PCA-kernel Estimation

Gérard Biau

LSTA & LPMA[*]
Université Pierre et Marie Curie – Paris VI
Boîte 158, Tour 15-25, 2ème étage
4 place Jussieu, 75252 Paris Cedex 05, France

DMA[†]
Ecole Normale Supérieure
45 rue d'Ulm
75230 Paris Cedex 05, France

gerard.biau@upmc.fr

André Mas[‡]

Institut de Mathématiques et de Modélisation de Montpellier
UMR CNRS 5149, Equipe de Probabilités et Statistique
Université Montpellier II, CC 051
Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

mas@math.univ-montp2.fr

Abstract

Many statistical estimation techniques for high-dimensional or functional data are based on a preliminary dimension reduction step, which consists in projecting the sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ onto the first D eigenvectors of the Principal Component Analysis (PCA) associated with the empirical projector $\hat{\Pi}_D$. Classical nonparametric inference methods such as kernel density estimation or kernel regression analysis are then performed in the (usually small) D-dimensional space. However, the mathematical analysis of this data-driven dimension reduction scheme raises technical problems, due to the fact that the random variables of the projected sample $(\hat{\Pi}_D \mathbf{X}_1, \ldots, \hat{\Pi}_D \mathbf{X}_n)$ are no more independent. As a reference for further studies, we offer in this paper several results showing the asymptotic equivalencies between important kernel-related quantities based on the empirical projector and its theoretical counterpart. As an illustration, we provide an in-depth analysis of the nonparametric kernel regression case.

Index Terms — Principal Component Analysis, Dimension reduction, Nonparametric kernel estimation, Density estimation, Regression estimation, Perturbation method.

2010 Mathematics Subject Classification: 62G05, 62G20.

1 Introduction

Nonparametric curve estimation provides a useful tool for exploring and understanding the structure of a data set, especially when parametric models are inappropriate. A large amount of progress has been made in the 90's in both the design and the study of inferential aspects of nonparametric estimates. There are too many references to be included here, but the monographs of Silverman [28], Scott [26], Simonoff [29] and Györfi et al. [13] will provide the reader with good introductions to the general subject area.

Among all the nonparametric methods which have been proposed so far, kernel estimation has gained favor from many data analysts, probably because of its simplicity to implement and good statistical properties—see for example Simonoff [29] for a variety of real data examples which illustrate the power of the approach. Kernel estimates were originally studied in density estimation by Rosenblatt [24] and Parzen [22], and were latter introduced in regression estimation by Nadaraya [20, 21] and Watson [32]. A compilation of the mathematical properties of kernel estimates can be found in Prakasa Rao [23] (for density estimation), Györfi et al. [13] (for regression) and Devroye et al. [7] (for classification and pattern recognition). To date, most of the results pertaining to kernel estimation have been reported in the finite-dimensional case, where it is assumed that the observation space is the standard Euclidean space $\mathbb{R}^d$. However, in an increasing number of practical applications, input data items are in the form of random functions (speech recordings, multiple time series, images...) rather than standard vectors, and this casts the problem into the general class of functional data analysis. Motivated by this broad range of potential applications, Ferraty and Vieu describe in [10] a possible route to extend kernel estimation to potentially infinite-dimensional spaces.

On the other hand, it has become increasingly clear over the years that the performances of kernel estimates deteriorate as the dimension of the problem increases. The reason for this is that, in high dimensions, local neighborhoods tend to be empty of sample observations unless the sample size is very large. Thus, in kernel estimation, there will be no local averages to take unless the bandwidth is very large. This general problem was termed the curse of dimensionality (Bellman [1]) and, in fact, practical and theoretical arguments suggest that kernel estimation beyond 5 dimensions is fruitless. The paper by Scott and Wand [27] gives a good account on the feasibility and difficulties of high-dimensional estimation, with examples and computations.

In order to circumvent the high-dimension difficulty and make kernel estimation simpler, a wide range of techniques have been developed. One of the most common approaches is a two-stage strategy: first reduce the dimension of the data and then perform—density or regression—kernel estimation. With this respect, a natural way to reduce dimension is to extract the largest D principal component axes (with D chosen to account for most of the variation in the data), and then operate in this D-dimensional space, thereby improving the ability to discover interesting structures (Jee [15], Friedman [11] and Scott [26], Chapter 7). To illustrate more formally this mechanism, let $(\mathcal{F}, \langle ., . \rangle, \|.\|)$ be a (typically high or infinite-dimensional) separable Hilbert space, and consider for example

the regression problem, where we observe a set $\mathcal{D}_n = \{(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)\}$ of independent $\mathcal{F} \times \mathbb{R}$-valued random variables with the same distribution as a generic pair $(\mathbf{X}, Y)$ satisfying $\mathbb{E}|Y| < \infty$. The goal is to estimate the regression function $r(\mathbf{x}) = \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$ using the data $\mathcal{D}_n$. The kernel estimate of the function r takes the form

$$r_n(\mathbf{x}) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{\|\mathbf{x} - \mathbf{X}_i\|}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{\|\mathbf{x} - \mathbf{X}_i\|}{h_n}\right)}$$

if the denominator is nonzero, and 0 otherwise. Here the bandwidth $h_n > 0$ depends only on the sample size n, and the function $K : [0, \infty) \to [0, \infty)$ is called a kernel. Usually, $K(v)$ is "large" if v is "small", and the kernel estimate is therefore a local averaging estimate. Typical choices for K are the naive kernel $K(v) = \mathbf{1}_{[0,1]}(v)$, the Epanechnikov kernel $K(v) = (1 - v^2)_+$, and the Gaussian kernel $K(v) = \exp(-v^2/2)$.

As explained earlier, the estimate r_n is prone to the curse of dimensionality, and the strategy advocated here is to first reduce the ambient dimension by the use of Principal Component Analysis (PCA, see for example Dauxois et al. [5] and Jolliffe [16]). More precisely, assume without loss of generality that $\mathbb{E}\mathbf{X} = 0$, $\mathbb{E}\|\mathbf{X}\|^2 < \infty$, and let $\Gamma(.) = \mathbb{E}[\langle \mathbf{X}, \cdot \rangle \mathbf{X}]$ be the covariance operator of $\mathbf{X}$ and Π_D be the orthogonal projector on the collection of the first D eigenvectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_D\}$ of Γ associated with the first D eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_D \geq 0$. In the sequel we will assume as well that the distribution of $\mathbf{X}$ is nonatomic.

In this context, the PCA-kernel regression estimate reads

$$r_n^D(\mathbf{x}) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{\|\Pi_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{\|\Pi_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n}\right)}.$$

The hope here is that the most informative part of the distribution of $\mathbf{X}$ should be preserved by projecting the observations on the first D principal component axes, so that the estimate should still do a good job at estimating r while performing in a reduced-dimensional space. Alas, on the practical side, the smoother r_n^D is useless since the distribution of $\mathbf{X}$ (and thus, the projector Π_D) is usually unknown, making of r_n^D what is called a "pseudo-estimate". However, the covariance operator Γ can be approximated by its empirical version

$$\Gamma_n(.) = \frac{1}{n} \sum_{i=1}^{n} \langle \mathbf{X}_i, \cdot \rangle \mathbf{X}_i, \tag{1.1}$$

and Π_D is in turn approximated by the empirical orthogonal projector $\hat{\Pi}_D$ on the (empirical) eigenvalues $\{\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_D\}$ of Γ_n. Thus, the operational version $\hat{r}_n^D$ of the pseudo-estimate r_n^D takes the form

$$\hat{r}_n^D(\mathbf{x}) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{\|\hat{\Pi}_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{\|\hat{\Pi}_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n}\right)}.$$

3

Unfortunately, from a mathematical point of view, computations involving the numerator or the denominator of the estimate $\hat{r}_n^D$ are difficult, since the random variables $(K(\|\hat{\Pi}_D(\mathbf{x} - \mathbf{X}_i)\|/h_n))_{1 \leq i \leq n}$ are identically distributed but clearly *not* independent. Besides, due to nonlinearity, the distribution of $\|\hat{\Pi}_D(\mathbf{x} - \mathbf{X}_i)\|$ is usually inaccessible, even when the $\mathbf{X}_i$'s have known and simple distributions. In short, this makes any theoretical calculation impossible, and it essentially explains why so few theoretical results have been reported so far on the statistical properties of the estimate $\hat{r}_n^D$, despite its wide use. On the other hand, we note that the random variables $(K(\|\Pi_D(\mathbf{x} - \mathbf{X}_i)\|/h_n))_{1 \leq i \leq n}$ are independent and identically distributed. Therefore, the pseudo-estimate r_n^D is amenable to mathematical analysis, and fundamental asymptotic theorems such that the law of large numbers and the central limit theorem may be applied.

In the present contribution, we prove that $\hat{r}_n^D$ and r_n^D have the same asymptotic behavior and show that nothing is lost in terms of rates of convergence when replacing $\hat{r}_n^D$ by r_n^D (Section 4). In fact, taking a more general view, we offer in Section 3 a thorough asymptotic comparison of the partial sums

$$S_n(\mathbf{x}) = \sum_{i=1}^n K\left(\frac{\|\Pi_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n}\right) \quad \text{and} \quad \hat{S}_n(\mathbf{x}) = \sum_{i=1}^n K\left(\frac{\|\hat{\Pi}_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n}\right)$$

with important consequences in kernel density estimation. As an appetizer, we will first carry out in Section 2 a preliminary analysis of the asymptotic proximity between the projection operators $\hat{\Pi}_D$ and Π_D. Our approach will strongly rely on the representation of the operators by Cauchy integrals, through what is classically known in analysis as perturbation method. For the sake of clarity, proofs of the most technical results are postponed to Section 5.

2   Asymptotics for PCA projectors

Here and in the sequel, we let $(\mathcal{F}, \langle ., . \rangle, \|.\|)$ be a separable Hilbert space and $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent random variables, distributed as a generic nonatomic and centered random $\mathbf{X}$ satisfying $\mathbb{E}\|\mathbf{X}\|^2 < \infty$. Denoting by Γ the covariance operator of $\mathbf{X}$, we let Π_D be the orthogonal projection operator on $\{\mathbf{e}_1, \ldots, \mathbf{e}_D\}$, the set of first D eigenvectors of Γ associated with the (nonnegative) eigenvalues $\{\lambda_1, \ldots, \lambda_D\}$ sorted by decreasing order. The empirical version Γ_n of Γ is defined in (1.1), and we denote by $\{\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_D\}$ and $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_D\}$ the associated empirical eigenvector and (nonnegative) eigenvalue sets, respectively, based on the sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$. To keep things simple, we will assume throughout that the projection dimension D is fixed and independent of the observations (for data-dependent methods regarding the choice of D, see for example Jolliffe [16]). Besides, and without loss of generality, it will also be assumed that $\lambda_1 > \ldots > \lambda_{D+1}$. This assumption may be removed at the expense of more tedious calculations taking into account the dimension of the eigenspaces (see for instance [17] for a generic method).

