Local Extremes, Runs, Strings and Multiresolution [1]

By P. L. Davies and A. Kovac

Universität Essen

[2] The paper considers the problem of nonparametric regression with emphasis on controlling the number of local extremes. Two methods, the run method and the taut-string-multiresolution method, are introduced and analysed on standard test beds. It is shown that the number and locations of local extreme values are consistently estimated. Rates of convergence are proved for both methods. The run method converges slowly but can withstand blocks as well as a high proportion of isolated outliers. The rate of convergence of the taut-string-multiresolution method is almost optimal. The method is extremely sensitive and can detect very low power peaks.

Section 1 contains an introduction with special reference to the number of local extreme values. The run method is described in Section 2 and the taut-string-multiresolution method in Section 3. Low power peaks are considered in Section 4. Section 5 contains a comparison with other methods and Section 6 a short conclusion. The proofs are given in Section 7 and the taut string algorithm is described in the appendix.

1. Introduction

1.1. *Approximating data and procedures* Consider a data set $(t_i, y(t_i)), i = 1, \ldots, n$ in $[0, 1] \times \mathbb{R}$ and a family of nonparametric regression models

$$(1.1) \qquad\qquad Y(t) = f(t) + \varepsilon(t), \quad 0 \le t \le 1,$$

indexed by f and ε. We always assume that the t_i are strictly ordered, $0 \le t_1 < t_2 < \ldots < t_n \le 1$. Nonparametric regression is concerned with specifying models, that is, specifying functions f and noise ε such that typical data sets $(t_i, Y(t_i)), i = 1, \ldots, n$ generated under the model "look like" the given data set $(t_i, y(t_i)), i = 1, \ldots, n$ (Donoho 1988; Davies 1995). Such models will be regarded as adequate approximations to the data. They are obtained by manipulating or, more formally, applying a functional to the data and so define a statistical procedure (Tukey 1993). Two such procedures, the run procedure and the taut-string-multiresolution procedure, are considered in this paper. Both are based on a decomposition of the data of the form

$$(1.2) \qquad\qquad y(t_i) = f_n(t_i) + r_n(t_i).$$

which is a special case of the general Tukey decomposition

$$(1.3) \qquad\qquad \mathrm{Data} = \mathrm{Signal} + \mathrm{Noise}.$$

In general Signal and Noise in (1.3) can be separated by assuming that the signal is simple and the noise complex. We measure the simplicity of a function f on $[0, 1]$ by the number of local extreme values in the open interval $(0, 1)$. This is not the standard definition of simplicity which usually refers to some form of smoothness. Complexity is related to randomness and we use a stochastically based concept of noise defined in terms of independently distributed random variables. The two procedures we consider depend on two different definitions of approximation to

white noise. The first is based on Bernoulli sequences and can be written as

$$\text{(1.4)} \qquad \text{max\_run}(\text{sgn}(r_1(t_1)), \ldots, \text{sgn}(r_n(t_n))) \le \rho_n$$

where $\text{sgn}(r_n(t_i))$ denotes the sign of the residual $r_n(t_i)$ and ρ_n is some given number. In other words, the residuals $r_n(t_i)$ may be adequately approximated by white noise or, more informally, look like white noise if (1.4) holds. The default value we suggest is

$$\rho_n = \lceil \log_2 n - 1.47 \rceil$$

which is justified below. The second definition of approximation is based on Gaussian white noise. Suppose for the moment that $n = 2^m$ is a power of two. The multiresolution coefficients $w_{i,k}$ are defined by

$$\text{(1.5)} \qquad w_{j,k} = 2^{-j/2} \sum_{i=k2^j+1}^{(k+1)2^j} r_n(t_i).$$

The residuals may be adequately approximated by Gaussian white noise if

$$\text{(1.6)} \qquad |w_{j,k}| \le \sigma_n \sqrt{2.5 \log n}$$

where σ_n is some measure of the scale of the residuals $r_n(t_i)$. The default functional we use is

$$\text{(1.7)} \qquad \sigma_n = \frac{1.48}{\sqrt{2}} \text{Median}\{|y(t_2) - y(t_1)|, \ldots, |y(t_n) - y(t_{n-1})|\}.$$

This is the same as hard thresholding for wavelets except that we use the factor 2.5 instead of the usual 2. This is for the simple pragmatic reason that 2.5 seems to give better results.

The nonparametric regression procedures we study may now be formulated as follows:

The run problem

Determine the smallest integer k_n for which there exists a function f_n on $[0, 1]$ with k_n local extreme values and such that the residuals $r_n(t_i)$ satisfy the run condition (1.4).

The multiresolution problem

Determine the smallest integer k_n for which there exists a function f_n on $[0, 1]$ with k_n local extreme values and such that the residuals $r_n(t_i)$ satisfy the multiresolution condition (1.6).

Both problems are well defined for any given data set. We give a complete solution of the run problem and make a laudable attempt at the multiresolution problem. The difficulty with the multiresolution problem is that of obtaining suitable approximating functions f_n with no unnecessary local extreme values. The method we use is that of the taut string which will be described below.

The approach given in this paper is neither Bayesian nor frequentist. It makes no assumption that the data are generated by a random mechanism. Indeed they may well be deterministic. We take the point of view that models are approximations to data. In particular we make no reference to "true" regression functions for real data as we do not think that these exist in the sense that, say, elephants exist. By formulating the problem in terms of approximation we avoid the embarrassment of using the word "true", whether in inverted commas or not. We refer to Davies (1995). Readers who do not like this may forget this paragraph, move on to the next section and assume or "know" that all data sets derive from some real existing "true but unknown" regression function f, whatever that might mean.

1.2. *Test beds*   Statistical procedures can be evaluated using real data sets and under the well controlled conditions of a stochastic model or test bed. Tukey uses the word "challenge" (Morgenthaler and Tukey 1991) in this context. The test beds we use are data sets generated by stochastic models of the form

$$(1.8) \qquad Y(t_i) = f(t_i) + \varepsilon(t_i), \quad i = 1, \ldots, n$$

where the $\varepsilon(t_i)$ are generally but not always taken to be independently and identically distributed random variables. On test beds one can talk about estimating functions. They have the advantage of allowing a direct comparison of the function f used to generate the data with functions f_n yielded by the procedures. The comparisons are often visual, indeed these may be the most convincing ones, but can include, as we shall do so, statements on consistency and rates of convergence. Traditional confidence sets are also possible on test beds but these are of a one sided nature (Donoho 1988). For any given sample size n it is possible to perturb f in such a way that it has an arbitrarily large number of local extremes without these being visible in the data. For this reason it is not possible to give a finite upper bound for the number of local extreme values of the function f. When considering rates of convergence we use the supremum norm

$$\|f - f_n\| = \sup_{0 \leq t \leq 1} |f(t) - f_n(t)|.$$

For certain test beds it can be shown that optimal rates of convergence exist. We show that the run based procedure has a slow rate of convergence namely $O(\frac{\log \log n}{\log n})$ whilst the taut-string-multiresolution procedure has the optimal rate of convergence of $O((\frac{\log n}{n})^{1/3})$.

1.3. *Smoothness*   Smoothness is not a consideration in this paper. Techniques for smoothing under shape and deviation constraints have been developed by Metzner (1997), Davies and Löwendick (1999) and Majidi (2000).

1.4. *Previous work*   Much work has been done on the problem of nonparametric regression. Of the different approaches we mention kernel estimation (Nadaraya 1964; Watson 1964), penalized likelihood (Silverman 1985; Green and Silverman 1994), wavelets (Donoho, Johnstone, Kerkyacharian and Picard 1995) and local polynomials (Fan and Gijbels 1995, 1996). None of these methods is directly concerned with local extremes but research has been done which explicitly takes the shape of the regression function into account. Mammen (1991) uses monotone least squares fits between local extrema whilst Mammen and Thomas-Agnan (1998), Mammen, Marron, Turlach and Wand (1998), Delecroix, Simioni and Thomas-Agnan (1995) and Ramsay (1998) modify classical estimators such as spline smoothers and kernel estimators to deal with monotonicity or convexity constraints. None of these papers is directly concerned with estimating the number or the positions of the local extreme values. Work in this direction has been done by Dümbgen (1998b) who applies linear rank tests to locate local extrema. Hengartner and Stark (1995) use the Kolmogoroff ball centred at the empirical distribution function to obtain nonparametric confidence bounds for shape restricted densities. Chaudhuri and Marron (1997) assess the significance of zero crossings of derivatives and use their results to provide a graphical device for displaying the significance the local extremes. Another approach is that of mode testing. We refer to Good and Gaskins (1980), Silverman (1986), Hartigan and Hartigan (1985), Fisher, Mammen and Marron (1994). The positions of the local extreme values are considered by Minotte (1997) using a procedure which decides for each mode in a mode tree

(Minotte and Scott 1993) whether it is significant or not. Mächler (1995) presents an approach using a roughness penalty to penalize points of inflection. The taut string method was used by Davies (1995) in the context of density estimation; there are connections with the excess mass approach of Müller and Sawitzki (1991). Further work in this direction is due to Polonik (1995a, 1995b, 1999). Other articles on shape restricted densities and regression functions are Groeneboom (1985) , Hartigan (1987), Robertson (1967), Sager (1979, 1982, 1986) and Wegman (1970). Order restricted inference is considered in Barlow, Bartholomew, Bremner and Brunk (1972) and Robertson, Wright and Dykstra (1988) . Finally in an article which appeared whilst the present paper was under revision Polzehl and Spokoiny (2000) give a method based on adaptive local means.

The run method (Davies 1995; Metzner, 1997) may be seen as the inversion of the run test for testing the independence of a sequence of observations. It yields the minimum number of local extremes consistent with the observations as well as approximation intervals for their location. Dümbgen (1998a, 1998b) inverts other tests and obtains better convergence rates on standard test beds but at the cost of greater computational complexity.

Taut strings are well understood in the context of fitting an isotone function. The greatest convex minorant of the integrated data is a taut string and its derivative is precisely the least squares isotone approximation (Barlow et al. 1972; Leurgans 1982). The idea of using taut strings for densities goes back to Hartigan and Hartigan (1985) who derived a test for the unimodality of a density. In Davies (1995) it was explicitly used to calculate approximate densities. It was first used in the general nonparametric regression problem by Mammen and van de Geer (1997). They showed that the taut string is a special case of a penalized least squares functional

8

where the penalty is based on the total deviation norm. The derivative of the taut string has the smallest number of local extremes of all functions whose integral lies in the supremum ball. Davies (2000) gives an application of the string method to spectral density functions.

2. The run method

2.1. *General description*   We illustrate the general method of solving the run problem of Section 1.1 using the data shown in Figure 1 with a maximum allowable run length $\rho_n = 2$. To ease the notation we set $t_i = i$.
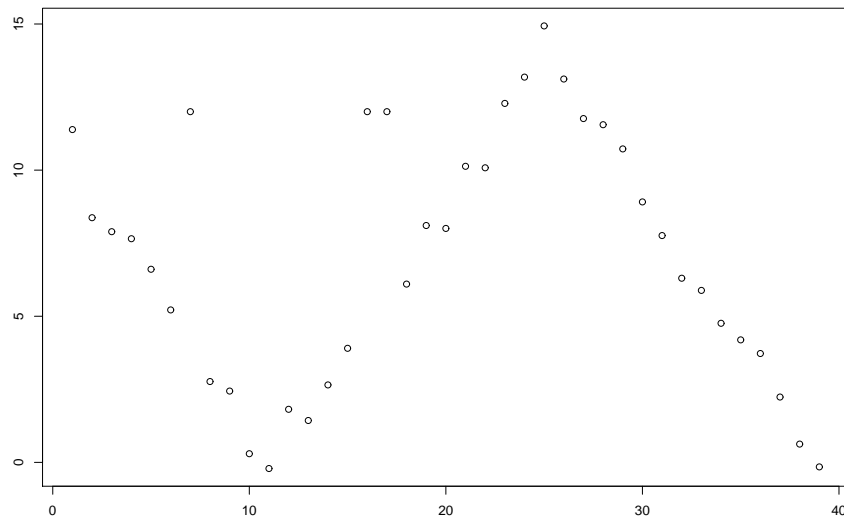

FIG. 1.   *An artificial data set to illustrate the run procedure.*

Consider a function f_n which is initially nonincreasing. The value of $f_n(3)$ cannot exceed $\max\{y(1), y(2), y(3)\}$ as otherwise the first three residuals would be negative

giving rise to a run of length 3. This gives the upper bound

(2.9)
$$f_n(3) \leq \max\{y(1), y(2), y(3)\}.$$

No nontrivial upper bounds are available for $f_n(1)$ and $f_n(2)$. Similarly a lower bound for $f_n(1)$ is given by

$$f_n(1) \geq \min\{y(1), y(2), y(3)\}.$$

As the function is taken to be initially nonincreasing the above argument extended to each point gives rise to a sequence $u(j)$ of upper bounds defined by

(2.10)
$$u(j) = \min\{u(j-1), \max\{y(j-2), y(j-1), y(j)\}\}$$

with $u(1) = u(2) = \infty$. To see this we note that the reasoning which lead to (2.9) gives

(2.11)
$$f_n(j) \leq \max\{y(j-2), y(j-1), y(j)\}.$$

On the other hand if $u(j-1)$ is an upper bound for $f_n(j-1)$ then

(2.12)
$$f_n(j) \leq f_n(j-1) \leq u(j-1).$$

Combining (2.11) and (2.12) leads to (2.10). The corresponding lower bounds $l(j)$ are given by

(2.13)
$$l(j) = \max\{l(j+1), \min\{y(j), y(j+1), y(j+2)\}\}.$$

If at some point i the lower bound $l(i)$ exceeds the upper bound $u(i)$ then it is not possible for the function f_n to be nonincreasing on $1, \ldots, i+2$ and for the maximum run length not to exceed 2. Thus at the latest at the point $i+1$ we must switch from a nonincreasing to a nondecreasing function. It turns out that it is not necessary to calculate the lower bounds in order to determine the point at which the switch must be made. The point $i+2$ is also the first point at which the last three values

$y(i), y(i + 1)$ and $y(i + 2)$ lie above the upper bounds $u(i), u(i + 1)$ and $u(i + 2)$ respectively. This can be shown by backward induction. The result can be seen in the upper panel of Figure 2. The values of y at the points $14, 15$ and 16 lie above the upper bounds. Thus at the latest at the point 15 a switch must be made to a nondecreasing function. This is done as follows. We set $l(14) = l(15) = -\infty$ and calculate lower bounds for the nondecreasing section by

$$l(j) = \max\{l(j - 1), \min\{y(j - 2), y(j - 1), y(j)\}\}, \quad j = 16, ...$$

The lower bounds are calculated until either the end of the data set is reached or at some point the last three y-values lie below the corresponding lower bounds. If the latter contingency occurs a switch is made to a nonincreasing function and the upper bounds are calculated as before. For the data shown in Figure 1 the switch must be made at the latest at the point 28 as can be seen from the upper panel of Figure 2. The results obtained are the following. If we start with a nonincreasing function and limit the allowable run length to 2 then at least two local extremes are required. The first, a local minimum, must be attained at a point i with $i \leq 15$. The second, a local maximum, must be attained at a point j with $j \leq 28$. The process described here is called "stretching to the right" in Davies (1995). Starting with a nondecreasing function leads to one extra local maximum: in general that initial behaviour is chosen which minimizes the number of local extremes. The opposite process, stretching to the left, starts with the last point of the data set and moves to the left calculating the bounds in the same manner. The result is shown in the lower panel of Figure 2. The number of local extremes is the same, in this case two, and the lower bounds for the positions of the local extremes are 8 and 15. The combined result for the data of Figure 1 is that the local minimum must be attained in the interval $[8, 15]$ and the local maximum in the interval $[22, 28]$. In an interval

bounding the position of a local minimum there is no nontrivial lower bound for the function without further specification of the location of the local minimum. Similarly, in an interval bounding the position of a local maximum there is no nontrivial upper bound for the function. At points where only one lower bound has been calculated (either from stretching to the right or to the left), this defines the lower bound. At points where two lower bounds have been calculated the minimum is taken. The upper bound is calculated in the corresponding manner. The final bounds are shown in Figure 3. The outliers at the points 7,16 and 17 are seen not to have lead to additional local extremes. A more detailed analysis of the effect of outliers is given below.

The results so far show that restricting the maximal run length to 2 forces any regression function to have at least two local extreme values. If f_n is a function with two local extreme values and whose residuals have a maximum run length of 2 then the function must initially be nonincreasing, the local minimum must be attained in the interval $[8, 15]$ and the local maximum in the interval $[22, 28]$. Furthermore the function must lie between the bounds at all data points.

