Final Year Project Work Record

Week 4:

* Finished Udemy SQL course and researched uploading csv and .mat files into a sql database.
* Set up github
* Added more detail to project gannt chart
* Attempted an alternate iteration method on Python code for selecting ram locations to see if pre-selecting matrix size would reduce computational effort & time to evaluate objective function
* Read about some machine learning techniques and explored code and python libraries for tensor flows, k nearest neighbours, k clustering, random forests etc. Also read about keras, scikitlearn etc.
* **Emailed course organiser regarding a digital workbook**
* **Was allocated Ignazio Maria Viola**
* Met up with James Davidson. After introducing FASTBLADE and the scope of the project, we discussed python compared to other programming languages, available solvers, toolboxes and libraries that could be used for the problem. James will look at the Python code written in more detail soon.

Week 5:

15th

* Looking at how to fix LSQ objective function
* Simplified some code and added some comments
* Read up about Dataframes from panda library and unsure of how this could be utilised to improve current code
* Observed very bizarre pattern from Objective functions created and also noticed discrepancy from Python to MATLAB code.

17th

* Working on interim report
* Reading related documents to industry, FB etc.

18th

* Fixed issue with objective function. It was giving clearly wrong plot when determining discrepancy between target and actual shear while iterating over different values of sol
* Explored INERA001 data and explored methods of exporting it into a csv format. Come across many issues, will likely need to discuss with Jasmina
* Met with Encarni, discussing interim report and next steps
* Met with James, resolved issue of objective function computational efficiency, identified next steps in developing solvers. Also explored difference in speed of different IDEs, Jupyter Notebook seems to be no slower than Spyder. Could also compile to C using Cython if really needed to speed things up

Week 6

* Interim Report research
* Further experimented with .mat data files and working out how to export to csvs as fast as possible, using as few files as possible
* Met with Jeff- experimented with .mat data organsing and exporting to csvs
* Met with Bruce from IT discussed solvers, went through existing code, discussed IDEs, speeds, priorities in coding etc etc. Also discussed possible use of dataframes. Will likely meet up to discuss solvers again

Week 7

* Read about system link offered by national instruments
* Finished and submitted interim report
* Prepared for meeting with Jasmina by experimenting with .mat files again
* Met with Jasmina, discussed autoregressive models and agreed that PCA would be useful to apply to models used. Due to sinusoidal continuous data rather than discrete. Cluster techniques not very relevant for this problem. Factor analysis and PCA should be applied after exploring regressive models. Regressive models could be multivariate or autoregressive as required.
  + Discussed pros and cons of MATLAB vs Python and whether transferring and restructuring the data from one to the other is a good use of time.
  + Concluded that clustering techniques are unlikely to be valuable on data
  + Discussed options for calling python from MATLAB or vice versa, allowing the csv file intermediary step to be skipped- will be valuable to determine how important Python-only code is
  + Would seriously need to plan how data would be structured and planned within Python- perhaps using classes
  + Consider matlab vs python and where value is-<https://uk.mathworks.com/products/matlab/matlab-and-python.html>

Week 9

* Looking into tsv vs csv, when data wrangling
* Looks like tsv is more efficient in java or for tensor flow, CSV files are used more in fields like machine learning, data analysis, deep learning. Looks like CSV may as well do the job. <https://www.it4nextgen.com/tsv-vs-csv-file/>
* Computed required field size of data in file (number of points)
* Planned a structure to use- considering 2 files or 2 sheets if cannot easily put headers and everything else in. 2 sheets is obvious play. <https://stackoverflow.com/questions/26521266/using-pandas-to-pd-read-excel-for-multiple-worksheets-of-the-same-workbook>
* Looked at xlswrite compared to csvwrite and dlmwrite
* Wrote a working script but not got 2nd sheet to work yet
* Comparing exporting XLS and CSV files, issue of writing data into two sheets. Also considering batch convert from XLS to CSV
* File seize<https://toggl.com/difference-between-csv-xls/>
* No need to convert xls files to csv since multiple sheets needed, also xlsx files take up less memoryyyyyyyyyyyyy- 7.5MB compared to 9.5MB
* Could alternatively fit all into single sheet CSV- but realistically harder to read, larger file space and cannot specifiy where to place data in csv write
* Xlsxwrite is temperamental with filenames though…
* Successfully got data read into xlsx and then into Python, now on to more exciting work 😊
* Started looking at how to import xlsx file data into PostGreSQL- v unclear how I can do this while avoiding the use of csv files and splitting up sheets. Could alternatively make separate csv files for this but it is clearly very desirable to avoid this.

