Adaptive minimax regression on the interval

L.M. Artiles Martínez [†]
B.Y. Levit [‡]

May 2, 2002

Abstract:
We consider the problem of adaptive point-wise estimation of an unknown regression function $f(x)$ whose observations at the points of a fixed design on a given interval $[a, b]$ are corrupted by a white Gaussian noise. The function f is assumed to belong to a the class $\mathcal{A}(\gamma, M)$ of analytic functions bounded by a constant M inside the ellipse having its foci at the end-points of the interval $[a, b]$ and a sum $(b - a) \exp(\gamma)$ of the semi-axes.
First, for two different designs - Legendre and Chebyshev design - we describe asymptotically minimax estimators for any of the fixed classes $\mathcal{A}(\gamma, M)$, as the number of observations n increases. A slight extension of this setting, with both γ, M allowed to depend on n, brings in the concept of non-parametric (NP) and pseudo-parametric (PP) functional scales characterized by the corresponding rates of convergence.
Finally, with γ and M unknown, we propose adaptive estimators 'tuning up' to the unknown smoothness of f. We prove them to be asymptotically adaptively minimax for large collections of NP functional scales, subject to being rate efficient for any of the PP scales.

Keywords: Non-parametric statistics, minimax estimation, adaptive estimation, orthogonal polynomials.
AMS Subject Classification: Primary 62G05, 62G20; secondary 62C20.

[†]EURANDOM, P.O. Box 513 - 5600 MB Eindhoven, The Netherlands
[‡]Department of Statistics, Queen's University, Kingston, Ontario, Canada, K7L 3N6

1 Introduction

In this paper we study adaptive non-parametric regression models with a fixed design in the case when the unknown regression function f is analytic in a vicinity $\mathcal{V}$ of the observation interval $[-1,1]$ of the complex plane. The smoothness of f is then characterized by the size of $\mathcal{V}$ and $\max_{\mathcal{V}} |f|$. A more concise description of such dependence – and more accurate results – become feasible when $\mathcal{V}$ is the region E_γ with boundary

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

This boundary set is the ellipse with foci at the end points of the interval $[-1,1]$ and the sum of its semi-axes equal to $\exp\gamma$. The family of such elliptic areas is natural in the sense that $\cap E_\gamma = [-1,1]$ and $\cup E_\gamma = \mathbb{C}$. Note, without loss of generality we have assumed that the regression interval is $[-1,1]$, but an obvious generalization can be made to any real interval $[a,b]$.

We will denote by $\mathcal{A}(\gamma, M)$ the set of functions which are analytic and bounded in E_γ with $|f(z)| \le M$ in that region. 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 [1982] have demonstrated point-wise asymptotically minimax estimators based on Legendre polynomials.

We consider the problem of discrete regression in the model

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

where the points x_k^n form the design knots and the ξ_k are independent identically distributed Gaussian random variables, with zero mean and given variance σ^2. Given the observations $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, 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. This method is easier to implement and to study. For instance, if we consider the design of equally spaced knots

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

one could use the so called Chebyshev discrete polynomials $C_r(x), r = 0, 1, ...,$ (cf. Bateman [1953], Sect. 10.23, p. 223). However, for this design we will find it more convenient to use a family $p_r(x), r = 0, 1, \ldots,$ of normalized Legendre polynomials which are asymptotically equivalent to C_r (cf. Bateman, Sect. 10.23, eq. 7). In particular, the normalized Legendre polynomials $p_r(x)$ are asymptotically orthonormal over the design knots (2). Thus we shall refer to these knots as the *Legendre knots* or *equidistant knots*, and to the set of these knots as the *Legendre design* or *equidistant design*.

As intuition suggests, when we use the equidistant design to estimate the unknown regression function at points which are close to the border of the interval, less information is gathered than when we are interested in estimation inside the interval. Although it might seem that the number of observations available at the end-points is just halved, we shall see that in fact the accuracy of estimation near the border becomes worse by a factor of order $\sqrt{\log n}$, compared to the accuracy obtained inside the interval.

This situation can be improved by using another – non-uniform – design which will balance the distribution of the design points, in favor of increasing the accuracy of the estimation at the end-points. A special classical design having this property is specified by the knots

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

which we will conveniently call the *Chebyshev knots* and the corresponding design the *Chebyshev design*. Remarkably, the classical orthornormal Chebyshev polynomials $t_r(x)$, $r = 0, 1, ...,$ are also orthonormal over the Chebyshev design. As we shall see later, with this polynomials the same rate of convergence is achieved inside the interval, whilst at the end-points the rate is only a factor 2 slower. [1]

Given that for the equidistant design we use the Legendre polynomials and for the Chebyshev design we use the Chebyshev polynomials, one can question whether the difference in rates of convergence is due to the particular method of estimation we are studying or indeed is 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 use of the equidistant design and that this problem does not present itself in the case of the Chebyshev design.

Several remarkable properties of the functions $p_r(x)$ and $t_r(x)$ make this approach attractive for practical purposes. The normalized Legendre polynomials $p_r(x)$ are asymptotically orthonormal over the equally spaced knots while the normalized Chebyshev polynomials $t_r(x)$ are orthonormal over the Chebyshev knots. This makes the evaluation of the projection polynomials straightforward. At the same time the orthonormality property allows an easy evaluation of the variance of the corresponding regression estimators in the statistical framework (1). In the case of known classes $\mathcal{A}(\gamma, M)$, the variance can be easily balanced against the systematic error, thus determining the optimal number N of polynomials in use.

A property that will play a major role in the application of Legendre or Chebyshev design for estimation is the behavior of the functions

$$\frac{1}{N}\sum_{r=0}^{N-1} t_r^2(x) \quad \text{and} \quad \frac{1}{N}\sum_{r=0}^{N-1} p_r^2(x)$$

both inside the interval $[-1, 1]$ and near the end-points (see Lemmas 1 and 2). These terms appear as variances of the corresponding estimators and to a great extent are important in shaping the results (see Theorems 1 and 2).

The structure of this paper is as follows. In Section 2 we introduce the functional classes $\mathcal{A}(\gamma, M)$ and discuss the Legendre and Chebyshev polynomials. In Section 3 asymptotically minimax estimators $\hat{f}_n$, $n \to \infty$, are described in the case when the unknown regression function f belongs to a given fixed class $\mathcal{A}(\gamma, M)$, using Legendre and Chebyshev polynomials for their corresponding designs. In both cases the polynomial estimates we consider are shown to be point-wise asymptotically efficient, for their corresponding designs. In Section 4 we introduce functional scales and construct asymptotically optimal adaptive estimators, under the assumption that the parameters γ and M are unknown.

[1] Motivated by this study B. Levit (2001) introduced a general theory of Optimal Designs in Non-parametric Regression.

2 The building blocks

The purpose of this section is to introduce classes $\mathcal{A}(\gamma, M)$ of analytic functions, as well as the Legendre and Chebyshev polynomials. We discuss their properties and the relation between them. Classes $\mathcal{A}(\gamma, M)$ will serve as the underlying functional classes in the regression problems that we will study, while Legendre and Chebyshev polynomials will be used, in corresponding designs, for constructing the estimators.