The aim of this section is to derive new asymptotic results regarding the empirical projector $\hat{\Pi}_D$ on $\{\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_D\}$ as the sample size n grows to infinity. Let

us first recall some elementary facts from complex analysis. The eigenvalues $\{\lambda_1, \ldots, \lambda_D\}$ are nonnegative real numbers, but we may view them as points in the complex plane $\mathbb{C}$. Denote by $\mathcal{C}$ a closed oriented contour in $\mathbb{C}$, that is a closed curve (for instance, the boundary of a rectangle) endowed with a circulation. Suppose first that $\mathcal{C} = \mathcal{C}_1$ contains λ_1 only. Then, the so-called formula of residues (Rudin [25]) asserts that

$$\int_{\mathcal{C}_1} \frac{\mathrm{d}z}{z - \lambda_1} = 1 \quad \text{and} \quad \int_{\mathcal{C}_1} \frac{\mathrm{d}z}{z - \lambda_i} = 0 \quad \text{for } i \neq 1.$$

In fact, this formula may be generalized to functional calculus for operators. We refer for instance to Dunford and Schwartz [8] or Gohberg et al. [12] for exhaustive information about this theory, which allows to derive integration formulae for functions with operator values, such as

$$\Pi_1 = \int_{\mathcal{C}_1} (zI - \Gamma)^{-1} \, \mathrm{d}z.$$

Thus, in this formalism, the projector on $\mathbf{e}_1$ is explicitly written as a function of the covariance operator. Clearly, the same arguments allow to express the empirical projector $\hat{\Pi}_1$ as

$$\hat{\Pi}_1 = \int_{\hat{\mathcal{C}}_1} (zI - \Gamma_n)^{-1} \, \mathrm{d}z,$$

where $\hat{\mathcal{C}}_1$ is a (random) contour which contains $\hat{\lambda}_1$ and no other eigenvalue of Γ_n. These formulae generalize and, letting $\mathcal{C}_D$ (respectively $\hat{\mathcal{C}}_D$) be contours containing $\{\lambda_1, \ldots, \lambda_D\}$ (respectively $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_D\}$) only, we may write

$$\Pi_D = \int_{\mathcal{C}_D} (zI - \Gamma)^{-1} \, \mathrm{d}z \quad \text{and} \quad \hat{\Pi}_D = \int_{\hat{\mathcal{C}}_D} (zI - \Gamma_n)^{-1} \, \mathrm{d}z.$$

The contours $\mathcal{C}_D$ may take different forms. However, to keep things simple, we let in the sequel $\mathcal{C}_D$ be the boundary of a rectangle as in Figure 1, with a right vertex intercepting the real line at $x = \lambda_1 + 1/2$ and a left vertex passing through $x = \lambda_D - \delta_D$, with

$$\delta_D = \frac{\lambda_D - \lambda_{D+1}}{2}.$$

With a slight abuse of notation, we will also denote by $\mathcal{C}_D$ the corresponding rectangle.

Thus, with this choice, $\mathcal{C}_D$ contains $\{\lambda_1, \ldots, \lambda_D\}$ an no other eigenvalue. Lemma 2.1 below, which is proved in Section 5, shows that, asymptotically, this assertion is also true with $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_D\}$ in place of $\{\lambda_1, \ldots, \lambda_D\}$. In the sequel, the letter C will denote a positive constant, the value of which may vary from line to line. Moreover, the notation $\|\cdot\|_\infty$ and $\|\cdot\|_2$ will stand for the classical operator and Hilbert-Schmidt norms, which are respectively defined by

$$\|T\|_\infty = \sup_{\mathbf{x} \in \mathcal{B}_1} \|T\mathbf{x}\| \quad \text{and} \quad \|T\|_2^2 = \sum_{p=1}^\infty \|T\mathbf{u}_p\|^2,$$
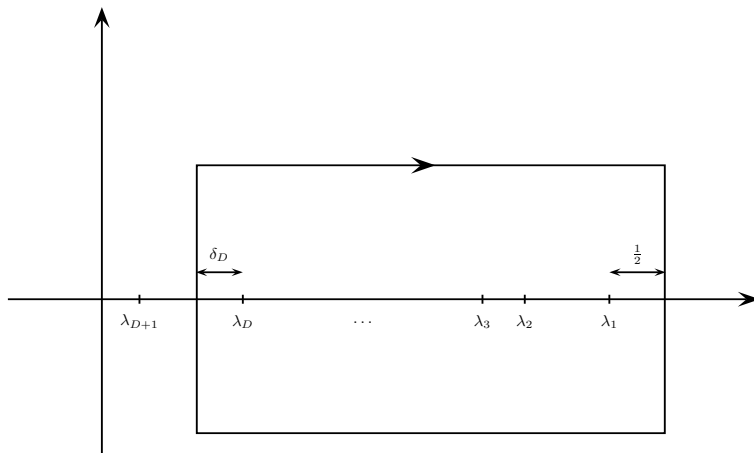
5

Figure 1: Oriented rectangle-contour $\mathcal{C}_D$, with a right vertex intercepting the real line at $x = \lambda_1 + 1/2$ and a left vertex passing through $x = \lambda_D - \delta_D$, $\delta_D = (\lambda_D - \lambda_{D+1})/2$.

where $\mathcal{B}_1$ denotes the closed unit ball of $\mathcal{F}$ and $(\mathbf{u}_p)_{p \geq 1}$ a Hilbertian basis of $\mathcal{F}$. It is known (Dunford and Schwartz [8]) that the value of $\|T\|_2$ does not depend on the actual basis and that $\|\cdot\|_\infty \leq \|\cdot\|_2$. The Hilbert-Schmidt norm is more popular, essentially because it yields simpler calculations than the sup-norm.

As promised, the next lemma ensures that the empirical eigenvalues are located in the rectangle $\mathcal{C}_D$ through an exponential concentration inequality.

Lemma 2.1 *For all $n \geq 1$, let the event*

$$\mathcal{A}_n = \left\{ \hat{\lambda}_i \in \mathcal{C}_D,\ i = 1, \ldots, D,\ \text{and } \hat{\lambda}_{D+1} \notin \mathcal{C}_D \right\}.$$

There exists a positive constant C such that

$$\mathbb{P}(\mathcal{A}_n^c) = \mathcal{O}\left(\exp(-Cn)\right).$$

Remark that the constants involved in the document depend on the actual dimension D and their values increase as D becomes large. To circumvent this difficulty, a possible approach is to let D depend on n. This is beyond the scope of the present paper, and we refer to Cardot et al. [4] for some perspectives in this direction.

We are now in a position to state the main result of the section. Theorem 2.1 below states the asymptotic proximity of the operators $\hat{\Pi}_D$ and Π_D, as n becomes large, with respect to different proximity criteria. We will make repeated use of this result throughout the document. We believe however that it is interesting by itself. For a sequence of random variables $(Z_n)_{n \geq 1}$ and a positive sequence $(v_n)_{n \geq 1}$, notation $Z_n = \mathcal{O}(v_n)$ a.s. means that each random draw of Z_n is $\mathcal{O}(v_n)$.

Theorem 2.1 *The following three assertions are true for all $n \geq 1$:*

(i) There exists a positive constant C such that, for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty \geq \varepsilon\right) = \mathcal{O}\left(\exp\left(-Cn\varepsilon^2\right)\right).$$

(ii) One has

$$\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty = \mathcal{O}\left(\sqrt{\frac{\log n}{n}}\right) \quad a.s.$$

(iii) One has

$$\mathbb{E}\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty^2 = \mathcal{O}\left(\frac{1}{n}\right).$$

Proof of Theorem 2.1 The proof will be based on arguments presented in Mas and Menneteau [17]. Using the notation of Lemma 2.1, we start from the decomposition

$$\hat{\Pi}_D - \Pi_D = \left(\hat{\Pi}_D - \Pi_D\right)\mathbf{1}_{\mathcal{A}_n^c} + \left(\hat{\Pi}_D - \Pi_D\right)\mathbf{1}_{\mathcal{A}_n}. \tag{2.1}$$

Consequently,

$$\mathbb{P}\left(\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty \geq \varepsilon\right)$$
$$\leq \mathbb{P}\left(\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty \mathbf{1}_{\mathcal{A}_n^c} \geq \varepsilon/2\right) + \mathbb{P}\left(\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty \mathbf{1}_{\mathcal{A}_n} \geq \varepsilon/2\right).$$

Observing that

$$\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty \mathbf{1}_{\mathcal{A}_n^c} \leq 2\mathbf{1}_{\mathcal{A}_n^c},$$

we conclude by Lemma 2.1 that

$$\mathbb{P}\left(\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty \mathbf{1}_{\mathcal{A}_n^c} \geq \varepsilon/2\right) \leq \mathbb{P}\left(\mathcal{A}_n^c\right)$$
$$= \mathcal{O}\left(\exp(-nC\varepsilon^2)\right). \tag{2.2}$$

With respect to the second term in (2.1), write

$$\left(\hat{\Pi}_D - \Pi_D\right)\mathbf{1}_{\mathcal{A}_n} = \mathbf{1}_{\mathcal{A}_n}\int_{\mathcal{C}_D}\left[(zI - \Gamma_n)^{-1} - (zI - \Gamma)^{-1}\right]\mathrm{d}z$$
$$= \mathbf{1}_{\mathcal{A}_n}\int_{\mathcal{C}_D}\left[(zI - \Gamma_n)^{-1}(\Gamma_n - \Gamma)(zI - \Gamma)^{-1}\right]\mathrm{d}z.$$

Let ℓ_D be the length of the contour $\mathcal{C}_D$. Using elementary properties of Riesz integrals (Gohberg et al. [12]), we obtain

$$\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty \mathbf{1}_{\mathcal{A}_n}$$
$$\leq \ell_D\left\|\Gamma_n - \Gamma\right\|_\infty \sup_{z \in \mathcal{C}_D}\left[\left\|(zI - \Gamma_n)^{-1}\right\|_\infty \left\|(zI - \Gamma)^{-1}\right\|_\infty\right]\mathbf{1}_{\mathcal{A}_n}.$$

Observing that the eigenvalues of the symmetric operator $(zI - \Gamma)^{-1}$ are the $\left\{ (z - \lambda_i)^{-1}, i \in \mathbb{N}^\star \right\}$, we see that $\left\| (zI - \Gamma)^{-1} \right\|_\infty = \mathcal{O}(\delta_D)$. The same bound is valid taking Γ_n instead of Γ, when $\mathcal{A}_n$ holds. In consequence,

$$\left\| \hat{\Pi}_D - \Pi_D \right\|_\infty \mathbf{1}_{\mathcal{A}_n} = \mathcal{O}\left(\|\Gamma_n - \Gamma\|_\infty \right). \tag{2.3}$$

The conclusion follows from the inequalities (2.1)-(2.2)-(2.3), the inequality $\|\Gamma_n - \Gamma\|_\infty \leq \|\Gamma_n - \Gamma\|_2$ and the asymptotic properties of the sequence $(\Gamma_n - \Gamma)_{n \geq 1}$ (Bosq [3], Chapter 4). ∎

3   Some asymptotic equivalencies

As for now, we assume $D > 2$ and let

$$S_n(\mathbf{x}) = \sum_{i=1}^n K\left(\frac{\|\Pi_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n} \right) \quad \text{and} \quad \hat{S}_n(\mathbf{x}) = \sum_{i=1}^n K\left(\frac{\left\| \hat{\Pi}_D(\mathbf{x} - \mathbf{X}_i) \right\|}{h_n} \right).$$

We note that $S_n(\mathbf{x})$ is a sum of independent and identically distributed random variables, whereas the terms in $\hat{S}_n(\mathbf{x})$ have the same distribution but are *not* independent. In light of the results of Section 2, our goal in this section will be to analyse the asymptotic proximity between $S_n(\mathbf{x})$ and $\hat{S}_n(\mathbf{x})$ under general conditions on K and the sequence $(h_n)_{n \geq 1}$. Throughout, we will assume that the kernel K satisfies the following set of conditions:

Assumption K

(**K**1) K is positive and bounded with compact support $[0, 1]$.

