Supplementary Materials for "Dynamic Multivariate Functional Data Modeling via Sparse Subspace Learning"

Chen Zhang[1], Hao Yan[2], Seungho Lee[3], Jianjun Shi[4]

[1]Tsinghua University, Beijing, China, 100084

[2]Arizona State University, Tempe, AZ, 85282

[3]Samsung Electronics, Suwon, South Korea, 443742

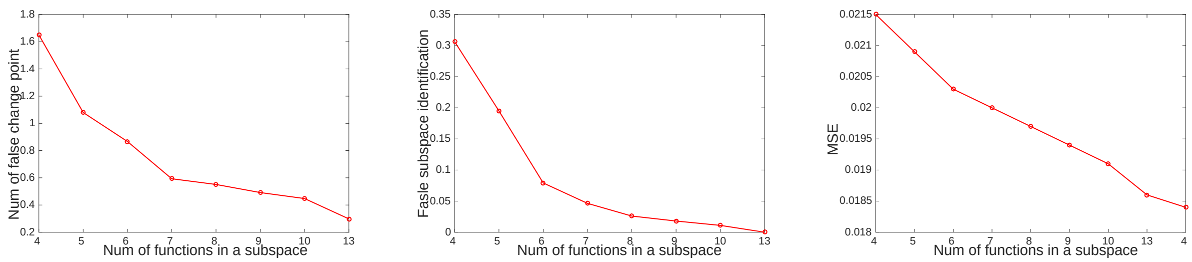
[4]Georgia Institute of Technology, Atlanta, GA, 30332

A Additional Results of Numerical Studies

A.1 Sensitivity Analysis

To have a complete understanding of DFSL, we evaluate its performance sensitivity to the number of functions per subspace for Model (I). The results are shown in Figure A.1. As the number of functions per subspace increases, DFSL has fewer false change points, fewer false subspace identification, and smaller representation MSE overall. This indicates that DFSL is particularly beneficial for higher dimensional cases with larger number of functions.

This phenomenon is consistent with Theorem 2. Because when $d_l = 3$, $p_l = 4$ just satisfies Assumption 2, its false subspace identification rate cannot be very small. As p_l increases, the number of false subspace identification decreases. This is because as the number of functions per subspace increases, the point density per subspace increases, indicating more information we can gather from each subspace. Consequently, the probability that one function can be well fitted by its neighbors from the same subspace increases, and the probability of being fitted by functions from other subspaces decreases.



(a) False change point detection (b) False subspace identification (c) Representation MSE

Figure A.1: Modeling results of DFSL as the number of functions per subspace increases for Model (I).

A.2 Additional Tables

Tables A1, A2, A3, A4 and A5 show the exact values of the estimated results together with their standard deviations presented in Figures 6 and 7 of the main text and Figure A.1. Here we add the results of KerGra. Specifically, we use the mean square of the difference between the original data and the kernel smoothed ones as the representation MSE of KerGra.

Table A1: Average false change point detection in Figure 6a, false subspace identification in Figure 6b and their standard deviations (in the parentheses) for Model (I).

	False subspace identification($\times 10^{-1}$)				False change point detection($\times 10^{-1}$)	
Noise std	DFSL	SFSL	FPCAGra	KerGra	DFSL	KerGra
0.025	0(0)	0.19(0.08)	4.08(0.01)	0.05(0.00)	0.08(0.01)	15.2(0.03)
0.05	0(0)	0.31(0.09)	2.63(0.01)	0.06(0.00)	1.87(0.05)	17.5(0.02)
0.1	0(0)	0.53(0.01)	2.17(0.00)	0.25(0.00)	2.75(0.04)	21.2(0.03)
0.15	0.25(0.01)	0.78(0.01)	1.54(0.00)	0.51(0.00)	6.83(0.07)	27.5(0.05)
0.25	0.79(0.01)	1.58(0.02)	0.83(0.00)	2.69(0.01)	15.1(0.01)	42.0(0.06)
0.5	2.86(0.02)	3.52(0.02)	0(0)	5.38(0.00)	16.5(0.09)	55.0(0.17)

Table A2: Average representation MSE in Figure 6c and their standard deviations (in the parentheses) for Model (I)($\times 10^{-3}$).

Noise std	DFSL	SFSL	MFPCA	FPCAGra	KerGra
0.025	0.4(0.01)	5.8(0.03)	17.3(0.00)	11.0(0.00)	2.10(0.01)
0.05	1.20(0.01)	8.02(0.04)	17.1(0.01)	11.3(0.00)	2.50(0.01)
0.1	3.95(0.01)	9.67(0.01)	16.9(0.00)	11.8(0.00)	5.02(0.01)
0.15	7.26(0.01)	12.7(0.02)	17.2(0.00)	12.6(0.00)	8.16(0.02)
0.25	13.3(0.01)	17.1(0.02)	17.6(0.00)	15.2(0.00)	13.9(0.02)
0.5	21.5(0.01)	21.2(0.01)	22.0(0.00)	22.5(0.00)	21.9(0.02)

Table A3: Average false change point detection in Figure 7a false subspace identification in Figure 7b and their standard deviations (in the parentheses) for Model (II).

	False subspace identification ($\times 10^{-1}$)				False change point detection($\times 1$)	
Noise std	DFSL	SFSL	FPCAGra	KerGra	DFSL	KerGra
0.025	0(0)	0(0)	0.15(0.01)	0(0)	0(0)	3.93(0.06)
0.05	0(0)	0(0)	0.10(0.01)	0(0)	0(0)	4.13(0.07)
0.1	0(0)	0(0)	0.04(0.01)	0(0)	0(0)	4.38(0.01)
0.15	0(0)	0(0)	0(0)	0.01(0.00)	0(0)	5.12(0.10)
0.25	0(0)	0(0)	0(0)	0.03(0.01)	0(0)	8.24(0.16)
0.5	0(0)	0(0)	0(0)	0.06(0.01)	0(0)	15.4(0.37)

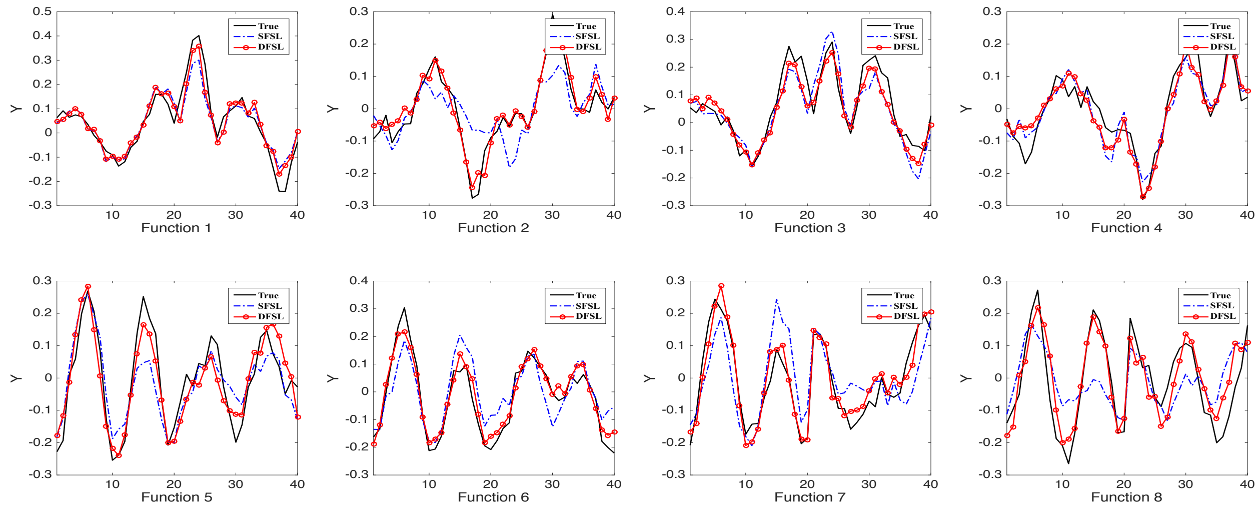
Table A4: Average representation MSE in Figure 7c and their standard deviations (in the parentheses) for Model (II)($\times 10^{-3}$).

Noise std	DFSL	SFSL	MFPCA	FPCAGra	KerGra
0.025	0.13(0.01)	0.47(0.01)	5.12(0.14)	0.46(0.01)	0.81(0.06)
0.05	0.29(0.01)	0.75(0.01)	5.36(0.15)	0.75(0.01)	1.41(0.07)
0.1	0.89(0.05)	1.38(0.03)	5.79(0.13)	1.37(0.03)	2.62(0.07)
0.15	1.61(0.07)	2.15(0.03)	6.98(0.19)	2.15(0.03)	3.86(0.07)
0.25	3.12(0.07)	3.62(0.04)	6.99(0.09)	3.68(0.04)	5.42(0.06)
0.5	5.21(0.07)	5.14(0.05)	7.43(0.04)	5.12(0.05)	7.45(0.03)

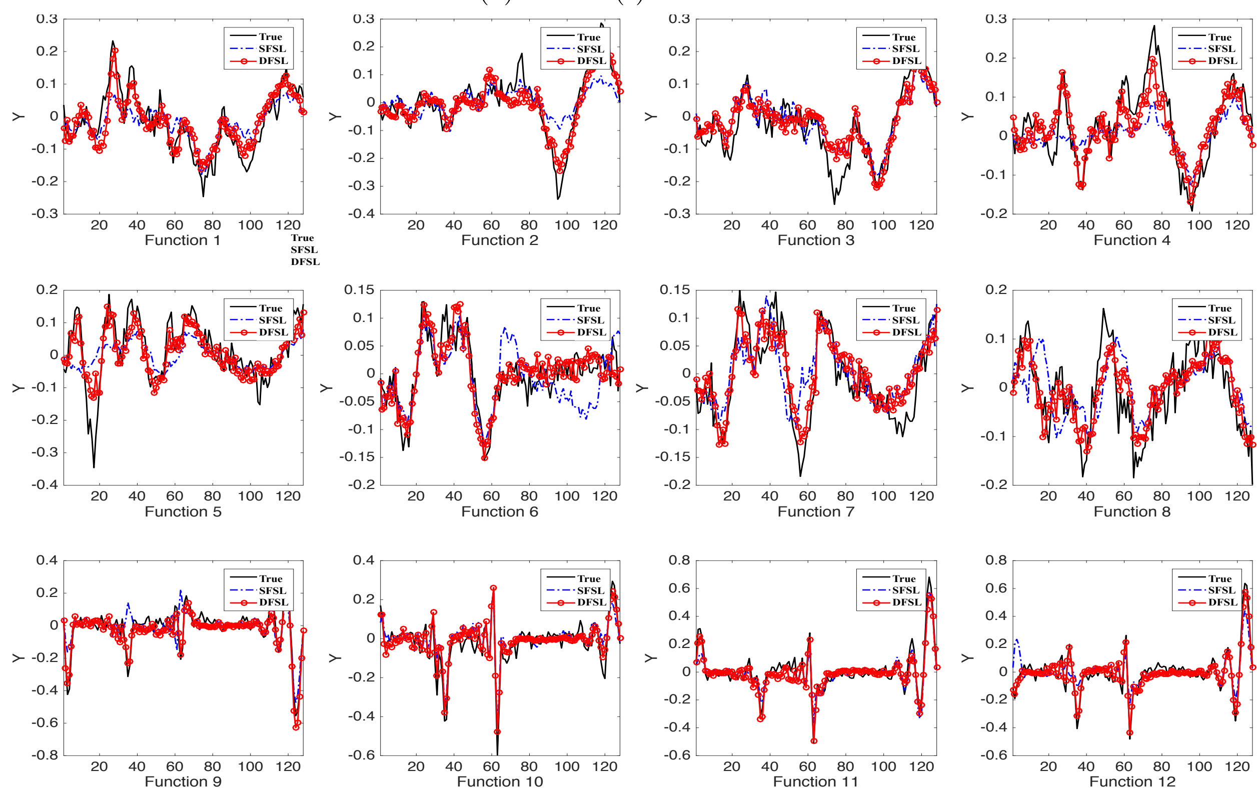
Table A5: Average representation MSE, false subspace identification, false change point detection in Figure A.1 and their standard deviations (in the parentheses) of the sensitivity analysis.

