$\mathbb{L}_2$-Boosting for sensitivity analysis with dependent inputs

Magali Champion[1,3], Gaelle Chastaing[1,2], Sébastien Gadat[4], Clémentine Prieur[2]

[1] *Institut de Mathématiques de Toulouse, CNRS and Université Paul Sabatier, Toulouse France*
[2] *Université Grenoble Alpes CNRS, and INRIA, Grenoble, France*
[3] *INRA, UR875 MIA-T, F-31326 Castanet-Tolosan, France*
[4] *Toulouse School of Economics - CNRS and Université Toulouse I Capitole, France*

Supplementary Material

We present here the proofs of Theorem 1 and 2 of the main document. Section S1 establishes the notation that will be used throughout the document. Section S2 provides a concentration inequality on random matrices that will be used in the rest of the document. We develop the proofs of Theorems 1 and 2 in Section S3 and S4, respectively.

S1 Notation and reminder

Let us first recall some standard notations on matricial norms. For any square matrix M, its spectral radius $\rho(M)$ will refer to the largest absolute value of the elements of its spectrum:

$$\rho(M) := \max_{\alpha \in Sp(M)} |\alpha|.$$

Moreover, $|||M|||_2$ is the Euclidean endomorphism norm and is given by:

$$|||M|||_2 := \sqrt{\rho({}^tMM)},$$

where ${}^tM$ is the transpose of M. Note that for self-adjoint matrices, $|||M|||_2 = \rho(M)$. At last, the Frobenius norm of M is given by:

$$\|M\|_F := \left(Tr({}^tMM)\right)^{1/2},$$

where $Tr(M)$ is the trace of the matrix M.

S2 Hoeffding-type Inequality for random bounded matrices

For the sake of completeness, we refer to Theorem 1.3 of Tropp (2012). $\preceq$ denotes the semi-definite order on self-adjoint matrices, which is defined for all self-adjoint matrices

M_1 and M_2 of size q as:

$$M_1 \preceq M_2 \text{ iff } \forall u \in \mathbb{R}^q, \ {}^t u M_1 u \leq {}^t u M_2 u.$$

Theorem 1 (Hoeffding's matrix concentration inequality: bounded case)**.** *Consider a finite sequence $(X_k)_{1\leq k\leq n}$ of independent random self-adjoint matrices with dimension d, and let $(A_k)_{1\leq k\leq n}$ be a deterministic sequence of self-adjoint matrices. Assume that:*

$$\forall 1 \leq k \leq n \qquad \mathbb{E}X_k = 0 \qquad \textit{and} \qquad X_k^2 \preceq A_k^2 \quad a.s.$$

Then, for all $t \geq 0$:

$$P\left(\lambda_{max}\left(\sum_{k=1}^n X_k\right) \geq t\right) \leq d e^{-t^2/8\sigma^2}, \qquad \textit{where} \qquad \sigma^2 = \rho\left(\sum_{k=1}^n A_k^2\right).$$

In our work, a more precise concentration inequality such as the Bernstein one (see Theorem 6.1 of Tropp (2012)) is useless since we do not consider any asymptotic on L (the number of basis functions for each variables X_j). Such an asymptotic setting is far beyond the scope of this paper and we leave this problem open for a future study.

S3 Proof of Theorem 1

Consider any subset $u = (u_1, ..., u_t) \in S_n^*$ with $t \geq 1$ and note that if $u = \{i\}$, i.e. $t = 1$, the *Initialization* of Algorithm 1 is such that:

$$\hat{\phi}^i_{l_i,n_1} = \phi^i_{l_i}, \quad \forall\, l_i \in [1:L],$$

Therefore, we obviously have that $\sup_{\substack{i\in[1:p] \\ l_i\in[1:L]}} \left\| \hat{\phi}^i_{l_i,n_1} - \phi^i_{l_i} \right\| = 0$.

Now, for $t = 2$, let $u = \{i, j\}$, with $i \neq j \in [1 : p]$, and $\boldsymbol{l_{ij}} = (l_i, l_j) \in [1 : L]^2$, and recall that $\phi^{ij}_{\boldsymbol{l_{ij}}}$ is defined as:

$$\phi^{ij}_{\boldsymbol{l_{ij}}}(x_i, x_j) = \phi^i_{l_i}(x_i) \times \phi^j_{l_j}(x_j) + \sum_{k=1}^L \lambda^i_{k,\boldsymbol{l_{ij}}} \phi^i_k(x_i) + \sum_{k=1}^L \lambda^j_{k,\boldsymbol{l_{ij}}} \phi^j_k(x_j) + C_{\boldsymbol{l_{ij}}},$$

where $(C_{\boldsymbol{l_{ij}}}, (\lambda^i_{k,\boldsymbol{l_{ij}}})_k, (\lambda^j_{k,\boldsymbol{l_{ij}}})_k)$ are given as the solutions of:

$$\begin{aligned}
&\langle \phi^{ij}_{\boldsymbol{l_{ij}}}, \phi^i_k \rangle = 0, \quad \forall\, k \in [1:L] \\
&\langle \phi^{ij}_{\boldsymbol{l_{ij}}}, \phi^j_k \rangle = 0, \quad \forall\, k \in [1:L] \\
&\langle \phi^{ij}_{\boldsymbol{l_{ij}}}, 1 \rangle = 0.
\end{aligned} \tag{S3.1}$$

When removing $C_{\boldsymbol{l_{ij}}}$, the resolution of (S3.1) leads to the resolution of a linear system of the type:

$$A^{ij} \boldsymbol{\lambda}^{\boldsymbol{l_{ij}}} = D^{\boldsymbol{l_{ij}}}, \tag{S3.2}$$

with $\boldsymbol{\lambda}^{\boldsymbol{l_{ij}}} = {}^t\left(\lambda^i_{1,\boldsymbol{l_{ij}}} \cdots \lambda^i_{L,\boldsymbol{l_{ij}}} \lambda^j_{1,\boldsymbol{l_{ij}}} \cdots \lambda^j_{L,\boldsymbol{l_{ij}}}\right)$ and

$$A^{ij} = \begin{pmatrix} B^{ii} & B^{ij} \\ {}^tB^{ij} & B^{jj} \end{pmatrix}, \quad B^{ij} = \begin{pmatrix} \langle \phi^i_1, \phi^j_1 \rangle & \cdots & \langle \phi^i_1, \phi^j_L \rangle \\ \vdots & & \\ \langle \phi^i_L, \phi^j_1 \rangle & \cdots & \langle \phi^i_L, \phi^j_L \rangle \end{pmatrix}, \quad D^{\boldsymbol{l_{ij}}} = -\begin{pmatrix} \langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^i_1 \rangle \\ \vdots \\ \langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^i_L \rangle \\ \langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^j_1 \rangle \\ \vdots \\ \langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^j_L \rangle \end{pmatrix}.$$

Consider now $\hat{\phi}^{ij}_{\boldsymbol{l_{ij}},n_1}$ that is decomposed on the dictionary as follows:

$$\hat{\phi}^{ij}_{\boldsymbol{l_{ij}},n_1}(x_i, x_j) \quad = \quad \phi^i_{l_i}(x_i) \times \phi^j_{l_j}(x_j) + \sum_{k=1}^L \hat{\lambda}^i_{k,\boldsymbol{l_{ij}},n_1} \phi^i_k(x_i) + \sum_{k=1}^L \hat{\lambda}^j_{k,\boldsymbol{l_{ij}},n_1} \phi^j_k(x_j) + \hat{C}^{n_1}_{\boldsymbol{l_{ij}}},$$

where $(\hat{C}^{n_1}_{\boldsymbol{l_{ij}}}, (\hat{\lambda}^i_{k,\boldsymbol{l_{ij}},n_1})_k, (\hat{\lambda}^j_{k,\boldsymbol{l_{ij}},n_1})_k)$ are given as solutions of the following *random* equalities:

$$\begin{array}{l} \langle \hat{\phi}^{ij}_{\boldsymbol{l_{ij}},n_1}, \phi^i_k \rangle_{n_1} = 0, \quad \forall\, k \in [1:L] \\ \langle \hat{\phi}^{ij}_{\boldsymbol{l_{ij}},n_1}, \phi^j_k \rangle_{n_1} = 0, \quad \forall\, k \in [1:L] \\ \langle \hat{\phi}^{ij}_{\boldsymbol{l_{ij}},n_1}, 1 \rangle_{n_1} = 0. \end{array} \tag{S3.3}$$

When removing $\hat{C}^{n_1}_{\boldsymbol{l_{ij}}}$, the resolution of (S3.3) can also lead to the resolution of a linear system of the type:

$$\hat{A}^{ij}_{n_1} \hat{\boldsymbol{\lambda}}^{\boldsymbol{l_{ij}}}_{n_1} = \hat{D}^{\boldsymbol{l_{ij}}}_{n_1}, \tag{S3.4}$$

where $\hat{\boldsymbol{\lambda}}^{\boldsymbol{l_{ij}}}_{n_1} = {}^t\left(\hat{\lambda}^i_{1,\boldsymbol{l_{ij}},n_1} \cdots \hat{\lambda}^i_{L,\boldsymbol{l_{ij}},n_1} \hat{\lambda}^j_{1,\boldsymbol{l_{ij}},n_1} \cdots \hat{\lambda}^j_{L,\boldsymbol{l_{ij}},n_1}\right)$ and $\hat{A}^{ij}_{n_1}$ (*resp.* $\hat{D}^{\boldsymbol{l_{ij}}}_{n_1}$) are obtained from A^{ij} (*resp.* $D^{\boldsymbol{l_{ij}}}$) by substituting the theoretical inner product with its empirical version.

