SOME INSIGHTS ABOUT THE SMALL BALL PROBABILITY FACTORIZATION FOR HILBERT RANDOM ELEMENTS

Enea G. Bongiorno and Aldo Goia

Università del Piemonte Orientale

enea.bongiorno@uniupo.it, aldo.goia@uniupo.it

Supplementary Material

This document collects the detailed proofs of the results presented in the paper "Some Insights About the Small Ball Probability Factorization for Hilbert Random Elements", by Enea G. Bongiorno and Aldo Goia.

Proof of Theorem 1

We are interested in the asymptotic behaviour, whenever ε tends to zero, of the SmBP of the process X, that is

$$\begin{aligned}\varphi(x,\varepsilon) &= \mathbb{P}\left(\|X-x\|\leq\varepsilon\right) = \mathbb{P}\left(\|X-x\|^2\leq\varepsilon^2\right)\\ &= \mathbb{P}\left(\sum_{j=1}^{+\infty}\langle X-x,\xi_j\rangle^2\leq\varepsilon^2\right)\\ &= \mathbb{P}\left(\sum_{j=1}^{+\infty}\left(\theta_j-x_j\right)^2\leq\varepsilon^2\right), \qquad \text{as } \varepsilon\to 0\end{aligned}$$

Let $S_1=\sum_{j\leq d}\left(\theta_j-x_j\right)^2$ and $S=\frac{1}{\varepsilon^2}\sum_{j\geq d+1}\left(\theta_j-x_j\right)^2$ be the truncated series and the scaled version of the remainder respectively. Thus, the SmBP is

$$\begin{aligned}\varphi(x,\varepsilon) =& \mathbb{P}\left(S_1+\varepsilon^2 S\leq\varepsilon^2\right) = \mathbb{P}\left(S_1\leq\varepsilon^2\left(1-S\right)\right)\\ =& \mathbb{P}\left(\left\{S_1\leq\varepsilon^2\left(1-S\right)\right\}\cap\left\{S\geq 1\right\}\right)+\\ &+\mathbb{P}\left(\left\{S_1\leq\varepsilon^2\left(1-S\right)\right\}\cap\left\{0\leq S<1\right\}\right)\\ =& \mathbb{P}\left(\left\{S_1\leq\varepsilon^2\left(1-S\right)\right\}\cap\left\{0\leq S<1\right\}\right)\\ =& \int_0^1\varphi(s|x,\varepsilon,d)dG\left(s\right) \qquad \text{(S1.1)}\end{aligned}$$

where G is the cumulative distribution function of S. At first, for any $s\in(0,1)$, let us consider $\varphi(s|x,\varepsilon,d)$, that is the SmBP about $\Pi_d x$ of the process $\Pi_d X$ in the space spanned by $\{\xi_j\}_{j\leq d}$. In terms of $f_d(\cdot)$, the probability density function of $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_d)'$, it can be written as

$$\varphi(s|x,\varepsilon,d) = \int_{D^x} f_d(\boldsymbol{\vartheta})\, d\boldsymbol{\vartheta},$$

where $D = D^x = \left\{\boldsymbol{\vartheta} \in \mathbb{R}^d : \sum_{j\leq d} (\vartheta_j - x_j)^2 \leq \varepsilon^2 (1-s)\right\}$ is a d–dimensional ball centered about $\Pi_d x = (x_1, \ldots, x_d)$ with radius $\varepsilon\sqrt{1-s}$. Now, consider the Taylor expansion of $f = f_d$ about $\Pi x = \Pi_d x$,

$$\begin{aligned} f(\boldsymbol{\vartheta}) =& f(x_1,\ldots,x_d) + \langle \boldsymbol{\vartheta} - \Pi x, \nabla f(x_1,\ldots,x_d)\rangle + \\ &+ \frac{1}{2}(\boldsymbol{\vartheta} - \Pi x)' H_f\left(\Pi x + (\boldsymbol{\vartheta} - \Pi x)t\right)(\boldsymbol{\vartheta} - \Pi x), \end{aligned}$$

for some $t \in (0,1)$ and with H_f denoting the Hessian matrix of f. (In general, t depends on $\boldsymbol{\vartheta} - \Pi x$, but we are not interested in the actual value of it because the boundedness of the second derivatives of f allows us to drop, in what follows, those terms depending on t). Then we can write

