoothing operation. Transfer functions are related to weighting functions in that they both describe how frequency components of a noisy data sequence are attenuated. They are generally defined in units of radians  $s^{-1}$ . These transfer functions have been evaluated for the various different parameters used in polynomial smoothing; a detailed accounting of them has been given by Wilson and Edwards.<sup>1</sup>

Figure 1 presents some examples of these transfer functions. By changing the window size and the number of terms used in the polynomial equation, the transfer function can be modified to determine which frequencies will be allowed through the digital filter or smoother. As the number of points in the window increases and the number of terms used in the polynomial equation decreases, the higher frequencies are attenuated more strongly. As with the rectangular weighting function, if the attenuation is too severe, and frequencies below the Nyquist limit are attenuated, the higher frequency components of the signal are removed, and the signal becomes distorted through aliasing. If the frequencies above the Nyquist limit are not attenuated properly, noise in these frequencies is not removed. Bromba and Ziegler<sup>11-13</sup> have discussed how various modifications to the transfer function can affect the resulting signal passed by a polynomial filter. They have shown that the best results occur for polynomial filtering when the true signal shape is known, or can be guessed, so that the order and window size of the polynomial filter can be set optimally. They offered no method of determining these parameters for an unknown signal, however.

**Adaptive Smoothing.** The second smoother algorithm of interest here is the adaptive smoothing method reported

by Kawata and Minami.<sup>25</sup> This method is similar to the first-order polynomial (running average) smoother. The average of a number of adjacent points in a window is calculated, and the center point of the window receives this average value as an estimate of the true signal value. The window is shifted by one point and the process is repeated. The term  $h_n$  in eq 6 for a running average is

$$h_n = \frac{1}{(2N + 1)} \quad (11)$$

which has a transfer function

$$H(\theta) = \frac{1}{(2N + 1)} \frac{\sin((2N + 1)\theta)}{\sin \theta} \quad (12)$$

The mean and variance of the data within the smoother window are calculated. This "local variance" estimate,  $\sigma_x^2$ , is then compared to the expected variance of the noise over the entire data set

$$a(\nu) = \frac{\sigma_x^2(\nu) - \sigma_n^2}{\sigma_n^2(\nu)} \quad (13)$$

where  $\sigma_n^2$  is the overall (global) variance of the noise in the data. This calculated ratio  $a(\nu)$  is then used as to determine how much smoothing the center point in the rectangular window receives

$$\hat{s}(\nu) = [a(\nu)(x(\nu) - \bar{x}(\nu))] + \bar{x}(\nu) \quad (14)$$

where  $\bar{x}(\nu)$  is the average value of the data in the window. When the local variance is greater than or equal to the global estimate of the variance of the noise in the data,  $a(\nu)$  is assigned a value of zero and the center point of the window is set to the calculated average. In this case, the smoothing provided by the adaptive smoother is equivalent to that given by a polynomial smoother of order 1, and extensive smoothing results. As the local variance becomes much less than the variance of the noise,  $a(\nu)$  becomes close to one, and no smoothing occurs. Thus the degree of smoothing provided adapts to the local noise variance. It is easy to generalize the treatment of Kawata and Minami to higher-order polynomial filters, too. In this way, polynomial filters that adapt to local noise variance conditions can be produced using the approach outlined in eqs 11-14 above. Only the adaptive, polynomial smoother of degree one will be discussed here because it offers the greatest range of smoothing for data with nonuniform noise, a situation which occurs, among other cases, when data collected in the time domain are presented in the frequency domain.

The transfer functions for adaptive running-average smoothing are similar to a polynomial smoother of degree one, but offset on the y-axis by  $a(\nu)$ , the ratio of the local variance to the global variance. As with polynomial smoothing, the size of the window determines how much of the higher frequencies are attenuated. In this smoother algorithm, window size is not quite as important in controlling signal distortion as in the nonadaptive polynomial smoother, however, due to the adaptive nature of the algorithm, but if the window size is too large or too small, very little smoothing is achieved. The ratio of variances raises or lowers the transfer function on the y-axis, thus attenuating or passing frequencies to a greater or lower extent, respectively. Thus, guesses for the estimated variances have a significant effect on the attenuation of higher frequencies. Overly optimistic estimates (where the estimated variance is too small) cause over smoothing where the local signal variance exceeds the global estimate, according to eq 13 above. Overly pessimistic estimates cause little smoothing to occur, on the other hand. For any noisy signal, there should be optimal estimates of window size and global signal variance. With these parameters, a smoother transfer function can be

created that will attenuate mainly noise. Kawata and Minami suggested no method for determining the optimal estimates of signal variance or smoother window size in their derivation of this smoother.

**Fourier Smoothing.** The simplest smoother to implement is based on truncation of the noisy signal in the Fourier domain. The direct application of eq 5, with the rectangular weighting function  $H(\Omega)$ , removes frequencies above some lower limit but preserves all frequency components below that limit.<sup>3</sup> The rectangular truncation can lead to spurious oscillations (Gibbs' effect) in the back-transformed data, especially if the data are not properly preprocessed, and it is common to modify the truncation so that attenuation of frequencies near the cutoff point is more gradual.<sup>24</sup> A trapezoidal weighting function is quite commonly substituted in these cases, but other windowing functions can also be used. When noise and signal are well-separated in the Fourier domain, finding the proper point for truncation is simple.<sup>3</sup> However, when the noise and signal components share frequencies, finding the optimal truncation point is not straightforward. The same tradeoffs mentioned with the other two smoothers: signal preservation versus noise attenuation must be balanced in the optimal placement of the truncation.

In all three of the methods discussed above, optimal values for those parameters that control the transfer function shape are not easily accessible from the equations describing the smoothing in either the time or the Fourier domain. Commonly, these values can only be guessed by the experimenter. The empirical nature of their selection makes smoothing techniques extremely difficult to use in an optimal manner and makes comparison of the different methods challenging.

