Volume 12, No. 1 (2003), pp. 62–94 Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

ADAPTIVE ESTIMATION OF ANALYTIC FUNCTIONS ON AN INTERVAL

L. M. Artiles Martínez[1] and B. Y. Levit[2]

[1]Eurandom
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

[2]Department of Mathematics and Statistics, Queen's University
Kingston, Ontario, K7L 3N6 Canada
E-mail: blevit@mast.queensu.ca

We consider the problem of adaptive point-wise estimation of an unknown regression function $f(x)$ observed in a white Gaussian noise at given n design points on a bounded interval. The function f is assumed to belong to a collection of classes $\mathcal{A}(\gamma, M)$, $\gamma, M > 0$, of functions analytic and bounded by M inside ellipses E_γ containing the observation interval.

Two particular designs, called Legendre and Chebyshev designs, are especially interesting. For each of these designs, and for any of the classes $\mathcal{A}(\gamma, M)$, we describe asymptotically minimax estimates $\hat{f}_n$ as $n \to \infty$. A slight extension of this setting in which the parameters γ, M themselves may depend on n, motivates the concepts of non-parametric (NP) and pseudo-parametric (PP) functional scales characterized by the minimax rates of convergence.

In the case of Chebyshev design we propose adaptive estimates "tuning up" point-wise to the unknown smoothness of f, when parameters γ and M are unknown. We prove these estimates to be asymptotically adaptively minimax for large collections of NP functional scales, subject to being rate efficient for any of the PP scales.

Key words: minimax estimation, adaptive estimation, analytic functions, orthogonal polynomials.

2000 Mathematics Subject Classification: Primary 62G05, 62G20; secondary 62C20.

1. Introduction

In this paper we study adaptive estimation in non-parametric regression models, for some classical discrete designs on a bounded observation interval which, without loss of generality, is assumed to be $[-1, 1]$. We assume that the unknown regression function f is analytic and bounded in a vicinity $\mathcal{V} \subset \mathbb{C}$ of $[-1, 1]$. In this case, the smoothness of f depends on the size of $\mathcal{V}$ and $\max_{\mathcal{V}} |f|$. A more concise description of such dependence – and more accurate results – can be obtained when $\mathcal{V}$ is an ellipse E_γ with boundary

$$\partial E_\gamma = \{z: \ z = \cosh\gamma\cos\phi + i\sinh\gamma\sin\phi, \ 0 \le \phi \le 2\pi\}.$$

Such an ellipse has its foci at the end-points of the interval $[-1,1]$ and the sum of semi-axes $\exp\gamma$. The family of such ellipses is convenient in describing classes of analytic functions of various degree of smoothness since $\cap E_\gamma = [-1,1]$ and $\cup E_\gamma = \mathbb{C}$.

We will denote by $\mathcal{A}(\gamma, M)$ the set of functions which are analytic and bounded in E_γ: $|f(z)| \le M, \ z \in E_\gamma$. For functions $f \in \mathcal{A}(\gamma, M)$ observed in the continuous-time Gaussian white noise on the interval $[-1,1]$, Ibragimov and Has'minskii [8] have demonstrated point-wise asymptotically minimax estimators based on Legendre polynomials.

We consider the problem of discrete regression in the model

$$y_k = f(x_k^n) + \xi_k, \quad k = 1, \dots, n, \tag{1}$$

where x_k^n are the design knots and ξ_k are i.i.d. Gaussian random variables, with zero mean and a known variance σ^2. Given the observations $\mathbf{y} = (y_1, y_2, \dots, y_n)$, the function $f \in \mathcal{A}(\gamma, M)$ can be estimated by the projection-type estimators

$$\hat{f}_{n,N}(x, \mathbf{y}) = \sum_{r=0}^{N-1} \hat{c}_r Q_r(x), \qquad \hat{c}_r = \frac{1}{n}\sum_{k=1}^{n} y_k Q_r(x_k^n),$$

where Q_r are polynomials orthonormal over the design points x_k^n. For instance, in the case of the equidistant design

$$x_k^n = \frac{2k-1}{n} - 1, \quad k = 1, \dots, n, \tag{2}$$

one could use the so-called Chebyshev discrete polynomials $C_r(x), r = 0, 1, \dots$ (cf. [1], Section 10.23, p. 223). However, we will find it technically more convenient in this case to use the family $p_r(x)$, $r = 0, 1, \dots$, of normalized Legendre polynomials, which are asymptotically equivalent to corresponding discrete Chebyshev polynomials C_r (cf. [1], Section 10.23, Eq. 7). In fact, the continuous time white noise regression model for which the Legendre polynomials have been first used in [8], can be seen as the limiting case of the discrete model (2). Thus we shall refer to the equidistant knots also as the *Legendre knots* and to the equidistant design as the *Legendre design*.

It is intuitively clear that, with the use of the equidistant design, less information about $f(x)$ is gathered from observations near the end-points of the interval than in the middle of the interval. Although it might seem that the number of "informative" observations at the end-points is only halved, in fact the accuracy of estimation near the border worsens by a factor of order $\sqrt{\log n}$, compared to the accuracy in the middle of the interval.

The situation can be improved by using another – non-uniform – design balancing the distribution of the design points in favor of the end-points. A classical design having this property is specified by the knots

$$x_k^n = \cos\frac{(2k-1)\pi}{2n}, \quad k = 1, \dots, n. \tag{3}$$

We will call these design points the *Chebyshev knots* and the corresponding design the *Chebyshev design*. Remarkably, the classical orthornormal Chebyshev polynomials $t_r(x)$, $r = 0, 1, \dots$, are also orthonormal over the Chebyshev design. We shall see that with the Chebyshev design the same rate is achievable over the whole interval $[-1, 1]$, while the mean square error at the end-points is only twice as large as in the middle of the interval.

One can ask whether this difference in the achievable rates of convergence is due to the particular method of estimation we are studying or indeed is a property of the design itself. To clarify this, we will demonstrate that in each of the corresponding designs, our estimators are asymptotically optimal among all possible estimators, at every point of the interval $[-1, 1]$. This leads us to the conclusion that the observed difference in the rates of convergence near the end-points is a direct consequence of the equidistant design, rather than of a particular method of estimation, and that the rate near the end-points substantially improves in the case of the Chebyshev design.

In Section 2 we introduce the functional classes $\mathcal{A}(\gamma, M)$ and discuss some properties of the Legendre and Chebyshev polynomials. In Section 3 asymptotically minimax estimators $\hat{f}_n$, $n \to \infty$, are described for any of the classes $\mathcal{A}(\gamma, M)$, in both cases of Legendre and Chebyshev designs. In Section 4 we introduce functional scales and, in case of the Chebyshev design, we construct asymptotically optimal adaptive estimators when the parameters γ and M are unknown. Here we restrict ourselves to the Chebyshev design due to its demonstrated advantage although the general construction can be used in other cases as well. Adaptive estimates with respect to the uniform norm can be found in the recent paper [5].

2. The Building Blocks

This section is mostly expository. First, we describe the functional scale of classes $\mathcal{A}(\gamma, M)$ of functions analytic in a vicinity of $[-1, 1]$. Then we recapitulate some of the basic properties of the classical Legendre and Chebyshev polynomials which will be used later in constructing the projection-type estimates for corresponding designs.

2.1. The classes $\mathcal{A}(\gamma, M)$. For $\gamma > 0$, let E_γ be the interior of the ellipse

$$\partial E_\gamma = \left\{z\colon\ z = \cosh\gamma\cos\phi + i\sinh\gamma\sin\phi,\ 0 \le \phi \le 2\pi\right\}$$

in the complex plane $\mathbb{C}$, with the sum of its semi-axes e^γ. The sets E_γ represent a convenient family of vicinities of the interval $[-1, 1]$, expanding from $[-1, 1]$ to $\mathbb{C}$, as γ increases from 0 to ∞. It is well known that ∂E_γ is the locus of points such that the sum of their distances from the focal points $\pm\sqrt{\cosh^2\gamma - \sinh^2\gamma} = \pm 1$ is constant and equals $2\cosh\gamma$. We denote by $\mathcal{A}(\gamma, M)$ the class of functions bounded and analytic inside E_γ: $|f(z)| \le M$ for $z \in E_\gamma$.

Let ρ_γ be the distance from the interval $[-1,1]$ to the boundary ∂E_γ. An elementary calculation shows that

$$\rho_\gamma = \cosh\gamma - 1. \tag{4}$$

According to the integral Cauchy formula, for any $\varepsilon > 0$ and any ball $B_{\rho_\gamma-\varepsilon}$ of radius $\rho_\gamma - \varepsilon$ centered at $x \in [-1,1]$,

$$f^{(m)}(x) = \frac{m!}{2\pi i}\int_{B_{\rho_\gamma-\varepsilon}} \frac{f(z)}{(z-x)^{m+1}}\,dz, \qquad m = 0,1,\dots$$

Thus, since ε is arbitrary, one obtains the following bound for the derivatives of a function $f \in \mathcal{A}(\gamma, M)$:

$$|f^{(m)}(x)| \le M m! \rho_\gamma^{-m}, \qquad x \in [-1,1]. \tag{5}$$

2.2. Legendre polynomials. Legendre polynomials form a complete system of orthogonal polynomials in the space $L^2[-1,1]$, with the inner product conveniently given by

$$\langle f \mid g\rangle := \frac{1}{2}\int_{-1}^{1} f(x)g(x)\,dx.$$

Normalized by the condition

$$P_r(1) = 1, \qquad P_r(-1) = (-1)^r, \tag{6}$$

they are explicitly given by (cf. [12], p. 68)

$$P_r(x) = 2^{-r}\sum_{\nu=0}^{r}\binom{r}{r-\nu}\binom{r}{\nu}(x-1)^\nu(x+1)^{r-\nu}. \tag{7}$$

For practical applications, the following recurrent form is handy (cf. [12], p. 71):

$$P_0 \equiv 1, \qquad P_1(x) = x,$$
$$rP_r(x) = (2r-1)xP_{r-1}(x) - (r-1)P_{r-2}(x), \qquad r \ge 2.$$

We will need some bounds on the polynomials $P_r(x)$ and their derivatives. Combining the well known in the approximation theory Markov inequality (cf. [13], Sect. 4.8.8) with the fact that the maximum of $|P_r(x)|$ is attained at the end points of the interval (cf. [12], Sect. 7.21), one obtains

$$|P_r^{(m)}(x)| \le r^{2m}, \qquad m = 1,2,\dots \tag{8}$$

Outside the interval $[-1,1]$, within the sets E_γ, one obtains the following bound using Theorem 2.9.11 in [13]:

$$|P_r(z)| \le |T_r(z)| \le e^{\gamma r}.$$

Here $T_r(x)$ are the Chebyshev polynomials to be discussed in the next section.

The orthonormal Legendre polynomials in $L^2[-1,1]$ are given by

$$p_r(x) = (2r+1)^{1/2} P_r(x), \qquad r = 0, 1, \dots \tag{9}$$

From the previous, we obtain we following bounds

$$|p_r^{(m)}(x)| \le (2r+1)^{1/2} r^{2m}, \qquad m = 1, 2, \dots, \quad x \in [-1,1], \tag{10}$$

and

$$|p_r(z)| \le (2r+1)^{1/2} e^{\gamma r}, \qquad z \in E_\gamma. \tag{11}$$

Finally, we present two results about the Legendre polynomials which later will be underlying their use in the construction of projection-type estimates. First, we show that the $L^2[-1,1]$ orthonormality of the polynomials $p_r(x)$ carries over – approximately – to the equidistant design

$$x_k^n = \frac{2k-1}{n} - 1, \quad k = 1, \dots, n. \tag{12}$$

To this end, let us denote the corresponding discrete inner product

$$(f \mid g) := \frac{1}{n} \sum_{k=1}^{n} f(x_k^n) g(x_k^n).$$

Lemma 1. *Uniformly in $r_1, r_2 \ge 0$ and $n \ge 1$,*

$$(p_{r_1} \mid p_{r_2}) = \langle p_{r_1} \mid p_{r_2} \rangle + O\left(\frac{(r_1^2 + r_2^2)^3}{n^2} \right).$$

Proof. The numerical integration method for approximating $\int_a^b f(x)\,dx$, in which the interval is divided into n equally spaced sub-intervals and the function $f \in C^2[a,b]$ is evaluated at the middle points of the sub-intervals, has the accuracy

$$\frac{(b-a)^2}{24n^2} \max_{a \le x \le b} \left| \frac{d^2}{dx^2} f(x) \right|, \tag{13}$$

cf., e.g., [11]. Thus, we have

$$\begin{aligned} \left|(p_{r_1} \mid p_{r_2}) - \langle p_{r_1} \mid p_{r_2} \rangle\right| &= \left| \frac{1}{n} \sum_{k=1}^{n} p_{r_1}(x_k^n) p_{r_2}(x_k^n) - \frac{1}{2} \int_{-1}^{1} p_{r_1}(x) p_{r_2}(x)\,dx \right| \\ &\le \frac{1}{3n^2} \max_{-1 \le x \le 1} \frac{d^2}{dx^2} \big(p_{r_1}(x) p_{r_2}(x)\big). \end{aligned} \tag{14}$$

Using the bound (10) for the derivatives of $p_r(x)$ we get

$$\left|(p_{r_1} \mid p_{r_2}) - \langle p_{r_1} \mid p_{r_2} \rangle\right| \le \frac{1}{3n^2} (2r_1+1)(2r_2+1)(r_1^2 + r_2^2)^2.$$

The second important property is manifested by the behavior of the *kernel* corresponding to the Legendre family p_r:

$$K_N(x,y) := \sum_{r=0}^{N-1} p_r(x)p_r(y).$$

Denote

$$\alpha_N^2(x) := \frac{1}{N}K_N(x,x). \tag{15}$$

Lemma 2. *Uniformly on any interval* $[a,b] \subset (-1,1)$,

$$\alpha_N^2(x) = \frac{2}{\pi\sqrt{1-x^2}}(1+o(1)), \qquad N \to \infty,$$

and $\alpha_N^2(\pm 1) = N$.

