Minimax Risk: Pinsker Bound

Michael Nussbaum
Cornell University

From: Encyclopedia of Statistical Sciences, Update Volume (S. Kotz, Ed.), 1999. Wiley, New York.

Abstract

We give an account of the Pinsker bound describing the exact asymptotics of the minimax risk in a class of nonparametric smoothing problems. The parameter spaces are Sobolev classes or ellipsoids, and the loss is of squared L_2-type. The result from 1980 turned out to be a major step in the theory of nonparametric function estimation.

Keywords:
COMMUNICATION THEORY, STATISTICAL
DENSITY ESTIMATION
FISHER INFORMATION
KERNEL ESTIMATORS
LINEAR ESTIMATORS, BAYES
LOCAL ASYMPTOTIC NORMALITY
METHOD OF SIEVES
MINIMAX ESTIMATION
NOISE (SIGNAL PROCESSING IN THE PRESENCE OF)
PREDICTION AND FILTERING; LINEAR
SIEVES, METHOD OF
SPECTRAL ANALYSIS
SHRINKAGE ESTIMATORS
SMOOTHNESS PRIORS
SOBOLEV SPACES
SPLINE FUNCTIONS
STATIONARY PROCESSES
STEIN EFFECT

Consider observation of a Gaussian white noise model

$$dy(t) = f(t)dt + n^{-1/2}\sigma\, dW(t),\ t \in [0,1] \tag{1}$$

where $dW(t)$ is the derivative of standard Brownian motion and $n \to \infty$ and σ is fixed. Assume the function f is in a function class Σ, defined as follows. Let $\|\cdot\|_2$ be the L_2-norm on $[0,1]$, and let $f^{(m)}$ be the generalized derivative of f in the sense that $f^{(m-1)}$ is absolutely continuous. The periodic Sobolev class $\tilde{W}_2^m(M)$ is the class of functions on $[0,1]$ which satisfy $\left\|f^{(m)}\right\|_2^2 \le M$ and where periodic boundary conditions hold: $D^k f(0) = D^k f(1),\ k = 0,\ldots,m-1$. Let $\hat{f}$ be an estimator of f based on observations of the process $y(t)$, $t \in [0,1]$ and consider a loss $\left\|\hat{f} - f\right\|_2^2$. Consider the minimax risk over all estimators:

$$R_n(\Sigma) = \inf_{\hat{f}} \sup_{f \in \Sigma} E_{n,f} \left\|\hat{f} - f\right\|_2^2. \tag{2}$$

Pinsker's theorem [44] says that for $\Sigma = \tilde{W}_2^m(M)$

$$\lim_{n\to\infty} n^{2m/(2m+1)} R_n(\Sigma)\ = (\sigma/\pi)^{2m/(2m+1)}\ M^{1/(2m+1)}\ P_m \tag{3}$$

where

$$P_m = \left(\frac{m}{(m+1)}\right)^{2m/(2m+1)} (2m+1)^{1/(2m+1)} \tag{4}$$

is the *Pinsker constant* (in the narrow sense). The importance of that results is that it provides the *exact asymptotic behaviour* of the minimax risk, i. e. not only the "optimal rate of convergence for estimators" $n^{-2m/(2m+1)}$, but also the "optimal constant" , i.e. the right hand side of (3). The rate $n^{-2m/(2m+1)}$ for the convergence $R_n(\tilde{W}_2^m(M)) \to 0$ had been established before by Ibragimov and Khasminskii ([35], chap. VII). Pinsker's bound represents a breakthrough in nonparametric estimation theory, by allowing comparison of estimators on the level of constants rather than just comparing rates of convergence. In parametric theory, such constants are given in the form of "Fisher's bound for asymptotic variances" and its modern version (the Hajek-LeCam asymptotic minimax theorem). Consider e. g. the case where f is constant:

$$\Sigma = \left\{f :\ f = \vartheta 1,\ \vartheta^2 \le M\right\}.$$

Then estimating f with the above loss means just estimating ϑ with squared loss, and from general parametric estimation theory

$$\lim_{n\to\infty} n\ R_n(\Sigma)\ = \sigma^2 \tag{5}$$

Here the rate is n^{-1} and the constant is σ^2. Thus, Pinsker's bound (3) can be seen as an analog of Fisher's bound for an ill-posed (non-$\sqrt{n}$-consistent) function estimation problem.