One difficulty in optimizing the smoothing or filtering produced by these methods is the problem of determining whether the change in shape of the smoothed signal produced by the signal processing is due to noise reduction or to signal distortion. In many cases, the change in the data produced by the signal processing is due to a change in *both* signal and noise. For most types of noise, there are two distinct regions in the frequency space representation of the data: one which, in effect, contains only noise and the other region which contains both noise and signal. Attenuation of frequencies in the region containing only noise results in decreased noise and no change in the signal shape because no frequency components of the signal are altered. When the transfer function of the smoother attenuates frequencies in the region where there is a significant amount of signal, however, the noise decreases, but the signal is distorted as well. It is possible to decide on the relative contributions of signal distortion and noise reduction by making use of Rayleigh's theorem.<sup>15,16</sup> Because the Fourier transform is a linear transform, moments of a function are preserved across the transformation, and the integral of the square modulus of a function is equal to the integral of the square modulus of its transform

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx = \int_{-\infty}^{+\infty} |F(\omega)|^2 d\omega \quad (15)$$

This theorem can be modified to show that for a given set of discrete, band-limited data in the frequency domain, the inverse Fourier transform can be obtained

$$\hat{s}(t) = \sum_{n=0}^b S_n e^{in\Delta\omega} \quad (16)$$

such that any integer value for  $b$  from 1 to  $N/2$  will give a vector  $\hat{S}$  that will have a constant area  $T$

$$T = \sum_{i=1}^N \hat{s}_i \quad (17)$$

where  $N$  is the number of points in the vector and where  $\hat{s}_i$  is the magnitude of the  $i$ th point in the vector  $\hat{S}$ . This relationship can be used to distinguish noise removal and signal distortion. When frequency components from the signal are removed, the intensity of the signal must decrease, and the signal must broaden to preserve the area  $T$ . This relation is exactly analogous to the effect of aliasing brought about by undersampling of a signal, as both are based on the representation of a wave form by a Fourier series.<sup>16</sup> Truncation or undersampling both broaden signals. Any empirical method which detects either the decrease in signal intensity or the increase in signal broadening resulting from filtering can therefore be used to improve the quality of the filtering or smoothing. The one used here is based on the Shannon information entropy.

**Entropy.** Entropy is a term which was first used to describe the amount of "disorder" in a system. Those states in a system which contain higher levels of entropy are favored over those which have less.<sup>25</sup> In a statistical sense these states are more probable. Informational entropy is defined mathematically in the discrete form as follows: A random experiment is performed that has  $n$  possible results. This experiment is performed  $j$  times, so that, assuming independence, there are  $N = n^j$  outcomes possible. The frequency  $f_i$  of each different outcome is

$$f_i = \frac{N_i}{N} \quad \text{for } i = 1-N \quad (18)$$

where  $i$  is one possible outcome of the experiment and where  $N_i$  is the number of times that outcome  $i$  occurred. The vector  $\mathbf{F} = \{f_1, f_2, \dots, f_n\}$  is a discrete probability distribution function.<sup>27,28</sup> The entropy of this distribution is defined as

$$H = - \sum_{i=1}^n f_i \log f_i \quad (19)$$

The base of the logarithm in eq 19 is arbitrary, but it is conventional to use base 2 logarithms, giving the entropy the units of bits.

By defining the information entropy  $H$  associated with data vector  $\mathbf{X}$  as

$$H = - \sum_{i=1}^n x_i \log x_i \quad (20)$$

where  $x_i$  is the value of  $i$ th point in the data vector scaled by the vector norm  $\|\mathbf{X}\|$ , an informational entropy for a spectrum can be calculated. When Gaussian noise with a mean of zero is randomly distributed on the noise-free data vector, the uncertainty associated with the data vector increases. It is convenient to think of this addition of noise as having the effect of broadening the confidence region in which the true, noise-free signal will fall.<sup>29</sup> The information entropy will decrease with the addition of noise; however, a direct reflection of decreased smoothness in the data.<sup>30</sup> Many possible "true" signals might be used to describe the noise-free part of the noisy composite signal. By calculating the information

(25) Kawata, S.; Minami, S. *Appl. Spectrosc.* 1984, 38, 49.

(26) Schrödinger, E. *Statistical Thermodynamics*, Cambridge University Press: Cambridge, England, 1948.

(27) Jaynes, E. T. *Proc. IEEE* 1982, 70, 939.

(28) Jaynes, E. T. *Annu. Rev. Phys. Chem.* 1980, 31, (1980) 579.

(29) Hurvich, C. M. *Technometrics* 1986, 28, 259.

(30) Shannon, C. E. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, 1949.

(24) Brown, S. D. In *Practical Guide to Chemometrics*; Haswell, S. J., Ed.; Dekker: New York, NY, 1992; Chapter 8.

entropy of a trial smoothed data vector obtained as a function of one or more adjustable parameters in the smoothing algorithm, a multivariate entropy function can be defined on the parameter space. Maximizing the information entropy of a smoothed representation of  $\mathbf{X}$ , as given by eq 20, yields the most probable smoothed version of  $\mathbf{S}$ . In maximizing the entropy, the amount of noise contained in the smoothed version of  $\mathbf{X}$  is, in effect, minimized subject to a constraint. The constraint is imposed by the nature of the unknown, noise-free signal itself, for which a minimal number of frequencies are required by the Nyquist criterion. The constraint can be specified in terms of the  $\chi^2$  statistic, where the fitting of the smoothed signal  $\hat{\mathbf{S}}$  to the noisy data  $\mathbf{X}$  is evaluated by

$$\chi^2 = \sum_{i=1}^n (x_i - \hat{s}_i)^2 / \sigma_i^2 \quad (21)$$

The estimated noise at point  $i$  is given by  $\sigma_i^2$ . The  $\chi^2$  distribution thus sets a restricted area of the confidence band of possible solutions wherein a maximum in informational entropy is sought.<sup>31</sup> Therefore, by maximizing the information entropy in this region, the most probable signal consistent with the unknown signal shape, but containing less noise than the original data, is selected from the many possible "true" signals. By the theorem of entropy concentration, it can be shown that many other highly probable solutions which are also consistent with the data lie very close to the one with maximum entropy.<sup>32</sup>

## EXPERIMENTAL SECTION

To compare the different methods and test the entropy maximization method in a systematic manner, a well-defined system is necessary. As a first approximation to experimental data, computer-generated Gaussian and Lorentzian peak shapes were used. These synthetic peaks described experimentally-observed spectral shapes fairly well, as most spectral peak shapes are either Lorentzian or Gaussian, while a few spectral peak shapes fall between that of a Lorentzian and that of a Gaussian.<sup>1</sup> By characterizing the smoothing of both shapes, a basis for estimating how well the method will perform on different types of experimental data can be established. Gaussian peaks were generated from the relation

$$S = A \exp \left[ \frac{4(\ln 2)t^2}{W^2} \right] \quad (22)$$

and Lorentzian peaks from the relation

$$S = \frac{A}{1 + 4t^2/W^2} \quad (23)$$

where  $A$  is the peak height,  $t$  is time or  $x$ -axis of the plot, and  $W$  is the full width at the half-height (fwhh) of the Gaussian or Lorentzian peak in the same units of the  $x$ -axis. Each of the synthetic peaks contained 200 points.

