TECHNOMETRICS, NOVEMBER 1992, VOL. 34, NO. 4

Edge-Preserving and Peak-Preserving Smoothing

Peter Hall

Centre for Mathematics and its Applications
Australian National University
Canberra, Australia

and

Division of Mathematics and Statistics
CSIRO
Sydney, Australia

D. M. Titterington

Department of Statistics
University of Glasgow
Glasgow
Scotland

An alternative procedure is developed to the smoothed linear fitting method of McDonald and Owen. The procedure is based on the detection of discontinuities by comparing, at any given position, three smooth fits. Diagnostics are used to detect discontinuities in the regression function itself (edge detection) or in its first derivative (peak detection). An application in electron microscopy is discussed.

KEY WORDS: Changepoint; Edge detection; Image processing; Nonparametric regression; Peak detection.

The objective of this article is to contribute to the methodology available for dealing with a very common statistical problem, the detection of, or accommodation of, discontinuities in regression functions when noisy observations thereof are available.

The methodology finds a niche in two existing, voluminous literatures, the statistical literature on changepoint problems and the engineering literature on edge detection. For recent discussion of these two branches of research, see Carlstein (1988) and Lee (1990). Our objective is largely to produce nonparametric estimates of discontinous regression functions, although we shall comment on the algorithm's ability to detect the points of discontinuity themselves; for another approach to the latter goal, see Wu and Chu (1990).

The scope for application of any methodology in this area is wide indeed. Lee (1990) listed, from the "engineering" point of view, the areas of "computer vision, computer graphics, signal processing, image processing, pattern recognition, geology, tomography, remote sensing, etc." (p. 321). Statistical changepoint analysis is a natural tool for quality control (Anonymous 1982). Edge detection is important in a wide variety of scientific contexts, such as the monitoring of sudden changes in texture, as well as object detection (Lee 1990). In Section 4, we shall illustrate our methodology with an example in electron microscopy.

As just remarked, we will be attempting to "recover" piecewise smooth regression functions from noisy data. The noise will be combated by smoothing, but an edge-accommodation stage will be developed to prevent discontinuities from being degraded, as would be the case with unmodified smoothing procedures.

Our aim is, therefore, the same as that of McDonald and Owen (1986), and we readily acknowledge the seminal influence of their article on our research. The fundamental key to their approach is to create, corresponding to any given point, three smoothed estimates of the function, based on data to the right, to the left, and on both sides of the point in question. These are called the right, left, and central fits and were taken by them to be linear fits. The quantity of data that contributes to a particular fit depends on the size of the so-called data window; the larger the window, the more data are included, but the less plausible might be the assumption of linearity of the function in general.

If local linearity obtains in the absence of discontinuities, then the three fits will be very similar in value except in the neighborhoods of breaks, and the disagreements among them will usually reach an extremum at a point of discontinuity. This basic feature was embedded by McDonald and Owen (1986) in a somewhat complicated algorithm for constructing a split-linear fit from an arbitrary set of data.

The objective of the present article is to derive an alternative, edge-preserving, smoothing algorithm with specific analytical properties that is less complicated to implement than is the method of McDonald and Owen (1986). Our algorithm is based loosely on kernel-type smoothing, whereas they used ordinary least

squares fitting; spline smoothing was used by Lee (1990, 1991) and others; for instance, see Laurent and Utreras (1986), Shiau, Wahba, and Johnson (1986), Shiau (1987), and Vercken and Potier (1990).

Section 1 includes the derivation of our algorithm for edge-preserving smoothing. Somewhat artificial illustrative examples are presented in Section 2. Modifications appropriate for preserving sharp peaks or troughs are discussed and illustrated in Section 3. Section 4 presents a real, illustrative example from electron microscopy, and a brief discussion concludes the article in Section 5.

1. AN ALGORITHM FOR EDGE-PRESERVING SMOOTHING

1.1 Introduction

The basic features of our algorithm are as follows:

1. At all *design points*, compute left, right, and central smooths according to the procedure to be described in Section 1.2.
2. Combine the three smooths into a diagnostic that identifies points of discontinuity; see Section 1.3 for details.
3. On the basis of feature 2, produce a single estimate of the regression function at each design point, as described in Section 1.4.

It will be assumed that the design points are equally spaced and that the observations $\{Y_i\}$ satisfy the model

$$Y_i = f(i/n) + \varepsilon_i \qquad (1.1)$$

($i = 1, \ldots, n$), where $f(\cdot)$ is a piecewise smooth function and the errors are assumed to be iid, each with variance σ^2. Note that from the point of view of the methodological development we are assuming that the design points all belong to the interval [0, 1]. This is not necessary in practice and is not the case in the illustrative examples used later, in which the design points are still equally spaced but over a different interval.

In Section 1.5, we discuss the possibility of extending the method to the case of unequally spaced design points.

1.2 The Left, Right, and Central Smooths

We shall use smooths that are linear in the data. They are local smooths in the sense that they involve only $2m + 1$ consecutive observations, for some *window size* $m \in \{0, 1, 2 \ldots\}$.

At a given design point, the central smooth is a linear combination of the observations at that point and those at the m nearest design points on either side; the left (right) smooth is a linear combination of the observation at that point and those at the $2m$ nearest design points to the left (right). To be specific, the central, right, and left smooths at design point i are defined respectively by

$$\hat{f}_c(i/n) = \sum_{j=-m}^{m} c_j Y_{i+j},$$

$$\hat{f}_r(i/n) = \sum_{j=0}^{2m} r_j Y_{i+j},$$

and

$$\hat{f}_l(i/n) = \sum_{j=-2m}^{0} l_j Y_{i+j},$$

where the c_j, r_j and l_j are constants, to be determined. In view of the equal spacing of the design points, it is appropriate to take $l_j = r_{-j}$ for all j.

In the work of McDonald and Owen (1986), the c_j, r_j, and l_j corresponded to ordinary least squares fitting. Our approach to defining the c_j and r_j is to provide equality in the leading terms of the Taylor expansions of $\mathbf{E}\{\hat{f}_c(i/n)\}$ and $\mathbf{E}\{\hat{f}_r(i/n)\}$. We shall see that this requirement leads to indeterminacy, which creates opportunities for considerable flexibility. In particular, one can propose *any* desired set of c_j and then construct a set of r_j that conform to the requirement, or vice versa.

