The Annals of Statistics
1993, Vol. 21, No. 2, 600–610

ADAPTIVELY LOCAL ONE-DIMENSIONAL SUBPROBLEMS WITH APPLICATION TO A DECONVOLUTION PROBLEM

BY JIANQING FAN

University of North Carolina

In this paper, a method for finding global minimax lower bounds is introduced. The idea is to adjust automatically the direction of a local one-dimensional subproblem at each location to the nearly hardest one, and to use locally the difficulty of the one-dimensional subproblem. This method has the advantages of being easily implemented and understood. The lower bound is then applied to nonparametric deconvolution to obtain the optimal rates of convergence for estimating a whole function. Other applications are also addressed.

1. Introduction. Nonparametric techniques provide a useful tool for investigating the structure of some interesting functions. A mathematical formulation is to think of estimating some function $T \circ f(x)$ (e.g., density or regression function) based on a random sample $X_1, \dots, X_n$ from a density f with a (smoothness) constraint $f \in \mathscr{F}$ under some global losses. The global loss functions are typically those induced by L_p-norm:

$$L(d, T \circ f) = \left(\int_a^b |T \circ f(x) - d(x)|^p w(x)\, dx \right)^{1/p},$$

where $w(x)$ is a weight function and $d(x)$ is a decision function estimating $T \circ f(x)$.

How can one measure the difficulty of estimating the function $T \circ f$ under the weighted L_p-loss? How to find a global minimax lower bound? The popular approach is that:

1. Specify a subproblem—estimating $T \circ f(x)$ on a specified subset $\mathscr{F}_n$ of the parameter space $\mathscr{F}$; the geometry of $\mathscr{F}_n$ is typically hyperrectangular.
2. Use the difficulty of the subproblem as a lower bound for the minimax risk of the full nonparametric problem.

In the second step, we first formulate problems of estimating a *functional* $T \circ f(x_0)$ at each location x_0, then adjust automatically the direction at the location x_0 to the nearly most difficult direction for estimating the *functional* $T \circ f(x_0)$, and finally add the difficulties of one-dimensional subproblems at all locations, according to their weights, to obtain a lower bound. See Section 2 for details.

Received November 1989; revised June 1992.

AMS 1991 *subject classifications*. Primary 62G07; secondary 62C20, 62G20.

Key words and phrases. Cubical lower bound, one-dimensional subproblems, global rates of convergence, minimax integrated risks, deconvolution.

The above approach is related to the illuminating ideas of Donoho and Liu (1987, 1991a, b) and Stone's two point testing argument [see Stone (1980)] as well as other approaches [Farrell (1972) and Has'minskii (1979)] for estimating a statistical *functional*. In that setting, Donoho and Liu (1987) shows that the difficulty of the hardest one-dimensional subproblem is hard enough to capture the difficulty of a full nonparametric problem. However, the hardest one-dimensional subproblem is *not* difficult enough for estimating a whole function (e.g., the whole density function). To bridge this gap, we use a growing number of dimensional subproblems, adjusting directions accordingly, to capture the difficulty of estimating a whole function. The current approach is inspired by the ideas of Has'minskii (1978), Bretagnolle and Huber (1979) and Stone (1982), who used a growing number of subproblems to construct lower bounds, and by the hardest hyperrectangular approach of Donoho, Liu and MacGibbon (1990), which finds good bounds on global minimax risks (rates and constants) for Gaussian white noise models.

In comparison with the existing methods of Stone (1982), Has'minskii (1978) and Birgé (1987), our approach is simpler in the second step above. An advantage of our approach is that it is easy to use: Finding a lower bound of minimax integrated risks can typically be reduced to that of a pointwise problem. Thus, obtaining a lower rate for estimating a function is as easy as that for estimating the function at a point. See Section 3 for further discussions. For related literature, see Efroimovich and Pinsker (1983), Nussbaum (1985), Donoho, Liu and MacGibbon (1990), Johnstone and Silverman (1990) and Low (1993).