The Gaussian noise added to the synthetic data was generated using either a C language random number generator supplied with the UNIX BSD4.3 operating system or a uniform, random number generator in MATLAB. To ensure a Gaussian distribution of noise, five random numbers were generated between -1 and 1, and their average was calculated. This number was used as a single noise value. A noise vector was generated by repeating this process for each element in the vector. The noise vector was scaled to the desired root mean square value and then added to the known signal vector to generate the noisy data vector to be smoothed. The signal to noise ratio was calculated by

dividing the maximum peak height of the signal vector by the root mean square of the noise contained in the noise vector. The problem of distinguishing signal broadening versus noise reduction can be eliminated by detecting changes in either intensity of the peaks or in broadening of the smoothed peaks. To keep the peak intensity constant throughout the procedure, so as to measure peak broadening at constant peak height, the maximum peak height after smoothing was set to a constant value of 0.1 unit. Every smoothed vector was normalized in this manner.

The smoothing was also evaluated using experimental UV-visible spectra. Solutions of nickel nitrate, cobalt nitrate, and copper nitrate in a 0.1 M nitric acid solution were made up using reagent-grade chemicals. Their spectra were obtained with a Hewlett-Packard 8452 diode array spectrometer with 2-nm resolution. The concentration of these solutions was adjusted to give peak absorbance readings near 1.0 absorbance unit. Gaussian noise was added to these spectra using the same method described above. The smoothing techniques were applied, and optimal smoothing parameters were determined by using least-squares fitting of the unaltered spectra and by location of the maximum of the entropy function.

For calculating the coefficients used in the polynomial (Savitzky-Golay) smoothing algorithm, the pseudoinverse of the Vandermonde matrix was used.<sup>33</sup> These coefficients were then used to determine the smoothed values for each point in the data vector. With this method, the points prior to the first and after the last smoothing window could also be smoothed if desired. Results reported here were obtained with midpoint smoothers only.

The informational entropy was calculated according to eq 20. The negative of the entropy was minimized with the aid of a simplex optimization program. Least-squares fitting of the known, noise-free peak to the smoothed peak was used to evaluate the smoothing. Programs were written in the Pascal and MATLAB languages. Calculations were performed in the Pascal programming language on a Celerity 1200 minicomputer under the Unix BSD 4.3 operating system and in the MATLAB language on an Apple Macintosh SE with 68020/68882 accelerator board and on a Sun SPARC 1+ workstation.

## RESULTS AND DISCUSSION

Several sets of synthetic data with known signal shapes were generated, and noise was added to provide known responses which could be used to test the accuracy and precision of the smoothers generated by entropy maximization method. Gaussian and Lorentzian peaks of 40, 20, 10, and 5 points at full width half-height (fwhh) were used to evaluate the smoothing. For all these peaks, zero-mean Gaussian noise was added to produce data with a specified maximum signal to noise (S/N) ratio. A series of 10 different noisy data sets, containing the same signal vector, but different noise vectors, was examined for each trial.

To provide a check on the extent of the smoothing possible given the noise, the exact signal shape was used to determine the best (least-squares) smoothing possible with each smoother applied to the noisy data vector. Results from the smoothing were evaluated by fitting the smoothed result to the true signal. With the true signal shape, the sum of the squared error (SSE) between the data that had been smoothed,  $\hat{\mathbf{S}}$ , and the true, noise-free data,  $\mathbf{S}$ , can be calculated directly

$$\text{SSE} = \sum_{i=1}^N (\hat{s}_i - s_i)^2 \quad (24)$$

By plotting the sum of squared error obtained from eq 24 as a function of the smoothing parameters, a function which exhibits a minimum is obtained. Here, the parameters which produce smoothed data with a minimum sum of squared fit error are taken as the "true" parameters. These parameters are optimal in a least-squares sense because they minimize a sum of squared error terms between the true and the smoothed data.

(31) Skilling, J.; Gull, S. F. In *Maximum Entropy and Bayesian Methods in Inverse Problems*; Smith, C. R., Grandy, W. T., Eds.; Reidel: Dordrecht, 1985.

(32) Jaynes, E. T. In *Maximum Entropy and Bayesian Methods in Inverse Problems*; Smith, C. R., Grandy, W. T., Eds.; Reidel: Dordrecht, 1985.

(33) Bialkowski, S. E. *Anal. Chem.* 1989, 61, 1308.

**Table I. Comparison of Best Smoothing Parameters for a Gaussian Peak<sup>a</sup> Containing Different Noise Vectors**

	least-squares method		entropy method	
	window size	SSE value of fit	window size	SSE value of fit
	6	0.143 74	6	0.143 74
	10	0.144 02	10	0.144 02
	6	0.144 00	4	0.144 05
	6	0.144 17	4	0.144 21
	5	0.143 89	4	0.143 96
	7	0.143 93	7	0.143 93
	9	0.143 71	5	0.143 74
	7	0.143 77	7	0.143 77
	7	0.143 96	7	0.143 96
	6	0.143 86	6	0.143 86
av	6.9	0.143 90	6.0	0.143 92
SD	1.5	0.000 14	1.9	0.000 15

<sup>a</sup> Ten noisy Gaussians with fwhh of 40 pts, and maximum S/N ratio of 10, were used as input. Fourier smoothing was done using a rectangular smoothing window.