First we develop the Taylor expansion of $\mathbf{E}\{\hat{f}_c(i/n)\}$ up to the qth order term ($q \geq 1$). Provided that there is sufficient regularity,

$$\begin{aligned}
\mathbf{E}\hat{f}_c(i/n) &= \sum_{j=-m}^{m} c_j f\{(i + j)/n\} \\
&= \sum_{j=-m}^{m} c_j \left\{ \sum_{k=0}^{q} (k!)^{-1} n^{-k_j k} f^{(k)}(i/n) \right. \\
&\qquad \left. + 0(n^{-(q+1)j(q+1)}) \right\} \\
&= \sum_{k=0}^{q} \left\{ (k!)^{-1} n^{-k} \right. \\
&\qquad \left. \times \left(\sum_{j=-m}^{m} j^k c_j \right) \right\} f^{(k)}(i/n) \\
&\qquad + 0\left(n^{-(q+1)} \sum_{j=-m}^{m} j^{q+1} c_j \right). \qquad (1.2)
\end{aligned}$$

The expansion for $\mathbf{E}\{\hat{f}_r(i/n)\}$ is quite analogous, so the c_j and r_j should be chosen to satisfy

$$\sum_{j=0}^{2m} j^k r_j = \sum_{j=-m}^{m} j^k c_j, \qquad k = 0, \ldots, q. \qquad (1.3)$$

To resolve the remaining indeterminacy, our approach is, for prescribed q, to propose a set of c_j and then determine the r_j from (1.3). This may be done

as follows. Define

$$s_k = \sum_{j=0}^{2m} j^k, \qquad k = 0, \ldots, 2q.$$

Construct the $(q + 1) \times (q + 1)$ matrix S defined by

$$S = \begin{bmatrix} s_0 & s_1 & \cdots & s_q \\ s_1 & s_2 & \cdots & s_{q+1} \\ \vdots & \vdots & & \vdots \\ s_q & s_{q+1} & \cdots & s_{2q} \end{bmatrix};$$

that is, $(S)_{kw} = s_{k+w}$, for $k, w = 0, \ldots, q$.

Next, compute the $(q + 1)$-dimensional vectors $\mathbf{u} = (u_0, \ldots, u_q)^T$ and $\mathbf{v} = (v_0, \ldots, v_q)^T$, where

$$v_k = \sum_{j=-m}^{m} j^k c_j, \qquad k = 0, \ldots, q,$$

and

$$\mathbf{u} = S^{-1}\mathbf{v}. \tag{1.4}$$

Finally, obtain

$$r_j = \sum_{w=0}^{q} u_w j^w, \qquad j = 0, \ldots, 2m,$$
$$= 0 \qquad \text{otherwise.} \tag{1.5}$$

To verify that (1.3) holds, as required, write (1.4) as $S\mathbf{u} = \mathbf{v}$. Identification of the kth elements of either side gives

$$\sum_{w=0}^{q} s_{k+w} u_w = v_k = \sum_{j=-m}^{m} j^k c_j \tag{1.6}$$

$(k = 0, \ldots, q)$. The left side of (1.6) is

$$\sum_{w=0}^{q} \sum_{j=0}^{2m} j^{k+w} u_w = \sum_{j=0}^{2m} j^k \left[\sum_{w=0}^{q} j^w u_w \right] = \sum_{j=0}^{2m} j^k r_j.$$

We illustrate the method for the case in which the central smooth is the simple $(2m + 1)$-point moving average and $q \leq 3$. Thus

$$c_j = (2m + 1)^{-1} \qquad \text{if } |j| \leq m$$
$$= 0 \qquad \text{otherwise,}$$

and

$$\sum_{j=-m}^{m} j^k c_j = 1 \qquad \text{if } k = 0$$
$$= 0 \qquad \text{if } k = 1$$
$$= m(m + 1)/3 \qquad \text{if } k = 2$$
$$= 0 \qquad \text{if } k = 3. \tag{1.7}$$

Expansion (1.2) then takes the form

$$\sum_{j=-m}^{m} c_j f\{(i + j)/n\} = f(i/n) + \frac{1}{2} f''(i/n) \times \frac{1}{3} m(m + 1)n^{-2} + 0\{(m/n)^4\}.$$

In our numerical work, we used $q = 1$, so that $\mathbf{v} = (1, 0)^T$, and it turned out that $m = 20$ gave satisfactory performance. For this combination, $u_0 = .0941$, $u_1 = -.0035$, and the resulting r_j are linear: $r_j = (27 - j)/287$ $(j = 0, \ldots, 40)$.

Note that approximately one-third of the r_j are negative and that this will show up in the edge-detection strategy that we discuss in Section 1.3.

For central smooths other than the moving average, one simply uses the corresponding c_j in (1.7) and proceeds as previously.

1.3 Identification of Points of Discontinuity

Breakpoints are sought by studying diagnostics based on comparisons among the left, right, and central smoothers. If there is no discontinuity and the regression function is locally linear, then the three smooths should be very similar.

There are various possibilities so far as diagnostics are concerned, as can be seen by considering Figure 1. Figure 1(a) depicts a discontinuous regression function, along with schematic versions of the three smooths, drawn in the form of their continuous versions. The smooths are drawn as if from noise-free data. Note that to the left of the breakpoint f and $\hat{f}_l$ coincide, whereas to the right f and $\hat{f}_r$ are identical.

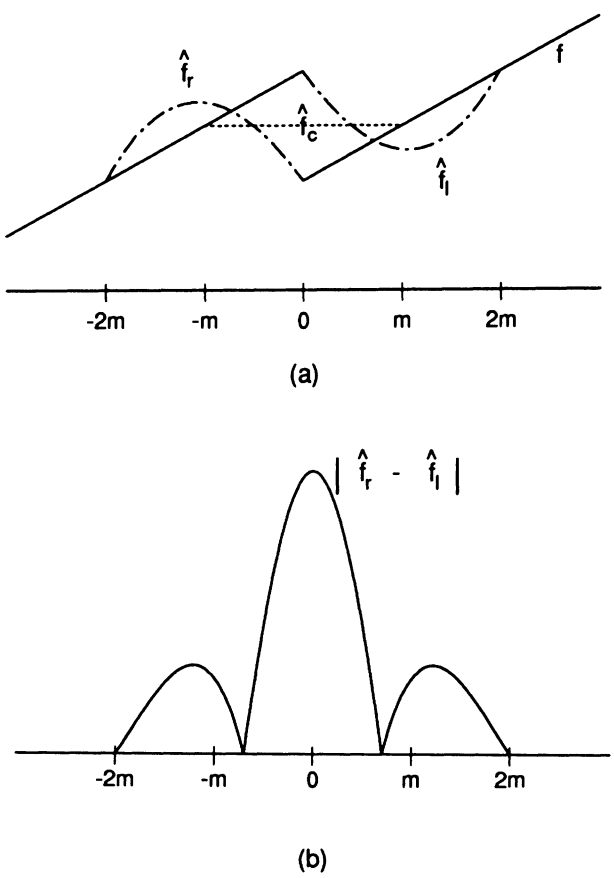

Figure 1. (a) Discontinuous Regression Function (———) with $\hat{f}_l$, (–·–·–·), $\hat{f}_c$ (·····), and $\hat{f}_r$ (– – –); (b) $|\hat{f}_l - \hat{f}_r|$ From (a).

The ways in which f, $\hat{f}_r$, and $\hat{f}_l$ differ are dictated by the form of the r_j (and therefore of the l_j) described at the end of Section 1.2. Various other qualitative statements can be made that motivate possible diagnostics:

1. At the breakpoint, $\hat{f}_c$ lies between $\hat{f}_l$ and $\hat{f}_r$.
2. At the breakpoint, $|\hat{f}_l - \hat{f}_r|$ is locally maximal, and generally $|\hat{f}_l - \hat{f}_r|$ takes the form shown in Figure 1(b).
3. Between m and $2m$ units of distance to the left of the breakpoint, $|\hat{f}_l - \hat{f}_c|$ is 0, but $\hat{f}_r$ is different from them both.
4. An obvious, corresponding version of statement 3 holds to the right of the breakpoint.

