

# Sharp Lower Bound for Regression with Measurement Errors and Its Implication for Ill-Posedness of Functional Regression

Sam Efromovich\*

*Department of Mathematical Sciences, UT Dallas, Richardson, TX 75080 USA*

Received April 5, 2023; revised July 15, 2023; accepted July 23, 2023

**Abstract**—Nonparametric regression estimation with Gaussian measurement errors in predictors is a classical statistical problem. It is well known that the errors dramatically slow down the rate of regression estimation, and this paper complement that result by presenting a sharp constant. Then an interesting example of using this sharp constant to discover a new curse of dimensionality in functional nonparametric regression is presented, and analysis of real data complements the theory.

**DOI:** 10.3103/S1066530723030031

**Keywords:** additive model, measurement error, MISE, sharp minimax

## 1. INTRODUCTION

Consider the following classical nonparametric regression problem with measurement errors in predictors. We are interested in estimation of regression function  $g(x) := \mathbb{E}\{Y|X\}$  over an interval  $[a, b]$ ,  $a < b$  under the mean integrated squared error (MISE) criterion when the predictor  $X$  is not observed directly due to measurement errors. Namely, estimation of the regression  $g(x)$  is based on a sample of size  $n$  from pair  $(Z, Y)$  and

$$Z = X + \sigma\eta. \tag{1}$$

Here  $\eta$  is independent from  $X$  standard normal random variable and  $\sigma > 0$ . There is a vast literature devoted to this problem, see [3, 4, 6, 7, 11, 24, 31, 35–37] where further references may be found.

The pathbreaking result in the theory of the nonparametric regression is the optimal rate of MISE convergence established by Fan and Truong in [10]. Let us formulate that result. Consider a positive integer  $\alpha$ , set  $f^{(\alpha)}(x) := d^\alpha f(x)/d^\alpha x$ , introduce a function class of design densities  $f^X$  of the predictor  $X$ ,

$$\mathcal{F}(\alpha, Q_0) := \{f : |f^{(\alpha)}(x)| < Q_0 < \infty, f(x) > Q_0^{-1}, x \in [a, b]\}, \tag{2}$$

and the Sobolev class of regression functions

$$\mathcal{S}(\alpha, Q) := \left\{ g : \int_a^b [g^{(\alpha)}(x)]^2 dx \leq Q \right\}. \tag{3}$$

Consider an oracle who knows a sample of size  $n$  from  $(Z, Y)$  as well as parameter  $\sigma$  in (1), design density  $f^X$  and the two function classes. Then the following lower bound holds for the oracle's estimators  $\tilde{g}$ ,

$$\liminf_{n \rightarrow \infty} [\ln(n)]^\alpha \inf_{\tilde{g}} \sup_{f^X \in \mathcal{F}(\alpha, Q_0), g \in \mathcal{S}(\alpha, Q)} \mathbb{E} \left\{ \int_a^b (\tilde{g}(x) - g(x))^2 dx \right\} > 0. \tag{4}$$

Further, the rate  $[\ln(n)]^{-\alpha}$  of the MISE convergence is attainable and accordingly it is called optimal. This is a remarkable result because it shows that the gaussian measurement error, which frequently

\*E-mail: efrom@utdallas.edu

occurs in applications, slows down the rate of MISE convergence to logarithmic while for the case of no measurement errors the rate is  $n^{-2\alpha/(2\alpha+1)}$ . Accordingly, the problem of regression with measurement errors in predictors is called ill-posed.

The aim of this paper is to complement the rate-optimal lower bound (4) by a sharp bound that reveals a minimal constant of the MISE convergence and how it depends on variance  $\sigma^2$  of the measurement error in (1) and parameters of the two function classes considered in (4). The sharp lower bound will be also complemented by an interesting example of its application in nonparametric functional regression.

The content of the article is as follows. The second section is devoted to the new sharp lower bound, the example of its application to analysis of functional regression is presented in Section 3, and proofs can be found in Section 4. Conclusions are in Section 5. While the article is pure theoretical in its nature, the interested reader can find a practical example of nonparametric functional regression in the online Supplementary Materials.

## 2. SHARP LOWER BOUND

We are considering the problem of estimating a regression function  $g(x) := \mathbb{E}\{Y|X = x\}$  over interval  $[a, b]$ ,  $a < b$  under the MISE criterion when available data is a sample of size  $n$  for a pair of random variables  $(Z, Y)$  and  $Z = X + \sigma\eta$  is defined in (1). Our aim is to complement the rate-optimal lower bound (4) by a sharp one.

In addition to function classes  $\mathcal{F}(\alpha, Q_0)$  and  $\mathcal{S}(\alpha, Q)$ , defined in (2) and (3) and used in the rate-optimal lower bound, let us introduce a sequence of increasing intervals for underlying standard deviations  $\sigma$  of the measurement errors. Let  $\gamma_n$  be a positive sequence such that  $\gamma_n \leq 1$  and  $\gamma_n \rightarrow 0$  as  $n \rightarrow \infty$ , and  $\sigma_0$  is a positive constant. Then we set

$$\mathcal{T}_n := \mathcal{T}(n, \sigma_0, \gamma_n) := \{\sigma : \sigma_0 \leq \sigma \leq [\ln(n)\gamma_n]^{1/2}\}. \tag{5}$$

In what follows  $o_n(1)$  denotes vanishing sequences in  $n$ .

**Theorem 1.** *Consider the problem of estimation of regression function  $g(x) = \mathbb{E}\{Y|X = x\}$  based on a sample of size  $n$  from  $(Z, Y)$  where  $Z$  is defined in (1). Then the following minimax lower bound holds for the MISE of oracle-estimators,*

$$\inf_{\sigma \in \mathcal{T}} \inf_{\tilde{g}} \sup_{f^X \in \mathcal{F}(\alpha, Q_0), g \in \mathcal{S}(\alpha, Q)} \mathbb{E} \left\{ \int_a^b [\sigma^{-\alpha}(\tilde{g}(x) - g(x))]^2 dx \right\} \geq Q[\ln(n)]^{-\alpha}(1 + o_n(1)). \tag{6}$$

Here the infimum in  $\tilde{g}$  is over all oracle-estimators  $\tilde{g}$  that know data,  $\sigma$ ,  $f^X$ ,  $\mathcal{T}_n$  and the function classes  $\mathcal{F}(\alpha, Q)$  and  $\mathcal{S}(\alpha, Q)$ .

Now let us show that the oracle's lower bound (6) is sharp. To show that we will solve a problem formulated by Hall and Qui in [18]. The problem and its solution are presented in the following proposition.