$$\begin{aligned} \varphi(s|x,\varepsilon,d) =& \int_D \Big(f(x_1,\ldots,x_d) + \langle \boldsymbol{\vartheta} - \Pi x, \nabla f(x_1,\ldots,x_d)\rangle + \\ &+ \frac{1}{2}(\boldsymbol{\vartheta} - \Pi x)' H_f\left(\Pi x + (\boldsymbol{\vartheta} - \Pi x)t\right)(\boldsymbol{\vartheta} - \Pi x)\Big)\, d\boldsymbol{\vartheta} \\ =& f(x_1,\ldots,x_d)\int_D d\boldsymbol{\vartheta} + \int_D \langle \boldsymbol{\vartheta} - \Pi x, \nabla f(x_1,\ldots,x_d)\rangle\, d\boldsymbol{\vartheta} + \\ &+ \frac{1}{2}\int_D (\boldsymbol{\vartheta} - \Pi x)' H_f\left(\Pi x + (\boldsymbol{\vartheta} - \Pi x)t\right)(\boldsymbol{\vartheta} - \Pi x) d\boldsymbol{\vartheta} \\ =& f(x_1,\ldots,x_d) I + \\ &+ \frac{1}{2}\int_D (\boldsymbol{\vartheta} - \Pi x)' H_f\left(\Pi x + (\boldsymbol{\vartheta} - \Pi x)t\right)(\boldsymbol{\vartheta} - \Pi x) d\boldsymbol{\vartheta} \qquad \text{(S1.2)} \end{aligned}$$

where $I = I(s,\varepsilon,d)$ denotes the volume of D that is

$$I = \frac{\varepsilon^d \pi^{d/2}}{\Gamma(d/2+1)} (1-s)^{d/2}$$

and, the addend $\int_D \langle \boldsymbol{\vartheta} - \Pi x, \nabla f(x_1,\ldots,x_d)\rangle\, d\boldsymbol{\vartheta}$ is null since the integrand is a linear functional integrated over the symmetric – with respect to the center $(x_1, \ldots, x_d)$ – domain D. Thus from (S1.2), thanks to: the boundedness of second derivatives (2.3), the fact that symmetry arguments lead to $\int_D (\vartheta_i - x_i)(\vartheta_j - x_j) d\boldsymbol{\vartheta} = 0$ for $i \neq j$ and monotonicity of eigenvalues, it follows

$$
\begin{aligned}
&|\varphi(s|x,\varepsilon,d) - f(x_1,\ldots,x_d)I| = \\
&\qquad = \left| \frac{1}{2} \int_D \sum_{i\leq d} \sum_{j\leq d} (\vartheta_i - x_i)(\vartheta_j - x_j) \frac{\partial^2 f}{\partial\vartheta_i \partial\vartheta_j} \left(\Pi x + (\boldsymbol{\vartheta} - \Pi x)t\right) d\boldsymbol{\vartheta} \right| \\
&\qquad \leq \frac{1}{2} C_2 f(x_1,\ldots,x_d) \left| \sum_{i\leq d} \sum_{j\leq d} \int_D \frac{(\vartheta_i - x_i)(\vartheta_j - x_j)}{\sqrt{\lambda_i}\sqrt{\lambda_j}} d\boldsymbol{\vartheta} \right| \\
&\qquad = \frac{1}{2} C_2 f(x_1,\ldots,x_d) \int_D \sum_{j\leq d} \frac{(\vartheta_j - x_j)^2}{\lambda_j} d\boldsymbol{\vartheta} \\
&\qquad \leq \frac{C_2}{2\lambda_d} f(x_1,\ldots,x_d) \int_D \sum_{j\leq d} (\vartheta_j - x_j)^2 d\boldsymbol{\vartheta}.
\end{aligned}
$$

