Spectral Clustering for Functional Data with Asymptotic Properties

Yang Ren[1], Jinguan Lin[1], Peijun Sang[2] and Jiangyan Wang[1,*]

1.Nanjing Audit University; 2.University of Waterloo

Supplementary Materials

The Supplementary Materials present more extensive numerical results, along with supporting lemmas and the proofs of the main results. Lemma 1 to Lemma 3 are preliminaries to prove Theorem 2.1, which shows the uniform convergence of the clustering coefficients. Lemma 4 shows both the original and alternative forms of the loss function $\mathcal{N}_{\text{cut}}$. Lemma 5 focuses on the convergence of the estimated matrix $\widehat{\boldsymbol{V}}$. Lemma 6 discusses the convergence of clusters obtained by TSFDSC. Based on Lemmas 4 - 6, one obtains the convergence of the loss function $\mathcal{N}_{\text{cut}}$. Consequently, Theorem 2.2 proves that the loss function of our proposed clustering algorithm converges to that of the optimal clustering.

S1 Proof of Theorem 2.1

Lemma 1 is the Bernstein's inequality used to prove technical Lemmas 2 $\sim$ 3.

Lemma 1. *Let $\varsigma_1, \varsigma_2, \cdots, \varsigma_n$ be independent real valued random variables with* $\mathrm{E}\,\varsigma_i = 0, \mathrm{E}\,\varsigma_i^2 = \sigma_i^2 < +\infty$, $i = 1, 2, \cdots, n$ *and let* $M_n = \sum_{i=1}^n \varsigma_i$,

$S_n^2 = \sum_{i=1}^n \sigma_i^2$. *If there exists* $c_0 > 0$ *such that for* $r \geq 3$,

$$\mathrm{E}\left|\varsigma_i\right|^r \leq c_0^{r-2} r! \sigma_i^2 < \infty, i = 1, 2, \cdots, n,$$

then for each $n > 1, t > 0$,

$$\mathrm{P}\left(\left|\frac{M_n}{S_n}\right| \geq t\right) \leq 2\exp\left\{-\frac{t^2}{4}\frac{1}{1 + c_0 t S_n^{-1}/2}\right\}.$$

Lemma 2. *Denote* $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \cdots, \epsilon_{im_i})^\top$, *as* $m \to +\infty$,

$$\left\|m_i^{-1}\boldsymbol{X}_i^\top \boldsymbol{\epsilon}_i\right\|_\infty = \max_{1 \leq J \leq N_s + p}\left|\frac{1}{m_i}\sum_{j=1}^{m_i} B_{J,p}\left(\frac{j}{m_i}\right)\epsilon_{ij}\right| = O_p\left(\sqrt{m^{-1}\log m}\right),$$

in which $m = \min\{m_1, m_2, \cdots, m_n\}$.

Proof of Lemma 2. Note that ϵ_{ij} are iid variables with $\mathrm{E}(\epsilon_{ij}) = 0$ and $\mathrm{E}(\epsilon_{ij}^2) = \sigma^2$, denote variables $\varsigma_{iJj} = m_i^{-1} B_{J,p}(j/m_i)\epsilon_{ij}$. Then, for all $1 \leq J \leq N_s + p$,

$$\begin{aligned}
&\mathrm{E}\,\varsigma_{iJj} = 0,\\
&\mathrm{E}\,\varsigma_{iJj}^2 = \frac{\sigma^2}{m_i^2} B_{J,p}^2\left(\frac{j}{m_i}\right),\\
&M_{m_i} = \frac{1}{m_i}\sum_{j=1}^{m_i} B_{J,p}\left(\frac{j}{m_i}\right)\epsilon_{ij},\\
&S_{m_i}^2 = \frac{\sigma^2}{m_i^2}\sum_{j=1}^{m_i} B_{J,p}^2\left(\frac{j}{m_i}\right) \sim m_i^{-1}.
\end{aligned}$$

Clearly, $c_0 = 2m_i^{-1/2}$ satisfies the moment condition in Lemma 1, thus we have

$$\begin{aligned}
\mathrm{P}\left(\left|\frac{M_{m_i}}{m^{-1/2}}\right| > \sqrt{\log m}\right) &\leq \mathrm{P}\left(\left|\frac{M_{m_i}}{S_{m_i}}\right| > \sqrt{\log m_i}\right) \\
&\leq 2\exp\left\{-\frac{\log m_i}{4}\frac{1}{1+m_i^{-1/2}\sqrt{\log m_i}{m_i}^{1/2}}\right\} \\
&= 2\exp\left\{-\frac{\log m_i}{4(1+\sqrt{\log m_i})}\right\},
\end{aligned}$$

that is to say,

$$\left\|m_i^{-1}\boldsymbol{X}_i^\top\boldsymbol{\epsilon}_i\right\|_\infty = O_p\left(\sqrt{m^{-1}\log m}\right).$$

Lemma 3. *Let* $\widetilde{\boldsymbol{\epsilon}}_i = \left(\boldsymbol{X}_i^\top\boldsymbol{X}_i\right)^{-1}\boldsymbol{X}_i^\top\boldsymbol{\epsilon}_i$*, as* $m \to +\infty$*,*

$$\|\widetilde{\boldsymbol{\epsilon}}_i\|_\infty = O_p\left(N_s\sqrt{m^{-1}\log m}\right).$$

Proof of Lemma 3. According to Burman (1991), $\left\|\left(\boldsymbol{X}_i^\top\boldsymbol{X}_i\right)^{-1}\right\|_\infty = O\left(N_s m_i^{-1}\right)$,

$$\begin{aligned}
\|\widetilde{\boldsymbol{\epsilon}}_i\|_\infty &= \left\|\left(\boldsymbol{X}_i^\top\boldsymbol{X}_i\right)^{-1}\boldsymbol{X}_i^\top\boldsymbol{\epsilon}_i\right\|_\infty \\
&= \left\|\left(m_i^{-1}\boldsymbol{X}_i^\top\boldsymbol{X}_i\right)^{-1}m_i^{-1}\boldsymbol{X}_i^\top\boldsymbol{\epsilon}_i\right\|_\infty \\
&\leq \left\|\left(m_i^{-1}\boldsymbol{X}_i^\top\boldsymbol{X}_i\right)^{-1}\right\|_\infty\left\|m_i^{-1}\boldsymbol{X}_i^\top\boldsymbol{\epsilon}_i\right\|_\infty \\
&= O_p\left(N_s\sqrt{m^{-1}\log m}\right).
\end{aligned}$$

Then we prove Theorem 2.1 in the following.

Proof of Theorem 2.1. Acording to Lemma 3,

$$\begin{aligned}\max_{1\le i\le n}\|\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\| &\le \sqrt{N_s+p}\max_{1\le i\le n}\|\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\|_\infty\\ &= \sqrt{N_s+p}\max_{1\le i\le n}\|\widetilde{\boldsymbol{\epsilon}}_i\|_\infty\\ &= O_p\left(\sqrt{N_s^3 m^{-1}\log m}\right).\end{aligned}$$

S2 Proof of Theorem 2.2

Lemma 4 shows both the original and alternative forms of the loss function $\mathcal{N}_{\mathrm{cut}}$. Lemma 5 focuses on the convergence of the matrix $\boldsymbol{V}$. Lemma 6 discusses the convergence of clusters obtained by TSFDSC. Ultimately, the convergence of the loss function $\mathcal{N}_{\mathrm{cut}}$ is obtained.