**Theorem 2.** *Consider estimation of a regression function  $g(x) := \mathbb{E}\{Y|X = x\}$  over interval  $[0, 1]$ . Predictor  $X$  is uniformly distributed on  $[0, 1]$ , regression function  $g \in \mathcal{S}(\alpha, Q)$  and periodic with period 1. Available sample of size  $n$  is from  $(Z, Y)$  where  $Z$  is defined in (1) with known parameter  $\sigma$ . Set  $\varphi_0(x) = 1, \varphi_s(x) = 2^{1/2} \cos(\pi s x), s = 1, 2, \dots, \hat{\theta}_s := n^{-1} \sum_{l=1}^n Y_l \varphi_s(Z_l) e^{(\pi\sigma s)^2/2}$  and*

$$\hat{g}(x) = \sum_{s=0}^{S_n} \hat{\theta}_s \varphi_s(x). \tag{7}$$

Here  $S_n$  is the smallest integer larger than  $(1/(\pi\sigma)[\ln(n+3) - (\ln(\ln(n+3)))]^{1/2}$ . Then

$$\sup_{g \in \mathcal{S}(\alpha, Q)} \mathbb{E}_g \left\{ \int_0^1 [\sigma^{-\alpha}(\hat{g}(x) - g(x))]^2 \right\} \leq Q[\ln(n)]^{-\alpha}(1 + o_n(1)). \tag{8}$$

Theorem 2 verifies that the lower bound in Theorem 1 is sharp and its constant cannot be increased. Furthermore, let us stress that the estimator (7) does not depend on smoothness of an estimated regression function  $g$  because it is constructed without using parameters  $(\alpha, Q)$ . This is an important conclusion because adaptation to smoothness of regression function is the most complicated problem in regression analysis. As we see, at least in this matter the measurement errors simplify regression estimation, and this is a welcome relief.

### 3. APPLICATION TO NONPARAMETRIC FUNCTIONAL REGRESSION

There is an interesting example of application of the sharp lower bound in nonparametric functional regression. Let us explain the connection, and the reader interested in functional regression can find comprehensive reviews in [1, 2, 5, 14, 16, 25, 37]. Consider a pair  $(\mathcal{X}, Y)$  where  $\mathcal{X} := \{X(t), 0 \leq t \leq 1\}$  is a random function (process, trajectory). The problem is to estimate the nonparametric functional regression  $\mathbb{E}\{Y|\mathcal{X}\}$ , but similarly to the classical regression with measurement errors in predictors, the available sample of size  $n$  is from  $(\mathcal{Z}', Y)$  where

$$\mathcal{Z}' := \left\{ Z'(t) = \int_0^t X(\tau) d\tau + \nu B(t), \quad 0 \leq t \leq 1 \right\}, \quad (9)$$

$B(t)$  is a standard Brownian motion and  $\nu$  is a positive constant.

A traditional functional regression methodology uses a two-stage procedure when on the first stage an underlying process  $\mathcal{X}$  is approximated by a Fourier series of order  $p := p_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Let us denote the corresponding Fourier coefficients as  $(U_1, U_2, \dots, U_p)$ . Then the second stage solves a multivariate regression problem of estimating  $\mathbb{E}\{Y|U_1 = u_1, U_2 = u_2, \dots, U_p = u_p\}$ , which in its turn, to remedy the curse of multidimensionality, is approximated by an additive regression

$$g(u_1, u_2, \dots, u_p) := q_1(u_1) + q_2(u_2) + \dots + q_p(u_p). \quad (10)$$

Now we can see a connection between functional regression and classical regression with measurement errors in predictors. Using an orthogonal basis on  $[0, 1]$ , say the cosine basis, we get from (9) that in place of  $U_j$  we observe

$$Z'_j = U_j + \nu \eta_j, \quad j = 1, 2, \dots, p, \quad (11)$$

and  $\eta_j, j = 1, 2, \dots, p$  are independent standard normal variables, see [34]. Accordingly, in (10) each additive component  $q_j$  is a univariate regression with measurement errors in predictors.

Now we are in a position to explain why the sharp constant is needed for analysis of the additive regression. On first glance, the additive components look alike and the problem is similar to those considered in the classical regression theory of additive models. But there is a dramatic difference. Only to be specific, assume that an underlying trajectory (functional predictor)  $X(t)$  is  $\beta$ -fold differentiable and its Fourier coefficients  $U_j$  decrease with the rate  $j^{-\beta}$ , see [7]. Fast decreasing covariates are highly sought-after in functional regression, but then we are dealing with atypical nonparametric regression setting where support of  $U_j$  vanishes as  $j \rightarrow \infty$ . This is not a setting considered in classical approximation analysis and nonparametric regression theory where supports for individual variables are assumed to be comparable and this allows us to introduce a feasible notion of smoothness of a multivariate function in each direction, see [19, 25, 38]. To remedy the issue, let us rescale each  $U_j$  by multiplying it by  $j^\beta$  and set  $X_j := j^\beta U_j, Z_j := j^\beta Z'_j$ . Then (11) becomes

$$Z_j = X_j + \nu j^\beta \eta_j, \quad j = 1, 2, \dots, p_n, \quad (12)$$

and we may approximate  $\mathbb{E}\{Y|X_1 = x_1, X_2 = x_2, \dots, X_{p_n} = x_{p_n}\}$  by an additive model

$$g(x_1, x_2, \dots, x_{p_n}) = g_1(x_1) + g_2(x_2) + \dots + g_{p_n}(x_{p_n}). \quad (13)$$

The model (12)–(13) allows us to use the same measure of smoothness for all  $p_n$  functions regardless of how large  $p_n$  is. For instance, we can make a traditional assumption that all functions  $g_j$  are  $\alpha$ -fold continuously differentiable. Further, because  $p_n \rightarrow \infty$  we need to understand how the variance  $(\nu j^\beta)$

affects estimation of  $g_j$ . The latter makes the functional regression a perfect example for using the sharp lower bound.

Let us apply results of the previous section to the nonparametric functional model (12)–(13). Recall that sequence  $\gamma_n$  is defined above line (5). Our first result is the following sharp lower bound.

**Theorem 3.** *Consider a functional model (12)–(13) and estimation of a particular additive component  $g_j$  with  $j \leq \lfloor \ln(n)\gamma_n/\nu \rfloor^{1/2\beta}$ . Then*

$$\inf_{\check{g}_j} \sup_{f^X \in \mathcal{F}(\alpha, Q_0), g_j \in \mathcal{S}(\alpha, Q)} \mathbb{E} \left\{ \int_a^b (\check{g}_j(x) - g_j(x))^2 dx \right\} \geq Q \frac{(\nu j^\beta)^{2\alpha}}{[\ln(n)]^\alpha} (1 + o_n^*(1)). \tag{14}$$

Here the infimum is over oracle-estimators  $\check{g}_j$  based on data, distribution of  $X$  and parameters  $(\alpha, Q, \beta, \nu)$ , and  $o_n^*(1) \rightarrow 0$  as  $n \rightarrow \infty$  uniformly over the considered indexes  $j$ . Further, the lower bound is sharp.