2.1 The class $\mathcal{A}(\gamma, M)$

For $\gamma > 0$ let E_γ be the open ellipse in the complex plane, with its boundary defined by

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

The ellipses 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 ∞. One can verify by simple algebra that the elliptic boundary ∂E_γ has its foci at the end-points of the interval $[-1, 1]$, thus

$$E_\gamma = \{z \in \mathbb{C} : |z-1| + |z+1| < e^\gamma + e^{-\gamma}\}.$$

Definition 1 *We denote by $\mathcal{A}(\gamma, M)$ the class of functions analytic inside E_γ such that $|f(z)| \le M$, for all $z \in E_\gamma$.*

Denote by ρ_γ the distance from the interval $[-1, 1]$ to the boundary ∂E_γ. From the integral Cauchy formula for the *mth* derivative of analytic functions we know that for any $\epsilon > 0$ and any ball $B_{\rho_\gamma - \epsilon}$ of radii $\rho_\gamma - \epsilon$ centered at $x \in [-1, 1]$,

$$f^{(m)}(x) = \frac{m!}{2\pi\mathbf{i}} \int_{B_{\rho_\gamma-\epsilon}} \frac{f(z)}{(z-x)^{m+1}}\, dz, \qquad m = 1, 2, \ldots.$$

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

$$|f^{(m)}(x)| \le M m!/\rho_\gamma^m \tag{4}$$

for all $x \in [-1, 1]$. An elementary calculation shows that

$$\rho_\gamma = \cosh\gamma - 1. \tag{5}$$

Equations (4) and (5) will be used later in Section 3, in obtaining some discrete-type approximations to analytic functions.

2.2 Legendre polynomials

Legendre polynomials form a complete system of orthogonal polynomials in $L^2([-1, 1])$. Their explicit definition is (cf. Szegö [1975], p. 68)

$$P_r(x) = 2^{-r} \sum_{\nu=0}^{r} \binom{r}{r-\nu} \binom{r}{\nu} (x-1)^\nu (x+1)^{r-\nu}, \tag{6}$$

and their recurrent form is (cf. Szegö, p. 71)

$$P_0 \equiv 1,$$

$$P_1(x) = x,$$
$$rP_r(x) = (2r-1)xP_{r-1}(x) - (r-1)P_{r-2}(x), \quad r \geq 2.$$

In particular, from the definition (6), it holds

$$P_r(1) = 1, \quad P_r(-1) = (-1)^r. \tag{7}$$

An important bound for the derivatives of Legendre polynomials can be obtained by combining the A.A. Markov inequality (cf. Timan [1963], Sect. 4.8.8)

$$|P_r^{(m)}(x)| \leq r^{2m} \max_{-1\leq x\leq 1} |P_r(x)|, \qquad m = 1, 2, \ldots; \tag{8}$$

with the fact that the maximum of $|P_r(x)|$ is attained at the end points of the interval (cf. Szegö, Sect. 7.21),

$$\max_{-1\leq x\leq 1} |P_r(x)| = |P_r(\pm 1)| = 1. \tag{9}$$

The normalized Legendre polynomials, given by

$$p_r(x) = (2r+1)^{1/2} P_r(x), \qquad r = 0, 1, \ldots, \tag{10}$$

satisfy, from (8)–(10),

$$\max_{-1\leq x\leq 1} |p_r^{(m)}(x)| \leq (2r+1)^{1/2} r^{2m} \qquad m = 1, 2, \ldots. \tag{11}$$

The defined normalized Legendre polynomials form an orthonormal basis in the space $L^2([-1,1])$ corresponding to the inner product

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

Besides that, they are asymptotically orthonormal with respect to a "discrete" inner product defined below which is a discrete version of the "continuous" inner product just mentioned. For a given design, x_k^n, $k = 1, 2, \ldots, n$, we define the corresponding discrete inner product of the functions f and g to be

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

In this subsection, we consider the discrete inner product with respect to the Legendre design, for which x_k^n represent the equidistant knots

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

Let us denote the *kernel* corresponding to the Legendre family p_r by

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

Underlying the quality of our estimators will be remarkable properties of the following type.

Lemma 1 *Let $N \in \mathbb{N}$. The normalized Legendre polynomials p_r satisfy*

(a) *Uniformly for $0 \leq r_1, r_2 \leq N$,*

$$(p_{r_1} | p_{r_2}) = \frac{1}{n} \sum_{k=1}^{n} p_{r_1}(x_k^n) p_{r_2}(x_k^n) = \delta_{r_1 r_2} + O\left(\frac{N^6}{n^2}\right), \qquad (n \to \infty).$$

(b) *If*

$$\alpha_N^2(x) := \frac{1}{N} K_N(x,x) = \frac{1}{N} \sum_{r=0}^{N-1} p_r^2(x), \tag{13}$$

then

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

uniformly on any interval $[a,b] \subset (-1,1)$, 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 will explain why the results presented below hold uniformly only on the compact subsets of $(-1,1)$ while at the extremes of the interval the accuracy of estimation, based on the equidistant design, will deteriorate, even to the extent of being of a different order!*

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

$$\frac{(b-a)^2}{24n^2} \max_{a \leq x \leq b} \left| \frac{d^2}{dx^2} f(x) \right| \tag{14}$$

when the function $f \in C^2[a,b]$ (cf. e.g. Stoer and Bulirsch). Thus, we have

$$\begin{aligned} | (p_{r_1} | p_{r_2}) - \langle p_{r_1} \,|\, p_{r_2} \rangle | &= \left| \frac{1}{n} \sum_{k=1}^{n} p_{r_1}(x_k^n) p_{r_2}(x_k^n) - \frac{1}{2} \int_{-1}^{1} p_{r_1}(x) p_{r_2}(x) dx \right| \\ &\leq \frac{1}{3n^2} \max_{-1 \leq x \leq 1} \frac{d^2}{dx^2} \left(p_{r_1}(x) p_{r_2}(x) \right). \end{aligned} \tag{15}$$

Applying L^2-orthonormality and bounds (11) for the derivatives of $p_r(x)$ we get

$$| (p_{r_1} | p_{r_2}) - \delta_{r_1, r_2} | \leq \frac{1}{3n^2} (2r_1+1)(2r_2+1)(r_1^2 + r_2^2)^2 = O\left(\frac{N^6}{n^2}\right)$$

as $n \to \infty$.

(b) Using the asymptotic formula of Laplace (cf. Szegö, p. 194)

$$p_r(x) \sim \frac{2}{\sqrt{\pi(1-x^2)^{1/2}}} \cos((r+1/2)\sqrt{1-x^2} - \frac{\pi}{4}) + O(r^{-1}), \qquad r \to \infty, |x| < 1 \tag{16}$$

and formula (cf. e.g. Gradshtein and Ryzhik, f. 1.341(1), p. 29)

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

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 compacts in $(-1, 1)$. At the end-points

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

□

Finally, let us mention the following bound on the growth of the Legendre polynomials outside the interval $[-1, 1]$. According to Timan, Theorem 2.9.11, for any polynomial P_r of order r and any $z \in \mathbb{C}$

$$|P_r(z)| \leq |T_r(z)| \max_{-1 \leq x \leq 1} |P_r(x)|.$$

Here $T_r(x)$ are the Chebyshev polynomials which will be discussed in the next section. In particular we will see that $|T_r(z)| \leq e^{\gamma r}, z \in E_\gamma$. Therefore according to (13),

$$|p_r(z)| \leq (2r+1)^{1/2} e^{\gamma r} \tag{17}$$

for every $z \in E_\gamma$.

