Group Project

Due: Friday, December 6, at 11:59pm.

Project Objectives

The project will focus on modeling the spread of an infectious disease using the SIR model (Susceptible-Infected-Recovered). The SIR model contains a set of ordinary differential equations (ODEs) that describe how individuals in a population transition between three states: susceptible, infected, and recovered.

The goal of this project is to provide students with a deeper understanding of how numerical methods are applied to solve real-world problems in epidemiology. By simulating the spread of infectious diseases using the SIR model, students will explore how mathematical equations can predict epidemic dynamics and inform public health decisions. Through the integration of solving ODEs, interpolation, data fitting, and spectral analysis, students will gain practical experience with tools that bridge theory and application, empowering them to tackle interdisciplinary challenges in science and engineering.

The project will integrate the following topics:

1. Numerical solution of the ODE systems.
2. Interpolation.
3. Least Squares.
4. Fourier Analysis.

Project Preparation and Submission

The students will work in the groups of four. The students are to divide the work equally between the team members. All programming need to be done in MATLAB. Built-in MATLAB functions for ODE solution, least square sand interpolation can not be used in this project. All programs need to be self-written. The students can use built-in MATLAB *fft* function for Fourier analysis. The project needs to be created and managed in GitHub. Please, see the instructions on how to use GitHub posted on Canvas.

How to submit

The projects need to be submitted through Gradescope. Only one submission per group is required. While submitting the project on Gradescope, please link all the group members during the group submission. All files need to be saved into a single .pdf document. One .pdf document from group needs to be uploaded. The .pdf documents needs to include:

1. All published MATLAB codes.
2. All plots as required by assignment.
3. All discussions and answers to questions (typed).
4. Documentation of GigHub collaboration, including:
   (a) A screenshot of the main page of the project showing all uploaded files.
   (b) A screenshot showing the content of the README file

(c) A screenshot of "Insights/Contributors", showing the contribution activity of all members of the group. Additionally, for each contributor, click on "commits" and include screenshots of detailed history of contributions for each member of the group.

Project Description

Part I: Modeling disease spread using SIR model (50 points).

The Susceptible-Infected-Recovered (SIR model) describes the dynamics of a population divided into three compartments: **S** (susceptible), **I** (infected), and **R** (recovered). The system of differential equations governing this model is:

$$\frac{dS(t)}{dt} = -\frac{\beta}{N}S(t)I(t) \quad (1)$$

$$\frac{dI(t)}{dt} = \frac{\beta}{N}S(t)I(t) - \gamma I(t) \quad (2)$$

$$\frac{dR(t)}{dt} = \gamma I(t) \quad (3)$$

Where:

1. $S(t)$ is the number of susceptible individuals at time t,
2. $I(t)$ is the number of infected individuals at time t,
3. $R(t)$ is the number of recovered individuals at time t,
4. $N = S(t) + I(t) + R(t)$ is the total population (assumed constant),
5. β is the transmission rate (how often a susceptible individual gets infected),
6. γ is the recovery rate (how often an infected individual recovers).

The task of Part I will be to implement the ODE solver and simulate the progression of the disease for different parameter values (β, γ). The solution will provide time-dependent functions for $S(t)$, $I(t)$, and $R(t)$, which describe how the disease spreads and eventually fades as more people recover. **Steps for this task**:

1. Solve the above ODE system numerically with 4^{th}-order Runge-Kutta method using the following parameters:
   (a) Time step $h = 1$ (day).
   (b) Initial conditions: $S(0) = 990$, $I(0) = 10$, $R(0) = 0$.
   (c) Parameter values (β, γ) equal to:
      i. **Seasonal Influenza**: $\beta = 0.3, \gamma = 0.1$
      ii. **COVID-19**: $\beta = 1, \gamma = 0.1$
      iii. **Measles**: $\beta = 2, \gamma = 0.2$
   (d) The total simulation time of $T = 100$ days.
2. Provide plots of $S(t)$, $I(t)$, and $R(t)$ for each parameter combination. For each (β, γ), include $S(t)$, $I(t)$, and $R(t)$ on the same plot - 3 plots total are required.
3. Provide a discussion of how the values of parameters β and γ effect the results. Explain whether your results make sense intuitively, taking into account what parameters β and γ physically represent in a disease spread model.

Part II: Interpolation (50 points).

The goal of this task is to observe how a lack of data recovered by interpolation compares to the data that can be obtained from a model itself, which is more accurate.

The steps for this task:

1. Take a baseline case of Seasonal Influenza ($\beta = 0.3, \gamma = 0.1$).
2. Run it for the total time of T=100 days but using the coarser time step of $h = 2$ days.
3. For every odd number of days (Day 1, 3, 5, etc.) for which you don't have a data from the model, use linear interpolation (you can use Newton or Lagrange form according to your preference) to recover the missing data.
4. Compare the interpolated data with the data obtained from the model in Part I and compute L_2 error as

$$E_{L_2} = \sqrt{\frac{\sum_{k=1}^{N_{int}} \left(V^{int}(t_{2k-1}) - V^{model}(t_{2k-1})\right)^2}{N_{int}}} \tag{4}$$

   for all $S(t)$, $I(t)$, and $R(t)$, where

   - N_{int} is the total number of interpolated points,
   - $V^{int}(t_{2k-1})$ is the interpolated value at the interpolated point t_{2k-1} (all interpolated points fall on odd days),
   - $V^{model}(t_{2k-1})$ is the value obtained by running the model with $h = 1$ in Part I evaluated at the same point t_{2k-1} as the interpolated point.

