Name: Ben Aoki-Sherwood

Collaborators: Altaf Barelvi and Violet Ross

Code: https://github.com/aoki-sherwoodb/CSCI-5897-InfectiousDiseases/blob/main/hw2/hw2.ipynb

1. The goal of this problem is to develop flexibility with your Forward Euler code, and to learn a bit about the effect of step size on the accuracy of the solution.

   a. Using your Forward Euler method, simulate the solution to the *normalized* SIS model discussed in class (Week 3) using $\beta = 3$ and $\gamma = 2$, and with $(s_0, i_0) = (0.99, 0.01)$. Create three plots ranging from $t = 0$ to $t = 25$. On the first, simulate using a step size $\Delta t = 2$. On the second, use $\Delta t = 1$. On the third, use $\Delta t = \frac{1}{2}$. In each plot, show only your solution's $I(t)$ in a red solid line, labeled as "Forward Euler", and then also plot the analytical solution from class in a black dashed line, labeled as "Analytical." Please also set the y-axis range to $[0, 0.5]$.
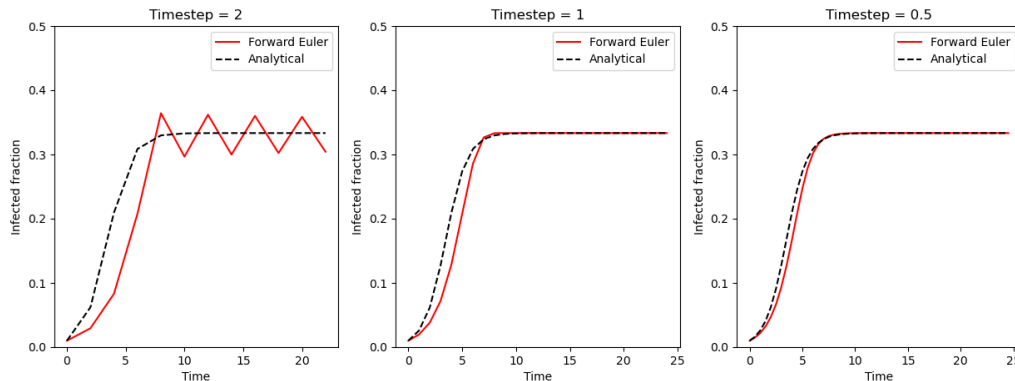

Figure 1: Normalized SIS model simulated using Forward Euler with time steps of 2, 1, and 0.5. The simulated infected fraction is plotted in red and the analytical solution for the infected compartment is plotted in black.

   b. Comment on what you see in your three plots. How does the step size affect our solution?

   The error of the simulated solution decreases as the step size decreases. The simulated solution systematically underestimates the analytical value during the exponential growth phase because the derivative is increasing, so FE is using an estimate of the derivative that is smaller than the actual derivative. Backward Euler would overestimate the solution during this phase. When $\Delta t = 2$, the simulated solution oscillates around the analytical equilibrium instead of converging to the equilibrium. However, for the smaller step sizes, the simulated solution does converge to the equilibrium.

c. Define the maximum absolute error for a simulation using a particular Δt as

$$E(\Delta t) = \max_t \left| I_{\text{Euler},\Delta t}(t) - I_{\text{analytical}}(t) \right| .$$

Write a function that runs the appropriate simulation, computes the analytical solution, and returns E without plotting. Share a link to your code for this problem.