The optimal smoothing parameters were obtained for the entropy maximization and the least-squares minimization. With these values, any apparent difference in the optimized smoothing parameters obtained from the two methods could be compared in terms of the quality of smoothing. Optimizing smoothing parameters by entropy maximization does not always result the same values as those calculated from the least-squares minimization, however. Table I summarizes results for smoothing of a series of 10 different, noisy signals made from the same signal vector, but different noise vectors. Here, the smoothing window of a rectangular, Fourier-domain smoother was to be optimized. The noise vectors all had approximately the same root mean square noise value (0.1) and mean ( $\sim 0$ ). The variation in the optimal windows found by the entropy method and the least-squares method results from variations in the noise vector and the effect of the noise on the two approaches to optimization. Noise acts in a well-defined manner on average, but it may not act in a well-defined manner for any one specific case. Therefore, to properly account for the effect of the noise, an average of 10 runs was taken, and the average optimal parameter set is given along with its standard deviations. These statistical indicators describe the random effects of the noise much more reliably than a one-case example.

In evaluating discrepancies in window sizes between the least-squares and the entropy methods, it must also be determined how much any difference in the predicted smoother parameters makes in the final, smoothed results. While the most probable signal shape determined from entropy maximization may not be the optimal, smoothed signal shape determined from the least-squares fits, the principle of entropy concentration suggests that these two should be close and that the entropy maximum should be rather broad. If the least-squares and entropy functions both show broad extrema, large differences in smoothing param-

eters will make little difference in the smoothed data and little difference in the sum of squared error (SSE) values obtained from fitting the smoothed data to the true signal, for example. If, however, the minima in the least-squares function and maxima in the entropy function are sharp, even small differences in the smoothing parameters, as determined from the two methods, will have a pronounced impact on the quality of the smoothing. To evaluate the significance of differences in sets of optimal parameters, SSE values are calculated by fitting the optimal, smoothed data produced from both functions to the true signal. In Table I, the average minimum SSE value produced by fitting data obtained from the smoothing using prior knowledge of the signal (the least-squares method) was 0.143 90. The average SSE value calculated by fitting the signal obtained from the information entropy-optimized smoothing (the entropy method) to the noisy data was 0.143 92. The difference of only 0.000 02 in the average minimum SSE values for the two methods clearly indicates that the entropy-based selection of smoothing window, on average, produces smoothing results which are essentially identical to those obtained from least-squares fitting of data using prior knowledge of the true signal shape. Unlike the least-squares smoothing method, however, the entropy-based selection of smoothing parameters presumes no prior knowledge of signal or noise.

**Polynomial Smoothing.** One set of studies evaluated the effectiveness of polynomial smoothing. Considerable work on polynomial smoothers has been reported by Willson and Edwards<sup>1</sup> and by others,<sup>15-22</sup> but no previous work has appeared on the optimization of these smoothers for a particular signal. It was of interest to determine the success with which the entropy could be used to optimize the two adjustable parameters in the polynomial smoother: smoothing window size and polynomial order. Both can be varied to generate a series of smoothed data, for which an entropy surface can be generated, and a maximum located. This was done for Gaussian peaks with varying peak widths and signal-to-noise ratios of 10. An exhaustive search was used to locate the optimal smoothing parameters for polynomial smoothing of Gaussian data at low (maximum S/N = 10) and medium (maximum S/N = 100) signal-to-noise ratios. Average results from smoothing of 10 noisy data each are summarized in Table II. While a correlation exists between the true optima for least-squares smoothing and those selected by entropy maximization, there is scatter in the optimal polynomial order and smoother window selected by maximizing the informational entropy. This result can be explained by considering the nature of the transfer function for polynomial smoothers. Because changes in smoother window and order cause relatively minor changes in the smoother transfer function, as compared to changing the length of the rectangular window transfer function for a Fourier domain smoother, the two-dimensional entropy surface for polynomial smoothing should not be as steeply pitched to the single, obvious minimum seen in plots of the SSE value as a function of the size of a

**Table II. Comparison of Best Smoothing Parameters for Gaussian Peaks, Variable-Order Polynomial Smoothing<sup>a</sup>**

	pts fwhh	optimum window				optimum order				sum of the squared error value			
		LSQ <sup>b</sup> method	SD <sup>c</sup>	entropy method	SD <sup>c</sup>	LSQ method	SD <sup>c</sup>	entropy method	SD <sup>c</sup>	LSQ method	SD <sup>c</sup>	entropy method	SD <sup>c</sup>
S/N	40	44.0	3.7	44.6	7.4	2.4	0.8	2.7	2.2	0.08	0.02	0.09	0.03
10	20	38.0	11.1	37.4	10.9	4.3	2.2	3.6	1.8	0.15	0.04	0.21	0.14
	10	26.8	10.2	36.8	11.7	5.1	3.1	5.1	3.0	0.19	0.05	0.5	0.5
	5	13.4	4.9	23.0	10.2	3.4	2.1	5.3	2.9	0.34	0.05	0.6	0.3
S/N	40	37.0	10.0	48.8	0.6	3.2	1.0	0.0	0.4	0.0008	0.0003	0.0159	0.0016
100	20	38.2	5.2	41.0	3.4	6.2	1.1	2.2	0.4	0.0014	0.0004	0.0119	0.005
	10	22.0	6.4	28.0	7.3	6.3	2.1	5.5	2.1	0.0029	0.0007	0.0142	0.006
	5	12.6	3.0	14.4	4.0	5.8	1.5	5.1	2.1	0.0051	0.0009	0.017	0.007

<sup>a</sup> Polynomial order varied from 2 to 9. Window size varied from 2\*(order) - 1 to 128. <sup>b</sup> Least-squares method. <sup>c</sup> Standard deviation.