Lemma 4. *For NCut, the object is to*

$$\min \mathcal{N}_{\mathrm{cut}}(SC_1, SC_2,\cdots,SC_Q)=\min\sum_{q=1}^{Q}\frac{\mathrm{cut}(SC_q,\overline{SC_q})}{\mathrm{vol}(SC_q)},$$

which can be rewritten as

$$\min_{SC_1,\cdots,SC_Q}\mathrm{tr}(\boldsymbol{U}^\top\boldsymbol{L}\boldsymbol{U}),\, s.t., \boldsymbol{U}^\top\boldsymbol{D}\boldsymbol{U}=\boldsymbol{I}, \tag{S2.1}$$

where $\boldsymbol{U}$ is a binary matrix. Denote $\boldsymbol{V}=\boldsymbol{D}^{1/2}\boldsymbol{U}$ and then relax the limitation that $\boldsymbol{U}$ is a binary matrix, (S2.1) is relaxed to

$$\min_{\boldsymbol{V}\in\mathbb{R}^{n\times Q}}\mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}\boldsymbol{V}),\, s.t., \boldsymbol{V}^\top\boldsymbol{V}=\boldsymbol{I}, \tag{S2.2}$$

(S2.2) holds when $\boldsymbol{V}$ *is formed by the first* Q *eigenvectors* $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_Q$ *of* $\boldsymbol{L}_{norm}$.

Proof of Lemma 4. Define vectors $\boldsymbol{u}_q = (u_{1q}, u_{2q}, \cdots, u_{nq})^\top$, $q = 1, 2, \cdots, Q$, where

$$u_{iq} = \begin{cases} \frac{1}{\sqrt{\mathrm{vol}(SC_q)}}, & \boldsymbol{\beta}_i \in SC_q \\ 0, & \text{else} \end{cases}.$$

Recall the decomposition, one has

$$\begin{aligned}
\boldsymbol{u}_q^\top \boldsymbol{L} \boldsymbol{u}_q &= \boldsymbol{u}_q^\top \boldsymbol{D} \boldsymbol{u}_q - \boldsymbol{u}_q^\top \boldsymbol{W} \boldsymbol{u}_q \\
&= \sum_{i=1}^n u_{iq}^2 d_i - \sum_{i,j} \omega_{ij} u_{iq} u_{jq} \\
&= \frac{1}{2} \left(\sum_{i=1}^n u_{iq}^2 d_i + \sum_{j=1}^n u_{jq}^2 d_j - 2 \sum_{i=1}^n \sum_{j=1}^n u_{iq} u_{jq} \omega_{ij} \right) \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} (u_{iq} - u_{jq})^2 \\
&= \frac{1}{2\,\mathrm{vol}(SC_q)} \left(\sum_{\boldsymbol{\beta}_i \in SC_q, \boldsymbol{\beta}_j \in \overline{SC_q}} \omega_{ij} u_{iq}^2 + \sum_{\boldsymbol{\beta}_j \in SC_q, \boldsymbol{\beta}_i \in \overline{SC_q}} \omega_{ij} u_{jq}^2 \right) \\
&= \frac{\mathrm{cut}(SC_q, \overline{SC_q})}{\mathrm{vol}(SC_q)}.
\end{aligned}$$

Consequently, the followings hold:

$$
\begin{aligned}
\mathcal{N}_{\text{cut}}(SC_1, SC_2, \cdots, SC_Q) &= \sum_{q=1}^{Q} \frac{\text{cut}(SC_q, \overline{SC_q})}{\text{vol}(SC_q)} \\
&= \sum_{q=1}^{Q} \boldsymbol{u}_q^\top \boldsymbol{L} \boldsymbol{u}_q \\
&= \sum_{q=1}^{Q} (\boldsymbol{U}^\top \boldsymbol{L} \boldsymbol{U})_{qq} \\
&= \text{tr}(\boldsymbol{U}^\top \boldsymbol{L} \boldsymbol{U})
\end{aligned}
$$

where $\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_Q) \in \mathbb{R}^{n \times Q}$.

Notice that

$$
\boldsymbol{u}_q^\top \boldsymbol{D} \boldsymbol{u}_q = \sum_{i=1}^{n} u_{iq}^2 d_i = \frac{1}{\text{vol}(SC_q)} \sum_{\boldsymbol{\beta}_i \in SC_q} d_i = 1,
$$

one has $\boldsymbol{U}^\top \boldsymbol{D} \boldsymbol{U} = \boldsymbol{I}$. Therefore, the NCut problem can be rewritten as

$$
\min_{SC_1, \cdots, SC_Q} \text{tr}(\boldsymbol{U}^\top \boldsymbol{L} \boldsymbol{U}), \text{s.t.,} \boldsymbol{U}^\top \boldsymbol{D} \boldsymbol{U} = \boldsymbol{I}.
$$

Recall that $\boldsymbol{u}_q$ is a binary vector with 2^n possible values, making it an NP-hard problem. By allowing u_{iq} to take any real values, the problem can be relaxed to

$$
\min_{\boldsymbol{U} \in \mathbb{R}^{n \times Q}} \text{tr}(\boldsymbol{U}^\top \boldsymbol{L} \boldsymbol{U}), \text{s.t.,} \boldsymbol{U}^\top \boldsymbol{D} \boldsymbol{U} = \boldsymbol{I}.
$$

Since $\boldsymbol{V} = \boldsymbol{D}^{1/2} \boldsymbol{U}$, we have

$$
\min_{\boldsymbol{V} \in \mathbb{R}^{n \times Q}} \text{tr}(\boldsymbol{V}^\top \boldsymbol{D}^{-1/2} \boldsymbol{L} \boldsymbol{D}^{-1/2} \boldsymbol{V}), \text{s.t.,} \boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}.
$$

