Project in Machine Learning in Structural Engineering

Background

We wish to monitor the behaviour of structures to help owners make informed decisions regarding safe operation and maintenance of infrastructure. Automatic and human-free monitoring shows promise in expanding monitoring to more structures, reducing its costs, and increasing its objectivity.

The common approach is to perform system identification on vibration acceleration data measured simultaneously at multiple points on a structure. The system identification gives an estimate of the modal features (frequency, damping, mode shape, …) of the structure. Using covariance-driven stochastic subspace identification (cov-SSI) as a system identification method requires a knowing the number of modes to detect in advance. This information is generally unavailable and differs from structure to structure. To counteract this problem, cov-SSI is performed for a range of model orders, and modal estimates are extracted from a stabilisation diagram. This is traditionally a manual task.

Automatic operational modal analysis (AOMA) algorithms are a class of algorithms capable of automatically extracting modal estimates from the output of cov-SSI over a range of orders. They rely heavily on clustering, a subset of unsupervised machine learning algorithms, to achieve this task.

Project aim

Familiarise oneself to the functionality of different clustering algorithms, such as k-means and hierarchical clustering, while building an AOMA algorithm to extract modal features from a shear frame.

Necessary tools

Strid Python package https://github.com/Gunnstein/strid, SciKit-Learn https://scikit-learn.org/stable/index.html

Tasks

1. Install and familiarise yourself with the strid package from operational modal analysis. Use the example of the shear frame to generate random vibration data for a shear frame of 8 storeys.

   Use the covariance-driven stochastic subspace identification example to perform the system identification on the shear frame vibration data. Create a stabilisation diagram with the detected system poles from the output of cov-SSI for a range of model orders.
2. In addition to the physical poles (representing structural modes), we see the apparition of many mathematical (spurious) poles which occur because of measurement and process noise. These spurious poles having no value and need to be removed.
   2.1. Note that physical poles create nice vertical lines in the stabilisation diagram. This is because we are detecting (very nearly) the same features for each pole at different model orders, because all these poles represent a physical manifestation in the data. The mathematical poles are scattered randomly throughout the diagram and no discernible pattern across the model orders is visible. Use this difference in characteristics between physical poles and mathematical poles to associate a value to each pole to quantify its likelihood of being physical or mathematical.

Note: Quantify (separately) a relative difference in frequency, damping, and mode shape from each pole to its nearest neighbour at the next inferior model order. Associate this value to the pole.

2.2. Use a partition clustering algorithm to separate all the poles into two groups based on their relative difference values. Determine which group of poles is physical and which is mathematical.

2.3. **Extra:** Basing yourself on published literature, use other or additional features to quantify the stability of a pole, and/or use another clustering algorithm to separate physical and mathematical poles, to improve the clearing of the stabilisation diagram.

3. Once most of the mathematical poles are removed, the vertical lines in the stabilisation diagram become clearer. Each line should represent a structural mode.

   3.1. Detect these structural modes by using an agglomerative clustering method on the remaining physical poles.
   Note: Use hierarchical clustering with a single linkage method with a cut-off distance of $d_c = 0.04$. *Define the distance between two poles as* $d_{c_{i,j}} = d\lambda_{i,j} + \left(1 - \mathrm{MAC}_{i,j}\right)$.

   3.2. Investigate the impact of the cut-off value and the linkage type on the clustering output.

   3.3. **Extra:** Investigate the impact of the definition of distance between two poles on the output of the agglomerative clustering. Use published literature to find additional or other features to associate to the poles.

4. Some of the clusters identified at the previous step are groupings of remaining mathematical poles. They generally have less members and contain a larger scatter in their components' features. These clusters are unwanted and need to be removed.

   4.1. Use a partition clustering method to remove these unwanted clusters.
   Note: Use k-means clustering with two clusters on the number of components of each hierarchical cluster.

   4.2. Extract the modal features of each detected mode as the average of the all the components' features within each hierarchical cluster.