The rate  $[\ln(n)]^{-\alpha}$  in (14) is well known in the functional regression literature thanks to the above-mentioned lower bound of [10]. The new here is the factor (traditionally referred to as the “constant”)  $Q(\nu j^\beta)^{2\alpha}$  which is sharp and quantifies the effect of: (i) Particular additive component in terms of its index  $j$ ; (ii) Smoothness of the function-predictor  $\mathcal{X}$  in terms of the parameter  $\beta$ ; (iii) Smoothness of the univariate regression function  $g_j(x)$  in terms of the parameter  $\alpha$ ; (iv) Standard deviation  $\nu$  of the Brownian measurement error-process.

The important corollary from Theorem 3 is that even if all additive components  $g_j$  in (13) have the same smoothness, for each  $g_j$  decrease of the MISE slows down by factor  $j^{2\beta\alpha}$ . This phenomenon defines the new *curse of dimensionality in ill-posedness* of nonparametric functional regression. To the best of my knowledge, there is no other example of such a curse in statistical literature. Another surprising outcome is that the smoother an underlying function-predictor  $\mathcal{X}$  is ( $\beta$  is larger), the larger increase in the MISE due to adding an extra covariate. This outcome is in contrary to known results in nonparametric curve estimation where smoother curves imply faster estimation.

Now let us present an insightful assertion which sheds light on the number  $p = p_n$  of feasible additive components in (13). We are considering two possible “extreme” scenarios about the underlying functions. The first one is when all additive components, regardless of  $p$ , are from the same Sobolev class  $\mathcal{S}(\alpha, Q)$ , that is they all may have the same Sobolev’s power. The second one is when the  $p$ -variate function  $g(\mathbf{x})$ ,  $\mathbf{x} := (x_1, \dots, x_p)$  belongs to a  $p$ -variate Sobolev class  $\mathcal{S}_p(\alpha, Q) := \{g : g(\mathbf{x}) = \sum_{\mathbf{j} \in \mathcal{N}_p} \theta_{\mathbf{j}} \varphi_{\mathbf{j}}(\mathbf{x}), \sum_{\mathbf{j} \in \mathcal{N}_p} [1 + \sum_{i=1}^p (\pi j_i)^{2\alpha}] \theta_{\mathbf{j}}^2 \leq Q, \mathbf{j} := (j_1, \dots, j_p), \mathcal{N}_p := \{0, 1, \dots\}^p, \varphi_{\mathbf{j}}(\mathbf{x}) := \prod_{i=1}^p \varphi_{j_i}(x_i)\}$  whose discussion can be found in Nikolskii (1975) and Hoffmann and Lepski (2002). Note that in the second case the total Sobolev’s power of an estimated functional regression  $g(\mathbf{x})$  is bounded and parameters  $(\alpha, Q)$  do not change with  $p = p_n$ . Accordingly, the additive components  $g_j$  share the total power. These two cases are the extremes, and this is why it is of interest to explore them.

**Theorem 4.** *Consider functional regression (8)–(9) where either each additive component  $g_j \in \mathcal{S}(\alpha, Q)$  or the additive regression  $g \in \mathcal{S}_p(\alpha, Q)$ . Then for consistent estimation of the functional regression the number  $p$  of additive components should not exceed*

$$p_n^* = o_n(1)[\ln(n)]^{\frac{\alpha}{2\alpha\beta+1}} \quad \text{or} \quad p_n' = o_n(1)[\ln(n)]^{\frac{1}{2\beta}} \tag{15}$$

for the two considered cases, respectively.

Let us make several comments about assertion of Theorem 4. First, keeping in mind that typical function-predictors are smooth ( $\beta$  is large), the two extreme cases imply surprisingly similar and extremely small bounds (15) for a feasible dimensionality of functional regression. Second, consider the case when it is known that the functions are differentiable, that is  $\min(\alpha, \beta) \geq 1$ . Then we get that that  $[\ln(n)]^{\frac{\alpha}{2\alpha\beta+1}} = [\ln(n)]^{\frac{1}{2\beta+1/\alpha}} = o_n(1)[\ln(n)]^{1/2\beta} = o_n(1)[\ln(n)]^{1/2}$ . Accordingly, even with respect to other classical ill-posed problems, the maximal number of components is small. Third, considered sets of  $j$  in Theorems 1–3 are sufficiently large for analysis of consistent functional regression estimators. Fourth, note that the smoother  $\mathcal{X}$ , the smaller the number of additive components regardless of smoothness of an underlying functional regression.

We finish this section with the following remark. The lower bound (14) is obtained for the model (9) of continuous in time observation of process  $Z(t), t \in [0, 1]$  which is a classical one in the functional regression literature. In some practical examples observations of  $Z(t)$  are made at discrete points in time, and then instead of a stochastic process we observe a time series  $\{Z(t_i), 0 \leq t_1 \leq \dots \leq t_m \leq 1\}$ . Apparently, the above-presented lower bounds still hold for a time series of observations with any  $m$ . On first glance, this conclusion contradicts [26] where it is asserted that ill-posedness disappears as  $m \rightarrow \infty$ . But the contradiction is due to different underlying models. In [26] it is assumed that available observations are  $Z(t_i) = X(t_i) + \nu\xi_i, i = 1, \dots, m_n$  where errors  $\xi_i$  are iid and  $m_n \rightarrow \infty$  sufficiently fast as  $n \rightarrow \infty$ . In other words, observations of the function-predictor follow a classical regression model with independent errors, and the observations may be made as frequently in time as desired. This is a nice model to study because you can restore an underlying  $X(t)$  as well as you wish and accordingly vanish measurement errors in predictors. Unfortunately, there is no such remedy for stochastic processes considered in this paper.

#### 4. PROOFS

**Proof of Theorem 1.** An interesting aspect of the proof is the used technique of two hypotheses. Traditionally more complicated techniques are used based on using Bayesian approaches, see a discussion in [7, 8, 10, 34]. Accordingly, let us begin with formally introducing the two hypotheses that the oracle uses in the proof.

To verify the lower bound, the oracle chooses the following parametric class of joint densities  $f^{X,Y}(x, y), (x, y) \in (-\infty, \infty)^2$  of predictor  $X$  and response  $Y$ .