In the examples discussed in Section 2 we experimented with various diagnostics based on 1–4 and eventually settled on the following procedure for *preliminary* identification of a breakpoint: Flag position is i if, at i, (a) $(\hat{f}_r - \hat{f}_c)(\hat{f}_c - \hat{f}_l) > 0$ and (b) $|(\hat{f}_r - \hat{f}_l)(i)| > |(\hat{f}_r - \hat{f}_l)(i + j)|$ for all $j = \pm m, (m + 1), \ldots, 2m$, where, for instance, $\hat{f}_r(i)$ denotes the right smooth at point i.

In view of Figure 1, this procedure is not enough, because clearly if i is flagged then so are numerous neighboring points. There may also be many other spurious flags. Two further steps are taken: First, only "significant" values of a diagnostic are flagged, and second, only one flag is retained in any given data window. We therefore supplement the preceding (a) and (b) with (c) and (d), as follows: (c) Retain i as a flagged position only if $U(i) > u$, where $U = |(\hat{f}_l - \hat{f}_r)(\hat{f}_c - \hat{f}_r)(\hat{f}_c - \hat{f}_l)|$ and u is a critical value, determined in the Appendix and (d) if as the data are scanned from left to right i is flagged, then look ahead locally and identify the position at which the corresponding U is maximum. Regard that point as identifying a breakpoint, and disregard any other flagged point in the data window.

Clearly, variations on (c) exist. For instance, one could consider (c*), based on $U^*(i) = |(\hat{f}_l - \hat{f}_r)(i)|$ with a corresponding critical value u^*.

1.4 Production of the Final Estimate of the Regression Function

Having identified breakpoints, we estimate the regression function by $\hat{f}$ as follows:

1. For i within $2m$ units to the left of a breakpoint, take $\hat{f}(i) = \hat{f}_l(i)$.
2. For i within $2m$ units to the right of a breakpoint, take $\hat{f}(i) = \hat{f}_r(i)$.
3. For all other i, take $\hat{f}(i) = \hat{f}_c(i)$.

This too could be varied. In particular, for i between $2m$ and m units to the left (right) of a breakpoint, one could take $\hat{f}(i) = \frac{1}{2}\{\hat{f}_l(i) + \hat{f}_c(i)\}[\frac{1}{2}\{\hat{f}_r(i) + \hat{f}_c(i)\}]$ and, instead of 3, one could suggest:

3′. For all other i, take $\hat{f}(i) = \frac{1}{3}\{\hat{f}_l(i) + \hat{f}_c(i) + \hat{f}_r(i)\}$.

In our numerical work, we used 1–3.

McDonald and Owen (1986) created a final smooth in the form of a weighted average of the left, right, and central smooths corresponding to one or more window sizes. The weights were somewhat arbitrarily, if sensibly, defined, and the window sizes were chosen with a view to the scale of structure they wished to detect in the data. We effectively use a degenerate form of their technique when combining the smooths, and we only use a single window size "somewhat arbitrarily, if sensibly, defined"! One could clearly experiment with their type of approach although our somewhat simpler procedure seems to work quite well.

Formally, the technique provides estimates of the regression function only at the design points. Any preferred interpolation method can be applied to "join up" the points to create a continuous curve. This appears to have been done in the figures of McDonald and Owen (1986), whose technique also, formally, only deals with the design points.

The procedure does not allow for breaks that are less than $2m$ units apart. In some examples this may lead to degradation of edges if m is too large, and experimentation with a variety of values for m may then be advisable. In the examples, we also decline to estimate the regression function at the boundaries of the design space and thereby may miss edges there. In principle, one could modify the procedure to deal with the boundaries along the lines of McDonald and Owen (1986).

1.5 The Case of Unequally Spaced Design Points

The crucial calculations in Section 1.2 were carried out only for the case in which the design points are equally spaced. There is no difficulty in principle, however, in extending the argument to the case of less regularly spaced design points. In that case, of course, one would require different sets of weights c_j, r_j, and l_j at *each* design point, and it would no longer be so sensible to take $l_j = r_{-j}$. For instance, at design point x_i, one would wish to specify

$$\hat{f}_c(x_i) = \sum_{j=-m}^{m} c_{ij} Y_{i+j}.$$

The core of the argument in Section 1.2 was to create, via Taylor expansions of two smooths, a set of matching conditions (1.3). Having prescribed one

set of weights, we used (1.4) and (1.5) to determine values for the other set of weights so as to satisfy the matching conditions. To show how the same technique can be applied in a general context, consider the following two smooths at design point x_i:

$$\hat{f}_g(x_i) = \sum_{j=m_1}^{m_2} g_j Y_{i+j}$$

and

$$\hat{f}_h(x_i) = \sum_{j=m_3}^{m_4} h_j Y_{i+j},$$

omitting the i subscript on g_j and h_j for clarity.

The matching conditions created by the Taylor expansion argument are

$$\sum_{j=m_1}^{m_2} \delta_{ij}^k g_j = \sum_{j=m_3}^{m_4} \delta_{ij}^k h_j, \qquad k = 0, \ldots, q, \quad (1.8)$$

where $\delta_{ij} = (x_j - x_i)$ for all relevant i and j. Equation (1.8) corresponds to (1.3).

Suppose that we prescribe the h_j and try to compute g_j to satisfy (1.8). We define

$$s_k = \sum_{j=m_1}^{m_2} \delta_{ij}^k, \qquad k = 0, \ldots, 2q,$$

define S as before by $(S)_{kw} = s_{k+w}$ for $k(w = 0, \ldots, q)$, and define

$$v_k = \sum_{j=m_3}^{m_4} \delta_{ij}^k h_j, \qquad k = 0, \ldots, q.$$

We construct $\mathbf{u} = S^{-1}\mathbf{v}$ as before and calculate

$$g_j = \sum_{w=0}^{q} u_w \delta_{ij}^w, \qquad j = m_1, \ldots, m_2.$$

The short argument following (1.5) verifies that (1.8) holds for this set of g_j.

2. EXAMPLES OF EDGE DETECTION

In our simulation study of the procedure described in Section 1, we consider two examples. The second is the sawtooth function used by McDonald and Owen (1986). First, however, we consider a less regular, but still piecewise linear, function designed to exhibit more than one type of break.

