Supplementary Document for the Manuscript entitled "Recovering the Underlying Trajectory from Sparse and Irregular Longitudinal Data"

Theorem 1. *For any given value of* M

$$\{\hat{\psi}_1, \ldots, \hat{\psi}_M\} = \arg\min \frac{1}{n}\sum_{i=1}^{n}\bigg(\int [x_i(t) - \sum_{m=1}^{M}\alpha_{mi}\psi_m(t)]^2 dt\bigg), \tag{1}$$

subject to $\langle\psi_i, \psi_j\rangle = \delta_{ij}$. *Then* $\hat{\psi}_1, \ldots, \hat{\psi}_M$ *are the first* M *eigenfunctions of* $\hat{K}(s,t) = \frac{1}{n}\sum_{i=1}[x_i(s)x_i(t)]$ *and* $\alpha_{ki} = \langle x_i, \psi_k\rangle$.

Proof. We start with $M = 1$, then the problem above becomes

$$\hat{\psi}_1 = \arg\min \frac{1}{n}\sum_{i=1}^{n}\bigg(\int [x_i(t) - \alpha_{1i}\psi_1(t)]^2 dt\bigg),$$

subject to $||\psi_1||^2 = 1$. For every $x_i(t)$, we can express it as $x_i(t) = \langle x_i, \psi_1\rangle\psi_1(t) + \eta_i(t)$, in which $\eta_i \perp \psi_1$.

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\bigg(\int [x_i(t) - \alpha_{1i}\psi_1(t)]^2 dt\bigg) &= \frac{1}{n}\bigg(\int [\langle x_i, \psi_1\rangle\psi_1(t) + \eta_i(t) - \alpha_{1i}\psi_1(t)]^2 dt\bigg) \\ &= \frac{1}{n}\sum_{i=1}^{n}\bigg(\int [(\langle x_i, \psi_1\rangle - \alpha_{1i})\psi_1(t) + \eta_i(t)]^2 dt\bigg) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int \bigg[(\langle x_i, \psi_1\rangle - \alpha_{1i})\psi_1(t)\bigg]^2 dt + \frac{1}{n}\sum_{i=1}^{n}\int \eta_i^2(t) dt\end{aligned}$$

The first term is minimized when $\alpha_{1i} = \langle x_i, \psi_1 \rangle$ and the second term is minimized only when $\psi_1(t)$ is the first eigenfunction of $\hat{K}(s,t)$. This is because

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\int [x_i(t) - \langle x_i, \psi_1\rangle \psi_1(t)]^2 dt\right) &= \frac{1}{n}\sum_{i=1}^{n}\left(\int x_i(t)^2 dt - 2\int \langle x_i, \psi_1\rangle x_i(t)\psi_1(t)dt + \langle x_i, \psi_1\rangle^2 \int \psi_1^2(t)dt\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int x_i^2(t)dt - \frac{1}{n}\sum_{i=1}^{n}\langle x_i, \psi_1\rangle^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\int x_i^2(t)dt - \frac{1}{n}\sum_{i=1}^{n}\int x_i(t)\psi_1(t)dt \int x_i(s)\psi_1(s)ds \\
&= \frac{1}{n}\sum_{i=1}^{n}\int x_i^2(t)dt - \frac{1}{n}\sum_{i=1}^{n}\int\int x_i(t)\psi_1(t)x_i(s)\psi_1(s)dsdt \\
&= \frac{1}{n}\sum_{i=1}^{n}\int x_i^2(t)dt - \int\int \psi_1(t)\hat{K}(s,t)\psi_1(s)dsdt.
\end{aligned}$$

We can easily see that the second term is maximized when $\psi_1(t)$ is the first eigenfunction of $\hat{K}(s,t)$.

When $M > 1$, we can write each $x_i(t) = \sum_{m=1}^{M}\langle x_i, \psi_m\rangle \psi_m(t) + \eta_i(t)$ and $\eta_i \perp span\{\psi_1, \ldots, \psi_M\}$. Following the same strategy, we can first show that $\alpha_{mi} = \langle x_i, \psi_m\rangle$. Then the problem becomes minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\left(\int \left[x_i(t) - \sum_{m=1}^{M}\langle x_i, \psi_m\rangle \psi_m(t)\right]^2 dt\right) = \frac{1}{n}\sum_{i=1}^{n}\int x_i^2(t)dt - \sum_{m=1}^{M}\int\int \psi_m(t)\hat{K}(s,t)\psi_m(s)dsdt,$$

which is equivalent to maximizing the second term. It is only when $\hat{\psi}_1, \ldots, \hat{\psi}_K$ are the first K leading eigenfunctions of the sample covariance function $\hat{K}(s,t)$ that the second term is maximized.

□

Let the Mercer expansion with kernel $K(s,t) = \mathtt{E}X(s)X(t)$ of the stochastic process $X(t)$ be $\{\lambda_k, \psi_k^0 : k = 1, 2, \cdots\}$. That is, there are r.v. a_k's and square integrable functions ψ_k^0's on $[0,1]$ such that $X(t) = \sum a_k \psi_k^0(t)$. We observe that $\mathtt{E}a_k a_l = \lambda_k \delta_{kl}$ with the λ_i's positive and strictly decreasing.

