On the Estimation of Quadratic Functionals

Jianqing Fan

Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

ABSTRACT

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 $\sigma \longrightarrow 0$) for estimating quadratic functionals under certain geometric constraints are found. Specially, the optimal rates of estimating $\int_0^1 [f^{(k)}(x)]^2\,dx$ under hyperrectangular constraints $\Sigma = \{f : x_j(f) \le Cj^{-p}\}$ and weighted l_p-body constraints $\Sigma_p = \{f : \sum_1^\infty j^r |x_j(f)|^p \le C\}$ are computed explicitly, where $x_j(f)$ is the jth Fourier-Bessel coefficient of the unknown function f. We invent a new method for developing lower bounds based on testing two highly composite hypercubes, and address its advantages. The attainable lower bounds are found by applying the hardest 1-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 infinite as $\sigma \longrightarrow 0$) hypercube subproblems.

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

Key Words and Phrases. Cubical subproblems, quadratic functionals, lower bounds, rates of convergence, difficulty of estimation, minimax risk, Gaussian white noise, estimation of bounded squared-mean, testing of hypotheses, hardest 1-dimensional subproblem, hyperrectangle, hypersphere, weighted l_p-body, hypercube.

1. Introduction

The problem of estimating a quadratic functional was considered by Bickel and Ritov (1988), Hall and Marron (1987b), and Ibragimov *et al* (1987). 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 difficulty of estimating a *nonlinear* functional nonparametrically? The problem itself is poorly understood and the pioneering works show 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

$$y_j = x_j + z_j, \tag{1.1}$$

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

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

with some geometric constraint Σ. The geometric shapes of Σ we will use are either a hyperrectangle

$$\Sigma = \{x : |x_j| \leq A_j\}, \tag{1.3}$$

or a weighted l_p-body

$$\Sigma = \{x : \sum_1^{\infty} \delta_j |x_j|^p \leq C\}. \tag{1.4}$$

These are two interesting geometric shapes of constraints, which appear quite often in the literature of non-parametric estimation (Donoho *el al* (1988), Efroimovich and Pinsker (1982), Parzen (1971), Pinsker (1980), Prakasa Rao (1983), etc.). The connections of such geometric constraints with the usual constraints on the bounded derivatives will be discussed in section 4.

An interesting feature of our study is the use of geometric idea, including hypercube subproblem, inner length, and hardest hyperrectangle subproblem. We use the difficulty of a hypercube subproblem to develop a lower bound. 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 hypercube subproblem. We compare the hypercube bound with the minimax risk of a truncated quadratic estimator, and show that the ratio of the lower bound and the upper bound is bounded away from 0. Thus, in minimax theory at least, there is little to be gained by nonquadratic procedures, and hence, consider quadratic estimators are good enough for estimating a quadratic functional.

A related approach to ours is the hypersphere method developed by Ibragimov *et al* (1987). The notion of their method is to use the difficulty of a hypersphere subproblem as that 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 symmetric functionals (see (2.1)) and other shapes of constraints to get attainable lower rates? Unforetunely, the answer is "No". We show in section 6 that the hypersphere method can not give attainable lower rates of convergence for some other kind of constraints (e.g. hyperrectangles) and some other kind of symmetric 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 our 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. Our arguments also indicate that the hypercube method has potential applications to some other symmetric functionals, as the value of a symmetric functional remains the same on the vertices of a hypercube.

Comparing our approach to the traditional approach of measuring the difficulty of a linear functional (see Donoho and Liu (1987 a, c, 1988), Fan (1989), Farrell (1972), Hall and Marron (1987a), Khas'minskii (1979), Stone (1980), Wahba (1975) and many others), the hypercube method uses the *difficulty of an* n_σ*-dimensional* ($n_\sigma \to \infty$) *subproblem, instead of 1-dimensional,* as the difficulty of the full nonparametric problem. It has been shown that for estimating *a linear functional,* the difficulty of a 1-dimensional subproblem can capture the difficulty of a full problem with great generality. However, totally phenomena occur if we are trying to estimate a quadratic functional. The difficulty of the hardest 1-dimensional subproblem can *only* capture the difficulty of a full non-parametric problem *for the regular cases* (the case that the regular rate can be achieved). For nonregular cases, the hardest 1-dimensional subproblem *can not capture* the difficulty of the full problem. Thus, any 1-dimensional based methods *fail* to give an attainable rate of convergence. The discrepancy is, however, resolved by using multi-dimensionally based hypercube method. Our hypercube method indicates that the difficulty of the full problem for a nonregular case is captured at an n_σ-dimensional subproblem.

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 *et al* (1988), Ibragimov *et al* (1987), Efroimovich and Pinsker (1982), Nussbaum (1985). Suppose we are interesting in estimating

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

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

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

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

Let us assume that $[a, b] = [0, 1]$. 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

$$T(f) = (2\pi)^{2k} \sum_{j=2}^{\infty} j^{2k} x_j^2 + \delta_k x_1^2, \tag{1.7}$$

and the model (1.6) is equivalent to

$$y_j = x_j + \sigma z_j, \tag{1.8}$$

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 a smoothness *priori* that the Fourier-Bessel coefficients of f decay rapidly:

$$|x_j| \le A_j,\ A_j \to 0, \text{ if } j \to \infty.$$

Then, the problem reduces to (1.1)-(1.3). Specially, if the $(p - 1)$ derivatives of f are bounded, and these derivatives satisfy periodic boundary conditions at 0, and 1. Then, $|x_j| \le Cj^{-p}$ for some C. Thus, $A_j = Cj^{-p}$ is a weakening condition that f have $(p - 1)$ bounded derivatives.

If a *priori* smoothness condition is $\Sigma = \{f : \int_0^1 [f^{(k)}(t)]^2 dt \le C\}$. Then, by Parseval's identity, Σ is an ellipsoid $\Sigma = \{x: \sum_{j=2}^{\infty} j^{2k} x_j^2 \le C/(2\pi)^{2k}\}$. Thus, we reduce the problem to (1.1), (1.2) with a constraint (1.4).

The white-noise model (1.6) is closely related to the problems of density estimation, and spectral density. It should be no surprise that the results allow one to attack certain asymptotic minimax problems for estimating the asymptotic variance of a R-estimate (Bickel and Ritov (1988), Hall and Marron (1987b)), and the asymptotic variance of similar problem in time series, and even some problems in bandwidth selection (Hall and Marron (1987a, 1987b)). Other comments on the applications of the white noise model (1.6) can be found in Donoho *et* *al* (1988).

Even though we discuss the possible applications on a bounded interval [0, 1], the notion above can be easily extended to an unbounded interval.

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 $\sigma \to 0$ is important, which is natural when we make connections with density estimation.

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, e.g. ellipsoids, l_p-bodies. In section 5, we demonstrate that our hypercube method is a better technique than the hypersphere method of Ibragimov *et al* (1987). In section 6, we give some further remarks to show that the hypercube method is strictly better than hypersphere method. Some comments are further discussed in section 7. Technical proofs are given in section 8.

2. The hypercube bound

Let's introduce some terminologies. Suppose that we want to estimate a functional $T(x)$ under a constraint $x \in \Sigma \subset R^{\infty}$. Let $\Sigma_0 \subset \Sigma$. We call estimating $T(x)$ on Σ_0 as a subproblem of the estimation, and estimating $T(x)$ on Σ as a full problem of the estimation. 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 symmetric functional $T(x)$, i.e.

$$T(\pm x_1, \pm x_2, \cdots) = T(x_1, , x_2, \cdots), \tag{2.1}$$

based on the observations (1.1) under a geometric constraint Σ. Assume without loss of generality that $T(0) = 0$, and $0 \in \Sigma$. Let $l_n(\Sigma)$ be the supremum of the half lengths of all n-dimensional hypercubes centered at the origin lying in Σ (Figure 1). We call it the n-dimensional inner length of Σ.

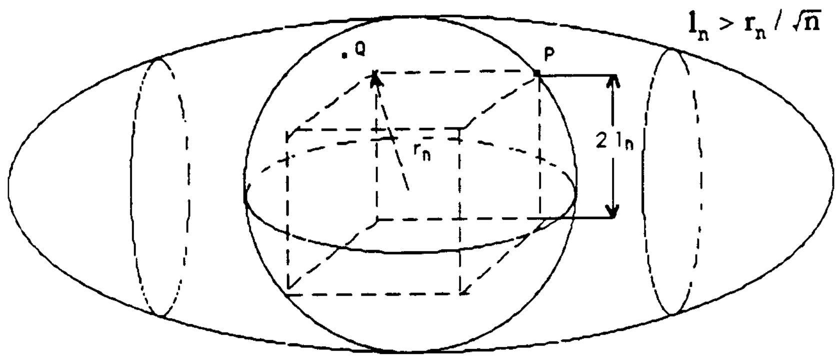

Figure 1. Testing a point P uniformly on the hypercube is as difficulty as testing a point Q uniformly on the sphere.