Note that

$$\int_D \sum_{j\leq d} (\vartheta_j - x_j)^2 d\boldsymbol{\vartheta} = \int_{\|\boldsymbol{\vartheta}\|^2_{\mathbb{R}^d} \leq \varepsilon^2(1-s)} \|\boldsymbol{\vartheta}\|^2_{\mathbb{R}^d} d\boldsymbol{\vartheta}$$

whose integrand is a radial function (i.e. a map $H : \mathbb{R}^d \to \mathbb{R}$ such that $H(\boldsymbol{\vartheta}) = h(\|\boldsymbol{\vartheta}\|_{\mathbb{R}^d})$ with $h : \mathbb{R} \to \mathbb{R}$), for which the following identity applies

$$\int_{\|\boldsymbol{\vartheta}\|_{\mathbb{R}^d} \leq R} H(\boldsymbol{\vartheta}) d\boldsymbol{\vartheta} = \omega_{d-1} \int_0^R h(\rho)\rho^{d-1} d\rho,$$

where ω_{d-1} denotes the surface area of the sphere of radius 1 in $\mathbb{R}^d$. Hence

$$\int_{\|\boldsymbol{\vartheta}\|^2_{\mathbb{R}^d} \leq \varepsilon^2(1-s)} \|\boldsymbol{\vartheta}\|^2_{\mathbb{R}^d} d\boldsymbol{\vartheta} = \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^{\varepsilon\sqrt{1-s}} \rho^{d+1} d\rho = \frac{d}{(d+2)} I\varepsilon^2(1-s) \leq I\varepsilon^2,$$

where the latter inequality follows from the fact that $s \in [0,1)$. This leads to

$$|\varphi(s|x,\varepsilon,d) - f(x_1,\ldots,x_d)I| \leq C_2 \frac{\varepsilon^2 I}{2\lambda_d} f(x_1,\ldots,x_d). \tag{S1.3}$$

Come back to the SmBP (S1.1),

$$\varphi(x,\varepsilon) = \int_0^1 f(x_1,\ldots,x_d) I dG(s) + \int_0^1 \left(\varphi(s|x,\varepsilon,d) - f(x_1,\ldots,x_d)I\right) dG(s), \tag{S1.4}$$

and note that, thanks to (S1.3) and because d is fixed, the second addend in the right–hand side of (S1.4) is infinitesimal with respect to the first addend

$$\left| \frac{\int_0^1 \left(\varphi(s|x,\varepsilon,d) - f(x_1,\dots,x_d)I\right) dG\left(s\right)}{\int_0^1 f(x_1,\dots,x_d) I dG\left(s\right)} \right| \leq$$
$$\leq \left| \frac{C_2 \frac{\varepsilon^2}{2\lambda_d} f(x_1,\dots,x_d) \int_0^1 I dG\left(s\right)}{f(x_1,\dots,x_d) \int_0^1 I dG\left(s\right)} \right| = C_2 \frac{\varepsilon^2}{2\lambda_d}.$$

Noting that

$$\int_0^1 I(s,\varepsilon,d) dG(s) = \frac{\varepsilon^d \pi^{d/2}}{\Gamma\left(d/2+1\right)} \mathbb{E}\left[(1-S)^{d/2}\, \mathbb{I}_{\{S\leq 1\}}\right],$$

we obtain

$$|\varphi(x,\varepsilon) - \varphi_d(x,\varepsilon)| \leq C_2 \frac{\varepsilon^2}{2\lambda_d} \varphi_d(x,\varepsilon) \tag{3.6}$$

where,

$$\varphi_d(x,\varepsilon) = f(x_1,\dots,x_d) \frac{\varepsilon^d \pi^{d/2}}{\Gamma\left(d/2+1\right)} \mathbb{E}\left[(1-S)^{d/2}\, \mathbb{I}_{\{S\leq 1\}}\right]. \tag{3.5}$$

Thus, since d is fixed, as ε tends to zero,

$$\varphi(x,\varepsilon) = \int_0^1 \varphi(s|x,\varepsilon,d) dG\left(s\right) = \varphi_d(x,\varepsilon) + o\left(\frac{\varphi_d(x,\varepsilon)}{f(x_1,\dots,x_d)}\right)$$

or, equivalently, $\varphi(x,\varepsilon) \sim \varphi_d(x,\varepsilon)$ that concludes the proof.

Proof Proofs of Proposition 1, and theorems 2 and 3

To prove Proposition 1 we need the following Lemma.