2.3 Chebyshev polynomials

Chebyshev polynomials appeared for the first time in the problem of finding polynomials $T_r(x) = x^r + a_1 x^{r-1} + \cdots + a_r$ least deviating from zero, in the uniform norm on the interval $[-1, 1]$; Chebyshev [1859]. Normed by $T_r(1) = 1$, they can be represented as

$$T_r(x) = \cos r \arccos x, \quad r = 0, 1, \dots , \tag{18}$$

or in the recurrent form

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

The Chebyshev polynomials are extensively used as an appropriate Fourier basis for approximating non-periodic functions. Consider the normalized family

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

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

$$\langle f \,|\, g\rangle := \frac{1}{\pi}\int_{-1}^{1} \frac{f(x)g(x)}{\sqrt{1-x^2}}\,dx, \tag{19}$$

i.e. they satisfy $\langle t_{r_1} \,|\, t_{r_2}\rangle = \delta_{r_1,r_2}$ 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

$$\sum_{r=0}^{\infty} \langle f \,|\, t_r\rangle t_r(x). \tag{20}$$

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

$$f_N(x) = \sum_{r=0}^{N-1} \langle f \,|\, t_r\rangle t_r(x) \;=\; \langle f \,|\, K_N(x,\cdot)\rangle \tag{21}$$

provides the best approximation to a function f, with respect to the weighted L^2-norm corresponding to (19), among all polynomials of degree less than N. The class $\mathcal{A}(\gamma, M)$ has the important property that the coefficients of the Chebyshev-Fourier series (20) decrease very fast (cf. Timan, Sect. 3.7.3). For all $r = 0, 1, \ldots$, the inequality

$$\sup_{f\in\mathcal{A}(\gamma,M)} |\langle f \,|\, t_r\rangle| \;\leq\; \sqrt{\pi}Me^{-\gamma r} \tag{22}$$

holds. From (21),(22) and the bound $|t_r(x)| \leq \sqrt{2}$ it follows that for every $f \in \mathcal{A}(\gamma, M)$

$$\max_{x\in[-1,1]} |f_N(x) - f(x)| \leq \sum_{r=N}^{\infty} |\langle f \,|\, t_r\rangle|\; |t_r(x)| \leq \frac{\sqrt{2\pi}M}{1-e^{-\gamma}} e^{-\gamma N}, \tag{23}$$

(cf. Timan, Sect. 3.7.3 and 5.4.1).

The function $f_N(x)$ is the polynomial of the best approximation in the weighted L^2-space. Remarkably, for analytic functions of the classes $\mathcal{A}(\gamma, M)$, the approximation $f_N(x)$ based on 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. Timan, Sect. 6.5.2).

According to their definition, the Chebyshev polynomials satisfy $|t_r(x)| \leq \sqrt{2}$ 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

$$\Big\{\omega \in \mathbb{C} : e^{-\gamma} < |\omega| < e^{\gamma}\Big\}$$

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

$$|t_r(z)| = \sqrt{2}\,|T_r(z)| \le \sqrt{2}\,e^{\gamma r}. \tag{24}$$

Denote the discrete inner product by

$$(\, f \,|\, g\,) := \frac{1}{n}\sum_{k=1}^{n} f(x_k^n) g(x_k^n) \tag{25}$$

where the points x_k^n correspond to the Chebyshev design[3]

$$x_k^n = \cos\frac{(2k-1)\pi}{2n}, \quad k = 1, \dots, n. \tag{26}$$

We can state next a lemma which is similar to Lemma 1. The first of the properties is usually referred to as 'double-orthogonality' (cf. e.g. Fox and Parker, Sect. 2.7) and is closely related to the corresponding property of the classical trigonometric polynomials. The second property follows from a standard calculation.

Lemma 2 *The normalized Chebyshev polynomials t_r satisfy*

(a) *For any $r_1, r_2 = 0, 1, ...$*

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

(b) *If*

$$\beta_N^2(x) := \frac{1}{N} K_N(x,x) \;=\; \frac{1}{N}\sum_{r=0}^{N-1} t_r^2(x) \tag{27}$$

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

$$\begin{aligned}\beta_N^2(x) &= 1 + \frac{1}{N}\frac{\cos(N\theta)\ \sin((N-1)\theta)}{\sin\theta} \\ &= 1 + \frac{O(1)}{N},\end{aligned} \tag{28}$$

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

[3]Given the parallel between our work with Legendre and Chebyshev polynomials we duplicate some of the notations, e.g. x_k^n, the inner products, the projection operator K_N, etc. The reader must just keep in mind whether we are working under the Chebyshev or the Legendre setting.

Remark 2 *Note the slightly different behavior at the end-points when compared with the inner points. Compare this with Lemma 1.*

Proof. (a) This is a consequence of the double orthogonality property of the trigonometric Fourier basis (cf. e.g. Gradshtein and Ryzhik, f. 1.351(1), p. 30).

(b) This is a classical identity (cf. e.g. Gradshtein and Ryzhik, 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 (known γ, M) setting.

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 (unknown γ, M) setting. Statistical estimation using the uniform norm as the quality criterion of estimators requires a different approach (cf. Golubev, Lepski and Levit [2001]).

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 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 $w : \mathbb{R} \to \mathbb{R}^+$ such that

$$w(x) = w(-x),$$

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

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 $\mathbf{Var}_f$ the distribution, the expectation and the variance corresponding to f. Sometimes the sub-index f will be dropped, when there is no possibility of confusion.

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

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

where $w \in \mathcal{W}$. The parameter σ_n defining 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 the notation, we will write

$$\hat{f}_{n,N}(x) \;=\; (\mathbf{y} \,|\, K_N(x,\cdot)) \;=\; \sum_{r=0}^{N-1} (\mathbf{y}|p_r)\, p_r(x). \tag{29}$$

Now consider two auxiliary functions:

$$f_N(x) \;=\; \langle f \,|\, K_N(x,\cdot)\rangle \;=\; \sum_{r=0}^{N-1} \langle f \,|\, p_r\rangle\, p_r(x), \tag{30}$$

and

$$f_{n,N}(x) \;=\; (f \,|\, K_N(x,\cdot)) \;=\; \sum_{r=0}^{N-1} (f|p_r)\, p_r(x). \tag{31}$$

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.

The following theorem holds.

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 (13), $\tilde{f}_n$ is an arbitrary estimator of f, $\hat{f}_n = \hat{f}_{n,N}$ is the projection estimator (29) with

$$N = N_n := \left\lfloor \frac{1}{2\gamma}\log n \right\rfloor \qquad \textit{and} \qquad \xi \sim \mathcal{N}(0,1). \tag{32}$$

Proof: the upper bound. Let N be given by (32). As usual we decompose the mean square error as

$$\mathbf{E}\big(\hat{f}_{n,N}(x) - f(x)\big)^2 = \mathbf{Var}\, v_N^2(x) + b_N^2(x) \tag{33}$$

where, according to (29) and (31),

$$v_N(x) \;=\; \hat{f}_{n,N}(x) - f_{n,N}(x) \;=\; \frac{1}{n}\sum_{k=1}^{n} \xi_k K_N(x, x_k^n) \tag{34}$$

is a zero-mean stochastic term and

$$b_N(x) \;=\; \big(f_{n,N}(x) - f_N(x)\big) \;+\; \big(f_N(x) - f(x)\big) \tag{35}$$

is the bias.