Remark 1. *Remark that each A^{ij} depends on (i,j) as well as $\boldsymbol{\lambda}^{\boldsymbol{l_{ij}}}$ and $D^{\boldsymbol{l_{ij}}}$ depend on (i,j) and $\boldsymbol{l_{ij}}$, but we will deliberately omit these indexes in the sequel for the sake of convenience (when no confusion is possible). For example, when a couple (i,j) is considered, we will frequently use the notation $A, \boldsymbol{\lambda}, D, C, \lambda^i_k, \lambda^j_k$ instead of $A^{ij}, \boldsymbol{\lambda}^{\boldsymbol{l_{ij}}}, D^{\boldsymbol{l_{ij}}}, C_{\boldsymbol{l_{ij}}}, \lambda^i_{k,\boldsymbol{l_{ij}}}$ and $\lambda^j_{k,\boldsymbol{l_{ij}}}$. This will be also the case for the estimators $\hat{A}_{n_1}, \hat{\boldsymbol{\lambda}}_{n_1}, \hat{D}_{n_1}, \hat{C}^{n_1}, \hat{\lambda}^i_{k,n_1}$ and $\hat{\lambda}^j_{k,n_1}$.*

The following useful lemma then compares the two matrices $\hat{A}_{n_1}$ and A.

Lemma 1. *Under Assumption $(\mathbf{H_b})$, and for any ξ given by $(\mathbf{H^2_b})$, we have:*

$$\sup_{1 \le i,j \le p_n} \left\| \hat{A}_{n_1} - A \right\|_2 = \mathcal{O}_P(n^{-\xi/2}).$$

Proof. First, consider a couple (i, j) and note that $\left\|\left\|\hat{A}_{n_1} - A\right\|\right\|_2 = \rho(\hat{A}_{n_1} - A)$, since $\hat{A}_{n_1} - A$ is self-adjoint. To obtain a concentration inequality on the matricial norm $\left\|\left\|\hat{A}_{n_1} - A\right\|\right\|_2$, we use the result of Tropp (2012) (see Theorem 1), which gives concentration inequalities for the largest eigenvalue of self-adjoint matrices (see Section 6.2).

Remark that $\hat{A}_{n_1} - A$ can be written as follows:

$$\hat{A}_{n_1} - A = \frac{1}{n_1}\sum_{r=1}^{n_1} \Theta_{r,ij}, \quad \Theta_{r,ij} = \begin{pmatrix} \Theta_r^{ii} & \Theta_r^{ij} \\ {}^t\Theta_r^{ij} & \Theta_r^{jj} \end{pmatrix}, \quad \forall\ r \in [1:n_1],$$

where, for all $k, m \in [1:L]$, $(\Theta_r^{i_1 i_2})_{k,m} = \phi_k^{i_1}(x_{i_1}^r)\phi_m^{i_2}(x_{i_2}^r) - \mathbb{E}[\phi_k^{i_1}(X_{i_1})\phi_m^{i_2}(X_{i_2})]$ with $i_1, i_2 \in \{i, j\}$. Since the observations $(\mathbf{x}^r)_{r=1,\cdots,n_1}$ are independent, $\Theta_{1,ij}, \cdots, \Theta_{n_1,ij}$ is a sequence of independent, random, centered, and self-adjoint matrices. Moreover, for all $u \in \mathbb{R}^{2L}$, all $r \in [1:n_1]$:

$${}^t u\Theta_{r,ij}^2 u = \|\Theta_{r,ij}u\|_2^2 \leq \|u\|_2^2 \|\Theta_{r,ij}\|_F^2,$$

where

$$\begin{aligned} \|\Theta_{r,ij}\|_F^2 &\leq (2L)^2 \left(\max_{k,m\in[1:L]} |(\Theta_{r,ij})_{k,m}|\right)^2 \\ &\leq (2L)^2 \left(\max_{\substack{k,m\in[1:L] \\ i_1,i_2\in\{i,j\}}} |\phi_k^{i_1}(x_{i_1}^r)\phi_m^{i_2}(x_{i_2}^r) - \mathbb{E}[\phi_k^{i_1}(X_{i_1})\phi_m^{i_2}(X_{i_2})]|\right)^2 \\ &\leq 16L^2M^4 \quad \text{by } (\mathbf{H_b^1}). \end{aligned}$$

We then deduce that each element of the sum satisfies $X_{l,ij}^2 \preceq 16L^2M^4\mathrm{I}_{L^2}$, where I_{L^2} designates the identity matrix of size L^2.

Applying the Hoeffding-type Inequality stated as Theorem 1.3 of Tropp (2012) to our sequence $\Theta_{1,ij}, \cdots, \Theta_{n_1,ij}$, with $\sigma^2 = 16n_1L^2M^4$, we obtain that:

$$\forall t \geq 0 \qquad P\left(\rho\left(\frac{1}{n_1}\sum_{r=1}^{n_1}\Theta_{r,ij}\right) \geq t\right) \leq 2Le^{-\frac{(n_1 t)^2}{8\sigma^2}},$$

Now considering the whole set of estimators $\hat{A}_{n_1}$, we obtain:

$$\forall t \geq 0 \qquad P\left(\sup_{1\leq i,j\leq p_n} \rho\left(\frac{1}{n_1}\sum_{r=1}^{n_1}\Theta_{r,ij}\right) \geq t\right) \leq 2Lp_n^2 e^{-\frac{(n_1 t)^2}{8\sigma^2}},$$

We take $t = \gamma n^{-\xi/2}$, where $\gamma > 0$, and $0 < \xi \leq 1$ is given in $(\mathbf{H_b^2})$. Then, the following inequality holds:

$$P\left(\sup_{1\leq i,j\leq p_n} \rho\left(\hat{A}_{n_1} - A\right) \geq \gamma n^{-\xi/2}\right) \leq 2Lp_n^2 e^{-\frac{{n_1}^{1-\xi}\gamma^2}{128L^2M^4}}. \tag{S3.5}$$

Since $n_1 = n/2$, and $p_n = \underset{n\to+\infty}{\mathcal{O}}(\exp(Cn^{1-\xi}))$ by Assumption $(\mathbf{H_b^2})$, the right-hand side of the previous inequality becomes arbitrarily small for n sufficiently large and $\gamma > 0$ large enough. The end of the proof follows using Inequality (S3.5). □

Similarly, we can show that the estimated quantity $\hat{D}_{n_1}$ is not far from the theoretical D, with high probability.

Lemma 2. *Under Assumptions* $(\mathbf{H_b})$, *and for any* ξ *given by* $(\mathbf{H_b^2})$, *one has*

$$\sup_{i,j,\boldsymbol{l_{ij}}} \left\| \hat{D}_{n_1} - D \right\|_2 = \mathcal{O}_P(n^{-\xi/2}).$$

Proof. First consider one couple (i,j). We aim to apply another concentration inequality on $\left\| \hat{D}_{n_1} - D \right\|_2$. Remark that $\left\| \hat{D}_{n_1} - D \right\|_2$ can be written as:

$$\begin{aligned}
\left\| \hat{D}_{n_1} - D \right\|_2 &= \Big(\textstyle\sum_{k=1}^{L} \left(\langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^i_k \rangle_{n_1} - \langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^i_k \rangle \right)^2 + \\
&\quad \textstyle\sum_{k=1}^{L} \left(\langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^j_k \rangle_{n_1} - \langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^j_k \rangle \right)^2 \Big)^{1/2} \\
&\leq \textstyle\sum_{k=1}^{L} \left| \frac{1}{n_1} \sum_{r=1}^{n_1} \phi^i_{l_i}({x_i}^r) \phi^j_{l_j}({x_j}^r) \phi^i_k({x_i}^r) - \langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^i_k \rangle \right| + \\
&\quad \textstyle\sum_{k=1}^{L} \left| \frac{1}{n_1} \sum_{r=1}^{n_1} \phi^i_{l_i}({x_i}^r) \phi^j_{l_j}({x_j}^r) \phi^j_k({x_j}^r) - \langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^j_k \rangle \right|.
\end{aligned}$$

Now, Bernstein's Inequality (see Birgé and Massart (1998) for instance) implies that, for all $\gamma > 0$,

$$\begin{aligned}
P\left(n_1^{\xi/2} \left\| \hat{D}_{n_1} - D \right\|_2 \geq \gamma \right) &\leq P\left(n_1^{\xi/2} \textstyle\sum_{k=1}^{L} \left| \frac{1}{n_1} \sum_{r=1}^{n_1} \phi^i_{l_i}(x_i^r) \phi^j_{l_j}(x_j^r) \phi^i_k(x_i^r) - \langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^i_k \rangle \right| > \gamma/2 \right) \\
&+ P\left(n_1^{\xi/2} \textstyle\sum_{k=1}^{L} \left| \frac{1}{n_1} \sum_{r=1}^{n_1} \phi^i_{l_i}(x_i^r) \phi^j_{l_j}(x_j^r) \phi^i_k(x_i^r) - \langle \phi^i_{l_i} \times \phi^j_{l_j}, \phi^i_k \rangle \right| > \gamma/2 \right) \\
&\leq 4L \exp\left(-\frac{1}{8} \frac{\gamma^2 n_1^{1-\xi}}{M^6 + M^3 \gamma / 6 n_1^{-\xi/2}} \right),
\end{aligned}$$

which gives:

$$P\left(\sup_{i,j,\boldsymbol{l_{ij}}} \left\| \hat{D}_{n_1} - D \right\|_2 \geq \gamma n_1^{-\xi/2} \right) \leq 4L \times L^2 p_n^2 \exp\left(-\frac{1}{8} \frac{\gamma^2 {n_1}^{1-\xi}}{M^6 + M^3 \gamma / 6 {n_1}^{-\xi/2}} \right). \quad \text{(S3.6)}$$

Now, since $n_1 = n/2$, Assumption $(\mathbf{H_b^2})$ implies that the right-hand side of Inequality (S3.6) can also become arbitrarily small for n sufficiently large, which concludes the proof. □

The next lemma then compares the estimated $\hat{\boldsymbol{\lambda}}_{n_1}$ with $\boldsymbol{\lambda}$.

