ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/224473494

Binomial smoothing filter: A way to avoid some pitfalls of least-squares polynomial smoothing

Article *in* Review of Scientific Instruments · September 1983

DOI: 10.1063/1.1137498 · Source: IEEE Xplore

CITATIONS
165

READS
124

2 authors, including:

Louis Marmet
York University
73 PUBLICATIONS **1,467** CITATIONS

SEE PROFILE

Available from: Louis Marmet
Retrieved on: 16 August 2016

Binomial smoothing filter: A way to avoid some pitfalls of least-squares polynomial smoothing

P. Marchand and L. Marmet

Laboratoire de Physique Atomique et Moléculaire, Département de Physique, et Centre de Recherches sur Les Atomes et Les Molécules, Université Laval, Québec, Canada G1K 7P4

(Received 26 August 1982; accepted for publication 27 March 1983)

Some pitfalls of least-squares polynomial (LSP) smoothing (better known to analytical chemists as Savitzky–Golay smoothing) are demonstrated and discussed, as well as some remedies. For instance, smoothing by long LSP sequences leads to transmission zeros, phase reversals, and overshoots that may be objectionable in some applications. An alternative method, the binomial smoothing filter, is described and some of its properties are presented. It is shown to be preferable to LSP smoothing in many cases. It is faster, better behaved in both the data and the frequency domains for many applications, and simpler to use and to program or to implement in hardware. Finally, the action of a filter of a given order is easy to predict.

PACS numbers: 84.30.Vn

INTRODUCTION

Least-squares polynomial (LSP) smoothing is a very old technique described in many manuals on numerical analysis to reduce high-frequency noise and improve the signal-to-noise ratio of data.[1–5] It is also often used to obtain smoothed derivatives.[1–5] It will be shown in Sec. VIII that smoothing can also be used to produce high-pass filtering useful for extracting small signals from large nonlinear backgrounds and removing low-frequency trends.

In 1964, Savitzky and Golay[6] published (somewhat erroneous) coefficients for LSP smoothings up to 25 points from the first- to the fifth-degree polynomials and for the first to the fifth derivative. As mentioned by Madden,[7] this paper has since been widely cited, from which it may be inferred that the technique has been and is still widely used. In fact, many manufacturers of data-acquisition instruments even implement the five-point third-degree LSP smoothing as part of standard hardware or software packages. Since its use is so widespread, it seems worthwhile to make its users aware of some of its dangers and limitations. The present paper does not dwell on the caveats of smoothing in general but underlines some of the pitfalls that may result from the indiscriminate use of this particular LSP technique, and describes another method which is more efficient computationally, is not subject to these pitfalls, and has other desirable properties.

I. GENERAL PROPERTIES OF SMOOTHING FORMULAS

Let $\{x_n\}(n = \ldots, -2, -1, 0, 1, 2, \ldots)$ be a given data sequence which one wishes to smooth. The computational algorithm or filter for the smoothing process is assumed to be a linear weighting of the input data sequence.[8] This is also called moving average smoothing or convolution of the data sequence by the filter coefficients. The weighting factors are the coefficients of the filter. Thus, the smoothed output data $\{y_n\}$ are given by

$$y_n = \sum_{k=-N_p}^{N_p} b_k x_{n-k}, \tag{1}$$

where b_k = filter coefficients symmetrical about the central coefficient b_0 such that $b_{-k} = b_k$, and $2N_p + 1$ = total number of coefficients in the filter, where N_p is any positive integer.

The problem is now to determine the coefficients b_k that provide a smoothing action having desirable properties. According to Schoenberg,[9] a sufficient condition for $\{b_k\}$ to be a smoothing formula is

$$\sum_{k=-N_p}^{N_p} b_k = 1$$

and

$$\sum_{k=-N_p}^{N_p} |b_k| < \infty,$$

while its characteristic function

$$\phi(u) = b_0 + 2b_1 \cos u + 2b_2 \cos 2u + \ldots \tag{2}$$

satisfies the condition

$$-1 \leqslant \phi(u) \leqslant 1 \quad (0 \leqslant u \leqslant 2\pi).$$

Replacing u by $2\pi j/N$, where $j = 0,1,2, \ldots, N-1$, it is seen that the characteristic function is nothing else than N times the real part of the digital Fourier transform of the sequence b_k evaluated over N points.[10] Schoenberg[9] also states that it is desirable for an efficient smoothing formula to have its characteristic function satisfy the more restrictive condition

$$0 \leqslant \phi(u) \leqslant 1. \tag{3}$$

It will be seen below why this condition is particularly significant in some cases.

II. FOURIER ANALYSIS

Explained now is why the above smoothing coefficients are chosen to have even symmetry and to be odd in number $(2N_p + 1)$. It is well known[10] that if the smoothing coefficients b_k satisfy these conditions, the Fourier transform of the coefficients or transfer function of the filter is purely real, i.e., as pointed out by Kaiser and Reed,[8] the phase function is

 0034-6748/83/081034-08$01.30 © 1983 American Institute of Physics

Downloaded 06 May 2009 to 132.246.170.171. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/rsi/copyright.jsp

identically zero for all frequencies and the phase shift between input and output is zero and independent of frequency. This seems evident since the phase response of a filter is the arctangent of the imaginary part of its transfer function divided by its real part. However, it is not mentioned in Ref. 8 that if the transfer function is real and negative in any given frequency range, the resulting phase shift is π instead of 0, thus leading to phase reversals at these frequencies.