(i) Marginal density  $f^X$  of the predictor is fixed (not changing with  $n$  or  $\sigma$ ) and known. It is  $\alpha$ -fold continuously differentiable on  $[a, b]$ , belongs to class (2),

$$\int_a^b f^X(x) dx = 1 - \alpha_1, \quad 0 < \alpha_1 < 1, \quad (16)$$

and for  $x \in (-\infty, \infty)$

$$f^X(x) \geq c_{\alpha_0}(1 + x^2)^{-\alpha_0}, \quad \alpha_0 > 1/2, c_{\alpha_0} > 0. \quad (17)$$

(ii) Regression function  $g(x) := \mathbb{E}\{Y|X = x\}$  has domain  $(-\infty, \infty)$  and for  $x \in [a, b]$  it belongs to the Sobolev class (3). Further, it is known that regression function is

$$g(x) = g_\theta(x) := u(n, \sigma, Q)[c^* + \theta v(n, \sigma, x)],$$

where the absolute constant  $c^*$  and functions  $u$  and  $v$  are chosen by the oracle and  $\theta$  is an unknown parameter taking on values  $-1$  and  $1$ . Accordingly, for the oracle the regression is parametric in  $\theta \in \{-1, 1\}$ . The absolute constant and the two functions will be defined shortly, and their choice implies that the range of  $g_\theta(x)$  is  $[0, 1]$  and this property is used below in (19).

(iii) Let  $p_0$  and  $p_1$  be two known densities on  $(-\infty, \infty)$  that have zero and unit means, respectively, and

$$\int_{-\infty}^{\infty} \frac{[p_0(y) - p_1(y)]^2}{\min(p_0(y), p_1(y))} dy < \infty. \quad (18)$$

Then the conditional density  $f^{Y|X}$  of  $Y$  given  $X$  is defined as a mixture of the above-introduced densities  $p_0$  and  $p_1$  with the weight being the regression function  $g_\theta(x)$ , namely

$$f^{Y|X}(y|x) := f_\theta^{Y|X}(y|x) = p_0(y)[1 - g_\theta(x)] + g_\theta(x)p_1(y), \quad (x, y) \in (-\infty, \infty)^2. \quad (19)$$

Accordingly, the conditional density  $f_\theta^{Y|X}(y|x)$  and the joint density  $f^{X,Y}(x, y) := f_\theta^{X,Y}(x, y) := f^X(x)f_\theta^{Y|X}(y|x)$  are known up to the parameter  $\theta \in \{-1, 1\}$ . This defines the two hypotheses used to verify the sharp lower bound for the MISE.

We need several more notations and auxiliary results. Consider a nonnegative and symmetric about zero mollifier  $\phi(t)$  which is infinitely differentiable on the real line, equal to zero beyond  $\{(-2, -1) \cup (1, 2)\}$  and  $\int_{-\infty}^{\infty} \phi^2(t) dt = 1$ , see Section 7.1 in [7]. Using this mollifier, introduce two infinitely differentiable on the real line functions

$$G_1(x) := \frac{1}{\pi} \int_1^2 \cos(tx) \phi(t) dt, \quad G_2(x) := -\frac{1}{\pi} \int_1^2 \sin(tx) \phi(t) dt. \quad (20)$$

For a positive integer  $J$  introduce an infinitely differentiable on the real line function

$$H_J(x) := \cos(Jx)G_1(x) + \sin(Jx)G_2(x). \quad (21)$$

Let us comment on properties of function  $H_J(x)$  that will be used shortly. We have  $|H_J(x)| < 2/\pi$  and

$$\int_{-1/2}^{1/2} H_J^2(x) dx = \int_{-1/2}^{1/2} [\cos^2(Jx)G_1^2(x) + \sin^2(Jx)G_2^2(x) + \sin(2Jx)G_1(x)G_2(x)] dx. \quad (22)$$

Using differentiability of  $H_J(x)$ , trigonometric formulae  $\cos^2(Jx) = 1/2 + (1/2)\cos(2Jx)$  and  $\sin^2(Jx) = 1/2 - (1/2)\cos(2Jx)$ , results of Section 2.2 in [7] on how fast Fourier transforms of differentiable functions decrease, and the Plancherel identity we conclude that

$$\begin{aligned} \int_a^b H_J^2(x) dx &= (1/2) \int_a^b [G_1^2(x) + G_2^2(x)] dx (1 + o_J(1)) \\ &\leq (1/2) [2 \int_{-\infty}^{\infty} [\phi(t)]^2 dt] (1 + o_J(1)) = 1 + o_J(1). \end{aligned} \quad (23)$$

Here and in what follows  $o_J(1)$ s are generic sequences such that  $o_J(1) \rightarrow 0$  as  $J \rightarrow \infty$ .

Now let us explore derivatives of  $H_J(x)$ . Denote by  $q^{(s)}(x)$  the  $s$ th derivative of a function  $q(x)$ , and write

$$\begin{aligned} H_J^{(1)}(x) &= J[-\sin(Jx)G_1(x) + \cos(Jx)G_2(x)] \\ &\quad + [\cos(Jx)G_1^{(1)}(x) + \sin(Jx)G_2^{(1)}(x)]. \end{aligned} \quad (24)$$

Similarly to (23)–(24), we can conclude that

$$\begin{aligned} \int_a^b [H_J^{(1)}(x)]^2 dx &= J^2(1/2) \int_a^b [G_1^2(x) + G_2^2(x)] dx (1 + o_J(1)) \\ &= J^2 \int_a^b H_J^2(x) dx (1 + o_J(1)). \end{aligned} \quad (25)$$

For  $\alpha > 1$ , we may repeat (24)–(25) and get that the  $\alpha$ th derivative of  $H_J(x)$  satisfies

$$\int_a^b [H_J^{(\alpha)}(x)]^2 dx = J^{2\alpha} \int_a^b H_J^2(x) dx (1 + o_J(1)). \quad (26)$$

Now recall that the probability density of  $X$  is  $f^X(x)$ . In what follows we may use notation  $f_0(x) := f^X(x)$  to simplify formulas. Using functions  $H_J$  and  $f_0$  we introduce a new function

$$m_J(x) := H_J(x)/f_0(x). \quad (27)$$

Note that the function is continuously differentiable at least  $\alpha$  times and

$$m_J^{(1)}(x) = \frac{H_J^{(1)}(x)}{f_0(x)} - \frac{H_J(x)f_0^{(1)}(x)}{f_0^2(x)}. \tag{28}$$

Using (24) and following (25), we conclude that

$$\int_a^b [m_J^{(1)}(x)]^2 dx = J^2 \int_a^b [H_J(x)/f_0(x)]^2 dx (1 + o_J(1)). \tag{29}$$

The reader may compare this result with (25) and also note that the relation (29) becomes plain if  $f_0(x)$  is constant on  $[a, b]$ . Repeating (28) and (29) we get a relation mimicking (26), namely

$$\int_a^b [m_J^{(\alpha)}(x)]^2 dx = J^{2\alpha} \int_a^b [H_J(x)/f_0(x)]^2 dx (1 + \gamma_J^*), \quad \gamma_J^* = o_J(1). \tag{30}$$

