Submitted to the Annals of Statistics

ADAPTIVE CONFIDENCE INTERVALS FOR REGRESSION FUNCTIONS UNDER SHAPE CONSTRAINTS

By T. Tony Cai, * Mark G. Low and Yin Xia *

University of Pennsylvania

Adaptive confidence intervals for regression functions are constructed under shape constraints of monotonicity and convexity. A natural benchmark is established for the minimum expected length of confidence intervals at a given function in terms of an analytic quantity, the local modulus of continuity. This bound depends not only on the function but also the assumed function class. These benchmarks show that the constructed confidence intervals have near minimum expected length for each individual function, while maintaining a given coverage probability for functions within the class. Such adaptivity is much stronger than adaptive minimaxity over a collection of large parameter spaces.

1. Introduction. The construction of useful confidence sets is one of the more challenging problems in nonparametric function estimation. There are two main interrelated issues which need to be considered together, coverage probability and the expected size of the confidence set. For a fixed parameter space it is often possible to construct confidence sets which have guaranteed coverage probability over the parameter space while controlling the maximum expected size. However such minimax statements are often thought to be too conservative and a more natural goal is to have the expected size of the confidence set reflect in some sense the difficulty of estimating the particular underlying function.

These issues are well illustrated by considering confidence intervals for the value of a function at a fixed point. Let Y be an observation from the white noise model:

$$dY(t) = f(t)dt + n^{-\frac{1}{2}}dW(t), \quad -\frac{1}{2} \le t \le \frac{1}{2} \tag{1}$$

*The research of Tony Cai and Yin Xia was supported in part by NSF FRG Grant DMS-0854973.

AMS 2000 subject classifications: Primary 62G99; secondary 62F12,62F35,62M99

Keywords and phrases: Adaptation, confidence interval, convex function, coverage probability, expected length, minimax estimation, modulus of continuity, monotone function, nonparametric regression, shape constraint, white noise model.

where $W(t)$ is standard Brownian motion and f belongs to some parameter space $\mathcal{F}$. Suppose that we wish to construct a confidence interval for f at some point $t_0 \in (-\frac{1}{2}, \frac{1}{2})$. Let CI be a confidence interval for $f(t_0)$ based on observing the process Y and let $L(CI)$ denote the length of the confidence interval. The minimax point of view can then be expressed by: subject to the constraint on the coverage probability $\inf_{f\in\mathcal{F}} P(f(t_0) \in CI) \geq 1-\alpha$, minimize the maximum expected length $\sup_{f\in\mathcal{F}} E_f(L(CI))$.

As an example it is common to consider the Lipschitz classes

$$\Lambda(\beta, M) = \left\{ f : |f(y) - f(x)| \leq M|y-x|^{\beta} \text{ for } x, y \in [-\tfrac{1}{2}, \tfrac{1}{2}] \right\}, \quad \text{if } 0 < \beta \leq 1$$

and for $\beta > 1$

$$\Lambda(\beta, M) = \left\{ f : |f^{(\lfloor\beta\rfloor)}(x) - f^{(\lfloor\beta\rfloor)}(y)| \leq M\,|x-y|^{\beta'} \text{ for } x, y \in [-\tfrac{1}{2}, \tfrac{1}{2}] \right\},$$

where $\lfloor\beta\rfloor$ is the largest integer less than β and $\beta' = \beta - \lfloor\beta\rfloor$. For these classes it easily follows from results of Donoho (1994), Low (1997), and Evans, Hansen and Stark (2005) that the minimax expected length of confidence intervals, which have guaranteed coverage of $1-\alpha$ over $\Lambda(\beta, M)$, is of order $M^{\frac{1}{1+2\beta}} n^{-\frac{\beta}{1+2\beta}}$.

It should however be stressed that confidence intervals which achieve such an expected length rely on the knowledge of the particular smoothness parameters β and M, which are not known in most applications. Unfortunately, Low (1997) and Cai and Low (2004) have shown that the natural goal of constructing an adaptive confidence interval which has a given coverage probability and has expected length that is simultaneously close to these minimax expected lengths for a range of smoothness parameters is not in general attainable. More specifically suppose that a confidence interval has guaranteed coverage probability of $1-\alpha$ over $\Lambda(\beta, M)$. Then for any f in the interior of $\Lambda(\beta, M)$ the expected length for this f must also be of order $n^{-\frac{\beta}{1+2\beta}}$. In other words the minimax rate describes the actual rate for all functions in the class other than those on the boundary of the set. For example in the case that a confidence interval has guaranteed coverage probability of $1-\alpha$ over the Lipschitz class $\Lambda(1, M)$ then even if the underlying function has two derivatives and first derivative smaller than M, the confidence interval for $f(x)$ *must* still have expected length of order $n^{-1/3}$ even though one would hope that an adaptive confidence interval would have a much shorter length of order $n^{-2/5}$.

Despite these very negative results there are some settings where some degree of adaptation has been shown to be possible. In particular under certain shape constraints Hengartner and Stark (1995) constructed confidence

bands which have a guaranteed coverage probability of at least $1-\alpha$ over the collection of all monotone densities and which have maximum expected length of order $(\frac{\log n}{n})^{\frac{\beta}{2\beta+1}}$ for those monotone densities which are in $\Lambda(\beta, M)$ for a particular choice of β where $0 < \beta \leq 1$. This construction relies on the selection of a tuning parameter and is thus not adaptive. Dümbgen (2003) however does provide adaptive confidence bands with optimal rates for both isotonic and convex functions under supremum norm loss on arbitrary compact subintervals. These results are however still framed in terms of the maximum length over particular large parameter spaces and the existence of such intervals raises the question of exactly how much adaption is possible. It is this question that is the focus of the present paper.

Rather than considering the maximum expected length over large collections of functions we study the problem of adaptation to each and every function in the parameter space. We examine this problem in detail for two commonly used collections of functions that have shape constraints, namely the collection of convex functions and the collection of monotone functions. We focus on these parameter spaces as it is for such shape constrained problems for which there is some hope for adaptation. Within this context we consider the problem of constructing a confidence interval for the value of a function at a fixed point under both the white noise with drift model given in (1) as well as a nonparametric regression model. We show that within the class of convex functions and the class of monotone functions it is indeed possible to *adapt to each individual function*, and not just to the minimax expected length over different parameter spaces in a collection. The notion of adaptivity to a single function is also discussed in Lepski, Mammen and Spokoiny (1997) and Lepski and Spokoiny (1997) for the related point estimation problem but in that context a logarithmic penalty of the noise level must be paid and so in that context the notion of adaptivity is somewhat different.

This result is achieved in two steps. First we study the problem of minimizing the expected length of a confidence interval assuming that the data is generated from a particular function f in the parameter space, subject to the constraint that the confidence interval has guaranteed coverage probability over the entire parameter space. The solution to this problem gives a benchmark for the expected length which depends on the function f considered. It gives a bound on the expected length of any adaptive interval because if the expected length is smaller than this bound for *any* particular function, the confidence interval cannot have the desired coverage probability. In applications it is more useful to express the benchmark in terms of a local modulus of continuity, an analytic quantity that can be easily calcu-

lated for individual functions. In situations where adaptation is not possible this local modulus of continuity does not vary significantly from function to function. Such is the case in the settings considered in Low (1997). However in the context of convex or monotone functions the resulting benchmark does vary significantly and this opens up the possibility for adaptation in those settings.

Our second step is to actually construct adaptive confidence intervals. This is done separately for monotone functions and convex functions, with similar results. For example, an adaptive confidence interval is constructed which is shown to have expected length uniformly within an absolute constant factor of the benchmark for every convex function, while maintaining coverage probability over the collection of all convex functions. In other words, this confidence interval has smallest expected length, up to a universal constant factor, for each and every convex function within the class of all confidence intervals which guarantee a $1-\alpha$ coverage probability over all convex functions. A similar result is established for a confidence interval designed for monotone functions.

The rest of the paper is organized as follows. In section 2 the benchmark for the expected length at each monotone function or each convex function is established under the constraint that the interval has a given level of coverage probability over the collection of monotone functions or the collection of convex functions. Section 3 constructs data driven confidence intervals for both monotone functions and convex functions and shows that these confidence intervals maintain coverage probability and have expected length within an absolute constant factor of the benchmark given in Section 2 for each monotone function and convex function. Section 4 considers the nonparametric regression model and Section 5 discusses connections of our results with other work in the literature. Proofs are given in Section 6.

2. Benchmark and Lower Bound on Expected Length.

As mentioned in the introduction, the focus in this paper is the construction of confidence intervals which have expected length that adapts to the unknown function. The evaluation of these procedures depends on lower bounds which are given here in terms of a local modulus of continuity first introduced by Cai and Low (2011) in the context of point estimation of convex functions under mean squared error loss. These lower bounds provide a natural benchmark for our problems.

2.1. *Benchmark and Lower Bound.*

We focus in this paper on estimating the function f at 0 since estimation at other points away from the boundary is similar. For a given function class $\mathcal{F}$, write $\mathcal{I}_\alpha(\mathcal{F})$ for the collection of all

confidence intervals which cover $f(0)$ with guaranteed coverage probability of $1-\alpha$ for all functions in $\mathcal{F}$. For a given confidence interval CI, denote by $L(CI)$ the length of CI and $L(CI,f) = E_f(L(CI))$ the expected length of CI at a given function f. The minimum expected length at f of all confidence intervals with guaranteed coverage probability of $1-\alpha$ over $\mathcal{F}$ is then given by

$$L^*_\alpha(f,\mathcal{F}) = \inf_{CI\in\mathcal{I}_\alpha(\mathcal{F})} L(CI,f). \tag{2}$$

A natural goal is to construct a confidence interval with expected length close to the minimum $L^*_\alpha(f,\mathcal{F})$ for every $f\in\mathcal{F}$ while maintaining the coverage probability over $\mathcal{F}$. However although $L^*_\alpha(f,\mathcal{F})$ is a natural benchmark for the expected length of confidence intervals it is not easy to evaluate exactly. Instead as a first step towards our goal, we provide a lower bound for the benchmark $L^*_\alpha(f,\mathcal{F})$ in terms of a local modulus of continuity $\omega(\epsilon,f,\mathcal{F})$ introduced by Cai and Low (2011). The local modulus is a quantity that is more easily computable and techniques for its analysis are similar to those given in Donoho and Liu (1991) and Donoho (1994) where a global modulus of continuity was introduced in the study of minimax theory for estimating linear functionals. See the examples in Section 2.2.

For a parameter space $\mathcal{F}$ and function $f\in\mathcal{F}$, the local modulus of continuity is defined by

$$\omega(\epsilon,f,\mathcal{F}) = \sup\{|g(0)-f(0)| : \ \|g-f\|_2\le\epsilon,\ g\in\mathcal{F}\}. \tag{3}$$

where $\|\cdot\|_2$ is the $L_2(-\frac{1}{2},\frac{1}{2})$ function norm. The following theorem gives a lower bound for the minimum expected length $L^*_\alpha(f,\mathcal{F})$ in terms of the local modulus of continuity $\omega(\epsilon,f,\mathcal{F})$. In this theorem and throughout the paper we write Φ for the cumulative distribution function and ϕ for the density function of a standard normal density and set $z_\alpha = \Phi^{-1}(1-\alpha)$.