To apply the Lagrange Multiplier, one defines the objective function as follows:

$$\begin{aligned} f(\boldsymbol{U}) &= \mathrm{tr}(\boldsymbol{U}^\top \boldsymbol{L}\boldsymbol{U}) - \sum_{q=1}^{Q} \widetilde{\lambda}_q(\boldsymbol{u}_q^\top \boldsymbol{D}\boldsymbol{u}_q - 1) \\ &= \mathrm{tr}(\boldsymbol{U}^\top \boldsymbol{L}\boldsymbol{U}) - \mathrm{tr}\left(\boldsymbol{\Lambda}(\boldsymbol{U}^\top \boldsymbol{D}\boldsymbol{U} - \boldsymbol{I})\right), \end{aligned}$$

where $\boldsymbol{\Lambda} = diag\{\widetilde{\lambda}_1, \widetilde{\lambda}_2, \cdots, \widetilde{\lambda}_Q\}$. Setting the partial derivative of the objective function to zero yields

$$\begin{aligned} \frac{\partial f}{\partial \boldsymbol{U}} &= \boldsymbol{L}\boldsymbol{U} + \boldsymbol{L}^\top \boldsymbol{U} - \boldsymbol{D}^\top \boldsymbol{U}\boldsymbol{\Lambda}^\top - \boldsymbol{D}\boldsymbol{U}\boldsymbol{\Lambda} \\ &= 2\boldsymbol{L}\boldsymbol{U} - 2\boldsymbol{D}\boldsymbol{U}\boldsymbol{\Lambda} = 0, \end{aligned}$$

one has $\boldsymbol{L}\boldsymbol{U} = \boldsymbol{D}\boldsymbol{U}\boldsymbol{\Lambda}$, i.e.,

$$\boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda}, \tag{S2.3}$$

premultiplying $\boldsymbol{V}^\top$, one obtains

$$\boldsymbol{V}^\top \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}\boldsymbol{V} = \boldsymbol{V}^\top \boldsymbol{V}\boldsymbol{\Lambda},$$

as $\boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}$, we have

$$\min \mathrm{tr}(\boldsymbol{V}^\top \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}\boldsymbol{V}) = \min \mathrm{tr}(\boldsymbol{\Lambda}),$$

According to (S2.3), each $\widetilde{\lambda}_q$ is an eigenvalue of $\boldsymbol{L}_{\mathrm{norm}}$, and the q-th column of $\boldsymbol{V}$ is its corresponding eigenvector. To minimize $\mathrm{tr}(\boldsymbol{V}^\top \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}\boldsymbol{V})$, one needs to find the first Q eigenvalues of $\boldsymbol{L}_{\mathrm{norm}}$, and $\boldsymbol{V}$ is constructed from the corresponding eigenvectors.

Remark 1. The k-means algorithm applied in Algorithm 1 is used to map the solution of the relaxed linear programming (S2.2) to the solution of the original problem.

Let $\widehat{\boldsymbol{W}}, \widehat{\boldsymbol{D}}, \widehat{\boldsymbol{L}}, \widehat{\boldsymbol{L}}_{\text{norm}}$ be the weighted adjacency matrix, degree matrix, Laplacian matrix, and normalized Laplacian matrix estimated based on $\{\widehat{\boldsymbol{\beta}}_i, i = 1, 2, \cdots, n\}$, respectively. Let $\widehat{\boldsymbol{V}}$ be formed by the first k eigenvectors of $\widehat{\boldsymbol{L}}_{\text{norm}}$. Similarly, let $\boldsymbol{W}, \boldsymbol{D}, \boldsymbol{L}, \boldsymbol{L}_{\text{norm}}, \boldsymbol{V}$ be the corresponding matrices computed by $\{\boldsymbol{\beta}_i, i = 1, 2, \cdots, n\}$. Denote $\{\widehat{SC}^{\star}_q, q = 1, 2, \cdots, Q\}$ as the spectral clusters obtained by Algorithm 1. Furthermore, denote the corresponding k-means clusters based on $\widehat{\boldsymbol{V}}$ as $\{\widehat{C}^{\star}_q, q = 1, 2, \cdots, Q\}$. The following lemma shows the convergence of $\widehat{\boldsymbol{V}}$.

Lemma 5. *Under Assumptions 2.1-2.5, $\widehat{\boldsymbol{V}}$ converges strongly to $\boldsymbol{V}$ as $m \to \infty$, in which $m = \min\{m_1, m_2, \cdots, m_n\}$.*

Proof of Lemma 5. Since we use the Gaussian kernel function as the radial basis function to compute the edge weight, i.e.,

$$\widehat{\omega}_{ij} = \exp(-\|\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_j\|^2/2\iota^2).$$

Thus, one has

$$
\begin{aligned}
|\omega_{ij} - \widehat{\omega}_{ij}| &= \left|\exp(-\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2/2\iota^2) - \exp(-\|\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_j\|^2/2\iota^2)\right| \\
&= \left|\exp\left(\frac{-\|\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_j\|^2}{2\iota^2}\right)\left\{\exp\left(\frac{\|\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_j\|^2 - \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2}{2\iota^2}\right) - 1\right\}\right| \\
&\leq \left|\exp\left(\frac{\|\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_j\|^2 - \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2}{2\iota^2}\right) - 1\right|.
\end{aligned}
$$

WLOG, assume $\|\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_j\|^2 \geq \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2$, notice that

$$
\begin{aligned}
\|\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_j\|^2 - \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2 &= \|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i + \boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j + \boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2 - \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2 \\
&\leq \left(\|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\| + \|\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j\| + \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|\right)^2 - \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2 \\
&= \|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\|^2 + \|\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j\|^2 + 2\|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\|\|\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j\| \\
&\quad + 2\|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\|\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\| + 2\|\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j\|\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|,
\end{aligned}
$$

according to Assumption 2.4, we have $\exp\{-\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2/2\iota^2\} > c$, thus $\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\| < \sqrt{-2\iota^2 \log c}$. As $\|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\|^2 = O_p\left(N_s^3 m^{-1} \log m\right)$, then, $\|\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_j\|^2 - \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2 = O_p\left(\sqrt{N_s^3 m^{-1} \log m}\right)$, it is evident that for a given ι^2 and constant c, $|\widehat{\omega}_{ij} - \omega_{ij}| = O_p\left(\sqrt{N_s^3 m^{-1} \log m}\right)$. Therefore,

$$
\begin{aligned}
\left\|\widehat{\boldsymbol{W}} - \boldsymbol{W}\right\|_F^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}(\widehat{\omega}_{ij} - \omega_{ij})^2 \\
&\leq n^2 \max_i O_p\left(N_s^3 m^{-1} \log m\right) \\
&= n^2 O_p\left(N_s^3 m^{-1} \log m\right) \\
&= O_p\left(N_s^3 n^2 m^{-1} \log m\right),
\end{aligned}
$$

according to Assumption 2.5, we have $N_s^3 n^2 m^{-1} \log m \to 0$. Thus $\widehat{\boldsymbol{W}}$ converges strongly to $\boldsymbol{W}$, similarly, $\widehat{\boldsymbol{D}}$, $\widehat{\boldsymbol{L}}$, $\widehat{\boldsymbol{L}}_{\text{norm}}$ converges strongly to $\boldsymbol{D}$, $\boldsymbol{L}$, $\boldsymbol{L}_{\text{norm}}$ respectively. Consequently, $\widehat{\boldsymbol{V}}$ converges strongly to $\boldsymbol{V}$ as $m \to \infty$.