Remark 1. Note the different behavior of α_N inside the interval and at the end-points. This explains the phenomenon, first observed in [8] in the continuous regression and derived below in the discrete case, according to which, in the Legendre case, the accuracy of estimation near the end-points deteriorates to the extent of being of a different order!

Proof. Using the asymptotic formula of Laplace (cf. [12], p. 194)

$$p_r(x) \sim \frac{2}{\sqrt{\pi(1-x^2)^{1/2}}} \cos\Big((r+\tfrac{1}{2})\sqrt{1-x^2} - \frac{\pi}{4}\Big) + O(r^{-1}), \tag{16}$$

as $r \to \infty$, $|x| < 1$, and the formula (cf., e.g., [6], f. 1.341(1), p. 29)

$$\sum_{r=0}^{N-1} \sin(r\theta_1 + \theta_2) = \sin\Big(\frac{N-1}{2}\theta_1 + \theta_2\Big) \sin\frac{N\theta_1}{2} \csc\frac{\theta_1}{2}$$

we obtain, with some algebra,

$$\begin{aligned}\frac{1}{N}\sum_{r=0}^{N-1} p_r^2(x) &= \frac{2}{\pi\sqrt{1-x^2}}\Big(1 - \frac{1}{N}\sum_{r=0}^{N-1}\sin((2r+1)\theta) + O(N^{-1}\log N)\Big) \\ &= \frac{2}{\pi\sqrt{1-x^2}}(1+o(1)), \qquad N \to \infty,\end{aligned}$$

uniformly on compact sets in $(-1,1)$. At the end-points

$$\frac{1}{N}\sum_{r=0}^{N-1} p_r^2(\pm 1) = \frac{1}{N}\sum_{r=0}^{N-1}(2r+1) = N. \qquad \square$$

2.3. Chebyshev polynomials. Chebyshev polynomials appeared for the first time in [2] in the problem of finding polynomials $T_r(x) = x^r + a_1x^{r-1} + \cdots + a_r$

least deviating from zero, in the uniform norm on the interval $[-1,1]$. They are very useful in the Approximation theory and Numerical mathematics as an appropriate Fourier basis for representing non-periodic functions. Normed by $T_r(1)=1$, they can be represented as

$$T_r(x)=\cos r\arccos x,\quad r=0,1,\dots, \tag{17}$$

or in the recurrent form

$$\begin{gathered} T_0(x)=1,\qquad T_1(x)=x,\\ T_{r+1}(x)=2xT_r(x)-T_{r-1}(x),\quad r=1,2,\dots \end{gathered}$$

Consider the normalized Chebyshev polynomials

$$t_r(x)=\begin{cases} T_0(x), & r=0,\\ \sqrt{2}\,T_r(x), & r\neq 0.\end{cases}$$

These polynomials constitute an orthonormal system in the weighted L^2-space with the scalar product

$$\langle f\mid g\rangle := \frac{1}{\pi}\int_{-1}^{1}\frac{f(x)g(x)}{\sqrt{1-x^2}}\,dx, \tag{18}$$

i.e., they satisfy $\langle t_{r_1}\mid t_{r_2}\rangle=\delta_{r_1,r_2}$ for all integers $r_1,r_2\geq 0$.

Denote by

$$K_N(x,y):=\sum_{r=0}^{N-1}t_r(x)t_r(y)$$

the kernel associated with the polynomials $t_r(x)$. For a given function f, the corresponding Chebyshev–Fourier series is given by

$$\sum_{r=0}^{\infty}\langle f\mid t_r\rangle t_r(x). \tag{19}$$

This expansion becomes just the classical trigonometric series if the change of variables $x=\cos\theta$ is made. The partial sum

$$f_N(x)=\sum_{r=0}^{N-1}\langle f\mid t_r\rangle t_r(x)=\langle f\mid K_N(x,\cdot)\rangle \tag{20}$$

is the polynomial of the best approximation to $f(x)$, of degree less than N, in the weighted L^2-space. The class $\mathcal{A}(\gamma,M)$ has the important property that the coefficients of the Chebyshev–Fourier series (19) decrease exponentially (cf. [13], Sect. 3.7.3), namely, the following inequality

$$\sup_{f\in\mathcal{A}(\gamma,M)}|\langle f\mid t_r\rangle|\leq\sqrt{\pi}Me^{-\gamma r} \tag{21}$$

holds for all $r = 0, 1, \dots$ From (20), (21), and the bound $|t_r(x)| \le \sqrt{2}$ it follows that for every $f \in \mathcal{A}(\gamma, M)$

$$\max_{x\in[-1,1]} |f_N(x) - f(x)| \le \sum_{r=N}^{\infty} |\langle f \mid t_r\rangle|\, |t_r(x)| \le \frac{\sqrt{2\pi}M}{1-e^{-\gamma}} e^{-\gamma N} \tag{22}$$

(cf. [13], Sections 3.7.3 and 5.4.1).

Remarkably, for analytic functions of the classes $\mathcal{A}(\gamma, M)$, the approximation $f_N(x)$ based on the Chebyshev polynomials is asymptotically also the polynomial of the best uniform approximation on $[-1, 1]$. More precisely,

$$\sup_{f\in\mathcal{A}(\gamma,M)} \limsup_{N\to\infty} \big(\inf_{p\in Q_N} \|f - p\|_\infty\big)^{1/N} = \sup_{f\in\mathcal{A}(\gamma,M)} \limsup_{N\to\infty} \big(\|f - f_N\|_\infty\big)^{1/N},$$

where Q_N is the class of all the polynomials of the form $p = \sum_{k=0}^{N-1} a_k x^k$ (cf. [13], Section 6.5.2).

According to their definition, the Chebyshev polynomials satisfy $|t_r(x)| \le \sqrt{2}$ for all $x \in [-1, 1]$. Now we shall exhibit an interesting bound that can be obtained in the whole region E_γ. From the identity

$$2\cos rt = (\cos t + i \sin t)^r + (\cos t - i \sin t)^r$$

it follows that

$$T_r(x) = \frac{1}{2}\Big(\big(x + \sqrt{x^2-1}\big)^r + \big(x - \sqrt{x^2-1}\big)^r\Big) = \frac{1}{2}(\omega^r + \omega^{-r}),$$

where $x = \frac{1}{2}(\omega + \omega^{-1})$. Further, the transformation $z = \frac{1}{2}(\omega + \omega^{-1})$ maps the ring

$$\big\{\omega \in \mathbb{C}\colon\, e^{-\gamma} < |\omega| < e^{\gamma}\big\}$$

into E_γ and therefore $T_r(z) = \frac{1}{2}(\omega^r + \omega^{-r})$. Thus the normalized Chebyshev polynomials are bounded in E_γ by

$$|t_r(z)| = \sqrt{2}\,|T_r(z)| \le \sqrt{2}\, e^{\gamma r}. \tag{23}$$

Denote the discrete inner product by

$$(f \mid g) := \frac{1}{n}\sum_{k=1}^{n} f(x_k^n) g(x_k^n), \tag{24}$$

where the points x_k^n correspond to the Chebyshev design[1]

$$x_k^n = \cos\frac{(2k-1)\pi}{2n}, \quad k = 1, \dots, n. \tag{25}$$

[1] Some notations such as x_k^n, the projection operator K_N, etc. duplicate each other in the cases of Legendre and Chebyshev polynomials. However at any given part of the text it should be clear from the context which of the approximation schemes is used.

The first of the properties formulated in the following lemma is usually referred to as "double-orthogonality" (cf., e.g., [4], Section 2.7).

Lemma 3. *The normalized Chebyshev polynomials t_r satisfy*:
(a) *For any $r_1, r_2 = 0, 1, \ldots$*

$$(t_{r_1} \mid t_{r_2}) = \frac{1}{n} \sum_{k=1}^{n} t_{r_1}(x_k^n) t_{r_2}(x_k^n) = \delta_{r_1 \, r_2}.$$

(b) *If*

$$\beta_N^2(x) := \frac{1}{N} K_N(x,x) = \frac{1}{N} \sum_{r=0}^{N-1} t_r^2(x) \tag{26}$$

and we denote $x = \cos\theta$ then, for $N \to \infty$,

$$\beta_N^2(x) = 1 + \frac{1}{N} \frac{\cos(N\theta)\sin((N-1)\theta)}{\sin\theta} = 1 + \frac{O(1)}{N}, \tag{27}$$

uniformly on any $[a,b] \subset (-1,1)$, and $\beta_N^2(x) = 2$ for $x = \pm 1$.

Remark 2. Note a slightly different kernel behavior at the end-points; compare this with Lemma 1.

Proof. (a) The identity follows from the double orthogonality of the usual trigonometric polynomials (cf., e.g., [6], f. 1.351(1), p. 30). (b) This is a standard identity (cf., e.g., [6], f. 1.351(2), p. 31); compare with the proof of Lemma 1. □

In the following section we will discuss the use of the Legendre and Chebyshev polynomials in constructing pointwise asymptotically minimax estimators for analytic functions, in the non-adaptive setting (known γ, M).

We shall see, in particular, that the best achievable rate of convergence at the end-points using the Chebyshev design is faster than that in the case of the Legendre design. Here we have only considered and compared two most important designs: one which often appears to be the natural choice – the equidistant design, and one which is actually more preferable – the Chebyshev design. There are of course many others designs; their importance and a more comprehensive study has only started recently, partly as a result of the study presented here.

In Section 4 we shall restrict our study to Chebyshev designs, in constructing minimax estimator in the adaptive setting (unknown γ, M). Statistical estimation using the uniform norm as the quality criterion of estimators requires a different approach, [5].

3. Minimax Regression in $\mathcal{A}(\gamma, M)$

3.1. The statistical setting. Our observation model is given by

$$y_k = f(x_k^n) + \xi_k, \quad k = 1, \ldots, n,$$

where the random variables ξ_k are independent identically normally distributed $\mathcal{N}(0, \sigma^2)$, and the design x_k^n is either Legendre or Chebyshev design. Throughout this paper the unknown regression function f belongs to $\mathcal{A}(\gamma, M)$. In this section we assume that the parameters γ and M which determine the class are fixed and known to the statistician. We prove that it is possible, asymptotically, to have as good minimax risk using projection-type estimators based on the Legendre–Fourier and Chebyshev–Fourier series, for their respective designs, as with any other estimator.

Let $\mathcal{W}$ be the class of loss functions $w\colon \mathbb{R} \to \mathbb{R}^+$ such that

$$w(x) = w(-x), \qquad w(x) \geq w(y) \quad \text{for } |x| \geq |y|,\ x, y \in \mathbb{R},$$

and for some $0 < \eta < \frac{1}{2}$

$$\int e^{-\eta x^2} w(x)\, dx < \infty.$$

Let $\tilde{f}_n(x) = \tilde{f}_n(x, \mathbf{y})$ be an arbitrary estimator of $f(x)$ based on the observation vector $\mathbf{y} = (y_1, \ldots, y_n)$, and denote by $\mathbf{P}_f$, $\mathbf{E}_f$ and Var_f the distribution, the expectation and the variance corresponding to f. Sometimes the subscript f will be dropped, when there is no possibility of confusion.

Our main interest will be in the asymptotic behavior of the minimax risk

$$\inf_{\tilde{f}_n} \sup_{f \in \mathcal{A}(\gamma, M)} \mathbf{E}_f\, w\big(\sigma_n^{-1}(\tilde{f}_n(x) - f(x))\big),$$

where $w \in \mathcal{W}$. The parameter σ_n determining the minimax rate of convergence for each of the corresponding designs, Legendre or Chebyshev, will be specified later in Theorems 1 and 2.