We need to establish one more property of function  $H_J$  that will allow us to study ratio  $H_J(x)/f_0(x)$  on the real line. Using integration by parts and the boundary property of mollifier  $\phi(t)$  we get for  $x \neq 0$ ,

$$\begin{aligned} G_1(x) &= (1/\pi) \int_1^2 \cos(tx)\phi(t)dt = (1/\pi x) \sin(tx)\phi(t)|_{t=1}^{t=2} - (1/\pi x) \int_1^2 \sin(tx)\phi^{(1)}(t)dt \\ &= -(1/\pi x) \int_1^2 \sin(tx)\phi^{(1)}(t)dt. \end{aligned} \tag{31}$$

We can continue the integration by parts and get (recall that  $\alpha_0$  was introduced in the definition of density  $f^X$  and  $C_s$  denote generic positive constants)

$$|G_1(x)| \leq \frac{C}{(1+x^2)^{\alpha_0+1}}. \tag{32}$$

Similarly we establish that  $|G_2(x)| \leq C/(1+x^2)^{\alpha_0+1}$ , and combining the results yields that for some constant  $C_H$  not depending on  $J$  we have

$$|H_J(x)| \leq \frac{C_H}{(1+x^2)^{\alpha_0+1}}. \tag{33}$$

In its turn, (33) and (17) yield that uniformly over all  $J$  and  $x \in (-\infty, \infty)$

$$\frac{|H_J(x)|}{f_0(x)} \leq C_H/c_{\alpha_0} =: c^* < \infty. \tag{34}$$

Our next step is to define the specific parametric regression function  $g_\theta(x)$  used in (19), that is we are going to define the constant  $c^*$  and functions  $u$  and  $v$ . Set

$$g_\theta(x) := a_J [c^* + \theta H_J(x)/f_0(x)], \quad \theta \in \{-1, 1\}. \tag{35}$$

Here  $a_J$  is a positive sequence in  $J$  such that

$$a_J^2 := \min \left( \frac{Q}{J^{2\alpha} \int_a^b (H_J(x)/f_0(x))^2 dx (1 + \gamma_J^*)} \frac{1}{[2c^*]^2} \right). \tag{36}$$

The used in (36) sequence  $\gamma_J^*$  is defined in (30), and positive constant  $c^*$  is defined in (34).

Note that we are dealing with just two underlying regression functions known up to a parameter  $\theta$  that takes either value  $-1$  or value  $1$ , recall (19). It is plain to see that  $0 \leq g_\theta(x) \leq 1$ , and using (30) we verify that the two regression functions belong to the Sobolev class  $\mathcal{S}(\alpha, Q)$ . Indeed, (30) and (36) yield

$$\int_a^b [g_\theta^{(\alpha)}(x)]^2 dx = \int_a^b [\theta a_J H_J(x)/f_0(x)]^2 J^{2\alpha} dx \leq Q. \tag{37}$$

Recall that the joint density of pair  $(X, Y)$  is

$$f_{\theta}^{X,Y}(x, y) = f^X(x)[p_0(y)(1 - g_{\theta}(x)) + p_1(y)g_{\theta}(x)]. \tag{38}$$

Now we are in a position to consider a corresponding minimax MISE for oracles that know everything apart of the value of parameter  $\theta \in \{-1, 1\}$ . Write,

$$\begin{aligned} R &:= \inf_{\tilde{g}} \sup_{g \in \mathcal{S}(\alpha, Q)} \mathbb{E}_g \left\{ \int_a^b [\sigma^{-\alpha}(\tilde{g}(x) - g(x))^2] dx \right\} \\ &\geq \inf_{\tilde{\theta}} \sup_{\theta \in \{-1, 1\}} \mathbb{E}_{\theta} \{ (\tilde{\theta} - \theta)^2 \} a_J^2 \sigma^{-2\alpha} \int_a^b [H_J^2(x)/f_0^2(x)] dx. \end{aligned} \tag{39}$$

Here  $\tilde{\theta}$  is an oracle-estimator of parameter  $\theta$ . Set

$$J = J_{n, \sigma} := \lceil \sigma^{-1} [\ln(n) + (\ln(\ln(n)))^2]^{1/2} \rceil, \tag{40}$$

where  $\lceil c \rceil$  denotes the smallest integer larger than  $c$ . Note that due to (17), the assumed  $\sigma \in \mathcal{T}_n$  and (23) we get

$$\inf_{\sigma \in \mathcal{T}_n} \min(J_{n, \sigma}, 1/a_J) \rightarrow \infty \quad \text{as } n \rightarrow \infty. \tag{41}$$

Accordingly, there exists a positive integer  $J_*$  such that  $a_J \leq 1/2c^*$  for all  $J \geq J_*$ . In other words, for  $J \geq J_*$  the sequence  $a_J^2$  is equal to the first term on the right side of (36).

We continue (39) for  $J \geq J_*$ ,

$$\begin{aligned} R &\geq J^{-2\alpha} \sigma^{-2\alpha} Q (1 + \gamma_J^*)^{-1} \inf_{\tilde{\theta}} \sup_{\theta \in \{-1, 1\}} \mathbb{E}_{\theta} \{ (\tilde{\theta} - \theta)^2 \} \\ &\geq J^{-2\alpha} \sigma^{-2\alpha} Q (1 + \gamma_J^*)^{-1} \inf_{\tilde{\theta}} \sup_{\theta \in \{-1, 1\}} [\mathbb{E}_{\theta} \{ |\tilde{\theta} - \theta| \}]^2. \end{aligned} \tag{42}$$

Consider the expectation on the right side of (42). Introduce a minimax parametric risk with absolute loss function

$$R_* := \inf_{\tilde{\theta}} \sup_{\theta \in \{-1, 1\}} \mathbb{E}_{\theta} \{ |\tilde{\theta} - \theta| \}. \tag{43}$$

Let us show that

$$R_* \geq (1 + o_n(1)), \tag{44}$$

and note that if (44) holds, then

$$\inf_{\tilde{\theta}} \sup_{\theta \in \{-1, 1\}} [\mathbb{E}_{\theta} \{ |\tilde{\theta} - \theta| \}]^2 \geq 1 + o_n(1). \tag{45}$$

To find a lower bound for  $R_*$  we bound it from below by a Bayes risk. Introduce a random variable  $\Theta$  taking two values  $-1$  and  $1$  with the equal probability  $0.5$ . The corresponding Bayes estimate is a sample median taking two values  $-1$  and  $1$ , see Section 2.4 in [23]. Write,

$$\begin{aligned} R_* &\geq \mathbb{E} \{ |\tilde{\Theta} - \Theta| \} = (1/2) 2\mathbb{P}(\tilde{\Theta} = 1 | \Theta = 1) + (1/2) 2\mathbb{P}(\tilde{\Theta} = -1 | \Theta = 1) \\ &\geq \inf_{\tau} [\mathbb{E} \{ \tau | \theta = -1 \} + \mathbb{E} \{ 1 - \tau | \theta = 1 \}]. \end{aligned} \tag{46}$$