The ellipsoid framework for sequence data

Actually Pinsker's result [44] was developed in a more general framework of a parameter space given as an ellipsoid. Consider countably many observations

$$y_j = \theta_j + \epsilon\, \xi_j,\ j = 1, 2, \ldots \tag{6}$$

where ξ_j are i.i.d. $N(0, \sigma^2)$, $\epsilon > 0$ is the noise size, and the sequence $\theta = (\theta_j)$ is in l_2. Consider a parameter space

$$\Theta = \left\{ \theta : \sum_{j=1}^{\infty} a_j \theta_j^2 \leq M \right\} \tag{7}$$

where $a = (a_j)$ is a sequence of nonnegative numbers, $a_j \to \infty$, i. e. Θ is an ellipsoid in l_2. Consider the problem of estimating the parameter θ with a loss given by the squared norm in l_2. A *linear filter* is a sequence $c = (c_j) \in l_2$ such that $0 \leq c_j \leq 1$ for all j. For such a c, a *linear filtering estimate* of θ is given by $\hat{\theta}^c = (c_j y_j)$. Pinsker's result is obtained by looking at the minimax estimator within this class: define

$$R_{L,\epsilon}(\Theta) = \inf_c \sup_{\theta \in \Theta} E_{\epsilon,\theta} \left\| \hat{\theta}^c - \theta \right\|_{l_2}^2 . \tag{8}$$

Along with this minimax risk over a restricted class of estimators, consider the risk over arbitrary estimators (analogous to 2)

$$R_{\epsilon}(\Theta) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} E_{\epsilon,\theta} \left\| \hat{\theta} - \theta \right\|_{l_2}^2 . \tag{9}$$

In this framework, Pinker's result takes the following remarkable form (notation $a \sim b$ means $a = b(1 + o(1))$): if $R_{L,\epsilon}(\Theta)/\epsilon^2 \to \infty$ then

$$R_{\epsilon}(\Theta) \sim R_{L,\epsilon}(\Theta),\ \epsilon \to 0 \tag{10}$$

In words, the minimax linear filtering estimate is asymptotically minimax among all estimators. The asymptotics of $R_{L,\epsilon}(\Theta)$ can often be found as regards rates and constants, and then gives rise to results like (3), (4). The minimax linear filter is easy to calculate in the above framework. For any $\hat{\theta}^c$ we have by the usual bias-variance decomposition

$$E_{n,\theta} \left\| \hat{\theta}^c - \theta \right\|_{l_2}^2 = \sum_{j=1}^{\infty} (1 - c_j)^2 \theta_j^2 + \epsilon^2 \sum_{j=1}^{\infty} c_j^2 = L_{\epsilon}(c, \theta), \tag{11}$$

say. Set $\theta^{(2)} = (\theta_j^2)$; then $L_{\epsilon}(c, \theta)$ is concave in $\theta^{(2)}$ and convex in c, and $\theta^{(2)}$ varies in a compact, convex subset of l_2 if $\theta \in \Theta$ while c varies in a closed convex subset of l_2. Hence

$$R_{L,\epsilon}(\Theta) = \inf_c \sup_{\theta \in \Theta} L_{\epsilon}(c, \theta) = \sup_{\theta \in \Theta} \inf_c L_{\epsilon}(c, \theta) \tag{12}$$

and there is a saddle point $(c_\epsilon^*, \theta_\epsilon^*)$. This saddle point is found as follows (see e. g. Belitser and Levit [1]. Observe that there exists a unique solution μ_ϵ of

$$\epsilon^2 \sum_{j=1}^{\infty} a_j^{1/2}(1-\mu_\epsilon a_j^{1/2})_+ = \mu_\epsilon M.$$

(where $x_+ = x \vee 0$). Then

$$L_\epsilon(c_\epsilon^*, \theta_\epsilon^*) = \epsilon^2 \sum_{j=1}^{\infty}(1-\mu_\epsilon a_j^{1/2})_+, \; c_\epsilon^* = ((1-\mu_\epsilon a_j^{1/2})_+) \tag{13}$$

and θ_ϵ^* has components $\epsilon^2(\mu_\epsilon^{-1} a_j^{-1/2} - 1)_+$ for $a_j > 0$, 0 otherwise. Thus the asymptotics of $R_{L,\epsilon}(\Theta) = L_\epsilon(c_\epsilon^*, \theta_\epsilon^*)$ is made explicit in its dependence upon a, M and ϵ. The principal case is that $a_j \sim (\pi j)^{2m}$, $j \to \infty$, where a calculation yields

$$R_{L,\epsilon}(\Theta) \sim \epsilon^{-4m/(2m+1)} \; (M/\pi^{2m})^{1/(2m+1)} \; P_m \tag{14}$$

with P_m from (4). This coincides with (3) for $\epsilon = n^{-1/2}\sigma$.