3.2. Estimation in the Legendre design. Given the observations $\mathbf{y}$ taken at the Legendre knots (12), and following the notation introduced in Section 2.2, define the estimator

$$\hat{f}_{n,N}(x) = \frac{1}{n} \sum_{k=1}^{n} y_k K_N(x, x_k^n) = \sum_{r=0}^{N-1} \left(\frac{1}{n} \sum_{k=1}^{n} y_k p_r(x_k^n) \right) p_r(x).$$

With a slight abuse of notation, we will write

$$\hat{f}_{n,N}(x) = \big(\mathbf{y} \mid K_N(x, \cdot)\big) = \sum_{r=0}^{N-1} (\mathbf{y} \mid p_r)\, p_r(x). \tag{28}$$

Now consider two auxiliary functions:

$$f_N(x) = \langle f \mid K_N(x, \cdot) \rangle = \sum_{r=0}^{N-1} \langle f \mid p_r \rangle\, p_r(x), \tag{29}$$

and

$$f_{n,N}(x) = \big(f \mid K_N(x, \cdot)\big) = \sum_{r=0}^{N-1} (f \mid p_r)\, p_r(x). \tag{30}$$

Notice that the projection-type estimator $\hat f_{n,N}(x)$ is an unbiased estimator of the finite expansion term $f_{n,N}(x)$ which, in turn, approximates the sum f_N of the first N terms of the Legendre–Fourier series.

Theorem 1. *For any $w \in \mathcal{W}$ and every $x \in [-1,1]$*

$$\lim_{n\to\infty} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\alpha_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\big(\hat f_n(x)-f(x)\big)\right)$$
$$= \lim_{n\to\infty} \inf_{\tilde f_n} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\alpha_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\big(\tilde f_n(x)-f(x)\big)\right) = \mathbf{E}\, w(\xi),$$

where $\alpha_N(x)$ is defined in (15), *$\hat f_n = \hat f_{n,N}$ is the projection estimator* (28) *with*

$$N = N_n := \left\lfloor \frac{1}{2\gamma}\log n \right\rfloor, \tag{31}$$

$\tilde f_n$ is an arbitrary estimator of f and $\xi \sim \mathcal{N}(0,1)$.

Proof: the upper bound. Let N be given by (31). As usual, we decompose the mean square error as

$$\mathbf{E}\big(\hat f_{n,N}(x)-f(x)\big)^2 = \mathrm{Var}\; v_N^2(x) + b_N^2(x), \tag{32}$$

where, according to (28) and (30),

$$v_N(x) = \hat f_{n,N}(x) - f_{n,N}(x) = \frac{1}{n}\sum_{k=1}^{n} \xi_k K_N(x, x_k^n) \tag{33}$$

is a zero-mean stochastic term and

$$b_N(x) = \big(f_{n,N}(x) - f_N(x)\big) + \big(f_N(x) - f(x)\big) \tag{34}$$

is the bias.

Let us first analyze the variance of $v_N(x)$. Applying Lemma 1 we get

$$\begin{aligned} \mathrm{Var}\; v_N(x) &= \frac{\sigma^2}{n^2}\sum_{k=1}^{n} K_N^2(x,x_k^n) = \frac{\sigma^2}{n^2}\sum_{k=1}^{n}\left(\sum_{r=0}^{N-1} p_r(x)p_r(x_k^n)\right)^2 \\ &= \frac{\sigma^2}{n}\sum_{r_1=0}^{N-1}\sum_{r_2=0}^{N-1} p_{r_1}(x)p_{r_2}(x)\,\frac{1}{n}\sum_{k=1}^{n} p_{r_1}(x_k^n)p_{r_2}(x_k^n) \\ &= \frac{\sigma^2}{n}\sum_{r_1=0}^{N-1}\sum_{r_2=0}^{N-1} p_{r_1}(x)p_{r_2}(x)\left(\delta r_1 r_2 + O\Big(\frac{N^6}{n^3}\Big)\right) \\ &= \frac{\sigma^2}{n}\sum_{r=0}^{N-1} p_r^2(x) + O\Big(\frac{N^6}{n^3}\Big)\sum_{r_1=0}^{N-1}\sum_{r_2=0}^{N-1} p_{r_1}(x)p_{r_2}(x). \end{aligned} \tag{35}$$

Now, applying the Cauchy–Schwarz inequality we see that

$$\begin{aligned}\Big|\sum_{r_1=0}^{N-1}\sum_{r_2=0}^{N-1} p_{r_1}(x)p_{r_2}(x)\Big| &= \Big(\sum_{r=0}^{N-1} p_r(x)\Big)^2 \le N\sum_{r=0}^{N-1} p_r^2(x)\\ &= N\,K_N(x,x) = N^2\,\alpha_N^2(x).\end{aligned}$$

Thus, according to the last two equations and (31),

$$\operatorname{Var}\, v_N(x) = \alpha_N^2(x)\,\frac{\sigma^2 N}{n}\,(1+o(1)) \tag{36}$$

for any $x\in[-1,1]$, as n goes to infinity.

Now let us consider the bias. First, we have

$$f_{n,N}(x) - f_N(x) = \sum_{r=0}^{N-1}\big((f\mid p_r) - \langle f\mid p_r\rangle\big)\,p_r(x).$$

By definition

$$\big|(f\mid p_r) - \langle f\mid p_r\rangle\big| = \Big|\frac{1}{2}\int_{-1}^{1} f(x)p_r(x)\,dx - \frac{1}{n}\sum_{k=1}^{n} f(x_k^n)p_r(x_k^n)\Big|. \tag{37}$$

Next, applying (13), this difference can be bounded by

$$\frac{1}{3n^2}\max_{x\in[-1,1]}\Big|\frac{d^2}{dx^2}f(x)p_r(x)\Big|. \tag{38}$$

Thus, applying the bounds for derivatives $|f^{(m)}(x)|\le M m!/\rho_\gamma^m$ (cf. Section 2.1) and $|p_r^{(m)}(x)|\le(2r+1)^{1/2}r^{2m}$ (cf. Eq. (10)), it follows that

$$\begin{aligned}\big|(f\mid p_r) - \langle f\mid p_r\rangle\big| &\le \frac{M}{3n^2}\big((2\rho_\gamma)^{-2} + 2\rho_\gamma^{-1}(2r+1)^{1/2}r^2 + (2r+1)^{1/2}r^4\big)\\ &= O\Big(\frac{r^5}{n^2}\Big), \qquad n\to\infty.\end{aligned} \tag{39}$$

Combining Cauchy–Schwarz inequality with the previous bound and using the fact that N is of order $O(\log n)$, cf. Eq. (31), we find

$$\begin{aligned}\big(f_{n,N}(x) - f_N(x)\big)^2 &\le \sum_{r=0}^{N-1}\big((f\mid p_r) - \langle f\mid p_r\rangle\big)^2\sum_{r=0}^{N-1} p_r^2(x)\\ &= \alpha_N^2(x)\,N\sum_{r=0}^{N-1}\big((f\mid p_r) - \langle f\mid p_r\rangle\big)^2 = \alpha_N^2(x)\,O\Big(\frac{N^{12}}{n^4}\Big)\\ &= \alpha_N^2(x)\frac{\sigma^2 N}{n}\,O\Big(\frac{N^{11}}{n^3}\Big) = o(1)\ \operatorname{Var}\, v_N(x).\end{aligned} \tag{40}$$

As demonstrated in [7], for functions $f \in \mathcal{A}(\gamma, M)$

$$|\langle f \mid p_r \rangle| \le C_1\, e^{-\gamma r}$$

for some constant $C_1 > 0$. According to the Laplace formula (16) the polynomials $p_r(x)$ are uniformly bounded on any interval $[a, b] \subset (-1, 1)$. Thus, from the previous inequality, for some $C_2 > 0$,

$$\begin{aligned} \big(f_N(x) - f(x)\big)^2 &\le \Big(\sum_{r=N}^{\infty} |\langle f \mid p_r \rangle|\, |p_r(x)| \Big)^2 \\ &\le C_2\, e^{-2\gamma N} \sim C_2 n^{-1} = o(1) \text{ Var } v_N(x). \end{aligned} \tag{41}$$

At the end-points of the interval we have $|p_r(\pm 1)| = (2r+1)^{1/2}$, see Eqs. (6) and (9), thus for $x = \pm 1$

$$\begin{aligned} |f_N(x) - f(x)| &\le C_1 \sum_{r=N}^{\infty} (2r+1)^{1/2} e^{-\gamma r} \le C_3 \sum_{r=N+1}^{\infty} r^{1/2} e^{-\gamma r} \\ &\le C_3\, e^{\gamma} \int_{N+1}^{\infty} r^{1/2} e^{-\gamma r}\, dr = C_3\, N^{1/2} e^{-\gamma N} (\, 1 + o(1)\,) \end{aligned}$$

as $N \to \infty$. Therefore for some $C_4 > 0$ and N large enough

$$\big(f_N(x) - f(x)\big)^2 \le C_4 N e^{-2\gamma N} \sim C_4 \frac{N}{n} = o(1) \text{ Var } v_N(x). \tag{42}$$

From (32), (36), (40), and (41) or (42), we can conclude that

$$\mathbf{E}\big(\hat{f}_{n,N}(x) - f(x)\big)^2 = \alpha_N^2(x)\, \frac{\sigma^2 N}{n} (1 + o(1)),$$

uniformly on $[-1, 1]$. It follows that

$$\alpha_N^{-1}(x) \sqrt{\frac{n}{\sigma^2 N}}\, \big(\hat{f}_{n,N_n}(x) - f(x)\big)$$

is normally distributed with mean of order $o(1)$ and variance equal to $1 + o(1)$, when n goes to infinity, uniformly with respect to $f \in \mathcal{A}(\gamma, M)$. Therefore using the dominated convergence theorem we obtain the following upper bound:

$$\limsup_{n\to\infty} \sup_{f \in \mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\bigg(\alpha_N^{-1}(x) \sqrt{\frac{n}{\sigma^2 N}}\, \big(\hat{f}_n(x) - f(x)\big) \bigg) = \mathbf{E}\, w(\xi). \tag{43}$$

Proof of the lower bound for the risk. For fixed $x \in [-1, 1]$ and any $z \in \mathbb{C}$ consider the following parametric sub-family of functions

$$f_\theta(z) = \theta \sqrt{\frac{\sigma^2}{n}}\, \frac{K_{\bar{N}}(x, z)}{\sqrt{K_{\bar{N}}(x, x)}}, \qquad |\theta| < \theta_n = \bar{N}^{\,1/2}, \tag{44}$$

where we will use

$$\bar{N} = \bar{N}_n = \lfloor N_n - 3\log N_n \rfloor , \tag{45}$$

see (31). Note that $\bar{N}$ is asymptotically equivalent to $N = N_n$ when $N \to \infty$. According to Lemma 2 this implies that

$$\frac{\alpha^2_{\bar{N}}(x)}{\alpha^2_N(x)} \to 1, \tag{46}$$

uniformly in $[-1, 1]$, when $n \to \infty$.

We need the following lemma.

Lemma 4. *For a given $x \in [-1, 1]$ and any $z \in E_\gamma$, let $f_\theta(z)$ be defined by* (44). *Then*:

(a) $f_\theta(x) = \theta\, \alpha_{\bar{N}}(x) \sqrt{\frac{\sigma^2 \bar{N}}{n}}$.

(b) $f_\theta \in \mathcal{A}(\gamma, M)$, $|\theta| < \theta_n$, *for all n large enough.*

(c) *The statistic*

$$T = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} y_k \,\frac{K_{\bar{N}}(x, x_k^n)}{\sqrt{K_{\bar{N}}(x, x)}}$$

has a normal distribution $\mathcal{N}(\theta\mathcal{I}_n, \mathcal{I}_n)$ under f_θ, where $\mathcal{I}_n = 1 + o(1)$.

(d) *The statistic T is sufficient and the log-likelihood ratio $\Lambda := \log \frac{d\mathbf{P}_\theta}{d\mathbf{P}_0}(\mathbf{y})$ satisfies*

$$\Lambda = \theta T - \frac{\theta^2}{2}\mathcal{I}_n,$$

where $\mathbf{P}_\theta$ and $\mathbf{P}_0$ denote the probabilities associated with f_θ and f_0 respectively.

Proof. (a) This follows directly from the definitions of f_θ and $\alpha_{\bar{N}}(x)$.

(b) Obviously $f_\theta(z)$ is analytic in the whole complex plane, thus also in E_γ. Using (11), applying the Cauchy–Schwarz inequality and recalling the definition of $\bar{N} = \bar{N}_n$, we obtain

$$\begin{aligned}
|f_\theta(z)| &\le \theta_{\bar{N}} \sqrt{\frac{\sigma^2}{n}} \left(\frac{K^2_{\bar{N}}(x, z)}{K_{\bar{N}}(x, x)}\right)^{1/2} \le \sqrt{\frac{\sigma^2 \bar{N}}{n}} K^{1/2}_{\bar{N}}(z, z) \\
&= \sqrt{\frac{\sigma^2 \bar{N}}{n}} \left(\sum_{r=0}^{\bar{N}-1} p_r^2(z)\right)^{1/2} \le \sqrt{\frac{\sigma^2 \bar{N}}{n}} \left(\sum_{r=0}^{\bar{N}-1} (2r+1) e^{2\gamma r}\right)^{1/2} \\
&= O(1) \frac{\bar{N}}{\sqrt{n}} e^{\gamma \bar{N}} = O(\bar{N}^{\,-1/2}) \le M,
\end{aligned}$$

in E_γ for all n large enough.

