Data-driven Physical Modelling (SEMTM0007) Coursework – Part 1

Released: 8 November 2024
Submission deadline: 28 November 2024

1 Coursework description

Formatting requirements

- The report for part 1 of the coursework is a single PDF file. Part 2 of the coursework is a separate PDF file. Note however that the two parts of the coursework (which are two PDF files) are submitted together on BlackBoard.
- The report is a combination of Python code, numerical results and diagrams with explanations, insights and conclusions.
- We prefer that you use a Jupyter notebook. The answers, insights and explanations to the questions can be written in Markdown, that allows the insertion of LaTeX formulae, should you need to explain the mathematics you are using.
- Your textual answers must be to the point and precise, we do not expect essay type answers. Excessive explanations can lead to confusion and lack of clarity.
- Assessment of your reports is not a tick-box exercise, and therefore we do not attach marks to each question. We use the university generic marking criteria for M-level units.

Intended learning outcomes From the unit catalogue:

1. Choose an appropriate data-driven modelling framework that aligns with the problem specification and provides the required accuracy and/or interpretability of the model.
2. Demonstrate mastery of a data-driven modelling technique through the analysis of a synthetic or experimental data set.
3. Use appropriate techniques to generate and prepare data for modelling purposes
4. Assess the quality of a mathematical model using metrics such as accuracy, repeatability, and generalisation

Marking criteria A pass mark (50) is achieved by demonstrating all intended learning outcomes across the two parts of the coursework. The relevant aspects of how the learning outcomes are achieved from the generic marking criteria are

1. **Content knowledge.** Whether you can make use of methods and skills explicitly taught. The precision and difficulty of skills demonstrated differentiates the marks. (20 %)
2. **Critical approach.** It is about convincing us whether you know what you are doing as opposed to following a recipe. You are able to explain why you are getting the results that the methods produce by referring to the underlying theory. For higher marks you are able to propose improvements. (15 %)
3. **Logical argument, explanation and evaluation of perspectives.** Your work follows a logical order and presents a convincing argument why you have approached the problem the way you did. It is all about the precision of your thinking. (20 %)
4. **Problem-solving.** This is strongly associated with content knowledge and demonstrated by solving the problems to various standards. (20 %)
5. **Insight.** You are able to delineate the strengths and weaknesses of your approach. You are able to come to a conclusion and summarise what you have learnt from solving the problem. (15 %)
6. **Decision-making.** When faced with a problem that has multiple solutions you are able to make a reasoned decision how to proceed. For higher marks you can make reasoned decisions about unseen problems and choose from solutions not explicitly taught. (10 %)

The weightings of the six marking criteria are approximate. As you notice we deem that content knowledge, logical explanation and evaluation and problem solving are the most important.

To achieve marks of 70 and above, you must demonstrate the use and mastery of methods and skills not explicitly taught. This does not need to be excessive and can be

- a realisation about the taught material that was not explicitly mentioned
- a method or skill that you have learnt in addition to the course material
- an insight about the problem that does not immediately follow from standard arguments.

Note that while striving for a mark of 70 and above, you should not forget about the basics. You can only get a high marks when the basics are already covered.