Lemma 6. *Denote h the Hausdorff metric, $\widehat{\alpha}^\star = \{\widehat{\boldsymbol{c}}_1^\star, \widehat{\boldsymbol{c}}_2^\star, \cdots, \widehat{\boldsymbol{c}}_Q^\star\}$ where $\widehat{\boldsymbol{c}}_q^\star$ is the center of cluster $\widehat{C}_q^\star$, $\alpha^\star = \{\boldsymbol{c}_1^\star, \boldsymbol{c}_2^\star, \cdots, \boldsymbol{c}_Q^\star\}$ where $\boldsymbol{c}_q^\star$ is the center of cluster $C_q^\star$. We have $\lim_{n\to\infty} h(\widehat{\alpha}^\star, \alpha^\star) = 0$, a.s., while for each q, there exists a unique q' such that $\lim_{n\to\infty} h(\widehat{C}_q^\star, C_{q'}^\star) = 0$ a.s.*

Proof of Lemma 6: Denote $\widehat{\boldsymbol{V}} = (\widehat{\boldsymbol{v}}_1, \cdots, \widehat{\boldsymbol{v}}_Q) = (\widehat{\boldsymbol{r}}_1, \cdots, \widehat{\boldsymbol{r}}_n)^\top$, $\boldsymbol{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_Q) = (\boldsymbol{r}_1, \cdots, \boldsymbol{r}_n)^\top$. For $\widehat{\alpha}^\star$ and $\{\widehat{C}_q^\star\}$, it minimizes

$$\frac{1}{n}\sum_{q=1}^{Q}\sum_{\widehat{\boldsymbol{r}}_i \in \widehat{C}_q} \|\widehat{\boldsymbol{r}}_i - \widehat{\boldsymbol{c}}_q\|^2. \tag{S2.4}$$

For $\alpha^\star$ and $\{C_q^\star\}$, it minimizes

$$\frac{1}{n}\sum_{q=1}^{Q}\sum_{\boldsymbol{r}_i \in C_q} \|\boldsymbol{r}_i - \boldsymbol{c}_q\|^2. \tag{S2.5}$$

According to Lemma 5, it is clear that $\widehat{\boldsymbol{r}}_i$ converges strongly to $\boldsymbol{r}_i$, combining Theorem 1 proposed by Abraham et al. (2003), we have $\lim_{n\to\infty} h(\widehat{\alpha}^\star, \alpha^\star) = 0$ a.s., furthermore, for each q, there exists a unique q', we have $\lim_{n\to\infty} h(\widehat{C}_q^\star, C_{q'}^\star) = 0$ a.s. Finally, one proves Theorem 2.2 in the following.

Proof of Theorem 2.2 According to Lemma4,

$$\begin{aligned}
&\left|\mathcal{N}_{\text{cut}}(\widehat{SC}^{\star}_1, \widehat{SC}^{\star}_2, \cdots, \widehat{SC}^{\star}_Q) - \mathcal{N}_{\text{cut}}(SC^{\star}_1, SC^{\star}_2, \cdots, SC^{\star}_Q)\right| \\
&= \left|\sum_{q=1}^{Q} \frac{\text{cut}(\widehat{SC}^{\star}_q, \overline{\widehat{SC}^{\star}_q})}{\text{vol}(\widehat{SC}^{\star}_q)} - \sum_{q=1}^{Q} \frac{\text{cut}(SC^{\star}_q, \overline{SC^{\star}_q})}{\text{vol}(SC^{\star}_q)}\right|,
\end{aligned}$$

as $\widehat{\boldsymbol{V}}$ converges strongly to $\boldsymbol{V}$, the eigenvalues of $\widehat{\boldsymbol{L}}_{\text{norm}}$ converges strongly to the eigenvalues of $\boldsymbol{L}_{\text{norm}}$.

Then one needs to prove for each q, $\lim_{n\to\infty} h(\widehat{SC}^{\star}_q, SC^{\star}_q) = 0$ a.s.. Recall the k-means clusters $\{\widehat{C}^{\star}_q, q = 1, \cdots, Q\}$ for $\{\widehat{\boldsymbol{r}}_i, i = 1, \cdots, n\}$ and the TSFDSC clusters $\{\widehat{SC}^{\star}_q, q = 1, \cdots, Q\}$ for $\{\widehat{\boldsymbol{\beta}}_i, i = 1, \cdots, n\}$, is one-to-one correspondence, i.e. if $\widehat{\boldsymbol{r}}_{i_1} \in \widehat{C}^{\star}_{q_1}$, $\widehat{\boldsymbol{\beta}}_{i_1} \in \widehat{SC}^{\star}_{q_1}$. The same is true for $\{SC^{\star}_q\}$ and $\{C^{\star}_q\}$. As $\max_{1\le i\le n} \|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\| = O_p\left(\sqrt{N_s^3 m^{-1}\log m}\right)$ where $\widehat{\boldsymbol{\beta}}_i \in \widehat{SC}^{\star}_{q_1}$ and $\boldsymbol{\beta}_i \in SC^{\star}_{q_2}$. Combining Lemma 6, for each q, there exists a unique q', such that $\lim_{n\to\infty} h(\widehat{SC}^{\star}_q, SC^{\star}_{q'}) = 0$ a.s., thus by rearranging the order of $\{\widehat{SC}^{\star}_1, \widehat{SC}^{\star}_2, \cdots, \widehat{SC}^{\star}_Q\}$, $\lim_{n\to\infty} h(\widehat{SC}^{\star}_q, SC^{\star}_q) = 0$. Similar to Lemma 5,

$$\text{cut}(\widehat{SC}^{\star}_q, \overline{\widehat{SC}^{\star}_q}) \to \text{cut}(SC^{\star}_q, \overline{SC^{\star}_q}),$$

$$\text{vol}(\widehat{SC}^{\star}_q) \to \text{vol}(SC^{\star}_q).$$

Obviously, one obtains that

$$\lim_{n\to\infty} \left|\sum_{q=1}^{Q} \frac{\text{cut}(\widehat{SC}^{\star}_q, \overline{\widehat{SC}^{\star}_q})}{\text{vol}(\widehat{SC}^{\star}_q)} - \sum_{q=1}^{Q} \frac{\text{cut}(SC^{\star}_q, \overline{SC^{\star}_q})}{\text{vol}(SC^{\star}_q)}\right| = 0.$$

The proof of Throrem 2.2 has been completed.

S3 Tables and Figures

Due to space limitations in the main text, this section provides more extensive numerical results, plots of the mean curves, smoothed generated trajectories, and stability scores across different clusters.