Num of functions in a subspace	MSE($\times 10^{-3}$)	False subspace identification ($\times 10^{-1}$)	Num of false change point ($\times 10^{-1}$)
4	21.5(0.01)	2.86(0.02)	16.5(0.09)
5	21.9(0.01)	1.95(0.02)	10.8(0.07)
6	20.3(0.01)	0.79(0.01)	8.61(0.05)
7	20.0(0.01)	0.46(0.01)	5.93(0.04)
8	19.6(0.01)	0.25(0.01)	5.02(0.04)
9	19.4(0.01)	1.81(0.01)	4.60(0.03)
10	19.1(0.01)	0.11(0.01)	2.98(0.02)

A.3 Additional Figures

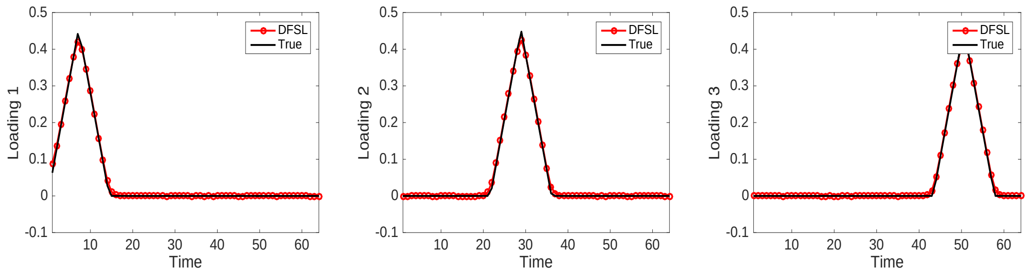


(a) Model (I)

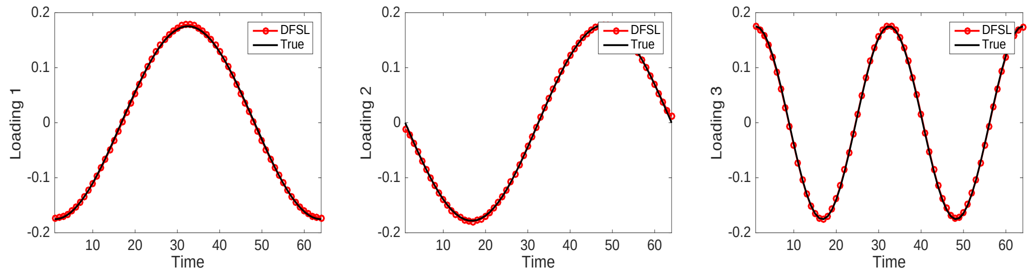


(b) Model (II)

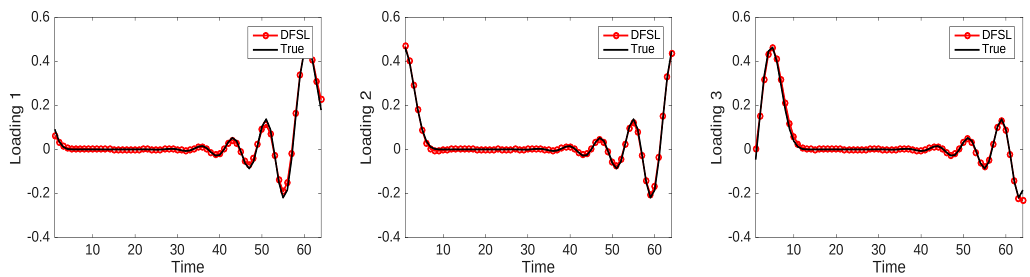
Figure A.2: Self-expression results of DFSL and SFSL for Model (I) and Model (II).



(a) Subspace (I): B-spline space



(b) Subspace (II): Fourier space



(c) Subspace (III): Wavelet space

Figure A.3: Extracted basis functions of the three subspaces for Model (II) segment 3.

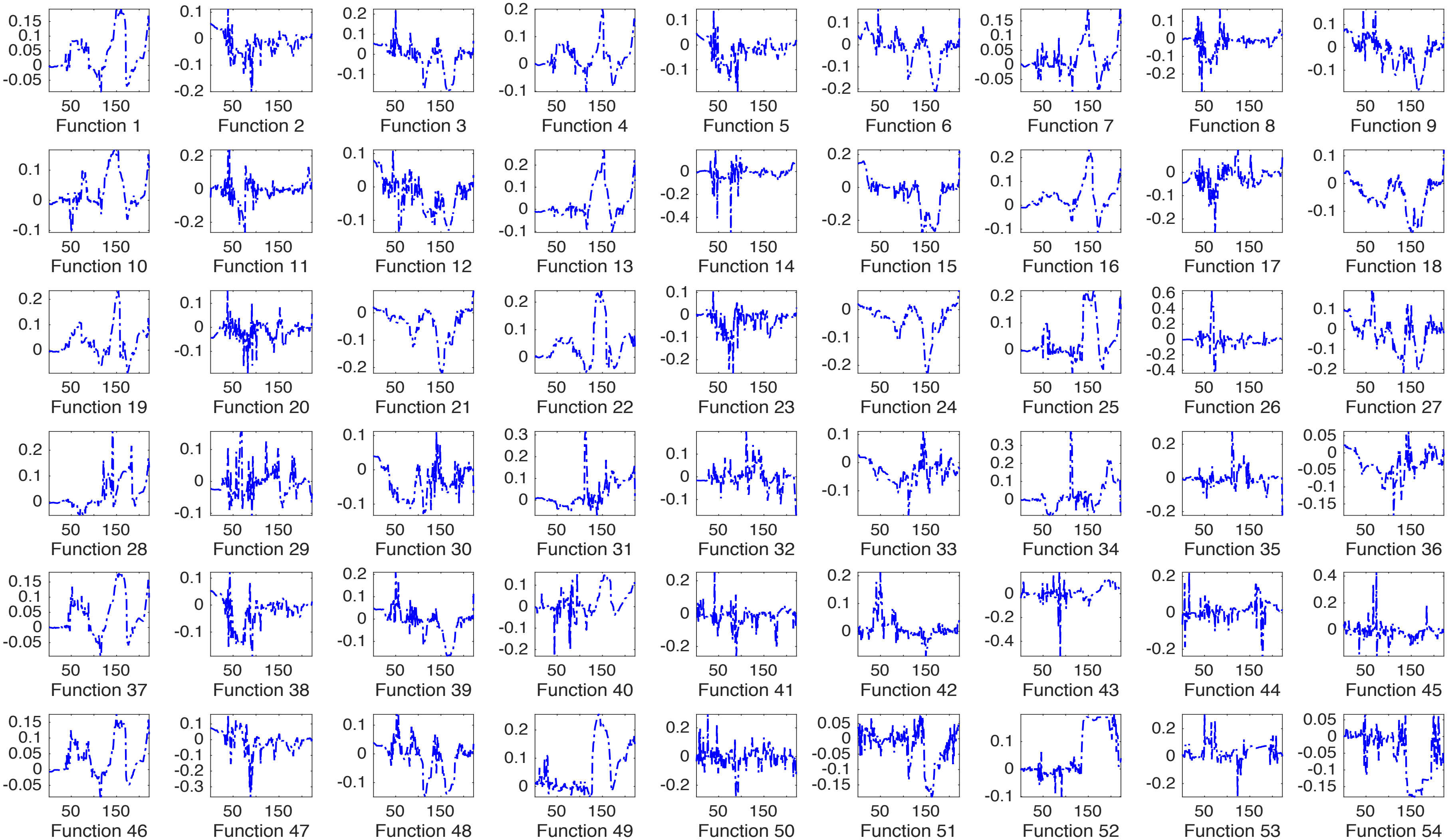


Figure A.4: Profiles of the 54 coordinates for one sample. In particular, Functions $\{3j-2, 3j-1, 3j\}$ above correspond to the functions of the $\{x, y, z\}$ coordinates of joint j, for $j = 1, \ldots, 18$.