For reference in the rest of this paper, we shall make use of the following standard notation. Let X_j be the digital Fourier transform of the N-point data sequence $\{x_k\}$. We write

$$x_k \leftrightarrow X_j \quad (j, k = 0, 1, 2, \ldots, N-1), \tag{4}$$

where

$$X_j = \left(\frac{1}{N}\right) \sum_{k=0}^{N-1} x_k \exp(-i2\pi jk/N) \tag{5}$$

is the frequency behavior of x_k

and

$$x_k = \sum_{j=0}^{N-1} X_j \exp(i2\pi jk/N) \tag{6}$$

is the data (or time) behavior resulting from a given frequency content X_j.

For convenience we shall make use of the normalized frequency variable ν.[8] Let f_s be the sampling frequency, i.e., the reciprocal of the interval T between each of the N data points: $f_s = 1/T$. The exponent in Eqs. (5) and (6) may be rewritten as $i2\pi kfT$, where the frequency variable f is j/NT. In terms of the sampling frequency f_s, this becomes $i2\pi kf/f_s$. Now, the so-called Nyquist frequency f_N is half the sampling frequency: $f_N = f_s/2$, so that the exponent becomes $i\pi kf/f_N$ when expressed in terms of f_N. Let ν be defined as the normalized frequency f/f_N. The exponent then becomes $i\pi k\nu$. This maps the Nyquist interval $0 \leqslant j \leqslant N/2$ to the interval $0 \leqslant \nu \leqslant 1$. The number of points per cycle in the data domain is hence given by $2/\nu$.

In the same fashion, let

$$b_k \leftrightarrow H(\nu), \tag{7}$$

with $N > 2N_p + 1$, the number of smoothing coefficients. The smoothing or convolution operation as defined in Eq. (1) is written

$$y_k = b_k * x_k$$

and is commutative and associative.[10] b_k is also the impulse response of the filter since if the impulse function $\delta(k)$ is defined as

$$\delta(k) = \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise,} \end{cases}$$

$$b_k * \delta(k) = b_k.$$

$H(\nu)$ is called the transfer function of the filter.

From the convolution theorem,[10]

$$b_k * x_k = y_k \leftrightarrow Y(\nu) = H(\nu)X(\nu), \tag{8}$$

so that in the frequency domain the convolution operation is transformed into a complex multiplication.

III. LEAST-SQUARES POLYNOMIAL SMOOTHING

A $(2N_p + 1)$-point LSP smoothing of degree n is performed by replacing the k th data point x_k by the value, at that point, of the calculated power series of degree $n < 2N_p + 1$ best fitting the data points x_{k-N_p} to x_{k+N_p} in the least-squares sense. The procedure is repeated for each k. This reduces to a single smoothing formula[1-6] depending only on n and N_p, except for the N_p points at each end of the data sequence.

Let us now examine some properties of these LSP smoothing formulas.

For $n = 1$ and $N_p = 1$,

$$b_k = \{1, 1, 1\}/3$$

and

$$H(\nu) = (1 + 2\cos \pi\nu)/3,$$

where the factor $1/N$ has been left out for simplicity. Figure 1(a) shows $H(\nu)$ as a function of ν.

For $n = 1$ and $N_p = 2$,

$$b_k = \{1, 1, 1, 1, 1\}/5$$

and

$$H(\nu) = (1 + 2\cos \pi\nu + 2\cos 2\pi\nu)/5$$

[see Fig. 1(b)].

For $n = 3$ and $N_p = 2$,

$$b_k = \{-3, 12, 17, 12, -3\}/35$$

and

$$H(\nu) = (17 + 24\cos \pi\nu - 6\cos 2\pi\nu)/35$$

[Fig. 1(c)].

For $n = 3$ and $N_p = 3$,

$$b_k = \{-2, 3, 6, 7, 6, 3, -2\}/21$$

and

$$H(\nu) = (7 + 12\cos \pi\nu + 6\cos 2\pi\nu - 4\cos 3\pi\nu)/21$$

[Fig. 1(d)].

It may be noticed that in all cases $H(1) \neq 0$, that there are transmissions zeros (or more accurately near zeros because of the sampled nature of digital data; true transmission zeros occur only if their frequency ν_0 is such that $2/\nu_0$ is an integer, which is relatively exceptional) at the frequency where $H(\nu)$ crosses the axis, and that there are frequency ranges where $H(\nu) < 0$, so that phase reversals will occur in these ranges. The number of zero crossings is a function of both n and N_p, and since $n = 1, 3, 5, \ldots$ and $N_p \geqslant (n+1)/2$, it is given by $N_p - (n-1)/2$. The percentage of frequencies where $H(\nu)$ is negative decreases with n but increases with N_p. Figure 2(a) shows the behavior of $H(\nu)$ vs ν for $n = 3$ and different values of N_p, while Fig. 2(b) shows the same for $n = 5$. Finally, it may also be noticed that some of the coefficients b_k are negative so that overshoots and undershoots will result since the b_k are the impulse response of the smoothing.