(c) Denote

$$\mathcal{I}_n = \frac{1}{n} \sum_{k=1}^{n} \frac{K^2_{\bar{N}}(x, x_k^n)}{K_{\bar{N}}(x, x)}.$$

We can see that T is normally distributed,

$$\mathbf{E}\,T = \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n} f_\theta(x_k^n)\,\frac{K_{\bar N}(x,x_k^n)}{\sqrt{K_{\bar N}(x,x)}} = \theta\frac{1}{n}\sum_{k=1}^{n}\frac{K_{\bar N}^2(x,x_k^n)}{K_{\bar N}(x,x)} = \theta\mathcal{I}_n,$$

$$\operatorname{Var} T = \frac{1}{n}\sum_{k=1}^{n}\frac{K_{\bar N}^2(x,x_k^n)}{K_{\bar N}(x,x)} = \mathcal{I}_n.$$

Thus $T \sim \mathcal{N}(\theta\mathcal{I}_n, \mathcal{I}_n)$. Now let us show that $\mathcal{I}_n \to 1$ when $n \to \infty$. Using Lemma 1 and the Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
(47)\quad \mathcal{I}_n &= \frac{1}{n}\,K_{\bar N}^{-1}(x,x)\sum_{k=1}^{n} K_{\bar N}^2(x,x_k^n) = \frac{1}{n}\,K_{\bar N}^{-1}(x,x)\sum_{k=1}^{n}\Bigl(\sum_{r=0}^{\bar N-1} p_r(x)p_r(x_k^n)\Bigr)^2 \\
&= K_{\bar N}^{-1}(x,x)\sum_{r_1=0}^{\bar N-1}\sum_{r_2=0}^{\bar N-1}\Bigl(p_{r_1}(x)p_{r_2}(x)\,\frac{1}{n}\sum_{k=1}^{n} p_{r_1}(x_k^n)p_{r_2}(x_k^n)\Bigr) \\
&= K_{\bar N}^{-1}(x,x)\sum_{r_1=0}^{\bar N-1}\sum_{r_2=0}^{\bar N-1}\Bigl(p_{r_1}(x)p_{r_2}(x)\Bigl(\delta_{r_1 r_2} + O\Bigl(\frac{N^6}{n^2}\Bigr)\Bigr)\Bigr) \\
&= K_{\bar N}^{-1}(x,x)\sum_{r=0}^{\bar N-1} p_r^2(x) + O\Bigl(\frac{\bar N^6}{n^2}\Bigr)K_{\bar N}^{-1}(x,x)\sum_{r_1=0}^{\bar N-1}\sum_{r_2=0}^{\bar N-1} p_{r_1}(x)p_{r_2}(x) \\
&= 1 + O\Bigl(\frac{\bar N^6}{n^2}\Bigr)K_{\bar N}^{-1}(x,x)\Bigl(\sum_{r=0}^{\bar N-1} p_r(x)\Bigr)^2 = 1 + o(1), \qquad n\to\infty.
\end{aligned}
$$

(d) It is easy to see that the log-likelihood

$$
\begin{aligned}
\Lambda &= \log\prod_{k=0}^{n-1}\exp\Bigl\{-\frac{1}{2\sigma^2}\bigl(y_k - f_\theta(x_k^n)\bigr)^2 + \frac{1}{2\sigma^2}y_k^2\Bigr\} \\
&= -\frac{1}{2\sigma^2}\sum_{k=1}^{n}\bigl(y_k - f_\theta(x_k^n)\bigr)^2 + \frac{1}{2\sigma^2}\sum_{k=1}^{n} y_k^2 \\
&= \theta\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n} y_k\,\frac{K_{\bar N}(x,x_k^n)}{\sqrt{K_{\bar N}(x,x)}} - \frac{\theta^2}{2n}\sum_{k=1}^{n}\frac{K_{\bar N}^2(x,x_k^n)}{K_{\bar N}(x,x)} = \theta T - \frac{\theta^2}{2}\mathcal{I}_n.
\end{aligned}
$$

This completes the proof of the lemma. □

Now we can continue the proof of the theorem. Given $\alpha_{\bar N}^2(x) \sim \alpha_N^2(x)$, see Eq. (46),

$$
\begin{aligned}
(48)\quad \mathcal{R} &:= \inf_{\tilde f_n}\sup_{f\in\mathcal{A}(\gamma,M)}\mathbf{E}_f\, w\Bigl(\alpha_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\bigl(\tilde f_n(x) - f(x)\bigr)\Bigr) \\
(49)\quad &= \inf_{\tilde f_n}\sup_{f\in\mathcal{A}(\gamma,M)}\mathbf{E}_f\, w\Bigl(\alpha_{\bar N}^{-1}(x)\sqrt{\frac{n}{\sigma^2 \bar N}}\bigl(\tilde f_n(x) - f(x)\bigr)(1+o(1))\Bigr) \\
&\ge \inf_{\tilde f_n}\sup_{f_\theta}\mathbf{E}_{f_\theta}\, w\Bigl((1+o(1))\alpha_{\bar N}^{-1}(x)\sqrt{\frac{n}{\sigma^2 \bar N}}\bigl(\tilde f_n(x) - f_\theta(x)\bigr)\Bigr), \quad \bar N\to\infty.
\end{aligned}
$$

Denote $\tilde\theta = \alpha_{\bar N}^{-1}(x)\sqrt{n/(\sigma^2\bar N)}\tilde f_n(x)$. Then applying Lemma 4 (a)

$$\mathcal{R} \geq \inf_{\tilde\theta} \sup_{|\theta|\leq\theta_n} \mathbf{E}_\theta\, w\big((\tilde\theta-\theta)(1+o(1))\big), \qquad n\to\infty.$$

Since $|\theta|\leq\theta_n$, we can restrict ourselves exclusively to estimators such that $|\tilde\theta|\leq\theta_n$; otherwise trimming $\tilde\theta$ at an appropriate level will produce a smaller risk. For such estimators $|\tilde\theta-\theta|\leq 2\theta_n$. Now, from equations (48) and (49), applying Lemma 1 and definition (45) of $\bar N$ we can verify that the term $o(1)$ in the previous equation is of order $(\log N)/N$. Thus $\theta_n o(1)\to 0$ and therefore the previously mentioned estimators satisfy $|\tilde\theta-\theta|o(1)\to 0$. Hence

$$\mathcal{R} \geq \inf_{\tilde\theta} \sup_{|\theta|\leq\theta_n} \mathbf{E}_\theta\, w\big((\tilde\theta-\theta)+o(1)\big), \qquad n\to\infty.$$

We can approximate any loss function $w\in\mathcal{W}$, by a sequence of bounded uniformly continuous functions $w_\delta\in\mathcal{W}$ such that $w_\delta \nearrow w$ when $\delta\to 0$ and see that for any δ

$$\mathcal{R} \geq \inf_{\tilde\theta} \sup_{|\theta|\leq\theta_n} \mathbf{E}_\theta\, w_\delta\big((\tilde\theta-\theta)+o(1)\big) = \inf_{\tilde\theta} \sup_{|\theta|\leq\theta_n} \mathbf{E}_\theta\, w_\delta(\tilde\theta-\theta)+o(1).$$

Now let us fix an arbitrary prior density λ on $(-\theta_n,\theta_n)$ with a finite Fisher information $I(\lambda)$. Then

$$\begin{aligned}\inf_{\tilde\theta} \sup_{|\theta|\leq\theta_n} \mathbf{E}_\theta\, w_\delta(\tilde\theta-\theta) &\geq \inf_{\tilde\theta}\int_{-\theta_n}^{\theta_n} \mathbf{E}_\theta\, w_\delta(\tilde\theta-\theta)\lambda(\theta)\,d\theta \\ &= \inf_{\tilde\theta(T)}\int_{-\theta_n}^{\theta_n} \mathbf{E}_\theta\, w_\delta\big(\tilde\theta(T)-\theta\big)\lambda(\theta)\,d\theta\end{aligned}$$

given that T is sufficient for θ, according to Lemma 4 (c). Applying results presented in [10], we get that

$$\inf_{\tilde\theta} \sup_{|\theta|\leq\theta_n} \mathbf{E}_\theta\, w_\delta(\tilde\theta-\theta) \geq \mathbf{E}\, w_\delta(\xi)+O(\theta_n^{-2}), \qquad (n\to\infty),$$

where $\xi\sim\mathcal{N}(0,1)$. Thus $\liminf_{n\to\infty}\mathcal{R}\geq \mathbf{E}\, w_\delta(\xi)$. Applying the dominated convergence theorem for $\delta\to 0$ we get

$$\liminf_{n\to\infty}\inf_{\tilde f_n}\sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\alpha_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\big(\tilde f_n(x)-f(x)\big)\right) \geq \mathbf{E}\, w(\xi). \tag{50}$$

Finally, from (43) and (50) the theorem is proved. $\square$

Corollary 1. *For any $[a,b]\subset(-1,1)$, uniformly in $x\in[a,b]$,*

$$\begin{aligned}&\lim_{n\to\infty}\sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\sqrt{(1-x^2)^{1/2}\frac{\pi n}{\sigma^2 N_n}}\big(\hat f_n(x)-f(x)\big)\right)\\ &\quad= \lim_{n\to\infty}\inf_{\tilde f_n}\sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\sqrt{(1-x^2)^{1/2}\frac{\pi n}{\sigma^2 N_n}}\,\big(\tilde f_n(x)-f(x)\big)\right) = \mathbf{E}\, w(\xi),\end{aligned}$$

where $\hat{f}_n$ *and* $\tilde{f}_n$ *are as in Theorem* 3.1. *For* $x = \pm 1$,

$$\lim_{n\to\infty} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\sqrt{\frac{n}{\sigma^2 N_n^2}}\big(\hat{f}_n(x) - f(x)\big)\right)$$
$$= \lim_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\sqrt{\frac{n}{\sigma^2 N_n^2}}\big(\tilde{f}_n(x) - f(x)\big)\right) = \mathbf{E}\, w(\xi).$$

3.3. Estimation in the Chebyshev design. Consider now the design given by the Chebyshev knots (25). Following the notation of Section 2.3 define the estimator

$$\hat{f}_{n,N}(x) = \frac{1}{n}\sum_{k=1}^{n} y_k K_N(x, x_k^n) = \sum_{r=0}^{N-1}\left(\frac{1}{n}\sum_{k=1}^{n} y_k t_r(x_k^n)\right) t_r(x).$$

As before, we will write, with a slight abuse of notation

$$\hat{f}_{n,N}(x) = (\mathbf{y} \mid K_N(x,\cdot)) = \sum_{r=0}^{N-1} (\mathbf{y}|t_r)\, t_r(x), \tag{51}$$

and consider the two functions

$$f_N(x) = \langle f \mid K_N(x,\cdot)\rangle = \sum_{r=0}^{N-1} \langle f \mid t_r\rangle\, t_r(x),$$

and

$$f_{n,N}(x) = \big(f \mid K_N(x,\cdot)\big) = \sum_{r=0}^{N-1} (f \mid t_r)\, t_r(x);$$

see the footnote on page 69 with regards to these notations. Then the following result holds.

Theorem 2. *For any* $w \in \mathcal{W}$ *and every* $x \in [-1,1]$

$$\lim_{n\to\infty} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\beta_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\,\big(\hat{f}_n(x) - f(x)\big)\right)$$
$$= \lim_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\beta_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\,\big(\tilde{f}_n(x) - f(x)\big)\right) = \mathbf{E}\, w(\xi),$$

where $\hat{f}_n = \hat{f}_{n,N}$ *is the projection estimator* (51) *with*

$$N = N_n := \left\lfloor \frac{1}{2\gamma}\log\big(M^2\gamma\,(1-e^{-\gamma})^{-2} n\big)\right\rfloor, \tag{52}$$

$\tilde{f}_n$ *is an arbitrary estimator of* f, $\beta_N^2(x)$ *is defined by* (26) *and* $\xi \sim \mathcal{N}(0,1)$.

Remark 3. Note that $\beta_N^2(x)$ plays the same role in the present context of estimation using Chebyshev design as played by $\alpha_N^2(x)$ in the previous Legendre case.

Proof: *the upper bound.* The proof of this theorem is similar to the proof of the equivalent result for Legendre polynomials, Theorem 1. However, notice that in the case of Chebyshev polynomials we have exact orthogonality, and not just asymptotic orthogonality as for the Legendre polynomials; compare Lemmas 1 and 3 (a). This will make some computations more straightforward. Some steps in this proof will be presented somewhat differently; we will keep track of the dependency in the variance and the bias on the parameters of the class, γ and M. This will be used in the next section for adaptive estimation.

Let $N \in \mathbb{N}$. Applying the same decomposition as in Theorem 1, cf. (33) and (34), we have

$$\mathbf{E}(\hat{f}_{n,N}(x) - f(x))^2 = \mathrm{Var}\ v_N^2(x) + b_N^2(x). \tag{53}$$

Let us first analyze the variance of $v_N(x)$. As before (cf. Eq. (35)), applying Lemma 3 (a) we obtain

$$\mathrm{Var}\ v_N(x) = \frac{\sigma^2}{n} \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} t_{r_1}(x) t_{r_2}(x) \delta r_1 r_2 = \beta_N^2(x)\, \frac{\sigma^2 N}{n} \tag{54}$$

for any $x \in [-1,1]$.

