IDENTIFICATION AND ESTIMATION OF GENERAL NONLINEAR STRUCTURED LATENT FACTOR MODEL FOR FUNCTIONAL DATA

Xiaorui Wang[1,2], Yimang Zhang[2] and Jian Qing Shi[2]

[1] *Nanjing University of Information Science and Technology*

[2] *Southern University of Science and Technology*

Supplementary Material

The online Supplementary Materials include the detailed proof of the three theorems.

We first provides the following suitable regularity conditions necessary for proving consistency.

A1 Let Π be a prior distribution for $f_j(\boldsymbol{x}^{(j)}(t_i))$. Let $M_n = O(n^\alpha)$ for some $\frac{1}{2} < \alpha < 1$ and define $\Theta_n = \{f_j : \|f_j\|_\infty < M_n, \|\frac{\partial}{\partial x_k(t_i)} f_j(\boldsymbol{x}^{(j)}(t_i))\|_\infty < M_n, k = 1, \ldots, K\}$, $j = 1, \ldots, J$. Assume that Π assigns an exponentially small probability to the set Θ_n.

A2 Assume that $f_j^*(\cdot)$ as a function of the latent factors $\boldsymbol{x}^{(j)}(t_i)$ is continuously differentiable on a compact set.

Xiaorui Wang and Yimang Zhang contributed equally.
Corresponding author, email: shijq@sustech.edu.cn

S1 Proof of Theorem 1

Proof of Theorem 1. Following Chen et al. (2020), we use a proof by contradiction to complete the proof. If $\{k\} = \bigcap_{k\in\Delta,\ |R_Q(\Delta)=\mathcal{O}(J)|}\Delta$, is not satisfied, theni there are two cases: (i) $\{k, k'\} \subset \bigcap_{k\in\Delta,\ |R_Q(\Delta)=\mathcal{O}(J)|}\Delta$ for some $k' \neq k$, and (ii) $\emptyset = \bigcap_{k\in\Delta,\ |R_Q(\Delta)=\mathcal{O}(J)|}\Delta$. We show below that, in both of the above cases, one can construct $\tilde{\boldsymbol{X}}, \boldsymbol{X}' \in \mathcal{S}$ such that $P_{\tilde{\boldsymbol{X}}} = P_{\boldsymbol{X}'}$, while $\sin_+ \angle\left(\tilde{\boldsymbol{X}}_{[k]}, \boldsymbol{X}'_{[k]}\right) > 0$.

Case (i) Without loss of generality, we assume $K = 3$, $k = 1$, $k' = 2$. Then we establish that $\widetilde{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{I}_3 \\ \boldsymbol{1}_3' \\ \vdots \end{pmatrix}$ and $\boldsymbol{X}' = \widetilde{\boldsymbol{X}} - \widetilde{\boldsymbol{X}}_{[k']}$. It is easy to verify that $\|\boldsymbol{x}(t_i)\| \leq C$, and the columns of $\boldsymbol{X}$ are linearly independent, then $\widetilde{\boldsymbol{X}}, \boldsymbol{X}' \in \mathcal{S}$. Since $\boldsymbol{Y}_j \mid \boldsymbol{X}, f_j \sim N\left(f_j(\boldsymbol{x}^{(j)}(t_i)), \sigma^2\right)$, when given the structured index matrix $\boldsymbol{Q}$ and link function f_j, this can lead to $P_{\tilde{\boldsymbol{X}}} = P_{\boldsymbol{X}'}$, where $P_{\boldsymbol{X}}$ denote the probability distribution of $\{Y_j(t_i), i \in \mathbb{Z}_+, j \in \{1, \ldots, J\}\}$ given the parameters $\boldsymbol{X}, \boldsymbol{Q}, f_j$. The key to the validity of the previous step is our assumption that the column vectors of $\boldsymbol{Q}$ are linearly independent. The following step requires verifying that $\sin_+ \angle\left(\tilde{\boldsymbol{X}}_{[k]}, \boldsymbol{X}'_{[k]}\right) \neq 0$. Unlike Chen et al. (2020), here we consider the dependency of $\boldsymbol{x}_k$ across different times t_i. Fortunately, this only affects the norms of $\tilde{\boldsymbol{X}}_{[k]}$ and $\boldsymbol{X}'_{[k]}$ without influencing the inner product of those. Therefore, as discussed in Chen et al. (2020), we can see that $\sin_+ \angle \left(\tilde{\boldsymbol{X}}_{[k]}, \boldsymbol{X}'_{[k]}\right) \neq 0$. This contradicts the definition of structured identifiability.