It is easy to show that not every function with the correct monotonicity behaviour and lying between the bounds satisfies the run condition. The construction of a function with only two local extreme values and which satisfies the run condition is not completely trivial. Suppose that it is possible, as indeed it is, to define a nonincreasing function which satisfies the run condition with $\rho_n = 2$ on the interval $[1, 15]$. The last two residuals are of necessity positive as 15 is an upper bound for the position of the local minimum. To prevent a run of length 3 the next residual must be negative and the function must therefore have a value exceeding $y(16) = 12$. It is clear that this choice of f_n will lead to four local extreme values. To construct
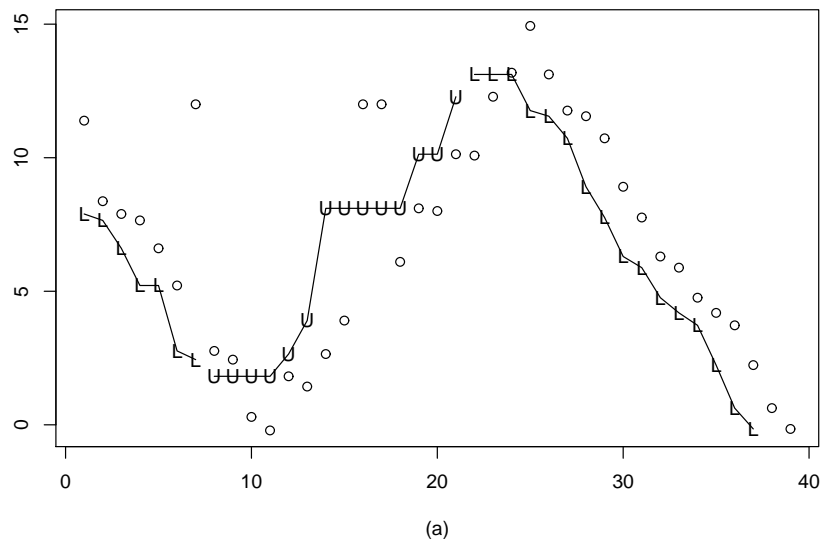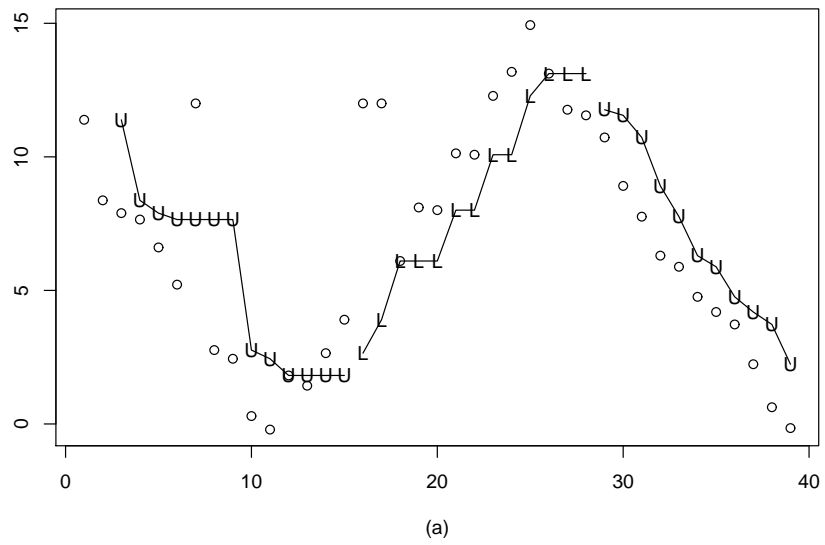
Fig. 2. *The data set of Figure 1 together with the bounds from stretching to the right (upper panel) and stretching to the left (lower panel)*

a function with only two local extreme values we proceed as follows. We take the upper bound 15 for the position of the first local extreme value. The upper and lower bounds u and l as defined by (2.10) and (2.13) respectively are calculated for the section $[1, 15]$ of the data. The lower bound is best calculated from right to left with initial values $l(14) = l(15) = -\infty$. The first 2 values are defined to be

$$f_n(1) = f_n(2) = \max\{y(1), y(2)\}$$

and the others are defined as follows. If the first i values of f_n have been defined the value at $i + 1$ is defined as follows. If $f_n(i) \leq u(i + 1)$ then we set $f_n(i + 1) = f_n(i)$. If $f_n(i) > u(i + 1)$ we define f_n as follows. If $y(i + 1) \geq l(i + 1)$ then we set $f_n(i + 1) = l(i + 1)$. If not we backtrack and determine the first j to the left of $i + 1$ such that $y(j) \geq l(j)$. That is

$$j = \max\{m : m \leq i + 1, y(m) \geq l(m)\}.$$

We then set $f_n(m) = l(j)$ for $j \leq m \leq i + 1$. At the latest we must switch to a nondecreasing function at the point 15. We do this as follows. We consider the upper bound of the position of the next local extreme value. For the present data the is the point 28. We now calculate the lower and upper bounds for the section of the data on $[14, 28]$. The first two values of the function are defined by

$$f_n(14) = f_n(15) = \min\{y(14), y(15)\}.$$

The function is now defined as on the first section but with the roles of the lower and upper bounds interchanged. The function f_n is continued as a constant until the first time it drops below the lower bound. At this point a switch is made to the upper bound but only if the value of the observation at this point does not exceed the upper bound. If not we backtrack to that first point where the upper bound exceeds the y-value. This process is continued in the obvious manner until the end

14

of the data set is reached. The function so defined satisfies the run condition, has the correct monotonicity behaviour and lies between the upper and lower bounds over the whole range. It is shown in Figure 3. This construction is due to Davies (1995) and Metzner (1997). The latter contains further details and variations on the run procedure.
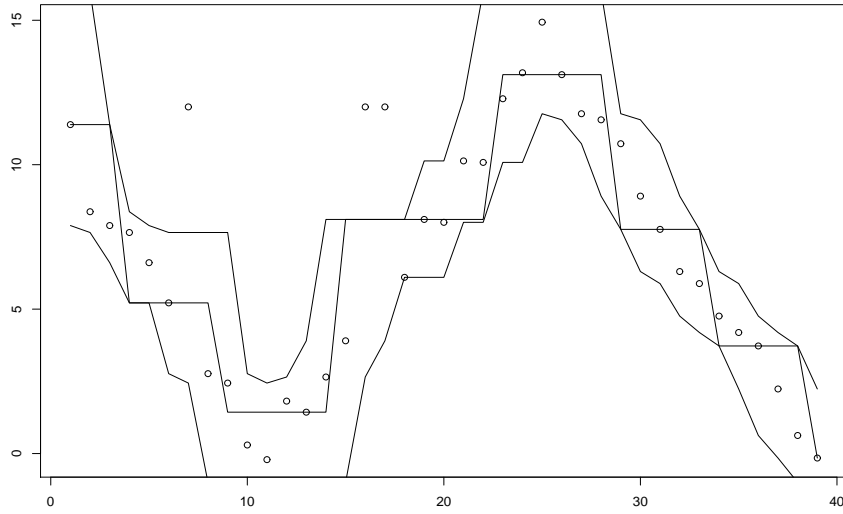

FIG. 3. *Final bounds for a run length of 2 together with a function satisfying the run condition and lying between the bounds.*

The general recurrence equations for calculating the upper and lower bounds for a given run length ρ_n are as follows:

$$(2.14) \quad u_n(i) = \begin{cases} \min\{u_n(i-1), \max\{y(j), i-\rho_n \leq j \leq i\}\}, \text{ nonincreasing} \\ \min\{u_n(i+1), \max\{y(j), i \leq j \leq i+\rho_n\}\}, \text{ nondecreasing.} \end{cases}$$

$$(2.15) \quad l_n(i) = \begin{cases} \max\{l_n(i+1), \min\{y(j), i \leq j \leq i+\rho_n\}\}, \text{ nonincreasing} \\ \max\{l_n(i-1), \min\{y(j), i-\rho_n \leq j \leq i\}\}, \text{ nondecreasing.} \end{cases}$$

Combining these results gives the following theorem.

THEOREM 1. *For a given data set* $(t_i, y(t_i)), i = 1, \ldots, n$ *and a given maximal run length* ρ_n *for the residuals*

$$r_n(t_i) = y(t_i) - f_n(t_i)$$

the lower and upper bounds described above specify the minimum number k_n *of local extreme values a function must have so that the residuals satisfy the run condition. Furthermore there exist functions with* k_n *local extreme values whose residuals satisfy the run condition. Any function with* k_n *local extreme values whose residuals satisfy the run condition must lie between the bounds over the whole range. The local extreme values of any such function must also be attained in the* k_n *intervals* $[t_i^l, t_i^r], 1 \leq i \leq k_n$, *determined by the stretching to the right and left procedures.*

The run procedure is easy to apply and requires only the specification of the maximal run length. To give a default value for the allowable run length ρ_n we consider a sequence of i.i.d. Bernoulli random variables taking the values 1 and -1 each with probability $1/2$. We denote by R_n the length of the longest subsequence which is composed entirely of 1s or -1s. For any given $\alpha, 0 < \alpha < 1$, we denote the α-quantile of R_n by $qu(n, \alpha, R_n)$ that is

$$qu(n, \alpha, R_n) = \min\{m : \mathbf{P}(R_n \leq m) \geq \alpha\}.$$

For large n we have the simple approximation

$$(2.16) \qquad \rho_n = qu(n, \alpha, R_n) \approx \lceil \log_2 n - 2 - \log_2(-\log(\alpha)) \rceil.$$

which can be deduced from the results given in Section XIII.7 of Feller (1968) or shown directly. The default choice of run length is (2.16) with $\alpha = 0.5$ which represents a form of median for the number of local extreme values. In this case

$$(2.17) \qquad \rho_n = qu(n, 0.5, R_n) \approx \lceil \log_2 n - 1.47 \rceil.$$

Nonparametric regression is an infinite dimensional problem. This is one reason why the bounds l_n and u_n are relatively wide and why the intervals $[t_i^l, t_i^r], i = 1, \ldots, k_n$, are relatively long. Another is the use of a rather crude measure of approximation to white noise, namely the run length. The effects can be seen in the upper panel of Figure 4 which shows 1000 points of a sine curve contaminated with Cauchy noise as in (2.18)

$$(2.18) \qquad y(t) = 10\sin(t) + \varepsilon(t), \quad \varepsilon(t) \quad \text{i.i.d. standard Cauchy.}$$

The bounds are based on a run length $\rho_n = 9$ which is the default value given by (2.17) and result in $k_n = 2$ local extreme values. One method of obtaining narrower bounds is to decrease the run length whilst maintaining just two local extreme values. For the sine curve the run length can be reduced to $\rho_n = 6$ without introducing any further local extremes. The lower panel of Figure 4 shows these new bounds.

The bounding intervals based on a run length of $\rho_n = 9$ are $[141, 376]$ for the local maximum and $[600, 878]$ for the local minimum. The corresponding intervals for the run length $\rho_n = 6$ are $[178, 340]$ and $[630, 823]$ respectively. These hold for any function which satisfies the run condition. It is however possible to specify points within these intervals near which the function is required to have a local extreme value. The data of Figure 1 show that it may not be possible to have local extremes at exactly these points and still fulfil the run condition. A modification of the construction given above of a function within the bounds which satisfies the run condition shows the following. Given any points t_i^c in the intervals $[t_i^l, t_i^r]$ there exists a function satisfying the run condition with exactly k_n local extremes which are attained at points differing from the prescribed t_i^c by at most the run length. Different choices lead to different functions. This can be interpreted as an attempt
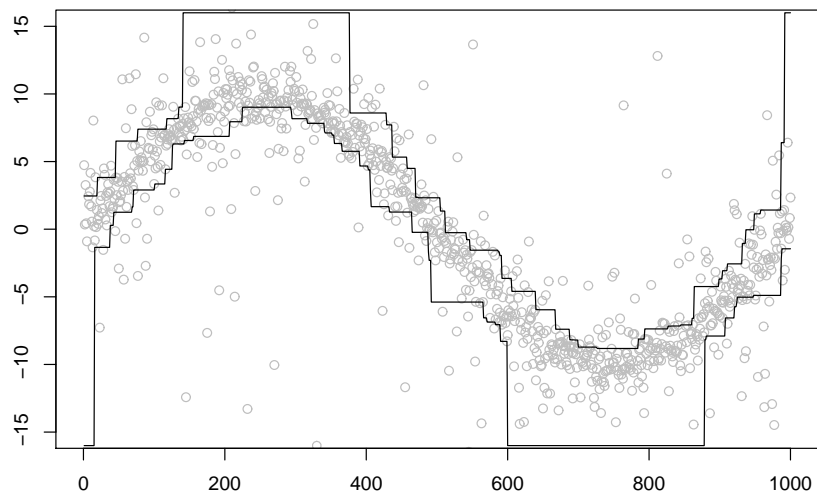
(a)
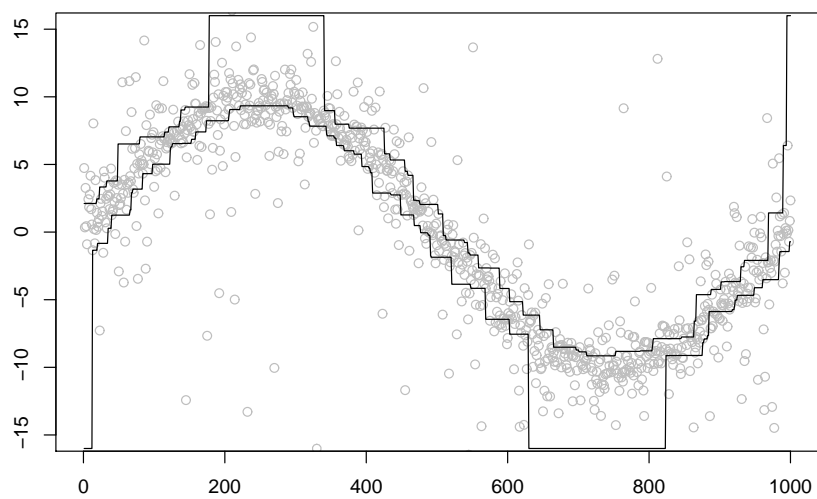
(b)

FIG. 4. *1000 data points generated under (2.18). Upper panel shows bounds with default run length $\rho_n = 9$ given by (2.17). Lower panel shows bounds with minimal run length $\rho_n = 6$ consistent with two local extremes.*

to visualize the variability of the set of adequate functions. One obvious choice for the points t_i^c is the centres of the intervals. The top panel of Figure 5 shows the result of doing this for the Cauchy sine data using a run length of $\rho_n = 6$. The local extreme points are required to be near the centres of the intervals $[178, 340]$ and $[630, 823]$ namely 259 and 726. The lower panel of Figure 5 shows the bounds when the local extremes are required to be near 180 (maximum) and 820 (minimum).

If a specific approximating function for the data is required there are several alternatives. One method is to calculate the bounds either with or without minimizing the run length and with or without specifying the positions of the local extremes. From this a specific function which satisfies the run condition and lies between these bounds can be obtained from the algorithm described above. The upper panel of Figure 6 shows this function together with the underlying sine curve. It is based on the bounds shown in the upper panel of Figure 5. Although bounds are constructed using the run criterion the statistician is not obliged to consider only functions which satisfy it. The simplest manner of obtaining a function is to take the average of the lower and upper bounds. The lower panel of Figure 6 shows the average of the bounds of the upper panel of Figure 5 together with the approximating sine curve.

That no nontrivial bounds are available at the local extreme values or at the start and finish of the data set is related to the ability of the run method to withstand outliers. We analyse the situation for a block of outliers. If the observation before the block of outliers lies below the upper bound then the method can withstand a block of large positive outliers of length equal to the run length. If however the last s observations before the block of outliers lie above the upper bound then only a block of length $\rho_n - s$ of arbitrarily large positive outliers can be tolerated.
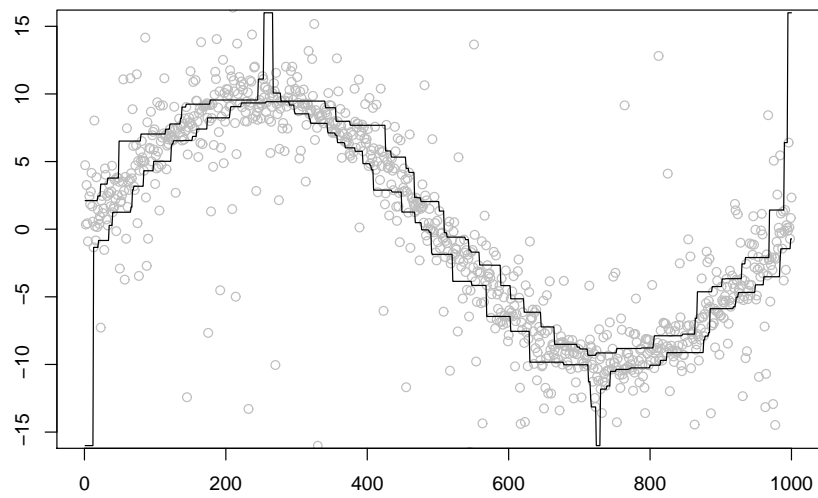
(a)
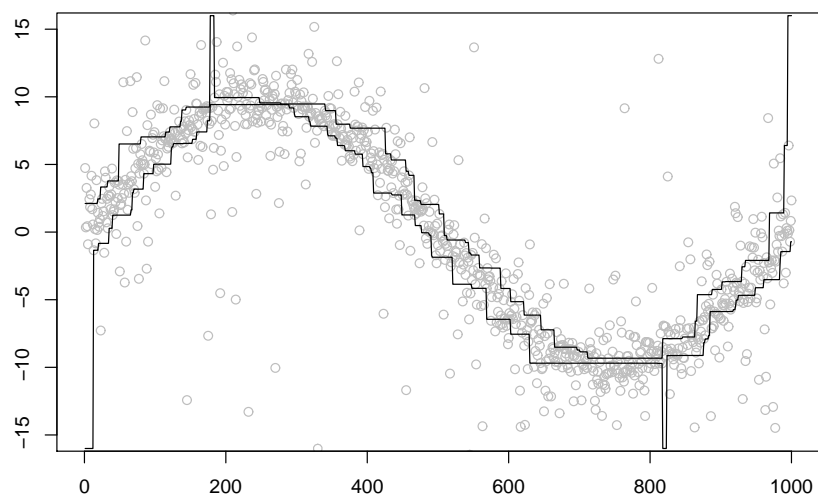
(b)