Lemma 3. *Under Assumptions* $(\mathbf{H_b})$ *with* $\vartheta < \xi/2$, *we have:*

$$\sup_{i,j,\boldsymbol{l_{ij}}} \left\| \hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda} \right\|_2 = \mathcal{O}_P(n^{\vartheta - \xi/2}).$$

Proof. Fix any couple (i,j), $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\lambda}}_{n_1}$ satisfy Equations (S3.2) and (S3.4). Hence,

$$\begin{aligned}
A(\hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda}) - A\hat{\boldsymbol{\lambda}}_{n_1} &= -D = \hat{D}_{n_1} - D - \hat{D}_{n_1} \\
&= (\hat{D}_{n_1} - D) - \hat{A}_{n_1} \hat{\boldsymbol{\lambda}}_{n_1} \\
\Leftrightarrow \quad A(\hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda}) &= (\hat{D}_{n_1} - D) + (A - \hat{A}_{n_1}) \hat{\boldsymbol{\lambda}}_{n_1} \\
\Leftrightarrow \quad \hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda} &= A^{-1}[(A - \hat{A}_{n_1}) \hat{\boldsymbol{\lambda}}_{n_1}] + A^{-1}(\hat{D}_{n_1} - D),
\end{aligned}$$

since the matrix A is positive definite. It follows that:

$$\hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda} = A^{-1}(A - \hat{A}_{n_1})(\hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda}) + A^{-1}(A - \hat{A}_{n_1})\boldsymbol{\lambda} + A^{-1}(\hat{D}_{n_1} - D),$$

and

$$\left(\mathrm{I} - A^{-1}(A - \hat{A}_{n_1})\right)(\hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda}) = A^{-1}(A - \hat{A}_{n_1})\boldsymbol{\lambda} + A^{-1}(\hat{D}_{n_1} - D), \tag{S3.7}$$

Remark that $\left|\left|\left|\hat{A}_{n_1} - A\right|\right|\right|_2 = \mathcal{O}_P(n^{-\xi/2})$ by Lemma 1. Hence, with a high probability and for n large enough $\mathrm{I} - A^{-1}(A - \hat{A}_{n_1})$ is invertible, and Inequality (S3.7) can be rewritten as:

$$\hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda} = \left(\mathrm{I} - A^{-1}(A - \hat{A}_{n_1})\right)^{-1}\left(A^{-1}(A - \hat{A}_{n_1})\boldsymbol{\lambda} + A^{-1}(\hat{D}_{n_1} - D)\right).$$

We then deduce that:

$$\begin{aligned}
\left\|\hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda}\right\|_2 &\leq \left|\left|\left|\left(\mathrm{I} - A^{-1}(A - \hat{A}_{n_1})\right)^{-1}\right|\right|\right|_2 \\
&\quad \times \left(\left|\left|\left|A^{-1}[A - \hat{A}_{n_1}]\right|\right|\right|_2 \|\boldsymbol{\lambda}\|_2 + \left\|A^{-1}(\hat{D}_{n_1} - D)\right\|_2\right) \\
&\leq \left|\left|\left|\left(\mathrm{I} - A^{-1}(A - \hat{A}_{n_1})\right)^{-1}\right|\right|\right|_2 \\
&\quad \times \left(|||A^{-1}|||_2 \left|\left|\left|A - \hat{A}_{n_1}\right|\right|\right|_2 \|\boldsymbol{\lambda}\|_2 + |||A^{-1}|||_2 \left\|\hat{D}_{n_1} - D\right\|_2\right).
\end{aligned} \tag{S3.8}$$

A uniform bound for $|||A^{-1}|||_2$ (over all the couples (i,j)) can be easily obtained since A (and obviously A^{-1}) is Hermitian.

$$|||A^{-1}|||_2 \leq \max_{(i',j')\in[1:p_n]^2} \rho\left(\left(A^{i'j'}\right)^{-1}\right)$$

Simple algebra then yields:

$$\rho\left(\left(A^{i'j'}\right)^{-1}\right) \leq Tr\left(\left(A^{i'j'}\right)^{-1}\right) = \frac{Tr\left(Com(A^{i'j'})^t\right)}{det(A^{i'j'})} = \frac{1}{det(A^{i'j'})}\sum_{k=1:2L} Com(A^{i'j'})_{k,k}$$

where $Com(A^{ij})$ is the cofactor matrix associated to A^{ij}. Now, recall the classical inequality (that can be found in Bullen (1998)): for any symetric definite positive matrix squared S of size $Q \times Q$:

$$\det(S) \leq \prod_{\ell=1}^{Q} |S_{\ell\ell}|.$$

This last inequality applied to the determinant involved in $Com(A^{i'j'})_{k,k}$ associated with $(\mathbf{H_b^1})$ implies:

$$\forall k \in [1:2L] \qquad \left|Com(A^{i'j'})_{k,k}\right| \leq \{M^2\}^{2L-1}.$$

We then deduce from ($\mathbf{H_b^{3,\vartheta}}$) that a constant $C > 0$ exists such that:

$$\begin{array}{rcl}\left\|\left|A^{-1}\right|\right\|_2 & \leq & \max_{(i,j)\in[1:p_n]^2} \frac{2LM^{4L-2}}{det(A^{i'j'})} \\ & \leq & 2C^{-1}LM^{4L-2}n^{\vartheta}.\end{array} \tag{S3.9}$$

Similarly, if we denote $\Delta_{n_1} = A - \hat{A}_{n_1}$, we have:

$$\begin{array}{rcl}\left\|\left|\left(\mathrm{I} - A^{-1}(A - \hat{A}_{n_1})\right)^{-1}\right|\right\|_2 & = & \rho\left(\left(I - A^{-1}\Delta_{n_1}\right)^{-1}\right) \\ & = & \max\limits_{\alpha\in Sp(A^{-1}\Delta_{n_1})} \dfrac{1}{|1-\alpha|},\end{array}$$

using the fact that $A - \hat{A}_{n_1}$ is self-adjoint. We have seen that $\rho(A^{-1}) \leq 2C^{-1}LM^{4L-2}n^{\vartheta}$ and Lemma 1 yields $\rho\left(\Delta_{n_1}\right) = \mathcal{O}_P(n^{-\xi/2})$. As a consequence, we have

$$\max_{\alpha\in Sp(A^{-1}\Delta_{n_1})} |\alpha| \leq \rho(A^{-1})\rho\left(\Delta_{n_1}\right) = \mathcal{O}_P(n^{\vartheta-\xi/2}).$$

Finally, it should be observed that:

$$\max_{\alpha\in Sp(A^{-1}\Delta_{n_1})} \frac{1}{|1-\alpha|} - 1 = \max_{\alpha\in Sp(A^{-1}\Delta_{n_1})} \frac{1-|1-\alpha|}{|1-\alpha|}$$

We know that for n large enough, each absolute value of $\alpha \in Sp(A^{-1}\Delta_{n_1})$ becomes smaller than $1/2$ with a probability tending to one. Hence, with probability tending to one, we have:

$$\max_{\alpha\in Sp(A^{-1}\Delta_{n_1})} \left|\frac{1-|1-\alpha|}{|1-\alpha|}\right| \leq \max_{\alpha\in Sp(A^{-1}\Delta_{n_1})} \frac{|\alpha|}{1-\alpha} \leq 2\rho(A^{-1}\Delta_{n_1}).$$

Since $\rho(A^{-1}\Delta_{n_1}) = \mathcal{O}_P(n^{\vartheta-\xi/2})$, we deduce that:

$$\sup_{i,j,\boldsymbol{l_{ij}}} \left\|\left|\left(\mathrm{I} - A^{-1}(A - \hat{A}_{n_1})\right)^{-1}\right|\right\|_2 \leq 1 + 2LM^{4L-2}C^{-1}\mathcal{O}_P(n^{\vartheta-\xi/2}). \tag{S3.10}$$

To conclude the proof, we can now apply the same argument as the one used in Lemmas 1 and 2 with Bernstein's Inequality, using Equations (S3.9), (S3.10) and the assumption on the uniform bound $\|\boldsymbol{\lambda}\|_2 < \Lambda$ over all the couples (i, j) for the norm $\left\|\boldsymbol{\lambda}^{l_{ij}}\right\|_2$. □

The last lemma finally compares the constant $\hat{C}^{n_1}$ with C.