Semester 2

Week 1

13/1

* Reviewed codes, reports, meeting notes etc etc
* Udemy course refreshes for SQL, Machine Learning and Python

14/1

* Experimented with SQL and successfully created database for *some* FASTBLADE data, need to further explore file types for importing data and decide what level of precision is required inside SQL data base
* Also require confirmation as to what architecture of database is required and understand how to best shape database structure
* Need to decide whether to automate xlsx into 2 csvs each time or what is best?
* New timescale plan- aiming to finish objectives 1 & 2 in January, 3&4 in Feb and then write all March
* Started final dissertation report

15/1

* Had meeting- decided rough plan for when to finish each objective, Encarni put me in touch with two of her pHds as well as David Garcia to discuss Machine Learning
* Researched connecting Python to SQL database
* Set up a Mendeley Account

16/1

* Looked at data and began planning structures
* Worked out what Data is still to be outputted
* Observed issue of automating import of csvs into SQL for each dataset given 283 tables
* Practiced exporting databases from SQL
* Checked which inputs changed under different datasets

17/1

* Further explored data looking at what input variables change and found accurate summary of changing variables and collated into a spreadsheet
* Tried exporting as csv as an alternative to xlsx but issue remains of including variables of different lengths, even if 0s are artificially inserted.
* Created list of all single value variables, so that it can be confirmed that none of these are required for the machine learning portion of the project

**Week 2**

20/1

* Extensive coding on MATLAB after successfully wrangling and organising data, put into csvs and ready for upload to Postgre SQL. Successfully automated uploading for different file names
* Achieved Objective 0
* Jeff meeting
  + Discussed transfer function somewhat
  + Discussed sigcomp
  + Discussed some next steps
  + It is acknowledged that full data sets are required to be uploaded to SQL for posterity and in the event of future additions to the database or to the Python Study
  + Discussed phase lag, gain, timelag etc and sigcomp- and a path to a transfer function, bode diagrams was made with what outputs was desired specified for statistical analysis comparing load desired and measured made

21/1

* Read through control 3 notes- specifically focusing on bode plots and frequency domain content while also looking at fast fourier content
* Read some of the textbook about linear systems
* Finally carried out mass export of data into csvs
* Uploaded first two datasets into sql- had issues with file permission and the number of columns in the dataset. Looking good though, awaiting confirmation form jeff as to desired total force term / measured term
* Repracticed exporting database just to make sure
* Upon confirmation from jeff, reuploaded all csvs overnight (correctly this time)

22/1

* Successfully uploaded INERA files (26) into SQL in 10 mins. Looking like can upload rest in 2 hours at some point
* Successfully created a Python script that connects to my SQL database and can perform sql queries on the data as well as extract the data as panda dataframe
* Worked some more on Python code- created initial populations, planned out next steps in code development and researched genetic algorithms

23/1

* Uploaded entire SQL database and exported backup and saved file within fastblade and my laptop.
* Objective 1 complete

Week 3

27/1

* Began writing genetic algorithm from scratch on python
* Encountered significant obstacles
* Began rewriting rams code functions to get into a form where it could all be called in a main function, with every function only having the argument (sol)

28/1

* Rewrote a number of functions within rams and tidied code significantly
* Got a number of solvers to work, and read up on each one. Plotted the best solver for computational time vs number of rams input for a fixed number of nodes
* Met with jeff, showed python code (objective 2), sql database (objective 1) before fixing sigcomp and determing a strategy for finding the lag between signals

29/1

* Sorted indexing problem with Target Bending improving efficiency and ensuring accuracy
* Practicing creating dataframes and exporting to csv
* Improved plots and added bending to output
* Confirmed the repeatability of results from solver at different inputs of number of rams
* Rerecorded values for computational time against rams now that its faster
* Confirmed that each soln is a local minima