Supplementary Material for "Estimation of Functional Sparsity in Nonparametric Varying Coefficient Models for Longitudinal Data Analysis"

Catherine Y. Tu[1], Juhyun Park[2], and Haonan Wang[1]

[1]Colorado State University, [2]Lancaster University

This supplementary material includes technical assumptions and proofs of the theoretical properties of our proposed method. The technical assumptions are given in Section A. The proofs of Theorems 1 and 2 are given in Section B and Section C respectively.

Appendix: Technical assumptions and proofs

A Technical assumptions

The following assumptions are made for Theorems 1 and 2 in Section 3. These are standard assumptions used to establish asymptotic properties of nonparametric estimation procedures for varying coefficient models; see Huang et al. [2] and Wang et al. [4] for more details.

(A1) The response and covariate processes $\{y_k(t), \boldsymbol{x}_k(t),\ k = 1, \ldots, n\}$ are iid as $\{y(t), \boldsymbol{x}(t)\}$. And the observation time points, t_{kl}, $l = 1, \ldots, n_k$, $k = 1, \ldots, n$, are iid from an unknown density, $f(t)$, on $[0, M]$, where $f(t)$ is uniformly bounded away from zero and infinity. That is, $0 < h_1 \leq f(t) \leq h_2 < \infty$ for some positive constants h_1 and h_2. Moreover, the observation time points are independent of the response and covariate processes $\{y_k(t), \boldsymbol{x}_k(t), k = 1, \ldots, n\}$.

(A2) The eigenvalues of the matrix $E[\boldsymbol{x}(t)\boldsymbol{x}^T(t)]$ are uniformly bounded away from zero and infinity for $t \in [0, M]$, that is, there exist positive constants M_1 and M_2 to be the lower and upper bound of the eigenvalues for all $t \in [0, M]$.

(A3) There exists a positive constant M_3 such that $|x_i(t)| \leq M_3$ for $t \in [0, M]$ and $i = 1, \ldots, p$.

(A4) There exists a positive constant M_4 such that $E\{\epsilon^2(t)\} \leq M_4$ for all $t \in [0, M]$

(A5) $\text{limsup}_n(\text{max}_i K_i / \text{min}_i K_i) < \infty$.

(A6) The process $\epsilon(t)$ can be decomposed as the sum of two independent stochastic processes, $\epsilon^{(1)}$ and $\epsilon^{(2)}$, where $\epsilon^{(1)}$ is an arbitrary mean zero process, and $\epsilon^{(2)}$ is a process of measurement errors that are independent at different time points and have mean zero and finite constant variance σ^2.

B Proof of Theorem 1

The following lemma from Lemma A.3 of Huang et al. [2] will be used in the proof.

Lemma 1. *Suppose that* $\lim_{n\to\infty} K_n \log K_n / n = 0$. *There are positive constants* C_1 *and* C_2 *such that, except on an event whose probability tends to zero, all eigenvalues of* $n^{-1} K_n \boldsymbol{U}^T\boldsymbol{U}$ *fall between* C_1 *and* C_2, *and consequently* $\boldsymbol{U}^T\boldsymbol{U}$ *is invertible.*

Proof of Theorem 1. Note that

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 \leq \|\widetilde{\boldsymbol{\beta}}^0 - \boldsymbol{\beta}\|_2 + \|\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}^0\|_2.$$

By B-spline property, $\|\beta_i - \widetilde{\beta}_i\|_2 = O_p(K_n^{-2})$ where $\widetilde{\beta}_i$ is an approximation in B-spline space as defined in (2.3). It can be shown that the same rate holds true if $\widetilde{\beta}_i$ is replaced by its sparse approximation of $\widetilde{\beta}_i^0$ (see Lemma 1 in Wang and Kai [3]). Thus, $\|\widetilde{\boldsymbol{\beta}}^0 - \boldsymbol{\beta}\|_2 = O_p(K_n^{-2})$.

For the second term, by (A5) and B-spline property, we have $\|\widetilde{\beta}_i\|_2^2 \leq D_i\|\boldsymbol{\alpha}_i\|_2^2/K_n$ for some positive constant D_i, $i = 1, \ldots, p$ [1, 2]. Denote $D_* = \max_i D_i$, and we have

$$\|\widetilde{\boldsymbol{\beta}}^0 - \widehat{\boldsymbol{\beta}}\|_2^2 = \sum_{i=1}^{p} \|\widetilde{\beta}_i^0 - \widehat{\beta}_i\|_2^2 \leq \sum_{i=1}^{p} \frac{D_i}{K_n}\|\widetilde{\boldsymbol{\alpha}}_i^0 - \widehat{\boldsymbol{\alpha}}_i\|_2^2 \leq D_* \frac{\|\widehat{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}^0\|_2^2}{K_n}$$

Therefore,

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2^2 = O_p\Big(K_n^{-4} + \frac{\|\widehat{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}^0\|_2^2}{K_n}\Big).$$

Below we concentrate on the term $\|\widehat{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}^0\|_2$ and in particular we show that $\|\widehat{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}}^0\|_2^2 = O_p(n^{-1}K_n^2)$.