The cubical method is especially useful for establishing optimal global rates for deconvolution problems. We use this to solve an open problem of the optimal global rates for nonparametric deconvolution. In this context, the cubical method provides a precise description of the difficulty, depending on the tail of characteristic functions of error distributions, of deconvolution. The usefulness of this cubical method is further illustrated by considering problems in density estimation, nonparametric regression, errors-in-variables regression [Fan and Truong (1993)] and positron emission tomography [Johnstone and Silverman (1990) and Zhang (1992)].

The paper is organized as follows. Section 2 introduces the cubical lower bound, which is applied to deconvolution, density estimation and nonparametric regression in Section 3. Technical arguments are deferred to Section 4.

2. Cubical lower bound of global rates. In this section, we give a lower bound for estimating a function $T \circ f(x)$. We discuss the problem for the one-dimensional case. Higher dimensional results follow naturally.

Let $[a, b]$ be an interval on the line and $x_{n,j} = a + j(b-a)/m_n$, $j = 1, \dots, m_n$. Denote $\boldsymbol{\theta}_{m_n} = (\theta_1, \dots, \theta_{m_n})$ and

$$f_{\boldsymbol{\theta}_{m_n}}(t) = f_0(t) + a_n^{-1} \sum_{j=1}^{m_n} \theta_j H\big(m_n(t - x_{n,j})\big), \tag{2.1}$$

where $f_0(t)$ is a density function, $H(\cdot)$ is a bounded function whose integral on the line is 0, and m_n and a_n are sequences tending to infinity. By suitable choice of m_n, a_n, f_0 and H, the function $f_{\boldsymbol{\theta}_{m_n}}$ will be a density function. Denote a class of densities by

$$\mathscr{F}_n = \left\{ f_{\boldsymbol{\theta}_{m_n}} : \boldsymbol{\theta}_{m_n} \text{ is a sequence of 0's and 1's} \right\} \tag{2.2}$$

and let

$$\begin{aligned} \boldsymbol{\theta}_{j_0} &= \left(\theta_1, \dots, \theta_{j-1}, 0, \theta_{j+1}, \dots, \theta_{m_n}\right), \\ \boldsymbol{\theta}_{j_1} &= \left(\theta_1, \dots, \theta_{j-1}, 1, \theta_{j+1}, \dots, \theta_{m_n}\right). \end{aligned} \tag{2.3}$$

Suppose that a_n and m_n are chosen such that we cannot distinguish the pair of densities $f_{\boldsymbol{\theta}_{j0}}$ and $f_{\boldsymbol{\theta}_{j1}}$:

$$\max_{1 \le j \le m_n} \max_{\boldsymbol{\theta}_{m_n} \in \{0,1\}^{m_n}} \chi^2\left(f_{\boldsymbol{\theta}_{j0}}, f_{\boldsymbol{\theta}_{j1}}\right) \le c/n, \qquad c > 0, \tag{2.4}$$

where $\chi^2(f, g) = \int (f - g)^2/f$. Then, we have the following lower bound for estimating $T \circ f(x)$.

THEOREM 1. *Suppose that $T \circ f(x)$ satisfies*

$$\left| T \circ f_{\boldsymbol{\theta}_{j0}}(x) - T \circ f_{\boldsymbol{\theta}_{j1}}(x) \right| \ge \left| B_n\left(m_n(x - x_{n,j})\right) \right|, \tag{2.5}$$

for some function B_n, depending on n and H. Let $w(x)$ be a nonnegative and continuous function on $[a, b]$. If condition (2.4) *holds, then for any estimator $\hat{T}_n(x)$ based on n i.i.d. observations from the unknown density f, we have*

$$\begin{aligned} &\inf_{\hat{T}_n(x)} \sup_{f \in \mathscr{F}_n} E_f \int_a^b \left| \hat{T}_n(x) - T \circ f(x) \right|^p w(x)\, dx \\ &\quad \ge \frac{1 - \sqrt{1 - e^{-c}}}{2^{p+1}(b-a)} \int_a^b w(x)\, dx \int_0^{b-a} \left| B_n(x) \right|^p dx (1 + o(1)). \end{aligned} \tag{2.6}$$

In particular when $T \circ f = f^{(k)}(x)$, the kth derivative of the unknown density function, condition (2.5) is satisfied with

$$|B_n(x)| = \left| a_n^{-1} m_n^k H^{(k)}(x) \right|.$$

We have the following lower bound for estimating $f^{(k)}(x)$.

