Probab. Theory Relat. Fields 99, 145–170 (1994)

Probability Theory and Related Fields

Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery

David L. Donoho
Statistics Department, Stanford University, Sequoia Hall, Stanford, CA 94305, USA

Received: 29 December 1992/In revised form: 4 February 1994

Summary. We study the problem of estimating an unknown function on the unit interval (or its k-th derivative), with supremum norm loss, when the function is observed in Gaussian white noise and the unknown function is known only to obey Lipschitz-β smoothness, $\beta > k \geqq 0$. We discuss an optimization problem associated with the theory of *optimal recovery*. Although optimal recovery is concerned with deterministic noise chosen by a clever opponent, the solution of this problem furnishes the kernel of the minimax linear estimate for Gaussian white noise. Moreover, this minimax linear estimator is asymptotically minimax among all estimates. We sketch also applications to higher dimensions and to indirect measurement (e.g. deconvolution) problems.

Mathematics Subject Classifications: 62G07, 62C20, 60G70, 41A25

1 Introduction

A.P. Korostelev, in Khas'minskii's seminar in Moscow, has recently presented an evaluation of the precise constants in the asymptotic minimax risk, with supremum norm loss, for estimating a β-Lipschitz regression function, $\beta \in (0, 1]$, from n noisy samples $y_i = f(i/n) + \sigma z_i$, $z_i \overset{\text{iid}}{\sim} N(0, 1)$ (Korostelev, 1991).

This result joins a very short list of exact asymptotic results about the minimax risk for estimating infinite-dimensional parameters. It bears especial comparison with results of Pinsker (1980), Efroimovich and Pinsker (1981, 1982), and Nussbaum (1984). Those results concern minimax risk with squared-error loss and smoothness assumed in an L^2-Sobolev sense. Korostelev's result concerns instead the L^∞-loss and L^∞-Sobolev smoothness. Hence the result may be viewed as the L^∞-analog of the L^2-work just mentioned.

The form of Korostelev's result is interesting for several reasons:

(a) The asymptotic minimax risk is attained by a certain linear (kernel) estimator $\hat{f}_n(t) = n^{-1} \sum_i y_i K_n(t - i/n)$.

Dedicated to R.Z. Khas'minskii for his 60th birthday

(b) The optimal kernel and the constants in Korostelev's solution were known previously as the optimal kernel and constants in some (apparently different) problems (Donoho and Liu 1991, Donoho 1994).
(c) The minimaxity for risk $E\ell(\|\hat{f}_n - f\|_\infty)$ does not depend on [reasonably chosen] $\ell(\cdot)$.

The subject area is appealing because L^∞-loss has special importance in connection with setting fixed-width simultaneous confidence bands for an unknown regression function.
The purpose of this article is threefold:

(A) To describe a more general result, encompassing estimation of k-th derivatives $k \geqq 0$ of β-Lipschitz regression functions $\beta > k$, of which Korostelev's result is the special case for $k=0$, $\beta \leqq 1$. This more general result derives from a general method, absent in Korostelev's solution; the general method may also be used to treat indirect observations, such as deconvolution, as well as higher-dimensional problems.
(B) To describe the family of optimal kernels which arise in the minimax linear estimators.
(C) To show that the constants in the asymptotics of the minimax risk are precisely the same as the constants arising in certain problems of optimal recovery, and that the optimal procedures are the same as well – when the noise levels in the two problems are calibrated appropriately.

The paper makes heavy use of ideas of renormalization and optimal recovery. See Donoho and Low (1992) and Donoho (1994).

2 Preliminaries: optimal recovery of linear functionals

We begin by turning away from statistical estimation and consider instead a problem of recovering f in the presence of *nonstochastic* noise. Suppose we observe

$$y(t) = f(t) + \varepsilon \cdot z(t), \quad t \in (-\infty, \infty), \tag{2.1}$$

where f is an unknown function, which is known to belong to the Lipschitz class $\mathscr{F}(\beta, C)$. This is the set of all functions f such that, with $m = \lfloor \beta \rfloor$ and $\alpha = \beta - m$,

$$|f^{(m)}(t) - f^{(m)}(u)| \leqq C|t-u|^\alpha, \quad t, u \in \mathbb{R}.$$

The term $z(t)$ is a nuisance term which is chosen by a malicious opponent, subject to the constraint that $\|z\|_2 \equiv \|z\|_{L^2(-\infty,\infty)} \leqq 1$; $\varepsilon > 0$ is a known "noise level".

We are interested in recovering the linear functional $T(f) = T_k(f) = f^{(k)}(0)$, and we evaluate performance of a rule $\hat{T}(y)$ by the worst-case principle. We assume that our opponent chooses z in the most skillful way possible, i.e. our measure of risk is

$$\mathrm{Err}_\varepsilon(\hat{T}, f) = \sup\{|\hat{T}(y) - T(f)| : \|y - f\|_2 \leqq \varepsilon\}.$$

Moreover, we assume f is chosen from $\mathscr{F}(\beta, C)$ to make life difficult. Our best possible performance is then the minimax error:

$$E^*(\varepsilon) = E^*(\varepsilon; T, \mathscr{F}) = \inf_{\hat{T}} \sup_{\mathscr{F}} \mathrm{Err}_\varepsilon(\hat{T}, f).$$

A rule $\hat{T}(y)$ attaining E^*, i.e.

$$\sup_{\mathscr{F}} \mathrm{Err}_\varepsilon(\hat{T}, f) = E^*(\varepsilon),$$

will be called *minimax*. [The general setting we are discussing is one of *optimal* recovery: Micchelli (1975), Micchelli and Rivlin (1977), Traub et al. (1988). In such a setting, when the available data are finite in number, a minimax rule is generally called an *optimal algorithm*.]
Applying the optimal recovery theorem of Micchelli (1975) in the present setting yields the formula

$$E^*(\varepsilon) = \mathrm{val}(P_{\varepsilon,C}),$$

where $(P_{\varepsilon,C})$ denotes the optimization problem:

$$(P_{\varepsilon,C}): \quad \sup T_k(f) \quad \text{subject to} \begin{cases} \|f\|_2 \leqq \varepsilon, \\ f \in \mathscr{F}(\beta, C). \end{cases}$$

Here we are suppressing mention of k and β from the problem label, but the actual problem being solved depends on those parameters. Also, $\mathrm{val}(P_{\varepsilon,C})$ is finite for all $\varepsilon > 0$ and $C > 0$ if and only if $\beta > k$.

2.1 Renormalization

Our assumption that data are available on the domain $t \in (-\infty, \infty)$ has the following pleasant consequence.

Lemma 2.1 *Let $k \geqq 0$ and $\beta > k$. Then*

$$val(P_{\varepsilon,C}) = val(P_{1,1}) C^{1-r} \varepsilon^r,$$

where $val(P_{1,1}) < \infty$ and

$$r = \frac{2\beta - 2k}{2\beta + 1}.$$

The importance of this lemma comes in isolating the single optimization problem $(P_{1,1})$ as of fundamental interest. Once $\mathrm{val}(P_{1,1})$ is known, $E^*(\varepsilon)$ is known for all C, ε.
[The constant $\mathrm{val}(P_{1,1})$ has an interpretation outside optimal recovery. Suppose we wish to know the smallest constant $A(k, \beta)$ such that the inequality

$$\|f^{(k)}\|_\infty \leqq A(k, \beta) \|f\|_2^r \|f\|_{\mathrm{Lip}(\beta)}^{1-r}$$

holds, where

$$\|f\|_{\mathrm{Lip}(\beta)} = \sup_{u,t} \frac{|f^{(m)}(t) - f^{(m)}(u)|}{|t-u|^\alpha}.$$

Then $A(k, \beta) = \mathrm{val}(P_{1,1})$. Compare also remarks in Donoho and Liu (1991), Sect. 4 and references there.]

Proof. This has essentially been proved in Lemma 1 of Donoho and Low (1992). The key idea is to define the renormalization operator $(\mathscr{U}_{a,b} f)(t) = af(bt)$ and note

that whenever f is feasible for $(P_{1,1})$, then, defining a and b by

$$\varepsilon = ab^{-1/2}, \qquad C = ab^{\beta},$$

the renormalized function $\mathscr{U}_{a,b}f$ is feasible for $(P_{\varepsilon,C})$. Hence,

$$\mathrm{val}(P_{\varepsilon,C}) \leqq ab^{k}\,\mathrm{val}(P_{1,1}) = \varepsilon^{r}C^{1-r}\,\mathrm{val}(P_{1,1}).$$

On the other hand, if f is feasible for $(P_{\varepsilon,C})$ then $\mathscr{U}_{a^{-1},b^{-1}}f$ is feasible for $(P_{1,1})$; so

$$\mathrm{val}(P_{1,1}) \geqq (\varepsilon^{r}C^{1-r})^{-1}\,\mathrm{val}(P_{\varepsilon,C}).$$

2.2 Solutions to $(P_{1,1})$

It is easier to discuss solutions to the optimization problem $(Q_{1,1})$ dual to $(P_{1,1})$:

$$(Q_{1,1}): \quad \inf \|f\|_2 \quad \text{subject to} \begin{cases} T_k(f) = 1, \\ f \in \mathscr{F}(\beta, 1). \end{cases}$$

If g_1 is a solution to $(Q_{1,1})$ then a solution f_1 to $(P_{1,1})$ is obtained by renormalization:

$$f_1 = \mathscr{U}_{a,b}g_1,$$

where

$$ab^{\beta} = 1, \qquad ab^{-1/2}\|g_1\|_2 = 1.$$

It follows that

$$\mathrm{val}(P_{1,1}) = (\mathrm{val}(Q_{1,1}))^{-r} = (\|g_1\|_2)^{(2k-2\beta)/(2\beta+1)}.$$

The case $k=0$ and $0<\beta\leqq 1$ is particularly easy to solve. Set

$$g_1(t) = (1-|t|^{\beta})_{+}.$$

Then $T_0(g_1) = g_1(0) = 1$, $g_1 \in \mathscr{F}(\beta, 1)$, and if f is some other element of $\mathscr{F}(\beta, 1)$, satisfying $T_0(f) = 1$, then

$$f(t) \geqq g_1(t), \quad t \in [-1, 1].$$

Consequently, $\int_{-1}^{1} f^2(t)\,dt \geqq \int_{-1}^{1} g_1^2(t)\,dt = \|g_1\|_2^2$, and so g_1 is the solution to $(Q_{1,1})$. Now $\|g_1\|_2^2 = 4\beta^2/[(2\beta+1)(\beta+1)]$, and so

$$\mathrm{val}(P_{1,1}) = ((2\beta+1)(\beta+1)/4\beta^2)^{r/2}.$$

The cases where $\beta > 1$ are by no means so simple as the case $\beta \leqq 1$. In the case $\beta = 2$, more precisely,

$$\mathscr{F}(2,1) = \{f: f^{(2)} \text{ exists in measure}, f^{(1)} \text{ continuous}, |f^{(2)}| \leqq 1 \text{ a.e. } [dx]\},$$

we have considered the cases $k=0$ (estimating a function) and $k=1$ (estimating its derivative). In both cases the solution g_1 has the following qualitative behavior: g_1 is piecewise quadratic; each piece has $|g_1''|(t) = 1$ throughout. However, we do not have explicit formulas for g_1. Correspondence with Linda Zhao, a student at Cornell University, and with A.P. Korostelev, reveals that an exact solution to the problem with $\beta = 2$, $k=0$ can be had. It is a piecewise polynomial with an infinite number of knots and it is not of compact support.