Fig. 5. *The upper panel shows bounds with run length $\rho_n = 6$, a local maximum near 259 and a local minimum near 726. The lower panel shows bounds with run length $\rho_n = 6$, a local maximum near 180 and a local minimum near 820.*
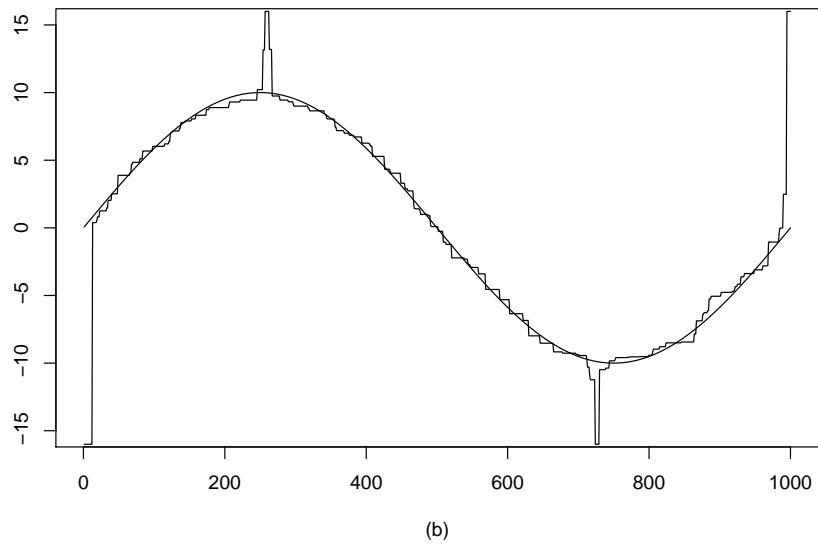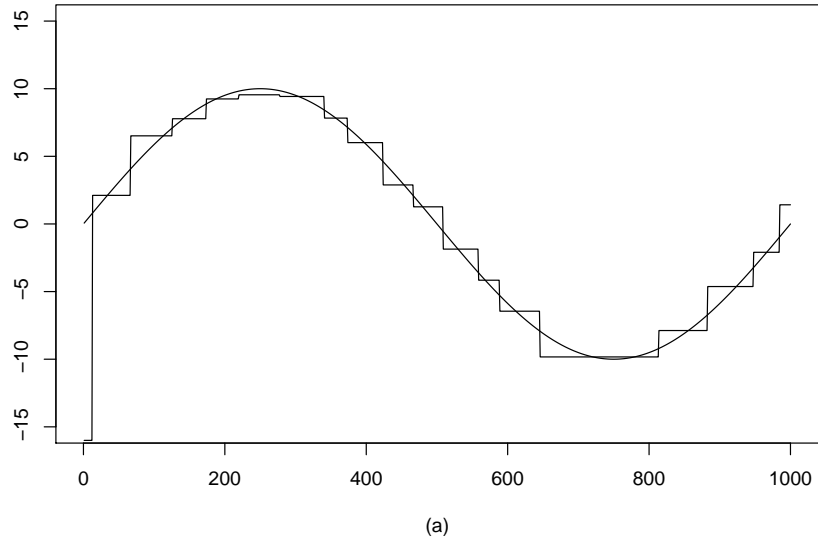
20

(a)

(b)

FIG. 6. *The upper panel shows a function satisfying the run condition with $\rho_n = 6$ and lying between the bounds of the upper panel of Figure 5. The lower panel shows the average of the bounds of the upper panel of Figure 5. The underlying sine curve is shown in both panels.*

Similar considerations apply to the lower bounds. Figure 7 shows Gaussian noise with 9 outliers in the centre. The upper panel shows that they are ignored if the run length is $\rho_n = 10$ (upper panel) but detected if the run length is $\rho_n = 9$ (lower panel). The reason is that the observation just before the block of outliers lies above the upper bound.

2.2. *Behaviour on test beds*   We consider a data set generated on a test bed of the form (1.8). Let K_n^α denote the number of local extreme values determined by the run procedure using the α-quantile of the length of the longest run (2.16). The following theorem holds.

THEOREM 2.   *Consider data generated on the test bed (1.8) where the function f has k local extremes and the errors $\varepsilon(t)$ are independently distributed with $\mathbf{P}(\varepsilon(t) < 0) = \mathbf{P}(\varepsilon(t) > 0) = \frac{1}{2}$. Then*

$$\mathbf{P}(k \geq K_n^\alpha) \geq \alpha.$$

The residuals $\varepsilon(t)$ fulfil the run condition with probability at least α and from Theorem 1 it follows that at least K_n^α local extreme values are required if the run length of the residuals is not to exceed $qu(n, \alpha, R_n)$.

As mentioned in Section 1.2 this form of result is best possible for fixed n. If however the sample size increases and the design points $t_i, 1 \leq i \leq n$, become dense in $[0, 1]$ then the procedure will determine the correct number of local extremes for sufficiently large n. The precise statement is contained in the next theorem where we assume $t_i = \frac{i}{n}, 1 \leq i \leq n$. The theorem shows that f can be consistently estimated and gives a rate of convergence. As the rate of convergence applies to the bounds it holds for any function f_n lying within the bounds as long as it has the same monotonicity behaviour as the bounds. It is not necessary for the function to give
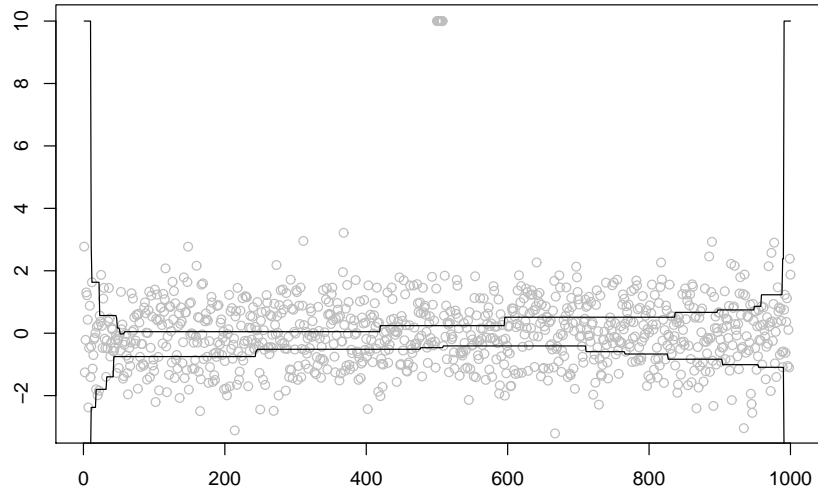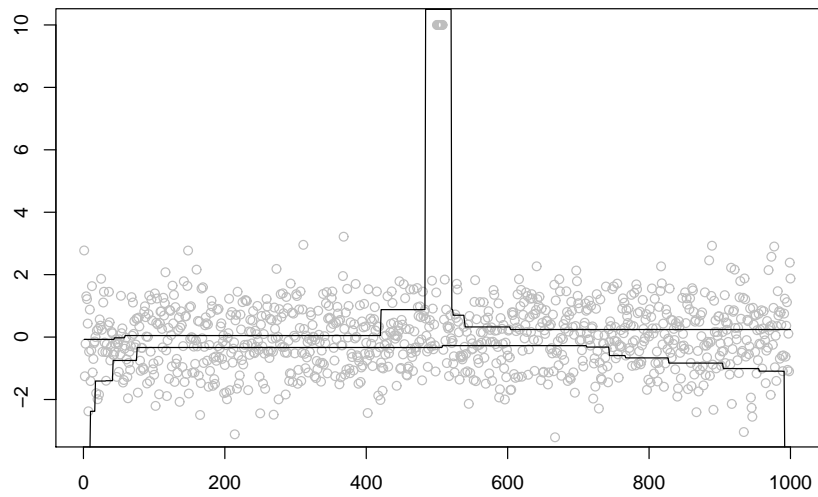
FIG. 7. *The upper panel shows the bounds for $n = 1000$ and run length $\rho_n = 10$ with Gaussian noise and a block of 9 outliers in the centre. The lower panel shows the effect of reducing the run length to $\rho_n = 9$.*

rise to residuals which satisfy the run condition. In particular the rate of convergence holds for the function defined to be the mean of the bounds. The intervals which contain the local extremes will be denoted by $I_i^e(n, \alpha), i = 1, \ldots, K_n^\alpha$ and their midpoints by $t_i^e(n, \alpha), i = 1, \ldots, K_n^\alpha$. The following theorem holds

THEOREM 3. *Consider the test bed (1.8) where*

- *f has a bounded continuous first derivative $f^{(1)}$ and exactly k local extreme values at the points $0 < t_1^e < \ldots < t_k^e < 1$*

- *$f^{(1)}(t) = 0$ only for $t \in \{t_1^e, \ldots, t_k^e\}$*

- *the $\varepsilon(t)$ are independently and identically distributed, have median zero and a continuously differentiable distribution function in a neighbourhood of zero.*

Then the following hold:

(a) For all $\delta > 0$

$$\liminf_{n \to \infty} \mathbf{P}(\{K_n^\alpha = k\} \cap \{\max_{1 \leq i \leq k} |I_i^e(n, \alpha)| \leq \delta\}$$

$$\cap \{\max_{1 \leq i \leq k} |t_i^e(n, \alpha) - t_i^e| \leq \delta\}) \geq \alpha.$$

(b) For a sequence of functions f_n within the bounds and with the same monotonicity behaviour as the bounds the locations of the extreme values of the f_n converge in probability to those of f.

(c) There exists a constant $b > 0$ such that for any sequence of functions f_n as in (b)

$$\lim_{n \to \infty} \mathbf{P}\left(\sup_{\{t:|f^{(1)}(t)| \geq \delta\}} |f(t) - f_n(t)| \leq b\frac{\log \log n}{\log n}\right) = 1$$

for all $\delta > 0$.

24

(d) There exists a constant $b > 0$ such that

$$\lim_{n \to \infty} \mathbf{P} \left(\inf_{\{t: |f^{(1)}(t)| \geq \delta\}} u_n(t) - f(t) \geq b \frac{\log \log n}{\log n} \right) = 1$$

and

$$\lim_{n \to \infty} \mathbf{P} \left(\inf_{\{t: |f^{(1)}(t)| \geq \delta\}} f(t) - l_n(t) \geq b \frac{\log \log n}{\log n} \right) = 1$$

for all $\delta > 0$.

3. The taut-string-multiresolution method

3.1. *Multiresolution and local averages*   The multiresolution problem of Section 1.1 is based on the white noise approximation defined by (1.5), (1.6) and (1.7). If the sample size n is not a power of 2 then the interval $[(k2^j + 1)/n, (k + 1)2^j/n]$ of (1.5) can be replaced by $[(k2^j + 1)/n, \min\{(k + 1)2^j/n, 1\}]$. This is legitimate because of the subadditivity of probability measures. An alternative is to use the techniques of Kovac and Silverman (2000).

The $w_{j,k}$ can be replaced by discrete wavelet coefficients or, more generally, co-efficients generated by any multiresolution scheme. Prior knowledge of the shape of the peaks to be detected can be incorporated into the shape of the mulitreso-lution functions. The advantages multiresolution schemes are that they work well in practice, that they result in almost optimal rates of convergence on appropriate test beds and that the calculations are of order n. The bound (1.6) can be replaced by the more general bound

(3.19) $$\sigma_n \sqrt{\tau \log n}$$

with σ_n again given by (1.7). The idea behind the bound (3.19) is that for a se-quence of random variables $Z_1, Z_2 \ldots$ with a common subgaussian distribution (see

Theorem 6 below)

$$\lim_{n \to \infty} \mathbb{P}(\max_{i=1,\ldots,n} |Z_i| < \sqrt{\tau \log n}) = 1$$

for some $\tau > 0$. This follows from the tail estimate

(3.20) $$\mathbb{P}(|Z_i| \geq x) \leq c \exp(-x^2/\tau)$$

for some $\tau > 0$ which continues to hold for sums $Z = \frac{1}{\sqrt{n}} \sum_1^n Z_i$ of i.i.d. random variables. The multiresolution coefficients are of this form and so we may use a threshold of the form (3.19).

3.2. *Strings and bounds*   Although the multiresolution problem of Section 1.1 is well posed the difficulty is that, in contrast to the run problem, there is no immediate connection between the number of local extremes on the one hand and the multiresolution definition of white noise approximation on the other. To overcome this is we use an unrelated method, the taut string, to produce candidate functions f_n^k with k local extreme values and then to take the smallest k for which the residuals

$$r_n^k \left(\frac{i}{n}\right) = y \left(\frac{i}{n}\right) - f_n^k \left(\frac{i}{n}\right)$$

approximate white noise. The success of this approach depends entirely on the efficacy of the taut string method to provide such candidate functions.

The left panel of Figure 8 shows a small simulated data set $y \left(\frac{i}{n}\right), i = 1, \ldots, n$: the right panel the integrated process y_n° defined by

(3.21) $$y_n^\circ \left(\frac{j}{n}\right) = \frac{1}{n} \sum_{i=1}^{j} y \left(\frac{i}{n}\right), \quad j = 0, \ldots, n.$$

Integrated or summed processes will always be distinguished by a superscript $^\circ$. Consider the lower l_n and upper bound u_n for y_n° defined by

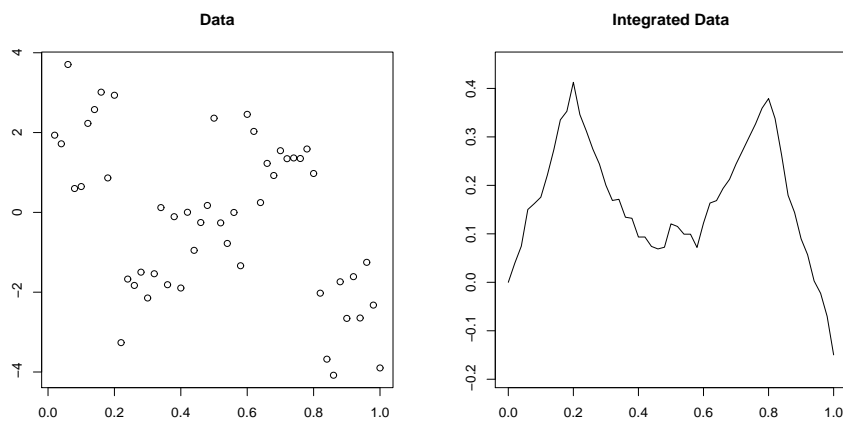(3.22) $$l_n = y_n^\circ - \frac{C}{\sqrt{n}} \quad \text{(lower bound)}$$

FIG. 8. *Integrating the data. The left panel shows a small simulated data set and the right panel their partial sums.*
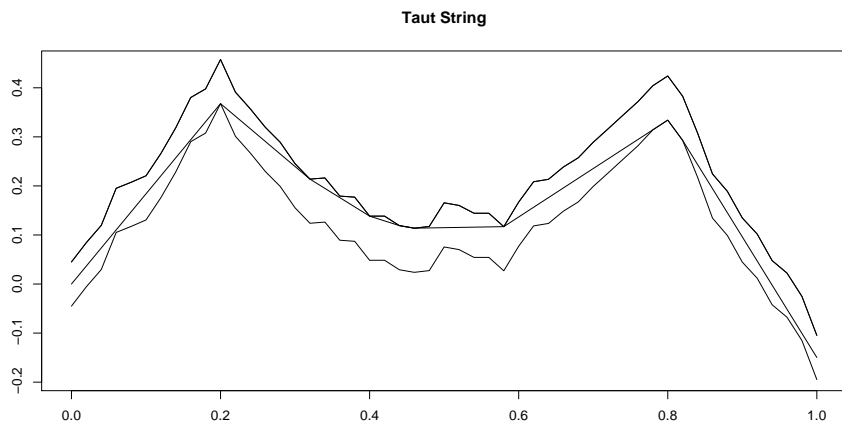
FIG. 9. *A taut string between two bounds*

$$(3.23) \qquad u_n = y_n^\circ + \frac{C}{\sqrt{n}} \quad \text{(upper bound)}$$

for some $C > 0$. Consider a piece of string attached with one end at the point $(0,0)$ and the other at the point $(1, y_n^\circ(1))$ and constrained to lie between l_n and u_n. Suppose that the string is now pulled until it is taut. It defines a function s_n° on $[0,1]$. This is shown in Figure 9. We shall use the derivative of the taut string as an approximation to the data as shown in Figure 10

FIG. 10. *The derivative of the taut string of Figure 9 together with the data.*

The string can be defined analytically as that function s_n° on $[0, 1]$ with with the smallest length

$$\text{length}(s_n^\circ) = \int_0^1 \sqrt{1 + s_n^{\circ\,(1)}(t)^2}\,dt.$$

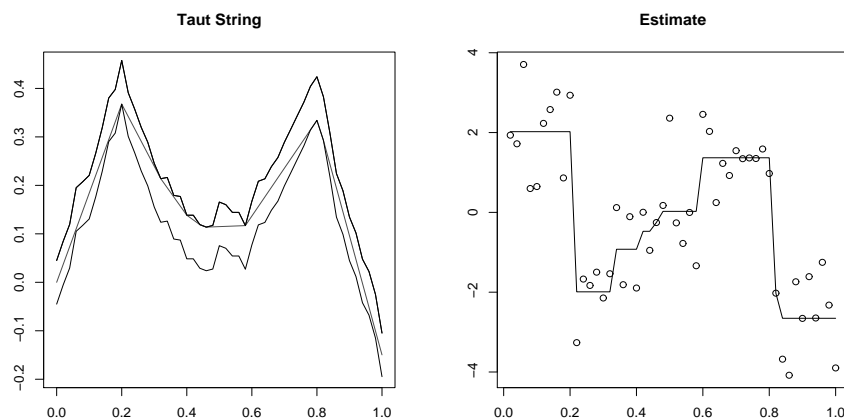which satisfies

$$(3.24) \qquad s_n^\circ(0) = 0, \quad s_n^\circ(1) = y_n^\circ(1), \quad l_n(t) \leq s_n^\circ(t) \leq u_n(t), \quad 0 \leq t \leq 1.$$

It can be shown that s_n° also has the smallest total variation under the side conditions (3.24). That is s° minimizes

$$(3.25) \qquad\qquad\qquad \text{TV}(s_n^\circ) = \int_0^1 |s_n^{\circ\,(1)}(t)|\,dt.$$

subject to (3.24). In practice the restrictions (3.24) only hold at a finite number of points $i/n, i = 0, \ldots, n$. The calculation of the ensuing string can be performed quickly and the complexity of the algorithm is $\text{O}(n)$. The algorithm we use is explained in the Appendix. The description of the taut string in the finite case is the following (see also Mammen and van de Geer 1997). The points at which the string coincides with either the lower or upper boundary are called knots. Between knots the function is linear. Between two knots x_i and x_j where the string touches

the upper bound u_n and between which it does not touch the lower bound l_n the string is the largest convex minorant of the upper bound u_n. Similarly if the knots are ones where the string touches the lower bound l_n and between which it does not touch the upper bound u_n then it is the smallest concave majorant of the lower bound l_n. At points where the string switches from the upper bound u_n to the lower bound l_n the derivative $s_n = s_n^{\circ(1)}$ has a local maximum. At points where the string switches from the lower bound l_n to the upper bound u_n the derivative s_n has a local minimum.