Lemma 1. *Assume (A-1) and (A-2). Then, it is possible to choose $d = d(\varepsilon)$ so that it diverges to infinity as ε tends to zero and*

$$\sum_{j\geq d+1} \lambda_j = o(\varepsilon^2). \tag{S1.5}$$

Moreover, as $\varepsilon \to 0$, $S(x,\varepsilon,d) \to 0$, where the convergence holds almost surely, in the L^1 norm and hence in probability.

Proof. A possible choice for $d = d(\varepsilon)$ satisfying (S1.5) can be, for a fixed $\delta > 0$, as follows

$$d = \min\left\{k \in \mathbb{N} : \sum_{j\geq k+1} \lambda_j \leq \varepsilon^{2+\delta}\right\}, \qquad \text{for any } \varepsilon > 0.$$

Such a minimum is well defined since eigenvalues series is convergent.
Let us prove that S converges to zero in probability. For any $k > 0$, by Markov inequality and, thanks to Assumption (A-2),

$$\mathbb{P}\left(|S| > k\right) = \mathbb{P}\left(S > k\right) = \mathbb{P}\left(\frac{1}{\varepsilon^2}\sum_{j \geq d+1}\left(\theta_j - x_j\right)^2 > k\right)$$
$$\leq \frac{\mathbb{E}\left[\frac{1}{\varepsilon^2}\sum_{j\geq d+1}\left(\theta_j - x_j\right)^2\right]}{k^2} \leq \frac{C_1}{k^2}\frac{\sum_{j\geq d+1}\lambda_j}{\varepsilon^2}. \tag{S1.6}$$

Thanks to (S1.5) we get the convergence in probability. Since $S = S(x, \varepsilon, d)$ is non–increasing when d increases,

$$\mathbb{P}\left(\sup_{j\geq d+1}|S(x,\varepsilon,j) - 0| \geq k\right) = \mathbb{P}\left(S(x,\varepsilon,d+1) \geq k\right)$$

holds for any $k > 0$ and any x. This fact, together with (S1.6), guarantees the almost sure convergence of S to zero (e.g. Shiryayev (1984, Theorem 10.3.1)) as ε tends to zero. Moreover, the monotone convergence theorem guarantees the L^1 convergence. ■

Proof of Proposition 1. Note that if $d(\varepsilon)$ satisfies $d\sum_{j\geq d+1}\lambda_j = o(\varepsilon^2)$, then (S1.5) and Lemma 1 hold. For a fixed $\delta > 0$, a possible choice of such $d = d(\varepsilon)$ can be

$$d = \min\left\{k \in \mathbb{N} : \quad k\sum_{j\geq k+1}\lambda_j \leq \varepsilon^{2+\delta}\right\},$$

where the minimum is achieved thanks to the eigenvalues hyperbolic decay assumption.
At this stage, note that

$$0 < \mathbb{E}\left[\left(1-S\right)^{d/2}\mathbb{I}_{\{S<1\}}\right] \leq 1$$

then, after some algebra, thanks to Bernoulli inequality (i.e. $(1+s)^r \geq 1+rs$ for $s \geq -1$ and $r \in \mathbb{R}\setminus(0,1)$), Markov inequality and Assumption (A-2), we have (for any $d \geq 2$)

$$\begin{aligned}0 \leq & 1-\mathbb{E}\left[(1-S)^{d/2}\,\mathbb{I}_{\{S<1\}}\right] \leq 1-\mathbb{E}\left[\left(1-\frac{d}{2}S\right)\mathbb{I}_{\{S<1\}}\right] \\ \leq & \mathbb{P}(S\geq 1)+\mathbb{E}\left[\frac{d}{2}S\,\mathbb{I}_{\{S<1\}}\right] \leq \mathbb{E}\left[\frac{(d+2)}{2\varepsilon^2}\sum_{j\geq d+1}(\theta_j-x_j)^2\right] \\ \leq & \frac{C_1(d+2)}{2\varepsilon^2}\sum_{j\geq d+1}\lambda_j.\end{aligned}$$

Choosing d according to $d\sum_{j\geq d+1}\lambda_j = o(\varepsilon^2)$ the thesis follows. ■