Here the infimum is over all possible critical functions  $\tau$  for testing two simple hypothesis  $\theta = 1$  versus  $\theta = -1$  based on a sample  $(Z_1, Y_1), \dots, (Z_n, Y_n)$  from  $(Z, Y)$  where  $Z := X + \sigma\eta$ ,  $\eta$  is a standard normal random variable, and the parametric joint density of  $(X, Y)$  is defined in (38). Denote by  $f_{\theta}^{Z,Y}$  the joint density of  $(Z, Y)$  corresponding to the joint density  $f_{\theta}^{X,Y}$ , that is the convolution joint density. Using [21] we can continue (46),

$$R_* \geq 1 - (1/2) \left( \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{[f_1^{Z,Y}(z, y)]^2}{f_{-1}^{Z,Y}(z, y)} dz dy \right]^n - 1 \right). \tag{47}$$



Suppose that

$$D := \sup_{\sigma \in \mathcal{T}_n} \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{[f_1^{Z,Y}(z,y)]^2}{f_{-1}^{Z,Y}(z,y)} dz dy \right]^n \leq 1 + o_n(1). \tag{48}$$

Then combining (39)–(47), we establish that

$$R \geq J^{-2\alpha} \sigma^{-2\alpha} (1 + \gamma_J^*)^{-1} Q(1 + o_n(1)). \tag{49}$$

Now plug  $J = J_{n,\sigma}$ , defined in (40), and this verifies the lower bound of Theorem 1.

We are left with proving (48). Introduce notation  $q_1 * q_2(z) := \int_{-\infty}^{\infty} q_1(z-t)q_2(t)dt$  for the convolution of two functions. Our next step is to establish several properties of the convolutions  $f^X * f^{\sigma\eta}$  and  $H_J * f^{\sigma\eta}$ . Note that this is the first time when we are dealing with the measurement error in predictor and its density  $f^{\sigma\eta}$ . We are analyzing these two convolutions in turn. Due to (17) the density  $f^X(x)$  is not smaller than  $c_{\alpha_0}(1+x^2)^{-\alpha_0}$ ,  $\alpha_0 > 1/2$ . Write,

$$f^X * f^{\sigma\eta}(z) = \int_{-\infty}^{\infty} f^X(z-t)f^{\sigma\eta}(t)dt \geq \int_{-1}^1 f^{\sigma\eta}(t)f^X(z-t)dt \geq \frac{C}{\sigma(1+z^2)^{\alpha_0}}. \tag{50}$$

Now we are considering the convolution  $H_J * f^{\sigma\eta}$ . Introduce a symmetric about zero function  $\phi_J(t)$  such that  $\phi_J(t) = \phi(t-J)I(t \geq J)$  for  $t \geq 0$ , here  $I(\cdot)$  is the indicator. Then we can write that

$$\begin{aligned} H_J(x) &= (1/\pi) \cos(Jx) \int_1^2 \cos(tx)\phi(t)dt - (1/\pi) \sin(Jx) \int_1^2 \sin(tx)\phi(t)dt \\ &= (1/\pi) \int_1^2 \cos(x(J+t))\phi(t)dt = (1/\pi) \int_{1+J}^{2+J} \cos(xt)\phi(t-J)dt \\ &= (1/\pi) \int_0^{\infty} \cos(xt)\phi_J(t)dt = (1/2\pi) \int_{-\infty}^{\infty} e^{-itx}\phi_J(t)dt. \end{aligned} \tag{51}$$

Accordingly, we get

$$H_J * f^{\sigma\eta}(z) = (1/2\pi) \int_{-\infty}^{\infty} e^{-itz}\phi_J(t)e^{-\sigma^2 t^2/2}dt = (1/\pi) \int_{1+J}^{2+J} \cos(tz)\phi_J(t)e^{-\sigma^2 t^2/2}dt. \tag{52}$$

For  $z = 0$  the right side of (52) is bounded from above by  $Ce^{-\sigma^2 J^2/2}$ . Using integration by parts and the boundary property of  $\phi_J(t)$  we continue the analysis for  $z \neq 0$ ,

$$\begin{aligned} H_J * f^{\sigma\eta}(z) &= (1/\pi z) \sin(tz)\phi_J(t)e^{-\sigma^2 t^2/2} \Big|_{t=1+J}^{t=2+J} - (1/\pi z) \int_{1+J}^{2+J} \sin(tz)(d[\phi_J(t)e^{-\sigma^2 t^2/2}]/dt)dt \\ &= -(1/\pi z) \int_{1+J}^{2+J} \sin(tz)(d[\phi_J(t)e^{-\sigma^2 t^2/2}]/dt)dt. \end{aligned} \tag{53}$$

Let  $k$  be a minimal integer larger than  $\alpha_0 + 1/2$  where  $\alpha_0$  was introduced in (17). Repeating  $k - 1$  times integration by parts of the integral on the right side of (53) we get that for all  $z \in (-\infty, \infty)$

$$|H_J * f^{\sigma\eta}(z)| \leq \frac{C(J\sigma^2)^k e^{-\sigma^2 J^2/2}}{(1+z^2)^{k/2}}. \tag{54}$$

We need two more technical results. To simplify formulas in what follows the integrals are over  $(-\infty, \infty)^2$ . The first one is a familiar and directly verified relation for two densities  $q_1$  and  $q_2$  such that the support of  $q_1$  is the subset of the support of  $q_2$ ,

$$\int \frac{q_1^2(z, y)}{q_2(z, y)} dz dy = \int \frac{[q_1(z, y) - q_2(z, y)]^2}{q_2(z, y)} dz dy + 1. \tag{55}$$

The second technical relation is based on (34), (38), (50), (54), and (55). Write,

$$\begin{aligned} d(J, \sigma) &:= \int \frac{[f_1^{Z,Y}(z, y) - f_{-1}^{Z,Y}(z, y)]^2}{f_1^{Z,Y}(z, y)} dz dy \\ &= \int \frac{4a_J^2 [H_J * f^{\sigma\eta}(z)]^2 [p_1(y) - p_2(y)]^2}{f_1^{Z,Y}(z, y)} dz dy. \end{aligned} \tag{56}$$

Let us write down  $f_1^{Z,Y}$  and evaluate it from below

$$\begin{aligned} f_1^{Z,Y}(z, y) &= p_0(y)[f^X * f^{\sigma\eta}(z)(1 - a_J c^*) - a_J H_J * f^{\sigma\eta}(z)] + p_1(y)a_J[c^* f^X * f^{\sigma\eta}(z) + H_J * f^{\sigma\eta}(z)] \\ &\geq p_0(y)[f^X * f^{\sigma\eta}(z)(1/2) - a_J H_J * f^{\sigma\eta}(z)] + p_1(y)a_J[c^* f^X * f^{\sigma\eta}(z) - |H_J * f^{\sigma\eta}(z)|]. \end{aligned} \tag{57}$$