The model (1) with parameter space $\tilde{W}_2^m(M)$ as a special case. Consider the trigonometric orthonormal basis in $L_2(0,1)$: put $\varphi_1(t) \equiv 1$, $\varphi_{2k}(t) = 2^{-1/2}\cos(2\pi k t)$, $\varphi_{2k+1}(t) = 2^{-1/2}\sin(2\pi k t)$ for $k \geq 1$. The model (1) can be mapped canonically to a sequence model (6) via $y_j = \int \varphi_j(t) dy(t)$. Then the components of the signal are $\theta_j = \int \varphi_j(t) f_j(t) dt$, i. e. the Fourier coefficients of f, and it is known that

$$\tilde{W}_2^m(M) = \left\{ f: \; f = \sum_{j=1}^{\infty} \theta_j \varphi_j, \; \sum_{k=1}^{\infty} (2\pi k)^{2m} (\theta_{2k}^2 + \theta_{2k+1}^2) \leq M \right\}. \tag{15}$$

By Parseval's identity we also have $\|f\|_2^2 = \|\theta\|_{l_2}^2$, so that the loss functions coincide. Then $R_n(\tilde{W}_2^m(M))$ from (2) coincides with $R_\epsilon(\Theta)$ from (9) for $\epsilon = n^{-1/2}\sigma$ and for an ellipsoid $\Theta = \Theta(a, M)$ given by $a_j = (\pi j)^{2m}$ for even j, $a_j = (\pi(j-1))^{2m}$ for uneven j. Thus the asymptotics is $a_j \sim (\pi j)^{2m}$ for $j \to \infty$ as above.

Estimating a bounded normal mean

Consider the following variant of the model (6), (7): we observe

$$y_j = \theta_j + \; \xi_j, \; j = 1, \ldots, n \tag{16}$$

and the problem is to estimate the n-dimensional parameter $\theta = (\theta_j)$ with normed squared Euclidean loss $n^{-1} \|\cdot\|^2$. The parameter space is $\Theta = \{\theta : n^{-1} \|\theta\|^2 \leq M\}$. (Strictly speaking this is not a special case of (6), (7), but can be transformed via $\tilde{\theta} = n^{-1/2}\theta$ into an extended ellipsoid model where $\epsilon = n^{-1/2}$ and a is allowed to depend on n, and $a_j = 1$, $j \leq n$, $a_j = \infty$ for $j > n$). Let $R_n(\Theta)$ be the minimax risk over all estimators.

Upper asymptotic bound. Consider the linear filter $c_j = M/(M+1)$, $j = 1, \ldots, n$. Then from (11) we obtain

$$E_{n,\theta} n^{-1} \left\| \hat{\theta}^c - \theta \right\|^2 = n^{-1} \sum_{j=1}^{n} \left((1/(M+1)^2)\theta_j^2 + M^2/(M+1)^2 \right) \leq M/(M+1).$$

Lower asymptotic bound. Note that we need only consider estimators $\hat{\theta}$ with values in Θ, i. e. such that $n^{-1} \left\| \hat{\theta} \right\|^2 \leq M$. Let Q_n be a prior distribution Q_n on $\mathbb{R}^n$ (not necessarily concentrated on Θ) and let $r_n(Q_n)$ be the associated Bayes risk. Then

$$\begin{aligned} R_n(\Theta) &\geq r_n(Q_n) - \sup_{\hat{\theta}} \int_{\Theta^c} E_{n,\theta} n^{-1} \left\| \hat{\theta} - \theta \right\|^2 Q_n(d\theta) \\ &\geq r_n(Q_n) - 2M Q_n(\Theta^c) - 2 \left(\int n^{-2} \left\| \theta \right\|^4 Q_n(d\theta) \right)^{1/2} Q_n^{1/2}(\Theta^c). \end{aligned}$$

Take Q_n such that θ_j are i.i.d. $N(0, \delta^2 M)$ for some $\delta < 1$. Then

$$Q_n(\Theta^c) = \Pr(n^{-1} \sum_{j=1}^{n} \theta_j^2 > M) \to 0 \tag{17}$$

from the law of large numbers, and

$$\int n^{-2} \left\| \theta \right\|^4 Q_n(d\theta) = E_{Q_n} \left(n^{-1} \sum_{j=1}^{n} \theta_j^2 \right)^2 = O(1).$$

A standard reasoning for Gaussian priors yields

$$r(Q_n) = n^{-1} \sum_{j=1}^{n} (\delta^2 M/(1+\delta^2 M)) = \delta^2 M/(1+\delta^2 M).$$

Letting $\delta \nearrow 1$ we obtain

$$\liminf_n R_n(\Theta) \geq M/(M+1).$$

□

In this simple model the Pinsker bound $R_n(\Theta) \sim M/(M+1)$, $n \to \infty$ is the result of a dimension asymptotics effect when estimating a bounded normal mean in Euclidean space. A connection with Stein estimation in this setting is discussed by Beran [3].