COROLLARY 1. *Under condition* (2.4),

$$\begin{aligned} &\inf_{\hat{T}_n(x)} \sup_{f \in \mathscr{F}_n} E_f \int_a^b \left| \hat{T}_n(x) - f^{(k)}(x) \right|^p w(x)\, dx \\ &\quad \ge \frac{1 - \sqrt{1 - e^{-c}}}{2^{p+1}(b-a)} \int_a^b w(x)\, dx \int_0^{b-a} \left| H^{(k)}(x) \right|^p dx \left(\frac{m_n^k}{a_n} \right)^p (1 + o(1)). \end{aligned}$$

The key to the applications of Theorem 1 is the choice of $\mathscr{F}_n$. The choice (2.2) is especially suitable for finding lower rates of the Lipschitz class

$$\mathscr{F}_{l,\alpha,B} = \{f : |f^{(l)}(x) - f^{(l)}(y)| \le B|x-y|^\alpha\}, \tag{2.7}$$

for some $0 \le \alpha < 1$. This will be illustrated by the examples in Section 3.

REMARK 2.1. The cubical method attempts only to find sharp lower bound in terms of rates of convergence. A similar idea has independently been used by Low (1993) to find a constant factor for a particular Lipschitz constraint under a global L_2-loss. Note that when $H(\cdot)$ has a bounded support on $[0, b-a]$, $f_{\theta_{m_n}}(x_{n,j}) = f_0(x_{n,j})$—a known constant. Thus in general, this family cannot be a least favorable subfamily. As a consequence, this approach may not yield a good lower bound for the constant factor which multiplies the rate.

REMARK 2.2. Condition (2.4) states essentially that any two vertices of a hypercube cannot be tested consistently. The χ^2-distance in (2.4) can be replaced by the Hellinger distance, and Theorem 1 holds with $(1-\sqrt{1-e^{-c}})$ replaced by $(1-\sqrt{1-e^{-2c}})$ [see pages 46–47 and 475–477 of Le Cam (1985)].

REMARK 2.3. With $a_n = m_n^{l+\alpha}$ and an appropriate choice of functions H and f_0, the class $\mathscr{F}_n$ defined by (2.2) will be a subset of smoothness constraint (2.7). Thus (2.6) is also a minimax lower bound for $\mathscr{F}_{l,\alpha,B}$.

It is not hard to obtain a lower bound for the minimax risk

$$\inf_{\hat{T}_n(x)} \sup_{f\in\mathscr{F}_n} E\left(\int_a^b |\hat{T}_n(x) - T\circ f(x)|^p w(x)\,dx\right)^{1/p}. \tag{2.8}$$

By renormalization of $w(x)$, without loss of generality, assume that the total weight of $w(x)$ on $[a,b]$ is 1. Thus using the fact that

$$\left(\int_a^b |\hat{T}_n(x) - T\circ f(x)|^p w(x)\,dx\right)^{1/p} \ge \int_a^b |\hat{T}_n(x) - T\circ f(x)| w(x)\,dx, \qquad p\ge 1,$$

a lower bound for (2.8) can be obtained via (2.6) with $p = 1$.

3. Applications.

In this section, we apply the lower bound to nonparametric deconvolution problems. We also illustrate that the global problem can easily be reduced to a local problem via condition (2.4). Thus, knowledge of the local problem can be used. Other applications of this cubical method are also discussed.

3.1. *Deconvolution.* Deconvolution arises when direct observation is not possible. The basic model is as follows. We wish to estimate the unknown density of a random variable X, but the only data available are observations $Y_1, \dots, Y_n$, which are contaminated with independent additive error ε, from the model $Y = X + \varepsilon$. In density function terms, we wish to estimate f_X using data $Y_1, \dots, Y_n$ from the density

$$f_Y(y) = \int f_X(y - x)\, dF_\varepsilon(x), \tag{3.1}$$

where F_ε is the known cumulative distribution function of ε.

