Math 9 2025 Spring Final Project (Section A & B)

Numerical simulation of a COVID-19 SIR model

This final project serves as the open note programming final exam of this course, counting 20 points (20%) in the total grades. The "step size" in grading is 0.5 points, meaning that whenever one requirement below is not satisfied, at least 0.5 points will be subtracted.

Due Date: 11:59 pm (PST) June 11th, Wednesday. NO EXTENSION. <u>Graded as 0% automatically for any submission after this deadline.</u>

Submission: write your report in ONE single MATLAB live script file and upload it to Canvas. The submission should be a well-organized and well-written report, which means:

(1) All sections are well-structured, with appropriate section titles.
(2) All figures are clear and have the necessary figure captions (or descriptions) as text.
(3) Have necessary descriptions/answers for the required parts/questions.
(4) Have basic comments for your MATLAB codes providing an overview of each code block.
(5) All code blocks should already be executed, have the required outputs, and suppress the unnecessary outputs.
(6) All formulas should be presented in the text blocks using the "Insert - Equation" function.
(7) All functions should be in-script functions (defined at the end of the live script file).

<u>Please include everything in one single live script file (.mlx file)</u>, any other formats (.pdf, .doc, .m file, Jupyter notebook, …) or redundant files are not valid and will be graded as 0% automatically. Submitting merely the codes and/or incomplete results will severely impact your grades.

Your report should address all the tasks listed in this file.

Task 1 (0.5pt)

1. (At the beginning of your report) Give your report an appropriate title reflecting its scientific contents (don't simply name it as "*Math9 Final Project*").
2. Write your student id and name under the title.
3. Rename your live script file to "*Math9_FinalProject_name_ID.mlx*".
4. Acknowledgeing the final project policy by reading and **copying** the following paragraph (displayed *<u>in Italic font</u>*) at the beginning of your report (as a separate section) and keeping the requirements in mind.

I hereby confirm that the present report is my own work. If any texts or codes from books, papers, the Web, or any other sources have been copied or in any other way used, all references have been acknowledged and cited. I am aware that the instructor has the authority to report and investigate any potential academic misconduct in this course, and such misconduct could lead to failing this course. If I have employed any AI models (including but not limited to ChatGPT), I have appropriately and explicitly acknowledged their use in the report whenever the results were generated with the aid of AI.

Task 2 (3pt)

1. Search and read online resources and articulate, then use your own words to briefly discuss what a mathematical model is (does not have to be the SIR model), and its significance in the study of life science (100~300 words).
2. Read the first 33 pages of the provided material (you can also find it here), which should provide you a good introduction of the SIR model.
   Additional materials (optional):
   2.1. Compartmental models in epidemiology –the SIR model part.
   2.2. SIR Modeling –the first four sections.
3. Assume the total population N is constant, and divide N into three groups: susceptibles, infectives, and recovered/removed. Then we denote:
   $S(t) =$ population of the susceptible at time t;
   $I(t) =$ population of the infective at time t;
   $R(t) =$ population of the recovered at time t.
   Write the following standard SIR model on your live script file using "Insert -- Equation" function:
   $$\frac{dS}{dt} = -\beta \frac{I}{N} S,$$
   $$\frac{dI}{dt} = \beta \frac{I}{N} S - \gamma I,$$
   $$\frac{dR}{dt} = \gamma I.$$
   And explain the biological meaning of all the terms, variables and parameters.
4. Answer the following questions:
   4.1. What is the biological meaning of the term $\beta \frac{I}{N} S$? Why does it have the negative sign?
   4.2. Why does the right-hand side of the third equation only have one term? What does it mean?
   4.3. One of the major assumptions of this standard SIR model is: $S(t) + I(t) + R(t) = N$ is a constant. Practically speaking, under what circumstances can we make such an assumption?
   4.4. The basic reproduction number is defined by $R_0 = \frac{\beta}{\gamma}$, what is its biological meaning? And what will happen if $R_0 > 1$?
   4.5. Search for online materials and provide an example (from news/articles/research papers) that during the COVID-19 pandemic, $R_0 > 1$ held for a quite long time.

Task 3 (6pt)

Consider the following situation: people in a small town (total population: 10000) are experiencing the flu season, and today's record shows that 10 people have already recovered, and 200 people are having the flu. Assume that active immunity is obtained once fully recovered, and the newborn and death can be neglected during a short period of time. Here we also assume that the rate of infection when contacting a patient is 75%, and every day a patient has a probability of 15% of becoming fully recovered (rate of recovery).

1. Based on the above information, write down the SIR model, clearly indicate the initial condition and parameter values.
2. Use *ode45* (a MATLAB built-in toolbox, very similar to the *ode23*) to simulate the trend of these three groups within the next 45 days (time unit: day). When writing the function for this SIR mode, name it "*simple_SIR*", and make sure the parameters β (beta) and γ (gamma) are also part of the function inputs.
3. Plot simulation results in one figure using the *plot* function.
4. Polish your plot, add a figure title, give appropriate names for both two axes, and change the labels of the legend to 'S', 'I', and 'R'.
5. Interpret your simulation result, briefly explain what may happen in the next 45 days.
6. What is the basic reproduction number in this case? What is the basic reproduction number if the infection rate is only 10%? What will your simulation be like? Plot the simulation results in a separate figure and provide a brief interpretation.
7. Assume we can reduce the infection rate to 40% by keeping social distance and increase the recovery rate to 25% by using better medications. What will your simulation be like? Plot the simulation results in a separate figure and provide a brief interpretation.