Assumption A0: $\sum_k \mathtt{E}a_k^4 < \infty$, $\psi_k^0(t) < M$ for all $t \in [0,1]$ and each $k = 1, 2, \cdots$.

Assumption A1: The parameter set $\Theta = \{((\alpha_i), (\beta_k)) \in C_{00} \oplus B_{\ell_2}\}$ is manageable (**?**), where $C_{00} = \{(c_i) : |c_i| < M \text{ for finite many } i's \text{ and other } c_i's \text{ are } 0 \text{ for some constant } M\}$ and $B_{\ell_2} = \{(b_i) : \sum |b_i|^2 \leq 1\}$.

Lemma 1. *Let $A = [a_{ij}]$ be an $n \times p$ matrix, $r \le \min(n,p)$, and define $\tilde{A} = \sum_{i=1}^{r} \boldsymbol{\alpha}_i \otimes \boldsymbol{\beta}_i$, where $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ are the r left and r right singular eigenvectors of A, i.e., the eigenvectors of $AA^\top$ and $A^\top A$, respectively. Then the Frobenius norm of $A - B$, where B is an $n \times p$ matrix of rank r, is minimized when $B = \tilde{A}$.*

Proof. By singular value decomposition of A, we can write $A = \sum_{k=1}^{\min(n,p)} d_k \mathbf{u}_k \otimes \mathbf{v}_k$, where $(\mathbf{u}_k)_{k=1}^{n}$ and $(\mathbf{v}_k)_{k=1}^{p}$ are orthogonal basis in $\mathbb{R}^n$ and $\mathbb{R}^p$. Hence $A - \sum_{i=1}^{r} \boldsymbol{\alpha}_i \otimes \boldsymbol{\beta}_i$ has minimum squared Frobenius norm $\sum_{i=r+1}^{\min(n,p)} d_i^2$ with minimizing $\boldsymbol{\alpha}_i = d_i \mathbf{u}_i$ and $\boldsymbol{\beta}_i = \mathbf{v}_i$. □

Given sparse observations of functional data $y_{ij} = x_i(t_{ij}) + \epsilon_{ij}$, where the observation times $t_{ij}, j = 1, \dots, n_i$, for subject i are uniformly drawn from $[0,1]$, recall the objective function

$$L_n(\boldsymbol{\alpha}, \psi) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} [y_{ij} - \alpha_i \psi(t_{ij})]^2$$

where $\boldsymbol{\alpha} = (\alpha_i) \in \mathbb{R}^n$ and $\psi(t)$ is a function on $[0,1]$ with constraint $\int_0^1 \psi^2(t) d\, t = 1$.

Theorem 3. *Under assumptions A0-A1, if $(\hat{\psi}(t), [\hat{\alpha}_i, i = 1, \dots, n])$ jointly minimize L_n, then as $n \to \infty$, $|||\hat{\psi} - \psi_1^0||| \to 0$ in probablity, and $n^{-1} \sum_{i=1}^{n} \hat{\alpha}_i \to < \psi_1^0, \mathtt{E}(X) >$. Every theorem must be numbered by hand.*

Proof. Since $\Psi^0 := \{\psi_k^0 : k = 1, 2, \cdots\}$ is a complete orthonormal system in $L^2[0,1]$, we represent $\psi(t)$ as $\psi(t) = \boldsymbol{\beta}^T \Psi^0$ for coefficient vector $\boldsymbol{\beta} = (\beta_i)$. We also represent $x_i(t) = \sum_{k=1}^{\infty} a_{ik} \psi_k^0(t)$ for $i = 1, \cdots, n$, then the objection function in Equation (3) of the main manuscript becomes

$$\begin{aligned}
L_n(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} [y_{ij} - \alpha_i \psi(t_{ij})]^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{\infty} (a_{ik} - \alpha_i \beta_k)^2 m_{ikk} + \sum_{k \neq l} (a_{ik} - \alpha_i \beta_k)(a_{il} - \alpha_i \beta_l) m_{ikl} \\
&\quad + \frac{-2}{n} \sum_{i=1}^{n} \sum_{k=1}^{\infty} (a_{ik} - \alpha_i \beta_k) \frac{1}{n_i} \sum_{j=1}^{n_i} \epsilon_{ij} \psi_k^0(t_{ij}) \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \epsilon_{ij}^2
\end{aligned}$$

where $m_{ikl} = \frac{1}{n_i}\sum_{j=1}^{n_i}\psi_k^0(t_{ij})\psi_l^0(t_{ij})$. Note that, since t_{ij} are uniformly drawn from $[0,1]$,

$$\mathtt{E}m_{ikl} = \frac{1}{n_i}\sum_{j=1}^{n_i}\int_0^1 [\psi_k^0(t_{ij})\psi_l^0(t_{ij})]\,dt_{ij} = \delta_{kl}.$$

We will show that

$$L_n(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{\infty}(a_{ik}-\alpha_i\beta_k)^2 + o_P(1;\Theta) + \sigma^2,$$