Proof of Theorem 2. Thanks to hyper–exponentiality (4.12), there exists $d_0 \in \mathbb{N}$ so that for any $d \geq d_0$

$$d\sum_{j\geq d+1}\lambda_j < \lambda_d.$$

Moreover, there exist $\delta_1, \delta_2 \in (0,1)$ (depending on d) for which, for any $d \geq d_0$

$$0 \leq d\sum_{j\geq d+1}\lambda_j \leq b(d,\{\lambda_j\}_{j\geq d+1},\delta_1) < B(d,\{\lambda_j\}_{j\leq d},\delta_2) \leq \lambda_d, \qquad \text{(S1.7)}$$

where

$$b(d,\{\lambda_j\}_{j\geq d+1},\delta_1) = \left(d\sum_{j\geq d+1}\lambda_j\right)^{1-\delta_1}, \qquad B(d,\{\lambda_j\}_{j\leq d},\delta_2) = \lambda_d^{1-\delta_2}.$$

As instance, for a given $d \geq d_0$, fix $\delta_1 \in (0,1)$ and solve (S1.7) with respect to δ_2, that is $\delta_2 \in (\min\{0,\beta(\delta_1)\},1)$ where $\beta(\delta_1) = 1-(1-\delta_1)\ln\left(d\sum_{j\geq d+1}\lambda_j\right)/\ln(\lambda_d)$. As a consequence, for any $\varepsilon > 0$ and for such a choice of δ_1, δ_2, the following minimum is well–defined

$$d(\varepsilon) = \min\left\{k\in\mathbb{N} : b(k,\{\lambda_j\}_{j\geq k+1},\delta_1) \leq \varepsilon^2 \leq B(k,\{\lambda_j\}_{j\leq k},\delta_2)\right\}.$$

This guarantees that the right–hand side of (4.10) vanishes as ε goes to zero. ■

To prove Theorem 3 we need the following Lemma.

Lemma 2. *Assume (A-1) and (A-2). Then, as* $\varepsilon \to 0$,

$$\mathcal{R}(x,\varepsilon,d)^{2/d} \to 1, \qquad \text{or,} \qquad \log(\mathcal{R}(x,\varepsilon,d)) = o(d). \qquad \text{(S1.8)}$$

Proof. Jensen inequality for concave functions (i.e. $\mathbb{E}[f(g)] \le f(\mathbb{E}[g])$ if f is a concave function) guarantees that

$$\begin{aligned}\mathbb{E}\left[\left((1-S)\mathbb{I}_{\{S<1\}}\right)^{\frac{d}{2}}\right] &= \mathbb{E}\left[\left((1-S)\mathbb{I}_{\{S<1\}}\right)^{\frac{d+1}{2}\frac{d}{d+1}}\right] \\ &\le \left\{\mathbb{E}\left[\left((1-S)\mathbb{I}_{\{S<1\}}\right)^{\frac{d+1}{2}}\right]\right\}^{\frac{d}{d+1}},\end{aligned}$$

noting that $S(x,\varepsilon,d+1) =: S_{d+1} \le S_d := S(x,\varepsilon,d)$ and $\mathbb{I}_{\{S_d<1\}} \le \mathbb{I}_{\{S_{d+1}<1\}}$, then

$$\mathbb{E}\left[\left((1-S_d)\mathbb{I}_{\{S_d<1\}}\right)^{\frac{d}{2}}\right] \le \left\{\mathbb{E}\left[\left((1-S_{d+1})1_{\{S_{d+1}<1\}}\right)^{\frac{d+1}{2}}\right]\right\}^{\frac{d}{d+1}}.$$

The latter guarantees that $\mathbb{E}\left[(1-S)^{d/2}\,\mathbb{I}_{\{S<1\}}\right]^{2/d}$ is a non–decreasing monotone sequence with respect to d whose values are in $(0,1]$ and eventually bounded away from zero. ■

Proof of Theorem 3. Given results in Theorem 1, thesis holds using the same arguments as in Delaigle and Hall (2010, Proof of Theorem 4.2.): the idea is to combine together (S1.8), the Stirling expansion of the Gamma function in V_d and the (super–)exponential eigenvalues decay. ■

Proof of Theorem 4

In what follows, as in Section 5, we simplify the notations dropping the dependence on d for the density estimators f_n and $\widehat{f}_n$. Moreover, C denotes a general positive constant. The proof of Theorem 4 uses similar arguments as in Biau and Mas (2012).