Now let us consider the bias

$$b_N(x) = \big(f_{n,N}(x) - f_N(x)\big) + \big(f_N(x) - f(x)\big). \tag{55}$$

Using the Cauchy–Schwarz inequality we see that

$$\begin{aligned}\big(f_{n,N}(x) - f_N(x)\big)^2 &\le \sum_{r=0}^{N-1} \big((f \mid t_r) - \langle f \mid t_r\rangle\big)^2 \sum_{r=0}^{N-1} t_r^2(x) \\ &= N\, \beta_N^2(x) \sum_{r=0}^{N-1} \big((f \mid t_r) - \langle f \mid t_r\rangle\big)^2.\end{aligned}$$

If we rewrite the inner products as

$$(f \mid t_r) = \frac{1}{\pi} \sum_{k=1}^{n} f\Big(\cos(k - \tfrac{1}{2})\frac{\pi}{n}\Big) \cos\Big(r(k - \tfrac{1}{2})\frac{\pi}{n}\Big) \frac{\pi}{n}$$

and

$$\langle f \mid t_r\rangle = \frac{1}{\pi} \int_0^{\pi} f(\cos\zeta) \cos(r\zeta)\, d\zeta$$

(cf. Eqs. (18) and (24)), we can apply the same arguments that we used in (37)–(39). Using the bounds for the derivatives of f given in Eq. (5) we find that

$$\begin{aligned}
(56)\qquad \big|(f \mid t_r) - \langle f \mid t_r\rangle\big| &\le \frac{\pi}{24}\left(\frac{\pi}{n}\right)^2 \max_{\zeta}\left|\frac{d^2}{d\zeta^2} f(\cos\zeta)\cos(r\zeta)\right| \\
&\le \frac{\pi^3}{24\, n^2} M\left(r^2 + \frac{(2r+1)}{\rho_\gamma} + \frac{2}{\rho_\gamma^2}\right) \\
&\le \frac{\pi^3 (r+1)^2}{6\, n^2}\, M\, \max(1, \rho_\gamma^{-1}, \rho_\gamma^{-2}) = M C_\gamma \frac{(r+1)^2}{n^2}
\end{aligned}$$

where, using (4), one can verify that

$$(57)\qquad C_\gamma = O(1 - e^{-\gamma})^{-4},$$

both at $\gamma = 0$ and $\gamma = \infty$ and it is bounded when γ is varying in compact subsets of $(0, \infty)$. Thus, both for $\gamma \to 0$ and for $\gamma \to \infty$, uniformly in N

$$(58)\qquad \big(f_{n,N}(x) - f_N(x)\big)^2 = \beta_N^2(x)\, O\bigg(M^2 (1 - e^{-\gamma})^{-8}\, \frac{N^6}{n^4}\bigg).$$

If we choose $N = N_n$

$$(59)\qquad \big(f_{n,N}(x) - f_N(x)\big)^2 = o(1) \text{ Var } v_N(x), \qquad n \to \infty.$$

In the previous section we saw that in order to bound the truncation error term $f_N(x) - f(x)$ it was necessary to consider separately two cases: $|x| < 1$ and $|x| = 1$ (cf. Eqs. (41) and (42)). Now, one can see that both cases can be considered simultaneously. From (22) one can see that for any x and $N = N_n$

$$\begin{aligned}
(60)\qquad \big(f_N(x) - f(x)\big)^2 &\le 2\pi M^2 (1 - e^{-\gamma})^{-2}\, e^{-2\gamma N} = O\Big(\frac{1}{\gamma n}\Big) \\
&= \beta_N^2(x) \frac{\sigma^2 N}{n} O\Big(\frac{1}{\gamma N}\Big) = o(1) \text{ Var } v_N(x),
\end{aligned}$$

when $n \to \infty$. From (53)–(55), (59), and (60) we have proved that

$$\mathbf{E}\big(\hat f_{n,N}(x) - f(x)\big)^2 = \beta_N^2(x)\, \frac{\sigma^2 N}{n}(1 + o(1)), \qquad n \to \infty,$$

which holds uniformly on $[-1, 1]$. It follows that

$$\beta_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\big(\hat f_{n,N}(x) - f(x)\big)$$

is normally distributed with mean of order $o(1)$ and variance equal $1 + o(1)$, $n \to \infty$, uniformly with respect to $f \in \mathcal{A}(\gamma, M)$. Therefore using the dominated convergence theorem we obtain the upper bound:

$$(61)\qquad \lim_{n\to\infty} \sup_{f \in \mathcal{A}(\gamma, M)} \mathbf{E}_f\, w\bigg(\beta_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\,\big(\hat f_n(x) - f(x)\big)\bigg) = \mathbf{E}\, w(\xi).$$

Proof of the lower bound for the risk. We can follow the same proof of the lower bound as we did in Theorem 1. For fixed $x \in [-1,1]$ and any $z \in \mathbb{C}$ consider again the parametric sub-family of functions

$$f_\theta(z) = \theta \sqrt{\frac{\sigma^2}{n}}\, \frac{K_{\bar N}(x,z)}{\sqrt{K_{\bar N}(x,x)}}, \qquad |\theta| < \theta_n = \bar N^{\,1/2}, \tag{62}$$

where $K_{\bar N}$ is now defined in terms of the Chebyshev polynomials and

$$\bar N = \bar N_n = \lfloor N_n - 3 \log N_n \rfloor \tag{63}$$

(cf. definition of N_n in Eq. (52)).

Lemma 5. *The following properties are satisfied for any $x \in [-1,1]$:*

(a) $f_\theta(x) = \theta\, \beta_{\bar N}(x) \sqrt{\frac{\sigma^2 \bar N}{n}}$.

(b) $f_\theta \in \mathcal{A}(\gamma, M)$, $|\theta| < \theta_n$, *for n large enough.*

(c) *The statistic*

$$T = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} y_k\, \frac{K_{\bar N}(x, x_k^n)}{\sqrt{K_{\bar N}(x,x)}}$$

has the normal distribution $\mathcal{N}(\theta,1)$ under f_θ, i.e., it can be represented as

$$T = \theta + \xi, \tag{64}$$

where $\xi \sim \mathcal{N}(0,1)$.

(d) *The statistic T is sufficient and the log-likelihood ratio satisfies*

$$\Lambda := \log \frac{d\mathbf{P}_\theta}{d\mathbf{P}_0} = \theta\, T - \frac{\theta^2}{2}, \tag{65}$$

where $\mathbf{P}_\theta$ and $\mathbf{P}_0$ denote the probabilities associated with f_θ and f_0 respectively.

Proof of the lemma. The proof is the same as that of Lemma 4. Nevertheless, a couple of remarks can be made. First, the bound (11) for Legendre polynomials is also a bound for the Chebyshev polynomials, thus the proof of (b) remains the same. Secondly, in the present case, $\mathcal{I}_n = 1$ given exact orthogonality of Chebyshev polynomials (cf. Eq. (47)). The rest of the proofs of the lemma and the theorem remain the same and we get

$$\liminf_{n\to\infty} \inf_{\tilde f_n} \sup_{f \in \mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\beta_N^{-1}(x) \sqrt{\frac{n}{\sigma^2 N}} \left(\tilde f_n(x) - f(x)\right)\right) \geq \mathbf{E}\, w(\xi). \tag{66}$$

The theorem follows from (61) and (66). $\square$

Corollary 2. *For any $[a,b] \subset (-1,1)$ uniformly in $x \in [a,b]$*

$$\begin{aligned} &\lim_{n\to\infty} \sup_{f \in \mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\sqrt{\frac{n}{\sigma^2 N_n}} \left(\hat f_n(x) - f(x)\right)\right) \\ &\quad = \lim_{n\to\infty} \inf_{\tilde f_n} \sup_{f \in \mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\sqrt{\frac{n}{\sigma^2 N_n}} \left(\tilde f_n(x) - f(x)\right)\right) = \mathbf{E}\, w(\xi), \end{aligned}$$

where $\hat{f}_n$ and $\tilde{f}_n$ are as in the previous theorem. For $x = \pm 1$,

$$\lim_{n\to\infty} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\sqrt{\frac{n}{2\sigma^2 N_n}}\,(\hat{f}_n(x) - f(x))\right)$$
$$= \lim_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\sqrt{\frac{n}{2\sigma^2 N_n}}\,(\tilde{f}_n(x) - f(x))\right) = \mathbf{E}\, w(\xi).$$

Till now we have proved, first, that the polynomial estimators we proposed, with the order of polynomials adequately chosen, are asymptotically minimax for fixed classes $\mathcal{A}(\gamma, M)$. Secondly, we have seen that the optimal rate of convergence may be affected by the chosen design. One is tempted to think that what we have proved is a consequence of the polynomials we used but in fact the optimal rate is intrinsic to the problem – and thus the design – and not the applied estimators.

In particular, we have shown that the rate of convergence at the end-points of the interval is worse for the Legendre design as compared to the Chebyshev design. For that reason, we will restrict ourselves to the study of the regression problem on a bounded interval under the Chebyshev design. In the next subsection we will make necessary steps towards the adaptive framework.

3.4. Estimation for non-fixed classes. In order to create an adaptive framework we follow a procedure which is based on the ideas introduced in [9]. The basic underlying idea is to allow the parameters of the model – in our case γ and M – take values from the broadest possible set, pushed to its "limits." Such "limits" can be taken to be the extreme values for which either there is no consistency or, on the other hand, a parametric rate $O(n^{-1})$ is possible. Since in both cases these extreme values are not some fixed values $(\gamma^{\text{extr}}, M^{\text{extr}})$, but rather should be thought as some sequences $(\gamma_n^{\text{extr}}, M_n^{\text{extr}})$, our first step towards the adaptive framework will be to look for corresponding results in the situation where the parameters of the model, though known, are allowed to depend on n.

Thus we will assume in this subsection that although the parameters $\gamma = \gamma_n > 0$ and $M = M_n > 0$ are still known, they may depend on the number of observations n. This is not yet a proper adaptive framework. However it will allow us to explore the "limits" of the model if the parameters have more freedom. Let N_n be as it was defined in Theorem 2. The dependence of N_n on n comes also from the parameters γ, M in the present situation. Nevertheless, the statement of Theorem 2 will still hold provided the appropriate assumptions are fulfilled.

Theorem 3. *Let $w \in \mathcal{W}$, $\gamma = \gamma_n$, $M = M_n$ and let $N = N_n$ be as defined in* (52). *If the following conditions are satisfied*

$$\lim_{n\to\infty} \gamma N = \infty, \tag{67}$$

$$\lim_{n\to\infty} M^2 (1 - e^{-\gamma})^{-8} N^5\, n^{-3} = 0, \tag{68}$$

$$\lim_{n\to\infty} N = \infty, \tag{69}$$

then

$$\lim_{n\to\infty} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\beta_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\,\left(\hat{f}_n(x)-f(x)\right)\right)$$
$$= \lim_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\beta_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\,\left(\tilde{f}_n(x)-f(x)\right)\right) = \mathbf{E}\, w(\xi),$$

for all $x \in [-1,1]$*. Here* $\hat{f}_n = \hat{f}_{n,N}$ *is the projection estimator* (51) *and* $\tilde{f}_n$ *is an arbitrary estimator of* f*.*

Proof. Note that the conditions of this theorem were automatically fulfilled in the case of fixed classes. The proof in the general case is similar to the proof of Theorem 2 and consists in checking that conditions (67) and (68) guarantee asymptotic unbiasness of the optimal estimator (cf. Eqs. (58) and (60)), while (69) allows us to prove the lower bound result. The rest of the proof is the same. □

Though conditions (67)–(69) are sufficient to prove optimality results in non-fixed classes, it may be more convenient to express them explicitly in terms of the parameters γ and M, as is done in the following theorem.