Numerical evaluations of $\mathrm{val}(P_{1,1})$ for a range of k and β are reported in Donoho (1991).

2.3 *Optimal kernels*

The solution of $(P_{1,1})$ is important not just for determination of $E^*(\varepsilon)$, but also because it yields optimal kernels. Let f_1 be a solution to $(P_{1,1})$, and set

$$\lambda = k! \Big/ \int f_1(t) t^k \, dt.$$

Then

$$\psi_1 = \lambda \cdot f_1$$

furnishes a minimax procedure in the sense that $\hat{T}_1(y) = \int \psi_1(t) y(t) \, dt$ has a worst-case error at the theoretical minimum level:

$$\sup_{\mathscr{F}(\beta,1)} \mathrm{Err}_1(\hat{T}_1, f) = E^*(1). \tag{2.2}$$

[This follows by applying the optimal recovery theorem of Micchelli (1975) and Micchelli and Rivlin (1977) to the case at hand: compare also Donoho (1994).]

Optimal kernels for other values of ε and C may be obtained by renormalization of ψ_1.

Lemma 2.2 *Set* $\gamma = 2/(2\beta+1)$ *and*

$$h = h(\varepsilon, C) = (\varepsilon/C)^{\gamma}.$$

The procedure $\hat{T}_\varepsilon(y) = \int \psi_h(t) y(t) \, dt$ *with kernel* $\psi_h(t) = h^{-k-1} \psi_1(t/h)$ *has*

$$\sup_{\mathscr{F}(\beta,C)} \mathrm{Err}_\varepsilon(\hat{T}_\varepsilon, f) = E^*(\varepsilon).$$

Proof. Let a and b satisfy

$$a b^{-1/2} \varepsilon = 1, \tag{2.3}$$

$$a b^{\beta} C = 1. \tag{2.4}$$

Set $\tilde{y}(t) = a y(bt)$. Then

$$\tilde{y}(t) = g(t) + \tilde{z}(t),$$

where $g(t) = a f(bt)$ and $\tilde{z}(t) = \varepsilon a z(bt)$. Because of (2.3) and (2.4), $g \in \mathscr{F}(\beta, 1)$ and $\| \tilde{z} \|_2 \leqq 1$. Consequently, if we apply $\hat{T}_1(\tilde{y}) = \int \psi_1(t) \tilde{y}(t) \, dt$, we have by (2.2) an error not larger than $E^*(1)$. Now note that $\hat{T}_\varepsilon(y) = a^{-1} b^{-k} \hat{T}_1(\tilde{y})$, $T_k(f) = a^{-1} b^{-k} T_k(g)$, and $a^{-1} b^{-k} = \varepsilon^r C^{1-r}$. Thus,

$$\begin{aligned} |\hat{T}_\varepsilon(y) - T_k(f)| &= a^{-1} b^{-k} \, |\hat{T}_1(\tilde{y}) - T_k(g)| \\ &= \varepsilon^r C^{1-r} |\hat{T}_1(\tilde{y}) - T_k(g)|, \end{aligned}$$

and

$$\sup_{\mathscr{F}(\beta,C)} \mathrm{Err}_\varepsilon(\hat{T}_\varepsilon, f) = E^*(1) \; \varepsilon^r C^{1-r} = E^*(\varepsilon).$$

2.4 Optimality for other purposes

The kernels obtained in this way are also optimal for the problem of optimal recovery with global L^∞-loss. Indeed, apply the kernel ψ_h of Lemma 2.2 in the convolutional form $\hat{f}(t)=\int y(u)\psi_h(u-t)\,du$. Then, because of the translation invariance of the class $\mathscr{F}(\beta, C)$ and the translation invariance of the L^2- and L^∞-norm,

$$\begin{aligned}\sup_{\mathscr{F}(\beta,C)} \sup_{\|z\|_2\leqq 1} \|\hat{f}-f\|_{L^\infty(-\infty,\infty)} &= \sup_{\mathscr{F}(\beta,C)} \sup_{\|z\|_2\leqq 1} \sup_t |\hat{f}(t)-f(t)| \\ &= \sup_t \sup_{\mathscr{F}(\beta,C)} \sup_{|z\|_2\leqq 1} |\hat{f}(t)-f(t)| \\ &= \sup_{\mathscr{F}(\beta,C)} \sup_{\|z\|_2\leqq 1} |\hat{f}(0)-f(0)| \\ &= E^*(\varepsilon).\end{aligned}$$

Hence the problem of optimal recovery at a point and in global L^∞-norm have the same minimax error and "same" optimal strategies.

Because it derives from optimal recovery, the family ψ_h is known to be optimal for several problems of statistical estimation at a point. The general reasoning is explained in Donoho (1994). Specific examples are given in Donoho and Liu (1991), Sect. 4.

3 Statistical estimation in global supremum norm

We now turn to problems of statistical estimation. Initially, we consider the "white noise" model, with continuous observations. Compare Ibragimov and Has'minskii (1984), Donoho and Liu (1991). Sections 6 and 7 below will explain that results in this model immediately yield similar results in the nonparametric regression model. We suppose that we are given data

$$Y(dt)=f(t)\,dt+\varepsilon W(dt), \quad t\in(-\infty,\infty), \tag{3.1}$$

where f is again an unknown element of $\mathscr{F}(\beta, C)$, W is now a two-sided Brownian motion $W(0)\sim N(0,\tau^2)$, $W(t)-W(0)\sim N(0,|t|)$, and the "noise level" is ε (presumably small).

Our goal is to estimate the whole object $(f^{(k)}(t); t\in[0,1])$, with supremum-norm loss

$$\|\hat{f}^{(k)}-f^{(k)}\|_\infty \equiv \sup_{t\in[0,1]} |\hat{f}^{(k)}-f^{(k)}(t)|. \tag{3.2}$$

The basic principle we wish to establish is the following. Set the *pseudo-noise level*

$$\tilde{\varepsilon}=\tilde{\varepsilon}(\varepsilon)=\sqrt{2\gamma}\sqrt{\log(\varepsilon^{-1})}\,\varepsilon, \tag{3.3}$$

where again $\gamma=\gamma(\beta)=2/(2\beta+1)$. Take the kernel family (ψ_h) designed for the optimal recovery problem, and use it in the statistical estimation problem to generate a curve estimate via the convolutional form

$$\tilde{f}^{(k)}(t)=\int \psi_{\tilde{h}}(u-t)\,Y(du), \tag{3.4}$$

where $\tilde{h}$ is the bandwidth *which would be optimal for dealing with deterministic noise of size* $\tilde{\varepsilon}$:

$$\tilde{h} = h(\tilde{\varepsilon}, C) = (\tilde{\varepsilon}/C)^{\gamma}.$$

Basic idea. The random variable $\| \tilde{f}^{(k)} - f^{(k)} \|_{\infty}$ is not essentially larger than $E^*(\tilde{\varepsilon})$ and for no estimator could this random variable be essentially smaller. The optimal procedure in this statistical problem is the same as in the optimal recovery problem after an appropriate recalibration of noise levels.

To make this idea precise, let $\ell(\cdot)$ be a continuous, monotone increasing function, satisfying, with $Z \sim N(0,1)$,

$$(L) \quad E\ell(1 + o(1) + o(1)|Z|) \to \ell(1).$$

Define the minimax risk

$$\mathscr{M}^*(\varepsilon) = \inf_{\hat{f}^{(k)}} \sup_{\mathscr{F}(\beta, C)} E\ell\left(\frac{\| \hat{f}^{(k)} - f^{(k)} \|_{\infty}}{E^*(\tilde{\varepsilon})} \right). \tag{3.5}$$

Theorem A.

$$\overline{\lim_{\varepsilon \to 0}} \, \mathscr{M}^*(\varepsilon) \leqq \ell(1) \tag{3.6}$$

and a procedure with worst-case risk obeying this bound is the kernel estimator (3.4) *deriving from the optimal recovery problem at noise level* $\tilde{\varepsilon}$.

Theorem B.

$$\underline{\lim}_{\varepsilon \to 0} \mathscr{M}^*(\varepsilon) \geqq \ell(1);$$

hence the optimal recovery kernel is asymptotically minimax among all estimates.

The remarkable aspect of this result, besides the connection with optimal recovery, is certainly the nonstochastic nature of the limiting result. It implies for example that for each $\eta > 0$

$$\sup_{\hat{f}^{(k)}} \inf_{\mathscr{F}(\beta, C)} P\{ \| \hat{f}^{(k)} - f^{(k)} \|_{\infty} < (1-\eta) E^*(\tilde{\varepsilon}) \} \to 0,$$

$$\inf_{\hat{f}^{(k)}} \sup_{\mathscr{F}(\beta, C)} P\{ \| \hat{f}^{(k)} - f^{(k)} \|_{\infty} > (1+\eta) E^*(\tilde{\varepsilon}) \} \to 0.$$

This quasi-deterministic nature of the loss is somehow responsible for the connection between optimal recovery (i.e. a problem with deterministic nuisance) and the problem of estimation with Gaussian white noise.