Table III. Comparison of Best Smoothing Parameters for Gaussian Peaks<sup>a</sup>

pts fwhh	optimum window				sum of the square error value				difference in SSE
	LSQ method	SD	entropy method	SD	LSQ method	SD	entropy method	SD	
Polynomial Smoothing, Third Order									
40	11.3	1.4	19.7	5.1	0.016 00	0.000 17	0.016 33	0.000 26	0.000 33
20	18.8	1.6	22.6	3.0	0.033 00	0.000 20	0.033 06	0.000 17	0.000 06
10	25.6	3.7	33.7	4.7	0.069 36	0.000 19	0.069 42	0.000 20	0.000 06
5	37.3	7.6	39.4	7.8	0.143 79	0.000 11	0.143 80	0.000 12	0.000 01
Fourier Smoothing, Rectangular Window									
40	6.3	1.8	6.6	2.0	0.143 82	0.000 08	0.143 83	0.000 08	0.000 01
20	8.9	1.0	8.8	1.1	0.069 30	0.000 15	0.069 31	0.000 17	0.000 01
10	15.7	1.8	14.6	2.0	0.033 10	0.000 14	0.033 13	0.000 16	0.000 03
5	24.3	1.7	20.4	5.1	0.015 92	0.000 14	0.016 05	0.000 26	0.000 13
Fourier Smoothing, Trapezoid Window									
40	5.1	1.2	4.5	1.4	0.143 83	0.000 09	0.143 84	0.000 10	0.000 01
20	7.9	1.3	7.2	1.6	0.069 23	0.000 18	0.069 25	0.000 18	0.000 02
10	14.6	1.4	13.6	2.4	0.032 93	0.000 15	0.032 96	0.000 14	0.000 03
5	24.3	3.0	18.6	5.8	0.015 93	0.000 23	0.016 05	0.000 36	0.000 12
Adaptive Smoothing									
40	13.2	3.9	16.1	5.7	0.143 97	0.000 08	0.143 97	0.000 08	0.000 00
20	17.8	6.3	21.6	5.0	0.069 59	0.000 15	0.069 60	0.000 16	0.000 01
10	18.2	1.9	20.4	3.8	0.032 91	0.000 21	0.032 93	0.000 22	0.000 02
5	15.6	2.7	20.0	4.5	0.015 51	0.000 31	0.015 54	0.000 32	0.000 03

<sup>a</sup> The maximum signal-to-noise ratio for all studies reported in this table was 10.Table IV. Comparison of Best Smoothing Parameters for Lorentzian Peaks<sup>a</sup>

pts fwhh	optimum window				sum of the square error value				difference in SSE
	LSQ method	SD	entropy method	SD	LSQ method	SD	entropy method	SD	
Polynomial Smoothing, Third Order									
40	13.8	2.8	24.8	4.0	0.016 85	0.000 13	0.017 05	0.000 21	0.000 20
20	20.0	2.6	26.2	4.6	0.035 10	0.000 11	0.035 20	0.000 09	0.000 10
10	21.4	9.3	23.9	11.5	0.157 65	0.000 07	0.157 65	0.000 07	0.000 00
5	19.7	6.4	9.3	8.6	0.150 91	0.000 06	0.150 94	0.000 08	0.000 03
Fourier Smoothing, Rectangular Window									
40	9.3	2.2	6.2	10.1	0.150 93	0.000 09	0.151 14	0.000 06	0.000 21
20	12.7	6.5	12.0	9.9	0.157 69	0.000 12	0.157 82	0.000 14	0.000 13
10	15.7	2.0	11.7	3.0	0.035 11	0.000 10	0.035 21	0.000 18	0.000 10
5	19.0	1.1	15.0	2.3	0.016 80	0.000 21	0.016 89	0.000 21	0.000 09
Fourier Smoothing, Trapezoid Window									
40	8.0	1.5	23.7	14.1	0.150 93	0.000 09	0.150 94	0.000 09	0.000 01
20	13.3	7.7	11.9	8.1	0.157 69	0.000 12	0.157 70	0.000 12	0.000 01
10	12.6	2.5	8.0	1.9	0.035 10	0.000 11	0.035 22	0.000 19	0.000 12
5	18.0	1.4	11.3	3.8	0.016 78	0.000 21	0.016 97	0.000 33	0.000 19
Adaptive Smoothing									
40	5.8	2.1	7.2	4.7	0.151 05	0.000 06	0.151 05	0.000 06	0.000 00
20	5.1	2.8	5.1	2.8	0.157 84	0.000 06	0.157 84	0.000 06	0.000 00
10	7.6	2.1	6.2	2.3	0.035 94	0.000 09	0.035 95	0.000 09	0.000 01
5	6.4	1.0	5.6	2.5	0.017 52	0.000 12	0.017 52	0.000 13	0.000 00