(**K**2) K is of class $\mathcal{C}^1$ on $[0, 1]$.

These assumptions are typically satisfied by the naive kernel $K(v) = \mathbf{1}_{[0,1]}(v)$. In fact, all the subsequent results also hold for kernels with an unbounded support, provided K is Lipschitz—we leave to the reader the opportunity to check the details and adapt the proofs, which turn out to be simpler in this case. For any integer $p \geq 1$, we set

$$M_{D,p} = D \int_0^1 v^{D-1} K^p(v)\, \mathrm{d}v$$

and, for all $\mathbf{x} \in \mathcal{F}$ and $h > 0$, we let

$$F_{\mathbf{x}}(h) = \mathbb{P}\left(\Pi_D \mathbf{X} \in \mathcal{B}_D(\Pi_D \mathbf{x}, h) \right),$$

where $\mathcal{B}_D(u, h)$ denotes the closed Euclidean ball of dimension D centered at u and of radius h. Observe that $F_{\mathbf{x}}(h)$ is positive for μ-almost all $\mathbf{x} \in \mathcal{F}$, where μ is the distribution of $\mathbf{X}$. Besides, by decreasing monotonicity, since $\mathbf{X}$ is nonatomic, we have

$$\lim_{h \downarrow 0} F_{\mathbf{x}}(h) = 0.$$

When the projected random variable $\Pi_D \mathbf{X}$ has a density f with respect to the Lebesgue measure λ on $\mathbb{R}^D$, then $F_{\mathbf{x}}(h) \sim \gamma_D f(\mathbf{x}) h^D$ as $h \to 0$, where γ_D is a

positive constant, for λ-almost all $\mathbf{x}$ (see for instance Wheeden and Zygmund [33]). Thus, in this case, the function $F_{\mathbf{x}}$ is regularly varying with index D. We generalize this property below.

Assumption $\mathbf{R_x}$ $\quad F_{\mathbf{x}}$ is regularly varying at 0 with index D.

Assumption $\mathbf{R_x}$ means that, for any $u > 0$,
$$\lim_{s \to 0^+} \frac{F_{\mathbf{x}}(su)}{F_{\mathbf{x}}(s)} = u^D.$$

The index of regular variation was fixed to D in order to alleviate the notation, but the reader should note that our results hold for any positive index, with different constants however. In fact this index is directly connected with the support of the distribution of $\mathbf{X}$. To see this, observe that by fixing the index to D we implicitly assume that $\Pi_D \mathbf{X}$ fills the whole space of dimension D. However, elementary calculations show that most distributions in $\mathbb{R}^D$, when concentrated on a subspace of smaller dimension $D' < D$, will match assumption $\mathbf{R_x}$ with D' instead of D. Moreover, representation theorems for regularly varying functions (see Bingham et al. [2]) show that, under $\mathbf{R_x}$, $F_{\mathbf{x}}$ may be rewritten as $F_{\mathbf{x}}(u) = u^D L(u)$, where the function L is slowly varying at 0, that is $\lim_{s \to 0^+} L(su)/L(s) = 1$. This enables to consider functions $F_{\mathbf{x}}$ with non-polynomial behaviour such as, for instance, $F_{\mathbf{x}}(u) \sim C u^D |\ln u|$ as $u \to 0^+$. Observe also that $F_{\mathbf{x}}(u)$ is negligible with respect to u^2 as soon as $D > 2$.

In the rest of the document, we fix $\mathbf{x} \in \mathcal{F}$ once and for all, and assume that $F_{\mathbf{x}}(h) > 0$ for all positive h with $\lim_{h \downarrow 0} F_{\mathbf{x}}(h) = 0$.

We will finally say that a real-valued random variable ξ satisfies the Assumption $\mathbf{C}$ (for **C**ramer's conditions) if:

Assumption C $\quad$ There exist two positive constants v and c such that, for all integers $k \geq 2$,
$$\mathbb{E}|\xi|^k \leq \frac{k!}{2} v c^{k-2}.$$

Observe that Assumption $\mathbf{C}$ holds for a Gaussian random variable and recall that if $\xi_1, \ldots, \xi_n$ are independent copies of ξ then, by Bernstein's inequality (see for example Massart [18], Chapter 2)
$$\mathbb{P}\left(\left|\sum_{i=1}^{n}(\xi_i - \mathbb{E}\xi_i)\right| \geq \varepsilon\right) \leq 2 \exp\left(-\frac{\varepsilon^2}{2(nv + c\varepsilon)}\right)$$
for all positive ε.

We start the analysis with two technical lemmas. Proof of Lemma 3.1 is deferred to Section 5, whereas Lemma 3.2 is an immediate consequence of Lemma 3.1 and Bennett's inequality. Its proof is therefore omitted.

Lemma 3.1 *Assume that Assumption $\mathbf{K}$ and Assumption $\mathbf{R_x}$ are satisfied. Then, if $h_n \downarrow 0$,*
$$\mathbb{E}K\left(\frac{\|\Pi_D(\mathbf{x} - \mathbf{X})\|}{h_n}\right) \sim M_{D,1} F_{\mathbf{x}}(h_n)$$

and

$$\mathbb{E}K^2\left(\frac{\|\Pi_D\left(\mathbf{x}-\mathbf{X}\right)\|}{h_n}\right) \sim M_{D,2}F_{\mathbf{x}}\left(h_n\right) \quad \text{as } n \to \infty.$$

Lemma 3.2 *Assume that Assumption* **K** *and Assumption* **R$_\mathbf{x}$** *are satisfied. Then, if $h_n \downarrow 0$ and $nF_{\mathbf{x}}(h_n)/\ln n \to \infty$,*

$$S_n(\mathbf{x}) \sim M_{D,1}nF_{\mathbf{x}}\left(h_n\right) \quad a.s.$$

and

$$\mathbb{E}S_n^2(\mathbf{x}) \sim [M_{D,1}nF_{\mathbf{x}}\left(h_n\right)]^2 \quad \text{as } n \to \infty.$$

The following proposition is the cornerstone of this section. It asserts that, asymptotically, the partial sums $S_n(\mathbf{x})$ and $\hat{S}_n(\mathbf{x})$ behave similarly.

Proposition 3.1 *Assume that Assumption* **K** *and Assumption* **R$_\mathbf{x}$** *are satisfied. Assume also that $\|\mathbf{x}-\mathbf{X}\|$ satisfies Assumption* **C**. *Then, if $h_n \downarrow 0$ and $nF_{\mathbf{x}}(h_n)/\ln n \to \infty$,*

$$\hat{S}_n(\mathbf{x}) \sim S_n(\mathbf{x}) \quad a.s.$$

and

$$\mathbb{E}\hat{S}_n^2(\mathbf{x}) \sim \mathbb{E}S_n^2(\mathbf{x}) \quad \text{as } n \to \infty.$$

Proof of Proposition 3.1 Recall that the letter C denotes a positive constant, the value of which may vary from line to line. To simplify notation a bit, we let, for $i = 1, \dots, n$, $V_i = \|\Pi_D(\mathbf{x}-\mathbf{X}_i)\|$ and $\hat{V}_i = \|\hat{\Pi}_D(\mathbf{x}-\mathbf{X}_i)\|$. Let the events $\mathcal{E}_i$ and $\hat{\mathcal{E}}_i$ be defined by

$$\mathcal{E}_i = \{V_i \le h_n\} \quad \text{and} \quad \hat{\mathcal{E}}_i = \left\{\hat{V}_i \le h_n\right\}.$$

Clearly,

$$\begin{aligned}
&\hat{S}_n(\mathbf{x}) - S_n(\mathbf{x}) \\
&= \sum_{i=1}^{n} K\left(\hat{V}_i/h_n\right) - \sum_{i=1}^{n} K\left(V_i/h_n\right) \\
&= \sum_{i=1}^{n}\left[K\left(\hat{V}_i/h_n\right) - K\left(V_i/h_n\right)\right]\mathbf{1}_{\hat{\mathcal{E}}_i \cap \mathcal{E}_i} + \sum_{i=1}^{n}K\left(\hat{V}_i/h_n\right)\mathbf{1}_{\hat{\mathcal{E}}_i \cap \mathcal{E}_i^c} \\
&\quad - \sum_{i=1}^{n}K\left(V_i/h_n\right)\mathbf{1}_{\hat{\mathcal{E}}_i^c \cap \mathcal{E}_i}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\left|\hat{S}_n(\mathbf{x}) - S_n(\mathbf{x})\right| \\
&\le \sum_{i=1}^{n}\left|K\left(\hat{V}_i/h_n\right) - K\left(V_i/h_n\right)\right|\mathbf{1}_{\hat{\mathcal{E}}_i \cap \mathcal{E}_i} + C\sum_{i=1}^{n}\left(\mathbf{1}_{\hat{\mathcal{E}}_i \cap \hat{\mathcal{E}}_i^c} + \mathbf{1}_{\mathcal{E}_i^c \cap \hat{\mathcal{E}}_i}\right) \\
&\le C\left[\frac{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty}{h_n}\sum_{i=1}^{n}\|\mathbf{x}-\mathbf{X}_i\|\,\mathbf{1}_{\mathcal{E}_i} + \sum_{i=1}^{n}\left(\mathbf{1}_{\mathcal{E}_i \cap \hat{\mathcal{E}}_i^c} + \mathbf{1}_{\mathcal{E}_i^c \cap \hat{\mathcal{E}}_i}\right)\right]. \quad (3.1)
\end{aligned}$$

Consequently, by Lemma 3.2, the result will be proved if we show that

$$\frac{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty}{nh_nF_{\mathbf{x}}\left(h_n\right)}\sum_{i=1}^{n}\|\mathbf{x} - \mathbf{X}_i\|\,\mathbf{1}_{\mathcal{E}_i} \to 0 \quad \text{a.s.}$$

and

$$\frac{1}{nF_{\mathbf{x}}\left(h_n\right)}\sum_{i=1}^{n}\left(\mathbf{1}_{\mathcal{E}_i\cap\hat{\mathcal{E}}_i^c} + \mathbf{1}_{\mathcal{E}_i^c\cap\hat{\mathcal{E}}_i}\right) \to 0 \quad \text{a.s.} \quad \text{as } n \to \infty.$$

The first limit is proved in technical Lemma 5.1 and the second one in technical Lemma 5.2.