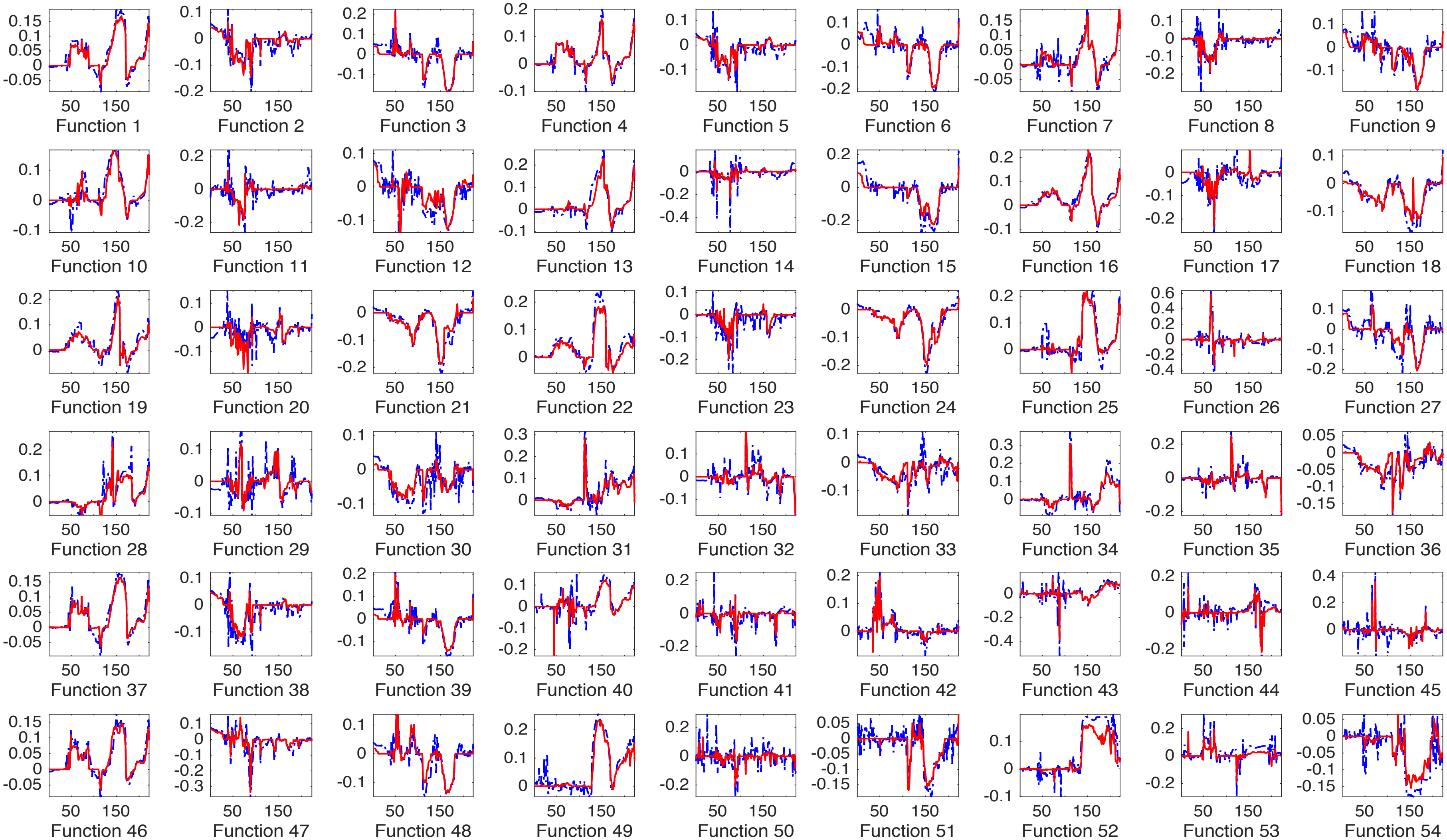


Figure A.5: Profiles of the 54 coordinates for one sample. The blue dashed lines are true curves, and the red solid lines are the self-expressed ones based on the proposed dynamic functional subspace learning (DFSL).

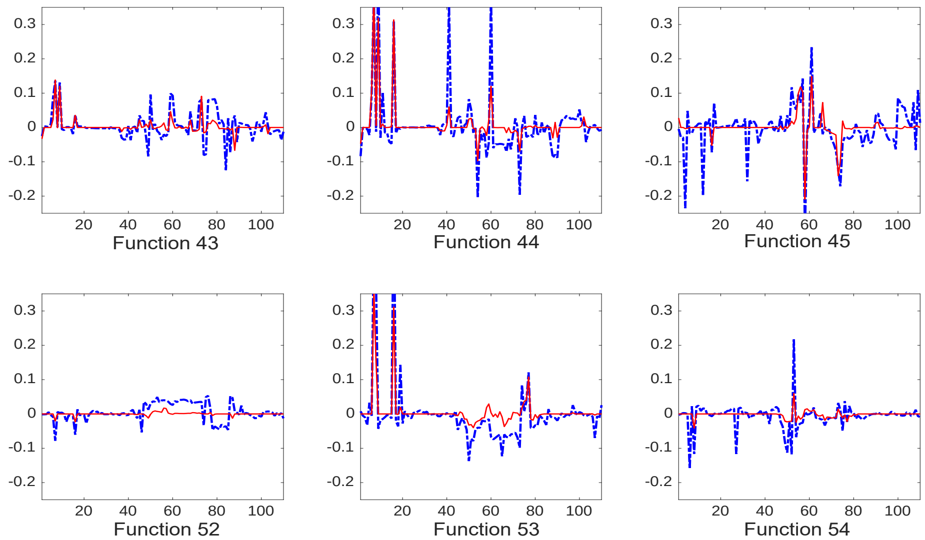


Figure A.6: The profiles of the three coordinates of the left foot (the first row), and the profiles of three coordinates of the right foot (the second row) for the first segment $[1, 110]$.

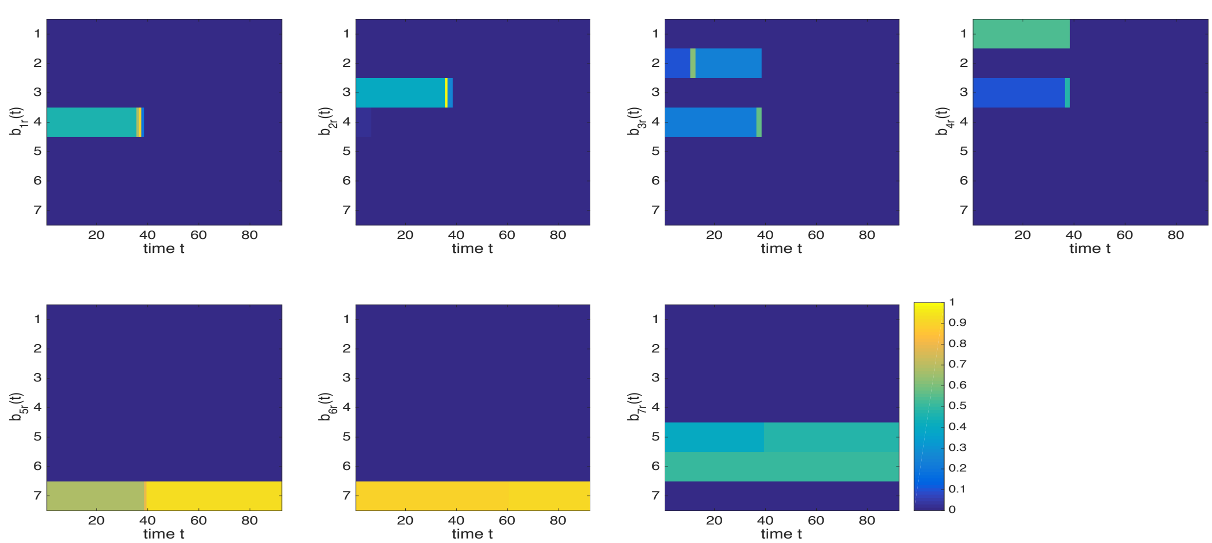


Figure A.7: The estimated $\hat{\mathbf{b}}_j (j = 1, \ldots, 7)$ for the 7 sensors.

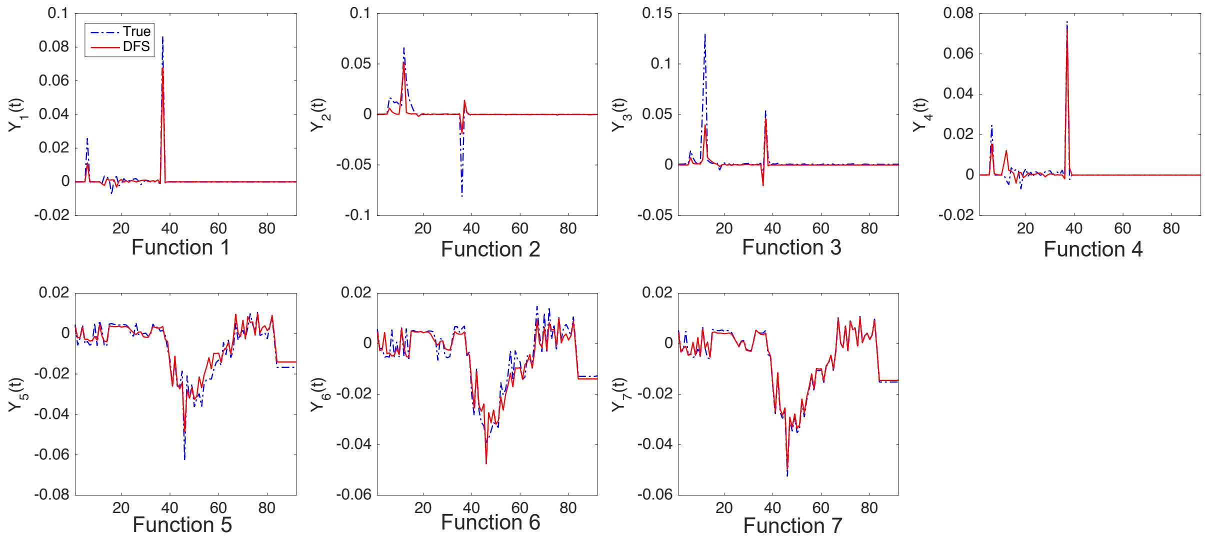


Figure A.8: Profiles of the seven sensors in an advanced manufacturing system. The blue dashed lines are true curves, and the red solid lines are the self-expressed ones based on the proposed dynamic functional subspace learning (DFSL).

