Supplementary Material

Proof for Theorem 2

Let X_{i_k} and $X_{i_{k'}}$ be pivot time series for P_k and $P_{k'}$ as defined in Theorem 1. Furthermore, let $c = c_{k,k'} = \rho_{i_k,i_{k'}}$. We first apply Theorem 1 to derive bounds for the correlation between $X_i \in P_k$ and $X_j \in P_{k'}$ via X_k and $X_{k'}$ for the two cases $c \geq 0$ and $c < 0$ separately.

Let $c \geq 0$. We first apply Theorem 1 to derive bounds for the correlation between X_i and $X_{i_{k'}}$ via X_{i_k} and obtain

$$\rho_{i,i_{k'}} \in \left[\alpha c - \sqrt{(1-c^2)(1-\alpha^2)}, c + \sqrt{(1-c^2)(1-\alpha^2)}\right].$$

We apply Theorem 1 again to bound the correlation between X_i and X_j via $X_{i_{k'}}$ using this result:

$$\begin{aligned}
\rho_{i,j} &\leq \rho_{i,i_{k'}}\rho_{i_{k'},j} + \sqrt{(1-\rho_{i,i_{k'}})(1-\rho_{i_{k'},j})} \\
&\leq c + \sqrt{(1-c^2)(1-\alpha^2)} \\
&\quad + \sqrt{\left(1-\left(\alpha c - \sqrt{(1-c^2)(1-\alpha^2)}\right)^2\right)(1-\alpha^2)} \\
&= c + \sqrt{1-\alpha^2}\Bigg(\sqrt{1-c^2} \\
&\quad + \sqrt{1-\left(\alpha c - \sqrt{(1-c^2)(1-\alpha^2)}\right)^2}\Bigg),
\end{aligned}$$

$$\begin{aligned}
\rho_{i,j} &\geq \rho_{i,i_{k'}}\rho_{i_{k'},j} - \sqrt{(1-\rho_{i,i_{k'}})(1-\rho_{i_{k'},j})} \\
&\geq \left(\alpha c - \sqrt{(1-c^2)(1-\alpha^2)}\right)\rho_{i_{k'},j} \\
&\quad - \sqrt{\left(1-\left(\alpha c - \sqrt{(1-c^2)(1-\alpha^2)}\right)^2\right)(1-\alpha^2)} \\
&= \alpha c \rho_{i_{k'},j} - \sqrt{(1-c^2)(1-\alpha^2)}\rho_{i_{k'},j} \\
&\quad - \sqrt{\left(1-\left(\alpha c - \sqrt{(1-c^2)(1-\alpha^2)}\right)^2\right)(1-\alpha^2)} \\
&\geq \alpha^2 c - \sqrt{(1-c^2)(1-\alpha^2)} \\
&\quad - \sqrt{\left(1-\left(\alpha c - \sqrt{(1-c^2)(1-\alpha^2)}\right)^2\right)(1-\alpha^2)}
\end{aligned}$$

$$= \alpha^2 c - \sqrt{1-\alpha^2}\Bigg(\sqrt{1-c^2} + \sqrt{1-\Big(\alpha c - \sqrt{(1-c^2)(1-\alpha^2)}\Big)^2}\Bigg).$$

Let $c < 0$. With the same arguments as before, we get

$$\rho_{i,i_{k'}} \in \Big[c - \sqrt{(1-c^2)(1-\alpha^2)}, \alpha c + \sqrt{(1-c^2)(1-\alpha^2)}\Big]$$

and

$$\rho_{i,j} \leq \alpha^2 c + \sqrt{1-\alpha^2}\Bigg(\sqrt{1-c^2} + \sqrt{1-\Big(\alpha c + \sqrt{(1-c^2)(1-\alpha^2)}\Big)^2}\Bigg),$$

$$\rho_{i,j} \geq c - \sqrt{1-\alpha^2}\Bigg(\sqrt{1-c^2} + \sqrt{1-\Big(\alpha c + \sqrt{(1-c^2)(1-\alpha^2)}\Big)^2}\Bigg).$$