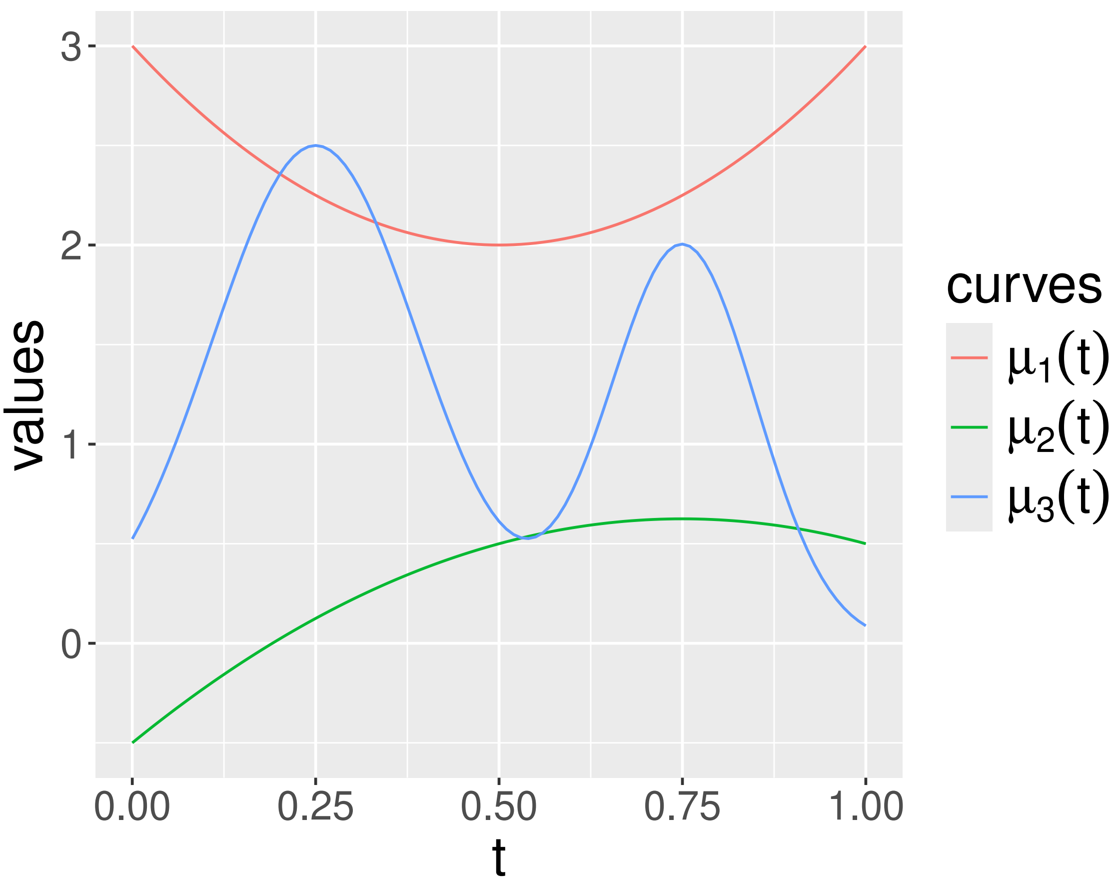

Figure S.1: Plot of generated mean curves $\mu_\ell(\cdot), \ell = 1, 2, 3$.

Table S.1: Clustering results for settings 1.a with empirical formula-based N_s over 100 replications (incorporate comparisons with FSC-S and kCFC).

m_i	Scenario	n																	
		50						100						200					
		FD k-means	TSFDSC	FSC-S(D_0)	FSC-S(D_1)	FSC-S(D_2)	kCFC	FD k-means	TSFDSC	FSC-S(D_0)	FSC-S(D_1)	FSC-S(D_2)	kCFC	FD k-means	TSFDSC	FSC-S(D_0)	FSC-S(D_1)	FSC-S(D_2)	kCFC
6	1	0.1319*	0.1308	-	-	-	-	0.1299*	0.1097	-	-	-	-	0.1389*	0.1129	-	-	-	-
	2	0.0623	0.0818*	-	-	-	-	0.0495	0.0686*	-	-	-	-	0.0400	0.0696*	-	-	-	-
	3	0.7699	0.9465*	-	-	-	0.4419	0.7711	0.9672*	-	-	-	0.1908	0.7623	0.9488*	-	-	-	0.0821
	4	0.7107	0.8041*	-	-	-	0.4136	0.7131	0.8521*	-	-	-	0.1731	0.7107	0.8776*	-	-	-	0.0606
	5	0.9518*	0.9295	-	-	-	0.9213	0.9482*	0.9453	-	-	-	0.8632	0.9479	0.9747	-	-	-	0.7823
	6	0.9303	0.9336*	-	-	-	0.8842	0.9221	0.9409*	-	-	-	0.8387	0.9279	0.9547*	-	-	-	0.7370
11	1	0.1303*	0.1258	-	-	-	-	0.1297*	0.1086	-	-	-	-	0.1362	0.1113	-	-	-	-
	2	0.0329	0.0344*	-	-	-	-	0.0732	0.0924*	-	-	-	-	0.0683	0.1125*	-	-	-	-
	3	0.7695	0.9539*	-	-	-	0.2930*	0.7705	0.9581*	-	-	-	0.0896	0.7622	0.9526*	-	-	-	0.0341
	4	0.7430	0.7855*	-	-	-	0.3423	0.7418	0.8549*	-	-	-	0.1618	0.7333	0.8936*	-	-	-	0.0606
	5	0.9502*	0.9343	-	-	-	0.8878*	0.9471*	0.9451	-	-	-	0.8318	0.9476	0.9705*	-	-	-	0.7084
	6	0.9427*	0.9388	-	-	-	0.8582	0.9382	0.9694*	-	-	-	0.8022	0.9345	0.9597*	-	-	-	0.7060
21	1	0.1318*	0.1258	0.1049	0.0303	0.0168	-	0.1272*	0.1038	0.0854	0.0176	0.0073	-	0.1335*	0.1115	0.0788	0.0129	0.0041	-
	2	0.0329	0.0344	0.0719*	0.0122	0.0134	-	0.0310	0.0285	0.0525*	0.0055	0.0067	-	0.0183	0.0311	0.0436*	0.0027	0.0038	-
	3	0.7695	0.9539*	0.6388	0.0001	0.0094	0.3155	0.7720	0.9415*	0.6332	0.0001	0.0050	0.0905	0.7625	0.9315*	0.6286	0.0000	0.0027	0.0272
	4	0.7430	0.7855*	0.6217	0.0110	0.0131	0.3252	0.7463	0.8131*	0.6206	0.0055	0.0065	0.1561	0.7402	0.8411*	0.6135	0.0025	0.0036	0.0549
	5	0.9502*	0.9343	0.9391	0.8483	0.8448	0.9055	0.9461*	0.9409	0.9395	0.8487	0.8469	0.8251	0.9471	0.9621*	0.9361	0.8445	0.8468	0.7084
	6	0.9427*	0.9388	0.9321	0.7364	0.0135	0.9042	0.9417	0.9646*	0.9293	0.7316	0.0072	0.8360	0.9398	0.9428*	0.9245	0.7331	0.0042	0.7006
51	1	0.1293*	0.1264	0.1048	0.0293	0.0166	-	0.1288*	0.1067	0.0862	0.0179	0.0080	-	0.1366*	0.1077	0.0801	0.0130	0.0041	-
	2	0.1286*	0.1268	0.0857	0.0132	0.0132	-	0.1192	0.1355*	0.0702	0.0063	0.0064	-	0.1205	0.1482*	0.0601	0.0033	0.0036	-
	3	0.7717	0.9495*	0.6380	0.0002	0.0096	0.2490	0.7720	0.9571*	0.6325	0.0002	0.0044	0.1036	0.7628	0.9358*	0.6280	0.0001	0.0024	0.0409
	4	0.7673	0.8838*	0.6282	0.0103	0.0138	0.2815	0.7647	0.9138*	0.6282	0.0050	0.0064	0.0827	0.7552	0.9378*	0.6214	0.0029	0.0035	0.0564
	5	0.9502*	0.9177	0.9393	0.8466	0.8200	0.9080	0.9458*	0.9410	0.9382	0.8482	0.8184	0.8265	0.9470*	0.9325	0.9349	0.8440	0.8157	0.6941
	6	0.9465*	0.9215	0.9348	0.6770	0.0150	0.8969	0.9445*	0.9425	0.9312	0.6801	0.0080	0.8406	0.9448	0.9472*	0.9289	0.6791	0.0048	0.6918
101	1	0.1309*	0.1264	0.1047	0.0309	0.0237	-	0.1272*	0.1125	0.0863	0.0182	0.0101	-	0.1377*	0.1009	0.0798	0.0133	0.0073	-
	2	0.1301	0.1401*	0.0959	0.0141	0.0138	-	0.1223	0.1395*	0.0763	0.0061	0.0064	-	0.1305	0.1410*	0.0710	0.0034	0.0033	-
	3	0.7701	0.9386*	0.6376	0.0002	0.0065	0.2391	0.7712	0.9615*	0.6319	0.0001	0.0033	0.0931	0.7627	0.9568*	0.6274	0.0001	0.0015	0.0338
	4	0.7652	0.9424*	0.6353	0.0093	0.0137	0.2422	0.7654	0.9366*	0.6296	0.0046	0.0065	0.1068	0.7564	0.9519*	0.6229	0.0024	0.0032	0.0485
	5	0.9505	0.9551*	0.9387	0.8482	0.8488	0.9154	0.9462	0.9662*	0.9378	0.8488	0.8449	0.8304	0.9467	0.9705*	0.9346	0.8449	0.8413	0.7139
	6	0.9481	0.9540*	0.9372	0.7525	0.0158	0.9131	0.9461*	0.9361	0.9355	0.7515	0.0081	0.8401	0.9455*	0.9363	0.9318	0.7473	0.0048	0.6953