To show to what extent these factors may affect some data Fig. 3(a) shows a simulated data sequence and the resulting smoothed sequence for a 17-point LSP smoothing sequence of degree 3, while Fig. 3(b) shows the same for a 25-point sequence of degree 3. Figure 3(c) shows the same data

Downloaded 06 May 2009 to 132.246.170.171. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/rsi/copyright.jsp

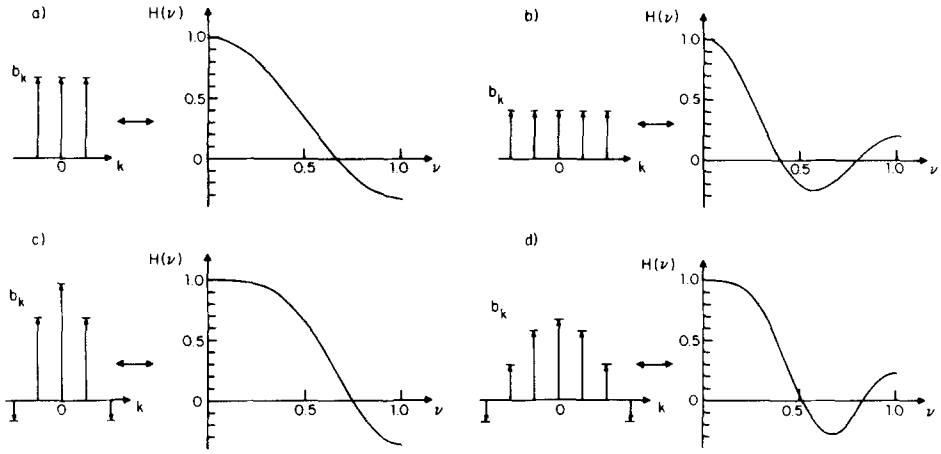

FIG. 1. (a) Filter coefficients b_k (or impulse response) and transfer function $H(\nu)$, for LSP smoothing with $n = 1$ and $N_p = 1$; (b) with $n = 1$ and $N_p = 2$; (c) with $n = 3$ and $N_p = 2$; (d) with $n = 3$ and $N_p = 3$.

smoothed by a 25-point binomial filter to be described below. The data sequence simply consists of two sine wave trains offset from the base line. The normalized frequencies are, respectively, $\nu = 0.187$ (~ 11 points per cycle) for the leftmost train and $\nu = 0.280$ (~ 7 points per cycle) for the other one. This is intended to roughly simulate series of equally spaced peaks similar to what might be encountered, for instance, in molecular vibrational or rotational spectra. It is clear in Figs. 3(a) and 3(b) that distortions of the data are introduced by the smoothings even in the absence of noise. Phase reversals at certain frequencies, as well as undershoots and overshoots, are readily noticeable. These defects may not be significant in all applications, but if for instance the smoothed curves of Fig. 3 had been obtained from noisy

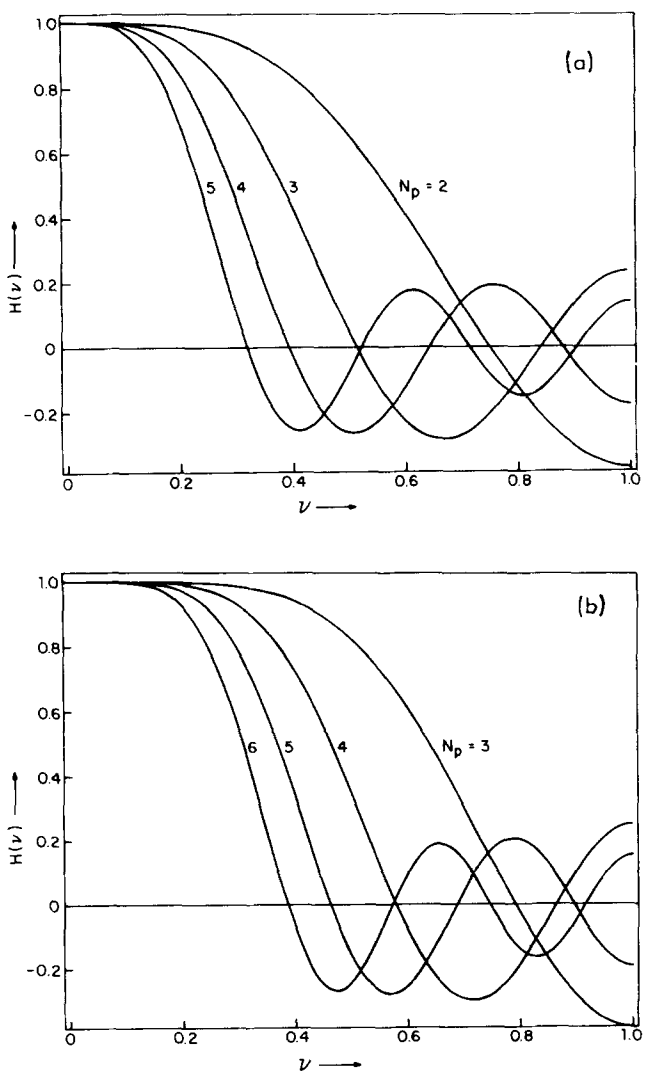

FIG. 2. (a) Transfer function $H(\nu)$ of the LSP smoothings with $n = 3$ for $N_p = 2,3,4,5$, showing the zero crossings and the regions where $H(\nu)$ is negative. (b) Same as (a) for the LSP smoothings with $n = 5$ for $N_p = 3,4,5,6$.