In total, this gives

$$\rho_{i,j} \leq \begin{cases} c + \sqrt{1-\alpha^2}\Bigg(\sqrt{1-c^2} + \sqrt{1-\Big(\alpha c - \sqrt{(1-c^2)(1-\alpha^2)}\Big)^2}\Bigg) & (c \geq 0) \\ \alpha^2 c + \sqrt{1-\alpha^2}\Bigg(\sqrt{1-c^2} + \sqrt{1-\Big(\alpha c + \sqrt{(1-c^2)(1-\alpha^2)}\Big)^2}\Bigg) & (c < 0) \end{cases}$$

$$\rho_{i,j} \geq \begin{cases} \alpha^2 c - \sqrt{1-\alpha^2}\Bigg(\sqrt{1-c^2} + \sqrt{1-\Big(\alpha c - \sqrt{(1-c^2)(1-\alpha^2)}\Big)^2}\Bigg) & (c \geq 0) \\ c - \sqrt{1-\alpha^2}\Bigg(\sqrt{1-c^2} + \sqrt{1-\Big(\alpha c + \sqrt{(1-c^2)(1-\alpha^2)}\Big)^2}\Bigg) & (c < 0) \end{cases}$$

which we abbreviate by $\rho_{i,j} \in [\mathrm{LB}, \mathrm{UB}]$. The diameter $D(\alpha, c)$ of these bounds, i.e. the difference between the upper and the lower bound, is given by:

$$D(\alpha,c) = (1-\alpha^2)|c| + 2\sqrt{1-\alpha^2}\Bigg(\sqrt{1-c^2} + \sqrt{1-\left(\alpha|c| - \sqrt{(1-c^2)(1-\alpha^2)}\right)^2}\Bigg).$$

The estimator from Equation 3 is the midpoint between the two bounds,

$$\hat{\rho}(i,j \mid \mathcal{P}, C) = \frac{1}{2}(\mathrm{UB} + \mathrm{LB}) = \frac{1}{2}(1+\alpha^2)c.$$

We guarantee $\ell_{ij} \leq \epsilon$ by setting the parameter α such that the diameter $D(\alpha, c)$ is smaller than 2ϵ for all possible values of c. To make the diameter sufficiently small, we observe

$$\begin{aligned} D(\alpha,c) &\leq (1-\alpha^2)|c| + 2\sqrt{1-\alpha^2}\left(\sqrt{1-c^2}+1\right) \\ &\leq \sqrt{1-\alpha^2}|c| + 2\sqrt{1-\alpha^2}\left(\sqrt{1-c^2}+1\right) \\ &= \sqrt{1-\alpha^2}\left(|c| + 2\sqrt{1-c^2}+2\right) \\ &\leq \sqrt{1-\alpha^2}\left(\sqrt{5}+2\right), \end{aligned}$$

where we obtained $|c| + 2\sqrt{1-c^2} \leq \sqrt{5}$ by maximizing the term for $c \in [-1, 1]$. Therefore, if

$$\sqrt{1-\alpha^2}\left(\sqrt{5}+2\right) \leq 2\epsilon \Leftrightarrow \alpha \geq \sqrt{1-\left(\frac{2\epsilon}{\sqrt{5}+2}\right)^2}$$

we have $D(\alpha, c) \leq 2\epsilon$ for all c and thus $\ell_{ij} \leq \epsilon$.

Evaluation on chlorine concentration data

We also evaluated COREQ-A against ApproxThresh on the chlorine concentration data used in the original publication by Mueen et al. (2010). The dataset consists of $N = 10624$ time series of length $T = 2155$. Details on the data can be found in their publication. We observe that the models obtained by COREQ-A are three orders of magnitude smaller and can be computed with much fewer correlation computations, while covering a similar range of average loss. The largest model obtained with COREQ-A (using $\alpha = 0.9$) has an average loss of 0.08 with a model size of 0.0003. ApproxThresh has lower losses only at the far right for $\tau \in \{0.3, 0.2, 0.1\}$, which leads to orders of magnitudes larger models of size 0.55, 0.69 and 0.84, respectively.