By the minimality of $\widehat{\boldsymbol{\alpha}}$, we have $\mathrm{pl}(\widehat{\boldsymbol{\alpha}}) \leq \mathrm{pl}(\widetilde{\boldsymbol{\alpha}}^0)$; that is,

$$\|\boldsymbol{y} - \boldsymbol{U}\widehat{\boldsymbol{\alpha}}\|_2^2 - \|\boldsymbol{y} - \boldsymbol{U}\widetilde{\boldsymbol{\alpha}}^0\|_2^2 \leq \lambda_n \sum_{i=1}^{p}\sum_{g=1}^{G_i} \|\widetilde{\boldsymbol{\alpha}}_{A_{ig}}^0\|_1^\gamma - \lambda_n \sum_{i=1}^{p}\sum_{g=1}^{G_i} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma. \tag{1}$$

Note that, the right hand side of (1) can be decomposed into two terms, $\lambda_n \sum_{i=1}^p \sum_{g\in\mathcal{A}_{i1}} \|\widetilde{\boldsymbol{\alpha}}_{A_{ig}}^0\|_1^\gamma - \lambda_n \sum_{i=1}^p \sum_{g\in\mathcal{A}_{i1}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma$ and $\lambda_n \sum_{i=1}^p \sum_{g\in\mathcal{A}_{i2}} \|\widetilde{\boldsymbol{\alpha}}_{A_{ig}}^0\|_1^\gamma - \lambda_n \sum_{i=1}^p \sum_{g\in\mathcal{A}_{i2}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma$. For the first term, applying the inequality $|b^\gamma - a^\gamma| \leq 2|b-a|b^{\gamma-1}$, for $a, b \geq 0$, and Cauchy-Schwarz inequality yields that

$$
\begin{aligned}
&\left|\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1^{\gamma}-\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^{\gamma}\right| \\
\leq\ & 2\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\left|\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1-\|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1\right|\cdot\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1^{\gamma-1} \\
\leq\ & 2\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}-\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1\cdot\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1^{\gamma-1} \\
\leq\ & 2(d+1)^{1/2}\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1^{\gamma-1}\cdot\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}-\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_2 \\
\leq\ & 2(d+1)^{1/2}\left(\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1^{2(\gamma-1)}\right)^{1/2}\left(\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}-\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_2^2\right)^{1/2}.
\end{aligned}
$$

For the second term, note that $\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1=0$ for $g\in\mathcal{A}_{i2}$. Thus, the second term is less than or equal to zero. Combining above results and (1), we have

$$
\begin{aligned}
&\|\boldsymbol{y}-\boldsymbol{U}\widehat{\boldsymbol{\alpha}}\|_2^2-\|\boldsymbol{y}-\boldsymbol{U}\widetilde{\boldsymbol{\alpha}}^0\|_2^2 \\
&\leq\lambda_n\left|\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1^{\gamma}-\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^{\gamma}\right|+\lambda_n\left(\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i2}}\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1^{\gamma}-\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i2}}\|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_2^{\gamma}\right) \\
&\leq\lambda_n\left|\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1^{\gamma}-\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^{\gamma}\right| \\
&\leq 2\lambda_n\phi_n\left(\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}-\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_2^2\right)^{1/2}
\end{aligned}
$$

It follows that

$$
\|\boldsymbol{y}-\boldsymbol{U}\widehat{\boldsymbol{\alpha}}\|_2^2-\|\boldsymbol{y}-\boldsymbol{U}\widetilde{\boldsymbol{\alpha}}^0\|_2^2\leq 2\lambda_n\phi_n(d+1)^{1/2}\|\widetilde{\boldsymbol{\alpha}}^0-\widehat{\boldsymbol{\alpha}}\|_2, \tag{2}
$$

where $\phi_n=(d+1)^{1/2}\left(\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}}\|\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}}\|_1^{2(\gamma-1)}\right)^{1/2}$.