Example 2.1. The solid line in Figure 2 depicts the true regression function, f, used in this example. Three breakpoints are present, each corresponding to a different type of change in f. All are *zeroth-order* discontinuities, but there are qualitative differences in the way the slope changes through the

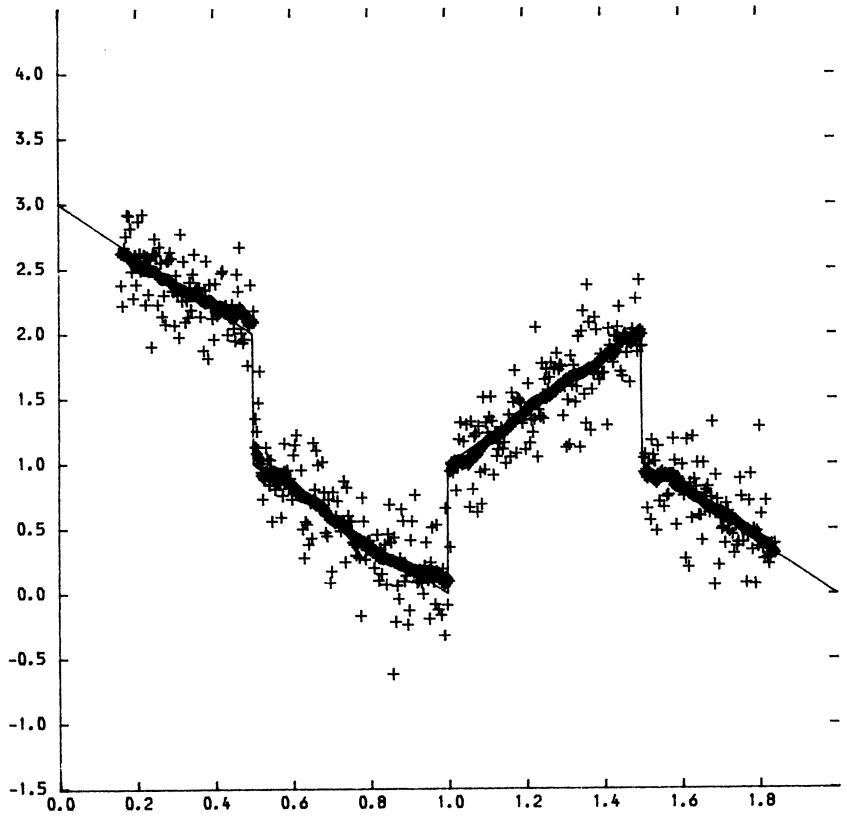

Figure 2. A Single Realization and Piecewise-Smooth Fit for Example 2.1: The True Regression Function Is Denoted by ———, the Data Are Denoted by (+), and the Fitted Values Are Denoted by (*).

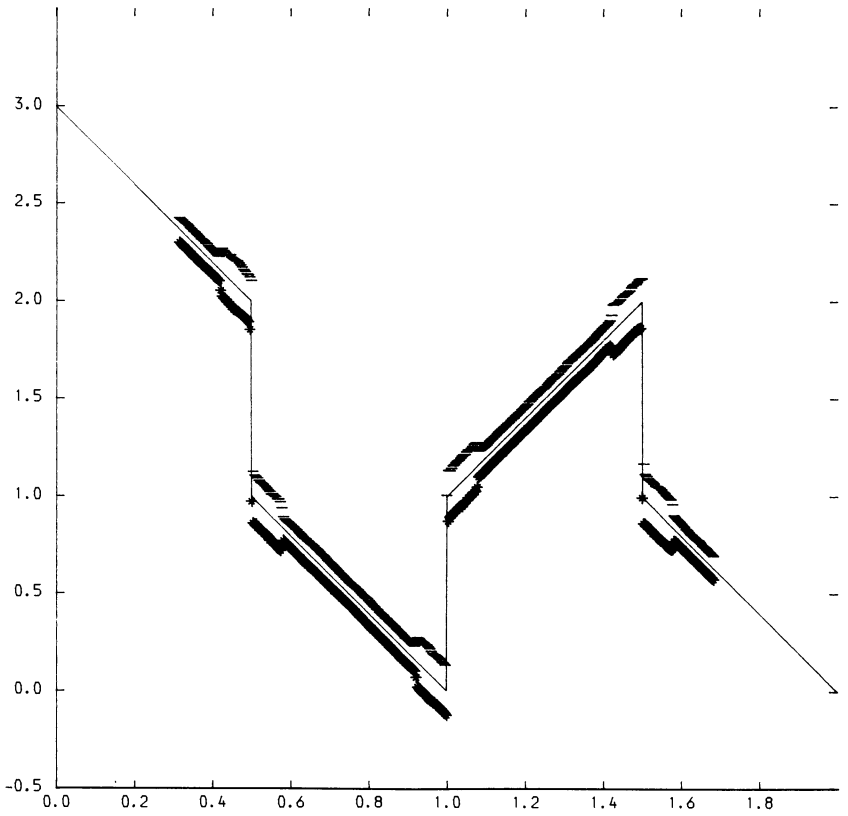

Figure 3. 5% and 95% Envelopes of $\hat{f}$ From 1,000 Realizations of Example 2.1: Lower Envelope Is (*); Upper Envelope is (−).

point of discontinuity. The number of data points was $n = 512$. Experiments were carried out using the methodology described previously for various choices of m and for several values of σ, the standard deviation of the observation noise. The results quoted here are for $\sigma = .25$, which corresponds to a reasonably demanding signal-to-noise ratio, similar to that used by McDonald and Owen (1986) in the context of their sawtooth function.

Also shown in Figure 2 are the data corresponding to a single realization and the piecewise smooth fit, as determined by the algorithm defined by (a)–(d) in Section 1.2 along with 1–3 in Section 1.3.

In Figure 3, the results of 1,000 replications of the fitting procedure are summarized in the form of the 5% and 95% envelopes of the values of $\hat{f}$. The values of σ and m are the same as in Figure 2.

The slight widening of the bands near the breakpoints is the penalty for having to use $\hat{f}_r$ and $\hat{f}_l$ there rather than $\hat{f}_c$. It is not caused by any meaningful inadequacy in the breakfinding part of the algorithm. This is made clear by Table 1, which indicates the frequencies with which discontinuities are identified by the 1,000 replications. Recall that the true breakpoints are at points 128, 256, and 384.

Example 2.2. The methodology of Section 1 was also applied to the sawtooth function of McDonald and Owen (1986). Figure 4, by analogy with Figure 2, shows the true function, along with the data and the fitted function corresponding to one realization, and Figure 5 provides the 5% and 95% envelopes of

Table 1. Breakpoint Identification Frequencies

i	*Frequencies*
125	1
126	2
127	37
*128	922
129	33
130	4
131	0
132	1
	1,000
255	11
*256	908
257	64
258	16
259	1
	1,000
382	1
383	8
*384	927
385	55
386	6
387	3
	1,000

*Location of true breakpoint.