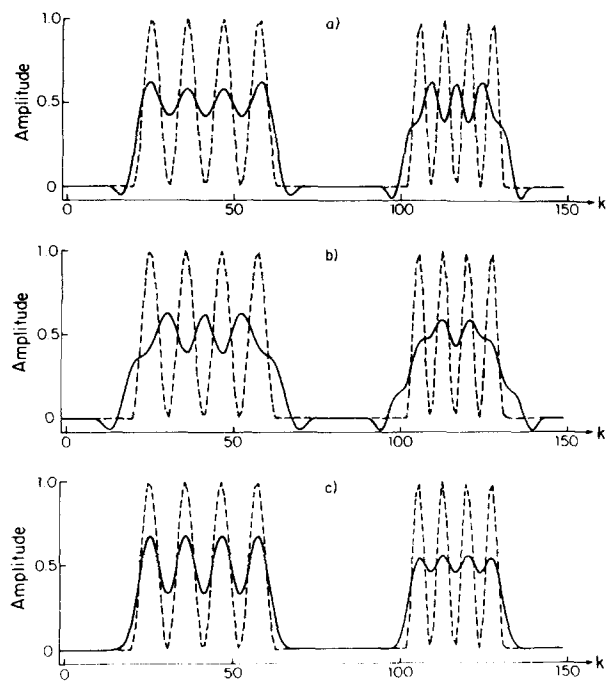

FIG. 3. (a) Hypothetical data (---) and the resulting smoothed curve (—) obtained with the 17-point third-degree LSP smoothing formula. Notice how the group on the right undergoes a 180° phase shift and severe distortions, while the one on the left, at a slightly lower frequency, is not affected except for some undershoots and an enhancement of the two outermost peaks. (b) Same data (---) and the resulting smoothed curve (—) obtained with the 25-point LSP third-degree smoothing formula. Here, it is the group on the left that is phase shifted and distorted. Notice that the group on the right is also distorted. (c) Same data (---) and the resulting smoothed curve (—) obtained with the 25-point binomial filter. Notice the absence of distortion or phase shift.

Downloaded 06 May 2009 to 132.246.170.171. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/rsi/copyright.jsp

experimental data, serious errors could occur in the interpretation of such results if the smoothing had been done to enable the determination of the positions of the peaks. The presence of overshoots may or may not be objectionable, depending on the application, but they could in some cases be mistaken for smaller peaks.

Of course, if the data have only low-frequency components so that none lie above the frequency where the transfer function becomes negative, there will be no phase reversals. The undershoots will also be much smaller. These distortions, therefore, stem mainly from the fact that a long smoothing sequence and a low-degree polynomial are used with rapidly varying data. This is an extreme example but serves to illustrate what might happen if LSP smoothing is used carelessly. Unfortunately, there is no clear-cut easily used criterion to determine the maximum safe length of an LSP smoothing sequence for a given degree of the polynomial as a function of the rate of change of the data. On the other hand, distortions of this type do not occur with the low-pass binomial filter, as exemplified in Fig. 3(c), so that no such criterion is needed.

Figure 4 shows that a 25-point LSP smoothing formula of degree 3 (which is the degree expected to perform the strongest smoothing action) applied to noisy data does not remove high-frequency noise (which is usually the primary aim of smoothing) as well as the 25-point binomial filter, because $H(1) \neq 0$.

IV. SOME REMEDIES

To reduce the effect of phase reversals, repeating a short smoothing sequence n times will yield the transfer function $H(\nu)^n$ and its negative terms will rapidly become negligible as n increases. Even small values of n will ensure $H(\nu)^n \geqslant 0$ if n is even. In this respect, this is better than using a long smoothing sequence once. However, the price paid for this is a much higher cutoff frequency, a broader transition region, and a lower slope in that region. For instance, to achieve the same cutoff frequency as the 25-point third-degree LSP smoothing formula, the 5-point third-degree formula must be applied approximately 170 times. This is naturally an extreme case and more advantageous tradeoffs between the number of points in the sequence to be repeated and the number of repetitions needed to obtain a given smoothing action are possible. However, it is not readily apparent which combination will yield the best results and this may be considered a major drawback of LSP smoothing. Whereas two applications of a $(2N_p + 1)$-point binomial smoothing are exactly equivalent to one application of the longer sequence having $[2(2N_p) + 1]$ points, this is not the case for LSP smoothing. Shorter sequences yield transfer functions that have much higher cutoff frequencies and a large number of applications is needed in order to obtain the same one as a longer sequence. Thus, shorter LSP sequences are, as it were, less "efficient," as demonstrated in Fig. 5. This figure shows that, whereas, the efficiency of the binomial filter is independent of N_p, that of LSP smoothing increases very rapidly with N_p.

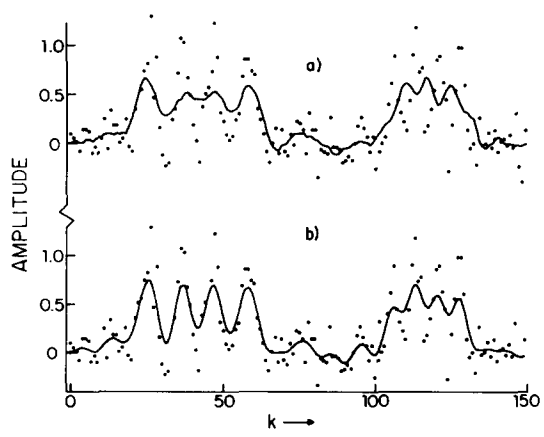

FIG. 4. (a) Same data as in Fig. 3 with added noise (· · ·) and the resulting smoothed curve (—) obtained with the 25-point third-degree LSP smoothing formula. (b) Same data (· · ·) and the resulting smoothed curve (—) obtained with the 25-point binomial filter.

The term "efficiency" used here must not be confused with the one used in the framework of optimum smoothing (see, for instance, Papoulis[11]), where it designates how well a filter maximizes the signal-to-noise ratio for the parameters to be measured. There are indeed other smoothing filters, some of which are described in Refs. 8, 9, and 11, that may be preferable to either LSP or binomial smoothing in certain applications. There are also cases not studied here where the LSP technique might prove superior to the binomial filter. One such case is that in which the data closely approximate a polynomial. Also, one drawback of the binomial filter is that it has only one parameter, so that the cutoff rate may not be adjusted independently of the cutoff frequency. LSP smoothing is more flexible in this respect since two parameters are available: the degree n of the polynomials and the length N of the smoothing sequence.