5. Repeat the same steps for quadratic interpolation (you can again use Newton or Lagrange form according to your preference) and compute the error.
6. Make a 3×2 table documenting error in each quantity for linear and quadratic interpolation (use $S(t)$, $I(t)$, and $R(t)$ as columns, and a form of interpolation – linear or quadratic – as rows of the table). Comment which form of interpolation provides smaller errors.

Part III: Least Squares (50 points).

The goal of this task is to learn how we can estimate the values of the model parameters based on the observed data. A slight difficulty is that the SIR model given by Eqs. (1)–(3) is non-linear. However, we can make an assumption that S is constant (valid early in an epidemic) and equal to S_0. Then Eq. (5) simplifies to:

$$\frac{dI(t)}{dt} = \left(\frac{\beta S_0}{N} - \gamma\right) I(t), \tag{5}$$

which has a simple analytical solution

$$I(t) = I(0)\, e^{kt}, \tag{6}$$

with

$$k = \left(\frac{\beta S_0}{N} - \gamma\right). \tag{7}$$

Taking the natural logarithm of both parts of Eq. (6) yields:

$$\ln I(t) = \ln I(0) + kt. \tag{8}$$

The objective of this task will be to apply linear least squares to Eq. (8) and estimate the value of the model parameter k as well as the initial conditions on the infected population $I(0)$ by comparing the linear model predictions to the observed data. To simplify the problem, the observed data will come from a **full non-linear model** (with a non-constant $S(t)$) from Part I.

The steps for this task:

1. Run a non-linear SIR model from Part I using $h = 1$ day for 30 days with the value of parameters $(\beta = 0.3, \gamma = 0.1)$. This will be your "true" data $I(t)$. Use the same initial conditions as in Part I.
2. Apply linear lest squares to the linear model of Eq. (8) with $t = [1, 2, 3, \ldots, 30]$ corresponding to 30 days of running the model, and $I(t)$ from the Step 1 of this task (previous item). Estimate $I(0)$. Assuming $N = 1000, S_0 = 990$ and $\gamma = 0.1$, estimate β.
3. Repeat Step 2 of this task, but using only 10 days of data (as opposed to 30 days). Do the estimates of $I(0)$ and β improve compared to the true parameters? Why or why not?

Part IV: Fourier Analysis (50 points).

In this part, we will investigate the influence of the periodic variation in the transmission rate on the disease spread dynamics. In particular, the students will perform Fourier Transform of $S(t)$, $I(t)$ and $R(t)$ and observe the frequency peaks.

The steps for this task:

1. Assume a periodic variation of the transmission rate as

$$\beta(t) = \beta_0(1 + A\sin(\omega\, t)) \tag{9}$$

   with $\beta_0 = 0.3$, $A = 5$ (amplitude), and $\omega = 2\pi \times 365/365$ (angular frequency). This variation correspond to a periodicity of $365/365 = 1$ day and may reflect daily versus nightly variations in the spread cycle (for example, virus spreads more during the day when people are awake and active and less during the night).
2. Run the SIR model with $\beta(t)$ given by Eq. (9), $\gamma = 0.1$, and $h = 0.1$ day (to resolve the periodic fluctuations better) for $T = 30$ days. Use initial conditions $S_0 = 990$, $I_0 = 10$, $R_0 = 0$.
3. Plot the signals $S(t)$, $I(t)$ and $R(t)$ as in Part I. Do you observe any periodic fluctuations in the signals to due periodicity of β?
4. Perform Fast Fourier Transform (the students are allowed to use built-in MATLAB *fft* routine for this task) of $S(t)$, $I(t)$ and $R(t)$.
5. Plot the spectrum for the infected cases $I(t)$ (i.e. the absolute value of the Fourier coefficients $\hat{I}$, versus frequency, f). You can use a similar program as the one that appeared in the worksheet on Signal Analysis. Define your frequency vector as $f = \frac{1}{T}(0 : \frac{N}{2})$, T is the total time length of the signal, and N is the number of samples (or number of samples minus one if the number of samples is odd). Only plot the half of the spectrum (i.e. until $f = \frac{N}{2T}$ as in the worksheet. Observe the frequency peak(s) and comment on what you see. Does it make sense physically?

6. Now repeat all the steps, but with using a lower value of $\omega = 2\pi \times 100/365$. This periodicity roughly corresponds to a period of $365/100 \sim 3$ days and may reflect variations in the disease spread on a weekly basis (for example, the transmission rate peaks twice weekly - in the middle of the week when many people are at work, and on Saturday when many people attend social gatherings). Observe the change in the peak frequency. Does it shift to lower or higher values? Discuss your observations.