Chapter 3

Adaptive regression on a bounded interval

In the previous chapter we discussed a regression model in which the unknown function $f(x)$ was assumed to be analytic on the whole real line. This model required that the observations of $f(x)$ were made on the whole line. In practice however observations of the regression function $f(x)$ are often available only on a bounded interval. This is exactly the case which will be discussed in the current chapter. To begin with, we will introduce relative classes $A(\gamma, M)$ of functions analytic in a vicinity of a given interval (Section 3.1.1).

Next, an important issue of the observation design will be highlighted. A seemingly natural, but somewhat naive approach is to use the simplest possible uniform, or equidistant, design. We will see however that such a design loses substantially in accuracy, near the end-points of the observation interval. We will explain that this is not a drawback of any specific method of estimation, but rather an in-built defect of the equidistant design itself.

A much more satisfactory design is the Chebyshev design. In Section 3.2 we describe, pointwisely, the quality of the best attainable accuracy of estimation for both designs. Finally, in Section 3.3 we will present our main results about adaptive estimation of functions $f \in A(\gamma, M)$. Here we restrict ourselves only to the Chebyshev designs, in view of their greater efficiency. Without any loss of generality, we will assume throughout the chapter that our observation interval is $[-1, 1]$; a generalization to an arbitrary bounded interval $[a, b]$ is straightforward; cf. e.g. Timan [1963], Sect. 3.7.

3.1 The building blocks

The purpose of this section is to introduce classes $A(\gamma, M)$ of analytic functions, as well as the Legendre and Chebyshev polynomials. We discuss their properties and the relation between them. Classes $A(\gamma, M)$ 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.

3.1.1 The class $A(\gamma, M)$

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

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

The ellipses $E_\gamma$ represent a convenient family of vicinities of the interval $[-1, 1]$, expanding from $[-1, 1]$ to $\mathbb{C}$, as $\gamma$ increases from 0 to $\infty$. One can verify by simple algebra that the elliptic boundary $\partial E_\gamma$ 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 3.1** We denote by $A(\gamma, M)$ the class of functions analytic inside $E_\gamma$ such that $|f(z)| \leq M$, for all $z \in E_\gamma$. Denote by $\rho_\gamma$ the distance from the interval $[-1, 1]$ to the boundary $\partial E_\gamma$. From the integral Cauchy formula for the $m$th 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 $x \in [-1, 1]$,

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

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

$$|f^{(m)}(x)| \leq M m! / \rho_\gamma^m \quad (3.1)$$

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

$$\rho_\gamma = \cosh \gamma - 1. \quad (3.2)$$

Equations (3.1) and (3.2) will be used later in Section 3.2, in obtaining some discrete-type approximations to analytic functions.

3.1.2 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)

$$P_r(x) = 2^{-r} \sum_{\nu=0}^{r} \binom{r}{\nu} \left( \binom{r}{\nu} \right) (x-1)^\nu (x+1)^{r-\nu}, \quad (3.3)$$

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

$$P_0 \equiv 1,$$
3.1. THE BUILDING BLOCKS

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

In particular, from the definition (3.3), it holds

\[ P_r(1) = 1, \quad P_r(-1) = (-1)^r. \]  \hspace{1cm} (3.4)

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)

\[ |P_r^{(m)}(x)| \leq r^{2m} \max_{-1 \leq x \leq 1} |P_r(x)|, \quad m = 1, 2, \ldots; \]  \hspace{1cm} (3.5)

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

\[ \max_{-1 \leq x \leq 1} |P_r(x)| = |P_r(\pm 1)| = 1. \]  \hspace{1cm} (3.6)

The normalized Legendre polynomials, given by

\[ p_r(x) = (2r + 1)^{1/2}P_r(x), \quad r = 0, 1, \ldots, \]  \hspace{1cm} (3.7)

satisfy, from (3.5)-(3.7),

\[ \max_{-1 \leq x \leq 1} |p_r^{(m)}(x)| \leq (2r + 1)^{1/2}r^{2m} \quad m = 1, 2, \ldots. \]  \hspace{1cm} (3.8)

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

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

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, \( x^n_k, \, k = 1, 2, \ldots, n \), we define the corresponding discrete inner product of the functions \( f \) and \( g \) to be

\[ (f | g) := \frac{1}{n} \sum_{k=1}^{n} f(x^n_k)g(x^n_k). \]

In this subsection, we consider the discrete inner product with respect to the Legendre design, for which \( x^n_k \) represent the equidistant knots

\[ x^n_k = \frac{2k - n - 1}{n}, \quad k = 1, \ldots, n. \]  \hspace{1cm} (3.9)

Let us denote the kernel corresponding to the Legendre family \( p_r \) by

\[ K_N(x, y) := \sum_{r=0}^{N-1} p_r(x)p_r(y). \]  \hspace{1cm} (3.10)

Underlying the quality of our estimators will be remarkable properties of the following type.
Lemma 3.1 Let \( N \in \mathbb{N} \). The normalized Legendre polynomials \( p_r \) satisfy

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

\[
(p_{r_1} | p_{r_2}) = \frac{1}{n} \sum_{k=1}^{n} p_{r_1}(x_k^n) p_{r_2}(x_k^n) = \delta_{r_1 r_2} + O\left(\frac{N^6}{n^2}\right), \quad (n \to \infty). \tag{3.11}
\]

(b) If

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

then

\[
\alpha_N^2(x) = \frac{2}{\pi \sqrt{1 - x^2}} (1 + o(1)), \quad \text{for} \quad N \to \infty, \tag{3.13}
\]

uniformly on any interval \([-1, 1]\), and \( \alpha_N^2(\pm 1) = N \).

Remark 3.1 Note the different behavior of \( \alpha_N \) 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!

The property (b) is illustrated by Figure 3.1.

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

\[
\frac{(b-a)^2}{24n^2} \max_{a \leq x \leq b} \left| \frac{d^2}{dx^2} f(x) \right| \tag{3.14}
\]

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

\[
| (p_{r_1} | p_{r_2}) - (p_{r_1} | p_{r_2}) | 
= \left| \frac{1}{n} \sum_{k=1}^{n} p_{r_1}(x_k^n) p_{r_2}(x_k^n) - \frac{1}{2} \int_{-1}^{1} p_{r_1}(x)p_{r_2}(x)dx \right|
\leq \frac{1}{3n^2} \max_{-1 \leq x \leq 1} \left| \frac{d^2}{dx^2} (p_{r_1}(x)p_{r_2}(x)) \right|. \tag{3.15}
\]

Applying \( L^2 \)-orthonormality and bounds (3.8) for the derivatives of \( p_r(x) \) we get

\[
| (p_{r_1} | p_{r_2}) - \delta_{r_1 r_2} | \leq \frac{1}{3n^2} (2r_1 + 1)(2r_2 + 1)(r_1^2 + r_2^2)^2 = O\left(\frac{N^6}{n^2}\right).
\]
3.1. THE BUILDING BLOCKS

as $n \to \infty$.

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

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

$r \to \infty, |x| < 1$ (3.16)

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,

$$\frac{1}{N} \sum_{r=0}^{N-1} p_r^2(x) = \frac{2}{\pi \sqrt{1-x^2}} \left(1 - \frac{1}{N} \sum_{r=0}^{N-1} \sin((2r+1)\theta) + O(N^{-1} \log N)\right)$$

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

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 (3.8),

\[
|p_r(z)| \leq (2r + 1)^{1/2} e^{\gamma r}
\]

for every \(z \in E_\gamma\).

### 3.1.3 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

\[
T_r(x) = \cos r \arccos x, \quad r = 0, 1, \ldots,
\]

or in the recurrent form

\[
T_0(x) = 1, \\
T_1(x) = x, \\
T_{r+1}(x) = 2xT_r(x) - T_{r-1}(x), \quad r = 1, 2, \ldots.
\]

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

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

i.e. they satisfy \(< t_{r_1} | t_{r_2} > = \delta_{r_1,r_2} \) 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

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

(3.20)

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

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

(3.21)

provides the best approximation to a function $f$, with respect to the weighted $L^2$-norm corresponding to (3.19), among all polynomials of degree less than $N$. The class $A(\gamma, M)$ has the important property that the coefficients of the Chebyshev-Fourier series (3.20) decrease very fast (cf. Timan, Sect. 3.7.3). For all $r = 0, 1, \ldots$, the inequality

$$\sup_{f \in A(\gamma, M)} \|f - t_r\| \leq \sqrt{\pi} Me^{-\gamma r}$$

(3.22)

holds. From (3.21), (3.22) and the bound $|t_r(x)| \leq \sqrt{2}$ it follows that for every $f \in A(\gamma, M)$

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

(3.23)

(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 $A(\gamma, M)$, the approximation $f_N(x)$ based on Chebyshev polynomials is asymptotically also the polynomial of the best uniform approximation on $[-1, 1]$. More precisely,

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

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 $E_\gamma$. From the identity

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

it follows that

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

\[ \{ \omega \in \mathbb{C} : e^{-\gamma} < |\omega| < e^\gamma \} \]

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

\[ |t_r(z)| = \sqrt{2} |T_r(z)| \leq \sqrt{2} e^{\gamma r}. \]  \hfill (3.24)