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

$$\begin{aligned}
\mathbf{Var}\, v_N(x) \;&=\; \frac{\sigma^2}{n^2}\sum_{k=1}^{n} K_N^2(x, x_k^n) \;=\; \frac{\sigma^2}{n^2}\sum_{k=1}^{n}\left(\sum_{r=0}^{N-1} p_r(x)p_r(x_k^n)\right)^2 \\
&=\; \frac{\sigma^2}{n}\sum_{r_1=0}^{N-1}\sum_{r_2=0}^{N-1} p_{r_1}(x)p_{r_2}(x)\,\frac{1}{n}\sum_{k=1}^{n} p_{r_1}(x_k^n)p_{r_2}(x_k^n) \\
&=\; \frac{\sigma^2}{n}\sum_{r_1=0}^{N-1}\sum_{r_2=0}^{N-1} p_{r_1}(x)p_{r_2}(x)\left(\delta r_1 r_2 + O\left(\frac{N^6}{n^3}\right)\right) \\
&=\; \frac{\sigma^2}{n}\sum_{r=0}^{N-1} p_r^2(x) + O\left(\frac{N^6}{n^3}\right)\sum_{r_1=0}^{N-1}\sum_{r_2=0}^{N-1} p_{r_1}(x)p_{r_2}(x).
\end{aligned} \tag{36}$$

Now, applying the Cauchy-Schwartz inequality we see that

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

Thus, according to the last two equations and (32),

$$\mathbf{Var}\, v_N(x) \;=\; \alpha_N^2(x)\,\frac{\sigma^2 N}{n}\,(1 + o(1)) \tag{37}$$

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

Now let us consider the bias. First, we have

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

By definition

$$\big|\,(f\,|\,p_r)\;-\;\langle f\,|\,p_r\rangle\big| \;=\; \left|\frac{1}{2}\int_{-1}^{1} f(x)p_r(x)\,dx - \frac{1}{n}\sum_{k=1}^{n} f(x_k^n)p_r(x_k^n)\right|. \tag{38}$$

Next, applying (14), this difference can be bounded by

$$\frac{1}{3n^2}\max_{x\in[-1,1]}\left|\frac{d^2}{dx^2}f(x)p_r(x)\right|. \tag{39}$$

Thus, applying the bounds for the derivatives $|f^{(m)}(x)| \le Mm!/\rho_\gamma^m$ (cf. Sect. 2.1) and $|p_r^{(m)}(x)| \le (2r+1)^{1/2}r^{2m}$ (cf. eq. (11)), it follows that

$$\begin{aligned}\big|\,(f\,|\,p_r) - \langle f\,|\,p_r\rangle\big| \;&\le\; \frac{M}{3n^2}\,(\,(2\rho_\gamma)^{-2} + 2\rho_\gamma^{-1}(2r+1)^{1/2}r^2 + (2r+1)^{1/2}r^4)\\ &=\; O\Big(\frac{r^5}{n^2}\Big), \qquad (n\to\infty).\end{aligned} \tag{40}$$

Combining Cauchy-Schwartz inequality with the previous bound and using the fact that N is of order $O(\log n)$, cf. eq. (32), we find

$$\begin{aligned}(\,f_{n,N}(x) - f_N(x)\,)^2 \;&\le\; \sum_{r=0}^{N-1}\big((f\,|\,p_r) - \langle f\,|\,p_r\rangle\big)^2\;\sum_{r=0}^{N-1}p_r^2(x)\\ &=\; \alpha_N^2(x)\,N\sum_{r=0}^{N-1}\,\big((f\,|\,p_r) - \langle f\,|\,p_r\rangle\big)^2 \;=\; \alpha_N^2(x)\,O\Big(\frac{N^{12}}{n^4}\Big)\\ &=\; \alpha_N^2(x)\frac{\sigma^2 N}{n}\,O\Big(\frac{N^{11}}{n^3}\Big) \;=\; o(1)\,\mathbf{Var}v_N(x).\end{aligned} \tag{41}$$

As demonstrated in Ibragimov and Has'minskii [1981], for functions $f \in \mathcal{A}(\gamma, M)$

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

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 previous inequality, for some $C_2 > 0$,

$$\begin{aligned}\big(f_N(x) - f(x)\big)^2 \;&\le\; \Big(\sum_{r=N}^{\infty}|\langle f\,|\,p_r\rangle|\,|p_r(x)|\Big)^2\\ &\le\; C_2\,e^{-2\gamma N} \;\sim\; C_2 n^{-1} \;=\; o(1)\,\mathbf{Var}\,v_N(x).\end{aligned} \tag{42}$$

At the end-points of the interval we have $|p_r(\pm 1)| = (2r+1)^{1/2}$, see eqs. (7) and (10), 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

$$\big(f_N(x) - f(x)\big)^2 \leq C_4\, N e^{-2\gamma N} \;\sim\; C_4 \frac{N}{n} \;=\; o(1)\, \mathbf{Var}\, v_N(x). \tag{43}$$

From (33), (37), (41) and (42) or (43) 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:

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w \left(\alpha_N^{-1}(x) \sqrt{\frac{n}{\sigma^2 N}}\; \big(\hat{f}_n(x) - f(x)\big)\right) \;=\; \mathbf{E}\, w(\xi). \tag{44}$$

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

$$f_\theta(z) = \theta \sqrt{\frac{\sigma^2}{n}}\, \frac{K_{\bar{N}}(x, z)}{\sqrt{K_{\bar{N}}(x, x)}}, \qquad |\theta| < \theta_n = \bar{N}^{1/2}, \tag{45}$$

where we will use

$$\bar{N} = \bar{N}_n = \lfloor N_n - 3\log N_n \rfloor\,, \tag{46}$$

see (32). Note that $\bar{N}$ is asymptotically equivalent to $N = N_n$ when $N \to \infty$. This implies, according to Lemma 1(b), that

$$\frac{\alpha_{\bar{N}}^2(x)}{\alpha_N^2(x)} \to 1, \tag{47}$$

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

We need the following lemma.

Lemma 3 *For a given $x \in [-1, 1]$ and any $z \in E_\gamma$, let $f_\theta(z)$ be defined by (45). Then*

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

(b) $f_\theta \in \mathcal{A}(\gamma, M)$, $\quad |\theta| < \theta_n$, *for all n big 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 of lemma.

(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 (17), applying the Cauchy-Schwartz inequality and recalling the definition of $\bar N = \bar N_n$, we obtain

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

in E_γ for all n large enough.

(c) Denote

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

We can see that T is normally distributed,

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

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

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(a) and the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned}
\mathcal{I}_n \;&=\; \frac{1}{n}K_{\bar N}^{-1}(x,x)\sum_{k=1}^{n}K_{\bar N}^2(x,x_k^n) \;=\; \frac{1}{n}K_{\bar N}^{-1}(x,x)\sum_{k=1}^{n}\left(\sum_{r=0}^{\bar N-1}p_r(x)p_r(x_k^n)\right)^2 \\
&=\; K_{\bar N}^{-1}(x,x)\sum_{r_1=0}^{\bar N-1}\sum_{r_2=0}^{\bar N-1}\left(p_{r_1}(x)p_{r_2}(x)\,\frac{1}{n}\sum_{k=1}^{n}p_{r_1}(x_k^n)p_{r_2}(x_k^n)\right) \\
&=\; K_{\bar N}^{-1}(x,x)\sum_{r_1=0}^{\bar N-1}\sum_{r_2=0}^{\bar N-1}\left(p_{r_1}(x)p_{r_2}(x)\,\left(\delta_{r_1\,r_2} + O\left(\frac{N^6}{n^2}\right)\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= K_{\bar N}^{-1}(x,x)\sum_{r=0}^{\bar N-1} p_r^2(x) + O\left(\frac{\bar N^6}{n^2}\right) K_{\bar N}^{-1}(x,x)\sum_{r_1=0}^{\bar N-1}\sum_{r_2=0}^{\bar N-1} p_{r_1}(x)p_{r_2}(x) \\
&= 1 + O\left(\frac{\bar N^6}{n^2}\right) K_{\bar N}^{-1}(x,x)\left(\sum_{r=0}^{\bar N-1} p_r(x)\right)^2 \\
&= 1 + o(1), \qquad\qquad (n\to\infty). \qquad (48)
\end{aligned}
$$