A.4 Additional Algorithm

Algorithm A.1 Estimate $\mathbf{b}_{jr}(r = 1, \ldots, p, r \neq j)$ and $\boldsymbol{\Sigma}_j$ for $j = 1, \ldots, p$ based on BCD

Data: $\mathbf{Y}_i, i = 1, \ldots, N, \rho_0$
Result: Estimated $\mathbf{b}_j^g, \boldsymbol{\Sigma}_j^g, j = 1, \ldots, p$
Initialization *(by the least square method for the ordinary linear regression)*
Set $\boldsymbol{\Gamma}_j^0 = \mathbf{I}_n$ for $j = 1, \ldots, p$
for $j = 1, \ldots, p$ **do**
  Initialize $\mathbf{b}_j^0(t_k)$ by the multivariate linear regression for $k = 1, \ldots, n$ as
  for $k = 1, \ldots, n$ **do**
    $\mathbf{b}_j^0(t_k)' = (\sum_{i=1}^N \tilde{\mathbf{Y}}_{i(-j)}(t_k)'\tilde{\mathbf{Y}}_{i(-j)}(t_k))^{-1}(\sum_{i=1}^N \tilde{\mathbf{Y}}_{i(-j)}(t_k)'\tilde{Y}_{ij}(t_k))$
  end
end
Estimation *(by the block coordinate descent method)*
while $\sum_{j=1}^p ||\mathbf{b}_j^g - \mathbf{b}_j^{g-1}||_2^2 > e_2$ *or* $\sum_{j=1}^p ||\boldsymbol{\Sigma}_j^g - \boldsymbol{\Sigma}_j^{g-1}||_2^2 > e_3$ **do**
  for $j = 1, \ldots, p$ **do**
    Calculate the residual of $\mathbf{Y}_{ij}, j = 1, \ldots, p$ as
    for $k = 1, \ldots, n$ **do**
      Estimate $\epsilon_{ij}(t_k)$ for every $i = 1, \ldots, N$ as

$$\epsilon_{ij}(t_k) = Y_{ij}(t_k) - \sum_{r \neq j} Y_{ir}(t_k) b_{jr}^g(t_k)$$

    end
    Set $\boldsymbol{\Sigma}_j^g = \sum_{i=1}^N \boldsymbol{\epsilon}_{ij}'\boldsymbol{\epsilon}_{ij}$
    Estimate $\mathbf{b}_j^g$ using FISTA, with input values as $\boldsymbol{\Gamma}_j = \boldsymbol{\Sigma}_j^g, \mathbf{b}_{jr}^0 = \mathbf{b}_{jr}^{g-1}, r = 1, \ldots, p, r \neq j$
  end
  Set $g = g + 1$
end

B Proofs of Theorems 1 and 2

For demonstration purpose, we work on the case that $\boldsymbol{\Gamma}_j = \mathbf{I}$, while all of our methods and theoretical results can be extended to cases with general $\boldsymbol{\Gamma}$ by following the same procedures. For notation convenience, we redefine $\mathbf{b}_j(t_k) \in \mathcal{R}^{(p-1)\times 1}$ as $[b_{j1}(t_k), \ldots, b_{j(j-1)}(t_k)$ $, b_{j(j+1)}(t_k), \ldots, b_{jp}(t_k)]$, indicating the true cross-correlations at t_k. $\hat{\mathbf{b}}_j(t_k)$ is its correspondingly estimated one. Furthermore, we define $\boldsymbol{\beta}_j^s = \mathbf{b}_j(t_k), k \in [\tau_{s-1}, \tau_s)$ as the true cross-correlations in the s time segment. Correspondingly, $\hat{\boldsymbol{\beta}}_j^s = \hat{\mathbf{b}}_j(t_k), k \in [\hat{\tau}_{s-1}, \hat{\tau}_s)$ are the estimated cross-correlations in the $\hat{s}$ time segment.

B.1 Property of the Static Functional Subspace Learning

In particular, assume n equally-spaced sampling points $0 \leq t_1 < t_2 < \cdots < t_n \leq T$ can be observed for $Y_{ij}(t), t \in \mathcal{T}$. Then Equation (5) in the manuscript can be reformulated in terms of $\mathbf{b}_j$ as

$$\min_{\mathbf{b}_j} \lambda \sum_{k=1}^{n} ||\mathbf{b}_j||_1 + \frac{1}{2}\sum_{i=1}^{N} \mathbf{Z}_{ij}' \boldsymbol{\Gamma}_j^{-1} \mathbf{Z}_{ij} \tag{S-1}$$
$$\text{subject to } Z_{ij}(t_k) = Y_{ij}(t_k) - \mathbf{Y}_i(t_k)\mathbf{b}_j, \quad b_{jj} = 0, \text{ for } k = 1, \ldots, n.$$

Assume $\mathbf{Y}_{ij} \in \mathcal{X}_l$ whose cardinality is p_l and subspace dimension d_l. This formuatlion is the same as the sparse subspace clustering with noise in Wang and Xu (2013). According to Theorem 3 of Wang and Xu (2013) or Theorem 10 of Wang and Xu (2016), there exists a non-empty rank of λ such that the probability that $\mathbf{b}_j$ only has nonzero values for components from the same subspace as $\mathbf{Y}_{ij}$ is

$$1 - \frac{9}{Np} - \frac{1}{L^2}\sum_{l\neq k}\frac{1}{(Np_l+1)Np_k}e^{-\frac{t}{4}} - 6\sum_{l=1}^{L}(e^{\gamma_1(n-d_l)} + e^{\gamma_2 d_l} + e^{-\sqrt{Np_l d_l}}),$$

as long as the noise level $\delta := \max_{i,j} ||\boldsymbol{\epsilon}_{ij}||_2$ obeys

$$\delta(1+\delta) \leq \max_{l,k} \sqrt{\frac{n - \max_{l'} d_{l'}}{6 \log Np}} \frac{\sqrt{\log \kappa}}{40 K_2 \sqrt{d d_l}} (1 - \frac{K_1 K_2 \mathrm{aff}(\mathcal{S}_l, \mathcal{S}_k)}{\sqrt{d_k}}),$$

where $K_1 := (t \log[(p_l N + 1) p_k N] + \log L)$, $K_2 := 4\sqrt{\frac{1}{\log \kappa_l}}$, $\kappa_l := \frac{N p_l}{d_l}$, $\frac{\log \kappa}{d} := \min_l \frac{\log \kappa_l}{d_l}$, and γ_1, γ_2 are absolute constants. The proof is achieved by duality. First a set of conditions on the optimal dual variable of the dual program of (S-1) is established corresponding to all primal solutions satisfying the self-expression property. Then such a dual variable is constructed as a certificate of proof. More detailed proof procedure can be refereed to Section 5 and Section 6 of Wang and Xu (2016).

B.2 Proof of Theorem 1

Lemma 1. *A matrix $\hat{\mathbf{b}}_j$ is optimal for the optimization of Equation (10) of the manuscript if and only if there exist subgradient vectors $\hat{\mathbf{u}}_j(t_k) \in \partial||\hat{\mathbf{b}}_j(t_k) - \hat{\mathbf{b}}_j(t_{k-1})||_1$ and $\hat{\mathbf{v}}_j(t_k) \in \partial||\hat{\mathbf{b}}_j(t_k)||_1$ for $k = 1, \ldots, n$ that satisfy*

$$\sum_{i=1}^{N} \sum_{k=l}^{n} \mathbf{Y}_{i(-j)}(t_k) \langle \mathbf{Y}_{i(-j)}(t_k), \hat{\mathbf{b}}_j(t_k) - \mathbf{b}_j(t_k) \rangle - \sum_{i=1}^{N} \sum_{k=l}^{n} \mathbf{Y}_{i(-j)}(t_k) e_{ij}(t_k) + \lambda_1 \hat{\mathbf{u}}_j(t_k) + \lambda_2 \sum_{k=l}^{n} \hat{\mathbf{v}}_j(t_k) = 0, \tag{S-2}$$

for all $k = 1, \ldots, n$. Here $e_{ij}(t_k) = \sum_{r \neq j} \epsilon_{ir}(t_k) b_{jr}(t_k) - \epsilon_{ij}(t_k)$, and $\hat{\mathbf{u}}_j(t_1) = \hat{\mathbf{u}}_j(t_n) = 0$.

We follow the proof of Proposition 5 in Harchaoui and Lévy-Leduc (2010) and Theorem 2 in Kolar and Xing (2012). In particular, based on the union bound, we have

$$P[\max_{s \in 1, \ldots, S-1} |\tau_s - \hat{\tau}_s| > n\delta_N^n] \leq \sum_{s=1}^{S-1} P[|\tau_s - \hat{\tau}_s| > n\delta_N^n].$$

Then the theorem will hold up if we can prove that $P[|\tau_s - \hat{\tau}_s| > \delta_N^n] \to 0$ for all $s =$

$1, \ldots, S-1$. Define the set $A_{n,s}$ as

$$A_{N,s} = \{|\tau_s - \hat{\tau}_s| > n\delta_N^n\},$$

and the set C_N as

$$C_N = \{\max_{1\leq s\leq S-1} |\tau_s - \hat{\tau}_s| < \frac{\Delta_{min}}{2}\},$$

where $\Delta_{min} = \min_{1\leq s\leq S} |\tau_s - \tau_{s-1}|$. Then it is enough to prove that $P[A_{N,s} \cap C_N] \to 0$ and that $P[A_{N,s} \cap \bar{C}_c] \to 0$.

Let us first consider the proof of $P[A_{N,s} \cap C_N] \to 0$. Note that C_N implies that $\tau_{s-1} \leq \hat{\tau}_s \leq \tau_{s+1}$ for all $s = 1, \ldots, S-1$. We first assume $\hat{\tau}_s < \tau_s$. By using (S-2) twice with $k = \tau_s$ and $k = \hat{\tau}_s$ respectively and applying the triangle inequality, we have

$$2\lambda_1 p \geq || \sum_{i=1}^{N} \sum_{k=\hat{\tau}_s}^{\tau_{s-1}} \mathbf{Y}_{i(-j)}(t_k)\langle \mathbf{Y}_{i(-j)}(t_k), \hat{\mathbf{b}}_j(t_k) - \mathbf{b}_j(t_k)\rangle - \sum_{i=1}^{N} \sum_{k=\hat{\tau}_k}^{\tau_k - 1} \mathbf{Y}_{i(-j)}(t_k) e_{ij}(t_k)||_1. \quad \text{(S-3)}$$

Recall the true $\mathbf{b}_j(t_k), k \in [\tau_{s-1}, \tau_s)$ as $\boldsymbol{\beta}_j^s$, and define the estimated $\hat{\mathbf{b}}_j(t_k), k \in [\hat{\tau}_{s-1}, \hat{\tau}_s)$ as $\hat{\boldsymbol{\beta}}_j^s$. This yields the event $C_{N,s}$, defined as follows, occurs with probability one:

$$\begin{aligned} C_{N,s} = \{2\lambda_1 p + (\tau_s - \hat{\tau}_s)\lambda_2 p \geq & || \sum_{i=1}^{N} \sum_{k=\hat{\tau}_s}^{\tau_s - 1} \mathbf{Y}_{i(-j)}(t_k)\langle \mathbf{Y}_{i(-j)}(t_k), \boldsymbol{\beta}_j^s - \boldsymbol{\beta}_j^{s+1}\rangle||_1 & \text{(S-4)} \\ & - | \sum_{i=1}^{N} \sum_{k=\hat{\tau}_s}^{\tau_s - 1} \mathbf{Y}_{i(-j)}(t_k)\langle \mathbf{Y}_{i(-j)}(t_k), \boldsymbol{\beta}_j^{s+1} - \hat{\boldsymbol{\beta}}_j^{s+1}\rangle||_1 & \\ & - || \sum_{i=1}^{N} \sum_{k=\hat{\tau}_k}^{\tau_k - 1} \mathbf{Y}_{i(-j)}(t_k) e_{ij}(t_k)||_1\} & \text{(S-5)} \\ & =: ||R_1||_1 - ||R_2||_1 - ||R_3||_1. & \end{aligned}$$

