Technical Computing Software project instructions 2023

General

The purpose of the project is to do an advanced study on one of hte topics. That is you will need to study the topic in slightly more depth than in weekly exercises. You can use tools of your choice, but if you use an advanced tool, to avoid the "I put numbers in this box and out came the answer" effect you will need to find out how the tool works or the very least its operating principle.

The project has two deliverables

1. Project report – background, method description, interpretation of results (returned using the TurnItIn tool)
2. Practical part (return source code and data files)

The intended order of completing the project is as follows: A brief literature study (1 – 2 sources on the topic is enough) looking for potential real-world solutions to the problem, and potentially looking for a method for the simplifications in this project. Don't reinvent the wheel if there exists a solution. Remember to take notes of your sources and cite them in your report. In general, we are not looking for very high level research on a freshman year course, but lecture slides are not a suitable source. (The textbook is available for "the public" and is suitable for citation – usually entry-level textbooks are not cited, but since this is an entry level project, they are topical.) In what ways the problem is simplified? How this might affect the solution? At a high level, describe a method you could solve the problem. At this point you have two chapters, titled "introduction" and "methods". Implement the method you have described (item 2, practical part) and describe your results. After that, consider the realism and reasonability of the method and results.

Your project report should contain the following parts (**titles in bold**):

Introduction: What this study is about? Has someone written about the same topic? What is the phenomenon that is being studied, and is there a need for it to be studied? How we could solve the problem or improve on the situation? What is a more accurate description of the problem and what kind of data might be available? In a software engineering point of view, you are doing requirements engineering. The problems are described in a rather open-ended manner, so also scope your problem. (That is, any assumptions you might want to make for your work.) If you do a longer literature review, it could be placed into a subsection such as 1.1 titled "background" or "literature review". For example, try searching the internet for the title of your work.

Methods: What method is being used to solve the problem? You should do an accurate description of the method here. If you did not invent the method yourself, you should try to find a source that supports your description of the method. In case of a more complex method (e.g. inner workings of the ode45 solver) a "popularized" explanation of what is happening is enough (e.g. "how Runge-Kutta methods work"). In software engineering terms, this is the design phase. Note that if your topic is e.g. "optimizing production", this chapter is not about telling about production methods (that would be in the introduction/background), but by telling about the optimization methods used. Also tell how you selected this specific method and what tools are you going to use to implement it. If you include figures or formulae, you should reference them in text.

If you notice during the practical implementation that you need to differ from the planned method, update your methods chapter and if necessary, your introduction to correspond to the real implementation. The following chapters discuss with the previous two – in "results" you should report what happens when the methods were implemented, and "discussion and conclusions" mirrors the results to the background presented in the introduction. For example, if you have the "effects of packaging on food carbon footprint" topic and have found out that food causes a certain amount of carbon dioxide emissions, how much would be the difference between the worst and best option on a global scale?

Results and analysis: What happened when the "methods" described were implemented? Depending on the kind of results you got, the analysis can be under "results", its own chapter (specially if quantitative analysis), or then some of the analysis could be in the "discussion" chapter (specially if qualitative analysis). Avoid single-paragraph chapters. Possible questions you may want to answer here: What is the best option? Why is it the best option? If you have fit a model, then under what kind of conditions the model works or does not work? Is there an explanation for the error cases (which usually are not your errors – all models are simplifications.) In software engineering terms, we are reporting on the implementation and testing of your project. Also note that "the fit function was $y = ax^2 + bx + c$" is not analysis of your solution, but "the fit model displays an increasing trend because $b > 0$, but the growth is slowing because $a < 0$" is (assuming supported by data, of course!) You may want to end with a short summary of the results, such as "due to simplification, it was assumed that matter X is not significant, which might have an impact on…" Figures and tables fit well in the results chapter, though if you have a lot of data, you should return it as an appendix (with the rest of your source and data files in the practical part).