4 An upper bound via optimal recovery

We now prove Theorem A. We are again in the white-noise model with data (3.1) and we wish to bound the worst-case risk of $\tilde{f}^{(k)}$, (3.4). Set

$$\mathrm{Bias}(t) = \mathrm{Bias}[\tilde{f}^{(k)}, f^{(k)}](t) = E\tilde{f}^{(k)}(t) - f^{(k)}(t).$$

Define the noise process

$$Z_{\varepsilon}(t) = [\tilde{f}^{(k)}(t) - E\tilde{f}^{(k)}(t)]/\varepsilon.$$

We have, trivially, the decomposition

$$\tilde{f}^{(k)}(t) - f^{(k)}(t) = \text{Bias}(t) + \varepsilon Z_\varepsilon(t),$$

into bias and noise. It follows that

$$\| \tilde{f}^{(k)} - f^{(k)} \|_\infty \leqq \| \text{Bias} \|_\infty + \varepsilon \| Z_\varepsilon \|_\infty . \tag{4.1}$$

The idea in this section is to show that the random variable $\| \tilde{f}^{(k)} - f^{(k)} \|_\infty$ satisfies

$$\text{med}(\| \tilde{f}^{(k)} - f^{(k)} \|_\infty) \leqq E^*(\tilde{\varepsilon})(1 + o(1))$$

and then to show that the deviation above its median is of negligible size. To show the relation between the median of the random variable and $E^*(\tilde{\varepsilon})$, we analyze the two right-hand side terms of (4.1), separately.

4.1 Bias term

Let $\tilde{T}(y) = \int \psi_{\tilde{h}}(u) y(u)\, du$ be a minimax procedure in the optimal recovery model of Sect. 2, with noise level $\tilde{\varepsilon}$, and define the pseudo-bias

$$B(\tilde{T}, f) = \tilde{T}(f) - T(f).$$

Due to the identities

$$E \tilde{f}^{(k)}(0) = E \int \psi_{\tilde{h}}(t) Y(dt) = \int \psi_{\tilde{h}}(t) f(t)\, dt = \tilde{T}(f)$$

we have the identity

$$B(\tilde{T}, f) = \text{Bias}[\tilde{f}^{(k)}, f^{(k)}](0)$$

relating a pseudo-bias in the optimal recovery model to a bias in the function estimation model. Now the class $\mathscr{F}(\beta, C)$ is translation invariant: $f \in \mathscr{F} \Rightarrow f(\cdot - t_0) \in \mathscr{F}$. It therefore follows that

$$\sup_{\mathscr{F}} \sup_t |\text{Bias}[\tilde{f}^{(k)}, f^{(k)}](t)| = \sup_{\mathscr{F}} |\text{Bias}[\tilde{f}^{(k)}, f^{(k)}](0)|.$$

(Indeed, if the worst bias for f is attained at t_0, the worst bias for $f(\cdot - t_0) \in \mathscr{F}$ is attained at 0.) Hence, we arrive at

$$\sup_{\mathscr{F}} \| \text{Bias}(\cdot) \|_\infty = \sup_{\mathscr{F}} |B(\tilde{T}, f)|, \tag{4.2}$$

equating a worst-case bias in the statistical estimation problem to a worst-case pseudo-bias in the optimal recovery problem.

4.2 Noise term

In the optimal recovery model, it is useful to define a pseudo-standard-deviation, reflecting the range of values a procedure can assume under various choices of noise by the opponent. For the procedure $\hat{T}(y) = \int \psi(t) y(t)\, dt$ we define

$$\begin{aligned} S_\varepsilon(\hat{T}, f) &= (\sup\{\hat{T}(y) \colon \| y - f \|_2 \leqq \varepsilon\} - \inf\{\hat{T}(y) \colon \| y - f \|_2 \leqq \varepsilon\})/2, \\ &= \varepsilon \| \psi \|_2, \end{aligned}$$

where the last step follows by Cauchy–Schwartz. In particular, the pseudo-standard-deviation does not depend on f, and we write simply $S_\varepsilon(\tilde{T})$.

We will now show that the stochastic term $\|Z_\varepsilon\|_\infty$ is essentially not bigger than $S_{\tilde{\varepsilon}}(\tilde{T})$.

Note that

$$\begin{aligned} Z_\varepsilon(t) &= \int \psi_{\tilde{h}}(u-t)W(du) \\ &= \tilde{h}^{-k-1}\int \psi_1\left(\frac{y-t}{\tilde{h}}\right)W(dy) \\ &= \tilde{h}^{-k-1/2}\int \psi_1(v-t/\tilde{h})W(dv) \\ &=_D \frac{\|\psi_{\tilde{h}}\|_2}{\|\psi_1\|_2} Z_1(t/\tilde{h}). \end{aligned}$$

Consequently,

$$\|Z_\varepsilon\|_\infty =_D \frac{\|\psi_{\tilde{h}}\|_2}{\|\psi_1\|_2}\|Z_1\|_{L^\infty[0,\tilde{h}^{-1}]}. \tag{4.3}$$

Now the norm $\|Z_1\|_{L^\infty[0,\tilde{h}^{-1}]}$ is easily understood using known results in the theory of extreme values of stochastic processes. For the following result, note that if X is a random variable, we denote by $\mathrm{med}(X)$ a largest median, i.e. a value such that $P\{X \geqq \mathrm{med}(X)\} \leqq \frac{1}{2}$.

Proposition 4.1 *Let* $\psi \in L_1 \cap L_2 \cap \mathscr{F}(\beta, C)$. *Then with* $\sigma = \|\psi\|_2$ *and* $Z_1(t) = \int \psi(u-t)W(du)$

$$\mathrm{med}(\|Z_1\|_{L^\infty[0,A]}) \sim \sqrt{2\log(A)}\cdot\sigma, \quad \text{as } A\to\infty. \tag{4.4}$$

This follows from Theorem 8.2.7 in Leadbetter et al. (1983, Chap. 8) and the regularity assumed for ψ.

It follows immediately from this proposition that

$$\mathrm{med}(\|Z_\varepsilon\|_\infty) = \sqrt{2\log\tilde{h}^{-1}}\,\|\psi_{\tilde{h}}\|_2\,(1+o(1)).$$

Moreover, with

$$\tilde{h} = h(\tilde{\varepsilon}, C) \quad (=(\tilde{\varepsilon}/C)^\gamma)$$

we have

$$\sqrt{\log(\tilde{h}^{-1})} = \sqrt{\gamma\log(\varepsilon^{-1})}(1+o(1)).$$

We conclude that

$$\mathrm{med}(\varepsilon\|Z_\varepsilon\|_\infty) = \tilde{\varepsilon}\|\psi_{\tilde{h}}\|_2(1+o(1)), \quad \varepsilon\to 0, \tag{4.5}$$

or, equivalently,

$$\mathrm{med}(\varepsilon\|Z_\varepsilon\|_\infty) \sim S_{\tilde{\varepsilon}}(\tilde{T}), \quad \varepsilon\to 0.$$

This equation is crucial: the definition (3.3) of the pseudo-noise-level $\tilde{\varepsilon}$ was chosen expressly to make this relation true. It is important to note that $(1+o(1))$ in (4.5) stands for a term which is $(1+o(1))$ independent of $f\in\mathscr{F}$.

4.3 Bounds on $\operatorname{med}(\|\tilde{f}^{(k)}-f^{(k)}\|_\infty)$

In the optimal recovery model, we have the general relation for $\hat{T}(y)=\int \psi(t)y(t)\,dt$ that

$$\operatorname{Err}_\varepsilon(\hat{T},f)=|B(\hat{T},f)|+S_\varepsilon(\hat{T}).$$

From the definition of $\psi_{\tilde{h}}$ as minimax for noise level $\tilde{\varepsilon}$, we have

$$E^*(\tilde{\varepsilon})=\sup_{\mathscr{F}} B(\tilde{T},f)+S_{\tilde{\varepsilon}}(\tilde{T}). \tag{4.6}$$

We now combine the analysis above:

$$\begin{aligned}\operatorname{med}(\|\tilde{f}^{(k)}-f^{(k)}\|_\infty) &\leqq \|\operatorname{Bias}\|_\infty+\operatorname{med}(\varepsilon\|Z_\varepsilon\|_\infty)\\ &\leqq \sup_{\mathscr{F}} B(\tilde{T},f)+S_{\tilde{\varepsilon}}(\tilde{T})(1+o(1))\\ &\leqq \left\{\sup_{\mathscr{F}} B(\tilde{T},f)+S_{\tilde{\varepsilon}}(\tilde{T})\right\}(1+o(1))\\ &= E^*(\tilde{\varepsilon})(1+o(1)).\end{aligned} \tag{4.7}$$

The second step follows from (4.2) and (4.5); the middle step follows from the fact that $(1+o(1))$ in (4.5) does not depend on $f\in\mathscr{F}$; the final step follows from our use of the optimal kernel, which guarantees (4.6).

4.4 Bounds on the tails of $\|\tilde{f}^{(k)}-f^{(k)}\|_\infty$

The median performance (4.7) of $\tilde{f}^{(k)}$ tells nearly the entire story. This is most easily seen using a special case of Borell's inequality; compare Talagrand (1988).

