

ON THE ESTIMATION OF QUADRATIC FUNCTIONALS

BY JIANQING FAN

University of North Carolina

We discuss the difficulties of estimating quadratic functionals based on observations $Y(t)$ from the white noise model

$$Y(t) = \int_0^t f(u) du + \sigma W(t), \quad t \in [0, 1],$$

where $W(t)$ is a standard Wiener process on $[0, 1]$. The optimal rates of convergence (as $\sigma \rightarrow 0$) for estimating quadratic functionals under certain geometric constraints are found. Specifically, the optimal rates of estimating $\int_0^1 [f^{(k)}(x)]^2 dx$ under hyperrectangular constraints $\Sigma = \{f: |x_j(f)| \leq Cj^{-\alpha}\}$ and weighted l_p -body constraints $\Sigma_p = \{f: \sum_1^\infty j^\alpha |x_j(f)|^p \leq C\}$ are computed explicitly, where $x_j(f)$ is the j th Fourier–Bessel coefficient of the unknown function f . We develop lower bounds based on testing two highly composite hypercubes and address their advantages. The attainable lower bounds are found by applying the hardest one-dimensional approach as well as the hypercube method.

We demonstrate that for estimating regular quadratic functionals [i.e., the functionals which can be estimated at rate $O(\sigma^2)$], the difficulties of the estimation are captured by the hardest one-dimensional subproblems, and for estimating nonregular quadratic functionals [i.e., no $O(\sigma^2)$ -consistent estimator exists], the difficulties are captured at certain finite-dimensional (the dimension goes to infinity as $\sigma \rightarrow 0$) hypercube subproblems.

1. Introduction. The problem of estimating a quadratic functional was considered by Bickel and Ritov (1988), Hall and Marron (1987) and Ibragimov, Nemirovskii and Khas'minskii (1986). Their results indicate the following phenomena: For estimating a quadratic functional, the regular rate of convergence can be achieved when the unknown density is smooth enough and otherwise a singular rate of convergence will be achieved. Naturally, one might ask: What is the insight of estimating a quadratic functional nonparametrically? The problem itself is poorly understood and the pioneering works shows that the new phenomena need to be discovered.

Let us consider the following problem of estimating a quadratic functional. Suppose that we observe $y = (y_j)$ with

$$(1.1) \quad y_j = x_j + z_j,$$

where z_1, z_2, \dots are i.i.d. random variables distributed as $N(0, \sigma^2)$ and $x = (x_j: j = 1, 2, \dots)$ is an unknown element of a set $\Sigma \subset R^\infty$. We are interested in

Received May 1989; revised July 1990.

AMS 1980 subject classifications. Primary 62C05; secondary 62M99.

Key words and phrases. Cubical subproblems, quadratic functionals, rates of convergence, minimax risk, Gaussian white noise, hardest one-dimensional subproblem, shapes of geometry.

estimating a quadratic functional

$$(1.2) \quad Q(x) = \sum_{j=1}^{\infty} \lambda_j x_j^2, \quad \lambda_j \geq 0,$$

with a geometric constraint Σ . The geometric shapes of Σ we will consider are either a hyperrectangular

$$(1.3) \quad \Sigma = \{x: |x_j| \leq A_j\}$$

or a weighted l_p -body

$$(1.4) \quad \Sigma = \left\{ x: \sum_1^{\infty} \delta_j |x_j|^p \leq C \right\}.$$

These are two interesting geometric shapes of constraints, which appear quite often in the literature of nonparametric estimation [Donoho, Liu and MacGibbon (1990), Efroimovich and Pinsker (1982), Nussbaum (1985), among others].

Let us indicate briefly how the problem (1.1)–(1.4) is related to estimating a quadratic functional of an unknown function; see also Donoho, Liu and MacGibbon (1990), Efroimovich and Pinsker (1982), Nussbaum (1985). Suppose that we are interested in estimating

$$(1.5) \quad T(f) = \int_a^b [f^{(k)}(t)]^2 dt$$

with a priori information that f is smooth, but f is observed in a white noise

$$(1.6) \quad Y(t) = \int_a^t f(u) du + \sigma \int_a^t dW(u), \quad t \in [a, b],$$

where $W(t)$ is a Wiener process.

Let us assume that $[a, b] = [0, 1]$ and the function f fulfills periodic boundary conditions at 0 and 1. Take an orthogonal basis to be the usual sinusoids: $\phi_1(t) = 1$, $\phi_{2j}(t) = \sqrt{2} \cos(2\pi jt)$ and $\phi_{2j+1}(t) = \sqrt{2} \sin(2\pi jt)$. Then, (1.5) can be rewritten as

$$(1.7) \quad T(f) = (2\pi)^{2k} \sum_{j=1}^{\infty} j^{2k} (x_{2j}^2 + x_{2j+1}^2) + \delta_k x_1^2$$

and the model (1.6) is equivalent to

$$(1.8) \quad y_j = x_j + \sigma z_j,$$

where $y_j = \int_0^1 \phi_j(t) dY(t)$, $x_j = \int_0^1 \phi_j(t) f(t) dt$, $z_j = \int_0^1 \phi_j(t) dW(t)$ and $\delta_k = 1$, if $k = 0$ and $\delta_k = 0$, otherwise.

Suppose that we know on a priori smoothness condition that the Fourier–Bessel coefficients of f decay rapidly:

$$|x_j| \leq A_j, \quad A_j \rightarrow 0, \quad \text{if } j \rightarrow \infty.$$

Then, the problem reduces to (1.1)–(1.3). Specifically, if the α th derivative of f is bounded, then $|x_j| \leq Cj^{-\alpha}$ for some C . Thus, $A_j = Cj^{-\alpha}$ is a weakening

condition that f has α bounded derivatives. For this specific problem with $A_j = Cj^{-\alpha}$, the optimal rate of estimating the functional $\int_0^1 [f^{(k)}(t)]^2 dt$ is $O(\sigma)$ when $\alpha > 2k + 0.75$ and $O(\sigma^{2-2(4k+1)/(4\alpha-1)})$, when $k + 0.5 < \alpha \leq 2k + 0.75$.

If a priori smoothness condition is $\Sigma = \{f: \int_0^1 [f^{(\alpha)}(t)]^2 dt \leq C\}$, then, by Parseval's identity, Σ is an ellipsoid $\Sigma = \{x: \sum_{j=1}^{\infty} j^{2\alpha}(x_{2j}^2 + x_{2j+1}^2) \leq C/(2\pi)^{2\alpha}\}$. Thus, we reduce the problem to (1.1), (1.2) with a constraint (1.4). The best rate of convergence is $O(\sigma)$, if $\alpha \geq 2k + 0.25$ and $O(\sigma^{8(\alpha-k)/(4\alpha+1)})$, if $k < \alpha < 2k + 0.25$.

Even though we discuss the possible applications on a bounded interval $[0, 1]$, the previous notion can be easily extended to an unbounded interval by using a suitable orthogonal basis.

In this paper, we consider only for observations (1.1) taking it for granted that the results have a variety of applications, such as those just mentioned. We also take it for granted that the behavior as $\sigma \rightarrow 0$ is important, which is natural when we make connections with density estimation.

An interesting feature of our study is the use of geometric idea, including hypercubical subproblems, inner lengths and hardest hyperrectangular subproblems. We relate a geometric concept—inner length with a lower bound by using the difficulty of a hypercubical subproblem. We show in Section 3 and 4 that for some geometric shapes of constraints (e.g., hyperrectangles, ellipsoids and weighted l_p -bodies), the difficulty of a full nonparametric problem is captured by a hypercubical subproblem. We compare the hypercube lower bound with the maximum risk of a truncated quadratic estimator and show that the ratio of the lower and upper bound is bounded away from 0. Thus, as far as the minimax criterion is concerned, there is little to be gained by nonquadratic procedures in terms of rates of convergence.

A related approach to ours is the hypersphere method developed by Ibragimov, Nemirovskii and Khas'minskii (1986). The notion of their method is to use the difficulty of a hypersphere subproblem as the difficulty of a full nonparametric problem. Their results indicate that for estimating a spherically symmetric functional with an ellipsoid constraint, the difficulty of the full problem is captured by a hypersphere subproblem. We might ask more generally: Can the hypersphere method apply to some other coordinatewise symmetric functionals [see (2.1)] and other shapes of constraints to get attainable lower rates? Unfortunately, the answer is no. We show in Section 5 that the hypersphere method cannot give attainable lower rates of convergence for some other interesting constraints (e.g., hyperrectangles) and some other interesting functionals [e.g., (1.5) with $k \neq 0$]. In contrast, our hypercube bound can give attainable rates in these cases. Indeed, in Section 5, we demonstrate that the hypercube method can give a lower bound at least as sharp as the hypersphere method, no matter what kinds of constraints and functionals are. In other words, the hypercube method is strictly better than the hypersphere method.

