NONPARAMETRIC REGRESSION, CONFIDENCE REGIONS AND REGULARIZATION

BY P. L. DAVIES,[1] A. KOVAC AND M. MEISE[1]

University of Duisburg–Essen and Technical University of Eindhoven, University of Bristol and University of Duisburg–Essen

In this paper we offer a unified approach to the problem of nonparametric regression on the unit interval. It is based on a universal, honest and nonasymptotic confidence region $\mathcal{A}_n$ which is defined by a set of linear inequalities involving the values of the functions at the design points. Interest will typically center on certain simplest functions in $\mathcal{A}_n$ where simplicity can be defined in terms of shape (number of local extremes, intervals of convexity/concavity) or smoothness (bounds on derivatives) or a combination of both. Once some form of regularization has been decided upon the confidence region can be used to provide honest nonasymptotic confidence bounds which are less informative but conceptually much simpler.

1. Introduction. Nonparametric regression on the unit interval is concerned with specifying functions $\tilde{f}_n$ which are reasonable representations of a data set $\mathbf{y}_n = \{(t_i, y(t_i)), i = 1, \ldots, n\}$. The design points t_i are assumed to be ordered. Here and below we use lower case letters to denote generic data and upper case letters to denote data generated under a specific stochastic model. The first approach to the problem used kernel estimators with a fixed bandwidth [Watson (1964)] but since then many other procedures have been proposed. We mention splines [Green and Silverman (1994), Wahba (1990)], wavelets [Donoho and Johnstone (1994)], local polynomial regression [Fan and Gijbels (1996)], kernel estimators with local bandwidths [Wand and Jones (1995)] very often with Bayesian and non-Bayesian versions.

The models on which the methods are based are of the form

$$(1) \qquad Y(t) = f(t) + \sigma(t)\varepsilon(t), \qquad t \in [0, 1],$$

with various assumptions being made about $\sigma(t)$, the noise $\varepsilon(t)$ as well as the design points $\{t_1, \ldots, t_n\}$. We shall restrict attention to the simplest case

$$(2) \qquad Y(t) = f(t) + \sigma Z(t), \qquad t \in [0, 1],$$

where Z is Gaussian white noise and the t_i are given by $t_i = i/n$. We mention that the same ideas can be used for the more general model (1) and that robust versions are available. The central role in this paper is played by a confidence region $\mathcal{A}_n$ which is defined below. It specifies all functions $\tilde{f}_n$ for which the model (2) is consistent (in a well-defined sense) with the data $\mathbf{y}_n$. By regularizing within $\mathcal{A}_n$ we can control both the shape and the smoothness of a regression function and provide honest nonasymptotic confidence bounds.

The paper is organized as follows. In Section 2 we define the confidence region $\mathcal{A}_n$ and show that it is universal, honest and nonasymptotic for data generated under (2). In Section 3 we consider shape regularization, in Section 4 regularization by smoothness and the combination of shape and smoothness regularization. Finally, in Section 5 we show how honest and nonasymptotic confidence bounds can be obtained both for shape and smoothness regularization.

2. The confidence region $\mathcal{A}_n$.

2.1. *Nonparametric confidence regions.* Much attention has been given to confidence sets in recent years. These sets are often expressed as a ball centred at some suitable estimate [Li (1989), Hoffmann and Lepski (2002), Baraud (2004), Cai and Low (2006), Robins and van der Vaart (2006)] with particular emphasis on adaptive methods where the radius of the ball automatically decreases if f is sufficiently smooth. The concept of adaptive confidence balls is not without conceptual difficulties as the discussion of Hoffmann and Lepski (2002) shows. An alternative to smoothness is the imposition of shape constraints such as monotonicity and convexity [Dümbgen (1998, 2003), Dümbgen and Spokoiny (2001), Dümbgen and Johns (2004), Dümbgen (2007)]. Such confidence sets require only that f satisfy the shape constraint which often has some independent justification.

We consider data $\mathbf{Y}_n = \mathbf{Y}_n(f)$ generated under (2) and limit attention to functions f in some family $\mathcal{F}_n$. We call a confidence set $\mathcal{C}_n(\mathbf{Y}_n(f), \alpha)$ exact if

$$(3) \qquad P\big(f \in \mathcal{C}_n(\mathbf{Y}_n(f), \alpha)\big) = \alpha \qquad \text{for all } f \in \mathcal{F}_n,$$

honest [Li (1989)] if

$$(4) \qquad P\big(f \in \mathcal{C}_n(\mathbf{Y}_n(f), \alpha)\big) \geq \alpha \qquad \text{for all } f \in \mathcal{F}_n,$$

and asymptotically honest if

$$(5) \qquad \liminf_{n \to \infty} \inf_{f \in \mathcal{F}_n} P\big(f \in \mathcal{C}_n(\mathbf{Y}_n(f), \alpha)\big) \geq \alpha$$

holds, but it is not possible to specify the n_0 for which the coverage probability exceeds $\alpha - \epsilon$ for all $n \geq n_0$. Finally, we call $\mathcal{C}_n(\mathbf{Y}_n(f), \alpha)$ universal if $\mathcal{F}_n = \{f : f : [0, 1] \to \mathbb{R}\}$.

2.2. *Definition of $\mathcal{A}_n$*. The confidence region $\mathcal{A}_n$ we use was first given in Davies and Kovac (2001). It is constructed as follows. For any function $g : [0, 1] \rightarrow \mathbb{R}$ and any interval $I = [t_j, t_k]$ of $[0, 1]$ with $j \le k$ we write

$$(6) \qquad w(\mathbf{y}_n, g, I) = \frac{1}{\sqrt{|I|}} \sum_{t_i \in I} (y(t_i) - g(t_i))$$

where $|I|$ denotes the number of points t_i in I. With this notation,

$$(7) \qquad \mathcal{A}_n = \mathcal{A}_n(\mathbf{y}_n, \mathcal{I}_n, \sigma, \tau_n) = \left\{ g : \max_{I \in \mathcal{I}_n} |w(\mathbf{y}_n, g, I)| \le \sigma \sqrt{\tau_n \log n} \right\},$$

where $\mathcal{I}_n$ is a family of intervals of $[0, 1]$ and for given α the value of $\tau_n = \tau_n(\alpha)$ is defined by

$$(8) \qquad P\left(\max_{I \in \mathcal{I}_n} \frac{1}{\sqrt{|I|}} \left| \sum_{t_i \in I} Z(t_i) \right| \le \sqrt{\tau_n \log n} \right) = \alpha.$$

If the data $\mathbf{y}_n$ were generated under (2), then (8) implies that $P(f \in \mathcal{A}_n) = \alpha$ with no restrictions on f so that $\mathcal{A}_n$ is a universal, exact and nonasymptotic α-confidence region. We mention that by using an appropriate norm [Mildenberger (2008)] $\mathcal{A}_n$ can also be expressed as a ball centered at the observations $\mathbf{y}_n$.

A function g belongs to $\mathcal{A}_n$ if and only if its vector of evaluations at the design points $(g(t_1), \ldots, g(t_n))$ belongs to the convex polyhedron in $\mathbb{R}^n$ which is defined by the linear inequalities

$$\frac{1}{\sqrt{|I|}} \left| \sum_{t_i \in I} (y(t_i) - g(t_i)) \right| \le \sigma_n \sqrt{\tau_n \log n}, \qquad I \in \mathcal{I}_n.$$

The remainder of the paper is in one sense nothing more than exploring the consequences of these inequalities for shape and smoothness regularization. They enforce both local and global adaptivity to the data and they are tight in that they yield optimal rates of convergence for both shape and smoothness constraints.

In the theoretical part of the paper we take $\mathcal{I}_n$ to be the set of all intervals of the form $[t_i, t_j]$. For this choice of $\mathcal{A}_n$, checking whether $g \in \mathcal{A}_n$ for a given g involves about $n^2/2$ linear inequalities. Surprisingly there exist algorithms which allow this to be done with algorithmic complexity $O(n \log n)$ [Bernholt and Hofmeister (2006)]. In practice we restrict $\mathcal{I}_n$ to a multiresolution scheme as follows. For some $\lambda > 1$, we set

$$\mathcal{I}_n = \left\{ \left[t_{l(j,k)}, t_{u(j,k)} \right] : l(j, k) = \lfloor (j-1)\lambda^k + 1 \rfloor, \right.$$
$$(9) \qquad\qquad u(j, k) = \min\{ \lfloor j\lambda^k \rfloor, n \},$$
$$\left. j = 1, \ldots, \lceil n\lambda^{-k} \rceil, k = 1, \ldots, \lceil \log n / \log \lambda \rceil \right\}.$$

For any $\lambda > 1$, we see that $\mathcal{I}_n$ now contains $O(n)$ intervals. For $\lambda = 2$, we get the wavelet multiresolution scheme which we use throughout the paper when doing the

calculations for explicit data sets. If $\mathcal{I}_n$ is the set of all possible intervals it follows from a result of Dümbgen and Spokoiny (2001) that $\lim_{n\to\infty} \tau_n = 2$ whatever the value of α. On the other hand, for any $\mathcal{I}_n$ which contains all the degenerate intervals $[t_j, t_j]$ (as will always be the case), then $\lim_{n\to\infty} \tau_n \geq 2$ whatever α. In the following, we simply take $\tau_n = 3$ as our default value. This guarantees a coverage probability of at least $\alpha = 0.95$ for all samples of size $n \geq 500$ and it tends rapidly to one as the sample size increases. The exact asymptotic distribution of $\max_{1 \leq i < j \leq n} (\sum_{l=i}^{j} Z_l)^2/(j - i + 1)$ has recently been derived by Kabluchko (2008).

As it stands, the confidence region (7) cannot be used as it requires σ. We use the following default estimate:

$$(10) \quad \sigma_n = \text{median}(|y(t_2) - y(t_1)|, \ldots, |y(t_n) - y(t_{n-1})|)/(\Phi^{-1}(0.75)\sqrt{2}),$$

where Φ^{-1} is the inverse of the standard normal distribution function Φ. It is seen that σ_n is a consistent estimate of σ for white noise data. For data generated under (2), σ_n is positively biased and consequently the coverage probability will not decrease. Simulations show that

$$(11) \qquad\qquad P\big(f \in \mathcal{A}_n(\mathbf{Y}_n, \mathcal{I}_n, \sigma_n, 3)\big) \geq 0.95$$

for all $n \geq 500$ and

$$(12) \qquad\qquad \lim_{n\to\infty} \inf_f P\big(f \in \mathcal{A}_n(\mathbf{Y}_n, \mathcal{I}_n, \sigma_n, 3)\big) = 1.$$

In other words, $\mathcal{A}_n$ is a universal, honest and nonasymptotic confidence region for f. To separate the problem of specifying the size of the noise from the problem of investigating the behavior of the procedures under the model (2) we shall always put $\sigma_n = \sigma$ for theoretical results. For real data and in all simulations, however, we use the σ_n of (10).