Now let us use the cubical bound to find optimal global rates under the smoothness constraint (2.7).

Take $f_0(x) = C_r(1 + x^2)^{-r}$, $r > 0.5$, and a $(k+1)$-time bounded differentiable function $H(x)$ and $a_n = m_n^{l+\alpha}$. With appropriate choice of r and $H(\cdot)$, $\mathscr{F}_n \subset \mathscr{F}_{l,\alpha,B}$, where $\mathscr{F}_n$ is defined by (2.2). Since data are observed from the density f_Y given by (3.1), it follows from Corollary 1 that

$$\max_{1 \le j \le m_n} \max_{\boldsymbol{\theta}_{m_n} \in \{0,1\}^{m_n}} \chi^2\left(f_{\boldsymbol{\theta}_{j0}} * F_\varepsilon, f_{\boldsymbol{\theta}_{j1}} * F_\varepsilon\right) \le c/n \tag{3.2}$$

entails

$$\liminf_{n \to \infty} \inf_{\hat{T}_n(x)} \sup_{f_X \in \mathscr{F}_{l,\alpha,B}} m_n^{p(l+\alpha-k)} E_{f_X} \int_a^b \left|\hat{T}_n(x) - f_X^{(k)}(x)\right|^p w(x)\, dx > 0. \tag{3.3}$$

Thus, it remains to determine m_n from (3.2).

We remark that when n is large enough

$$f_{\boldsymbol{\theta}_{j0}}(\cdot) \ge \tfrac{1}{2} f_0(\cdot), \qquad \forall\, \boldsymbol{\theta}_{m_n} \in \{0,1\}^{m_n},$$

and that there exists a constant C such that

$$f_0(\cdot) \ge C \max_{x_{n,j} \in [a,b]} f_0(\cdot - x_{n,j}).$$

Combining the last two displays, we have

$$f_{\boldsymbol{\theta}_{j0}} * F_\varepsilon \ge \tfrac{1}{2} f_0 * F_\varepsilon \ge \frac{C}{2} f_0(\cdot - x_{n,j}) * F_\varepsilon.$$

This together with a change of variable lead to

$$\chi^2\left(f_{\boldsymbol{\theta}_{j0}} * F_\varepsilon, f_{\boldsymbol{\theta}_{j1}} * F_\varepsilon\right) \le \frac{2}{C} m_n^{-2(l+\alpha)} \int_{-\infty}^{+\infty} \frac{\left(H(m_n x) * F_\varepsilon(x)\right)^2}{(f_0 * F_\varepsilon)}. \tag{3.4}$$

Therefore, we need only to determine m_n so that (3.4) is of order $O(1/n)$.

If we use a similar argument for estimating $f_X(0)$, we also end up with the same problem: Finding m_n from

$$\chi^2(f_0 * F_\varepsilon, f_1 * F_\varepsilon) = m_n^{-2(l+\alpha)} \int_{-\infty}^{+\infty} \frac{\left(H(m_n x) * F_\varepsilon(x)\right)^2}{(f_0 * F_\varepsilon)} = O\left(\frac{1}{n}\right), \tag{3.5}$$

where $f_1(x) = f_0(x) + a_n^{-(l+\alpha)} H(m_n x)$. Thus, we have reduced the global problem to a local problem. Using the local result from Fan (1991a), the solution to

(3.5) is given by

$$m_n = \begin{cases} c_1 n^{1/(2(l+\alpha+\beta)+1)}, & \text{for ordinary smooth (3.6)}, \\ \gamma^{1/\beta}(\log n + c_2 \log\log n)^{1/\beta}, & \text{for super smooth (3.8)}, \end{cases}$$

where c_1 and c_2 are constants. Hence condition (3.2) holds, and (3.3) leads to:

