The Annals of Statistics
1998, Vol. 26, No. 1, 273–278

SIMULTANEOUS SHARP ESTIMATION OF FUNCTIONS AND THEIR DERIVATIVES[1]

BY SAM EFROMOVICH

University of New Mexico

A data-driven estimate is given that, over a Sobolev space, is simultaneously asymptotically sharp minimax for estimating both the function and its derivatives under integrated squared error loss. It is also shown that linear estimates cannot be simultaneously asymptotically sharp minimax over a given Sobolev space.

1. Introduction. There are a number of reasons for wishing to estimate both a function and its derivatives; for instance, first and second derivatives may be of intrinsic interest as measures of slope and curvature. So far the attention has focused on estimation where typically derivatives of a rate optimal estimate are rate optimal estimates of the corresponding derivatives; see, for instance, Stone (1982) as well as an interesting discussion in Rice and Rosenblatt (1983).

In this paper we consider the filtering statistical model where on the interval $[0, 1]$ we observe random processes $Y_n(t)$ having stochastic differential

$$dY_n(t) = f(t)\,dt + n^{-1/2}\,dW(t), \qquad 0 \le t \le 1 \tag{1.1}$$

and $W(t)$ is a standard Brownian motion. We assume that f is one-periodic, including its α derivatives, and it belongs to the Sobolev class $W_2^\alpha Q = \{f\colon \int_0^1 [f^{(\alpha)}(t)]^2\,dt \le Q,\ \alpha \ge 2,\ 0 < Q < \infty\}$ where $f^{(k)}$ denotes the kth derivative and $f^{(0)} = f$. Recall that the filtering model is equivalent to a density estimation and a nonparametric regression where n is the sample size. Then, Tsybakov (1997) has shown that the linear estimate

$$\begin{aligned}\tilde{f}_n&(t, k, \alpha, Q)\\ &= \hat{\theta}_0 \varphi_0^{(k)}(t) + \sum_{1 \le j \le J(n,k,\alpha,Q)} \Big[1 - (j/J(n,k,\alpha,Q))^{(\alpha-k)}\Big]\\ &\qquad\qquad \times \Big(\hat{\theta}_{2j-1}\varphi_{2j-1}^{(k)}(t) + \hat{\theta}_{2j}\varphi_{2j}^{(k)}(t)\Big)\end{aligned} \tag{1.2}$$

Received July 1996; revised June 1997.
[1]Research partially supported by NSF Grant DMS-96-25412.
AMS 1991 *subject classifications*. Primary 62C05; secondary 62E20, 62J02, 62G05, 62M99.
Key words and phrases. Adaptation, Sobolev functions, minimax, linear estimates, filtering.

is the asymptotically sharp minimax estimate of the kth derivative $f^{(k)}$; that is, for any $k = 0, 1, \dots, \alpha - 1$ the following relations hold:

$$\begin{aligned} &\sup_{f \in W_2^\alpha Q} E_f \left\{ \int_0^1 \left(\tilde{f}_n(t, k, \alpha, Q) - f^{(k)}(t) \right)^2 dt \right\} \\ (1.3) \qquad &= \inf_{\bar{f}_n} \sup_{f \in W_2^\alpha Q} E_f \left\{ \int_0^1 \left(\bar{f}_n(t, k, \alpha, Q) - f^{(k)}(t) \right)^2 dt \right\} (1 + o(1)) \\ &= P(\alpha, Q, k) n^{-2(\alpha - k)/(2\alpha + 1)} (1 + o(1)). \end{aligned}$$

Here $\{\varphi_0(t) = 1,\ \varphi_{2j-1}(t) = \sqrt{2}\sin(2\pi jt)$ and $\varphi_{2j}(t) = \sqrt{2}\cos(2\pi jt),\ j = 1, 2, \dots\}$ is the classical trigonometric basis in $L_2[0,1]$; $\hat{\theta}_j = \int_0^1 \varphi_j(t)\, dY_n(t) = \theta_j + n^{-1/2}\xi_j$ is the sample Fourier coefficient that estimates the Fourier coefficient $\theta_j = \int_0^1 f(t)\varphi_j(t)\, dt$, and ξ_j are iid standard normal; $J(n, k, \alpha, Q) = n^{1/(2\alpha+1)}[Q(2\pi)^{-2\alpha}(\alpha + k + 1)(2\alpha + 1)/2(\alpha - k)]^{1/(2\alpha+1)}$, and $P(\alpha, Q, k) = (2k+1)^{-1}[(\alpha - k)/\pi(\alpha + k + 1)]^{2(\alpha-k)/(2\alpha+1)}(Q(2\alpha + 1))^{(2k+1)/(2\alpha+1)}$.