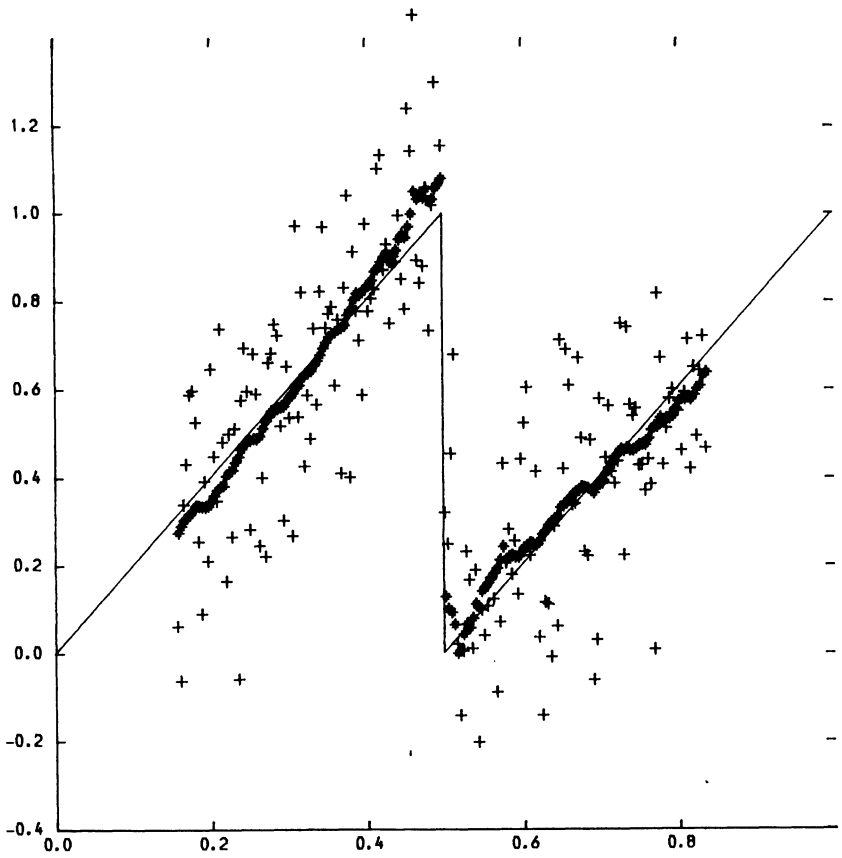

Figure 4. A Single Realization and Piecewise-Smooth Fit for Example 2.2: The True Regression Function Is Denoted by ———, the Data Are Denoted by (+), and the Fitted Values by (*).

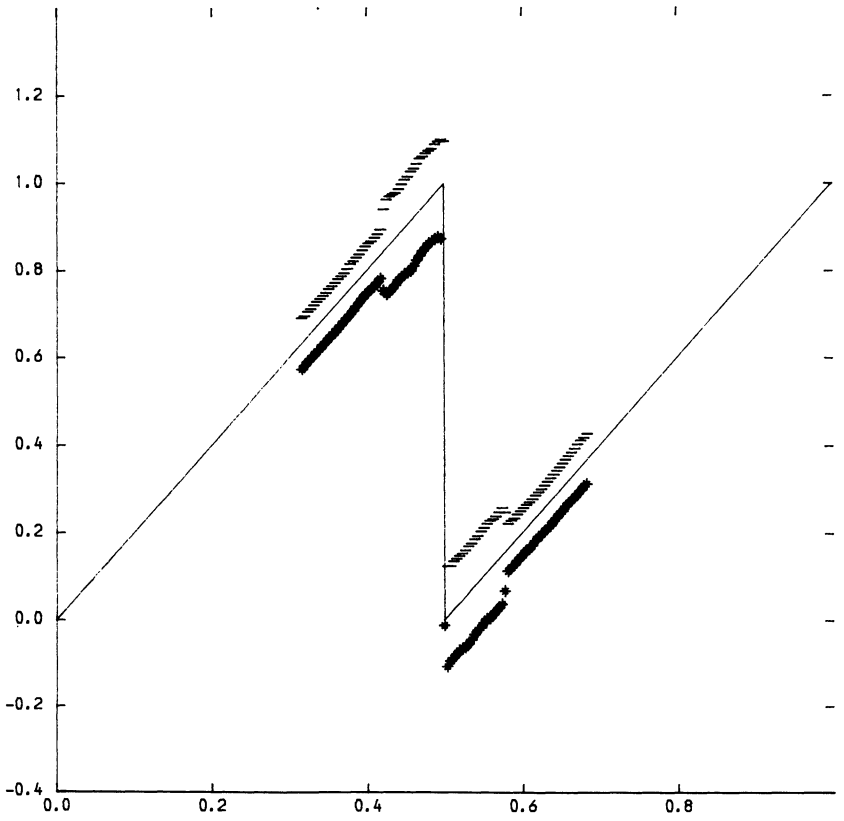

Figure 5. 5% and 95% Envelopes of $\hat{f}$ From 1,000 Realizations of Example 2.2: Lower Envelope Is (*); Upper Envelope is (−).

the fitted function from 1,000 realizations; compare with Figure 3.

3. PEAK IDENTIFICATION

3.1 Methodology

So far, we have discussed the development of smoothing procedures that cope with discontinuity in functions, called zeroth-order discontinuities by Lee (1990). It is natural, therefore, to seek adaptations to identify higher-order discontinuities. We shall restrict attention to the first-order case, in which we envisage abrupt changes in the first derivative of the regression function. In less technical terms, we are attempting to identify and adapt to peaks (or troughs) in f. Once again, we base our methodology on differences among $\hat{f}_c$, $\hat{f}_l$, $\hat{f}_r$, and the true f.

As before, it is helpful to consider the noise-free version, depicted in Figure 6, for a peak. Figure 6(a) shows the three smooths and the true f, whereas Figure 6, (b) and (c), shows the two diagnostic functions $|\hat{f}_r + \hat{f}_l - 2\hat{f}_c|$ and $|\hat{f}_r - \hat{f}_l|$.

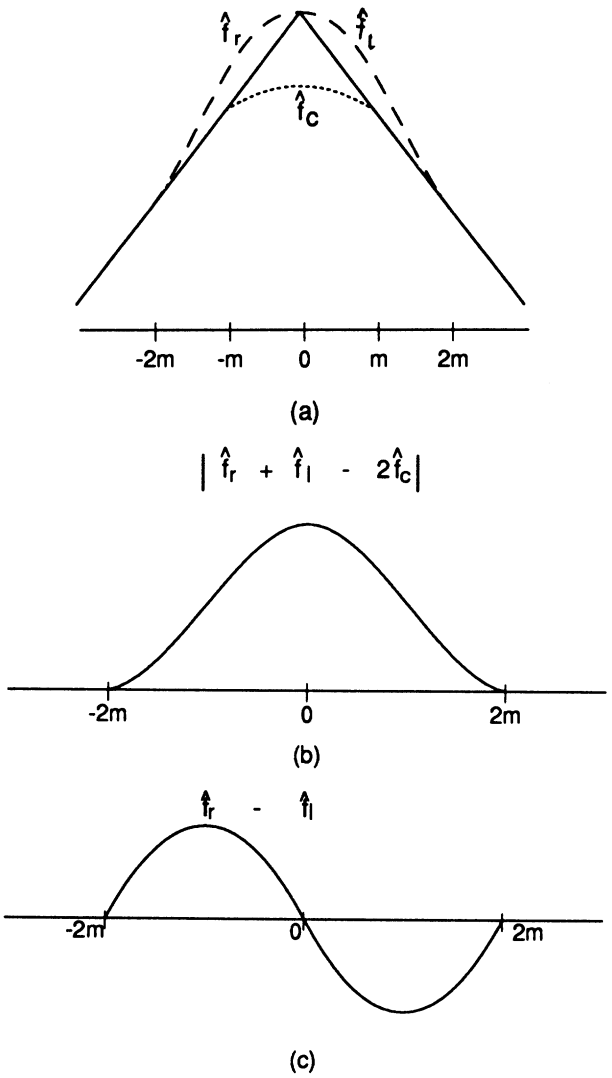

Figure 6. (a) Sharply Peaked Function (———) with $\hat{f}_l$(– – –), $\hat{f}_c$ (·····), and $\hat{f}_r$ (– – – –), (b) $|\hat{f}_r + \hat{f}_l - 2\hat{f}_c|$ From (a); (c) $(\hat{f}_r - \hat{f}_l)$ From (a).