Table S.2: Clustering results for settings 1.a and 2.a with empirical formula-based N_s over 100 replications. The best performance in each simulation scenario is indicated with an asterisk "*".

m_i	Scenario	n														
		50					100					200				
		FD k-means	TSFDSC				FD k-means	TSFDSC				FD k-means	TSFDSC			
			ι					ι					ι			
			adaptive	0.1	0.5	0.9		adaptive	0.1	0.5	0.9		adaptive	0.1	0.5	0.9
6	1	0.1319	0.1308	0.1029	0.1451	0.1522*	0.1299	0.1097	0.0901	0.1535	0.1582*	0.1097	0.0901	0.1535	0.1582	0.1591*
	2	0.0623	0.0818*	0.0331	0.0676	0.0809	0.0495	0.0686	0.0219	0.0659	0.0843*	0.0399	0.0696	0.0149	0.0702	0.0975*
	3	0.7699	0.9465*	0.7636	0.9159	0.9362	0.7711	0.9672*	0.7585	0.9195	0.9224	0.7623	0.9488	0.7503	0.9306	0.9728*
	4	0.7107	0.8041	0.7041	0.7946	0.8094*	0.7131	0.8521*	0.7001	0.8025	0.8486	0.7107	0.8776*	0.6997	0.8056	0.8594
	5	0.9518*	0.9295	0.9443	0.9505	0.9259	0.9482*	0.9453	0.9346	0.9192	0.9199	0.9479	0.9747*	0.9425	0.9733	0.9663
	6	0.9303	0.9336	0.9240	0.9473*	0.9379	0.9221	0.9409*	0.9097	0.9305	0.9314	0.9249	0.9547	0.9177	0.9597	0.9613*
11	1	0.1303	0.1326	0.1057	0.1435	0.1547*	0.1297	0.1086	0.0922	0.1545	0.1579*	0.1362	0.1113	0.0878	0.1772*	0.1591
	2	0.0868	0.0925	0.0457	0.0913	0.0927*	0.0732	0.0924	0.0303	0.0866	0.0978*	0.0683	0.1125	0.0258	0.1107	0.1279*
	3	0.7715	0.9299*	0.7612	0.9026	0.9068	0.7705	0.9581	0.7592	0.9274	0.9639*	0.7622	0.9526*	0.7495	0.9303	0.9519
	4	0.7447	0.8265	0.7338	0.8359	0.8455*	0.7418	0.8549	0.7266	0.8407	0.8815*	0.7333	0.8936	0.7209	0.8411	0.8966*
	5	0.9508*	0.9340	0.9436	0.9423	0.9174	0.9471	0.9451	0.9499	0.9650*	0.9536	0.9476	0.9705*	0.9424	0.9478	0.9452
	6	0.9395	0.9424*	0.9314	0.9387	0.9199	0.9382	0.9694*	0.9446	0.9674	0.9637	0.9345	0.9597	0.9285	0.9604*	0.9489
21	1	0.1318	0.1258	0.0990	0.1419	0.1526*	0.1272	0.1038	0.0904	0.1515	0.1572*	0.1335	0.1115	0.0876	0.1777*	0.1584
	2	0.0329	0.0344	0.0239	0.0356	0.0425*	0.0310	0.0285	0.0150	0.0251	0.0381*	0.0183	0.0311	0.0080	0.0229	0.0499*
	3	0.7695	0.9539*	0.7620	0.9080	0.9413	0.7720	0.9415	0.7599	0.9274	0.9515*	0.7625	0.9315	0.7503	0.9308	0.9730*
	4	0.7430	0.7855	0.7321	0.8282	0.8381*	0.7463	0.8131	0.7339	0.8497	0.8606*	0.7402	0.8411	0.7310	0.8588	0.9035*
	5	0.9502*	0.9343	0.9307	0.9460	0.9213	0.9461*	0.9409	0.9261	0.9312	0.9199	0.9471	0.9621*	0.9423	0.9562	0.9494
	6	0.9427*	0.9388	0.9296	0.9405	0.9370	0.9417	0.9646*	0.9233	0.9403	0.9389	0.9398	0.9428	0.9281	0.9498*	0.9415
51	1	0.1293	0.1264	0.1020	0.1427	0.1473*	0.1288	0.1067	0.0916	0.1546	0.1563*	0.1366	0.1077	0.0873	0.1771*	0.1586
	2	0.1286*	0.1268	0.0787	0.1221	0.1219	0.1192	0.1355	0.0761	0.1340	0.1362*	0.1205	0.1482	0.0699	0.1685*	0.1567
	3	0.7717	0.9495	0.7620	0.9232	0.9696*	0.7720	0.9571*	0.7593	0.9188	0.9397	0.7628	0.9358	0.7504	0.9313	0.9604*
	4	0.7673	0.8838	0.7538	0.8886	0.9402*	0.7647	0.9138	0.7506	0.8891	0.9372*	0.7552	0.9378	0.7433	0.9026	0.9451*
	5	0.9502	0.9177	0.9397	0.9544*	0.9463	0.9458	0.9410	0.9562	0.9603*	0.9409	0.9470	0.9325	0.9382	0.9478*	0.9368
	6	0.9465*	0.9215	0.9418	0.9307	0.9281	0.9445	0.9425	0.9538*	0.9502	0.9400	0.9448	0.9472	0.9390	0.9639*	0.9408
101	1	0.1309	0.1264	0.1041	0.1409	0.1529*	0.1272	0.1125	0.0901	0.1546	0.1589*	0.1377	0.1009	0.0885	0.1765*	0.1586
	2	0.1301	0.1401*	0.0981	0.1368	0.1356	0.1223	0.1395	0.0857	0.1487	0.1529*	0.1305	0.1410	0.0835	0.1732*	0.1578
	3	0.7701	0.9386*	0.7632	0.9194	0.9361	0.7712	0.9615*	0.7594	0.9229	0.9437	0.7627	0.9568*	0.7504	0.9274	0.9519
	4	0.7652	0.9424*	0.7566	0.9062	0.9412	0.7654	0.9366	0.7560	0.9105	0.9394*	0.7564	0.9519	0.7480	0.9103	0.9615*
	5	0.9505	0.9551*	0.9398	0.9379	0.9424	0.9462	0.9662*	0.9375	0.9479	0.9325	0.9467	0.9705*	0.9447	0.9607	0.9452
	6	0.9481	0.9540*	0.9388	0.9334	0.9174	0.9461*	0.9361	0.9369	0.9424	0.9362	0.9455	0.9363	0.9432	0.9602*	0.9451