The result shows that derivatives of the sharp minimax estimate $\tilde{f}_n(t, 0, \alpha, Q)$ are not sharp minimax estimates of the corresponding derivatives of f.

We show in the next section that the Efromovich–Pinsker data-driven estimator [Efromovich and Pinsker (1984)], which adapts to underlying f but not to $W_2^\alpha Q$, is simultaneously sharp optimal. In Section 3 we show that a linear estimate cannot be simultaneously sharp minimax and therefore linear estimates are outperformed by nonlinear ones. This is an interesting addition to the known list of such examples because, for the Sobolev classes and mean integrated squared error, a sharp minimaxity is the trademark of linear estimates.

2. Simultaneously sharp estimate.

Set

$$\hat{f}_n(t) = \hat{\theta}_0 + \sum_{1 \le s \le n^{1/10}\ln(n)} \hat{\Lambda}_s \sum_{j=(s-1)s+1}^{s(s+1)} \hat{\theta}_j \varphi_j(t)$$

where

$$\hat{\Lambda}_s = \hat{\Theta}_s \left[\hat{\Theta}_s + n^{-1} \right]^{-1} I \frac{\hat{\Theta}_s > 1}{n \ln(s+3)}, \qquad \hat{\Theta}_s = (2s)^{-1} \sum_{j=(s-1)s+1}^{s(s+1)} \left(\hat{\theta}_j^2 - n^{-1} \right)$$

and $I(\cdot)$ is the indicator function.

THEOREM 2.1. *The data-driven estimate $\hat{f}_n$ is simultaneously asymptotically sharp minimax for estimating $f \in W_2^\alpha Q$ and its derivatives, that is, for any $k = 0, 1, \dots, \alpha - 1$,*

$$\begin{aligned} (2.1) \qquad &\sup_{f \in W_2^\alpha Q} E_f \left\{ \int_0^1 \left(\hat{f}_n^{(k)}(t) - f^{(k)}(t) \right)^2 dt \right\} \\ &= P(\alpha, Q, k) n^{-2(\alpha-k)/(2\alpha+1)} (1 + o(1)). \end{aligned}$$

The underlying idea of our estimate is as follows. We notice that for all k the minimax estimates (1.2) are linear in $\hat{\theta}_j$ with smoothing weights depending on k. However, if we consider the best smoothing pseudoweights, that may depend on f, then it is not difficult to show [see the proof below or line (3.1)] that for all k they are the same and equal to $\theta_j^2/(\theta_j^2 + n^{-1})$. Thus, a simultaneously sharp pseudoestimate exists. Of course, we do not know f and we cannot precisely estimate the optimal pseudoweights; therefore we employ instead a procedure of Efromovich and Pinsker (1984) that mimics that optimal pseudoweights via a special grouping Fourier coefficients.

PROOF OF THEOREM 2.1. First, direct calculation shows that the estimate

$$\begin{aligned}\check{f}_n(t, k, \alpha, Q))\\ &= \hat{\theta}_0 \varphi_0^{(k)}(t) + \sum_{1 \le s \le [2J(n,k,\alpha,Q)]^{1/2}} \left[1 - ((s-1)s/2J(n,k,\alpha,Q))^{\alpha - k}\right]\\ &\qquad \times \sum_{j=(s-1)s+1}^{s(s+1)} \hat{\theta}_j \varphi_j^{(k)}(t)\end{aligned}$$

is also a sharp minimax estimate of $f^{(k)}$. Here we simply use the same weights (dependent on k) for subsets of Fourier coefficients. Then the elementary facts that $2J(n, k, \alpha, Q) < n^{1/5}\ln^2(n)$ for sufficiently large n and that $E_f\{(\lambda\tilde{\theta} - \theta)^2\}$ takes on its minimum value when $\lambda = E_f\{\tilde{\theta}\theta\}/E\{\tilde{\theta}^2\}$ imply that the pseudoestimate