Discussion: Discuss your results. Does the problem statement and results match a realistic case? Could the method be generalized? If you had a different scenario, how it would affect the solution? If you did not find a solution or think your solution is wrong, what should the correct solution look like and can you explain why the method did not work? If you found an existing study on the same topic, what differences and similarities were there between the project work and "real research"? These topics and concepts should be introduced in the introduction chapter.

Finally write a short **abstract** of your project. This is placed in the beginning of the report. Your first part of the report is thus written last. Return the report and practical source and data files. The report follows so-called IMRAD format (Introduction, Methods, Results, (Analysis) and Discussion). Your discussion might include as a final paragraph the conclusions of your work which corresponds to your abstract. In some templates, the conclusions are its own, short chapter.

The project reports will be automatically checked for plagiarism. There is no fixed passing score for plagiarism, but in practice you should not have directly copied paragraphs (unless it's a quote – ask for instructions how to quote an entire paragraph). Neither you should have a "patchwork" chapter (entirely composed of text copy-pasted from different chapters).

You may use the thesis template, but you may format the project report as you wish. Font size should be a reasonable 11 – 12 points. Write all formulae using the formula editor, not by taking photographs from a textbook etc. You don't need to have a title page (but you may have one). On the first page you should name your title, name, student number and course. (In an actual thesis, you don't mention your student number.) Assistance on how to complete the project will be available in exercises and office hours (remote exercise). The length of the report should be 3 – 5 pages (approximately 2 – 3 pages for introduction and methods, 2 - 3 pages for results and discussion). Possible title page, abstract page, table of contents, list of symbols, terminology, list of figures, list of tables or list of references are not counted on the page count.

When do I get my data?

You may start working on your project introduction and methods immediately. On the project work week we will open the initial value generator.

Grading

The project will be graded 0 – 10 points. Pass limit is 5 points. 10 project work points corresponds to 50% of your course grade.

In "introduction" you must tell, what is the problem you are working on, and what significance it might have. (1 point) You should also find out how similar "real" problems are solved (+1 point). Use this chapter also to scope your problem: verbally describe what exactly you are going to do.

In "methods" you must tell what are your variables, and at least generally describe your method. If the method needs initial values and stopping conditions, you should also mention this information. (1 point) You can also present a more accurate and formal description of the method you are using (+1 point).

In "results" you must present your results and at least briefly analyse them. (1 point) You also should compare your results to another method, perform error estimation or otherwise do further analysis. (+1 point)

In "discussion" you should have a wider analysis of the results (if it is not in "results and analysis"). You should consider the significance and realism of the results and the problem statement that lead to these results. (1 point) You should also look back to your introduction, e.g. if your introduction claims that it is possible to save 20 % of something, do your results support this claim? (+1 point) If you suspect that your solution is deficient, this is also good to mention in your discussion along with your reasoning leading to this conclusion.

The structure of the project work is graded yes/no (does it follow IMRAD and is approximately the required length, 1 point) as well as if appropriate sources have been used (1 point).

Topic 1: Simulating a RLC circuit

Brief overview of the problem: You have been given the properties of a circuit connected in series: Resistance, capacitance, impedance. The components are fed alternating current with the given voltage and frequency. The voltage signal is of sine wave form, $V_{rms} = \frac{V_{peak}}{\sqrt{2}}$. Simulate the properties of the current (amplification/dampening, phase shift) compared to the voltage signal.

In practice: Simulate the circuit using any (numerical) solution method. Learning to use SimScape Electric (Simulink add-in) is one possibility, or you may solve the differential equation numerically. The circuit is extremely simple: Just one resistor, coil, and capacitor, connected in series. Your data will consist of individually selected resistance, impedance, capacitance, voltage amplitude and phase at the beginning of the simulation.

Things to pay attention to your report: You should describe the method (mathematics involved) used to estimate the behaviour of the system, as well as why you selected this method. If you did not program the solution yourself but used some "commercially available software" you should find out what tools were used to develop the software or what is its operating principle.