Comparing our approach to the traditional approach of measuring the difficulty of a linear functional [see Donoho and Liu (1987 a, b), Farrell (1972), Khas'minskii (1979), Stone (1980)], the hypercube method uses the difficulty

of an n_σ -dimensional ($n_\sigma \rightarrow \infty$) subproblem, instead of a one-dimensional subproblem, as the difficulty of the full nonparametric problem. It has been shown that for estimating a linear functional, the difficulty of a one-dimensional subproblem can capture the difficulty of a full problem with great generality. However, totally new phenomena occur if we are trying to estimate a quadratic functional. The difficulty of the hardest one-dimensional subproblem can only capture the difficulty of a full nonparametric problem for the regular cases [the case that the regular rate $O(\sigma)$ can be achieved]. For nonregular cases (no σ -consistent estimate exists), the hardest one-dimensional subproblem cannot capture the difficulty of the full problem. Thus, any one-dimensional based methods fail to give an attainable rate of convergence. The discrepancy is, however, resolved by using multidimensionally based hypercube method.

CONTENT. We begin by introducing the hypercube method of developing a lower bound in Section 2 and then show that the hypercube method gives an attainable rate of convergence for hyperrectangular constraint in Section 3. The estimator that achieves the optimal rate of convergence is a truncated estimator. In Section 4, we extend the results to some other shapes of constraints, for example, ellipsoids, l_p -bodies. In Section 5, we demonstrate that the hypercube method is a better technique than the hypersphere method of Ibragimov, Nemirovskii and Khas'minskii (1986). Some comments are further discussed in Section 6. Technical proofs are given in Section 7.

2. The hypercube bound. Let us introduce some terminology. Let Σ be a subset of R^∞ and $\Sigma_0 \subset \Sigma$. Suppose that we want to estimate a functional $T(x)$ under the constraint $x \in \Sigma \subset R^\infty$. We call estimating $T(x)$ on Σ as a full problem of the estimation and estimating $T(x)$ on Σ_0 as a subproblem. We say that the difficulty of a subproblem captures the difficulty of the full problem, if the best attainable rates of convergence for both problems are the same. In terms of minimax risk, the minimax risks for the subproblem and the full problem are the same within a factor of constant.

Now, suppose that we want to estimate a coordinatewise symmetric functional $T(x)$, that is,

$$(2.1) \quad T(\pm x_1, \pm x_2, \dots) = T(x_1, x_2, \dots),$$

based on the observations (1.1) under a geometric constraint Σ . Assume that $0 \in \Sigma$. Let $l_n(\Sigma)$ be the supremum on the half lengths of all n -dimensional hypercubes centered at the origin lying in Σ . We call it the n -dimensional inner length of Σ .

The idea of constructing a lower bound of estimating T is to use the difficulty of estimating T on a hypercube as a lower bound of the difficulty of the full problem. More precisely, take the largest hypercube of dimension n (which depends on σ) in the constraint Σ and assign probability $1/2^n$ to each vertex of the hypercube and then test the vertices against the origin. When no perfect test exists by choosing some critical value n , depending on σ , the

difference in functional values at vertices of two hypercubes supplies a lower bound.

To carry out the idea, we want to test the origin against the vertices of the largest hypercube with a uniform prior based on the observations (1.1). The problem is equivalent to the testing problem

$$(2.2) \quad H_0: y_i \sim N(0, \sigma^2) \leftrightarrow H_1: y_i \sim \frac{1}{2} [N(l_{n_\sigma}, \sigma^2) + N(-l_{n_\sigma}, \sigma^2)],$$

$$i = 1, \dots, n_\sigma,$$

where $l_{n_\sigma} = l_{n_\sigma}(\Sigma)$. The result of the testing problem can be summarized as follows:

LEMMA 1. *If $n_\sigma^{1/2}(l_{n_\sigma}/\sigma)^2 \rightarrow c$ (as $\sigma \rightarrow 0$), then the sum of type I and type II errors of the best testing procedure for the problem (2.2) is*

$$(2.3) \quad 2\Phi\left(-\frac{\sqrt{n_\sigma}(l_{n_\sigma}/\sigma)^2}{\sqrt{8}}\right)(1 + o(1)),$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Choose the dimension of the hypercube $n_{\sigma,d}$ (if it exists) to be the smallest integer satisfying

$$(2.4) \quad \sqrt{n} [l_n(\Sigma)]^2 / \sigma^2 \leq d,$$

where d is a positive constant. By Lemma 1, as $\sigma \rightarrow 0$,

$$(2.5) \quad \min_{0 \leq \phi(y) \leq 1} \{E_0\phi(y) + E_1(1 - \phi(y))\} \geq 2\Phi(-d/\sqrt{8})(1 + o(1)),$$

where E_0 and E_1 mean take the expectation with respect to y distributed as (1.1) with the prior $x = 0$ and the prior of x distributed uniformly on the vertices of the hypercube, respectively. Let

$$(2.6) \quad r_n = \frac{1}{2} |T(H_1) - T(H_0)| = \frac{1}{2} |T(x_n) - T(0)|,$$

where $x_n = (l_n(\Sigma), \dots, l_n(\Sigma), 0, 0, \dots)$ is a vertex of the hypercube.

THEOREM 1. *Suppose that $T(x)$ is a coordinatewise symmetric function and $0 \in \Sigma$. If $n_{\sigma,d}$ defined by (2.4) exists, then for any estimator $\delta(y)$ based on the observations (1.1),*

$$\sup_{x \in \Sigma} P_x\{|\delta(y) - T(x)| \geq r_{n_{\sigma,d}}\} \geq \Phi\left(-\frac{d}{\sqrt{8}}\right)(1 + o(1))$$

and for any symmetric nondecreasing loss function $L(|\cdot|)$,

$$\sup_{x \in \Sigma} E_x \left[L\left(\frac{|\delta(y) - T(x)|}{r_{n_{\sigma,d}}}\right) \right] \geq L(1)\Phi\left(-\frac{d}{\sqrt{8}}\right) + o(1).$$

PROOF. By (2.5) and (2.6), for any estimator δ ,

$$\begin{aligned} & \sup_{x \in \Sigma} P_x \{ |\delta(y) - T(x)| \geq r_{n_{\sigma,d}} \} \\ & \geq \frac{1}{2} \{ P_0(|\delta(y) - T(0)| \geq r_{n_{\sigma,d}}) + P_1(|\delta(y) - T(x_{n_{\sigma,d}})| \geq r_{n_{\sigma,d}}) \} \\ & \geq \frac{1}{2} \{ P_0(|\delta(y) - T(0)| \geq r_{n_{\sigma,d}}) + P_1(|\delta(y) - T(0)| \leq r_{n_{\sigma,d}}) \} \\ & \geq \Phi \left(-\frac{d}{\sqrt{8}} \right) + o(1), \end{aligned}$$

where P_0 and P_1 are the probability measures generated by y distributed according to (1.1) with the prior $x = 0$ and the prior of x distributed uniformly on the vertices of the hypercube, respectively. Now for any symmetric nondecreasing loss function

$$\sup_{x \in \Sigma} E_x \left[L \left(\frac{|\delta(y) - T(x)|}{r_{n_{\sigma,d}}} \right) \right] \geq L(1) \sup_{x \in \Sigma} P_x \{ |\delta(y) - T(x)| \geq r_{n_{\sigma,d}} \}.$$

The second conclusion follows.

In particular, under the assumptions of Theorem 1, we have for any estimator

$$(2.7) \quad \sup_{x \in \Sigma} E_x (\delta(y) - T(x))^2 \geq \Phi \left(-\frac{d}{\sqrt{8}} \right) r_{n_{\sigma,d}}^2 (1 + o(1)).$$

Thus, $\Phi(-d/\sqrt{8})r_{n_{\sigma,d}}^2$ is a minimax lower bound under the quadratic loss. \square

3. Truncation estimators. Let us start with the model (1.1) with a hyperrectangular type of constraint (1.3). An intuitive class of quadratic estimators to estimate the quadratic functional (1.2) is the class of estimators defined by

$$(3.1) \quad q_B(y) = y'By + c,$$

where B is a symmetric matrix and c is a constant. Simple algebra shows that the risk of $q_B(y)$ under the quadratic loss is

$$\begin{aligned} (3.2) \quad R(B, x) & \triangleq E_x (q_B(y) - Q(x))^2 \\ & = (x'Bx + \sigma^2 \text{tr } B + c - Q(x))^2 + 2\sigma^4 \text{tr } B^2 + 4\sigma^2 x'B^2x. \end{aligned}$$

The following proposition tells us that the class of quadratic estimators with diagonal matrices is a complete class among all estimators defined by (3.1).