See the code cell under markdown header 1c: `https://github.com/aoki-sherwoodb/` `CSCI-5897-InfectiousDiseases/blob/main/hw2/hw2.ipynb`

d. Create a plot on log-log axes showing $E(\Delta t)$ vs Δt for values

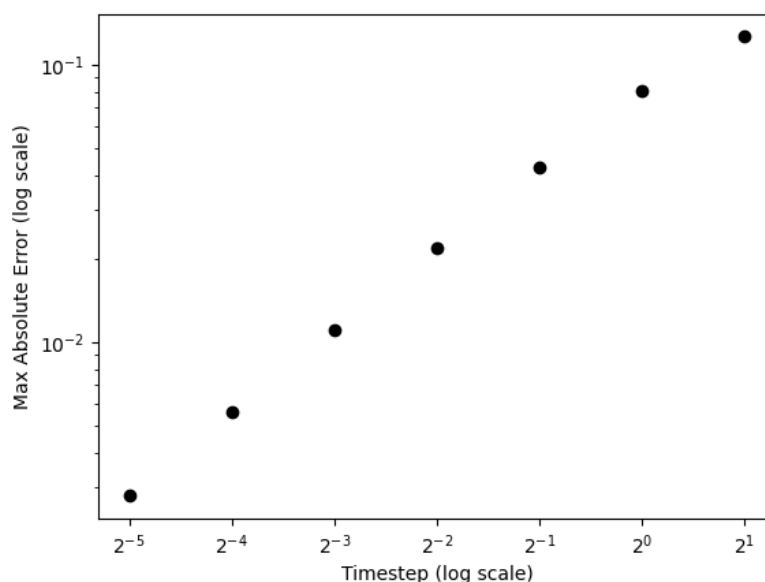$$\Delta t \in \{2, 1, \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{16}, \tfrac{1}{32}\}$$

Figure 2: Maximum absolute error $E(\Delta t)$ vs step size Δt on log-log axes.

e. Comment on what you observe in this plot, and comment on cases when you would want a larger or smaller step size, and why? Imagining yourself in an advisory position in your community, can you think of any scenario where there is a connection between the step size of your simulation and the ethics of your advice?

This plot shows that error increases roughly linearly with step size. In all cases, we want to use as small a step size as is computationally feasible to minimize error. A good check

is to halve the step size and see if the solution changes significantly; if not, then the step size is small enough. The only case where we might have to use a larger step size is if the model is extremely complex so that simulating many steps takes a long time. In an advisory position, we would always want to ensure that our advice is based on *real* (model) dynamics rather than *solver-induced* dynamics. We could do this by checking the step size as described above or trying a different solver algorithm.

2. The goal of this problem is to get you thinking about the constraints on population contact structure and contact matrices, as well as sensitivity analyses.

  a. As one who is interested in modeling disease transmission on college campuses, you hire two teams to measure contact patterns on a nearby campus. The first team, led by Dan Pemic, tells you that there are 200 faculty and 1800 students, with a contact matrix of

$$C_{\text{Pemic}} = \begin{pmatrix} 3.1 & 43.5 \\ 4.7 & 25.0 \end{pmatrix}$$

  The second team, led by Flynn Uenza, tells you that there are 210 faculty and 1750 students, with a contact matrix of

$$C_{\text{Uenza}} = \begin{pmatrix} 3.0 & 44.5 \\ 4.8 & 25.1 \end{pmatrix}$$

  Whom do you trust more, Dan Pemic or Flynn Uenza? To answer this question, consider the self-consistency (or lack thereof) of each dataset. Explain your reasoning in words and include any calculations used to arrive at your conclusions.

  Dan Pemic's data is more self-consistent. We can calculate the total number student-faculty and faculty-student contacts by multiplying the off-diagonal entries in the contact matrix by the appropriate population sizes. These numbers should be equal because contacts are symmetric. For Dan Pemic's data, this yields

$$4.7 \cdot 1800 = 8460, 43.5 \cdot 200 = 8700,$$

  while for Flynn Uenza's data this yields

$$4.8 \cdot 1750 = 8400, 44.5 \cdot 210 = 9345.$$

  The discrepancy between student-faculty and faculty-student contacts is smaller for Dan Pemic.

  b. A straightforward fix to self-consistency issues is to "symmetrize" the rates. First, we compute the implied total number of intergroup contacts from faculty to students, and then compute the same from students to faculty. After averaging those two counts, divide by the appropriate population size to get per-person rates. Use this approach to symmetrize the two contact matrices.

