

# APPENDIX B: CONVERGENCE RESULTS

In this part, we restate and prove Proposition 2 of Section 5 of the paper. For sake of clarity, we first define and recall some notation that will be used further.

## Reminder

First, as mentioned in Section 4 of the paper, we assume that  $Y$  is centered. Also, we assume that a functional ANOVA of order  $d$ , with  $d \leq p$ , substitutes properly the initial model. Let us denote  $S^d$  the collection of all subsets of  $[1 : p]$  of size at most  $d$ , where the empty set has been removed. Thus,  $S^d = \{u, \emptyset \subset u \subseteq [1 : p] \text{ and } |u| \leq d\}$ .

Recall that,  $\forall i \in [1 : p]$ ,  $L_i := L_{\{i\}}$  is the dimension of the spaces  $H_i^{0,L}$  and  $G_{i,n}^{0,L}$ . More generally,  $\phi_{\{i\}} := \phi_i$ . Also,  $L_u := \prod_{i \in u} L_i$  is the dimension of the spaces  $H_u^{0,L}$  and  $G_{u,n}^{0,L}$ .

For  $u \in S^d$ ,  $\mathbf{l}_u = (l_u^i)_{i \in u}$  is a multi-index of  $\times_{i \in u} [1 : L_i]$ , where  $\times_{i \in u} [1 : L_i]$  is the Cartesian product of the sets  $[1 : L_i]$ , for  $i \in u$ .

We refer  $(\phi_{\mathbf{l}_u}^u)_{\mathbf{l}_u \in \times_{i \in u} [1 : L_i]}$  as the basis of  $H_u^{0,L}$  and  $(\varphi_{\mathbf{l}_u,n}^u)_{\mathbf{l}_u \in \times_{i \in u} [1 : L_i]}$  as the basis of  $G_{u,n}^{0,L}$  constructed according to HOGS Procedure of Section 3 of the main document. Thus, these functions all lie in  $L_{\mathbb{R}}^2(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$ .

$\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$  are used as the inner product and norm on  $L_{\mathbb{R}}^2(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$ ,

$$\langle h_1, h_2 \rangle = \int h_1(\mathbf{x})h_2(\mathbf{x})p_{\mathbf{X}}d\nu(\mathbf{x}), \quad \|h\|^2 = \langle h, h \rangle$$

while  $\langle \cdot, \cdot \rangle_n$  and  $\|\cdot\|_n$  denote the empirical inner product and norm, that is

$$\langle g_1, g_2 \rangle_n = \frac{1}{n} \sum_{s=1}^n g_1(\mathbf{x}^s)g_2(\mathbf{x}^s), \quad \|g\|_n^2 = \langle g, g \rangle_n$$

when  $(y^s, \mathbf{x}^s)_{s=1, \dots, n}$  is the  $n$ -sample of observations from the distribution of  $(Y, \mathbf{X})$ .

## New settings

We set  $m := \sum_{u \in S^d} L_u$  the number of parameters in the regression model. Denote, for all  $u \in S^d$ ,  $\Phi_u(\mathbf{X}_u) \in (L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_{\mathbf{X}}))^{L_u}$ , with  $(\Phi_u(\mathbf{X}_u))_{\mathbf{l}_u} = \phi_{\mathbf{l}_u}^u(\mathbf{X}_u)$ , and by  $\boldsymbol{\beta}$  any vector of  $\Theta \subset \mathbb{R}^m$ , where  $(\boldsymbol{\beta})_{\mathbf{l}_u, u} = \beta_{\mathbf{l}_u}^u, \forall \mathbf{l}_u \in \times_{i \in u} [1 : L_i]$ .

Recall that, for  $a, b \in \mathbb{N}^*$ ,  $\mathcal{M}_{a,b}(\mathbb{R})$  denotes the set of all real matrices with  $a$  rows and  $b$  columns. Set  $\mathbb{X}_{\varphi} = \left( \varphi_1 \quad \dots \quad \varphi_u \quad \dots \right) \in \times_{u \in S^d} \mathcal{M}_{n, L_u}(\mathbb{R})$ , where, for  $u \in S^d$ ,  $(\varphi_u)_{s, \mathbf{l}_u} = \varphi_{\mathbf{l}_u, n}^u(\mathbf{x}_u^s), \forall s \in$

$[1 : n]$ ,  $\forall \mathbf{l}_u \in \times_{i \in u} [1 : L_i]$ . Also, we set  $\mathbb{X}_\phi = \begin{pmatrix} \mathfrak{C}_1 & \mathfrak{C}_2 & \dots \end{pmatrix} \in \times_{u \in S^d} \mathcal{M}_{n, L_u}(\mathbb{R})$ , where, for  $u \in S^d$ ,  $(\mathfrak{C}_u)_{s, \mathbf{l}_u} = \phi_{\mathbf{l}_u}^u(\mathbf{x}_u^s)$ ,  $\forall s \in [1 : n]$ ,  $\forall \mathbf{l}_u \in \times_{i \in u} [1 : L_i]$ .

Denote by  $A_\phi$  be the  $m \times m$  Gramian matrix whose block entries are  $(\mathbb{E}(\Phi_u(\mathbf{X}_u)^t \Phi_v(\mathbf{X}_v)))_{u, v \in S^d}$ .

At last, we define the functions

$$M_n(\boldsymbol{\beta}) = \|\mathbb{Y} - \mathbb{X}_\phi \boldsymbol{\beta}\|_n^2 \quad (1)$$

where  $\mathbb{Y}_s = y^s$ ,  $\forall s \in [1 : n]$ .

The Euclidean norm will be denoted  $\|\cdot\|_2$ .

