A Connected World in Future

Clustering is an important primitive step, and time series data will often be high-dimensional: the number of time series $d$ (i.e., number of sensors) will be much larger than the length of each time series $T$.

Challenge of High-dimensional Data

High-dimensional data result in an under-constrained problem: there are $T$ equations for estimating $d$ parameters, where $T << d$.

Clustering High-dimensional Data from VAR

We assume the high-dimensional data follow a vector autoregressive model (VAR) and do the clustering based on the degree to which a future value in each time series is predicted by past values of the others. We call this “cross-predictability”.

Methodology

The input:

$$X = [X_1, \ldots, X_T]^T \in \mathbb{R}^{T \times d} \quad (d \text{ time series of length } T)$$

$$X_S = [X_1, \ldots, X_{T-1}]^T \in \mathbb{R}^{(T-1) \times d}, \quad X_T = [X_2, \ldots, X_T]^T \in \mathbb{R}^{(T-1) \times d}$$

$$\Sigma = X_S X_S^T (T-1), \quad \hat{\gamma}_i = X_S^T X_T^T (T-1)$$

The VAR model:

$$X_{t+1} = A X_t + Z_t, \quad \text{for } t = 1, 2, \ldots, T - 1.$$  

The estimator:

$$\hat{\beta}_i = \arg \min_{\beta_i} \lambda \| \hat{\Sigma} \beta_i - \hat{\gamma}_i \|_{\infty, \infty} + \| \beta_i \|_1$$

Set $\hat{A} = [\hat{\beta}_1, \ldots, \hat{\beta}_d]^T$, and then apply a spectral clustering algorithm.

Main Results

**Main Theorem** Under the assumption of VAR model with a block diagonal transition matrix, we compactly denote $P_0 = P(Q_{S_0} S_0^T), \quad P_i = P(Q_{S_i} S_i^T), \quad r_0^i = r(P_0), \quad r_i^i = r(P_i)$, and $r_0 r_1 = \min_i r_i^i r_1^i$ for $l = 1, 2, \ldots, k$, and let

$$\rho = \frac{16 \| \Sigma \|_2 \max_i \Sigma_{j,j} \sqrt{6 \log d + 4}}{\min_i \| \Sigma_{j,j} \|_2 (1 - \| A \|_2) \sqrt{T}}.$$  

Furthermore, if

$$r_0 r_1 > \| \Sigma_{S_i, S_i} \|_{\infty, \infty} + 2 \rho,$$

where $\gamma_S \in \mathbb{R}^d$ is a column of $\Sigma_1$, then with probability at least $1 - 6d^{-1}$ the cluster recovery property holds for all the values of the regularization parameter $\lambda$ in the range:

$$\frac{1}{\rho r_1 (\| \gamma_S \|_{\infty, \infty} - 2 \rho)} < \lambda < \frac{1}{\rho + \| \Sigma_{S_i, S_i} \|_{\infty, \infty}},$$

which is guaranteed to be non-empty.