The confidence region $\mathcal{A}_n$ can be interpreted as the inversion of the multiscale tests that the mean of the residuals is zero on all intervals $I \in \mathcal{I}_n$. A similar idea is to be found in Dümbgen and Spokoiny (2001) who invert tests to obtain confidence regions. Their tests derive from kernel estimators with different locations and bandwidths where the kernels are chosen to be optimal for certain testing problems for given shape hypotheses. The confidence region may be expressed in terms of linear inequalities involving the weighted residuals with the weights determined by the kernels. The confidence region we use corresponds to the uniform kernel on $[0, 1]$. Because of their multiscale character all these confidence regions allow any lack of fit to be localized [Davies and Kovac (2001), Dümbgen and Spokoiny (2001)] and under shape regularization they automatically adapt to a certain degree of local smoothness. Universal, exact and nonasymptotic confidence regions based on the signs of the residuals $\text{sign}(y(t_i) - g(t_i))$ rather than the residuals themselves are to be found implicitly in Davies (1995) and explicitly

in Dümbgen (2003), Dümbgen (2007) and Dümbgen and Johns (2004). These require only that under the model the errors $\varepsilon(t)$ be independently distributed with median zero. As a consequence, they do not require an auxiliary estimate of scale such as (10). Estimates and confidence bounds based on such confidence regions are less sensitive but much more robust.

3. Shape regularization and local adaptivity.

3.1. *Generalities.* In this section we consider shape regularization within the confidence region $\mathcal{A}_n$. Two simple possibilities are to require that the function be monotone or that it be convex. Although much has been written about monotone or convex regression, we are not concerned with these particular cases. Given any data set $\mathbf{y}_n$ it is always possible to calculate a monotone regression function, for example, monotone least squares. In the literature the assumption usually made is that the f in (2) is monotone and then one examines the behavior of a monotone regression function. Although this case is included in the following analysis, we are mainly concerned with determining the minimum number of local extreme points or points of inflection required for an adequate approximation. This is STEP 2 of Mammen (1991). We shall investigate how pronounced a peak or a point of inflection must be before it can be detected on the basis of a sample of size n. These estimates are, in general, conservative but they do reflect the real finite sample behavior of our procedures. We shall also investigate rates of convergence between peaks and between points of inflection. We show that these are local in the strong sense that the rate of convergence at a point t depends only on the behavior of f in a small neighborhood of t. Furthermore, we show that in a certain sense shape regularization automatically adapts to the smoothness of f. All the calculations we perform use only the shape restrictions of the regularization and the linear inequalities which determine $\mathcal{A}_n$. The mathematics are extremely simple, involving no more than a Taylor expansion, and are of no intrinsic interest. We give one such calculation in detail and refer to the Appendix for the remainder.

3.2. *Local extreme values.* The simplest form of shape regularization is to minimize the number of local extreme values subject to membership of $\mathcal{A}_n$. We wish to determine this minimum number and exhibit a function in $\mathcal{A}_n$ which has this number of local extreme values. This is an optimization problem and the taut string algorithm of Davies (1995) and Davies and Kovac (2001) was explicitly developed to solve it. A short description of the algorithm used in Kovac (2007) is given in Appendix A.3. We analyze the properties of any such solution and, in particular, the ability to detect peaks or points of inflection. To do this we consider data generated under the model (2) and investigate how pronounced a peak of the generating function f of (2) must be before it is detected on the basis of a sample of size n. We commence with the case of one local maximum and assume that it is

located at $t = 1/2$. Let I_c denote an interval which contains $1/2$. For any $\tilde{f}_n$ in $\mathcal{A}_n$ we have

$$\frac{1}{\sqrt{|I_c|}} \sum_{t_i \in I_c} \tilde{f}_n(t_i) \geq \frac{1}{\sqrt{|I_c|}} \sum_{t_i \in I_c} f(t_i) - \sigma \sqrt{3 \log n} + \sigma Z(I_c),$$

and hence

$$\text{(13)} \qquad \max_{t_i \in I_c} \tilde{f}_n(t_i) \geq \frac{1}{|I_c|} \sum_{t_i \in I_c} f(t_i) - \sigma \frac{\sqrt{3 \log n} - Z(I_c)}{\sqrt{|I_c|}}$$

where

$$Z(I_c) = \frac{1}{\sqrt{|I_c|}} \sum_{t_i \in I_c} Z(t_i) \overset{D}{=} N(0, 1).$$

Let I_l and I_r be intervals to the left and right of I_c, respectively. A similar argument gives

$$\text{(14)} \qquad \min_{t_i \in I_l} \tilde{f}_n(t_i) \leq \frac{1}{|I_l|} \sum_{t_i \in I_l} f(t_i) + \sigma \frac{\sqrt{3 \log n} + Z(I_l)}{\sqrt{|I_l|}}$$

and

$$\text{(15)} \qquad \min_{t_i \in I_r} \tilde{f}_n(t_i) \leq \frac{1}{|I_r|} \sum_{t_i \in I_r} f(t_i) + \sigma \frac{\sqrt{3 \log n} + Z(I_r)}{\sqrt{|I_r|}}.$$

If now

$$\frac{1}{|I_c|} \sum_{t_i \in I_c} f(t_i) - \sigma \frac{\sqrt{3 \log n} - Z(I_c)}{\sqrt{|I_c|}}$$

$$\text{(16)} \qquad \geq \max\left\{ \frac{1}{|I_l|} \sum_{t_i \in I_l} f(t_i) + \sigma \frac{\sqrt{3 \log n} + Z(I_l)}{\sqrt{|I_l|}}, \right.$$

$$\left. \frac{1}{|I_r|} \sum_{t_i \in I_r} f(t_i) + \sigma \frac{\sqrt{3 \log n} + Z(I_r)}{\sqrt{|I_r|}} \right\},$$

then any function in $\mathcal{A}_n$ must have a local maximum in $I_l \cup I_c \cup I_r$. The random variables $Z(I_c)$, $Z(I_l)$ and $Z(I_r)$ are independently and identically distributed $N(0, 1)$ random variables. With probability at least 0.99 we have $Z(I_c) \geq -2.72$, $Z(I_l) \leq 2.72$ and $Z(I_r) \leq 2.72$, and hence we can replace (16) by

$$\frac{1}{|I_c|} \sum_{t_i \in I_c} f(t_i) - \sigma \frac{\sqrt{3 \log n} + 2.72}{\sqrt{|I_c|}}$$

$$\text{(17)} \qquad \geq \max\left\{ \frac{1}{|I_l|} \sum_{t_i \in I_l} f(t_i) + \sigma \frac{\sqrt{3 \log n} + 2.72}{\sqrt{|I_l|}}, \right.$$

$$\frac{1}{|I_r|} \sum_{t_i \in I_r} f(t_i) + \sigma \frac{\sqrt{3 \log n} + 2.72}{\sqrt{|I_r|}} \Bigg\}.$$

If we now regularize by considering those functions in $\mathscr{A}_n$ with the minimum number of local extreme values we see that this number must be at least one. As f itself has one local extreme value and belongs to $\mathscr{A}_n$ with probability rapidly approaching one we see that, with high probability, the minimum number is one and that this local maximum lies in $I_l \cup I_c \cup I_r$.

Condition (17) quantifies a lower bound for the power of the peak so that it will be detected with probability of at least 0.94 on the basis of a sample of size $n \geq 500$. The precision of the location is given by the interval $I_l \cup I_c \cup I_r$. We apply this to the specific function

(18) $$f_b(t) = b\big((t - 1/2)/0.01\big)$$

where

(19) $$b(t) = \begin{cases} 1, & |t| \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

We denote by f_{bn}^* a function in $\mathscr{A}_n$ which has the smallest number of local extreme values. As the function f_b of (18) lies in $\mathscr{A}_n$ with probability rapidly tending to one and has exactly one local extreme, it follows than any such f_{bn}^* must have exactly one local extreme. Suppose we wish to detect the local maximum of f_b with a precision of $\delta = 0.01$. As all points in the interval $[0.49, 0.51]$ are in a sense the same local maximum, we require the local maximum of f_{bn}^* to lie in the interval $[0.48, 0.52]$. A short calculation with $\sigma = 1$ shows that the smallest value of n for which (17) is satisfied is approximately 19500. A small simulation study using the taut string resulted in the peak being found with the prescribed accuracy in 99.6% of the 10000 simulations.

We now consider a function f which has exactly one local maximum situated in $t = 1/2$ and for which

(20) $$-c_2 \leq f^{(2)}(t) \leq -c_1 < 0, \qquad t \in I_0,$$

for some open interval I_0 which contains the point $t = 1/2$. We denote by f_n^* a function in $\mathscr{A}_n$ which minimizes the number of local extremes. For large n, any such function f_n^* will have exactly one local extreme value which is a local maximum situated at t_n^* with

(21) $$|t_n^* - 1/2| = O_f\left(\left(\frac{\log n}{n}\right)^{1/5}\right).$$

An explicit upper bound for the constant in O_f in terms of c_1 and c_2 of (20) is available. We also have

(22) $$f_n^*(t_n^*) \geq f(1/2) - O_f\left(\left(\frac{\log n}{n}\right)^{2/5}\right)$$

with again an explicit constant available. In the other direction,

$$(23) \qquad f_n^*(t_n^*) \leq f(1/2) + \sigma\left(\sqrt{3\log n} + 2.4\right).$$

The proofs are given in the Appendix.

More generally, suppose that f has a continuous second derivative and κ local extreme values situated at $0 < t_1^e < \cdots < t_\kappa^e < 1$ with $f^{(2)}(t_k^e) \neq 0$, $k = 1, \ldots, \kappa$. If $f_n^* \in \mathcal{A}_n$ now denotes a function which has the smallest number of local extreme values of all functions in $\mathcal{A}_n$ it follows that, with probability tending to one, f_n^* will have κ local extreme values located at the points $0 < t_{n1}^{*e} < \cdots < t_{n\kappa}^{*e} < 1$ with

$$(24) \qquad |t_{nk}^{*e} - t_k^e| = O_f\left(\left(\frac{\log n}{n}\right)^{1/5}\right), \qquad k = 1, \ldots, \kappa.$$

Furthermore, if t_k^e is the position of a local maximum of f then

$$(25) \qquad f_n^*(t_{nk}^{*e}) \geq f(t_k^e) - O_f\left(\left(\frac{\log n}{n}\right)^{2/5}\right)$$

whereas, if t_k^e is the position of a local minimum of f then

$$(26) \qquad f_n^*(t_{nk}^{*e}) \leq f(t_k^e) + O_f\left(\left(\frac{\log n}{n}\right)^{2/5}\right).$$

In the other direction, we have

$$(27) \qquad f_n^*(t_{nk}^{*e}) \leq f(t_k^e) + \sigma\left(\sqrt{3\log n} + \sqrt{3\log(8 + \kappa)}\right),$$

$$(28) \qquad f_n^*(t_{nk}^{*e}) \geq f(t_k^e) - \sigma\left(\sqrt{3\log n} + \sqrt{3\log(8 + \kappa)}\right).$$

More precise bounds cannot be attained on the basis of monotonicity arguments alone.