We proceed now to prove the second statement of the proposition. We have to show that

$$\frac{\mathbb{E}\hat{S}_n^2(\mathbf{x})}{\mathbb{E}S_n^2(\mathbf{x})} \to 1 \quad \text{as } n \to \infty.$$

Using the decomposition

$$\frac{\mathbb{E}U^2}{\mathbb{E}V^2} = 1 + \frac{\mathbb{E}\left[U - V\right]^2}{\mathbb{E}V^2} + 2\frac{\mathbb{E}\left[V\left(U - V\right)\right]}{\mathbb{E}V^2},$$

and the bound

$$\frac{|\mathbb{E}\left[V\left(U - V\right)\right]|}{\mathbb{E}V^2} \leq \sqrt{\frac{\mathbb{E}\left[U - V\right]^2}{\mathbb{E}V^2}},$$

it will be enough to prove that

$$\frac{\mathbb{E}\left[\hat{S}_n(\mathbf{x}) - S_n(\mathbf{x})\right]^2}{\mathbb{E}S_n^2(\mathbf{x})} \to 0,$$

which in turn comes down to prove that

$$\frac{\mathbb{E}\left[\hat{S}_n(\mathbf{x}) - S_n(\mathbf{x})\right]^2}{n^2F_{\mathbf{x}}^2\left(h_n\right)} \to 0,$$

since $\mathbb{E}S_n^2(\mathbf{x}) \sim \left[M_{D,1}nF_{\mathbf{x}}\left(h_n\right)\right]^2$ by Lemma 3.2.

Starting from inequality (3.1), we obtain

$$\left[\hat{S}_n(\mathbf{x}) - S_n(\mathbf{x})\right]^2 \leq C\left[\frac{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty}{h_n}\left(\sum_{i=1}^{n}\|\mathbf{x} - \mathbf{X}_i\|\,\mathbf{1}_{\mathcal{E}_i}\right)\right]^2$$

$$+ C\left[\sum_{i=1}^{n}\left(\mathbf{1}_{\mathcal{E}_i\cap\hat{\mathcal{E}}_i^c} + \mathbf{1}_{\mathcal{E}_i^c\cap\hat{\mathcal{E}}_i}\right)\right]^2. \qquad (3.2)$$

Consequently, the result will be proved if we show that

$$\mathbb{E}\left[\frac{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty}{nh_nF_{\mathbf{x}}(h_n)}\left(\sum_{i=1}^{n}\|\mathbf{x} - \mathbf{X}_i\|\,\mathbf{1}_{\mathcal{E}_i}\right)\right]^2 \to 0$$

and

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\left(\mathbf{1}_{\mathcal{E}_i\cap\hat{\mathcal{E}}_i^c}+\mathbf{1}_{\mathcal{E}_i^c\cap\hat{\mathcal{E}}_i}\right)\right]^2\to 0\quad\text{as }n\to\infty.$$

The first limit is established in technical Lemma 5.3 and the second one in technical Lemma 5.4. ∎

The consequences of Proposition 3.1 in terms of kernel regression estimation will be thoroughly explored in Section 4. However, it has already important repercussions in density estimation, which are briefly sketched here and may serve as references for further studies. Suppose that the projected random variable $\Pi_D\mathbf{X}$ has a density f with respect to the Lebesgue measure λ on $\mathbb{R}^D$. In this case, the PCA-kernel density estimate of f—based on the sample $(\hat{\Pi}_D\mathbf{X}_1,\ldots,\hat{\Pi}_D\mathbf{X}_n)$— reads

$$\hat{f}_n(\mathbf{x})=\frac{\hat{S}_n(\mathbf{x})}{nh_n^D}$$

and the associated pseudo-estimate—based on $(\Pi_D\mathbf{X}_1,\ldots,\Pi_D\mathbf{X}_n)$— takes the form

$$f_n(\mathbf{x})=\frac{S_n(\mathbf{x})}{nh_n^D}.$$

An easy adaptation of the proof of Corollary 4.1 in Section 4 shows that, under the conditions of Proposition 3.1,

$$\mathbb{E}\left[\hat{f}_n(\mathbf{x})-f_n(\mathbf{x})\right]^2=\mathcal{O}\left(\frac{\log\left(nh_n^2\right)}{nh_n^2}\right).$$

To illustrate the importance of this result, suppose for example that the target density f belongs to the class $\mathcal{G}_p$ of p-times continuously differentiable functions. In this context (Stone [30, 31]), the optimal rate of convergence over $\mathcal{G}_p$ is $n^{-2p/(2p+D)}$ and the kernel density estimate with a bandwidth $h_n^\star\asymp n^{-1/(2p+D)}$ achieves this minimax rate. Thus, letting $\hat{f}_n^\star$ (respectively $f_n^\star$) be the PCA-kernel density estimate (respectively pseudo-density estimate) based on this optimal bandwidth, we are led to

$$\frac{\mathbb{E}\left[\hat{f}_n^\star(\mathbf{x})-f_n^\star(\mathbf{x})\right]^2}{n^{-2p/(2p+D)}}\to 0$$

as soon as $D>2$. Thus, the L_2-rate of convergence of $\hat{f}_n^\star$ towards $f_n^\star$ is negligible with respect to the L_2-rate of convergence of $f_n^\star$ towards f. In consequence, replacing $\hat{f}_n^\star$ by $f_n^\star$ has no effect on the asymptotic rate. The same ideas may be transposed without further effort to asymptotic normality and other error criteria.

Note that the restriction $D>2$, which is essentially due to a methodological artefact, is not severe, since in all practical situations we have in mind the dimension D is usually much larger than 1 or 2. Besides, simpler tools—such as single index techniques, see McCullagh and Nelder [19] and Horowitz [14]—exist to deal with lower dimensions. Nevertheless, the cases $D=1$ and $D=2$ may

be directly addressed via an ad hoc analysis. It is for instance routine to prove that for, say $D = 1$ and $p = 1$, setting $\Delta_i = \langle \mathbf{X}_i - \mathbf{x}, \hat{\mathbf{e}}_1 - \mathbf{e}_1 \rangle$,

$$\mathbb{E}\left[K\left(V_i/h_n\right) - K\left(\left(V_i - \Delta_i\right)/h_n\right)\right]^2 = \mathcal{O}\left(\frac{(\log h_n)^2}{n^2}\right)$$

from which it immediately follows that

$$\mathbb{E}\left[\hat{f}_n^\star(\mathbf{x}) - f_n^\star(\mathbf{x})\right]^2 = \mathrm{o}\left(n^{-2/3}\right).$$

To conclude this section, we would like to mention that the high-dimensional density estimation problem via PCA-reduction has been recently addressed by Delaigle and Hall [6]. These authors are interested in estimating the density of the individual principal component scores—a problem which is close, though different, of ours.

4 Regression analysis

As framed in the introduction, we study in this final section the PCA-kernel regression procedure, which was our initial motivation. Recall that, in this context, we observe a set $\{(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)\}$ of independent $\mathcal{F} \times \mathbb{R}$-valued random variables with the same distribution as a generic pair $(\mathbf{X}, Y)$, where $\mathcal{F}$ is a separable Hilbert space, $\mathbf{X}$ is nonatomic centered, and Y satisfies $\mathbb{E}Y^2 < \infty$. The goal is to estimate the regression function $r^D(\mathbf{x}) = \mathbb{E}[Y | \Pi_D \mathbf{X} = \Pi_D \mathbf{x}]$ via the PCA-kernel estimate, which takes the form

$$\hat{r}_n^D(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\|\hat{\Pi}_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\|\hat{\Pi}_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n}\right)}.$$

This estimate is mathematically intractable and we plan to prove that we can substitute without damage to $\hat{r}_n^D$ the pseudo-estimate

$$r_n^D(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\|\Pi_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\|\Pi_D(\mathbf{x} - \mathbf{X}_i)\|}{h_n}\right)}.$$

To this aim, observe first that, with the notation of Section 3,

$$\hat{r}_n^D(\mathbf{x}) = \frac{\hat{Z}_n(\mathbf{x})}{\hat{S}_n(\mathbf{x})}$$

and

$$r_n^D(\mathbf{x}) = \frac{Z_n(\mathbf{x})}{S_n(\mathbf{x})},$$

where, for all $n \geq 1$,

$$\hat{Z}_n(\mathbf{x}) = \sum_{i=1}^n Y_i K\left(\frac{\left\|\hat{\Pi}_D\left(\mathbf{x} - \mathbf{X}_i\right)\right\|}{h_n}\right)$$

13

and

$$Z_n(\mathbf{x}) = \sum_{i=1}^{n} Y_i K \left(\frac{\|\Pi_D (\mathbf{x} - \mathbf{X}_i)\|}{h_n} \right).$$

Consider now the regression model

$$Y = \mathbb{E}\left[Y | \|\Pi_D (\mathbf{x} - \mathbf{X})\|\right] + \varepsilon,$$

where $\varepsilon = Y - \mathbb{E}[Y | \|\Pi_D(\mathbf{x} - \mathbf{X})\|]$ may be viewed as the noise part in the model and will support the moment assumptions in the proposition below.

Proposition 4.1 *Assume that Assumption* $\mathbf{K}$ *and Assumption* $\mathbf{R_x}$ *are satisfied, that* $r^D(\mathbf{x}) \neq 0$ *and* r^D *is Lipschitz in a neighborhood of* $\mathbf{x}$. *Assume also that* $\mathbf{X}$ *is bounded and* ε *satisfies Assumption* $\mathbf{C}$, *with*

$$\mathbb{E}\left[|\varepsilon|^k \,|\, \|\Pi_D (\mathbf{x} - \mathbf{X})\|\right] = \mathbb{E}\,|\varepsilon|^k$$

for all integers $k \geq 2$. *Then, if* $h_n \downarrow 0$ *and* $nF_{\mathbf{x}}(h_n)/\ln n \to \infty$,

$$\hat{Z}_n(\mathbf{x}) \sim Z_n(\mathbf{x}) \quad a.s.$$

and

$$\mathbb{E}\hat{Z}_n^2(\mathbf{x}) \sim \mathbb{E}Z_n^2(\mathbf{x}) \quad as \ n \to \infty.$$

Proof of Proposition 4.1 Using the Lipschitz property of r^D, we easily obtain by following the lines of Lemma 3.1 that

$$\mathbb{E}\left[Y K \left(\frac{\|\Pi_D (\mathbf{x} - \mathbf{X})\|}{h_n} \right)\right] \sim r^D(\mathbf{x})F_{\mathbf{x}}(h_n)$$

and

$$\mathbb{E}\left[Y^2 K^2 \left(\frac{\|\Pi_D (\mathbf{x} - \mathbf{X})\|}{h_n} \right)\right] \sim CF_{\mathbf{x}}(h_n) \quad \text{as } n \to \infty.$$

Moreover,

$$Z_n(\mathbf{x}) \sim n\mathbb{E}\left[Y K \left(\frac{\|\Pi_D (\mathbf{x} - \mathbf{X})\|}{h_n} \right)\right] \quad \text{a.s.}$$

and

$$\mathbb{E}Z_n^2(\mathbf{x}) \sim \left[n\mathbb{E}\left[Y K \left(\frac{\|\Pi_D (\mathbf{x} - \mathbf{X})\|}{h_n} \right)\right]\right]^2 \quad \text{as } n \to \infty.$$

The first equivalence is a consequence of Bennett's inequality and the fact that, for all large enough n and $i = 1, 2$,

$$\mathbb{E}\left[|Y|^i K \left(\frac{\|\Pi_D (\mathbf{x} - \mathbf{X})\|}{h_n} \right)\right] \geq CF_{\mathbf{x}}(h_n),$$

which itself follows from the requirement $r^D(\mathbf{x}) > 0$.