Theorem 1 *Suppose $\mathcal{F}$ is a nonempty convex set. Let $0<\alpha<\frac{1}{2}$ and $f\in\mathcal{F}$. Then for confidence intervals based on (1),*

$$L^*_\alpha(f,\mathcal{F}) \ge (1-\frac{1}{\sqrt{2\pi}z_\alpha}+\frac{\phi(z_\alpha)}{z_\alpha}-\alpha)\omega(\frac{z_\alpha}{\sqrt{n}},f,\mathcal{F}). \tag{4}$$

In particular,

$$L^*_\alpha(f,\mathcal{F}) \ge (1-\frac{1}{\sqrt{2\pi}z_\alpha})\omega(\frac{z_\alpha}{\sqrt{n}},f,\mathcal{F}). \tag{5}$$

The lower bounds given in Theorem 1 can be viewed as benchmarks for the evaluation of the expected length of confidence intervals when the true function is f for confidence intervals which have guaranteed coverage probability over all of $\mathcal{F}$. The bound depends on the underlying true function f as well as the parameter space $\mathcal{F}$.

The bounds from Theorem 1 are general. In some settings they can be used to rule out the possibility of adaptation, whereas in other settings they provide bounds on how much adaptation is possible. In particular the result ruling out adaptation over Lipschitz classes mentioned in the introduction easily follows from this theorem. For example, consider the Lipschitz class $\Lambda(\beta, M)$ and suppose that f is in the interior of $\Lambda(\beta, M)$. Straightforward calculations similar to those given in Section 2.2 show that

$$\omega(\epsilon, f, \Lambda(\beta, M)) \sim C\epsilon^{\frac{2\beta}{2\beta+1}}. \tag{6}$$

Now consider two Lipschitz classes $\Lambda(\beta_1, M_1)$ and $\Lambda(\beta_2, M_2)$ with $\beta_1 > \beta_2$. A fully adaptive confidence interval in this setting would have guaranteed coverage of $1-\alpha$ over $\Lambda(\beta_1, M_1) \cup \Lambda(\beta_2, M_2)$ and maximum expected length over $\Lambda(\beta_i, M_i)$ of order $n^{\frac{\beta_i}{2\beta_i+1}}$ for $i = 1$ and 2. However, it follows from Theorem 1 and (6) that for all confidence intervals with coverage probability of $1-\alpha$ over $\Lambda(\beta_2, M_2)$, for every $f \in \Lambda(\beta_2, M')$ with $M' < M_2$,

$$L^*_\alpha(f, \Lambda(\beta_2, M_2)) \geq C(\alpha) n^{-\frac{\beta_2}{2\beta_2+1}}$$

for some constant $C(\alpha)$ not depending on f. In particular this holds for all $f \in \Lambda(\beta_1, M_1) \cap \Lambda(\beta_2, M')$ and hence

$$\sup_{f \in \Lambda(\beta_1, M_1)} \inf_{CI \in \mathcal{I}_\alpha(\Lambda(\beta_1, M_1) \cup \Lambda(\beta_2, M_2))} L(CI, f) \geq C(\alpha) n^{-\frac{\beta_2}{2\beta_2+1}} \gg n^{-\frac{\beta_1}{2\beta_1+1}}.$$

Therefore it is not possible to have confidence intervals with adaptive expected length over two Lipschitz classes with different smoothness parameters.

In the present paper Theorem 1 will be used to provide benchmarks in the setting of shape constraints. Denote by F_m and F_c respectively the collection of all monotonically nondecreasing functions and the collection of all convex functions on $[-\frac{1}{2}, \frac{1}{2}]$. We shall now show that in these cases the modulus and the associated lower bounds vary significantly from function to function.

2.2. *Examples of Bounds For Monotone Functions and Convex Functions.* We now turn to the application of the lower bound given in Theorem 1 in the

case of monotone functions and convex functions. Here we shall evaluate the lower bound for four particular families of functions yielding different rates at which the expected length decreases to zero as the noise level decreases in contrast to the situation just described where the parameter space did not have an order constraint. Two of the functions will be both monotonically nondecreasing and convex. In this case the lower bound can also be quite different depending on whether we assume the knowledge that f is convex or monotonically nondecreasing.

The key quantity that is needed in any application of Theorem 1 is the local modulus. We follow the same approach as given in Donoho (1994) where a global modulus of continuity is considered for minimax estimation. In each case, for a given function f, we first minimize the L_2 norm between a function $g \in \mathcal{F}$ and the function f subject to the constraint that $|g(0) - f(0)| = a$ for some given value $a > 0$. From here it is easy to invert and thus maximize $|g(0) - f(0)|$ given a constraint on the L_2 norm between f and g.

Example 1 As a first example consider the linear function $f_k(t) = kt$ where $k \geq 0$ is a constant. This function is both monotonically nondecreasing and convex.

First consider the collection of monotonically nondecreasing functions F_m. We shall treat separately the case $k > 0$ and the case $k = 0$. For the moment we shall take $k > 0$. Suppose that $0 < a \leq \frac{k}{2}$. In this case $f_k \in F_m$ and a function g that minimizes $\|g - f_k\|_2$ subject to the constraint that $|g(0) - f_k(0)| = a$ is given by $g(t) = f_k(t)$ if $t < 0$, $g(t) = a$ if $0 \leq t \leq b$, and $g(t) = f_k(t)$ if $t > b$, where b satisfies $f_k(b) = a$. The assumption that $a \leq \frac{k}{2}$ guarantees $b \leq \frac{1}{2}$. We then have $\|g - f_k\|_2 = a^{\frac{3}{2}}/(3k)^{\frac{1}{2}}$. It follows that if $\epsilon^2 \leq \frac{1}{24}k^2$

$$\omega(\epsilon, f_k, F_m) = (3k)^{\frac{1}{3}}\epsilon^{\frac{2}{3}}$$

and consequently for $n \geq \frac{24z_\alpha^2}{k^2}$, if $k > 0$

$$L_\alpha^*(f_k, F_m) \geq (1 - \frac{1}{\sqrt{2\pi}z_\alpha})(3k)^{\frac{1}{3}}z_\alpha^{\frac{2}{3}}n^{-\frac{1}{3}}.$$

In the case that $k = 0$ a function g that minimizes $\|g - f_0\|_2$ subject to the constraint that $|g(0) - f_0(0)| = a$ is given by $g(t) = f_0(t)$ if $t < 0$, $g(t) = a$ if $0 \leq t \leq \frac{1}{2}$. In this case it is easy to check that $\|g - f_0\|_2 = \frac{1}{\sqrt{2}}\epsilon$ and hence

$$\omega(\epsilon, f_0, F_m) = \sqrt{2}\epsilon$$

and hence

$$L_\alpha^*(f_0, F_m) \geq (1 - \frac{1}{\sqrt{2\pi}z_\alpha})\sqrt{2}z_\alpha n^{-\frac{1}{2}}.$$

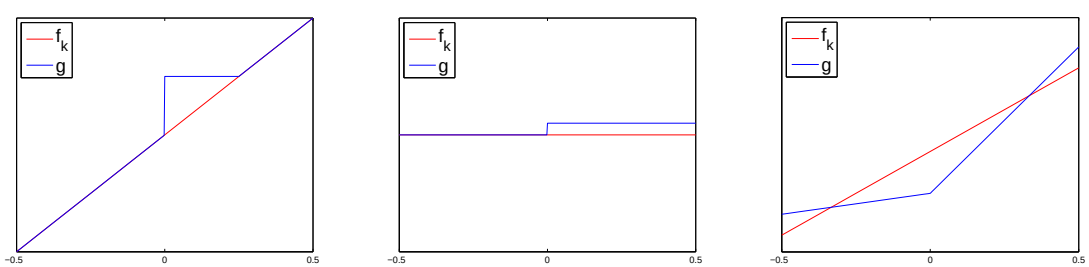
We now consider the bound for the length of the confidence interval for f_k belonging to the collection of convex functions. In this case we do not need to treat the cases $k > 0$ and $k = 0$ separately. The function g that minimize $\|g - f_k\|_2$ subject to the constraint that g is convex and $|g(0) - f_k(0)| = a$ is given by $g(t) = (k + 3a)t - a$ if $t \geq 0$ and $g(t) = (k - 3a)t - a$ if $t < 0$. In this case $\|g - f\|_2 = \frac{1}{2}a$. It then immediately follows that

$$\omega(\epsilon, f_k, F_c) = 2\epsilon$$

and so

$$L^*_\alpha(f_k, F_c) \geq (1 - \frac{1}{\sqrt{2\pi} z_\alpha})2z_\alpha n^{-\frac{1}{2}}.$$

It is important to note that for $k > 0$ the minimum expected lengths $L^*_\alpha(f_k, F_m)$ and $L^*_\alpha(f_k, F_c)$ are different, one of order $n^{-\frac{1}{3}}$ and another of order $n^{-\frac{1}{2}}$, although the function f_k is the same. It is also interesting to note that the expected length of the confidence for monotone functions is an increasing function of k whereas the expected length of the confidence for convex functions does not depend on k. Since we shall show that these bounds are achievable within a constant factor it follows that the minimum expected length of the confidence interval when f_k is the true function depends strongly on whether we specify that the underlying collection of functions is convex or monotone. Plots illustrating shapes of functions f_k and a least favorable function g are shown as below in Fig 1.



(a) $k > 0$ monotone (b) $k = 0$ monotone (c) convex

Fig 1: Plots of f_k and a least favorable function g in Example 1 with the constraints $|g(0) - f_k(0)| = a$.

Example 2 As a second example which is also both monotonically nondecreasing and convex consider the function $f(t) = k_1 t + k_2 t^r I(0 < t \leq \frac{1}{2})$ where $r \geq 1$ and $k_1 \geq 0$ and $k_2 > 0$ are constants.