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 $\frac{1}{2^n}$ 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. The approach we use is related to the one of Ibragimov *et al* (1987), who, however use a hypersphere rather than a hypercube.

To carry out the idea, we formulate a testing problem

$$H_0: \ x_i = 0 \ (i = 1, \cdots, n) \ \longleftrightarrow H_1: \ x_i = \pm \, l_n(\Sigma) \ (i = 1, \cdots, n) \tag{2.2}$$

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

$$H_0: y_i \sim N(0, \sigma^2)\ (i = 1, \cdots, n) \Longleftrightarrow \tag{2.3}$$

$$H_1: y_i \sim \frac{1}{2}[\phi(y, l_n, \sigma) + \phi(y, -l_n, \sigma)],\ (i = 1, \cdots, n),$$

where $\phi(y, t, \sigma)$ is the density function of $N(t, \sigma^2)$, and $l_n = l_n(\Sigma)$. The result of the testing problem can be summarized as follows:

Lemma 1. *The sum of type I and type II errors of the best testing procedure is*

$$2\,\Phi(-\frac{\sqrt{n}\,(l_n / \sigma)^2}{\sqrt{8}})(1 + o(1)), \tag{2.4}$$

if $n^{1/2}(l_n / \sigma)^2 \to c$ *(as* $\sigma \to 0$*), where* $\Phi(\cdot)$ *is the standard normal distribution function.*

Choose the dimension of the hypercube n_σ to be the smallest integer satisfying

$$\sqrt{n}\ [l_n(\Sigma)]^2/\sigma^2 \le d,\ (d > 0). \tag{2.5}$$

By Lemma 1, there is no perfect test for problem (2.2), i.e. (as $\sigma \to 0$)

$$\min_{0 \le \phi(y) \le 1} \{E_0\, \phi(y) + E_1\,(1 - \phi(y))\} \ge 2\Phi(-d/\sqrt{8})(1 + o(1)) \tag{2.6}$$

(the sum of the type I and type II errors is bounded away from 0), where E_0, and E_1 mean that take the expectation of 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

$$r_n = |T(H_1) - T(H_0)|/2 = |T(x_n)|/2 \tag{2.7}$$

be the half of the difference of the value of $T(x)$ on the vertices from that of the origin, where $x_n = (l_n(\Sigma), \cdots, l_n(\Sigma), 0, 0, \cdots)$ is a vertex of the hypercube.

Theorem 1. *Suppose that* $T(x)$ *is a symmetric function with* $T(0) = 0$, *and* $0 \in \Sigma$. *For any estimator* $\delta(y)$ *based on the observations (1.1),*

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

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

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

where n_σ *is the smallest integer* n *satisfying (2.5).*

Proof. By (2.6) and (2.7), for any estimator δ,

$$\sup_{x \in \Sigma} P_x\{\, |\delta(y) - T(x)| \ge r_{n_\sigma} \}$$

$$\ge \frac{1}{2}\{P_0(|\delta(y)| \ge r_{n_\sigma}) + P_1(|\delta(y) - T(x_{n_\sigma})| \ge r_{n_\sigma})\}$$

$$\ge \frac{1}{2}\{P_0(|\delta(y)| \ge r_{n_\sigma}) + P_1(|\delta(y)| \le r_{n_\sigma})\}$$

$$\ge \Phi(-\frac{d}{\sqrt{8}}) + o(1)\ ,$$

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

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

The second conclusion follows.

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

$$\sup_{x \in \Sigma} E_x(\delta(y) - T(x))^2 \ge \Phi(-\frac{d}{\sqrt{8}}) r_{n_\sigma}^2 (1 + o(1)). \tag{2.8}$$

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

Corollary 1. *Suppose that* $T(x)$ *is symmetric and*

$$\liminf_{n \to \infty} |T(l_n(\Sigma), \cdots, l_n(\Sigma), 0, 0, \cdots) - T(0)| > 0.$$

If $\liminf_n \sqrt{n}\,(l_n(\Sigma))^2 > 0$, *no uniform consistent estimator exists for* $T(x)$ *on the basis of the observations (1.1).*

For any regular symmetric functionals, in order to have a uniformly consistent estimator, the inner length $l_n(\Sigma)$, a geometric quantity, must be of order $o(n^{-\frac{1}{4}})$. Thus, we give a geometrical explanation when there is no uniformly consistent estimator exists.

3. Truncation estimators

Let us start with the model (1.1). Suppose we observe

$$y = x + \sigma z \tag{3.1}$$

with a hyperrectangular type of constraint (1.3). An intuitive class of quadratic estimators to estimate

$$Q(x) = \sum_1^\infty \lambda_i \, x_i^2, \tag{3.2}$$

(where $\lambda_j \geq 0$) is the the class of estimators defined by

$$q_B(y) = y'By + c, \tag{3.3}$$

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

$$R(B, x) \triangleq E_x\,(q_B(y) - Q(x))^2 \tag{3.4}$$

$$= (x'Bx + \sigma^2\, trB + c - Q(x))^2 + 2\sigma^4\, trB^2 + 4\sigma^2\, x'B^2x. \tag{3.5}$$

The following proposition tells us that the class of quadratic estimators with diagonal matrices is a complete class among all estimators defined by (3.3).

Proposition 1. *Let* D_B *be a diagonal matrix, whose diagonal elements are those of* B*. Then for each symmetric* 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 only diagonal matrices are needed to be considered. For a diagonal matrix

$$B = diag(b_1, b_2, \cdots), \tag{3.6}$$

the estimator (3.3) has risk

$$R(B, x) = (\sum_1^\infty b_j x_j^2 + \sigma^2 \sum_1^\infty b_j + c - \sum_1^\infty \lambda_j x_j^2)^2 + \sum_1^\infty b_j^2 (2\sigma^4 + 4\sigma^2 x_j^2). \tag{3.7}$$

A natural question is when the estimator (3.3) with B defined by (3.6) converges almost surely.

Proposition 2. *Under model (3.1), $q_B(y)$ converges almost surely for each $x \in \Sigma$ iff*

$$\sum_1^\infty b_i (A_i^2 + \sigma^2) < \infty. \tag{3.8}$$

Even for the diagonal matrices, it is hard to find the exactly optimal quadratic estimator (see Sacks and Ylvisaker (1981)). For the infinite dimensional estimation problem, usually the bias is a major contribution to the risk. 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 it does converge, it might contribute too much in variance term. Thus, we consider a truncated quadratic estimator

$$q_{UT}(y) = \sum_1^m \lambda_j (y_j^2 - \sigma^2), \tag{3.9}$$

and choose m to minimize its maximum MSE. For the estimator $q_{UT}(y)$, the maximum MSE is

$$\max_{x \in \Sigma} R(q_{UT}, x) = (\sum_m^\infty \lambda_j A_j^2)^2 + \sum_1^m \lambda_j^2 (2\sigma^4 + 4\sigma^2 A_j^2). \tag{3.10}$$

Our main result of estimating quadratic functionals under hyperrectangular constraints can be summarized as follows. The lower bound will be established in Theorem 3 below.

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 $n \longrightarrow \infty$, and

$$\limsup_{n \longrightarrow \infty} \frac{A_{n-1}}{A_n} < \infty, \quad \sum_1^n \lambda_j \sim dn\lambda_n \ \text{ for some } d > 0.$$

ii) $\sum_n^\infty \lambda_j A_j^2 = O(n\lambda_n A_n^2)$, and if $\limsup_{n \longrightarrow \infty} n^{1.5}\lambda_n^2 A_n^2 = \infty$, 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 optimal rate ($\sigma \longrightarrow 0$) of estimating $Q(x) = \sum_1^\infty \lambda_j x_j^2$ under the quadratic loss with a hyperrectangular constraint (1.3) is*

$$O(\sigma^2 + (\sum_1^{n_\sigma}\lambda_j)^2 A_{n_\sigma}^4), \tag{3.11}$$

where n_σ is the smallest integer such that

$$\sqrt{n}\ (A_n/\sigma)^2 \le c, \tag{3.12}$$

for some $c > 0$. Moreover, the optimal rate is achieved by the truncated estimator (3.9) with $m = n_\sigma$.