Theorem 4. *Let* $w \in \mathcal{W}$ *and the parameters* $\gamma = \gamma_n$ *and* $M = M_n$ *be such that*

$$\limsup_{n\to\infty} \frac{M^2}{\log n} = 0, \tag{70}$$

$$\liminf_{n\to\infty} M^2 \log n = \infty, \tag{71}$$

$$\limsup_{n\to\infty} \frac{\gamma}{\log\log n} = 0, \tag{72}$$

$$\liminf_{n\to\infty} \gamma \log n = \infty, \tag{73}$$

then, with $N = N_n$ *defined by* (52)*,*

$$\lim_{n\to\infty} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\beta_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\,\left(\hat{f}_n(x)-f(x)\right)\right)$$
$$= \lim_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\beta_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\,\left(\tilde{f}_n(x)-f(x)\right)\right) = \mathbf{E}\, w(\xi),$$

for all $x \in [-1,1]$*. Here* $\hat{f}_n = \hat{f}_{n,N}$ *is the projection estimator* (51) *and* $\tilde{f}_n$ *is an arbitrary estimator of* f*.*

Proof. In order to prove the theorem, we only need to verify that the hypotheses of Theorem 3 are satisfied, i.e., we just need to assure that the limits (67)–(69) are still valid (cf. Eqs. (58) and (60)). If γ and M are bounded then trivially (67)–(69) hold. Let us consider the two extreme cases $\gamma \to 0$ and $\gamma \to \infty$. Remember that

$$N = N_n = \left\lfloor \frac{1}{2\gamma} \log\left(M^2\,\gamma(1-e^{-\gamma})^{-2} n\right) \right\rfloor .$$

Case $\gamma \to 0$: Applying some asymptotics and conditions (71) and (72), we see that for n large enough

$$M^2\gamma(1-e^{-\gamma})^{-2}n \sim M^2\gamma^{-1}n \geq \gamma^{-1}\log n \to \infty.$$

Thus γN and N go to infinity. Using (70) and (73)

$$\begin{aligned} M^2(1-e^{-\gamma})^{-8}N^5n^{-3} &= O\left(M^2\gamma^{-13}n^{-3}\log^5(M^2\gamma^{-1}n)\right) \\ &= O\left(n^{-3}\log^{14} n\log^5(n\log^2 n)\right) = o(1). \end{aligned}$$

Case $\gamma \to \infty$: Applying (71) and (72)

$$N \geq \frac{\log M^2 n}{2\gamma} = O\left(\frac{\log n}{\log\log n}\right) \to \infty, \qquad n \to \infty,$$

thus N and γN go to infinity. From (70) and (73)

$$\begin{aligned} M^2(1-e^{-\gamma})^{-8}N^5n^{-3} &= O\left(M^2\gamma^{-5}n^{-3}\log^5(M^2\gamma n)\right) \\ &= O\left(n^{-3}\log n\log^5(n\log n)\right) = o(1), \qquad n \to \infty. \end{aligned}$$

Thus the theorem is proved. □

4. Adaptive Minimax Regression

4.1. Adaptive estimation in functional scales. In the previous section we described asymptotically minimax estimators for classes $\mathcal{A}(\gamma, M)$ where the parameters γ and M were known. However, in practice we do not usually know to which class the unknown function belongs, in other words, we do not know the smoothness parameters. A data-dependent method for choosing an estimator in the presence of the unknown smoothness parameters is then necessary. In order to create the adaptive framework in a situation where γ and M are unknown we consider the following class of parameters. Let $v = (\gamma, M)$ where v belongs to the region $\Gamma_n \subset \mathbb{R}^2_+$. Let $\mathcal{A}(v) = \mathcal{A}(\gamma, M)$ and define the functional scale $\mathcal{A}_{\Gamma_n}$,

$$\mathcal{A}_{\Gamma_n} := \left\{\mathcal{A}(v) \mid v \in \Gamma_n\right\},$$

corresponding to the *parameter class* Γ_n. As our scales $\mathcal{A}_{\Gamma_n}$ can be identified with corresponding subsets Γ_n, we will speak sometimes about a scale Γ_n, instead of $\mathcal{A}_{\Gamma_n}$, when there is no risk it could lead to a confusion.

From now on we will restrict ourselves to the loss functions $w(x) = |x|^p$, $p > 0$. Let $\mathcal{A}_{\Gamma_n}$ be a functional scale, and $\mathcal{F}$ a class of estimators $\tilde{f}_n$, both possibly depending on n.

Definition 1. An estimator $\hat{f}_n \in \mathcal{F}$ is called $(p, \Gamma_n, \mathcal{F})$-adaptively minimax, at a point $x \in \mathbb{R}$, if for any other estimator $\tilde{f}_n \in \mathcal{F}$

$$\limsup_{n\to\infty} \sup_{v\in\Gamma_n} \frac{\sup_{f\in\mathcal{A}(v)} \mathbf{E}_f|\hat{f}_n(x) - f(x)|^p}{\sup_{f\in\mathcal{A}(v)} \mathbf{E}_f|\tilde{f}_n(x) - f(x)|^p} \leq 1.$$

This property depends crucially on which classes Γ_n and $\mathcal{F}$ are considered. The rate of convergence in estimating $f(x)$ over the whole scale $\mathcal{A}_{\mathbb{R}^2_+}$ can be of any order; it can vary from extremely fast parametric rates to extremely slow non-parametric ones, even to no consistency at all. We thus define a type of scales, so-called *regular-pseudo-parametric* scales, for which the parametric rate $n^{-1/2}$ can be achieved, consider estimators which are rate efficient on these scales, and build an adaptive minimax estimator in *regular-non-parametric* ones.

Definition 2. A functional scale $\mathcal{A}_{\Gamma_n}$ (or the corresponding scale Γ_n) is called a regular or an R scale, if the condition

$$\lim_{n\to\infty} \sup_{v\in\Gamma_n} M^2(1-e^{-\gamma})^{-8} N_n^5(v)\, n^{-3} = 0, \tag{74}$$

where $N_n(v)$ was defined in (52), is satisfied.

This condition is aimed to guarantee that the approximation arguments which were used in (58) and (59) are still applicable. Let us remark that in this condition the powers of the terms are not so relevant as far as we have $N_n(v)$ at most of order $\log n$.

We shall restrict our study to regular scales. Two special cases of regular scales are:

Definition 3. A functional scale $\mathcal{A}_{\Gamma_n}$ (a scale Γ_n) is called a regular-pseudo-parametric, or RPP functional scale (regular-pseudo-parametric, or RPP scale) if there exist finite constants M_+ and C_+ such that for all $(\gamma, M) \in \Gamma_n$ uniformly

$$\limsup_{n\to\infty} \sup_{v\in\Gamma_n} M \le M_+, \qquad \text{and} \tag{75}$$

$$\limsup_{n\to\infty} \sup_{v\in\Gamma_n} \gamma^{-1}\log n \le C_+. \tag{76}$$

Regular-pseudo-parametric scales are regular, in the sense of Definition 2, and uniformly on them, we have parametric rates, i.e., the rate $n^{-1/2}$ is achieved given

$$\limsup_{n\to\infty} \sup_{v\in\Gamma_n} N_n(v) < \infty.$$

Definition 4. A functional scale $\mathcal{A}_{\Gamma_n}$ (a scale Γ_n) is called a regular-non-parametric, or RNP functional scale (regular-non-parametric, or RPP scale) if

$$\limsup_{n\to\infty} \sup_{v\in\Gamma_n} \frac{M^2}{\log n} = 0, \tag{77}$$

$$\liminf_{n\to\infty} \inf_{v\in\Gamma_n} M^2 \log n = \infty, \tag{78}$$

$$\liminf_{n\to\infty} \sup_{v\in\Gamma_n} \frac{\gamma}{\log\log n} = 0, \tag{79}$$

$$\liminf_{n\to\infty} \inf_{v\in\Gamma_n} \gamma \log n = \infty. \tag{80}$$

Note that conditions for regular-non-parametric scales require that the assumptions of Theorem 4 hold uniformly on RNP scales. Thus, according to the proof of Theorem 4, the conditions of Theorem 3 also hold uniformly in RNP scales; in particular,

$$\liminf_{n\to\infty} \inf_{v\in\Gamma_n} N_n(v) = \infty.$$

Also note that regular-non-parametric scales are regular in the sense of Definition 2.

Let $\mathcal{F}_p = \mathcal{F}_p(x)$ be the class of all estimators $\tilde{f}_n$ that satisfy

$$\limsup_{n\to\infty} \sup_{v\in\Gamma_n} \sup_{f\in\mathcal{A}(v)} \mathbf{E}_f \left| n^{1/2}\big(\tilde{f}_n(x) - f(x)\big)\right|^p < \infty$$

for any RPP functional scale $\mathcal{A}_{\Gamma_n}$ and let $\mathcal{F}_p^0 = \mathcal{F}_p^0(x)$ be the class of all estimators such that

$$\limsup_{n\to\infty} \mathbf{E}_0 \left| n^{1/2}\, \tilde{f}_n(x)\right|^p < \infty.$$

One can see that $\mathcal{F}_p \subset \mathcal{F}_p^0$, since $f \equiv 0$ belongs to any of the classes $\mathcal{A}(\gamma, M)$. Below we present an adaptive estimator $\hat{f}_n \in \mathcal{F}_p$ and prove an upper bound on the quality of the estimator in RNP functional scales. Then we prove a lower bound with the same rate for any estimator in $\mathcal{F}_p^0$. Finally we shall conclude that our adaptive estimator is $(p, \Gamma_n, \mathcal{F}_p)$-adaptive minimax for RNP functional scales.

4.2. Upper bound on the quality of adaptive estimators.

Theorem 5. *For any $p > 0$ there exists an adaptive estimator $\hat{f}_n$ such that for any $x \in \mathbb{R}$ and for any RNP functional scale $\mathcal{A}_{\Gamma_n}$, $\hat{f}_n \in \mathcal{F}_p$ and*

$$\limsup_{n\to\infty} \sup_{v\in\Gamma_n} \sup_{f\in\mathcal{A}(v)} \mathbf{E}_f \left| \psi_n^{-1}(v)\big(\hat{f}_n(x) - f(x)\big)\right|^p \le 1.$$

Here

$$\psi_n^2(v) = p\,(\log N_n) \cdot \beta_{N_n}^2(x) \frac{\sigma^2 N_n}{n},$$

where N_n was defined in (52) *for any $v \in \Gamma_n$.*

The estimator. Let us first describe our adaptive estimator. Fix the parameters $\frac{1}{2} < l < 1$, $\frac{1}{2} < \delta < 1$, $p_1 > 0$, $l_1 = \delta l$ and consider the sequence of truncation orders $N_0 = 0$, $N_i = \lfloor \exp(i^l) \rfloor$ for $i = 1, 2, \ldots$ Two consecutive elements of this sequence satisfy

$$N_{i+1} - N_i \sim l(\log N_i)^{1-1/l} N_i \to \infty, \qquad i \to \infty, \tag{81}$$

but, at the same time, they are close enough so that they are asymptotically equivalent,

$$\frac{N_{i+1}}{N_i} \sim e^{l i^{l-1}} \sim 1, \qquad i \to \infty. \tag{82}$$

For each n we will consider the subsequence $\mathcal{S}_n = \{N_0, N_1, \ldots, N_{I_n}\}$, where

$$I_n = \arg\max_i \{N_i \le n^{1/2}\}. \tag{83}$$

Since for any δ, $0 < \delta < \frac{1}{2}$, and for n large enough, $N_n(\upsilon) \le n^{1/2-\delta}$ for all υ in any RPP scale as well as any RNP scales, one can always find $i(\upsilon) \le I_n$ such that

$$N_{i(\upsilon)-1} < N_n(\upsilon) \le N_{i(\upsilon)}. \tag{84}$$

For fixed $x \in [-1, 1]$ denote

$$\begin{aligned} \hat f_i(x) &= \hat f_{n,N_i}(x), & b_i &= \mathbf{E}_f\ \hat f_i(x) - f(x),\\ \sigma_i^2 &= \mathrm{Var}_f\ \hat f_i(x), & \hat\sigma_i^2 &= \beta_{N_i}^2(x)\frac{\sigma^2 N_i}{n},\\ \sigma_{i,j}^2 &= \mathrm{Var}_f\ \big(\hat f_j(x) - \hat f_i(x)\big), & \hat\sigma_{i,j}^2 &= \hat\sigma_j^2 - \hat\sigma_i^2, \end{aligned}$$

and define the sequence of thresholds

$$\lambda_j^2 = p \log N_j + p_1 \log^\delta N_j.$$

Adaptive procedure. Define

$$\hat\imath = \min\Big\{1 \le i \le I_n :\ \big|\hat f_j(x) - \hat f_i(x)\big| \le \lambda_j\, \hat\sigma_{i,j}\ \ \forall j\ (i \le j \le I_n)\Big\}.$$

We will prove that the estimator

$$\hat f_n(x) = \hat f_{\hat\imath}(x)$$

satisfies Theorem 5. First, however, we derive some inequalities which are necessary for the proof.

Lemma 6. *Using the previous notation, uniformly with respect to υ in any RPP or RNP scale, and uniformly with respect to $1 \le i, j \le I_n$, as $n \to \infty$,*

(a) $b_j^2 = o(1)\, \hat\sigma_j^2$ *for all j such that $i(\upsilon) \le j \le I_n$;*

(b) $\sigma_j^2 = \hat\sigma_j^2$ *for all j;*

(c) $(b_j - b_i)^2 = O(1)\, \hat\sigma_{i,j}^2$ *for all i, j such that $i(\upsilon) \le i \le j \le I_n$;*

(d) $\sigma_{i,j}^2 = \hat\sigma_{i,j}^2$ *for all i, j.*

Proof. (a) As we saw before

$$b_j^2 \le 2\big(f_{n,N_j}(x) - f_{N_j}(x)\big)^2 + 2\big(f_{N_j}(x) - f(x)\big)^2.$$

From equations (58), (83), and conditions for RPP scales, as well as conditions for RNP scales (cf. Definitions 3 and 4), we have

$$\begin{aligned} \big(f_{n,N_j}(x) - f_{N_j}(x)\big)^2 &\le \beta_{N_j}^2(x)\frac{\sigma^2 N_j}{n}\, O\left(M^2(1-e^{-\gamma})^{-8} N_j^5 n^{-3}\right)\\ &\le \beta_{N_j}^2(x)\frac{\sigma^2 N_j}{n}\, O\left(M^2(1-e^{-\gamma})^{-8} n^{-1/2}\right) = o(1)\, \hat\sigma_j^2. \end{aligned}$$

From (60),

$$\begin{aligned}\big(f_{N_j}(x)-f(x)\big)^2 &\le 2\pi M^2(1-e^{-\gamma})^{-2}\,e^{-2\gamma N_j} \le 2\pi M^2(1-e^{-\gamma})^{-2}\,e^{-\gamma N_n}\\ &= O\Big(\frac{1}{\gamma n}\Big) = O\Big(\frac{1}{\gamma N_j}\Big)\,\hat\sigma_j^2.\end{aligned}$$

In RPP scales γ goes to infinity uniformly, thus γN_j goes to infinity uniformly for all $N_j \ge N_1$. In RNP scales $\gamma N_j \ge \gamma N_n \to \infty$, thus

$$\big(f_{N_j}(x)-f(x)\big)^2 = o(1)\,\hat\sigma_j^2,$$

as $n\to\infty$. From previous equations we have that $b_j^2 = o(1)\,\hat\sigma_j^2$ for all $j \ge i(\upsilon)$, uniformly in RPP as well as RNP functional scales.