It is clear from the description of the taut string that its derivative s_n has the smallest number of local extremes amongst all functions g_n whose integral $g_n^{\circ}(x) = \int_0^x g_n(t)dt$ satisfies the conditions (3.24). It is precisely this property which makes the taut sring so efficacious in controlling the number of local extremes. It is possible that the chosen initial and final conditions of (3.24) may themselves cause additional local extremes. This can be overcome by attatching the ends of string either to the upper or to the lower boundary at the start and at the end. Those of the four strings whose derivative has the smallest of local extremes minizies the number of local extremes of all functions whose integral lies within the bounds. Dümbgen has pointed out to us that the effect of the initial conditions disappears in the limit. The advantage of the string defined by (3.24) is that end effects are reduced. We therefore use it in the examples although the theorems are stated in terms of the string which minimizes the number of local extremes.

Bounds of the form (3.22) and (3.23) will be referred to as a tube with centre y_n° and radius $\frac{C}{\sqrt{n}}$: it will be denoted by $T\left(y_n^{\circ}, \frac{C}{\sqrt{n}}\right)$,

$$(3.26) \qquad\qquad T\left(y_n^{\circ}, \frac{C}{\sqrt{n}}\right) = \{g_n : l_n \leq g \leq u_n\}.$$

Other bounds derived from local squeezing will be used later. As already mentioned the derivative $s_n(C)$ of the taut string $s_n^\circ(C)$ is used as a candidate regression function. As $s_n^\circ(C)$ is piecewise linear $s_n(C)$ is piecewise constant. It is perhaps worth mentioning some properties of the derivative $s_n(C)$. Consider two knots of the string at the points $\frac{i}{n} < \frac{j}{n}$ where the taut string touches the upper bound u_n and are such that the string does not touch the lower bound between the two knots. The derivative $s_n(C)$ is then the least squares monotone increasing approximation to the y-values between the knots. In other words the $s_n(C)\left(\frac{m}{n}\right), m = i, \dots, j$ solve

$$\text{minimize} \quad \sum_i^j \left(y\left(\frac{m}{n}\right) - a_m\right)^2 \quad \text{subject to} \quad a_i \le \dots \le a_j.$$

(see Barlow et al. 1972). On sections where the string is the smallest concave majorant of the lower bound l_n the derivative s_n of the string solves the corresponding problem where the values are nonincreasing. This implies the following. Between two successive knots either both on the upper bound or both on the lower bound the derivative is the mean of the observations between the knots. Between knots defining a local minimum or maximum the derivative is again the mean of the observations but increased or decreased respectively by the diameter of the tube at that point (see also Proposition 8 of Mammen and van de Geer 1997). Compared to the data local maxima will tend to be too small and local minima too large.

3.3. *Previous work*   The first use of the taut string method in statistics would seem to be due to Barlow et al. (1972). Hartigan and Hartigan (1985) were the first to use it in connection with modality, it being an integral part of their dip test for unimodality. It was explicitly used by Davies (1995) to construct k-modal densities. Mammen and van de Geer (1997) extended its use to the nonparametric

regression problem but do not mention the connection with the number of local extreme values.

3.4. *Behaviour on test beds*   The next three theorems are concerned with the asymptotic behaviour of the taut string on the test bed (1.8). In particular it is shown to have an optimal rate of convergence away from the local extremes. The first theorem corresponds to Theorem 2 for the run procedure. We denote the number of local extremes of the derivative of the taut string in the supremum tube with upper and lower bounds given by (3.22) and (3.23) by K_n^C. Theorems 4, 5 and 6 refer to the taut string which minimizes the number of local extremes. Dümgen has proved that the theorems continue to hold with sutiable modifications, for example replacing the Brownian motion on $[0, 1]$ of Theorem 4 by the Browninan Bridge on $[0, 1]$, if the initial conditions (3.24) are imposed.

THEOREM 4.   *Consider the test bed (1.8) but where the errors $\varepsilon(t)$ are independently and identically distributed with mean 0 and finite variance σ^2. Then*

$$\lim_{n \to \infty} \mathbf{P}(k \geq K_n^C) = H\left(\frac{C}{\sigma}\right).$$

where H denotes the distribution of $\sup_{0 \leq t \leq 1} |W(t)|$ where W denotes a standard Wiener process.

It is worth pointing out that the rate of convergence in Theorem 4 does not depend on the function f of the test bed but only on the rate of convergence of the partial sums

$$\sqrt{n}\varepsilon_n^\circ(t) = \frac{1}{\sqrt{n}} \sum_1^{[nt]} \varepsilon\left(\frac{j}{n}\right)$$

to a Wiener process. To make the theorem applicable without knowledge of the variance σ^2 an estimate of it is required. We use (1.7).

The taut string based on the radius $C/\sqrt{n}$ will be denoted by $S_n^\circ = S_n^\circ(C)$ with derivative $S_n = S_n(C)$. We write $I_i^e(n, C), 1 \le i \le K_n^C$, for the intervals where S_n attains it local extreme values and denote the midpoints of these intervals by $\tau_n^e(n, C), 1 \le i \le K_n^C$. The next theorem corresponds to Theorem 3(a) for the run procedure.

THEOREM 5.   *Consider the test bed (1.8) where f satisfies the conditions of Theorem 3 and where the errors $\varepsilon(t)$ are identically and independently distributed with mean 0 and finite variance σ^2. Then for all $\delta > 0$*

$$\lim_{C \to \infty} \lim_{n \to \infty} \mathbf{P}(\{K_n^C = k\} \cap \{\max_{1 \le i \le k} |I_i^e(n, C)| \le \delta\} \cap \{\max_{1 \le i \le k} |\tau_i^e(n, C) - t_i^e| \le \delta\}) = 1.$$

We note that the rate of convergence to the correct number of local extremes depends much more heavily on the function f than the rate of convergence in Theorem 4. It is for this reason that we state Theorem 5 separately.

In the following the length of an interval I will be denoted by $|I|$.

THEOREM 6.   *Consider the test bed (1.8) where f satisfies the conditions of Theorem 3 and additionally*

- *f has a bounded second derivative $f^{(2)}$ which is nonzero at the k local extremes,*

- *the errors $\varepsilon(t)$ are independently and identically distributed subgaussian random variables i.e. $\mathbf{E} \exp(\lambda \varepsilon(t)) < \exp(c\lambda^2)$ for all λ for some $c > 0$.*

Then

(a) $$\lim_{C \to \infty} \lim_{n \to \infty} \mathbf{P}(t_i^e \in I_i^e(n, C), 1 \le i \le k) = 1.$$

(b) $\lim_{C \to \infty} \lim_{n \to \infty} \mathbf{P}\left(1 - \delta \le \dfrac{n^{1/6} |f^{(2)}(t_i^e)|^{1/3} |I_i^e(n, C)|}{(24C)^{1/3}} \le 1 + \delta\right) = 1 \, for \, all \, \delta > 0.$

(c) Let x_i and x_{i+1} denote successive knots which are either both on the upper or both on the lower bound. Then there exists an $A > 0$ such that

$$\lim_{C \to \infty} \lim_{n \to \infty} \mathbf{P}\left(\max_i (x_{i+1} - x_i)|f^{(1)}(x_i)|^{2/3} \le A\left(\frac{\log n}{n}\right)^{1/3}\right) = 1.$$

(d) There exists an $A > 0$ such that

$$\lim_{C \to \infty} \lim_{n \to \infty} \mathbf{P}\left(\sup_{t \in A_n} \frac{|f(t) - S_n(t)|}{|f^{(1)}(t)|^{1/3}} \le A\left(\frac{\log n}{n}\right)^{1/3}\right) = 1$$

where $A_n = [A\left(\frac{\log n}{n}\right)^{1/3}, 1 - A\left(\frac{\log n}{n}\right)^{1/3}] \backslash \cup_i^n I_i^e(n, C)$.

(e) There exists a constant $A > 0$ such that

$$\lim_{C \to \infty} \lim_{n \to \infty} \mathbf{P}\left(\sup_{t \in B_n} \frac{|f(t) - S_n(t)|}{|f^{(2)}(t)|^{1/3}} \le A C^{2/3} n^{-1/3}\right) = 1$$

where $B_n = \cup_1^k I_i^e(n, C)$.

The definition of subgaussian given above implies that the mean is zero (see Kahane 1968, page 62, Exercise 10). We note that (d) implies that S_n is optimal at points t with $|f^{(1)}(t)| > \delta > 0$. At points t with $f^{(1)}(t) = 0$ its rate of convergence is $n^{-1/3}$ compared with the optimal rate of $n^{-2/5}$ (Leurgans 1982). As the length of the interval is $n^{-1/6}$ as against the optimal $n^{-1/5}$ this cannot be alleviated by replacing $S_n(t)$ for $t \in \cup_1^k I_i^e(n, C)$ by the mean of the $y-$values. Nevertheless the change of boundaries at local extremes does alter the behaviour of the multiresolution coefficients described in the following section.

Donoho et al. (1995) consider four sets of simulated data, the Blocks data, the Bumps data, the Heavisine data and the Doppler data. They were also analysed by Fan and Gijbels (1995), (1996). Figure 11 shows the taut string method to the Blocks signal contaminated with Gaussian white noise. The tube radius is $1.149\sigma/\sqrt{n}$. The choice $C = 1.149$ corresponds to the 0.5 quantile of the maxi-
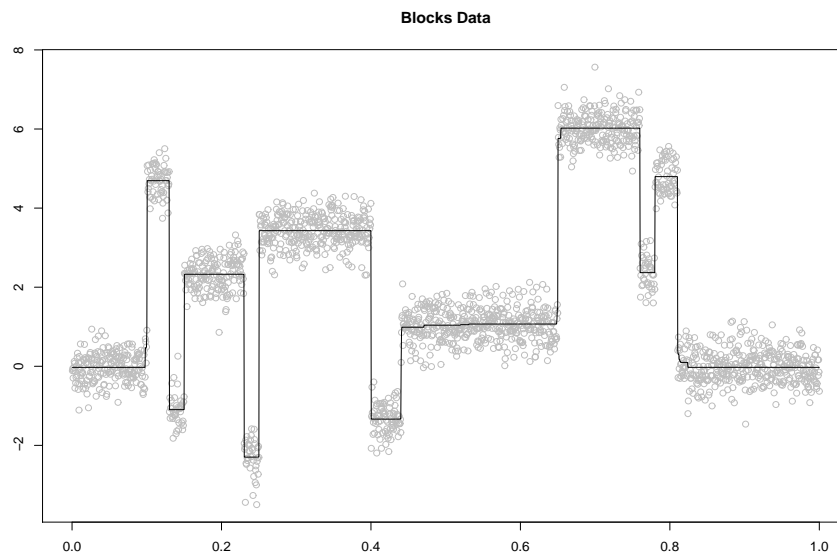
Blocks Data

FIG. 11. *The Blocks signal. The signal was contaminated with white noise and afterwards the taut string approximation calculated using the $C/\sqrt{n}$ tube with $C = 1.149$.*

mum of the absolute value of a standard Brownian motion on $[0, 1]$ (see Freedman 1971, (34) Proposition, page 27).

3.5. *Strings and multiresolution* As stated above the derivative of the taut string tends to be too small at local maxima and too large at local minima. A simple analysis shows that this effect alone will cause some of the multiresolution coefficients of the residuals to exceed the bound as $n \to \infty$ regardless of the value of τ in (3.19). The problem may be rectified to some extent by the simple expedient of replacing S_n by $\tilde{S}_n$ where $\tilde{S}_n$ is constant between knots where it is equal to the mean of the y−values between the knots. The integrated $\tilde{S}_n$ will be denoted by $\tilde{S}_n^\circ$. It may be seen that $\tilde{S}_n^\circ$ coincides with Y_n° at the knots and is linear in between. This alteration does not effect the number of local extremes but it improves the rate of

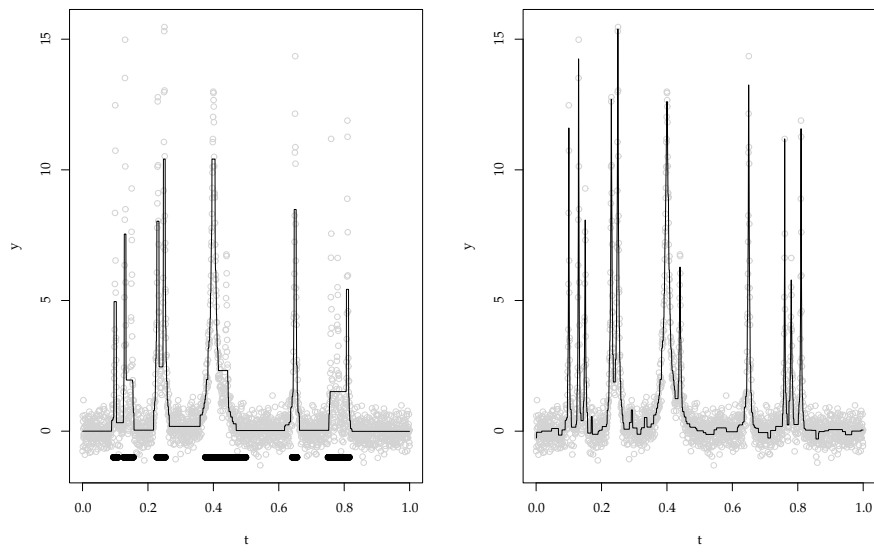FIG. 12. *These figures show the application of the taut string methods to the Bumps signal. In the left panel the tube radius is $1.149\sigma/\sqrt{n}$ and the lines at the bottom indicate regions where multiresolution coefficients are too large. The right panel shows the global squeezing procedure where the tube is squeezed until all multiresolution coefficients are below the threshold value.*

convergence at local extremes. The behaviour of the multiresolution coefficients for candidate functions based on $\tilde{S}_n$ is covered by the next theorem.

THEOREM 7.   *Suppose the assumptions of Theorem 6 hold and consider the multiresolution coefficients of the residuals of $\tilde{S}_n$ i.e.*

$$W_{i,j} = 2^{-i/2} \sum_{m=j2^i+1}^{(j+1)2^i} \left(Y\left(\frac{m}{n}\right) - \tilde{S}_n\left(\frac{m}{n}\right) \right).$$

with supports $[j2^i + 1, (j+1)2^i]$. Then for each constant $A > 0$ there exists a $\tau > 0$ such that the following hold:

(a) All multiresolution coefficients $W_{i,j}$ taken from intervals whose support is of length at most $A(\log n/n)^{1/3}$ are eventually smaller than the bound (3.19).

(b) All multiresolution coefficients $W_{i,j}$ the endpoints of whose supports do not lie in $\cup_1^k I_i^e(n,C)$ are eventually smaller than the bound (3.19).

(c) At each local extremum t_i^e of f there exists an interval $J_i^e(n,C) \supset I_i^e(n,C)$ of length at most $O\left(\frac{1}{\log n}\right)$ for which the following holds. All multiresolution coefficients $W_{i,j}$ with one endpoint in $I_i^e(n,C)$ for some i and the other outside of the corresponding $J_i^e(n,C)$ are eventually smaller than the bound (3.19).

(d) For each interval $I_i^e(n,C)$ of (b) there exists a multiresolution coefficient $W_{i,j}$ with support in $I_i^e(n,C)$ and whose absolute value eventually exceeds the threshold (3.19) whatever the value of τ.

Theorem 7 shows that for an appropriate choice of τ in (3.19) all multiresolution coefficients will eventually be below the threshold (3.19) apart from some in shrinking intervals which contain the local extreme points of the function f. Furthermore there are multiresolution coefficients which will eventually exceed the bound (3.19) whatever the choice of τ. In other words, the suboptimal rate of convergence at local

extremes will be detected by the multiresolution analysis of the residuals. Figure 12 shows the application of the taut string method to the Bumps data of Donoho et al. (1995). The left panel shows the intervals where the multiresolution coefficients of the residuals exceed the threshold (3.19). The right panel shows the effect of squeezing globally until all multiresolution coefficients are below the threshold. It results in several spurious local extreme values. In the next section we describe how local squeezing may be used to eliminate the spurious local extremes.