On the other hand, straightforward calculation gives that

$$\begin{aligned}\|\boldsymbol{y}-\boldsymbol{U}\widehat{\boldsymbol{\alpha}}\|_2^2-\|\boldsymbol{y}-\boldsymbol{U}\widetilde{\boldsymbol{\alpha}}^0\|_2^2 &= (\boldsymbol{U}\widehat{\boldsymbol{\alpha}})^T\boldsymbol{U}\widehat{\boldsymbol{\alpha}}-(\boldsymbol{U}\widetilde{\boldsymbol{\alpha}}^0)^T\boldsymbol{U}\widetilde{\boldsymbol{\alpha}}^0-2\boldsymbol{y}^T\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0)\\ &= (\boldsymbol{U}\widehat{\boldsymbol{\alpha}}+\boldsymbol{U}\widetilde{\boldsymbol{\alpha}}^0-2\boldsymbol{U}\widetilde{\boldsymbol{\alpha}}^0-2\boldsymbol{\epsilon}_*)^T\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0)\\ &= \|\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0)\|_2^2-2\boldsymbol{\epsilon}_*^T\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0)\\ &\geq \|\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0)\|_2^2-2\left|\boldsymbol{\epsilon}_*^T\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0)\right|,\end{aligned}$$

where $\boldsymbol{\epsilon}_*=\boldsymbol{\epsilon}-\boldsymbol{e}$, $\boldsymbol{e}=(\boldsymbol{e}_1^T,\ldots,\boldsymbol{e}_n^T)^T$ with

$$\boldsymbol{e}_k=(\mathcal{U}_k(t_{k1})\widetilde{\boldsymbol{\alpha}}^0-\boldsymbol{x}_k(t_{k1})^T\boldsymbol{\beta}(t_{k1}),\ldots,\mathcal{U}_k(t_{kn_k})\widetilde{\boldsymbol{\alpha}}^0-\boldsymbol{x}_k(t_{kn_k})^T\boldsymbol{\beta}(t_{kn_k}))^T\,.$$

Let $\delta_n=\|\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0\|_2$, then by Lemma 1, $\|\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0)\|_2^2\geq C_1nK_n^{-1}\delta_n^2$ with probability approaching 1. In addition, applying Cauchy-Schwarz inequality yields that $(\boldsymbol{\epsilon}_*^T\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0))^2\leq\delta_n^2(\boldsymbol{\epsilon}_*^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\epsilon}_*)$. Further, $E(\boldsymbol{\epsilon}_*^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\epsilon}_*)=E(\boldsymbol{\epsilon}^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\epsilon})+E(\boldsymbol{e}^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{e})$. As a consequence of Lemma A.3 of [4], we have $E(\boldsymbol{\epsilon}^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\epsilon})=O(n)$ with n_k uniformly bounded. Similarly, we have $E(\boldsymbol{e}^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{e})=O(n)$ since $E(e(t_{kl})e(t_{kl'}))\leq C\|\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}^0\|_\infty^2$ for some constant C and $\|\boldsymbol{\beta}-\widetilde{\boldsymbol{\beta}}^0\|_\infty$ is bounded by $O(K_n^{-2})$. Therefore, $E(\boldsymbol{\epsilon}_*^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{\epsilon}_*)=O(n)$. Thus, we have

$$\|\boldsymbol{y}-\boldsymbol{U}\widehat{\boldsymbol{\alpha}}\|_2^2-\|\boldsymbol{y}-\boldsymbol{U}\widetilde{\boldsymbol{\alpha}}^0\|_2^2\geq C_1nK_n{}^{-1}\delta_n^2-\delta_nO_p(n^{1/2}). \tag{3}$$

Combining (2) and (3), we have

$$\frac{nC_1}{K_n}\delta_n^2-\delta_nO_p(n^{1/2})\leq 2\lambda_n\phi_n(d+1)^{1/2}\delta_n,$$

and by (S1) we have $\|\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0\|_2^2=O_p(n^{-1}K_n^2)$.

□

C Proof of Theorem 2

Proof. First, for any i, define $\widehat{\alpha}^*_{ij}$ in the following way. Let $\widehat{\alpha}^*_{ij} = 0$ if $\{j-d, \ldots, j\} \cap \mathcal{A}_{i2} \neq \emptyset$, otherwise, $\widehat{\alpha}^*_{ij} = \widehat{\alpha}_{ij}$. Note that $\widehat{\boldsymbol{\alpha}}^*_{A_{ig}} = \mathbf{0}$ for $g \in \mathcal{A}_{i2}$.