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

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

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. (47),

$$
\mathcal{R} := \inf_{\tilde f_n}\sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\alpha_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\big(\tilde f_n(x) - f(x)\big)\right) \qquad (49)
$$

$$
= \inf_{\tilde f_n}\sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\alpha_{\bar N}^{-1}(x)\sqrt{\frac{n}{\sigma^2 \bar N}}\big(\tilde f_n(x) - f(x)\big)(1+o(1))\right) \qquad (50)
$$

$$
\geq \inf_{\tilde f_n}\sup_{f_\theta}\mathbf{E}_{f_\theta}\, w\left((1+o(1))\alpha_{\bar N}^{-1}(x)\sqrt{\frac{n}{\sigma^2\bar N}}\big(\tilde f_n(x) - f_\theta(x)\big)\right), \quad (\bar N\to\infty).
$$

Denote $\tilde\theta = \alpha_{\bar N}^{-1}(x)\sqrt{\frac{n}{\sigma^2\bar N}}\tilde f_n(x)$. Then applying Lemma 3(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 (49) and (50), applying Lemma 1(b) and definition (46) 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

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

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

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

given that T is sufficient for θ, according to Lemma 3(c). Applying results presented in Levit [1980], 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 dominate convergence theorem for $\delta \to 0$ we get

$$\liminf_{n\to\infty} \inf_{\tilde f_n} \sup_{f\in\mathcal{A}(\gamma,M)} \mathbf{E}_f\; w\left(\alpha_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\big(\tilde f_n(x) - f(x)\big)\right) \geq \mathbf{E}\, w(\xi). \tag{51}$$

Finally, from (44) and (51) the theorem is proved. □

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

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

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

where $\hat f_n$ and $\tilde f_n$ are as in the 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 (26). 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 the notation

$$\hat{f}_{n,N}(x) \;=\; (\mathbf{y} \,|\, K_N(x,\cdot)) \;=\; \sum_{r=0}^{N-1} (\mathbf{y}|t_r)\, t_r(x), \tag{52}$$

and consider the two functions

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

and

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

see the footnote on page 9 with regards to these notations. Then the following result holds.

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

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

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

where $\tilde{f}_n$ is an arbitrary estimator of f, $\hat{f}_n = \hat{f}_{n,N}$ is the projection estimator (52) with

$$N = N_n := \left\lfloor \frac{1}{2\gamma}\log\left(M^2\,\gamma\,(1-e^{-\gamma})^{-2} n\right)\right\rfloor, \tag{53}$$

$\beta_N^2(x)$ is defined by (27) 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 the Lemmas 1(a) and 2(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. (34) and (35), we have

$$\mathbf{E}(\hat{f}_{n,N}(x) - f(x))^2 = \mathbf{Var}\, v_N^2(x) + b_N^2(x). \tag{54}$$

Let us first analyze the variance of $v_N(x)$. As before (cf. eq. (36)), applying Lemma 2(a) we obtain

$$\mathbf{Var}\, v_N(x) \;=\; \frac{\sigma^2}{n} \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} t_{r_1}(x) t_{r_2}(x) \delta r_1 r_2 \;=\; \beta_N^2(x)\, \frac{\sigma^2 N}{n} \tag{55}$$

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

Now let us consider the bias

$$b_N(x) \;=\; \big(f_{n,N}(x) - f_N(x)\big) \;+\; \big(f_N(x) - f(x)\big). \tag{56}$$

Using Cauchy-Schwartz inequality we see that

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

If we rewrite the inner products as

$$(f \,|\, t_r) = \frac{1}{\pi} \sum_{k=1}^{n} f\left(\cos(k - 1/2)\frac{\pi}{n}\right) \cos\left(r(k - 1/2)\frac{\pi}{n}\right) \frac{\pi}{n}$$

and

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

(cf. eqs. (19) and (25)), we can apply the same arguments that we used in (38)–(40). Using the bounds for the derivatives of f given in eq. (4) we find that

$$\begin{aligned}
\big|(f \,|\, t_r) - \langle f \,|\, t_r \rangle\big| \;&\leq\; \frac{\pi}{24} \left(\frac{\pi}{n}\right)^2 \max_{\zeta} \left| \frac{d^2}{d\zeta^2} f(\cos\zeta) \cos(r\zeta) \right| \\
&\leq\; \frac{\pi^3}{24\, n^2} M \left(r^2 + \frac{(2r+1)}{\rho_\gamma} + \frac{2}{\rho_\gamma^2} \right) \\
&\leq\; \frac{\pi^3 (r+1)^2}{6\, n^2} M\, \max(1, \rho_\gamma^{-1}, \rho_\gamma^{-2}) \;=\; M C_\gamma \frac{(r+1)^2}{n^2}
\end{aligned} \tag{57}$$

where, using (5), one can verify that

$$C_\gamma = O(1 - e^{-\gamma})^{-4}, \tag{58}$$

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

$$\big(f_{n,N}(x) - f_N(x)\big)^2 \;=\; \beta_N^2(x)\, O\left(M^2\, (1 - e^{-\gamma})^{-8} \frac{N^6}{n^4} \right). \tag{59}$$

If we choose $N = N_n$

$$\big(f_{n,N}(x) - f_N(x)\big)^2 \;=\; o(1)\,\mathbf{Var} v_N(x), \qquad (n \to \infty). \tag{60}$$

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. (42) and (43)). Now, one can see that both cases can be considered simultaneously. From (23) one can see that for any x and $N = N_n$

$$\begin{aligned} \big(f_N(x) - f(x)\big)^2 \;&\le\; 2\pi M^2 (1 - e^{-\gamma})^{-2}\, e^{-2\gamma N} \;=\; O\Big(\frac{1}{\gamma n}\Big) \\ &=\; \beta_N^2(x)\frac{\sigma^2 N}{n} O\Big(\frac{1}{\gamma N}\Big) \;=\; o(1)\,\mathbf{Var}\, v_N(x), \end{aligned} \tag{61}$$

when $n \to \infty$. From (54)–(56), (60) and (61) we have proved that

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

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:

$$\lim_{n\to\infty} \sup_{f \in \mathcal{A}(\gamma,M)} \mathbf{E}_f\, w\left(\beta_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\;\big(\hat{f}_n(x) - f(x)\big)\right) \;=\; \mathbf{E}\, w(\xi). \tag{62}$$

Proof of the lower bound for the risk. We can follow the same proof of the lower bound 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

$$f_\theta(z) = \theta\sqrt{\frac{\sigma^2}{n}}\,\frac{K_{\bar{N}}(x, z)}{\sqrt{K_{\bar{N}}(x, x)}} \qquad |\theta| < \theta_n = \bar{N}^{1/2} \tag{63}$$

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

$$\bar{N} = \bar{N}_n = \lfloor N_n - 3\log N_n \rfloor \tag{64}$$

(cf. definition of N_n in eq. (53)).

Lemma 4 *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) $f_\theta \in \mathcal{A}(\gamma, M), \quad |\theta| < \theta_n$, *for n big 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

$$T = \theta + \xi \tag{65}$$

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

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

$$\Lambda \;:=\; \log \frac{d\mathbf{P}_\theta}{d\mathbf{P}_0} \;=\; \theta T - \frac{\theta^2}{2}. \tag{66}$$

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 3. Nevertheless, a couple of remarks can be made. First, the bound (17) for Legendre polynomials is also a bound for the Chebyshev polynomials, thus the proof of (b) remaines the same. Second, in the present case, $\mathcal{I}_n = 1$ given exact orthogonality of Chebyshev polynomials (cf. eq. (48)). The rest of the proofs of the lemma and the theorem remain the same and we get

$$\liminf_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f \in \mathcal{A}(\gamma,M)} \mathbf{E}_f \; w\left(\beta_N^{-1}(x)\sqrt{\frac{n}{\sigma^2 N}}\big(\tilde{f}_n(x) - f(x)\big)\right) \geq \mathbf{E}\, w(\xi). \tag{67}$$

The theorem follows from (62) and (67). □

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

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

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

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

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

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

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 Lepski and Levit [1998]. 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 $(\gamma^{extr}, M^{extr})$, but rather should be thought as some sequences $(\gamma_n^{extr}, M_n^{extr})$, 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 (53). If the following conditions are satisfied*

$$\lim_{n\to\infty} \gamma N = \infty, \tag{68}$$

$$\lim_{n\to\infty} M^2(1-e^{-\gamma})^{-8} N^5\, n^{-3} = 0, \tag{69}$$

$$\lim_{n\to\infty} N = \infty, \tag{70}$$

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 (52) and $\tilde{f}_n$ is an arbitrary estimator of f.