Lemma 4. *Under Assumptions* ($\mathbf{H_b}$)*, we have:*

$$\sup_{i,j,\boldsymbol{l_{ij}}} \left|\hat{C}^{n_1} - C\right| = \mathcal{O}_P(n^{-\xi/2}).$$

Proof. For any couple (i,j), remark that constants $\hat{C}^{n_1}$ and C satisfy:

$$C = -\langle \phi^i_{l_i} \times \phi^j_{l_j}, 1\rangle \quad \text{and} \quad \hat{C}^{n_1} = -\langle \phi^i_{l_i} \times \phi^j_{l_j}, 1\rangle_{n_1}.$$

If we designate

$$\Delta_{i,j,\boldsymbol{l_{ij}}} := \frac{1}{n_1}\sum_{r=1}^{n_1} \phi^i_{l_i}(x_i{}^r)\phi^j_{l_j}(x_j{}^r) - \mathbb{E}(\phi^i_{l_i}(X_i)\phi^j_{l_j}(X_j)),$$

we can again apply Bernstein's Inequality on $(\phi^i_{l_i}(x_i{}^r)\phi^j_{l_j}(x_j{}^r))_{r=1,\cdots,n_1}$. From $(\mathbf{H^1_b})$, these independent random variables are bounded by M^2 and:

$$\begin{aligned}
P\left(\sup_{i,j,\boldsymbol{l_{ij}}} \left|\Delta_{i,j,\boldsymbol{l_{ij}}}\right| \geq \gamma n_1^{-\xi/2}\right) &\leq \sum_{i,j,\boldsymbol{l_{ij}}} P\left(\left|\Delta_{i,j,\boldsymbol{l_{ij}}}\right| \geq \gamma n_1^{-\xi/2}\right) \\
&\leq \sum_{i,j,\boldsymbol{l_{ij}}} 2\exp\left(-\frac{1}{2}\frac{\gamma^2 n_1^{1-\xi}}{M^4 + M^2\gamma/3n_1^{-\xi/2}}\right) \\
&\leq 2L^2 p_n^2 \exp\left(-\frac{1}{2}\frac{\gamma^2 n_1^{1-\xi}}{M^4 + M^2\gamma/3n_1^{-\xi/2}}\right).
\end{aligned}$$

Under Assumption $(\mathbf{H^2_b})$, the right-hand side of this inequality can be arbitrarly small for n large enough, which ends the proof. □

To finish the proof of Theorem 1, remark that:

$$\begin{aligned}
\left\|\hat{\phi}^{ij}_{\boldsymbol{l_{ij}},n_1} - \phi^{ij}_{\boldsymbol{l_{ij}}}\right\| &= \left\|\sum_{k=1}^L (\hat{\lambda}^i_{k,n_1} - \lambda^i_k)\phi^i_k + \sum_{k=1}^L (\hat{\lambda}^j_{k,n_1} - \lambda^j_k)\phi^j_k + (\hat{C}^{n_1} - C)\right\| \\
&\leq \underbrace{\left\|\sum_{k=1}^L (\hat{\lambda}^i_{k,n_1} - \lambda^i_k)\phi^i_k + \sum_{k=1}^L (\hat{\lambda}^j_{k,n_1} - \lambda^j_k)\phi^j_k\right\|}_{I} + \left|\hat{C}^{n_1} - C\right|.
\end{aligned}$$

Moreover,

$$\begin{aligned}
I^2 &= \int \left(\sum_{k=1}^L (\hat{\lambda}^i_{k,n_1} - \lambda^i_k)\phi^i_k + \sum_{k=1}^L (\hat{\lambda}^j_{k,n_1} - \lambda^j_k)\phi^j_k\right)^2 p_{X_i,X_j}(x_i,x_j)dx_i dx_j \\
&= \underbrace{\int \left(\sum_{k=1}^L (\hat{\lambda}^i_{k,n_1} - \lambda^i_k)\phi^i_k\right)^2 p_{X_i}(x_i)dx_i}_{I_1} + \underbrace{\int \left(\sum_{k=1}^L (\hat{\lambda}^j_{k,n_1} - \lambda^j_k)\phi^j_k\right)^2 p_{X_j}(x_j)dx_j}_{I_2} \\
&\quad \underbrace{+2\int \left(\sum_{k=1}^L (\hat{\lambda}^i_{k,n_1} - \lambda^i_k)\phi^i_k\right)\left(\sum_{k=1}^L (\hat{\lambda}^i_{k,n_1} - \lambda^i_k)\phi^i_k\right) p_{X_i,X_j}(x_i,x_j)dx_i dx_j}_{I_3}.
\end{aligned}$$

Using the inequality $2ab \leq a^2 + b^2$, we deduce that $I_3 \leq I_1 + I_2$, and:

$$\begin{array}{lll} I_1 & = & \int \sum_{k=1}^{L} \sum_{m=1}^{L} (\hat{\lambda}^i_{k,n_1} - \lambda^i_k)(\hat{\lambda}^i_{m,n_1} - \lambda^i_m)\phi^i_k(x_i)\phi^i_m(x_i)p_{X_i}(x_i)dx_i \\ & = & \sum_{k=1}^{L} (\hat{\lambda}^i_{k,n_1} - \lambda^i_k)^2 \quad \text{by orthonormality.} \end{array}$$

The same equality is satisfied for I_2: $I_2 = \sum_{k=1}^{L} (\hat{\lambda}^j_{k,n_1} - \lambda^j_k)^2$.

Consequently, we obtain:

$$\begin{array}{lll} \left\|\hat{\phi}^{ij}_{\boldsymbol{l_{ij}},n_1} - \phi^{ij}_{\boldsymbol{l_{ij}}}\right\| & \leq & \sqrt{2\left[\sum_{k=1}^{L} (\hat{\lambda}^i_{k,n_1} - \lambda^i_k)^2 + \sum_{k=1}^{L} (\hat{\lambda}^j_{k,n_1} - \lambda^j_k)^2\right]} + \left|\hat{C}^{n_1} - C\right| \\ & = & \sqrt{2}\left\|\hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda}\right\|_2 + \left|\hat{C}^{n_1} - C\right|. \end{array} \tag{S3.11}$$

The end of the proof follows with Lemmas 3 and 4.

□

S4 Proof of Theorem 2

We recall first that $\langle , \rangle$ designates the theoretical inner product based on the law $P_{\mathbf{X}}$ (and $\|\|$ is the derived Hilbertian norm). A careful inspection of the Gram-Schmidt procedure used to build the HOFD shows that:

$$M^* := \sup_{u,\boldsymbol{l_u}} \left\|\phi^u_{\boldsymbol{l_u}}(\mathbf{X}_u)\right\|_\infty < \infty,$$

provided that $(\mathbf{H^1_b})$ holds.

We can now observe that the EHOFD is obtained through the first sample $\mathcal{O}_1$ which determines the first empirical inner product $\langle , \rangle_{n_1}$, although the $\mathbb{L}^2$-boosting depends on the second sample $\mathcal{O}_2$. Indeed, $\mathcal{O}_2$ determines the second empirical inner product $\langle , \rangle_{n_2}$. Hence, $\langle , \rangle_{n_2}$ uses observations which are *independent* to the ones used to build the HOFD.

We begin this section with a lemma that establishes that the estimated functions $\hat{\phi}^u_{\boldsymbol{l_u},n_1}$ (which result in the EHOFD) are bounded.

Lemma 5. *Under Assumption* $(\mathbf{H_b})$, *define*

$$N_{n_1} := \sup_{u,\boldsymbol{l_u}} \left\|\hat{\phi}^u_{\boldsymbol{l_u},n_1}(\mathbf{X}_u)\right\|_\infty.$$

We then have:

$$N_{n_1} - M^* = \mathcal{O}_P(n^{\vartheta - \xi/2}).$$

Proof. Using the decomposition of $\hat{\phi}^u_{\boldsymbol{l_u},n_1}$ on the dictionary, Assumption ($\mathbf{H^2_b}$) and Cauchy-Schwarz Inequality, a fixed constant $C>0$ exists such that for all $u \in S$, $\boldsymbol{l_u}$:

$$\forall x \in \mathbb{R}^p \qquad |\hat{\phi}^u_{\boldsymbol{l_u},n_1}(x) - \phi^u_{\boldsymbol{l_u}}(x)| \leq CM\sqrt{L}\sqrt{\left\|\hat{\boldsymbol{\lambda}}_{n_1} - \boldsymbol{\lambda}\right\|_2 + \left\|\hat{C}^{n_1}_{\boldsymbol{l_u}} - C_{\boldsymbol{l_u}}\right\|}.$$

The conclusion then follows using Lemmas 3 and 4. □

We now present a key lemma that compares the elements $(\phi^u_{\boldsymbol{l_u}})_{\boldsymbol{l_u},u}$ with their estimated version $(\hat{\phi}^u_{\boldsymbol{l_u},n_1})_{\boldsymbol{l_u},u}$.