Denote the discrete inner product by

\[ (f \mid g) := \frac{1}{n} \sum_{k=1}^{n} f(x^n_k)g(x^n_k) \]  \hfill (3.25)

where the points \( x^n_k \) correspond to the Chebyshev design

\[ x^n_k = \cos \left( \frac{2k-1}{2n} \pi \right), \quad k = 1, \ldots, n. \]  \hfill (3.26)

We can state next a lemma which is similar to Lemma 3.1. 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 3.2** The normalized Chebyshev polynomials \( t_r \) satisfy

(a) For any \( r_1, r_2 = 0, 1, \ldots \)

\[ (t_{r_1} \mid t_{r_2}) = \frac{1}{n} \sum_{k=1}^{n} t_{r_1}(x^n_k)t_{r_2}(x^n_k) = \delta_{r_1 r_2}, \]  \hfill (3.27)

(b) If

\[ \beta_N^2(x) := \frac{1}{N} K_N(x, x) = \frac{1}{N} \sum_{r=0}^{N-1} t_r^2(x) \]  \hfill (3.28)

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

\[ \beta_N^2(x) = 1 + \frac{1}{N} \cos(N\theta) \frac{\sin((N-1)\theta)}{\sin \theta} \]

\[ = 1 + \frac{O(1)}{N}, \]  \hfill (3.29)

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

---

\[ ^3 \text{Given the parallel between our work with Legendre and Chebyshev polynomials we decided to duplicate some of the notations, e.g. } x^n_k, \text{ the inner products, the projection operator } K_N, \text{ etc. The reader must just keep in mind whether we are working under the Chebyshev or the Legendre setting.} \]
Remark 3.2 Note the slightly different behavior at the end-points when compared with the inner points. Compare this with Lemma 3.1.

The second property is illustrated in Figure 3.2.

![Figure 3.2: Averaged sum of squared Chebyshev polynomials](image)

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 3.1.

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 $\gamma, M$) 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 is 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 in this chapter.

In Section 3.3 we shall restrict our study to Chebyshev designs, in constructing minimax estimator in the adaptive (unknown $\gamma, M$) setting. Statistical estimation using the uniform
CHAPTER 3. ADAPTIVE REGRESSION ON A BOUNDED INTERVAL

norm as the quality criterion of estimators requires a different approach (cf. Golubev, Lepski and Levit [2001]).

3.2 Minimax regression in $A(\gamma, M)$

3.2.1 The statistical setting

Our observation model in this chapter is given by

$$y_k = f(x_k^n) + \xi_k, \quad k = 1, \ldots, n, \quad (3.30)$$

where the random variables $\xi_k$ are independent identically distributed $\mathcal{N}(0, \sigma^2)$ and the design $x_k^n$ is either Legendre and Chebyshev design. Throughout this chapter the unknown regression function $f$ belongs to $A(\gamma, M)$. In this section we assume that the parameters $\gamma$ 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

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

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

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

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

Let $\hat{f}_n(x) = \hat{f}_n(x, y)$ be an arbitrary estimator of $f(x)$ based on the observation vector $y = (y_1, \ldots, y_n)$, and denote by $P_f$, $E_f$ and $\text{Var}_f$ 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

$$\inf_{\hat{f}_n} \sup_{f \in A(\gamma, M)} E_f w\left(\sigma_n^{-1}(\hat{f}_n(x) - f(x))\right)$$

where $w \in \mathcal{W}$. The parameter $\sigma_n$ defining the minimax rate of convergence, for each of the corresponding designs, Legendre or Chebyshev, will be specified later in Theorems 3.1 and 3.2.

3.2.2 Estimation in the Legendre design

Given the observations $y$ taken at the Legendre knots (3.9), and following the notation introduced in Section 3.1.2, define the estimator

$$\hat{f}_{n,N}(x) = \frac{1}{n} \sum_{k=1}^{n} y_k K_N(x, x_k^n) = \sum_{r=0}^{N-1} \left( \frac{1}{n} \sum_{k=1}^{n} y_k p_r(x_k^n) \right) p_r(x). \quad (3.31)$$
With a slight abuse of the notation, we will write

\[ \hat{f}_{n,N}(x) = (y \mid K_N(x, \cdot)) = \sum_{r=0}^{N-1} (y \mid p_r) p_r(x). \]  

(3.32)

Now consider two auxiliary functions:

\[ f_N(x) = (f \mid K_N(x, \cdot)) = \sum_{r=0}^{N-1} (f \mid p_r) p_r(x), \]  

(3.33)

and

\[ f_{n,N}(x) = (f \mid K_N(x, \cdot)) = \sum_{r=0}^{N-1} (f \mid p_r) p_r(x). \]  

(3.34)

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 3.1** For any \( w \in \mathcal{W} \) and every \( x \in [-1, 1] \)

\[
\lim_{n \to \infty} \sup_{f \in \mathcal{A}(\gamma, M)} E_w \left( \alpha_N^{-1}(x) \sqrt{\frac{n}{\sigma^2 N}} \left( \hat{f}_n(x) - f(x) \right) \right) = \lim_{n \to \infty} \inf_{f \in \mathcal{A}(\gamma, M)} E_w \left( \alpha_N^{-1}(x) \sqrt{\frac{n}{\sigma^2 N}} \left( \hat{f}_n(x) - f(x) \right) \right) = E_w(\xi)
\]

where \( \alpha_N(x) \) is defined in (3.12), \( \hat{f}_n \) is an arbitrary estimator of \( f \), \( \hat{f}_n = \hat{f}_{n,N} \) is the projection estimator (3.32) with

\[
N = N_n := \left\lfloor \frac{1}{2 \gamma} \log n \right\rfloor \quad \text{and} \quad \xi \sim \mathcal{N}(0, 1).
\]

(3.35)

**Proof: the upper bound.** Let \( N \) be given by (3.35). As usual we decompose the mean square error as

\[
E(\hat{f}_{n,N}(x) - f(x))^2 = \text{Var} v_N^2(x) + b_N^2(x)
\]

(3.36)

where, according to (3.32) and (3.34),

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

(3.37)

is a zero-mean stochastic term and

\[
b_N(x) = (f_{n,N}(x) - f_N(x)) + (f_N(x) - f(x))
\]

(3.38)
is the bias.

Let us first analyze the variance of $v_N(x)$. Applying Lemma 3.1(a) we get

$$\text{Var} v_N(x) = \frac{\sigma^2}{n^2} \sum_{k=1}^{n} K_N^2(x, x_k^n) = \frac{\sigma^2}{n^2} \sum_{k=1}^{n} \left( \sum_{r=0}^{N-1} p_r(x) p_r(x_k^n) \right)^2$$

$$= \frac{\sigma^2}{n} \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} p_{r_1}(x) p_{r_2}(x) \frac{1}{n} \sum_{k=1}^{n} p_{r_1}(x_k^n) p_{r_2}(x_k^n)$$

$$= \frac{\sigma^2}{n} \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} p_{r_1}(x) p_{r_2}(x) \left( \delta r_1 r_2 + O \left( \frac{N^6}{n^3} \right) \right)$$

$$= \frac{\sigma^2}{n} \sum_{r=0}^{N-1} p^2_r(x) + O \left( \frac{N^6}{n^3} \right) \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} p_{r_1}(x) p_{r_2}(x).$$

(3.39)

Now, applying the Cauchy-Schwartz inequality we see that

$$\left| \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} p_{r_1}(x) p_{r_2}(x) \right| = \left( \sum_{r=0}^{N-1} p_r(x) \right)^2 \leq N \sum_{r=0}^{N-1} p^2_r(x)$$

$$= N K_N(x, x) = N^2 \alpha^2_N(x).$$

(3.40)

Thus, according to the last two equations and (3.35),

$$\text{Var} v_N(x) = \alpha^2_N(x) \frac{\sigma^2 N}{n} (1 + o(1))$$

(3.41)

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

Now let us consider the bias. First, we have

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

(3.42)

By definition

$$\left| (f | p_r) - \langle f | p_r \rangle \right| = \left| \frac{1}{2} \int_{-1}^{1} f(x) p_r(x) \, dx - \frac{1}{n} \sum_{k=1}^{n} f(x_k^n) p_r(x_k^n) \right|. \quad (3.43)$$

Next, applying (3.14), this difference can be bounded by

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

(3.44)
3.2. MINIMAX REGRESSION IN $\mathcal{A}(\gamma, M)$

Thus, applying the bounds for the derivatives $|f^{(m)}(x)| \leq Mm!\rho^{m}_\gamma$ (cf. Sect. 3.1.1) and $|p^{(m)}_r(x)| \leq (2r + 1)^{1/2}r^{2m}$ (cf. eq. (3.8)), it follows that

$$
| \langle f | p_r \rangle - \langle f | p_r \rangle | \leq \frac{M}{3n^2} ((2\rho) - 2 + 2\rho^{-1}(2r + 1)^{1/2}r^2 + (2r + 1)^{1/2}r^4)
$$

$$
= O\left(\frac{r^5}{n^2}\right), \quad (n \to \infty).
$$

Combining Cauchy-Schwartz inequality with the previous bound and using the fact that $N$ is of order $O(\log n)$, cf. eq. (3.35), we find

$$
(f_{n,N}(x) - f_N(x))^2 \leq \sum_{r=0}^{N-1} \left(\langle f | p_r \rangle - \langle f | p_r \rangle \right)^2 \sum_{r=0}^{N-1} p_r^2(x)
$$

$$
= \alpha_N^2(x) N \sum_{r=0}^{N-1} \left(\langle f | p_r \rangle - \langle f | p_r \rangle \right)^2 = \alpha_N^2(x) O\left(\frac{N^{12}}{n^4}\right)
$$

$$
= \alpha_N^2(x) \frac{\sigma^2 N \log n}{n} O\left(\frac{N^{11}}{n^3}\right) = o(1) \text{Var} \nu_N(x).
$$