**Proposition .** *Assume that*

$$Y = \eta^R(\mathbf{X}) + \varepsilon, \quad \text{where } \eta^R(\mathbf{X}) = \sum_{u \in S^d} \sum_{\mathbf{l}_u \in \times_{i \in u} [1 : L_i]} \beta_{\mathbf{l}_u}^{u,0} \phi_{\mathbf{l}_u}^u(\mathbf{X}_u) \in H_u^{0,L}$$

with  $\mathbb{E}(\varepsilon) = 0$ ,  $\mathbb{E}(\varepsilon^2) = \sigma_*^2$ ,  $\mathbb{E}(\varepsilon \cdot \phi_{\mathbf{l}_u}^u(\mathbf{X}_u)) = 0$ ,  $\forall \mathbf{l}_u \in \times_{i \in u} [1 : L_i]$ ,  $\forall u \in S^d$ . ( $\boldsymbol{\beta}_0 = (\beta_{\mathbf{l}_u}^{u,0})_{\mathbf{l}_u, u}$  is the true parameter).

Further, let us consider the least-squares estimation  $\hat{\eta}^R$  of  $\eta^R$  using the sample  $(y^s, \mathbf{x}^s)_{s \in [1:n]}$  and the functions  $(\varphi_{\mathbf{l}_u}^u)_{\mathbf{l}_u, u}$ , that is

$$\hat{\eta}^R(\mathbf{X}) = \sum_{u \in S^d} \hat{\eta}_u^R(\mathbf{X}_u), \quad \text{where } \hat{\eta}_u^R(\mathbf{X}_u) = \sum_{\mathbf{l}_u \in \times_{i \in u} [1 : L_i]} \hat{\beta}_{\mathbf{l}_u}^u \varphi_{\mathbf{l}_u}^u(\mathbf{X}_u) \in G_{u,n}^{0,L}$$

where  $\hat{\boldsymbol{\beta}} = \text{Argmin}_{\boldsymbol{\beta} \in \Theta} M_n(\boldsymbol{\beta})$ . If we assume that

(H) The distribution  $P_{\mathbf{X}}$  is equivalent to  $\otimes_{i=1}^p P_{X_i}$ ,

then

$$\|\hat{\eta}^R - \eta^R\| \xrightarrow{a.s.} 0 \quad \text{when } n \rightarrow +\infty. \quad (2)$$

The sketch of proof in Appendix A shows that the proof of Proposition is broken up into Lemmas 1-5. Here, we enunciate and prove them.

## Preliminary results

**Lemma 1.** *If (H) holds, then  $A_\phi$  is a non singular matrix.*

*Proof of Lemma 1*

First of all notice that when we consider a Gramian matrix, by a classical argument on the associated quadratic form, the full rank of this matrix holds if and only if the associated functional vector has full rank in  $L^2$  [1].

To begin with, set, for all  $i \in [1 : p]$ ,  $\Psi_i = \begin{pmatrix} 1 \\ \Phi_i \end{pmatrix}$  and  $G_i := \mathbb{E}(\Psi_i^t \Psi_i)$ . As  $(\phi_{l_i}^i)_{l_i=1}^{L_i}$  is an orthonormal system, we obviously get  $G_i = I_{(L_i+1) \times (L_i+1)}$ . Now we may rewrite the tensor product  $\otimes_{i=1}^p G_i$  as

$$\otimes_{i=1}^p G_i = \otimes_{i=1}^p \mathbb{E}(\Psi_i^t \Psi_i) = \int \otimes_{i=1}^p [\Psi_i(x_i)^t \Psi_i(x_i)] dP_{X_1}(x_1) \cdots dP_{X_p}(x_p) \quad (3)$$

We obviously have  $\otimes_{i=1}^p G_i = I$ . So that, using the remark of the beginning of the proof, the system  $\otimes_{i=1}^p [\Psi_i^t \Psi_i] = \left( 1 \quad (\otimes_{i \in u} \Phi_i)_{\substack{u \subseteq [1:p] \\ u \neq \emptyset}} \right)$  is linearly independent  $(\otimes_{i=1}^p P_{X_i})$ -a.e.

As we assumed that  $\otimes_{i=1}^p P_{X_i}$  and  $P_{\mathbf{X}}$  are equivalent (H), we get that  $\left( 1 \quad (\otimes_{i \in u} \Phi_i)_{\substack{u \subseteq [1:p] \\ u \neq \emptyset}} \right)$  is linearly independent  $P_{\mathbf{X}}$  - a.e..

Now, we may conclude as in the classical Gram-Schmidt construction. Indeed, the construction of the system  $(\Phi_u)_{u \in S^d}$  involves an invertible triangular matrix. ■

**Lemma 2.** Let  $u, v \in S^d$ . Assume that  $\|\varphi_{l_{u,n}}^u - \phi_{l_u}^u\| \xrightarrow{a.s.} 0$  and  $\|\varphi_{l_{v,n}}^v - \phi_{l_v}^v\| \xrightarrow{a.s.} 0$  for any  $l_u \in \times_{i \in u} [1 : L_i]$ ,  $l_v \in \times_{i \in v} [1 : L_i]$ . Then, the following results hold:

(i)  $\|\phi_{l_{u,n}}^u\| \xrightarrow{a.s.} \|\phi_{l_u}^u\|$  and  $\|\varphi_{l_{u,n}}^u\|_n \xrightarrow{a.s.} \|\phi_{l_u}^u\|$ ;

(ii)  $\langle \phi_{l_u}^u, \varphi_{l_{v,n}}^v \rangle \xrightarrow{a.s.} \langle \phi_{l_u}^u, \phi_{l_v}^v \rangle$  and  $\langle \phi_{l_{u,n}}^u, \varphi_{l_{v,n}}^v \rangle_n \xrightarrow{a.s.} \langle \phi_{l_u}^u, \phi_{l_v}^v \rangle$ ;

(iii)  $\langle \varphi_{l_{u,n}}^u, \varphi_{l_{v,n}}^v \rangle \xrightarrow{a.s.} \langle \phi_{l_u}^u, \phi_{l_v}^v \rangle$  and  $\langle \varphi_{l_{u,n}}^u, \varphi_{l_{v,n}}^v \rangle_n \xrightarrow{a.s.} \langle \phi_{l_u}^u, \phi_{l_v}^v \rangle$ .