Background: Bayes-minimax problems

In the model (16) $R_n(\Theta)$, $R_{L,n}(\Theta)$ and $r_n(Q)$ are the minimax risk, the minimax risk among linear filters and the Bayes risk, respectively. Let us consider the case $n = 1$; for this we omit the subscript n. Thus we look at the univariate problem of estimating

θ from data y having distribution $N(\theta, 1)$, with quadratic loss and parameter space $\Theta = \{\theta : \theta^2 \leq M\}$. A linear estimator $\hat{\theta}^c$ is given by$\hat{\theta}^c = cy$ where c is a real number. Its risk is

$$E_\theta(\hat{\theta}^c - \theta)^2 = (1-c)^2\theta^2 + c^2 = L(c, \theta),$$

say. For given θ, the best linear estimator is given by $c(\theta^2) = \theta^2/(\theta^2+1)$, $c(\theta^2)$ is a linear filter, and the risk is $\theta^2/(\theta^2+1)$. In view of the minimax theorem (12), $\hat{\theta}^{c(M)}$ is minimax among linear estimators and

$$R_L(\Theta) = M/(M+1).$$

Note that $\hat{\theta}^{c(\theta^2)}$ has another interpretation as a Bayes estimator: for a prior distribution Q on θ having $E_Q\theta^2 = \sigma^2$, not necessarily concentrated on Θ, the mixed risk is again

$$E_Q E_\theta(\hat{\theta}^c - \theta)^2 = (1-c)^2 E_Q\theta^2 + c^2 = L(c, \sigma).$$

Hence $\hat{\theta}^{c(\sigma^2)}$ is the Bayesian-among-linear estimator for Q, with risk $\sigma^2/(\sigma^2+1)$. This estimator is actually Bayesian if $Q = N_{\sigma^2} = N(0, \sigma^2)$. Hence

$$r(N_M) = M/(M+1) = R_L(\Theta).$$

Moreover, Donoho and Johnstone [4] establish the following:

$$\sup_{E_Q\theta^2 \leq M} r(Q) = r(N_M). \tag{18}$$

Thus $R_L(\Theta)$ is also the solution of a Bayes-minimax problem: it is a least favorable Bayes risk over Q: $E_Q\theta^2 \leq M$. Consider now the model (16) for general n; it is obvious by reasons of symmetry that again

$$R_{L,n}(\Theta) = R_L(\Theta) = r(N_M).$$

and that for any Q we have $r_n(Q^{\otimes n}) = r(Q)$. Since $R_{L,n}(\Theta) \geq R_n(\Theta)$, the minimax risk $R_n(\Theta)$ is bracketed

$$\sup_{suppQ \subset \Theta} r_n(Q) \leq R_n(\Theta) \leq \sup_{E_Q\theta^2 \leq M} r_n(Q^{\otimes n}).$$

This gives the basic heuristics for the validity of the Pinsker bound. Distributions $Q^{\otimes n}$ with $E_Q\theta^2 < M$ do not have support in Θ in general, but as $n \to \infty$ they tend to be concentrated on Θ (cp. the law of large numbers (17)), so that asymptotically the upper and lower brackets coincide.

The special role of Gaussian priors in the symmetric setting (16) is determined by (18); in the general "oblique" ellipsoid case (6) product priors with non-identical components are appropriate. The proof in [44] employs also non-Gaussian components, in dependence on the size of a_j. Bayes-minimax problems in relation to the Pinsker bound are discussed by Heckman and Woodroofe [34], Donoho, MacGibbon and Liu [6], and Donoho and Johnstone [4].

Renormalization and continuous minimax problem. Let us sketch a derivation of the asymptotics (14) by a renormalization technique. Suppose that $a_j = (\pi j)^{2m}$ and consider linear filters $c_j = c(hj)$, where c: $[0,\infty) \mapsto [0,1]$ is a "filter function" (assumed Riemann integrable) and h is a bandwidth parameter, tending to 0 for $\epsilon \to 0$. Consider also a Gaussian prior measure $N(0,\sigma_j^2)$ for θ_j where $\sigma_j = M\pi^{-2m}h^{2m+1}\sigma(jh)$ for a (continuous) function $\sigma: \ [0,\infty) \mapsto \mathbb{R}$, fulfilling

$$\int_0^\infty x^{2m}\sigma^2(x)dx \leq 1. \tag{19}$$

Then the restriction $\sigma \in \Theta$ is asymptotically satisfied since

$$\begin{aligned} M &\geq \sum_{j=1}^\infty a_j\sigma_j^2 = Mh\sum_{j=1}^\infty (jh)^{2m}\sigma^2(jh) \\ &\to M\int_0^\infty x^{2m}\sigma^2(x)dx, \ h \to 0. \end{aligned}$$