V. THE BINOMIAL SMOOTHING FILTER

In the general case, a smoothing formula should have the following properties:

(1) Zero phase shift at all frequencies unless the transfer function is negligible at frequencies where there is some phase shift. One good way to ensure this is to have $H(\nu) \geqslant 0$ at all frequencies in addition to $b_{-k} = b_k$.

(2) The sequence of smoothing coefficients b_k should be such that the smoothing introduces no undesirable side effects such as multiple peaks when only one is present in the original data, or overshoots and undershoots in the response to an impulse or to a step function. One good way to ensure this is to have $b_0 > b_1 > \ldots > b_{N_p} > 0$.

(3) Nowhere should the transfer function become > 1, especially if repeated uses of the formula on the same data

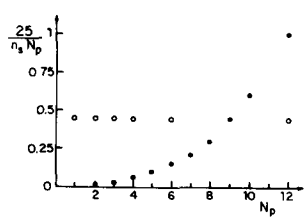

FIG. 5. Relative efficiency of LSP smoothings (○) for different values of N_p compared with that of the binomial filter (●). n_s is the number of applications of N_p-point third-degree LSP smoothings needed in order to obtain the same cutoff frequency as with the 25-point third-degree smoothing formula.

Downloaded 06 May 2009 to 132.246.170.171. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/rsi/copyright.jsp

have to be performed, since some frequency components would thereby be unduly enhanced.

Note that these are general criteria that are sufficient but not necessary in all possible cases. There may indeed be particular experimental cases where they may not apply to the smoothing algorithm which optimizes the measurement error of the desired parameters.[8,11] Still, they are applicable in a wide range of situations and, therefore, useful.

One class of filters that has the above properties plus others to be described below is the binomial filter, so called because a $(2N_p + 1)$-point smoothing sequence is defined by the binomial coefficients

$$b_k = \binom{2N_p}{N_p + k} \Big/ 4^{N_p} \quad (k = 0, 1, \ldots, N_p) \tag{9}$$

and $b_{-k} = b_k$.

For example, for $N_p = 1$, we obtain $b_{-1,0,1} = \{1, 2, 1\}/4$, for $N_p = 2, b_{-2,-1,0,1,2} = \{1, 4, 6, 4, 1\}/16$, etc. The coefficients are found using Eq. (9) or by every other row of the Pascal triangle. The transfer function for n repeated applications of the $N_p = 1$ binomial filter has been shown[12–14] to be

$$H(\nu) = \cos^{2n} \pi\nu/2 \quad (0 \leqslant \nu \leqslant 1). \tag{10}$$

Thus, the transfer function is always positive and there are no phase reversals. Figure 6 compares the response to a frequency sweep for the 17-point third-degree LSP smoothing with that of the binomial filter with $n = 8$.

We now show that the binomial smoothing formula with $2N_p + 1$ coefficients yields a maximally flat transfer function with a zero of order 2 of $H(\nu) - 1$ at $\nu = 0$ and a zero of order $2N_p - 1$ of $H(\nu)$ at $\nu = 1$. This filter is, furthermore, the maximally flat filter of order N_p with the lowest cutoff frequency and the greatest slope in the transition region (see Fig. 7). Finally, its coefficients b_k are all positive and decrease monotonically with k.

According to Hermann,[15] the maximally flat filter of order n is given by

$$P_{n,k} = (1 - x)^k \sum_{m=0}^{n-k} \binom{k + m - 1}{m} x^m, \tag{11}$$

where

$$x = (1 - \cos \pi\nu)/2 \quad (0 \leqslant \nu \leqslant 1). \tag{12}$$

There are exactly n possible values of k, for which $P_{n,k}(x)$ has a zero of order k at $x = 1$ and $P_{n,k}(x) - 1$ has a zero of order $n - k + 1$ at $x = 0$. The value of k yielding the maximally flat filter of order n having the lowest cutoff frequency and steepest slope in the transition region is $k = n$ (Fig. 7).

In this case

$$P_{n,n} = (1 - x)^n. \tag{13}$$

Substituting the value of x from Eq. (12), the transfer function $H(\nu)$ is obtained:

$$H(\nu) = [1 - (1 - \cos \pi\nu)/2]^n \tag{14}$$

$$= \cos^{2n}(\pi\nu/2) \quad (0 \leqslant \nu \leqslant 1). \tag{15}$$

This is exactly the transfer function of a three-point binomial smoothing applied n times.[12–14] $H(\nu) - 1$ has a zero of order 2 at $\nu = 0$ and $H(\nu)$ has a zero of order $2n$ at $\nu = 1$, which means that

$$H(1) = H'(1) = H''(1) = \ldots = H^{(2n-1)}(1) = 0.$$

It remains to be shown that this is identical to a $(2N_p + 1)$-coefficient binomial smoothing sequence. To do this, we show that the three-point binomial smoothing of any