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

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

for some constant $C_1 > 0$. According to the Laplace formula (3.16) the polynomials $p_r(x)$ are uniformly bounded, on any interval $[a, b] \subset (-1, 1)$. Thus, from previous inequality, for some $C_2 > 0$,

$$
(f_N(x) - f(x))^2 \leq \left(\sum_{r=\infty}^{N} |\langle f | p_r \rangle||p_r(x)|\right)^2
$$

$$
\leq C_2 e^{-2\gamma N} \sim C_2 n^{-1} = o(1) \text{Var} \nu_N(x).
$$

At the end-points of the interval we have $|p_r(\pm 1)| = (2r + 1)^{1/2}$, see eqs. (3.4) and (3.7), thus for $x = \pm 1$

$$
|f_N(x) - f(x)| \leq C_1 \sum_{r=\infty}^{N} (2r + 1)^{1/2}e^{-\gamma r} \leq C_3 \sum_{r=\infty}^{N} r^{1/2}e^{-\gamma r}
$$

$$
\leq C_3 e^{-\gamma} \int_{N+1}^{\infty} r^{1/2}e^{-\gamma r}dr = C_3 N^{1/2}e^{-\gamma N}(1 + o(1))
$$

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

$$
(f_N(x) - f(x))^2 \leq C_4 Ne^{-2\gamma N} \sim C_4 \frac{N}{n} = o(1) \text{Var} \nu_N(x).
$$
CHAPTER 3. ADAPTIVE REGRESSION ON A BOUNDED INTERVAL

From (3.36), (3.41), (3.46) and (3.47) or (3.48) we can conclude that

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

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

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

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 $f \in A(\gamma, M)$. Therefore using the dominated convergence theorem we obtain the following upper bound:

$$\limsup_{n \to \infty} \sup_{f \in A(\gamma, M)} \mathbb{E} w \left( \alpha_N^{-1}(x) \sqrt{\frac{n}{\sigma^2 N}} (\hat{f}_n(x) - f(x)) \right) = \mathbb{E} w(\xi). \quad (3.49)$$

**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

$$f_\theta(z) = \theta \sqrt{\frac{\sigma^2}{n}} \frac{K_N(x, z)}{\sqrt{K_N(x, x)}} \quad |\theta| < \theta_n = N^{1/2}; \quad (3.50)$$

where we will use

$$\bar{N} = N_n = [N_n - 3 \log N_n], \quad (3.51)$$

see (3.35). Note that $\bar{N}$ is asymptotically equivalent to $N = N_n$ when $N \to \infty$. This implies, according to Lemma 3.1(b), that

$$\frac{\alpha_N^2(x)}{\alpha_{\bar{N}}^2(x)} \to 1, \quad (3.52)$$

uniformly in $[-1, 1]$, when $n \to \infty$. We need the following lemma.

**Lemma 3.3** For a given $x \in [-1, 1]$ and any $z \in E_\gamma$, let $f_\theta(z)$ be defined by (3.50). Then

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

(b) $f_\theta \in A(\gamma, M), \quad |\theta| < \theta_n$, for all $n$ big enough.

(c) The statistic

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

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

(d) The statistic $T$ is sufficient and the log-likelihood $\Lambda := \log \frac{P_\theta}{P_0}(y)$ satisfies

$$\Lambda = \theta T - \frac{\theta^2}{2} T_n$$

where $P_\theta$ and $P_0$ denote the probabilities associated with $f_\theta$ and $f_0$ respectively.

Proof of lemma.

(a) This follows directly from the definitions of $f_\theta$ and $\alpha_N(x)$.

(b) Obviously $f_\theta(z)$ is analytic in the whole complex plane, thus also in $E_\gamma$. Using (3.17), applying the Cauchy-Schwartz inequality and recalling the definition of $\tilde{N} = \tilde{N}_n$, we obtain

$$|f_\theta(z)| \leq \theta_N \sqrt{\frac{\sigma^2}{n}} \left( \frac{K_N^2(x, z)}{K_N(x, x)} \right)^{1/2} \leq \sqrt{\frac{\sigma^2 \tilde{N}}{n}} K_N^{1/2}(z, z) = \sqrt{\frac{\sigma^2 \tilde{N}}{n}} \left( \sum_{r=0}^{\tilde{N}-1} \rho_r^2(z) \right)^{1/2}$$

$$\leq \sqrt{\frac{\sigma^2 \tilde{N}}{n}} \left( \sum_{r=0}^{\tilde{N}-1} (2r + 1) e^{2\gamma r} \right)^{1/2} = O(1) \frac{N\sqrt{n}}{\sqrt{\tilde{N}}} e^{\gamma \tilde{N}} = O(\tilde{N}^{-1/2}) \leq M,$$

in $E_\gamma$ for all $n$ large enough.

(c) Denote

$$T_n = \frac{1}{n} \sum_{k=1}^{n} \frac{K_N^2(x, x_k^n)}{K_N(x, x)}.$$

We can see that $T$ is normally distributed,

$$E T = \frac{1}{\sigma \sqrt{n}} \sum_{k=1}^{n} f_\theta(x_k^n) \frac{K_N(x, x_k^n)}{\sqrt{K_N(x, x)}} = \theta \frac{1}{n} \sum_{k=1}^{n} \frac{K_N^2(x, x_k^n)}{K_N(x, x)} = \theta T_n,$$

and

$$\text{Var} T = \frac{1}{n} \sum_{k=1}^{n} \frac{K_N^2(x, x_k^n)}{K_N(x, x)} = T_n.$$

Thus $T \sim N(\theta T_n, T_n)$. Now let us show that $T_n \to 1$ when $n \to \infty$. Using Lemma 3.1(a)
and the Cauchy-Schwartz inequality, we find that

\[
\mathcal{I}_n = \frac{1}{n} K_N^{-1}(x, x) \sum_{k=1}^{n} K_N^2(x, x^n_k) = \frac{1}{n} K_N^{-1}(x, x) \sum_{k=1}^{n} \left( \sum_{r=0}^{N-1} p_r(x)p_r(x^n_k) \right)^2 \\
= K_N^{-1}(x, x) \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} \left( p_{r_1}(x)p_{r_2}(x) \frac{1}{n} \sum_{k=1}^{n} p_{r_1}(x^n_k)p_{r_2}(x^n_k) \right) \\
= K_N^{-1}(x, x) \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} \left( p_{r_1}(x)p_{r_2}(x) \left( \delta_{r_1 r_2} + O \left( \frac{N^6}{n^2} \right) \right) \right) \\
= K_N^{-1}(x, x) \sum_{r=0}^{N-1} p_r^2(x) + O \left( \frac{N^6}{n^2} \right) K_N^{-1}(x, x) \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} p_{r_1}(x)p_{r_2}(x) \\
= 1 + O \left( \frac{N^6}{n^2} \right) K_N^{-1}(x, x) \left( \sum_{r=0}^{N-1} p_r(x) \right)^2 \\
= 1 + o(1), \quad (n \rightarrow \infty). \quad (3.53)
\]

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

\[
\Lambda = \log \prod_{k=0}^{n-1} \exp \left\{ -\frac{1}{2\sigma^2} \left( y_k - f_{\theta}(x^n_k) \right)^2 + \frac{1}{2\sigma^2} y_k^2 \right\} \\
= -\frac{1}{2\sigma^2} \sum_{k=1}^{n} \left( y_k - f_{\theta}(x^n_k) \right)^2 + \frac{1}{2\sigma^2} \sum_{k=1}^{n} y_k^2 \\
= \theta \frac{1}{\sigma \sqrt{n}} \sum_{k=1}^{n} y_k \frac{K_N(x, x^n_k)}{\sqrt{K_N(x, x)}} = \frac{\theta^2}{2n} \sum_{k=1}^{n} \frac{K_N^2(x, x^n_k)}{K_N(x, x)} \\
= \theta T - \frac{\theta^2}{2} \mathcal{I}_n.
\]