When $\lambda_j = j^q$ and $A_j = Cj^{-p}$, $(p > (q+1)/2)$, the conditions of Theorem 2 are satisfied. In order to know how efficient the truncated estimator is, let's evaluate its MSE more carefully as $\sigma \to 0$. By (3.10), the estimator

$$\sum_1^m j^q (y_j^2 - \sigma^2) \tag{3.13}$$

has its maximum risk

$$R(m) \triangleq \frac{2m^{2q+1}}{2q+1}\sigma^4(1 + o(1)) + 4C^2\sigma^2 \sum_1^m j^{2q-2p} + C^4(\frac{m^{-(2p-q-1)}}{2p-q-1})^2(1 + o(1)).$$

When $(q+1)/2 < p \le q + 0.75$, simple algebra shows that the optimal m, which minimizes $R(m)$, is

$$m_0 = [(C^4/(2p - q - 1))^{\frac{1}{4p-1}} \sigma^{-\frac{4}{4p-1}}], \tag{3.14}$$

and the maximum risk of the optimal truncated estimator (3.13) with $m = m_0$ is

$$C^{4(2q+1)/(4p-1)} \left\{ \frac{2c_0^{2q+1}}{2q+1} + \frac{c_0^{-2(2p-q-1)}}{(2p-q-1)^2} + 4\sum_1^\infty j^{-1.5}\, \delta_{p,\, q+0.75} \right\} \sigma^{4 - \frac{4(2q+1)}{4p-1}} (1 + o(1)), \tag{3.15}$$

where $c_0 = (2p - q - 1)^{-\frac{1}{4p-1}}$, and $\delta_{p,\, q+0.75} = 1$, if $p = q + 0.75$, and $= 0$, otherwise.

When $p > q + 0.75$, the optimal $m_0 = d\sigma^{-\frac{4}{4p-1}}$ for a positive constant d, with the risk

$$4C^2 \sum_1^\infty j^{2q-2p} \sigma^2 (1 + o(1)). \tag{3.16}$$

In summary,

Corollary 2. *Suppose that $\lambda_j = j^q$ and $A_j = Cj^{-p}$, $(p > (q+1)/2)$. Then the best truncated estimator is given by (3.13) with $m = m_0$. The optimal m_0 is defined by (3.14) when $(q+1)/2 < p \le q + 0.75$ with the maximum risk given by (3.15), and the optimal $m_0 = d\sigma^{-\frac{4}{4p-1}}$ when $p > q + 0.75$ with the maximum risk given by (3.16). Moreover, the estimator achieves the optimal rate of convergence.*

When $p \ge q + 0.75$, the regular rate $O(\sigma^2)$ is achieved by the best truncated estimator, and hence the difficulty of the full problem of estimating $Q(x)$ is captured by a 1-dimensional subproblem. However, the situation changes when $p < q + 0.75$. The difficulty of the hardest 1-dimensional subproblem can not capture the difficulty of the full problem any more (Compare (7.2) with (3.15)). Thus, we need to establish a larger lower bound for this case by applying Theorem 1. By our method of construction, intuitively we need an n_σ-dimensional subproblem, not 1-dimensional subproblem, in order to capture the difficulty of the full problem for this case.

Theorem 3. *Suppose that $\sqrt{j}A_j^2$ is decreasing in j, when j is large. Let n_σ be the smallest integer such that*

$$\sqrt{n}\,(A_n/\sigma)^2 \le c.$$

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

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

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

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

When $A_j = Cj^{-p}$ and $\lambda_j = j^q$, we can calculate the rate in Theorem 3 explicitly.

Corollary 3. *When $A_j = Cj^{-p}$ and $\lambda_j = j^q$, for any estimator, the maximum risk under the quadratic loss is no smaller than*

$$(2q+2)^{-2} C^{4(2q+1)/(4p-1)} \xi_{p,q}\, \sigma^{4 - \frac{4(2q+1)}{4p-1}} (1 + o(1)). \tag{3.17}$$

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

$$\sup_{x \in \Sigma} P_x\left\{ |T(y) - Q(x)| \ge (2q+2)^{-1} (d\sigma^2/C^2)^{-\frac{2q+1}{4p-1}} d\sigma^2 \right\}$$

$$\ge \Phi\left(-\frac{d}{\sqrt{8}}\right)(1 + o(1)),$$

where

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

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

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 $\sum_1^{m_0} \lambda_j (y_j^2 - \sigma^2)$, where

$$m_0 = \begin{cases} O(\sigma^{-\frac{4}{4p-1}}), & \text{if } p > 0.75 \\ [(C^4/(2p-1))^{\frac{1}{4p-1}} \sigma^{-\frac{4}{4p-1}}], & \text{if } 0.5 < p \le 0.75 \end{cases}$$

Moreover, when $p > 0.75$, the estimator is an asymptotic minimax estimator (see (7.3)). For $0.5 < p \le 0.75$, 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

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

The estimator which achieves the optimal rate of convergence is $(2\pi)^{2k} \sum_{j=2}^{m_0} j^{2k} (y_j^2 - \sigma^2)$, where

$$m_0 = \begin{cases} O(\sigma^{-\frac{4}{4p-1}}), & \text{if } p > 2k + 0.75 \\ [(C^4/(2p-2k-1))^{\frac{1}{4p-1}} \sigma^{-\frac{4}{4p-1}}], & \text{if } k + 0.5 < p \le 2k + 0.75 \end{cases}$$

Moreover, the estimators achieve the optimal rates given by

$$\begin{cases} O(\sigma^2), & \text{if } p > 2k + 0.75 \\ O(\sigma^{4 - \frac{4(4k+1)}{4p-1}}), & \text{if } k + 0.5 < p \le 2k + 0.75 \end{cases} \qquad (3.18)$$

Example 3. (Inverse Problem) Suppose that we are interested in recovering the indirectly observed function $f(t)$ from data of the form

$$Y(u) = \int_0^u \int_0^s K(t, s) f(t) dt\, ds + \sigma \int_0^u dW, \quad u \in [0, 1],$$

where W is a Wiener process and K is known. Let $K: L_2[0, 1] \to L_2[0, 1]$ have a singular system decomposition (Bertero *et al* (1982)):

$$\int_0^u K(t, u) f(t) dt = \sum_{i=1}^{\infty} \lambda_i (f, \xi_i)\, \eta_i(u),$$

where the $\{\lambda_i\}$ are singular values, the $\{\xi_i\}$ and $\{\eta_i\}$ are orthogonal sets in $L_2[0, 1]$. Then the observations are equivalent to

$$y_i = \lambda_i \theta_i + \sigma\, \varepsilon_i, \; i = 1, 2, \cdots,$$

where $y_i = \int_0^1 \eta_i(u) dY(u)$ is the Fourier-Bessel coefficient of the observed data, $\theta_i = (f, \xi_i)$, and $\varepsilon_i = \int_0^1 \eta_i(u) dW(u)$ are i.i.d. $N(0, 1)$, $i = 1, 2, \cdots$. Now suppose that we want to estimate

$$\int_0^1 f^2(t)\, dt = \sum_1^{\infty} \theta_i^2 = \sum_1^{\infty} \lambda_i^{-2} x_i^2,$$

where $x_i = \lambda_i\, \theta_i$. If the non-parametric constraint on θ is a hyperrectangle, then the constraint on x is also a hyperrectangle in R^{∞}. Applying Theorem 2, we can get an optimal estimator in terms of rate of convergence. Moreover, we will know roughly how efficient the estimator is if we apply Table 7.1 and 7.2. If, instead, the constraint on θ is a weighted l_p-body defined by (1.4), then the constraint on x is also a weighted l_p-body, and we can apply the results in section 4 to get the best possible estimator in terms of rate of convergence.

4 Extension to quadratically convex sets

In this section, we will find the optimal rate of convergence for estimating the quadratic functional

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

under a geometric constraint

$$\Sigma_p = \{x : \sum_1^\infty \delta_j \; |x_j|^p \le C\}, \tag{4.2}$$

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 will make some general assumptions:

Assumption B.

i) The sequences $\{\lambda_n\}$, and $\{\delta_n\}$ are positive nondecreasing sequences.

ii) There exists a positive constant c such that

$$\sum_1^n \lambda_j > cn\lambda_n, \qquad \sum_1^n \delta_j > cn\delta_n \; .$$

Let's study the weighted l_2-bodies (ellipsoids) first. Consider the truncated estimator (3.9), whose risk under the quadratic loss is

$$R(q_{UT}, x) = (\sum_m^\infty \lambda_j x_j^2)^2 + \sum_1^m \lambda_j^2 (2\sigma^4 + 4\sigma^2 x_j^2). \tag{4.3}$$

Assumption C.

a) $\limsup_{n \to \infty} \sqrt{n}\,\lambda_n^2/\delta_n = \liminf_{n \to \infty} \sqrt{n}\,\lambda_n^2/\delta_n$.

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

c) $\{\lambda_j/\delta_j\}$ is a non-increasing sequence.

Theorem 4. *Under Assumption B & C, the optimal rate of convergence for estimating $Q(x)$ under the quadratic loss with an ellipsoid constraint Σ_2 defined by (4.2) is*

$$\begin{cases} O(\sigma^2), & \text{if } \limsup_{n \to \infty} n\lambda_n^4/\delta_n^2 < \infty \\ n_\sigma \lambda_{n_\sigma}^2 \sigma^4, & \text{if } \limsup_{n \to \infty} n\lambda_n^4/\delta_n^2 = \infty \end{cases} \quad (4.4)$$

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

$$n\delta_n^2\sigma^4 < d, \quad (4.5)$$

and d is a positive constant.