$$\hat{f}_n(t, k, \{\Lambda_{ks}\}) = \hat{\theta}_0 \varphi_0^{(k)}(t) + \sum_{1 \le s \le n^{1/10}\ln(n)} \Lambda_{ks} \sum_{j=(s-1)s+1}^{s(s+1)} \hat{\theta}_j \varphi_j^{(k)}(t) \tag{2.2}$$

is also sharp minimax whenever $\Lambda_{ks} = (\sum \theta_j^2 \sigma_{k,j}^2)/(\sum(\theta_j^2 + n^{-1})\sigma_{k,j}^2)$ where hereafter the summation is taken over $j \in \{(s-1)s + 1, \dots, s(s+1)\}$ and $\sigma_{k,j}^2 = \int_0^1 [\varphi_j^{(k)}(t)]^2\, dt$.

Now we would like to modify the pseudoestimate (2.2) in such a way that it remains sharp minimax and its weights do not depend on k. Let us consider weights $\Lambda_s = \Theta_s/(\Theta_s + n^{-1})$, $\Theta_s = (2s)^{-1}\sum_{j=(s-1)s+1}^{s(s+1)} \theta_j^2$ in the hope that Λ_s are close to Λ_{ks} for sufficiently large s. Note that $\hat{f}_n^{(k)}(t, 0, \{\Lambda_s\}) = \hat{f}_n(t, k, \{\Lambda_s\})$ and therefore, to finish the proof, we need to establish two facts: (i) the pseudoestimate $\hat{f}_n(t, k, \{\Lambda_s\})$ is sharp minimax; (ii) this pseudoestimate is well approximated (in the minimax sense) by $\hat{f}_n^{(k)}(t)$. The latter fact is proved following along lines of the proof in Efromovich and Pinsker (1984) where the case $k = 0$ is considered. We leave the details to the interested reader. The fact in (i) is verified below.

It suffices to show that

$$\begin{aligned}&\sup_{f \in W_2^\alpha Q} E_f\left\{\int_0^1 \left(\hat{f}_n(t, k, \{\Lambda_s\}) - f^{(k)}(t)\right)^2 dt\right\}\\ &\qquad \le (1 + o(1)) \sup_{f \in W_2^\alpha Q} E_f\left\{\int_0^1 \left(\hat{f}_n(t, k, \{\Lambda_{ks}\}) - f^{(k)}(t)\right)^2 dt\right\}.\end{aligned} \tag{2.3}$$

Direct calculations, based on the Parseval identity, show that

$$
\begin{aligned}
&E_f\left\{\int_0^1 \left(\hat{f}_n(t,k,\{\Lambda_s\}) - f^{(k)}(t)\right)^2 dt\right\} \\
&\quad\le n^{-1} + \sum_{1\le s\le n^{1/10}\ln(n)} n^{-1}\left[\Theta_s^2\left(\sum \sigma_{k,j}^2\right) + n^{-1}\left(\sum \theta_j^2\sigma_{k,j}^2\right)\right]\left(\Theta_s + n^{-1}\right)^{-2} \\
&\qquad + \sum_{s>n^{1/10}\ln(n)} (2s)\Theta_s \\
&\quad\le n^{-1} + \sum_{1\le s\le n^{1/10}\ln(n)} \sigma_{k,\,s(s+1)}^2\; n^{-1}(2s)\Theta_s/\left(\Theta_s + n^{-1}\right) \\
&\qquad + \sum_{s>n^{1/10}\ln(n)} (2s)\Theta_s.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&E_f\left\{\int_0^1 \left(\hat{f}_n(t,k,\{\Lambda_{ks}\}) - f^{(k)}(t)\right)^2 dt\right\} \\
&\quad\ge \sum_{1\le s\le n^{1/10}\ln(n)} n^{-1}\left(\sum \sigma_{k,j}^2\right)\left(\sum \theta_j^2\sigma_{k,j}^2\right)/\left(\sum\left(\theta_j^2 + n^{-1}\right)\sigma_{k,j}^2\right) \\
&\qquad + \sum_{s>n^{1/10}\ln(n)} (2s)\Theta_s \\
&\quad\ge \sum_{1\le s\le n^{1/10}\ln(n)} \left[\sigma_{k,(s-1)s+1}^4/\sigma_{k,\,s(s+1)}^2\right] n^{-1}(2s)\Theta_s/\left(\Theta_s + n^{-1}\right) \\
&\qquad + \sum_{s>n^{1/10}\ln(n)} (2s)\Theta_s.
\end{aligned}
$$