We consider the cases $r = 1$ and $r > 1$ separately. When $r = 1$ the function is piecewise linear with the change of slope at 0. In this case suppose $0 < a \leq \frac{k_1+k_2}{2}$. A monotonically nondecreasing function $g \in F_m$ that minimize $\|g-f\|_2$ subject to the constraint that $|g(0)-f(0)| = a$ is given by $g(t) = f(t)$ if $t < 0$, $g(t) = a$ if $0 \leq t \leq b$, and $g(t) = f(t)$ if $t > b$, where b satisfies $f(b) = a$. The constraint $a \leq \frac{k_1+k_2}{2}$ is to guarantee that such a b exists with $b \leq \frac{1}{2}$. Then we have $\|g - f\|_2 = a^{\frac{3}{2}}(3(k_1 + k_2))^{-\frac{1}{2}}$ and it follows that if $\epsilon^2 \leq \frac{1}{24}(k_1 + k_2)^2$

$$\omega(\epsilon, f, F_m) = (3(k_1 + k_2))^{\frac{1}{3}}\epsilon^{\frac{2}{3}}$$

and consequently for $n \geq \frac{24z_\alpha^2}{(k_1+k_2)^2}$,

$$L^*_\alpha(f, F_m) \geq (1 - \frac{1}{\sqrt{2\pi}z_\alpha})(3(k_1 + k_2))^{\frac{1}{3}}z_\alpha^{\frac{2}{3}}n^{-\frac{1}{3}}.$$

We can also give a lower bound on the expected length for this same function for confidence intervals which guarantee coverage over the class of convex functions. Suppose $0 < a \leq \frac{k_2}{4}$. Here we need to find the convex h that minimizes $\|h - f\|_2$ subject to the constraints that $|h(0) - f(0)| = a$. It is given by $h(t) = f(t)$ if $t \leq -\frac{2a}{k_2}$, $h(t) = (\frac{k_2}{2} + k_1)t + a$ if $-\frac{2a}{k_2} \leq t \leq \frac{2a}{k_2}$ and $h(t) = f(t)$ if $t \geq \frac{2a}{k_2}$. Then $\|f - g\|_2 = 2a^{\frac{3}{2}}/(3k_2)^{\frac{1}{2}}$ and it follows that if $\epsilon^2 \leq \frac{k_2^2}{48}$,

$$\omega(\epsilon, f, F_c) = (3k_2/4)^{\frac{1}{3}}\epsilon^{\frac{2}{3}}.$$

Hence, for $n \geq \frac{48z_\alpha^2}{k_2^2}$,

$$L^*_\alpha(f, F_c) \geq (1 - \frac{1}{\sqrt{2\pi}z_\alpha})(3k_2/4)^{\frac{1}{3}}z_\alpha^{\frac{2}{3}}n^{-\frac{1}{3}}.$$

We now turn to the case where $r > 1$. Suppose $0 < a \leq \frac{k_1}{2} + k_2(\frac{1}{2})^r$. In this case the monotonically nondecreasing function g that minimizes $\|g - f\|_2$ subject to the constraints that $|g(0) - f(0)| = a$ is given by $g(t) = f(t)$ if $t < 0$, $g(t) = a$ if $0 \leq t \leq b$, and $g(t) = f(t)$ if $t > b$, where b satisfies $f(b) = a$. As before the condition $0 < a \leq \frac{k_1}{2} + k_2(\frac{1}{2})^r$ guarantees that b exists with $b < \frac{1}{2}$. In this case $a^{\frac{3}{2}}(3k_1)^{-\frac{1}{2}} - ca^s \leq \|g-f\|_2 \leq a^{\frac{3}{2}}(3k_1)^{-\frac{1}{2}} + ca^s$ for some constant $c > 0$ and $s > 3/2$. It follows that if $\epsilon^2 \leq \frac{1}{24}k_1^2 + (1 + \frac{1}{2r+1} - \frac{2}{r+1})(\frac{1}{2})^{2r+1}k_2^2 + (\frac{1}{2} - \frac{1}{r+1} + \frac{1}{r+2})(\frac{1}{2})^{r+1}k_1k_2$, then

$$\omega(\epsilon, f, F_m) = (3k_1)^{\frac{1}{3}}\epsilon^{\frac{2}{3}}(1 + o(1)).$$

Hence,

$$L_\alpha^*(f, F_m) \geq (1 - \frac{1}{\sqrt{2\pi} z_\alpha})(3k_1)^{\frac{1}{3}} z_\alpha^{\frac{2}{3}} n^{-\frac{1}{3}}(1 + o(1)).$$

For a bound on the expected length of this same function for confidence intervals with coverage guaranteed over the collection of convex functions we suppose $0 < a \leq k_2(\frac{1}{2})^{r+1}$. In this case the convex function h that minimizes $\|h - f\|_2$ subject to the constraints that $|h(0) - f(0)| = a$, is given by $h(t) = kt + a$, $k > k_1$, if $x_0 \leq t \leq x_1$ and $h(t) = f(t)$ otherwise, where (x_0, cx_0) and $(x_1, cx_1 + x_1^r)$ are the intersection points of $f(t)$ and the line $kt + a$. Then the function h with slope k_0 that minimize $\|h - f\|_2$ would be the least favorable function. It follows that, if $\epsilon^2 \leq \frac{k_2^2}{24}(\frac{1}{2})^{2r}$,

$$\omega(\epsilon, f, F_c) = C(r) k_2^{\frac{1}{2r+1}} \epsilon^{\frac{2r}{2r+1}}$$

and consequently for $n \geq \frac{24 z_\alpha^2 2^{2r}}{k_2^2}$,

$$L_\alpha^*(f, \mathcal{F}_c) \geq (1 - \frac{1}{\sqrt{2\pi} z_\alpha}) C(r) k_2^{\frac{1}{2r+1}} z_\alpha^{\frac{2r}{2r+1}} n^{-\frac{r}{2r+1}},$$

where $C(r) > 0$ is a constant depending on r only.

It is interesting to note that in this example the rates of convergence for $L_\alpha^*(f, \mathcal{F}_m)$ and $L_\alpha^*(f, \mathcal{F}_c)$ are the same for the case $r = 1$, and are different when $r > 1$. Plots illustrating shapes of functions f and a least favorable function g are shown as below in Fig 2.

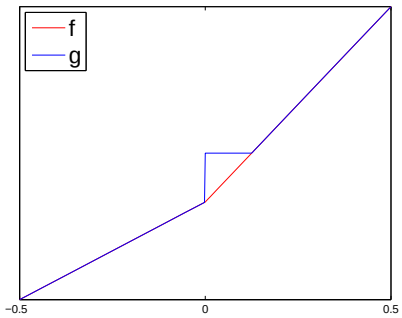
Next we consider a function which is monotonically nondecreasing but not convex.

Example 3 Let $f(t) = kt^r$ for some constant $k > 0$ and $r = 2l + 1$ or $r = \frac{1}{2l+1}$ for $l = 0, 1, 2, \cdots$. Suppose that $a < (\frac{1}{2})^r k$. In this case a function g that minimizes $||g - f||_2$ subject to the constraint that $|g(0) - f(0)| = a$ is given by $g(t) = f(t)$ if $t < 0$, $g(t) = a$ if $0 \leq t \leq b$, and $g(t) = f(t)$ if $t > b$, where b satisfies $f(b) = a$. As before the condition $a < (\frac{1}{2})^r k$ guarantees that b exists with $b < \frac{1}{2}$. Then $||g - f||_2 = a^{1+\frac{1}{2r}} / k^{\frac{1}{2r}} (2r^2/(r+1)(2r+1))^{\frac{1}{2}}$ and it follows that when $\epsilon^2 \leq (\frac{1}{2})^{2r+1} k^2 \frac{2r^2}{(r+1)(2r+1)}$,

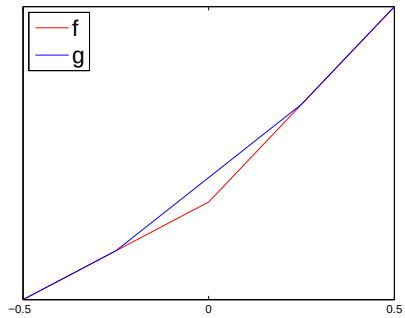
$$\omega(\epsilon, f, F_m) = (\frac{(r+1)(2r+1)k}{2r^2})^{\frac{r}{2r+1}} \epsilon^{\frac{2r}{2r+1}}.$$

Hence for $n \geq \frac{2^{2r+1}(r+1)(2r+1) z_\alpha^2}{2r^2 k^2}$,

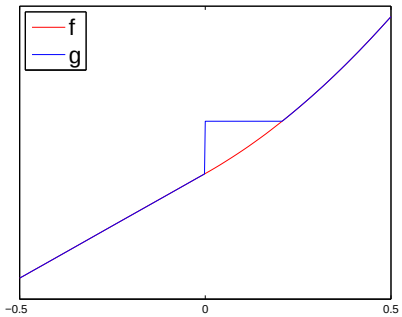
$$L_\alpha^*(f, \mathcal{F}_c) \geq (1 - \frac{1}{\sqrt{2\pi} z_\alpha})(\frac{(r+1)(2r+1)k}{2r^2})^{\frac{r}{2r+1}} z_\alpha^{\frac{2r}{2r+1}} n^{-\frac{r}{2r+1}}$$



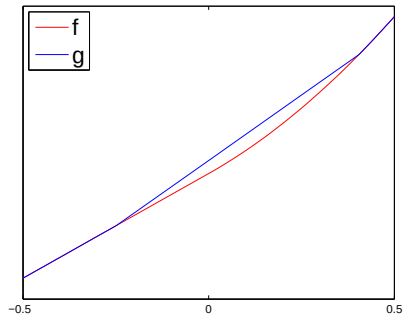
(a) $r = 1$ monotone



(b) $r = 1$ convex



(c) $r > 1$ monotone



(d) $r > 1$ convex

Fig 2: Plots of f and a least favorable function g in Example 2 with the constraints $|g(0) - f(0)| = a$.

and once again it is clear that the rate at which the expected length decreases to zero depends strongly on the value of r.

As a final example we consider a function which is convex but not monotonically nondecreasing.

Example 4 Let $f(t) = t^2$. Suppose that $a < 1/2$. In this case the function g that minimizes $||g - f||_2$ subject to the constraint that $|g(0) - f(0)| = a$ is given by $g(t) = -3\sqrt{a/2}t - a$ if $-\sqrt{2a} \leq t \leq 0$, $g(t) = 3\sqrt{a/2}t - a$ if $0 \leq t \leq \sqrt{2a}$, and $g(t) = f(t)$ otherwise. Then $||g - f||_2 = 2^{5/4}/\sqrt{15}a^{5/4}$ and it follows that when $\epsilon^2 \leq 1/\sqrt{15}$,

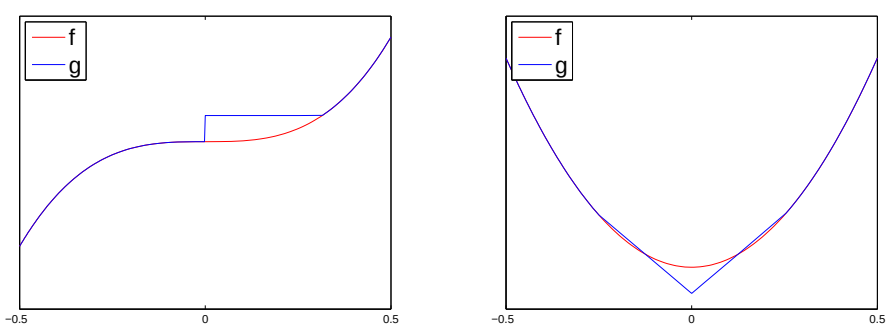
$$\omega(\epsilon, f, F_m) = \frac{15^{2/5}}{2}\epsilon^{\frac{4}{5}}.$$

Hence for $n \geq \sqrt{15}z_\alpha^2$,

$$L_\alpha^*(f, \mathcal{F}_c) \geq (1 - \frac{1}{\sqrt{2\pi}z_\alpha})\frac{15^{2/5}}{2}z_\alpha^{\frac{4}{5}}n^{-\frac{2}{5}}.$$

A similar minimization problem is solved in Dümbgen (2003).