Using that $P[A_{N,s} \cap C_N] = P[A_{N,s} \cap C_N \cap C_{N,s}]$, we can get

$$\begin{aligned} P[A_{N,s} \cap C_N] \leq & P[A_{N,s} \cap C_N \cap \{2\lambda_1 p + (\tau_s - \hat{\tau}_s)p\lambda_2 \geq \frac{1}{3}||R_1||_1\}] \\ & + P[A_{N,s} \cap C_N \cap \{||R_2||_1 \geq \frac{1}{3}||R_1||_1\}] \\ & + P[A_{N,s} \cap C_N \cap \{||R_3||_1 \geq \frac{1}{3}||R_1||_1\}] \\ & =: P[A_{N,s,1}] + P[A_{N,s,2}] + P[A_{N,s,3}]. \end{aligned} \tag{S-6}$$

According to Lemma 3, we can get the upper bound $P[A_{N,s,1}]$ with

$$P[2p\lambda_1 + (\tau_s - \hat{\tau}_s)\sqrt{p}\lambda_2 \geq N\frac{\phi_{s0}}{27}(\tau_s - \hat{\tau}_s)\xi_{min}] + 2\exp(-n\delta_N^n N + 2\log N + 2\log n).$$

With the assumption $(nN\delta_N^n p\xi_{min})^{-1}\lambda_1 \to 0$ and $(N\sqrt{p}\xi_{min})^{-1}\lambda_2 \to 0$ as either $N \to \infty$ or $n \to \infty$, we have that $P[A_{N,s,1}] \to 0$ as either $N \to \infty$ or $n \to \infty$.

To show $P[A_{N,s,2}]$ converges to zero, define $\bar{\tau}_s = \lfloor 2^{-1}(\tau_s + \tau_{s+1}) \rfloor$. Note C_N implies $\hat{\tau}_{s+1} > \bar{\tau}_s$. Consequently, we have $\mathbf{b}_j(t_k) = \hat{\boldsymbol{\beta}}_j^{s+1}$ for $k \in [\tau_s, \bar{\tau}_s]$. Using Lemma 1 twice with $k = \tau_s$ and $k = \bar{\tau}_s$ respectively, we have

$$2p\lambda_1 + (\bar{\tau}_s - \tau_s)\sqrt{p}\lambda_2 \geq ||\sum_{i=1}^{N}\sum_{k=\tau_s}^{\bar{\tau}_s - 1}\mathbf{Y}_{i(-j)}(t_k)\langle \mathbf{Y}_{i(-j)}(t_k), \boldsymbol{\beta}_j^{s+1} - \hat{\boldsymbol{\beta}}_j^{s+1}\rangle||_1 - ||\sum_{i=1}^{N}\sum_{k=\tau_s}^{\bar{\tau}_s - 1}\mathbf{Y}_{i(-j)}(t_k)e_{ij}(t_k)||_1.$$

With Lemma 3 on the display above, we have

$$||\boldsymbol{\beta}_j^{s+1} - \hat{\boldsymbol{\beta}}_j^{s+1}||_1 \leq \sqrt{p}\frac{36p\lambda_1 + 18(\bar{\tau}_s - \tau_s)\sqrt{p}\lambda_2 + 18||\sum_{i=1}^{N}\sum_{k=\tau_s}^{\bar{\tau}_s - 1}\mathbf{Y}_{i(-j)}(t_k)e_{ij}(t_k)||_1}{(\tau_{s+1} - \tau_s)\phi_{s0}N}, \tag{S-7}$$

which holds with probability at least $1 - 2\exp(-\Delta_{min}N/4 + 2\log N + 2\log n)$.

Furthermore, based on Lemma 3, we have $N\phi_{s0}(\tau_s - \hat{\tau}_s)\xi_{min}/9 \leq \sqrt{p}||R_1||_1$ and $||R_2||_1 \leq N(\tau_s - \hat{\tau}_s)9\phi_{s1}\sqrt{p}||\boldsymbol{\beta}_j^{s+1} - \hat{\boldsymbol{\beta}}_j^{s+1}||_1$ with probability at least $1 - 4\exp(-Nn\delta_N^n/2 + 2\log N +$

$2\log n$). Consequently, with (S-7), $P[A_{n,s,2}]$ has the upper bound

$$\begin{aligned}
&P[\lambda_1 \geq c_1 p^{-2}\phi_{s0}^2\phi_{s1}^{-1}\Delta_{min}N\xi_{min}] + P[\lambda_2 \geq c_2 p^{-1.5}\phi_{s0}^2\phi_{s1}^{-1}\xi_{min}N] \\
&+ P[(\bar{\tau}_j^s - \tau_j^s)^{-1}N^{-1}||\sum_{i=1}^{N}\sum_{k=\tau_s}^{\bar{\tau}_s-1}\mathbf{Y}_{i(-j)}(t_k)e_{ij}(t_k)||_1 \geq c_3\xi_{min}p^{-1}\phi_{s0}^2\phi_{s1}^{-1}] \\
&+ c_4\exp(-Nn\delta_N^n/2 + 2\log n + 2\log N).
\end{aligned} \tag{S-8}$$

Under $A_{N,s}\cap C_N$, we have $\Delta_{min} \geq n\delta_N^n$. Then as $\lambda_1/(N\xi_{min}n) \to 0$ when either $n \to \infty$ or $N \to \infty$, and $\lambda_2/(\xi_{min}N) \to 0$ when either $n \to \infty$ or $N \to \infty$. the first, second and the last term converge to zero as either $n \to \infty$ or $N \to \infty$. For the third term, it converges to zero with the rate $\exp(-c_6\log N)$ since $\log N/N \to 0$ as $N \to \infty$, and converges to zero with the rate $\exp(-c_6 \log n)$ since $\log n/n \to 0$ as $n \to \infty$.

Now we show $P[A_{N,s,3}]$ converges to zero. Because $N\phi_{s0}(\tau_s - \hat{\tau}_s)\xi_{min}/9 \leq \sqrt{p}||R_1||_1$ with probability $1 - 2\exp(-Nn\delta_N^n/2 + 2\log n + 2\log N)$, we have the upper bound of $P[A_{N,s,3}]$ as

$$P[\frac{\phi_{s0}\xi_{min}}{27} \leq \sqrt{p}\frac{||\sum_{i=1}^{N}\sum_{k=\hat{\tau}_k}^{\tau_k-1}\mathbf{Y}_{i(-j)}(t_k)e_{ij}(t_k)||_1}{N(\tau_s - \hat{\tau}_s)}] + 2\exp(-Nn\delta_N^n/2 + 2\log n + 2\log N),$$

which, according to Lemma 4, converges to zero as either $n \to \infty$ or $N \to \infty$, since $\log N/N \to 0$. As to the case with $\hat{\tau}_s > \tau_s$, we can show the above proof in a similar way. Consequently, we have $P[A_{N,s}\cap C_N] \to 0$ as either $n \to \infty$ or $N \to \infty$.

We proceed to show that $P[A_{N,s}\cap \bar{C}_N] \to 0$ as either $n \to \infty$ or $N \to \infty$. Recall $\bar{C}_N = \{\max_s |\hat{\tau}_s - \tau_s| \geq \Delta_{min}/2\}$, We now split it into three events,

$$\begin{aligned}
D_N^l &= \{\text{there exists } s, \hat{\tau}_s \leq \tau_{s-1}\}\cap \bar{C}_N, \\
D_N^m &= \{\text{for all } s, \tau_{s-1} < \hat{\tau}_s < \tau_{s+1}\}\cap \bar{C}_N, \\
D_N^r &= \{\text{there exists } s, \hat{\tau}_s \geq \tau_{s+1}\}\cap \bar{C}_N.
\end{aligned}$$

Then we have $P[A_{N,s}\cap \bar{C}_N] = P[A_{N,s}\cap D_N^l] + P[A_{N,s}\cap D_N^m] + P[A_{N,s}\cap D_N^r]$.

We first focus on $P[A_{N,s} \cap D_N^m]$ and consider the case where $\hat{\tau}_s \leq \tau_s$, since the case with $\tau_s \leq \hat{\tau}_s$ can be addressed in a similar way. Note that

$$P[A_{N,s} \cap D_N^m] \leq P[A_{N,s} \cap \{(\hat{\tau}_{s+1} - \tau_s) \geq \frac{\Delta_{min}}{2}\} \cap D_n^m] \\ + \sum_{g=s+1}^{S-1} P[\{(\tau_g - \hat{\tau}_g) \geq \frac{\Delta_{min}}{2}\} \cap \{(\hat{\tau}_{g+1} - \tau_g) \geq \frac{\Delta_{min}}{2}\} \cap D_N^m]. \tag{S-9}$$

We first bound the first term in (S-9). Using (S-2) twice with $k = \hat{\tau}_s$ and $k = \tau_s$ respectively, we have

$$\frac{||\boldsymbol{\beta}_j^s - \hat{\boldsymbol{\beta}}_j^{s+1}||_1}{p} \leq \frac{18p\lambda_1 + 9(\tau_s - \hat{\tau}_s)\sqrt{p}\lambda_2 + 9||\sum_{i=1}^N \sum_{k=\hat{\tau}_k}^{\tau_k - 1} \mathbf{Y}_{i(-j)}(t_k)e_{ij}(t_k)||_1}{N\phi_{s0}(\tau_s - \hat{\tau}_s)}, \tag{S-10}$$

with probability at least $1 - 2\exp(-\delta_N^n N/2 + 2\log N + 2\log n)$ Define $\bar{\tau}_s = \lfloor(\tau_s + \tau_{s+1})/2\rfloor$. Using (S-2) twice with $k = \bar{\tau}_s$ and $k = \tau_s$ respectively, we have

$$\frac{||\boldsymbol{\beta}_j^s - \boldsymbol{\beta}_j^{s+1}||_1}{p} \leq \frac{18p\lambda_1 + 9(\bar{\tau}_s - \tau_s)\sqrt{p}\lambda_2 + 9||\sum_{i=1}^N \sum_{k=\tau_k}^{\bar{\tau}_k - 1} \mathbf{Y}_{i(-j)}(t_k)e_{ij}(t_k)||_1}{N\phi_{s0}(\bar{\tau}_s - \tau_s)} \\ + 81\phi_{s0}^{-1}\phi_{s1}||\boldsymbol{\beta}_j^s - \hat{\boldsymbol{\beta}}_j^{s+1}||_1, \tag{S-11}$$

with probability at least $1 - c_1\exp(-n\delta_N^n N + 2\log n + 2\log N)$. Combining the two above inequalities and using Lemma 4, we can get the upper bound for the first term in (S-9) with

$$P[\xi_{min}\phi_{s0}Nn\delta_N^n \leq c_1p^2\lambda_1] + P[\xi_{min}\phi_{s0}N \leq c_2p^{1.5}\lambda_2] + P[\xi_{min}\sqrt{Nn\delta_N^n} \leq c_3p^2\sqrt{\log N}] \\ + c_4\exp(-n\delta_N^n N + 2\log n + 2\log N).$$