Case (ii) Similar to Chen et al. (2020), we can construct $\tilde{\boldsymbol{X}}, \boldsymbol{X}' \in \mathcal{S}$ such that $P_{\tilde{\boldsymbol{X}}} = P_{\boldsymbol{X}'}$, while $\sin_+ \angle \left(\tilde{\boldsymbol{X}}_{[k]}, \boldsymbol{X}'_{[k]}\right) > 0$. The proof is complete. □

S2 Proof of Theorem 2

Proof of Theorem 2. Note that f_j^* is the true link function, σ_*^2 is the true variance of noise, $\boldsymbol{\Theta}_{xk}^*$ are the true values of model parameters, and $\boldsymbol{\Theta}_{fj}^*$ are the true values of hyper-parameters. From Equation (2.17) in the main article, it follows that we only need to prove the consistency of $\widehat{\boldsymbol{x}}(t_i)$ for each i, given f_j^*, σ_*^2 and $\boldsymbol{\Theta}_*$. Based on Equation (2.17) in the main article, given the true link function f_j^*, the likelihood function of $\boldsymbol{x}(t_i)$ is as follows:

$$\begin{aligned} l(\boldsymbol{x}(t_i)) &= \sum_{j=1}^{J} \frac{1}{\sigma_*^2} \left[Y_j(t_i) f_j^*(\boldsymbol{x}^\top(t_i)\boldsymbol{R}_j) - \frac{1}{2}\{f_j^*(\boldsymbol{x}^\top(t_i)\boldsymbol{R}_j)\}^2 \right] - J \log \sigma_* \\ &\quad + \sum_{k=1}^{K} \log \widetilde{p}\left(x_k(t_i) \mid N(0, \widetilde{d}_{ik})\right), \end{aligned}$$

where $\widetilde{d}_{ik}$ is defined in Equation (2.16). It is sufficient to show that for any given $\eta > 0$, there exists a sufficient small constant ϵ such that

$$P\left\{\sup_{\|\boldsymbol{x}(t_i)-\boldsymbol{x}^*(t_i)\|<\epsilon} l(\boldsymbol{x}(t_i)) < l(\boldsymbol{x}^*(t_i))\right\} \geq 1-\eta.$$

Based on Taylor expansion, we can get

$$\begin{aligned}
& l(\boldsymbol{x}(t_i)) - l(\boldsymbol{x}^*(t_i)) \\
= & \sum_{j=1}^{J} \frac{1}{\sigma_*^2}[Y_j(t_i) f_j^*(\boldsymbol{x}^\top(t_i)\boldsymbol{R}_j) - \frac{1}{2}\{f_j^*(\boldsymbol{x}^\top(t_i)\boldsymbol{R}_j)\}^2] \\
& + \sum_{k=1}^{K} \log \widetilde{p}\left(x_k(t_i) \mid N(0, \widetilde{d}_{ik})\right) - \sum_{j=1}^{J} \frac{1}{\sigma_*^2}[Y_j(t_i) f_j^*(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j) \\
& -\frac{1}{2}\{f_j^*(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j)\}^2] - \sum_{k=1}^{K} \log \widetilde{p}\left(x_k^*(t_i) \mid N(0, \widetilde{d}_{ik})\right)
\end{aligned}$$

$$\begin{aligned}
= & \left[\sum_{j=1}^{J} \frac{1}{\sigma_*^2}\{Y_j(t_i) - f_j^*(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j)\}\left\{\frac{\partial f_j(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j)}{\partial x_1(t_i)}, \ldots, \frac{\partial f_j(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j)}{\partial x_k(t_i)}, \ldots,\right.\right. \\
& \left.\left.\frac{\partial f_j(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j)}{\partial x_K(t_i)}\right\} \boldsymbol{R}_j\{\boldsymbol{x}(t_i) - \boldsymbol{x}^*(t_i)\} - \sum_{k=1}^{K} \frac{x_k^*(t_i)}{\widetilde{d}_{ik}^2}\{x_k(t_i) - x_k^*(t_i)\}\right]\{1+o_p(1)\} \\
= & \left[\sum_{j=1}^{J} K_j - \sum_{k=1}^{K} \frac{x_k^*(t_i)}{\widetilde{d}_{ik}^2}\{x_k(t_i) - x_k^*(t_i)\}\right]\{1+o_p(1)\}.
\end{aligned}$$