This completes the proof of the lemma. \qed
Now we can continue the proof of the theorem. Given $\alpha_N^2(x) \sim \alpha_N^2(x)$, see eq. (3.52),

$$
\mathcal{R} := \inf_{f_n} \sup_{f \in \mathcal{A}(\gamma, M)} \mathbb{E}_f w\left(\alpha_N^{-1}(x) \frac{n}{\sigma^2 N} (\tilde{f}_n(x) - f(x))\right) \quad (3.54)
$$

$$
= \inf_{f_n} \sup_{f \in \mathcal{A}(\gamma, M)} \mathbb{E}_f w\left(\alpha_N^{-1}(x) \frac{n}{\sigma^2 N} (\tilde{f}_n(x) - f(x))(1 + o(1))\right) \quad (3.55)
$$

$$
\geq \inf_{f_n} \sup_{f \neq \theta_n} \mathbb{E}_f \left((1 + o(1))\alpha_N^{-1}(x) \frac{n}{\sigma^2 N} (\tilde{f}_n(x) - f_\theta(x))\right), \quad (N \to \infty).
$$

Denote $\tilde{\theta} = \alpha_N^{-1}(x) \frac{n}{\sigma^2 N} \tilde{f}_n(x)$. Then applying Lemma 3.3(a)

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

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 (3.54) and (3.55), applying Lemma 3.1(b) and definition (3.51) of $\tilde{N}$ 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

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

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 $w_\delta \not\to w$ when $\delta \to 0$ and see that for any $\delta$

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

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

$$
\inf_{\tilde{\theta}} \sup_{|\theta| \leq \theta_n} \mathbb{E}_{\tilde{\theta}} w_\delta(\tilde{\theta} - \theta) \geq \inf_{\tilde{\theta}} \int_{-\theta_n}^{\theta_n} \mathbb{E}_{\tilde{\theta}} w_\delta(\tilde{\theta} - \theta) \lambda(\theta) d\theta
$$

$$
= \inf_{\theta(T)} \int_{-\theta_n}^{\theta_n} \mathbb{E}_{\tilde{\theta}} w_\delta(\tilde{\theta}(T) - \theta) \lambda(\theta) d\theta
$$

given that $T$ is sufficient for $\theta$, according to Lemma 3.3(c). Applying results presented in Levit [1980], we get that

$$
\inf_{\tilde{\theta}} \sup_{|\theta| \leq \theta_n} \mathbb{E}_{\tilde{\theta}} w_\delta(\tilde{\theta} - \theta) \geq \mathbb{E} w_\delta(\xi) + O(\theta_n^{-2}), \quad (n \to \infty),
$$
where $\xi \sim \mathcal{N}(0, 1)$. Thus $\lim_{n \to \infty} \mathcal{R} \geq \mathbb{E} w_0(\xi)$. Applying the dominate convergence theorem for $\delta \to 0$ we get

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

(3.56)

Finally, from (3.49) and (3.56) the theorem is proved. \hfill \square

**Corollary 3.1** For any $[a, b] \subset (-1, 1)$, uniformly in $x \in [a, b]$,

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

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

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

### 3.2.3 Estimation in the Chebyshev design

Consider now the design given by the Chebyshev knots (3.26). Following the notation of Section 3.1.3 define the estimator

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

(3.57)

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

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

(3.58)

and consider the two functions

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

(3.59)

and

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

(3.60)

see the footnote on page 50 with regards to these notations. Then the following result holds.
3.2. MINIMAX REGRESSION IN $A(\gamma, M)$

**Theorem 3.2** For any $w \in W$ and every $x \in [-1,1]$

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

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

where $\hat{f}_n$ is an arbitrary estimator of $f$, $\hat{f}_n = \hat{f}_{n,N}$ is the projection estimator (3.58) with

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

(3.61)

$\beta_N^2(x)$ is defined by (3.28) and $\xi \sim N(0,1)$.

**Remark 3.3** 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 3.1. 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 3.1(a) and 3.2(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, $\gamma$ 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 3.1, cf. (3.37) and (3.38), we have

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

(3.62)

Let us first analyze the variance of $v_N(x)$. As before (cf. eq. (3.39)), applying Lemma 3.2(a) we obtain

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

(3.63)

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

Now let us consider the bias

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

(3.64)

Using Cauchy-Schwartz inequality we see that

$$\left( f_{n,N}(x) - f_N(x) \right)^2 \leq \sum_{r=0}^{N-1} ((f \mid t_r) - (f \mid t_r))^2 \sum_{r=0}^{N-1} t_r^2(x)$$

$$= N \beta_N^2(x) \sum_{r=0}^{N-1} ((f \mid t_r) - (f \mid t_r))^2.$$
If we rewrite the inner products as
\[
(f | t_r) = \frac{1}{\pi} \sum_{k=1}^{n} f \left( \cos \left( k - \frac{1}{2} \frac{\pi}{n} \right) \cos \left( r \left( k - \frac{1}{2} \frac{\pi}{n} \right) \frac{\pi}{n} \right) \right)
\]
and
\[
(f | t_r) = \frac{1}{\pi} \int_{0}^{\pi} f(\cos \zeta) \cos(r\zeta) d\zeta
\]
(cf. eqs. (3.19) and (3.25)), we can apply the same arguments that we used in (3.43)–(3.45). Using the bounds for the derivatives of \( f \) given in eq. (3.1) we find that
\[
\left| (f | t_r) - \langle f | t_r \rangle \right| \leq \frac{\pi}{24} \left( \frac{n}{\pi} \right)^2 \max_{\zeta} \left| \frac{d^2}{d\zeta^2} f(\cos \zeta) \cos(r\zeta) \right|
\]
\[
\leq \frac{\pi^3}{24 n^2} M \left( r^2 + \frac{2r + 1}{\rho_{\gamma}} + \frac{2}{\rho_{\gamma}^2} \right)
\]
\[
\leq \frac{\pi^3(r + 1)^2}{6 n^2} M \max(1, \rho_{\gamma}^{-1}, \rho_{\gamma}^{-2}) = MC_{\gamma} \frac{(r + 1)^2}{n^2}
\]  
(3.66)

where, using (3.2), one can verify that
\[
C_{\gamma} = O(1 - e^{-\gamma})^{-4},
\]  
(3.67)

both at \( \gamma = 0 \) and \( \gamma = \infty \) and it is bounded when \( \gamma \) is varying in compact subsets of \((0, \infty)\). Thus, both for \( \gamma \to 0 \) and for \( \gamma \to \infty \), uniformly in \( N \)
\[
(f_{n,N}(x) - f_N(x))^2 = \beta_N^2(x) O \left( M^2(1 - e^{-\gamma})^{-8} \frac{N^6}{n^4} \right).
\]  
(3.68)

If we choose \( N = N_n \)
\[
(f_{n,N}(x) - f_N(x))^2 = o(1) \text{Var} v_N(x), \quad (n \to \infty).
\]  
(3.69)

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. (3.47) and (3.48)). Now, one can see that both cases can be considered simultaneously. From (3.23) one can see that for any \( x \) and \( N = N_n \)
\[
(f_N(x) - f(x))^2 \leq 2\pi M^2(1 - e^{-\gamma})^{-2} e^{-2\gamma N} = O \left( \frac{1}{\gamma_N} \right)
\]  
(3.70)
\[
= \beta_N^2(x) \sigma_N^2 O \left( \frac{1}{\gamma_N} \right) = o(1) \text{Var} v_N(x),
\]  
(3.71)
when \( n \to \infty \). From (3.62)–(3.64), (3.69) and (3.71) we have proved that

\[
E(\hat{f}_{n,N}(x) - f(x))^2 = \beta^2_N(x) \frac{\sigma^2 N}{n} (1 + o(1)), \quad (n \to \infty),
\]

which holds uniformly on \([-1, 1]\). It follows that

\[
\beta^{-1}_N(x) \sqrt{\frac{n}{\sigma^2 N}} (\hat{f}_{n,N}(x) - f(x))
\]

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:

\[
\lim_{n \to \infty} \sup_{f \in \mathcal{A}(\gamma, M)} E_w \left( \beta^{-1}_N(x) \sqrt{\frac{n}{\sigma^2 N}} (\hat{f}_{n,N}(x) - f(x)) \right) = E(w(\xi)). \quad (3.72)
\]

**Proof of the lower bound for the risk.** We can follow the same proof of the lower bound we did in Theorem 3.1. For fixed \( x \in [-1, 1] \) and any \( z \in \mathbb{C} \) consider again the parametric sub-family of functions

\[
f_\theta(z) = \theta \sqrt{\frac{\sigma^2}{n} \frac{K_N(x, z)}{K_N(x, x)}} \quad |\theta| < \theta_n = \bar{N}^{1/2} \quad (3.73)
\]

where \( K_N \) is now defined in terms of the Chebyshev polynomials and

\[
\bar{N} = \bar{N}_n = \lfloor N_n - 3 \log N_n \rfloor \quad (3.74)
\]

(cf. definition of \( N_n \) in eq. (3.61)).

**Lemma 3.4** The following properties are satisfied for any \( x \in [-1, 1] \):

(a) \( f_\theta(x) = \theta \beta_N(x) \sqrt{\frac{\sigma^2 N}{n}} \).