The only change in the algorithm defined by (a)–(d) of Section 1.2 and 1–3 of Section 1.3 pertains to (c). Various diagnostics are available as replacements for U. Two possibilities are $U_1(i) = |(\hat{f}_r + \hat{f}_l - 2\hat{f}_c)(i)|$ and $U_2(i) = |(\hat{f}_r + \hat{f}_l - 2\hat{f}_c)(i)(\hat{f}_r - \hat{f}_c)(i)(\hat{f}_l - \hat{f}_c)(i)|$.

The numerical work summarized next was based on use of U_1. Critical values for both U_1 and U_2 are discussed in the Appendix.

3.2 Example 3.1

The underlying regression function selected was the simple function $f(x) = -c|x|$. As in Section 2, we took $m = 20$ (after some experimentation) and $\sigma = .25$. The design space was $[-2, 2]$, and 256 equally spaced observations were taken. The value of c dictates the strength of the discontinuity in the first derivative, which should be more accurately detected for larger values of c. Table 2 illustrates this by listing the frequencies of discontinuity identification from 1,000 realizations and for various values of c. Even for $c = 4$, the ability to identify the correct point ($i = 128$) is not overwhelmingly impressive. This is symptomatic of the general fact that peaks

Table 2. Peak Identification Frequencies From 1,000 Realizations

	Frequencies		
i	$c = 2$	$c = 3$	$c = 4$
116	1		
117	1		
118	2		
119	9	2	
120	11	3	
121	23	5	
122	22	13	4
123	39	28	16
124	72	68	54
125	95	87	82
126	104	121	125
127	134	184	233
*128	135	174	213
129	94	113	126
130	78	81	80
131	60	54	40
132	48	38	19
133	29	14	6
134	20	11	2
135	7	3	
136	9	1	
137	4		
138	2		
139	1		

*Location of true peak.

are harder to identify than are breaks—the difficulty of identifying an rth order discontinuity increases with increasing r.

Figure 7 shows, for one realization with $c = 2$, the true function, the data, and the fitted function. Figure 8 depicts the now-familar 5% and 95% envelopes from 1,000 realizations, also corresponding to $c = 2$. The fit is good everywhere, even at the peak, although there is slight flattening, associated with the imprecision identified in Table 2.

4. AN EXAMPLE INVOLVING TRANSMISSION–ELECTRON MICROSCOPY

In this section, we apply our methodology to data from the field of transmission–electron microscopy. In truth, our one-dimensional techniques will provide only a partial treatment of what is essentially a two-dimensional problem. Figure 9 displays data corresponding to a single line of 512 pixels extracted from a rectangular, pixellated image of a continuous thin film of permalloy. The measured quantity is the strength of vertical induction, measured in microradians, assessed in the so-called differential phase contrast imaging mode; for instance, see Chapman, McVitie, and Hefferman (1991).

The sample contains two domains, and it is of interest to locate the domain wall, which in one dimension amounts to finding an edge. Within each domain, there is considerable variability, caused by random noise and, in addition, by elastic scattering from the individual crystallites within the material. This latter is apparent in Figure 9 in the form of the isolated peaks and troughs on either side of the domain wall, and it is important for the eventual smoothed fit not to be dominated by the effects of the crystallites. One could, of course, try to model this effect, but it turns out that a reasonable analysis can be achieved without this extra complication.

Along with the data, Figure 9 also displays the result of applying our methodology. Figure 10 shows the corresponding left, right, and central fits. From this it is clear that, apart from near the domain wall, the central fit is retained.

To achieve the smoothed fit in Figure 9, it was necessary to modify the procedure slightly by recognizing the different levels of variability within the two domains. Examination of residuals about the central fit within the body of the two domains led to the realization that the standard deviation of the data in the left domain was approximately $4^{1/3}$ times that in the right half. As a result, the critical u used in the left half was taken to be four times that in the right half. Experimentation showed that some recognition had to be made of heteroscedasticity. If a

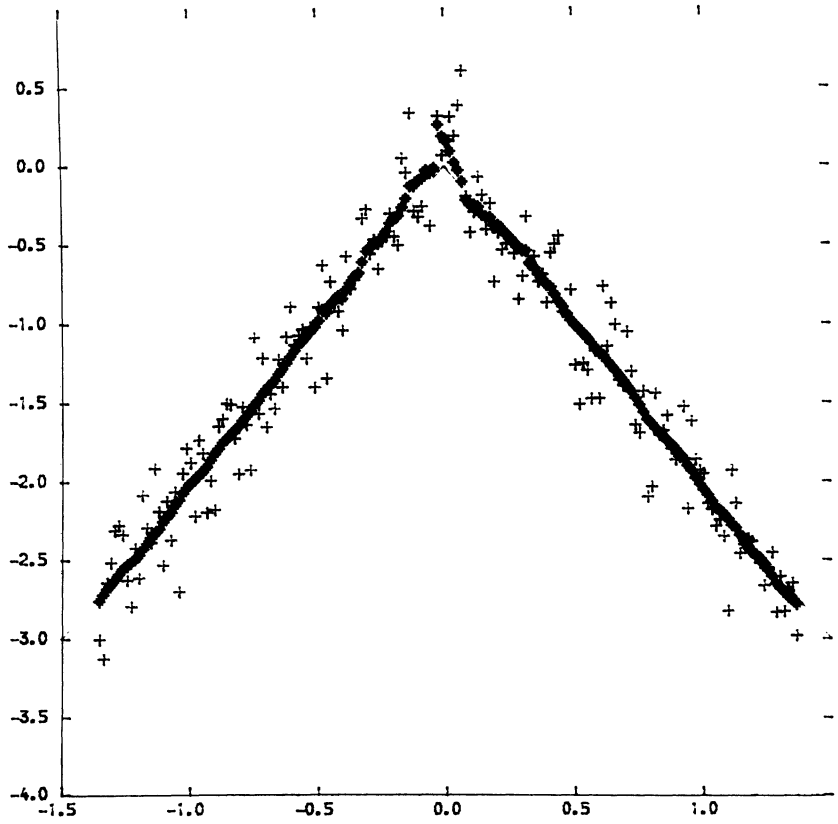

Figure 7. A Single Realization and Piecewise-Smooth Fit for Example 3.1: The True Regression Function Is Denoted by ———, the Data are Denoted by (+), and the Fitted Values Are Denoted by (*).