THEOREM 2. (a) *Suppose that the characteristic function ϕ_ε of ε satisfies*

$$|t^{\beta+j}\phi_\varepsilon^{(j)}(t)| \le d_j, \quad \text{as } t \to \infty, \text{ for } j = 0,1,2, \tag{3.6}$$

where d_j is a nonnegative constant and $\phi_\varepsilon^{(j)}$ is the jth derivative of ϕ_ε. Then for any $0 \le p < \infty$,

$$\liminf_{n\to\infty} \inf_{\hat{T}_n(x)} \sup_{f_X \in \mathscr{F}_{l,\alpha,B}} n^{(p(l+\alpha-k))/(2(l+\alpha+\beta)+1)} \times E_{f_X}\int_a^b |\hat{T}_n(x) - f_X^{(k)}(x)|^p w(x)\,dx > 0. \tag{3.7}$$

(b) *Assume that*

$$\limsup_{|t|\to\infty} |\phi_\varepsilon(t)|\,|t|^{-\beta_1}\exp(|t|^\beta/\gamma) < \infty \quad \textit{with constants } \beta, \gamma > 0 \textit{ and } \beta_1, \tag{3.8}$$

and that $P\{|\varepsilon - u| \le |u|^{\alpha_0}\} = O(|u|^{-(a-\alpha_0)})$, as $|u| \to \infty$, for some $0 < \alpha_0 < 1$ and $a > 1 + \alpha_0$. Then

$$\liminf_{n\to\infty} \inf_{\hat{T}_n(x)} \sup_{f_X \in \mathscr{F}_{l,\alpha,B}} (\log n)^{p(l+\alpha-k)/\beta} \times E_{f_X}\int_a^b |\hat{T}_n(x) - f_X^{(k)}(x)|^p w(x)\,dx > 0. \tag{3.9}$$

The error distributions satisfying (3.6) include gamma and symmetric gamma, and those satisfying (3.8) include normal, Cauchy and their mixture distributions. Result (3.7) answers an open question of Zhang (1990) on the lower rate for the ordinary smooth case. A particular result of (3.9), $l = 2$, $\alpha = k = 0$, for normal and Cauchy error distribution was obtained by Zhang (1990).

REMARK 3.1. Combining with upper bound results [Fan (1991b)], we have demonstrated that the global rates in Theorem 2 are the best attainable ones for the ordinary smooth error distributions under L_p-norm, $1 \le p < \infty$. Specifically, for estimating $f_X^{(k)}(x)$ under the constraint $f \in \mathscr{F}_{l,\alpha,B}$, we have the optimal global rates of convergence $(q = l + \alpha)$ given in Table 1. In particular, the optimal global rate for estimating $f_X^{(k)}(x)$ is $O(n^{-(q-k)/(2q+5)})$ when error distribution is double exponential.

REMARK 3.2. For deconvolution with a supersmooth distribution (3.8), the difficulty of estimating a whole density function can actually be captured by a

TABLE 1
Optimal global rates of convergence

	Error distributions		
	$\varepsilon \sim$ Gamma (β)	$\varepsilon \sim$ Symmetric gamma (β) $\beta \neq 2j+1$ (j integer)	$\varepsilon \sim$ Symmetric gamma (β) $\beta = 2j+1$ (j integer)
Optimal global rates	$O(n^{-(q-k)/[2(q+\beta)+1]})$	$O(n^{-(q-k)/[2(q+\beta)+1]})$	$O(n^{-(q-k)/[2(q+\beta)+3]})$

one-dimensional subproblem [Fan (1989) and Zhang (1990)]. Using the cubical method, without much effort, we easily obtain the same result by reducing the global problem to a pointwise estimation problem.

3.2. *Density estimation and nonparametric regression.* In this section, we illustrate the usefulness of the lower bound by considering problems in density estimation and nonparametric regression. With a few lines of arguments, we obtain a sharp rate for the global minimax risk.