Proposition 4.2 *Let* $Z(t)$ *be a stationary, zero-mean Gaussian process, with* $\sigma=\sqrt{\operatorname{Var} Z(t)}$.

$$P\{\|Z\|_{L^\infty[0,A]}>\operatorname{med}(\|Z\|_{L^\infty[0,A]})+\sigma z\}\leqq 2P\{N(0,1)>z\},\quad z>0. \tag{4.8}$$

To apply this, note the following relationship: set

$$\sigma_{\tilde{h}}^2=\operatorname{Var}(\tilde{f}^{(k)}(t)-f^{(k)}(t))$$

(this does not depend on t). Then

$$\begin{aligned}\sigma_{\tilde{h}} &= \|\psi_{\tilde{h}}\|\cdot\varepsilon\\ &= \|\psi_1\|\tilde{h}^{-k-1/2}\varepsilon\\ &= \operatorname{Const}\cdot(\tilde{\varepsilon}^\gamma)^{-k-1/2}\cdot\varepsilon\\ &= \operatorname{Const}\cdot\varepsilon^r[\log(\varepsilon^{-1})]^{(-k-1/2)\gamma/2}\\ &= o(1)\varepsilon^r.\end{aligned}$$

Hence,

$$\sigma_{\tilde{h}}/E^*(\tilde{\varepsilon})=o(1). \tag{4.9}$$

Combining this with (4.7) gives, with Z a standard normal random variable, that

$$E\ell\left(\frac{\|\tilde{f}^{(k)}-f^{(k)}\|_\infty}{E^*(\tilde{\varepsilon})}\right)\leqq E\ell\left(\frac{E^*(\tilde{\varepsilon})(1+o(1))+\sigma_{\tilde{h}}|Z|}{E^*(\tilde{\varepsilon})}\right)$$
$$=E\ell(1+o(1)+o(1)|Z|) \tag{4.10}$$
$$\to\ell(1)$$

for loss functions $\ell(\cdot)$ satisfying assumption (L). As the $o(1)$ terms are independent of $f\in\mathscr{F}$, this proves Theorem A.
As an example, set $\ell(t)=|t|^p$, $p>0$; we have

$$\sup_{\mathscr{F}(\beta,C)} E\|\tilde{f}^{(k)}-f^{(k)}\|_\infty^p\leqq E^*(\tilde{\varepsilon})^p(1+o(1)).$$

5 Lower bounds via hypercubes

We now prove Theorem B. Our approach is to find the hardest cubical subproblem of $\mathscr{F}(\beta,\mathscr{C})$; we develop an optimization problem to find the hardest cubical subproblem and then apply Korostelev's lemma on the difficulty of standard hypercubes for the max-norm loss.

5.1 Optimization under support constraints

Consider now an optimization problem akin to $(P_{\varepsilon,C})$ only with the additional constraint that f be supported in $[-A,A]$.

$$(P_{\varepsilon,C,A}):\quad \sup f^{(k)}(0)\quad\text{subject to}\quad\begin{cases}\|f\|_2\leqq\varepsilon,\\ f\in\mathscr{F}(\beta,C),\\ \operatorname{supp}(f)\subset[-A,A].\end{cases} \tag{5.1}$$

Evidently, as every f feasible for $(P_{\varepsilon,C,A})$ is also feasible for $(P_{\varepsilon,C})$,

$$\operatorname{val}(P_{\varepsilon,C,A})\leqq\operatorname{val}(P_{\varepsilon,C}), \tag{5.2}$$

but not necessarily vice versa.
By methods of Donoho and Low (1992, Theorem 3) we can establish asymptotic equality.

Lemma 5.1 *Let $0\leqq k<\beta$. Then*

$$\operatorname{val}(P_{1,C,A})\to\operatorname{val}(P_{1,C}),\quad A\to\infty. \tag{5.3}$$

5.2 Minimax risk for hypercubes

Consider the following problem: we observe

$$y_i=\theta_i+\sigma_N Z_i,\quad i=0,\dots,N-1, \tag{5.4}$$

where $|\theta_i|\leqq\tau_N$ and $Z_i\sim\text{iid}\,N(0,1)$. We wish to estimate $(\theta_i)_{i=0}^{N-1}$, and evaluate success in max-norm loss $\|\hat{\theta}-\theta\|_{\ell_N^\infty}$. What is the minimax risk over the hypercube?

Let $\ell_N(t)=\ell_1(t/\tau_N)$ and set

$$\mathbf{m}^*(N)=\inf_{\hat{\theta}} \sup_{\|\theta\|_\infty \leqq \tau_N} E\ell_N(\|\hat{\theta}-\theta\|_{\ell_N^\infty}). \tag{5.5}$$

Proposition 5.2 (Korostelev 1991). *Let (τ_N, σ_N) be a sequence satisfying*

$$\frac{\tau_N}{\sigma_N} \leqq \sqrt{2-\eta}\sqrt{\log N} \tag{5.6}$$

for all sufficiently large N and some $\eta>0$. Then

$$\mathbf{m}^*(N) \to \ell_1(1). \tag{5.7}$$

In words, if τ_N is not too large relative to σ_N, the max-norm of any estimator's error is almost certain to be at least of size τ_N as well.

Korostelev's proof is analytical, but we sketch an intuitive argument which can easily be made rigorous. Because the coordinates are independent and the Gaussian law has monotone likelihood ratio, the minimax behavior is attained within the class of coordinatewise rules $\hat{\theta}_i=\delta_i(y_i)$ which are odd and monotone increasing. Hence, $y_i \geqq 0 \Rightarrow \hat{\theta}_i \geqq 0$. For any such rule $\hat{\theta}$, the event

$$\{\text{some } y_i \text{ has the opposite sign of the corresponding } \theta_i\}$$

implies, if all $|\theta_i|=\tau_N$,

$$\{\|\hat{\theta}-\theta\|_{\ell_N^\infty} \geqq \tau_N\}.$$

If we use the coin tossing prior θ_i iid with $\{+\tau_N, -\tau_N\}$ equiprobable, then the former event is essentially the same as

$$\{\sigma_N \|Z\|_{\ell_N^\infty} \geqq \tau_N\},$$

where $Z=(Z_i)_{i=0}^N$ iid $N(0,1)$. Elementary results for Normal extremes (see again Leadbetter et al. 1983) show that if $c_N \leqq \sqrt{2-\eta}\sqrt{\log N}$ then

$$P\{\|Z\|_{\ell_N^\infty} \geqq c_N\} \to 1$$

as $N \to \infty$. Hence, if $\tau_N/\sigma_N \leqq \sqrt{2-\eta}\sqrt{\log N}$ we are practically certain, for large N, that a coordinatewise application of odd, monotone rules will make an error of size τ_N in some coordinate.

5.3 *Cubical subproblems of $\mathscr{F}(\beta, C)$*

Fix $A>0$. Let f_0 be a function supported in $[-A/2, A/2]$ and satisfying $f_0 \in \mathscr{F}(\beta, C)$. Set $\phi_i(t)=f_0(t-A(i+1/2))$. Then for any sequence (s_i) of multipliers $|s_i| \leqq 1$ we have

$$\sum_{i=-\infty}^{\infty} s_i \phi_i \in \mathscr{F}(\beta, C)$$

as well. Let $\phi_{i,M}(t)=M^{-\beta}\phi_i(Mt)$, for M, an integer $\geqq 1$. Then, similarly

$$\sum_{i=-\infty}^{\infty} s_i \phi_{i,M} \in \mathscr{F}(\beta, C).$$

In words, $\mathscr{F}(\beta, C)$ contains the hypercube generated by the vertices $\phi_{i,M}$.

Suppose now that $N \cdot A/M \leqq 1$. There is an N-dimensional hypercube $\mathscr{C}(f_0, N, \beta)$ defined by the finite sum

$$f = \sum_{i=0}^{N-1} \theta_i \phi_{i,M}, \tag{5.8}$$

where each $|\theta_i| \leqq 1$. Suppose we have white-noise data (2.1), and that we wish to estimate $f^{(k)}$ in supremum norm. By an argument based on sufficiency, a complete class of estimators for estimating $f \in \mathscr{C}(f_0, N, \beta)$ consists of all procedures of the form

$$\hat{f}^{(k)} = \sum_{i=0}^{N-1} \hat{\theta}_i \phi_{i,M}^{(k)},$$

where

$$\hat{\theta}_i = \delta_i(y_0, \ldots, y_{N-1}), \quad i = 0, \ldots, N-1$$

for measurable functions $\delta_i(\cdot)$ and

$$y_i = \int \xi_i(t) Y(dt), \quad i = 0, \ldots, N-1$$

with $\xi_i = \phi_{i,M} / \| \phi_{i,M} \|_2^2$.
Let $\hat{f}^{(k)}$ be any such estimate. Then putting $t_i = A(i+1/2)/M$, $i = 0, \ldots, N-1$,

$$\begin{aligned} \| \hat{f}^{(k)} - f^{(k)} \|_\infty &\geqq \max_{i=0,\ldots,N-1} |\hat{f}^{(k)}(t_i) - f^{(k)}(t_i)| \\ &= |\phi_{0,M}^{(k)}(t_0)| \, \| \hat{\theta} - \theta \|_{\ell_N^\infty} \\ &= M^{k-\beta} T_k(f_0) \| \hat{\theta} - \theta \|_{\ell_N^\infty}. \end{aligned}$$

If we define the loss function

$$\ell_M(t) = \ell_1(M^{\beta - k} t)$$

then, provided ℓ_1 is increasing and continuous, we get

$$\inf_{\hat{f}} \sup_{\mathscr{C}(N, f_0, \beta)} E \ell_M(\| \hat{f}^{(k)} - f^{(k)} \|_\infty) \geqq \inf_{\hat{\theta}} \sup_{|\theta| \leqq 1} E \ell_1(T_k(f_0) \| \hat{\theta} - \theta \|_{\ell_N^\infty}). \tag{5.9}$$

In short, the risk over the Hypercube gives a lower bound on the minimax risk over $\mathscr{C}(f_0, N, \beta)$. Suppose now that with A fixed, we pick M depending on ε in such a way that $M(\varepsilon) \to \infty$ as $\varepsilon \to 0$. Let $N(\varepsilon)$ be the largest integer satisfying $N \cdot A/M(\varepsilon) \leqq 1$. We may apply Korostelev's Lemma with

$$\tau_N = \sup\{ \| \theta \|_\infty \} = 1, \qquad \sigma_N^2 = \mathrm{Var}(y_i) = \varepsilon^2 / \| \phi_{i,M} \|_2^2;$$

the condition

$$\sigma_N^{-1} \leqq \sqrt{2 - \eta} \sqrt{\log N}$$

implies, by Proposition 5.2 above,

$$\underline{\lim}_{\varepsilon \to 0} \inf_{\hat{f}} \sup_{\mathscr{C}(N, f_0, \beta)} E \ell_M(\| \hat{f}^{(k)} - f^{(k)} \|_\infty) \geqq \ell_1(T_k(f_0)). \tag{5.10}$$

5.4 Hardest cubical subproblems

We now choose f_0 so that (5.10) is most effective. We first address the value of $M(\varepsilon)$. We need to have $\sigma_N^{-1} \leqq \sqrt{2-\eta}\sqrt{\log(N)}$. Equivalently,

$$(\varepsilon/\|\phi_{0,M}\|_2)^{-1} \leqq \sqrt{2-\eta}\sqrt{\log N}$$

or

$$\varepsilon^{-1} \cdot M^{-\beta-1/2} \cdot \|f_0\|_2 \leqq \sqrt{2-\eta}\sqrt{\log N}.$$