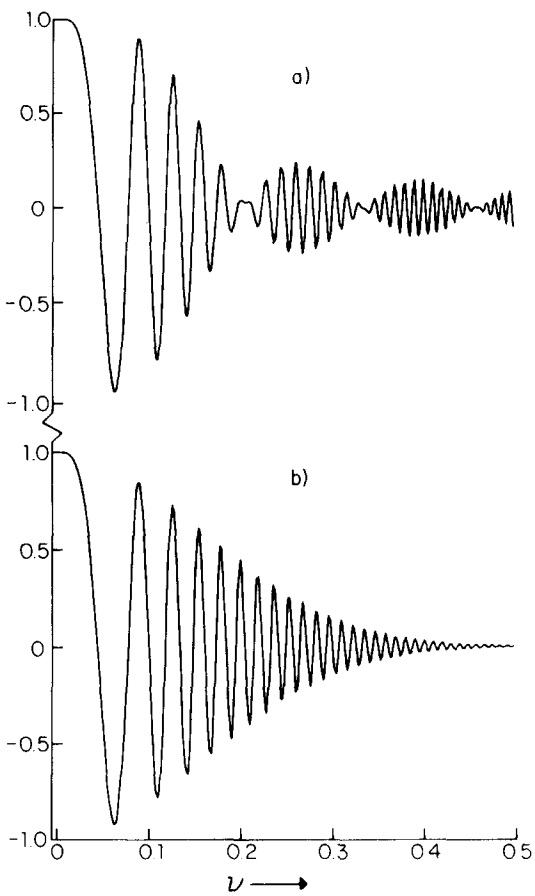

FIG. 6. Comparison of the response to a frequency sweep of (a) the third-order 17-point LSP smoothing and (b) the binomial filter with $n = 8$. Notice the transmission zeros and phase reversals in (a) and the smooth decrease in (b).

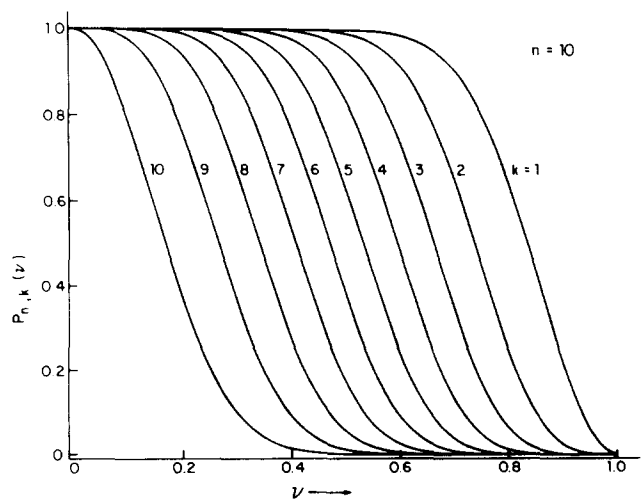

FIG. 7. Transfer functions of the ten possible maximally flat filters of order $n = 10$.

Downloaded 06 May 2009 to 132.246.170.171. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/rsi/copyright.jsp

binomial smoothing sequence of order $n = N_p$ is again a binomial smoothing but of order $n + 1$. That is, from Eq. (9):

$$\{1, 2, 1\} * \binom{2n}{m} = \binom{2n+2}{m+1} \quad (m = 0, 1, \ldots, k). \tag{16}$$

Indeed, a little algebra shows that

$$\binom{2n}{m-1} + 2\binom{2n}{m} + \binom{2n}{m+1} = \binom{2n+2}{m+1}. \tag{17}$$

Thus, a given lowest-cutoff maximally flat filter of order n may be obtained in two ways: one application of the binomial filter of order n, i.e., $2n + 1$ coefficients b_k given by Eq. (9) with $n = N_p$, or better, n applications of the three-point binomial smoothing formula. The results are identical except for end effects. One way of handling end effects is to keep the end points fixed, i.e., $y_1 = x_1$ and $y_N = x_N$. With the three-point binomial smoothing formula, this may be done directly. In the case of any linear averaging filter, there are two widely used ways of handling end effects: one is to use N_p special formulas to calculate the N_p first and last points of the smoothed data sequence to avoid losing them[2]; another is to keep the end points fixed by extending the data sequence by N_p points at both ends so that

$$x_{1-i} = x_1 - (x_i - x_1) = 2x_1 - x_i,$$

and

$$x_{N+i} = 2x_N - x_{N-i} \quad (i = 1, 2, \ldots, N_p),$$

provided that the original data sequence contains $N > N_p + 1$ points. This will ensure that $y_1 = x_1$ and $y_N = x_N$.

Binomial smoothings of any order n_i can be applied successively in any order with easily predictable properties such as half-transmission frequency and maximum slope. The resulting transfer function will always be a binomial filter of order $n = \Sigma n_i$ with the transfer function of Eq. (15):

$$\cos^{2n} \pi\nu/2.$$

The required order n for a given half-transmission frequency $\nu_{1/2}$ is easily calculated and is given by

$$n = 1/\log_2[(\cos \pi\nu_{1/2}/2)^{-2}] \quad (0 \leqslant \nu_{1/2} \leqslant 0.5), \tag{18}$$

while the half-transmission frequency for a given n is

$$\nu_{1/2} = (2/\pi)\arccos 0.5^{1/2n}. \tag{19}$$

The frequency of maximum slope is given by

$$\nu_0 = (2/\pi)\arctan(2n - 1)^{-1/2} \tag{20}$$

and the slope at that frequency is

$$H'(\nu_0) = -\pi n(2n)^{-1/2}(1 - 1/2n)^{n-1/2}. \tag{21}$$

The importance of choosing the right cutoff frequency for efficient smoothing of a signal having a given frequency content has been pointed out by Kaiser and Reed.[8] Finally, as n (or N_p, which is equivalent in the case of the binomial filter) increases, both b_k and $H(\nu)$ tend extremely rapidly toward a Gaussian shape. This is a consequence of the central limit theorem for protracted self-convolution.[10] Rapid convergence is a desirable property because the Gaussian function is very well behaved: it is always positive and decreases very rapidly and monotonically toward zero. This tendency is much weaker in the case of repeated LSP smoothings and occurs only for very large numbers n of repetitions (many hundreds). There is obviously no such tendency as a function of N_p.