Task 4 (2pt)

In the earlier scenario, we assumed that the infection and recovery rates were constant, which is not entirely realistic. For example, the rate of infection is influenced by various factors such as mask usage, proximity, the health condition of individuals, and more. Instead of introducing new variables to account for these uncertainties, we incorporate randomness into our model to capture these fluctuating elements. Here, we consider the infection rate as a random variable ranging from 30% to 90%, and the recovery rate as a random variable spanning from 10% to 30%. Both rates follow a uniform distribution within their respective ranges.

1. Based on the above assumptions, write a new function for the improved SIR mode and name it "*simple_SIR_2*". You should generate the random infection and recovery rates <u>inside</u> your "*simple_SIR_2*" function.
2. Use *ode45* to simulate the trend of these three groups (S, I, and R) within the next 45 days (time unit: day).
3. Run your simulation 20 times and plot all results in the same figure. Color your curves of the three groups (S, I, and R) in blue, red, and yellow, respectively. (Your figure should have an appropriate title, axes labels, and a legend.)
4. Interpret your simulation results.

Task 5 (4pt)

The provided file "ArizonaData.mat" records the COVID-19 data in Arizona from 5/23/20 to 8/30/20: column 1 – daily new cases; column 2 – total death. (If you want to learn more about this data set, you can visit here.)

For simplicity, we assume this COVID-19 pandemic happened in an area with a total of 50,000 population (fixed). In Task 5, we do not consider the total death.

1. Input the above data into MATLAB (your live script) and visualize the daily new cases data using bar plot (you can use the 'bar' function). In your figure, the labels of x and y axes should be "Day" and "New cases", respectively.
2. We assume all of those who tested positive were considered susceptible from the beginning until the end of that day before becoming isolated and taking medications. Then the total number of infectives people on day x equals to the same day's total new cases.
3. Use the standard SIR model ("*simple_SIR*") and *ode45* to simulate the trend of these three groups (S, I, and R) from day 1 to day 100 (time unit: day). The initial conditions: total population = 50,000, total infectives = new cases on 5/23/20 (day 1), total recovered/removed = 200. Infection rate = 0.25, recovery rate = 0.15. Plot your simulation result.
4. Plot the "daily new cases" data and your simulation result of the infectives group in one figure. (When plotting the "daily new cases" data, the first point should be at day 1, the second point is at day 2, …, and the last point is at day 100.) This is also called "data fitting", briefly discuss your fitting results.
5. Calculate the absolute error between your simulation and the "daily new cases" data at each corresponding time point and summarize all the errors (name it "fitting_err").
6. Report the fitting_err of the following simulations: (1). infection rate = 0.25, recovery rate = 0.13; (2). infection rate = 0.24, recovery rate = 0.15; (3). infection rate = 0.24, recovery rate = 0.13. Do not plot the figures.

Task 6 (3pt)

If we consider the death caused by COVID-19, we can improve our SIR model in a following way:

$$\frac{dS}{dt} = -\beta \frac{I}{N} S,$$
$$\frac{dI}{dt} = \beta \frac{I}{N} S - \gamma I - \delta I,$$
$$\frac{dR}{dt} = \gamma I,$$
$$\frac{dD}{dt} = \delta I.$$

Here $D(t)$ = total number of deaths at time t, δ is the death rate. Similarly, we consider this COVID-19 pandemic to happen in an area with a total of 50,000 population (fixed).

1. Improve the standard SIR model accordingly and name the new function as "*simple_SIRD*". Then use *ode45* to simulate the trend of these four groups (S, I, R, and D) from day 1 to day 100 (time unit: day). The initial conditions: total population = 50,000, total infectives = new cases on 5/23/20 (day 1), total recovered/removed = 200, total death = total death on 5/23/20 (day 1). Infection rate = 0.22, recovery rate = 0.12, death rate = 0.016.
2. Plot the "daily new cases" data and your simulation result of the infectives group in one figure.

3. Plot the "total death cases" data and your simulation result of the death group in one figure.
4. Calculate the fitting_err of 2. & 3., respectively.

Task 7 (1pt)

Use 100~300 words to briefly discuss what you have learned about the SIR model, and how numerical simulations help you understand the model and the control of an infectious disease. And more importantly, raise at least three limitations of this model (i.e., how to improve the model to make it more practical). You do not need to write any MATLAB code in this section.

Task 8 (0.5pt)

1. The title should use "Title" font.
2. Format each task (from Task 2 to 6) into one section, separate by section break, and give an appropriate section title (using "Heading 1" font).
3. Add one more section titled "References", and list all the materials (including AI models) that you used (you do not need to list the materials provided in Task 2).
4. Add one more section titled "Acknowledgement", and list all the people (besides the instructor and TAs) that helped you on your final project.
5. Execute all your code and only show the required outputs.

Optional Task (Ungraded, but will be taken into consideration for an A+)

1. In **Task 5**, we experienced how to manually tune the parameter values to make our simulation a better fit (i.e., a smaller fitting error) for the data. However, it is almost impossible for us to find the parameter values that can minimize the fitting error, especially when we have multiple parameters. Luckily, MATLAB has a built-in toolbox named *fminsearch*, that can help us minimize the fitting error. Learn how to use *fminsearch* and apply it to our **Task 5** to find the parameter values of β and γ that minimize the fitting error and plot the corresponding result.
2. Recall in **Task 2**, we mentioned that one of the major assumptions of the standard SIR model is that the total population is a constant, which is also called "SIR model without demography".
   2.1. Learn the "SIR model with demography" via online materials and develop the corresponding model with "birth rate = mortality/death rate = μ" and interpret the model (similar to **Task 2**).
   2.2. Use *ode45* to simulate the trend of these three groups (S, I, and R) from day 0 to day 100 (time unit: day).
   2.3. Use *fminsearch* to find the optimal parameter values that provide the best fit of the "daily new cases" data (**Task 5**) and plot your result.