Lemma 6. *Assume that* ($\mathbf{H_b}$) *holds with* $\xi \in (0,1)$*, that the noise* ε *satisfies* ($\mathbf{H}_{\varepsilon,\mathbf{q}}$) *with* $q > 4/\xi$ *and that* ($\mathbf{H}_{\mathbf{s},\alpha}$) *is fullfilled. Then, the following inequalities hold:*

(i)
$$\sup_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} |\langle \hat{\phi}^u_{\boldsymbol{l_u},n_1}, \hat{\phi}^v_{\boldsymbol{l_v},n_1}\rangle - \langle \phi^u_{\boldsymbol{l_u}}, \phi^v_{\boldsymbol{l_v}}\rangle| = \zeta_{n,1} = \mathcal{O}_P(n^{\vartheta-\xi/2})$$

(ii)
$$\sup_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} |\langle \hat{\phi}^u_{\boldsymbol{l_u},n_1}, \hat{\phi}^v_{\boldsymbol{l_v},n_1}\rangle_{n_2} - \langle \phi^u_{\boldsymbol{l_u}}, \phi^v_{\boldsymbol{l_v}}\rangle| = \zeta_{n,2} = \mathcal{O}_P(n^{\vartheta-\xi/2})$$

(iii)
$$\sup_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} |\langle \varepsilon, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2}| = \zeta_{n,3} = \mathcal{O}_P(n^{-\xi/2})$$

(iv)
$$\sup_{u,\boldsymbol{l_u}} \left|\langle \tilde{f}, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2} - \langle \tilde{f}, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle\right| = \|\boldsymbol{\beta^0}\|_{L^1}\mathcal{O}_P(n^{-\xi/2})$$

In the sequel, we will designate $\zeta_n := \max_{i\in[1:3]}\{\zeta_{n,i}\}$.

Proof. **Assertion** (i) Let $u,v \in S$, $\boldsymbol{l_u} \in [1:L]^{|u|}$ and $\boldsymbol{l_v} \in [1:L]^{|v|}$. We then have

$$\begin{aligned}
\left|\langle \hat{\phi}^u_{\boldsymbol{l_u},n_1}, \hat{\phi}^v_{\boldsymbol{l_v},n_1}\rangle - \langle \phi^u_{\boldsymbol{l_u}}, \phi^v_{\boldsymbol{l_v}}\rangle\right| &\leq \left|\langle \hat{\phi}^u_{\boldsymbol{l_u},n_1} - \phi^u_{\boldsymbol{l_u}}, \hat{\phi}^v_{\boldsymbol{l_v},n_1}\rangle - \langle \phi^u_{\boldsymbol{l_u}}, \phi^v_{\boldsymbol{l_v}} - \hat{\phi}^v_{\boldsymbol{l_v},n_1}\rangle\right| \\
&\leq \left\|\hat{\phi}^u_{\boldsymbol{l_u},n_1} - \phi^u_{\boldsymbol{l_u}}\right\| \left\|\hat{\phi}^v_{\boldsymbol{l_v},n_1}\right\| + \|\phi^u_{\boldsymbol{l_u}}\| \left\|\hat{\phi}^v_{\boldsymbol{l_v},n_1} - \phi^v_{\boldsymbol{l_v}}\right\| \\
&\leq \left\|\hat{\phi}^u_{\boldsymbol{l_u},n_1} - \phi^u_{\boldsymbol{l_u}}\right\| \left(\left\|\hat{\phi}^v_{\boldsymbol{l_v},n_1} - \phi^v_{\boldsymbol{l_v}}\right\| + 1\right) + \left\|\hat{\phi}^v_{\boldsymbol{l_v},n_1} - \phi^v_{\boldsymbol{l_v}}\right\|,
\end{aligned}$$

and the conclusion holds applying Theorem 1.

Assertion (ii) We break down the term into two parts:

$$\begin{aligned}
\left|\langle \hat{\phi}^u_{\boldsymbol{l_u},n_1}, \hat{\phi}^v_{\boldsymbol{l_v},n_1}\rangle_{n_2} - \langle \phi^u_{\boldsymbol{l_u}}, \phi^v_{\boldsymbol{l_v}}\rangle\right| \quad &\leq \underbrace{\left|\langle \hat{\phi}^u_{\boldsymbol{l_u},n_1}, \hat{\phi}^v_{\boldsymbol{l_v},n_1}\rangle_{n_2} - \langle \hat{\phi}^u_{\boldsymbol{l_u},n_1}, \hat{\phi}^v_{\boldsymbol{l_v},n_1}\rangle\right|}_{I} \\
&+ \underbrace{\left|\langle \hat{\phi}^u_{\boldsymbol{l_u},n_1}, \hat{\phi}^v_{\boldsymbol{l_v},n_1}\rangle - \langle \phi^u_{\boldsymbol{l_u}}, \phi^v_{\boldsymbol{l_v}}\rangle\right|}_{II}.
\end{aligned}$$

Assertion (i) implies that:

$$\sup_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} |II| = \mathcal{O}_P(n^{\vartheta-\xi/2}).$$

To control $\sup_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} |I|$, we apply Bernstein's inequality to the family of independent random variables $\left(\hat{\phi}^u_{\boldsymbol{l_u},n_1}(\mathbf{x}^s_u)\hat{\phi}^v_{\boldsymbol{l_v},n_1}(\mathbf{x}^s_v)\right)_{s=1\ldots n_2}$ and we denote:

$$\Delta_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} = \left|\frac{1}{n_2}\sum_{s=1}^{n_2}\hat{\phi}^u_{\boldsymbol{l_u},n_1}(\mathbf{x}^s_u)\hat{\phi}^v_{\boldsymbol{l_v},n_1}(\mathbf{x}^s_v) - \mathbb{E}(\hat{\phi}^u_{\boldsymbol{l_u},n_1}(\mathbf{X}_u)\hat{\phi}^v_{\boldsymbol{l_v},n_1}(\mathbf{X}_v))\right|.$$

Bernstein's inequality then implies that:

$$\begin{aligned}
P\left(\sup_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} \Delta_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} \geq \gamma n_2^{-\xi/2}\right) &\leq P\left(\sup_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} \Delta_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} \geq \gamma n_2^{-\xi/2} \& N_{n_1} < M^*+1\right)\\
&\quad +P\left(\sup_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} \Delta_{u,v,\boldsymbol{l_u},\boldsymbol{l_v}} \geq \gamma n_2^{-\xi/2} \& N_{n_1} > M^*+1\right)\\
&\leq 64L^4 p_n^4 \exp\left(-\frac{1}{2}\frac{\gamma^2 {n_2}^{1-\xi}}{(M^*+1)^4 + (M^*+1)^2\gamma/3{n_2}^{-\xi/2})}\right)\\
&\quad +P\left(N_{n_1} > M^*+1\right)
\end{aligned}$$

Lemma 5 and Assumption $(\mathbf{H_b^2})$ yields (ii).

Assertion (iii) The proof follows the roadmap of (ii) of Lemma 1 of Bühlmann (2006). We define the truncated variable ε_t for all $s \in [1:n_2]$:

$$\varepsilon^s_t = \begin{cases} \varepsilon^s & \text{if } |\varepsilon^s| \leq K_n \\ sg(\varepsilon^s)K_n & \text{if } |\varepsilon^s| > K_n \end{cases}$$

where $sg(\varepsilon)$ is the sign of ε. Then, for $\gamma > 0$, we have:

$$\begin{aligned}
P\left(n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\langle\hat{\phi}^u_{\boldsymbol{l_u},n_1},\varepsilon\rangle_{n_2}\right| > \gamma\right) &\leq P\left(n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\langle\hat{\phi}^u_{\boldsymbol{l_u},n_1},\varepsilon_t\rangle_{n_2} - \langle\hat{\phi}^u_{\boldsymbol{l_u},n_1},\varepsilon_t\rangle\right| > \gamma/3\right)\\
&\quad +P\left(n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\langle\hat{\phi}^u_{\boldsymbol{l_u},n_1},\varepsilon-\varepsilon_t\rangle_{n_2}\right| > \gamma/3\right)\\
&\quad +P\left(n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\langle\hat{\phi}^u_{\boldsymbol{l_u},n_1},\varepsilon_t\rangle\right| > \gamma/3\right)\\
&= I + II + III
\end{aligned}$$

Term II: We can bound II using the following simple inclusion:

$$\begin{aligned}\left\{n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\langle\hat{\phi}^u_{\boldsymbol{l_u},n_1},\varepsilon_t\rangle_{n_2}-\langle\hat{\phi}^u_{\boldsymbol{l_u},n_1},\varepsilon_t\rangle\right|>\gamma/3\right\} &\subset \{s\text{exists such that}\, \varepsilon^s-\varepsilon^s_t\neq 0\}\\ &= \{s\text{exists such that}\, |\varepsilon^s|>K_n\}\end{aligned}$$

Hence,

$$\begin{aligned} II &\leq P(\text{some } |\varepsilon^s|>K_n)\\ &\leq n_2 P(|\varepsilon|>K_n)\leq n_2 K_n^{-q}\mathbb{E}(|\varepsilon|^q)=\underset{n\to+\infty}{\mathcal{O}}(n^{1-q\xi/4}),\end{aligned}$$

where $n_2=n/2$ with the choice $K_n:=n^{\xi/4}$, since $q>4/\xi$ by Assumption of the Lemma. Hence, II can become arbitrarily small.