We now adopt the normalization $\|f_0\|_2 = 1$. Hence, we need to choose $M(\varepsilon)$ so that

$$M^{-\beta-1/2} \leqq \sqrt{2-\eta}\sqrt{\log N}\varepsilon. \tag{5.11}$$

Let, as before, $\gamma = 1/(\beta+1/2)$. For an appropriate choice of δ, depending on $\eta > 0$, and γ, (5.11) can be arranged by picking

$$M^{-1} \sim \left(\sqrt{2\gamma-\delta}\sqrt{\log(\varepsilon^{-1})}\varepsilon\right)^{\gamma}.$$

If M is chosen in this way, then for all sufficiently small $\varepsilon > 0$ and each fixed $A > 0$ it satisfies (5.11). Moreover, there is no essentially better choice of M: anything essentially smaller (e.g. by constant multiples significantly different from 1) would fail to satisfy (5.11).
This choice of $M(\varepsilon)$ gives

$$M^{k-\beta} \sim \tilde{\varepsilon}^r \left(\frac{2\gamma}{2\gamma-\delta}\right)^{-r/2}$$

and so

$$M^{k-\beta}\left(\frac{2\gamma}{2\gamma-\delta}\right)^{r/2} \operatorname{val}(P_{1,c}) \sim E^*(\tilde{\varepsilon}), \quad \varepsilon \to 0. \tag{5.12}$$

Suppose that, given a loss function $\ell(t)$, we define $\ell_1(t)$ so that

$$\ell_1\left(t \cdot \left(\frac{2\gamma}{2\gamma-\delta}\right)^{-r/2} \cdot \operatorname{val}(P_{1,c})^{-1}\right) = \ell(t).$$

Then, as $M(\varepsilon)$ obeys (5.12),

$$\ell\left(\frac{\|\hat{f}^{(k)} - f^{(k)}\|_\infty}{E^*(\tilde{\varepsilon})}\right) \approx \ell_1(M^{\beta-k}\|\hat{f}^{(k)} - f^{(k)}\|_\infty).$$

Recalling the convention

$$\ell_1(M^{\beta-k}\|\hat{f}^{(k)} - f^{(k)}\|_\infty) = \ell_M(\|\hat{f}^{(k)} - f^{(k)}\|_\infty)$$

and applying the bound (5.10), we reach the following conclusion:

Let $\|f_0\|_2 = 1$, let f_0 be supported in $[-A/2, A/2]$, and let M satisfy (5.12). *Let ℓ be a continuous increasing loss function. Then we have the lower bound*

$$\underline{\lim}_{\varepsilon \to 0} \inf_{\hat{f}^{(k)}} \sup_{\mathcal{F}} E\ell\left(\frac{\|\hat{f}^{(k)} - f^{(k)}\|_\infty}{E^*(\tilde{\varepsilon})}\right) \geqq \ell_1(|T_k(f_0)|)$$

$$= \ell\left(\frac{|T_k(f_0)|}{\operatorname{val}(P_{1,c})} \cdot \left(\frac{\sqrt{2\gamma-\delta}}{\sqrt{2\gamma}}\right)^r\right).$$

To make this bound most effective, we should choose f_0 to maximize $T_k(f_0)$ subject to the constraints we have assumed for it, namely, $\|f\|_2 \leqq 1$, $f \in \mathscr{F}(\beta, C)$, $\text{supp}(f) \in [-A/2, A/2]$. When we do this, we will have achieved the most clever application of the hypercube lower bound.

However, the optimization required is the same as requiring that f_0 solve $(P_{1,C,A/2})$. When we choose f_0 in this way, we get the bound

$$\underline{\lim}_{\varepsilon \to 0} \mathscr{M}^*(\varepsilon) \geqq \ell\left(\frac{\text{val}(P_{1,C,A/2})}{\text{val}(P_{1,C})} \cdot \left(\frac{\sqrt{2\gamma - \delta}}{\sqrt{2\gamma}}\right)^r\right). \tag{5.13}$$

This bound is valid for each $\delta > 0$; continuity of $\ell(\cdot)$ furnishes the conclusion

$$\underline{\lim}_{\varepsilon \to 0} \mathscr{M}^*(\varepsilon) \geqq \ell\left(\frac{\text{val}(P_{1,C,A/2})}{\text{val}(P_{1,C})}\right). \tag{5.14}$$

Picking A large, Lemma 5.1 shows that the right-hand side can be made arbitrarily close to $\ell(1)$. Theorem B is now proved.

6 Data available only on $[0, 1]$

The alert reader may have noted that the result of Sect. 3 applies to the problem where data $Y(t)$ are available for all $t \in (-\infty, \infty)$, but only for $t \in [0, 1]$ do we attempt to reconstruct f. We now suppose that data are available only on $[0, 1]$:

$$Y(dt) = f(t)\,dt + \varepsilon W(dt), \quad t \in [0, 1]. \tag{6.1}$$

Theorem C. *For the problem of estimation of $f(t)$, $t \in [0, 1]$, from data $Y(t)$, $t \in [0, 1]$, we have, under the same interpretation of $\|\cdot\|_\infty$, $\ell(\cdot)$, and E^* as in Theorem A, the same conclusions about the risk asymptotics and bounds. However, the estimator which attains the asymptotics is no longer of pure convolutional form but instead takes the form*

$$\hat{f}^{(k)}(t) = \int \Psi_\varepsilon(s, t)\, Y(ds) \tag{6.2}$$

for a kernel Ψ_ε to be specified below, which is supported on $(s, t) \in [0, 1]^2$ for sufficiently small ε.

Theorem C is the counterpart for this setting of Theorem A; it furnishes upper bounds. There is no need for a counterpart of Theorem B, because Theorem B holds in the present setting (6.1) – intuitively, this is because "throwing away" data $Y(t)$, $t \notin [0, 1]$, only makes the estimation problem more difficult. The next two subsections describe the construction of the kernel Ψ_ε. The proof of Theorem C is given in the Appendix.

6.1 Optimal kernels under support constraints

A modification of $(P_{1,1})$ allows us to derive kernels which are optimal under support constraints. Compare also the section on "boundary kernels" in Donoho

and Low (1992). Let $a \leqq 0$, $A \geqq 0$, $A - a > 0$. Define the optimization problem

$$(P_{\varepsilon,C}[a,A]): \quad \sup T_k(f) \text{ subject to } \begin{cases} \int_a^A f^2 \leqq \varepsilon^2, \\ f \in \mathscr{F}(\beta, C). \end{cases}$$

As $\int_a^A f^2 \leqq \int_{-\infty}^{\infty} f^2$, every f feasible for $(P_{\varepsilon,C})$ is also feasible for $(P_{\varepsilon,C}[a,A])$ and

$$\text{val}(P_{\varepsilon,C}) \leqq \text{val}(P_{\varepsilon,C}[a,A]). \tag{6.3}$$

In the cases we usually have in mind, $[a,A]$ is a compact interval. However, we remark that $a = -\infty$, $A = +\infty$ gives us $(P_{\varepsilon,C})$; and $a = 0$, $A = +\infty$ gives us an optimization problem on the half-line.

Note well that there is no restriction for the solution to be supported in $[a,A]$. This problem is therefore different from (5.1). Compare (5.2) and (6.3).

Let f_1 be a solution to the problem. f_1 may be used to construct an optimal kernel $\psi_1^{[a,A]} = \lambda f_1 1_{[a,A]}$ for a certain $\lambda = \lambda(\beta, C, k, a, A)$. Optimality is in the sense of the optimal recovery theorem, see Donoho (1994) and literature referred to there, such as Micchelli and Rivlin (1977). Suppose we are given data $y(t)$, $t \in [a,A]$ and that

$$y(t) = f(t) + \varepsilon z(t), \quad t \in [a,A]$$

where f is an unknown element of $\mathscr{F}(\beta, C)$ and $\int_a^A z^2(t)\,dt \leqq 1$. Our objective is to develop a procedure $\hat{T}_\varepsilon(y)$ attaining the minimax error

$$E^*(\varepsilon; [a,A]) = \inf_{\hat{T}(y)} \sup_{\mathscr{F}(\beta,C)} \text{Err}_\varepsilon(\hat{T}, f).$$

The kernel ψ_1 is optimal for this problem at noise level 1: with $\hat{T}_1(y) = \int_a^A \psi_1(t) y(t)\,dt$ we have

$$\sup_{\mathscr{F}(\beta,C)} \text{Err}_1(\hat{T}_1, f) = E^*(1; [a,A]).$$

Moreover, $\text{val}(P_{1,C}[a,A])$ is the minimax error $E^*(1;[a,A]) = \text{val}(P_{1,C}[a,A])$. An easy renormalization argument shows that the kernel $\psi_h^{[a,A]}(t) = h^{-k-1}\psi_1^{[a,A]}(t/h)$ with bandwidth $h = \varepsilon^\gamma$ gives a procedure $\hat{T}_\varepsilon = \int_{ah}^{Ah} \psi_h(t) y(t)\,dt$ with worst-case error

$$\sup_{\mathscr{F}(\beta,C)} \text{Err}_\varepsilon(\hat{T}_\varepsilon, f) = \text{val}(P_{1,C}[a,A])\varepsilon^r, \quad \varepsilon > 0.$$

Moreover, this behavior is optimal among kernels with support limited to $[ah, Ah]$, since renormalization shows that

$$E^*(\varepsilon; [ah, Ah]) = \text{val}(P_{1,C}[a,A])\varepsilon^r.$$

Hence, the family of kernels $\psi_h^{[a,A]}$ is in some sense optimal among compactly supported kernels.

In this connection, it is interesting to study the behavior of $\text{val}(P_{1,C}[a,A])$ as a function of $[a,A]$. By methods of Donoho and Low (1992, Theorem 3) one may prove the following lemma.