Plots illustrating shapes of functions f and a least favorable function g for both Examples 3 and 4 are shown as below in Fig 3.



(a) Example 3 (b) Example 4

Fig 3: Plots of f and a least favorable function g in Examples 3 and 4 with the constraints $|g(0) - f(0)| = a$.

3. Confidence Procedures. In this section we both construct and give an analysis of adaptive confidence intervals for monotone functions and

convex functions. The procedures are easily implementable. We consider the class of monotonically nondecreasing functions and the class of convex functions. Concave functions and monotonically nonincreasing functions can be handled similarly.

3.1. *Construction.* The construction is split into two steps. In the first step a countable collection of confidence intervals is created each of which has guaranteed coverage probability. These intervals are based on a collection of pairs of linear estimators. For each interval one of the estimators has nonnegative bias and the other nonpositive bias. The one sided control of the bias of these estimators is a key special feature in these problems and an important part of what makes it possible to adapt to every individual function. Moreover for each function f this collection has at least one interval with expected length within a constant factor of the local modulus bound given in Theorem 1. The second step is to select from this collection a particular interval.

In the case of monotonically nondecreasing functions we take for each $j \geq 2$, pairs of estimators $\delta_j^R = 2^j(Y(2^{-j}) - Y(0))$ and $\delta_j^L = 2^j(Y(0) - Y(-2^{-j}))$. Then for estimating $f(0)$ it is easy to check that δ_j^R has nonnegative and monotonically nonincreasing biases while δ_j^L have nonpositive and monotonically nondecreasing biases. The one sided control of the biases of these estimators over the class of all monotonically nondecreasing functions easily allows for the construction of a confidence interval. For that we shall need the standard deviation of δ_j^R and δ_j^L. In order to give a unified treatment in both the monotone and convex case it is useful to establish a common notation. Here we shall set $\sigma_j^2 = \frac{2^{j-1}}{n}$. It is then easy to check that both δ_j^R and δ_j^L have a standard deviation of $\sqrt{2}\sigma_j$. It is then also easy to see that for each $j \geq 2$, the confidence interval $CI_j^m(\alpha)$ given by

$$CI_j^m(\alpha) = [\delta_j^L - z_{\alpha/2}\sqrt{2}\sigma_j, \delta_j^R + z_{\alpha/2}\sqrt{2}\sigma_j] \tag{7}$$

has guaranteed coverage of $1-\alpha$. We should however note that in (7) the left endpoint of the interval may be larger than the right end point in which case we adopt the convention that the confidence interval is just the empty set. The length of this confidence interval is then $\max(\delta_j^R - \delta_j^L + 2\sqrt{2}z_{\alpha/2}\sigma_j, 0)$.

In the case of convex functions for $j \geq 1$ let $\delta_j = 2^{j-1}(Y(2^{-j}) - Y(-2^{-j}))$ and let $\tilde{\delta}_j = 2\delta_{j+1} - \delta_j$.

The following Lemma shows that for convex functions δ_j have nonnegative and monotonically nonincreasing biases and that $\tilde{\delta}_j$ have nonpositive and monotonically nondecreasing biases.

Lemma 1 *For any convex function f,*

$$0 \le Bias(\delta_{j+1}) \le \frac{1}{2} Bias(\delta_j) \tag{8}$$

$$E\delta_j - 3E\delta_{j+1} + 2E\delta_{j+2} \ge 0. \tag{9}$$

It is also easy to check that the standard deviation of δ_j is equal to σ_j where $\sigma_j^2 = \frac{2^{j-1}}{n}$ and that $2\delta_{j+1} - \delta_j$ has a standard deviation of $\sqrt{5}\sigma_j$. It then follows from the signs of the biases of δ_{j+1} and $2\delta_{j+1} - \delta_j$ that for any given j

$$CI_j^c(\alpha) = [2\delta_{j+1} - \delta_j - z_{\alpha/2}\sqrt{5}\sigma_j,\ \delta_{j+1} + z_{\alpha/2}\sigma_{j+1}] \tag{10}$$

gives a confidence interval with coverage probability of at least $1 - \alpha$. We should also note once again that the left endpoint of the interval may be larger than the right end point in which case the confidence interval is taken to be the empty set and so in this case the length of this confidence interval is $\max(\delta_j - \delta_{j+1} + (\sqrt{5} + \sqrt{2})z_{\alpha/2}\sigma_j, 0)$.

These results, for which a more formal proof is given in Section 6 are summarized in the following proposition.

Proposition 1 *For every $j \ge 2$, the confidence interval CI_j^m defined in (7) has coverage probability of at least $1-\alpha$ for all monotonically nondecreasing function $f \in F_m$ and for every $j \ge 1$, the confidence interval CI_j^c defined in (10) has coverage probability of at least $1 - \alpha$ for all convex functions $f \in F_c$.*

The second stage in the construction is that of selecting from these collections of intervals the one to be used. First note that one should not select the shortest interval since the collections defined in (7) and (10) will always contain one which corresponds to the empty set. A more sensible goal is to try to select the interval with the smallest expected length or at least one which has expected length close to the smallest expected length.

The approach we take here is to choose an interval for which the expected length is of the same order of magnitude as the standard deviation of the length. Such an interval will always have expected length close to the shortest expected length. For the case of monotonically nondecreasing functions the selection of the interval from the countable collection in (7) can be done by creating another collection of estimators which can be used to estimate the expected length of the intervals.

More specifically set $\xi_j = 2^{j-1}(Y(2^{-j+1}) - Y(2^{-j})) - 2^{j-1}(Y(-2^{-j}) - Y(-2^{-j+1}))$. Then for $j \ge 2$, ξ_j's are independent of each other and both δ_j^R and δ_j^L are independent of ξ_k for every $k \le j$. We should note that

the estimators ξ_j are similar to $\delta_j^R - \delta_j^L$ in that they are both differences of averages of Y to the left and right of the origin and thus estimate the average local change of the function. However $\delta_j^R - \delta_j^L$ are not independent for different j whereas the ξ_j are independent. It is thus natural to view the ξ_j as a surrogate for $\delta_j^R - \delta_j^L$ with the technical advantage that they are independent. The selection of a j for which ξ_j has expected value close to σ_j will then result in a confidence interval CI_j^m close to the one with the smallest expected length. The independence properties of the ξ_j allows us to guarantee a $1-\alpha$ coverage probability while making this selection.

More specifically the construction proceeds as follows. Let

$$\hat{j} = \inf_j \left\{ j : \xi_j \leq \frac{3}{2} z_\alpha \sigma_j \right\} \tag{11}$$

and define the final confidence interval by

$$CI_*^m = CI_{\hat{j}}^m(\alpha) \tag{12}$$

Before we turn to the analysis of this procedure we also introduce here a related confidence procedure in the convex case. Here rather than introducing an independent estimate of the difference between the two estimators used in constructing the confidence interval we proceed more directly. The basic idea is similar but the dependence between the estimates of j and the confidence interval constructed from this estimate requires that we adjust the original coverage level of our CI_j^c.

More specifically let $T_j = \delta_j - \delta_{j+1}$. When the expected value of T_j is the same order as σ_j, the confidence interval CI_j^c will then be close to the one with the smallest expected length. Our estimate of j is given by an empirical version, namely

$$\hat{j} = \inf_j \{ j : T_j \leq z_\alpha \sigma_j \} \tag{13}$$

Although this estimate can be used to select the appropriate CI_j^c to use, as just mentioned, care also needs to be taken to make sure that the resulting selected interval maintains the required coverage probability. The analysis given below shows that a choice of $\alpha/6$ in the construction of the original collection of intervals guarantees an overall coverage probability of α. Thus in the case of convex functions we define our interval by

$$CI_*^c = CI_{\hat{j}}^c(\frac{\alpha}{6}) \tag{14}$$

3.2. *Analysis of the Confidence Intervals.* In this section we present the properties of the confidence intervals CI_*^m and CI_*^c defined by (12) and (14) focusing on the coverage and the expected length of these intervals.

We begin with the confidence interval CI_*^m. In this case it is easy to check the coverage probability of CI_*^m by the independence of the interval CI_j^m and ξ_k for every k satisfying $2 \le k \le j$.

The key to the analysis of the expected length is the introduction of j_*^m where

$$j_*^m = \arg\min_j \{j : \ E\xi_j \le z_\alpha \sigma_j\}. \tag{15}$$

The analysis of the expected length relies on showing that $\hat{j}$ is highly concentrated around j_*^m. The concentration of $\hat{j}$ around j_*^m then provides a bound on the expected length of CI^*. These results, for which a proof is given in Section 6 are summarized in the following theorem.

Theorem 2 *Let $0 < \alpha \le 0.2$. The confidence interval CI_*^m defined in (12) has coverage probability of at least $1-\alpha$ for all monotonically nondecreasing functions $f \in F_m$ and satisfies*

$$E_f L(CI_*^m) \le 1.21(3z_\alpha + 2\sqrt{2} z_{\frac{\alpha}{2}})\sigma_{j_*^m} \le c_0 z_\alpha \sigma_{j_*^m}, \tag{16}$$

where c_0 is a constant and can be taken to be 8.85 for all $0 < \alpha \le 0.2$.

Remark 1 The constant c_0 in Theorem 2 depends on the upper limit of α. c_0 can be smaller if the upper limit on α is reduced. For example, for common choices of $\alpha = 0.05$ or 0.01, $c_0 \le 7.71$ for $\alpha = 0.05$, and $c_0 \le 7.42$ for $\alpha = 0.01$.

Theorem 2 shows that the coverage probability is attained and also provides an upper bound on the expected length in terms of $\sigma_{j_*^m}$. In order to to establish that this expected length is within a constant factor of the lower bound given in Theorem 1, we need to provide a lower bound for $L_\alpha^*(f, F_m)$ in terms of $z_\alpha \sigma_{j_*^m}$. This connection is given in the following theorem.

Theorem 3 *Let $0 < \alpha \le 0.2$ and let $f \in F_m$. Then*

$$L_\alpha^*(f, F_m) \ge (1 - \frac{1}{\sqrt{2\pi} z_\alpha})\frac{1}{\sqrt{2}} z_\alpha \sigma_{j_*^m}. \tag{17}$$

Combining Theorems 2 and 3, we have

$$E_f L(CI_*^m) \le c_1 L_\alpha^*(f, F_m) \tag{18}$$

for all monotonically nondecreasing function $f \in F_m$, where c_1 is a constant depending on α only. For example, c_1 can be taken to be 14.40 for $\alpha = 0.05$ and 12.67 for $\alpha = 0.01$. Hence, the confidence interval CI_*^m is uniformly within a constant factor of the benchmark $L_\alpha^*(f, F_m)$ for all monotonically nondecreasing functions f and all confidence level $1 - \alpha \geq 0.8$.