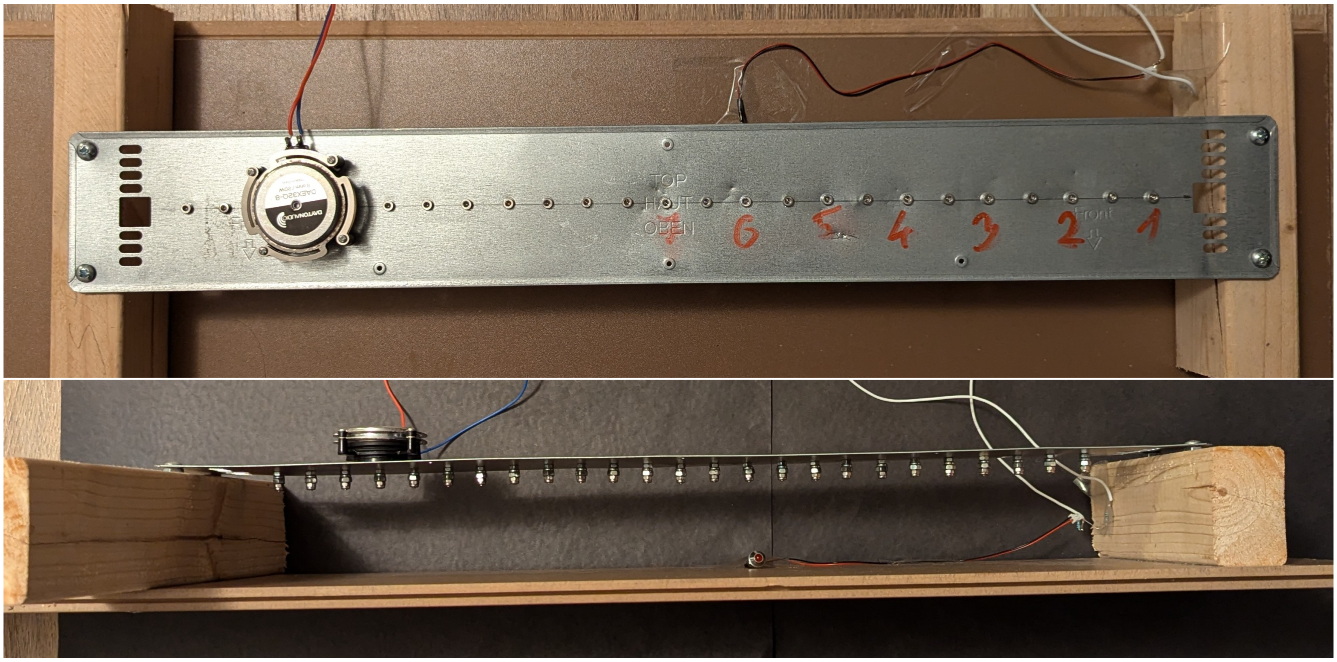

Fig. 1: Experimental rig. A soft steel plate with bolts attached. The vibration is measured using a video camera with 720p resolution running at 240 frames per second. Vibration data was extracted using digital image correlation (DIC).

1.1 Problem descriptions

The data comes from impact tests of a plate made out of soft steel. As such, the plate has high material damping. The plate has multiple bolts attached along its centre to increase the mass, that lowers the natural frequencies. There is also a transducer attached to the plate which is shorted for the duration of the experiment and therefore introduces additional damping. The plate was hit by a hammer at 8 different points three times each. The resulting vibration was captured using an action camera that has a 240 frames per seconds capture rate. The videos were post-processed and the vertical position of 17 bolts were tracked starting from the last frame of each video. This means the all data is recorded with time reversed. Each recording starts while the plate is stationary and last between 30-40 seconds. The level of noise can be deduced from the recordings of the stationary state.

Each trajectory is encoded as a matrix. The rows correspond to the bolts being tracked, the columns correspond to time. The start of the video is the last column of the matrix. The data is encoded as a Numpy *.npz* file. To open the file one can use the commands

data_dict = np.load('AutonomousTrajectoriesBig.npz')

data_all = [it[1] for it in data_dict.items()]

The variable *data_all* is the list of all trajectories.

Problem 1. Initial data processing.

1. Make sure that the data is in chronological order.

2. Remove the initial steady state (without vibrations) and eliminate the decayed part of each trajectory to increase the signal - noise ratio.

3. Make sure that each trajectory has zero mean. Explain why we cannot trust that the steady state is at the same numerical value for all trajectories.

4. Scale the data such that all entries of the data matrices have at most magnitude one. Explain why this scaling might be beneficial.

5. *Not part of the marking scheme*: characterise the power spectrum of the noise from the apparent steady state signals. Is there anything surprising?

Problem 2. Use delay embedding to find a linear model of the system.

1. Separate the data into a training and testing set. Explain why we need to do this.

2. Choose the largest delay such that the testing error is at most 10% greater than the training error. The error should be calculated for each data point using the L_2 norm and averaged over all data points to arrive at the mean error. Make sure that you take into account the dimensionality of each data point when calculating the error, so that the delay value does not scale the error and therefore errors remain comparable irrespective of the delay.

3. Calculate the standard deviation of the testing and the training error for each tested delay value. What can you conclude from high values standard deviation as opposed to low values of standard deviation?

4. Illustrate in a diagram how the testing and training error and the standard deviations vary as functions of the delay.

Problem 3. Extend the analysis in Problem 2, with dimension reduction using Singular Value Decomposition (SVD). Keep the length of delay constant at 8 or whatever the your analysis in Problem 1.2 recommends.

1. Use diagrams to illustrate how the rank of the SVD affects the training/testing accuracy and standard deviation of errors in case of a linear model.

2. Use diagrams to illustrate how the rank of the SVD affects the training/testing accuracy and standard deviation of errors in case of a quadratic polynomial model.

3. Use diagrams to illustrate how the rank of the SVD affects the training/testing accuracy and standard deviation of errors in case of a cubic polynomial model.

4. Explain what it means when the standard deviation of the testing error increases with increasing SVD rank.

Problem 4. Dynamic mode decomposition (DMD).

1. Calculate the dynamic mode decomposition of the linear model identified in Problem 2.

2. Order the eigenvalues of the DMD matrix starting with the smallest residual. Explain what the residual measures.

3. Plot these eigenvalues in the complex plane and colour code their residual values. Create two plots, one indicating their position relative to the complex unit circle, and another displaying the frequencies they correspond to on the vertical axis (in rad/sec).

4. Pick the eigenvalue with the smallest residual and non-zero vibration frequency. Use bagging to illustrate the uncertainty of this eigenvalue. Fit a Gaussian distribution to sampled values of this eigenvalue, calculate the mean and the covariance matrix of this distribution. Illustrate the distribution on a diagram by plotting contours of the fitted probability distribution function in the complex plane.

5. What can you say about the uncertainty of the vibration frequency corresponding to the dynamic mode selected in part 3 of this Problem?