A choice $h = (\epsilon^2\pi^{2m}/M)^{1/(2m+1)}$ and a similar reasoning for the functional $L_\epsilon(c,\sigma)$ gives

$$L_\epsilon(c,\sigma) \sim \epsilon^{4m/(2m+1)}(M/\pi^{2m})^{1/(2m+1)}L_0(c,\sigma) \tag{20}$$

where

$$L_0(c,\sigma) = \int_0^\infty (1-c(x))^2\sigma^2(x)dx + \int_0^\infty c^2(x)dx.$$

The saddle point problem (12) for each ϵ is thus asymptotically expressed in terms of a fixed continuous problem. The solution is as follows, cf. Golubev [17]. There is a unique solution λ^* of the equation

$$\int_0^\infty ((\lambda x)^m - (\lambda x)^{2m})_+ dx = 1. \tag{21}$$

Then the saddle point (c^*, σ^{2*}) is given by

$$c^*(x) = (1-(\lambda^* x)^m)_+, \ \sigma^{2*}(x) = ((\lambda^* x)^{-m} - 1)_+ \tag{22}$$

and the Pinsker constant P_m from (4) is the value of the game:

$$P_m = L_0(c^*,\sigma^{2*}) = \inf_c \sup_{\int x^{2m}\sigma^2(x)dx \leq 1} L_0(c,\sigma).$$

The function $c^*(x)$ in (22) has sometimes been called the *Pinsker filter* (cp. c_ϵ^* in (13)). The continuous saddle point problem arises naturally in a continuous Gaussian white noise setting (1) and a parameter space described in terms of the continuous Fourier transform (cf. Golubev [17]), e. g. a Sobolev class of functions on the whole real line. Using the basic structure of the above renormalization argument, it is easy to mimick the rigorous proof in the model (16) above for a proof of (3). The Gaussian prior distribution then should be taken for the saddle point function σ^* with respect to a restriction $\int x^{2m}\sigma^2(x)dx \leq \delta < 1$; cf. [41] for details.

Statistical applications and further developments

The result of Pinsker [44] for the signal in white noise model (1) or (6) gave rise to a multitude of results in related nonparametric curve estimation problems having a similar structure.

Nonparametric density estimation. Efromovich and Pinsker [8] treated the case of observed i.i.d. random variables y_j, $j = 1, \ldots, n$ with values in $[0, 1]$ having a density f. This density is assumed to be in a set Σ which has an ellipsoid representation in terms of the Fourier basis (cp. (15)):

$$\Sigma = \left\{ f : \; f = \sum_{j=1}^{\infty} \theta_j \varphi_j, \; \theta_1 = 1, \; \sum_{j=2}^{\infty} a_j \theta_j^2 \leq M \right\};$$

then $\Sigma^+ = \Sigma \cap \{f : \; f(x) \geq 0, x \in [0,1]\}$ is a set of densities. Let $R_n(\Sigma^+)$ be the minimax risk for the density problem defined analogously to (2) and let $R_{L,\epsilon}(\Theta)$ be the minimax linear filtering risk (8) in a discrete white noise model with $\epsilon = n^{-1/2}$ and ellipsoid given by coefficients a_j where $a_1 = 0$; then under a condition $a_j / \log j \to \infty$ for $j \to \infty$ we have, similarly to (10)

$$R_n(\Sigma^+) \sim R_{L,n^{-1/2}}(\Theta), \; n \to \infty$$

and in the main case of a periodic Sobolev class one obtains again (14). Indeed for $a_{2k} = a_{2k+1} = (2\pi k)^{2m}$, $k \geq 1$ the class Σ^+ above coincides with the set of densities in $\tilde{W}_2^m(M)$. The proof relies essentially on a kind of uniform LAN property, individually for each Fourier coefficient $\theta_k = \int f \varphi_k$ considered as a functional of f. Similar results were obtained for spectral density estimation for an observed Gaussian stationary sequence, cf. Efromovich and Pinsker [7], Golubev [24], [25].