Finally, since Y is bounded, an inspection of the proof of Proposition 3.1 reveals that displays (3.1) and (3.2) may be verbatim repeated with S replaced by Z.∎

Corollary 4.1 *Under the assumptions of Proposition 4.1, the estimate $\hat{r}_n^D$ and the pseudo-estimate r_n^D satisfy*

$$\hat{r}_n^D(\mathbf{x}) \sim r_n^D(\mathbf{x}) \quad a.s. \ as \ n \to \infty.$$

Moreover,

$$\mathbb{E}\left[\hat{r}_n^D(\mathbf{x}) - r_n^D(\mathbf{x})\right]^2 = \mathcal{O}\left(\frac{\log\left(nh_n^2\right)}{nh_n^2}\right).$$

Proof of Corollary 4.1 We start with the decomposition

$$\hat{r}_n^D(\mathbf{x}) - r_n^D(\mathbf{x}) = \frac{\hat{Z}_n(\mathbf{x})}{\hat{S}_n(\mathbf{x})}\left(1 - \frac{\hat{S}_n(\mathbf{x})}{S_n(\mathbf{x})}\right) + \frac{1}{S_n(\mathbf{x})}\left(\hat{Z}_n(\mathbf{x}) - Z_n(\mathbf{x})\right),$$

which comes down to

$$\frac{\hat{r}_n^D(\mathbf{x}) - r_n^D(\mathbf{x})}{\hat{r}_n^D(\mathbf{x})} = \left(1 - \frac{\hat{S}_n(\mathbf{x})}{S_n(\mathbf{x})}\right) + \frac{\hat{S}_n(\mathbf{x})}{S_n(\mathbf{x})}\left(1 - \frac{Z_n(\mathbf{x})}{\hat{Z}_n(\mathbf{x}))}\right).$$

The first part of the corollary is then an immediate consequence of Proposition 2.1 and Proposition 4.1.

We turn to the second part. Note that the quotient $\hat{Z}_n(\mathbf{x})/\hat{S}_n(\mathbf{x})$ is bounded whenever Y is bounded. In consequence, we just need to provide upper bounds for the terms $\mathbb{E}\left[1 - \hat{S}_n(\mathbf{x})/S_n(\mathbf{x})\right]^2$ and $\mathbb{E}\left[\left(\hat{Z}_n(\mathbf{x}) - Z_n(\mathbf{x})\right)/S_n(\mathbf{x})\right]^2$. Besides, classical arguments (see e.g. Fan [9], pages 210-211) show that the latter two expectations may be replaced by the quantities $\mathbb{E}\left[S_n(\mathbf{x}) - \hat{S}_n(\mathbf{x})\right]^2/\mathbb{E}S_n^2(\mathbf{x})$ and $\mathbb{E}\left[\hat{Z}_n(\mathbf{x}) - Z_n(\mathbf{x})\right]^2/\mathbb{E}S_n^2(\mathbf{x})$, respectively. It turns out that the analysis of each of these terms is similar, and we will therefore focus on the first one only. Given the result of Proposition 3.2, this comes down to analyse $\mathbb{E}\left[S_n(\mathbf{x}) - \hat{S}_n(\mathbf{x})\right]^2/[M_{D,1}nF_\mathbf{x}(h_n)]^2$ and to refine the bound.

By inequality (3.2), we have

$$\frac{\mathbb{E}\left[S_n(\mathbf{x}) - \hat{S}_n(\mathbf{x})\right]^2}{n^2 F_\mathbf{x}^2(h_n)} \leq C\mathbb{E}\left[\frac{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty}{nh_n F_\mathbf{x}(h_n)}\left(\sum_{i=1}^n \|\mathbf{x} - \mathbf{X}_i\|\mathbf{1}_{\mathcal{E}_i}\right)\right]^2$$

$$+ C\left[\frac{1}{nF_\mathbf{x}(h_n)}\sum_{i=1}^n\left(\mathbf{1}_{\mathcal{E}_i\cap\hat{\mathcal{E}}_i^c} + \mathbf{1}_{\mathcal{E}_i^c\cap\hat{\mathcal{E}}_i}\right)\right]^2. \quad (4.1)$$

With respect to the first term, Lemma 5.3 asserts that

$$\mathbb{E}\left[\frac{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty}{nh_n F_\mathbf{x}(h_n)}\left(\sum_{i=1}^n \|\mathbf{x} - \mathbf{X}_i\|\mathbf{1}_{\mathcal{E}_i}\right)\right]^2 = \mathcal{O}\left(\frac{1}{nh_n^2}\right).$$

The second term in inequality (4.1) is of the order $\mathcal{O}(\log(nh_n^2)/(nh_n^2))$, as proved in technical Lemma 5.5. This completes the proof. ∎

15

To illustrate the usefulness of Corollary 4.1, suppose that the regression function r^D belongs to the class $\mathcal{G}_p$ of p-times continuously differentiable functions. In this framework, it is well-known (Stone [30, 31]) that the optimal rate of convergence on the class $\mathcal{G}_p$ is $n^{-2p/(2p+D)}$ and that the kernel estimate with a bandwidth $h_n^\star \asymp n^{-1/(2p+D)}$ achieves this minimax rate. Plugging this optimal $h_n^\star$ into the rate of Corollary 4.1, we obtain

$$\mathbb{E}\left[\hat{r}_n^D\left(\mathbf{x}\right) - r_n^D\left(\mathbf{x}\right)\right]^2 = \mathcal{O}\left(n^{-\alpha}\right),$$

with $\alpha = \left(2p + D - 2\right)/\left(2p + D\right)$. This rate is strictly faster than the minimax rate $n^{-2p/(2p+D)}$ provided $\alpha > 2p/\left(2p + D\right)$ or, equivalently, when $D > 2$. In this case,

$$\lim_{n\to\infty}\frac{\mathbb{E}\left[\hat{r}_n^D\left(\mathbf{x}\right) - r_n^D\left(\mathbf{x}\right)\right]^2}{n^{-2p/(2p+D)}} = 0,$$

and Corollary 4.1 claims in fact that the rate of convergence of $\hat{r}_n^D$ towards r_n^D is negligible with respect to the rate of convergence of r_n^D towards r^D. In conclusion, even if $\hat{r}_n^D$ is the only possible and feasible estimate, carrying out its asymptotics from the pseudo-estimate r_n^D is permitted.

5 Proofs

5.1 Proof of Lemma 2.1

Observe first, since $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_D$, that

$$\mathcal{A}_n = \left\{\hat{\lambda}_1 \leq \lambda_1 + 1/2, \hat{\lambda}_D \geq \lambda_D - \delta_D, \text{and } \hat{\lambda}_{D+1} < \lambda_D - \delta_D\right\}.$$

Therefore

$$\mathbb{P}\left(\mathcal{A}_n^c\right)$$
$$\leq \mathbb{P}\left(\hat{\lambda}_1 - \lambda_1 > 1/2\right) + \mathbb{P}\left(\hat{\lambda}_D - \lambda_D < -\delta_D\right) + \mathbb{P}\left(\hat{\lambda}_{D+1} - \lambda_{D+1} \geq \delta_D\right)$$
$$\leq \mathbb{P}\left(\left|\hat{\lambda}_1 - \lambda_1\right| \geq 1/2\right) + \mathbb{P}\left(\left|\hat{\lambda}_D - \lambda_D\right| \geq \delta_D\right) + \mathbb{P}\left(\left|\hat{\lambda}_{D+1} - \lambda_{D+1}\right| \geq \delta_D\right).$$

The inequality

$$\sup_{i\geq 1}\left|\hat{\lambda}_i - \lambda_i\right| \leq \|\Gamma_n - \Gamma\|_2$$

shifts the problem from $|\hat{\lambda}_i - \lambda_i|$ to $\|\Gamma_n - \Gamma\|_2$. An application of a standard theorem for Hilbert-valued random variables (see for instance Bosq [3]) leads to

$$\mathbb{P}\left(\|\Gamma_n - \Gamma\|_2 \geq \varepsilon\right) = \mathcal{O}\left(\exp\left(-c_1\frac{n\varepsilon^2}{c_2 + c_3\varepsilon}\right)\right),$$

for three positive constants c_1, c_2 and c_3. Consequently, for fixed D,

$$\mathbb{P}\left(\mathcal{A}_n^c\right) = \mathcal{O}\left(\exp\left(-nC\varepsilon^2\right)\right),$$

where C is a positive constant depending on D.

16

5.2 Proof of Lemma 3.1

The proof will be based on successive applications of Fubini's theorem. Denoting by $\mu_{D,\mathbf{x},h_n}$ the probability measure associated with the random variable $\|\Pi_D(\mathbf{x}-\mathbf{X})\|/h_n$, we may write

$$\mathbb{E}K\left(\frac{\|\Pi_D(\mathbf{x}-\mathbf{X})\|}{h_n}\right) = \int_0^1 K(v)\,\mu_{D,\mathbf{x},h_n}(\mathrm{d}v).$$

Thus

$$\begin{aligned}
\mathbb{E}K\left(\frac{\|\Pi_D(\mathbf{x}-\mathbf{X})\|}{h_n}\right) &= \int_0^1 \left[K(1) - \int_v^1 K'(s)\,\mathrm{d}s\right]\mu_{D,\mathbf{x},h_n}(\mathrm{d}v) \\
&= K(1)F_\mathbf{x}(h_n) - \int_0^1 K'(s)\int_{[0\le v\le s]}\mu_{D,\mathbf{x},h_n}(\mathrm{d}v)\,\mathrm{d}s \\
&= K(1)F_\mathbf{x}(h_n) - \int_0^1 F_\mathbf{x}(h_ns)K'(s)\,\mathrm{d}s \\
&= F_\mathbf{x}(h_n)\left[K(1) - \int_0^1 \frac{F_\mathbf{x}(h_ns)}{F_\mathbf{x}(h_n)}K'(s)\,\mathrm{d}s\right].
\end{aligned}$$

Using the fact that $F_\mathbf{x}$ is increasing regularly varying of order D, an application of Lebegue's dominated convergence theorem yields

$$\mathbb{E}K\left(\frac{\|\Pi_D(\mathbf{x}-\mathbf{X})\|}{h_n}\right) \sim F_\mathbf{x}(h_n)\left[K(1) - \int_0^1 s^D K'(s)\,\mathrm{d}s\right]$$

i.e.,

$$\mathbb{E}K\left(\frac{\|\Pi_D(\mathbf{x}-\mathbf{X})\|}{h_n}\right) \sim F_\mathbf{x}(h_n)\,D\int_0^1 s^{D-1}K(s)\,\mathrm{d}s \quad \text{as } n\to\infty.$$

This shows the first statement of the lemma. Proof of the second statement is similar.