For the density estimation problem, we wish to estimate $f_X(\cdot) \in \mathscr{F}_{l,\alpha,B}$ from its sample of size n. To find a lower bound, let $a_n = m_n^{l+\alpha}$. Choose a density function $f_0(x)$ and a function $H(x)$ having a bounded support on $[0, b-a]$ such that $\mathscr{F}_n \subset \mathscr{F}_{l,\alpha,B}$. Then, it is easy to see that if $m_n = cn^{1/(2(l+\alpha)+1)}$, $c>0$,

$$\chi^2\left(f_{\theta_{J0}}, f_{\theta_{J1}}\right) \le D^{-1} m_n^{-2(l+\alpha)-1} \int_a^b |H(x-a)|^2\,dx = O(1/n),$$

where $D = \min_{a \le x \le b} f_0(x) - \max_{0 \le x \le b-a} |H(x)|/a_n$. Thus, condition (2.4) is satisfied and Corollary 1 assures

$$\liminf_{n\to\infty} \inf_{\hat{T}_n(x)} \sup_{f \in \mathscr{F}_{l,\alpha,B}} n^{p(l+\alpha-k)/(2(l+\alpha)+1)} \times E_f \int_a^b \left|\hat{T}_n(x) - f^{(k)}(x)\right|^p w(x)\,dx > 0. \tag{3.10}$$

For nonparametric regression, we proceed as follows. We wish to estimate the conditional mean $m(\cdot) \equiv T \circ f(\cdot) = E(Y \mid X = \cdot)$ based on a bivariate random sample from $f(\cdot, \cdot)$. Assume that

$$f(\cdot,\cdot) \in \mathscr{F}_{l,\alpha,B}^2 \equiv \Big\{ f(\cdot,\cdot)\colon \left|m^{(l)}(x) - m^{(l)}(y)\right| \le B\,|x-y|^\alpha, \quad \operatorname{var}(Y \mid X = x) \le B,\ \min_{x\in[a,b]} f_X(x) \ge 1/B \Big\}.$$

Without loss of generality, assume $[a,b] = [0,1]$. Let f_0 and g_0 be symmetric densities, h_0 be a function satisfying $\int x^j h_0(y)\,dy = j$ for $j = 0, 1$ and $H(\cdot)$ has $l+1$ continuous derivatives having support $[0,1]$. Define a bivariate subfamily of $\mathscr{F}_{l,\alpha,B}^2$ by

$$\mathscr{F}_n = \left\{ f_{\theta_{m_n}}(x,y) = f_0(x) g_0(y) + m_{\theta_{m_n}}(x) h_0(y), \theta_{m_n} \in \{0,1\}^{m_n} \right\},$$

where $m_{\boldsymbol{\theta}_{m_n}}(x) = m_n^{-(l+\alpha)}\sum_{j=1}^{m_n}\theta_j H(m_n(x - x_{n,j}))$. Then, condition (2.5) holds:

$$|T \circ f_{\boldsymbol{\theta}_{j0}}(x) - T \circ f_{\boldsymbol{\theta}_{j1}}(x)| \geq m_n^{-(l+\alpha)}|H(m_n(x - x_{n,j}))| \Big/ \max_{x\in[0,1]} f_0(x). \tag{3.11}$$

Note that when n is large, $f_{\boldsymbol{\theta}_{m_n}}(x, y) \geq f_0(x)g_0(y)/2$. Using this, we have

$$\chi^2(f_{\boldsymbol{\theta}_{j0}}, f_{\boldsymbol{\theta}_{j1}}) \leq 2m_n^{-2(l+\alpha)}\int_{-\infty}^{+\infty} H^2(m_n(x - x_{n,j}))/f_0(x)\,dx$$
$$\times \int_{-\infty}^{+\infty} h_0^2(y)/g_0(y)\,dy = O(n^{-1}),$$

if $m_n = cn^{1/(2(l+\alpha)+1)}$, $c > 0$. Therefore, (3.11) and Theorem 1 entail

$$\liminf_{n\to\infty} \inf_{\hat{m}(x)} \sup_{f\in\mathscr{F}_{l,\alpha,B}^2} n^{p(l+\alpha)/(2(l+\alpha)+1)} \times \int_0^1 E(\hat{m}(x) - m(x))^p w(x)\,dx > 0. \tag{3.12}$$

Remark that the rates (3.10) and (3.12) are optimal [Stone (1982)].

3.3. *Summary.* It appears that the cubical bound is an easily-implemented method for finding global minimax rates. Condition (2.4) can usually be checked based on knowledge of the pointwise estimation problem. This technique has also been used in finding optimal global rates of convergence for problems such as positron emission tomography [Zhang (1992)] and errors-in-variables regression [Fan and Truong (1993)]. Optimal global rates of convergence are found via a few lines of simple arguments.