Table S.3: Clustering results for settings 1.a and 2.a, with BIC-based N_s selection over 100 replications.

m_i	Scenario	n					
		50		100		200	
		FD k-means	TSFDSC	FD k-means	TSFDSC	FD k-means	TSFDSC
6	1	0.1067	0.1242*	0.0913	0.1121*	0.0834	0.1154*
	2	0.0329	0.0518*	0.0304	0.0442*	0.0223	0.0422*
	3	0.8450	0.9524*	0.8371	0.9571*	0.8350	0.9532*
	4	0.7690	0.8146*	0.7746	0.8785*	0.7721	0.8840*
	5	0.9938*	0.9114	0.9937*	0.9579	0.9931*	0.9537
	6	0.9786*	0.9377	0.9751*	0.9463	0.9747*	0.9469
11	1	0.1271	0.1346*	0.1238*	0.1073	0.1264*	0.1079
	2	0.0677*	0.0637	0.0569	0.0708*	0.0557	0.0924*
	3	0.7881	0.9215*	0.7883	0.9282*	0.7798	0.9782*
	4	0.7612	0.8325*	0.7556	0.8520*	0.7492	0.8890*
	5	0.9561*	0.9340	0.9538	0.9744*	0.9540	0.9579*
	6	0.9453*	0.9439	0.9446	0.9691*	0.9401	0.9599*
21	1	0.1200	0.1297*	0.1314*	0.1084	0.1541*	0.1045
	2	0.0149	0.0206*	0.0064	0.0080*	0.0032*	0.0031
	3	0.7569	0.9564*	0.7516	0.9456*	0.7333	0.9261*
	4	0.7304	0.7454*	0.7189	0.7416*	0.4371	0.5242*
	5	0.9473	0.9507*	0.9421*	0.9366	0.9395	0.9578*
	6	0.9401*	0.9354	0.9364	0.9471*	0.9248	0.9313*
51	1	0.0973	0.1166*	0.0827	0.1148*	0.0783	0.1099*
	2	0.0887	0.1048*	0.0812	0.0925*	0.0700	0.0984*
	3	0.8524	0.9470*	0.8431	0.9689*	0.8433	0.9702*
	4	0.8387	0.9405*	0.8397	0.9656*	0.8340	0.9552*
	5	1.0000*	0.9033	1.0000*	0.9117	1.0000*	0.9369
	6	0.9997*	0.9579	1.0000*	0.9537	1.0000*	0.9664
101	1	0.0972	0.1237*	0.0827	0.1110*	0.0776	0.1095*
	2	0.0918	0.1053*	0.0817	0.0991*	0.0784	0.0936*
	3	0.8505	0.9401*	0.8436	0.9644*	0.8425	0.9574*
	4	0.8480	0.9531*	0.8419	0.9351*	0.8358	0.9615*
	5	1.0000*	0.9243	1.0000*	0.9369	1.0000*	0.9159
	6	1.0000*	0.9327	1.0000*	0.9327	1.0000*	0.9411

Table S.4: Clustering results for 1.b and 2.b, with BIC-based N_s selection over 100 replications.

m_i	Scenario	n = 50 FD k-means	50 TSFDSC	100 FD k-means	100 TSFDSC	200 FD k-means	200 TSFDSC
4-11	1	0.1371*	0.1296	0.1330*	0.1177	0.1404*	0.1059
	2	0.0727	0.0771*	0.0492	0.0699*	0.0469	0.0805*
	3	0.7802	0.8886*	0.7800	0.9333*	0.7699	0.9488*
	4	0.7334	0.8365*	0.7190	0.8374*	0.7230	0.8830*
	5	0.8572*	0.8000	0.8828*	0.8348	0.8931*	0.8125
	6	0.8426	0.8756*	0.8867*	0.8787	0.8851	0.8888*
4-21	1	0.1391*	0.1271	0.1473*	0.1156	0.1523*	0.1021
	2	0.0396	0.0458*	0.0343*	0.0244	0.0823*	0.0481
	3	0.7529	0.9141*	0.7446	0.9529*	0.7273	0.9837*
	4	0.7230	0.7279*	0.6984*	0.6695	0.7018	0.7560*
	5	0.8696	0.9122*	0.8928*	0.8753	0.8852*	0.8520
	6	0.8668	0.8884*	0.8602	0.9066*	0.9084*	0.8898
21-101	1	0.0926	0.1242*	0.0835	0.1115*	0.0799	0.1044*
	2	0.0893	0.0940*	0.0809	0.0882*	0.0784	0.0928*
	3	0.8374	0.9538*	0.8495	0.9349*	0.8356	0.9070*
	4	0.8319	0.9439*	0.8418	0.9662*	0.8253	0.9507*
	5	0.9009	0.9324*	0.9117	0.9495*	0.8825	0.9579*
	6	0.9240	0.9411*	0.9534	0.9621*	0.9368	0.9369*