When $\lambda_j = j^q$ and $\delta_j = j^r$, the conditions of Theorem 4 are satisfied. We can compute the optimal rate explicitly. In this case, $n_\sigma = [d^{\frac{1}{2r+1}} \sigma^{-\frac{4}{2r+1}}]$, and $n_\sigma \lambda_{n_\sigma}^2 \sigma^4 = O(\sigma^{\frac{8(r-q)}{2r+1}})$, where $[a]$ denotes the largest integer which does not exceed a. In summary,

Corollary 4. *When $\lambda_j = j^q$ and $\delta_j = j^r$, the optimal rate of estimating $Q(x)$ under the weighted l_2-body constraint (4.2) is*

$$\begin{cases} O(\sigma^2), & \text{if } r \geq 2q + 0.5 \\ O(\sigma^{\frac{8(r-q)}{2r+1}}), & \text{if } 2q + 0.5 > r > q \end{cases} \quad (4.6)$$

Moreover, the truncated estimator $\sum_1^{n_\sigma} j^q \, (y_j^2 - \sigma^2)$ achieves the optimal rate of convergence, where $n_\sigma = [d^{\frac{1}{2r+1}} \sigma^{-\frac{4}{2r+1}}]$, d is a positive constant.

Example 4. (estimating integrated squared derivatives) Suppose that we want to estimate $T(f) = \int_0^1 [f^{(k)}(t)]^2 \, dt$ under the non-parametric constraint that

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

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

$$T(f) = (2\pi)^{2k} \left(\sum_{j=2}^{\infty} j^{2k} x_j^2 + \eta_k x_1^2 \right) ,$$

where $\eta_k = 1$, if $k = 0$, and $\eta_k = 0$, if $k \ne 0$, and x_j is the jth Fourier coefficient. Assume additionally that $|x_1| \le B$, a finite constant, when $k = 0$. The non-parametric constraint Σ can be rewritten as an ellipsoid

$$\Sigma = \{ x : \sum_{j=2}^{\infty} j^{2\alpha} x_j^2 \le \frac{C}{(2\pi)^{2\alpha}} \} .$$

By Corollary 4, the truncated estimator $(2\pi)^{2k} \left(\sum_2^{n_\sigma} j^{2k} (y_j^2 - \sigma^2) + \eta_k (y_1^2 - \sigma^2) \right)$ with

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

achieves the optimal rate of convergence given by

$$\begin{cases} O(\sigma^2), & \text{if } \alpha \ge 2k + 0.25 \\ \sigma^{\frac{16(\alpha - k)}{4\alpha + 1}}, & \text{if } 2k + 0.25 > \alpha > k \end{cases} . \tag{4.7}$$

Now, let's give the optimal rate for the weighted l_p-body constraints ($p > 2$).

Assumption D.

a) $\limsup_n n^{(3p-4)/(2p)} \lambda_n^2 \delta_n^{-2/p} = \liminf_n n^{(3p-4)/(2p)} \lambda_n^2 \delta_n^{-2/p}$.

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

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

d) $\delta_n^{4/p} n^{4/p - 1}$ increases to infinite as $n \longrightarrow \infty$.

Theorem 5. *Under the Assumption B & D, the optimal rate of convergence of estimating $Q(x)$ under the weight l_p-body constraint (4.2) is given by*

$$\begin{cases} O(\sigma^2), & \text{if } \limsup_n \ n^{(3p-4)/(2p)} \lambda_n^2 \delta_n^{-2/p} < \infty \\ n_\sigma \lambda_{n_\sigma}^2 \sigma^4, & \text{if } \limsup_n \ n^{(3p-4)/(2p)} \lambda_n^2 \delta_n^{-2/p} = \infty \end{cases} \qquad (4.8)$$

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

$$\delta_n^{4/p} n^{4/p - 1} \sigma^4 < d \qquad (4.9)$$

for some positive constant d.

Corollary 5. *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) ($p > 2$) is*

$$\begin{cases} O(\sigma^2), & \text{if } r \ge 0.75p - 1 + pq \\ O(\sigma^{\frac{8(2(r+1) - p(q+1))}{4(r+1) - p}}), & \text{if } p(q+1)/2 - 1 < r < 0.75p - 1 + pq \end{cases} \qquad (4.10)$$

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 $p \ge 2$, and is convex when $1 \le p < 2$, and is not convex when $0 < p < 1$ (*Donoho et al* (1988)). To understand the difficulty of estimation problem under some constraint, it is good to try to study such a kind of geometric constraint first. 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.

Remark 2. The constraints on the condition that the *kth* derivative is bounded, and the constraints that the Fourier coefficients fall in certain R^∞ set are both smoothness constraints. A simple connection is that

$$\{f(t): \int_0^1 [f^{(\alpha)}(t)]^2\, dt \le C\} = \{x: \sum_{j=2}^{\infty} j^{2\alpha} x_j^2 \le \frac{C}{(2\pi)^{2\alpha}}\}.$$

Thus, the optimal rates under the bounded derivative constraints should be the same as those under l_2-body constraints. Indeed, in density estimation setting, Bickel and Ritov (1988) give the optimal rate of convergence for estimating the functional discussed in Example 4 under certain constraint on the boundedness of derivatives of a density, and the optimal rate is precisely the same as (4.7) (with $\sigma = n^{-1/2}$). Thus, the special l_2-body constraints are the same as constraints on the boundedness of derivatives. Also, for estimating quadratic functionals, optimal rates under a hyperrectangular constraint $\{x: |x_j| \le Cj^{-p}\}$ and the optimal rates under a bounded α-derivative constraint agree when $p = \alpha + 0.5$. (compare (3.18) with Bickel and Ritov (1988) or (4.7)).

5. Comparison with Ibragimov-Nemirovskii-Khas'minskii

Our method of developing a lower bound is similar to that of Ibragimov *et al* (1987). Their method is based on testing the largest inner *sphere* instead of testing the vertices of a *hypercube*. Let's walk through the main steps of Ibragimov *et al*'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 such functionals remain the same on the sphere (see Remark 4). Furthermore, it is not hard to argue that *if the method of Ibragimov et al (1987) gives an attainable lower bound (sharp in rate) for some symmetric functionals under some geometric constraints, our method does in the same setting, and on the other hand, even though the method of Ibragimov et al (1987) can not give an attainable lower bound for some geometric constraints (e.g. hyperrectangle constraints, estimating a spherically symmetric functional* $T(x) = \sum_1^\infty x_j^2$; *see Remark 3) and for some symmetric functionals (e.g.* $T(x) = \sum_1^\infty j^q x_j^2$, $q \neq 0$; *see Remark 4), our method does. Therefore, it turns out that our method has much broader applications not only in the shapes of geometric constraints (e.g. hyperrectangle; see Remark 3) but also in the classes of symmetric functionals being estimated* (see Remark 4).

The argument of the above statement is as follows. Let $r_n(\Sigma)$ be the n-dimensional inner radius of a set Σ (Figure 1), namely, the supremum of the radii of all n-dimensional discs centered at 0 lying in Σ (see Ibragimov *et al* (1987) and Chentsov (1980)). Then it is easy to see that our n-dimensional inner length $l_n(\Sigma) \geq r_n(\Sigma)/\sqrt{n}$ because if one can inscribe an n-dimensional inner disc into Σ, then one can also inscribe an n-dimensional inner hypercube inside the disc (Figure 1). The key Lemma (Lemma 3.1) used by Ibragimov *et al* (1987) to develop a lower bound is that

Lemma I (Ibragimov *et al* (1987)). *Suppose we want to test the hypothesis:*

$$H_0\colon y_n \sim N(0, \sigma^2 I_n) \Longleftrightarrow H_1\colon y_n \sim N(x_n, \sigma^2 I_n), \tag{5.1}$$

with a prior of x_n uniform on the sphere $\{ \|x_n\| = r_n \}$.

Then the sum of the type I and type II errors of the best testing procedure for testing problem

(5.1) is

$$2\,\Phi(-\,r_n^2/(\sqrt{8n}\,\sigma^2))(1+o(1)),$$

if $\sigma/r_n \longrightarrow 0$ *(Note that* n *depends on* σ *in the current setting).*

Comparing Lemma I and the Lemma 1 of the present paper, we find out that *testing an* n*-dimensional sphere with the uniform prior against the origin is as difficult as testing the vertices (with uniform prior) of the largest inner hypercube of the sphere* (Figure 1), i.e. the sum of type I and type II errors of the best testing procedures for the both testing problems is asymptotically the same (as $r_n = \sqrt{n}\, l_n$). For simplicity of notations, assume without loss of generality that the largest n-dimensional hypersphere is attained at

$$S^n = \{x : x_1^2 + \cdots + x_n^2 = r_n^2(\Sigma),\ x_{n+1}=0,\ x_{n+2}=0,\ \cdots\},$$

and assume that we want to estimate a functional $T(x)$ (not necessary a symmetric functional) based on observations (1.1) with $T(0) = 0$. Let

$$C^n = \{(\pm r_n(\Sigma)/\sqrt{n},\ \cdots,\ \pm r_n(\Sigma)/\sqrt{n},\ 0,\ 0,\ \cdots)\}$$

be the vertices of the largest inner hypercube of S^n. Then, by Lemma 1 and Lemma I, choosing the dimension n_σ^* (the smallest one) such that

$$r_n^2/(\sqrt{n}\,\sigma^2) \le d, \tag{5.2}$$

the sum of type I and type II errors of testing problems (5.1) and (2.3) (with $l_n = r_n/\sqrt{n}$) is no smaller than $\Phi(-d/\sqrt{8})$.