3.3. *Between the local extremes.* We investigate the behavior of f_n^* between the local extremes where f_n^* is monotone. For any function $g : [0, 1] \to \mathbb{R}$ we define

$$(29) \qquad \|g\|_{I,\infty} = \sup\{|g(t)| : t \in I\}.$$

Consider a point $t = i/n$ between two local extreme values of f and write $I_{nk}^r = [i/n, (i + k)/n]$ with $k > 0$. Then,

$$(30) \qquad f_n^*(i/n) - f(i/n) \leq \min_{1 \leq k \leq k_n^{*r}}\left\{\frac{k}{n}\|f^{(1)}\|_{I_{nk}^r,\infty} + 2\sigma\sqrt{\frac{3\log n}{k}}\right\},$$

where k_n^{*r} denotes the largest value of k for which f_n^* is nondecreasing on I_{nk}^r. It follows from (30) and the corresponding inequality on the left that as long as f_n^* has the correct global monotonicity behavior its behavior at a point t with

$f^{(1)}(t) \neq 0$ depends only on the behavior of f in a small neighborhood of t. In particular, we have asymptotically

$$(31) \qquad |f(t) - f_n^*(t)| \leq 3^{4/3} \sigma^{2/3} |f^{(1)}(t)|^{1/3} \left(\frac{\log n}{n} \right)^{1/3}.$$

Furthermore, if $f^{(1)}(t) = 0$ on a nondegenerate interval $I = [t_l, t_r]$ between two local extremes, then for $t_l < t < t_r$ we have $I_l^* = [t_l, t]$ and $I_r^* = [t, t_r]$ which results in

$$(32) \qquad |f(t) - f_n^*(t)| \leq \frac{3^{1/2} \sigma}{\min\{\sqrt{t - t_l}, \sqrt{t_r - t}\}} \left(\frac{\log n}{n} \right)^{1/2}.$$

The same argument shows that if

$$|f(t) - f(s)| \leq L|t - s|^\beta$$

with $0 < \beta \leq 1$, then

$$(33) \qquad |f(t) - f_n^*(t)| \leq cL^{1/(2\beta+1)} (\sigma/\beta)^{2\beta/(2\beta+1)} (\log n/n)^{\beta/(2\beta+1)}$$

where

$$c \leq (2\beta + 1) 3^{\beta/(2\beta+1)} \left(\frac{1}{\beta + 1} \right)^{1/(2\beta+1)} \leq 4.327.$$

Apart from the value of c, this corresponds to Theorem 2.2 of Dümbgen and Spokoiny (2001).

3.4. *Convexity and concavity.* We now turn to shape regularization by concavity and convexity. We take an f which is differentiable with derivative $f^{(1)}$ which is strictly increasing on $[0, 1/2]$ and strictly decreasing on $[1/2, 1]$. We put $I_{nk}^c = [1/2 - k/n, 1/2 + k/n]$, $I_{nk}^l = [t_l - k/n, t_l + k/n]$ with $t_l + k/n < 1/2 - k/n$ and $I_{nk}^l = [t_r - k/n, t_r + k/n]$ with $t_r - k/n > 1/2 + k/n$. Corresponding to (17), if f satisfies

$$\min_{t \in I_{nk}^c} f^{(1)}(t)/n - \left(2\sigma \left(\sqrt{3 \log n} + 2.72/\sqrt{2} \right) \right)/k^{3/2}$$

$$(34) \qquad \geq \max \left\{ \max_{t \in I_{nk}^l} f^{(1)}(t)/n + \left(2\sigma \left(\sqrt{3 \log n} + 2.72/\sqrt{2} \right) \right)/k^{3/2}, \right.$$

$$\left. \max_{t \in I_{nk}^r} f^{(1)}(t)/n + \left(2\sigma \left(\sqrt{3 \log n} + 2.72/\sqrt{2} \right) \right)/k^{3/2} \right\},$$

then it follows that with probability tending to at least 0.99 the first derivative of every differentiable function $\tilde{f}_n \in \mathcal{A}_n$ has at least one local maximum. Let f_n^* be a differentiable function in $\mathcal{A}_n$ whose first derivative has the smallest number of local extreme values. Then, as f belongs to $\mathcal{A}_n$ with probability tending to one,

it follows that $f_n^{*(1)}$ has exactly one local maximum with probability tending to at least 0.99. Suppose now that f has a continuous third derivative and κ points of inflection located at $0 < t_1^i < \cdots < t_\kappa^i$ with

$$f^{(2)}(t_j^i) = 0 \quad \text{and} \quad f^{(3)}(t_j^i) \neq 0, \qquad j = 1, \ldots, \kappa.$$

If f_n^* has the smallest number of points of inflection in $\mathcal{A}_n$ then, as $f \in \mathcal{A}_n$ with probability tending to one, it follows that with probability tending to one f_n^* will have κ points of inflection located at $0 < t_{n1}^{*i} < \cdots < t_{n\kappa}^{*i} < 1$. Furthermore, corresponding to (24) we have

$$(35) \qquad |t_{nk}^{*i} - t_k^i| = O_f\left(\left(\frac{\log n}{n}\right)^{1/7}\right), \qquad k = 1, \ldots, \kappa.$$

Similarly, if t_k^i is a local maximum of $f^{(1)}$ then corresponding to (25) we have

$$(36) \qquad f_n^{*(1)}(t_{nk}^{*e}) \geq f^{(1)}(t_k^e) - O_f\left(\left(\frac{\log n}{n}\right)^{2/7}\right)$$

and if t_k^i is a local minimum of $f^{(1)}$ then corresponding to (26) we have

$$(37) \qquad f_n^{*(1)}(t_{nk}^{*e}) \leq f^{(1)}(t_k^e) + O_f\left(\left(\frac{\log n}{n}\right)^{2/7}\right).$$

3.5. *Between points of inflection.* Finally, we consider the behavior of f_n^* between the points of inflection where it is then either concave or convex. We consider a point $t = i/n$ and suppose that f_n^* is convex on $I_{nk}^r = [i/n, (i+2k)/n]$. Corresponding to (30) we have

$$(38) \qquad f_n^{*(1)}(i/n) - f^{(1)}(i/n) \leq \min_{1 \leq k \leq k_n^{*r}} \left\{ \frac{k}{n} \|f^{(2)}\|_{I_{nk}^r, \infty} + 4\sigma n \sqrt{\frac{3\log n}{k^3}} \right\}$$

where k_n^{*r} is the largest value of k such that f_n^* is convex on $[i/n, (i+2k)/n]$. Similarly, corresponding to (77) we have

$$(39) \qquad f^{(1)}(i/n) - f_n^{*(1)}(i/n) \leq \min_{1 \leq k \leq k_n^{*l}} \left\{ \frac{k}{n} \|f^{(2)}\|_{I_{nk}^l, \infty} + 4\sigma n \sqrt{\frac{3\log n}{k^3}} \right\}$$

where $I_{nk}^l = [i/n - 2k/n, i/n]$ and k_n^{*l} is the largest value of k for which f_n^* is convex on I_{nk}^l. If $f^{(2)}(t) \neq 0$ we have corresponding to (31)

$$(40) \qquad \left| f_n^{*(1)}(t) - f^{(1)}(t) \right| \leq 4.36\sigma^{2/5}\left| f^{(2)}(t) \right|^{3/5}\left(\frac{\log n}{n}\right)^{1/5},$$

as n tends to infinity. If $f^{(2)}(t) = 0$ on the nondegenerate interval $I = [t_l, t_r]$, then for $t_l < t < t_r$ we have corresponding to (32)

$$(41) \qquad \left| f_n^{*(1)}(t) - f^{(1)}(t) \right| \leq \frac{4\sqrt{3}\sigma}{\min\{(t-t_l)^{3/2}, (t_r-t)^{3/2}\}}\left(\frac{\log n}{n}\right)^{1/2}.$$

The results for f_n^* itself are as follows. For a point t with $f^{(2)}(t) \neq 0$ and an interval $I_{nk}^r = [t, t + 2k/n]$ where f_n^* is convex we have

$$f_n^*(t) \leq f(t) + c_1(f, t)\left(\frac{k}{n}\right)\left(\frac{\log n}{n}\right)^{1/5} + \frac{k^2}{2n^2}\|f^{(2)}\|_{I_{nk}^r, \infty} + 4\sigma\sqrt{\frac{3 \log n}{k}}$$

where $c_1(f, t) = 4.36\sigma^{2/5}|f^{(2)}(t)|^{3/5}$. If we minimize over k and repeat the argument for a left interval we have corresponding to (31)

$$(42) \qquad |f_n^*(t) - f(t)| \leq 11.58\sigma^{4/5}|f^{(2)}(t)|^{1/5}\left(\frac{\log n}{n}\right)^{2/5}.$$

Finally, if $f^{(2)}(t) = 0$ for t in the nondegenerate interval $[t_l, t_r]$ we have corresponding to (32) for $t_l < t < t_r$

$$(43) \qquad |f_n^*(t) - f(t)| \leq \frac{14\sigma}{\min\{\sqrt{t - t_l}, \sqrt{t_r - t}\}}\left(\frac{\log n}{n}\right)^{1/2}.$$

If the derivative $f^{(1)}$ of f satisfies $|f^{(1)}(t) - f^{(1)}(s)| \leq L|t - s|^\beta$ with $0 < \beta \leq 1$, then corresponding to (33) we have

$$\left|f_n^{*(1)}(t) - f^{(1)}(t)\right| \leq cL^{3/(2\beta+3)}(\sigma/\beta)^{2\beta/(2\beta+3)}\left(\frac{\log n}{n}\right)^{\beta/(2\beta+3)}$$

with

$$c \leq 2^\beta\left(\frac{6\sqrt{3}}{2^\beta}\right)^{(\beta+2)/(2\beta+3)} + 4\sqrt{3}\beta\left(\frac{2^\beta}{6\sqrt{3}}\right)^{3/(2\beta+3)} \leq 8.78.$$

There is, of course, a corresponding result for f_n^* itself.

4. Regularization by smoothness.

4.1. *Minimizing total variation.* We define the total variation of the kth derivative of a function g evaluated at the design point $t_i = i/n$ by

$$(44) \qquad TV(g^k) := \sum_{i=k+2}^{n} \left|\Delta^{(k+1)}(g(i/n))\right|, \qquad k \geq 0,$$

where

$$(45) \qquad \Delta^{(k+1)}(g(i/n)) = \Delta^{(1)}\left(\Delta^{(k)}(g(i/n))\right)$$

with

$$\Delta^{(1)}(g(i/n)) = n\left(g(i/n) - g((i-1)/n)\right).$$

Similarly, the supremum norm $\|g^{(k)}\|_\infty$ is defined by

$$(46) \qquad \|g^{(k)}\|_\infty = \max_i\left|\Delta^{(k)}(g(i/n))\right|.$$

Minimizing either $TV(g^k)$ or $\|g^{(k)}\|_\infty$ subject to $g \in \mathcal{A}_n$ leads to a linear programming problem. Minimizing the more traditional measure of smoothness

$$\int_0^1 g^{(k)}(t)^2 \, dt$$

subject to $g \in \mathcal{A}_n$ leads to a quadratic programming problem which is numerically much less stable [cf. Davies and Meise (2008)] so we restrict attention to minimizing $TV(g^k)$ or $\|g^{(k)}\|_\infty$.