We now turn to an analysis of the properties of the confidence interval CI_*^c defined in (14). The key to this analysis is the introduction of j_*^c where

$$j_*^c = \arg\min_j \left\{ j : \ ET_j \leq \frac{2}{3} z_\alpha \sigma_j \right\}. \tag{19}$$

The analysis of both the coverage probability and the expected length relies on showing that $\hat{j}$ is highly concentrated around j_*^c. The probability of not covering $f(0)$ can be bounded by

$$P(f(0) \notin CI_*^c) \leq P(\hat{j} \leq j_*^c - 3) + P(\hat{j} \geq j_*^c + 3) + \sum_{l=-2}^{2} P(f(0) \notin CI_{j_*^c + l}). \tag{20}$$

The first two terms are controlled by the high concentration of $\hat{j}$ around j_*^c and the last term is controlled by Proposition 1 which bounds the coverage probability of any given j. The concentration of $\hat{j}$ around j_*^c also allows control on the expected length of CI_*^c which leads to the following theorem.

Theorem 4 *Let $0 < \alpha \leq 0.2$. The confidence interval CI_*^c defined in (14) has coverage probability of at least $1 - \alpha$ for all convex f and satisfies*

$$E_f L(CI_*^c) \leq 1.25(z_\alpha + (\sqrt{5} + \sqrt{2}) z_{\frac{\alpha}{12}}) \sigma_{j_*^c} \leq c_0 z_\alpha \sigma_{j_*^c} \tag{21}$$

where c_0 is a constant and can be taken to be 12.79 for all $0 < \alpha \leq 0.2$.

Remark 2 The constant c_0 in Theorem 4 depends on the upper limit of α. c_0 can be smaller if the upper limit on α is reduced. For example, for common choices of $\alpha = 0.05$ or 0.01, $c_0 \leq 8.57$ for $\alpha = 0.05$, and $c_0 \leq 7.42$ for $\alpha = 0.01$.

Theorem 4 shows that the coverage probability is attained and also provides an upper bound on the expected length in terms of $\sigma_{j_*^c}$. As was the case for monotone functions, in order to to establish that this expected length for convex functions is within a constant factor of the lower bound given in Theorem 1, we need to provide a lower bound for $L_\alpha^*(f, F_c)$ in terms of $z_\alpha \sigma_{j_*^c}$. This connection is given in the following theorem.

Theorem 5 *Let $0 < \alpha \leq 0.2$ and let $f \in F_c$. Then*

$$L_\alpha^*(f, F_c) \geq (1 - \frac{1}{\sqrt{2\pi} z_\alpha}) \frac{\sqrt{2}}{3} z_\alpha \sigma_{j_*^c}. \tag{22}$$

Theorems 4 and 5 together yield

$$E_f L(CI_*^c) \le c_2 L_\alpha^*(f, F_c) \tag{23}$$

for all convex function $f \in F_c$, where c_2 is a constant depending on α only. For example, c_2 can be taken to be 24 for $\alpha = 0.05$ and 19 for $\alpha = 0.01$. Hence, the confidence interval CI_*^c is uniformly within a constant factor of the benchmark $L_\alpha^*(f, F_c)$ for all convex functions f and all confidence level $1 - \alpha \ge 0.8$.

4. Nonparametric Regression. We have so far focused on the white noise model. The theory presented in the earlier sections can also easily be extended to nonparametric regression. Consider the regression model

$$y_i = f(x_i) + \sigma z_i, \quad i = -n, -(n-1), -1, 0, 1, ..., n \tag{24}$$

where $x_i = \frac{i}{2n}$ and $z_i \overset{iid}{\sim} N(0,1)$ and where for notational convenience we index the observations from $-n$ to n. Note that the noise level σ can be accurately estimated easily, as in Hall, et al (1990) or Munk, et al (2005). See also Wang, et al (2008). We shall thus assume it is known in this section. Then under the assumption that f is convex or monotone, we wish to provide a confidence interval for $f(0)$.

4.1. *Monotone Regression.* Let $J = \lfloor \log_2 n \rfloor$. For $1 \le j \le J$ define the local average estimators

$$\bar{\delta}_j^R = 2^{-j+1} \sum_{k=1}^{2^{j-1}} y_k \quad \text{and} \quad \bar{\delta}_j^L = 2^{-j+1} \sum_{k=1}^{2^{j-1}} y_{-k}. \tag{25}$$

We should note that the indexing scheme is the reverse of that given for the white noise with drift process. Here estimators $\bar{\delta}_j^R$ (or $\bar{\delta}_j^L$) with small values of j have smaller bias (or larger bias) and larger variance than those with larger values of j.

As in the white noise model it is easy to check that $\bar{\delta}_j^R$ has nonnegative bias and $\bar{\delta}_j^L$ has nonpositive bias. Simple calculations show that the variance of $\bar{\delta}_j^R$ and $\bar{\delta}_j^L$ are both $2\sigma_j^2$, where $\sigma_j^2 = 2^{-j}\sigma^2$. It is also important to introduce $\bar{\xi}_j$ as in the white noise case, where $\bar{\xi}_j = 2^{-j} \sum_{k=2^{j-1}+1}^{2^j} (y_k - y_{-k})$. It is easy to check that $E\bar{\xi}_j \le E\bar{\xi}_{j+1}$, $\bar{\xi}_j$'s are independent with each other and both $\bar{\delta}_j^R$ and $\bar{\delta}_j^L$ are independent with $\bar{\xi}_k$ for every $k \ge j$.

It then follows that $CI_j^m = [\bar{\delta}_j^L - z_{\frac{\alpha}{2}}\sqrt{2}\sigma_j, \bar{\delta}_j^R + z_{\frac{\alpha}{2}}\sqrt{2}\sigma_j]$ has guaranteed coverage probability of at least $1 - \alpha$ over all monotonically nondecreasing functions.

Now set

$$\hat{j} = \begin{cases} \max_j \left\{ j : \bar{\xi}_j \leq \frac{3}{2} z_\alpha \sigma_j \right\}, & \text{if } \bar{\xi}_1 \leq \frac{3}{2} z_\alpha \sigma_1; \\ 1, & \text{otherwise.} \end{cases}, \tag{26}$$

and define the confidence interval to be

$$CI_*^m = CI_{\hat{j}}^m. \tag{27}$$

The properties of this confidence interval can then be analyzed in the same way as before and can be shown to be similar to those for the white noise model. In particular, the following result holds.

Theorem 6 *Let $0 < \alpha \leq 0.2$. The confidence interval CI_*^m defined in* (27) *has coverage probability of at least $1 - \alpha$ for all monotone function f and satisfies*

$$E_f L(CI_*^m) \leq C_1 L_\alpha^*(f, F_m) \tag{28}$$

for all monotonically nondecreasing function $f \in F_m$, where $C_1 > 0$ is a constant depending on α only.

4.2. *Convex Regression.* As in the monotone case, set $J = \lfloor \log_2 n \rfloor$. For $1 \leq j \leq J$ define the local average estimators

$$\bar{\delta}_j = 2^{-j} \sum_{k=1}^{2^{j-1}} (y_{-k} + y_k). \tag{29}$$

We should note that this indexing scheme is the reverse of that given for the white noise with drift process. Here estimators $\bar{\delta}_j$ with small values of j have smaller bias and larger variance than those with larger values of j.

As in the white noise model it is easy to check that $\bar{\delta}_j$ has nonnegative bias. It is also important to introduce an estimate which has a similar variance but is guaranteed to have nonpositive bias. The key step is to introduce

$$T_j = \bar{\delta}_j - \bar{\delta}_{j-1}, \tag{30}$$

as an estimate of the bias of $\bar{\delta}_j$. The following lemma gives the required properties of $\bar{\delta}_j$ and T_j.

Lemma 2 *For any convex function f,*

$$2ET_j \leq ET_{j+1} \tag{31}$$

$$0 \leq \text{Bias}(\bar{\delta}_j) \leq \frac{2^{j-1}+1}{2^j+1} \text{Bias}(\bar{\delta}_{j+1}) \tag{32}$$

From (32) it is clear that the biases of the estimators $\bar{\delta}_j$ are nonnegative and monotonically nondecreasing. In addition straightforward calculations using both (31) and (32) show that the estimators

$$\delta_j^L = (2 + 2^{-(j-1)})\bar{\delta}_j - (1 + 2^{-(j-1)})\bar{\delta}_{j+1} = \bar{\delta}_j - (1 + 2^{-(j-1)})T_{j+1}$$

have a nonpositive and monotonically nonincreasing biases. Simple calculations show that the variance of δ_j^L is $\tau_j^2 = (5 + 2^{-j+3} + 2^{-2j+2})2^{-j-1}\sigma^2$.

It then follows that $CI_j^c = [\bar{\delta}_j - (1 + 2^{-(j-1)})T_{j+1} - z_{\alpha/12}\tau_j, \bar{\delta}_j + z_{\alpha/12}\sigma_j]$ has coverage over all convex functions.

Now set

$$\hat{j} = \begin{cases} \max_j \{j : T_j \le z_\alpha \sigma_j\}, & \text{if } T_2 \le z_\alpha \sigma_2; \\ 1, & \text{otherwise.} \end{cases}, \tag{33}$$

and define the confidence interval to be

$$CI_*^c = CI_{\hat{j}}^c. \tag{34}$$

This confidence interval shares similar properties as the one for the white noise model. In particular, the following result holds.

Theorem 7 *Let $0 < \alpha \le 0.2$. The confidence interval CI_*^c defined in* (34) *has coverage probability of at least $1-\alpha$ for all convex function f and satisfies*

$$E_f L(CI_*^c) \le C_2 L_\alpha^*(f, F_c) \tag{35}$$

for all convex function $f \in F_c$, where $C_2 > 0$ is a constant depending on α only.

5. Discussions. The major emphasis of the paper has been to show that with shape constraints it is possible to construct confidence intervals that have expected length that adapts to individual functions. In this section we shall discuss briefly the maximum expected lengths of our procedures over Lipschitz classes that are either monotone or convex in a way that is similar to that provided in Dümbgen (1998) and Dümbgen (2003) for the maximum width of a confidence band. We shall also explain how our results can be extended to the problem of estimating the value of f at points other than 0.

5.1. *Minimax results.* Although the focus of the present paper has been on the construction of a confidence interval with the expected length adaptive to each individual convex or monotone function, these results do yield

immediately adaptive minimax results for the expected length in the conventional sense. Define

$$F_c(\beta, M) = F_c \cap \Lambda(\beta, M) \quad \text{and} \quad F_m(\beta, M) = F_m \cap \Lambda(\beta, M).$$

The following results are direct consequence of Theorems 2 and 4.

Corollary 1 *(i). The confidence interval CI_*^m defined in (12) satisfies*

$$\sup_{f \in F_m(\beta, M)} E_f L(CI_*^m) \leq C_1 M^{\frac{1}{1+2\beta}} n^{-\frac{\beta}{1+2\beta}}. \tag{36}$$

simultaneously for all $0 \leq \beta \leq 1$ and $1 < M < \infty$, for some absolute constant $C_1 > 0$.