Let $R_n^S = \inf_{x \in S^n} |T(x)|/2$ be the half difference between the functional values on S^n and the functional value on the origin. Similarly, let $R_n^C = \inf_{x \in C^n} |T(x)|/2$. Then according to the proof of Theorem 1, the lower bound of Ibragimov *et al* (1987) is that for any estimator $\delta(y)$ based on the observations (1.1),

$$\sup_{x \in \Sigma} P_x\{|\delta(y) - T(x)| \ge R_{n_\sigma^*}^S\} \ge \Phi(-d/\sqrt{8})(1+o(1)), \tag{5.3}$$

and in terms of MSE

$$\sup_{x \in \Sigma} E_x \mid \delta(y) - T(x) \mid^2 \geq \Phi(-d/\sqrt{8})(R^S_{n^*_\sigma})^2(1 + o(1)). \tag{5.4}$$

Our hypercubical lower bound is exactly the same as (5.3) and (5.4) except replacing $R^S_{n^*_\sigma}$ by $R^C_{n^*_\sigma}$. Note that $R^C_{n^*_\sigma} \geq R^S_{n^*_\sigma}$. The claim follows.

Note that we didn't use the fact that hypercubes are easier to inscribe into geometric regions than hyperspheres in the above argument; but use the fact that the difference of functional values on the sphere from the value of the origin is no larger than the difference of functional values on the vertices of the hypercube from the value of the origin, i.e. $R^S_n \leq R^C_n$. For some geometric constraints Σ (e.g. hyperrectangle), even though the functional may be spherically symmetric (e.g. $T(x) = \sum_1^\infty x_j^2$; the differences of functional values R^S_n and R^C_n defined above are the same), one can inscribe an n-dimensional inner hypercube with the inner length $l_n(\Sigma)$, but it is impossible to inscribe an n-dimensional inner sphere with radius $\sqrt{n}\ \ l_n(\Sigma)$ (Figure 2). Hence, the hypersphere method can not give an attainable lower bound in these cases (see Remark 3) even though the functional may be spherically symmetric.

$l_n > r_n / \sqrt{n}$

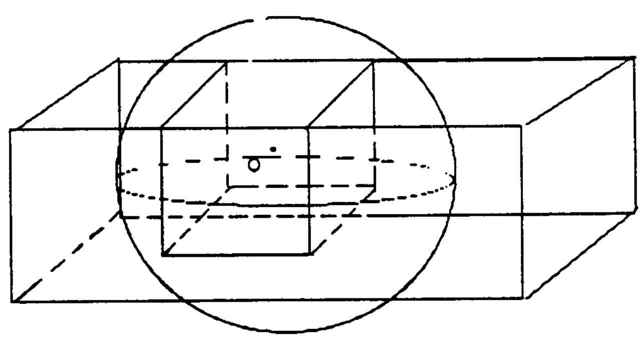

Figure 2. The hypercube is easier than to inscribe into the hypercube than a hypersphere. It is impossible to inscribe a hypersphere with radius $r_n = l_n \sqrt{n}$ inside the hyperrectangle.

In summary, the hypercube method is strictly better than the hypersphere approach. Comparing with the hypersphere method, the hypercube method has the following advantages: i) larger in difference of functional values ($R_n^C \geq R_n^S$; broader application in terms of estimating functionals),

ii) easier to inscribe ($l_n(\Sigma) \geq r_n(\Sigma)/\sqrt{n}$; broader application in terms of constraints).

Another advantage of our method is that it is easy to inscribe a hypercube into a nonsymmetric constraint as we only require the vertices of the hypercube lying in the constraint instead of the entire hypercube (see Example 5 below).

6. Further Remarks.

Remark 3: It appears that the hypersphere bound of Ibragimov *et al* (1987) would not give a lower bound of the same order as we are able to get via hypercubes for estimating the quadratic functional $Q(x) = \sum_1^\infty j^q x_j^2$ with the hyperrectangular constraint $\Sigma = \{x \in R^\infty: |x_j| \leq Cj^{-p}\}$. The proof is simple. Assume that $C = 1$. The n-dimensional inner radius of the hyperrectangle is $r_n(\Sigma) = n^{-p}$. By the Lemma 3.1 of Ibragimov *et al* (1987) (see Lemma I), the smallest dimension that we can not test the origin and the sphere with the uniform prior consistently is $n_\sigma^* = [(\sqrt{d}\,\sigma)^{-\frac{4}{4p+1}}]$ (see (5.2)), which is of small order of $n_\sigma = [(\sqrt{d}\,\sigma)^{-\frac{4}{4p-1}}]$, the dimension that we can not test the vertices of the hypercube against the origin perfectly, where d is a positive constant. By Corollary 6 of Ibragimov *et al* (1987), the lower bound under the quadratic loss is of order $(R_{n_\sigma^*}^S)^2 = O(\sigma^{4-\frac{4}{4p+1}})$, which is of lower order $(R_{n_\sigma}^C)^2$ given by (3.17). It is clear that in the current setting a hypercube is easier to inscribe into a hyperrectangle than a hypersphere, and hence hypersphere's method cannot give an attainable lower rate, while we can get the attainable lower bound via the hypercube method.

Remark 4. Using the method of Ibragimov *et al* (1987) 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

$$O(\sigma^{-\frac{8(2r+2-p)}{4r+4-p}}),$$

which is not an attainable rate when $q \neq 0$. Thus, the hypersphere method does not work in the current setting. The reason for this is that the value of the functional $Q(x) = \sum_1^\infty j^q x_j^2$ changes when x lies on an n-dimensional sphere (note that when $q = 0$, the hypersphere's method can also give an attainable lower rate as $Q(x)$ remains the same when x is on the sphere). Note that in the current setting, the largest n-dimensional inner hypercube lies in the largest n-dimensional hypersphere (see Figure 1). Thus, the reason for our method to give a larger lower bound *is not due to* the fact that the hypercube is easier to inscribe than the hypersphere, but *is due to* the fact that for estimating a symmetric functional ((2.1)), the value of the functional remains the same when x is on vertices of the hypercube.

7. Discussions

a) Possible Applications

We have demonstrated that for special kinds of constraints of the hyperrectangles and the weighted l_p-bodies ($p \geq 2$), the difficulties of estimating quadratic functionals are captured by the difficulties of the hypercube subproblems. The notions inside the problems can be explained as follows. For hypercube-typed hyperrectangles (i.e. the lengths of a hyperrectangle satisfy Assumption A), the difficulties of estimating the quadratic functionals are captured by the difficulties of the cubical subproblems (Theorem 2). Now, for estimating quadratic functionals under the constraints of weighted l_p-bodies (Theorem 4 & 5), the difficulties of the estimating quadratic functionals are actually captured by rectangular subproblems, which happen to be hypercube-typed. Thus, the difficulties of the estimating quadratic functionals under the weighted l_p-body constraints (Theorem 4 & 5) 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:

$$\min_{q(y)} \max_{x \in \Sigma} E_x(q(y) - Q(x))^2 \le C \max\{ \min_{q(y)} \max_{x \in \Theta(\tau)} E_x(q(y) - Q(x))^2 : \Theta(\tau) \in \Sigma\},$$

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 phenomenon above is true (Donoho *et* (1988)).

We can apply our hypercube bound to a non-symmetric constraint, and also to an unsymmetric functional. Let's give an example involved the use of our theory.