Nonparametric regression. Consider observations

$$y_i = f(t_i) + \zeta_i, \; i = 1, \ldots, n \tag{23}$$

where ζ_i are i. i. d. $N(0,1)$, $t_i = i/n$ and f is a smooth function on $[0, 1]$. The Sobolev class $W_2^m(M)$ is the class of functions on $[0, 1]$ which satisfy $\left\| f^{(m)} \right\|_2^2 \leq M$ (without periodic boundary conditions). Consider a semi scalar product $(f, g)_n = \sum_{i=1}^n f(t_i) g(t_i)$ and the associated seminorm $\|f\|_{2,n} = (f, f)_n^{1/2}$, and define a minimax risk $R_n(\Sigma)$ as in (2) but for a "design loss" $\left\| \hat{f} - f \right\|_{2,n}^2$. Then in the case $\Sigma = W_2^m(M)$ the asymptotics (3) obtains, cf. [41]. The key for this result is the representation of the model in the ellipsoid form (6), (7). This can be achieved using the Demmler-Reinsch spline basis, which is an orthonormal set of functions $\varphi_{j,n}$, $j = 1, \ldots, n$ with respect to $(\cdot, \cdot)_n$ and which simultaneously diagonalizes the quadratic form $(f^{(m)}, g^{(m)})$ (the expression $(\cdot, \cdot)$ denotes scalar product in $L_2(0, 1)$). The numbers $a_{jn} = (\varphi_{j,n}^{(m)}, \varphi_{j,n}^{(m)})$ represent the coefficients a_j in (7). Then the analytic result is required that $a_{jn} \sim (\pi j)^{2m}$ with appropriate uniformity in n, so that again (14) can be inferred. The optimal estimator

of f then is of the linear filtering type in terms of the Demmler-Reinsch spline basis and the Pinsker filter c^* from (22).
Speckman [46] independently found this estimator as minimax linear and gave its risk asymptotics; he used the following setting. Call an estimator $\hat{f}$ of f in (23) linear if it is linear in the n-dimensional data vector y; then $\hat{f} = Ay$ where A is a nonrandom linear operator. The estimator $\hat{f}$ is minimax linear if it minimizes $\sup_{f\in\Sigma} E_{n,f} \left\| \hat{f} - f \right\|_{2,n}^2$ among all linear estimators. In (8) only linear filtering estimates are admitted; it turns out that in the ellipsoid case the minima coincide (cf. Pilz [43]). Thus another paraphrase of (10) is that the minimax linear estimator is asymptotically minimax among all estimators.
The spectral asymptotics of differential quadratic forms like $(f^{(m)}, f^{(m)})$ turns out to be crucial, since it governs the behaviour of the ellipsoid coefficients a_j. If spectral values are calculated with respect to (f, f) rather than to $(f, f)_n$ (which corresponds to observations (1) with parameter space $W_2^m(M)$) then the appropriate basis consists of eigenfunctions of a differential operator, cf. [28], sec. 5.1. The spectral asymptotics is known to be $a_j \sim (\pi j)^{2m}$. The spectral theory for differential operators allows to obtain the Pinsker bound for quite general Sobolev smoothness classes on domains of $\mathbb{R}^k$; for the periodic case on a hypercube domain cf. [40].

Asymptotically gaussian models. The proof for the cases of density and spectral density estimation ([7], [8]) is based on the asymptotic Gaussianity of those models, in the problem of estimating one individual Fourier coefficient. Inspired by this, Golubev [21] formulated a general LAN type condition for a function estimation problem for the validity of the lower bound part of the Pinsker bound. The regression case (23) with nongaussian noise ζ_i in (23) was treated in [28]; for random design regression cf. Efromovich [13].

Analytic functions. The case of m-smooth functions where $a_j \sim (\pi j)^{2m}$ was treated as a standard example here, but another important case in the ellipsoid asymptotics is $a_j \sim \exp(\beta j)$. Then (14) is replaced by

$$R_{L,n}(\Theta) \sim (\epsilon^2 \log \epsilon^{-1})\ \beta^{-1}.$$

The exponential increase of a_j corresponds to the case of analytic functions; cf. Golubev, Levit, Tsybakov [27]. Ibragimov and Khasminskii [36] obtained an exact risk asymptotics in a case where the functions are even smoother (entire functions of exponential type on the real line) and the rate is ϵ^2, even though the problem is still nonparametric.

Adaptive Estimation. The minimax linear filtering estimate attaining the bound (10) depends on the ellipsoid via the set of coefficients a and M. A siginificant result of Efromovich and Pinsker [9] is that this attainment is possible even when a and M are not known, provided a varies in some large class of coefficients. The *Efromovich-Pinsker algorithm* of adaptive estimation (cf. also Efromovich [10]) thus allows to attain the bound (3) for periodic Sobolev classes by an estimator which does not depend on the degree of smoothness m and on the bound M. This represented a considerable advance in adaptive smoothing theory, improving respective rate of convergence results; for further developments and related theory cf. results in [19], [20], [29], [30], [24] and the discussion in [28].