3.6. *Local squeezing and multiresolution*   The results of the last section show that although the taut string is locally adaptive (Mammen and van de Geer 1997) it is not sufficiently so. Furthermore, as the Bumps data in Figure 12 show, the problem cannot be solved by decreasing the radius of the tube globally. The calculations used in the proof of Theorem 6 show that if the radius is of order $n^{-3/5}$ instead of $n^{-1/2}$ then the length of the interval which contains a local extreme value is of order $(\log n/n)^{1/5}$ and the rate of convergence is of order $(\log n/n)^{2/5}$. This is almost optimal and calculations show that all multiresolution coefficients will now be smaller than the bound (3.19). However a global radius of order $n^{-3/5}$ for the tube will eventually give rise to spurious local extremes. This may be countered by decreasing the radius of the tube only in an $n^{-1/5}$-neighbourhood of those local extremes found with the tube radius $Cn^{-1/2}$. As it stands this is not feasible as the taut string with radius $n^{-1/2}$ locates the local extremes only with an accuracy of order $n^{-1/6}$. One possibility is to decrease the radius of the tube on the intervals of order $n^{-1/6}$ which contain the local extreme points. For subgaussian random variables with variance σ^2 we have

$$\mathbf{P}\left(\max_{l \leq i \leq l+m-1} \frac{1}{\sqrt{m}} \left| \sum_{j=l}^{i} \varepsilon\left(\frac{j}{m}\right) \right| \geq \sigma\lambda \right) \leq a\exp(-c\lambda^2)$$

for all $\lambda > 0$ for some constants a and c. Translating this inequality into one for the integrated process Y_n° gives the following

$$(3.27) \quad \mathbf{P}\left(\max_{l \leq i \leq l+m-1}\left|Y_n^\circ\left(\frac{i}{n}\right) + \Delta_n - f^\circ\left(\frac{i}{n}\right)\right| \leq \gamma(n,m)\right) \geq 1 - a\exp(-c\lambda^2)$$

where

$$(3.28) \qquad\qquad \Delta_n = f^\circ\left(\frac{l}{n}\right) - Y_n^\circ\left(\frac{l}{n}\right).$$

and

$$(3.29) \qquad\qquad \gamma(n,m) = \lambda\frac{\sqrt{m}}{n}$$

The inequality (3.27) implies that with high probability depending on λ the integrated process f° lies in the locally squeezed tube

$$(3.30) \qquad T(Y_n^\circ + \Delta, \gamma, I) = \{g : \max_{t \in I}|Y_n^\circ(t) + \Delta - g(t)| \leq \gamma\}$$

with $\Delta = \Delta_n$ given by (3.28), $\gamma = \gamma(n,m)$ given by (3.29) and $I = I(l,m,n) = \left[\frac{l}{n}, \frac{l+m-1}{n}\right]$. In other words local squeezing on the interval I is accomplished by shifting the integrated process Y_n° by an amount Δ on I and then using the modified lower and upper bounds

$$l_n(t,I) = Y_n^\circ(t) + \Delta - \gamma \text{ and } u_n(t,I) = Y_n^\circ(t) + \Delta + \gamma$$

for $t \in I$. The size of the shift Δ is not known in advance; (3.27) and (3.28) say that for some shift, namely that given by (3.28), the integrated process f° lies in the locally squeezed tube with high probability. This means that Δ should be chosen to minimize the number of local extreme values of the derivative of the taut string through the modified tube. This may be accomplished by calculating the taut string over a grid of the possible values of Δ.

The effect of local squeezing on an interval where the derivative of the taut string has a local extreme value may be analysed by setting $m = m_n = cn^{5/6}$

(Theorem 6 (b)) in (3.27). This results in $\gamma_n = \lambda\sqrt{c}n^{-7/12}$. The radius of the tube is now $n^{-7/12}$ compared with the optimal radius of $n^{-3/5}$. The proportional difference is $n^{-1/60}$. Repeating this procedure will always improve the asymptotic rate of convergence but no finite number of steps will result in the optimal radius of $n^{-3/5}$. Suppose however we know the points where f takes on its local extreme values. Under these circumstances it possible to squeeze the tube locally at the local extremes over intervals which contain $cn^{4/5}$ points. The result is described by (3.27) with $\gamma_n = \lambda\sqrt{c}n^{-3/5}$. The taut string through the modified tube will behave asymptotically as before but will have an improved and indeed optimal rate of convergence at the local extreme values of f. If this is taken into account then the proof of Theorem 7 shows that *all* multiresolution coefficients will now be set to zero. The problem is that we cannot at the moment locate the local extremes of f with the required accuracy. We now show how this can be done, at least in principle, using a multiresolution analysis of the residuals.

We consider the taut string using a tube of radius $Cn^{-1/2}$ and the modified derivative $\tilde{S}_n$. As shown above this will, in the limit, result in the correct number of local extremes and all the extreme points will lie in the corresponding intervals of $\tilde{S}_n$ whose lengths will be of order $n^{-1/6}$. Furthermore all multiresolution coefficients will be smaller than the bound (3.19) apart from some whose support is strictly contained in shrinking neighbourhoods of the local extreme points. We now gradually squeeze the tube locally at the intervals where $\tilde{S}_n$ has local extreme values. This is continued until all the multiresolution coefficients of the residuals drop below the threshold (3.19).

We show firstly that this process will not terminate too soon i.e. before the optimal rate of convergence is obtained. To demonstrate this we consider a function g

defined on $[0, 1]$ and a sequence of piecewise constant functions g_n used to approximate g. The change of notation is to emphasize that the following arguments relate only to the approximation of g by the piecewise constant functions g_n. We assume the following:

(i) g has a finite number of local extreme values in $(0, 1)$.

(ii) g has a first derivative $g^{(1)}$ which is zero only at the local extremes.

(iii) g has a continuous second derivative $g^{(2)}$ which is nonzero at the local extremes.

(iv) g_n is piecewise constant and with the same monotonicity behaviour as g on each interval.

(v) if g has a local extreme at t_e then g_n is constant on an interval which contains t_e and is of length at least $c \left(\frac{\log n}{n} \right)^{1/5}$ for some constant $c > 0$.

The condition (iv) means that if g is monotone on an interval I then so is g_n and with the same form of monotonicity. The converse need not hold.

Suppose now that the multiresolution coefficients of the differences $g - g_n$ are all smaller than the bound (3.19). We consider what implications this has for the maximal deviation of g_n from g. Consider firstly a point t with $d_n = |g(t) - g_n(t)|$ and where g and g_n are nondecreasing with $g^{(1)}(t) > 0$. We write $d_n = |g(t) - g_n(t)|$ and suppose that $g(t) > g_n(t)$. The other case is treated similarly. Let the interval I_n of length λ_n be the support of some multiresolution coefficient with $I_n \subset [t - d_n/g^{(1)}(t), t]$. The multiresolution coefficient is

$$\frac{1}{\sqrt{\lambda_n}} \int_{I_n} (g(u) - g_n(u)) du \geq \frac{1}{2} \lambda_n^{3/2} g^{(1)}(t).$$

If this coefficient does not exceed the threshold (3.19) then

$$\frac{1}{2}\lambda_n^{3/2}g^{(1)}(t) \leq \sqrt{\frac{\tau\log n}{n}}$$

which gives

$$\lambda_n \leq \left(\frac{4\tau\log n}{ng^{(1)}(t)^2}\right)^{1/3}.$$

As there exists some multiresolution coefficient with $\lambda_n \geq d_n/(2g^{(1)}(t))$ this implies

$$\frac{d_n}{2g^{(1)}(t)} \leq \left(\frac{4\tau\log n}{ng^{(1)}(t)^2}\right)^{1/3}$$

or

$$(3.31) \qquad d_n \leq 2^{5/3}|g^{(1)}(t)|^{1/3}\left(\frac{\tau\log n}{n}\right)^{1/3}.$$

At the local extremes of the function g we can no longer exploit the assumption (iv) but use instead the assumption (v). Let g have a local extreme value at t_e and set $d_n = |g(t_e) - g_n(t_e)|$. The function g is locally quadratic at t_e and arguments similar to the ones just used show the following. If none of the multiresolution coefficients with support close to the local extreme point t_e exceed the threshold (3.19) then

$$(3.32) \qquad |d_n| \leq 2|g^{(2)}(t_e)|^{1/5}\left(\frac{\tau\log n}{n}\right)^{2/5}.$$

We now apply the above results to the nonparametric regression problem. The residuals are

$$r_n(t) = Y(t) - f_n(t) = f(t) - f_n(t) + \varepsilon(t).$$

On setting $g = f$ and $g_n = f_n$. We note firstly that if the $\varepsilon(t)$ are i.i.d. subgaussian then their contribution to the multiresolution coefficients will lie below the threshold for some value of τ. If f_n and f satisfy the assumptions (i) to (v) above and if all the multiresolution coefficients of the residuals r_n do not exceed the threshold it follows that multiresolution coefficients of the differences $f - f_n$ will also lie below

the threshold. This implies that the rate of convergence of the f_n to f will be optimal away from the local extremes in the sense of uniform convergence. At the local extremes themselves the rate of convergence will be at least $\left(\frac{\log n}{n}\right)^{2/5}$ as against the optimal rate of $n^{-2/5}$. Let now $f_n = \tilde{S}_n$ where $\tilde{S}_n$ denotes the modified taut string through the tube which is now locally squeezed at the local extremes with $m_n = C(\log n)^{1/5}n^{4/5}$ in (3.6). The proof of Theorem 7 shows that all multiresolution coefficients of the residuals will now be below the threshold. This implies the following. Suppose we start with a tube of radius $C/\sqrt{n}$ and then gradually locally squeeze the tube at the local extremes of $\tilde{S}_n$. As the lengths of the intervals where $\tilde{S}_n$ has its local extreme values are initially of order $n^{5/6}$ (Theorem 6) we see that this process will eventually lead to local squeezing over intervals containing $m_n = C(\log n)^{1/5}n^{4/5}$ observations. At this point all multiresolution coefficients will be below the threshold. This implies that if we stop squeezing at the local extremes of $\tilde{S}_n$ as soon as all multiresolution coefficients drop below the threshold then $g = f$ and $g_n = \tilde{S}_n$ will satisfy the conditions (i) to (v) above. The resulting rate of uniform convergence will be $\left(\frac{\log n}{n}\right)^{1/3}$ away from the local extreme values of f. At the local extreme values of f the rate of convergence will be $\left(\frac{\log n}{n}\right)^{2/5}$. We have therefore shown that in principle local squeezing based on a multiresolution analysis of the residuals results in almost optimal rates of convergence at the local extreme values. In practice local squeezing must be done by an algorithm and to complete the proof it must be shown that the algorithm attains what is in principle possible. We do not pursue this any further and we simply describe the algorithm we use without a proof that it is asymptotically correct. In practice it works very well.

3.7. *A practical implementation*   Local squeezing is described by (3.30) and requires the specification of the shift Δ, the radius of the tube γ and the interval I. The intervals on which local squeezing is performed are determined by the multiresolution analysis of the residuals. As mentioned in in the discussion of local squeezing given above the choice of shift Δ seems not to be critical and is set to zero. Finally the radius of the tube γ must be specified and this is critical. It is implemented by reducing the radius of the tube locally by a constant factor $\rho, 0 < \rho < 1$. If ρ is small, say $\rho = 0.5$, then the procedure terminates quickly as all multiresolution coefficients soon drop below the threshold. This occasionally leads to too many local extreme values. If ρ is nearly 1, say $\rho = 0.95$, then the calculations require more time but this choice of ρ has in our experience not lead to superfluous local extreme values. The following simple procedure turned out to be very reliable and was used for all the examples.

Firstly, the scale of the noise is calculated using (1.7). The next step is to choose a large initial global tube radius γ_0. A reasonable choice is such that the straight line connecting $(0,0)$ and $(1, y_n^\circ(1))$ lies inside the tube. An alternative is to use Theorem 4 and use an initial tube radius $\gamma_0 = C/\sqrt{n}$ where C is set to 2.242, the 0.95-quantile of the distribution of the maximum of a Brownian motion. Given the initial radius the taut string is calculated as described in the appendix.

We use the modified version $\tilde{s}_n$ of the derivative s_n as described just before Theorem 7. If all multiresolution coefficients of the residuals $\tilde{r}_n \left(\frac{i}{n} \right) = y \left(\frac{i}{n} \right) - \tilde{s}_n \left(\frac{i}{n} \right)$ are smaller than the bound (3.19) the algorithm terminates. If not we define local tube radii γ_i^1 $(i = 0, \ldots, n)$ by $\gamma_i^1 = \rho \gamma_0$ if a multiresolution coefficient which depends on y_i or y_{i+1} exceeds the bound (3.19) and by $\gamma_i^1 = \gamma_0$ elsewhere. The
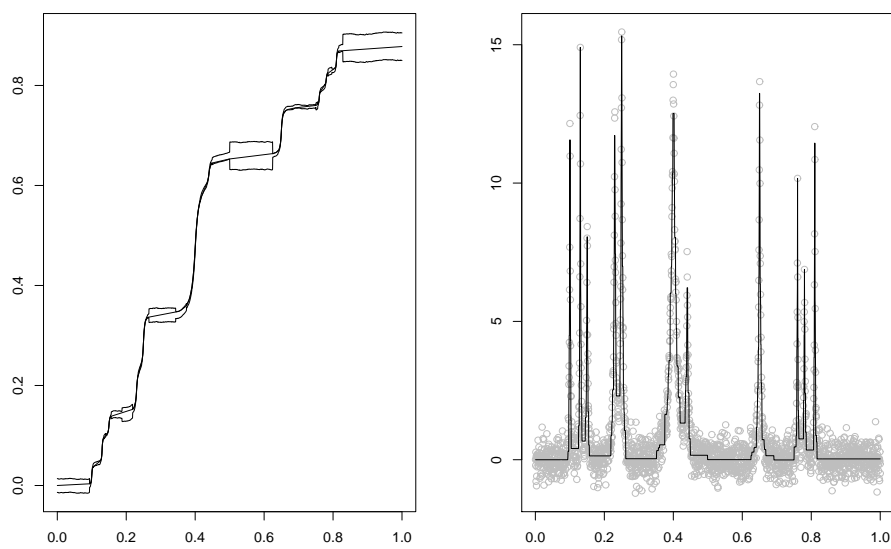
FIG. 13. *Local squeezing and the Bumps data. The left panel shows the final upper and lower bounds together with the taut string. The Bumps data and the derivative of the taut string are plotted in the right panel.*

tube is squeezed centrally i.e. $\Delta = 0$. The squeezing factor ρ is set to 0.95. A new candidate function is finally defined by the result of the taut string procedure.

The algorithm proceeds by calculating the multiresolution analysis of the new residuals. If again some coefficients are still too large further squeezing is applied by reducing the tube radius in the relevant intervals by a factor of ρ until eventually all coefficients are below the threshold. As indicated in Section 3.1 the value of τ we use is 2.5.

3.8. *Examples*   The effect of local squeezing on the bounds is shown in Figure 13. Displayed are the final bounds for the Bumps signal of Figure 12 in the left panel and the reconstruction in the right panel. This should be compared with Figure 12.

4. Low power peaks   Most work on nonparametric regression and density problems evaluates procedures in terms of rates of convergence on test beds of the form (1.8). The distribution of the errors $\varepsilon(t)$ is specified and the regression function f is kept fixed whilst the number n of observations is increased. In this situation there exist optimal rates of convergence (Khas'minski 1978; Ibragimov and Khas'sminski, 1980; Stone, 1982). As shown the taut string method based on a ball of radius $Cn^{-1/2}$ attains the optimal rate away from the local extremes for functions with a specified number of local extremes. The run method falls well short with a rate of convergence of order $(\log\log n)/\log n$. In spite of this the run based procedure can give better results than an optimal taut string method based on a tube of constant radius. To investigate this phenomenon we consider a different form of test bed. Let f be a continuous function with k peaks. We consider an interval $[a_n, b_n]$ which does not contain a peak and graft a peak onto the function f. The height of the peak is h_n and we denote the new function by f^n. The power
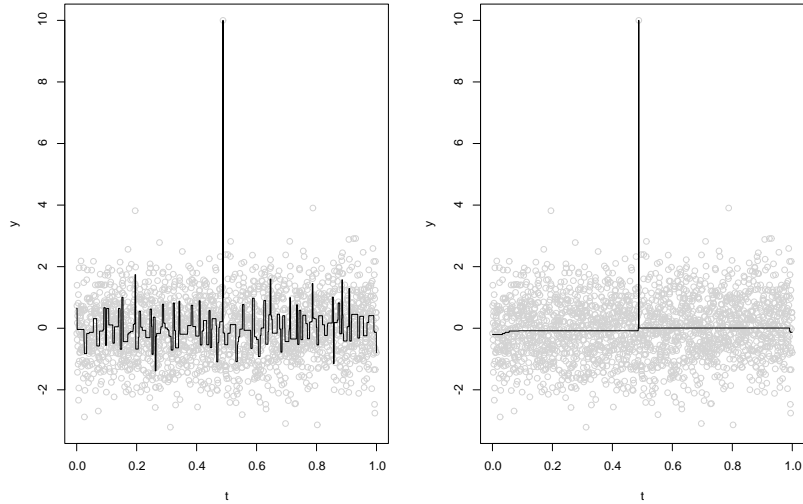
FIG. 14. *A regression function corrupted with Gaussian noise and the reconstructions based on n = 2048 observations. The left panel shows the effect of globally squeezing until the peak near 0.5 is detected. The right panel shows the result of the local squeezing procedure.*

of the peak is defined to be

$$(4.33) \qquad P_n = \int_0^1 |f - f^n| = \int_{a_n}^{b_n} |f - f^n|.$$

Consider now the asymptotic test bed

$$(4.34) \qquad Y_{n,i} = f^n \left(\frac{i}{n} \right) + \varepsilon_{n,i}$$

with $P_n = \mathrm{o}\left(n^{-1/2}\right)$ and $\lim_{n \to \infty} h_n = \infty$. For large C the tube $T(f^{n\circ}, C)$ will contain f° with large probability. As f has k local extremes so will the taut string through the ball. From this it follows that Theorems 5 and 6 will continue to hold for the asymptotic test bed (4.34). In other words the taut string method will fail to identify the peak at $[a_n, b_n]$. Similar considerations apply to kernel estimation. If the optimal global window is used a low powered peak will not be detected if $P_n = \mathrm{o}\left(n^{-1/5}\right)$.