*Proof of Lemma 2*

The first point (i) is trivial. By continuity, the first result of (ii) holds. The second result of (ii) holds by the law of large numbers, and because  $\|\phi_{l_u}^u\| < +\infty$ ,

$$\begin{aligned} \left| \langle \phi_{l_u}^u, \varphi_{l_{v,n}}^v \rangle_n - \langle \phi_{l_u}^u, \phi_{l_v}^v \rangle \right| &\leq \left| \langle \phi_{l_u}^u, \varphi_{l_{v,n}}^v \rangle_n - \langle \phi_{l_u}^u, \varphi_{l_{v,n}}^v \rangle \right| + \left| \langle \phi_{l_u}^u, \varphi_{l_{v,n}}^v \rangle - \langle \phi_{l_u}^u, \phi_{l_v}^v \rangle \right| \\ &\leq \left| \langle \phi_{l_u}^u, \varphi_{l_{v,n}}^v \rangle_n - \langle \phi_{l_u}^u, \varphi_{l_{v,n}}^v \rangle \right| + \|\phi_{l_u}^u\| \|\varphi_{l_{v,n}}^v - \phi_{l_v}^v\| \\ &\xrightarrow{a.s.} 0 \end{aligned}$$

(iii) follows from

$$\begin{aligned} \left| \langle \varphi_{\mathbf{l}_u, n}^u, \varphi_{\mathbf{l}_v, n}^v \rangle - \langle \phi_{\mathbf{l}_u}^u, \phi_{\mathbf{l}_v}^v \rangle \right| &= \left| \langle \varphi_{\mathbf{l}_u, n}^u - \phi_{\mathbf{l}_u}^u, \varphi_{\mathbf{l}_v, n}^v \rangle + \langle \phi_{\mathbf{l}_u}^u, \varphi_{\mathbf{l}_v, n}^v - \phi_{\mathbf{l}_v}^v \rangle \right| \\ &\leq \left\| \varphi_{\mathbf{l}_u, n}^u - \phi_{\mathbf{l}_u}^u \right\| \left\| \varphi_{\mathbf{l}_v, n}^v \right\| + \left\| \phi_{\mathbf{l}_u}^u \right\| \left\| \varphi_{\mathbf{l}_v, n}^v - \phi_{\mathbf{l}_v}^v \right\| \end{aligned}$$

$\left\| \varphi_{\mathbf{l}_u, n}^u - \phi_{\mathbf{l}_u}^u \right\| \xrightarrow{a.s.} 0$  and  $\left\| \varphi_{\mathbf{l}_v, n}^v - \phi_{\mathbf{l}_v}^v \right\| \xrightarrow{a.s.} 0$ . As  $\left\| \varphi_{\mathbf{l}_u, n}^u \right\| \xrightarrow{a.s.} \left\| \phi_{\mathbf{l}_u}^u \right\|$  (by (i)) and  $\left\| \phi_{\mathbf{l}_u}^u \right\| < +\infty$ . By using the technique to show the second result of (ii), we deduce the second result of (iii).  $\blacksquare$

**Lemma 3.** Let  $u \in S^d$ . For any  $l_u^i \in [1 : L_i]$ , assume that  $\left\| \varphi_{l_u^i, n}^i - \phi_{l_u^i}^i \right\| \xrightarrow{a.s.} 0$ . Moreover, we also assume that there exists  $v \in S^d$  and  $\mathbf{l}_v \in \times_{i \in v} [1 : L_i]$  such that  $\left\| \varphi_{\mathbf{l}_v, n}^v - \phi_{\mathbf{l}_v}^v \right\| \xrightarrow{a.s.} 0$ . Then, we have

$$(i) \left\| \prod_{i \in u} \varphi_{l_u^i, n}^i - \prod_{i \in u} \phi_{l_u^i}^i \right\| \xrightarrow{a.s.} 0;$$

$$(ii) \langle \prod_{i \in u} \varphi_{l_u^i, n}^i, \varphi_{\mathbf{l}_v, n}^v \rangle_n \xrightarrow{a.s.} \langle \prod_{i \in u} \phi_{l_u^i}^i, \phi_{\mathbf{l}_v}^v \rangle.$$

*Proof of Lemma 3*

Let  $u \in S^d$ . By induction on the cardinality of  $u$ , set

$$(\mathcal{H}_{|u|}) \left\| \prod_{i \in u} \varphi_{l_u^i, n}^i - \prod_{i \in u} \phi_{l_u^i}^i \right\| \xrightarrow{a.s.} 0, \quad \forall l_u^i \in [1 : L_i]$$

- If  $u = \{i\}$ , i.e.  $|u| = 1$ ,  $\left\| \prod_{i \in u} \varphi_{l_u^i, n}^i - \prod_{i \in u} \phi_{l_u^i}^i \right\| \xrightarrow{a.s.} 0$ , so  $(\mathcal{H}_1)$  holds.
- Let  $k \in [1 : d]$  and suppose that  $(\mathcal{H}_{|\tilde{u}|})$  is true for any  $1 \leq |\tilde{u}| \leq k-1$ . For  $u$  such that  $|u| = k$ , for given  $\mathbf{l}_u = (l_u^i) \in \times_{i \in u} [1 : L_i]$ , we have,

$$\begin{aligned} \left\| \prod_{i \in u} \varphi_{l_u^i, n}^i - \prod_{i \in u} \phi_{l_u^i}^i \right\| &= \left\| \varphi_{l_u^j, n}^j \left( \prod_{|\tilde{u}|=k-1}^{i \in \tilde{u}} \varphi_{l_u^i, n}^i - \prod_{|\tilde{u}|=k-1}^{i \in \tilde{u}} \phi_{l_u^i}^i \right) + \prod_{|\tilde{u}|=k-1}^{i \in \tilde{u}} \phi_{l_u^i}^i (\varphi_{l_u^j, n}^j - \phi_{l_u^j}^j) \right\| \\ &\leq \left\| \varphi_{l_u^j, n}^j \right\| \left\| \prod_{|\tilde{u}|=k-1}^{i \in \tilde{u}} \varphi_{l_u^i, n}^i - \prod_{|\tilde{u}|=k-1}^{i \in \tilde{u}} \phi_{l_u^i}^i \right\| + \left\| \prod_{|\tilde{u}|=k-1}^{i \in \tilde{u}} \phi_{l_u^i}^i \right\| \left\| \varphi_{l_u^j, n}^j - \phi_{l_u^j}^j \right\| \\ &= \left\| \prod_{|\tilde{u}|=k-1}^{i \in \tilde{u}} \varphi_{l_u^i, n}^i - \prod_{|\tilde{u}|=k-1}^{i \in \tilde{u}} \phi_{l_u^i}^i \right\| + \left\| \varphi_{l_u^j, n}^j - \phi_{l_u^j}^j \right\| \end{aligned}$$