PROPOSITION 1. Let D_B be a diagonal matrix, whose diagonal elements are those of B . Then for each symmetric matrix B ,

$$\max_{x \in \Sigma} R(B, x) \geq \max_{x \in \Sigma} R(D_B, x),$$

where Σ is defined by (1.3).

Thus, it is enough to consider only quadratic estimators with diagonal matrices. For a diagonal matrix $B = \text{diag}(b_1, b_2, \dots)$, the estimator (3.1) has risk

$$\begin{aligned} R(B, x) &= \left(\sum_1^\infty b_j x_j^2 + \sigma^2 \sum_1^\infty b_j + c - \sum_1^\infty \lambda_j x_j^2 \right)^2 \\ (3.3) \quad &+ \sum_1^\infty b_j^2 (2\sigma^4 + 4\sigma^2 x_j^2). \end{aligned}$$

Even for the diagonal matrices, it is hard to find the exactly optimal quadratic estimator [see Sacks and Ylvisaker (1981)]. For an infinite-dimensional estimation problem, usually bias is a major contributor to the risk of an estimator. Thus, we would prefer to use the unique unbiased quadratic estimator $\sum_1^\infty \lambda_j (y_j^2 - \sigma^2)$, but it might not converge in L^2 and even if it does converge, it might contribute too much in its variance. Thus, we consider a truncated quadratic estimator

$$(3.4) \quad q_T(y) = \sum_1^m \lambda_j (y_j^2 - \sigma^2)$$

and choose m to minimize its maximum MSE. Note that the naive truncated estimator $\sum_1^m \lambda_j y_j^2$ will not attain optimal rate of convergence, as it contributes too much bias. For the estimator $q_T(y)$, the maximum MSE is

$$(3.5) \quad \max_{x \in \Sigma} R(q_T, x) = \left(\sum_m^\infty \lambda_j A_j^2 \right)^2 + \sum_1^m \lambda_j^2 (2\sigma^4 + 4\sigma^2 A_j^2).$$

The following theorem shows that the truncated quadratic estimator (3.4) achieves a certain rate, which will be further justified to be optimal by Theorem 3.

ASSUMPTION A. Assume that when n is large, the following conditions hold.

(i) nA_n^4 is a strictly decreasing sequence, which goes to 0 as $n \rightarrow \infty$, and λ_n is a nondecreasing sequence. Moreover,

$$\limsup_{n \rightarrow \infty} \frac{A_{n-1}}{A_n} < \infty, \quad \liminf_{n \rightarrow \infty} \sum_1^n \frac{\lambda_j}{n\lambda_n} > c \quad \text{for some } c > 0.$$

(ii) $\sum_n^\infty \lambda_j A_j^2 = O(n\lambda_n A_n^2)$ and if $\limsup_{n \rightarrow \infty} n^{1.5} \lambda_n^2 A_n^2 = \infty$, then $\sum_1^n \lambda_j^2 A_j^2 = O(n^{1.5} \lambda_n^2 A_n^2)$.

THEOREM 2. *Under Assumption A, the achievable rate ($\sigma \rightarrow 0$) of estimating $Q(x) = \sum_1^\infty \lambda_j x_j^2$ under the quadratic loss with a hyperrectangular constraint (1.3) is*

$$(3.6) \quad O\left(\sigma^2 + \left(\sum_1^{n_{\sigma,d}} \lambda_j\right)^2 A_{n_{\sigma,d}}^4\right),$$

where $n_{\sigma,d}$ is the smallest integer such that

$$(3.7) \quad \sqrt{n} (A_n/\sigma)^2 \leq d$$

for some $d > 0$. Moreover, the rate is achieved by the truncated estimator (3.4) with $m = n_{\sigma,d}$.

When $\lambda_j = j^q$ and $A_j = Cj^{-\alpha}$, ($\alpha > (q + 1)/2$), the conditions of Theorem 2 are satisfied. In this case, the mean square error can be more precisely evaluated. Let

$$(3.8) \quad m_\sigma = \begin{cases} (C^4/(2\alpha - q - 1))^{1/(4\alpha-1)} \sigma^{-1/(4\alpha-1)}, & \text{when } (q + 1)/2 < \alpha \leq q + 0.75, \\ D\sigma^{-4/(4\alpha-1)}, & \text{when } \alpha > q + 0.75, \end{cases}$$

which optimizes the truncated estimator (3.4), where D is a positive constant. The maximum risk of the truncated estimator is given by

$$(3.9) \quad \begin{aligned} & C^{(4(2q+1))/(4\alpha-1)} \left\{ \frac{2c_0^{2q+1}}{2q+1} + \frac{c_0^{-2(2\alpha-q-1)}}{(2\alpha-q-1)^2} + D_{\alpha,q} \right\} \sigma^{4-(4(2q+1))/(4\alpha-1)}, \\ & \text{when } \frac{q+1}{2} < \alpha \leq q + 0.75, \\ & 4C^2 \sum_1^\infty j^{2q-2\alpha} \sigma^2, \quad \text{when } \alpha > q + 0.75, \end{aligned}$$

where $c_0 = (2\alpha - q - 1)^{-1/(4\alpha-1)}$ and $D_{\alpha,q} = 4\sum_1^\infty j^{-1.5}$, if $\alpha = q + 0.75$, and $= 0$ otherwise. In summary:

COROLLARY 1. *Suppose that $\lambda_j = j^q$ and $A_j = Cj^{-\alpha}$, ($\alpha > (q + 1)/2$). Then the best truncated estimator is given by (3.4) with $m = m_\sigma$. The asymptotic maximum risk of the estimator is given by (3.9). Moreover, the estimator achieves the optimal rate of convergence.*

Note that the condition $\alpha > (q + 1)/2$ is a necessary condition to make the quadratic function $Q(x)$ bounded on Σ . As soon as the finiteness condition is fulfilled, the functional can be estimated consistently. A similar remark applies to the weighted l_p -bodies discussed in the next section.

When $\alpha \geq q + 0.75$, the regular rate $O(\sigma^2)$ is achieved by the best truncated estimator and hence the difficulty of the full problem of estimating $Q(x)$ can be captured by a one-dimensional subproblem. However, the situation changes when $\alpha < q + 0.75$. The difficulty of the hardest one-dimensional subproblem cannot capture the difficulty of the full problem any more [compare (6.3) with (3.9)]. Thus, we need to establish a larger lower bound for this case by applying Theorem 1. By our method of construction, we need an $n_{\sigma,d}$ -dimensional subproblem, not just a one-dimensional subproblem, in order to capture the difficulty of the full problem for this case.

THEOREM 3. *Suppose that nA_n^4 is a sequence decreasing to 0. Let Σ be defined by (1.3) and $n_{\sigma,d}$ be the smallest integer such that*

$$\sqrt{n} (A_n/\sigma)^2 \leq d.$$

Then, for any estimator $T(y)$, the maximum MSE of estimating $Q(x)$ on Σ is no smaller than

$$\frac{\Phi(-d/\sqrt{8})}{4} \left(\sum_1^{n_{\sigma,d}} \lambda_j \right)^2 A_{n_{\sigma,d}}^4 (1 + o(1)) \quad \text{as } \sigma \rightarrow 0.$$

Moreover, for any estimator $T(y)$,

$$\sup_{x \in \Sigma} P_x \left\{ |T(y) - Q(x)| \geq \frac{(\sum_1^{n_{\sigma,d}} \lambda_j) A_{n_{\sigma,d}}^2}{2} \right\} \geq \Phi \left(-\frac{d}{\sqrt{8}} \right) (1 + o(1)).$$

Combining Theorem 2 and 3, the rates given by (3.6) are optimal. When $A_j = Cj^{-\alpha}$ and $\lambda_j = j^q$, we can calculate the rate in Theorem 3 explicitly.