5.3 Some technical lemmas

In this subsection, for all $i=1,\ldots,n$, we let $V_i = \|\Pi_D(\mathbf{x}-\mathbf{X}_i)\|$ and $\hat{V}_i = \|\hat{\Pi}_D(\mathbf{x}-\mathbf{X}_i)\|$. The events $\mathcal{E}_i$ and $\hat{\mathcal{E}}_i$ are defined by

$$\mathcal{E}_i = \{V_i \le h_n\} \quad \text{and} \quad \hat{\mathcal{E}}_i = \left\{\hat{V}_i \le h_n\right\}.$$

Recall that we fix $\mathbf{x}\in\mathcal{F}$ once and for all, and assume that $F_\mathbf{x}(h)>0$ for all positive h with $\lim_{h\downarrow 0}F_\mathbf{x}(h)=0$. To avoid concentrating too much on technicalities, we first state the results in the easier case where the random variables $\mathbf{X}$ and/or Y are bounded. We finally explain in Lemma 5.6 and Lemma 5.7 how to transpose the results to the more involved situation where $\|\mathbf{x}-\mathbf{X}\|$ and/or Y satisfy the general Assumption **C**, and leave to the reader the opportunity to work out the details.

Lemma 5.1 *Assume that* $\mathbf{X}$ *is bounded. Then*

$$\frac{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty}{nh_nF_\mathbf{x}(h_n)}\sum_{i=1}^n \|\mathbf{x}-\mathbf{X}_i\|\,\mathbf{1}_{\mathcal{E}_i} = \mathcal{O}\left(\sqrt{\frac{\log n}{nh_n^2}}\right) \quad a.s.$$

Proof of Lemma 5.1   According to statement (ii) of Theorem 2.1,

$$\frac{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty}{nh_n F_{\mathbf{x}}(h_n)} = \mathcal{O}\left(\sqrt{\frac{\log n}{n^3 h_n^2 F_{\mathbf{x}}^2(h_n)}}\right) \quad \text{a.s.}$$

Moreover, since $\mathbf{X}$ is bounded, there exists a positive constant M such that

$$\sum_{i=1}^n \|\mathbf{x} - \mathbf{X}_i\| \, \mathbf{1}_{\mathcal{E}_i} \leq M \sum_{i=1}^n \mathbf{1}_{\mathcal{E}_i} \quad \text{a.s.}$$

Clearly, $\sum_{i=1}^n \mathbf{1}_{\mathcal{E}_i}$ has a Binomial distribution with parameters n and $F_{\mathbf{x}}(h_n)$ and consequently, by Bennett's inequality,

$$\sum_{i=1}^n \|\mathbf{x} - \mathbf{X}_i\| \, \mathbf{1}_{\mathcal{E}_i} = \mathcal{O}\left(nF_{\mathbf{x}}(h_n)\right) \quad \text{a.s.}$$

This completes the proof of the lemma. ∎

Lemma 5.2 *Assume that Assumption $\mathbf{R_x}$ is satisfied and $\mathbf{X}$ is bounded. Then, if $h_n \downarrow 0$ and $nh_n^2/\log n \to \infty$,*

$$\frac{1}{nF_{\mathbf{x}}(h_n)} \sum_{i=1}^n \left(\mathbf{1}_{\mathcal{E}_i \cap \hat{\mathcal{E}}_i^c} + \mathbf{1}_{\mathcal{E}_i^c \cap \hat{\mathcal{E}}_i}\right) \to 0 \quad a.s. \quad as \ n \to \infty.$$

Proof of Lemma 5.2   Define $\kappa_n = C_\kappa \sqrt{\frac{\log n}{nh_n^2}}$ and $\eta_n = \kappa_n h_n$, where C_κ is a constant which will be chosen later. Observe that

$$\mathcal{E}_i \cap \hat{\mathcal{E}}_i^c = \left[\{h_n - \eta_n < V_i \leq h_n\} \cap \hat{\mathcal{E}}_i^c\right] \cup \left[\{V_i \leq h_n - \eta_n\} \cap \hat{\mathcal{E}}_i^c\right].$$

Similarly

$$\hat{\mathcal{E}}_i^c = \left\{\hat{V}_i > h_n\right\} = \left\{\hat{V}_i - V_i > h_n - V_i\right\}.$$

Consequently, we may write

$$
\begin{aligned}
\sum_{i=1}^n \mathbf{1}_{\mathcal{E}_i \cap \hat{\mathcal{E}}_i^c} &\leq \sum_{i=1}^n \mathbf{1}_{\{h_n - \eta_n < V_i \leq h_n\}} + \sum_{i=1}^n \mathbf{1}_{\{V_i \leq h_n - \eta_n\} \cap \hat{\mathcal{E}}_i^c} \\
&= \sum_{i=1}^n \mathbf{1}_{\{h_n - \eta_n < V_i \leq h_n\}} + \sum_{i=1}^n \mathbf{1}_{\left\{V_i \leq h_n - \eta_n, \hat{V}_i - V_i > h_n - V_i\right\}} \\
&\leq \sum_{i=1}^n \mathbf{1}_{\{h_n - \eta_n < V_i \leq h_n\}} + \sum_{i=1}^n \mathbf{1}_{\left\{\hat{V}_i - V_i > \eta_n\right\}} \\
&\leq \sum_{i=1}^n \mathbf{1}_{\{h_n - \eta_n < V_i \leq h_n\}} + \sum_{i=1}^n \mathbf{1}_{\left\{|\hat{V}_i - V_i| > \eta_n\right\}}. \quad (5.1)
\end{aligned}
$$

By Bennett's inequality, we have

$$\sum_{i=1}^n \mathbf{1}_{\{h_n - \eta_n < V_i \leq h_n\}} \sim n\mathbb{P}\left(h_n - \eta_n < V_1 \leq h_n\right) \quad \text{a.s. as } n \to \infty,$$

whence

$$\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\{h_n-\eta_n<V_i\leq h_n\}} \sim \frac{F_{\mathbf{x}}(h_n)-F_{\mathbf{x}}(h_n-\eta_n)}{F_{\mathbf{x}}(h_n)} \quad \text{a.s.} \qquad (5.2)$$

But, using the fact that $F_{\mathbf{x}}$ is regularly varying with index D, we may write

$$1-\frac{F_{\mathbf{x}}(h_n-\eta_n)}{F_{\mathbf{x}}(h_n)} \sim 1-(1-\kappa_n)^D \sim D\kappa_n \to 0 \quad \text{as } n\to\infty. \qquad (5.3)$$

(Note that the value of the index D influences constants only.)

Next, since $\mathbf{X}$ is bounded,

$$\sum_{i=1}^{n}\mathbf{1}_{\{|\hat{V}_i-V_i|>\eta_n\}} \leq \sum_{i=1}^{n}\mathbf{1}_{\{\|\hat{\Pi}_D-\Pi_D\|_{\infty}\|\mathbf{x}-\mathbf{X}_i\|>\eta_n\}}$$

$$\leq n\mathbf{1}_{\{\|\hat{\Pi}_D-\Pi_D\|_{\infty}>\eta_n/M\}} \quad \text{a.s.}$$

for some positive constant M. By statement (i) of Theorem 2.1, we have

$$\mathbb{P}\left(\left\|\hat{\Pi}_D-\Pi_D\right\|_{\infty}\geq\varepsilon\right) = \mathcal{O}\left(\exp\left(-Cn\varepsilon^2\right)\right).$$

Therefore,

$$\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\{|\hat{V}_i-V_i|>\eta_n\}} \to 0 \quad \text{a.s.} \qquad (5.4)$$

whenever

$$\sum_{n=1}^{\infty}\exp\left(-Cn\eta_n^2/M^2\right) < \infty.$$

Observing that $n\eta_n^2 = C_{\kappa}^2\log n$, we see that the summability condition above is fulfilled as soon as C_{κ} is large enough.

Combining inequality (5.1) with (5.2)-(5.3) and (5.4), we conclude that

$$\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\mathcal{E}_i\cap\hat{\mathcal{E}}_i^c} \to 0 \quad \text{a.s. as } n\to\infty.$$

One shows with similar arguments that

$$\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\mathcal{E}_i^c\cap\hat{\mathcal{E}}_i} \to 0 \quad \text{a.s. as } n\to\infty.$$

$\blacksquare$

Lemma 5.3 *Assume that* $\mathbf{X}$ *is bounded. Then, if* $nF_{\mathbf{x}}(h_n)\to\infty$,

$$\mathbb{E}\left[\frac{\left\|\hat{\Pi}_D-\Pi_D\right\|_{\infty}}{nh_nF_{\mathbf{x}}(h_n)}\left(\sum_{i=1}^{n}\|\mathbf{x}-\mathbf{X}_i\|\,\mathbf{1}_{\mathcal{E}_i}\right)\right]^2 = \mathcal{O}\left(\frac{1}{nh_n^2}\right).$$

19

Proof of Lemma 5.3 For $i = 1, \ldots, n$, let

$$U_i = \|\mathbf{x} - \mathbf{X}_i\| \, \mathbf{1}_{\mathcal{E}_i} - \mathbb{E}\left[\|\mathbf{x} - \mathbf{X}_i\| \, \mathbf{1}_{\mathcal{E}_i}\right],$$

and write

$$\mathbb{E}\left[\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty \left(\sum_{i=1}^n \|\mathbf{x} - \mathbf{X}_i\| \, \mathbf{1}_{\mathcal{E}_i}\right)\right]^2$$

$$\leq 2\mathbb{E}\left[\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty \sum_{i=1}^n U_i\right]^2$$

$$+ 2n^2 \mathbb{E}^2\left[\|\mathbf{x} - \mathbf{X}\| \, \mathbf{1}_{\mathcal{E}_1}\right] \mathbb{E}\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty^2.$$

By Theorem 2.1 (iii),

$$\mathbb{E}\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty^2 = \mathcal{O}\left(\frac{1}{n}\right),$$

and, since $\mathbf{X}$ is bounded,

$$\mathbb{E}^2\left[\|\mathbf{x} - \mathbf{X}\| \, \mathbf{1}_{\mathcal{E}_1}\right] = \mathcal{O}\left(F_{\mathbf{x}}^2(h_n)\right).$$

Consequently,

$$n^2 \mathbb{E}^2\left[\|\mathbf{x} - \mathbf{X}\| \, \mathbf{1}_{\mathcal{E}_1}\right] \mathbb{E}\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty^2 = \mathcal{O}\left(nF_{\mathbf{x}}^2(h_n)\right).$$

One easily shows, with methods similar to the ones used to prove (iii) of Theorem 2.1, that

$$\mathbb{E}\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty^4 = \mathcal{O}\left(\frac{1}{n^2}\right).$$