By  $(\mathcal{H}_{|\tilde{u}|})$ ,  $\left\| \prod_{|\tilde{u}|=k-1}^{i \in \tilde{u}} \varphi_{l_u^i, n}^i - \prod_{|\tilde{u}|=k-1}^{i \in \tilde{u}} \phi_{l_u^i}^i \right\| \xrightarrow{a.s.} 0$  and by  $(\mathcal{H}_1)$ ,  $\left\| \varphi_{l_u^j, n}^j - \phi_{l_u^j}^j \right\| \xrightarrow{a.s.} 0$ . Hence  $(\mathcal{H}_k)$  holds.

Therefore, (i) follows. Now, let us prove (ii).

$$\begin{aligned} \left| \langle \prod_{i \in u} \varphi_{l_u^i, n}^i, \varphi_{\mathbf{l}_v, n}^v \rangle_n - \langle \prod_{i \in u} \phi_{l_u^i}^i, \phi_{\mathbf{l}_v}^v \rangle \right| &\leq \left| \langle \prod_{i \in u} \varphi_{l_u^i, n}^i, \varphi_{\mathbf{l}_v, n}^v \rangle_n - \langle \prod_{i \in u} \varphi_{l_u^i, n}^i, \varphi_{\mathbf{l}_v, n}^v \rangle \right| \\ &\quad + \left| \langle \prod_{i \in u} \varphi_{l_u^i, n}^i, \varphi_{\mathbf{l}_v, n}^v \rangle - \langle \prod_{i \in u} \phi_{l_u^i}^i, \phi_{\mathbf{l}_v}^v \rangle \right| \end{aligned}$$

We have,

$$\begin{aligned} \left| \langle \Pi_{i \in u} \varphi_{l_{i,n}}^i, \varphi_{l_{v,n}}^v \rangle - \langle \Pi_{i \in u} \phi_{l_i}^i, \phi_{l_v}^v \rangle \right| &\leq \left| \langle \Pi_{i \in u} \varphi_{l_{i,n}}^i, \varphi_{l_{v,n}}^v - \phi_{l_v}^v \rangle + \langle \Pi_{i \in u} \varphi_{l_{i,n}}^i - \Pi_{i \in u} \phi_{l_i}^i, \phi_{l_v}^v \rangle \right| \\ &\leq \left\| \Pi_{i \in u} \varphi_{l_{i,n}}^i \right\| \left\| \varphi_{l_{v,n}}^v - \phi_{l_v}^v \right\| + \left\| \Pi_{i \in u} \varphi_{l_{i,n}}^i - \Pi_{i \in u} \phi_{l_i}^i \right\| \left\| \phi_{l_v}^v \right\| \end{aligned}$$

(ii) follows from (i) and from (i) of Lemma 2. ■

## Main convergence results

**Lemma 4.** Remind that the true regression function is

$$\eta^R(\mathbf{X}) = \sum_{u \in S^d} \eta_u^R(\mathbf{X}_u), \text{ where } \eta_u^R(\mathbf{X}_u) = \sum_{\substack{\mathbf{l}_u \in \times [1:L_i] \\ i \in u}} \beta_{\mathbf{l}_u}^{u,0} \phi_{\mathbf{l}_u}^u(\mathbf{X}_u) \in H_u^{0,L}$$

Further, let  $\bar{\eta}^R$  be the approximation of  $\eta^R$ ,

$$\bar{\eta}^R(\mathbf{X}) = \sum_{u \in S^d} \bar{\eta}_u^R(\mathbf{X}_u), \text{ where } \bar{\eta}_u^R(\mathbf{X}_u) = \sum_{\substack{\mathbf{l}_u \in \times [1:L_i] \\ i \in u}} \beta_{\mathbf{l}_u}^{u,0} \varphi_{\mathbf{l}_u,n}^u(\mathbf{X}_u) \in G_{u,n}^{0,L}$$

Then  $\|\bar{\eta}_u^R - \eta_u^R\| \xrightarrow{a.s.} 0 \quad \forall u \in S^d$ , and  $\|\bar{\eta}^R - \eta^R\| \xrightarrow{a.s.} 0$ .

*Proof of Lemma 4*

For any  $u \in S^d$ ,  $\|\bar{\eta}_u^R - \eta_u^R\| = \left\| \sum_{\mathbf{l}_u=1}^{L_u} \beta_{\mathbf{l}_u}^{u,0} \varphi_{\mathbf{l}_u,n}^u - \sum_{\mathbf{l}_u=1}^{L_u} \beta_{\mathbf{l}_u}^{u,0} \phi_{\mathbf{l}_u}^u \right\| \leq \sum_{\mathbf{l}_u=1}^{L_u} |\beta_{\mathbf{l}_u}^{u,0}| \cdot \|\varphi_{\mathbf{l}_u,n}^u - \phi_{\mathbf{l}_u}^u\|$ . Let us show that  $\|\varphi_{\mathbf{l}_u,n}^u - \phi_{\mathbf{l}_u}^u\| \xrightarrow{a.s.} 0$  by induction on  $|u|$ . Set

$$(\mathcal{H}_k) \quad \forall u, |u| = k \quad \|\varphi_{\mathbf{l}_u,n}^u - \phi_{\mathbf{l}_u}^u\| \xrightarrow{a.s.} 0, \quad \forall \mathbf{l}_u = (l_i)_{i \in u} \in \times_{i \in u} [1 : L_i]$$