Proof. Note that the previous conditions 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 on checking that conditions (68) and (69) guarantee asymptotic unbiasness of the optimal estimator (cf. eqs. (59) and (61)), while (70) allows us to prove the lower bound result. The rest of the proof is the same.

□

Though conditions (68)–(70) 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*

$$\limsup_{n\to\infty} \frac{M^2}{\log n} = 0, \tag{71}$$

$$\liminf_{n\to\infty} M^2 \log n = \infty, \tag{72}$$

$$\limsup_{n\to\infty} \frac{\gamma}{\log\log n} = 0, \tag{73}$$

$$\liminf_{n\to\infty} \gamma \log n = \infty, \tag{74}$$

then, with $N = N_n$ defined by (53),

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

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

for all $x \in [-1,1]$. Here $\hat{f}_n = \hat{f}_{n,N}$ is the projection estimator (52) and $\tilde{f}_n$ is an arbitrary estimator of f.

Proof. In order to prove the theorem, we only need to verify that hypothesis of the Theorem 3 are satisfied, i.e. we just need to assure that the limits (68)–(70) are still valid (cf. eqs. (59) and (61)). If γ and M are bounded then trivially (68)–(70) 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 (72) and (73), 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 (71) and (74)

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

Case $\gamma \to \infty$: Applying (72) and (73)

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

thus N and γN go to infinity. From (71) and (74)

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

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 $\upsilon = (\gamma, M)$ where υ belongs to the region $\Gamma_n \subset \mathbb{R}^2_+$. Let $\mathcal{A}(\upsilon) = \mathcal{A}(\gamma, M)$ and define the functional scale $\mathcal{A}_{\Gamma_n}$,

$$\mathcal{A}_{\Gamma_n} := \left\{\mathcal{A}(\upsilon) \;\middle|\; \upsilon \in \Gamma_n\right\},$$

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 2 *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 $\tilde{f}_n \in \mathcal{F}$*

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

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

$$\lim_{n\to\infty} \sup_{\upsilon\in\Gamma_n} M^2(1-e^{-\gamma})^{-8} N_n^5(\upsilon)\, n^{-3} = 0, \tag{75}$$

where $N_n(\upsilon)$ was defined in (53), is satisfied.

The previous condition is aimed to guarantee that the approximation arguments which were used in (59) and (60) are still applicable. Let us remark that in this condition the powers of the terms are not so relevant as far as we have $N_n(\upsilon)$ of orden $\log n$ at most.

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

Definition 4 *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 exit finite constants M_+ and C_+ such that for all $(\gamma, M) \in \Gamma_n$ uniformly*

$$\limsup_{n\to\infty} \sup_{\upsilon\in\Gamma_n} M \leq M_+, \quad \textit{and} \tag{76}$$

$$\limsup_{n\to\infty} \sup_{\upsilon\in\Gamma_n} \gamma^{-1} \log n \leq C_+. \tag{77}$$

Regular-pseudo-parametric scales are regular, in the sense of Definition 3, and uniformly on them, we have parametric rates, i.e. the rate $n^{-1/2}$ is achieved given

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

Definition 5 *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*

$$\limsup_{n\to\infty} \sup_{\upsilon\in\Gamma_n} \frac{M^2}{\log n} = 0, \tag{78}$$

$$\liminf_{n\to\infty} \inf_{\upsilon\in\Gamma_n} M^2 \log n = \infty, \tag{79}$$

$$\liminf_{n\to\infty} \sup_{\upsilon\in\Gamma_n} \frac{\gamma}{\log\log n} = 0, \tag{80}$$

$$\liminf_{n\to\infty} \inf_{\upsilon\in\Gamma_n} \gamma \log n = \infty. \tag{81}$$

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

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

Also note that regular-non-parametric scales are regular, in the sense of Definition 3.

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

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

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*

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

Here

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

where N_n was defined in (53) for any $\upsilon \in \Gamma_n$.

The estimator. Let us first describe our adaptive estimator. Fix the parameters, $1/2 < l < 1$, $1/2 < \delta < 1$, $p_1 > 0$, $l_1 = \delta l$ 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

$$N_{i+1} - N_i \sim l(\log N_i)^{1-\frac{1}{l}} N_i \to \infty \qquad (i \to \infty) \tag{82}$$

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

$$\frac{N_{i+1}}{N_i} \sim e^{li^{l-1}} \sim 1 \qquad (i \to \infty). \tag{83}$$

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

$$I_n = \arg\max_i \left\{N_i \le n^{1/2}\right\}. \tag{84}$$

Since for any δ, $(0 < \delta < 1/2)$ 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

$$N_{i(\upsilon)-1} < N_n(\upsilon) \le N_{i(\upsilon)}. \tag{85}$$

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

$$\hat{f}_i(x) = \hat{f}_{n,N_i}(x), \qquad b_i = \mathbf{E}_f\,\hat{f}_i(x) - f(x),$$

$$\sigma_i^2 = \mathbf{Var}_f\,\hat{f}_i(x), \qquad \hat{\sigma}_i^2 = \beta_{N_i}^2(x)\frac{\sigma^2 N_i}{n},$$

$$\sigma_{i,j}^2 = \mathbf{Var}_f\left(\hat{f}_j(x) - \hat{f}_i(x)\right), \qquad \hat{\sigma}_{i,j}^2 = \hat{\sigma}_j^2 - \hat{\sigma}_i^2,$$

and define the sequence of thresholds

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

Adaptive procedure. Define

$$\hat{i} = \min\left\{1 \le i \le I_n : \left|\hat{f}_j(x) - \hat{f}_i(x)\right| \le \lambda_j\,\hat{\sigma}_{i,j} \quad \forall j\ (i \le j \le I_n)\right\}.$$

We will prove that the estimator

$$\hat{f}_n(x) = \hat{f}_{\hat{i}}(x)$$

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

Lemma 5 *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 of lemma. (a) As we saw before

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

From equations (59), (84), and conditions for RPP scales, or as well, conditions for RNP scales (cf. Definitions 4 and 5), we have

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

From (61),

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

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

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

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 (55), taking $N = N_j$, we obtain

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

(c) We have

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

Now,

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

Applying the Cauchy-Schwartz inequality, (57) and (58) we see that for regular scales, as we did in (a),

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

Also, applying the Cauchy-Schwartz 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 (22), the definition (53) of N_n and condition (81) one can verify that

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

Now,

$$\begin{aligned}
b_2^2 &= O(1)\,\frac{1}{n}\Big(\sum_{r=0}^{N_j-1} t_r^2(x) - \sum_{r=0}^{N_i-1} t_r^2(x)\Big) \\
&= O(1)\Big(\beta^2_{N_j}(x)\frac{\sigma^2 N_j}{n} - \beta^2_{N_i}(x)\frac{\sigma^2 N_i}{n}\Big) \\
&= O(1)\,(\hat{\sigma}_j^2 - \hat{\sigma}_i^2), \qquad (n\to\infty).
\end{aligned}$$

Thus $\big(b_j - b_i\big)^2 = O(1)\,(\hat{\sigma}_j^2 - \hat{\sigma}_i^2)$ for any $x \in [-1,1]$, when $n \to \infty$.