In the case of the run method on the test bed (4.34) the peak will be detected if

$$\liminf \frac{n(b_n - a_n)}{\log_2 n} > 1.$$

If

$$\liminf \frac{n(b_n - a_n)}{\log_2 n} < 1.$$

it will not be detected. Figure 7 shows this effect.

The local squeezing method using local averages will clearly identify the peak of f^n. Indeed the method will pick up a signal confined to one single observation as long $h_n \geq B\sqrt{\log n}$ where B depends on f and the bound parameter τ. This is demonstrated by Figure 14 where a very low power peak defined by a single observation is detected without introducing any spurious peaks.

5. Examples and comparisons

5.1. *Artificial data* Figure 15 shows the results of applying the taut-string-multiresolution procedure to the data sets of Donoho et al. (1994). The Doppler signal is displayed in the upper left corner. Local squeezing detects the oscillations near the origin very well without introducing spurious extreme values at other positions. The main feature of the Heavisine signal in the upper right panel is the presence of discontinuities near 1/3 and 2/3. The taut string reproduces them, again without introducing spurious extreme values elsewhere. The Blocks data signal is piecewise constant so it is not surprising that the taut string performs very well. Indeed the reconstruction can hardly be distinguished from the original signal. Finally the Bumps data can be compared with Figure 12. Local squeezing has identified all the relevant peaks without introducing spurious ones.
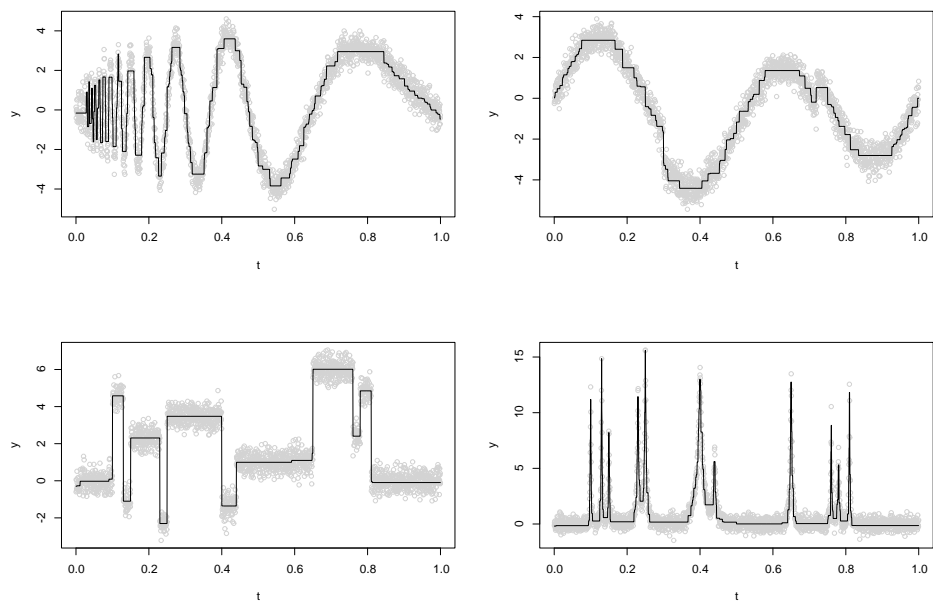
FIG. 15. *Four noisy test signals and their reconstructions using the taut string procedure with local squeezing.*

5.2. *Real data*   Figure 16 shows the results of a spectroscopical analysis of a gall stone. The chemists involved informed us that for this particular data all the peaks found were real ones and no spurious peaks were introduced. We note that between peaks the reconstruction is almost constant. This data set, although real, is not much different to the Bumps data sets of Figure 13.

5.3. *Comparison with other methods*   Comparison with other methods is always difficult especially when the aims may be different. This paper is concerned with the number of local extremes and only secondly with rates of convergence and not at all with L^2 errors. Other methods are concerned with smoothness and rates of convergence in spaces of smooth functions. The problem of comparison is made more difficult by the lack of available software. We therefore restrict the comparisons to wavelets for which standard software is available. Even here many options are open such as the choice of the wavelet and the threshold level. The left panel of Figure 17 shows the block data and the wavelet reconstruction. The Haar wavelet was used which is presumably the best for this data. Nevertheless pseudo Gibbs effects are apparent and lead to 47 local extreme values. The corresponding result for the taut-string-multiresolution method is shown in the right panel of Figure 17. It gives the correct number of local extremes namely 9.

6. Conclusion   We have introduced two methods for obtaining regression functions whilst keeping the number of local extremes under control. Each method has advantages and disadvantages. The run method can be calculated quickly in $O(n)$ operations and it has certain desirable robustness properties. It can withstand many isolated outliers and can also be tuned to detect blocks of outliers of a specified length. The disadvantage is a slow rate of convergence. The
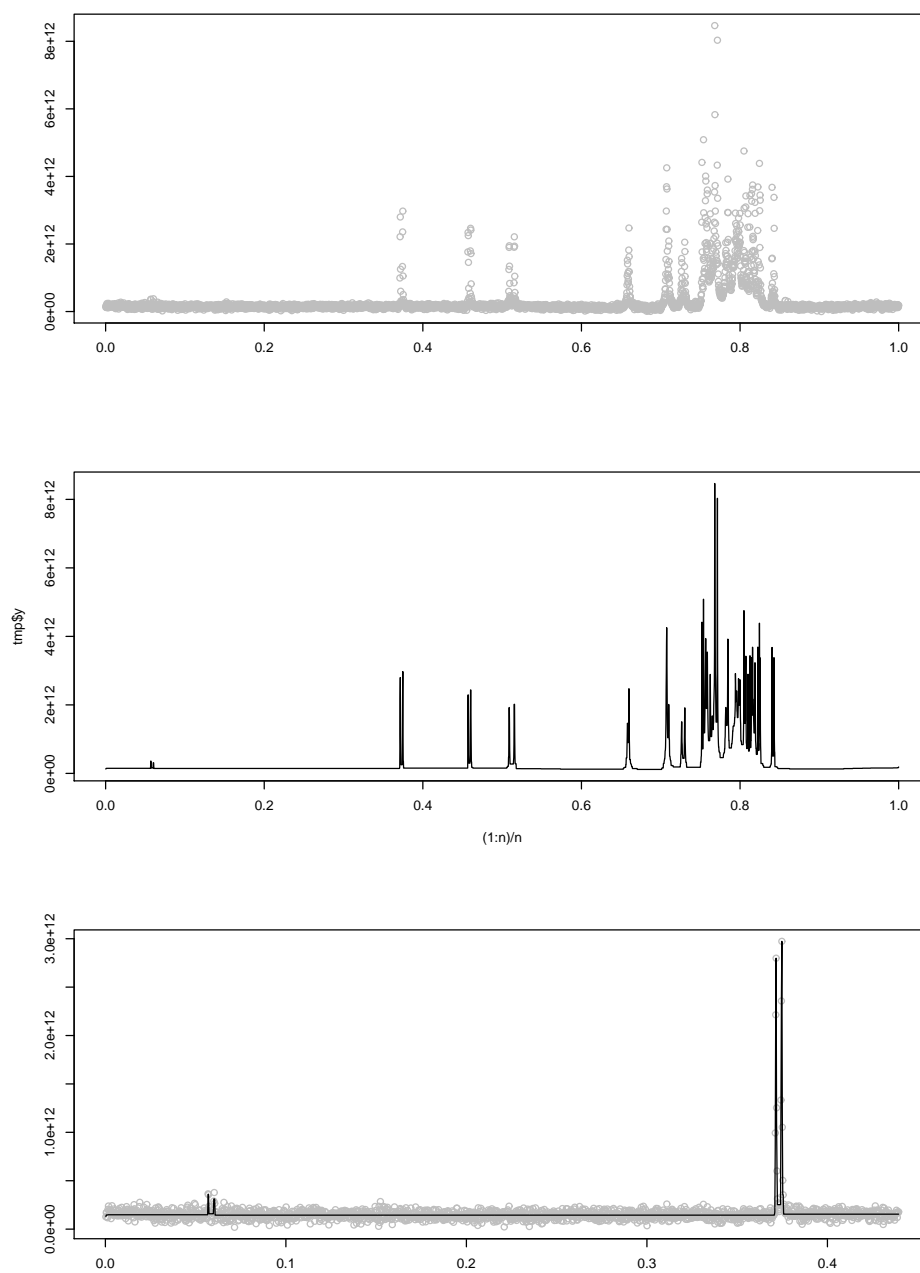
FIG. 16. *Data from Spectroscopy. The upper panel shows data that were gathered during a spectroscopical analysis of a gall stone. The reconstruction is shown in the second panel and an excerpt in the bottom panel.*
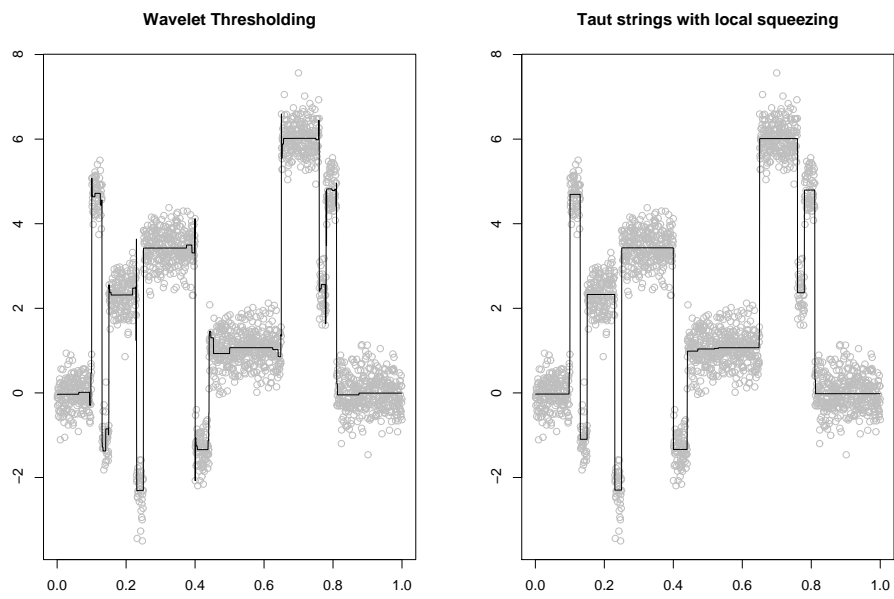
FIG. 17. *Wavelet Thresholding and local squeezing. Both methods were applied to the Blocks data. Wavelet Thresholding tends to produce pseudo Gibbs effects near discontinuities leading to 47 local extreme values. The taut string method with local squeezing attains the correct modality of 9.*

taut-string-multiresolution method has a complexity of $O(n \log n)$ and has almost optimal rates of convergence on standard test beds. It is extremely sensitive being able to identify peaks of very low power. This very sensitivity makes it susceptible to outliers. Implementations are available from our home page at

http://www.stat-math.uni-essen.de.

We are indebted to Lutz Dümbgen and Martin Löwendick for useful comments. We would also like to thank two anonymous referees and in particular an Associate Editor whose comments resulted in an increase in accuracy and clarity.

7. Proofs

7.1. *Proof of Theorem 3* To prove (c) we analyse the behaviour of the upper bound $u_n(\cdot) = u_n(\cdot, \alpha)$ defined by (2.14). We have from (2.17)

$$q_n = qu(n, \alpha, R_n) = \log_2 n + O(1).$$

Let

$$\Gamma_i = \max_{(i-1)q_n+1 \leq j \leq iq_n} \varepsilon(t_j), \quad i = 1, \ldots, \left\lceil \frac{n}{q_n} \right\rceil.$$

If F denotes the common distribution function of the $\varepsilon\left(\frac{i}{n}\right)$ then the common distribution function of the Γ_i is F^{q_n}. We set $a_n = \lceil \frac{n \log \log n}{(\log_2 n)^2} \rceil$ and define

$$\Theta_i = \min_{(i-1)a_n+1 \leq j \leq ia_n} \Gamma_j, \quad i = 1, \ldots, \left\lfloor \frac{\log n}{\log \log n} \right\rfloor.$$

The common distribution function of the Θ_i is $1 - (1 - F^{q_n})^{a_n}$. On using $F(x) = \frac{1}{2} + ax$ for small x with $a > 0$ it follows that

$$\mathbf{P}\left(\Theta_i \geq \frac{b \log \log n}{\log n}\right) = O(\exp(-c(\log n)^{b'}))$$

where $c > 0$ and where b may be chosen so that $b' > 0$. For this choice of b we have

$$\mathbf{P}\left(\max_{1 \leq i \leq \lfloor \log n / \log \log n \rfloor} \Theta_i \geq \frac{b \log \log n}{\log n}\right) = O\left((\log n) \exp(-c(\log n)^{b'})\right).$$

The upper bound u_n is nonincreasing and we firstly analyse its behaviour on an interval where f where $f^{(1)}(t) > \delta > 0$. Without loss of generality we set $I = [0, \frac{j}{n}]$. The above estimates show that

$$(7.35) \qquad u_n(t) \leq f(0) + \frac{A \log \log n}{\log n}, t = \frac{\log \log n}{\log n}.$$

The corresponding result for the lower bound l_n is

$$(7.36) \qquad l_n(t) \geq f\left(\frac{j}{n}\right) - \frac{A \log \log n}{\log n}, t = \frac{j}{n} - \frac{\log \log n}{\log n}.$$

As both the upper and lower bounds are nonincreasing (7.35) and (7.36) imply

$$(7.37) \qquad A \frac{\log \log n}{\log n} + f(0) \geq f\left(\frac{j}{n}\right) - A \frac{\log \log n}{\log n}.$$

As $f^{(1)}(t) > \delta > 0$ on this interval we have

$$(7.38) \qquad \frac{j}{n} \leq A \frac{\log \log n}{\log n}.$$

It follows that if the bounds have the wrong monotonic behaviour this will be detected at latest on an interval of length $\mathrm{O}\left(\frac{\log \log n}{\log n}\right)$ where the constant depends on the size of the derivative of f on the interval.

The case where the bounds have the same monotonic behaviour as f is as follows. Because of the construction of the intervals it is clear that f will remain below the upper bound and above the lower bound with probability at least α. On combining these two results we see the monotonic behaviour of the bounds which minimizes the number of local extreme values will coincide with the monotonic behaviour of f with probability at least α as n tends to infinity. In other words the number of local extreme values will be determined correctly for large n with probability at least α. The reasoning also shows that the lengths of the intervals $I_i^e(n, \alpha)$ tend to zero with n and that the midpoints converge to the local extreme points of f. Finally the reasoning which lead to (7.35) and (7.36) implies the rate of convergence of (b) of the theorem.