4. Proofs.

Proof of Theorem 1. Assign the prior $\theta_1, \dots, \theta_{n_m}$ to be i.i.d. with

$$P(\theta_j = 0) = P(\theta_j = 1) = 1/2, \quad \text{for } j = 1, \dots, m_n.$$

Let $E_\theta g(\boldsymbol{\theta}_{m_n})$ denote the expectation of $g(\boldsymbol{\theta}_{m_n})$ with respect to the prior distribution of $\theta_1, \dots, \theta_{m_n}$. By Fubini's theorem,

$$\inf_{\hat{T}_n(x)} \sup_{f\in\mathscr{F}_n} E_f \int_a^b |\hat{T}_n(x) - T \circ f(x)|^p w(x)\,dx \geq \int_a^b \inf_{\hat{T}_n(x)} E_\theta E_{f_\theta}|\hat{T}_n(x) - T \circ f(x)|^p w(x)\,dx. \tag{4.1}$$

Let $a_{nj}(x) = |T \circ f_{\boldsymbol{\theta}_{j0}}(x) - T \circ f_{\boldsymbol{\theta}_{j1}}(x)|/2$, where $\boldsymbol{\theta}_{j_0}$ and $\boldsymbol{\theta}_{j_1}$ are given by (2.3). Then, it follows that

$$\inf_{\hat{T}_n(x)} E_\theta E_{f_\theta}|\hat{T}_n(x) - T \circ f(x)|^p \geq \max_{1\leq j\leq m_n} \inf_{\hat{T}_n(x)} E_\theta\left(E_{\theta_j} E_{f_\theta}|\hat{T}_n(x) - T \circ f(x)|^p\right). \tag{4.2}$$

Let $P_{\boldsymbol{\theta}_{j0}}$ and $P_{\boldsymbol{\theta}_{j1}}$ denote the probability measures generated by the density functions $f_{\boldsymbol{\theta}_{j0}}$ and $f_{\boldsymbol{\theta}_{j1}}$, respectively. Then the left-hand side of (4.2) is greater than or equal to

$$\max_{1\le j\le m_n} E_\theta a_{nj}^p(x)\frac{1}{2}\inf_{\hat{T}_n(x)}\Big(P_{\boldsymbol{\theta}_{j0}}\Big\{\Big|\hat{T}_n(x)-T\circ f_{\boldsymbol{\theta}_{j0}}(x)\Big|\ge a_{nj}(x)\Big\}$$

$$+P_{\boldsymbol{\theta}_{j1}}\Big\{\Big|\hat{T}_n(x)-T\circ f_{\boldsymbol{\theta}_{j1}}(x)\Big|\ge a_{nj}(x)\Big\}\Big) \tag{4.3}$$

$$\ge \max_{1\le j\le m_n}\frac{a_{nj}^p(x)}{2}E_\theta S_{n,j}(\boldsymbol{\theta}_{m_n}),$$

where

$$S_{n,j}(\boldsymbol{\theta}_{m_n}) = \inf_{\hat{T}_n(x)}\left(P_{\boldsymbol{\theta}_{j0}}\{R\} + P_{\boldsymbol{\theta}_{j1}}\{R^c\}\right)$$

and $R = \{|\hat{T}_n(x) - T\circ f_{\boldsymbol{\theta}_{j0}}(x)| \ge a_{nj}(x)\}$. Note that $S_{n,j}$ can be viewed as the sum of type I and II errors of a testing procedure with reject region R for the problem

$$H_0\colon f = f_{\boldsymbol{\theta}_{j0}} \leftrightarrow H_1\colon f = f_{\boldsymbol{\theta}_{j1}}. \tag{4.4}$$

Since the χ^2-distance for the pair of densities is no larger than c/n, it follows that [see, e.g., Lemma 1.3 of Fan (1989), page 14]

$$S_{n,j}(\boldsymbol{\theta}_{m_n}) \ge 1-\sqrt{1-e^{-c}} \equiv s_c. \tag{4.5}$$