Minimizing the total variation of g itself, $k = 0$, leads to piecewise constant solutions which are very similar to the taut string solution. In most cases the solution also minimizes the number of local extreme values but this is not always the case. The upper panel of Figure 1 shows the result of minimizing $TV(g)$ for the Doppler data of Donoho and Johnstone (1994). It has the same number of peaks as the taut string reconstruction. The lower panel of Figure 1 shows the result of minimizing $TV(g^{(1)})$. The solution is a linear spline. Figure 1 and the following figures were obtained using the software of Kovac (2007). Just as minimizing $TV(g)$ can be used for determining the intervals of monotonicity so can we use the solution of minimizing $TV(g^{(1)})$ to determine the intervals of concavity and convexity. Minimizing $TV(g^{(k)})$ or $\|g^{(k)}\|_\infty$ for larger values of k leads to very smooth functions, but the numerical problems increase.

4.2. *Smoothness and shape regularization.* Regularization by smoothness alone may lead to solutions which do not fulfill obvious shape constraints. Figure 2 shows the effect of minimizing the total variation of the second derivative without further constraints and the minimization with the imposition of the taut string shape constraints.

4.3. *Rates of convergence.* Let $\tilde{f}_n$ be such that

$$(47) \qquad \|\tilde{f}_n^{(2)}\|_\infty \le \|g^{(2)}\|_\infty \qquad \forall g \in \mathcal{A}_n.$$

For data generated under (2) with f satisfying $\|f^{(2)}\|_\infty < \infty$ it follows that, with probability rapidly tending to one,

$$(48) \qquad \|\tilde{f}_n^{(2)}\|_\infty \le \|f^{(2)}\|_\infty.$$

A Taylor expansion and a repetition of arguments already used leads to

$$(49) \qquad |\tilde{f}_n(i/n) - f(i/n)| \le 3.742 \|f^{(2)}\|_\infty^{1/5} \sigma^{4/5} \left(\frac{\log n}{n}\right)^{2/5}$$

on an interval

$$[0.58\sigma^{2/5}(\log n)^{1/5}/(\|f^{(2)}\|_\infty^{2/5} n^{1/5}), 1 - 0.58\sigma^{2/5}(\log n)^{1/5}/(\|f^{(2)}\|_\infty^{2/5} n^{1/5})]$$
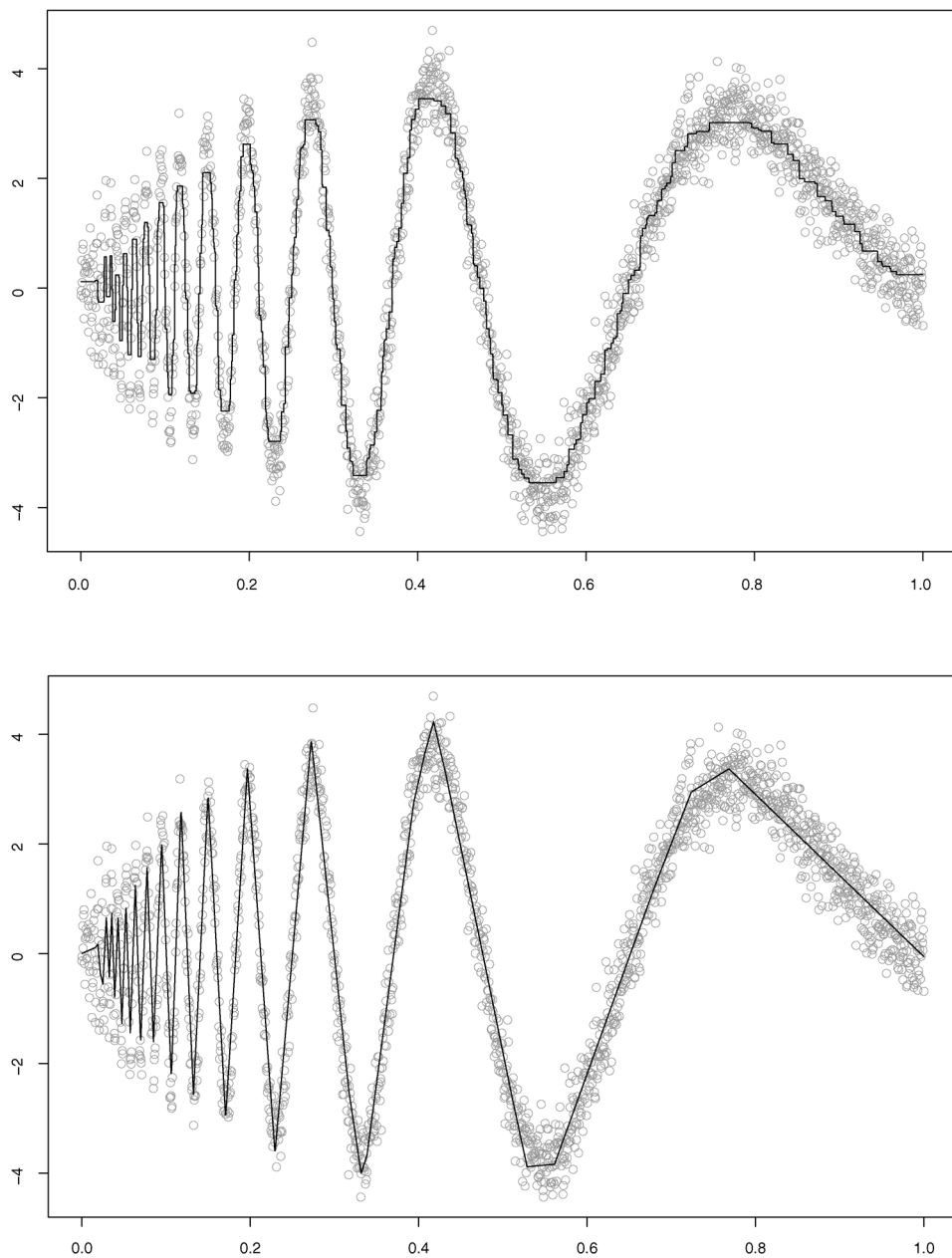
FIG. 1. *Minimization of TV(g) (upper panel) and TV($g^{(1)}$) (lower panel) subject to $g \in \mathcal{A}_n$ for a noisy Doppler function.*
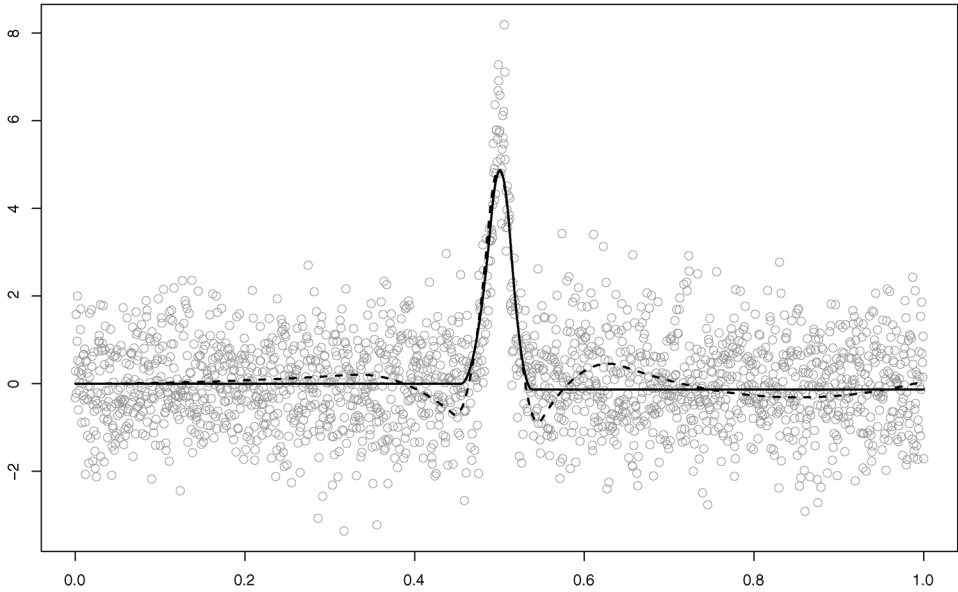
FIG. 2. *The minimization of the total variation of the second derivative with* (*solid line*) *and without* (*dashed line*) *the shape constraints derived from the taut string. The solution subject to the shape constraints was also forced to assume the same value at the local maximum as the unconstrained solution.*

with a probability rapidly tending to one. A rate of convergence for the first derivative may be derived in a similar manner and results in

$$(50) \qquad |\tilde{f}_n(i/n) - f^{(1)}(i/n)| \leq 4.251 \|f^{(2)}\|_\infty^{3/5} \sigma^{2/5} \left(\frac{\log n}{n}\right)^{1/5}$$

on an interval

$$[2.15\sigma^{2/5}(\log n)^{1/5}/(\|f^{(2)}\|_\infty^{2/5} n^{1/5}), 1 - 2.15\sigma^{2/5}(\log n)^{1/5}/(\|f^{(2)}\|_\infty^{2/5} n^{1/5})].$$

5. Confidence bands.

5.1. *The problem.*   Confidence bounds can be constructed from the confidence region $\mathcal{A}_n$ as follows. For each point t_i we require a lower bound $lb_n(\mathbf{y}_n, t_i) = lb_n(t_i)$ and an upper bound $ub_n(\mathbf{y}_n, t_i) = ub_n(t_i)$, such that

$$(51) \qquad \mathcal{B}_n(\mathbf{y}_n) = \{g : lb_n(\mathbf{y}_n, t_i) \leq g(t_i) \leq ub_n(\mathbf{y}_n, t_i), i = 1, \ldots, n\}$$

is an honest nonasymptotic confidence region

$$(52) \qquad\qquad P\big(f \in \mathcal{B}_n(\mathbf{Y}_n(f))\big) \geq \alpha \qquad \text{for all } f \in \mathcal{F}_n$$

for data $\mathbf{Y}_n(f)$ generated under (2). In a sense, the problem has a simple solution. If we put

$$(53) \qquad lb_n(t_i) = y(t_i) - \sigma_n\sqrt{3\log n}, \qquad ub_n(t_i) = y(t_i) + \sigma_n\sqrt{3\log n},$$

then $\mathcal{A}_n \subset \mathcal{B}_n$ and (52) for all holds with $\mathcal{F}_n = \{f \mid f : [0, 1] \to \infty\}$. Such universal bounds are too wide to be of any practical use and are consequently not acceptable. They can only be made tighter by restricting $\mathcal{F}_n$ by imposing shape or quantitative smoothness constraints. A qualitative smoothness assumption such as

$$(54) \qquad \mathcal{F}_n = \{f : \|f^{(2)}\|_\infty < \infty\}$$

does not lead to any improvement of the bounds (53). They can only be improved by replacing (54) by a quantitative assumption such as

$$(55) \qquad \mathcal{F}_n = \{f : \|f^{(2)}\|_\infty < 60\}.$$

5.2. *Shape regularization.*

5.2.1. *Monotonicity.* As an example of a shape restriction we consider bounds for nondecreasing approximations. If we denote the set of nonincreasing functions on $[0, 1]$ by

$$\mathcal{M}^+ = \{g : g : [0, 1] \to \mathbb{R}, g \text{ nondecreasing}\}$$

then there exists a nondecreasing approximation if and only if

$$(56) \qquad \mathcal{M}^+ \cap \mathcal{A}_n \neq \varnothing.$$

This is the case when the set of linear inequalities which define $\mathcal{A}_n$ together with $g(t_1) \leq \cdots \leq g(t_n)$ are consistent. This is once again a linear programming problem. If (56) holds then the lower and upper bounds are given, respectively, by

$$(57) \qquad lb_n(t_i) = \min\{g(t_i) : g \in \mathcal{M}^+ \cap \mathcal{A}_n\},$$

$$(58) \qquad ub_n(t_i) = \max\{g(t_i) : g \in \mathcal{M}^+ \cap \mathcal{A}_n\}.$$