Since $H_n = h_n^2 I$, it holds $K_{H_n}(u) = h_n^{-d} K(u)$. Consider

$$S_n(x) = \sum_{i=1}^{n} K\left(\frac{\|\Pi_d(X_i - x)\|}{h_n}\right), \qquad \widehat{S}_n(x) = \sum_{i=1}^{n} K\left(\frac{\left\|\widehat{\Pi}_d(X_i - x)\right\|}{h_n}\right),$$

then the pseudo-estimator and the estimator are given by

$$f_n(x) = \frac{S_n(x)}{nh_n^d}, \qquad \widehat{f}_n(x) = \frac{\widehat{S}_n(x)}{nh_n^d},$$

and, hence,

$$\mathbb{E}\left[f_n(x) - \widehat{f}_n(x)\right]^2 = \frac{1}{(nh_n^d)^2}\mathbb{E}\left[S_n(x) - \widehat{S}_n(x)\right]^2.$$

Set $V_i = \left\| \Pi_d \left(X_i - x \right) \right\|$, $\widehat{V}_i = \left\| \widehat{\Pi}_d \left(X_i - x \right) \right\|$, consider the events

$$A_i = \left\{ V_i \le h_n \right\}, \qquad B_i = \left\{ \widehat{V}_i \le h_n \right\},$$

then we have the decomposition

$$\begin{aligned} S_n(x) - \widehat{S}_n(x) = & \sum_{i=1}^{n} \left[K\left(\frac{V_i}{h_n} \right) - K\left(\frac{\widehat{V}_i}{h_n} \right) \right] \mathbb{I}_{A_i \cap B_i} + \\ & + \sum_{i=1}^{n} K\left(\frac{V_i}{h_n} \right) \mathbb{I}_{A_i \cap \overline{B}_i} - \sum_{i=1}^{n} K\left(\frac{\widehat{V}_i}{h_n} \right) \mathbb{I}_{\overline{A}_i \cap B_i}. \end{aligned}$$

Since $(a+b)^2 \le 2a^2 + 2b^2$,

$$\begin{aligned} \mathbb{E}\left[S_n(x) - \widehat{S}_n(x) \right]^2 \le & 2\mathbb{E}\left[\sum_{i=1}^{n} \left(K\left(\frac{V_i}{h_n} \right) - K\left(\frac{\widehat{V}_i}{h_n} \right) \right) \mathbb{I}_{A_i \cap B_i} \right]^2 + \\ & + 2\mathbb{E}\left[\left(\sum_{i=1}^{n} K\left(\frac{V_i}{h_n} \right) \mathbb{I}_{A_i \cap \overline{B}_i} \right)^2 + \right. \\ & \left. + \left(\sum_{i=1}^{n} K\left(\frac{\widehat{V}_i}{h_n} \right) \mathbb{I}_{\overline{A}_i \cap B_i} \right)^2 \right]. \end{aligned} \tag{S1.9}$$

Consider now the first addend in the right–hand side of (S1.9): Assumption (B-3) and the fact that $\left| V_i - \widehat{V}_i \right| \le \left\| \Pi_d - \widehat{\Pi}_d \right\|_\infty \left\| X_i - x \right\|$, where $\left\| \cdot \right\|_\infty$ denotes the operator norm, lead to

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{n} \left(K\left(\frac{V_i}{h_n} \right) - K\left(\frac{\widehat{V}_i}{h_n} \right) \right) \mathbb{I}_{A_i \cap B_i} \right]^2 \le & \\ \le C\mathbb{E}\left[\left\| \Pi_d - \widehat{\Pi}_d \right\|_\infty \sum_{i=1}^{n} \left\| X_i - x \right\| \mathbb{I}_{A_i \cap B_i} \right]^2. & \end{aligned}$$

Thanks to the Cauchy-Schwartz inequality we control the previous bound by

$$C\mathbb{E}\left[\left\| \Pi_d - \widehat{\Pi}_d \right\|_\infty^2 \right] \mathbb{E}\left[\left(\sum_{i=1}^{n} \left\| X_i - x \right\| \mathbb{I}_{A_i \cap B_i} \right)^2 \right]. \tag{S1.10}$$