(b) \( f_\theta \in \mathcal{A}(\gamma, M), \quad |\theta| < \theta_n, \text{ for } n \text{ big enough.} \)

(c) The statistic

\[
T = \frac{1}{\sigma \sqrt{n}} \sum_{k=1}^{n} y_k \frac{K_N(x, x^n)}{\sqrt{K_N(x, x)}}
\]

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

\[
T = \theta + \xi \quad (3.75)
\]

where \( \xi \sim \mathcal{N}(0, 1) \).
(d) The statistic $T$ is sufficient and the log-likelihood satisfies
\[
\Lambda := \log \frac{dP_\theta}{dP_0} = \theta T - \frac{\theta^2}{2}.
\] (3.76)

where $P_\theta$ and $P_0$ denote the probabilities associated with $f_\theta$ and $f_0$ respectively.

Proof of the lemma. The proof is the same as that of Lemma 3.3. Nevertheless, a couple of remarks can be made. First, the bound (3.17) for Legendre polynomials is also a bound for the Chebyshev polynomials, thus the proof of (b) remains the same. Second, in the present case, $I_n = 1$ given exact orthogonality of Chebyshev polynomials (cf. eq. (3.53)). The rest of the proofs of the lemma and the theorem remain the same and we get
\[
\liminf_{n \to \infty} \inf_{f_n} \sup_{f \in \mathcal{A}(\gamma, M)} E_f \left( \frac{n}{\sigma^2 N_n} \left( \tilde{f}_n(x) - f(x) \right) \right) \geq E w(\xi). 
\] (3.77)

The theorem follows from (3.72) and (3.77).

Corollary 3.2 For any $[a, b] \subset (-1, 1)$ uniformly in $x \in [a, b]$
\[
\lim_{n \to \infty} \sup_{f \in \mathcal{A}(\gamma, M)} E_f \left( \sqrt{\frac{n}{\sigma^2 N_n}} (\tilde{f}_n(x) - f(x)) \right) = 
\]
\[
\liminf_{n \to \infty} \sup_{f \in \mathcal{A}(\gamma, M)} E_f \left( \sqrt{\frac{n}{\sigma^2 N_n}} (\tilde{f}_n(x) - f(x)) \right) = E w(\xi)
\]

where $\tilde{f}_n$ and $\hat{f}_n$ are as in the previous Theorem. For $x = \pm 1$,
\[
\lim_{n \to \infty} \sup_{f \in \mathcal{A}(\gamma, M)} E_f \left( \sqrt{\frac{n}{2\sigma^2 N_n}} (\hat{f}_n(x) - f(x)) \right) = 
\]
\[
\liminf_{n \to \infty} \sup_{f \in \mathcal{A}(\gamma, M)} E_f \left( \sqrt{\frac{n}{2\sigma^2 N_n}} (\tilde{f}_n(x) - f(x)) \right) = E w(\xi)
\]

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; in particular 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.
3.2.4 Estimation for non-fixed classes

In order to create an adaptive framework we follow the same procedure as in the previous chapter. This procedure 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 \( \gamma \) and \( M \) – to 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 \( n \). As we saw in the previous chapter, 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 3.2. The dependence of \( N_n \) on \( n \) comes also from the parameters \( \gamma, M \) in the present situation. Nevertheless, the statement of Theorem 3.2 will still hold provided the appropriate assumptions are fulfilled.

**Theorem 3.3** Let \( w \in \mathcal{W} \), \( \gamma = \gamma_n \), \( M = M_n \) and let \( N = N_n \) be as defined in (3.61). If the following conditions are satisfied

\[
\lim_{n \to \infty} \gamma N = \infty, \tag{3.78}
\]
\[
\lim_{n \to \infty} M^2 (1 - e^{-\gamma})^{-8} N^5 n^{-3} = 0, \tag{3.79}
\]
\[
\lim_{n \to \infty} N = \infty, \tag{3.80}
\]

then

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

for all \( x \in [-1, 1] \). Here \( \tilde{f}_n \) is an arbitrary estimator of \( f \) and \( \hat{f}_n = \hat{f}_{n,N} \) is the projection estimator (3.58).

**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 3.2, and consists in checking that the conditions (3.78) and (3.79) guarantee asymptotic unbiasedness of the optimal estimator (cf. eqs. (3.68) and (3.71)), while (3.80) allow us to prove the lower
bound result. The rest of the proof is the same.

Though conditions (3.78)–(3.80) are sufficient to prove optimality results in non-fixed classes, it may be more convenient to express them explicitly in terms of the parameters \( \gamma \) and \( M \), as is done in the following theorem.

**Theorem 3.4** Let \( w \in W \) and the parameters \( \gamma = \gamma_n \) and \( M = M_n \) be such that

\[
\limsup_{n \to \infty} \frac{M^2}{\log n} = 0, \tag{3.81}
\]

\[
\liminf_{n \to \infty} M^2 \log n = \infty, \tag{3.82}
\]

\[
\limsup_{n \to \infty} \frac{\gamma}{\log \log n} = 0, \tag{3.83}
\]

\[
\liminf_{n \to \infty} \gamma \log n = \infty, \tag{3.84}
\]

then, with \( N = N_n \) defined by (3.61),

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

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

for all \( x \in [-1, 1] \). Here \( \hat{f}_n \) is an arbitrary estimator of \( f \) and \( \hat{f}_n = \hat{f}_{n,N} \) is the projection estimator (3.58).

**Proof.** In order to prove the theorem, we only need to verify that hypothesis of the Theorem 3.3 are satisfied, i.e. we just need to assure that the limits (3.78)–(3.80) are still valid (cf. eqs. (3.68) and (3.71)). If \( \gamma \) and \( M \) are bounded then trivially (3.78)–(3.80) 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 (3.82) and (3.83), 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 \( \gamma N \) and \( N \) go to infinity. Using (3.81) and (3.84)

\[
M^2 (1 - e^{-\gamma})^{-8} N^5 n^{-3} = O \left( M^2 \gamma^{-13} n^{-3} \log^5 (M^2 \gamma^{-1} n) \right)
\]

\[
= O \left( n^{-3} \log^{14} n \log^5 (n \log^2 n) \right) = o(1).
\]
3.3. ADAPTIVE MINIMAX REGRESSION

Case $\gamma \to \infty$: Applying (3.82) and (3.83)

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

thus $N$ and $\gamma N$ go to infinity. From (3.81) and (3.84)

$$M^2 (1 - e^{-\gamma})^{-8} N^5 n^{-3} = O \left( M^2 \gamma^{-5} n^{-3} \log^5 (M^2 \gamma n) \right)$$

$$= O \left( n^{-3} \log n \log^5 (n \log n) \right) = o(1), \quad (n \to \infty).$$

Thus the theorem is proved. \qed

3.3 Adaptive minimax regression

3.3.1 Adaptive estimation in functional scales

In the previous section we described asymptotically minimax estimators for classes $A(\gamma, M)$ where the parameters $\gamma$ 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. We follow here the same procedure that we used in the previous chapter in order to create the adaptive framework in a situation where $\gamma$ and $M$ are unknown.

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

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

corresponding to the parameter class $\Gamma_n$. As our scales $A_{\Gamma_n}$ can be identified with corresponding subsets $\Gamma_n$, we will speak sometimes about a scale $\Gamma_n$, instead of $A_{\Gamma_n}$, 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 $A_{\Gamma_n}$ be a functional scale, and $F$ a class of estimators $\hat{f}_n$, both possibly depending on $n$.

**Definition 3.2** An estimator $\hat{f}_n \in F$ is called $(p, \Gamma_n, F)$-adaptively minimax, at a point $x \in \mathbb{R}$, if for any other estimator $\tilde{f}_n \in F$

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

As it was discussed in the previous chapter, this property depends crucially on which classes $\Gamma_n$ and $F$ are considered. The rate of convergence in estimating $f(x)$ over the whole scale $A_{\mathbb{R}^2_+}$ 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 3.3** A functional scale $A_{\Gamma_n}$ (or the corresponding scale $\Gamma_n$) is called a regular, or an R-scale if the condition

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

where $N_n(v)$ was defined in (3.61), is satisfied.

The previous condition is aimed to guarantee that the approximation arguments which were used in (3.68) and (3.69) are still applicable. Let us remark that in this condition the powers of the terms are not so relevant as far as we have $N_n(v)$ of orden $\log n$ at most.

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

**Definition 3.4** A functional scale $A_{\Gamma_n}$ (a scale $\Gamma_n$) 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

$$\limsup_{n \to \infty} \sup_{v \in \Gamma_n} M \leq M_+, \quad \text{and}$$

$$\limsup_{n \to \infty} \sup_{v \in \Gamma_n} \gamma^{-1} \log n \leq C_+.$$  \hspace{1cm} (3.86)

Regular-pseudo-parametric scales are regular, in the sense of Definition 3.3, and uniformly on them, we have parametric rates, i.e. the rate $n^{-1/2}$ is achieved given

$$\limsup_{n \to \infty} \sup_{v \in \Gamma_n} N_n(v) < \infty.$$  \hspace{1cm} (3.88)

**Definition 3.5** A functional scale $A_{\Gamma_n}$ (a scale $\Gamma_n$) is called a regular-non-parametric, or RNP-functional scale (regular-non-parametric, or RPP-scale) if

$$\limsup_{n \to \infty} \sup_{v \in \Gamma_n} \frac{M^2}{\log n} = 0,$$  \hspace{1cm} (3.89)

$$\liminf_{n \to \infty} \inf_{v \in \Gamma_n} M^2 \log n = \infty,$$  \hspace{1cm} (3.90)

$$\liminf_{n \to \infty} \sup_{v \in \Gamma_n} \frac{\gamma}{\log \log n} = 0,$$  \hspace{1cm} (3.91)

$$\liminf_{n \to \infty} \inf_{v \in \Gamma_n} \gamma \log n = \infty.$$  \hspace{1cm} (3.92)
Note that conditions for regular-non-parametric scales require that the assumptions of Theorem 3.4 hold uniformly on RNP-scales. Thus, according to the proof of Theorem 3.4, the conditions of Theorem 3.3 also hold uniformly in RNP-scales; in particular

$$\liminf_{n \to \infty} \inf_{v \in \Gamma_n} N_n(v) = \infty. \quad (3.93)$$

Also notice that regular-non-parametric scales are regular, in the sense of Definition 3.3.