Let us show that  $(\mathcal{H}_k)$  is true for any  $k \leq d$  :

- Let  $u = \{i\}$ , so  $k = 1$ . We used the Gram-Schmidt procedure on  $(\phi_{l_i}^i)_{l_i=1}^{L_i}$  to construct  $(\varphi_{l_i,n}^i)_{l_i=1}^{L_i}$ . Let us show by induction on  $l_i$  that  $\|\varphi_{l_i,n}^i - \phi_{l_i}^i\| \xrightarrow{a.s.} 0, \forall s_i \in [1 : L_i]$ . Set

$$(\mathcal{H}'_{l_i}) \quad \|\varphi_{l_i,n}^i - \phi_{l_i}^i\| \xrightarrow{a.s.} 0.$$

– For  $l_i = 1$ ,  $\varphi_{1,n}^i = \frac{\phi_1^i - \langle \phi_1^i, \varphi_{0,n}^i \rangle_n \varphi_{0,n}^i}{\underbrace{\|\phi_1^i - \langle \phi_1^i, \varphi_{0,n}^i \rangle_n \varphi_{0,n}^i\|_n}_{D_n}}$ , with  $\varphi_{0,n}^i = \phi_0^i = 1$ . So,

$$\begin{aligned} \|\varphi_{1,n}^i - \phi_1^i\| &\leq \frac{1-D_n}{D_n} \|\phi_1^i\| + \frac{|\langle \phi_1^i, \varphi_{0,n}^i \rangle_n|}{D_n} \|\varphi_{0,n}^i\| \\ &= \frac{D_n-1}{D_n} + \frac{|\langle \phi_1^i, 1 \rangle_n|}{D_n} \end{aligned}$$

As  $(\phi_1^i)_{l_i=1}^{L_i}$  is an orthonormal system, we get  $|\langle \phi_1^i, 1 \rangle_n| \xrightarrow{a.s.} \mathbb{E}(\phi_1^i) = 0$  and  $D_n \xrightarrow{a.s.} \|\phi_1^i\| = 1$ .

Therefore,  $\|\varphi_{1,n}^i - \phi_1^i\| \xrightarrow{a.s.} 0$ , and  $(\mathcal{H}'_1)$  is true.

– Let  $l_i \in [1 : L_i]$ . Suppose that  $(\mathcal{H}'_k)$  is true for any  $k \leq l_i$ . Let us show  $(\mathcal{H}'_{l_i+1})$  holds. By construction, we get,

$$\varphi_{l_i+1}^i = \frac{\phi_{l_i+1}^i - \sum_{k=0}^{l_i} \langle \phi_{l_i+1}^i, \varphi_{k,n}^i \rangle_n \cdot \varphi_{k,n}^i}{\underbrace{\left\| \phi_{l_i+1}^i - \sum_{k=0}^{l_i} \langle \phi_{l_i+1}^i, \varphi_{k,n}^i \rangle_n \cdot \varphi_{k,n}^i \right\|_n}_{D_n}}$$

So,

$$\|\varphi_{l_i+1,n}^i - \phi_{l_i+1}^i\| \leq \frac{1 - D_n}{D_n} \|\phi_{l_i+1}^i\| + \frac{1}{D_n} \sum_{k=0}^{l_i} |\langle \phi_{l_i+1}^i, \varphi_{k,n}^i \rangle_n| \cdot \|\varphi_{k,n}^i\|$$

For all  $k \leq l_i$ ,  $\|\varphi_{k,n}^i\| \xrightarrow{\text{a.s.}} \|\phi_k^i\| = 1$  ((i) of Lemma 2), and

$$\begin{aligned} |\langle \phi_{l_i+1}^i, \varphi_{k,n}^i \rangle_n| &= |\langle \phi_{l_i+1}^i, \phi_k^i \rangle_n + \langle \phi_{l_i+1}^i, \varphi_{k,n}^i - \phi_k^i \rangle_n| \\ &\leq |\langle \phi_{l_i+1}^i, \phi_k^i \rangle_n| + |\langle \phi_{l_i+1}^i, \varphi_{k,n}^i - \phi_k^i \rangle_n| \\ &\leq |\langle \phi_{l_i+1}^i, \phi_k^i \rangle_n| + \|\phi_{l_i+1}^i\|_n \cdot \|\varphi_{k,n}^i - \phi_k^i\|_n \end{aligned}$$

By induction,  $\|\varphi_{k,n}^i - \phi_k^i\| \xrightarrow{\text{a.s.}} 0$ . By the law of large numbers,  $\|\varphi_{k,n}^i - \phi_k^i\|_n \xrightarrow{\text{a.s.}} 0$  and  $|\langle \phi_{l_i+1}^i, \phi_k^i \rangle_n| \xrightarrow{\text{a.s.}} |\langle \phi_{l_i+1}^i, \phi_k^i \rangle| = 0$  as the system  $(\phi_{l_i}^i)_{l_i=1}^{L_i}$  is orthonormal. As  $\|\phi_{l_i+1}^i\|_n \xrightarrow{\text{a.s.}} \|\phi_{l_i+1}^i\| = 1$ , we deduce that  $|\langle \phi_{l_i+1}^i, \varphi_{k,n}^i \rangle_n| \xrightarrow{\text{a.s.}} 0$ . Also,

$$D_n \xrightarrow{\text{a.s.}} \|\phi_{l_i+1}^i\| = 1 \Rightarrow \|\varphi_{l_i+1,n}^i - \phi_{l_i+1}^i\| \xrightarrow{\text{a.s.}} 0$$

Hence,  $(\mathcal{H}'_{l_i+1})$  is true.

We conclude that for any  $l_i \in [1 : L_i]$ ,  $\|\phi_{l_i}^i - \varphi_{l_i,n}^i\| \xrightarrow{\text{a.s.}} 0$ , and that  $(\mathcal{H}_1)$  is checked.