Since $Y_j(t_i)$, $j = 1, \ldots, J$ are independent and identically distributed with respect to j, and the expectation of $Y_j(t_i)|\boldsymbol{x}^*(t_i), f_{j*}, \sigma_*^2, \boldsymbol{\Theta}_*$ is $f_j(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j)$, we have $E(K_j|\boldsymbol{x}^*(t_i)) = 0$. Due to $\|\boldsymbol{x}(t_i) - \boldsymbol{x}^*(t_i)\| < \epsilon$, when $J \to \infty$, for a sufficiently small constant η, $l_i(\boldsymbol{x}(t_i)) - l_i(\boldsymbol{x}^*(t_i)) < 0$ with probability $1-\eta$. The consistency of $\widehat{\boldsymbol{x}}(t_i)$ is proved.

Note that $\boldsymbol{x}_k = (x_k(t_1), \ldots, x_k(t_N))^\top \sim \mathcal{NNGP}_k\left(0, \widetilde{\boldsymbol{\Sigma}}_k(\cdot, \cdot; \boldsymbol{\Theta}_{xk})\right)$. Based on Theorem 2.1 of Basawa and Rao (1980) (Choi et al. (2011)), it follows that the conclusion of the second part of the theorem holds. Next, we verify the conditions in Basawa and Rao (1980). Denote $\boldsymbol{x}_k^s = (x_k(t_1), \ldots, x_k(t_s))^\top$ and $\boldsymbol{x}_k^s$ has nonsingular distribution $\mathcal{NNGP}_k\left(0, \widetilde{\boldsymbol{\Sigma}}_k^s(\cdot, \cdot; \boldsymbol{\Theta}_{xk})\right)$. Let $p_s(\boldsymbol{\Theta}_{xk}) = p(\boldsymbol{x}_k^s; \boldsymbol{\Theta}_{xk})/p(\boldsymbol{x}_k^{s-1}; \boldsymbol{\Theta}_{xk})$ for every $s \geq 1$. It is easy to show that $p_s(\boldsymbol{\Theta}_{xk})$ follows a normal distribution with mean $\mu_k^s(\boldsymbol{\Theta}_{xk})$ and variance $V_k^s(\boldsymbol{\Theta}_{xk})$, where $\mu_k^s(\boldsymbol{\Theta}_{xk})$ and $V_k^s(\boldsymbol{\Theta}_{xk})$ are functions of $\boldsymbol{\Theta}_{xk}$. Then by calculation, $\phi_s(\boldsymbol{\Theta}_{xk})$ and its derivatives are given by