(ii). The confidence interval CI_*^c defined in (14) satisfies

$$\sup_{f \in F_c(\beta, M)} E_f L(CI_*^c) \leq C_2 M^{\frac{1}{1+2\beta}} n^{-\frac{\beta}{1+2\beta}}. \tag{37}$$

simultaneously for all $1 \leq \beta \leq 2$ and $1 < M < \infty$, for some absolute constant $C_2 > 0$.

We should note that these ranges of Lipschitz classes are the only ones of interest in these cases. In particular suppose that CI is a confidence interval with guaranteed coverage over the class of monotonically nondecreasing functions. Then for any $\beta > 1$ the class $\Lambda(\beta, M)$ includes the linear function $f_k(t) = kt$. As shown in Example 1 in Section 2.2

$$L_\alpha^*(f_k, F_m) \geq (1 - \frac{1}{\sqrt{2\pi} z_\alpha})(3k)^{\frac{1}{3}} z_\alpha^{\frac{2}{3}} n^{-\frac{1}{3}}.$$

Hence,

$$\sup_{f \in F_m(\beta, M)} E_f L(CI) \geq \sup_k L_\alpha^*(f_k, F_m) = \sup_k (1 - \frac{1}{\sqrt{2\pi} z_\alpha})(3k)^{\frac{1}{3}} z_\alpha^{\frac{2}{3}} n^{-\frac{1}{3}} = \infty. \tag{38}$$

A similar results holds for convex functions assumed to belong to $\Lambda(\beta, M)$ with $\beta > 2$. On the other hand suppose f is convex and assumed to belong to $\Lambda(\beta, M)$ with $\beta < 1$. Then from the assumption that f is in $\Lambda(\beta, M)$ it follows that $|f(1/2) - f(-1/2)| \leq M$. Convexity then shows that $f \in \Lambda(1, M)$ and the maximum expected length over this class is given above.

5.2. *Confidence interval at other points.* The focus of the present paper has been on the problem of estimating the value of $f(0)$. The basic development is similar for any other point t in the interior of the interval $[-1/2, 1/2]$ unless t is near to the boundary. More specifically for any $0 \le t < 1/2$ we can consider estimators $\delta_j^R(t) = 2^j(Y(t+2^{-j}) - Y(t))$ and $\delta_j^L(t) = 2^j(Y(t) - Y(t-2^{-j}))$ where $j \ge -\log_2(\frac{1}{4} - \frac{t}{2})$ for monotone functions and $\delta_j(t) = 2^{j-1}(Y(t+2^{-j}) - Y(t-2^{-j}))$ where $j \ge -\log_2(\frac{1}{2} - t)$ for convex functions. The basic theory is the same as before.

For monotonically nondecreasing functions, the confidence interval CI_j^m is replaced by

$$CI_j^m(t) = [\delta_j^L(t) - z_{\frac{\alpha}{2}}\sqrt{2}\sigma_j, \delta_j^R(t) + z_{\frac{\alpha}{2}}\sqrt{2}\sigma_j]$$

and the choice of $\hat{j}$ is given by

$$\hat{j}(t) = \inf_{j \ge -\log_2(\frac{1}{4}-\frac{t}{2})} \left\{ j : \xi_j(t) \le \frac{3}{2} z_\alpha \sigma_j \right\},$$

where $\xi_j(t) = 2^{j-1}(Y(t+2^{-j+1}) - Y(t+2^{-j})) - 2^{j-1}(Y(t-2^{-j}) - Y(t-2^{-j+1}))$. The final confidence interval is defined by

$$CI_*^m = CI_{\hat{j}(t)}^m. \tag{39}$$

For convex functions, the confidence interval CI_j^c is replaced by

$$CI_j^c(t) = [\delta_{j+1}(t) - (\delta_j(t) - \delta_{j+1}(t))_+ - z_{\frac{\alpha}{12}}\sqrt{5}\sigma_j,\ \delta_{j+1}(t) + z_{\frac{\alpha}{12}}\sigma_{j+1}]$$

and $\hat{j}$ is chosen to be

$$\hat{j}(t) = \inf_{j \ge -\log_2(\frac{1}{2}-t)} \{ j : T_j(t) \le z_\alpha \sigma_j \}.$$

Define the final confidence interval by

$$CI_*^c = CI_{\hat{j}(t)}^c.$$

The modulus of continuity defined in (40) is replaced by

$$\omega(\epsilon, f, t, \mathcal{F}) = \sup\{|g(t) - f(t)| :\ \|g - f\|_2 \le \epsilon,\ g \in \mathcal{F}\}. \tag{40}$$

The earlier analysis then yields

$$E_f L(CI_*^m(t)) \le c_1 L_\alpha^*(f, t, F_m),$$

and

$$E_f L(CI_*^c(t)) \leq c_2 L_\alpha^*(f, t, F_c),$$

where we now have

$$L_\alpha^*(f, t, \mathcal{F}) \geq (1 - \frac{1}{\sqrt{2\pi} z_\alpha})\omega(\frac{z_\alpha}{\sqrt{n}}, f, t, \mathcal{F}).$$

Finally we should note that at the boundary the construction of a confidence interval must be unbounded. For example any honest confidence interval for $f(1/2)$ must be of the form $[\hat{f}(1/2), \infty)$, otherwise it cannot have guaranteed coverage probability.

6. Proofs. We prove the main results in this section. We shall omit the proofs for Theorems 6 and 7 as they are analogous to those for the corresponding results in the white noise model.

6.1. *Proof of Lemma 1.* Set $f_s(t) = \frac{f(t)+f(-t)}{2} - f(0)$. Now note that $f_s(tx)$ is convex in x for all $0 \leq t \leq 1$. Hence $g(x) = \int_0^1 f_s(tx)dt$ is also convex with $g(0) = 0$. For $x > 0$ set $z = xt$ and it follows that $g(x) = \frac{1}{x}\int_0^x f_s(z)dz = \frac{1}{2x}\int_{-x}^x (f(z) - f(0))dz$. Equation (8) follows from the fact that $g(x) \leq \frac{1}{2}g(2x)$ for $x = 2^{-(j+1)}$ and equation (9) follow from the fact that $g(2x) \leq 2/3g(x) + 1/3g(4x)$. ∎

6.2. *Proof of Lemma 2.* For any convex function f, let $f_s(x) = \frac{1}{2}(f(x) + f(-x)) - f(0)$. Then $f_s(x)$ is convex, increasing in $|x|$ and $f_s(0) = 0$. Convexity of f_s yields that for $0 < x \leq y$

$$\frac{f_s(x)}{x} \leq \frac{f_s(y)}{y}. \tag{41}$$

Note that $E\delta_j = 2^{-(j-1)}\sum_{k=1}^{2^{j-1}} f_s(\frac{k}{n})$ and

$$ET_j = 2^{-(j-1)}\left\{\sum_{k=2^{j-2}+1}^{2^{j-1}} f_s(\frac{k}{n}) - \sum_{k=1}^{2^{j-2}} f_s(\frac{k}{n})\right\}.$$

So $ET_j \geq 2ET_{j-1}$ is equivalent to

$$\sum_{k=2^{j-2}+1}^{2^{j-1}} f_s(\frac{k}{n}) - \sum_{k=1}^{2^{j-2}} f_s(\frac{k}{n}) \geq 4\sum_{k=2^{j-3}+1}^{2^{j-2}} f_s(\frac{k}{n}) - 4\sum_{k=1}^{2^{j-3}} f_s(\frac{k}{n})$$

which is the same as

$$\sum_{k=2^{j-2}+1}^{2^{j-1}} f_s(\frac{k}{n}) + 3\sum_{k=1}^{2^{j-3}} f_s(\frac{k}{n}) \geq 5 \sum_{k=2^{j-3}+1}^{2^{j-2}} f_s(\frac{k}{n}). \tag{42}$$

Now note that for $x \geq 0$ and $u \geq 0$,

$$f_s(x)+f_s(x+3u) \geq f_s(x+u)+f_s(x+2u) \quad \text{and} \quad f_s(x)+f_s(x+2u) \geq 2f_s(x+u)$$

and consequently $f_s(x+3u) + f_s(x+2u) + 3f_s(x) \geq 5f_s(x+u)$. Then (42) follows by taking $u = \frac{2^{j-3}}{n}$ and $x = \frac{k}{n}$ and then summing over $k = 1, ..., 2^{j-3}$.

Denote the bias of $\bar{\delta}_j$ by $\bar{b}_j = E\bar{\delta}_j - f(0)$. Then

$$\bar{b}_j = 2^{-(j-1)} \sum_{k=1}^{2^{j-1}} f_s(\frac{k}{n}) = 2^{-(j-1)} \left\{ \sum_{k=2^{j-2}+1}^{2^{j-1}} f_s(\frac{k}{n}) + \sum_{k=1}^{2^{j-2}} f_s(\frac{k}{n}) \right\}.$$

It follows from (41) that for $k > 2^{j-2}$, $f_s(\frac{k}{n}) \geq \frac{k}{2^{j-2}} f_s(\frac{2^{j-2}}{n})$, and for $k \leq 2^{j-2}$, $f_s(\frac{k}{n}) \leq \frac{k}{2^{j-2}} f_s(\frac{2^{j-2}}{n})$. Hence

$$\begin{aligned} \sum_{k=2^{j-2}+1}^{2^{j-1}} f_s(\frac{k}{n}) &\geq \sum_{k=2^{j-2}+1}^{2^{j-1}} \frac{k}{2^{j-2}} \cdot f_s(\frac{2^{j-2}}{n}) \geq \frac{\sum_{k=2^{j-2}+1}^{2^{j-1}} \frac{k}{2^{j-2}}}{\sum_{k=1}^{2^{j-2}} \frac{k}{2^{j-2}}} \sum_{k=1}^{2^{j-2}} f_s(\frac{k}{n}) \\ &= \frac{3 \cdot 2^{j-2}+1}{2^{j-2}+1} \sum_{k=1}^{2^{j-2}} f_s(\frac{k}{n}). \end{aligned}$$

Hence,

$$\bar{b}_j \geq 2^{-(j-1)} \cdot (\frac{3 \cdot 2^{j-2}+1}{2^{j-2}+1} + 1) \sum_{k=1}^{2^{j-2}} f_s(\frac{k}{n}) = \frac{2^{j-1}+1}{2^{j-2}+1} \bar{b}_{j-1}. \quad \blacksquare$$

6.3. *Proof of Theorem 1.* Suppose that $X \sim N(\theta, \sigma^2)$ where it is known that $\theta \in [0, a\sigma]$. The confidence interval for θ which has guaranteed coverage over the interval $\theta \in [0, a\sigma]$ and which minimizes the expected length when $\theta = 0$ is given by

$$[0, \max(0, \min(X + z_\alpha \sigma, a\sigma))] \tag{43}$$

It follows that

$$L = \sigma \int_{-z_\alpha}^{a-z_\alpha} z\phi(z)dz + \sigma(z_\alpha P(-z_\alpha \leq Z \leq a - z_\alpha) + aP(Z \geq a - z_\alpha)) \tag{44}$$

and hence

$$\frac{L}{\sigma} = (\phi(z_\alpha) - \phi(a - z_\alpha)) + z_\alpha(\Phi(a - z_\alpha) - \Phi(-z_\alpha)) + a(1 - \Phi(a - z_\alpha)) \tag{45}$$

In particular when $a = z_\alpha$

$$\frac{L}{\sigma} \geq z_\alpha(1 - \frac{\phi(0)}{z_\alpha} + \frac{\phi(z_\alpha)}{z_\alpha} - \alpha) \tag{46}$$

In particular we have

$$\frac{L}{\sigma} \geq z_\alpha(1 - \frac{\phi(0)}{z_\alpha}) \tag{47}$$

Write $L^*_\alpha(f, \mathcal{F})$ for the smallest expected length at f when we have guaranteed coverage over $\mathcal{F}$. In particular let P_θ be a subfamily of $\mathcal{F}$ then $L^*_\alpha(f, \mathcal{F}) \geq L^*_\alpha(f, P_\theta)$.