- Let  $k \in [1 : d]$ . Suppose now that  $(\mathcal{H}_{|\tilde{u}|})$  is true for any  $1 \leq |\tilde{u}| \leq k-1$ , and any  $l_{\tilde{u}} \in \times_{i \in \tilde{u}} [1 : L_i]$ . Show that  $(\mathcal{H}_k)$  is satisfied. Let  $u$  be such that  $|u| = k$ . We have seen that, for any  $\mathbf{l}_u = (l_u^i)_{i \in u} \in \times_{i \in u} [1 : L_i]$ ,

$$\varphi_{\mathbf{l}_u,n}^u = \prod_{i \in u} \varphi_{l_u^i,n}^i + \sum_{\substack{v \subset u \\ v \neq \emptyset}} \sum_{\substack{\mathbf{l}_v \in \times_{j \in v} [1 : L_j]}} \lambda_{\mathbf{l}_v, \mathbf{l}_u}^{v,n} \varphi_{\mathbf{l}_v,n}^v + C_{\mathbf{l}_u}^{n,u}$$

where  $(C_{\mathbf{l}_u}^{n,u}, (\lambda_{\mathbf{l}_v, \mathbf{l}_u}^{v,n})_{\mathbf{l}_v, v \subset u})$  are computed by the resolution of the following equations

$$\begin{cases} \langle \varphi_{\mathbf{l}_u,n}^u, \varphi_{\mathbf{m},n}^v \rangle_n = 0, & \forall v \subset u, \forall \mathbf{m} = (m^j)_{j \in v} \in \times_{j \in v} [1 : L_j] \\ \langle \varphi_{\mathbf{l}_u,n}^u, 1 \rangle_n = 0 \end{cases} \quad (4)$$

We have,

$$\begin{aligned}
\left\| \varphi_{\mathbf{l}_u, n}^u - \phi_{\mathbf{l}_u}^u \right\| &= \left\| \prod_{i \in u} \varphi_{l_u^i, n}^i - \prod_{i \in u} \phi_{l_u^i}^i + \sum_{\substack{v \subset u \\ v \neq \emptyset}} \sum_{\substack{\mathbf{l}_v \in \times_{j \in v} [1:L_j]}} (\lambda_{\mathbf{l}_v, \mathbf{l}_u}^{v, n} \varphi_{\mathbf{l}_v, n}^v - \lambda_{\mathbf{l}_v, \mathbf{l}_u}^v \phi_{\mathbf{l}_v}^v) \right. \\
&\quad \left. + C_{\mathbf{l}_u}^{n, u} - C_{\mathbf{l}_u}^u \right\| \\
&\leq \left\| \prod_{i \in u} \varphi_{l_u^i, n}^i - \prod_{i \in u} \phi_{l_u^i}^i \right\| + \sum_{\substack{v \subset u \\ v \neq \emptyset}} \sum_{\substack{\mathbf{l}_v \in \times_{j \in v} [1:L_j]}} \left\| \lambda_{\mathbf{l}_v, \mathbf{l}_u}^{v, n} \varphi_{\mathbf{l}_v, n}^v - \lambda_{\mathbf{l}_v, \mathbf{l}_u}^v \phi_{\mathbf{l}_v}^v \right\| \\
&\quad + |C_{\mathbf{l}_u}^{n, u} - C_{\mathbf{l}_u}^u|
\end{aligned} \tag{5}$$

Lemma 3 states that  $\left\| \prod_{i \in u} \varphi_{l_u^i, n}^i - \prod_{i \in u} \phi_{l_u^i}^i \right\| \xrightarrow{a.s.} 0$ ,  $\forall l_u^i \in [1:L_i]$ . Moreover, the resolution of (4) leads to the resolution of a linear system of the type  $A^{u, n} \Lambda^{u, n} = D^{\mathbf{l}_u, n}$ , when removing  $C_{\mathbf{l}_u}^{n, u}$ .  $\Lambda^{u, n}$  is the vector of unknown parameters  $(\lambda_{\mathbf{l}_v, \mathbf{l}_u}^{v, n})_{\substack{\mathbf{l}_v \in \times_{j \in v} [1:L_j], v \subset u}}$ ,  $A^{u, n}$  is the matrix whose block entries are  $(\langle \varphi_{v_1}, \varphi_{v_2} \rangle_n)_{v_1, v_2 \subset u}$ , and  $D^{\mathbf{l}_u, n}$  involves block entries  $(-\langle \otimes_{i \in u} \varphi_{l_u^i}^i, \varphi_{v_i} \rangle_n)_{v_i \subset u}$ . (iii) of Lemma 2 claims that  $A^{u, n}$  tends to the theoretical matrix  $A_\phi^u$  defined in HOGS Procedure of Section 3. Also, by (ii) of Lemma 3,  $D^{\mathbf{l}_u, n} \xrightarrow{a.s.} D^{\mathbf{l}_u}$ . Thus,  $\Lambda^{u, n} \xrightarrow{a.s.} \lambda^u$ ,  $\lambda^u$  being defined in HOGS Procedure of Section 3. We also deduce that  $C_{\mathbf{l}_u}^{n, u} \xrightarrow{a.s.} C_{\mathbf{l}_u}^u$ .

Consequently, we deduce that every piece of the right-hand side of (5) tends to 0, so is  $\left\| \varphi_{\mathbf{l}_u, n}^u - \phi_{\mathbf{l}_u}^u \right\|$ . Hence,  $(\mathcal{H}_k)$  is satisfied.