COROLLARY 2. *When $A_j = Cj^{-\alpha}$ and $\lambda_j = j^q$ for any estimator, the maximum risk under the quadratic loss is no smaller than*

$$(3.10) \quad (2q + 2)^{-2} C^{(4(2q+1))/(4\alpha-1)} \xi_{\alpha,q} \sigma^{4-(4(2q+1))/(4\alpha-1)} (1 + o(1)),$$

where

$$\xi_{\alpha,q} = \max_{d>0} d^{2-(4q+2)/(4\alpha-1)} \Phi \left(-\frac{d}{\sqrt{8}} \right).$$

In the following examples, we assume that the constraint is $\Sigma = \{x: |x_j| \leq Cj^{-\alpha}\}$.

EXAMPLE 1. Suppose that we want to estimate $T(f) = \int_0^1 f^2(t) dt$ from the model (1.6). Let $\{\phi_j(t)\}$ be a fixed orthonormal basis. Then $T(f) = \sum_{j=1}^{\infty} x_j^2$. Thus, the optimal truncated estimator is $\sum_1^{m_{\sigma}} \lambda_j (y_j^2 - \sigma^2)$, where m_{σ} is given by (3.8) with $q = 0$. Moreover, when $\alpha > 0.75$, the estimator is an asymptotic minimax estimator [see (6.4)]. For $0.5 < \alpha \leq 0.75$, the optimal rates are achieved.

EXAMPLE 2. Let orthonormal basis be $\{\phi_j(t)\}$, where $\phi_1 = 1$, $\phi_{2j} = \sqrt{2} \cos 2\pi jt$ and $\phi_{2j+1} = \sqrt{2} \sin 2\pi jt$. We want to estimate

$$T(f) = \int_0^1 [f^{(k)}(t)]^2 dt = (2\pi)^{2k} \sum_{j=1}^{\infty} j^{2k} (x_{2j}^2 + x_{2j+1}^2), \quad k \geq 1.$$

An estimator which achieves the optimal rate of convergence is $(2\pi)^{2k} \sum_{j=1}^{m_\sigma} j^{2k} (y_{2j}^2 + y_{2j+1}^2 - 2\sigma^2)$, where m_σ is given by (3.8) with $q = 2k$. Moreover, the estimators achieve the optimal rates given by

$$(3.11) \quad \begin{aligned} O(\sigma^2) & \quad \text{if } \alpha > 2k + 0.75, \\ O(\sigma^{4 - (4(4k+1))/(4\alpha-1)}) & \quad \text{if } k + 0.5 < \alpha \leq 2k + 0.75. \end{aligned}$$

4. Extension to quadratically convex sets. In this section, we will find the optimal rate of convergence for estimating the quadratic functional

$$(4.1) \quad Q(x) = \sum_1^{\infty} \lambda_j x_j^2$$

under a geometric constraint

$$(4.2) \quad \Sigma_p = \left\{ x: \sum_1^{\infty} \delta_j |x_j|^p \leq C \right\},$$

called the *weighted l_p -body*. We use the hypercube method to develop the attainable rate of convergence. From these studies, we demonstrate that the hypercubical subproblem captures the difficulty of the full problem with great generality. We begin with making some assumptions. Note that all the following assumptions are fulfilled by the sequence $\lambda_j = j^q$, $\delta_j = j^r$ ($q, r \geq 0$).

ASSUMPTION B. (i) The sequence $\{\lambda_n\}$ is nonnegative and satisfies $\sum_1^n \lambda_j^2 = O(n\lambda_n^2)$; $\{\delta_n\}$ are positive nondecreasing sequences.

(ii) There exists a positive constant c such that $\sum_1^n \lambda_j > cn\lambda_n$, $\sum_1^n \delta_j > cn\delta_n$, $\delta_n > c\delta_{n+1}$.

(iii) The sequence $\delta_n^{4/p} n^{4/p-1}$ increases to infinity as $n \rightarrow \infty$.

ASSUMPTION C. When $p = 2$, the following conditions are fulfilled.

(i) $\{\lambda_j/\delta_j\}$ is a nonincreasing sequence.

(ii) $\max_{1 \leq j \leq n} \lambda_j^2/\delta_j = O(\sqrt{n} \lambda_n^2/\delta_n)$, if $\lambda_n^2/\delta_n \rightarrow \infty$.

When $p > 2$, the following conditions are fulfilled.

(i') $\sum_n \lambda_j^{p/(p-2)} \delta_j^{-2/(p-2)} = O(n \lambda_n^{p/(p-2)} \delta_n^{-2/(p-2)})$.

(ii') $\sum_1^n \lambda_j^{2p/(p-2)} \delta_j^{-2/(p-2)} = O(n^{(3p-4)/(2(p-2))} \lambda_n^{2p/(p-2)} \delta_n^{-2/(p-2)})$, if $\limsup_n n^{(3p-4)/2p} \lambda_n^2 \delta_n^{-2/p} = \infty$.

THEOREM 4. Under Assumptions B and C, the optimal rate of convergence for estimating $Q(x)$ defined by (4.1) under the weighted l_p -body ($p \geq 2$)

constraint (4.2) is given by

$$(4.3) \quad \begin{aligned} &O(\sigma^2), && \text{if } \limsup_n n^{(3p-4)/(2p)} \lambda_n^2 \delta_n^{-2/p} < \infty, \\ &n_{\sigma,d} \lambda_{n_{\sigma,d}}^2 \sigma^4, && \text{if } \limsup_n n^{(3p-4)/(2p)} \lambda_n^2 \delta_n^{-2/p} = \infty. \end{aligned}$$

Moreover, the truncated estimator $\sum_1^{n_{\sigma,d}} \lambda_j (y_j^2 - \sigma^2)$ achieves the optimal rate of convergence, where $n_{\sigma,d}$ is the largest integer such that

$$(4.4) \quad \delta_n^{4/p} n^{4/p-1} \sigma^4 < d$$

for some positive constant d .

COROLLARY 3. When $\lambda_j = j^q$ and $\delta_j = j^r$, the optimal rate of estimating $Q(x)$ under the weighted l_p -body constraint (4.2) ($p \geq 2$) is

$$(4.5) \quad \begin{aligned} &O(\sigma^2), && \text{if } r \geq 0.75p - 1 + pq, \\ &O(\sigma^{8(2(r+1)-p(q+1))/(4(r+1)-p)}), && \text{if } p(q+1)/2 - 1 < r < 0.75p - 1 + pq. \end{aligned}$$

Moreover, the truncated estimator $\sum_1^{n_\sigma} j^q (y_j^2 - \sigma^2)$ achieves the optimal rate of convergence, where $n_\sigma = [(d/\sigma^4)^{p/(4r+4-p)}]$, d is a positive constant.

REMARK 1. Geometrically, the weighted l_p -body is quadratically convex (convex in x_j^2) when $p \geq 2$ and is convex when $1 \leq p < 2$ and is not convex when $0 < p < 1$ [Donoho, Liu and MacGibbon (1990)]. Our results in this section show that for the special quadratically convex constraints, the difficulty of estimating a quadratic functional is captured by a hypercubical subproblem. As the hardest hyperrectangular subproblem is at least as difficult as a hypercubical subproblem, the difficulty of estimating $Q(x)$ under a weighted l_p -body is further captured by the hardest hyperrectangular subproblem. More general phenomena might be true: the difficulties of estimating quadratic functionals under a quadratically convex constraint are captured by the hardest hyperrectangular subproblems (see Discussion for detail).