Example 5. (Estimation of Fisher Information) Suppose that we want to estimate the Fisher information

$$I(f) = \int_0^1 \frac{f'^2(x)}{f(x)}\, dx$$

based on the data observed from (1.6) under the nonparametric constraint that $f \in \Sigma^* = \Sigma \cap \{f : \int_0^1 f(x) = 1, f(x) \ge 0\}$, where Σ is a subset of R^∞. Take the same orthogonal set as Example 2. Let $l_n(\Sigma)$ be the n-dimensional inner length of Σ (not Σ^*). Inscribe the largest n-dimensional inner hypercube into Σ. The functions whose Fourier coefficients are on the vertices of the hypercube are the set of functions

$$f_n(x) = 1 + l_n(\Sigma) \sum_2^n \pm\phi_j(x)$$

for all possible choices of signs $\pm$, where $\phi_j(t)$ are sinusoid functions given by Example 2. If $nl_n(\Sigma) \longrightarrow 0$, all of these functions are positive and hence in the set of the our positivity constraints. By our hypercubical approach, for any estimator $T(Y)$ based on the observations (1.6) (see the proof of Theorem 1),

$$\sup_{f \in \Sigma \cap \{f \geq 0\}} E_f(T(Y) - I(f))^2 \geq \Phi(-\frac{d}{\sqrt{8}})\,(I(f_{n_\sigma}) - I(1))^2/4(1 + o(1))$$

$$= \frac{\Phi(-\frac{d}{\sqrt{8}})}{4}\left[\int_0^1 (f_{n_\sigma}'(x))^2 dx\right]^2 (1 + o(1))$$

$$= \frac{\Phi(-\frac{d}{\sqrt{8}})}{4}\left[(2\pi)^2 l_{n_\sigma}^2(\Sigma)\sum_2^{n_\sigma} j^2\right]^2 (1 + o(1)),$$

where n_σ is the smallest integer satisfying (2.5). Thus, if the nonparametric constraint Σ is defined as Example 4, then the minimax lower bound is at least as large as

$$\begin{cases} O(\sigma^2), & \text{if } \alpha \geq 2.25 \\ O(\sigma^{\frac{16(\alpha-1)}{4\alpha+1}}), & \text{if } 2.25 > \alpha > 1 \end{cases}.$$

and for the hyperrectangle constraint $\Sigma = \{|x_j(f)| \leq Cj^{-p}\}$, the minimax lower bound is

$$\begin{cases} O(\sigma^2), \text{ if} & p > 2.75 \\ O(\sigma^{\frac{16p-24}{4p-1}}), & \text{if } 1.5 < p \leq 2.75 \end{cases}.$$

Thus, the lower bound for estimating Fisher information is at least as large as that of estimating $\int_0^1 [f'(x)]^2\, dx$, and the lower bound is attainable if f is bounded away from 0. Note that Σ^* is an unsymmetric set in this example and $T(f)$ is an unsymmetric functional (see (2.1)). Our hypercube method can also apply to estimating an unsymmetric functionals with an unsymmetric positivity constraints.

The functional $T(f)$ above can also be replaced by $T(f) = \int_0^1 [f^{(k)}(t)]^2\, dt$.

b) Constants

By using the hardest 1-dimension trick, we can prove the following lower bound (see Fan (1989) for details):

Theorem 6. If Σ *is a symmetric, convex set containing the origin, then minimax risk of estimating* $Q(x)$ *from the observations (1.1) is at least*

$$\sup_{x \in \Sigma} \frac{Q(x)^2 \sigma^4}{\|x\|^4} \rho(\|x\| / \sigma, 1), \tag{7.1}$$

and as $\sigma \to 0$, *for any estimator* $\delta(y)$,

$$\sup_{x \in \Sigma} E_x(\delta(y) - Q(x))^2 \geq \sup_{x \in \Sigma, \|x\| > 0} \frac{4Q^2(x)}{\|x\|^2} \sigma^2 (1 + o(1)), \tag{7.2}$$

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

Comparing the lower bound (7.2) and the upper bound (3.16) for estimating $\sum_1^\infty j^q x_j^2$ with a hyperrectangular constraint $\{x \in R^\infty : |x_j| \leq C j^{-p}\}$, we have that when $p > q + 0.75$,

$$1 \geq \frac{\text{Lower Bound}}{\text{Upper Bound}} \geq \frac{C_{2p-q}^2}{C_{2p}\, C_{2p-2q}}, \quad (\text{as } \sigma \longrightarrow 0), \tag{7.3}$$

where $C_r = \sum_1^\infty j^{-r}$ can be calculated numerically. Table 7.1 shows results of the right hand side in (7.3).

From Table 7.1, we know that the best truncated estimator is very efficient. Thus, the difficulty of the hardest 1-dimensional subproblem captures the difficulty of the full problem pretty well in the case $p > q + 0.75$.

Table 7.1: Comparison of the lower bound and upper bound
p = 1 + q + 0.5 i

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

When $(q+1)/2 < p \le q + 0.75$, by comparing the lower bound (3.17) and the upper bound (3.15), again we show that the best truncated estimator attains the optimal rate. The following table shows how close the lower bound and the upper bound are.

Table 7.2: Comparison of the lower bound and the upper bound

q = 0		q = 1		q = 2		q = 3		q = 4	
p =	ratio	p =	ratio	p =	ratio	p=	ratio	p=	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

where $ratio = \sqrt{\dfrac{\text{hypercube lower bound}}{\text{upper bound by } q_{UT}(y)}}$

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

c) 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 $x \in \Sigma = \{x: |x_j| \le j^{-p}\}$.

It is well known that the minimax risk is attained at the worst Bayes risk. 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. Thus, finding an attainable rate of estimating a quadratic functional is a non-trivial job.

By an *intuitive* prior, we mean that assign the prior uniformly on the hyperrectangle, which is equivalent to that independently assign the prior uniformly on each coordinate, or more generally we mean that assign the prior $x_j \sim \pi_j(\theta)$ independently. To see why the Bayesian method with an intuitive prior can not give an attainable rate, let $\rho_{\pi_j}(j^{-p}, \sigma)$ be the Bayes risk of estimating θ^2 from $Y \sim N(\theta, \sigma^2)$ with a prior $\pi_j(\theta)$ concentrated on $[-j^{-p}, j^{-p}]$. Then,

$$\rho_{\pi_j}(j^{-p}, \sigma) \le \min(j^{-4p}, 3\sigma^4 + 4\sigma^2 j^{-2p}), \tag{7.4}$$

as the Bayes risk is no larger than the maximum risk of estimators 0 and Y^2.

Let $\hat{\delta}(y) = E(\sum_1^\infty j^q x_j^2 \mid y)$ be the Bayes solution of the problem. By independent assumption of the prior,

$$\hat{\delta}(y) = \sum_1^\infty j^q E(x_j^2 \mid y_j).$$

Thus, by (7.4) the Bayes risk is

$$E_\pi E_x (\hat{\delta}(y) - \sum_1^\infty j^q x_j^2)^2 = \sum_1^\infty j^{2q} \rho_{\pi_j}(j^{-p}, \sigma)$$

$$\le \sum_1^n j^{2q} (3\sigma^4 + 4\sigma^2 j^{-2p}) + \sum_{n+1}^\infty j^{2q-4p},$$

$$\le O(n^{2q+1}\sigma^4 + n^{2q-2p+1}\sigma^2 + n^{-4p+2q+1})$$

$$= O(\sigma^{4-(2q+1)/p} + \sigma^2)$$

$$= o(\sigma^{4 - \frac{4(2q+1)}{4p-1}}), \quad (\text{when } (q+1)/2 < p < q + 0.75)$$

by choosing $n = \sigma^{-\frac{1}{p}}$. Hence, the *intuitive* Bayes method gives a too small lower bound in terms of rate of convergence.

8. Proofs

Proof of Lemma 1. Without loss of generality, assume that $\sigma = 1$. Then the likelihood ratio of the density under H_0 and H_1 is

$$L_n = \prod_1^n L_{n,i}, \tag{8.1}$$

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,

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

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

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

under H_0. Note that under H_1,

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

Now under H_1,

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

and

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

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

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

Consequently, the sum of type I and type II errors is

$$P_{H_0}[L_n > 1] + P_{H_1}[L_n \le 1] = 2\,\Phi(-\frac{\sqrt{n}\ l_n^2}{\sqrt{8}})(1 + o(1)).$$

Proof of Proposition 1. For any subset $S \subset \{1, 2, \cdots\}$, 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 0 for $j \notin S$. Then by Jensen's inequality and (3.5),

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

$$\ge \max_S \{(E_{\mu^S}(x'Bx) + \sigma^2\, trB + c - E_{\mu^S}\, [Q(x)])^2 + 2\sigma^4\, trB^2 + 4\sigma^2\, E_{\mu^S}(x'B^2x)\} \quad (8.2)$$

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

$$E_{\mu^s}(x'Bx) = tr(BD_A) = tr(D_B D_A),$$

$$E_{\mu^s}(x'B^2x) = tr(B^2 D_A) \geq tr(D_B^2 D_A),$$

$$trB^2 = trB'B \geq trD_B^2.$$

Thus by (8.2) and the last 3 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 (3.7) is convex in x_j^2, and consequently attains its maximum at either $x_j^2 = 0$ or $x_j^2 = A_j^2$.

Proof of Proposition 2. Sufficiency follows from the monotone convergence theorem:

$$Eq_B(x) = \sum_1^\infty b_i(x_i^2 + \sigma^2) + c < \infty.$$

Suppose that $q_B(x)$ with B given by (3.6) converges a.s. for each $x \in \Sigma$. Then according to the Kolmogrov's 3-series theorem,

$$\sum_1^\infty P\{b_i y_i^2 \geq 1\} < \infty, \quad \sum_1^\infty E b_i y_i^2 \, 1_{(b_i y_i^2 \leq 1)} < \infty. \tag{8.3}$$

As the distribution of y_i is normal, it is easy to check that $E\, y_i^4 \leq 3(E\, y_i^2)^2$. Thus by Cauchy-Schwartz inequality and (8.3),

$$E\ b_i y_i^2 \, 1_{(b_i y_i^2 \geq 1)} = o(\,E\ b_i y_i^2\,). \tag{8.4}$$

Hence the assertion follows from (8.3) and (8.4).