By Karush-Kuhn-Tucker conditions, for $\widehat{\alpha}_{ij} \neq 0$ we have

$$2(\boldsymbol{y} - \boldsymbol{U}\widehat{\boldsymbol{\alpha}})^T U_{(ij)} = \sum_{g=j-d}^{j} \gamma\lambda_n \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^{\gamma-1} \mathrm{sgn}(\widehat{\alpha}_{ij}),$$

where $U_{(ij)}$ is the column of $\boldsymbol{U}$ corresponding to $\widehat{\alpha}_{ij}$. Multiplying both sides by $(\widehat{\alpha}_{ij} - \widehat{\alpha}^*_{ij})$ yields

$$\begin{aligned}
2(\boldsymbol{y} - \boldsymbol{U}\widehat{\boldsymbol{\alpha}})^T \boldsymbol{U}(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*) &= \sum_{i,j} \sum_{g=j-d}^{j} \gamma\lambda_n \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^{\gamma-1} \mathrm{sgn}(\widehat{\alpha}_{ij})(\widehat{\alpha}_{ij} - \widehat{\alpha}^*_{ij}) \\
&= \gamma\lambda_n \sum_{i,j} \sum_{g \in \mathcal{A}_{i2} \cap \{j-d,\ldots,j\}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^{\gamma-1} |\widehat{\alpha}_{ij}| \cdot \\
&= \gamma\lambda_n \sum_{i=1}^{p} \sum_{g=1}^{G_i} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^{\gamma-1} (\|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1 - \|\widehat{\boldsymbol{\alpha}}^*_{A_{ig}}\|_1).
\end{aligned}$$

Note that, $(\widehat{\alpha}_{ij} - \widehat{\alpha}^*_{ij})\mathrm{sgn}(\widehat{\alpha}_{ij}) = |\widehat{\alpha}_{ij}|$ if $\{j-d, \ldots, j\} \cap \mathcal{A}_{i2} \neq \emptyset$.

Since $\gamma b^{\gamma-1}(b-a) \leq b^\gamma - a^\gamma$ for $0 \leq a \leq b$, we have, for $g \in \mathcal{A}_{i1}$,

$$\gamma \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^{\gamma-1} (\|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1 - \|\widehat{\boldsymbol{\alpha}}^*_{A_{ig}}\|_1) \leq \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma - \|\widehat{\boldsymbol{\alpha}}^*_{A_{ig}}\|_1^\gamma.$$

Consequently, we have

$$2\left|(\boldsymbol{y} - \boldsymbol{U}\widehat{\boldsymbol{\alpha}})^T \boldsymbol{U}(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*)\right| \leq \lambda_n \sum_{i=1}^{p} \sum_{g \in \mathcal{A}_{i1}} (\|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma - \|\widehat{\boldsymbol{\alpha}}^*_{A_{ig}}\|_1^\gamma) + \gamma\lambda_n \sum_{i=1}^{p} \sum_{g \in \mathcal{A}_{i2}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma. \quad (4)$$

By the minimality of $\widehat{\boldsymbol{\alpha}}$, we have

$$\lambda_n \sum_{i=1}^{p} \sum_{g=1}^{G_i} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma - \lambda_n \sum_{i=1}^{p} \sum_{g=1}^{G_i} \|\widehat{\boldsymbol{\alpha}}^*_{A_{ig}}\|_1^\gamma \leq \|\boldsymbol{y} - \boldsymbol{U}\widehat{\boldsymbol{\alpha}}^*\|_2^2 - \|\boldsymbol{y} - \boldsymbol{U}\widehat{\boldsymbol{\alpha}}\|_2^2.$$

Note that $\|\widehat{\boldsymbol{\alpha}}^*_{A_{ig}}\|_1 = 0$ for $g \in \mathcal{A}_{i2}$. Thus, we have $\sum_{i=1}^{p} \sum_{g=1}^{G_i} \|\widehat{\boldsymbol{\alpha}}^*_{A_{ig}}\|_1^\gamma = \sum_{i=1}^{p} \sum_{g \in \mathcal{A}_{i1}} \|\widehat{\boldsymbol{\alpha}}^*_{A_{ig}}\|_1^\gamma$, and