Under the conditions of the theorem, as long as $n\delta_N^n \to \infty$ as $n \to \infty$, all the terms converge to zero as $n \to \infty$. As long as $N\delta_N^n/\log(N) \to \infty$ as $N \to \infty$, all the terms converge to zero as $N \to \infty$. Using the similar way, we can prove that the other items in (S-9) converge to zero. Finally, we can conclude $P[A_{N,s} \cap D_N^m] \to 0$ as $N \to \infty$. As to $P[A_{N,s} \cap D_N^l]$, it has the upper bound:

$$P[D_N^l] \le \sum_{s=1}^{S} 2^{s-1} P[\{\max_{l=1,\dots,S} \hat{\tau}_l \le \tau_{l-1}\} = s] \tag{S-12}$$

$$\le 2^{S-1} \sum_{s=1}^{S-1} \sum_{l \le s} P[\{\tau_l - \hat{\tau}_l \ge \frac{\Delta_{min}}{2}\} \cap \{\hat{\tau}_{l+1} - \tau_l \ge \frac{\Delta_{min}}{2}\}]. \tag{S-13}$$

With the same arguments as those used to bound (S-9), we can prove that $P[D_N^l] \to 0$ and $P[D_N^r] \to 0$ as either $N \to \infty$ or $n \to \infty$. Consequently, we can show $P[A_{N,s} \cap \bar{C}_N] \to 0$.

B.3 Proof of Theorem 2

With Theorem 1 satisfied, we are working on the event

$$\mathcal{E} = \{\max_{s=1,\dots,S} |\hat{\tau}_s - \tau_s| \le n\delta_N^n\}.$$

Define $\mathcal{B}^s = [\tau_{s-1}, \tau_s)$ and the corresponding estimated $\hat{\mathcal{B}}^s = [\hat{\tau}_{s-1}, \hat{\tau}_s)$, then for $k \in \hat{\mathcal{B}}^s$, we have

$$Y_{ij}(t_k) = \mathbf{Y}_{i(-j)}(t_k)\boldsymbol{\beta}_j^s + w_{ij}(t_k) + e_{ij}(t_k),$$

where $w_{ij}(t_k) = \mathbf{Y}_{i(-j)}(t_k)(\mathbf{b}_j(t_k) - \boldsymbol{\beta}_j^s)$ is the bias. For $k \in \mathcal{B}^s \cap \hat{\mathcal{B}}^s$, the bias $w_{ij}(t_k) = 0$, otherwise the bias is distributed with zero mean and bounded variance under the Assumption *(A6)*. Because $\hat{\boldsymbol{\beta}}_j^s$ is an optimal solution of Equation (10) of the manuscript, it satisfies

$$\sum_{i=1}^{N} \mathbf{Y}_{i(-j)}^{\hat{\mathcal{B}}^s\prime} \mathbf{Y}_{i(-j)}^{\hat{\mathcal{B}}^s} (\hat{\boldsymbol{\beta}}_j^s - \boldsymbol{\beta}_j^s) - \sum_{i=1}^{N} \mathbf{Y}_{i(-j)}^{\hat{\mathcal{B}}^s\prime} (\mathbf{w}_{ij}^{\hat{\mathcal{B}}^s} + \mathbf{e}_{ij}^{\hat{\mathcal{B}}^s}) + \lambda_1(\hat{\mathbf{u}}_j(t_{\hat{\tau}_{s-1}}) - \hat{\mathbf{u}}_j(t_{\hat{\tau}_s})) + \lambda_2 |\hat{\mathcal{B}}^s| \hat{\mathbf{v}}_j(t_{\hat{\tau}_{s-1}}) = 0, \tag{S-14}$$

where $\mathbf{Y}_{i(-j)}^{\hat{\mathcal{B}}^s} \in \mathcal{R}^{|\hat{\mathcal{B}}^s| \times (p-1)}$ are the observations in the s^{th} estimated segment $\hat{\mathcal{B}}^s$. $\mathbf{w}_{ij}^{\hat{\mathcal{B}}^s}, \mathbf{e}_{ij}^{\hat{\mathcal{B}}^s} \in \mathcal{R}^{|\hat{\mathcal{B}}^s| \times 1}$ are the stacked $w_{ij}(t_k)$ and $e_{ij}(t_k)$ in $\hat{\mathcal{B}}^s$ respectively. $\hat{\mathbf{u}}_j(t_{\hat{\tau}_{s-1}}) \in \partial ||\hat{\boldsymbol{\beta}}_j^s - \hat{\boldsymbol{\beta}}_j^{s-1}||_1$, $\hat{\mathbf{u}}_j(t_{\hat{\tau}_s}) \in \partial ||\hat{\boldsymbol{\beta}}_j^{s+1} - \hat{\boldsymbol{\beta}}_j^s||_1$, and $\hat{\mathbf{v}}_j(t_{\hat{\tau}_{s-1}}) = \text{sign}(\hat{\boldsymbol{\beta}}_j^s)$.

Lemma 2. *Suppose $\hat{\mathbf{b}}_j$ is a solution of Equation (10) of the manuscript, with the associated segment points $\hat{T} = \{\hat{\tau}_s, s = 1, \ldots, S-1\}$. Suppose that the subgradient vectors satisfy $|\hat{v}_{jr}(t_k)| < 1$ for all $r \notin S(\hat{\mathbf{b}}_j(t_k))$ where $S(\hat{\mathbf{b}}_j(t_k))$ denotes the set of nonzero elements of $\hat{\mathbf{b}}_j(t_k)$. Then any other solution $\tilde{\mathbf{b}}_j$ with the same time segment points as $\hat{T}$ satisfies $\tilde{b}_{jr}(t_k) = 0$ for $r \notin \mathcal{S}(\hat{\mathbf{b}}_j(t_k))$.*

According to Lemma 2, we consider designing vectors $\check{\boldsymbol{\beta}}_j^s, \check{\mathbf{u}}_j(t_{\hat{\tau}_{s-1}}), \check{\mathbf{u}}_j(t_{\hat{\tau}_s})$ and $\check{\mathbf{v}}_j(t_{\hat{\tau}_{s-1}})$ with the associated $\hat{T}$ that satisfy (S-14) and $\check{\beta}_{jr}^s = 0$ for $r \notin S(\boldsymbol{\beta}_j^s)$. Then if we can verify the subdifferential vectors, $\check{\mathbf{u}}_j(t_{\hat{\tau}_{s-1}}), \check{\mathbf{u}}_j(t_{\hat{\tau}_s})$ and $\check{\mathbf{v}}_j(t_{\hat{\tau}_{s-1}})$ are dual feasible, we can prove that any other solution $\tilde{\boldsymbol{\beta}}_j^s$ with the same time segment points as $\hat{T}$ satisfies $\tilde{\beta}_{jr}^s = 0$ for $r \notin \mathcal{S}(\boldsymbol{\beta}_j^s)$.

Denote $M_j^s = \mathcal{S}(\boldsymbol{\beta}_j^s)$ as the set of functions that belong to the same subspace as $\mathbf{Y}_{ij}^{\mathcal{B}^s}$, and N_j^s as the set of functions that do not belong to M_j^s. If we assume $\mathbf{Y}_{ij}^{\mathcal{B}^s} \in \mathcal{X}_{l_s}^s$ with the cardinality p_l^s. We have the cardinality of M_j^s equals $p_l^s - 1$. Now we consider the following restricted optimization problem,

$$\min_{\substack{\boldsymbol{\beta}_j^s, s=1,\ldots,S \\ \boldsymbol{\beta}_{j,N_j^s}^s = 0}} \sum_{i=1}^{N} \sum_{s=1}^{S} ||\mathbf{Y}_{ij}^{\hat{\mathcal{B}}^s} - \mathbf{Y}_{i(-j)}^{\hat{\mathcal{B}}^s} \boldsymbol{\beta}_j^s||_2^2 + 2\lambda_1 \sum_{s=2}^{S} ||\boldsymbol{\beta}_j^s - \boldsymbol{\beta}_j^{s-1}||_1 + 2\lambda_2 \sum_{s=1}^{S} |\hat{\mathcal{B}}^s| ||\boldsymbol{\beta}_j^s||_1, \quad \text{(S-15)}$$

where the vector $\boldsymbol{\beta}_{j,N_j^s}^s$ is constrained to be $\mathbf{0}$. Let $\{\check{\boldsymbol{\beta}}_j^s\}$ be a solution to the restricted optimization problem (S-15). Its subgradient vectors are $\check{\mathbf{u}}_j(t_{\hat{\tau}_{s-1}}) \in \partial||\check{\boldsymbol{\beta}}_j^s - \check{\boldsymbol{\beta}}_j^{s-1}||_1$, $\check{\mathbf{u}}_j(t_{\hat{\tau}_s}) \in \partial||\check{\boldsymbol{\beta}}_j^{s+1} - \check{\boldsymbol{\beta}}_j^s||_1$, and $\check{\mathbf{v}}_j(t_{\hat{\tau}_{s-1}}) = \text{sign}(\check{\boldsymbol{\beta}}_j^s)$.

It is obvious that the vectors $\check{\boldsymbol{\beta}}_j^s, \check{\mathbf{u}}_j(t_{\hat{\tau}_{s-1}}), \check{\mathbf{u}}_j(t_{\hat{\tau}_s})$ and $\check{\mathbf{v}}_j(t_{\hat{\tau}_{s-1}})$ satisfy (S-14). Furthermore, $\check{\mathbf{u}}_j(t_{\hat{\tau}_{s-1}})$ and $\check{\mathbf{u}}_j(t_{\hat{\tau}_s})$ are elements of the subdifferential and hence dual feasible. To show $\check{\boldsymbol{\beta}}_j^s$ is also a solution to Equation (10) of the manuscript, we need to show that $\check{\mathbf{v}}_j(t_{\hat{\tau}_{s-1}})$ is also dual feasible, i.e., $||\check{\mathbf{v}}_{j,N_j^s}(t_{\hat{\tau}_{s-1}})||_\infty < 1$. Then any other solution $\tilde{\boldsymbol{\beta}}_j^s$ to Equation (10) of the manuscript will satisfy $\tilde{\boldsymbol{\beta}}_{j,N_j^s}^s = 0$.