These two relations yield (2.3). $\square$

3. Linear estimates are not simultaneously sharp. The previous section shows that the simultaneously sharp minimax estimation of a function and its derivatives is possible. On the other hand, the suggested estimate is nonlinear while the sharp minimax estimates (1.2) are linear [of course, the parameters (α, Q) are assumed to be given for the linear estimates]. We now show that linear estimates cannot be simultaneously asymptotically sharp minimax. Such an issue is of a special interest in the nonparametrics literature; see, for instance, the discussion in Nemirovskii (1985) and Donoho, Johnstone, Kerkyacharian and Picard (1996).

THEOREM 3.1. *Assume that $f \in W_2^\alpha Q$ with a given (α, Q), and let k and r be integer and $0 \le k < r < \alpha$. Then there is no linear estimate that is simultaneously asymptotically sharp minimax for estimating $f^{(k)}$ and $f^{(r)}$.*

PROOF. First, consider a subclass of smoothing linear estimates $\hat{f}(t,\{\lambda_j\}) = \sum_{j=0}^{\infty} \lambda_j \hat{\theta}_j \varphi_j(t)$. Direct calculation shows that for $m \in \{k, r\}$,

$$E_f\left\{\int_0^1 \left(\hat{f}^{(m)}\left(t,\{\lambda_j\}\right) - f^{(m)}(t)\right)^2 dt\right\}$$

$$= \sum_{j=0}^{\infty} \left[\sigma_{mj}^2 n^{-1}\theta_j^2/\left(\theta_j^2 + n^{-1}\right) + \sigma_{mj}^2\left(\theta_j^2 + n^{-1}\right)\left(\lambda_j - \theta_j^2/\left(\theta_j^2 + n^{-1}\right)\right)^2\right], \tag{3.1}$$

where $\sigma_{mj}^2 = \int_0^1 (\varphi_j^{(m)}(t))^2\,dt \asymp (j+1)^{2m}$. One can verify directly [or refer to Tsybakov (1997)] that for estimating the mth derivative, the sharp smoothing weights in (1.2) are equal to $\lambda_{mj} = \theta_{mj}^2/(\theta_{mj}^2 + n^{-1})$ where $\{\theta_{mj}^2\} = \arg\max_{\{\{\theta_j^2\}:\, f \in W_2^{\alpha}(Q)\}}(\sum_{j=0}^{\infty} \sigma_{mj}^2 n^{-1}\theta_j^2/(\theta_j^2 + n^{-1}))$; the right side of (1.3) (i.e., the minimax risk) is equal to $\sum_{j=0}^{\infty} \sigma_{mj}^2 n^{-1}\theta_{mj}^2/(\theta_{mj}^2 + n^{-1})(1 + o(1))$. [The analysis of these results reveals the familiar saddlepoint property of (3.1) that is the key idea of the sharp minimax estimation.] Also note that $J_k \asymp J_r$ and $J_k < J_r$; here J_m is the shorthand for $J(n, m, \alpha, Q)$.

Thus, if a smoothing linear estimate is simultaneously sharp minimax then

$$\sum_{j=1}^{J_r} j^{2m}\left(\lambda_j - \lambda_{mj}\right)^2 = o(1)\,J_r^{2m+1}, \qquad m \in \{k, r\}. \tag{3.2}$$

On the other hand, direct calculation shows that for some $c_1 > 0$,

$$\sum_{j=1}^{J_r} j^{2m}\left(\lambda_{kj} - \lambda_{rj}\right)^2 > c_1 J_r^{2m+1}, \qquad m \in \{k, r\}. \tag{3.3}$$

Thus, if a simultaneously minimax smoothing estimate exists then

$$\begin{aligned} c_1 J_r^{2r+1} &\le \sum_{j=1}^{J_r} j^{2r}\left(\lambda_{kj} - \lambda_{rj}\right)^2 \\ &\le 2\sum_{j=1}^{J_r} j^{2r}\left(\lambda_j - \lambda_{rj}\right)^2 + 2J_r^{2(r-k)}\sum_{j=1}^{J_r} j^{2k}\left(\lambda_j - \lambda_{kj}\right)^2 = o(1)\,J_r^{2r+1}. \end{aligned}$$

This contradiction proves the desired assertion for the smoothing linear estimates.