Now suppose that f_0 is the "true" function.Fix $\epsilon > 0$. There is a function $f_1 \in \mathcal{F}$ such that

$$|f_1(0) - f_0(0)| = \omega(\frac{\epsilon}{\sqrt{n}}, f, \mathcal{F})$$

and such that

$$\|f_1 - f_0\|_2 = \frac{\epsilon}{\sqrt{n}}.$$

Now for $0 \leq \theta \leq 1$, let $f_\theta = f_0 + \theta(f_1 - f_0)$. Let P_θ be this collection of f_θ. Now for the process

$$dY(t) = f_\theta(t)dt + \frac{1}{\sqrt{n}}dW(t), \quad -\frac{1}{2} \leq t \leq \frac{1}{2}$$

there is a sufficient statistic S_n given by

$$S_n = f_0(0) + (f_1(0) - f_0(0))\frac{1}{\int(f_1 - f_0)^2}\int(f_1(t) - f_0(t))(dY(t) - f_0(t)dt).$$

Note that S_n has a normal distribution $S_n \sim N(f_\theta(0), \frac{(f_1(0)-f_0(0))^2}{n\int(f_1-f_0)^2})$ or more specifically $S_n \sim N(f_\theta(0), \frac{1}{\epsilon^2}\omega^2(\frac{\epsilon}{\sqrt{n}}, f_0, \mathcal{F}))$.

Note that $a = \epsilon$. Now take $\epsilon = z_\alpha$. It then follows that

$$L^*_\alpha(f_0, P_\theta) \geq \omega(\frac{z_\alpha}{\sqrt{n}}, f_0, \mathcal{F})(1 - \frac{\phi(0)}{z_\alpha} + \frac{\phi(z_\alpha)}{z_\alpha} - \alpha). \quad \blacksquare$$

6.4. *Proof of Proposition 1.* For monotone functions, we have

$$\begin{aligned}
\mathsf{P}(f(0) \in CI_j^m) &= \mathsf{P}(\delta_j^L - z_{\frac{\alpha}{2}}\sqrt{2}\sigma_j \le f(0) \le \delta_j^R + z_{\frac{\alpha}{2}}\sqrt{2}\sigma_j) \\
&\ge 1 - \mathsf{P}(\delta_j^R < f(0) - z_{\frac{\alpha}{2}}\sqrt{2}\sigma_j) - \mathsf{P}(\delta_j^L > f(0) + z_{\frac{\alpha}{2}}\sqrt{2}\sigma_j) \\
&= 1 - \mathsf{P}(Z < \frac{f(0) - E(\delta_j^R)}{\sqrt{2}\sigma_j} - z_{\frac{\alpha}{2}}) - \mathsf{P}(Z > \frac{f(0) - E(\delta_j^L)}{\sqrt{2}\sigma_j} + z_{\frac{\alpha}{2}}),
\end{aligned}$$

where Z is a standard normal random variable. Because $f(0) - E(\delta_j^R) \le 0$ and $f(0) - E(\delta_j^L) \ge 0$, we have

$$\mathsf{P}(f(0) \in CI_j^m) \ \ge \ 1 - \mathsf{P}(Z < -z_{\frac{\alpha}{2}}) - \mathsf{P}(Z > z_{\frac{\alpha}{2}}) = 1 - \alpha.$$

For convex functions, let $b_j = \text{Bias}(\delta_j)$. It follows from Lemma 1 that $b_j - 2b_{j+1} > 0$, and hence we have

$$\begin{aligned}
P(f(0) \in CI_j^c) &\ge P(2\delta_{j+1} - \delta_j - z_{\frac{\alpha}{2}}\sqrt{5}\sigma_j \le f(0) \le \delta_{j+1} + z_{\frac{\alpha}{2}}\sigma_{j+1}) \\
&\ge 1 - P(\delta_{j+1} < f(0) - z_{\frac{\alpha}{2}}\sigma_{j+1}) - P(2\delta_{j+1} - \delta_j > f(0) + z_{\frac{\alpha}{2}}\sqrt{5}\sigma_j) \\
&= 1 - P(\frac{\delta_{j+1} - E\delta_{j+1}}{\sigma_{j+1}} < -\frac{b_{j+1}}{\sigma_{j+1}} - z_{\frac{\alpha}{2}}) \\
&\quad -P(\frac{2\delta_{j+1} - \delta_j - E(2\delta_{j+1} - \delta_j)}{\sqrt{5}\sigma_j} > \frac{b_j - 2b_{j+1}}{\sqrt{5}\sigma_j} + z_{\frac{\alpha}{2}}) \\
&\ge 1 - P(Z < -z_{\frac{\alpha}{2}}) - P(Z > z_{\frac{\alpha}{2}}) \\
&= 1 - \alpha. \ \blacksquare
\end{aligned}$$

6.5. *Proof of Theorem 2.* We shall first prove that the confidence interval CI_*^m has guaranteed coverage probability of $1 - \alpha$ over F_m and then prove the upper bound for the expected length.

Note that

$$\mathsf{P}(f(0) \in CI_*^m) = \sum_{j=2}^{\infty} \mathsf{P}(f(0) \in CI_j^m | \hat{j} = j)\mathsf{P}(\hat{j} = j).$$

Because both δ_j^R and δ_j^L are independent of ξ_k for $k \le j$ and the event $\{\hat{j} = j\}$ depends only on ξ_k for $k \le j$, then by Proposition 1 we have

$$\mathsf{P}(f(0) \in CI_*^m) = \sum_{j=2}^{\infty} \mathsf{P}(f(0) \in CI_j^m)\mathsf{P}(\hat{j} = j) \ge \sum_{j=2}^{\infty}(1-\alpha)\mathsf{P}(\hat{j} = j) = 1-\alpha.$$

We now turn to the upper bound for the expected length. Note that for $s \ge 0$, $E\xi_{j_*^m+s} \le z_\alpha \sigma_{j_*^m} = \frac{1}{2^{\frac{s}{2}}} z_\alpha \sigma_{j_*^m+s}$, and so we have

$$\mathsf{P}(\hat{j} \ge j_*^m + k) \le \prod_{s=0}^{k-1} \mathsf{P}(\xi_{j_*^m+s} > \frac{3}{2} z_\alpha \sigma_{j_*^m+s}) \le \prod_{s=0}^{k-1} \mathsf{P}(Z > z_\alpha(\frac{3}{2} - \frac{1}{2^{\frac{s}{2}}})).$$

It follows from $E(\delta_j^R - \delta_j^L) \le 2E\xi_j$ that $E(\delta_{\hat{j}}^R - \delta_{\hat{j}}^L) \le 2E\xi_{\hat{j}}$, and hence we have

$$\begin{aligned} E_f L(CI^*) &= E_f(\delta_{\hat{j}}^R - \delta_{\hat{j}}^L + 2\sqrt{2}z_{\frac{\alpha}{2}}\sigma_{\hat{j}}) \le E_f(2\xi_{\hat{j}} + 2\sqrt{2}z_{\frac{\alpha}{2}}\sigma_{\hat{j}}) \\ &\le E_f((3z_\alpha + 2\sqrt{2}z_{\frac{\alpha}{2}})\sigma_{\hat{j}}) = \sum_{j=2}^{\infty}(3z_\alpha + 2\sqrt{2}z_{\frac{\alpha}{2}})\sigma_j \cdot \mathsf{P}(\hat{j} = j). \end{aligned}$$

Thus,

$$E_f L(CI^*) \le (3z_\alpha + 2\sqrt{2}z_{\frac{\alpha}{2}})\sigma_{j_*^m}\Big(\mathsf{P}(\hat{j} \le j_*^m) + \sum_{k=1}^{\infty} 2^{k/2}\mathsf{P}(\hat{j} = j_*^m + k)\Big). \tag{48}$$

Set $w_k = 2^{k/2} - 2^{(k-1)/2}$ for $k \ge 1$. Then it is easy to see that

$$S = \mathsf{P}(\hat{j} \le j_*^m) + \sum_{k=1}^{\infty} 2^{k/2}\mathsf{P}(\hat{j} = j_*^m + k) = 1 + \sum_{k=1}^{\infty} w_k \mathsf{P}(\hat{j} \ge j_*^m + k).$$

Thus,

$$S = 1 + \sum_{k=1}^{\infty} w_k \prod_{s=0}^{k-1} \mathsf{P}(Z > z_\alpha(\frac{3}{2} - \frac{1}{2^{\frac{s}{2}}})).$$

The right hand side is increasing in α. Through numerical calculations, we can see that, for $\alpha = 0.2$,

$$\sum_{k=1}^{\infty} w_k \prod_{s=0}^{k-1} \mathsf{P}(Z > z_\alpha(\frac{3}{2} - \frac{1}{2^{\frac{s}{2}}})) \le 0.21.$$

Thus, by equation (48), we have

$$E_f L(CI^*) \le 1.21(3z_\alpha + 2\sqrt{2}z_{\frac{\alpha}{2}})\sigma_{j_*^m}. \quad \blacksquare$$

6.6. *Proof of Theorem 3.* Note that if $j_*^m > 2$, then $E\xi_{j_*^m - 1} \ge z_\alpha \sigma_{j_*^m - 1} = \frac{1}{\sqrt{2}} z_\alpha \sigma_{j_*^m}$ and hence there is a $t_* \le 2^{-j_*^m + 2}$ such that we have either $f(t_*) - f(0) \ge \frac{1}{\sqrt{2}} z_\alpha \sigma_{j_*^m}$ or $f(0) - f(-t_*) \ge \frac{1}{\sqrt{2}} z_\alpha \sigma_{j_*^m}$. If $f(t_*) \ge \frac{1}{\sqrt{2}} z_\alpha \sigma_{j_*^m} + f(0)$, let

$$g(t) = \begin{cases} \max\{\frac{1}{\sqrt{2}} z_\alpha \sigma_{j_*^m} + f(0), f(t)\}, & \text{if } t \ge 0; \\ f(t), & \text{otherwise,} \end{cases}$$

and if $f(-t_*) \le -\frac{1}{\sqrt{2}} z_\alpha \sigma_{j_*^m} + f(0)$, let

$$g(t) = \begin{cases} \min\{-\frac{1}{\sqrt{2}} z_\alpha \sigma_{j_*^m} + f(0), f(t)\}, & \text{if } t \le 0; \\ f(t), & \text{otherwise.} \end{cases}$$