The calculation of $lb_n(t_i)$ and $ub_n(t_i)$ requires solving a linear programming problem and, although this can be done, it is practically impossible for larger sample sizes using standard software because of exorbitantly long calculation times. If the family of intervals $\mathcal{I}_n$ is restricted to a wavelet multiresolution scheme then samples of size $n = 1000$ can be handled. Fast, honest bounds can be obtained as follows. If $g \in \mathcal{M}^+ \cap \mathcal{A}_n$ then for any i and k with $i + k \leq n$ it follows that

$$\sqrt{k+1}\, g(t_i) \geq \frac{1}{\sqrt{k+1}} \sum_{j=0}^{k} Y_n(t_{i-j}) - \sigma\sqrt{3 \log n}.$$

From this we may deduce the lower bound

$$(59) \qquad lb_n(t_i) = \max_{0 \leq k \leq i-1}\left(\frac{1}{k+1} \sum_{j=0}^{k} Y_n(t_{i-j}) - \sigma\sqrt{\frac{3 \log n}{k+1}} \right)$$

with the corresponding upper bound

$$(60) \qquad ub_n(t_i) = \min_{0 \le k \le n-i} \left(\frac{1}{k+1} \sum_{j=0}^{k} Y_n(t_{i+j}) + \sigma \sqrt{\frac{3 \log n}{k+1}} \right).$$

Both these bounds are of algorithmic complexity $O(n^2)$. Faster bounds can be obtained by putting

$$(61) \qquad lb_n(t_i) = \max_{0 \le \theta(k) \le i-1} \left(\frac{1}{\theta(k)+1} \sum_{j=0}^{\theta(k)} Y_n(t_{i-j}) - \sigma \sqrt{\frac{3 \log n}{\theta(k)+1}} \right),$$

$$(62) \qquad ub_n(t_i) = \min_{0 \le \theta(k) \le n-i} \left(\frac{1}{\theta(k)+1} \sum_{j=0}^{\theta(k)} Y_n(t_{i+j}) + \sigma \sqrt{\frac{3 \log n}{\theta(k)+1}} \right)$$

where $\theta(k) = \lfloor \theta^k - 1 \rfloor$ for some $\theta > 1$. These latter bounds are of algorithmic complexity $O(n \log n)$. The fast bounds are not necessarily nondecreasing, but can be made so by putting

$$ub_n(t_i) = \min(ub_n(t_i), ub_n(t_{i+1})), \qquad i = n-1, \dots, 1,$$
$$lb_n(t_i) = \max(lb_n(t_i), lb_n(t_{i-1})), \qquad i = 2, \dots, n.$$

The upper panel of Figure 3 shows data generated by

$$(63) \qquad\qquad\qquad Y(t) = \exp(5t) + 5Z(t)$$

evaluated on the grid $t_i = i/1000, i = 1, \dots, 100$, together with the three lower and three upper bounds with σ replaced by σ_n of (10). The lower bounds are those given by (57) with $\mathcal{I}_n$ a dyadic multiresolution scheme, (59) and (61) with $\theta = 2$. The times required for were about 12 hours, 19 seconds and less than one second, respectively, with corresponding times for the upper bounds (58), (60) and (62). The differences between the bounds are not very large: it is not the case that one set of bounds dominates the others. The methods of Section 3 can be applied to show that all the uniform bounds are optimal in terms of rates of convergence.

5.2.2. *Convexity.* Convexity and concavity can be treated similarly. If we denote the set of convex functions on $[0, 1]$ by $\mathcal{C}^+$, then there exists a convex approximation if and only if

$$\mathcal{C}^+ \cap \mathcal{A}_n \ne \varnothing.$$

Assuming that the design points are of the form $t_i = i/n$ this will be the case if and only if the set of linear constraints

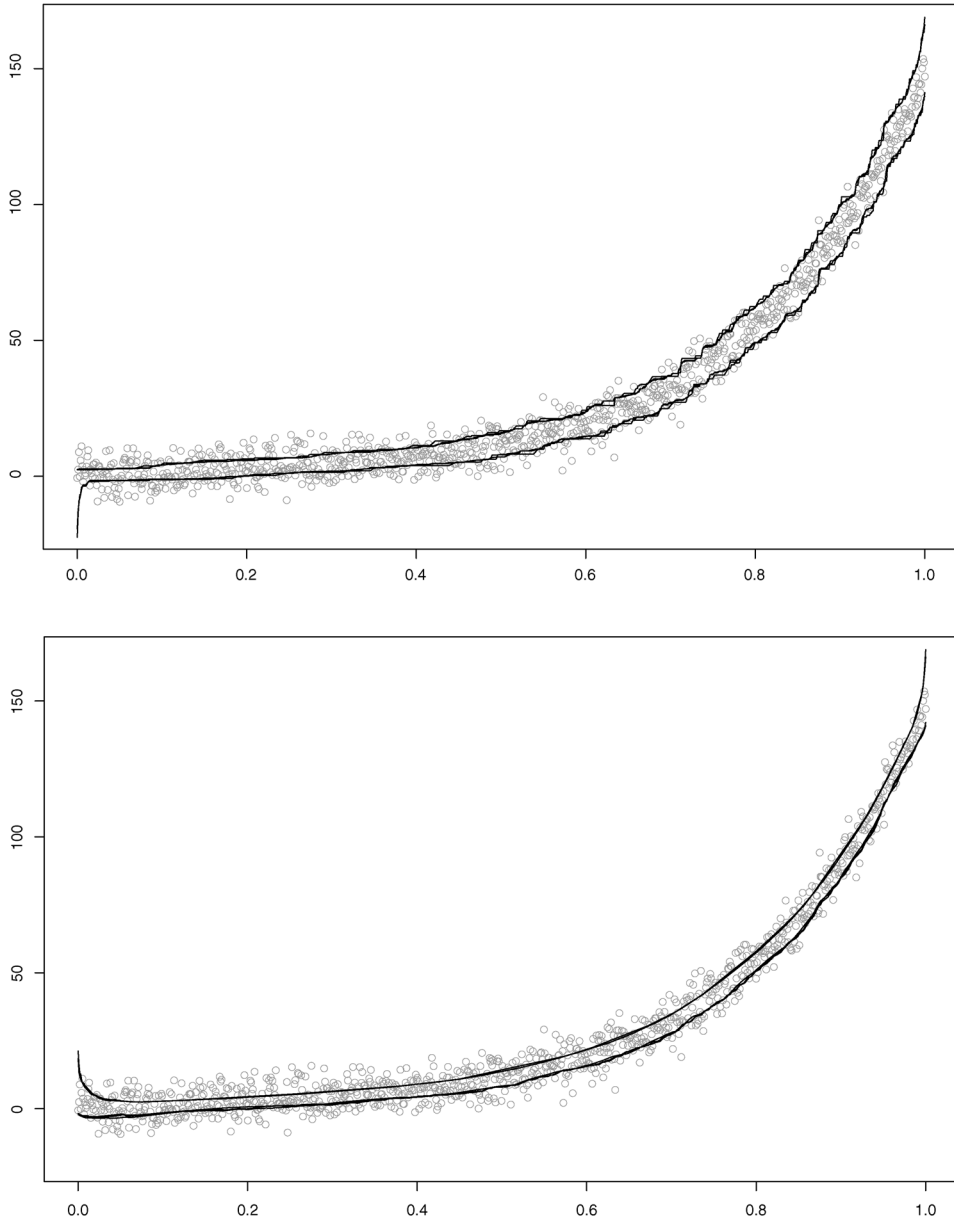$$g(t_{i+1}) - g(t_i) \ge g(t_i) - g(t_{i-1}), \qquad i = 2, \dots, n-1,$$

FIG. 3. *The function $f(t) = \exp(5t)$ degraded with $N(0, 25)$ noise together with monotone confidence bounds (upper panel) and convex confidence bounds (lower panel). The three lower bounds in the upper panel are derived from* (57), (59) *and* (61) *and the corresponding upper bounds are* (58), (60) *and* (62). *The lower bounds for the lower panel are* (64), (68) *and* (70) *and the corresponding upper bounds* (65), (66) *and* (69).

are consistent with the linear constraints which define $\mathcal{A}_n$. Again, this is a linear programming problem. If this is the case then lower and upper bounds are given, respectively, by

$$(64) \qquad lb_n(t_i) = \min\{g(t_i) : g \in \mathcal{C}^+ \cap \mathcal{A}_n\},$$

$$(65) \qquad ub_n(t_i) = \max\{g(t_i) : g \in \mathcal{C}^+ \cap \mathcal{A}_n\}$$

which again is a linear programming problem which can only be solved for relatively small values of n. An honest but faster upper bound can be obtained by noting that

$$g(i/n) \le \frac{1}{2k+1} \sum_{j=-k}^{k} g((i+j)/n), \qquad k \le \min(i-1, n-i)$$

which gives rise to

$$(66) \qquad ub_n(t_i) = \min_{0 \le k \le \min(i-1,n-i)} \left(\frac{1}{2k+1} \sum_{j=-k}^{k} Y_n(t_{i+j}) + \sigma \sqrt{\frac{3 \log n}{2k+1}} \right).$$