(b) From (54), taking $N = N_j$, we obtain

$$\sigma_j^2 = \mathrm{Var}\ \hat f_j(x) = \beta_{N_j}^2(x)\,\frac{\sigma^2 N_j}{n} = \hat\sigma_j^2.$$

(c) We have

$$\begin{aligned}\big(b_j-b_i\big)^2 &= \big(f_{n,N_j}(x)-f_{n,N_i}(x)\big)^2\\ &\le 2\big((f_{n,N_j}(x)-f_{N_j}(x))-(f_{n,N_i}(x)-f_{N_i}(x))\big)^2\\ &\quad + 2\big(f_{N_j}(x) - f_{N_i}(x)\big)^2 := 2\,b_1^2(x)+2\,b_2^2(x).\end{aligned}$$

Now,

$$b_1 = \big(f_{n,N_j}(x)-f_{N_j}(x)\big)-\big(f_{n,N_i}(x)-f_{N_i}(x)\big) = \sum_{r=N_i}^{N_j-1}\big((f\mid t_r)-\langle f\mid t_r\rangle\big)\,t_r(x).$$

Applying the Cauchy–Schwarz inequality, (56), and (57) we see that for regular scales, as we did in (a),

$$\begin{aligned}b_1^2 &= O\Big(M^2(1-e^{-\gamma})^{-8}N_j^5 n^{-4}\Big)\Big(\sum_{r=0}^{N_j-1} t_r^2(x) - \sum_{r=0}^{N_i-1} t_r^2(x)\Big)\\ &= O\Big(M^2(1-e^{-\gamma})^{-8}n^{-1/2}\Big)\,\Big(\beta_{N_j}^2(x)\frac{\sigma^2 N_j}{n} - \beta_{N_i}^2(x)\frac{\sigma^2 N_i}{n}\Big)\\ &= o(1)\,(\hat\sigma_j^2-\hat\sigma_i^2), \qquad n\to\infty.\end{aligned}$$

Also, applying the Cauchy–Schwarz inequality,

$$b_2^2 \le \Big(\sum_{r=N_i}^{N_j-1}|\langle f\mid t_r\rangle|\,|t_r(x)|\Big)^2 \le \sum_{r=N_n}^{\infty}|\langle f\mid t_r\rangle|^2\sum_{r=N_i}^{N_j-1}t_r^2(x),$$

where using (21), the definition (52) of N_n, and condition (80) one can verify that

$$\sum_{r=N_n}^{\infty} |\langle f \mid t_r\rangle|^2 = O\left(M^2 \frac{e^{-2\gamma N_n}}{1-e^{-2\gamma}}\right) = O\left(\frac{(1-e^{-\gamma})^2}{\gamma(1-e^{-2\gamma})}\right)\frac{1}{n} = O(1)\,\frac{1}{n}.$$

Now,

$$\begin{aligned} b_2^2 &= O(1)\,\frac{1}{n}\left(\sum_{r=0}^{N_j-1} t_r^2(x) - \sum_{r=0}^{N_i-1} t_r^2(x)\right) = O(1)\left(\beta_{N_j}^2(x)\frac{\sigma^2 N_j}{n} - \beta_{N_i}^2(x)\frac{\sigma^2 N_i}{n}\right) \\ &= O(1)\,(\hat\sigma_j^2 - \hat\sigma_i^2), \qquad n\to\infty. \end{aligned}$$

Thus $\big(b_j - b_i\big)^2 = O(1)\big(\hat\sigma_j^2 - \hat\sigma_i^2\big)$ for any $x\in[-1,1]$ when $n\to\infty$.

(d) Applying again the Cauchy–Schwarz inequality together with Lemma 3 (a) we see that

$$\begin{aligned} \mathrm{Var}\,\big(\hat f_j(x) - \hat f_i(x)\big) &= \frac{\sigma^2}{n^2}\sum_{k=1}^{n}\big(K_{N_j}(x,x_k^n) - K_{N_i}(x,x_k^n)\big)^2 \\ &= \frac{\sigma^2}{n}\sum_{r_1=N_i}^{N_j-1}\sum_{r_2=N_i}^{N_j-1}\Big(t_{r_1}(x)t_{r_2}(x)\,\frac{1}{n}\sum_{k=1}^{n} t_{r_1}(x_k^n)t_{r_2}(x_k^n)\Big) \\ &= \frac{\sigma^2}{n}\sum_{r_1=N_i}^{N_j-1}\sum_{r_2=N_i}^{N_j-1} t_{r_1}(x)t_{r_2}(x)\delta r_1 r_2 = \frac{\sigma^2}{n}\sum_{r=N_i}^{N_j-1} t_r^2(x) \\ &= \hat\sigma_j^2 - \hat\sigma_i^2. \qquad \square \end{aligned}$$

Proof of the theorem. For an arbitrary scale of parameters Γ_n and for any $f\in\mathcal{A}(\upsilon)$ for some $\upsilon\in\Gamma_n$,

$$\begin{aligned} R_n(f) = \mathbf{E}\,\big|\hat f_{\hat\imath}(x) - f(x)\big|^p &= \mathbf{E}\left\{\mathbf{1}_{\{\hat\imath\le i(\upsilon)\}}\big|\hat f_{\hat\imath}(x) - f(x)\big|^p\right\} \\ &\quad + \mathbf{E}\left\{\mathbf{1}_{\{\hat\imath> i(\upsilon)\}}\big|\hat f_{\hat\imath}(x) - f(x)\big|^p\right\} := R_n^-(f) + R_n^+(f). \end{aligned}$$

Let us examine $R_n^-(f)$ first. We have that

$$\big\{\hat\imath\le i(\upsilon)\big\}\subset\big\{\,\big|\hat f_{\hat\imath}(x) - \hat f_{i(\upsilon)}(x)\big|\le\hat\sigma_{\hat\imath,i(\upsilon)}\,\lambda_{i(\upsilon)}\,\big\}\subset\big\{\,\big|\hat f_{\hat\imath}(x) - \hat f_{i(\upsilon)}(x)\big|\le\hat\sigma_{i(\upsilon)}\,\lambda_{i(\upsilon)}\,\big\},$$

given the definition of $\hat\imath$ and the property $\hat\sigma_{i,j}^2 = \hat\sigma_j^2 - \hat\sigma_i^2$. Therefore

$$\begin{aligned} R_n^-(f) &\le \mathbf{E}\left\{\mathbf{1}_{\{\hat\imath\le i(\upsilon)\}}\Big(\big|\hat f_{\hat\imath}(x) - \hat f_{i(\upsilon)}(x)\big| + \big|\hat f_{i(\upsilon)}(x) - f(x)\big|\Big)^p\right\} \\ &\le \mathbf{E}\left(\hat\sigma_{i(\upsilon)}\lambda_{i(\upsilon)} + \big|\hat f_{i(\upsilon)}(x) - f(x)\big|\right)^p \\ &\le \mathbf{E}\left(\hat\sigma_{i(\upsilon)}\lambda_{i(\upsilon)} + |b_{i(\upsilon)}| + \sigma_{i(\upsilon)}|\xi|\right)^p, \end{aligned} \tag{85}$$

where $\xi\sim\mathcal{N}(0,1)$.

In RPP scales, the family of $N_n(v)$, the optimum bandwidths, is uniformly bounded with respect to v. Thus, the families of $N_{i(v)}$ and $\lambda_{i(v)}$ are also uniformly bounded in Γ_n, and we can see that the variance satisfies

$$\sigma^2_{i(v)} = \frac{\sigma^2}{n} \sum_{r=0}^{N_{i(v)}-1} t_r^2(x) \leq 2\frac{\sigma^2 N_{i(v)}}{n} = O(n^{-1}),$$

uniformly in such scales, when $n \to \infty$. From Lemma 6 we know that $b^2_{i(v)} = o(1)\,\hat{\sigma}^2_{i(v)}$, thus $b^2_{i(v)} = o(n^{-1})$. Using the above in (85) we have that for any RPP scale, uniformly,

$$\sup_{f \in \mathcal{A}(v)} R_n^-(f) = O(n^{-p/2}), \qquad n \to \infty. \tag{86}$$

From (85), applying Lemma 6, the dominated convergence theorem and relation (82), uniformly in any RNP scale

$$\sup_{f \in \mathcal{A}(v)} R_n^-(f) \leq \psi_n^p(v)\big(1 + o(1)\big), \qquad n \to \infty. \tag{87}$$

Now let us examine $R_n^+(f)$. Consider the auxiliary event

$$A_i = \Big\{\omega : \big|\hat{f}_i(x) - f(x)\big| \leq \sqrt{2}\,\hat{\sigma}_i\,\lambda_i\Big\}.$$

Applying the Hölder and Cauchy–Schwarz inequalities we obtain

$$\begin{aligned}
R_n^+(f) &= \mathbf{E}\Big\{\mathbf{1}_{\{\hat{\imath}>i(v)\}}\,\big|\hat{f}_{\hat{\imath}}(x) - f(x)\big|^p\Big\} = \sum_{i=i(v)+1}^{I_n} \mathbf{E}\Big\{\mathbf{1}_{\{\hat{\imath}=i\}}\,\big|\hat{f}_i(x) - f(x)\big|^p\Big\} \\
&= \sum_{i=i(v)+1}^{I_n} \mathbf{E}\Big\{\big|\hat{f}_i(x) - f(x)\big|^p\,\big(\mathbf{1}_{\{\hat{\imath}=i\}\cap A_i} + \mathbf{1}_{\{\hat{\imath}=i\}\cap A_i^c}\big)\Big\} \\
&\leq \sum_{i=i(v)+1}^{I_n} \mathbf{E}\Big\{\big|\hat{f}_i(x) - f(x)\big|^p\,\mathbf{1}_{\{\hat{\imath}=i\}\cap A_i}\Big\} + \sum_{i=i(v)+1}^{I_n} \mathbf{E}\Big\{\big|\hat{f}_i(x) - f(x)\big|^p\,\mathbf{1}_{A_i^c}\Big\} \\
&\leq R_{n,1}^+(f) + R_{n,2}^+(f),
\end{aligned}$$

where

$$R_{n,1}^+(f) = \sum_{i=i(v)+1}^{I_n} (2\hat{\sigma}_i^2\lambda_i^2)^{p/2}\,\mathbf{P}(\hat{\imath} = i)$$

and

$$R_{n,2}^+(f) = \sum_{i=i(v)+1}^{I_n} \mathbf{E}^{1/2}\,\big|\hat{f}_i(x) - f(x)\big|^{2p}\,\mathbf{P}^{1/2}(A_i^c).$$