One way to achieve your results is to draw the voltage and current graphs and compare the graphs to find your values. Error estimation is possible by using another method or tool, or solve the behaviour of the circuit analytically, and comparing the solutions.

If you want to sees a stronger shift, you may change the order of magnitude of your (randomly assigned) values. For example, the phase shift depends on the ratio of the capacitance and impedance of the circuit, and with just a resistor you would have $U = RI$. The reasonability of the values could also provide some material for your discussion chapter – due to random selection, some initial values given to you might not really make that much sense.

In the past years, the theory section in this topic has been done excellently. Do remember to define your units (in SI system) as well as to report the solution method (e.g. the principle of Runge-Kutta methods). Simulink documentation is available e.g. https://www.mathworks.com/help/simulink/ug/simulating-dynamic-systems.html.

Topic 2: Predicting weather

Brief overview of the problem: You are given a Finnish Meteorological Institute Site ID (FMISID). You can find the corresponding site name:

https://en.ilmatieteenlaitos.fi/observation-stations

Using the station name you can find the measurements at:

https://en.ilmatieteenlaitos.fi/download-observations

Consider:

1. How to deal with missing data?
2. I've made the list of stations based on "operating since" statistic – occasionally there's a lot of missing data. You may ask for a new station.
3. What will you try to predict? Maximum temperature, mean temperature, lowest temperature, something else?

The objective is to fit the model to historical data (training set, use data before April) and predict roughly what happens during April (test set, also acting as our control for the model). (If you want to make a prediction using some other data set, please contact the teacher.)

Practical part: Load a roughly 10-year data set from the FMI web pages (or other data source) and predict what happens in April 2022. (This project goes well with getting introduced to fitting tools in commercial packages such as Matlab, Excel and more.) Pick something you want to predict and see if you can predict it using earlier data. Use actual data from April to see what kind of an error your prediction had.

The baseline challenge in this work is to predict:

1) long duration variance – how years differ from each other and how this year is different
2) variation within the year (the cycle between seasons, summer and winter)

There are very many approaches to the task and different methods produce different results. The real weather prediction methods work by physically modelling the atmosphere, calibrate this model using measurements from several stations and then trying to extrapolate how the atmosphere moves. Simpler methods can, however, produce somewhat accurate results as long as we limit our prediction window. (One month, admittedly, is really pushing it.)

At least from your long duration model you should try a few different methods or at least degrees of functions. You should be able to measure somehow how good your fit is or what is your error. The common way is calculating mean square error (MSE): Calculate the error in every point of comparison, take the square

of the errors, and calculate the arithmetic mean of the squares. Sum of squares (SSE) is also used, but is not very well suited to measurements with different number of data points.

The variance within the year might be doable to fitting some model to an "average year", or trying to fit some periodical function (sines, cosines, or a combination) to the entire time series data. By combining your long duration model and yearly variation model you should be able to have some prediction on your phenomenon. It is probably also possible to use Fourier series (sum of sines and cosines) to model the phenomenon.

A possibility is also to consider weather as a set of probabilities and try to deduce the likelihood of some scenario, such as "does the sun shine on May Day?"

Remember to tell, what you are actually predicting, based on what information. What kind of data you can use? What kind of a model are you using? How are you taking into account the differences between years? How about difference between summer and winter? How do can you estimate the error? How do you fit the coefficients of your model? Can you estimate how long you can predict to the future?

Topic 3: Optimizing production

Problem in short: A machine workshop has several different machining tools, that can be used in the production of a certain product. When the objective is to maximise production, how should the various phases of production be organized for the machines?

Practical part: To produce a product, certain operations need to be performed using the machines. The given operation can be done using one or more machines. The time taken by each operation depends on the machine. In addition, it takes a certain time to set the product up for the machine operation. Define the optimization problem and its limitations. Familiarise yourself with on optimization method and implement it yourself, or familiarise yourself with the functioning of a optimization tool.

In those cases where one of the production phases requires very little time, you may consider how it might effect future development in a realistic case, but you don't need to consider e.g. profit from other machines needing to cover the loans taken to spend the machine with much slack. Of course, in a realistic situation the company might either want to specialize to a specific product or to increase its product portfolio, depending on the circumstances.