EXAMPLE 3 (Estimating integrated squared derivatives). Suppose that we want to estimate $T(f) = \int_0^1 [f^{(k)}(t)]^2 dt$ based on (1.6) under the nonparametric constraint that

$$\Sigma = \left\{ f(t) : \int_0^1 [f^{(\alpha)}(t)]^2 dt \leq C \right\}.$$

Let the orthonormal basis $\{\phi_j(t)\}$ be defined by Example 2. Then

$$T(f) = (2\pi)^{2k} \sum_{j=1}^\infty j^{2k} (x_{2j}^2 + x_{2j+1}^2) + \eta_k x_1^2.$$

where $\eta_k = 1$, if $k = 0$, and $\eta_k = 0$, if $k \neq 0$, and x_j is the j th Fourier coefficient. Assume additionally that when $k = 0$, $|x_1| \leq B$, a finite constant. The nonparametric constraint Σ can be rewritten as an ellipsoid (weighted l_2 -body)

$$\Sigma = \left\{ x: \sum_{j=1}^{\infty} j^{2\alpha} (x_{2j}^2 + x_{2j+1}^2) \leq \frac{C}{(2\pi)^{2\alpha}} \right\}.$$

By Corollary 3, the truncated estimator $(2\pi)^{2k} \Sigma_1^{n\sigma} j^{2k} (y_{2j}^2 + y_{2j+1}^2 - 2\sigma^2) + \eta_k (y_1^2 - \sigma^2)$ with

$$\eta_\sigma = [d^{-1/(4\alpha+1)} \sigma^{-4/(4\alpha+1)}], \quad d > 0,$$

achieves the optimal rate of convergence given by

$$(4.6) \quad \begin{array}{ll} O(\sigma^2) & \text{if } \alpha \geq 2k + 0.25, \\ \sigma^{16(\alpha-k)/(4\alpha+1)} & \text{if } 2k + 0.25 > \alpha > k. \end{array}$$

5. Comparison with Ibragimov, Nemirovskii and Khas'minskii.

Our method of developing a lower bound is similar to that of Ibragimov, Nemirovskii and Khas'minskii (1986). Their method is based on testing the largest inner sphere instead of testing the vertices of a hypercube. Let us walk through the main steps of Ibragimov, Nemirovskii and Khas'minskii's method: (i) inscribe the largest n -dimensional hypersphere S^n into the constraint Σ ; (ii) test the origin against S^n based on the observations (1.1); (iii) choose dimension n (depending on σ) such that no perfect testing procedure exists; (iv) compute the difference of functional $\inf_{x \in S^n} |T(x) - T(0)|$ and use it as the rate of a lower bound.

We expect that their method can only apply to spherically symmetric functionals, as the values of the functionals remain the same on the sphere (see Remark 3). Furthermore, it is not hard to argue that if the method of Ibragimov, Nemirovskii and Khas'minskii (1986) gives an attainable lower bound (sharp in rate) for some symmetric functionals under some geometric constraints, our method does in the the same setting, and on the other hand, even though the method of Ibragimov, Nemirovskii and Khas'minski (1986) cannot give an attainable lower bound for some geometric constraints and for some symmetric functionals, our method can. Therefore, it turns out that our method has much broader applications not only in the shapes of geometric constraints but also in the classes of symmetric functionals being estimated.

The key points of the argument can be highlighted as follows [see Fan (1989b) for detail].

1. It is easier to inscribe a hypercube into a set Σ than a hypersphere. In terms of geometric concepts, the inner length and inner radius satisfy $l_n(\Sigma) \geq r_n(\Sigma) / \sqrt{n}$, where $r_n(\Sigma)$ is the n -dimensional inner radius, namely,

the supremum of the radii of all n -dimensional discs centered at 0 lying in Σ [see Ibragimov, Nemirovskii and Khas'minskii (1986) and Chentsov (1980)].

2. Let C^n be the vertices of the largest inner hypercube of an n -dimensional hypersphere S^n centered at the origin. Under the model (1.1), testing the S^n (with a uniform prior) against the origin is as difficult as testing C^n (with a uniform prior) against the origin that is, the sum of type I and type II errors of the best testing procedures for both testing problems are asymptotically the same.
3. The difference of functional values satisfies

$$\inf_{x \in S^n} |T(x) - T(0)| \leq \inf_{x \in C^n} |T(x) - T(0)|.$$

Let us justify the previous claim by examples.

REMARK 2. Consider estimating the quadratic functional $Q(x) = \sum_1^\infty j^q x_j^2$ with the hyperrectangular constraint $\Sigma = \{x \in \mathbb{R}^\infty: |x_j| \leq Cj^{-\alpha}\}$. It appears that the hypersphere bound of Ibragimov, Nemirovskii and Khas'minskii (1986) would not give a lower bound of the same order as we are able to get via hypercubes. The proof is simple. Assume that $C = 1$. By Corollary 6 of Ibragimov, Nemirovskii and Khas'minskii (1986), the lower bound under the quadratic loss is of order $O(\sigma^{4-4/(4\alpha+1)})$, which is of lower order of (3.10)—the lower bound developed by the hypercube method. It is clear that in the current setting a hypercube is easier to inscribe into a hyperrectangular than a hypersphere: $l_n(\Sigma) > r_n(\Sigma)/\sqrt{n}$. Hence, a hypersphere's method cannot give an attainable lower rate.

REMARK 3. Using the method of Ibragimov, Nemirovskii and Khas'minskii (1986) to develop the lower bound for the weighted l_p -body ($\lambda_j = j^q, \delta_j = j^r$), we find that the lower bound is of order $O(\sigma^{-16r/(4r+p)})$, which is the attainable rate only when $p = 2$ and $q = 0$. When $p = 2$, the largest n -dimensional inner hypercube lies in the largest n -dimensional hypersphere: $l_n(\Sigma_2) = r_n(\Sigma_2)/\sqrt{n}$. The reason that the hypersphere method gives a too small lower bound is that the functional $Q(x) = \sum_1^\infty j^q x_j^2$ ($q > 0$) does not remain constant when x is on an n -dimensional sphere. For the case the $p > 2$, even when $q = 0$ (the functional is spherical symmetry), the hypersphere does not give an attainable rate, due to the fact that the hypercube is easier to inscribe than a hypersphere: $l_n(\Sigma_p) > r_n(\Sigma_p)/\sqrt{n}$.

In summary, the hypercube method is strictly better than the hypersphere approach and has the following advantages: (i) larger in difference of functional values (broader application in terms of functionals being estimated, for example, coordinatewise symmetric functionals), (ii) easier to inscribe (broader application in terms of constraints, for example, hyperrectangles, weighted

l_p -bodies with $p > 2$). (iii) more applicable to a nonsymmetric constraint and even to a nonsymmetric functionals [see Example 5 of Fan (1989b)].

6. Discussion.

Possible applications.

1. We have demonstrated that for geometric constraints of hyperrectangles and weighted l_p -bodies ($p \geq 2$), the difficulties of estimating quadratic functionals are captured by hypercubical subproblems. The notions inside can be explained as follows. For hypercube-typed hyperrectangles (i.e., the lengths of a hyperrectangle satisfy Assumption A), the difficulties of estimating quadratic functionals are captured by hypercubical subproblems (Theorem 2). Now for weighted l_p -bodies (Theorem 4), the difficulties of the estimating quadratic functionals are actually captured by rectangular subproblems, which happen to be hypercube-typed. Thus, the difficulties of estimating quadratic functionals under the weighted l_p -body constraints (Theorem 4) are also captured by the cubical subproblems. More general phenomena might be true: the difficulties of estimating quadratic functionals under quadratically convex (convex in x_j^2) constraints Σ (see Remark 1) are captured by the hardest hyperrectangular subproblems

$$\begin{aligned} & \min_{q(y)} \max_{x \in \Sigma} E_x(q(y) - Q(x))^2 \\ & \leq C \max \left\{ \min_{q(y)} \max_{x \in \Theta(\tau)} E_x(q(y) - Q(x))^2 : \Theta(\tau) \in \Sigma \right\}, \end{aligned}$$

where $\Theta(\tau)$ is a hyperrectangle with the coordinates τ , $q(y)$ is an estimator based on our model (1.1) and C is a finite constant. For estimating linear functionals, the previous phenomenon is true [Donoho, Liu and MacGibbon (1990)].

2. A useful geometric concept is the Kolmogorov n -width of a set Σ , denoted by $d_n(\Sigma)$, which has a strong tie with the achievability. Recall that the inner length $l_n(\Sigma)$ was related with a lower bound by Theorem 1. A referee observed the interesting phenomena: For the interesting examples under our consideration, the two geometric measures satisfy

$$(6.1) \quad d_n(\Sigma) = \sqrt{n} l_n(\Sigma),$$

where Σ is either a hyperrectangular $\{x: |x_j| \leq Cj^{-\alpha}\}$ or a weighted l_p -body $\{x: \sum_1^\infty j^\alpha |x_j|^p \leq C\}$. This suggests that for a certain class of functionals, if the geometric constraint satisfies (6.1), the hypercube method would give an attainable lower bound.

Constants. By using the hardest one-dimensional trick, we can prove the following lower bound [see Fan (1989a) for details].

