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

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.