For all large  $J$ , and accordingly for all large  $n$ , we have

$$c^* f^X * f^{\sigma\eta}(z) - |H_J * f^{\sigma\eta}(z)| > (1/2)c^* f^X * f^{\sigma\eta}(z). \tag{58}$$

Using this inequality we continue (57) and get that for all large  $n$

$$f_1^{Z,Y}(z, y) \geq (1/4)[\min(p_0(y), p_1(y))]f^X * f^{\sigma\eta}(z). \tag{59}$$

Now we use this inequality on the right side of (58) and get for all large  $J$  that

$$\begin{aligned} d(J, \sigma) &\leq \int \frac{16a_J^2 [H_J * f^{\sigma\eta}(z)]^2 [p_1(y) - p_2(y)]^2}{[\min(p_0(y), p_1(y))]f^X * f^{\sigma\eta}(z)} dz dy \\ &\leq C a_J^2 \sigma (J\sigma^2)^{2k} e^{-\sigma^2 J^2} \leq C \sigma^{4k+1} J^{-2(\alpha-k)} e^{-\sigma^2 J^2}. \end{aligned} \tag{60}$$

Now recall definition (40) of  $J_{n,\sigma}$ , plug  $J = J_{n,\sigma}$  in the right side of (60) and get

$$\sup_{\sigma \in \mathcal{T}} d(J_{n,\sigma}, \sigma) = o_n(1)n^{-1}. \tag{61}$$

Using this relation,  $[1 + o_n(1)n^{-1}]^n = 1 + o_n(1)$  and (56) verify (48). Theorem 1 is proved.

**Proof of Theorem 2.** First, it is straightforward to check that the mean squared error  $\mathbb{E}\{(\hat{\theta}_s - \theta_s)^2\} \leq Cn^{-1}e^{(\pi\sigma s)^2}$ . Second, we have  $\sup_{g \in \mathcal{S}(\alpha, Q)} \sum_{s > S_n} \theta_s^2 \leq Q/(\pi S_n)^{2\alpha}(1 + o_n(1))$ , see [34]. Third, a straightforward calculation, based on the Parseval identity and the Dawson formula, shows that for  $g \in \mathcal{S}(\alpha, Q)$  and  $\sigma \in \mathcal{T}_n$  the MISE satisfies the following relations,

$$\begin{aligned} \mathbb{E}_g \left\{ \int_0^1 [\sigma^{-\alpha}(\hat{g}(x) - g(x))]^2 \right\} &= \sigma^{-2\alpha} \sum_{s=0}^{S_n} \mathbb{E}_g \{(\hat{\theta}_s - \theta_s)^2\} + \sigma^{-2\alpha} \sum_{s > S_n} \theta_s^2 \\ &\leq C \sigma^{-2\alpha} n^{-1} \sum_{s=1}^{S_n} e^{(\pi\sigma s)^2} + \sigma^{-2\alpha} Q/(\pi S_n)^{2\alpha}(1 + o_n(1)) \\ &\leq C n^{-1} \sigma^{-2\alpha} S_n^{-1} e^{(\pi\sigma S_n)^2} + \sigma^{-2\alpha} Q[\sigma^2 / \ln(n)]^\alpha (1 + o_n(1)) = Q[\ln(n)]^{-\alpha} (1 + o_n(1)). \end{aligned} \tag{62}$$

Theorem 2 is proved.

**Proof of Theorem 3.** Set  $\sigma := \nu j^\beta$ . For the considered indexes  $j$  we have  $\sigma^2 \leq \ln(n)\gamma_n$ . Then Theorems 1 and 2 verify the assertion of Theorem 3.

**Proof of Theorem 4.** We begin with the case when all  $g_j \in \mathcal{S}(\alpha, Q)$ . Consider  $p \leq [\ln(n)\gamma'_n]^{1/2}$  where positive  $\gamma'_n$  tends to zero as slow as desired as  $n \rightarrow \infty$ . Then the Bessel inequality and Theorem 3 yield that even for studied oracle-estimators the following lower bound holds,

$$\mathbb{E} \left\{ \int_{[0,1]^p} (g(\mathbf{x}) - \hat{g}(\mathbf{x}))^2 d\mathbf{x} \right\} \geq C[\ln(n)]^{-\alpha} \sum_{j=1}^p j^{2\alpha\beta} \geq C[\ln(n)]^{-\alpha} p^{2\alpha\beta+1}. \quad (63)$$

Here  $C$ 's are generic positive constants. Now note that

$$[\ln(n)]^{\alpha/(2\alpha\beta+1)} = o_n(1)[\ln(n)]^{1/2}. \quad (64)$$

This verifies the first part of the theorem.

Now consider the case  $g \in \mathcal{S}_p(\alpha, Q)$ . Due to the additive structure of the  $p$ -variate Sobolev class, we can convert the setting into the former one by considering  $g_j \in \mathcal{S}(\alpha, Q/p)$ . Then the assertion follows from  $p^{-1} \sum_{j=1}^p j^{2\alpha\beta} \geq Cp^{2\alpha\beta}$ . Interestingly, it is also possible to consider  $g_j = 0$  for  $j \leq p-1$  and  $g_p \in \mathcal{S}(\alpha, Q)$ , and this choice also yields the verified assertion. In short, there is a large class of least favorable functions  $g \in \mathcal{S}_p(\alpha, Q)$  that yield the verified bound on dimensionality. Theorem 4 is proved.

## 5. CONCLUSIONS

Mathematical statistical theory of nonparametric regression with measurement errors in predictors is about 30 years old, and the seminal paper of Fan and Tryong [10] created its foundation by developing the theory of rate optimal estimation. Specifically, it was shown that under the MISE criterion a traditional rate  $n^{-2\alpha/(2\alpha+1)}$ , for estimating an  $\alpha$ -fold differentiable regression function  $g(x) = \mathbb{E}\{Y|X = x\}$  based on a sample of size  $n$  from pair  $(X, Y)$ , slows down to the logarithmic  $[\ln(n)]^{-\alpha}$  when we have a sample of size  $n$  from pair  $(Z, Y)$ ,  $Z = X + \sigma\eta$  and  $\eta$  is independent of  $X$  standard normal. Accordingly, it was established that regression with measurement errors in predictors is severely ill-posed. There is a rich statistical literature devoted to this important topic, but still that rate is the only mathematical result known in the literature.

This paper solves a long-standing problem of finding a sharp constant for the MISE convergence and this result complements the known optimal rate. Namely, it is shown how the standard deviation  $\sigma$  of the measurement error and Sobolev's power of the regression function affect the sharp constant of the MISE convergence. It is of interest to note that sharp constants for the MISE convergence have been known since 1980s for the classical problems of nonparametric regression, density and spectral density estimation based on direct observations. But the known techniques are not readily applicable to regression with measurement errors in predictors. Instead, this paper uses a two hypotheses technique that previously was successful for analysis of pointwise risks and developing optimal rates.