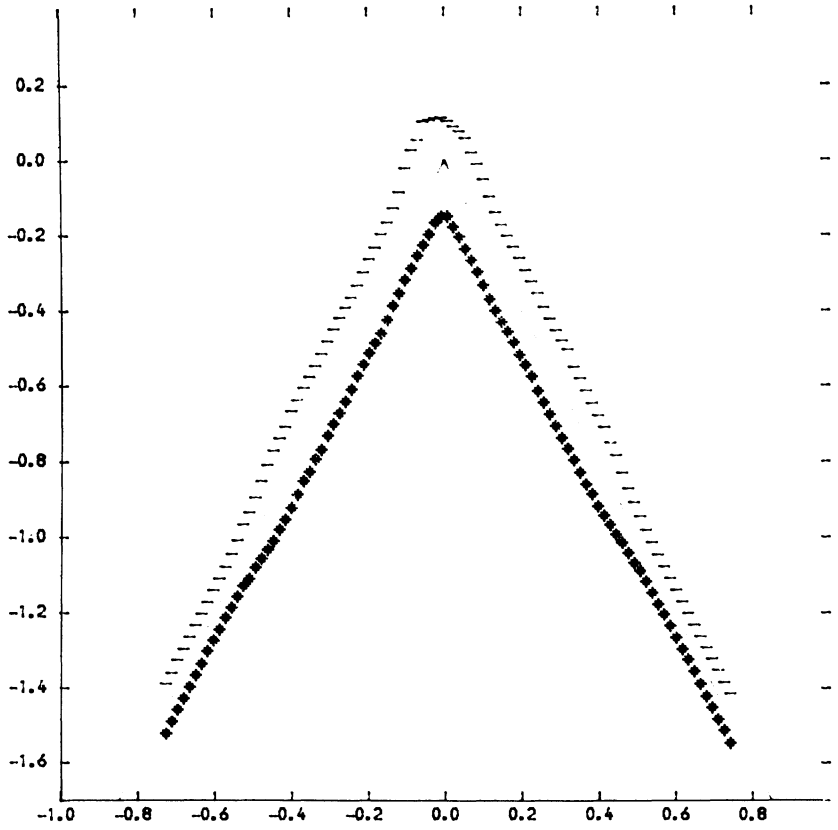

Figure 8. 5% and 95% Envelopes of $\hat{f}$ from 1,000 Realizations of Example 3.1: Lower Envelope Is (*); Upper Envelope is (−).

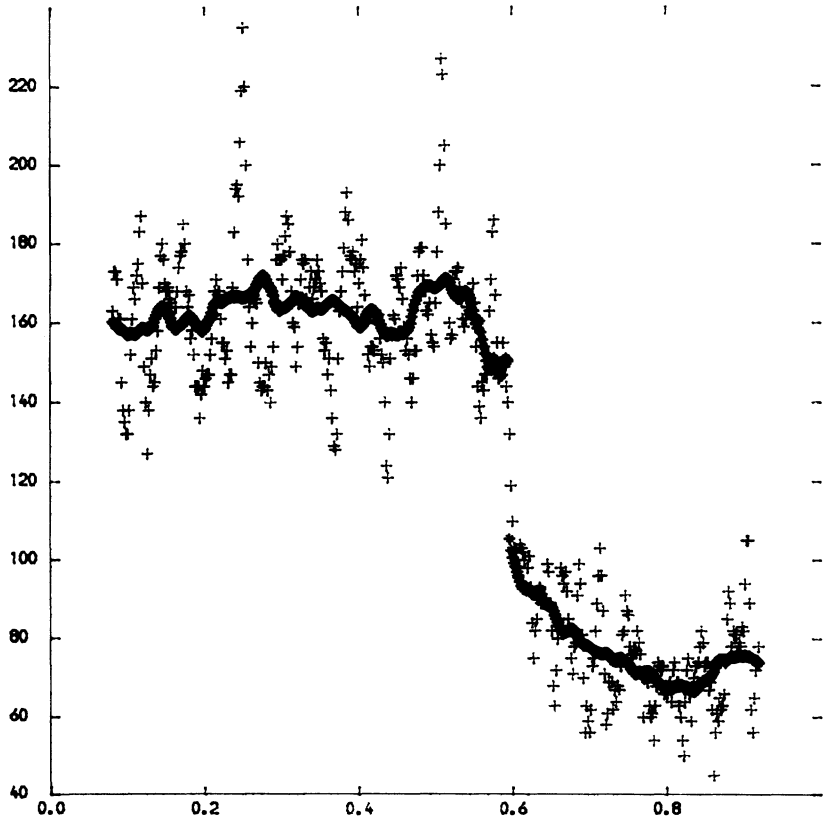

Figure 9. Data and Piecewise-Smooth Fit for Electron-Microscopy Example: The data are denoted by (+) and the fitted values by (*).

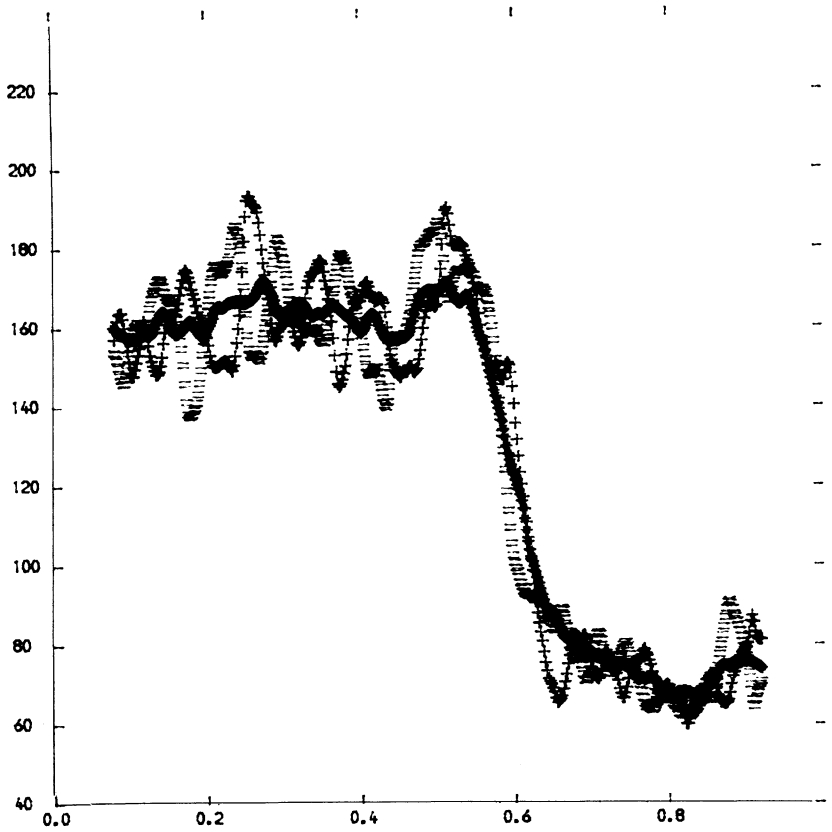

Figure 10. Left (+), Right (−), and Central (*) Smooths for Electron-Microscopy Data.

constant value of u was employed, then either strange effects occurred in the left half, resulting from the crystallites, or the central smooth was adopted across the domain wall, thereby failing to pick up the edge.