Term I: Applying Bernstein's Inequality again to the family of independent random variables $(\hat{\phi}^u_{\boldsymbol{l_u},n_1}(\mathbf{x}^s_u)\varepsilon^s_t)_{s=1,\cdots,n_2}$ and considering the two events $\{N_{n_1}>M^*+1\}$ and $\{N_{n_1}<M^*+1\}$, we have:

$$I\leq 2Lp_n\exp\left(-\frac{1}{2}\frac{(\gamma^2/9){n_2}^{1-\xi}}{(M^*+1)^4\sigma^2+(M^*+1)K_n\gamma/9{n_2}^{-\xi/2}}\right)+P(N_{n_1}>M^*+1),$$

where $\sigma^2:=\mathbb{E}(|\varepsilon|^2)$. We can then make the right-hand side of the previous inequality arbitrarily small owing to $(\mathbf{H^2_b})$ with $K_n=n^{\xi/2}$.

Term III: by assumption, $\mathbb{E}(\phi^u_{\boldsymbol{l_u}}(\mathbf{X}_u)\varepsilon)=0$. We then have:

$$\begin{aligned} III &\leq P\left(n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\mathbb{E}[(\hat{\phi}^u_{\boldsymbol{l_u},n_1}-\phi^u_{\boldsymbol{l_u}})(\mathbf{X}_u)\varepsilon_t]\right|>\gamma/6\right)+P\left(n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\mathbb{E}[\phi^u_{\boldsymbol{l_u}}(\mathbf{X}_u)(\varepsilon-\varepsilon_t)]\right|>\gamma/6\right)\\ &= III_1+III_2,\end{aligned}$$

with,

$$\begin{aligned} III_1 &= P\left(n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\mathbb{E}[(\hat{\phi}^u_{\boldsymbol{l_u},n_1}-\phi^u_{\boldsymbol{l_u}})(\mathbf{X}_u)]\right|\,|\mathbb{E}(\varepsilon_t)|>\gamma/6\right)\\ &\leq P\left(n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\mathbb{E}[(\hat{\phi}^u_{\boldsymbol{l_u},n_1}-\phi^u_{\boldsymbol{l_u}})(\mathbf{X}_u)]\right|\,|\mathbb{E}(\varepsilon_t)|>\gamma/6\right)\\ &\leq \mathbb{1}_{\{n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\mathbb{E}[(\hat{\phi}^u_{\boldsymbol{l_u},n_1}-\phi^u_{\boldsymbol{l_u}})(\mathbf{X}_u)]\right||\mathbb{E}(\varepsilon_t)|>\gamma/6\}}\end{aligned}$$

Moreover, we have:

$$\begin{aligned}
|\mathbb{E}(\varepsilon_t)| &= \left|\int_{|x|\leq K_n} x\mathrm{d}P_\varepsilon(x) + \int_{|x|>K_n} sg(x)K_n\mathrm{d}P_\varepsilon(x)\right| = \left|\int_{|x|>K_n}(sg(x)K_n - x)\mathrm{d}P_\varepsilon(x)\right| \\
&\leq \int \mathbb{1}_{|x|>K_n}(K_n + |x|)\mathrm{d}P_\varepsilon(x) \\
&\leq K_n P_\varepsilon(|\varepsilon| > K_n) + \int |x|\,\mathbb{1}_{|x|>K_n}\mathrm{d}P_\varepsilon(x) \\
&\leq K_n^{1-t}\mathbb{E}(|\varepsilon|^t) + \mathbb{E}(\varepsilon^2)^{1/2}K_n^{-t/2}\mathbb{E}(|\varepsilon|^t)^{1/2} \quad \text{by the Tchebychev Inequality} \\
&\leq \mathcal{O}(K_n^{1-t}) + \mathcal{O}(K_n^{-t/2}) = o(K_n^{-2})
\end{aligned} \tag{S4.1}$$

since $0 < \xi < 1$ and $t > 4/\xi > 4$. With the choice $K_n = n^{\xi/4}$, we obtain:

$$n_2^{\xi/2}\left\|\hat{\phi}^u_{\boldsymbol{l_u},n_1} - \phi^u_{\boldsymbol{l_u}}\right\| |\mathbb{E}(\varepsilon_t)| \leq n_2^{\xi/2}o(1)o(n^{-\xi/2}) = o(1),$$

when o is the usual Landau notation of relative insignificance.

Hence, $III_1 = 0$ for large enough n. For III_2, we have:

$$III_2 \leq \mathbb{1}_{\{n_2^{\xi/2}\sup_{u,\boldsymbol{l_u}}\left|\mathbb{E}[\phi^u_{\boldsymbol{l_u}}(\mathbf{X}_u)(\varepsilon-\varepsilon_t)]\right|>\gamma/6\}},$$

and, by independance:

$$\left|\mathbb{E}[\phi^u_{\boldsymbol{l_u}}(\mathbf{X}_u)(\varepsilon - \varepsilon_t)]\right| = \left|\mathbb{E}[\phi^u_{\boldsymbol{l_u}}(\mathbf{X}_u)]\right| |\mathbb{E}(\varepsilon - \varepsilon_t)| \leq M^* |\mathbb{E}(\varepsilon - \varepsilon_t)|.$$

Equation (S4.1) then implies:

$$|\mathbb{E}(\varepsilon - \varepsilon_t)| = \left|\int_{|x|>K_n}(sg(x)K_n - x)\mathrm{d}P_\varepsilon(x)\right| \leq o(K_n^{-2}) = o(n^{-\xi/2})$$

Thus, III is arbitrarily small for n and γ large enough and (iii) holds.

Assertion (iv) Note that:

$$\sup_{u,\boldsymbol{l_u}}\left|\langle\tilde{f},\hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2} - \langle\tilde{f},\hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle\right| \leq \|\boldsymbol{\beta^0}\|_{L^1}\sup_{u,\boldsymbol{l_u}}\left|\langle\phi^v_{\boldsymbol{l_v}},\hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2} - \langle\phi^v_{\boldsymbol{l_v}},\hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle\right|.$$

Now, $(\mathbf{H}_{\mathbf{s},\alpha})$ and Bernstein's Inequality imply:

$$\begin{aligned}
P\left(\sup_{u,\boldsymbol{l_u}}\left|\langle\phi^v_{\boldsymbol{l_v}},\hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2} - \langle\phi^v_{\boldsymbol{l_v}},\hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle\right| \geq \gamma n_2^{-\xi/2}\right) \quad &\leq \quad P(N_{n_1} > M^* + 1) \\
+2Lp_n\exp\left(-\frac{1}{2}\frac{\gamma^2 {n_2}^{1-\xi}}{(M^*+1)^4 + (M^*+1)^2\gamma/3{n_2}^{-\xi/2}}\right),&
\end{aligned}$$

which implies with Assumption $(\mathbf{H_b^2})$ that:

$$\sup_{u,\boldsymbol{l_u}}\left|\langle\phi^v_{\boldsymbol{l_v}},\hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2} - \langle\phi^v_{\boldsymbol{l_v}},\hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle\right| = \mathcal{O}_P(n^{-\xi/2}).$$

□

The following lemma, similar to Lemma 2 of Bühlmann (2006), holds:

Lemma 7. *Under Assumptions* $(\mathbf{H_b})$, $(\mathbf{H}_{\varepsilon,\mathbf{q}})$ *with* $q > 4/\xi$, *a constant* $C > 0$ *exists such that, on the set* $\Omega_n = \{\omega, |\zeta_n(\omega)| < 1/2\}$*:*

$$\sup_{u,\boldsymbol{l_u}} |\langle Y - G_k(\bar{f}), \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2} - \langle \tilde{R}_k(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle| \leq \left(\frac{5}{2}\right)^k (1 + C\|\boldsymbol{\beta^0}\|_{L^1})\zeta_n.$$

Proof. Denote $A_n(k,u) = \langle Y - G_k(\bar{f}), \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2} - \langle \tilde{R}_k(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle$. Assume first that $k = 0$:

$$\begin{aligned}
\sup_{u,\boldsymbol{l_u}} |A_n(0,u)| &= \sup_u |\langle Y, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2} - \langle \bar{f}, \phi^u_{\boldsymbol{l_u}}\rangle| \\
&\leq \sup_{u,\boldsymbol{l_u}} \left\{ \left|\langle \tilde{f}, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2} - \langle \tilde{f}, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle\right| + \left|\langle \tilde{f} - \bar{f}, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle\right| + \left|\langle \bar{f}, \hat{\phi}^u_{\boldsymbol{l_u},n_1} - \phi^u_{\boldsymbol{l_u}}\rangle\right| \right\} \\
&\quad + \sup_{u,\boldsymbol{l_u}} \left|\langle \varepsilon, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2}\right| \\
&\leq (1 + 4\|\boldsymbol{\beta^0}\|_{L^1})\zeta_n \quad \text{by } (iii)-(iv) \text{ of Lemma 6 and Theorem 1}
\end{aligned}$$