The paper also presents an interesting example of application of the developed sharp constant for analysis of nonparametric functional additive regression. A consistent functional regression requires an increasing number of additive components, and accordingly the problem is converted into an increasing series of classical regressions with measurement errors in predictors. The important specific of the problem is that variances of measurement errors are different, and this is why the developed theory of sharp constant sheds a new light on the theory of functional regression. In particular, a new curse of dimensionality in ill-posedness is discovered and it states that no longer an additive model is the remedy for the curse of dimensionality. On a positive side, the developed mathematical theory points upon a natural order of components in an additive functional regression.

Finally, the developed mathematical methodology of sharp statistical analysis of regression with measurement errors opens an opportunity for further analysis of classical ill-posed problems where so far only optimal rates are known. In particular, it will be of interest to consider cases of dependent observations and models with missing and censored observations discussed in [8].

## FUNDING

The research was supported by NSF grant DMS-1915845.

## ACKNOWLEDGMENTS

Valuable comments of the Editor-in-Chief, Prof. Balakrishnan, Associate Editor and the reviewers are greatly appreciated.

## CONFLICT OF INTEREST

The author declares that there are no conflicts of interest.

## REFERENCES

1. G. Aneiros, R. Cao, R. Fraiman, C. Genest, and P. Vieu, "Recent advances in functional data analysis and high-dimensional statistics," *J. Multiv. Anal.* **170**, 3–9 (2019).
2. G. Aneiros, S. Novo, and P. Vieu, "Variable selection in functional regression: A review," *J. Multiv. Anal.* **188**, 1–13 (2022).
3. R. Carroll, D. Ruppert, L. Stefanski, and C. Crainiceanu, *C. Measurement Error in Nonlinear Models: A Modern Prospective* (Chapman Hall, New York, 2006).
4. R. Carroll, A. Delaigle, and P. Hall, "Nonparametric prediction in measurement error models," *JASA* **104**, 993–1014 (2009).
5. A. Cuevas, "A partial overview of the theory of statistics with functional data," *J. Statist. Plann. Infer.* **147**, 1–23 (2014).
6. A. Delaigle and A. Meister, "Nonparametric regression with heteroscedastic measurement errors-in-variables problem," *JASA* **102**, 1416–1426 (2007).
7. S. Efromovich, *Nonparametric Curve Estimation: Methods, Theory, and Applications* (Springer, New York, 1999).
8. S. Efromovich, *Missing and Modified Data in Nonparametric Estimation* (Chapman Hall, Boca Raton, 2018).
9. S. Efromovich, S. "On sharp nonparametric estimation of differentiable functions," *Statist. Probab. Letters* **152**, 9–14 (2019).
10. J. Fan, and Y. Truong, "Nonparametric regression with errors in variables," *An. Statist.* **21**, 1900–1925 (1993).
11. J. Fan, and E. Masry, "Multivariate regression estimatin with errors-in-variables: Asymptotic normality for mixing processes," *Journal of Multivariare Anal.* **43**, 237–271 (1992).
12. Y. Fan, G. James, and P. Radchenko, "Functional additive regression," *Ann. Statist.* **43**, 2296–2325 (2015).
13. M. Febrero-Bande, W. Gonzalez-Manteiga, and M. de la Fuente, Variable selection in functional additive regression models. *Comput. Statist.* **34**, 469–487 (2019).
14. F. Ferraty, and P. Vieu, *Nonparametric Functional Data Analysis Theory and Practice* (Springer, New York, 2006).
15. J. Friedman, and W. Stuetzle, "Projection pursuit regression," *JASA* **76**, 817–823 (1981).
16. G. Geenens, "Curse of dimensionality and related issues in nonparametric functional regression," *Statist. Surveys* **5**, 30–43 (2011).
17. K. Gregory, E. Mammen, and M. Wahl, "Statistical inference in sparse high-dimensional additive models," *Ann. Statist.* **49**, 1514–1536 (2021).
18. P. Hall and P. Qiu, P. "Discrete-transform approach to deconvolution problems," *Biometrika* **93**, 135–148 (2005).
19. M. Hoffmann and O. Lepski, "Random rates in anisotropic regression," *Ann. Statist.* **30**, 325–396 (2002).
20. J. Horowitz, J. Klemela, and E. Mammen, "Optimal estimation in additive regression models," *Bernoulli* **12**, 271–298 (2006).
21. Y. Ingster, "Minimax esting nonparametric statistical hypotheses about density distribution in  $L_p$  metrics," *Theory Probab. Applic.* **31**, 384–389 (1986).
22. I. Johnstone, *Gaussian Estimation: Sequence and Wavelet Models* (Stanford, 2019).
23. E. Lehmann, *Theory of Point Estimation* (Wiley, New York, 1983)
24. A. Meister, *Deconvolution Problems in Nonparametric Statistics* (Springer, New York, 2009)
25. J. Morris, "Functional regression," *Annual Rev. Statist. Appl.* **2**, 321–359 (2015).
26. H.-G. Müller and F. Yao, "Functional additive models," *JASA* **103**, 1534–1544 (2008).
27. H.-G. Müller, "Peter Hall, functional data analysis and random objects," *Ann. Statist.* **44**, 1867–1887 (2016).
28. S. M. Nikolskii, *Approximation of Functions of Several Variables and Imbedding Theorems* (Springer, Berlin, 1975).
29. S. Novo, G. Aneiros, and P. Vieu, "Sparse semiparametric regression when predictors are mixture of functional and high-dimensional variables," *Test* **30**, 481–504 (2021).

30. V. Sadhanala, and R. Tibshirani, “Additive models with trend filtering,” *Ann. Statist.* **47**, 3032–2068 (2019).
31. S. Schennach, “Nonparametric regression in the presence of measurement error,” *Econom. Theory* **20**, 1046–1093 (2004).
32. C. Stone, “Additive regression and other nonparametric models,” *Ann. Statist.* **13**, 689–705 (1985).
33. Z. Tan and C.-H. Zhang, “Doubly penalized estimation in additive regression with high-dimensional data,” *Ann. Statist.* **47**, 2567–2600 (2019).
34. A. Tsybakov, *Introduction to Nonparametric Estimation* (Springer, New York, 2009).
35. G. Yi, A. Delaigle, and P. Gustafson, (2021). *Handbook of Measurement Error Models* (CRC, Boca Raton, 2021).
36. Z. Yin, F. Liu, and Y. Xie, “Nonparametric regression estimation with mixed measurement errors,” *Appl. Mathem.* **7**, 2269–2284 (2016).
37. J. Wang, J. Chiou, and H. Müller, “Functional data analysis,” *Annual Rev. Statist. Applic.* **3**, 257–295 (2016).
38. L. Wasserman, *All About Nonparametric Statistics* (Springer, New York, 2006).