As a conclusion,  $\left\| \varphi_{\mathbf{l}_u, n}^u - \phi_{\mathbf{l}_u}^u \right\| \xrightarrow{a.s.} 0$ ,  $\forall \mathbf{l}_u \in \times_{i \in u} [1:L_i]$ ,  $\forall u \in S^d$ . Hence, we deduce that  $\left\| \bar{\eta}^R - \eta^R \right\| \xrightarrow{a.s.} 0$ ,  $\forall u \in S^d$ , and

$$\left\| \bar{\eta}^R - \eta^R \right\| \leq \sum_{u \in S^d} \left\| \bar{\eta}_u^R - \eta_u^R \right\| \implies \left\| \bar{\eta}^R - \eta^R \right\| \xrightarrow{a.s.} 0$$

■

**Lemma 5.** Recall that  $\hat{\beta} = \text{Arg min}_{\beta \in \Theta} M_n(\beta)$ . If (H) holds, then

$$\left\| \hat{\beta} - \beta_0 \right\|_2 \xrightarrow{a.s.} 0 \tag{6}$$

Moreover,

$$\left\| \hat{\eta}^R - \eta^R \right\| \xrightarrow{a.s.} 0. \tag{7}$$

*Proof of Lemma 5*

First, we remind the true regression model,

$$Y = \eta^R(\mathbf{X}) + \varepsilon, \quad \text{where } \eta^R(\mathbf{X}) = \sum_{u \in S^d} \sum_{i \in u} \beta_{l_u^i}^{u, 0} \phi_{l_u^i}^u(\mathbf{X}_u) \tag{8}$$

with  $\mathbb{E}(\varepsilon) = 0$ ,  $\mathbb{E}(\varepsilon^2) = \sigma_*^2$ ,  $\mathbb{E}(\varepsilon \cdot \phi_{\mathbf{l}_u}^u(\mathbf{X}_{\mathbf{u}})) = 0$ ,  $\forall \mathbf{l}_u \in \times_{i \in u} [1 : L_i]$ ,  $\forall u \in S^d$ , and  $\boldsymbol{\beta}_0 = (\beta_{\mathbf{l}_u}^{u,0})_{\mathbf{l}_u, u \in S^d}$  the true parameter.

Let  $\tilde{\boldsymbol{\beta}}$  be the solution of the least-squared problem

$$\min_{\boldsymbol{\beta} \in \Theta} \|\mathbb{Y} - \mathbb{X}_\phi \boldsymbol{\beta}\|_n^2 \quad (9)$$

Due to Lemma 1,  $({}^t \mathbb{X}_\phi \mathbb{X}_\phi)^{-1}$  is well defined. Thus,

$$\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \left( \frac{{}^t \mathbb{X}_\phi \mathbb{X}_\phi}{n} \right)^{-1} \cdot \frac{{}^t \mathbb{X}_\phi \cdot \varepsilon}{n}$$

By the law of large numbers,  $(\frac{{}^t \mathbb{X}_\phi \cdot \varepsilon}{n})_u \xrightarrow{a.s.} \mathbb{E}(\varepsilon \cdot \phi_{\mathbf{l}_u}^u(\mathbf{X}_{\mathbf{u}})) = 0$ ,  $\forall u \in S^d$ . Moreover,  $\frac{{}^t \mathbb{X}_\phi \mathbb{X}_\phi}{n} \xrightarrow{a.s.} A_\phi$ , where  $A_\phi$  is defined in the new settings. Thus,  $\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \xrightarrow{a.s.} 0$ .

Now, we are going to use the fact that  $\|\varphi_{\mathbf{l}_{u,n}}^u - \phi_{\mathbf{l}_u}^u\| \xrightarrow{a.s.} 0$ , (see the Proof of Lemma 4) to show that  $\|\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\|_2 \xrightarrow{a.s.} 0$ .

As  $\hat{\boldsymbol{\beta}}$  is the solution of the ordinary least squares problem, we get  $\hat{\boldsymbol{\beta}} = ({}^t \mathbb{X}_\varphi \mathbb{X}_\varphi)^{-1} {}^t \mathbb{X}_\varphi \mathbb{Y}$  because, as seen later  ${}^t \mathbb{X}_\varphi \mathbb{X}_\varphi \xrightarrow{a.s.} A_\phi$ , that is invertible.

We define the usual matrix norm  $\|\cdot\|_2$  as  $\|A\|_2 := \sup_{\|x\|_2=1} \|Ax\|_2$ . The Frobenius matrix norm is defined as  $\|A\|_F := \sqrt{\text{Trace}(A^t A)}$ , and  $\|A\|_2 \leq \|A\|_F$  [2]. We use this inequality to get

$$\begin{aligned} \|\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\|_2 &= \|(({}^t \mathbb{X}_\varphi \mathbb{X}_\varphi)^{-1} {}^t \mathbb{X}_\varphi - ({}^t \mathbb{X}_\phi \mathbb{X}_\phi)^{-1} {}^t \mathbb{X}_\phi) \mathbb{Y}\|_2 \\ &\leq \|({}^t \mathbb{X}_\varphi \mathbb{X}_\varphi)^{-1} {}^t \mathbb{X}_\varphi - ({}^t \mathbb{X}_\phi \mathbb{X}_\phi)^{-1} {}^t \mathbb{X}_\phi\|_2 \cdot \|\mathbb{Y}\|_2 \\ &= \sqrt{n} \|({}^t \mathbb{X}_\varphi \mathbb{X}_\varphi)^{-1} {}^t \mathbb{X}_\varphi - ({}^t \mathbb{X}_\phi \mathbb{X}_\phi)^{-1} {}^t \mathbb{X}_\phi\|_F \cdot \left\| \frac{\mathbb{Y}}{\sqrt{n}} \right\|_2 \end{aligned}$$

First, by (8),

$$\left\| \frac{\mathbb{Y}}{\sqrt{n}} \right\|_2 = \left\| \frac{\mathbb{X}_\phi \boldsymbol{\beta}_0 + \varepsilon}{\sqrt{n}} \right\|_2 \leq \left\| \frac{\mathbb{X}_\phi \boldsymbol{\beta}_0}{\sqrt{n}} \right\|_2 + \left\| \frac{\varepsilon}{\sqrt{n}} \right\|_2$$

and  $\left\| \frac{\mathbb{X}_\phi \boldsymbol{\beta}_0}{\sqrt{n}} \right\|_2 = \sqrt{\frac{1}{n} \sum_{s=1}^n \left[ \sum_{u \in S^d} \sum_{\mathbf{l}_u} \beta_{\mathbf{l}_u}^{u,0} \phi_{\mathbf{l}_u}^u(\mathbf{x}_{\mathbf{u}}^s) \right]^2} \xrightarrow{a.s.} ({}^t \boldsymbol{\beta}_0 A_\phi \boldsymbol{\beta}_0)^{1/2}$ . Also,  $\left\| \frac{\varepsilon}{\sqrt{n}} \right\|_2 \xrightarrow{a.s.} \sqrt{\mathbb{E}(\varepsilon^2)} = \sigma_*$ .