Referring to the main document, we recall that:

$$G_k(\bar{f}) = G_{k-1}(\bar{f}) + \gamma\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}\rangle_{n_2} \cdot \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \tag{S4.2}$$

$$\begin{aligned}
R_k(\bar{f}) &= \bar{f} - G_k(\bar{f}) \\
&= \bar{f} - G_{k-1}(\bar{f}) - \gamma\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}\rangle_{n_2} \cdot \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}
\end{aligned} \tag{S4.3}$$

and

$$\left\{ \begin{array}{l} \tilde{R}_0(\bar{f}) = \bar{f} \\ \tilde{R}_k(\bar{f}) = \tilde{R}_{k-1}(\bar{f}) - \gamma\langle \tilde{R}_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}\rangle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}. \end{array} \right. \tag{S4.4}$$

The recursive relations (S4.2) and (S4.4) leads to, for any $k \geq 0$:

$$\begin{aligned}
A_n(k,u) &= \langle Y - G_{k-1}(\bar{f}) - \gamma\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}\rangle_{n_2} \cdot \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_n \\
&\quad - \langle \tilde{R}_{k-1}(\bar{f}) - \gamma\langle \tilde{R}_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}\rangle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \phi^u_{\boldsymbol{l_u}}\rangle \\
&\leq A_n(k-1,u) \\
&\quad -\gamma \underbrace{\left(\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}\rangle_{n_2} - \langle \tilde{R}_{k-1}(\bar{f}), \phi^{u_k}_{\boldsymbol{l_{u_k}}}\rangle\right) \langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2}}_{I} \\
&\quad +\gamma \underbrace{\langle \tilde{R}_{k-1}(\bar{f}), \phi^{u_k}_{\boldsymbol{l_{u_k}}}\rangle \left(\langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \phi^u_{\boldsymbol{l_u}}\rangle - \langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle_{n_2}\right)}_{II} \\
&\quad +\gamma \underbrace{\langle \tilde{R}_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1} - \phi^{u_k}_{\boldsymbol{l_{u_k}}}\rangle \langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \phi^u_{\boldsymbol{l_u}}\rangle}_{III}.
\end{aligned}$$

On the one hand, using assertion (ii) of Lemma 6, and the Cauchy-Schwarz inequality (with $\left\|\phi^u_{\boldsymbol{l_u}}\right\| = 1$), we can deduce that:

$$\begin{aligned}
\sup_{u,\boldsymbol{l_u}}|I| &\leq \sup_{u,\boldsymbol{l_u}}|\langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \hat{\phi}^{u}_{\boldsymbol{l_u},n_1}\rangle_{n_2}|\sup_{u,\boldsymbol{l_u}}|A_n(k-1,u)| \\
&\leq (\sup_{u,\boldsymbol{l_u}}|\langle \phi^{u_k}_{\boldsymbol{l_{u_k}}}, \phi^{u}_{\boldsymbol{l_u}}\rangle| + \zeta_n)\sup_{u,\boldsymbol{l_u}}|A_n(k-1,u)| \\
&\leq (1+\zeta_n)\sup_{u,\boldsymbol{l_u}}|A_n(k-1,u)|.
\end{aligned}$$

Now consider now the phantom residual on the basis of its recursive relationship. We can show that: $\left\|\tilde{R}_k(\bar{f})\right\|^2 = \left\|\tilde{R}_{k-1}(\bar{f})\right\|^2 - \gamma(2-\gamma)\langle \tilde{R}_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}\rangle^2 \leq \left\|\tilde{R}_{k-1}(\bar{f})\right\|^2$ and we deduce

$$\left\|\tilde{R}_k(\bar{f})\right\|^2 \leq \left\|\bar{f}\right\|^2. \tag{S4.5}$$

Then,

$$\begin{aligned}
\sup_{u,\boldsymbol{l_u}}|II| &\leq \left\|\tilde{R}_{k-1}(\bar{f})\right\| \left\|\phi^{u_k}_{\boldsymbol{l_{u_k}}}\right\| \sup_{u,\boldsymbol{l_u}}|\langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \phi^{u}_{\boldsymbol{l_u}}\rangle - \langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \hat{\phi}^{u}_{\boldsymbol{l_u},n_1}\rangle_{n_2}| \\
&\leq \left\|\bar{f}\right\| \sup_{u,\boldsymbol{l_u}}|\langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \phi^{u}_{\boldsymbol{l_u}}\rangle - \langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \hat{\phi}^{u}_{\boldsymbol{l_u},n_1}\rangle_{n_2}|,
\end{aligned}$$

where

$$\begin{aligned}
|\langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \phi^{u}_{\boldsymbol{l_u}}\rangle - \langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \hat{\phi}^{u}_{\boldsymbol{l_u},n_1}\rangle_{n_2}| &\leq |\langle \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \hat{\phi}^{u}_{\boldsymbol{l_u},n_1}\rangle_{n_2} - \langle \phi^{u_k}_{\boldsymbol{l_{u_k}}}, \phi^{u}_{\boldsymbol{l_u}}\rangle| \\
&\quad + |\langle \phi^{u_k}_{\boldsymbol{l_{u_k}}} - \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}, \phi^{u}_{\boldsymbol{l_u}}\rangle|.
\end{aligned}$$

Using assertion (ii) from Lemma 6 and Theorem 1 again, we obtain the following bound for II:

$$\begin{aligned}
\sup_{u,\boldsymbol{l_u}}|II| &\leq \left\|\bar{f}\right\| (\zeta_n + \sup_{u,\boldsymbol{l_u}}\left\|\phi^{u}_{\boldsymbol{l_u}} - \hat{\phi}^{u}_{\boldsymbol{l_u},n_1}\right\|) \\
&\leq 2\zeta_n \left\|\bar{f}\right\|.
\end{aligned}$$

Finally, Theorem 1 gives:

$$\begin{aligned}
\sup_{u,\boldsymbol{l_u}}|III| &\leq \sup_{u,\boldsymbol{l_u}}\left\|\tilde{R}_{k-1}(\bar{f})\right\| \left\|\hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1} - \phi^{u_k}_{\boldsymbol{l_{u_k}}}\right\| \left\|\hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}\right\| \left\|\phi^{u}_{\boldsymbol{l_u}}\right\| \\
&\leq \left\|\bar{f}\right\| \zeta_n.
\end{aligned}$$

Our bounds on I, II and III, and $\gamma < 1$ on the set $\Omega_n = \{\zeta_n < 1/2\}$ yields that

$$\begin{aligned}
\sup_{u,\boldsymbol{l_u}}|A_n(k,u)| &\leq \sup_{u,\boldsymbol{l_u}}|A_n(k-1,u)| + (1+\zeta_n)\sup_{u,\boldsymbol{l_u}}|A_n(k-1,u)| + 3\zeta_n \left\|\bar{f}\right\| \\
&\leq \frac{5}{2}\sup_{u,\boldsymbol{l_u}}|A_n(k-1,u)| + 3\zeta_n \left\|\bar{f}\right\|.
\end{aligned}$$

A simple induction indicates that:

$$\begin{aligned}\sup_{u,\boldsymbol{l_u}}|A_n(k,u)| &\leq \left(\frac{5}{2}\right)^k \underbrace{\sup_{u,\boldsymbol{l_u}}|A_n(0,u)|}_{\leq(1+4\|\boldsymbol{\beta^0}\|_{L^1})\zeta_n} +3\zeta_n \|\bar{f}\| \sum_{\ell=0}^{k-1}\left(\frac{5}{2}\right)^\ell \\ &\leq \left(\frac{5}{2}\right)^k \zeta_n \left(1+\|\boldsymbol{\beta^0}\|_{L^1}\left(4+6\sum_{\ell=1}^{\infty}\left(\frac{5}{2}\right)^{-\ell}\right)\right),\end{aligned}$$

which ends the proof with $C = 14$. □

We then aim at applying Theorem 2.1 from Champion et al. (2013) to the phantom residuals $(\tilde{R}_k(\bar{f}))_k$. Using the notation of Champion et al. (2013), this will be possible if we can show that the phantom residuals follow a theoretical boosting with a shrinkage parameter $\nu \in [0,1]$. Thanks to Lemma 7 and by definiton of $\hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}$, we have:

$$\begin{aligned}|\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}\rangle_{n_2}| &= \sup_{u,\boldsymbol{l_u}}|\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u}_{\boldsymbol{l_u},n_1}\rangle_{n_2}| \\ &\geq \sup_{u,\boldsymbol{l_u}}\left\{|\langle \tilde{R}_{k-1}(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle| - C\left(\frac{5}{2}\right)^{k-1}\zeta_n\|\boldsymbol{\beta^0}\|_{L^1}\right\}. \quad \text{(S4.6)}\end{aligned}$$

Applying again Lemma 7 on the set Ω_n, we obtain:

$$\begin{aligned}|\langle \tilde{R}_{k-1}(\bar{f}), \phi^{u_k}_{\boldsymbol{l_{u_k}}}\rangle| &\geq |\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l_{u_k}},n_1}\rangle_{n_2}| - C\left(\frac{5}{2}\right)^{k-1}\zeta_n\|\boldsymbol{\beta^0}\|_{L^1} \\ &\geq \sup_{u,\boldsymbol{l_u}}|\langle \tilde{R}_{k-1}(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle| - 2C\left(\frac{5}{2}\right)^{k-1}\zeta_n\|\boldsymbol{\beta^0}\|_{L^1}. \quad \text{(S4.7)}\end{aligned}$$