5. DISCUSSION

We have developed and illustrated simple-minded approaches to edge-preserving and peak-preserving smoothing. The former aim, in particular, appears to have been satisfactorily achieved. Directions for further research include the combination of the two types of diagnostic into a coordinated program and the more detailed investigation of similar methods in the case of nonequally spaced data and in surface-fitting problems, in which the design points are two-dimensional. This will afford more satisfactory processing of images such as the electron-microscopy example described in Section 4.

There is clearly a degree of arbitrariness in the procedure as reported here in terms of the choice of window size, the choice of the c_j, the choice of threshold for identifying points of discontinuity, and so on. No doubt, theorems could be proved related to some of these aspects in particular contexts, but we perceive the technique as a flexible data-analytic tool, within which the scope for adjustment of certain features allows for informative exploration of a data set, as in the case with many alternative approaches such as that of McDonald and Owen (1986).

ACKNOWLEDGMENTS

This research was made possible by a Visiting Fellowship Research Grant for Peter Hall awarded by the U.K. Science and Engineering Research Council. We are most grateful to J. N. Chapman, Department of Physics and Astronomy, University of Glasgow, for providing the data analyzed in Section 4 and to W. Qian for preparing some of the figures. We are also appreciative of the very helpful comments of the associate editor and the referees.

APPENDIX: CRITICAL VALUES FOR U, U_1, AND U_2

1. *The case of* $U = |(\hat{f}_l - \hat{f}_r)(\hat{f}_c - \hat{f}_r)(\hat{f}_c - \hat{f}_l)|$. Under the assumption that the true regression mean f is linear and the errors ε_i are normal $N(0, \sigma^2)$, $\hat{f}_c - \hat{f}_r$ and $\hat{f}_c - \hat{f}_l$ are both normal $N(0, \tau^2)$, where

$$\tau^2 = \sigma^2\left\{\sum_{j=0}^{2m} r_j^2 - 2(2m+1)^{-1}\sum_{j=0}^{m} r_j + (2m+1)^{-1}\right\}.$$

Given $0 < p < 1$, let $z = z_p > 0$ be the solution of $4\Phi(z)\{1 - \Phi(z)\} = p$, where Φ is the standard normal distribution function. (Thus, $z = \Phi^{-1}[1/2\{1 + (1 - p)^{1/2}\}]$.) Put $u = 2(\tau z)^3$. We claim that $\Pr(U \geq u) \leq p$. Thus u forms a convenient critical value for assessment of U. In practice, the value of σ^2 can be estimated from parts of the data where the regression mean is smooth.

To prove that $\Pr(U \geq u) \leq p$, first define $X_1 = (\hat{f}_c - \hat{f}_r)/\tau$ and $X_2 = (\hat{f}_c - \hat{f}_l)/\tau$. Since $U \geq 2 \max(|\hat{f}_c - \hat{f}_r|^3, |\hat{f}_c - \hat{f}_l|^3)$ then $\Pr(U > u) \leq \Pr\{\max(|X_1|, |X_2|) > z\} = 1 - \Pr(|X_1| < z, |X_2| < z)$. The last-written probability is maximized (over all possible correlations of X_1, X_2) when X_1 and X_2 are independent, as may be shown by writing the probability in terms of an integral with respect to a bivariate normal measure and looking for a turning point with respect to the correlation. Therefore,

$$\Pr(U > u) \leq 1 - \{2\Phi(z) - 1\}^2 = 4\Phi(z)\{1 - \Phi(z)\} = p. \quad \text{(A.1)}$$

Since we are using U as a diagnostic simultaneously over the design space, it is appropriate to use, as u, a value somewhat larger than the value defined by (A.1) for a prescribed p. For $p = .05$, (A.1) gives $z = 2.24$, and we found that taking $u = 6(\tau z)^3$ produced good performance in our simulations.

2. *The case of* $U_1 = |(\hat{f}_r + \hat{f}_l - 2\hat{f}_c)|$. Let z be as in 1, and put $u_1 = 2\tau z$. We claim that $\Pr(U_1 \geq u_1) \leq p$, which follows from the argument in 1 on noting that $U_1/\tau \leq 2 \max(|X_1|, |X_2|)$.

3. *The case of* $U_2 = |(\hat{f}_r + \hat{f}_l - 2\hat{f}_c)(\hat{f}_r - \hat{f}_c)(\hat{f}_l - \hat{f}_c)|$. Let $u = 2(\tau z)^3$ be as in 1. Then it follows as in 1 that $\Pr(U \geq u) \leq p$.

[*Received September 1990. Revised February 1992.*]

REFERENCES

Anonymous (1982), "Change-point Model," in *Encyclopedia of Statistical Sciences* (Vol. 1), eds. S. Kotz, N. L. Johnson, and C. R. Read, New York: John Wiley, p. 415.

Carlstein, E. (1988), "Nonparametric Change-point Estimation," *The Annals of Statistics*, 16, 188–197.

Chapman, J. N., McVitie, S., and Hefferman, S. J. (1991), "Mapping Induction Distributions by Transmission Electron Microscopy," *Journal of Applied Physics*, 69, 6078–6083.

Laurent, P. J., and Utreras, F. I. (1986), "Optimal Smoothing of Noisy Broken Data Using Spline Functions," *Journal of Approximation Theory and Applications*, 2, 71–94.

Lee, D. (1990), "Coping With Discontinuities in Computer Vision: Their Detection, Classification and Measurement," *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 12, 321–344.

——— (1991), "Discontinuity Detection, Classification and Measurement," *SIAM Journal of Scientific and Statistical Computing*, 12, 311–341.

McDonald, J. A., and Owen, A. B. (1986), "Smoothing With Split Linear Fits," *Technometrics*, 28, 195–208.

Shiau, J. H. (1987), "A Note on MSE Coverage Interval of a Partial Spline Model," *Communications in Statistics—Theory and Methods*, 16, 1851–1866.

Shiau, J. H., Wahba, G., and Johnson, D. R. (1986), "Partial Spline Models for the Inclusion of Tropopause and Frontal Boundary Information in Otherwise Smooth Two and Three Dimensional Objective Analysis," *Journal of Atmospheric and Oceanic Technology*, 3, 714–725.

Vercken, C., and Potier, C. (1990), "Spline Fitting Numerous Noisy Data with Discontinuities," in *Curves and Surfaces*, eds. P. J. Laurent et al., New York: Academic Press, pp. 477–480.

Wu, J. S. and Chu, C. K. (1990), "Kernel Type Estimators of Jump Points and Values of a Regression Function," unpublished manuscript.