Lemma 6.1.

$$\mathrm{val}(P_{1,C}[-A,A]) \to \mathrm{val}(P_{1,C}), \quad A \to \infty, \tag{6.4a}$$

$$\mathrm{val}(P_{1,C}[0,A]) \to \mathrm{val}(P_{1,C}[0,\infty)), \quad A \to \infty, \tag{6.4b}$$

$$\mathrm{val}(P_{1,C}[-A,0]) \to \mathrm{val}(P_{1,C}(-\infty,0]), \quad A \to \infty. \tag{6.4c}$$

Since $E^*(\varepsilon) = \mathrm{val}(P_{1,C})\varepsilon^r$, these relations show that for sufficiently large A,

$$E^*(\varepsilon; [-A\varepsilon^\gamma, A\varepsilon^\gamma]) < (1+\eta)E^*(\varepsilon), \quad \varepsilon > 0, \tag{6.5a}$$

$$E^*(\varepsilon; [0, A\varepsilon^\gamma]) < (1+\eta)E^*(\varepsilon; [0,\infty)), \quad \varepsilon > 0, \tag{6.5b}$$

$$E^*(\varepsilon; [-A\varepsilon^\gamma, 0]) < (1+\eta)E^*(\varepsilon; [0,\infty)), \quad \varepsilon > 0. \tag{6.5c}$$

6.2 Construction of the kernel Ψ_ε

We now describe the kernel Ψ_ε of Theorem C. With parameters $\eta \in (0, 1/2)$ and $A \gg 0$ selected as described later, we set $h = \varepsilon^\gamma$ and

$$\Psi_\varepsilon(s,t) = \begin{cases} \psi_h^{[0,A]}(s-t), & 0 \leqq t < \eta, \\ \psi_h^{[-A,A]}(s-t), & \eta \leqq t \leqq 1-\eta, \\ \psi_h^{[-A,0]}(s-t), & 1-\eta < t \leqq 1. \end{cases}$$

The kernel results from splicing together the boundary kernels $\psi_h^{[0,A]}$ and $\psi_h^{[-A,0]}$ at the "edges" of $[0,1]$ with a traditional compactly supported kernel $\psi_h^{[-A,A]}$ in the "middle" of $[0,1]$.

The key point is that as soon as $\varepsilon^\gamma A < \eta$, the kernel Ψ_ε is supported in $[0,1]^2$. The key parameters η and A are arrived at as follows. First, we pick η so small that

$$\eta^{(k+1)r}\,\mathrm{val}(P_{1,C}[0,\infty]) < (1-\eta)\mathrm{val}(P_{1,C}). \tag{6.6}$$

Second, we pick $A = A(\eta)$ so that

$$\mathrm{val}(P_{1,C}[-A,A]) < (1+\eta)\mathrm{val}(P_{1,C}) \tag{6.7}$$

and also

$$\eta^{(k+1)r}\,\mathrm{val}(P_{1,C}[0,A]) < (1-\eta)\mathrm{val}(P_{1,C}). \tag{6.8}$$

These choices are all possible because of the Lemma 6.1 given above, (6.4) and (6.5).

6.3 Proof of Theorem C

The proof, given in the Appendix, uses a set of simple calculations based heavily on earlier techniques, and (6.6)–(6.8).

7 Nonparametric regression

We now consider an apparently different problem: we observe n noisy samples

$$y_i = f(t_i) + \sigma z_i, \quad i = 1, \ldots, n \tag{7.1}$$

with $z_i \overset{\text{iid}}{\sim} N(0,1)$ and $t_i = i/n$. We wish to estimate f with sup-norm loss. In the special case $k=0$, $\beta \leqq 1$, this is the problem originally considered by Korostelev (1991).

Despite appearances, we will see that this is essentially the same as the white-noise model with $\varepsilon = \sigma/\sqrt{n}$. For this section only, set $\tilde{\varepsilon}_n = \tilde{\varepsilon}(\varepsilon) = \sqrt{2\gamma}\sqrt{\log(\sqrt{n}/\sigma)}\sigma/\sqrt{n}$, and define the minimax risk

$$\begin{aligned}\mathscr{M}(n) &\equiv \mathscr{M}(n; \sigma, \beta, C, k) \\ &\equiv \inf_{\hat{f}_n} \sup_{\mathscr{F}(\beta,C)} E\ell\left(\frac{\|\hat{f}_n - f\|_\infty}{E^*(\tilde{\varepsilon}_n)}\right).\end{aligned}$$

Theorem D. $0 \leqq k < \beta$, $\beta > 1/2$.

$$\lim_{n\to\infty} \mathscr{M}(n) = \ell(1).$$

Informally, there exists an estimator having

$$\|\hat{f}_n^{(k)} - f^{(k)}\|_\infty \leqq E^*(\tilde{\varepsilon}_n)(1 + o(1)) \tag{7.2}$$

with very high probability, and no estimator can do substantially better than this.

Written in the form (7.2), the risk asymptotics are seen to be again closely connected with the problem of optimal recovery.
It is instructive to express this result in terms of n.

$$\begin{aligned}E^*(\tilde{\varepsilon}_n) &= \mathrm{val}(P_{1,1}) C^{1-r} \tilde{\varepsilon}_n^r \\ &\sim \mathrm{val}(P_{1,1}) C^{1-r} (\sqrt{\gamma})^r \left(\frac{\log(n)}{n}\right)^{r/2} \cdot \sigma^r \\ &= \Delta_{\beta,k} C^{1-r} \cdot \left(\frac{\log(n)}{n}\right)^{r/2} \cdot \sigma^r, \text{ say.}\end{aligned}$$

Corollary *Let* $0 \leqq k < \beta$, $1/2 < \beta$. *Set*

$$\Delta_{\beta,k} = \mathrm{val}(P_{1,1})(\sqrt{\gamma})^r \tag{7.3}$$

as above. Then

$$\inf_{\hat{f}_n} \sup_{\mathscr{F}(\beta,C)} E\ell\left(\|\hat{f}_n - f\|_\infty \left(\frac{n}{\sigma^2 \log(n)}\right)^{\beta - k/(2\beta+1)}\right) \sim \ell(\Delta_{\beta,k} C^{1-r}).$$

In the case $k=0$, $1/2 < \beta \leqq 1$, we get

$$\begin{aligned}\Delta_{\beta,0} &= \mathrm{val}(P_{1,1})(\sqrt{\gamma})^r \\ &= [(2\beta+1)(\beta+1)/4\beta^2]^{r/2} \cdot \left(\frac{2}{2\beta+1}\right)^{r/2} \\ &= \left(\frac{(\beta+1)}{2\beta^2}\right)^{\beta/(2\beta+1)},\end{aligned}$$

and so we recover the result of Korostelev (1991):

$$\inf_{\hat{f}_n} \sup_{\mathscr{F}(\beta,C)} E\ell\left(\|\hat{f}_n - f\|_\infty \left(\frac{n}{\sigma^2 \log(n)}\right)^{\beta/(2\beta+1)}\right) \sim \ell\left(\left(\frac{(\beta+1)}{2\beta^2}\right)^{\beta/(2\beta+1)} C^{\frac{1}{2\beta+1}}\right).$$

We now recognize that Korostelev's "optimal constant" $((\beta+1)/2\beta^2)^{\beta/(2\beta+1)}$ for the problem of nonparametric regression derives from the optimal constant for optimal recovery. Of course, our approach also gives results for a general range of k and β; the optimal constant for any of these problems is determined, through (7.3), by $\mathrm{val}(P_{1,1})$, the optimal constant in the corresponding optimal recovery problem.

Theorem D follows from a general principle: that white-noise data (3.1) and nonparametric regression data (7.1) are essentially equivalent, when noise levels are calibrated so that $\varepsilon = \sigma/\sqrt{n}$.

Theorem 7.1 *Let $0 \leq k < \beta$, $\beta > 1/2$. Then for every bounded loss function ℓ,*

$$|\mathscr{M}(n) - \mathscr{M}^*(\sigma/\sqrt{n})| \to 0, \quad n \to \infty. \tag{7.4}$$

This equivalence principle may be established as follows. The minimax risk for observations on $[0,1]$ is asymptotically equivalent to the minimax risk for observations on $(-\infty, \infty)$, by Theorem C. So it is enough to show that observations (6.1) are equivalent to observations (7.1), from the point of view of minimax risk.

To do this we use the following criterion, due to Brown and Low (1994). Given a continuous function f on $[0,1]$, let f_n denote the step function approximation

$$f_n(t) = \sum_{i=1}^{n} f(t_i)\, 1_{\{t_{i-1} < t \leq t_i\}}.$$

Proposition 7.2 (Brown and Low 1994). *Suppose that $\mathscr{F}$ is a class of continuous functions obeying*

$$\sup_{\mathscr{F}} n \int_0^1 (f(t) - f_n(t))^2\, dt \to 0, \quad n \to \infty$$

Let $\ell_n(\cdot,\cdot)$ be a measurable functional, bounded by M independently of n. Let

$$\mathbf{m}_n = \inf_{\hat{f}_n} \sup_{\mathscr{F}} E\ell_n(\hat{f}_n, f)$$

denote the minimax value in the nonparametric regression model (7.1); *and let*

$$\mathbf{m}_n^* = \inf_{\hat{f}_\varepsilon} \sup_{\mathscr{F}} E\ell_n(\hat{f}_\varepsilon, f)$$

denote the minimax value in the white-noise model (6.1), *at noise level $\varepsilon = \sigma/\sqrt{n}$. Then*

$$|\mathbf{m}_n - \mathbf{m}_n^*| \to 0, \quad n \to \infty.$$

We briefly mention the principle behind the proof. Define the process Y_n by

$$Y_n(i/n) = \frac{1}{n} \sum_{j \leq i} y_i.$$

Interpolate between these values with independent Brownian Bridges $W_{0,i}$: if $i/n < t < (i+1)/n$

$$Y_n(t) = Y_n(i/n) + (t - i/n) y_i + \frac{\sigma}{\sqrt{n}} W_{0,i}(n(t - i/n)).$$

Let P_n denote the probability law of $\{Y(t),\ t\in[0,1]\}$ at noise level $\varepsilon=\sigma/\sqrt{n}$: let Q_n denote the probability law of $\{Y_n(t),\ t\in[0,1]\}$. The Hellinger distance between these probabilities is

$$H^2(P_n,Q_n)=2-2\exp(-\|f-f_n\|_2^2/8\varepsilon^2).$$

Hence sufficient L^2-closeness of f and f_n, uniformly in $f\in\mathscr{F}$, implies the Hellinger closeness of the two probability laws P_n and Q_n, uniformly in $\mathscr{F}$. In particular, the Brown–Low condition $n\int(f(t)-f_n(t))^2\,dt\to 0$ uniformly in $\mathscr{F}$ implies $H^2(P_n,Q_n)\to 0$ uniformly in $\mathscr{F}$.