Let \( F_p = F_p(x) \) be the class of all estimators \( \tilde{f}_n \) that satisfy

$$\limsup_{n \to \infty} \sup_{v \in \Gamma_n} \sup_{f \in A(v)} E_f\left[ n^{1/2} (\tilde{f}_n(x) - f(x)) \right]^p < \infty$$

for any RPP-functional scales \( A_{\Gamma_n} \) and let \( F_p^0 = F_p^0(x) \) be the class of all estimators such that

$$\limsup_{n \to \infty} E_{0} \left[ n^{1/2} \tilde{f}_n(x) \right]^p < \infty.$$

One can see that \( F_p \subset F_p^0 \), since \( f \equiv 0 \) belongs to any of the classes \( A(\gamma, M) \). Below we present an adaptive estimator \( \hat{f}_n \in F_p \) 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 \( F_p^0 \). Finally we shall conclude that our adaptive estimator is \((p, \Gamma_n, F_p)\)-adaptive minimax for RNP-functional scales.

### 3.3.2 Upper bound on the quality of adaptive estimators

**Theorem 3.5** 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 \( A_{\Gamma_n} \), \( \hat{f}_n \in F_p \) and

$$\limsup_{n \to \infty} \sup_{v \in \Gamma_n} \sup_{f \in A(v)} E_f \left[ \psi_n^{-1}(v) \left( \frac{\tilde{f}_n(x)}{f(x)} \right) \right]^p \leq 1.$$

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 (3.61) for any \( v \in \Gamma_n \).

**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 \( N_0 = 0 \), \( N_i = \lceil \exp(i^l) \rceil \) for \( i = 1, 2, \ldots \). Two consecutive elements of this sequence satisfy

$$N_{i+1} - N_i \sim l (\log N_i)^{1-\frac{1}{l}} N_i \to \infty \quad (i \to \infty) \quad (3.94)$$

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

$$\frac{N_{i+1}}{N_i} \sim e^{l_1 i^l - 1} \quad (i \to \infty). \quad (3.95)$$
For each \( n \) we will consider the subsequence \( S_n = \{N_0, N_1, \ldots, N_n\} \), where
\[
I_n = \arg \max_i \{N_i \leq n^{1/2}\}.
\] (3.96)

Since for any \( \delta, \ (0 < \delta < 1/2) \) and for \( n \) large enough, \( N_n(v) \leq n^{1/2-\delta} \) for all \( v \) in any RPP scale as well as any RNP scales, one can always find \( i(v) \leq I_n \) such that
\[
N_{i(v)-1} < N_n(v) \leq N_{i(v)}.
\] (3.97)

Let us denote
\[
\hat{f}_i(x) = f_{n,N_i}(x), \quad b_i = \mathbb{E}_f \hat{f}_i(x) - f(x),
\]
\[
\sigma_i^2 = \text{Var}_f \hat{f}_i(x), \quad \hat{\sigma}_i^2 = \beta_{N_i}^2(x) \frac{\sigma^2 N_i}{n},
\]
\[
\sigma_{i,j}^2 = \text{Var}_f (\hat{f}_j(x) - \hat{f}_i(x)), \quad \hat{\sigma}_{i,j}^2 = \hat{\sigma}_j^2 - \hat{\sigma}_i^2,
\]
and define the sequence of thresholds
\[
\lambda_j^2 = p \log N_j + p_1 \log^\delta N_j.
\]

**Adaptive procedure.** Define
\[
\hat{i} = \min \left\{ 1 \leq i \leq I_n : |\hat{f}_j(x) - \hat{f}_i(x)| \leq \lambda_j \hat{\sigma}_{i,j} \quad \forall j \ (i \leq j \leq I_n) \right\}.
\]

We will prove that the estimator
\[
\hat{f}_n(x) = \hat{f}_{\hat{i}}(x)
\]
satisfies Theorem 3.5. First, however, we derive some inequalities which are necessary for the proof.

**Lemma 3.5** Using the previous notation, uniformly with respect to \( v \) in any RPP or RNP-scale, and uniformly with respect to \( 1 \leq i, j \leq I_n \), as \( n \to \infty \),

\( a \) \( b_j^2 = o(1) \hat{\sigma}_j^2 \quad \text{for all } j \text{ such that } i(v) \leq j \leq I_n; \)

\( b \) \( \sigma_j^2 = \hat{\sigma}_j^2 \quad \text{for all } j; \)

\( c \) \( (b_j - b_i)^2 = O(1) \hat{\sigma}_{i,j}^2 \quad \text{for all } i, j \text{ such that } i(v) \leq i \leq j \leq I_n; \)

\( d \) \( \sigma_{i,j}^2 = \hat{\sigma}_{i,j}^2 \quad \text{for all } i, j. \)
Proof of lemma. (a) As we saw before

\[ b_j^2 \leq 2 \left( f_{n,N_j}(x) - f_{N_j}(x) \right)^2 + 2 \left( f_{N_j}(x) - f(x) \right)^2. \]

From equations (3.68), (3.96), and conditions for RPP scales, or as well, conditions for RNP-scales (cf. Definitions 3.4 and 3.5), we have

\[ (f_{n,N_j}(x) - f_{N_j}(x))^2 \leq \beta_{N_j}^2(x) \frac{\sigma^2 N_j}{n} O \left( M^2 (1 - e^{-\gamma})^{-8} N_j^5 n^{-3} \right) \]
\[ \leq \beta_{N_j}^2(x) \frac{\sigma^2 N_j}{n} O \left( M^2 (1 - e^{-\gamma})^{-8} n^{-1/2} \right) = o(1) \hat{\sigma}_j^2. \]

From (3.71),

\[ (f_{N_j}(x) - f(x))^2 \leq 2\pi M^2 (1 - e^{-\gamma})^{-2} e^{-2\gamma N_j} \leq 2\pi M^2 (1 - e^{-\gamma})^{-2} e^{-\gamma N_n} \]
\[ = O \left( \frac{1}{\gamma n} \right) = O \left( \frac{1}{\gamma N_j} \right) \hat{\sigma}_j^2. \]

In RPP-scales \( \gamma \) goes to infinity uniformly, thus \( \gamma N_j \) goes to infinity uniformly for all \( N_j \geq N_1 \). In RNP-scales \( \gamma N_j \geq \gamma N_n \to \infty \). Thus

\[ (f_{N_j}(x) - f(x))^2 = o(1) \hat{\sigma}_j^2, \]

as \( n \to \infty \). Thus from previous equations we have that \( b_j^2 = o(1) \hat{\sigma}_j^2 \) for all \( j \geq i(v) \), uniformly in RPP- as well as RNP-functional scales.

(b) From (3.63), taking \( N = N_j \), we obtain

\[ \sigma_j^2 = \text{Var} \hat{f}_j(x) = \beta_{N_j}^2(x) \frac{\sigma^2 N_j}{n} = \hat{\sigma}_j^2. \]

(c) We have

\[ (b_j - b_i)^2 = (f_{n,N_j}(x) - f_{n,N_i}(x))^2 \]
\[ \leq 2((f_{n,N_j}(x) - f_{N_j}(x)) - (f_{n,N_i}(x) - f_{N_i}(x)))^2 \]
\[ + 2(f_{N_j}(x) - f_{N_i}(x))^2 \]
\[ := 2 b_1^2(x) + 2 b_2^2(x). \]

Now,

\[ b_1 = (f_{n,N_j}(x) - f_{N_j}(x)) - (f_{n,N_i}(x) - f_{N_i}(x)) = \sum_{r=N_i}^{N_j-1} ((f \mid t_r) - (f \mid t_r)) t_r(x). \]
Applying the Cauchy-Schwartz inequality, (3.66) and (3.67) we see that, as we did in (a),

\[ b_1^2 = O\left(M^2(1 - e^{-\gamma})^{-8}N_j^5n^{-4}\right)\left(\sum_{r=0}^{N_j-1} t_r^2(x) - \sum_{r=0}^{N_i-1} t_r^2(x)\right) \]

\[ = O\left(M^2(1 - e^{-\gamma})^{-8}n^{-1/2}\right) \left( \beta_{N_j}^2(x)\frac{\sigma^2N_j}{n} - \beta_{N_i}^2(x)\frac{\sigma^2N_i}{n} \right) \]

\[ = o(1) (\hat{\sigma}_j^2 - \hat{\sigma}_i^2), \quad (n \to \infty). \]

