Week 1: Principal Component Analysis of Structural Health Monitoring Data

Artificial Intelligence for Aerospace Engineering

AE-2224II - Q3 2022/3

Dimitrios Zarouchas

February 13, 2023

This laboratory instruction will guide you into the development of a Principal Component Analysis algorithm in order to reduce the dimension of structural health monitoring (SHM) data. Among other SHM techniques, Acoustic Emission (AE) has been extensively used to monitor aerospace composite structures the last 3 decades. AE records elastic wave signals generated during the formation and the propagation of cracks. With the use of special sensors, these elastic wave signals are transformed to electrical signals, which are stored using an acquisition card. Each electrical signal represents an AE event which carries information about the health status of a structure, i.e. the location of damage and its type. Composite structures exhibit different failure modes, i.e. matrix cracking, fibre breakage, delamination etc., during operation and each mode generates a vast amount of data. In order to retrieve useful information, it is important to create a manageable dataset, giving the opportunity to the operator to analyse the data in real-time..

Aim: Thus, the aim of this exercise is to reduce the dimensionality of the AE data, by using Principal Component Analysis, and create a manageable dataset for damage analysis .

Dataset information You will find the files necessary to complete this instruction in your docker container under directory `week1/principal-component-analysis`. From now on, all the paths will be relative to that directory.

Dataset: The dataset has been generated in the Aerospace Structures & Materials Laboratory of the Aerospace Engineering Faculty of TU Delft and relative information can be found in this Journal Article. The dataset consists of Acoustic Emission data collected from twelve samples (single CFRP stiffeners) which were subjected to compression-compression fatigue. The data is stored in a python dictionary with keys being ["Sample1", "Sample2",...,"Sample12"] and values being corresponding to data frames. A preliminary statistical analysis of the raw data has been performed and each sample includes 201 statistical features. Thus, each sample contains a matrix with m x 201 elements, m being the number of measurements taken for each sample.

Part 1: Preprocessing

Before feeding the data into the PCA algorithm you should...

- Transfer the data from the .mat file to python dictionary using the given code
- Get acquainted with the data: Select one of the twelve samples and plot 2D scatter graphs by choosing arbitrary 2 columns. Try 5 different graphs.

Part 2: Develop a PCA algorithm

1. Develop the PCA algorithm (as a function) using only numpy and utilizing the theory of eigenvalues and eigenvectors as presented during the lecture.
2. Apply the built function to one out of the twelve samples (your choice which one).
3. Calculate how many principal components are needed in order to cover 90% of the variance for sample number 1.
4. Calculate how much variance of the data is covered if you use the 3 first principal components. You should perform that for all 12 samples and it is expected to deliver 12 variances (one per sample).
5. Use the pca function available in the sklearn to validate your results

Submit your answers to Questions 2.1, 2.2, 2.3, and 2.4 in weblab.