VI. COMPUTATION

The $(2N_p + 1)$-point binomial filter is most effectively computed by using N_p repeated three-point binomial filters. Each three-point binomial smoothing $\{y_k\}$ of an N-point data sequence $\{x_k\}$ can be performed in only $2(N - 1)$ additions and divisions by 2 if it is done in two passes of the $\{1, 1\}/2$ smoothing as follows:

$$z_1 = (x_1 + x_2)/2, \ldots,$$
$$z_k = (x_k + x_{k+1})/2, \ldots,$$
$$z_{N-1} = (x_{N-1} + x_N)/2,$$
$$y_2 = (z_1 + z_2)/2, \ldots,$$
$$y_k = (z_{k-1} + z_k)/2, \ldots,$$
$$y_{N-1} = (z_{N-2} + z_{N-1})/2,$$

and finally $y_1 = x_1$ and $y_N = x_N$. This keeps the endpoints fixed and avoids cumulative end effects. The total number of operations for the equivalent of a $(2N_p + 1)$-point filter is thus $2(N - 1)N_p$ additions plus $2(N - 1)N_p$ divisions by 2, which are merely shifts in binary arithmetic and are performed very rapidly. This algorithm may be performed in place without additional memory except for x_1 and x_N and lends itself extremely well to very efficient machine-language programming or hardware implementation. This is much faster than using longer binomial sequences directly or LSP smoothings, where $(2N_p + 1)N$ integer multiplications, $2N_pN$ additions, and N integer divisions are needed. A further advantage of this algorithm is that the coefficients of an n-point sequence do not have to be calculated or kept in memory, which is necessarily the case with LSP smoothing. The speed difference involved is far from being academic. Depending on the word length and the architecture of the processor used, execution times differing by a factor of 100 are easily reached for sequences of the same length.

When only relatively weak smoothing is needed, computation time is usually irrelevant except in real-time applications. In some cases, however, such as in the extraction of small signals from large backgrounds to be described below, strong smoothing is often needed, sometimes repeatedly. One such situation occurs when curve fitting of filtered signals is performed, since the fitting function (and often its derivatives) must be repeatedly evaluated with different parameters. In order for a meaningful comparison with the filtered data to be achieved, the function must be filtered with the same filter. So must its derivatives if they are used in the optimization procedure.[16] The computation time thus sometimes becomes important, even with large mainframe computers, so that the computational efficiency of the algorithm is not negligible in such cases.

All the filters described in the present work may also be performed using digital Fourier transform techniques.[13,16] This may be computationally more efficient using Fast Fourier Transform (FFT) algorithms if the smoothing sequence is extremely long. It, however, requires the data se-

Downloaded 06 May 2009 to 132.246.170.171. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/rsi/copyright.jsp

quence to have a number of points that is usually a power of 2. A special technique to maintain the endpoints fixed using FFT methods has been described by Marchand and Veillette.[13]

VII. FURTHER PROPERTIES OF THE BINOMIAL FILTER

There are an infinity of possible N-point formulas that do not have a maximally flat frequency response but have the properties $H(0) = 1$, $H(1) = 0$, $H(\nu) \leqslant 1$ and have no zeros, real or complex in the Nyquist interval. However

THEOREM: Among all the possible N-point smoothing formulas, the one with the sharpest cutoff having no zeros (real or complex) in the Nyquist interval $0 \leqslant \nu \leqslant 1$ and such that $H(0) = 1$ and $H(1) = 0$, is the binomial smoothing formula.

Proof. Let $N = 2N_p + 1$ and let $n = N_p$. The theorem is true for $n = 1$ since it can easily be shown that the three-point binomial smoothing formula is the only three-point smoothing formula yielding $H(0) = 1$ and $H(1) = 0$.

In the general case,

$$H(\nu) = b_0 + 2b_1 \cos \pi\nu + 2b_2 \cos 2\pi\nu + \ldots + 2b_{N_p} \cos N_p \pi\nu.$$

Using the multiple angle formulas, a power series of order $n = N_p$ in $\cos \pi\nu$ is obtained. Using the substitution of Eq. (12), $x = (1/2)(1 - \cos \pi\nu)$, the $\nu = 0$ to 1 interval is mapped in the $x = 1$ to 0 interval. We, therefore, have a power series $P_n(x)$ of order n in x. To satisfy the conditions $H(0) = 1$ and $H(1) = 0$, $P_n(x)$ must have at least one real zero at $x = 1$ and $P_n(x) - 1$ must have at least one real zero at $x = 0$. In the case of the binomial filter, $P_{n,n}(x) = (1 - x)^n$ from Eq. (13), the slope at $x = 0$ is $-n$, and $P_{n,n}(x)$ has a zero of order n at $x = 1$. We now show that if any $P_n(x)$ has a zero of order $k < n$ at $x = 1$ and its remaining $(n - k)$ zeros are outside the Nyquist interval, the slope at $x = 0$ will have a magnitude smaller than n, i.e., $P_n(x)$ will decrease more slowly than $P_{n,n}(x)$ so that $P_{n,n}$ may be said to have a sharper cutoff than P_n. Let

$$P_{n+1}(x) = p_n(x)(z_r - x)/z_r,$$

where $p_n(x)$ has $p_n(0) = 1$ and $p_n(1) = 0$, and z_r is a real zero outside the Nyquist interval $0 \leqslant x \leqslant 1$.