$$\begin{aligned}
\phi_s(\boldsymbol{\Theta}_{xk}) = \log p_s(\boldsymbol{\Theta}_{xk}) &= -\log\left((2\pi)^{s/2}|V_k^s(\boldsymbol{\Theta}_{xk})|^{1/2}\right) \\
&\quad - \frac{\{\boldsymbol{x}_k^s - \mu_k^s(\boldsymbol{\Theta}_{xk})\}^\top \{V_k^s(\boldsymbol{\Theta}_{xk})\}^{-1} \{\boldsymbol{x}_k^s - \mu_k^s(\boldsymbol{\Theta}_{xk})\}}{2}, \\
\dot{\phi}_s(\boldsymbol{\Theta}_{xk}) = \frac{\partial \phi_s(\boldsymbol{\Theta}_{xk})}{\partial \boldsymbol{\Theta}_{xk}} &= -\frac{1}{2} tr\left(\{V_k^s(\boldsymbol{\Theta}_{xk})\}^{-1} V_k^{s\prime}(\boldsymbol{\Theta}_{xk})\right) \\
&\quad + \{\boldsymbol{x}_k^s - \mu_k^s(\boldsymbol{\Theta}_{xk})\}^\top \{V_k^s(\boldsymbol{\Theta}_{xk})\}^{-1} \mu_k^{s\prime}(\boldsymbol{\Theta}_{xk}) \\
- \frac{1}{2}\{\boldsymbol{x}_k^s - \mu_k^s(\boldsymbol{\Theta}_{xk})\}^\top &\{V_k^s(\boldsymbol{\Theta}_{xk})\}^{-1} V_k^{s\prime}(\boldsymbol{\Theta}_{xk}) \{V_k^s(\boldsymbol{\Theta}_{xk})\}^{-1} \{\boldsymbol{x}_k^s - \mu_k^s(\boldsymbol{\Theta}_{xk})\}, \\
\ddot{\phi}_s(\boldsymbol{\Theta}_{xk}) = \frac{\partial \dot{\phi}_s(\boldsymbol{\Theta}_{xk})}{\partial \boldsymbol{\Theta}_{xk}} &= \{\boldsymbol{x}_k^s - \mu_k^s(\boldsymbol{\Theta}_{xk})\}^\top A_k(\boldsymbol{\Theta}_{xk}) \{\boldsymbol{x}_k^s - \mu_k^s(\boldsymbol{\Theta}_{xk})\} \\
&\quad + B_k(\boldsymbol{\Theta}_{xk}) \{\boldsymbol{x}_k^s - \mu_k^s(\boldsymbol{\Theta}_{xk})\} + C_k(\boldsymbol{\Theta}_{xk}),
\end{aligned}$$

where $A_k(\boldsymbol{\Theta}_{xk}), B_k(\boldsymbol{\Theta}_{xk})$, and $C_k(\boldsymbol{\Theta}_{xk})$ are some functions of $\boldsymbol{\Theta}_{xk}$, made up of first and second derivatives of $\mu_k^s(\boldsymbol{\Theta}_{xk})$ and $V_k^s(\boldsymbol{\Theta}_{xk})$. Subsequently, we apply a similar methodology as used in Choi et al. (2011) to verify the conditions C1-C4 in Basawa and Rao (1980), thus establishing the consistency of $\widehat{\boldsymbol{\Theta}}_{xk}$.

When $(\boldsymbol{x}^*(t_i), \boldsymbol{\Theta}^*_{xk})$ are given, σ^2_* can be estimated by maximizing

$$\ell(\sigma^2|\boldsymbol{x}^*(t_i), \boldsymbol{\Theta}^*_{xk}) = \sum_{j=1}^{J} \Big[-\frac{1}{2}\log\left|\sigma^2\boldsymbol{I}_N + \boldsymbol{C}_N\right| - \frac{1}{2}\mathrm{tr}\big((\sigma^2\boldsymbol{I}_N + \boldsymbol{C}_N)^{-1}\boldsymbol{Y}_j\boldsymbol{Y}_j^{\top}\big)\Big].$$

In Theorem A1 of the appendix in Choi et al. (2011), let $\theta_0 = \sigma^2_*$, similar to the proof of $\widehat{\boldsymbol{\Theta}}_{xk}$ above, then the consistency of $\widehat{\sigma}^2$ can be obtained as $N, J \to \infty$. □

S3 Proof of Theorem 3

Proof of Theorem 3. Choi and Schervish (2007) established posterior consistency in nonparametric regression problems with Gaussian errors when suitable prior distributions are used for the unknown regression function and the noise variance. Here, we will use Theorem 1 from Choi and Schervish (2007) to prove the posterior consistency of f_j for the true vectors $\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j$. Next, we verify the conditions of Theorem 1 in Choi and Schervish (2007). For each j, define $\boldsymbol{\gamma} = f_j$, $\boldsymbol{\gamma}_* = f_j^*$ and $p_i(\cdot;\boldsymbol{\gamma})$ as the density function of $Y_j(t_i)|\boldsymbol{\gamma}, \boldsymbol{R}_j, \boldsymbol{x}^*(t_i)$. Define $\Lambda(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) = \log\frac{p_i(Y_j(t_i)|\boldsymbol{x}^*(t_i);\boldsymbol{\gamma}^*)}{p_i(Y_j(t_i)|\boldsymbol{x}^*(t_i);\boldsymbol{\gamma})}$, $K_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) = E_{\boldsymbol{\gamma}^*}(\Lambda(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}))$ and $V_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) = \mathrm{Var}_{\boldsymbol{\gamma}^*}(\Lambda(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}))$. Due to $Y_j(t_i)|f_j(\cdot) \sim N\left(f_j(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j), \sigma^2\right)$ independently for $i = 1, \ldots, N$, then by calculation,