A fast lower bound is somewhat more complicated. Consider a function $\tilde{f}_n \in \mathcal{C}^+ \cap \mathcal{A}_n$, and two points $(i/n, \tilde{f}_n(i/n))$ and $((i+k)/n, ub_n((i+k)/n))$. As $\tilde{f}_n((i+k)/n) \le ub_n((i+k)/n)$ and $\tilde{f}_n$ is convex it follows that $\tilde{f}_n$ lies below the line joining $(i/n, \tilde{f}_n(i/n))$ and $((i+k)/n, ub_n((i+k)/n))$. From this and $\tilde{f}_n \in \mathcal{A}_n$ we may derive a lower bound by noting

$$(67) \qquad \begin{aligned} lb_n(t_i) &\le lb_n(t_i, k) \\ &:= \max_{1 \le j \le k} \left(\frac{1}{j} \sum_{l=1}^{j} Y_n(t_{i+j}) - ub_n(t_{i+k})(j+1)/(2k) - \sigma\sqrt{3 \log n / j} \right) \end{aligned}$$

for all $k, -i+1 \le k \le n-i$. An honest lower bound is therefore given by

$$(68) \qquad lb_n(t_i) = \max_{-i+1 \le k \le n-i} lb_n(t_i, k).$$

The algorithmic complexity of ub_n as given by (66) is $O(n^2)$ while that of the lower bound (68) is $O(n^3)$. Corresponding to (62) we have

$$(69) \qquad \begin{aligned} ub_n(t_i) = \min_{0 \le \theta(k) \le \min(i-1,n-i)} \bigg(&\frac{1}{2\theta(k)+1} \sum_{j=-\theta(k)}^{\theta(k)} Y_n(t_{i+j}) \\ &+ \sigma \sqrt{\frac{3 \log n}{2\theta(k)+1}} \bigg), \end{aligned}$$

and to (61)

$$(70) \qquad lb_n(t_i) = \max_{-i+1 \le \theta(k) \le n-i} lb_n(t_i, \theta(k)),$$

where

$$lb_n(t_i) \leq lb_n(t_i, \theta(k))$$

$$(71) \qquad := \max_{1 \leq \theta(j) \leq \theta(k)} \left(\frac{1}{\theta(j)} \sum_{l=1}^{\theta(j)} Y_n(t_{i+j}) \right.$$

$$\left. - ub_n(t_{i+\theta(k)})(\theta(j)+1)/(2\theta(k)) - \sigma\sqrt{3\log n/\theta(j)} \right)$$

with $\theta(k) = \lfloor \theta^k \rfloor$ for some $\theta > 1$. The algorithmic complexity of (69) is $O(n \log n)$ and that of (70) is $O(n(\log n)^2)$.

The lower panel of Figure 3 shows the same data as in the upper panel but with the lower bounds given by (64) with $\mathcal{I}_n$ a dyadic multiresolution scheme, (68) and (70) and the corresponding upper bounds (65), (66) and (69). The calculation of each of the bounds (64) and (65) took about 12 hours. The lower bound (68) took about 210 minutes, while (70) was calculated in less than 5 seconds. The lower bound (64) is somewhat better than (68) and (70), but the latter two are almost indistinguishable.

5.2.3. *Piecewise monotonicity.* We now turn to the case of functions which are piecewise monotone. The possible positions of the local extremes can in theory be determined by solving the appropriate linear programming problems. The taut string methodology is, however, extremely good and very fast so we can use this solution to identify possible positions of the local extremes. The confidence bounds depend on the exact location of the local extreme. If we take the interval of constancy of the taut string solution which includes the local maximum, we may calculate confidence bounds for any function which has its local maximum in this interval. The result is shown in the top panel of Figure 4 where we used the fast bounds (61) and (62), (61) and (62) with $\theta = 1.5$. If we use the midpoint of the taut string interval as a default choice for the position of a local extreme we obtain confidence bounds as shown in the lower panel of Figure 4. The user can of course specify these positions and the program will indicate if they are consistent with the linear constraints which define the approximation region $\mathcal{A}_n$.

5.2.4. *Piecewise concave–convex.* We can repeat the idea for functions which are piecewise concave–convex. There are fast methods for determining the intervals of convexity and concavity based on the algorithm devised by Groeneboom (1996), but in this section we use the intervals obtained by minimizing the total variation of the first derivative [Kovac (2007)]. The upper panel of Figure 5 shows the result for convexity/concavity which corresponds to Figure 4. Finally,
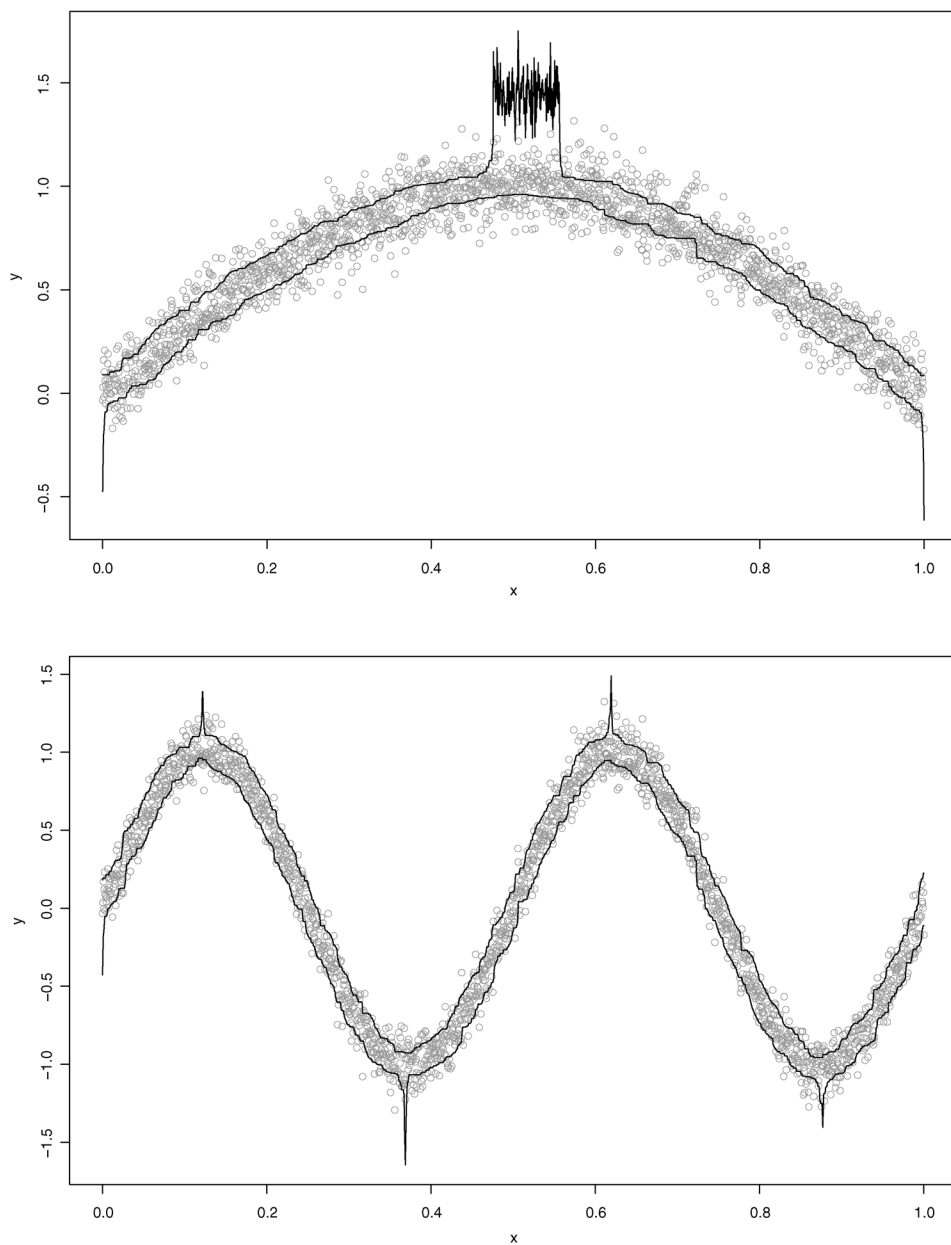
FIG. 4.   *Confidence bounds without* (*upper panel*) *and with* (*lower panel*) *the specification of the precise positions of the local extreme values. The positions in the lower panel are the default choices obtained from the taut string reconstruction* [Kovac (2007)]. *The bounds are the fast bounds* (61) *and* (62) *with* $\theta = 1.5$.
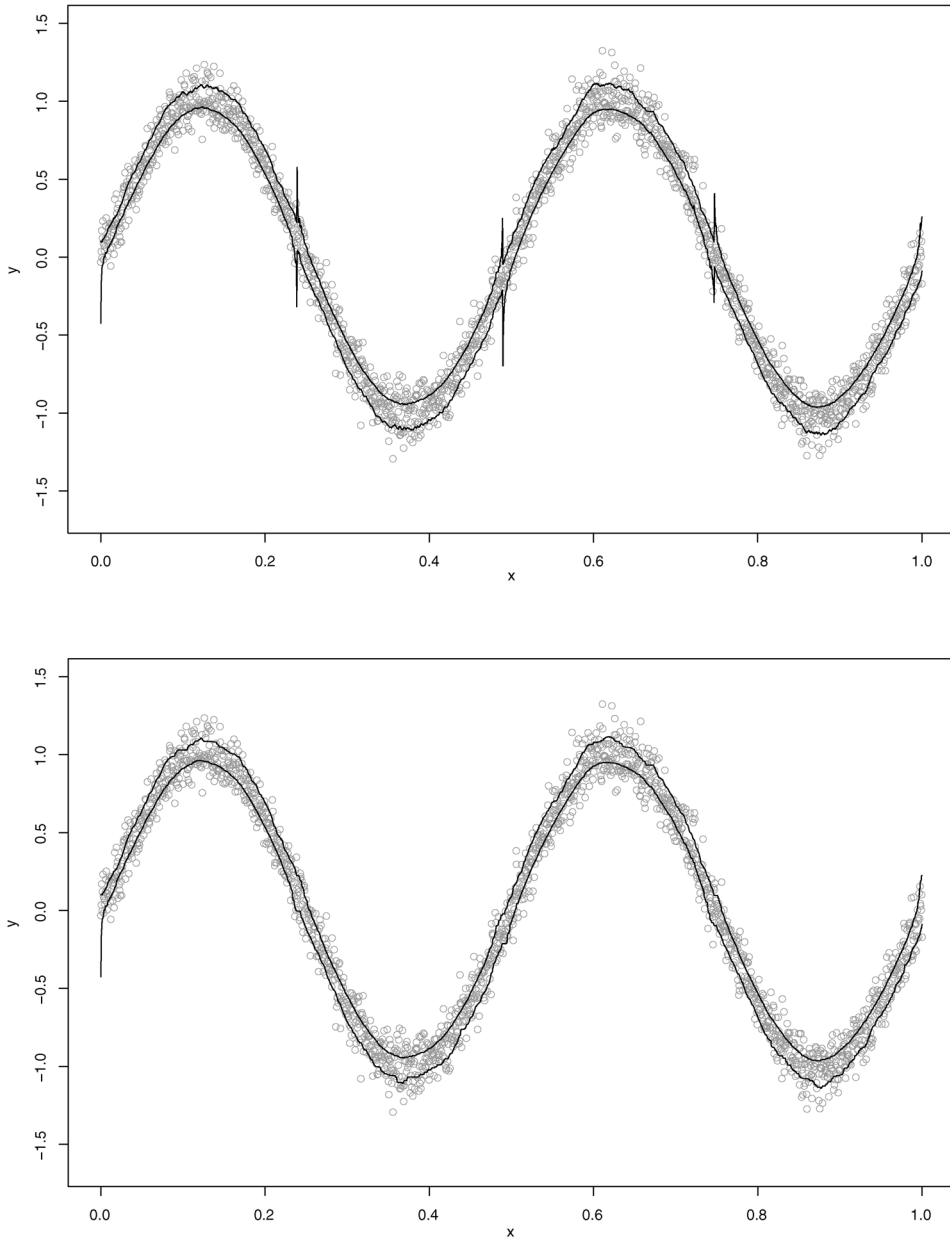
FIG. 5. *Confidence bounds with default choices for the intervals of convexity/concavity (upper panel based on* (69) *and* (70) *with* $\theta = 1.5$) *and combined confidence bounds for default choices of intervals of monotonicity and convexity/concavity.*

the lower panel of Figure 5 shows the result of imposing both monotonicity and convexity/concavity constraints. In both cases the bounds used are the fast bounds (69) and (70) with $\theta = 1.5$.