Differentiating, one obtains

$$P'_{n+1}(x) = [p'_n(x)(z_r - x) - p_n(x)]/z_r,$$

and at $x = 0$,

$$P'_{n+1}(0) = p'_n(0) - p_n(0)/z_r = p'_n(0) - 1/z_r.$$

Since if z_r is positive, it is by hypothesis greater than 1, then $P'_{n+1}(0) > p'_n(0) - 1$. If two complex conjugate roots z_1 and z_2 with real parts outside the Nyquist interval are added instead, one obtains

$$P_{n+2}(x) = p_n(x)(z_1 - x)(z_2 - x)/z_1 z_2.$$

Differentiating,

$$P'_{n+2}(x) = \{p'_n(x)(z_1 - x)(z_2 - x) - p_n(x)[(z_1 - x) + (z_1 - x)]\}/z_1 z_2$$

and

$$P'_{n+2}(0) = p'_n(0) - (z_1 + z_2)/z_1 z_2.$$

Since the last term is always < 2 ($z_1 + z_2$ and $z_1 z_2$ are real because z_1 and z_2 are complex conjugate),

$$P'_{n+2} > p'_n(0) - 2.$$

Finally, since $P_1(x) = p_1(x)$, then $P'_1(0) = -1$, and $P'_2(0) > -2$ unless $P_2(x) = P_{2,2}(x)$ from Eq. (13), for which $P'_{2,2}(0) = -2$. Increasing the order of $P_n(x)$ by adding real or complex conjugate roots in this way, $P'_n(0)$ is always larger than $P'_{n,n}(0) = -n$. ■

It must be remembered that when $\cos \pi\nu = 1 - 2x$ (Eq. 12) is substituted back into Eq. (11), the derivative of $H(\nu)$ at $\nu = 0$ becomes 0, so that the sharpest cutoff mentioned in the theorem does not occur at zero frequency but at that given by Eq. (20) as illustrated in Fig. 6(b).

VIII. APPLICATIONS

The binomial filter may be applied wherever easily predictable smoothing is required. The application of smoothing to improve signal-to-noise ratio has been discussed by Kaiser and Reed[8] and Enke and Nieman.[17] It is useful in making data easier to interpret provided that proper perspective relative to measurement errors is maintained.

Some scientists also use differentiating smoothing in an attempt to compensate for the fact that differentiation reduces the signal-to-noise ratio. This can easily be performed with the binomial filter by convolving either the filter coefficients or the data by the sequence $\{1, 0, -1\}/2$. This may be repeated any number of times to obtain the second-, third-, and higher-order derivatives, after which the data are low-pass filtered in the usual manner.

A high-pass filter useful for removing slowly varying backgrounds or trends is easily obtained using the transfer function

$$H(\nu) = 1 - \cos^{2n} \pi\nu/2.$$

This may be performed by subtracting the data, low-pass filtered by the binomial filter of order n, from the original data.[12–14,16,18–19] This technique has proven very powerful and has been used extensively in this laboratory for over ten years by Marmet and co-workers and Marchand and co-workers to extract very small narrow structures from much larger nonlinear backgrounds as described in Refs. 12–14, 16, 18–19, and references given therein. This may indeed be one of most useful applications of the binomial filter. It could not be performed using LSP smoothing because severe distortion of the small narrow structures to be enhanced would be caused by overshoots and phase reversals, since the technique often needs strong smoothing. It is indeed necessary in this application that the smoothing sequence be longer than the width of the features to be extracted, which is exactly the situation in which LSP smoothing has been shown in Sec. III to introduce such distortions.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research

Downloaded 06 May 2009 to 132.246.170.171. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/rsi/copyright.jsp

Council of Canada and of the Ministère de l'Éducation du Québec.

[1]E. Whittaker and G. Robinson, *The Calculus of Observations* (Blackie, London, 1944).
[2]F. B. Hildebrand, *Introduction To Numerical Analysis* (McGraw-Hill, New York, 1956).
[3]K. L. Nielsen, *Methods in Numerical Analysis* (Macmillan, New York, 1956).
[4]P. G. Guest, *Numerical Methods of Curve Fitting* (Cambridge University, London, 1961).
[5]M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Hafner, New York, 1966).
[6]A. Savitsky and M. J. E. Golay, Anal. Chem. **36**, 1627 (1964).
[7]H. H. Madden, Anal. Chem. **50**, 1383 (1978).
[8]J. F. Kaiser and W. A. Reed, Rev. Sci. Instrum. **48**, 1447 (1977).
[9]I. J. Schoenberg, Q. Appl. Math. **4**, 45 (1946).
[10]R. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1965).
[11]A. Papoulis, *Signal Analysis* (McGraw-Hill, New York, 1977).
[12]R. Shapiro, Math. Comput. **29**, 1094 (1975).
[13]P. Marchand and P. Veillette, Can. J. Phys. **54**, 1309 (1976).
[14]H. H. Arsenault and P. Marmet, Rev. Sci. Instrum. **48**, 512 (1977).
[15]O. Hermann, IEEE Trans. Circuit Theory **CT-18**, 411 (1971). [Reprinted in L. R. Rabiner and C. M. Rader, *Digital Signal Processing* (IEEE, New York, 1972).]
[16]P. Marchand and G. Boulet, Can. J. Phys. **58**, 619 (1980).
[17]C. G. Enke and T. A. Nieman, Anal. Chem. **48**, 705 (1976).
[18]E. Bolduc, J. J. Quemener, and P. Marmet, J. Chem. Phys. **57**, 1957 (1972).
[19]R. Carbonneau, E. Bolduc, and P. Marmet, Can. J. Phys. **51**, 505 (1973).

Downloaded 06 May 2009 to 132.246.170.171. Redistribution subject to AIP license or copyright; see http://rsi.aip.org/rsi/copyright.jsp