Other constants. Korostelev [39] obtained an analog of (3) when the squared L_2-loss $\|\cdot\|_2^2$ is substituted by the sup-norm loss and the Sobolev function class $\tilde{W}_2^m(M)$ is replaced by a Hölder class of smoothness m (a class where f satisfies a Hölder condition with exponent $m \in (0,1]$ uniformly on $[0,1]$). The rate in n then changes to include a logarithmic term and naturally the constant in (3) is another one; this *Korostelev constant* represents a further breakthrough and stimulated the search for constants in nonparametric function estimation. Tsybakov [47] was able to extend the realm of the Pinsker theory to loss functions $l(\|\cdot\|_2)$ where l is monotone and possibly bounded. An analog of the Pinsker bound for nonparametric hypothesis testing was established by Ermakov [14]; cf. also Ingster [37].

Besov bodies and wavelet estimation. Above it was seen that the case of data (16) and parameter space $\Theta = \{\theta : \sum_{n=1}^{n} \theta_j^2 \leq M\}$ is in some sense the simplest model where the Pinsker phenomenon (10) occurs. Donoho, MacGibbon and Liu [6] set out to investigate more general parameter spaces like $\Theta = \{\theta : \sum_{j=1}^{n} \theta_j^p \leq M\}$ (p-bodies); further results were obtained by Donoho and Johnstone [4]. It was found that (10) occurs only for $p = 2$; linear estimators were found to be asymptotically nonoptimal for $p < 2$, and threshold rules were described as nonlinear alternatives. The limitation of the Pinsker phenomenon to a Hilbertian setting thus became apparent; however this stimulated the development of nonlinear wavelet smoothing for function classes representable as Besov bodies (cf. Donoho and Johnstone [5]).

Remark. Several developments and facets of the theory have not been discussed here; these include applications in deterministic settings ([31], [32], [33]), inverse problems ([15], [16]), design of experiments ([28], [22]), discontinuities at unknown points ([42]).

References

[1] Belitser, E., Levit, B., (1995). On minimax filtering on ellipsoids. *Math. Meth. Statist.* **4** 259-273

[2] Belitser, E., Levit, B., Asymptotically minimax nonparametric regression in L_2. Manuscript

[3] Beran, R. (1996). Stein estimation in high dimensions: a retrospective. In: *Madan Puri Festschrift* (E. Brunner and M. Denker, eds.) 91-110 . VSP Editors, Zeist (Netherlands).

[4] Donoho, D. L. and Johnstone, I, M. (1994). Minimax risk over l_p-balls for l_q-error. *Prob. Theor. Rel. Fields* 99 277-303

[5] Donoho, D. L. and Johnstone, I, M. (1996). Neo-classical minimax problems, thresholding and adaptive function estimation. *Bernoulli* 2 (1) 39-62

[6] Donoho, D. L., Liu, R. and MacGibbon, B. (1990). Minimax risk for hyperrectangles. *Ann. Statist.* **18** 1416-1437.

[7] Efromovich, S. Yu. and Pinsker, M. S. (1981). Estimation of a a square integrable spectral density based on a sequence of observations. *Problems Inform. Transmission* **17,** 182-196

[8] Efromovich, S. Yu. and Pinsker, M. S. (1982). Estimating a square integrable probability density of a random variable (in Russian). *Problems Inform. Transmission* **18,** No. 3, 19-38

[9] Efromovich, S. and Pinsker, M. S. (1984). An adaptive algorithm of nonparametric filtering. *Automat. Remote Control* **11** 1434-1440

[10] Efromovich, S. (1985). Nonparametric estimation of a density with unkown smoothness. *Theor. Probab. Appl.* **30** 557-568

[11] Efromovich, S. (1989). On sequential nonparametric estimation of a density. *Theor. Probab. Appl.* **34** 228-239

[12] Efromovich, S. (1995). Sequential nonparametric estimation with assigned risk. *Ann. Statist.* **23** 1376-1392

[13] Efromovich, S. (1996). On nonparametric regression for iid observations in a general setting. *Ann. Statist.* **24** 1126-1144

[14] Ermakov, M. S. (1990). Minimax detection of a signal in Gaussian white noise. *Theory Probab. Appl.* **35** 667-679

[15] Ermakov, M. (1990). On optimal solutions of the deconvolution problem. *Inverse Problems* **6** 863-872

[16] Ermakov, M. (1992). Minimax estimation in a deconvolution problem. *J. Phys. A: Math. Gen.* 1273-1282