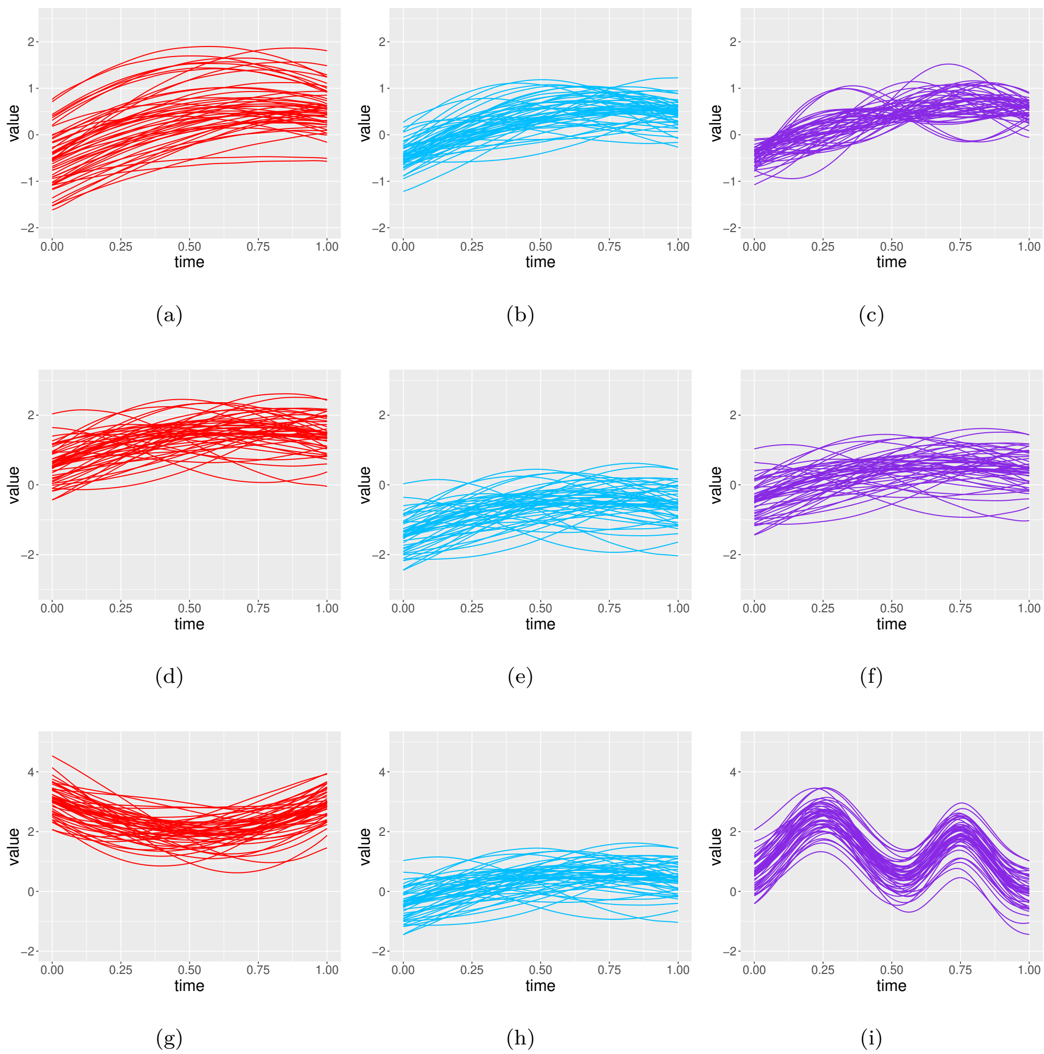

Figure S.2: B-spline smoothing for gengrated trajectories of Scenario 1, 3 and 5 with $n = 50$ and $m_i = 101$ in one replication. Panels (a)-(c) represent the B-spline smoothing for S1a, S1b, S1c respectively. Panels (d)-(f) represent the B-spline smoothing for S3a, S3b, S3c respectively. Panels (g)-(i) represent the B-spline smoothing for S5a, S5b, S5c respectively.

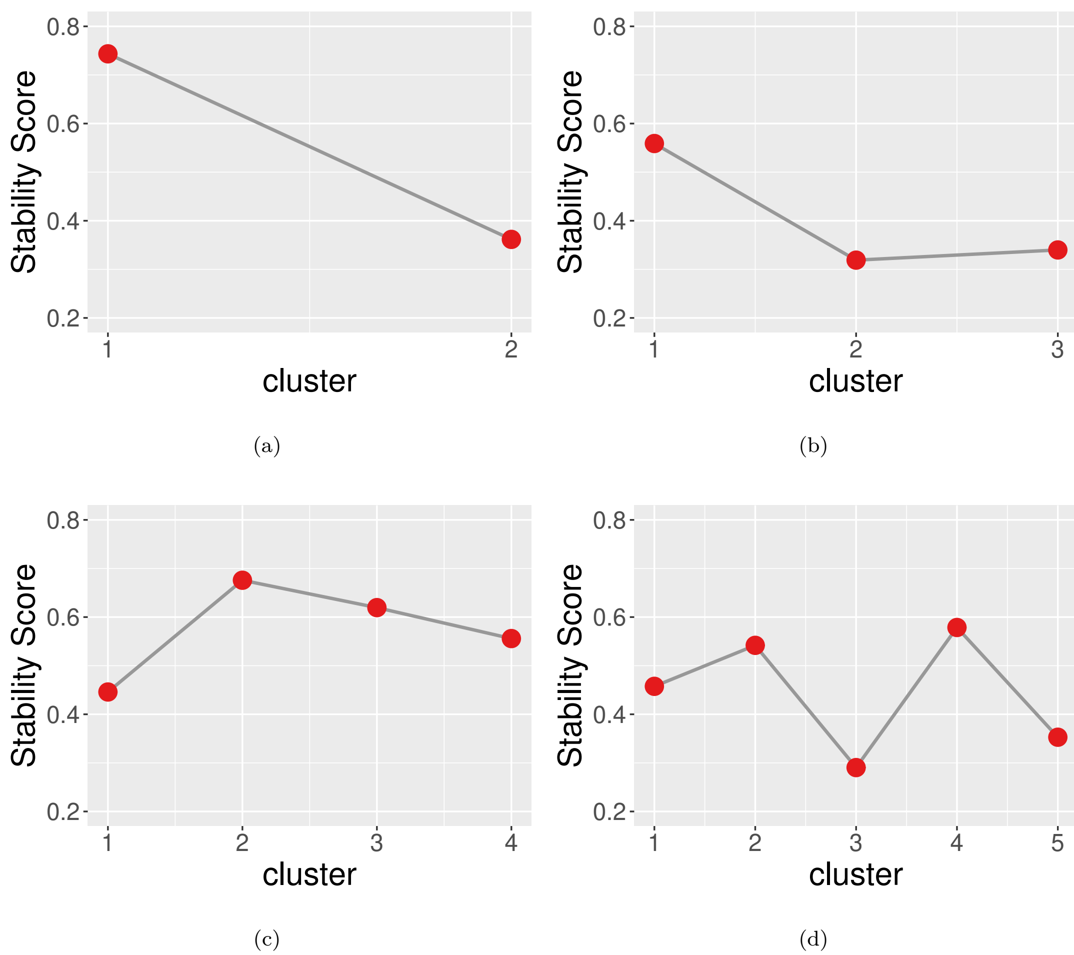

Figure S.3: Stability scores of different clusters.

References

Abraham, C., P. A. Cornillon, E. Matzner-Løber, and N. Molinari (2003). Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics 30*(3), 581–595.

Burman, P. (1991). Regression function estimation from dependent observations. *Journal of Multivariate Analysis 36*(2), 263–279.