  From part a., we have the total number of intergroup contacts for each dataset. The average number of intergroup contacts from Pemic is $\frac{8460+8700}{2} = 8580$, and from Uenza is $\frac{8400+9345}{2} = 8872.5$. Dividing by the number of students for the lower left entry and the number of faculty for the upper right entry yields the contact matrices

$$C_{\text{Pemic}} = \begin{bmatrix} 3.1 & 42.9 \\ 4.77 & 25.0 \end{bmatrix}$$

and

$$C_{\textbf{Uenza}} = \begin{bmatrix} 3.0 & 42.25 \\ 5.07 & 25.1 \end{bmatrix}$$

c. How different are these two symmetrized matrices, really? Answer the question by computing the ratio of R_0 under Pemic's data to R_0 under Uenza's data, assuming SIR models with otherwise identical parameters.

We compute R_0 as the largest eigenvalue of the next-generation matrix evaluated at the disease-free equilibrium. Assigning faculty subscript 1 and students subscript 2, this matrix is given by

$$G_0 = \begin{bmatrix} \frac{N_1 p c_{11}}{N_1 \gamma} & \frac{N_1 p c_{12}}{N_2 \gamma} \\ \frac{N_2 p c_{21}}{N_1 \gamma} & \frac{N_2 p c_{22}}{N_2 \gamma} \end{bmatrix}$$

and we compute its largest eigenvalue using the quadratic formula as

$$R_0 = \frac{\textbf{Tr}(G_0) + \sqrt{\textbf{Tr}(G_0)^2 - 4\textbf{det}(G_0)}}{2}$$

where $\textbf{Tr}(G_0) = \frac{p}{\gamma}(c_{11} + c_{22})$ is the trace and $\textbf{det}(G_0) = \frac{p^2}{\gamma^2}(c_{11}c_{22} - c_{12}c_{21})$ is the determinant.

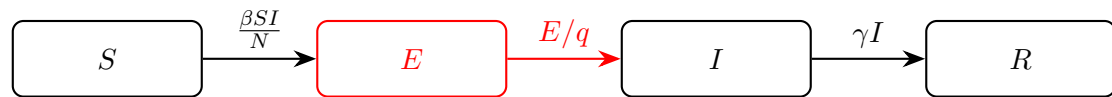**Thus the ratio of R_0 under Pemic's data to R_0 under Uenza's data is**

$$\frac{R_{0,\text{Pemic}}}{R_{0,\text{Uenza}}} = \frac{-\text{Tr}(G_{0,\text{Pemic}}) + \sqrt{\text{Tr}(G_{0,\text{Pemic}})^2 - 4\text{det}(G_{0,\text{Pemic}})}}{-\text{Tr}(G_{0,\text{Uenza}}) + \sqrt{\text{Tr}(G_{0,\text{Uenza}})^2 - 4\text{det}(G_{0,\text{Uenza}})}}$$

$$= \frac{\frac{p}{\gamma}(3.1 + 25.0) + \sqrt{\left(\frac{p}{\gamma}(3.1 + 25.0)\right)^2 - 4\frac{p^2}{\gamma^2}(3.1 \cdot 25.0 - 42.9 \cdot 4.77)}}{\frac{p}{\gamma}(3.0 + 25.1) + \sqrt{\left(\frac{p}{\gamma}(3.0 + 25.1)\right)^2 - 4\frac{p^2}{\gamma^2}(3.0 \cdot 25.1 - 42.25 \cdot 5.07)}}$$

$$= \frac{28.1 + \sqrt{28.1^2 - 4(77.5 - 204.633)}}{28.1 + \sqrt{28.1^2 - 4(75.3 - 214.2075)}} \quad [\text{factoring out } \frac{p}{\gamma}]$$

$$\approx 0.99,$$

so the values of R_0 under either dataset are virtually identical.