5.2.5. *Sign-based confidence bounds.* As mentioned in Section 2.2, work has been done on confidence regions based on the signs of the residuals. These can also be used to calculate confidence bands for shape-restricted functions. We refer to Davies (1995), Dümbgen (2003), Dümbgen (2007) and Dümbgen and Johns (2004).

5.3. *Smoothness regularization.* We turn to the problem of constructing lower and upper confidence bounds under some restriction on smoothness. For simplicity, we take the supremum norm $\|g^{(2)}\|_\infty$ to be the measure of smoothness for a function g. The discussion in Section 5.1 shows that honest bounds are attainable only if we restrict f to a set $\mathcal{F}_n = \{g : \|g^{(2)}\|_\infty \le K\}$ with a specified K. We illustrate the idea using data generated by (2) with $f(t) = \sin(4\pi t)$ and $\sigma = 1$. The minimum value of $\|g^{(2)}\|_\infty$ is 117.7 which compares with $16\pi^2 = 157.9$ for f itself. The upper panel of Figure 6 shows the data together with the resulting function f_n^*. The bounds under the restriction $\|\tilde{f}_n^{(2)}\|_\infty \le 117.2$ coincide with the function f_n^* itself. The middle panel of Figure 6 show the bounds based on $\|g^{(2)}\|_\infty \le K$ for

$$K = 137.8\big(= (117.7 + 157.9)/2\big), \quad 157.9 \quad \text{and} \quad 315.8(= 2 \times 157.9).$$

Just as before, fast bounds are also available. We have for the lower bound for given K

$$(72) \quad lb(i/n) \le \min_k \left(\frac{1}{2k+1} \sum_{j=-k}^{k} Y((i+j)/n) + \left(\frac{k}{n}\right)^2 K + \sigma \sqrt{\frac{3\log n}{2k+1}} \right)$$

and for the upper bound

$$(73) \quad ub(i/n) \ge \max_k \left(\frac{1}{2k+1} \sum_{j=-k}^{k} Y((i+j)/n) - \left(\frac{k}{n}\right)^2 K - \sigma \sqrt{\frac{3\log n}{2k+1}} \right).$$

As it stands, the calculation of these bounds is of algorithmic complexity $O(n^2)$, but this can be reduced to $O(n \log n)$ by restricting k to be of the form θ^m. The method also gives a lower bound for $\|g^{(2)}\|_\infty$ for g to be consistent with the data. This is the smallest value of K for which the lower bound lb lies beneath the upper bound ub. If we do this for the data of Figure 6 with $\theta = 1.5$ then the smallest value is 104.5 as against the correct bound of 115.0. The lower panel of Figure 6 shows the fast bounds for the same data and values of K.
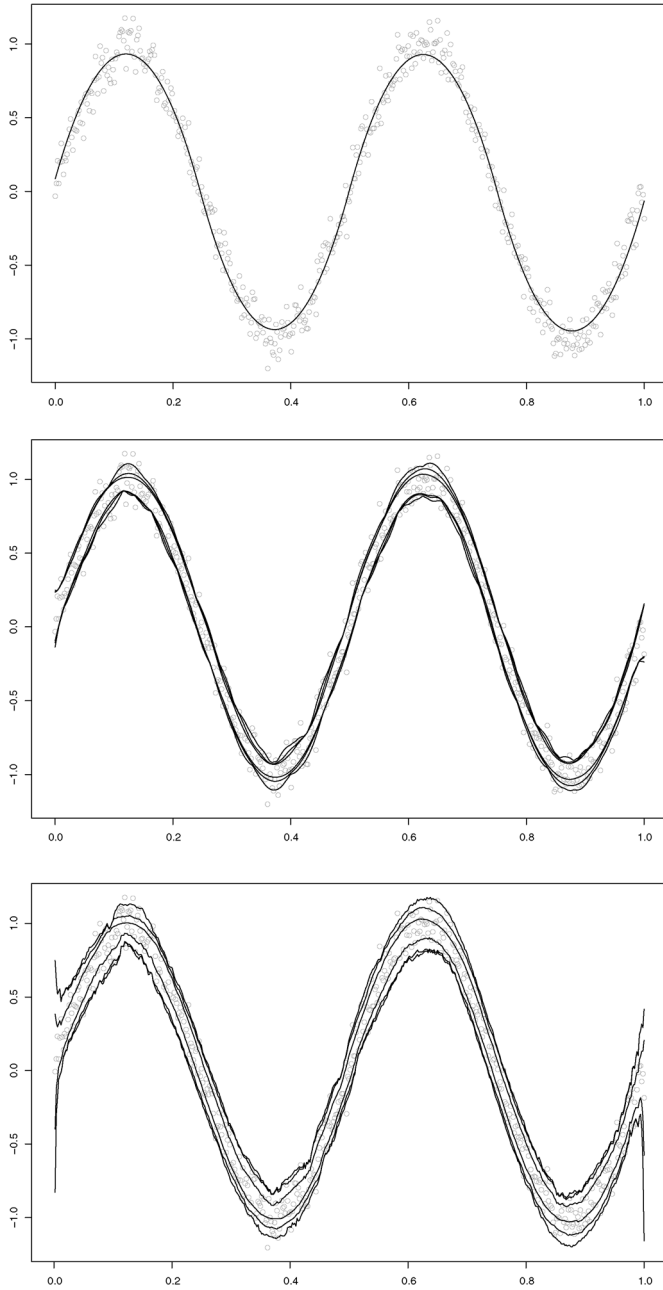
FIG. 6.   *Smoothness confidence bounds for $f \in \mathcal{F}_n = \{f : \|\tilde{f}_n^{(2)}\|_\infty \leq K\}$ for data generated according to* (2) *with* $f(t) = \sin(4\pi t)$, $\sigma = 0.2$ *and* $n = 500$. *The top panel shows the function which minimizes* $\|g^{(2)}\|_\infty$. *The minimum is* 117.7 *compared with* $16\pi^2 = 157.9$ *for* $f(t)$. *For this value of* K *the bounds are degenerate. The center panel shows the confidence bounds for* $K = 137.8$, 157.9 *and* 315.8. *The bottom panel shows the corresponding fast bounds* (72) *and* (73) *with* $\theta = 1.5$ *for the same values of* K.

APPENDIX

A.1. Proofs of Section 3.2.

A.1.1. *Proof of* (21). Let k be such that $I_c = [1/2 - k/n, 1/2 + k/n] \subset I_0$. A Taylor expansion together with (20) implies, after some manipulation,

$$\frac{1}{2k+1} \sum_{t_i \in I_c} f(t_i) - \sigma \frac{\sqrt{3 \log n} + 2.72}{\sqrt{2k+1}}$$

$$\geq f(1/2) - \frac{k^2}{2n^2} c_2 - \sigma \frac{\sqrt{3 \log n} + 2.72}{\sqrt{2k}}$$

and, on minimizing the right-hand side of the inequality with respect to k, we obtain

(74)
$$\frac{1}{|I_c|} \sum_{t_i \in I_c} f(t_i) - \sigma \frac{\sqrt{3 \log n} + 2.72}{\sqrt{|I_c|}}$$

$$\geq f(1/2) - 1.1 c_2^{1/5} \sigma^{4/5} \left(\sqrt{3 \log n} + 2.72\right)^{4/5} / n^{2/5}.$$

This inequality holds as long as $I_c = [1/2 - k_n/n, 1/2 + k_n/n] \subset I_0$ with

(75)
$$k_n = \lfloor 0.66 c_2^{-2/5} \sigma^{2/5} n^{4/5} \left(\sqrt{3 \log n} + 2.72\right)^{2/5} \rfloor.$$

If we put $I_l = [1/2 - (\eta + 1)k_n/n, 1/2 - \eta k_n/n]$, similar calculations give

$$\frac{1}{2k+1} \sum_{t_i \in I_l} f(t_i) + \sigma \frac{\sqrt{3 \log n} + 2.72}{\sqrt{2k+1}}$$

$$\leq f(1/2) - \frac{k^2}{2n^2} c_1 + \sigma \frac{\sqrt{3 \log n} + 2.72}{\sqrt{2k}},$$

and hence

$$\frac{1}{|I_l|} \sum_{t_i \in I_l} f(t_i) + \sigma \frac{\sqrt{3 \log n} + 2.72}{\sqrt{|I_l|}}$$

$$\geq f(1/2) - \frac{c_2^{1/5} \sigma^{4/5} (\sqrt{3 \log n} + 2.72)^{4/5}}{n^{2/5}} [0.2178 \eta^2 c_1/c_2 - 1.23]$$

with the same estimate for $I_r = [1/2 + \eta k_n/n, 1/2 + (\eta + 1)k_n/n]$. If we put $\eta = 3.4\sqrt{c_2/c_1}$ and

(76)
$$I_n := [1/2 - (\eta + 1)k_n/n, 1/2 + (\eta + 1)k_n/n] \subset I_0$$

then all estimates hold. Because of (75) this will be the case for n sufficiently

large. This implies that (17) holds for sufficiently large n and in consequence any function $\tilde{f}_n \in \mathcal{A}_n$ has a local maximum in I_n.