To prove (d) of the theorem it is sufficient to consider the upper bound on a stretch where f is monotone decreasing. On writing $t = \frac{i}{n}$ and $h = \frac{j}{n}$ we have

(7.39) $$u_n(t) = \min\{u_n(t-h), M(i), \ldots, M(i-j+1)\}$$

where

$$M(m) = \max\left\{Y\left(\frac{m}{n}\right), \ldots, Y\left(\frac{m-q_n}{n}\right)\right\}.$$

As $Y(t) = f(t) + \varepsilon(t)$ and f is non-increasing it follows that

$$M(m) \geq f(t) + \Lambda(m)$$

where

(7.40) $$\Lambda(m) = \max\left\{\varepsilon\left(\frac{m}{n}\right), \ldots, \varepsilon\left(\frac{m-q_n}{n}\right)\right\}.$$

On substituting this into (7.39) we obtain

$$u_n(t) \geq \min\{u_n(t-h), f(t) + \Lambda(i-j+1), \ldots, f(t) + \Lambda(i)\}.$$

As $f(t-h) \leq u_n(t-h)$ and $f(t-h) = f(t) - hf^{(1)}(t)(1+o(1))$ this implies

(7.41) $$u_n(t) \geq f(t) + \min\{-hf^{(1)}(t)(1+o(1)), \Lambda(i-j+1), \ldots, \Lambda(i)\}.$$

We have

$$\mathbf{P}\left(\Lambda(m) \geq \frac{b\log\log n}{\log n}\right) = 1 - \mathbf{P}\left(\Lambda(m) \leq \frac{b\log\log n}{\log n}\right) = 1 - F\left(\frac{b\log\log n}{\log n}\right)^{q_n}$$

which implies

$$\mathbf{P}\left(\Lambda(m) \geq \frac{b\log\log n}{\log n}\right) \approx 1 - \left(\frac{1}{2} + \frac{ab\log\log n}{\log n}\right)^{q_n}$$

$$\approx 1 - 2^{-q_n}\exp\left(q_n\frac{2ab\log\log n}{\log n}\right)$$

$$\geq 1 - \frac{1}{n}\exp(3ab\log\log n)$$

for large n. From this it follows

$$\mathbf{P}\left(\cup_{m=i-j+1}^{i}\left\{\Lambda(m) \leq \frac{b \log \log n}{\log n}\right\}\right) \leq \frac{j}{n} \exp(3ab \log \log n).$$

If we set $j = \lceil \frac{n \log \log n}{\log n} \rceil$ and choose b so that $3ab < \frac{1}{2}$ it follows that $h \approx \frac{\log \log n}{\log n}$ and

$$\mathbf{P}\left(\min_{i-j+1 \leq m \leq i} \Lambda(m) \geq \frac{b \log \log}{\log n}\right) \geq 1 - \mathrm{O}\left(\frac{\log \log n}{(\log n)^{1/2}}\right).$$

On using this in (7.41) it follows that

$$u_n(t) \geq f(t) + \min\left\{-f^{(1)}(t)\frac{\log \log n}{\log n}, \frac{b \log \log n}{\log n}\right\}$$

with $b > 0$ and $f^{(1)}(t) < 0$. This completes the proof of for the upper bound on stretches where f is monotone decreasing. The other cases follow by appropriate changes of sign. $\quad\square$

7.2. *Proof of Theorem 4*   The assumptions of the theorem imply that the integrated error process

$$\varepsilon_n^\circ(t) = \frac{1}{n}\sum_{1}^{[nt]} \varepsilon\left(\frac{j}{n}\right)$$

converges weakly to a Wiener process on $[0,1]$:

$$\sqrt{n}\varepsilon_n^\circ \Rightarrow \sigma W$$

where W denotes the standard Wiener process. In particular we have

$$\lim_{n \to \infty} \mathbf{P}\left(\max_{0 \leq t \leq 1} |\sqrt{n}\varepsilon_n^\circ(t)| \leq x\right) = \mathbf{P}\left(\max_{0 \leq t \leq 1} |W(t)| \leq \frac{x}{\sigma}\right) = H\left(\frac{x}{\sigma}\right).$$

It follows that on the test bed (1.8)

$$\lim_{n \to \infty} \mathbf{P}\left(\max_{0 \leq t \leq 1} |Y_n^\circ(t) - f(t)| \leq \frac{C}{\sqrt{n}}\right) = H\left(\frac{C}{\sigma}\right).$$

As n tends to infinity the probability that the function f lies in the tube $T(Y_n^\circ, C/\sqrt{n})$ tends to $H\left(\frac{C}{\sigma}\right)$. As the taut string minimizes the number of local

extremes in $T(Y_n^\circ, C/\sqrt{n})$ we see that

$$\lim_{n\to\infty} \mathbf{P}(K_n^C \le k) = H\left(\frac{C}{\sigma}\right).$$

7.3. *Proof of Theorem 5* We note that for f satisfying the assumptions of the theorem

$$\inf_{g\in M_{k-1}} \sup_{0\le t\le 1} |f^\circ(t) - g^\circ(t)| > 0$$

where M_j denotes the set of functions on $[0,1]$ with at most j local extremes. This implies

$$\lim_{n\to\infty} \mathbf{P}(K_n^C < k) = 0 \quad \text{for all} \quad C > 0.$$

In the other direction Theorem 4 implies

$$\lim_{C\to\infty} \lim_{n\to\infty} \mathbf{P}(K_n^C \le k) = 1.$$

and hence

$$\lim_{C\to\infty} \lim_{n\to\infty} \mathbf{P}(K_n^C = k) = 1.$$

The other claims are proved similarly. If the lengths of the intervals $I_i^e(n, C)$ do not converge in probability to to zero then

$$\max_{t\in I_i^e(n,C)} |S_n(t) - f(t)|$$

does not converge in probability to zero and this carries over to the integrated functions. A similar argument applies to the convergence of the location of the local extreme points. $\square$

7.4. *Proof of Theorem 6* The proof of Theorem 6 relies on the modulus of continuity of the integrated process ε_n° as expressed by

$$(7.42) \quad \lim_{\delta\to 0} \mathbf{P}\left(\sup_{0\le t, t+h\le 1, h\le\delta} \sqrt{n}|\varepsilon_n^\circ(t+h) - \varepsilon_n^\circ(t)| \le A\sqrt{-\delta\log\delta}\right) = 1$$

for some $A > 0$. This follows from the subgaussian form of the $\varepsilon(t)$. This will be used for $\delta = \delta_n$ of the form $(\log n)^\alpha n^{-\beta}$ with $\alpha \geq 0$ and $\beta > 0$ in which case we have

$$(7.43) \quad \lim_{n \to \infty} \mathbf{P} \left(\sup_{0 \leq t, t+h \leq 1, h \leq \delta_n} \sqrt{n} |\varepsilon_n^\circ(t + h) - \varepsilon_n^\circ(t)| \leq A \sqrt{-\delta_n \log \delta_n} \right) = 1.$$

The proofs given below require that the taut string through the tube of radius $C/\sqrt{n}$ has the correct shape i.e. it has the same number of local extremes as f. In order for this to happen in probability as n tends to infinity we have to let C tend to infinity. The order is first n to infinity and then C to infinity. As the distribution of the maximum of a Brownian motion on the interval $[0, 1]$ behaves exactly like the tails of the standard normal distribution the probabilities tend very quickly to one as C increases. In practice $C = 3\sigma_n$ with σ_n a robust scale functional such as (1.7) is a large value of C.

The proofs can probably be best understood by taking a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and considering a sequence of events $\Omega^0(C, n)$ which are chosen as the need arises but which satisfy

$$\lim_{C \to \infty} \lim_{n \to \infty} \mathbf{P}(\Omega^0(C, n)) = 1.$$

We can for example put

$$\Omega^1(C, n) = \{\omega : \sup_{0 \leq t \leq 1} |Y_n^\circ(t) - f^\circ(t)| \leq \frac{C}{\sqrt{n}}\}.$$

Then clearly

$$\lim_{C \to \infty} \lim_{n \to \infty} \mathbf{P}(\Omega^1(C, n)) = 1.$$

From Theorem 5 it follows that there exists a sequence $(\delta_n)_1^\infty$ tending to zero such that

$$\lim_{C\to\infty}\lim_{n\to\infty}\mathbf{P}(\Omega^2(C,n))=1$$

where

$$\Omega^3(C,n)=\{\omega:K_n^C=k,\max_{1\le i\le k}|I_i^e(n,C)|\le\delta_n,\max_{1\le i\le k}|\tau_i^e(n,C)-t_i^e|\le\delta_n\}.$$

Similarly for another given sequence $(\delta_n)_1^\infty$ tending to zero we may define

$$\Omega^3(A,n)=\{\omega:\sup_{0\le t,t+h\le1,h\le\delta}\sqrt{n}|\varepsilon_n^\circ(t+h)-\varepsilon_n^\circ(t)|\le A\sqrt{-\delta_n\log\delta_n}\}.$$

For some appropriate $A>0$ we have $\lim_{n\to\infty}\mathbf{P}(\Omega^3(n))=1$. Such choices are made according to our m needs and finally we set

$$\Omega^0(C,n)=\Omega^1(C,n)\cap\Omega^2(C,n)\cap\ldots\cap\Omega^m(C,n).$$

For all $\omega\in\Omega^0(C,n)$ all relevant properties hold such as, for example, $\tilde{S}_n$ having the same monotonicity behaviour as f. The terms O and o used below can be translated into inequalities will hold for each n and for all $\omega\in\Omega^0(C,n)$.

Proof of (a):

Suppose S_n° is initially convex. Then S_n° is the largest convex minorant of $Y_n^\circ+C/\sqrt{n}$ until it reaches the left endpoint $t_1^l(n,C)$ of $I_1^e(n,C)=(t_1^l(n,C),t_1^r(n,C))$. Then with $h_0=t_1^r(n,C)-t_1^l(n,C)$ we have

$$(7.44)\qquad h_0=\mathrm{argmax}_{0\le h\le\delta}\frac{Y_n^\circ(t_1^l(n,C)+h)-Y_n^\circ(t_1^l(n,C))-\frac{2C}{\sqrt{n}}}{h}$$

for arbitrarily small δ as n tends to infinity. On writing $t_1^l(n,C)=t_1^e-\kappa$ we may rewrite (7.44) to obtain

$$h_0=\mathrm{argmax}_{0\le h\le\delta}\frac{Y_n^\circ(t_1^e-\kappa+h)-Y_n^\circ(t_1^e-\kappa)-\frac{2C}{\sqrt{n}}}{h}.$$

58

A Taylor expansion together with the modulus of continuity of ε_n° gives

$$h_0 = \operatorname{argmax}_{0 \le h \le \delta} \left(-\frac{1}{6} h(3\kappa - h) f^{(2)}(t_1^e)(1 + \mathrm{o}(1)) - \frac{2C}{\sqrt{n} h}(1 + \mathrm{o}(1)) \right).$$

This implies

(7.45) $$\qquad -\frac{1}{6}(3\kappa - 2h_0) f^{(2)}(t_1^e)(1 + \mathrm{o}(1)) = -\frac{2C}{\sqrt{n} h_0^2}(1 + \mathrm{o}(1))$$

and as $f^{(2)}(t_1^e) < 0$ we may conclude $3\kappa \le 2h_0(1 + \mathrm{o}(1))$ which implies $t_1^e < t_1^r(n, C)$.
In the other direction we have

$$h_0 = \operatorname{argmax}_{0 \le h \le \delta} \frac{Y_n^\circ(t_1^r(n, C)) - Y_n^\circ(t_1^r(n, C) - h) - \frac{2C}{\sqrt{n}}}{h}.$$

On writing $t_1^r(n, C) = t_1^e + \kappa^*$ we obtain

(7.46) $$\qquad -\frac{1}{6}(3\kappa^* - 2h_0) f^{(2)}(t_1^e)(1 + \mathrm{o}(1)) = -\frac{2C}{\sqrt{n} h_0^2}(1 + \mathrm{o}(1)).$$

This implies $3\kappa \le 2h_0(1 + \mathrm{o}(1))$ and hence $t_1^e > t_1^l(n, C)$ so that t_1^e lies in the interval
$[t_1^l(n, C), t_1^r(n, C)]$ as was to be proved.

Proof of (b):

On adding (7.45) and (7.46) we obtain

$$-(3(\kappa + \kappa^*) - 2h_0) f^{(2)}(t_1^e)(1 + \mathrm{o}(1)) = -\frac{24C}{\sqrt{n} h_0^2}(1 + \mathrm{o}(1))$$

and as $\kappa + \kappa^* = h_0$ this implies

$$h_0 \sim (24C)^{1/3} |f^{(2)}(t_1^e)|^{-1/3} n^{-1/6}.$$

Proof of (c):

It is sufficient to consider x_1 and x_2 and to suppose that f° and S_n° are both convex
on (x_1, x_2). On writing $x_1 = i_1/n$ and $x_2 = (i_1 + l_1)/n$ we see that l_1 is the local
argmin of

$$\frac{Y_n^\circ(x_1 + \frac{l}{n}) - Y_n^\circ(x_1)}{\frac{l}{n}}.$$

A Taylor series expansion gives

$$\frac{Y_n^{\circ}(x_1 + \frac{l}{n}) - Y_n^{\circ}(x_1)}{\frac{l}{n}} = f(x_1) + \frac{1}{2}f^{(1)}(x_1)\frac{l}{n} + \frac{\varepsilon_n^{\circ}(x_1 + \frac{l}{n}) - \varepsilon_n^{\circ}(x_1)}{\frac{l}{n}} + O\left(\left(\frac{l}{n}\right)^2\right).$$

The modulus of continuity (7.42) of $\sqrt{n}\varepsilon_n^{\circ}$ implies

$$\sqrt{n}\left|\varepsilon_n^{\circ}(t + \frac{l}{n}) - \varepsilon_n^{\circ}(t)\right| \leq A\sqrt{\frac{l\log n}{n}}$$

uniformly in t and $l, 1 \leq l \leq n^{9/10}$, say. On setting

$$l = a|f^{(1)}(x_1)|^{-2/3}n^{2/3}(\log n)^{1/3}$$

we have

$$\frac{Y_n^{\circ}(x_1 + \frac{l}{n}) - Y_n^{\circ}(x_1)}{\frac{l}{n}} \geq f(x_1) + \frac{1}{2}a|f^{(1)}(x_1)|^{1/3}\left(\frac{\log n}{n}\right)^{1/3}$$

$$-\frac{A}{\sqrt{a}}|f^{(1)}(x_1)|^{1/3}\left(\frac{\log n}{n}\right)^{1/3}$$

$$+O\left(|f^{(1)}(x_1)|^{-4/3}\left(\frac{\log n}{n}\right)^{2/3}\right).$$

From (b) of the theorem $|f^{(1)}(x_1)| \geq An^{-1/6}$ for $C \geq C_0$ and consequently the term

$$O\left(|f^{(1)}(x_1)|^{-4/3}\left(\frac{\log n}{n}\right)^{2/3}\right)$$

may be neglected. This implies that for a sufficiently large

$$(7.47) \qquad \frac{Y_n^{\circ}(x_1 + \frac{l}{n}) - Y_n^{\circ}(x_1)}{\frac{l}{n}} \geq f(x_1) + \frac{1}{4}a|f^{(1)}(x_1)|^{1/3}\left(\frac{\log n}{n}\right)^{1/3}.$$

The lower bound

$$(7.48) \qquad \frac{Y_n^{\circ}(x_1 + \frac{l}{n}) - Y_n^{\circ}(x_1)}{\frac{l}{n}} \leq f(x_1) + a|f^{(1)}(x_1)|^{1/3}\left(\frac{\log n}{n}\right)^{1/3}.$$

is obtained analogously. On putting $a = a_1$ in (7.47) and $a = a_2$ in (7.48) with $a_1 = 4a_2$ with a_2 sufficiently large it follows that the local minimum is attained at

a point $x_1 + \frac{l}{n}$ with

$$\frac{l}{n} = O\left(|f^{(1)}(x_1)|^{-2/3}\left(\frac{\log n}{n}\right)^{1/3}\right).$$

This proves (c) of the theorem.

Proof of (d):

We consider first the case where $t = x_i$ is a knot which does not delimit the position

of a local extreme value of S_n. We take S_n° to be convex at x_i. We have

$$S_n(x_i) \le \frac{Y_n^\circ(x_i + \frac{l}{n}) - Y_n^\circ(x_i)}{\frac{l}{n}}.$$

A Taylor expansion of order two combined with (7.42) gives for

$$l = Af^{(1)}(x_i)^{-2/3}n^{2/3}(\log n)^{1/3}$$

(7.49) $$S_n(x_i) \le f(x_i) + A|f^{(1)}(x_i)|^{1/3}\left(\frac{\log n}{n}\right)^{1/3}.$$

Using

$$S_n(x_i) \ge \frac{Y_n^\circ(x_i) - Y_n^\circ(x_i - \frac{l}{n})}{\frac{l}{n}}$$

a similar argument gives

$$S_n(x_i) \ge f(x_i) - A|f^{(1)}(x_i)|^{1/3}\left(\frac{\log n}{n}\right)^{1/3}$$

which when combined with (7.49) gives

$$|f(x_i) - S_n(x_i)| = O\left(|f^{(1)}(x_i)|^{1/3}\left(\frac{\log n}{n}\right)^{1/3}\right)$$

at all knots x_i which do not delimit a local extreme value of S_n. For a point t not

in $[A\left(\frac{\log n}{n}\right)^{1/3}, 1 - A\left(\frac{\log n}{n}\right)^{1/3}]\backslash\cup_{i=1}^{k} I_i^e(n, C)$ we have

$$|f(t) - S_n(t)| = |f(t) - S_n(x_i)|$$

$$\le |f(x_i) - S_n(x_i)| + |f(t) - f(x_i)|$$

$$\le |f(x_i) - S_n(x_i)| + A|f^{(1)}(x_i)||f^{(1)}(x_i)|^{-2/3}\left(\frac{\log n}{n}\right)^{1/3}$$

$$\leq A|f^{(1)}(t)|^{1/3} \left(\frac{\log n}{n}\right)^{1/3}$$

where we have used

$$\sup_{x_i \leq t \leq x_{i+1}} \left|\frac{f^{(1)}(x_i)}{f^{(1)}(t)} - 1\right| \leq A$$

for all intervals $[x_i, x_{i+1}]$ which do not delimit a local extreme value. This follows from (b) of the theorem.