About the first factor in (S1.10), Biau and Mas (2012, Theorem 2.1 (ii)) established that

$$\mathbb{E}\left[\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty^2\right] = O\left(\frac{1}{n}\right). \tag{S1.11}$$

Consider now the second term in (S1.10). Thanks to the Chebyshev's algebraic inequality (see, for instance, Mitrinović et al. (1993, page 243)) and since $\mathbb{E}\left[\mathbb{I}_{A_i \cap B_i}\right] \leq \mathbb{E}\left[\mathbb{I}_{A_i}\right]$, for any $k \geq 1$ it holds

$$\mathbb{E}\left[\|X - x\|^k \mathbb{I}_{A_i \cap B_i}\right] \leq \mathbb{E}\left[\|X - x\|^k\right] \mathbb{E}\left[\mathbb{I}_{A_i}\right].$$

The fact that $\mathbb{E}\left[\mathbb{I}_{A_i}\right] \sim h_n^d$ and Assumption (B-4) give

$$\mathbb{E}\left[\|X - x\|^k \mathbb{I}_{A_i \cap B_i}\right] \leq C\frac{k!}{2} b^{k-2} h_n^d,$$

with $b > 0$. Hence, the Bernstein inequality (see e.g. Massart (2007)) can be applied: for any $M > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n \|X_i - x\| \mathbb{I}_{A_i \cap B_i} - \mathbb{E}\left[\sum_{i=1}^n \|X_i - x\| \mathbb{I}_{A_i \cap B_i}\right]\right| \geq Mnh^d\right) \leq \\ \leq \exp\left(-CM^2 nh^d\right).$$

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

$$\sum_{i=1}^n \|X_i - x\| \mathbb{I}_{A_i \cap B_i} \leq Cnh^d \qquad a.s.$$

and therefore,

$$\mathbb{E}\left[\left(\sum_{i=1}^n \|X_i - x\| \mathbb{I}_{A_i \cap B_i}\right)^2\right] \leq Cn^2 h^{2d}. \tag{S1.12}$$

Finally, combining results (S1.11) and (S1.12), we obtain:

$$\frac{1}{(nh_n^d)^2}\mathbb{E}\left[\sum_{i=1}^n \left(K\left(\frac{V_i}{h_n}\right) - K\left(\frac{\widehat{V}_i}{h_n}\right)\right)\mathbb{I}_{A_i \cap B_i}\right]^2 \leq C\frac{1}{nh_n^2}. \tag{S1.13}$$

Consider now the second addend in the right–hand side of (S1.9). We only look at

$$\mathbb{E}\left[\sum_{i=1}^n K\left(\frac{V_i}{h_n}\right)\mathbb{I}_{A_i \cap \overline{B}_i}\right]^2, \tag{S1.14}$$

because the behaviour of the other addend is similar. Define the sequence κ_n so that $\kappa_n \to 0$ as $n \to \infty$, the following inclusions hold:

$$
\begin{aligned}
A_i \cap \overline{B}_i = & \{V_i \leq h_n\} \cap \left\{\widehat{V}_i > h_n\right\} \\
= & (\{h_n (1-\kappa_n) < V_i \leq h_n\} \cup \{V_i \leq h_n (1-\kappa_n)\}) \cap \\
& \cap \left\{\widehat{V}_i - V_i > h_n - V_i\right\} \\
\subseteq & \{h_n (1-\kappa_n) < V_i \leq h_n\} \cup \left\{V_i \leq h_n (1-\kappa_n), \widehat{V}_i - V_i > h_n - V_i\right\} \\
\subseteq & \{h_n (1-\kappa_n) < V_i \leq h_n\} \cup \left\{\widehat{V}_i - V_i > \kappa_n h_n\right\}.
\end{aligned}
$$

The latter inclusion and Assumption (B-3) allow to control (S1.14) by

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{I}_{A_i \cap \overline{B}_i}\right]^2 \leq & 2\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{I}_{\{h_n(1-\kappa_n)<V_i\leq h_n\}}\right]^2 + \\
& + 2\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{I}_{\left\{\left\|\widehat{\Pi}_d - \Pi_d\right\| \|X_i - x\| > C\kappa_n h_n\right\}}\right]^2 . \qquad \text{(S1.15)}
\end{aligned}
$$