We have that

$$\mathbf{P}(\hat{\imath} = i) \leq \mathbf{P}(\hat{\imath} \geq i) \leq \sum_{j=i+1}^{\infty} \mathbf{P}\Big(\big|\hat{f}_{j-1}(x) - \hat{f}_{i-1}(x)\big| > \hat{\sigma}_{i-1,j-1}\,\lambda_{j-1}\Big), \tag{88}$$

but $\hat f_j(x) - \hat f_i(x) = \sigma_{i,j}\xi + b_j - b_i$, where $\xi \sim \mathcal{N}(0,1)$. Therefore applying Lemma 6, (c) and (d), and the well-known bound for the tails of the normal distribution (cf. [3], Lemma 2) we find that

$$\mathbf{P}\big(|\hat f_j(x) - \hat f_i(x)| > \hat\sigma_{i,j}\lambda_j\big) \le \mathbf{P}\Big(|\xi| > \lambda_j - \frac{|b_j - b_i|}{\hat\sigma_{i,j}}\Big) \le \exp\Big\{-\frac{1}{2}(\lambda_j - C_1)^2\Big\}$$
$$\le \exp\Big\{-\frac{1}{2}\lambda_j^2 + C_1\lambda_j\Big\},$$

for some $C_1 > 0$ and n large enough. Returning to (88) we obtain that

$$\begin{aligned}\mathbf{P}(\hat\imath = i) &\le \sum_{j=i+1}^{\infty} \exp\Big\{-\frac{1}{2}\lambda_{j-1}^2 + C_1\lambda_{j-1}\Big\} = \sum_{j=i}^{\infty}\exp\Big\{-\frac{1}{2}\lambda_j^2 + C_1\lambda_j\Big\}\\ &= \sum_{j=i}^{\infty}\exp\Big\{-\frac{pj^l + p_1 j^{l_1}}{2} + C_1\sqrt{pj^l + p_1 j^{l_1}}\Big\}\\ &\le \sum_{j=i}^{\infty}\exp\Big\{-\frac{pj^l}{2} - \frac{p_1 j^{l_1}}{3}\Big\} \sim \frac{2}{pl} i^{1-l}\exp\Big\{-\frac{pi^l}{2} - \frac{p_1 i^{l_1}}{3}\Big\}\\ &= \frac{2}{pl} i^{1-l} N_i^{-p/2}\exp\Big\{-\frac{p_1 i^{l_1}}{3}\Big\} \le C_2 N_i^{-p/2}\exp\Big\{-\frac{p_1 i^{l_1}}{4}\Big\}\end{aligned}$$

for some $C_2 > 0$ and all $i \ge i(v)$, when n is sufficiently large. Therefore uniformly in Γ_n

$$\sup_{f\in\mathcal{A}(v)} R_1^+(f) = O(n^{-p/2})\sum_{i=1}^{\infty} i^{pl/2}\exp\{-p_1 i^{l_1}/4\} = O(n^{-p/2}), \tag{89}$$

when $n \to \infty$. In order to bound $R_2^+(f)$ note that $\hat f_i - f(x) = b_i + \sigma_i\xi$, $\xi \sim \mathcal{N}(0,1)$. Then applying Lemma 6, (a) and (b), in the same way as before, we have

$$\begin{aligned}\mathbf{P}(A_i^c) &\le \mathbf{P}\Big(|\xi| > \sqrt{2}\,\lambda_i - \frac{|b_i|}{\sigma_i}\Big) \le \mathbf{P}\left(|\xi| > \sqrt{2}\,\lambda_i - \sqrt{2}\right)\\ &\le \exp\Big\{-\frac{1}{2}\left(\sqrt{2}\,\lambda_i - \sqrt{2}\right)^2\Big\} \le \exp\{-\lambda_i^2 + 2\,\lambda_i\}\\ &\le \exp\{-pi^l - p_1 i^{l_1}/2\} \sim N_i^{-p}\exp\{-p_1 i^{l_1}/2\},\end{aligned}$$

for all $i \ge i(v)$, n large enough. Thus, applying again Lemma 6, (a) and (b), and the previous bound, we obtain

$$\begin{aligned}R_{n,2}^+(f) &= \sum_{i=i(v)+1}^{I_n}\mathbf{E}^{1/2}\big|\hat f_i(x) - f(x)\big|^{2p}\,\mathbf{P}^{1/2}(A_i^c)\\ &\le \sum_{i=i(v)+1}^{I_n}\hat\sigma_i^p\,\mathbf{E}^{1/2}\big|o(1) + \xi\big|^{2p}\mathbf{P}^{1/2}(A_i^c) = O\Big(\beta_{N_i}^2\frac{\sigma^2}{n}\Big)^{p/2}\sum_{i=1}^{\infty}\exp\{-p_1 i^{r_1}/4\}\end{aligned}$$

and finally

$$\sup_{f\in\mathcal{A}(v)} R^{+}_{n,2}(f) = O\big(n^{-p/2}\big). \tag{90}$$

Finally we can conclude from (86), (87), (89), and (90) that $\hat{f}_n \in \mathcal{F}_p(x)$ and

$$\limsup_{n\to\infty} \sup_{v\in\Gamma_n} \sup_{f\in\mathcal{A}(v)} \mathbf{E}\left|\psi_n^{-1}(v)\big(f_n(x)-f(x)\big)\right|^p \le 1,$$

in RNP scales, thus ending the proof of the theorem. □

4.3. Lower bound. In the previous section we have established an upper bound for the risk of adaptive procedures, by evaluating the quality of the proposed adaptive estimator. In this section we will establish a lower bound for arbitrary such estimator, which will allow us to establish optimality of the proposed procedure in the sense of Definition 1.

Theorem 6. *Let $p>0$. Let $\mathcal{A}_{\Gamma_n}$ be an arbitrary RNP functional scale. For each $v\in\Gamma_n$, define*

$$\psi_n(v) = \sigma_n(v)\phi_n(v),$$

where

$$\sigma_n^2(v) = \beta^2_{N_n}(x)\frac{\sigma^2 N_n}{n}, \qquad \phi_n^2(v) = p\log N_n,$$

and N_n is the same as in Theorem 5. Then, for any estimator $\tilde{f}_n \in \mathcal{F}_p^0(x)$

$$\liminf_{n\to 0} \inf_{v\in\Gamma_n} \sup_{f\in\mathcal{A}(v)} \mathbf{E}\left|\psi_n^{-1}(v)\big(\tilde{f}_n(x)-f(x)\big)\right|^p \ge 1.$$

Proof. Denote for brevity $\psi_v=\psi_n(v)$, $\phi_v=\phi_n(v)$ and $\sigma_v=\sigma_n(v)$. Choose $\bar{N}$ as it was defined in (63), and define $\bar{\psi}_v=\bar{\sigma}_v\bar{\phi}_v$ where

$$\bar{\sigma}_v^2 = \beta^2_{\bar{N}}(x)\frac{\sigma^2\bar{N}}{n} \quad\text{and}\quad \bar{\phi}_v^2 = p\log\bar{N}.$$

Define $f_0\equiv 0$ and $f_1=f_\theta$ for $\theta=\bar{\phi}_v-\bar{\phi}_v^{1/2}$, where f_θ belongs to the parametric family defined in (62). Notice that $|\theta|<\bar{N}^{1/2}$ for all n large enough. According to Lemma 5, $f_1\in\mathcal{A}(v)$ and

$$f_1(x) = \theta\,\beta_{\bar{N}}(x)\sqrt{\frac{\sigma^2\bar{N}}{n}}.$$

For an arbitrary estimator $\tilde{f}_n\in\mathcal{F}_p^0(x)$ denote $f_n^*=\bar{\psi}_v^{-1}\tilde{f}_n(x)$ and $L=\bar{\phi}_v^{-1}\theta$. Then

$$\bar{\psi}_v^{-1}\big(\tilde{f}_n(x)-f_1(x)\big) = f_n^* - \bar{\psi}_v^{-1}f_1(x) = f_n^* - \bar{\phi}_v^{-1}\theta = f_n^* - L, \tag{91}$$

whereas

$$\frac{\sqrt{n}}{\sigma}\big(\tilde{f}_n(x)-f_0(x)\big)=\frac{\sqrt{n}}{\sigma}\,\bar{\psi}_v\,f_n^*(x)=\sqrt{\bar{N}}\,\bar{\phi}_v\,f_n^*(x)$$
$$=f_n^*\exp\Big\{\frac{\log\bar{N}}{2}+\log\bar{\phi}_v\Big\}. \tag{92}$$

Denote $\mathbf{P}_0$ and $\mathbf{P}_1$ the probabilities associated with f_0 and f_1 respectively. From equations (64) and (65),

$$\frac{d\mathbf{P}_0}{d\mathbf{P}_1}(y)=\exp\Big\{-\frac{\theta^2}{2}-\theta\xi\Big\} \tag{93}$$

with respect to $\mathbf{P}_1$, where $\xi\overset{\mathbf{P}_1}{\sim}\mathcal{N}(0,1)$. Denote $q=\exp\{-\bar{\phi}_v\}$, so that $q\to 0$ since $\bar{N}\to\infty$ ($n\to\infty$) in NP scales. Now, given $f_1\in\mathcal{A}(v)$, for any $\tilde{f}_n\in\mathcal{F}_p^0(x)$, uniformly in $v\in\Gamma_n$ as n goes to infinity, we have

$$\overline{\mathcal{R}}:=\sup_{f\in\mathcal{A}(v)}\mathbf{E}^{(n)}\Big|\,\bar{\psi}_v^{-1}\big(\,\tilde{f}_n(x)-f(x)\big)\Big|^p\geq\mathbf{E}_1\Big|\,\bar{\psi}_v^{-1}\big(\,\tilde{f}_n(x)-f_1(x)\big)\Big|^p \tag{94}$$
$$\geq q\,\mathbf{E}_0\Big|\,\frac{\sqrt{n}}{\sigma}\big(\,\tilde{f}_n(x)-f_0(x)\big)\Big|^p$$
$$+(1-q)\,\mathbf{E}_1\Big|\,\bar{\psi}_v^{-1}\big(\,\tilde{f}_n(x)-f_1(x)\big)\Big|^p+O(q).$$

According to (91)–(94),

$$\overline{\mathcal{R}}\geq q\exp\Big\{\frac{\bar{\phi}_v}{2}+p\log\bar{\phi}_v\Big\}\mathbf{E}_0\left|f_n^*(x)\right|^p+(1-q)\mathbf{E}_1\left|f_n^*(x)-L\right|^p+O(q) \tag{95}$$
$$\geq(1-q)\,\mathbf{E}_1\left(Z|f_n^*(x)|^p+|f_n^*(x)-L|^p\right)+O(q)$$
$$\geq(1-q)\,\mathbf{E}_1\,\inf_x\left(Z\,|x|^p+|x-L|^p\right)+O(q),$$

where

$$Z=q\,\exp\Big\{\frac{\bar{\phi}_v}{2}+p\log\bar{\phi}_v\Big\}\frac{d\mathbf{P}_0}{d\mathbf{P}_1}.$$

From (93) and definition of θ we have

$$Z=\exp\Big\{-\bar{\phi}_v+\frac{\bar{\phi}_v^2}{2}+p\log\bar{\phi}_v-(\bar{\phi}_v-\bar{\phi}_v^{1/2})\xi-\frac{1}{2}(\bar{\phi}_v-\bar{\phi}_v^{1/2})^2\Big\}\overset{\mathbf{P}_1}{\to}\infty$$

given $\bar{\phi}_v\to\infty$. Now consider the same optimization problem as before:

$$\min_x\left\{g(x):=Z|x|^p+|L-x|^p\right\}.$$

We saw in the previous section that

$$g(x_{\min})=\chi\,L^p, \tag{96}$$

where $\chi \xrightarrow{\mathbf{P}_1} 1$. Therefore according to equations (95) and (96), uniformly in $v \in \Gamma_n$,

$$\overline{\mathcal{R}} \geq (1-q)L^p \mathbf{E}_1 \chi + O(q) = 1 + o(1).$$

Finally, uniformly in Γ_n

$$\sup_{f \in \mathcal{A}(v)} \mathbf{E}^{(n)} \left| \psi_v^{-1} \big(\tilde{f}_n(x) - f(x) \big) \right|^p = \sup_{f \in \mathcal{A}(v)} \mathbf{E}^{(n)} \left| \bar{\psi}_v^{-1} \big(\tilde{f}_n(x) - f(x) \big) \right|^p (1 + o(1)) \\ \geq 1 + o(1).$$

This completes the proof of the theorem. □

Corollary 3. *Let $\mathcal{A}_{\Gamma_n}$ be an arbitrary RNP scale. Then for any $p > 0$ and $x \in \mathbb{R}$, the estimator $\hat{f}_n$ of Theorem 5 is $(p, \Gamma_n, \mathcal{F}_p(x))$-adaptively minimax at x.*

Proof. This is a consequence of Theorems 5 and 6.

References

[1] H. Bateman, *Higher Transcendental Functions*, Vol. II, McGraw-Hill, 1953.

[2] P. L. Chebyshev, *Problems on the least values connected with the approximate representation of functions.* In: *Sochineniya*, Vol. II, 1859, pp. 151–235. (In Russian.)

[3] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd ed., Wiley, New York, 1968.

[4] L. Fox, and I. B. Parker, *Chebyshev Polynomials in Numerical Analysis*, Oxford Univ. Press, London, 1968.

[5] G. K. Golubev, O. V. Lepski, and B. Y. Levit, *On adaptive estimation for the sup-norm losses*, Math. Methods Statist., 10 (2001), 23–37.

[6] I. S. Gradshtein and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York, 1965.

[7] I. A. Ibragimov and R. I. Has'minskii, *Statistical Estimation: Asymptotic Theory*, Springer, New York, 1981.

[8] I. A. Ibragimov and R. I. Has'minskii, *On nonparametric estimation of the value of a functional in Gaussian white noise*, Theory Probab. Appl., 29 (1984), 18–32.

[9] O. V. Lepski and B. Y. Levit, *Adaptive minimax estimation of infinitely differentiable functions*, Math. Methods Statist., 7 (1998), 123–156.

[10] B. Y. Levit, *On the asymptotic minimax estimates of the second order*, Theory Probab. Appl., 25 (1980), 552–568.

[11] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 2nd ed., Springer, 1992.

[12] G. Szegö, *Orthogonal Polynomials*, Vol. XXIII, 4th ed., American Math. Soc., 1975.

[13] A. F. Timan, *Theory of Approximation of Functions of a Real Variable*, Pergamon Press, 1963.

[*Received September* 2002]