[17] Golubev, G. K. (1982). On minimax filtering of functions in L_2 (in Russian). *Problems Inform. Transmission* **18,** No. 4, 67-75

[18] Golubev, G. K. (1984). On minimax estimation of regression (in Russian). *Problems Inform. Transmission* **20,** No. 1, 56-64

[19] Golubev, G. K. (1987). Adaptive asymptotically minimax estimates of smooth signals. (in Russian). *Problems Inform. Transmission* **23,** No. 1, 57-67

[20] Golubev, G. K. (1990). Quasilinear estimates for signals in L_2. *Problems Inform. Transmission* **26**, 1, 15-20

[21] Golubev, G. K. (1991). Local asymptotic normality in problems of nonparametric estimation of functions, and lower bounds for quadratic risk. *Theory Probab. Appl.* **36,** 1, 152-157

[22] Golubev, G. K. (1992). Sequential design of an experiment in the nonparametric estimation of smooth regression functions. *Problems Inform. Transmission* **28**, 3, 265-268

[23] Golubev, G. K. (1992). Asymptotically minimax estimation of a regression function in an additive model. *Problems Inform. Transmission* **28**, 2, 101-112

[24] Golubev, G. K. (1992). Quasilinear filtering of Gaussian stationary sequences. *Theory Probab. Appl.* 37, 4, 793-800

[25] Golubev, G. K. (1993). Nonparametric estimation of smooth spectral densities of Gaussian stationary sequences. *Theory Probab. Appl.* **38** 4 630-639

[26] Golubev, G. K., Levit, B. (1995). On the second order minimax estimation of distribution functions. *Math. Methods of Statistics,* to appear

[27] Golubev, G. K., Levit, B., Tsybakov, A. B. (1996). Asymptotically efficient estimation of analytic functions in Gaussian white noise. *Bernoulli* **2** 167-181

[28] Golubev, G. K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression. *Ann. Statist.* **18** 758-778

[29] Golubev, G. K. and Nussbaum, M. (1990). Nonparametric estimation of a regression function in L_2. *Problems Inform. Transmission* **26,** 3, 213-255

[30] Golubev, G. K. and Nussbaum, M. (1992). Adaptive spline estimates in a nonparametric regression model. *Theory Probab. Appl.* **37**, 3, 521-529

[31] Golubev, G. K. and Pinsker, M. S. (1983). Minimax extrapolation of sequences. *Problems Inform. Transmission* **19**, 4, 275-283

[32] Golubev, G. K. and Pinsker, M. S. (1984). Minimax extrapolation of functions. *Problems Inform. Transmission* **20**, 2, 27-43

[33] Golubev, G. K. and Pinsker, M. S. (1985). Extremal properties of minimax estimation of sequences. *Problems Inform. Transmission* **21**, 3, 36-52

[34] Heckman, N. E. and Woodroofe, M. (1991). Minimax Bayes estimation in nonparametric regression. *Ann. Statist.* **19** 2003-2014

[35] Ibragimov, I. A. and Khasminskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

[36] Ibragimov, I. A. and Khasminskii, R. Z. (1982). On density estimation within a class of entire functions. *Theory Probab. Appl.* **27** 514-524

[37] Ingster, Yu. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. *Math. Meth. Statist.,* Part I: **2** 2 85-114, Part II: **2** 3 171-189, Part III: **2** 4 249-268

[38] Kazbaras, A. (1986). An adaptive kernel type estimator for a square integrable distribution density. *Lithuanian Math. Journal* **26** 318-324

[39] Korostelev, A. P. (1993). Exact asymptotically minimax estimator for nonparametric regression in uniform norm. *Theory Probab. Appl.* **38**, no. 4, 775–782

[40] Nussbaum, M. (1983). Optimal filtering of a function of many variables in Gaussian white noise. *Probl. Information Transmission* **19** (2) 23-29 (1983)

[41] Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984-997

[42] Oudshoorn, K. (1996). Minimax estimation of a regression function with jumps: attaining the optimal constant. To appear, *Bernoulli.*

[43] Pilz, J. (1986). Minimax linear regression estimation with symmetric parameter restrictions. *J. Statist. Plann. Inference* **13** 297-318

[44] Pinsker, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* (1980) 120-133

[45] Rudzkis, R. (1985). On an estimate of the spectral density. *Lithuanian Math. Journal* **25** 273-280

[46] Speckman, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression. *Ann. Statist.* 13 970-983

[47] Tsybakov, A. B. (1994). Asymptotically efficient signal estimation in L_2 under general loss functions. Submitted to *Problems of Information Transmission.*