To see the implications, we use the language of Le Cam (1986). The problem of estimating properties of f from data $Y(t)$ according to (6.1) is a statistical experiment $\mathscr{E}_n^0=(Y,P_n,\mathscr{F})$. The problem of estimating from data (7.1) is another statistical experiment $\mathscr{E}_n^1=(Y_n,Q_n,\mathscr{F})$, with other data, but the same parameter space $\mathscr{F}$. Suppose that $H^2(P_n,Q_n)\to 0$ *uniformly* in $\mathscr{F}$. Then it follows from Le Cam's theory that the experiments are asymptotically indistinguishable: for every risk function available in one problem, there is an estimator giving essentially the same risk function in the other problem. In particular, the minimax risks must coincide asymptotically.

To apply this, we invoke the following lemma.

Lemma 7.3 *Let* $\beta\in(1/2,1]$. *Then*

$$\sup_{\mathscr{F}(\beta,C)} n\int(f(t)-f_n(t))^2\,dt\leqq C^2n^{-2(\beta-1/2)}.$$

Proof.

$$n\int(f-f_n)^2\,dt=n\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}(f(t)-f(t_i))^2\,dt$$

$$\leqq n\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}(Cn^{-\beta})^2\,dt=C^2n^{-2\beta}\cdot n.$$

We conclude that Brown and Low's criterion holds for $\beta\in(1/2,1]$, and Theorem 7.1 follows in this case.

For $\beta>1$, however, the criterion of Brown and Low (1994) does not hold. Intuitively, polynomials of extraordinarily large energy but vanishing β-Lipschitz seminorm prevent the criterion from working.

We develop a different argument. Every $f\in\mathscr{F}(\beta,C)$ can be written $f=\pi_0+f_0$ where π_0 is a polynomial of degree m, and f_0 is orthogonal (with respect to Lebesgue measure) to all polynomials of degree $\leqq m$. Let $\mathscr{F}_0(\beta,C)$ be the collection of all such f_0 arising from an $f\in\mathscr{F}(\beta,C)$. The following lemma is proved in the technical report of Donoho (1991).

Lemma 7.4 *For each* $\beta>1$, *there exists* $C'<\infty$ *with*

$$\mathscr{F}_0(\beta,C)\subset\mathscr{F}(1,C'C).$$

We conclude immediately that for estimating $f^{(k)}$, the experiments $\mathscr{E}_n^{0,0}=(Y,P_n,\mathscr{F}_0(\beta,C))$ and $\mathscr{E}_n^{1,0}=(Y_n,Q_n,\mathscr{F}_0(\beta,C))$ – with parameter spaces $\mathscr{F}_0(\beta,C)$ – are risk-equivalent as $n\to\infty$. Hence, they have the same asymptotic minimax risk for every bounded loss function.

To extend this conclusion to the experiments with full parameter space $\mathscr{F}(\beta, C)$, we need the concept of a polynomially equivariant estimator. Suppose that we have two types of data: $Y(dt)=f(t)\,dt+\varepsilon W(dt)$ and $\tilde{Y}(dt)=(f(t)+\pi_m(t))\,dt+\varepsilon W(dt)$, where π_m is a polynomial of degree m, and the two noise terms are the *same realization*. Then the estimator $\hat{T}(Y)$ is called *polynomially equivariant* if

$$\hat{T}(\tilde{Y})-\hat{T}(Y)=T_k(\pi_m)$$

for every such polynomial π_m of degree m. Similarly, if we have sampled data

$$y_i=f(i/n)+\sigma z_i,$$

$$\tilde{y}_i=f(i/n)+\pi_m(i/n)+\sigma z_i$$

for $i=1,\ldots,n$, and again the noise realizations are the *same*, and if

$$\widehat{T_n}(\tilde{y})-\widehat{T_n}(y)=T_k(\pi_m)$$

for all such polynomials, then we say that $\widehat{T_n}$ is polynomially equivariant.

Now the minimax risk, both for experiment $\mathscr{E}_n^0=(Y, P_n, \mathscr{F}(\beta, C))$ and for experiment $\mathscr{E}_n^1=(Y, Q_n, \mathscr{F}(\beta, C))$ is attained within the class of polynomially equivariant estimates. (The argument is the usual one based on placing increasingly diffuse priors on π_m in the decomposition $f=\pi_m+f_0$. Compare the abstract discussion of the Hunt–Stein Theorem of Strasser (1985, Sect. 39).)

But polynomially equivariant estimates have the same worst-case risk over $\mathscr{F}_0(\beta, C)$ as they do over $\mathscr{F}(\beta, C)$! We have seen that the two experiments with $\mathscr{F}_0(\beta, C)$ as parameter space are asymptotically risk equivalent. The experiments with parameter space $\mathscr{F}(\beta, C)$ have the same minimax risk as the $\mathscr{F}_0(\beta, C)$ counterparts, so we conclude that the minimax risk of the experiments with parameter space $\mathscr{F}(\beta, C)$ are also asymptotically equivalent. This proves (7.4) for all bounded loss functions ℓ. Theorem 7.1 is now proved.

8 Generalizations

8.1 Deconvolution

Suppose that in the white-noise model we have observations

$$Y(dt)=(K*f)(t)\,dt+\varepsilon W(dt), \quad t\in(-\infty, \infty),$$

with

$$(K*f)(t)=\int_{-\infty}^{\infty} K(t-u)f(u)\,du$$

for some convolutional kernel $K(u)$. We wish to recover $f^{(k)}(t)$, $t\in[0,1]$, with L^∞-loss.

To analyze this, consider the analogous optimal recovery problem: we observe

$$y(t)=(K*f)(t)+\varepsilon z(t), \quad t\in(-\infty, \infty),$$

we know that $f\in\mathscr{F}(\beta, 1)$, and we wish to recover $T_k(f)=f^{(k)}(0)$. The minimax error $E^*(\varepsilon)$ is the value of a certain optimization problem $(P_{\varepsilon,C})$:

$$(P_{\varepsilon,C}): \quad \sup T_k(f) \quad \text{subject to} \begin{cases} \|K*f\|_2 \leqq \varepsilon, \\ f\in\mathscr{F}(\beta, C). \end{cases}$$

The problem, however, no longer renormalizes exactly. As in Donoho and Low (1992), if the Kernel K behaves like a power law in the frequency domain, $\hat{K}(\omega) \sim |\omega|^{-a}$, as $|\omega| \to \infty$, then we have a kind of asymptotic renormalization, and

$$\mathrm{val}(P_{\varepsilon,C}) \sim \mathrm{val}(P^*_{1,1}) \varepsilon^r C^{1-r}, \quad \varepsilon \to 0,$$

where now

$$r = \frac{2\beta - 2k - 2a}{2\beta + 1}$$

and we use the optimization problem

$$(P^*_{1,1}): \quad \sup T_k(f) \quad \text{subject to} \begin{cases} \| K * f \|_2 \leqq 1, \\ f \in \mathscr{F}(\beta, 1), \end{cases}$$

with $\hat{K}^*(\omega) = |\omega|^{-a}$ the Fourier transform of a homogeneous Schwartz distribution which one might call "the equivalent renormalizing kernel".
In general, one obtains, without difficulty, results such as

$$\sup_{\mathscr{F}(\beta,C)} E \| \tilde{f}^{(k)} - f^{(k)} \|_\infty \leqq E^*(\tilde{\varepsilon})(1 + o(1)),$$

only now for an estimator $\tilde{f}$ which is not defined by renormalization of a fixed kernel; instead, one needs a kernel which derives from the solution of $(P_{\tilde{\varepsilon},C})$, and so is changing slightly in shape as ε decreases. The asymptotic shape of the kernel is proportional to the solution of $(P^*_{1,1})$.

Other questions, such as optimality of the estimator, are problematic, and require treatment elsewhere.

8.2 Higher dimensions

Theorem A generalizes easily to the case of observations

$$Y(dt) = f(t)dt + \varepsilon W(dt), \quad t \in R^d$$

with W a d-dimensional Brownian sheet, where we know that $f \in \mathscr{F}_d(\beta, C)$, the β-Lipschitz class on R^d. We let $f^{(k)}$ denote some specific partial derivative of index k. In this case the optimal recovery model is

$$y(t) = f(t) + \varepsilon z(t), \quad t \in R^d,$$

and $E^*(\varepsilon)$ is the value of the optimization problem

$$(P_{\varepsilon,C}): \quad \sup f^{(k)}(0) \quad \text{subject to} \begin{cases} \| f \|_{L^2(R^d)} \leqq \varepsilon, \\ f \in \mathscr{F}_d(\beta, C). \end{cases}$$

First, the renormalization lemma goes through for d-dimensional t, giving the same results, but with different values for the exponents $\gamma = 2/(2\beta + d)$, $r = (2\beta - 2k)/(2\beta + d)$. Second, we should define our estimate in convolutional form, using a kernel which derives from the optimal recovery problem at a certain noise level

$$\tilde{\varepsilon}_d(\varepsilon) = \varepsilon \sqrt{2d\gamma} \sqrt{\log(\varepsilon^{-1})}.$$

Without much additional work we get that the estimator defined in this way obtains the analog of Theorem A, and so for example

$$\sup_{\mathscr{F}(\beta,C)} E\|\tilde{f}^{(k)}-f^{(k)}\|_\infty \leqq E^*(\tilde{\varepsilon}_d)(1+o(1)).$$

This claim can be justified as follows. The generalization of Proposition 4.1 to maxima of stationary Gaussian random fields has been made by Bickel and Rosenblatt (1973) and by Qualls and Watanabe (1973); see Adler (1981) for an expository treatment. For a stationary Gaussian field $Z(t)=\int \psi(u-t)W(du)$, W a Brownian sheet on R^d, ψ sufficiently regular, we have

$$\mathrm{med}(\|Z\|_{L^\infty[0,A]^d}) \sim \sqrt{2\log(A^d)}\sigma$$

as $A\to\infty$, where $\sigma=\|\psi\|_2$. The key equation which must hold for our pattern of reasoning to go through is

$$\mathrm{med}(\varepsilon\|Z_\varepsilon\|_\infty) \sim S_{\tilde{\varepsilon}_d}(\tilde{T})$$

which can be rewritten as

$$\varepsilon\sqrt{2\,d\gamma\log(\varepsilon^{-1})}\|\psi_{\tilde{h}}\|_2 \sim \tilde{\varepsilon}_d\|\psi_{\tilde{h}}\|_2,$$

hence the justification of our calibration.

