Supplementary Material:
Similarity Preserving Representation Learning for Time Series Clustering

1 Proof of Theorem 2

Using the definition of matrix $\mathbf{A}$ and assumptions $A1$-$A3$, we have

$$\mathbf{A}_{ij} = \begin{cases} \frac{d_{a0}^2 + d_{b0}^2 - 2d_{ab}^2}{2} + \mathcal{O}(\epsilon) & \text{if } T_i \in C_a,\ T_j \in C_b,\ a \neq b \\ d_{a0}^2 + \mathcal{O}(\epsilon) & \text{if } T_i,\ T_j \in C_a \end{cases}$$

Let $N_{ij} = \mathcal{O}(\epsilon)$, and

$$\mathbf{L}_{ij} = \begin{cases} \frac{d_{a0}^2 + d_{b0}^2 - 2d_{ab}^2}{2} & \text{if } T_i \in C_a,\ T_j \in C_b,\ a \neq b \\ d_{a0}^2 & \text{if } T_i,\ T_j \in C_a, \end{cases}$$

then $\mathbf{A} = \mathbf{L} + \mathbf{N}$.

Let I_a be the index of the time series in cluster C_a, $a = 1, 2, \cdots, k$, then the matrix L could be divided into $k \times k$ blocks: $\mathbf{L}_{I_a, I_b}, 1 \leq a, b \leq k$, where each block has the same values:

$$\mathbf{L}_{I_a, I_b} = \begin{cases} \frac{d_{a0}^2 + d_{b0}^2 - 2d_{ab}^2}{2} & \text{if } a \neq b \\ d_{a0}^2 & \text{if } a = b. \end{cases}$$

Let $\mathbf{x}$ be a vector that satisfies $\mathbf{x}_{I_a} = \frac{d_{a0}^2}{2}$. For any index set I, let $\mathbf{e}^I$ be the indicator vector:

$$\mathbf{e}_i^I = \begin{cases} 1 & \text{if } i \in I \\ 0 & \text{otherwise} \end{cases}$$

Also, let $\mathbf{1}$ be the all 1 vector. Then we have:

$$\mathbf{L}_{ab} = \mathbf{x}\mathbf{1}^\top + \mathbf{1}\mathbf{x}^\top + \sum_{a \neq b} d_{ab}^2 \mathbf{e}^{I_a} (\mathbf{e}^{I_b})^\top.$$

In this sense, $\mathbf{L}$ is the summation of $2 + k(k-1)$ rank 1 matrices. Thus its rank is at most $2 + k(k-1)$.

2 Proof of Theorem 3

Let $f(\mathbf{X}) := \|P_\Omega(\tilde{\mathbf{A}} - \mathbf{X}\mathbf{X}^\top)\|_F^2$, and $\bar{\mathbf{X}}$ be the unique accumulation point of the sequence $(\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \dots)$. With the coordinate descent algorithm, the generated sequence $f(\mathbf{X}^{(i)}), i = 1, 2, \cdots$ is monotonically non-increasing and bounded below. Note that only one variable has been updated between $\mathbf{X}^{(k+1)}$ and $\mathbf{X}^{(k)}$.

We prove the Theorem 2 by contradiction. Suppose $\bar{\mathbf{X}}$ is not a stationary point of problem (6), then there exists a pair (i, j) satisfying

$$f(\bar{\mathbf{X}} + \alpha \mathbf{E}^{ij}) = f(\bar{\mathbf{X}}) - \epsilon < f(\bar{\mathbf{X}}),$$

where $\alpha \neq 0$, $\epsilon > 0$, and $\mathbf{E}^{ij}$ is the one-hot matrix with all zero entries except that the (i, j)'s entry equals to 1.

Let $(\mathbf{X}^{(n_0)}, \mathbf{X}^{(n_1)}, \dots)$ be a subsequence of $(\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \dots)$ and n_k is the number of iterations that updates the entry (i, j) for $k + 1$ times.

Note that f is continuous and $\mathbf{X}^{(n_k)} + \alpha \mathbf{E}^{ij} \to \bar{\mathbf{X}} + \alpha \mathbf{E}^{ij}$ when $k \to \infty$. There exists a sufficiently large K so that for all $k > K$, we have

$$\begin{aligned} f(\mathbf{X}^{(n_k)} + \alpha \mathbf{E}^{ij}) &\leq f(\bar{\mathbf{X}} + \alpha \mathbf{E}^{ij}) + \frac{\epsilon}{2} \\ &= f(\bar{\mathbf{X}}) - \frac{\epsilon}{2} \end{aligned}$$

By flipping the sign of the above formula and adding $f(\mathbf{X}^{(n_k)})$ to both sides, we have

$$f(\mathbf{X}^{(n_k)}) - f(\mathbf{X}^{(n_k)} + \alpha \mathbf{E}^{ij}) \geq f(\mathbf{X}^{(n_k)}) - f(\bar{\mathbf{X}}) + \frac{\epsilon}{2}.$$

By constructing the subsequence, the (i, j)th entry of $\mathbf{X}^{(n_k)}$ is updated earlier than the other entries to obtain $\mathbf{X}^{(n_{k+1})}$, which implies that

$$f(\mathbf{X}^{(n_{k+1})}) \leq f(\mathbf{X}^{(n_k+1)}) \leq f(\mathbf{X}^{(n_k)} + \alpha \mathbf{E}^{ij}).$$

Hence we have

$$\begin{aligned} f(\mathbf{X}^{(n_k)}) - f(\mathbf{X}^{(n_{k+1})}) &\geq f(\mathbf{X}^{(n_k)}) - f(\mathbf{X}^{(n_k)} + \alpha \mathbf{E}^{ij}) \\ &\geq f(\mathbf{X}^{(n_k)}) - f(\bar{\mathbf{X}}) + \frac{\epsilon}{2} \\ &\geq \frac{\epsilon}{2}. \end{aligned}$$

Since $f(\bar{\mathbf{X}}) \leq f(\mathbf{X}^{(n_k)})$, for all $k > K$ we have

$$f(\mathbf{X}^{(n_{k+1})}) \leq f(\mathbf{X}^{(n_k)}) - \frac{\epsilon}{2},$$

which leads to a contradiction since f is bounded below.