$$K_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) = \frac{1}{2}\frac{\left[f_j^*\left(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j\right) - f_j\left(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j\right)\right]^2}{\sigma^2}$$

and

$$V_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) = \left[f_j\left(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j\right) - f_j^*\left(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j\right)\right]^2.$$

Then, from the calculations of $K_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma})$ and $V_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma})$, it is easily shown that (i) for every $\delta > 0$, there exists $\epsilon > 0$ such that $\forall\, \boldsymbol{\gamma} \in W_\epsilon^0$, $K_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma}) < \delta$ for all i and that (ii) $\sum\limits_{i=1}^{\infty} \frac{V_i(\boldsymbol{\gamma}^*, \boldsymbol{\gamma})}{i^2} < \infty$, $\forall \boldsymbol{\gamma} \in W_\epsilon^0$, where

$$W_\epsilon^0 = \left\{ f_j : \int \left| f_j\left(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j\right) - f_j^*\left(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j\right)\right| d\boldsymbol{\lambda}(\boldsymbol{x}^*(t_i)) < \epsilon \right\}.$$

Hence, under the consistent estimator $\widehat{\boldsymbol{\Theta}}_{fj}$ of hyperparameters, if the prior, Π assigns positive probability to W_ϵ^0 for each $\epsilon > 0$, then condition (A1) of Theorem 1 in Choi and Schervish (2007) holds.

Following the test function construction method and verification methods for the existence of tests in Choi and Schervish (2007), then we can show that the subconditions (i) and (ii) of (A2) satisfies. The subcondition (iii) of (A2) requires that there exists a constant C_2 such that $\Pi(\boldsymbol{\gamma}_n^c) \le C_2 e^{-C_2 n}$. As in the construction of seives, if the prior distribution for the regression function, Π, assigns exponentially small probability to the two sets $\boldsymbol{\gamma}_{n,0}^c = \{f_j : \|f_j\|_\infty > M_n\}$ and $\boldsymbol{\gamma}_{n,i}^c = \{f_j : \|\frac{\partial f_j}{\partial x_k(t_i)}\|_\infty > M_n\}$, then the condition (iii) of (A2) holds.

Note that

$$
\begin{aligned}
&\left| f_j\left(\widehat{\boldsymbol{x}}^{\top}(t_i)\boldsymbol{R}_j\right) - f_j^*\left(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j\right)\right| \\
\leq\ &\left| f_j\left(\widehat{\boldsymbol{x}}^{\top}(t_i)\boldsymbol{R}_j\right) - f_j^*\left(\widehat{\boldsymbol{x}}^{\top}(t_i)\boldsymbol{R}_j\right)\right| + \left| f_j^*\left(\widehat{\boldsymbol{x}}^{\top}(t_i)\boldsymbol{R}_j\right) - f_j^*\left(\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j\right)\right|.
\end{aligned}
$$

Since $\widehat{\boldsymbol{x}}^{\top}(t_i)\boldsymbol{R}_j$ converges to $\boldsymbol{x}^{*\top}(t_i)\boldsymbol{R}_j$ in probability and f_j^* is continuous, by combining the above results, the theorem is proved. □

References for Supplementary Material

Basawa, I. and Rao, B. (1980). *Statistical Inference for Stochastic Processes.* Academic Press, London.

Chen, Y., Li, X. and Zhang, S. (2020). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *Journal of the American Statistical Association* **115**(532), 1756–1770.

Choi, T. and Schervish, M. J. (2007). On posterior consistency in nonparametric regression problems. *Journal of Multivariate Analysis* **98**(10), 1969–1987.

Choi, T., Shi, J. Q. and Wang, B. (2011). A gaussian process regression approach to a single-index model. *Journal of Nonparametric Statistics* **23**(1), 21–36.