(d) Applying again the Cauchy-Schwartz inequality together with Lemma 2(a) we see that

$$\begin{aligned}
\mathbf{Var}\big(\hat{f}_j(x) - \hat{f}_i(x)\big) &= \frac{\sigma^2}{n^2}\sum_{k=1}^{n}\big(K_{N_j}(x,x_k^n) - K_{N_i}(x,x_k^n)\big)^2 \\
&= \frac{\sigma^2}{n}\sum_{r_1=N_i}^{N_j-1}\sum_{r_2=N_i}^{N_j-1}\Big(t_{r_1}(x)t_{r_2}(x)\,\frac{1}{n}\sum_{k=1}^{n} t_{r_1}(x_k^n)t_{r_2}(x_k^n)\Big) \\
&= \frac{\sigma^2}{n}\sum_{r_1=N_i}^{N_j-1}\sum_{r_2=N_i}^{N_j-1} t_{r_1}(x)t_{r_2}(x)\delta r_1 r_2 = \frac{\sigma^2}{n}\sum_{r=N_i}^{N_j-1} t_r^2(x) \\
&= \hat{\sigma}_j^2 - \hat{\sigma}_i^2.
\end{aligned}$$

$\square$

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

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

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

$$\begin{aligned} \left\{\hat{i} \le i(\upsilon)\right\} &\subset \left\{\left|\hat{f}_{\hat{i}}(x) - \hat{f}_{i(\upsilon)}(x)\right| \le \hat{\sigma}_{\hat{i},i(\upsilon)}\,\lambda_{i(\upsilon)}\right\} \\ &\subset \left\{\left|\hat{f}_{\hat{i}}(x) - \hat{f}_{i(\upsilon)}(x)\right| \le \hat{\sigma}_{i(\upsilon)}\,\lambda_{i(\upsilon)}\right\}, \end{aligned}$$

given the definition of $\hat{i}$ and the property $\hat{\sigma}^2_{i,j} = \hat{\sigma}^2_j - \hat{\sigma}^2_i$. Therefore

$$\begin{aligned} R_n^-(f) &\le \mathbf{E}\left\{\mathbb{1}_{\{\hat{i}\le i(\upsilon)\}}\Big(\left|\hat{f}_{\hat{i}}(x) - \hat{f}_{i(\upsilon)}(x)\right| + \left|\hat{f}_{i(\upsilon)}(x) - f(x)\right|\Big)^p\right\} \\ &\le \mathbf{E}\left(\hat{\sigma}_{i(\upsilon)}\lambda_{i(\upsilon)} + \left|\hat{f}_{i(\upsilon)}(x) - f(x)\right|\right)^p \\ &\le \mathbf{E}\left(\hat{\sigma}_{i(\upsilon)}\lambda_{i(\upsilon)} + |b_{i(\upsilon)}| + \sigma_{i(\upsilon)}|\xi|\right)^p \end{aligned} \tag{86}$$

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

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

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

uniformly in such scales, when $n \to \infty$. From Lemma 5 we know that $b^2_{i(\upsilon)} = o(1)\,\hat{\sigma}^2_{i(\upsilon)}$, thus $b^2_{i(\upsilon)} = o(n^{-1})$. Using the above in (86) we have that for any RPP scale, uniformly,

$$\sup_{f\in\mathcal{A}(\upsilon)} R_n^-(f) = O(n^{-p/2}), \qquad (n\to\infty). \tag{87}$$

From (86), applying Lemma 5, the dominated convergence theorem and asymptotic (83), uniformly in any RNP scale

$$\sup_{f\in\mathcal{A}(\upsilon)} R_n^-(f) \le \psi_n^p(\upsilon)\big(1+o(1)\big), \qquad (n\to\infty). \tag{88}$$

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

$$A_i = \left\{\omega : \left|\hat{f}_i(x) - f(x)\right| \le \sqrt{2}\,\hat{\sigma}_i\,\lambda_i\right\}.$$

Applying the Hölder and Cauchy-Schwartz inequalities we obtain

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

where

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

and

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

We have that

$$
\begin{aligned}
\mathbf{P}(\hat{\imath} = i) &\leq \mathbf{P}(\hat{\imath} \geq i) \\
&\leq \sum_{j=i+1}^{\infty} \mathbf{P}\Big(\left|\hat{f}_{j-1}(x) - \hat{f}_{i-1}(x)\right| > \hat{\sigma}_{i-1,j-1}\, \lambda_{j-1} \Big), \qquad (89)
\end{aligned}
$$

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 5, (c) and (d), and a well known bound for the tails of the normal distribution (cf. Feller [1968], Lemma 2) we find that

$$
\begin{aligned}
\mathbf{P}(\,|\,\hat{f}_j(x) - \hat{f}_i(x)| > \hat{\sigma}_{i,j}\lambda_j\,) &\leq \mathbf{P}\left(|\xi| > \lambda_j - \frac{|b_j - b_i|}{\hat{\sigma}_{i,j}} \right) \\
&\leq \exp\left\{ -\frac{1}{2} (\lambda_j - C_1)^2 \right\} \leq \exp\left\{ -\frac{1}{2}\lambda_j^2 + C_1 \lambda_j \right\},
\end{aligned}
$$

for some $C_1 > 0$ and n large enough. Returning to (89) we obtain that

$$
\begin{aligned}
\mathbf{P}(\hat{\imath} = i) &\leq \sum_{j=i+1}^{\infty} \exp\left\{ -\frac{1}{2}\lambda_{j-1}^2 + C_1\lambda_{j-1} \right\} = \sum_{j=i}^{\infty} \exp\left\{ -\frac{1}{2}\lambda_j^2 + C_1\lambda_j \right\} \\
&= \sum_{j=i}^{\infty} \exp\left\{ -\frac{pj^l + p_1 j^{l_1}}{2} + C_1\sqrt{pj^l + p_1 j^{l_1}} \right\}
\end{aligned}
$$

$$
\begin{aligned}
&\le \sum_{j=i}^{\infty} \exp\Big\{-\frac{pj^l}{2}-\frac{p_1 j^{l_1}}{3}\Big\} \sim \frac{2}{pl} i^{1-l} \exp\Big\{-\frac{pi^l}{2}-\frac{p_1 i^{l_1}}{3}\Big\} \\
&= \frac{2}{pl} i^{1-l} N_i^{-p/2} \exp\Big\{-\frac{p_1 i^{l_1}}{3}\Big\} \le C_2 N_i^{-p/2} \exp\Big\{-\frac{p_1 i^{l_1}}{4}\Big\}
\end{aligned}
$$

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

$$
\sup_{f\in\mathcal{A}(\upsilon)} R_1^+(f) = O(n^{-p/2}) \sum_{i=1}^{\infty} i^{pl/2} \exp\Big\{-p_1 i^{l_1}/4\Big\} = O(n^{-p/2}), \tag{90}
$$

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 5, (a) and (b), in the same way as before, we have

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

for all $i \ge i(\upsilon)$, n large enough. Thus, applying again Lemma 5, (a) and (b), and previous bound

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

and finally

$$
\sup_{f\in\mathcal{A}(\upsilon)} R_{n,2}^+(f) = O\big(n^{-p/2}\big). \tag{91}
$$

Finally we can conclude from (87), (88), (90) and (91) that $\hat f_n \in \mathcal{F}_p(x)$ and