About the first term in the right–hand side of the latter, the Cauchy-Schwartz inequality gives

$$
\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{I}_{\{h_n(1-\kappa_n)<V_i\leq h_n\}}\right]^2 \leq n^2 \mathbb{P}\left(h_n (1-\kappa_n) < V \leq h_n\right).
$$

Since $\mathbb{P}\left(h_n (1-\kappa_n) < V \leq h_n\right) \sim h_n^d \left(1 - (1-\kappa_n)^d\right)$, performing a first order Taylor expansion of $(1-\kappa_n)^d$ in $\kappa_n = 0$, we get asymptotically

$$
\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{I}_{\{h_n(1-\kappa_n)<V_i\leq h_n\}}\right]^2 \leq C n^2 h_n^d \kappa_n.
$$

Similarly, for what concerns the other addend in the right–hand side of (S1.15), we have

$$
\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{I}_{\left\{\left\|\widehat{\Pi}_d - \Pi_d\right\| \|X_i - x\| > C\kappa_n h_n\right\}}\right]^2 \leq n^2 \mathbb{P}\left(\left\|\widehat{\Pi}_d - \Pi_d\right\| \|X - x\| > C\kappa_n h_n\right).
$$

Thanks to the Markov inequality, Biau and Mas (2012, Theorem 2.1 (iii)) and Assumption (B-4), it follows

$$
\mathbb{P}\left(\left\|\widehat{\Pi}_d - \Pi_d\right\| \|X - x\| > C\kappa_n h_n\right) = O\left(\frac{1}{n^{1/2} h_n \kappa_n}\right).
$$

Combining the previous results we obtain:

$$\frac{1}{\left(nh_n^d\right)^2}\mathbb{E}\left[\left(\sum_{i=1}^{n} K\left(\frac{V_i}{h_n}\right)\mathbb{I}_{A_i\cap\overline{B}_i}\right)\right]^2 = O\left(\frac{\kappa_n}{h_n^d}\right) + O\left(\frac{1}{n^{1/2}h_n\kappa_n}\right).$$

If we choose $\kappa_n = \left(n^{5/2}h_n^{2d}\right)^{-1/2}$ with $n^{5/4}h_n^d \to \infty$, as $n \to \infty$, we obtain:

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} K\left(\frac{V_i}{h_n}\right)\mathbb{I}_{A_i\cap\overline{B}_i}\right)^2 + \left(\sum_{i=1}^{n} K\left(\frac{\widehat{V}_i}{h_n}\right)\mathbb{I}_{\overline{A}_i\cap B_i}\right)^2\right] \leq C\frac{1}{n^{5/4}h_n^{2d}}. \tag{S1.16}$$

In conclusion, (S1.13) and (S1.16) lead to:

$$\frac{1}{\left(nh_n^d\right)^2}\mathbb{E}\left[S_n(x) - \widehat{S}_n(x)\right]^2 = O\left(\frac{1}{nh_n^2}\right) + O\left(\frac{1}{n^{5/4}h_n^{2d}}\right).$$

Choose the optimal bandwidth (5.20) and $p > \max\{2, 3d/10\}$, then, as n goes to infinity, the first addend becomes negligible compared to the second one that turns to be $O\left(n^{-(10p-3d)/4(2p+d)}\right)$. Moreover, a direct computation shows that such bound is definitively negligible when compared to the "optimal bound" $n^{-2p/(2p+d)}$, for any $p > \max\{2, 3d/2\}$ and $d \geq 1$. This concludes the proof.

References

Biau, G., Mas, A., 2012. PCA-kernel estimation. Stat. Risk Model. 29 (1), 19–46.

Delaigle, A., Hall, P., 2010. Defining probability density for a distribution of random functions. Ann. Statist. 38 (2), 1171–1193.

Massart, P., 2007. Concentration inequalities and model selection. Vol. 1896 of Lecture Notes in Mathematics. Springer, Berlin, lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

Mitrinović, D. S., Pečarić, J. E., Fink, A. M., 1993. Classical and new inequalities in analysis. Vol. 61 of Mathematics and its Applications (East European Series). Kluwer Academic Publishers Group, Dordrecht.

Shiryayev, A. N., 1984. Probability. Vol. 95 of Graduate Texts in Mathematics. Springer-Verlag, New York.