Incidentally, the analog of Theorem B can be carried through simply by following the proof in Sect. 5 step-by-step and making obvious adjustments here and there. Hence analogs of Theorems A and B hold in dimensions $d>1$.

9 Discussion

9.1 Comparison to Korostelev's work

Korostelev achieved an important breakthrough by deriving precise asymptotics for the minimax risk in L^∞-norm loss in the nonparametric regression problem. His lower bound via hypercubes is a particularly striking innovation.

However, Korostelev did not have available to him the concept of the optimization problem $(P_{1,1})$, of optimal recovery, of optimal kernels, and of renormalization. By introducing these concepts into the subject we have been able to get results covering $(k\geqq 0,\ \beta>1)$, and to establish that the optimal constants in this statistical estimation problem derive from the theory of optimal recovery.

9.2 Directions for Improvement

9.2.1 Solutions of $(P_{1,1})$ It would also be useful to have more information about solutions of $(P_{1,1})$ – explicit formulas for the knots of the piecewise polynomials, and explicit formulas for the coefficients of the polynomials.

9.2.2 Other problems The success obtained in this setting makes it seem plausible that a wide variety of minimax problems in sup-norm could be treated by these methods. A next candidate for treatment would be the case where the noise process Z_ε is nonstationary and possesses a unique point of maximal variance, in which case the results of Talagrand (1988) would apply.

Appendix: proof of Theorem C

We consider behavior in the three separate zones $[0,\eta],[\eta,1-\eta],[1-\eta,1]$ of $[0,1]$.
First, the middle zone. By repeating the argument of Sect. 4, we easily conclude that

$$\limsup_{\varepsilon\to 0}\sup_{\mathscr{F}} E\ell\left(\frac{\|\hat{f}^{(k)}-f^{(k)}\|_{L^\infty(\eta,1-\eta)}}{E^*(\tilde{\varepsilon};[-A\varepsilon^\gamma,A\varepsilon^\gamma])}\right)\leqq\ell(1).$$

As $E^*(\tilde{\varepsilon};[-A\varepsilon^\gamma,A\varepsilon^\gamma])\leqq(1+\eta)E^*(\tilde{\varepsilon})$, we conclude that

$$\limsup_{\varepsilon\to 0}\sup_{\mathscr{F}} E\ell\left(\frac{\|\hat{f}^{(k)}-f^{(k)}\|_{L^\infty(\eta,1-\eta)}}{E^*(\tilde{\varepsilon})}\right)\leqq\ell(1+\eta). \tag{A.1}$$

Now we consider the zone $[0,\eta)$. We remark that the process $\tilde{Y}(t)=Y(t/\eta)$ lives on $[0,1]$, and that

$$\tilde{Y}(dt)=_D f(t/\eta)dt/\eta+\varepsilon\sqrt{\eta}\,W(dt).$$

Hence, putting $g(t)=f(t/\eta)/\eta$ and $\delta=\varepsilon\sqrt{\eta}$ we may write

$$\tilde{Y}(dt)=_D g(t)dt+\delta W(dt).$$

Consider now the problem of estimating $g^{(k)}$ from the data $\tilde{Y}$. As $g\in\mathscr{F}(\beta,C\eta^{-\beta-1})$ and the noise level is δ, we argue as in Sect. 4 that if we used a one-sided kernel supported in $[0,A]$, with bandwidth $h=(\delta/\eta^{-\beta-1})^\gamma$, and if all data $\tilde{Y}(t)$, $t\geqq 0$ were available, we would have

$$\overline{\lim}_{\varepsilon\to 0}\sup_{\mathscr{F}} E\ell\left(\frac{\|\hat{g}^{(k)}-g^{(k)}\|_\infty}{\bar{E}(\tilde{\delta})}\right)\leqq\ell(1),$$

where $\bar{E}(\varepsilon)=\mathrm{val}(P_{1,C}[0,A])(\eta^{-\beta-1})^{1-r}\cdot\varepsilon^r$ and $\tilde{\delta}=\sqrt{2\gamma}\sqrt{\log(\delta^{-1})}\cdot\delta$. On the other hand

$$\eta^{k+1}\|\hat{g}^{(k)}-g^{(k)}\|_{L^\infty[0,1]}=\|\hat{f}^{(k)}-f^{(k)}\|_{L^\infty[0,\eta]}$$

and so

$$\overline{\lim}_{\varepsilon\to 0}\sup_{\mathscr{F}} E\ell\left(\frac{\|\hat{f}^{(k)}-f^{(k)}\|_{L^\infty[0,\eta]}}{\eta^{k+1}\bar{E}(\tilde{\delta})}\right)\leqq\ell(1). \tag{A.2}$$

By the assumption (6.6),

$$\eta^{k+1}\,\mathrm{val}(P_{1,C}[0,A])(\eta^{-\beta-1})^{1-r}\tilde{\delta}^r\leqq(1-\eta)\mathrm{val}(P_{1,C})\tilde{\varepsilon}^r(1+o(1))$$

as $\varepsilon\to 0$, and so

$$\overline{\lim}_{\varepsilon\to 0}\sup_{\mathscr{F}} E\ell\left(\frac{\|\hat{f}^{(k)}-f^{(k)}\|_{L^\infty[0,\eta)}}{E^*(\tilde{\varepsilon})}\right)\leqq\ell(1-\eta).$$

An exactly parallel argument handles the case $t \in (1-\eta, 1]$;

$$\overline{\lim}_{\varepsilon \to 0} \sup_{\mathscr{F}} E\ell\left(\frac{\|\hat{f}^{(k)} - f^{(k)}\|_{L^\infty(1-\eta,1]}}{E^*(\tilde{\varepsilon})}\right) \leqq \ell(1-\eta).$$

By technical means paralleling the development near (4.9) and (4.10), we can get that for $A_1 = [0, \eta)$, $A_2 = [\eta, 1-\eta]$, $A_3 = (1-\eta, 1]$

$$\overline{\lim}_{\varepsilon \to 0} \sup_{\mathscr{F}} E\ell\left(\frac{\max_{i=1,2,3} \|\hat{f}^{(k)} - f^{(k)}\|_{L^\infty(A_i)}}{E^*(\tilde{\varepsilon})}\right) \leqq \ell(1+\eta).$$

As $\eta > 0$ was arbitrary, this completes the proof of Theorem C.

Acknowledgements. R.Z. Khas'minskii, during his visit to Berkeley in September 1991, approached the author with the manuscript of Korostelev and posed the main questions answered in this article. The author thanks Catherine Carol-Huber, David Siegmund, and Joel Zinn for helpful information. This work was partially completed at the University of California, Berkeley, where it was supported by NASA Contract NCA2-488 and by NSF grant DMS-88-10192.

References

Adler, R.J.: The geometry of random fields. New York: Wiley 1981

Bickel, P.J., Rosenblatt, M.: On some global measures of the deviation of density function estimates. Ann. Stat. (6) **1**, 1071–1095 (1973)

Borell, C.: The Brunn–Minkowski inequality in Gauss space. Inventiones Mathematicae **20**, 205–216 (1975)

Brown, L.D., Low, M.G.: Asymptotic equivalence of nonparametric regression and white noise. Ann. Stat. **22** (March) (to appear)

Donoho, D.L.: Asymptotic Minimax Risk for Sup-Norm Loss: Solution via Optimal Recovery. Technical Report, Department of Statistics, Stanford University 1991

Donoho, D.L.: Statistical Estimation and Optimal Recovery. Ann. Stat. **22** (March) (to appear)

Donoho, D.L., Liu, R.C.: Geometrizing rates of convergence, III. Ann. Stat. **19**, 668–701 (1991)

Donoho, D.L., Low, M.G.: Renormalization exponents and optimal pointwise rates of convergence. Ann. Stat. **20**, 944–970 (1992)

Efroimovich, S.Y., Pinsker, M.S.: Estimation of square-integrable [spectral] density on the basis of a sequence of observations. Problems Inform. Transmission **17**, 50–68 (1981)

Efroimovich, S.Y., Pinsker, M.S.: Estimation of square-integrable probability density of a random variable. Problems Inform. Transmission **18**, 175–182 (1982)

Ibragimov, I.A., Has'minskii, R.Z.: On nonparametric estimation of the value of a linear functional in a Gaussian white noise. Teor. Verojatnost. I Primenen. **29**, 19–32 (1984)

Korostelev, A.P.: Exact asymptotic minimax estimate for a nonparametric regression in the uniform norm. Theory Probab. Appl. (to appear, 1992)

Le Cam, L.: Asymptotic methods in statistical decision theory. Berlin Heidelberg New York: Springer 1986

Leadbetter, M.R., Lindgren, G., Rootzen, H.: Extremes and related properties of random sequences and processes. Berlin Heidelberg New York: Springer 1983

Micchelli, C.: Optimal recovery of linear functionals. IBM Technical Report 1975.

Micchelli, C., Rivlin, T.J.: A survey of optimal recovery. In: Micchelli and Rivlin (eds.) Optimal estimation in approximation theory, pp. 1–54. New York: Plenum 1977

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

Pinsker, M.S.: Optimal filtering of square integrable signals in Gaussian white noise. Problems Inform. Transmission **16**, 52–68 (1980)

Strasser, H.: Mathematical Theory of Statistics. New York: de Gruyter 1985

Talagrand, M.: Small tails for the supremum of a Gaussian process. Annales de L'Institut Henri Poincaré **24**, 307–315 (1988)
Traub, J., Wasilkowski, G., Woźniakowski: Information-Based Complexity. Reading, MA, Addison–Wesley 1988
Watanabe, H., Qualls, C.: Asymptotic properties of Gaussian Random Fields. Trans. Am. Math. Soc. **177**, 155–171 (1973).