An arbitrary linear estimate $\hat{f}(t)$ may be written as $\hat{f}(t) = f(t,\{\lambda_j\}) + \sum_{j=0}^{\infty} \hat{b}_j \varphi_j(t) = \sum_{j=0}^{\infty}(\lambda_j \hat{\theta}_j + \hat{b}_j)\varphi_j(t)$ where $\hat{b}_j$ is a linear combination of 1 and $\{\hat{\theta}_l,\, l \ne j\}$.

Using the previous notation, let ζ_{mj} be independent Bernoulli random variables that take on the values $\pm\theta_{mj}$ with equal probabilities. Then we use a traditional method of estimating a minimax risk from below by a Bayes risk with $\theta_j = \zeta_{mj}$. Write

$$\begin{aligned}
&\sup_{f\in W_2^{\alpha}Q} E_f\left\{\int_0^1 \left(\hat{f}^{(m)}(t) - f^{(m)}(t)\right)^2 dt\right\}\\
&\quad\geq \sum_{j=0}^{\infty} \sigma_{mj}^2 E\left\{\left(\lambda_j\hat{\theta}_j + \hat{b}_j - \zeta_{mj}\right)^2\right\}\\
&\quad= \sum_{j=0}^{\infty} \sigma_{mj}^2 E\left\{\left(\lambda_j\hat{\theta}_j - \zeta_{mj}\right)^2\right\} + \sum_{j=0}^{\infty} \sigma_{mj}^2 E\left\{\hat{b}_j^2\right\}\\
&\quad\geq \sum_{j=0}^{\infty} \sigma_{mj}^2 E\left\{\left(\lambda_j\hat{\theta}_j - \zeta_{mj}\right)^2\right\}\\
&\quad= \sum_{j=0}^{\infty} \Big[\sigma_{mj}^2 n^{-1}\theta_{mj}^2/\left(\theta_{mj}^2 + n^{-1}\right)\\
&\qquad\qquad + \sigma_{mj}^2\left(\theta_{mj}^2 + n^{-1}\right)\left(\lambda_j - \theta_{mj}^2/\left(\theta_{mj}^2 + n^{-1}\right)\right)^2\Big],
\end{aligned} \tag{3.4}$$

where $\hat{\theta}_j = \zeta_{mj} + n^{-1/2}\xi_j$. To get the first equality, the independence of $(\lambda_j\hat{\theta}_j - \zeta_{mj})$ and $\hat{b}_j$ as well as the equality $E\{\lambda_j\hat{\theta}_j - \zeta_{mj}\} = 0$ have been used, and the last line has been obtained by a direct calculation.

The last line in (3.4) is identical to (3.1) with the minimax θ_{mj}^2 in place of θ_j^2; thus the general case is converted to the case studied above of the smoothing linear estimates. □

Acknowledgments. I thank an Associate Editor and the referees for helpful comments.

REFERENCES

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539.

EFROMOVICH, S. and PINSKER, M. S. (1984). Learning algorithm for nonparametric filtering. *Automat. Remote Control* **11** 1434–1440.

NEMIROVSKII, A. (1985). Nonparametric estimation of smooth regression function. *J. Comput. System Sci.* **23** 1–11.

RICE, J. and ROSENBLATT, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Ann. Statist.* **11** 141–156.

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

TSYBAKOV, A. B. (1997). Asymptotically efficient nonparametric estimation in L_2 with general loss. *Problems Inform. Trans.* **33**. To appear.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF NEW MEXICO
ALBUQUERQUE, NEW MEXICO 87131
E-MAIL: efrom@math.umn.edu