THEOREM 5. *If Σ is a symmetric, convex set containing the origin, then the minimax risk of estimating $Q(x)$ from the observations (1.1) is at least*

$$(6.2) \quad \sup_{x \in \Sigma} \frac{Q^2(x)\sigma^4}{\|x\|^4} \rho\left(\frac{\|x\|}{\sigma}, 1\right)$$

and as $\sigma \rightarrow 0$, for any estimator $\delta(y)$.

$$(6.3) \quad \sup_{x \in \Sigma} E_x(\delta(y) - Q(x))^2 \geq \sup_{x \in \Sigma, \|x\| > 0} \left\{ \frac{4Q^2(x)}{\|x\|^2} \right\} \sigma^2(1 + o(1)),$$

where $\rho(\tau, 1) = \inf_{\delta} \sup_{|\theta| \leq \tau} E_{\theta}(\delta(z) - \theta^2)^2$ and $z \sim N(\theta, 1)$.

It is proved by Fan (1989a) that $\rho(\tau, 1) = 4\tau^2(1 + o(1))$ (as $\tau \rightarrow \infty$) by using a Bayesian approach with a sequence of prior densities given by

$$(6.4) \quad g_m(\theta, \tau) = \frac{(2m + 1)\theta^{2m}}{2\theta^{2m+1}} 1_{[|\theta| \leq \tau]}.$$

Comparing the lower bound (6.3) with the upper bound (3.9) for estimating $\sum_1^{\infty} j^q x_j^2$ with a hyperrectangular constraint $\{x \in R^{\infty} \mid |x_j| \leq Cj^{-\alpha}\}$, we have that when $\alpha > q + 0.75$,

$$(6.5) \quad 1 \geq \frac{\text{lower bound}}{\text{upper bound}} \geq \frac{C_{2\alpha-q}^2}{C_{2\alpha} C_{2\alpha-2q}}, \text{ as } \sigma \rightarrow 0,$$

where $C_r = \sum_1^{\infty} j^{-r}$ can be calculated numerically. Table 1 shows results of the right-hand side in (6.5).

From Table 1, we know that the best truncated estimator is very efficient, when α is a little bit away from q . Thus, the difficulty of the hardest one-dimensional subproblem captures the difficulty of the full problem pretty well in the case $\alpha > q + 0.75$.

TABLE 1
Comparison of the lower bound and upper bound ($\alpha > q + 0.75$) $\alpha = 1 + q + 0.5i$

q	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$
0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.910	0.976	0.991	0.996	0.998	0.999	1.000	1.000	1.000
1.0	0.812	0.940	0.977	0.990	0.995	0.998	0.999	0.999	1.000
1.5	0.744	0.910	0.963	0.984	0.992	0.996	0.998	0.999	0.999
2.0	0.700	0.887	0.952	0.979	0.990	0.995	0.998	0.999	0.999
2.5	0.671	0.871	0.944	0.975	0.988	0.995	0.997	0.998	0.999
3.0	0.651	0.859	0.938	0.972	0.987	0.994	0.997	0.998	0.999
3.5	0.638	0.851	0.934	0.970	0.986	0.993	0.996	0.998	0.999
4.0	0.629	0.845	0.931	0.968	0.985	0.993	0.996	0.998	0.999
4.5	0.622	0.842	0.929	0.967	0.984	0.992	0.996	0.998	0.999
5.0	0.618	0.839	0.928	0.966	0.984	0.992	0.996	0.998	0.999

TABLE 2
 Comparison of the lower bound and the upper bound ($(q + 1)/2 < \alpha < q + 0.75$)

$q = 0$		$q = 1$		$q = 2$		$q = 3$		$q = 4$	
α	ratio	α	ratio	α	ratio	α	ratio	α	ratio
0.55	0.0333	1.15	0.0456	1.75	0.0485	2.35	0.0495	2.95	0.0499
0.60	0.0652	1.30	0.0836	2.00	0.0864	2.70	0.0864	3.40	0.0858
0.65	0.0949	1.45	0.1159	2.25	0.1170	3.05	0.1153	3.85	0.1132
0.70	0.1218	1.60	0.1431	2.50	0.1419	3.40	0.1383	4.30	0.1345

$$\text{ratio} = \sqrt{\frac{\text{hypercube lower bound}}{\text{upper bound by } q_T(y)}}$$

When $(q + 1)/2 < \alpha \leq q + 0.75$, by comparing the lower bound (3.10) and the upper bound (3.9), again we show that the best truncated estimator attains the optimal rate. Table 2 shows how close the lower bound and the upper bound are.

Table 2 tells us that there is a large discrepancy at the level of constants between the upper and lower bounds for the case that $(q + 1)/2 < \alpha \leq q + 0.75$.

Bayesian approach. The following discussion will focus on estimating the quadratic functional $Q(x) = \sum_1^\infty j^q x_j^2$ with the constraint $x \in \Sigma = \{x: |x_j| \leq j^{-\alpha}\}$.

The traditional method of finding a minimax lower bound is using Bayesian method with an intuitive prior. However, in the current setting, all intuitive Bayesian methods fail to give an attainable (sharp in rate) lower bound [see Fan (1989b) for a proof]. Thus, finding an optimal rate of estimating a quadratic functional is a nontrivial job. Here, by an intuitive prior, we mean that it assigns the prior to each coordinate $x_j \sim \pi_j(\theta)$ independently.

The structure of our hypercube method suggests that it may be possible to improve the constant factor of the lower bound via a Bayesian method: take the largest n -dimensional hypercube in Σ and then assign the prior $\pi(\theta)$ on a diagonal line segment starting from the origin to a vertex of the hypercube, with probability 2^{-n} to each diagonal line segment. Then, the minimax lower bound of estimating a functional $\sum_1^\infty \lambda_j x_j^2$ is bounded from below by

$$\left(\sum_1^n \lambda_j \right)^2 \sigma^2 B_\pi(A_n/\sigma, n),$$

where $B_\pi(\tau, n)$ is the Bayes risk of estimating θ^2 based on n i.i.d. observations from $0.5N(\theta, 1) + 0.5N(-\theta, 1)$ with a prior $\pi(\theta)$ concentrated on $|\theta| \leq \tau$, [see Section 2.5 of Fan (1989a) for detail]. We suggest to take a sequence of prior $\pi(\theta)$ given by (6.4).

7. Proofs.

PROOF OF LEMMA 1. Without loss of generality, assume that $\sigma = 1$, and write n_σ as n . Then the likelihood ratio of the joint densities under H_0 and H_1 is

$$(7.1) \quad L_n = \prod_1^n L_{n,i},$$

where

$$L_{n,i} = \exp(-l_n^2/2)[\exp(l_n y_i) + \exp(-l_n y_i)]/2.$$

Denote $\phi_{n,i} = \log L_{n,i}$. Then

$$\phi_{n,i} = -\frac{l_n^2}{2} + \frac{l_n^2 y_i^2}{2} - \frac{l_n^4 y_i^4}{12} + O_p(l_n^6).$$

Consequently,

$$\frac{\log L_n + nl_n^4/4}{\sqrt{n/2} l_n^2} = \frac{\sum_1^n [(y_i^2 - 1)l_n^2/2 - l_n^4(y_i^4 - 3)/12]}{\sqrt{n/2} l_n^2} + O_p(\sqrt{n} l_n^4).$$

By invoking the central limit theorem for the i.i.d. case, we conclude that

$$\frac{\log L_n + nl_n^4/4}{\sqrt{n/2} l_n^2} \rightarrow_L N(0, 1)$$

under H_0 . Note that under H_1 ,

$$\begin{aligned} \frac{\log L_n - nl_n^4/4}{\sqrt{n/2} l_n^2} &= \frac{\sum_1^n [(y_i^2 - 1 - l_n^2)/2 - l_n^2(y_i^4 - Ey_i^4)/12] - nl_n^4(2 + l_n^2)}{\sqrt{n/2}} \\ &+ O_p(\sqrt{n} l_n^4). \end{aligned}$$

Now under H_1 ,

$$nE(|y_i^2 - 1 - l_n^2|/\sqrt{n})^4 = O(n^{-1}) \quad \text{and} \quad nE(|y_i^4 - Ey_i^4|/\sqrt{n})^4 = O(n^{-1}).$$

Hence, the Lyapounov's condition holds for the triangular arrays. By triangular array central limit theorem, under H_1 ,

$$\frac{\log L_n - nl_n^4/4}{\sqrt{n/2} l_n^2} \rightarrow_L N(0, 1).$$

Consequently, the sum of type I and type II errors of the best testing procedure is

$$P_{H_0}[L_n > 1] + P_{H_1}[L_n \leq 1] = 2\Phi\left(-\frac{\sqrt{n} l_n^2}{\sqrt{8}}\right)(1 + o(1)). \quad \square$$