Proof of Theorem 2. We will prove that the estimator (3.9) with $m = n_\sigma$ achieves the rate given by (3.11); the lower bounds are proved by Theorem 1 & Theorem 3.

Note that n_σ increases to infinite as σ decreases to 0. By the assumptions, the right hand side of (3.10) is

$$O\left(\left(\sum_1^m \lambda_j\right)^2 (A_m^4 + \sigma^4/m)\right) + 4\sigma^2 \sum_1^m \lambda_j^2 A_j^2. \tag{8.5}$$

Taking $m = n_\sigma$, by (3.12) we have

$$\sigma^4/(n_\sigma - 1) \le A_{n_\sigma - 1}^4/c^2 = O(A_{n_\sigma}^4).$$

Thus for $m = n_\sigma$, (8.5) becomes

$$O\left(\left(\sum_1^{n_\sigma} \lambda_j\right)^2 A_{n_\sigma}^4\right) + 4\sigma^2 \sum_1^{n_\sigma} \lambda_j^2 A_j^2 \tag{8.6}$$

Case I: If $\limsup_{\sigma \to 0} n_\sigma \lambda_{n_\sigma}^2 \sigma^2 < \infty$, then by (3.12) and Assumption A i)

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

To prove (3.11), we need only to show that $\sum_1^\infty \lambda_j^2 A_j^2 < \infty$.

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

$$n_\sigma^{1.5} \lambda_{n_\sigma}^2 A_{n_\sigma}^2 \le D, \text{ when } \sigma \le \sigma_0. \tag{8.7}$$

By Assumption A i), it can be shown that as σ decreases from σ_0 to 0, n_σ should increase from n_{σ_0} to ∞ consecutively. Thus (8.7) implies that

$$\lambda_j^2 A_j^2 \le D\ j^{-1.5}, \text{ when } j \ge n_{\sigma_0}.$$

Consequently,

$$\sum_{j=1}^\infty \lambda_j^2 A_j^2 < \infty$$

and

$$(8.6) = \sum_{1}^{\infty} \lambda_j^2 A_j^2 O(\sigma^2).$$

Hence, the regular rate is the optimal one by Theorem 6.

Case II: If $\limsup_{\sigma \to 0} n_\sigma \lambda_{n_\sigma}^2 \sigma^2 = \infty$, then $\limsup_{n \to \infty} n^{1.5} \lambda_n^2 A_n^2 = \infty$. Thus, by (3.12) and Assumption A ii)

$$\frac{\sigma^2 \sum_{1}^{n_\sigma} \lambda_j^2 A_j^2}{(\sum_{1}^{n_\sigma} \lambda_j)^2 A_{n_\sigma}^4} \leq O\left(\frac{\sqrt{n_\sigma} \sum_{1}^{n_\sigma} \lambda_j^2 A_j^2}{n_\sigma^2 \lambda_{n_\sigma}^2 A_{n_\sigma}^2}\right),$$

which is bounded. Hence, (8.6) is bounded by its first term. Consequently, the truncated estimator (3.9) with $m = n_\sigma$ achieves the rate given by (3.11), and by Theorem 3, the rate of $(\sum_{1}^{n_\sigma} \lambda_j)^2 A_{n_\sigma}^4$ is the best attainable one.

Proof of Theorem 3. Note that $\{A_n\}$ is a decreasing sequence by the assumption and $l_n(\Sigma) = A_n$. The r_n defined by (2.7) is

$$r_n = A_n^2 \sum_{1}^{n} \lambda_j / 2.$$

Thus, by Theorem 1,

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

The conclusion follows by taking the supreme of d.

Proof of Theorem 4. First, we prove the truncated estimator (3.9) achieves the rate given by (4.4). For the truncated estimator (3.9), the maximum risk

$$\max_{x \in \Sigma_2} R(q_{UT}, x)$$

$$\leq \max_{x \in \Sigma_2} (\sum_m^\infty \lambda_j x_j^2)^2 + \max_{x \in \Sigma_2} \sum_1^m \lambda_j^2 (2\sigma^4 + 4\sigma^2 x_j^2) \tag{8.8}$$

$$\leq C^2 (\lambda_m / \delta_m)^2 + 2\sum_1^m \lambda_j^2 \sigma^4 + 4C\sigma^2 \max_{1 \leq j \leq m} \lambda_j^2/\delta_j . \tag{8.9}$$

Take $m = n_\sigma$. Note that as $\sigma \longrightarrow 0$, $n_\sigma \longrightarrow \infty$. By (4.5) and Assumption B ii), there exists a constant σ_0 such that when $\sigma \leq \sigma_0$,

$$\delta_{n_\sigma}^2 > \frac{c}{2} \delta_{n_\sigma + 1}^2 \geq \frac{dc/2}{(n_\sigma + 1)\, \sigma^4} . \tag{8.10}$$

By (8.10),

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

Consequently, by (8.9) and the fact that λ_j is non-decreasing, we have

$$\max_{x \in \Sigma_2} R(q_{UT}, x) = O(n_\sigma \lambda_{n_\sigma}^2 \sigma^4) + O(\sigma^2 \max_{1 \leq j \leq n_\sigma} \lambda_j^2/\delta_j). \tag{8.11}$$

Case I: If $\limsup_{n \longrightarrow \infty} n\ \lambda_n^4/\delta_n^2 < \infty$, then the sequence $\{\lambda_n^2/\delta_n\}$ stays bounded, and by (4.5)

$$n_\sigma \lambda_{n_\sigma}^2 \sigma^2 \leq \sqrt{d}\, \sqrt{n_\sigma} \lambda_{n_\sigma}^2 / \delta_{n_\sigma} = O(1).$$

By (8.11),

$$\max_{x \in \Sigma_2} R(q_{UT}, x) = O(\sigma^2).$$

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

Case II: If $\limsup_{n \longrightarrow \infty} n\ \lambda_n^4/\delta_n^2 = \infty$, then $\sigma^2 = o(n_\sigma \lambda_{n_\sigma}^2 \sigma^4)$. If the sequence $\{\lambda_n^2/\delta_n\}$ stays bounded, then (8.11) is of order $O(n_\sigma \lambda_{n_\sigma}^2 \sigma^4)$. Otherwise, by assumption C ii) and (8.10),

$$\limsup_{\sigma \longrightarrow 0} \frac{\max_{1 \leq j \leq n_\sigma} \lambda_j^2/\delta_j}{n_\sigma \lambda_{n_\sigma}^2 \sigma^2} \leq \limsup_{\sigma \longrightarrow 0} \frac{1}{\sqrt{n_\sigma} \delta_{n_\sigma} \sigma^2} < \infty.$$

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

To prove the the rate is the best attainable one, we need to show that no estimator can estimate $Q(x)$ faster than the rate given by (4.4). For the first case, by Theorem 6, we know that $O(\sigma^2)$ is the optimal rate. To prove the second case, let us inscribe an n-dimensional inner hypercube. The largest inner hypercube we can inscribe into the weighted l_2-body is the hypercube, which is symmetric about the origin, with length $2A_n$, where $A_n^2 = C/(\sum_1^n \delta_j)$, i.e. the n-dimensional inner length $l_n(\Sigma) = A_n$. Some simple algebra shows that when σ is small,

$$\sqrt{n_\sigma}\,[l_{n_\sigma}(\Sigma_2)]^2/\sigma^2 \le \frac{2C}{c\sqrt{n_\sigma + 1}\,\delta_{n_\sigma + 1}\sigma^2}\,(\frac{\sqrt{n_\sigma + 1}\,\delta_{n_\sigma + 1}}{\sqrt{n_\sigma}\,\delta_{n_\sigma}}) \le D.$$

for some $D > 0$. By Theorem 1, we conclude for any estimator $T(y)$,

$$\sup_{x \in \Sigma_2} E_x(T(y) - Q(x))^2 \ge \Phi(-\frac{D}{\sqrt{8}})\, r_{n_\sigma}^2 (1 + o(1)) ,$$

where $r_{n_\sigma} = A_{n_\sigma}^2 \sum_1^{n_\sigma} \lambda_j/2$. Now, it is easy to compute that

$$r_{n_\sigma} \ge \frac{C}{2}\,\frac{c\lambda_{n_\sigma}}{\delta_{n_\sigma}} \ge \frac{C\;c}{2\sqrt{d}}\sqrt{n_\sigma}\lambda_{n_\sigma}\sigma^2 .$$

The conclusion follows.

Proof of Theorem 5. We will prove the truncated estimator achieves the rate given by (4.8) and then use the hypercube approach to prove the lower bound.