Form (S-14), we obtain an explicit formula of $\boldsymbol{\beta}^s_{j,M_j^s}$,

$$\check{\boldsymbol{\beta}}^s_{j,M_j^s} = \boldsymbol{\beta}^s_{j,M_j^s} + \left(\sum_{i=1}^{N} \mathbf{Y}^{\hat{\mathcal{B}}^s\prime}_{i(-j),M_j^s}\mathbf{Y}^{\hat{\mathcal{B}}^s}_{i(-j),M_j^s}\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{Y}^{\hat{\mathcal{B}}^s\prime}_{i(-j),M_j^s}(\mathbf{w}^{\hat{\mathcal{B}}^s}_{ij,M_j^s} + \mathbf{e}^{\hat{\mathcal{B}}^s}_{ij,M_j^s}) \right.$$
$$\left. -\lambda_1(\check{\mathbf{u}}_{j,M_j^s}(t_{\hat{\tau}_{s-1}}) - \check{\mathbf{u}}_{j,M_j^s}(t_{\hat{\tau}_s})) - \lambda_2|\hat{\mathcal{B}}^s|\check{\mathbf{v}}_{j,M_j^s}(t_{\hat{\tau}_{s-1}})\right).$$

Then plug the above equation into (S-14), we have that $||\check{\mathbf{v}}_{j,N_j^s}(t_{\hat{\tau}_{s-1}})||_\infty < 1$, if $\max_{r\in N_j^s}|Y_r| < 1$ where

$$Y_r = \sum_{i=1}^{N} \mathbf{Y}^{\hat{\mathcal{B}}^s\prime}_{ir}\mathbf{Y}^{\hat{\mathcal{B}}^s}_{i(-j),M_j^s}\left(\sum_{i=1}^{N} \mathbf{Y}^{\hat{\mathcal{B}}^s\prime}_{i(-j),M_j^s}\mathbf{Y}^{\hat{\mathcal{B}}^s}_{i(-j),M_j^s}\right)^{-1}\left(\check{\mathbf{v}}_{j,M_j^s}(t_{\hat{\tau}_{s-1}}) + \frac{\lambda_1(\hat{\mathbf{u}}_{j,M_j^s}(t_{\hat{\tau}_{s-1}}) - \hat{\mathbf{u}}_{j,M_j^s}(t_{\hat{\tau}_s}))}{\lambda_2|\hat{\mathcal{B}}^s|}\right.$$
$$\left. - \sum_{i=1}^{N} \mathbf{Y}^{\hat{\mathcal{B}}^s\prime}_{i(-j),M_j^s}\frac{\mathbf{w}^{\hat{\mathcal{B}}^s}_{ij} + \mathbf{e}^{\hat{\mathcal{B}}^s}_{ij}}{\lambda_2|\hat{\mathcal{B}}^s|}\right) + \sum_{i=1}^{N} \mathbf{Y}^{\hat{\mathcal{B}}^s\prime}_{ir}\frac{\mathbf{w}^{\hat{\mathcal{B}}^s}_{ij} + \mathbf{e}^{\hat{\mathcal{B}}^s}_{ij}}{\lambda_2|\hat{\mathcal{B}}^s|} - \frac{\lambda_1(\check{\mathbf{u}}_{j,r}(t_{\hat{\tau}_{s-1}}) - \check{\mathbf{u}}_{j,r}(t_{\hat{\tau}_s}))}{\lambda_2|\hat{\mathcal{B}}^s|}. \tag{S-16}$$

As either $n \to \infty$ or $N \to \infty$, we have $\sum_{i=1}^{N} \mathbf{Y}^{\hat{\mathcal{B}}^s\prime}_{i(-j),M_j^s}\mathbf{Y}^{\hat{\mathcal{B}}^s}_{i(-j),M_j^s} \to N\frac{|\hat{\mathcal{B}}^s|}{n}(1+\sigma^2)\mathbf{I}$, and hence its inverse goes to $n(N|\hat{\mathcal{B}}^s|)^{-1}(1+\sigma^2)^{-1}\mathbf{I}$. As such, the first term of (S-16) has the upper bound

$$\frac{n}{N|\hat{\mathcal{B}}^s|(1+\sigma^2)}\sum_{i=1}^{N} ||\mathbf{Y}^{\hat{\mathcal{B}}^s\prime}_{ir}\mathbf{Y}^{\hat{\mathcal{B}}^s}_{i(-j),M_j^s}||_\infty\left(||\check{\mathbf{v}}_{j,M_j^s}(t_{\hat{\tau}_{s-1}})||_\infty + ||\frac{\lambda_1(\hat{\mathbf{u}}_{j,M_j^s}(t_{\hat{\tau}_{s-1}}) - \hat{\mathbf{u}}_{j,M_j^s}(t_{\hat{\tau}_s}))}{\lambda_2|\hat{\mathcal{B}}^s|}||_\infty\right.$$
$$\left. + ||\sum_{i=1}^{N} \mathbf{Y}^{\hat{\mathcal{B}}^s\prime}_{i(-j),M_j^s}\frac{\mathbf{w}^{\hat{\mathcal{B}}^s}_{ij} + \mathbf{e}^{\hat{\mathcal{B}}^s}_{ij}}{\lambda_2|\hat{\mathcal{B}}^s|}||_\infty\right).$$

We have

$$
\begin{aligned}
||\mathbf{Y}_{ir}^{\hat{\mathcal{B}}^s\prime}\mathbf{Y}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s}||_\infty &\le \sum_{g:\hat{\mathcal{B}}^s\cap\mathcal{B}^g\neq\emptyset} ||\mathbf{Y}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g\prime}\mathbf{Y}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_\infty \\
&\le \sum_{g:\hat{\mathcal{B}}^s\cap\mathcal{B}^g\neq\emptyset} ||\mathbf{X}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g\prime}\mathbf{X}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_\infty + \sum_{g:\hat{\mathcal{B}}^s\cap\mathcal{B}^g\neq\emptyset} ||\mathbf{X}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g\prime}\boldsymbol{\epsilon}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_\infty \\
&+ \sum_{g:\hat{\mathcal{B}}^s\cap\mathcal{B}^g\neq\emptyset} ||\boldsymbol{\epsilon}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g\prime}\mathbf{X}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_\infty + \sum_{g:\hat{\mathcal{B}}^s\cap\mathcal{B}^g\neq\emptyset} ||\boldsymbol{\epsilon}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g\prime}\boldsymbol{\epsilon}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_\infty.
\end{aligned}
$$

According to Lemmas 5 and 6, with probability at least $1-\frac{c_1}{(p_l^s)^2}$, we have the following inequalities,

$$
\begin{aligned}
||\mathbf{X}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g\prime}\mathbf{X}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_\infty &\le \frac{|\hat{\mathcal{B}}^s\cap\mathcal{B}^g|}{n}\sqrt{24}\log p_l^s \max_{r,l=1,\dots,L^g} \frac{\mathrm{aff}(\mathcal{S}_r^g,\mathcal{S}_l^g)}{\sqrt{\max(d_l^g,d_r^g)}}, \\
||\mathbf{X}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g\prime}\boldsymbol{\epsilon}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_\infty &\le \frac{|\hat{\mathcal{B}}^s\cap\mathcal{B}^g|}{n}2\sigma\sqrt{\frac{2\log|\hat{\mathcal{B}}^s\cap\mathcal{B}^g|}{|\hat{\mathcal{B}}^s\cap\mathcal{B}^g|}}, \\
||\boldsymbol{\epsilon}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g\prime}\mathbf{X}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_\infty &\le \sqrt{\frac{|\hat{\mathcal{B}}^s\cap\mathcal{B}^g|}{n}}\sqrt{24}\frac{\log p_l^s}{\sqrt{p_l^s}}||\boldsymbol{\epsilon}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_2, \\
||\boldsymbol{\epsilon}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g\prime}\boldsymbol{\epsilon}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_\infty &\le 2\sigma\sqrt{\frac{2\log|\hat{\mathcal{B}}^s\cap\mathcal{B}^g|}{n}}||\boldsymbol{\epsilon}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_2.
\end{aligned}
$$

According to Lemma 7, we have

$$
P\left[||\boldsymbol{\epsilon}_{ir}^{\hat{\mathcal{B}}^s\cap\mathcal{B}^g}||_2 \ge \sqrt{\frac{(1+\omega)\sigma^2}{n}|\hat{\mathcal{B}}^s\cap\mathcal{B}^g|}\right] \le \exp(-\frac{1}{8}|\hat{\mathcal{B}}^s\cap\mathcal{B}^g|\omega^2).
$$

As such, we have

$$
\begin{aligned}
\frac{n}{N|\hat{\mathcal{B}}^s|(1+\sigma^2)}\sum_{i=1}^N ||\mathbf{Y}_{ir}^{\hat{\mathcal{B}}^s\prime}\mathbf{Y}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s}||_\infty \le \frac{1}{1+\sigma^2}\Bigg(&\sqrt{24}\log p_l^s \max_{\substack{r,l=1,\dots,L^g\\ g:\hat{\mathcal{B}}^s\cap\mathcal{B}^g\neq\emptyset}} \frac{\mathrm{aff}(\mathcal{S}_r^g,\mathcal{S}_l^g)}{\sqrt{\max d_r^g, d_l^g}} + 2c_0\sigma \\
&+\sqrt{24}\frac{\log p_l^s}{\sqrt{p_l^s}}\sqrt{1+\omega}\sigma + 2c_0\sigma^2\sqrt{1+\omega}\Bigg),
\end{aligned}
$$

with probability $1-c_1/(p_l^s)^2 - c_2\exp(-\Delta_{min}\omega^2/8)$, under the event $\mathcal{E}$. (Here c_0 is smaller than 1). Furthermore, we have

$$||\check{\mathbf{v}}_{j,M_j^s}(t_{\hat{\tau}_{s-1}})||_\infty \leq 1,$$
$$||\frac{\lambda_1(\hat{\mathbf{u}}_{j,M_j^s}(t_{\hat{\tau}_{s-1}}) - \hat{\mathbf{u}}_{j,M_j^s}(t_{\hat{\tau}_s}))}{\lambda_2|\hat{\mathcal{B}}^s|}||_\infty \leq \frac{2\lambda_1}{\lambda_2(\hat{\tau}_s - \hat{\tau}_{s-1})}. \quad \text{(S-17)}$$

As long as $\lambda_1/(\lambda_2|\hat{\mathcal{B}}^s|) \to 0$ as either $n \to \infty$ or $N \to \infty$, we can bound (S-17) to 0.