Moreover, simple computations lead to

$$\mathbb{E}\left[\sum_{i=1}^n U_i\right]^4 = \mathcal{O}\left(nF_{\mathbf{x}}(h_n) + n^2 F_{\mathbf{x}}^2(h_n)\right) = \mathcal{O}\left(n^2 F_{\mathbf{x}}^2(h_n)\right),$$

when $nF_{\mathbf{x}}(h_n) \to \infty$. Consequently, by Cauchy-Schwarz inequality,

$$\mathbb{E}\left[\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty \sum_{i=1}^n U_i\right]^2 = \mathcal{O}\left(\frac{1}{n}\left(\mathbb{E}\left[\sum_{i=1}^n U_i\right]^4\right)^{1/2}\right)$$

$$= \mathcal{O}\left(F_{\mathbf{x}}(h_n)\right).$$

Putting all the pieces together, we obtain

$$\mathbb{E}\left[\frac{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty}{nh_n F_{\mathbf{x}}(h_n)}\left(\sum_{i=1}^n \|\mathbf{x} - \mathbf{X}_i\| \, \mathbf{1}_{\mathcal{E}_i}\right)\right]^2 = \mathcal{O}\left(\frac{1}{nh_n^2}\right) \quad \text{as } n \to \infty.$$

$\blacksquare$

Lemma 5.4 *Assume that Assumption* $\mathbf{R_x}$ *is satisfied and* $\mathbf{X}$ *is bounded. Then, if* $h_n \downarrow 0$ *and* $nh_n^2/\log n \to \infty$,

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\left(\mathbf{1}_{\mathcal{E}_i\cap\hat{\mathcal{E}}_i^c} + \mathbf{1}_{\mathcal{E}_i^c\cap\hat{\mathcal{E}}_i}\right)\right]^2 \to 0 \quad \text{as } n \to \infty.$$

Proof of Lemma 5.4 The proof is close to the derivation of Lemma 5.2—almost sure convergence is replaced here by convergence in mean square. Therefore, we go quickly through it.

Because of $(a+b)^2 \le 2a^2 + 2b^2$, it is enough to prove that

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\mathcal{E}_i\cap\hat{\mathcal{E}}_i^c}\right]^2 \to 0$$

and

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\mathcal{E}_i^c\cap\hat{\mathcal{E}}_i}\right]^2 \to 0 \quad \text{as } n \to \infty.$$

We will focus on the first limit only—proof of the second one is similar. With the notation of Lemma 5.2, by inequality (5.1),

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\mathcal{E}_i\cap\hat{\mathcal{E}}_i^c}\right]^2 \le 2\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\{h_n-\eta_n<V_i\le h_n\}}\right]^2$$

$$+ 2\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\{|\hat{V}_i-V_i|>\eta_n\}}\right]^2, \qquad (5.5)$$

where η_n is a tuning parameter which will be fixed later.

The first term on the right of (5.5) is handled exactly as in Lemma 5.2 and tends to zero. We just require that $\eta_n = \kappa_n h_n$, with $\kappa_n \to 0$.

With respect to the second term, write

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\{|\hat{V}_i-V_i|>\eta_n\}}\right]^2 \le \mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\{\|\hat{\Pi}_D-\Pi_D\|_\infty\|\mathbf{x}-\mathbf{X}_i\|>\eta_n\}}\right]^2$$

$$\le \mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\{\|\hat{\Pi}_D-\Pi_D\|_\infty>\eta_n/M\}}\right]^2$$

$$\le \frac{1}{F_{\mathbf{x}}^2(h_n)}\mathbb{P}\left(\left\|\hat{\Pi}_D-\Pi_D\right\|_\infty > \eta_n/M\right),$$

for some positive constant M. Applying finally statement (i) of Theorem 2.1, we obtain

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\{|\hat{V}_i-V_i|>\eta_n\}}\right]^2 = \mathcal{O}\left(\frac{\exp\left(-Cn\kappa_n^2/M^2\right)}{F_{\mathbf{x}}^2(h_n)}\right)$$

which tends to zero whenever $\kappa_n = C_\kappa\sqrt{\frac{\log n}{nh_n^2}}$ for a sufficiently large C_δ. ∎

21

Lemma 5.5 *Assume that Assumption* $\mathbf{R_x}$ *is satisfied and* $\mathbf{X}$ *is bounded. Then, if* $nF_{\mathbf{x}}(h_n) \to \infty$,

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\left(\mathbf{1}_{\mathcal{E}_i \cap \hat{\mathcal{E}}_i^c} + \mathbf{1}_{\mathcal{E}_i^c \cap \hat{\mathcal{E}}_i}\right)\right]^2 = \mathcal{O}\left(\frac{\log(nh_n^2)}{nh_n^2}\right).$$

Proof of Lemma 5.5   We deal only with the term

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\mathcal{E}_i \cap \hat{\mathcal{E}}_i^c}\right]^2,$$

since the other one may be addressed the same way. At this point, we have to get sharper into the bounds derived in Lemma 5.4. Let $(\kappa_n)_{n\geq 1}$ be a positive sequence which tends to 0, and recall that

$$\begin{aligned}
&\mathcal{E}_i \cap \hat{\mathcal{E}}_i^c \\
&\quad = \{V_i \leq h_n\} \cap \left\{\hat{V}_i > h_n\right\} \\
&\quad \subset \{V_i \leq h_n\} \cap \left\{h_n < V_i + \left\|\left(\hat{\Pi}_D - \Pi_D\right)(\mathbf{x} - \mathbf{X}_i)\right\|\right\} \\
&\quad \subset \{V_i \leq h_n\} \cap \left[\{h_n(1 - \kappa_n) < V_i\} \cup \left\{\left\|\left(\hat{\Pi}_D - \Pi_D\right)(\mathbf{x} - \mathbf{X}_i)\right\| > \kappa_n h_n\right\}\right] \\
&\quad = \{h_n(1 - \kappa_n) < V_i \leq h_n\} \\
&\qquad \cup \left[\{V_i \leq h_n\} \cap \left\{\left\|\left(\hat{\Pi}_D - \Pi_D\right)(\mathbf{x} - \mathbf{X}_i)\right\| > \kappa_n h_n\right\}\right].
\end{aligned}$$

Thus

$$\begin{aligned}
&\sum_{i=1}^{n}\mathbf{1}_{\mathcal{E}_i \cap \hat{\mathcal{E}}_i^c} \\
&\quad \leq \sum_{i=1}^{n}\mathbf{1}_{\{h_n(1-\kappa_n)<V_i\leq h_n\}} + \sum_{i=1}^{n}\mathbf{1}_{\left\{\left\|(\hat{\Pi}_D-\Pi_D)(\mathbf{x}-\mathbf{X}_i)\right\|>\kappa_n h_n\right\}}\mathbf{1}_{\{V_i\leq h_n\}},
\end{aligned}$$

and therefore

$$\begin{aligned}
\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\mathcal{E}_i \cap \hat{\mathcal{E}}_i^c}\right]^2 &\leq 2\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\{h_n(1-\kappa_n)<V_i\leq h_n\}}\right]^2 \\
&\quad + 2\left[\sum_{i=1}^{n}\mathbf{1}_{\left\{\left\|(\hat{\Pi}_D-\Pi_D)(\mathbf{x}-\mathbf{X}_i)\right\|>\kappa_n h_n\right\}}\mathbf{1}_{\{V_i\leq h_n\}}\right]^2.
\end{aligned}$$
(5.6)

Taking expectations and mimicking the method used in the proof of Lemma 5.2, we easily obtain

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^{n}\mathbf{1}_{\{h_n(1-\kappa_n)<V_i\leq h_n\}}\right]^2 \leq C\kappa_n^2. \tag{5.7}$$

It remains to bound the last term on the right-hand side of (5.6). To this aim, using the fact that $\mathbf{X}$ is bounded, we may write

$$\mathbf{1}_{\left\{\left\|(\hat{\Pi}_D - \Pi_D)(\mathbf{x} - \mathbf{X}_i)\right\| > \kappa_n h_n\right\}} \mathbf{1}_{\{V_i \leq h_n\}}$$
$$\leq \mathbf{1}_{\left\{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty > \kappa_n h_n / M\right\}} \mathbf{1}_{\{V_i \leq h_n\}}$$
$$(\text{for some positive } M)$$
$$\leq \mathbf{1}_{\mathcal{A}_n^c} \mathbf{1}_{\{V_i \leq h_n\}} + \mathbf{1}_{\left\{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty > \kappa_n h_n / M\right\} \cap \mathcal{A}_n} \mathbf{1}_{\{V_i \leq h_n\}},$$

where the set $\mathcal{A}_n$ is the same as in Lemma 2.1. Clearly, by Cauchy-Schwarz inequality,

$$\mathbb{E}\left[\mathbf{1}_{\mathcal{A}_n^c}\left(\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^n \mathbf{1}_{\{V_i \leq h_n\}}\right)\right]^2$$
$$\leq \mathbb{P}^{1/2}\left(\mathcal{A}_n^c\right)\mathbb{E}^{1/2}\left[\left(\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^n \mathbf{1}_{\{V_i \leq h_n\}}\right)^4\right]$$
$$= \mathcal{O}\left(\exp\left(-Cn\right)\right), \tag{5.8}$$

where the last inequality arises from Lemma 2.1. It remains to bound the term

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^n \mathbf{1}_{\left\{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty > \kappa_n h_n / M\right\} \cap \mathcal{A}_n} \mathbf{1}_{\{V_i \leq h_n\}}\right]^2.$$

We have

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^n \mathbf{1}_{\left\{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty > \kappa_n h_n / M\right\} \cap \mathcal{A}_n} \mathbf{1}_{\{V_i \leq h_n\}}\right]^2$$
$$= \mathbb{E}\left[\mathbf{1}_{\left\{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty > \kappa_n h_n / M\right\} \cap \mathcal{A}_n}\left(\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^n \mathbf{1}_{\{V_i \leq h_n\}}\right)^2\right].$$

Using again Cauchy-Schwarz inequality and a bound on the fourth moment of $(1/nF_{\mathbf{x}}(h_n))\sum_{i=1}^n \mathbf{1}_{\{V_i \leq h_n\}}$, it suffices to bound accurately

$$\mathbb{P}\left(\left\{\left\|\hat{\Pi}_D - \Pi_D\right\|_\infty > \kappa_n h_n / M\right\} \cap \mathcal{A}_n\right) \leq \mathbb{P}\left(\left\|\Gamma_n - \Gamma\right\|_\infty > \kappa_n h_n C / M\right)$$
$$= \mathcal{O}\left(\exp\left(-Cn\kappa_n^2 h_n^2\right)\right), \tag{5.9}$$

where we used the bound (2.3) and statement (i) in Theorem 2.1.