A.1.2. *Proofs of* (22) *and* (23). From (13) and (74) we have

$$f_n^*(t_n^*) \geq f(1/2) - 1.1 c_2^{1/5} \sigma^{4/5} \left(\sqrt{3\log n} + 2.72\right)^{4/5} / n^{2/5}$$

which is the required estimate (22). To prove (23) we simply note

$$f_n^*(t_n^*) \leq f(t_n^*) + \sigma Z(t_n^*) + \sigma\sqrt{3\log n} \leq f(1/2) + \sigma\left(\sqrt{3\log n} + 2.4\right).$$

A.1.3. *Proof of* (30) *and* (31). As $f_n^* \in \mathcal{A}_n$ by definition and $f \in \mathcal{A}_n$ with probability tending to one, we have for the interval $I_{nk}^r = [i/n, (i+k-1)/n]$

$$\frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} f_n^*((i+j)/n) \leq \frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} f((i+j)/n) + 2\sigma\sqrt{3\log n}$$

from which it follows that

$$f_n^*(i/n) \leq f(i/n) + \frac{k}{n} \|f^{(1)}\|_{I_{nk}^r, \infty} + 2\sigma\sqrt{\frac{3\log n}{k}}$$

which proves (30). Similarly, for the intervals $I_{nk}^l = [(i-k+1)/n, i/n]$ we have

$$(77) \qquad f(i/n) - f_n^*(i/n) \leq \min_{1 \leq k \leq k_n^{*l}} \left\{ \frac{k}{n} \|f^{(1)}\|_{I_{nk}^l, \infty} + 2\sigma\sqrt{\frac{3\log n}{k}} \right\}.$$

We note that (30) and (77) imply that f_n^* adapts automatically to f to give optimal rates of convergence. If $f^{(1)}(t) \neq 0$ then it may be checked that the lengths of the optimal intervals I_{nk}^{r*} and I_{nk}^{l*} tend to zero and consequently

$$\|f^{(1)}\|_{I_{nk}^{l*}, \infty} \approx |f^{(1)}(t)| \approx \|f^{(1)}\|_{I_{nk}^{r*}, \infty}.$$

The optimal choice of k is then

$$k_n^{*l} \approx \left(\frac{3\sigma^2 n^2 \log n}{|f^{(1)}(t)|^2} \right)^{1/3} \approx k_n^{*r}$$

which gives

$$\lambda(I_{nk}^{l*}) \approx \frac{3^{1/3}\sigma^{2/3}}{|f^{(1)}(t)|^{2/3}} \left(\frac{\log n}{n} \right)^{1/3} \approx \lambda(I_{nk}^{l*})$$

from which (31) follows.

A.2. Proofs of Section 3.4.

A.2.1. *Proof of* (34). Then adapting the arguments used above we have, for any differentiable function $\tilde{f}_n \in \mathcal{A}_n$,

$$\frac{1}{\sqrt{k}} \sum_{i=1}^{k} (\tilde{f}_n(1/2 + i/n) - \tilde{f}_n(1/2 - k/n + i/n))$$

$$\geq \frac{1}{\sqrt{k}} \sum_{i=1}^{k} (f(1/2 + i/n) - f(1/2 - k/n + i/n))$$

$$- 2\sigma \left(\sqrt{3 \log n} + Z(I_{nk}^c)/\sqrt{2} \right)$$

which implies

(78) $$\max_{t \in I_{nk}^c} f_n^{*(1)}(t)/n \geq \min_{t \in I_{nk}^c} f^{(1)}(t)/n - \left(2\sigma \left(\sqrt{3 \log n} + Z(I_{nk}^c)/\sqrt{2} \right) \right)/k^{3/2}.$$

Similarly, if $I_{nk}^l = [t_l - k/n, t_l + k/n]$ with $t_l + k/n < 1/2 - k/n$ we have

(79) $$\min_{t \in I_{nk}^l} f_n^{*(1)}(t)/n \leq \max_{t \in I_{nk}^l} f^{(1)}(t)/n + \left(2\sigma \left(\sqrt{3 \log n} + Z(I_{nk}^l)/\sqrt{2} \right) \right)/k^{3/2}$$

and for $I_{nk}^l = [t_r - k/n, t_r + k/n]$ with $t_r - k/n > 1/2 + k/n$ we have

(80) $$\min_{t \in I_{nk}^r} f_n^{*(1)}(t)/n \leq \max_{t \in I_{nk}^r} f^{(1)}(t)/n + \left(2\sigma \left(\sqrt{3 \log n} + Z(I_{nk}^r)/\sqrt{2} \right) \right)/k^{3/2}.$$

Again, following the arguments given above we may deduce from (78), (79) and (80), that for sufficiently large n, it is possible to choose I_{nk}^l, I_{nk}^c and I_{nk}^r so that (34) holds.

A.2.2. *Proof of* (38). We have

$$\frac{1}{\sqrt{k}} \sum_{j=1}^{k} (f_n^*(k/n + i/n) - f_n^*(i/n))$$

$$\leq \frac{1}{\sqrt{k}} \sum_{j=1}^{k} (f(k/n + i/n) - f(i/n)) + 2\sigma \sqrt{3 \log n}$$

and $f_n^{*(1)}$ is nondecreasing on I_{nk}^r, we deduce

$$\frac{k^{3/2}}{n} f_n^{*(1)}(t) \leq \frac{1}{\sqrt{k}} \sum_{j=1}^{k} (f(k/n + i/n) - f(i/n)) + 2\sigma \sqrt{3 \log n}.$$

A Taylor expansion for f yields

$$f_n^{*(1)}(t) \leq f^{(1)}(t) + \frac{k}{n} \| f^{(2)} \|_{I_{nk}^r, \infty} + 2\sigma n \sqrt{\frac{3 \log n}{k^3}}$$

from which (38) follows.

A.3. The taut string algorithm of Kovac (2007). We suppose that data $y_1, \ldots, y_n$ at time points $t_1 < t_2 < \cdots < t_n$ are given and first describe how to calculate the taut string approximation given some tube widths $\lambda_0, \lambda_1, \ldots, \lambda_n$. Subsequently, we describe how to determine these tube widths using a multiresolution criterion. Lower and upper bounds of a tube on $[0, n]$ are constructed by linear interpolation of the points $(i, Y_i - \lambda_i), i = 0, \ldots, n$ and $(i, Y_i + \lambda_i), i = 0, \ldots, n$, respectively, where $Y_0 = 0$ and $Y_k = Y_{k-1} + y_k$ for $k = 1, \ldots, n$. We consider a string $\tilde{F}_n$ forced to lie in this tube which passes through the points $(0, 0)$ and (n, Y_n) and is pulled tight. An explicit algorithm for doing this with computational complexity $O(n)$ is described in the Appendix of Davies and Kovac (2001). The taut string $\tilde{F}_n$ is linear on each interval $[i - 1, i]$ and its derivative $\tilde{f}_i = \tilde{F}_n(i) - \tilde{F}_n(i - 1)$ is used as an approximation for the data at t_i.

Our initial tube widths are $\lambda_0 = \lambda_n = 0$ and $\lambda_1 = \lambda_2 = \cdots = \lambda_n = \max(Y_0, \ldots, Y_n) - \min(Y_0, \ldots, Y_n)$. The default family $\mathcal{I}_n$ is the dyadic index set family

$$\mathcal{I}_n = \bigcup_{j,k \in \mathbb{N}_0} \{\{2^j k + 1, \ldots, 2^j (k+1)\} \cap \{1, \ldots, n\}\} \setminus \varnothing$$

which consists of at most $2n$ subsets of $\{1, \ldots, n\}$. Given some taut string approximation $\tilde{f}_1, \ldots, \tilde{f}_n$ using tube widths $\lambda_0, \ldots, \lambda_n$ we check whether

$$(81) \qquad \frac{1}{\sqrt{|I|}} \left| \sum_{i \in I} (y_i - f_i) \right| < \sigma_n \sqrt{\tau_n \log(n)}$$

is satisfied for each $I \in \mathcal{I}_n$. If this is not the case we generate new tube widths $\tilde{\lambda}_0, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ by setting $\tilde{\lambda}_0 = \tilde{\lambda}_n = 0$ and for $i = 1, \ldots, n - 1$

$$\tilde{\lambda}_i = \begin{cases} \lambda_i, & \text{if (81) is satisfied for all } I \in \mathcal{I} \text{ with } i \in I \text{ or } i + 1 \in I, \\ \lambda_i/2, & \text{otherwise.} \end{cases}$$

Then we calculate the taut string approximation corresponding to these new tube widths, check (81), possibly determine yet another set of tube widths and repeat this process until eventually (81) is satisfied for all $I \in \mathcal{I}_n$.

Acknowledgments. The authors gratefully acknowledge talks with Lutz Dümbgen which in particular lead to the smoothness regularization described in Section 4. We also acknowledge helpful comments made by two referees, an Associate Editor and an Editor, which have lead to a more focused article.

REFERENCES

BARAUD, Y. (2004). Confidence balls in Gaussian regression. *Ann. Statist.* **32** 528–551. MR2060168

BERNHOLT, T. and HOFMEISTER, T. (2006). An algorithm for a generalized maximum subsequence problem. In *LATIN 2006: Theoretical informatics*. *Lecture Notes in Comput. Sci.* **3887** 178–189. Springer, Berlin. MR2256330

CAI, T. T. and LOW, M. G. (2006). Adaptive confidence balls. *Ann. Statist.* **34** 202–228. MR2275240

DAVIES, P. L. (1995). Data features. *Statist. Neerlandica* **49** 185–245. MR1345378

DAVIES, P. L. and KOVAC, A. (2001). Local extremes, runs, strings and multiresolution (with discussion). *Ann. Statist.* **29** 1–65. MR1833958

DAVIES, P. L. and MEISE, M. (2008). Approximating data with weighted smoothing splines. *J. Nonparametr. Stat.* **20** 207–228. MR2421766

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089

DÜMBGEN, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26** 288–314. MR1611768

DÜMBGEN, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9** 423–449. MR1997491

DÜMBGEN, L. (2007). Confidence bands for convex median curves using sign-tests. In *Asymptotics: Particles, Processes and Inverse Problems* (E. Cator, G. Jongbloed, C. Kraaikamp, R. Lopuhaä and J. Wellner, eds.). *IMS Lecture Notes—Monograph Series* **55** 85–100. IMS, Hayward, USA. MR2459932

DÜMBGEN, L. and JOHNS, R. (2004). Confidence bands for isotonic median curves using sign-tests. *J. Comput. Graph. Statist.* **13** 519–533. MR2063998

DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London. MR1383587

GREEN, P. J. and SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, London. MR1270012

GROENEBOOM, P. (1996). Inverse problems in statistics. In *Proceedings of the St. Flour Summer School in Probability*. *Lecture Notes in Math.* **1648** 67–164. Springer, Berlin. MR1600884

HOFFMANN, M. and LEPSKI, O. (2002). Random rates in anisotropic regression. *Ann. Statist.* **30** 325–396. MR1902892

KABLUCHKO, Z. and MUNK, A. (2008). Exact convergence rate for the maximum of standardized Gaussian increments. *Electron. Commun. Probab.* **13** 302–310. MR2415138

KOVAC, A. (2007). *ftnonpar*. The R Project for Statistical Computing, Contributed Packages.

LI, K.-C. (1989). Honset confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001–1008. MR1015135

MAMMEN, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19** 741–759. MR1105842

MILDENBERGER, T. (2008). A geometric interpretation of the multiresolution criterion. *J. Nonparametr. Stat.* **20** 599–609. MR2454614

ROBINS, J. and VAN DER VAART, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229–253. MR2275241

WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia, PA. MR1045442

WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London. MR1319818

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā* **26** 101–116. MR0184336

P. L. DAVIES
UNIVERSITY OF DUISBURG–ESSEN
TECHNICAL UNIVERSITY EINDHOVEN
GERMANY
E-MAIL: laurie.davies@uni-due.de

A. KOVAC
UNIVERSITY OF BRISTOL
UNITED KINGDOM
E-MAIL: a.kovac@bristol.ac.uk

M. MEISE
UNIVERSITY OF DUISBURG–ESSEN
GERMANY
E-MAIL: monika.meise@uni-due.de