Then we have

$$\int_{-1/2}^{1/2}(f(t)-g(t))^2dt \le \frac{1}{2}z_\alpha^2\frac{2^{j_*^m-1}}{n}\cdot 2^{-j_*^m+2} = \frac{z_\alpha^2}{n}.$$

If $j_*^m = 2$, let

$$g(t) = \begin{cases} \max\{\frac{1}{\sqrt{2}}z_\alpha\sigma_{j_*^m} + f(0), f(t)\}, & \text{if } t \ge 0; \\ f(t), & \text{otherwise,} \end{cases}$$

then we have

$$\int_{-1/2}^{1/2}(f(t)-g(t))^2dt \le \frac{1}{2}z_\alpha^2\frac{2^{j_*^m-1}}{n}\cdot\frac{1}{2} \le \frac{z_\alpha^2}{n}.$$

It then follows that

$$\omega(\frac{z_\alpha}{\sqrt{n}}, f, F_m) \ge \frac{1}{\sqrt{2}}z_\alpha\sigma_{j_*^m},$$

and so

$$L_\alpha^*(f, F_m) \ge (1-\frac{1}{\sqrt{2\pi}z_\alpha})\frac{1}{\sqrt{2}}z_\alpha\sigma_{j_*^m}. \quad \blacksquare$$

6.7. *Proof of Theorem 4.* We shall first prove that the confidence interval CI_*^c has guaranteed coverage probability of $1-\alpha$ over F_c and then prove the upper bound for the expected length.

Note that if $j_*^c > 1$, then $ET_{j_*^c-1} \ge \frac{2}{3}z_\alpha\sigma_{j_*^c-1} = \frac{\sqrt{2}}{3}z_\alpha\sigma_{j_*^c}$. It follows that for $k \ge 1$, $ET_{j_*^c-k} \ge 2^{k-1/2}\frac{1}{3}z_\alpha\sigma_{j_*^c} = 2^{(3k-1)/2}\frac{1}{3}z_\alpha\sigma_{j_*^c-k}$. Hence

$$P(\hat{j} = j_*^c - k) \le P(T_{j_*^c-k} \le z_\alpha\sigma_{j_*^c-k}) \le P(Z \ge (\frac{2^{(3k-1)/2}}{3}-1)z_\alpha) \tag{49}$$

Also for $m \ge 0$, $ET_{j_*^c+m} \le 2^{-m}\cdot\frac{2}{3}z_\alpha\sigma_{j_*^c} = 2^{-3m/2}\cdot\frac{2}{3}z_\alpha\sigma_{j_*^c+m}$ and hence

$$P(\hat{j} \ge j_*^c + k) \le \prod_{m=0}^{k-1}P(T_{j_*^c+m} > z_\alpha\sigma_{j_*^c+m}) \le \prod_{m=0}^{k-1}P(Z > z_\alpha(1-\frac{2}{3}2^{-3m/2})). \tag{50}$$

To bound the coverage probability note that

$$P(f(0) \notin CI_*^c) \le \sum_{m\ge 3}P(\hat{j} = j_*^c - m) + P(\hat{j} \ge j_*^c + 3) + \sum_{k=-2}^{2}P(f(0) \notin CI_{j_*^c+k}). \tag{51}$$

It then follows from Equation (49) that

$$P(\hat{j} = j_*^c - 3) \le P(Z \ge \frac{13}{3}z_\alpha) \le \frac{7\alpha}{10000}$$

for all $0 < \alpha \le 0.2$. It is easy to verify directly that for all $z \ge 1$, $P(Z \ge 2z) \le (1/6)P(Z \ge z)$. Furthermore, it is easy to see that for $k \ge 1$, $\frac{2^{(3(k+3)-1)/2}}{3} - 1 \ge 2^k \frac{13}{3}$ and so

$$\begin{aligned} P(\hat{j} = j_*^c - 3 - k) &\le P(Z \ge (\frac{2^{(3(k+3)-1)/2}}{3} - 1)z_\alpha) \le P(Z \ge 2^k \frac{13}{3} z_\alpha) \\ &\le 6^{-k} P(Z \ge \frac{13}{3} z_\alpha) \le 6^{-k} \frac{7\alpha}{10000}. \end{aligned}$$

Hence,

$$\sum_{m \ge 3} P(\hat{j} = j_*^c - m) = \sum_{k \ge 0} P(\hat{j} = j_*^c - 3 - k) \le \frac{7\alpha}{10000} \sum_{k \ge 0} 6^{-k} \le \frac{7\alpha}{5000}.$$

Note that (50) yields that

$$P(\hat{j} \ge j_*^c + 3) \le P(Z \ge \frac{1}{3} z_\alpha) \cdot P(Z \ge (1 - \frac{1}{3\sqrt{2}}) z_\alpha) \cdot P(Z \ge \frac{11}{12} z_\alpha) \le \frac{\alpha}{6.4}$$

for all $0 < \alpha \le 0.3$. It now follows from (51) that

$$P(f(0) \in CI_*^c) = 1 - P(f(0) \notin CI_*^c) \ge 1 - (\frac{7\alpha}{5000} + \frac{\alpha}{6.4} + 5 \times \frac{\alpha}{6}) \ge 1 - \alpha. \quad \blacksquare$$

We now turn to the upper bound for the expected length. Note that

$$E_f L(CI_*^c) \le \sum_{j=1}^{\infty} (z_\alpha + (\sqrt{5} + \sqrt{2}) z_{\frac{\alpha}{12}}) \sigma_j \cdot P(\hat{j} = j) \tag{52}$$

Hence

$$E_f L(CI_*^c) \le (z_\alpha + (\sqrt{5} + \sqrt{2}) z_{\frac{\alpha}{12}}) \sigma_{j_*^c} \left(P(\hat{j} \le j_*^c) + \sum_{k=1}^{\infty} 2^{k/2} P(\hat{j} = j_*^c + k) \right) \tag{53}$$

Set $w_k = 2^{k/2} - 2^{(k-1)/2}$ for $k \ge 1$. Then it is easy to see that

$$S = P(\hat{j} \le j_*^c) + \sum_{k=1}^{\infty} 2^{k/2} P(\hat{j} = j_*^c + k) = 1 + \sum_{k=1}^{\infty} w_k P(\hat{j} \ge j_*^c + k).$$

It then follows from (50) that

$$S \le 1 + \sum_{k=1}^{\infty} w_k \prod_{m=0}^{k-1} P(Z > z_\alpha (1 - \frac{2}{3} \frac{1}{2^{3m/2}}))$$

The right hand side is clearly increasing in α. Direct numerical calculations show that for $\alpha = 0.2$

$$\sum_{k=1}^{\infty} w_k \prod_{m=0}^{k-1} P(Z > z_\alpha(1 - \frac{2}{3}\frac{1}{2^{3m/2}})) \leq 0.25.$$

It then follows directly from (53) that

$$EL(CI_*^c) \leq 1.25(z_\alpha + (\sqrt{5} + \sqrt{2})z_{\frac{\alpha}{12}})\sigma_{j_*^c}. \quad \blacksquare$$

6.8. *Proof of Theorem 5.* Note that if $j_*^c > 1$. then $ET_{j_*^c-1} \geq \frac{2}{3}z_\alpha\sigma_{j_*^c-1} = \frac{\sqrt{2}}{3}z_\alpha\sigma_{j_*^c}$ and hence there is a t_* satisfying $0 < t_* \leq 2^{-j_*^c+1}$ such that $f_s(t_*) = \frac{\sqrt{2}}{3}z_\alpha\sigma_{j_*^c}$, where $f_s(t) = \frac{f(t)+f(-t)}{2} - f(0)$. Let g be defined by

$$g(t) = f(t)1(|t| > t_*) + (f_s(t_*) + \frac{f(t_*) - f(-t_*)}{2t_*}t)1(|t| \leq t_*)$$

There is also a g as in the proof of lemma 5 in our other paper with $g(0) = f_s(t_*)$ for which

$$\int_{-1/2}^{1/2}(g(t) - f(t))^2 dt \leq \frac{9}{4}f_s^2(t_*)t_* \leq \frac{z_\alpha^2}{n}.$$

If $j_*^c = 1$, then let $g(t) = f(t) + \frac{\sqrt{2}}{3}z_\alpha\sigma_{j_*^c}$, then we have

$$\int_{-1/2}^{1/2}(g(t) - f(t))^2 dt \leq \frac{2}{9}z_\alpha^2\sigma_1^2 \leq \frac{z_\alpha^2}{n}.$$

It then follows that

$$\omega(\frac{z_\alpha}{\sqrt{n}}, f, F_c) \geq \frac{\sqrt{2}}{3}z_\alpha\sigma_{j_*^c},$$

and so

$$L_\alpha^*(f, F_c) \geq (1 - \frac{1}{\sqrt{2\pi}z_\alpha})\frac{\sqrt{2}}{3}z_\alpha\sigma_{j_*^c}. \quad \blacksquare$$

REFERENCES

[1] Cai, T. T. and Low, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32**, 1805-1840.

[2] Cai, T. T. and Low, M. G. (2011). A framework for estimation of convex functions. Technical report.

[3] Donoho, D. L. (1994). Statistical estimation and optimal recovery. *Ann. Statist.* **22**, 238-270.

[4] Donoho, D. L. and Liu, R. G. (1991). Geometrizing rates of convergence III. *Ann. Statist.* **19**, 668-701.

[5] Dümbgen, L. (1998). New goodness-of-fit tests and their applications to nonparametric confidence sets. *Ann. Statist.* **26**, 288-314.

[6] Dümbgen, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* 9(3), 423-449.

[7] Evans, S.A., Hansen, B.B. and Stark, P.B. (2005). Minimax expected measure confidence sets for restricted location parameters. *Bernoulli* **11**, 571-590.

[8] Hall, P., Kay, J. and Titterington, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77**, 521-528.

[9] Hengartner, N. W. and Stark, P. B. (1995). Finite-sample confidence envelopes for shape-restricted densities. *Ann. Statist.* **23**, 525-550.

[10] Lepski, O. V., Mammen, E. and Spokoiny V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25**, 929-947.

[11] Lepski, O. V. and Spokoiny V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25**, 2512-2546.

[12] Low, M. G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25**, 2547-2554.

[13] Munk, A., Bissantz, N., Wagner, T. and Freitag, G. (2005). On difference based variance estimation in nonparametric regression when the covariate is high dimensional. *J. Roy. Statist. Soc. B* **67**, 19-41.

[14] Wang, L., Brown, L. D., Cai, T. and Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *Ann. Statist.* **36**, 646-664.

Department of Statistics
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104.
E-mail: tcai@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/ tcai
E-mail: lowm@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/ lowm
E-mail: xiayin@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/ yin