Proof of (e):

This follows as in the other cases but using the next term of the Taylor expansion as $f^{(1)}(t) = 0$ for some point in the interval. $\quad\square$

7.5. *Proof of Theorem 7   Proof of (a):*

The multiresolution coefficient $w_{j,k}$ is given by

$$(7.50)\ w_{j,k} = \frac{\sqrt{n}}{2^{j/2}} \cdot \left(Y_n^\circ\left(\frac{k2^j}{n}\right) - \tilde{S}_n^\circ\left(\frac{k2^j}{n}\right) - Y_n^\circ\left(\frac{(k+1)2^j}{n}\right) + \tilde{S}_n^\circ\left(\frac{(k+1)2^j}{n}\right)\right).$$

To show that all multiresolution coefficients of the noise are smaller than the bound (3.19) for some $\tau > 0$ it is sufficient to show that each coefficient $w_{i,k}$ is at most $\sigma\sqrt{\tau \log n}$. This in turn follows from the inequality

$$(7.51) \qquad |Y_n^\circ(t_l) - \tilde{S}_n^\circ(t_l) - (Y_n^\circ(t_i) - \tilde{S}_n^\circ(t_i))| \leq A\sqrt{|t_l - t_i|}\sqrt{\frac{\log n}{n}}$$

where (t_i, t_l) is one of the pairs $((k2^j - 2^{j-1})/n, k2^j/n), (k2^j/n, (k2^j + 2^{j-1})/n)$. We have

$$Y_n^\circ - \tilde{S}_n^\circ = \varepsilon_n^\circ + f^\circ - \tilde{S}_n^\circ.$$

and on using the modulus of continuity of $\sqrt{n}\varepsilon_n^\circ$ as given by (7.42) we obtain for any points t_i and t_l with $|t_l - t_i| \geq 1/n$

$$|Y_n^\circ(t_l) - \tilde{S}_n^\circ(t_l) - (Y_n^\circ(t_i) - \tilde{S}_n^\circ(t_i))| \leq A\sqrt{|t_l - t_i|}\sqrt{\frac{\log n}{n}}$$

$$+|f^\circ(t_l) - \tilde{S}_n^\circ(t_l) - (f^\circ(t_i) - \tilde{S}_n^\circ(t_i))|.$$

(7.52)

As

$$|f^\circ(t_l) - \tilde{S}_n^\circ(t_l) - (f^\circ(t_i) - \tilde{S}_n^\circ(t_i))| = |\int_{t_i}^{t_l}(f(t) - \tilde{S}_n(t))dt|$$

$$\leq |t_l - t_i| \sup |f(t) - \tilde{S}_n(t)|$$

(c) and (d) of Theorem 6 imply

$$|f^\circ(t_m) - \tilde{S}_n^\circ(t_m) - (f^\circ(t_i) - \tilde{S}_n^\circ(t_i))| \leq A|t_m - t_i| \left(\frac{\log n}{n}\right)^{1/3}$$

$$\leq A\sqrt{|t_l - t_i|}\sqrt{\frac{\log n}{n}}$$

if

$$(7.53) \qquad\qquad |t_l - t_i| \leq A\left(\frac{\log n}{n}\right)^{1/3}.$$

Thus (7.51) holds for all intervals satisfying (7.53).

Proof of (b):

This follows from (a) and and Theorem 6 (c).

Proof of (c):

Consider a multiresolution coefficient whose support is $[t_1, t_2]$ with $t_1 \in I_i^e(n, C)$ and $t_2 \notin I_i^e(n, C)$ for some i. the other outside of the same interval. Let x_1 be the right endpoint of $I_i^e(n, C)$. It follows from the definition of $\tilde{S}_n^\circ$ that $|Y_n^\circ(t) - \tilde{S}_n^\circ(t)| \leq C/\sqrt{n}$ for all $t \in I_i^e(n, C)$. In particular we have

$$|Y_n^\circ(x_1) - \tilde{S}_n^\circ(x_1) - (Y_n^\circ(t_1) - \tilde{S}_n^\circ(t_1))| \leq \frac{2C}{\sqrt{n}}.$$

This is also the worst case for the other point t_2 and the result is the inequality

$$|Y_n^\circ(t_2) - \tilde{S}_n^\circ(t_2) - (Y_n^\circ(t_1) - \tilde{S}_n^\circ(t_1))| \leq \frac{4C}{\sqrt{n}}.$$

The multiresolution coefficient will satisfy (7.51) if

$$t_2 - t_1 \geq \frac{16C^2}{\tau \log n}$$

as was to be shown.

Proof of (d):

Consider a multiresolution coefficient whose support is contained in $I_1^e(n,c)$ with length $h \geq \frac{1}{2}|I_1^e(n,C)|$. From Theorem 6 we have $h \approx c_1 n^{-1/6}$ and $|f(t) - \tilde{s}_n(t)| \geq c_2 n^{-1/3}$ for some constants $c_1 > 0$ and $c_2 > 0$. Without loss of generality we may assume that $f(t) - \tilde{s}_n(t) \geq c_2 n^{-1/3}$ on the support of the multireslution coefficient. This implies that the absolute value of the multiresolution coefficient is at least $c_3 n^{-1/2}$ for some $c_3 > 0$ which eventually exceeds the bound

$$\sqrt{\frac{\tau h \log n}{n}} \leq c_4 n^{-7/12} \sqrt{\tau \log n}$$

whatever the value of τ. $\square$

8. Appendix: the taut string algorithm We give a short description of the taut string algorithm. Without loss of generality we suppose the data points are $(i, y(i)), i = 0, \ldots, n$. Let l and u denote the lower and upper bounds for the integrated process y° given by

$$y^\circ(k) = \frac{1}{n} \sum_{i=1}^{k} y(i), \quad 0 \leq k \leq n.$$

We assume that the endpoints of the string are fixed i.e. $l(0) = u(0)$ and $l(n) = u(n)$. Starting from the point $(0, l(0))$ and for given i the greatest convex minorant sx_i of the upper bound $(j, u(j)), j = 0, \ldots, i$ and the smallest concave majorant sv_i of the lower bound $(j, l(j)), j = 1, \ldots, i$ are both calculated. The greatest convex minorant sx_i may be calculated as follows (see also Barlow et al. 1972). Suppose sx_i has been calculated. It is defined by knots $(k(j), u(k(j))), 1 \leq j \leq K(i)$ where it touches the

(a)

(b)

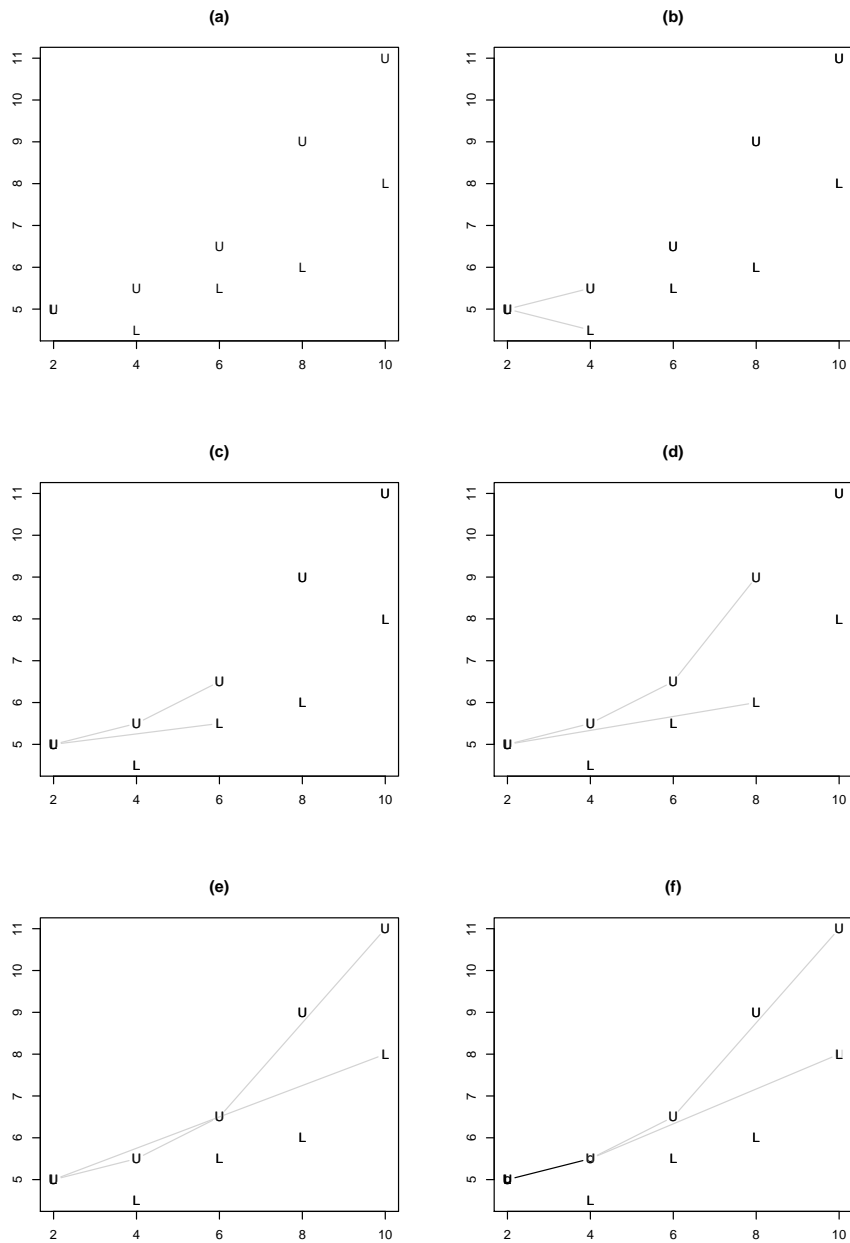**(c)**

(d)

(e)

(f)

Fig. 18. *Example of the taut string method. Moving from the left to the right, the greatest convex minorants of the upper bounds and the smallest concave majorant of the lower bounds up to the current position are calculated until both curves intersect. The leftmost knot is added to the taut string.*

upper bound and it is linear in between. The first and last points are knots. The line joining $(i+1, u(i+1))$ is now included as a knot. If the resulting linear interpolation is convex then this is sx_{i+1}. If not then the knot to the left of $(i+1, u(i+1))$ is permanently eliminated. If the resulting linear interpolation is convex this is sx_{i+1}. If it is not convex then again the knot to the left of $(i+1, u(i+1))$ is eliminated. This process is continued until a convex interpolation is attained and this defines sx_{i+1}. This algorithm has complexity $O(i)$ for calculating all the functions sx_j and sv_j $0 \leq j \leq i$. It is clear that the taut string must lie between the sx_i and sv_i. The functions sx_i and sv_i are linear between the knots where they touch the upper and lower bounds respectively. After sx_i and sv_i have been calculated it is checked whether

$$(8.54) \qquad\qquad sx_i^{(1)}(0+) > sv_i^{(1)}(0+)$$

holds where $sx_i^{(1)}(0+)$ and $sv_i^{(1)}(0+)$ denote respectively the right hand derivatives at 0 of sx_i and sv_i. These are nothing more than the gradients of the first sections of the functions sx_i and sv_i. If (8.54) holds then sx_{i+1} and sv_{i+1} are calculated. This process is continued until for some i (8.54) does not hold. At this point the leftmost knot of the set of knots of both the sx_i and sv_i is determined whereby the very first knot $(0, u(0))$ is not counted. This leftmost knot is the first knot of the taut string and we denote it by $(k_1, s(k_1))$. It may be either a knot from the upper bound or from the lower bound. Figure 18 demonstrates this procedure.

The process is now repeated. The origin is moved to the point $(k_1, 0)$ with translated upper bounds $(j, u(k_1 + j))$, $j = 0, \ldots n - k_1$ and lower bounds $(j, l(k_1 + j))$, $j = 0, \ldots n - k_1$. Starting from the point $(0, s(k_1))$ the greatest convex minorant of the upper bounds and the smallest concave majorant of the lower bounds are

calculated as before. The important point to notice is that these calculations have already been done for the first $i - k_1$ points. If the first knot of the string is on the upper bound i.e. $(k_1, s(k_1)) = (k_1, u(k_1))$ then the convex minorant for the section $0 \leq j \leq i - k_1$ of the translated bounds coinicides with the translation of convex minorant already calculated. The concave majorant is simply the straight line joining $(0, s(k_1))$ and $(i - k_1, l(i))$. A corresponding statement holds if the first knot is on the lower bound. It is this which makes the calculation of the taut string $O(n)$.

The calculations just described are continued until the last point is reached. The taut string is now determined by linear interpolation between its knots.

REFERENCES

Barlow, R., Bartholomew, D., Bremner, J., and Brunk, H. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York.

Chaudhuri, P. and Marron, J. S. (1999). Sizer for exploration of structures in curves. *Journal of the American Statistical Association*, 94:807–823.

Davies, P. (1995). Data features. *Statistica Neerlandica*, 49:185–245.

Davies, P. L. (2000). Hidden periodicities and strings. Department of Mathematics and Computuer Science, University of Essen, Germany.

Davies, P. L. and Löwendick, M. (1999). On smoothing under bounds and geometric constraints. Technical report, University of Essen.

Delecroix, M., Simioni, M., and Thomas-Agnan, C. (1995). A shape constrained smoother: simulation study. *Computational Statistics*, 10:155–175.

Donoho, D. (1988). One-sided inference about functionals of a density. *Annals of Statistics*, 16:1390–1420.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Pic ard, D. (1995). Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society*, 57:371–394.

Dümbgen, L. (1998a). Application of local rank tests to nonparametric regression. Medical University, Lübeck.

Dümbgen, L. (1998b). New goodness-of-fit tests and their application to nonparametric confidence sets. *Annals of Statistics*, 26:288–314.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaption. *Journal of the Royal Statistical Society B*, 57:371–394.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications: Volume 1*. Wiley, New York, 3 edition.

Fisher, N. I., Mammen, E., and Marron, J. S. (1994). Testing for multimodality. *Computational Statistics and Data Analysis*, 18:499–512.

Freedman, D. (1971). *Browniam Motion and Diffusion*. Holden-Day, San Francisco, 3 edition.

Good, I. and Gaskins, R. (1980). Density estimating and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. *Journal of the American Statistical Association*, 75:42–73.

Green, P. and Silverman, B. (1994). *Nonparametric regression and Generalized Linear Models: a roughness penalty approach*. Number 58 in Monographs on Statistics and Applied Probability. Chapman and Hall, London.

Groeneboom, P. (1985). Esimating a monotone density. In LeCam, L. and Olshen, R., editors, *Proceedings, Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume II, Monterey, CA. Wadsworth.

Hartigan, J. and Hartigan, P. (1985). The dip test of unimodality. *Annals of Statistics*, 13:70–84.

Hartigan, J. A. (1987). Esimation of a convex density contour in two dimensions. *Journal of the American Statistical Association*, 82:267–270.

Hengartner, N. and Stark, P. (1995). Finite-sample confidence envelopes for shape-restricted densities. *Annals of Statistics*, 23:525–550.

Ibragimov, I. and Khas'minskii, R. (1980). On non-parametric estimation of regression. *Soviet Math. Dockl.*, 21:810–814.

Kahane, J.-P. (1968). *Some random series of functions*. Heath Mathematical Monographs. Heath, Lexington, Massachusetts.

68

Khas'minskii, R. (1978). A lower bound on the risks of non-parametric estimates of densities in the uniform metric. *Theory Prob. Appl.*, 23:794–798.

Kovac, A. and Silverman, B. W. (2000). Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *Journal of the American Statistical Association*, 95:172–183.

Leurgans, S. (1982). Asymptotic distributions of slope-of-greatest-convex-minorant estimators. *Annals of Statistics*, 10:287–296.

Majidi, A. (2000). Nichtparametrische regression unter modalitätskontrolle. Diplomarbeit, Fachbereich Mathematik und Informatik, Universität Essen.

Mammen, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Annals of Statistics*, 19:741–759.

Mammen, E., Marron, J., Turlach, B., and Wand, M. (1998). A general framework for constrained smoothing. unpublished.

Mammen, E. and Thomas-Agnan, C. (1998). Smoothing splines and shape restrictions. unpublished.

Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Annals of Statistics*, 25:387–413.

Mächler, M. (1995). Variational solution of penalized likelihood problems and smooth curve estimation. *Annals of Statistics*, 23:1496–1517.

Metzner, L. (1997). *Facettierte nichtparametrische Regression*. PhD thesis, Universität Essen, Essen, Germany.

Minotte, M. C. (1997). Nonparametric testing of the existence of modes. *Annals of Statistics*, 25:1646–1660.

Minotte, M. C. and Scott, D. W. (1993). The mode tree: a tool for visualization of nonparametric density features. *Journal of Computational and Graphical Statistics*, 2:51–68.

Morgenthaler, S. and Tukey, J. (1991). *Configural Polysampling: A Route to Practical Robustness*. Wiley, New York.

Müller, D. and Sawitzki, G. (1991). Excess mass estimates and tests of multimodality. *Journal of the American Statistical Association*, 86:738–746.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, 10:186–190.

Polonik, W. (1995a). Density estimation under qualative assumptions in higher dimensions. *Journal of Mulitivariate Analysis*, 55:61–81.

Polonik, W. (1995b). Measuring mass concentrations and estimating density contour clusters–an excess mass approach. *Annals of Staistics*, 23:855–881.

Polonik, W. (1999). Concentration and goodness of fit in higher dimensions: (aymptotically) distribution-free methods. *Annals of Statistics*, 27:1210–1229.

Polzehl, J. and Spokoiny, G. (2000). Adaptive weights smoothing with applications to image resoration. *Journal of the Royal Statistical Society, Series B*, 62:335–354.

Ramsay, J. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society*, 60:365–375.

Robertson, T. (1967). On estimating a density measurable with respect to a σ–lattice. *Annals of Mathematical Statistics*, 38:482–493.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order restricted statistical inference.* Wiley, New York.

Sager, T. W. (1979). An iterative method for estimating a multivariate mode and isopleth. *Journal of the American Statistical Association*, 74:329–339.

Sager, T. W. (1982). Nonparametric maximum likelihood estimation of spatial patterns. *Annals of Statistics*, 10:1125–1136.

Sager, T. W. (1986). An application of isotonic regression to multivariate density estimation. In Dykstra, R. L., Robertson, T., and Wright, F. T., editors, *Advances in order restricted statistical inference*. Springer-Verlag.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society*, 47:1–52.

Silverman, B. W. (1986). *Density estimation for statistics amd data analysis*. Chapman and Hall, London.

Stone, C. (1982). Optimal rates of convergence for nonparametric regression. *Annals of Statistics*, 10:1040–1053.

Tukey, J. (1993). Exploratory analysis of variance as providing examples of strategic choices. In
S.Morgenthaler, E.Ronchetti, and W.A.Stahel, editors, *New Directions in Statistical
Data Analysis and Robustness*, Basel. Birkhäuser.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā*, 26:101–116.

Wegman, E. J. (1970). Maximum likelihood estimation of a unimodal density. *Annals of
Mathematical Statistics*, 41:457–471.