PROOF OF PROPOSITION 1. For any subset $S \subset \{1, 2, \dots\}$, let prior μ^S be the probability measure of independently assigning $x_j = \pm A_j$ with probability $\frac{1}{2}$ each, for $j \in S$ and assigning probability 1 to the point $x_j = 0$ for $j \notin S$. Then by Jensen's inequality and (3.2),

$$(7.2) \quad \max_{x \in \Sigma} R(B, x) \geq \max_s \left\{ \left(E_{\mu^S}(x' B x) + \sigma^2 \operatorname{tr} B + c - E_{\mu^S}[Q(x)] \right)^2 + 2\sigma^4 \operatorname{tr} B^2 + 4\sigma^2 E_{\mu^S}(x' B^2 x) \right\},$$

where $\operatorname{tr} A$ is the trace of a matrix A . Let $D_A = E_{\mu^S}(x x')$, which is a diagonal matrix. Simple calculation shows that

$$\begin{aligned} E_{\mu^S}(x' B x) &= \operatorname{tr}(B D_A) = \operatorname{tr}(D_B D_A), \\ E_{\mu^S}(x' B^2 x) &= \operatorname{tr}(B^2 D_A) \geq \operatorname{tr}(D_B^2 D_A), \\ \operatorname{tr} B^2 &= \operatorname{tr} B' B \geq \operatorname{tr} D_B^2. \end{aligned}$$

Thus by (7.2) and the last three displays,

$$\max_{x \in \Sigma} R(B, x) \geq \max_S E_{\mu^S} R(D_B, x) = \max_{x \in \Sigma} R(D_B, x).$$

The last equality holds because (3.3) is convex in x_j^2 and consequently attains its maximum at either $x_j^2 = 0$ or $x_j^2 = A_j^2$. \square

PROOF OF THEOREM 2. We will prove that the estimator (3.4) with $m = n_{\sigma, d}$ achieves the rate given by (3.6).

Note that n_{σ} increases to infinity as σ decreases to 0. By the assumptions, the right-hand side of (3.5) is

$$(7.3) \quad O \left(\left(\sum_1^m \lambda_j \right)^2 (A_m^4 + \sigma^4/m) \right) + 4\sigma^2 \sum_1^m \lambda_j^2 A_j^2.$$

Taking $m = n_{\sigma, d}$, by (3.7) we have

$$\sigma^4 / (n_{\sigma, d} - 1) \leq A_{n_{\sigma, d} - 1}^4 / d^2 = O(A_{n_{\sigma}}^4).$$

Thus for $m = n_{\sigma, d}$, (7.3) becomes

$$(7.4) \quad O \left(\left(\sum_1^{n_{\sigma, d}} \lambda_j \right)^2 A_{n_{\sigma, d}}^4 \right) + 4\sigma^2 \sum_1^{n_{\sigma, d}} \lambda_j^2 A_j^2.$$

CASE 1. If $\limsup_{\sigma \rightarrow 0} n_{\sigma, d} \lambda_{n_{\sigma, d}}^2 \sigma^2 < \infty$, then by (3.7) and Assumption A(i),

$$\left(\sum_1^{n_{\sigma, d}} \lambda_j \right)^2 A_{n_{\sigma, d}}^4 = O(n_{\sigma, d} \lambda_{n_{\sigma, d}}^2 \sigma^4) = O(\sigma^2).$$

To prove (3.6), we need only to show that $\sum_1^{\infty} \lambda_j^2 A_j^2 < \infty$.

Note that $\limsup_{\sigma \rightarrow 0} n_{\sigma,d} \lambda_{n_{\sigma,d}}^2 \sigma^2 < \infty$ implies that there exist constants σ_0 and D (fixed) such that

$$(7.5) \quad n_{\sigma}^{1.5} \lambda_{n_{\sigma}}^2 A_{n_{\sigma}}^2 \leq D, \quad \text{when } \sigma \leq \sigma_0.$$

By Assumption A(i), it can be shown that as σ decreases from σ_0 to 0, $n_{\sigma,d}$ should increase from $n_{\sigma_0,d}$ to ∞ consecutively. Thus (7.5) implies that $\lambda_j^2 A_j^2 \leq Dj^{-1.5}$, when $j \geq n_{\sigma_0,d}$. Consequently, (7.4) is of order $O(\sigma^2)$. By Theorem 5, the rate is the optimal.

CASE 2. If $\limsup_{\sigma \rightarrow 0} n_{\sigma,d} \lambda_{n_{\sigma,d}}^2 \sigma^2 = \infty$, then $\limsup_{n \rightarrow \infty} n^{1.5} \lambda_n^2 A_n^2 = \infty$. Thus, by Assumption A(ii)

$$\limsup_{\sigma \rightarrow \infty} \frac{\sigma^2 \sum_1^{n_{\sigma,d}} \lambda_j^2 A_j^2}{\left(\sum_1^{n_{\sigma,d}} \lambda_j\right)^2 A_{n_{\sigma}}^4} \leq O(1) \limsup_{\sigma \rightarrow \infty} \frac{\sqrt{n_{\sigma,d}} \sum_1^{n_{\sigma,d}} \lambda_j^2 A_j^2}{n_{\sigma,d}^2 \lambda_{n_{\sigma,d}}^2 A_{n_{\sigma,d}}^2} < \infty.$$

Hence, (7.4) is bounded by its first term. Consequently, the truncated estimator (3.4) with $m = n_{\sigma,d}$ achieves the rate given by (3.6), which will be further justified to be optimal by Theorem 3. \square

PROOF OF THEOREM 3. Note that $\{A_n\}$ is a decreasing sequence and $l_n(\Sigma) = A_n$. The r_n defined by (2.6) is $A_n^2 \sum_1^n \lambda_j / 2$. Thus, by Theorem 1 [see (2.7)],

$$\sup_{x \in \Sigma} E_x(\delta(y) - T(x))^2 \geq (\Phi(-d/\sqrt{8}) + o(1)) A_{n_{\sigma,d}}^4 \left(\sum_1^{n_{\sigma,d}} \lambda_j\right)^2 / 4. \quad \square$$

PROOF OF THEOREM 4. First we prove the truncated estimator (3.4) achieves the rate given by (4.3). We treat the case $p = 2$ and $p > 2$ separately. When $p = 2$, maximum risk of the truncated estimator (3.4) is

$$(7.6) \quad \max_{x \in \Sigma_2} R(q_T, x) \leq C^2(\lambda_m/\delta_m)^2 + 2 \sum_1^m \lambda_j^2 \sigma^4 + 4C\sigma^2 \max_{1 \leq j \leq m} \lambda_j^2/\delta_j.$$

Take $m = n_{\sigma,d}$. Note that as $\sigma \rightarrow 0$, $n_{\sigma,d} \rightarrow \infty$. By (4.4) and Assumption B(ii), there exists a constant σ_0 such that when $\sigma \leq \sigma_0$,

$$(7.7) \quad \delta_{n_{\sigma,d}}^2 > c\delta_{n_{\sigma,d}+1}^2 \geq \frac{dc}{(n_{\sigma,d} + 1)\sigma^4}.$$

By (7.7),

$$(\lambda_{n_{\sigma,d}}/\delta_{n_{\sigma,d}})^2 \leq O(n_{\sigma,d} \lambda_{n_{\sigma,d}}^2 \sigma^4).$$

Consequently, by (7.6) and Assumption B(i), we have

$$(7.8) \quad \max_{x \in \Sigma_2} R(q_T, x) = O(n_{\sigma,d} \lambda_{n_{\sigma,d}}^2 \sigma^4) + O\left(\sigma^2 \max_{1 \leq j \leq n_{\sigma,d}} \lambda_j^2/\delta_j\right).$$

CASE 1. If $\limsup_{n \rightarrow \infty} n\lambda_n^4/\delta_n^2 < \infty$, then the sequence $\{\lambda_n^2/\delta_n\}$ stays bounded and by (4.4) and (7.8),

$$\max_{x \in \Sigma_{2^*}} R(q_T, x) = O(\sigma^2).$$

Hence, the rate $O(\sigma^2)$ is the attainable one.