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

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

4.3 Lower bound

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

$$
\psi_n(\upsilon) = \sigma_n(\upsilon)\phi_n(\upsilon)
$$

where

$$\sigma_n^2(\upsilon) \;= \beta_{N_n}^2(x)\frac{\sigma^2 N_n}{n}, \qquad\qquad \phi_n^2(\upsilon) \;=\; p\log N_n,$$

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

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

Proof. Denote for shortness $\psi_\upsilon = \psi_n(\upsilon)$, $\phi_\upsilon = \phi_n(\upsilon)$ and $\sigma_\upsilon = \sigma_n(\upsilon)$. Choose $\bar{N}$ as it was defined in (64), and define $\bar{\psi}_\upsilon \;= \bar{\sigma}_\upsilon\bar{\phi}_\upsilon$ where

$$\bar{\sigma}_\upsilon^2 \;= \beta_{\bar{N}}^2(x)\frac{\sigma^2\bar{N}}{n} \qquad \text{and} \qquad \bar{\phi}_\upsilon^2 \;=\; p\log\bar{N}.$$

Define $f_0 \equiv 0$ and $f_1 = f_\theta$ for $\theta = \bar{\phi}_\upsilon - \bar{\phi}_\upsilon{}^{1/2}$, where f_θ belongs to the parametric family defined in (63). Notice that $|\theta| < \bar{N}^{1/2}$ for all n big enough. According to Lemma 4, $f_1 \in \mathcal{A}(\upsilon)$ 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 $f_n^* = \bar{\psi}_\upsilon^{-1}\tilde{f}_n(x)$ and $L = \bar{\phi}_\upsilon^{-1}\theta$. Then

$$\bar{\psi}_\upsilon^{-1}\big(\tilde{f}_n(x) - f_1(x)\big) = f_n^* - \bar{\psi}_\upsilon^{-1}f_1(x) = f_n^* - \bar{\phi}_\upsilon^{-1}\theta = f_n^* - L \tag{92}$$

whereas

$$\begin{aligned}\frac{\sqrt{n}}{\sigma}\big(\tilde{f}_n(x) - f_0(x)\big) \;&=\; \frac{\sqrt{n}}{\sigma}\,\bar{\psi}_\upsilon\, f_n^*(x) = \sqrt{\bar{N}}\,\bar{\phi}_\upsilon\, f_n^*(x)\\ &=\; f_n^*\exp\left\{\frac{\log\bar{N}}{2} + \log\bar{\phi}_\upsilon\right\}.\end{aligned} \tag{93}$$

Denote $\mathbf{P}_0$ and $\mathbf{P}_1$ the probabilities associated with f_0 and f_1 respectively. From equations (65) and (66),

$$\frac{d\mathbf{P}_0}{d\mathbf{P}_1}(y) = \exp\left\{-\frac{\theta^2}{2} - \theta\xi\right\} \tag{94}$$

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

$$\begin{aligned}\bar{\mathcal{R}} \;&:=\; \sup_{f\in\mathcal{A}(\upsilon)} \mathbf{E}^{(n)}\left|\,\bar{\psi}_\upsilon^{-1}\big(\,\tilde{f}_n(x) - f(x)\,\big)\,\right|^p \geq \mathbf{E}_1\left|\,\bar{\psi}_\upsilon^{-1}\big(\,\tilde{f}_n(x) - f_1(x)\,\big)\,\right|^p\\ &\geq\; q\,\mathbf{E}_0\left|\,\frac{\sqrt{n}}{\sigma}\big(\,\tilde{f}_n(x) - f_0(x)\,\big)\,\right|^p +\\ &\qquad (1-q)\,\mathbf{E}_1\left|\,\bar{\psi}_\upsilon^{-1}\big(\,\tilde{f}_n(x) - f_1(x)\,\big)\,\right|^p + O(q).\end{aligned} \tag{95}$$

According to (92)–(95),

$$\begin{aligned}
\bar{\mathcal{R}} &\geq q\,\exp\left\{\frac{\bar{\phi}_\upsilon}{2}+p\log\bar{\phi}_\upsilon\right\}\mathbf{E}_0\left|f_n^*(x)\right|^p+(1-q)\,\mathbf{E}_1\left|f_n^*(x)-L\right|^p+O(q)\\
&\geq (1-q)\,\mathbf{E}_1\;(\,Z\mid f_n^*(x)\mid^p+\mid f_n^*(x)-L\mid^p)+O(q)\\
&\geq (1-q)\,\mathbf{E}_1\;\inf_x(\,Z\,|x|^p+|\,x-L\,|^p)+O(q) \qquad (96)
\end{aligned}$$

where

$$Z=q\,\exp\left\{\frac{\bar{\phi}_\upsilon}{2}+p\log\bar{\phi}_\upsilon\right\}\frac{d\mathbf{P}_0}{d\mathbf{P}_1}.$$

From (94) and definition of θ we have

$$Z=\exp\left\{-\bar{\phi}_\upsilon+\frac{\bar{\phi}_\upsilon^2}{2}+p\log\bar{\phi}_\upsilon-\left(\bar{\phi}_\upsilon-\bar{\phi}_\upsilon^{1/2}\right)\xi-\frac{1}{2}\left(\bar{\phi}_\upsilon-\bar{\phi}_\upsilon^{1/2}\right)^2\right\}\xrightarrow{\mathbf{P}_1}\infty$$

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

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

We saw in the previous chapter that

$$g(x_{min})=\chi\,L^p \qquad (97)$$

where $\chi\xrightarrow{\mathbf{P}_1}1$. Therefore according to equations (96) and (97), uniformly in $\upsilon\in\Gamma_n$,

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

Finally, uniformly in Γ_n

$$\begin{aligned}
\sup_{f\in\mathcal{A}(\upsilon)}\mathbf{E}^{(n)}\left|\psi_\upsilon^{-1}\big(\tilde{f}_n(x)-f(x)\big)\right|^p &= \sup_{f\in\mathcal{A}(\upsilon)}\mathbf{E}^{(n)}\left|\bar{\psi}_\upsilon^{-1}\big(\tilde{f}_n(x)-f(x)\big)\right|^p(1+o(1))\\
&\geq 1+o(1).
\end{aligned}$$

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

Bateman, H. [1953] *Higher Transcendental Functions*, volume II (McGraw-Hill).

Chebyshev, P.L. [1859] "Problems on the least values connected with the approximate representation of functions. (voprosy naimen'shikh velichinakh, svyazannye s priblizhennym predstavleniyem funktsii)," *Sochineniya* **II**, 151–235.

Feller, W. [1968] *An Introduction to Probability Theory and its Applications*, volume I (Wiley, New York), 3rd edition.

Fox, L. & Parker, I.B. [1968] *Chebyshev Polynomials in Numerical Analysis* (Oxford University Press, London).

Golubev, G.K., Lepski, O.V. & Levit, B.Y. [2001] "On adaptive estimation using the *sup*-norm losses," *Math. Meth. Statist.* **10**, 23–37.

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

Ibragimov, I.A. & Has'minskii, R.I. [1981] *Statistical Estimation, Asymptotic Theory* (Springer, New York).

Ibragimov, I.A. & Has'minskii, R.I. [1982] "Bounds for the risks of non-parametric regression estimates," *Theor. Probab. Appl.* **27**, 84–99.

Lepski, O.V. & Levit, B.Y. [1998] "Adaptive minimax estimation of infinitely differentiable functions," *Math. Meth. Statist.* **7**, 123–156.

Levit, B.Y. [1980] "On the asymptotic minimax estimates of the second order," *Theory Prob. Appl.* **25**, 552–568.

Stoer, J. & Bulirsch, R. [1992] *Introduction to Numerical Analysis* (Springer-Verlag), 2nd edition.

Szegö, G. [1975] *Orthogonal Polynomials*, volume XXIII (American Math. Soc.), 4th edition.

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