Collecting the bounds (5.6)-(5.7)-(5.8)-(5.9), we finally obtain

$$\mathbb{E}\left[\frac{1}{nF_{\mathbf{x}}(h_n)}\sum_{i=1}^n \mathbf{1}_{\mathcal{E}_i \cap \hat{\mathcal{E}}_i^c}\right]^2 = \mathcal{O}\left(\kappa_n^2\right) + \mathcal{O}\left(\exp\left(-Cn\kappa_n^2 h_n^2\right)\right).$$

The choice $\kappa_n^{\star 2} \asymp \exp\left(-Cn\kappa_n^{\star 2} h_n^2\right)$, i.e.,

$$\kappa_n^{\star 2} \asymp \frac{\log\left(nh_n^2\right)}{nh_n^2}$$

leads to the desired result. ■

As promised, we conclude this section with two lemmas, were we relax the boundedness constraints and replace them by higher moment conditions. The first lemma is devoted to density and should be inserted within Lemma 5.1 and echoes the first assessment in Proposition 3.1. The second lemma deals with regression and is connected with Proposition 4.1. The proof of the first lemma may be easily derived from the second one, and it is therefore omitted.

Lemma 5.6 *Assume that* $\|\mathbf{x} - \mathbf{X}\|$ *satisfies Assumption* **C**. *Then, provided* $nF_{\mathbf{x}}(h_n)/\log n \to \infty$,

$$\sum_{i=1}^{n} \|\mathbf{x} - \mathbf{X}_i\| \, \mathbf{1}_{\mathcal{E}_i} = \mathcal{O}\left(nF_{\mathbf{x}}(h_n)\right) \quad a.s.$$

Lemma 5.7 *Let* $\varepsilon = Y - \mathbb{E}\left[Y \mid \|\Pi_D(\mathbf{x} - \mathbf{X})\|\right]$ *be the noise part in the regression model of* Y *on* $\|\Pi_D(\mathbf{x} - \mathbf{X})\|$. *Assume that, for all integers* $k \geq 2$,

$$\mathbb{E}\left[|\varepsilon|^k \,|\, \|\Pi_D(\mathbf{x} - \mathbf{X})\|\right] = \mathbb{E}|\varepsilon|^k.$$

Assume also that $\mathbf{X}$ *is bounded,* ε *satisfies Assumption* **C**, *and* $nF_{\mathbf{x}}(h_n)/\log n \to \infty$. *Then*

$$\sum_{i=1}^{n} Y_i \|\mathbf{x} - \mathbf{X}_i\| \, \mathbf{1}_{\mathcal{E}_i} = \mathcal{O}\left(nF_{\mathbf{x}}(h_n)\right) \quad a.s.$$

Proof of Lemma 5.7 Set $V = \|\Pi_D(\mathbf{x} - \mathbf{X})\|$ and $\mathcal{E} = \{V \leq h_n\}$. Our goal is to prove that

$$\mathbb{E}\left[Y^2 \|\mathbf{x} - \mathbf{X}\|^2 \mathbf{1}_{\mathcal{E}}\right] \leq C_1 F_{\mathbf{x}}(h_n) \tag{5.10}$$

for some positive C_1 and that, for all integers $k \geq 3$,

$$\mathbb{E}\left[|Y|^k \|\mathbf{x} - \mathbf{X}\|^k \mathbf{1}_{\mathcal{E}}\right] \leq \frac{k!}{2} sb^{k-2}, \tag{5.11}$$

with $s = C_1 F_{\mathbf{x}}(h_n)$ and b is some positive constant. From these bounds, we will deduce by Bernstein's exponential inequality that, for all positive M,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i \|\mathbf{x} - \mathbf{X}_i\| \, \mathbf{1}_{\mathcal{E}_i} - \mathbb{E}\left[\sum_{i=1}^{n} Y_i \|\mathbf{x} - \mathbf{X}_i\| \, \mathbf{1}_{\mathcal{E}_i}\right]\right| \geq MnF_{\mathbf{x}}(h_n)\right)$$
$$\leq \exp\left(-CM^2 nF_{\mathbf{x}}(h_n)\right)$$
$$= \mathcal{O}\left(n^{-M}\right).$$

This inequality, together with the Borel-Cantelli lemma, leads to the result.

Set $\varepsilon = Y - \mathbb{E}[Y|V]$ and consider the regression model

$$Y = r_D(V) + \varepsilon,$$

where $r_D(V) = \mathbb{E}[Y|V]$. Since $\mathbb{E}[Y|\Pi_D \mathbf{X} = \Pi_D \mathbf{u}]$ is bounded when $\mathbf{u}$ lies in a neighborhood of $\mathbf{x}$, the function r_D is also bounded in a neighborhood of zero by, say, some positive constant τ. Then, it is easily seen that $\mathbb{E}[|r_D(V)|^k \mathbf{1}_{\mathcal{E}}] \leq$

$\tau^k F_{\mathbf{x}}(h_n)$. From this inequality and the assumptions of the lemma, we deduce that

$$\mathbb{E}\left[|Y|^k \mathbf{1}_{\mathcal{E}}\right] \leq 2^k \left(\mathbb{E}\left[|\varepsilon|^k \mathbf{1}_{\mathcal{E}}\right] + \tau^k F_{\mathbf{x}}(h_n)\right)$$
$$\leq 2^k F_{\mathbf{x}}(h_n) \left(\frac{k!}{2} v c^{k-2} + \tau^k\right). \tag{5.12}$$

Thus, we obtain

$$\mathbb{E}\left[|Y|^2 \|\mathbf{x} - \mathbf{X}\|^2 \mathbf{1}_{\mathcal{E}}\right] \leq \Lambda^2 \mathbb{E}\left[|Y|^2 \mathbf{1}_{\mathcal{E}}\right]$$
$$\leq 4\Lambda^2 \left(v + \tau^2\right) F_{\mathbf{x}}(h_n)$$
$$= C_1 F_{\mathbf{x}}(h_n),$$

where $C_1 = 4\Lambda^2 \left(v + \tau^2\right)$. This shows inequality (5.10). We now proceed to prove (5.11). To this aim, keep (5.12) in mind, and consider the quantity $\mathbb{E}[|Y|^k \|\mathbf{x} - \mathbf{X}\|^k \mathbf{1}_{\mathcal{E}}]$. We have, for all $k \geq 3$,

$$\mathbb{E}\left[|Y|^k \|\mathbf{x} - \mathbf{X}\|^k \mathbf{1}_{\mathcal{E}}\right]$$
$$\leq \Lambda^k \mathbb{E}\left[|Y|^k \mathbf{1}_{\mathcal{E}}\right]$$
$$\leq \Lambda^k 2^k F_{\mathbf{x}}(h_n) \left(\frac{k!}{2} v c^{k-2} + \tau^k\right)$$
$$= \frac{\Lambda^k 2^k \left(\frac{k!}{2} v c^{k-2} + \tau^k\right)}{C_1} C_1 F_{\mathbf{x}}(h_n)$$
$$\leq \frac{k!}{2} C_1 F_{\mathbf{x}}(h_n) \left(2\Lambda\right)^{k-2} \frac{v c^{k-2} + \tau^k}{v + \tau^2}$$
$$\leq \frac{k!}{2} C_1 F_{\mathbf{x}}(h_n) \left(2\Lambda \max(c, \tau)\right)^{k-2}.$$

This proves (5.11).

We are now ready to apply Bernstein's exponential inequality. Assumption **C** holds with $\xi = Y \|\mathbf{x} - \mathbf{X}\| \mathbf{1}_{\mathcal{E}}$ by setting $b = 2\Lambda \max(c, \tau)$ and $s = C_1 F_{\mathbf{x}}(h_n)$. ∎

References

[1] Bellman, R.E. (1961). *Adaptive Control Processes: A Guided Tour*, Princeton University Press, Princeton.

[2] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variations*, Encyclopedia of Mathematics and Its Applications, Cambridge University Press, Cambridge.

[3] Bosq, D. (2000). *Linear Processes in Function Spaces*, Lecture Notes in Statistics **149**, Springer-Verlag, New York.

[4] Cardot, H., Mas, A. and Sarda, P. (2007). CLT in functional linear regression models, *Probability Theory and Related Fields*, **138**, 325-361.

[5] Dauxois, J.-Y., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a random vector function: Some applications to statistical inference, *Journal of Multivariate Analysis*, **12**, 136-154.

[6] Delaigle, A. and Hall, P. (2010). Defining probability density for a distribution of random functions, *The Annals of Statistics*, **38**, 1171-1193.

[7] Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*, Springer-Verlag, New York.

[8] Dunford, N. and Schwartz, J.T. (1988). *Linear Operators. Part I, II, III*, John Wiley and Sons, Hoboken.

[9] Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies, *The Annals of Statistics*, **21**, 196-216.

[10] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, Springer-Verlag, New York.

[11] Friedman, J.H. (1987). Exploratory projection pursuit, *Journal of the American Statistical Association*, **82**, 259-266.

[12] Gohberg, I., Goldberg, S. and Kaashoek, M.A. (1991). *Classes of Linear Operators. Vol. I, II*, Birkhäuser, Basel.

[13] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*, Springer-Verlag, New York.

[14] Horowitz, J.L. (1998). *Semiparametric Methods in Econometrics*, Springer-Verlag, New York.

[15] Jee, J.E. (1987). Exploratory projection pursuit using nonparametric density estimation, *Proceedings of the Statistical Computing Section of the American Statistical Association*, 335-339.

[16] Jolliffe, I.T. (2002). *Principal Component Analysis*, 2nd edition, Springer-Verlag, New York.

[17] Mas, A. and Menneteau, L. (2003). Perturbation approach applied to the asymptotic study of random operators, *Progress in Probability*, **55**, 127-134.

[18] Massart, P. (2007). *Concentration Inequalities and Model Selection*, Springer-Verlag, Berlin.

[19] McCullagh, P. and Nelder, J.A. (1983). *Generalized Linear Models*, Chapman and Hall, London.

[20] Nadaraya, E.A. (1964). On estimating regression, *Theory of Probability and Its Applications*, **9**, 141-142.

[21] Nadaraya, E.A. (1970). Remarks on nonparametric estimates for density functions and regression curves, *Theory of Probability and Its Applications*, **15**, 134-137.

[22] Parzen, E. (1962). On the estimation of a probability density function and the mode, *The Annals of Mathematical Statistics*, **33**, 1065-1076.

[23] Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation (Probability and Mathematical Statistics)*, Academic Press, London.

[24] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *The Annals of Mathematical Statistics*, **27**, 832-837.

[25] Rudin, W. (1987). *Real and Complex Analysis*, Mc Graw-Hill, New York.

[26] Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley and Sons, Hoboken.

[27] Scott, D.W. and Wand, M.P. (1991). Feasibility of multivariate density estimates, *Biometrika*, **78**, 197-205.

[28] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

[29] Simonoff, J.S. (1996). *Smoothing Methods in Statistics*, Springer-Verlag, New York.

[30] Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators, *The Annals of Statistics*, **8**, 1348-1360.

[31] Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression, *The Annals of Statistics*, **10**, 1040-1053.

[32] Watson, G.S. (1964). Smooth regression analysis, *Sankhyā Series A*, **26**, 359-372.

[33] Wheeden, R.L. and Zygmund, A. (1977). *Measure and Integral. An Introduction to Real Analysis*, Marcel Dekker, New York.