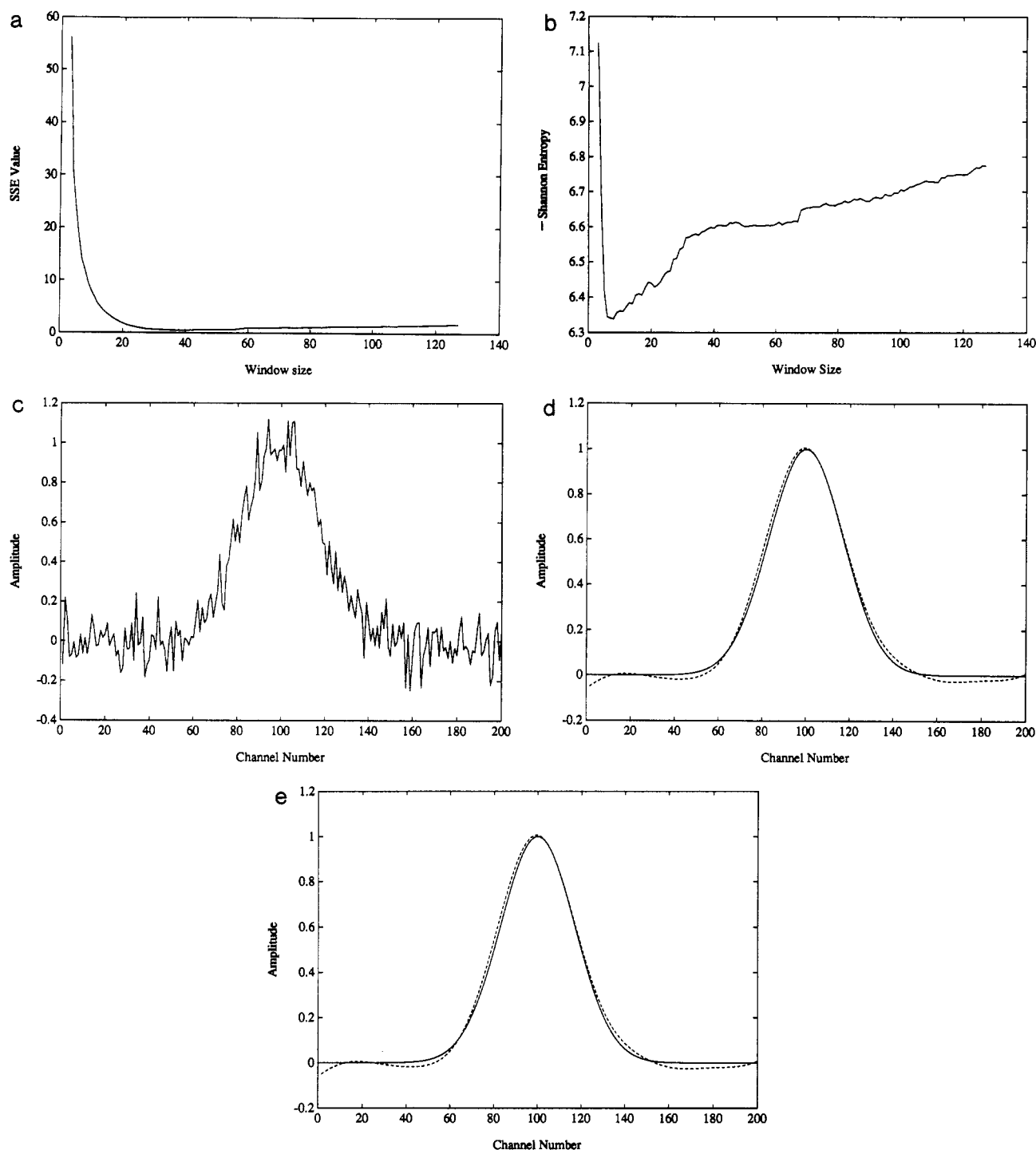
<sup>a</sup> The maximum signal-to-noise ratio for all studies reported in this table was 10.

rectangular smoother window. Instead, a broad, shallow entropy maximum exists. While, as expected, one combination of smoother window and polynomial order was found to give the best smoothing, several other combinations of smoothing window and polynomial order often gave smoothed signals with entropies that were only a little smaller, especially when narrow peaks with high noise levels were smoothed.

To reduce the scatter in the selection of optimal smoothing parameters, a more restricted set of studies was carried out. In these, a polynomial smoother with fixed order was used to process the data, and the entropy was used to optimize only the smoothing window. The order of the polynomial smoother determines the moments of the data that pass unaffected by the smoothing. For example, a third-order smoother will preserve the zeroth through the third moments of a noisy signal—the area, the mean, the variance, and the kurtosis—while a first-order smoother preserves only the zeroth and first moments (area and mean) of the signal.<sup>13</sup> For this reason, smoothers of second-degree and higher were

evaluated. The entropy was calculated for the smoothed data, as discussed above, and the plot of entropy versus smoothing window size was used to locate the optimal smoothing window size. Only results from application of the third-order polynomial smoother are reported here, as other smoothers behaved similarly.

Results from the application of the third-order polynomial smoother applied to noisy Gaussian signals of varying peak widths are given in Table III. This table is divided into three vertical sections. The column in the first section identifies the true signal shape and peak half-width. The middle columns of these tables compare the average optimum window obtained by the matched smoothing (least-squares) method and the entropy maximization method. The standard deviation for these optimal window widths are also given in this section. Ten trials were taken to determine the mean and standard deviation for each of the extrema located. The right-hand columns compare the averages of the sum of the squared error obtained at the optima. The standard deviation of these



**Figure 2.** Least squares and Shannon entropy functions for Fourier smoothing. (a) The SSE value determined as a function of the size of the smoothing window in the Fourier domain. (b) The information entropy determined as a function of the size of the smoothing window in the Fourier domain. (c) The noisy simulated data used to calculate these functions. The channel number indicates the ordinal number of the data. (d) The estimated results compared to the true signal using SSE optimization. (e) The estimated results compared to the true signal using information entropy optimization.

SSE values are also given in this section. The last column in this section gives the difference between the two averaged sums of the squared error values obtained by the two different methods. The results indicate that, on average, entropy-optimized smoothing identifies the same window size as that selected with prior knowledge of the signal. Note that the window size increases as the peak width of the Gaussian peak decreases. This is consistent with the fact that narrower peaks contain higher frequency components (and more non-zero signal frequencies in the Fourier domain) than broad peaks.

The third-order polynomial smoother was also applied to noisy Lorentzian peaks with various fwhh values. The results are summarized in Table IV. In Table IV there is the same

inverse correlation between the number of points used in the best smoothing window and the fwhh of the peak. The smaller the number of points at the fwhh for the peak, the larger the number of points necessary in the window used by the polynomial smoother to obtain optimal smoothing.<sup>1</sup> For the Lorentzian peaks, however, entropy-based smoothing gave more variable estimates for the optimal window size, and the entropy optimization of the polynomial smoother was not as accurate as for the Gaussian peaks. The lower quality of the results was found to result from the method used for normalizing the smoothed peaks to constant maximum intensity for comparisons of smoothing results, however, and not from any inherent limitation in smoothing or in optimizing