CASE 2. If $\limsup_{n \rightarrow \infty} n\lambda_n^4/\delta_n^2 = \infty$, then $\sigma^2 = o(n_{\sigma,d}\lambda_{n_{\sigma,d}}^2\sigma^4)$. If the sequence $\{\lambda_n^2/\delta_n\}$ stays bounded, then (7.8) is of order $O(n_{\sigma,d}\lambda_{n_{\sigma,d}}^2\sigma^4)$. Otherwise, by Assumption C(ii) and (7.7),

$$\limsup_{\sigma \rightarrow 0} \frac{\max_{1 \leq j \leq n_{\sigma,d}} \lambda_j^2/\delta_j}{n_{\sigma,d}\lambda_{n_{\sigma,d}}^2\sigma^2} \leq O(1) \limsup_{\sigma \rightarrow 0} \frac{1}{\sqrt{n_{\sigma,d}}\delta_{n_{\sigma,d}}\sigma^2} < \infty.$$

Thus, we conclude that the truncated estimator $q_T(y)$ achieves the rate $n_{\sigma,d}\lambda_{n_{\sigma,d}}^2\sigma^4$.

Now, let us consider the case $p > 2$. The maximum risk of the truncated estimator is given by

$$\max_{x \in \Sigma_p} R(q_T, x) \leq \max_{x \in \Sigma_p} \left(\sum_m^\infty \lambda_j x_j^2 \right)^2 + \max_{x \in \Sigma_p} \sum_1^m \lambda_j^2 (2\sigma^4 + 4\sigma^2 x_j^2).$$

Let $q = p/(p - 2)$ be the conjugate number of $p/2$. Then by Hölder's inequality, we have for any $x \in \Sigma_p$,

$$\begin{aligned} \sum_m^\infty \lambda_j x_j^2 &\leq \left(\sum_m^\infty (\lambda_j \delta_j^{-2/p})^q \right)^{1/q} \left(\sum_m^\infty \delta_j |x_j|^p \right)^{2/p} \\ (7.9) \qquad &\leq C^{2/p} \left(\sum_m^\infty \lambda_j^{p/(p-2)} \delta_j^{-2/(p-2)} \right)^{(p-2)/p} \end{aligned}$$

Similarly, we have

$$(7.10) \qquad \sum_1^m \lambda_j^2 x_j^2 \leq C^{2/p} \left(\sum_1^m \lambda_j^{2p/(p-2)} \delta_j^{-2/(p-2)} \right)^{(p-2)/p}$$

By (7.9) and (7.10),

$$\begin{aligned} \max_{x \in \Sigma_p} R(q_T, x) &\leq C^{4/p} \left(\sum_m^\infty \lambda_j^{p/(p-2)} \delta_j^{-2/(p-2)} \right)^{2(p-2)/p} \\ &\quad + 2 \sum_1^m \lambda_j^2 \sigma^4 + 4\sigma^2 C^{2/p} \left(\sum_1^m \lambda_j^{2p/(p-2)} \delta_j^{-2/(p-2)} \right)^{(p-2)/p} \\ (7.11) \qquad &= O \left(\lambda_m^2 \delta_m^{-4/p} m^{2(p-2)/p} + m \lambda_m^2 \sigma^4 \right. \\ &\quad \left. + \sigma^2 \left(\sum_1^m \lambda_j^{2p/(p-2)} \delta_j^{-2p/(p-2)} \right)^{(p-2)/p} \right), \end{aligned}$$

by Assumption C(i') and Assumption B(i). Now, by taking $m = n_{\sigma,d}$ and using the fact that [see (4.4)]

$$(7.12) \quad \delta_{n_{\sigma,d}+1}^{-4/p} < \sigma_*^4 (n_{\sigma,\alpha+1})^{-(p-4)/p} / d,$$

(7.11) is of order

$$(7.13) \quad O \left(n_{\sigma,d} \lambda_{n_{\sigma,d}}^2 \sigma^4 + \sigma^2 \left(\sum_1^{n_{\sigma,d}} \lambda_j^{2p/(p-2)} \delta_j^{-2/(p-2)} \right)^{(p-2)/p} \right).$$

CASE 1. If $\limsup_{n \rightarrow \infty} n^{(3p-4)/(2p)} \lambda_n^2 \delta_n^{-2/p} > \infty$, then it is easy to show that $\lambda_n^{2p/(p-2)} \delta_n^{-2/(p-2)} = O(n^{-(3p-4)/2(p-2)}) = o(n^{-1.5})$.

Consequently,

$$\sum_1^\infty \lambda_n^{2p/(p-2)} \delta_n^{-2/(p-2)} < \infty$$

and (7.13) is of order $O(\sigma^2)$.

CASE 2. If $\limsup_{n \rightarrow \infty} n^{(3p-4)/(2p)} \lambda_n^2 \delta_n^{-2/p} = \infty$, then by Assumption C(ii')

$$\left(\sum_1^{n_{\sigma,d}} \lambda_j^{2p/(p-2)} \delta_j^{-2/(p-2)} \right)^{(p-2)/p} \leq O \left(\lambda_{n_{\sigma}}^2 \delta_{n_{\sigma}}^{-2/p} n_{\sigma}^{(3p-4)/(2p)} \right) = O \left(n_{\sigma} \lambda_{n_{\sigma}}^2 \sigma^2 \right).$$

Hence, (7.13) is of order $O(n_{\sigma} \lambda_{n_{\sigma}}^2 \sigma^4)$. The truncated estimator achieves the rate given by (4.3).

Now, let us show that the rate of convergence is the optimal one. Note that the n -dimensional inner length $l_n(\Sigma_p) = (C/\sum_1^n \delta_j)^{1/p}$. For the $n_{\sigma,d}$ defined by (4.4), we have

$$\sqrt{n_{\sigma}} (l_{n_{\sigma,d}}(\Sigma_p) / \sigma)^2 \leq D,$$

where $D > 0$. Thus, by Theorem 1 for any estimator $T(y)$,

$$\sup_{x \in \Sigma_p} E_x(T(y) - Q(x))^2 \geq \Phi(-D/\sqrt{8}) r_{n_{\sigma,d}}^2,$$

where

$$r_{n_{\sigma,d}} = \sum_1^{n_{\sigma,d}} \lambda_j (l_{n_{\sigma,d}}(\Sigma_p))^2 / 2 \geq a \sqrt{n_{\sigma,d}} \lambda_{n_{\sigma,d}} \sigma^2$$

and $a > 0$. Thus, the conclusion follows. \square

Acknowledgments. This work is part of the author's Ph.D. dissertation at the University of California, Berkeley, written under the supervision of Professors D. L. Donoho and P. J. Bickel, whose generous guidance and suggestions are gratefully acknowledged and appreciated. Many thanks to

Professor L. Le Cam for his valuable remarks and discussions. The author would like to thank the referees for their helpful comments and suggestions that improved the presentation of the paper.

REFERENCES

- BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393.
- CHENTSOV, N. N. (1980). *Statistical Decision Rules and Optimal Inference*. Amer. Math. Soc., Providence, R. I.
- DONOHO, D. L. and LIU, R. C. (1987a). On minimax estimation of linear functionals. Technical Report 105, Dept. Statistics, Univ. California, Berkeley.
- DONOHO, D. L. and LIU, R. C. (1987b). Geometrizing rates of convergence, I. Technical Report 137a, Dept. Statistics, Univ. California, Berkeley.
- DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk for hyperrectangles, and implications. *Ann. Statist.* **18** 1416–1437.
- EFROIMOVICH, S. Y. and PINSKER, M. S. (1982). Estimation of square-integrable probability density of a random variable. *Problemy Peredachi Informatsii* **18** 19–38 (in Russian); *Problems Inform. Transmission* (1983) 175–189.
- FAN, J. (1989a). Contributions to the estimation of nonregular functionals. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.
- FAN, J. (1989b). On the estimation of quadratic functionals. Inst. Statistics, Mimeo Series 2005, Univ. North Carolina, Chapel Hill.
- 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.
- HALL, P. and MARRON, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **6** 109–115.
- IBRAGIMOV, I. A., NEMIROVSKII, A. S. and KHAS'MINSKII, R. Z. (1986). Some problems on nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31** 391–406.
- KHAS'MINSKII, R. Z. (1979). Lower bound for the risks of nonparametric estimate of the mode. In *Contributions to Statistics: Jaroslav Hájek Memorial Volume* (J. Jureckova, ed.) 91–97. Academia, Prague.
- LE CAM, L. (1985). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.
- NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2 . *Ann. Statist.* **13** 984–997.
- PINSKER, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16** 52–68.
- SACKS, J. and YLVIKAKER, D. (1981). Variance estimation for approximately linear models. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 147–162.
- SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970–983.
- STONE, C. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

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