$$
\begin{aligned}
&2\left|(\boldsymbol{y}-\boldsymbol{U}\widehat{\boldsymbol{\alpha}})^T\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widehat{\boldsymbol{\alpha}}^*)\right| + (1-\gamma)\lambda_n \sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i2}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma \\
&\le \lambda_n \sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}} (\|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma - \|\widehat{\boldsymbol{\alpha}}^*_{A_{ig}}\|_1^\gamma) + \lambda_n \sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i2}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma \\
&= \lambda_n \sum_{i=1}^{p}\sum_{g=1}^{G_i} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma - \lambda_n \sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i1}} \|\widehat{\boldsymbol{\alpha}}^*_{A_{ig}}\|_1^\gamma \\
&\le \|\boldsymbol{y}-\boldsymbol{U}\widehat{\boldsymbol{\alpha}}^*\|_2^2 - \|\boldsymbol{y}-\boldsymbol{U}\widehat{\boldsymbol{\alpha}}\|_2^2 \\
&= \|\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}^*-\widehat{\boldsymbol{\alpha}})\|_2^2 + 2(\boldsymbol{y}-\boldsymbol{U}\widehat{\boldsymbol{\alpha}})^T\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widehat{\boldsymbol{\alpha}}^*).
\end{aligned}
$$

By Lemma 1 we have

$$
\begin{aligned}
&(1-\gamma)\lambda_n \sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i2}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma \\
&\le \|\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}^*-\widehat{\boldsymbol{\alpha}})\|_2^2 + 2(\boldsymbol{y}-\boldsymbol{U}\widehat{\boldsymbol{\alpha}})^T\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widehat{\boldsymbol{\alpha}}^*) - 2\left|(\boldsymbol{y}-\boldsymbol{U}\widehat{\boldsymbol{\alpha}})^T\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}-\widehat{\boldsymbol{\alpha}}^*)\right| \\
&\le \|\boldsymbol{U}(\widehat{\boldsymbol{\alpha}}^*-\widehat{\boldsymbol{\alpha}})\|_2^2 \\
&\le \frac{nC_2}{K_n}\|\widehat{\boldsymbol{\alpha}}^*-\widehat{\boldsymbol{\alpha}}\|_2^2
\end{aligned}
$$

Note that $\widetilde{\boldsymbol{\alpha}}^0_{A_{ig}} = \mathbf{0}$ for $g \in \mathcal{A}_{i2}$. Thus, we have $\|\widehat{\boldsymbol{\alpha}}^*-\widehat{\boldsymbol{\alpha}}\|_2^2 \le \|\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0\|_2^2$, and

$$
(1-\gamma)\lambda_n \sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i2}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma \le \frac{nC_2}{K_n}\|\widehat{\boldsymbol{\alpha}}-\widetilde{\boldsymbol{\alpha}}^0\|_2^2 = O_p(K_n)
$$

Since

$$
\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i2}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma \ge \left(\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i2}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1\right)^\gamma \ge \|\widehat{\boldsymbol{\alpha}}^*-\widehat{\boldsymbol{\alpha}}\|_2^\gamma,
$$

then if $\|\widehat{\boldsymbol{\alpha}}^*-\widehat{\boldsymbol{\alpha}}\|_2 > 0$, we have

$$
\begin{aligned}
(1-\gamma)\lambda_n &\le \frac{nC_2}{K_n}\|\widehat{\boldsymbol{\alpha}}^*-\widehat{\boldsymbol{\alpha}}\|_2^2 \left\{\sum_{i=1}^{p}\sum_{g\in\mathcal{A}_{i2}} \|\widehat{\boldsymbol{\alpha}}_{A_{ig}}\|_1^\gamma\right\}^{-1} \\
&\le \frac{nC_2}{K_n}\|\widehat{\boldsymbol{\alpha}}^*-\widehat{\boldsymbol{\alpha}}\|_2^{2-\gamma} \\
&\le O_p(n^{\gamma/2}K_n^{1-\gamma}),
\end{aligned}
$$

and thus

$$\Pr\left\{\|\widehat{\boldsymbol{\alpha}}^* - \widehat{\boldsymbol{\alpha}}\|_2^2 > 0\right\} \le \Pr\left\{\frac{\lambda_n}{n^{\gamma/2}K_n^{1-\gamma}} \le O_p(1)\right\}.$$

By assumption (S2), the right hand side converges to zero as n goes to infinity, which implies that $(\widehat{\boldsymbol{\alpha}}_{A_{ig}} : g \in \mathcal{A}_{i2}) = \mathbf{0}$ with probability approaching to one. □

References

[1] de Boor, C. (2001). *A Practical Guide to Splines*. Springer.

[2] Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14:763–788.

[3] Wang, H. and Kai, B. (2015). Functional sparsity: Global versus local. *Statistica Sinica*, 25:1337–1354.

[4] Wang, L., Li, H., and Huang, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association*, 103:1556–1569.