Describing the optimization method as "we used Excel solver" or "we used fmincon" is not considered to be enough. You should look up adaptive gradient methods or the simplex method. Excel "GRG nonlinear" solver uses a so-called "generalized reduced gradient". In the case of Fmincon see Matlab documentation https://www.mathworks.com/help/optim/ug/choosing-the-algorithm.html ("interior point method" described in https://www.mathworks.com/help/optim/ug/constrained-nonlinear-optimization-algorithms.html#brnpd5f). It is a so-called primal-dual method. The abbreviation KKT refers to gentlemen named Karush, Kuhn and Tucker.

Topic 4: Effect of packaging and waste to food CO2 emissions

"Brief" overview of the topic: You are given the daily consumption of the product for the expected target group (in 100g portions per day). The goal is to select a packaging material which keeps the products edible at least the intended amount of days. There are three available types of material, which have preservation coefficients for an unopened package. The food would keep in an opened package for 2 days, so with a preservation coefficient of 4, the package would keep for 8 days. If opened after 1 day, it would keep opened

for one more day. (To simplify, the package controls the amount of oxygen seeping into its contents, reducing organic activity.)

The food itself has a CO2 equivalency factor (how many grams of CO2 is produced for one gram of food). If portions are spoiled, these of course contribute to the CO2 per portion eaten.

The standard packaging sizes are 1, 2, 3, 4, 6 and 8 portion packages, but you may also recommend an ideal package size for the given parameters. The amount of packaging material (g) is listed along with the package size and packaging material CO2 equivalency.

Practical part: Producing the food causes a certain amount of CO2 emissions per portion. In addition, manufacturing the package causes some CO2 emissions. In general, the better the food has been packaged, the better it keeps, though some food spoils before it can be consumed. Compare the package material and size options and deduce the option that causes least CO2 emissions. So both the material and the food cause emissions, but you should calculate emissions per consumed portion of food. Spoiled food increases emissions while not providing consumed portions.

The relative preservation provided by the package is an actual statistic used. If the factor for material 1 is 4, food keeps $2 \times 4 = 8$ days from when it was bought. To pass, you don't necessarily need to take into account what happens when the package is opened, but typically this additional preservation depends on either some sort of protective modified atmosphere or vacuum, so in real cases opening causes to food to spoil at an accelerated (or normal?) rate. For the reference, home-made vacuum packaging machines have a factor of 3 − 5, depending on how well the package is sealed and how it is being kept.

Clearly describe the function you use to model how the food keeps and explain its factors.

You do not need to compute all options, if you notice that one option is clearly unfeasible and can explain your reasoning to abandon that option.

You may discuss what factors (better preservation, different needs for food) would change the package material. What would be the ideal package size? The values are random-generated, so it might be that some material is simply superior to the others.

In the previous iterations of the course, the background information has been done well. Note that there may be values which allow you to ignore some options or entire materials – but remember to write down your reasoning for this. You probably can deduce or use brute force to compute the remaining options. It might be possible to apply continuous methods and the round (for example when considering ideal package size). How do you describe your objective function?

Topic 5: Neural network artificial intelligences

This topic requires you to introduce yourself to a special field in technical computing we don't have time to deal with on the demos, but may be of interest to some students. You should chiefly conduct a literature study and perform a small demo using some pre-existing implementation.

Neural networks are a form of computing which emulates our current understanding of how the neural system of animals works. These are used for many pattern detection and classification tasks, colloquially known as "artificial intelligence". Introduce yourself to their operating principle (describe how an artificial neuron works and how to create networks out of them) and define yourself a simple problem that you solve. You may use some ready algorithm (such as, but not necessarily Matlab Deep Learning Toolbox) and try to find out how the network solves the problem.

Topic 6: Suggest your own topic

If you are interested to implement some modelling or computing task of your own, please contact the teacher.