For the truncated estimator (3.9), the maximum risk

$$\max_{x \in \Sigma_p} R(q_{UT}, x)$$

$$\le \max_{x \in \Sigma_p} (\sum_m^\infty \lambda_j x_j^2)^2 + \max_{x \in \Sigma_p} \sum_1^m \lambda_j^2 (2\sigma^4 + 4\sigma^2 x_j^2)$$

Let $q = \dfrac{p}{p-2}$ be the conjugate number of $\dfrac{p}{2}$. Then by Holder's inequality, we have for any $x \in \Sigma_p$,

$$\sum_m^{\infty} \lambda_j x_j^2 \le (\sum_m^{\infty} (\lambda_j \ \delta_j^{-2/p})^q)^{1/q} \ (\sum_m^{\infty} \delta_j \ |x_j|^p)^{2/p}$$

$$\le C^{2/p} \ (\sum_m^{\infty} \lambda_j^{p/(p-2)} \ \delta_j^{-2/(p-2)})^{(p-2)/p}. \tag{8.12}$$

Similarly, we have

$$\sum_1^m \lambda_j^2 x_j^2 \ \le \ C^{2/p} \ (\sum_1^m \lambda_j^{2p/(p-2)} \ \delta_j^{-2/(p-2)})^{(p-2)/p}. \tag{8.13}$$

Hence by (8.12) and (8.13),

$$\max_{x \in \Sigma_p} R(q_{UT}, x)$$

$$\le C^{4/p} (\sum_m^{\infty} \lambda_j^{p/(p-2)} \ \delta_j^{-2/(p-2)})^{2(p-2)/p} + 2\sum_1^m \lambda_j^2 \ \sigma^4$$

$$+ 4\sigma^2 C^{2/p} \ (\sum_1^m \lambda_j^{2p/(p-2)} \ \delta_j^{-2/(p-2)})^{\frac{p-2}{p}}$$

$$\le O\Big(\lambda_m^2 \ \delta_m^{-4/p} m^{2(p-2)/p} + m\lambda_m^2 \sigma^4 + \sigma^2 (\sum_1^m \lambda_j^{2p/(p-2)} \delta_j^{-2/(p-2)})^{(p-2)/p}\Big), \tag{8.14}$$

by Assumption D b) and Assumption B ii). Now, by taking $m = n_\sigma$ and using the fact that (c.f. (4.9))

$$\delta_{n_\sigma + 1}^{-\frac{4}{p}} < \sigma^4 \ (n_\sigma + 1)^{-(p-4)/p} / d, \tag{8.15}$$

(8.14) is of order

$$O\Big(n_\sigma \lambda_{n_\sigma}^2 \sigma^4 \ + \ \sigma^2 (\sum_1^{n_\sigma} \lambda_j^{2p/(p-2)} \delta_j^{-2/(p-2)})^{(p-2)/p}\Big). \tag{8.16}$$

Case I: If $\limsup_{n \to \infty} \ n^{(3p-4)/(2p)} \lambda_n^2 \delta_n^{-2/p} < \infty$, then it is easy to show that

$$\lambda_n^{2p/(p-2)} \ \delta_n^{-2/(p-2)} = O(n^{-\frac{3p-4}{2(p-2)}}) = o(n^{-1.5})$$

Consequently,

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

and (8.16) is of order $O(\sigma^2)$.

Case II: If $\limsup_{n \to \infty} n^{(3p-4)/(2p)} \lambda_n^2 \delta_n^{-2/p} = \infty$, then by Assumption D c)

$$\left[\sum_1^{n_\sigma} \lambda_j^{2p/(p-2)} \delta_j^{-2/(p-2)} \right]^{(p-2)/p} \le O\left[\lambda_{n_\sigma}^2 \delta_{n_\sigma}^{-2/p} n_\sigma^{(3p-4)/(2p)} \right]$$

$$= O(n_\sigma \lambda_{n_\sigma}^2 \sigma^2).$$

Hence, (8.16) is of order $O(n_\sigma \lambda_{n_\sigma}^2 \sigma^4)$. Thus, the truncated estimator achieves the rate given by (4.8).

To prove the lower bound result, similar to the proof of Theorem 4, we need only to consider the second case. Note that the n-dimensional inner length $l_n(\Sigma_p) = (C/\sum_1^n \delta_j)^{1/p}$. For the n_σ defined by (4.9), we have

$$\sqrt{n_\sigma}(l_{n_\sigma}(\Sigma_p)/\sigma)^2 \le C^{2/p}\sqrt{d} \triangleq D.$$

Thus by Theorem 1, for any estimator $T(y)$

$$\sup_{x \in \Sigma_2} E_x(T(y) - Q(x))^2 \ge \Phi(-D/\sqrt{8}) r_{n_\sigma}^2$$

where

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

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

Acknowledgements. This work is part of the author's Ph.D. thesis at University of California, Berkeley, written under the supervision of Professor D.L. Donoho and Professor 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.

REFERENCES

1. Bickel, P.J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Stat.*, **9**, 1307-1309.
2. Bickel, P.J. and Ritov, Y.(1988). Estimating integrated squared density derivatives. *Technical Report 146*, Department of Statistics, University of California, Berkeley.
3. Bertero, M., Baccoaci, P., and Pike, E.R. (1982). On the recovery and resolution of exponential relaxation rates from experimental data I. *Proc. Roy. Sco. Lond.*, A **383**, 15-.
4. Chentsov, N.N. (1980). *Statistical Decision Rules and Optimal Inference.* American Mathematical Society, Providence, RI.
5. Donoho, D.L. and Liu, R.C. (1987a). On minimax estimation of linear functionals. *Tech. Report 105*, Dept. of Stat., University of California, Berkeley.
6. Donoho, D.L. and Liu, R.C. (1987b). Geometrizing rates of convergence, I. *Tech. Report 137a*, Department of Statistics, University of California, Berkeley.
7. Donoho, D.L. and Liu, R.C. (1987c). Geometrizing rates of convergence, II. *Technical Report 120*, Department of Statistics, University of California, Berkeley.
8. Donoho, D.L. and Liu, R.C. (1988). Geometrizing rates of convergence, III. *Technical Report 138*, Department of Statistics, University of California, Berkeley.
9. Donoho, D.L., MacGibbon, B., Liu, R.C. (1987). Minimax risk for hyperrectangles. *Technical Report 123*, Department of Statistics, University of California, Berkeley.

10. Efroimovich, S.Y. and Pinsker, M.S. (1982). Estimation of square-integrable probability density of a random variable. *Problemy Peredachi Informatsii*, **18**, **3**, 19-38 (in Russian); *Problems of Information Transmission* (1983), 175-189.

11. Fan, J.(1988). On the optimal rates of convergence for nonparametric deconvolution problem. *Tech. Report 157*, Dept. of Stat., Univ. of California, Berkeley.

12. Fan, J. (1989). *Contributions to the estimation of nonregular functionals. Dissertation*, Department of Statistics, University of California, Berkeley.

13. Farrell, R.H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Stat.*, **43, #1**, 170-180.

14. Hall, P. and Marron, J.S. (1987a). On the amount of noise inherent in bandwidth selection. *Ann. Statist.*, **15**, 163-181.

15. Hall, P. and Marron, J.S. (1987b). Estimation of integrated squared density derivatives. *Statist. & Probab. letters*, **6**, 109-115.

16. Ibragimov, I.A. and Khas'minskii (1981). *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, New York-Berlin-Heidelberg.

17. Ibragimov, I.A., Nemirovskii, A.S. and Khas'minskii, R.Z. (1987). Some problems on nonparametric estimation in Gaussian white noise, *Theory Prob. Appl.*, **31, 3**, 391-406.

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

19. Khas'minskii, R.Z. (1979). Lower bound for the risks of nonparametric estimate of the mode. *Contribution to Statistics (J. Hajek memorial volume)*, Academia, Prague, 91-97.

20. Le Cam, L. (1973). Convergence of estimates under dimensionality restrictions, *Ann. Stat.* **1**, 38-53.

21. Le Cam, L. (1985). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York-Berlin-Heidelberg.

22. Nussbaum, M. (1985). Spline smoothing in regression models and Asymptotic efficiency in L_2. *Ann. Statist.*, **13**, 984-997.

23. Parzen, E. (1971). Statistical inference on time series by RKHS methods. *Proc. 12th Biennial Seminar of the Canadian Mathematical Congress* (R. Pyke, ed.), 1-37.

24. Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation*, Academic Press.

25. Pinsker, M.S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems of Information Transmission*, **16, 2**, 52-68.

26. Sacks, J. and Ylvisaker, D. (1981). Variance estimation for approximately linear models. *Math. Operationsforch. Statist., Ser. Statistics*, **12**, 147-162.

27. Speckman, P. (1985). Spline smoothing and optimal rates of convergence in non-parametric regression models. *Ann. Statist.*, **13**, 970-983.

28. Stone, C. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Stat.* **8**, 1348-1360.

29. Wahba, G. (1975). Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation. *Ann. Statist*, **3**, 15-29.