Also, applying the Cauchy-Schwartz inequality,

\[ b_2^2 \leq \left( \sum_{r=0}^{N_j-1} |(f | t_r) \ | | t_r(x)| \right)^2 \leq \sum_{r=0}^{N_j-1} |(f | t_r)|^2 \sum_{r=0}^{N_i-1} t_r^2(x), \]

where using (3.22), the definition (3.61) of \( N_n \) and condition (3.92) one can verify that

\[ \sum_{r=N_n}^{\infty} |(f | t_r)|^2 = O\left(M^2 e^{-2\gamma N_n} \right) = O\left(\frac{(1 - e^{-\gamma})^2}{\gamma(1 - e^{-2\gamma})} \right) \frac{1}{n} = O(1) \frac{1}{n}. \]

Now,

\[ b_2^2 = O(1) \frac{1}{n} \left( \sum_{r=0}^{N_j-1} t_r^2(x) - \sum_{r=0}^{N_i-1} t_r^2(x) \right) \]

\[ = O(1) \left( \beta_{N_j}^2(x)\frac{\sigma^2N_j}{n} - \beta_{N_i}^2(x)\frac{\sigma^2N_i}{n} \right) \]

\[ = O(1) (\hat{\sigma}_j^2 - \hat{\sigma}_i^2), \quad (n \to \infty). \]

Thus

\[ (b_j - b_i)^2 = O(1) (\hat{\sigma}_j^2 - \hat{\sigma}_i^2) \]

for any \( x \in [-1, 1] \), when \( n \to \infty \).

(d) Applying again the Cauchy-Schwartz inequality together with Lemma 3.2(a) we see that

\[ \text{Var}(\hat{f}_j(x) - \hat{f}_i(x)) = \frac{\sigma^2}{n^2} \sum_{k=1}^{n} (K_{N_j}(x, x_k^u) - K_{N_i}(x, x_k^u))^2 \]

\[ = \frac{\sigma^2}{n} \sum_{r_1=N_1}^{N_j-1} \sum_{r_2=N_i}^{N_j-1} \left( t_{r_1}(x)t_{r_2}(x) \frac{1}{n} \sum_{k=1}^{n} t_{r_1}(x_k^u)t_{r_2}(x_k^u) \right) \]
3.3. ADAPTIVE MINIMAX REGRESSION

\[ \sigma_n^2 = \frac{\sigma^2}{n} \sum_{r_1 = N_i}^{N_j} \sum_{r_2 = N_i}^{N_j} t_{r_1}(x) t_{r_2}(x) \delta r_1 r_2 = \frac{\sigma^2}{n} \sum_{r = N_i}^{N_j} t_r^2(x) \]

\[ \hat{\sigma}^2_\delta \geq \hat{\sigma}^2_i \]

**Proof of the theorem.** For arbitrary scales of parameters \( \Gamma_n \) and for any \( f \in \mathcal{A}(v) \) for some \( v \in \Gamma_n \),

\[
R_n(f) = \mathbf{E} \left| \hat{f}_i(x) - f(x) \right|^p
\]

\[
= \mathbf{E} \left\{ \mathbb{1}_{\{i \leq i(v)\}} \left| \hat{f}_i(x) - f(x) \right|^p \right\} + \mathbf{E} \left\{ \mathbb{1}_{\{i > i(v)\}} \left| \hat{f}_i(x) - f(x) \right|^p \right\}
\]

\[
\leq R_n^-(f) + R_n^+(f).
\]

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

\[
\left\{ i \leq i(v) \right\} \subset \left\{ \left| \hat{f}_i(x) - \hat{f}_i(v) \right| \leq \hat{\sigma}_i \lambda_i \right\}
\]

\[
\subset \left\{ \left| \hat{f}_i(x) - \hat{f}_i(v) \right| \leq \hat{\sigma}_i \lambda_i \right\},
\]

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

\[
R_n^-(f) \leq \mathbf{E} \left\{ \mathbb{1}_{\{i \leq i(v)\}} \left( \left| \hat{f}_i(x) - \hat{f}_i(v) \right| + \left| \hat{f}_i(v) - f(x) \right| \right)^p \right\}
\]

\[
\leq \mathbf{E} \left( \hat{\sigma}_i \lambda_i + \left| b_i(v) \right| + \sigma_i \xi \right)^p
\]

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 \( \Gamma_n \), 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 3.5 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 (3.98) we have that for any RPP-scale, uniformly,

\[
\sup_{f \in \mathcal{A}(v)} R_n^-(f) = O(n^{-p/2}), \quad (n \to \infty).
\]

(3.100)
From (3.98), applying Lemma 3.5, the dominated convergence theorem and asymptotic (3.95), uniformly in any RNP-scale
\[
\sup_{f \in \mathcal{A}(v)} R_n^-(f) \leq \psi_n^p(v) (1 + o(1)), \quad (n \to \infty).
\] (3.101)

Now let us examine \( R_n^+(f) \). Consider the auxiliary event
\[
A_i = \left\{ \omega : |\hat{f}_i(x) - f(x)| \leq \sqrt{2} \hat{\sigma}_i \lambda_i \right\}.
\]

Applying the H"older and Cauchy-Schwartz inequalities we obtain
\[
R_n^+(f) = \mathbb{E} \left\{ \mathbb{I}_{\{i > i(v)\}} |\hat{f}_i(x) - f(x)|^p \right\} = \sum_{i=i(v)+1}^{I_n} \mathbb{E} \left\{ |\hat{f}_i(x) - f(x)|^p \mathbb{I}_{\{i=i\} \cap A_i} \right\}
\]
\[
\leq \sum_{i=i(v)+1}^{I_n} \mathbb{E} \left\{ |\hat{f}_i(x) - f(x)|^p \mathbb{I}_{\{i=i\} \cap A_i} \right\} + \sum_{i=i(v)+1}^{I_n} \mathbb{E} \left\{ |\hat{f}_i(x) - f(x)|^p \mathbb{I}_{A_i} \right\}
\]
\[
:= R_{n,1}^+(f) + R_{n,2}^+(f).
\]

where
\[
R_{n,1}^+(f) = \sum_{i=i(v)+1}^{I_n} (2\hat{\sigma}_i^2 \lambda_i^2)^{p/2} \mathbb{P}(i = i)
\]
and
\[
R_{n,2}^+(f) = \sum_{i=i(v)+1}^{I_n} \mathbb{E}^{1/2} |\hat{f}_i(x) - f(x)|^{2p} \mathbb{P}^{1/2}(A_i^c).
\]

We have that
\[
\mathbb{P}(i = i) \leq \mathbb{P}(i \geq i)
\]
\[
\leq \sum_{j=i+1}^{\infty} \mathbb{P} \left( \left| \hat{f}_{j-1}(x) - \hat{f}_{i-1}(x) \right| > \hat{\sigma}_{i-1,j-1} \lambda_{j-1} \right),
\] (3.102)

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 3.5, (c) and (d), and a well known bound for the tails of the normal distribution (cf. Feller [1968], Lemma 2) we find that
\[
\mathbb{P}( \left| \hat{f}_{j}(x) - \hat{f}_{i}(x) \right| > \hat{\sigma}_{i,j} \lambda_j ) \leq \mathbb{P} \left( \left| \xi \right| > \lambda_j - \frac{|b_j - b_i|}{\hat{\sigma}_{i,j}} \right)
\]
\[
\leq \exp \left\{ -\frac{1}{2} (\lambda_j - C_1)^2 \right\} \leq \exp \left\{ -\frac{1}{2} \lambda_j^2 + C_1 \lambda_j \right\},
\]
for some $C_1 > 0$ and $n$ large enough. Returning to (3.102) we obtain that
\[
P(\hat{i} \geq i) \leq \sum_{j=i+1}^{\infty} \exp \left\{ -\frac{1}{2} \lambda_{j-1}^2 + C_1 \lambda_{j-1} \right\} = \sum_{j=i}^{\infty} \exp \left\{ -\frac{1}{2} \lambda_j^2 + C_1 \lambda_j \right\}
\]
\[
= \sum_{j=i}^{\infty} \exp \left\{ -\frac{p_j t^1}{2} + C_1 \sqrt{p_j t^1 + p_1 t^1} \right\}
\]
\[
\leq \sum_{j=i}^{\infty} \exp \left\{ -\frac{p_j t^1}{2} - \frac{p_1 t^1}{3} \right\} \sim \frac{2}{pl} i^{-1} \exp \left\{ -\frac{p_1 t^1}{4} \right\}
\]
\[
= \frac{2}{pl} i^{-1} N_{i-1}^{-p/2} \exp \left\{ -\frac{p_1 t^1}{3} \right\} \leq C_2 N_i^{-p/2} \exp \left\{ -\frac{p_1 t^1}{4} \right\}
\] (3.103)