Now consider the set:

$$\tilde{\Omega}_n = \left\{\omega, \quad \forall k \leq k_n, \quad \sup_{u,\boldsymbol{l_u}}|\langle \tilde{R}_{k-1}(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle| > 4C\left(\frac{5}{2}\right)^{k-1}\zeta_n\|\boldsymbol{\beta^0}\|_{L^1}\right\}.$$

We deduce from Equation (S4.7) the following inequality on $\Omega_n \cap \tilde{\Omega}_n$:

$$|\langle \tilde{R}_{k-1}(\bar{f}), \phi^{u_k}_{\boldsymbol{l_{u_k}}}\rangle| \geq \frac{1}{2}\sup_{u,\boldsymbol{l_u}}|\langle \tilde{R}_{k-1}(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle|. \quad \text{(S4.8)}$$

Consequently, on $\Omega_n \cap \tilde{\Omega}_n$, the family $(\tilde{R}_k(\bar{f}))_k$ satisfies a theoretical boosting, given by Algorithm 1 of Champion et al. (2013), with constant $\nu = 1/2$ and we have:

$$\left\|\tilde{R}_k(\bar{f})\right\| \leq C'\left(1+\frac{1}{4}\gamma(2-\gamma)k\right)^{-\frac{2-\gamma}{2(6-\gamma)}}. \quad \text{(S4.9)}$$

Now consider the complementary set:

$$\tilde{\Omega}_n^C = \left\{\omega, \quad \exists\, k \le k_n \quad \sup_{u,\boldsymbol{l_u}} |\langle \tilde{R}_{k-1}(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle| \le 4C\left(\frac{5}{2}\right)^{k-1} \zeta_n \|\boldsymbol{\beta^0}\|_{L^1}\right\}.$$

It should be noted that:

$$\begin{aligned}\left\|\tilde{R}_k(\bar{f})\right\|^2 &= \langle \tilde{R}_k(\bar{f}), \bar{f} - \gamma \textstyle\sum_{j=0}^{k-1} \langle \tilde{R}_j(\bar{f}), \hat{\phi}^{u_j}_{\boldsymbol{l_{u_j}},n_1}\rangle \hat{\phi}^{u_j}_{\boldsymbol{l_{u_j}},n_1}\rangle \\ &\le \|\boldsymbol{\beta^0}\|_{L^1} \sup_{u,\boldsymbol{l_u}} \left|\langle \tilde{R}_k(\bar{f}), \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle\right| + \gamma \textstyle\sum_{j=0}^{k-1} \left|\langle \tilde{R}_j(\bar{f}), \hat{\phi}^{u_j}_{\boldsymbol{l_{u_j}},n_1}\rangle\right| \sup_{u,\boldsymbol{l_u}} \left|\langle \tilde{R}_k(\bar{f}), \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle\right|.\end{aligned}$$

Moreover,

$$\begin{aligned}\sup_{u,\boldsymbol{l_u}} \left|\langle \tilde{R}_k(\bar{f}), \hat{\phi}^u_{\boldsymbol{l_u},n_1}\rangle\right| &\le \sup_{u,\boldsymbol{l_u}} \left|\langle \tilde{R}_k(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle\right| + \sup_{u,\boldsymbol{l_u}} \left|\langle \tilde{R}_k(\bar{f}), \hat{\phi}^u_{\boldsymbol{l_u},n_1} - \phi^u_{\boldsymbol{l_u}}\rangle\right| \\ &\le \sup_{u,\boldsymbol{l_u}} \left|\langle \tilde{R}_k(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle\right| + 2\|\boldsymbol{\beta^0}\|_{L^1}\zeta_n \quad \text{by Theorem 1 and (S4.5).}\end{aligned}$$

We therefore have:

$$\begin{aligned}\left\|\tilde{R}_k(\bar{f})\right\|^2 &\le \left(\|\boldsymbol{\beta^0}\|_{L^1} + \gamma \sum_{j=0}^{k-1} \left|\langle \tilde{R}_j(\bar{f}), \hat{\phi}^{u_j}_{\boldsymbol{l_{u_j}},n_1}\rangle\right|\right)\left(\sup_{u,\boldsymbol{l_u}} \left|\langle \tilde{R}_k(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle\right| + 2\|\boldsymbol{\beta^0}\|_{L^1}\zeta_n\right) \\ &\le \|\boldsymbol{\beta^0}\|_{L^1}(1+2\gamma k)\left(\sup_{u,\boldsymbol{l_u}} \left|\langle \tilde{R}_k(\bar{f}), \phi^u_{\boldsymbol{l_u}}\rangle\right| + 2\|\boldsymbol{\beta^0}\|_{L^1}\zeta_n\right) \\ &\le 4C\|\boldsymbol{\beta^0}\|^2_{L^1}\zeta_n(1+2\gamma k)\left(\frac{5}{2}\right)^k \quad \text{on } \tilde{\Omega}_n^C. \qquad \text{(S4.10)}\end{aligned}$$

Finally, using Equations (S4.9) and (S4.10), we have on the set $(\Omega_n \cap \tilde{\Omega}_n) \cup \tilde{\Omega}_n^C$:

$$\left\|\tilde{R}_k(\bar{f})\right\|^2 \le C'^2\left(1 + \frac{1}{4}\gamma(2-\gamma)k\right)^{-\frac{2-\gamma}{6-\gamma}} + 4C\|\boldsymbol{\beta^0}\|^2_{L^1}\zeta_n(1+2\gamma k)\left(\frac{5}{2}\right)^k. \qquad \text{(S4.11)}$$

To conclude the first part of the proof, it should be noted that:

$$P\left((\Omega_n \cap \tilde{\Omega}_n) \cup \tilde{\Omega}_n^C\right) \ge P(\Omega_n) \underset{n\to+\infty}{\longrightarrow} 1.$$

On this set, Inequality (S4.11) holds almost surely, and for $k_n < c\log(n)$ with $c < \frac{\xi/2-\vartheta-2\alpha}{2\log(3)}$, we obtain:

$$\left\|\tilde{R}_{k_n}(\bar{f})\right\| \xrightarrow[n\to+\infty]{P} 0. \qquad \text{(S4.12)}$$

Consider now $A_k := \left\| R_k(\bar{f}) - \tilde{R}_k(\bar{f}) \right\|$ for $k \geq 1$. By definitions reminded in (S4.3)-(S4.4), we have:

$$\begin{aligned} A_k &\leq A_{k-1} + \gamma |\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l}_{\boldsymbol{u_k}}, n_1} \rangle_{n_2} - \langle \tilde{R}_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l}_{\boldsymbol{u_k}}, n_1} \rangle | \\ &\leq A_{k-1} + \gamma |\langle Y - G_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l}_{\boldsymbol{u_k}}, n_1} \rangle_{n_2} - \langle \tilde{R}_{k-1}(\bar{f}), \phi^{u_k}_{\boldsymbol{l}_{\boldsymbol{u_k}}} \rangle | \\ &\quad + \gamma |\langle \tilde{R}_{k-1}(\bar{f}), \hat{\phi}^{u_k}_{\boldsymbol{l}_{\boldsymbol{u_k}}, n_1} - \phi^{u_k}_{\boldsymbol{l}_{\boldsymbol{u_k}}} \rangle |. \end{aligned} \tag{S4.13}$$

By Lemma 7, we then deduce the following inequality on Ω_n:

$$A_k \leq A_{k-1} + \gamma \left(\frac{5}{2} \right)^{k-1} \left(1 + C \|\boldsymbol{\beta^0}\|_{L^1} \right) \zeta_n + 2\gamma \|\boldsymbol{\beta^0}\|_{L^1} \zeta_n. \tag{S4.14}$$

Since $A_0 = 0$, we deduce recursively from Equation (S4.14) that, on Ω_n:

$$A_{k_n} \xrightarrow[n \to +\infty]{P} 0.$$

Finally, since:

$$\left\| \hat{f} - \tilde{f} \right\| = \left\| G_{k_n}(\bar{f}) - \tilde{f} \right\| \leq \left\| \bar{f} - \tilde{f} \right\| + \left\| R_{k_n}(\bar{f}) - \tilde{R}_{k_n}(\bar{f}) \right\| + \left\| \tilde{R}_{k_n}(\bar{f}) \right\|,$$

it remains to deal with the term $\left\| \bar{f} - \tilde{f} \right\|$. However, it should be noted that:

$$\left\| \bar{f} - \tilde{f} \right\| \leq \|\boldsymbol{\beta^0}\|_{L^1} \left\| \phi^u_{\boldsymbol{l_u}} - \hat{\phi}^u_{\boldsymbol{l_u}, n_1} \right\|,$$

and the proof follows using $(\mathbf{H}_{\mathbf{s},\alpha})$ with $\alpha < \xi/4 - \vartheta/2$ and Theorem 1. □

Bibliography

Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375.

Bühlmann, P. (2006). Boosting for high-dimensional linear models. *The Annals of Statistics*, 34(2):559–583.

Bullen, P. (1998). *A dictionary of Inequalities*. Longman.

Champion, M., Cierco-Ayrolles, C., Gadat, S., and Vignes, M. (2013). Sparse regression and support recovery with $\mathbb{L}_2$-boosting algorithm.

Tropp, J. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434.