At last, for the term of $|| \sum_{i=1}^N \mathbf{Y}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\prime} \frac{\mathbf{w}_{ij}^{\hat{\mathcal{B}}^s} + \mathbf{e}_{ij}^{\hat{\mathcal{B}}^s}}{\lambda_2|\hat{\mathcal{B}}^s|}||_\infty$, because $\mathbf{w}_{ij}(t_k) \neq 0$ only on $k \in \hat{\mathcal{B}}^s \setminus \mathcal{B}^s$ and $n\delta_N^n/|\hat{\mathcal{B}}^s| \to 0$ as either $n \to \infty$ or $N \to \infty$, the term involving $\mathbf{w}_{ij}^{\hat{\mathcal{B}}^s}$ is stochastically dominated by the term involving $\mathbf{e}_{ij}^{\hat{\mathcal{B}}^s}$ and can be ignored. Consequently, we need to bound $|| \sum_{i=1}^N \mathbf{Y}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s\prime} \frac{\mathbf{e}_{ij}^{\hat{\mathcal{B}}^s}}{\lambda_2|\hat{\mathcal{B}}^s|}||_\infty$. According to Lemma 4, we have

$$P\left[|| \sum_{i=1}^N \mathbf{Y}_{i(-j),M_j^s}^{\hat{\mathcal{B}}^s} \frac{\mathbf{e}_{ij}^{\hat{\mathcal{B}}^s}}{\lambda_2|\hat{\mathcal{B}}^s|}||_\infty \geq c_1 \frac{\sqrt{N(1 + c\log N + c\log n)}}{\lambda_2\sqrt{|\hat{\mathcal{B}}^s|}}\right] \leq c_1 \exp(-c\log N - c\log n).$$

As $\lambda_2/\sqrt{N\log N} \to \infty$, we have the above item bounded to 0 as $N \to \infty$. As $n \to \infty$, $|\hat{\mathcal{B}}^s| \to O(n + n\delta_N^n)$, as long as $\lambda_2\sqrt{|\hat{\mathcal{B}}^s|/\log n} \to \infty$, we can bond the item.

Similarly, we can bound the second item in (S-16) to 0. Consequently, we have that as long as there is a positive constant ω that satisfies

$$\sqrt{1+\omega} \leq \frac{1 + \sigma^2 - \sqrt{24}\log p_l^s(\max_{\substack{r,l=1,\ldots,L^g \\ g=1,\ldots,S}} \frac{\text{aff}(\mathcal{S}_r^g, \mathcal{S}_l^g)}{\sqrt{\max(d_r^g, d_l^g)}}) - 2c_0\sigma}{\sqrt{24}\frac{\log p_l^s}{\sqrt{p_l^s}}\sigma + 2c_0\sigma^2}, \quad \text{(S-18)}$$

where $0 < c_0 < 1$. we have with probability at least $1 - c_1/(p_l^s)^2 - c_2\exp(-\Delta_{min}\omega^2/8)$, $\max_{r\in N_j^s}|Y_r| < 1$ and $\check{\boldsymbol{\beta}}_{j,N_j^s}^s = 0$. Consequently, we can prove that any solution $\tilde{\boldsymbol{\beta}}_j^s$ with the same segment points $\check{T}$ satisfies $\tilde{\beta}_{jr}^s = 0$ for $r \notin \mathcal{X}_l^s$.

B.4 Other Useful Results

Here we show some additional results that are useful for the derivations in Appendices B.2 and B.3.

Denote the functional data in the s^{th} time segment as $\mathbf{Y}_i^s = \left[\mathbf{Y}_{i(1)}^s, \ldots, \mathbf{Y}_{i(L^s)}^s\right]$ with the l^{th} subspace data $\mathbf{Y}_{i(l)}^s \in \mathcal{R}^{n_s \times p_l^s}$. Denote $\mathbf{\Phi}^s = \left[\mathbf{\Phi}_{(1)}^s, \ldots, \mathbf{\Phi}_{(L^s)}^s\right], l = 1, \ldots, L^s$, $\mathbf{A}_i^s = \text{diag}(\mathbf{A}_{i(1)}^s, \ldots, \mathbf{A}_{i(L^s)}^s)$ and $\mathbf{E}_i^s = \left[\epsilon_{i1}^s, \ldots, \epsilon_{ip}^s\right]$. Then we generate

$$\mathbf{Y}_i^s = \mathbf{X}_i^s + \mathbf{E}_i^s,$$
$$\mathbf{X}_i^s = \mathbf{\Phi}^s \mathbf{A}_i^s,$$

for $k = \tau_{s-1}, \ldots, \tau_s - 1$. $\mathbf{E}_{i(l)}^s = \left[\boldsymbol{\epsilon}_{i1}^s, \ldots, \boldsymbol{\epsilon}_{ip_l^s}^s\right]$ is the noise function, with $\boldsymbol{\epsilon}_{ij}^s \sim N_{n_s}(\mathbf{0}, \sigma^2/n\mathbf{I})(j = 1, \ldots, p)$. Denote

$$\mathbf{K}^s = \mathrm{E}(\mathbf{Y}_i^{s'} \mathbf{Y}_i^s) = \mathrm{E}(\mathbf{A}_i^{s'} \mathbf{\Phi}^{s'} \mathbf{\Phi}_k^s \mathbf{A}_i^s) + \frac{n_s}{n}\sigma^2 \mathbf{I}. \tag{S-19}$$

When every column of $\mathbf{A}_{il}^s$ is sampled at random from the unit sphere of $\mathcal{R}^{d_l \times 1}$, we have $\mathbf{K}^s = \frac{n_s}{n}(1+\sigma^2)\mathbf{I}$. Define $\phi_{s1} = \Lambda_{max}(\mathbf{K}^s)$ and $\phi_{s0} = \Lambda_{min}(\mathbf{K}^s)$. Let $\hat{\mathbf{K}}^s = N^{-1}(m-l+1)^{-1}\sum_{i=1}^N \sum_{k=l}^m \mathbf{Y}_i(t_k)' \mathbf{Y}_i(t_k)$ be the empirical estimation of $\mathbf{K}^s$ with $\tau_{s-1} \le l < m < \tau_s$. There are two crude bounds for the eigenvalues of $\hat{\mathbf{K}}^s$

$$P(\Lambda_{\max}(\hat{\mathbf{K}}^s) \ge 9\phi_{s1}) \le 2\exp(-N(m-l+1)/2),$$
$$P(\Lambda_{\min}(\hat{\mathbf{K}}^s) \le \phi_{s0}/9) \le 2\exp(-N(m-l+1)/2).$$

Lemma 3. *For any $v_n N > p$, we have*

$$P\left[\max_{\substack{\tau_{s-1} \le l < m < \tau_s \\ m-l > v_n}} \Lambda_{max}\left(\frac{\sum_{i=1}^N \sum_{k=l}^m \mathbf{Y}_i^{s'}(t_k) \mathbf{Y}_i^s(t_k)}{N(m-l+1)}\right) \ge 9\phi_{s1}\right] \le \exp(-v_n N/2 + 2\log(\tau_s - \tau_{s-1}) + 2\log(N)),$$

and

$$P\left[\max_{\substack{\tau_{s-1}\le l<m<\tau_s\\ m-l>v_n}} \Lambda_{min}\left(\frac{\sum_{i=1}^{N}\sum_{k=l}^{m}\mathbf{Y}_i^{s'}(t_k)\mathbf{Y}_i^{s}(t_k)}{N(m-l+1)}\right)\le \phi_{s0}/9\right]\le \exp(-v_nN/2+2\log(\tau_s-\tau_{s-1})+2\log(N)).$$

Lemma 4. *Recall* $e_{ij}(t_k)=\sum_{r\neq j}\epsilon_{ir}(t_k)b_{jr}(t_k)-\epsilon_{ij}(t_k)$*, we know* $e_{ij}(t_k)$ *is also normal distributed with mean* 0 *and bounded variance, denoted as* γ^2*. Then if* $v\ge C\log(N(\tau_s-\tau_{s-1}))$ *for some constant* $C>16$,

$$P\Bigg[\bigcap_{s=1,\dots,S}\bigcap_{\substack{\tau_{s-1}\le l<m<\tau_s\\ m-l>v}}\{\frac{1}{N(m-l+1)}||\sum_{i=1}^{N}\sum_{k=l}^{m}\mathbf{Y}_i(t_k)e_{ij}(t_k)||_1\le\frac{p\gamma\phi_1^{1/2}\sqrt{1+C}}{\sqrt{N(m-l+1)}} \tag{S-20}$$

$$\sqrt{1+C\log N+C\log(\tau_s-\tau_{s-1})}\,\}\Bigg]\ge 1-c_1\exp(-c_2\log N-c_2\log(\tau_s-\tau_{s-1})),$$

for some constants $c_1,c_2>0$*. Here* $\phi_1=\max_s\phi_{s1}$*.*

B.5 Standard Inequalities in Probability

Lemma 5. *Suppose the unit-norm vector* $\mathbf{x}_{ij}\in\mathcal{R}^{n\times 1}$ *is uniformly drawn at random from* $\mathcal{S}_r$*, and* $\mathbf{X}_{i(l)}\in\mathcal{R}^{n\times p_l}$ *are* p_l *unit-norm vectors drawn uniformly at random from* $\mathcal{S}_l$*. Then we have*

$$||\mathbf{X}_{i(l)}'\mathbf{x}_{ij}||_\infty\le\sqrt{\log a\log b}\frac{\sqrt{d_l\wedge d_r}\,\mathit{aff}(\mathcal{S}_l,\mathcal{S}_r)}{\sqrt{d_ld_r}}, \tag{S-21}$$

with probability at least $1-\frac{2}{\sqrt{a}}-\frac{2p_l}{\sqrt{b}}$*.*

Lemma 6. *Suppose* $\mathbf{Z}\in\mathcal{R}^{n\times p}$ *has iid* $N(0,1)$ *entries and let* $\mathbf{x}\in\mathcal{R}^{n\times 1}$ *be a unit-norm vector. Then*

$$||\mathbf{Z}'\mathbf{x}||_\infty\le 2\sqrt{2\log n},$$

with probability at least $1 - 2/p^2$.

Lemma 7. *For a* χ_n^2 *distribution with* n *degrees of freedom, it obeys*

$$P[\chi_n^2 \geq (1+\omega)n] \leq \exp(-n\omega^2/8).$$

References

Harchaoui, Z. and Lévy-Leduc, C. (2010), "Multiple change-point estimation with a total variation penalty," *Journal of the American Statistical Association*, 105, 1480–1493.

Kolar, M. and Xing, E. P. (2012), "Estimating networks with jumps," *Electronic journal of statistics*, 6, 2069.

Wang, Y.-X. and Xu, H. (2013), "Noisy sparse subspace clustering," in *Proc. Int. Conf. Mach. Learn.*, pp. 1–8.

— (2016), "Noisy sparse subspace clustering," *The Journal of Machine Learning Research*, 17, 320–360.