Hence,  $\left\| \frac{\mathbb{Y}}{\sqrt{n}} \right\|_2 \leq ({}^t \boldsymbol{\beta}_0 A_\phi \boldsymbol{\beta}_0)^{1/2} + \sigma_* < \infty$ .

Now, let us consider  $\sqrt{n} \|({}^t \mathbb{X}_\varphi \mathbb{X}_\varphi)^{-1} {}^t \mathbb{X}_\varphi - ({}^t \mathbb{X}_\phi \mathbb{X}_\phi)^{-1} {}^t \mathbb{X}_\phi\|_F$ . After computation, we get

$$\begin{aligned} n \|({}^t \mathbb{X}_\varphi \mathbb{X}_\varphi)^{-1} {}^t \mathbb{X}_\varphi - ({}^t \mathbb{X}_\phi \mathbb{X}_\phi)^{-1} {}^t \mathbb{X}_\phi\|_F^2 &= \text{Trace}[(\frac{{}^t \mathbb{X}_\phi \mathbb{X}_\phi}{n})^{-1}] + \text{Trace}[(\frac{{}^t \mathbb{X}_\varphi \mathbb{X}_\varphi}{n})^{-1}] \\ &\quad - 2 \text{Trace}[(\frac{{}^t \mathbb{X}_\phi \mathbb{X}_\phi}{n})^{-1} \cdot \frac{{}^t \mathbb{X}_\phi \mathbb{X}_\varphi}{n} \cdot (\frac{{}^t \mathbb{X}_\varphi \mathbb{X}_\varphi}{n})^{-1}] \end{aligned}$$

- $\frac{{}^t \mathbb{X}_\phi \mathbb{X}_\phi}{n} \xrightarrow{a.s.} A_\phi$
- $\frac{{}^t \mathbb{X}_\varphi \mathbb{X}_\varphi}{n} \xrightarrow{a.s.} A_\phi$ , by (iii) of Lemma 2



- $\frac{{}^t\mathbb{X}_\phi\mathbb{X}_\varphi}{n} \xrightarrow{a.s.} A_\phi$  as,  $\langle \phi_{\mathbf{l}_u}^u, \varphi_{\mathbf{l}_v, n}^v \rangle \xrightarrow{a.s.} \langle \phi_{\mathbf{l}_u}^u, \phi_{\mathbf{l}_v}^v \rangle$ ,  $\forall u, v \in S^d$ ,  $\forall \mathbf{l}_u \in \times_{i \in u} [1 : L_i]$ ,  $\mathbf{l}_v \in \times_{j \in v} [1 : L_j]$  by Lemma 2.

Under (H), using the result of Lemma 1,  $A_\phi$  is invertible. Then,

$$n \left\| \left( {}^t\mathbb{X}_\phi\mathbb{X}_\phi \right)^{-1} {}^t\mathbb{X}_\phi - \left( {}^t\mathbb{X}_\varphi\mathbb{X}_\varphi \right)^{-1} {}^t\mathbb{X}_\varphi \right\|_F^2 \xrightarrow{a.s.} \text{Trace}(A_\phi^{-1}) + \text{Trace}(A_\phi^{-1}) - 2\text{Trace}(A_\phi^{-1}) = 0$$

Thus,  $\left\| \hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} \right\|_2 \xrightarrow{a.s.} 0$ . We conclude that

$$\left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|_2 \leq \left\| \hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} \right\|_2 + \left\| \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|_2 \xrightarrow{a.s.} 0$$

At last,  $\left\| \hat{\boldsymbol{\eta}}^R - \bar{\boldsymbol{\eta}}^R \right\| \leq \left\| \mathbb{X}_\varphi \right\|_2 \left\| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|_2 \xrightarrow{a.s.} 0$ . Finally, as a consequence of Lemma 4,

$$\left\| \hat{\boldsymbol{\eta}}^R - \boldsymbol{\eta}^R \right\| \leq \left\| \hat{\boldsymbol{\eta}}^R - \bar{\boldsymbol{\eta}}^R \right\| + \left\| \bar{\boldsymbol{\eta}}^R - \boldsymbol{\eta}^R \right\| \xrightarrow{a.s.} 0.$$

■

**Notes:** When adding a penalization  $\frac{\lambda_n}{n} J(\boldsymbol{\beta})$ ,  $\lambda_n \geq 0$ , to the function  $M_n(\boldsymbol{\beta})$  defined in New settings, the convergence of  $\hat{\boldsymbol{\eta}}^R$  to  $\boldsymbol{\eta}^R$  is not straightforward. If  $J(\boldsymbol{\beta}) = \sum_{u \in S^d} \sum_{\mathbf{l}_u} (\beta_{\mathbf{l}_u}^u)^2$ , the result is immediate. However, the  $\ell_0$  or the  $\ell_1$  penalization is of our interest. In this case, some strong assumptions of uniform convergence on the penalized function  $M_n(\boldsymbol{\beta})$  would be necessary to get the consistency of  $\hat{\boldsymbol{\beta}}$  (to mimic the convergence on the M-estimators [3]).

## References

- [1] HOLMES, R.B.(1988) : *On random correlation matrices*, Massachusetts Institute of Technology (803).
- [2] GOLUB, G.H., VAN LOAN, C.F.(1996) : *Matrix computations*, The Johns Hopkins University Press, Baltimore.
- [3] VAN DER VAART, A.W.(1998) : *Asymptotic Statistics*, Cambridge University Press, Cambridge.