Consequently, by (2.5), (4.1), (4.3) and (4.5), we have

$$\inf_{\hat{T}_n(x)}\sup_{f\in\mathscr{F}_n} E_f\int_a^b\Big|\hat{T}_n(x)-T\circ f_{j0}(x)\Big|^p w(x)\,dx$$

$$\ge\frac{s_c}{2}\int_a^b\max_{1\le j\le m_n} a_{nj}^p(x)w(x)\,dx \tag{4.6}$$

$$\ge\frac{s_c}{2^{p+1}}\sum_{j=1}^{m_n}\int_{x_{n,j}}^{x_{n,j+1}}\Big|B_n(m_n(x-x_{n,j}))\Big|^p w(x)\,dx.$$

Now, we need to calculate the summation in the last expression. By the uniform continuity of $w(x)$, it follows that for any given ε, there exists an n_0 such that when $n\ge n_0$,

$$\inf_{x\in[x_{n,j},\,x_{n,j+1}]} w(x)\ge(1-\varepsilon)w(x_{n,j}).$$

Consequently, when $n \geq n_0$, the summation in (4.6) is greater than or equal to

$$\frac{1-\varepsilon}{m_n} \sum_{j=1}^{m_n} w(x_{n,j}) \int_a^b |B_n(x-a)|^p \, dx$$

$$= \frac{1-\varepsilon}{b-a} \sum_{j=1}^{m_n} w(x_{n,j})(x_{n,j+1} - x_{n,j}) \int_a^b |B_n(x-a)|^p \, dx.$$

The conclusion follows by letting $n \to \infty$ and then letting $\varepsilon \to 0$. □

Acknowledgments. The author wishes to thank Professor David L. Donoho for valuable conversations. Thanks also go to anonymous referees for helpful comments.

REFERENCES

BIRGÉ, L. (1987). Estimating a density under the order restrictions: Nonasymptotic minimax risk. *Ann. Statist.* **15** 995–1012.

BRETAGNOLLE, J. and HUBER, C. (1979). Estimation des densities, risque minimax. *Z. Wahrsch. Verw. Gebiete* **47** 119–137.

DONOHO, D. L. and LIU, R. C. (1987). Hardest 1-dimensional subproblems. Technical report 105, Dept. Statistics, Univ. California, Berkeley.

DONOHO, D. L. and LIU, R. C. (1991a). Geometrizing rates of convergence II. *Ann. Statist.* **19** 633–667.

DONOHO, D. L. and LIU, R. C. (1991b). Geometrizing rate of convergence III. *Ann. Statist.* **19** 668–701.

DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **18** 1416–1437.

EFROIMOVICH, S. Y. and PINSKER, M. S. (1983). Estimation of square-integrable probability density of a random variable. *Problems Inform. Transmission* **18** 175–189.

FAN, J. (1989). Contributions to the estimation of nonregular functionals. Dissertation, Dept. Statistics, Univ. California, Berkeley.

FAN, J. (1991a). On the optimal rates of convergence for the nonparametric deconvolution problem. *Ann. Statist.* **19** 1257–1272.

FAN, J. (1991b). Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1** 541–551.

FAN, J. and TRUONG, Y. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21**.

FARRELL, R. H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43** 170–180.

HAS'MINSKII, R. Z. (1978). A lower bound on the risks of nonparametric estimates densities in the uniform metric. *Theory. Probab. Appl.* **23** 794–798.

HAS'MINSKII, R. Z. (1979). Lower bound for the risks of nonparametric estimates of the mode. In *Contribution to Statistics: J. Hájek Memorial Volume* (J. Jurečková, ed.) 91–97. Academia, Prague.

JOHNSTONE, I. M. and SILVERMAN, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.* **18** 251–280.

LE CAM, L. (1985). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LOW, M. G. (1993). Renormalizing upper and lower bounds for integrated risk in the white noise model. *Ann. Statist.* **21** 577–589.

NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984–997.

STONE, C. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

STONE, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

ZHANG, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–830.

ZHANG, H. P. (1992). On deconvolution using time of flight information in positron emission tomography. *Statist. Sinica* **2** 553–575.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-3260