for some $C_2 > 0$ and all $i \geq i(v)$, when $n$ is sufficiently large. Therefore uniformly in $\Gamma_n$
\[
\sup_{f \in \mathcal{A}(v)} R^+_i(f) = O(n^{-p/2}) \sum_{i=1}^{\infty} i^{p/2} \exp \left\{ -p_1 t^1 / 4 \right\} = O(n^{-p/2}),
\] (3.104)
when $n \to \infty$. In order to bound $R^+_i(f)$ note that $\hat{f}_i - f(x) = b_i + \sigma_i \xi$, $\xi \sim \mathcal{N}(0, 1)$. Then applying Lemma 3.5, (a) and (b), in the same way as before, we have
\[
P(A_i^c) \leq \mathbb{P} \left( |\xi| > \sqrt{2} \lambda_i - \frac{|b_i|}{\sigma_i} \right) \leq \mathbb{P} \left( |\xi| > \sqrt{2} \lambda_i - \sqrt{2} \right)
\]
\[
\leq \exp \left\{ -\frac{1}{2} \left( \sqrt{2} \lambda_i - \sqrt{2} \right)^2 \right\} \leq \exp \left\{ -\lambda_i^2 + 2 \lambda_i \right\}
\]
\[
\leq \exp \left\{ -p_i t^1 - p_1 t^1 / 2 \right\} \sim N_i^{-p} \exp \left\{ -p_1 t^1 / 2 \right\},
\]
for all $i \geq i(v)$, $n$ large enough. Thus, applying again Lemma 3.5, (a) and (b), and previous bound
\[
R^+_{i,2}(f) = \sum_{i=i(v)+1}^{I_n} E^{1/2} \left| \hat{f}_i(x) - f(x) \right|^{2p} P^{1/2}(A_i^c)
\]
\[
\leq \sum_{i=i(v)+1}^{I_n} \sigma_i^2 \left\| \hat{f}_i \right\| \left\| f \right\|^{2p} P^{1/2}(A_i^c)
\]
\[
= O \left( \beta_{i,2}^2 \frac{\sigma_i^2}{n} \right) \sum_{i=1}^{\infty} \exp \left\{ -p_1 t^1 / 4 \right\}
\]

and finally
\[
\sup_{f \in \mathcal{A}(v)} R^+_{i,2}(f) = O(n^{-p/2}).
\] (3.105)
Finally we can conclude from (3.100), (3.101), (3.104) and (3.105) that \( \hat{f}_n \in \mathcal{F}_p(x) \) and

\[
\limsup_{n \to \infty} \sup_{v \in \Gamma_n} \sup_{f \in A(v)} E \left| \psi_n^{-1}(v)(f_n(x) - f(x)) \right|^p \leq 1,
\]

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

### 3.3.3 Lower bound

**Theorem 3.6** Let \( p > 0 \). Let \( A_{\Gamma_n} \) be an arbitrary RNP-functional scale. For each \( v \in \Gamma_n \), define

\[
\psi_n(v) = \sigma_n(v) \phi_n(v)
\]

where

\[
\sigma_n^2(v) = \beta_n^2(x) \frac{\sigma^2 N_n}{n}, \quad \phi_n^2(v) = p \log N_n,
\]

and \( N_n \) is the same as in Theorem 3.5. Then, for any estimator \( \hat{f}_n \in \mathcal{F}_p^0(x) \)

\[
\liminf_{n \to 0} \inf_{v \in \Gamma_n} \sup_{f \in A(v)} E \left| \psi_n^{-1}(v)(\hat{f}_n(x) - f(x)) \right|^p \geq 1.
\]

**Proof.** This proof is similar to the proof of Theorem 2.5 in Ch. 2. Denote for shortness \( \psi = \psi_n(v), \phi = \phi_n(v) \) and \( \sigma = \sigma_n(v) \). Choose \( \tilde{N} \) as it was defined in (3.74), and define \( \tilde{\psi}_v = \tilde{\sigma} \phi_v \) where

\[
\tilde{\sigma}_v^2 = \beta_n^2(x) \frac{\sigma^2 \tilde{N}}{n} \quad \text{and} \quad \tilde{\phi}_v^2 = p \log \tilde{N}.
\]

Define \( f_0 = 0 \) and \( f_1 = f_\theta \) for \( \theta = \phi_v - \phi_v^{-1/2} \), where \( f_\theta \) belongs to the parametric family defined in (3.73). Notice that \( |\theta| < \tilde{N}^{1/2} \) for all \( n \) big enough. According to Lemma 3.4, \( f_1 \in A(v) \) and

\[
f_1(x) = \theta \beta_N(x) \sqrt{\frac{\sigma^2 \tilde{N}}{n}}.
\]

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

\[
\tilde{\psi}_v^{-1}(\hat{f}_n(x) - f_1(x)) = f_n^* - \tilde{\psi}_v^{-1} f_1(x) = f_n^* - \tilde{\phi}_v^{-1} \theta = f_n^* - L
\]

whereas

\[
\frac{\sqrt{n}}{\sigma}(\hat{f}_n(x) - f_0(x)) = \sqrt{\frac{n}{\sigma}} \tilde{\psi}_v f_n^*(x) = \sqrt{\tilde{N}} \tilde{\phi}_v f_n^*(x)
\]

\[
= f_n^* \exp \left\{ \frac{\log \tilde{N}}{2} + \log \tilde{\phi}_v \right\}.
\]

Denote \( P_0 \) and \( P_1 \) the probabilities associated with \( f_0 \) and \( f_1 \) respectively. From equations (3.75) and (3.76),

\[
\frac{dP_0}{dP_1}(y) = \exp \left\{ -\frac{\theta^2}{2} - \theta \xi \right\}
\]

(3.108)
3.3. ADAPTIVE MINIMAX REGRESSION

with respect to \( P_1 \), where \( \xi \sim \mathcal{N}(0, 1) \). Denote \( q = \exp \{-\bar{\phi}_v\} \) so that \( q \to 0 \) since \( \tilde{N} \to \infty \) \((n \to \infty)\) in NP-scales. Now, given \( f_1 \in \mathcal{A}(\nu) \), for any \( f_n \in \mathcal{F}_p^0(x) \), uniformly in \( v \in \Gamma_n \) as \( n \) goes to infinity, we have

\[
\mathcal{R} := \sup_{f \in \mathcal{A}(\nu)} E^{(n)} \left[ \| \tilde{\psi}_v^{-1}(\tilde{f}_n(x) - f(x)) \|_p^p \right] \geq E_1 \left[ \| \tilde{\psi}_v^{-1}(\tilde{f}_n(x) - f_1(x)) \|_p^p \right]
\]

\[
\geq q E_0 \left\{ \frac{\sqrt{n}}{\sigma} (\tilde{f}_n(x) - f_0(x)) \right\}^p + (1 - q) E_1 \left[ \| \tilde{\psi}_v^{-1}(\tilde{f}_n(x) - f_1(x)) \|_p^p + O(q) \right].
\]

According to (3.106)–(3.109),

\[
\mathcal{R} \geq q \exp \left\{ \frac{\bar{\phi}_v}{2} + p \log \bar{\phi}_v \right\} E_0 \left[ f_n^*(x) \right]_p + (1 - q) E_1 \left[ f_n^*(x) - L \right]_p + O(q)
\]

\[
\geq (1 - q) E_1 (Z \inf x |f_n^*(x)|_p + |f_n^*(x) - L|_p) + O(q)
\]

(3.110)

where

\[
Z = q \exp \left\{ \frac{\bar{\phi}_v}{2} + p \log \bar{\phi}_v \right\} dP_0 \frac{dP_0}{dP_1}.
\]

From (3.108) and definition of \( \theta \) we have

\[
Z = \exp \left\{ -\bar{\phi}_v + \frac{\bar{\phi}_v^2}{2} + p \log \bar{\phi}_v - \left( \bar{\phi}_v - \bar{\phi}_v^{1/2} \right) - \frac{1}{2} \left( \bar{\phi}_v - \bar{\phi}_v^{1/2} \right)^2 \right\} P_1 \to \infty
\]

given \( \bar{\phi}_v \to \infty \). 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

\[
g(x_{\min}) = \chi L^p
\]

(3.111)

where \( \chi \to 1 \). Therefore according to equations (3.110) and (3.111), uniformly in \( v \in \Gamma_n \),

\[
\mathcal{R} \geq (1 - q)L^p E_1 \chi + O(q) = 1 + o(1).
\]

Finally, uniformly in \( \Gamma_n \)

\[
\sup_{f \in \mathcal{A}(\nu)} E^{(n)} \left[ \tilde{\psi}_v^{-1}(\tilde{f}_n(x) - f(x)) \right]_p = \sup_{f \in \mathcal{A}(\nu)} E^{(n)} \left[ \tilde{\psi}_v^{-1}(\tilde{f}_n(x) - f(x)) \right]_p (1 + o(1))
\]

\[
\geq 1 + o(1).
\]

This completes the proof of the theorem. \( \square \)
Corollary 3.3 Let \( \mathcal{A}_n \) be an arbitrary RNP-scale. Then for any \( p > 0 \) and \( x \in \mathbb{R} \), the estimator \( \hat{f}_n \) of Theorem 3.5 is \( (p, \Gamma_n, \mathcal{F}_p(x)) \)-adaptively minimax at \( x \).

**Proof.** This is a consequence of Theorems 3.5 and 3.6. \( \square \)