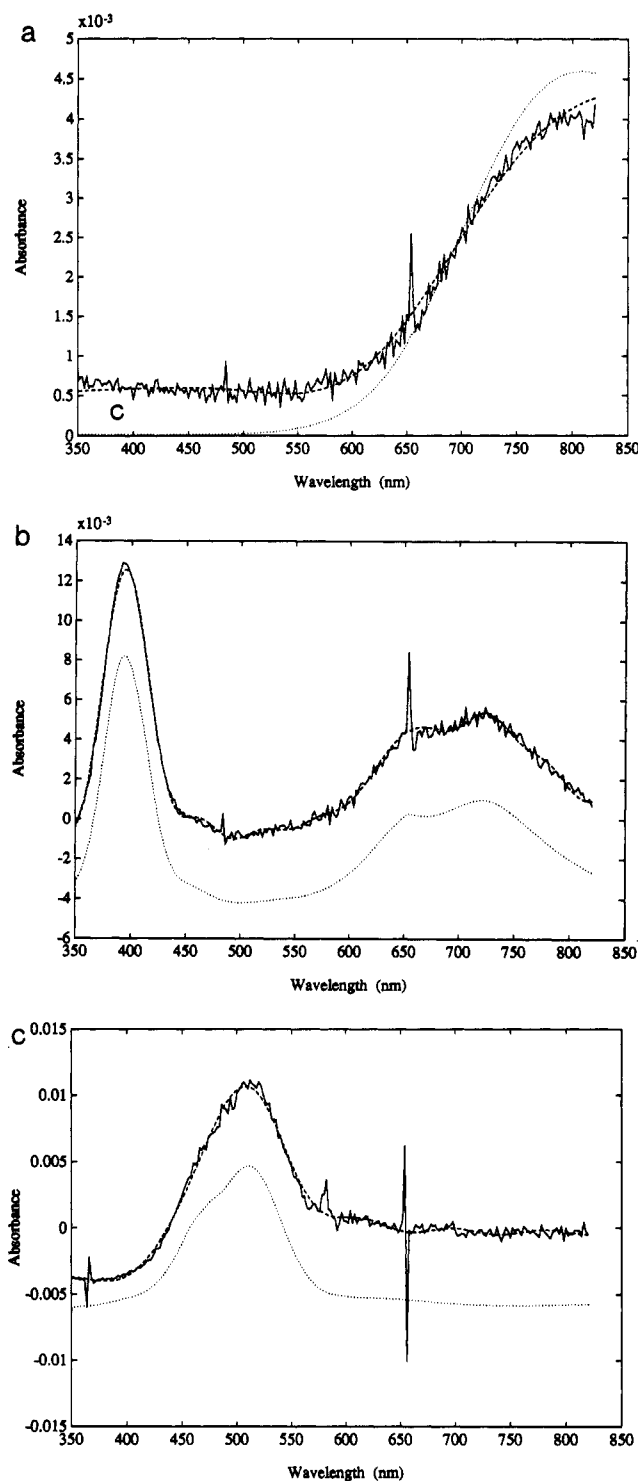


the smoother. Taking the point of maximum peak height of a smoothed Lorentzian as the scaling factor means that any error in the point determining the peak maximum translates to error in scaling the entire peak, which shows in the SSE of the fit to the true peak. By using a different normalizing procedure, where several points are averaged to find the peak maximum, noisy Lorentzian peaks can be smoothed very accurately. That normalization procedure is not as well suited for Gaussian peaks as the single point normalization, however.

**Fourier Smoothing.** The rectangular Fourier-domain smoother has only one adjustable parameter—the smoothing window size. Smoothing was investigated as a function of that parameter. Fourier smoothing was performed following methods previously reported:<sup>3</sup> the Fourier transform of the data vector was obtained, then both the real and imaginary parts of the transformed data were multiplied by a rectangular weighting function, and finally the inverse Fourier transform was taken. The rectangular function was constructed so that frequency components were passed unchanged (where the rectangular function was one) or were attenuated to zero (where the rectangular function was zero). To reduce the noise in a systematic fashion, the number of points set to a value of one in the rectangular function was increased from one to the total number of points in the data vector. The entropy of these smoothed vectors was calculated and plotted as a function of the number of points used in the rectangular portion of the weighting function. The SSE values from fitting the smoothed vectors to the true peaks were also calculated and plotted as discussed above. The optimal window sizes obtained from the least-squares and entropy functions were then compared. Typical plots of the SSE and entropy as functions of the smoothing window size are given in Figure 2. The agreement was found to be very good. Two different types of weighting functions have been used with Fourier-domain smoothing by others.<sup>3</sup> One is the rectangular window discussed above. Entropy maximization was evaluated using Fourier smoothing with a rectangular window applied to the same peaks as those used for the polynomial smoothing. Tables III and IV list the results of smoothing the noisy Gaussian and Lorentzian peaks. In general, Fourier-domain smoothing gave slightly better results than those obtained from polynomial smoothing. There again was a correlation between the width of the peak and the number of points necessary for optimal smoothing in the rectangular portion of the weighting function: the wider the peak, the smaller the number of points necessary. Again, the entropy optimization method proved to be more successful with Gaussian peaks than with Lorentzian peaks because the same normalization method was used.

The other common weighting function recommended for Fourier-domain smoothing has a trapezoidal shape.<sup>3</sup> As Tables III and IV show, for the noisy Gaussian and Lorentzian peaks examined here, a simple trapezoidal function gave results similar to those observed with the rectangular window. These results can be explained by examining Table I more closely. The Gaussian signal can be described exactly using only the first 10 frequencies, and a very reasonable description can be obtained using only 6 frequency components. Therefore, any weighting function that does not attenuate the first 6–10 frequencies will give very similar results, so long as most of the higher frequency noise is attenuated. The aliasing expected from truncation<sup>3</sup> is not a problem if signal frequencies have zero intensity already. Aliasing only occurs where non-zero signal frequencies are truncated, a situation that is relatively easily avoided in these data sets. By finding the optimal rectangular function, the use of other types of weighting functions is not necessary to obtain satisfactory results in smoothing these simple peak shapes.

**Adaptive Polynomial Smoothing.** Finally, the entropy optimization method was also used to examine the optimal window size for use with the adaptive, running-average



**Figure 3.** (a) The spectrum of copper(II) nitrate measured in dilute solution (—), with smoothed spectrum (---), and a normalized spectrum measured at high signal to noise ratio (···). Smoothing was optimized with Shannon entropy. The normalized spectrum has been offset for clarity. (b) The spectrum of nickel(II) nitrate measured in dilute solution (—), with smoothed spectrum (---), and a normalized spectrum measured at high signal to noise ratio (···). Smoothing was optimized with Shannon entropy. The normalized spectrum has been offset for clarity. (c) The spectrum of cobalt(II) nitrate measured in dilute solution (—), with smoothed spectrum (---), and a normalized spectrum measured at high signal to noise ratio (···). Smoothing was optimized with Shannon entropy. The normalized spectrum has been offset for clarity.