3. (Grad / EC): The goal of this problem is to get you to think about additional flavors of models that build on the SIR model backbone, and practice writing down flow diagrams and systems of differential equations. For each of the following situations please (i) draw a flow diagram with the SIR backbone in black, (ii) include any modifications in a second color, and (iii) write out your differential equations using the same color scheme.

   (a) Suppose that we want to model the possibility that the natural history of infection means that a person is infected *but not infectious* before becoming infected-and-infectious. Let the typical *latent* period—the time between exposure and infectiousness—last for q days. Use the letter E for this new compartment. Draw a flow diagram for the non-normalized system, and write a set of corresponding differential equations.

   Flow diagram:

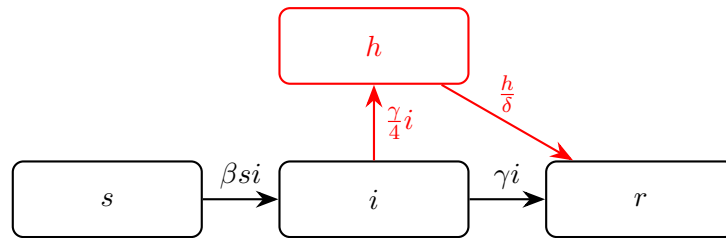   **ODEs:**

   $$\dot{S} = \frac{-\beta SI}{N}$$

   $$\dot{E} = \frac{\beta SI}{N} - \frac{E}{q}$$

   $$\dot{I} = \frac{E}{q} - \gamma I$$

   $$\dot{R} = \gamma I$$

   (b) Suppose that we want to model Hospitalization, with the following assumption: Infected folks either recover directly *or* they are hospitalized first and then recover. Let the direct recovery rate be γ, and suppose that there are 4 direct recoveries for every 1 hospitalization. Let the typical duration of a hospitalization be δ days. Use the letter h for the hospitalized compartment. Draw a flow diagram for the normalized system, and write a set of corresponding differential equations. You should assume that folks in the hospital do not come into contact with anyone else during their hospital stay.

   Flow diagram:

ODEs:

$$\dot{s} = -\beta si$$

$$\dot{i} = \beta si - \gamma i - \frac{\gamma}{4}i$$

$$\dot{h} = \frac{\gamma}{4}i - \frac{h}{\delta}$$
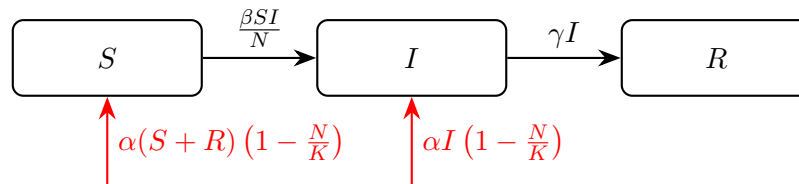
$$\dot{r} = \gamma i + \frac{h}{\delta}$$

(c) Suppose that we want to model an infectious disease that afflicts a growing population of bacteria, such that the infection follows an SIR model, the bacteria grow according to a logistic growth model with intrinsic growth rate α and carrying capacity K. Let all bacteria reproduce, but suppose that susceptible and recovered bacteria produce susceptible progeny, while infected bacteria produce infected progeny. The carrying capacity is shared by all three bacteria.

The total bacteria population grows according to

$$\dot{N} = \alpha N \left(1 - \frac{N}{K}\right),$$

so the number of bacteria born into S will be proportional to $(S + R)/N$ and the number born into I will be proportional to I/N.

Flow diagram:

ODEs:

$$\dot{S} = -\frac{\beta SI}{N} + \alpha(S + R)\left(1 - \frac{N}{K}\right)$$

$$\dot{I} = \frac{\beta S I}{N} - \gamma I + \alpha I \left(1 - \frac{N}{K} \right)$$

$$\dot{R} = \gamma I$$