By LLN,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\epsilon_{ij}^2 = \sigma^2 + o_P(1;\Theta).$$

where $o_P(1;\Theta)$ is some random quantity uniformly small over the parameter set Θ in probability. In order to show

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{\infty}(a_{ik}-\alpha_i\beta_k)^2 m_{ikk} - \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{\infty}(a_{ik}-\alpha_i\beta_k)^2 = o_P(1;\Theta),$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k\neq l}(a_{ik}-\alpha_i\beta_k)(a_{il}-\alpha_i\beta_l)m_{ikl} = o_P(1,\Theta),$$

$$\text{and}\ \frac{-2}{n}\sum_{i=1}^{n}\sum_{k=1}^{\infty}(a_{ik}-\alpha_i\beta_k)\frac{1}{n_i}\sum_{j=1}^{n_i}\epsilon_{ij}\psi_k^0(t_{ij}) = o_P(1;\Theta);$$

we check Pollard's ULLN conditions

$$\sum_i \mathtt{E}\{\sup_\Theta \sum_{k=1}^{\infty}(a_{ik}-\alpha_i\beta_k)^2 m_{ikk}\}^2/i^2 < \infty$$

$$\sum_i \mathtt{E}\{\sup_\Theta \sum_{k\neq l}(a_{ik}-\alpha_i\beta_k)(a_{il}-\alpha_i\beta_l)m_{ikl}\}/i^2 < \infty$$

$$\sum_i \mathtt{E}\{\sup_\Theta \sum_{k=1}^{\infty}(a_{ik}-\alpha_i\beta_k)\frac{1}{n_i}\sum_{j=1}^{n_i}\epsilon_{ij}\psi_k^0(t_{ij})\}^2/i^2 < \infty.$$

By the boundedness of the eigenfunction ψ_k^0 and Cauchy-Schwartz inequality, it suffices to check

$$\mathtt{E}\{\sup_\Theta \sum_{k=1}^{\infty}(a_{ik}-\alpha_i\beta_k)^2\}^2 < \infty.$$

which follows from assumptions A0-A1. If we denote

$$\tilde{\boldsymbol{\alpha}}_n, \tilde{\boldsymbol{\beta}}_n = \arg\min M_n \equiv \arg\min \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{\infty}(a_{ik}-\alpha_i\beta_k)^2$$

then

$$\hat{\boldsymbol{\alpha}}_n = \tilde{\boldsymbol{\alpha}}_n + o_P(1;\Theta); \hat{\boldsymbol{\beta}}_n = \tilde{\boldsymbol{\beta}}_n + o_P(1;\Theta)$$

By the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} a_{ik}a_{il} \to \mathtt{E}a_{ik}a_{il} = \lambda_k\delta_{kl}.$$

With $A = (\frac{1}{\sqrt{n}}a_{ik})_{i=1,\ldots,n}^{k=1,\cdots,\infty}$, and $(A^TA)_{ij} \to \lambda_k\delta_{kl}$,

$$\|A^TA - \text{diag}[\lambda_1, \cdots, \lambda_p]\|_F \to 0.$$

Hence the eigenvectors (u_k) of A^TA converges to those of $\text{diag}[\lambda_1, \cdots, \lambda_p]$; and the eigen-vectors $\{\mathbf{v}_k\}$ of AA^T are $v_k = A^Tu_k$, which will converges to the columns of A in the space c_{00} of finitely support sequencer equipped with the ℓ_2 norm. Hence the estimate $(\tilde{\alpha}_i)$, the first eigen-vector of AA^T, will converge to the first column of A in ℓ_2 norm: $(\frac{1}{\sqrt{n}}\tilde{\alpha}_i) - (\frac{1}{\sqrt{n}}a_{i1})$ converge to 0 in ℓ_2. Also, $\tilde{\boldsymbol{\beta}}$ converges to $\mathbf{e}_1$, the first unit vector. Hence $\hat{\psi}(t) := \hat{\boldsymbol{\beta}}_n^{(R)}\Psi^0(t)$ converges to $\psi_1^0(t)$ in $L^2[0,1]$, and $\frac{1}{n}\sum_{i=1}^n \hat{\alpha}_i \to \mathtt{E}a_{i1}$.

□

Theorem 2. *The estimators* $\{\hat{\psi}_k : k = 2, 3, \cdots\}$ *as obtained in Section 3.2 of* $\{\psi_k^0 : k = 2, 3, \cdots\}$ *are consistent in* $L_2(0,1)$.

Proof. Since the estimates $(\hat{\alpha}_i)$ and $(\hat{\beta}_k)$ are such that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}[(a_{i1}\psi_1^0(t_{ij}) - \hat{\alpha}_i\hat{\psi}_1(t_{ij})]^2 = o_P(1)$$

hence we can rewrite the model as

$$y_{ij} - \hat{\alpha}_i\hat{\psi}_1(t_{ij}) = \sum_{k=2}^{\infty} a_{ik}\psi_k^0(t_{ij}) + o_P(1),$$

by the same argument as in the proof of Theorem 1, the proof is complete. □