smoothing method of Minami and Kawata.<sup>25</sup> This method for adaptive smoothing also has two adjustable parameters: the smoothing window size and the estimate of the global signal variance. While both parameters can be varied, a detailed study of the effects of changing only the smoothing

Table V. Comparison of Best Smoothing Parameters for Metals

metal	optimum window				sum of the square error value				difference in SSE
	LSQ method	SD	entropy method	SD	LSQ method	SD	entropy method	SD	
Polynomial Smoothing, Third Order									
Co	36.6	9.5	47.5	13.7	0.073 06	0.000 08	0.073 07	0.000 08	0.000 01
Cu	57.0	9.1	56.8	9.4	0.122 46	0.000 05	0.122 46	0.000 05	0.000 00
Ni	20.0	4.3	26.8	10.0	0.066 75	0.000 04	0.066 76	0.000 04	0.000 01
Fourier Smoothing, Rectangular Window									
Co	7.0	1.5	5.1	1.5	0.073 07	0.000 06	0.073 09	0.000 07	0.000 02
Cu	4.4	2.1	4.0	1.6	0.122 43	0.000 05	0.122 43	0.000 05	0.000 00
Ni	1.1	2.1	14.5	9.1	0.066 75	0.000 03	0.066 76	0.000 03	0.000 01

window size is reported here. The approach taken in these studies was identical to that used for the polynomial smoothing studies. As in those studies, smoothing of a Gaussian peak by adaptive smoothers with varying smoothing window size and varying estimates of the global noise variance was performed. The estimates of the global noise variance were permitted to range from one-tenth of the true noise level to 10 times the true noise level.

Tables III and IV give the results for adaptive smoothing of Gaussian and Lorentzian peaks, respectively. This method produced essentially equivalent results from smoothing the Lorentzian peaks and the Gaussian peaks. The sum of the squared error values for the various peak were slightly lower than those obtained in the Fourier smoothing and polynomial smoothing, but the small difference in smoothing falls well within the confidence intervals set by the fitting. This similarity of performance indicates that all three smoothing methods give similar results, once their smoothing parameters are optimized.

**Comparing the Smoothing of Noisy, Visible Spectra.** As an additional check of the adequacy of smoothing by entropy-selected smoothing windows used with these smoother algorithms, noisy, experimental UV-visible spectra were also smoothed. The spectra of nickel(II), copper(II), and cobalt(II), obtained by measurement of dilute solutions of metal nitrate in 0.1 M HNO<sub>3</sub>, are shown in Figure 3. These spectra provide a somewhat representative sample of UV-visible spectra found in the literature: the spectrum of copper(II) is very simple, while that of cobalt(II) is a little more complex, and nickel(II) the most complex, with two significant peaks with different peak widths. Each was smoothed by several of the methods discussed above, but only the results of Fourier-domain smoothing are shown. As Figure 3 shows, the Fourier-domain smoothing is successful, but small changes in the metal spectra with concentration occur as a result of shifts in the hydrolysis equilibria, even in 0.1 M HNO<sub>3</sub>.<sup>34</sup> To permit an objective evaluation of the smoothing in the absence of concentration-dependent hydrolysis, noise of a form similar to that observed in the dilute spectra shown in Figure 3 was added to noise-free spectra of the three metals in 0.1 M HNO<sub>3</sub>. As before, sets of noisy spectra were generated by creating random sets of noise vectors. The sets of noisy spectra were then smoothed by using both Fourier smoothing and third-order polynomial smoothing methods, as summarized in Table V. The results, on average, are similar for both smoothing techniques. Fourier smoothing with a rectangular weighting function gave slightly more consistent results, as shown by the smaller standard deviation for the extrema obtained. This is not surprising, considering that all of the smoothing methods are very closely related. The Fourier smoothing might be expected to perform better because no high-frequency noise is passed above the cutoff frequency, while the polynomial smoother and adaptive smoother pass some small amounts of high-frequency noise, as a comparison of transfer functions shows.

## CONCLUSIONS

All of the smoothing methods discussed here remove noise by attenuating the high frequencies. The critical aspect for these methods lies in setting those parameters which eliminate noise frequencies. To improve the polynomial smoothing method, more information concerning those frequencies which contain both noise and signal must be obtained. As shown above, it is possible to "fine-tune" the shape of the transfer function somewhat by simultaneously optimizing the smoothing window and the degree of the smoothing polynomial. With the adaptive smoother, the additional optimization involves selecting a suitable value for the global noise variance, the parameter that turns the adaptive smoother off and on as noisy regions are encountered. As with the single-parameter optimizations reported above, all multiparameter optimizations can be done by exhaustive searches so that the nature of the extremum is identified, but other methods for locating optima (e.g., a simple search or simulated annealing of the entropy as a function of the adjustable smoothing parameters) could be used to speed routine optimizations.

While the entropy optimization method is general, the optimum obtained is specific to filtering or smoothing for a particular combination of smoother, noise, and signal. It is not possible to optimize the filtering or smoothing for a set of unknown signals in advance because the entropy constraint—the shape of the noisy signal—is lacking. Unless the noisy signal shape is known (or can be approximated) in advance, it is necessary to reoptimize the filtering or smoothing parameters for each signal and noise combination.

Finally, it should be noted that the entropy optimization method reported here can be applied to data other than spectra and smoothers other than the ones selected for discussion here. The maximum entropy solution is the optimal solution when the data are corrupted by noise due to undersampling<sup>27</sup> but is not optimal for removing white noise overlapped with data at all frequencies. Because truncation of noise-containing data amounts to an undersampling of those data, when there are two regions in frequency space, one with data and noise, and the other with only noise, the entropy-optimized smoothing should be nearly optimal. Given the overlap of noise and signal, however, it should be apparent that the smoothing, even when optimized, is not a panacea for careless data acquisition; even with optimal smoothing, some noise remains imbedded in the smoothed result.

## ACKNOWLEDGMENT

This work was supported by the Division of Chemical Sciences, Office of Basic Energy Research, of the U.S. Department of Energy, under Grant DE-FG02-86ER13542.

(34) Hartley, F. R.; Burgess, C.; Alcock, R. *Solution Equilibria*; Halstead Press: New York, NY, 1980.