Problems.

Q1 (model fitting). The SIR model, in its simplest form, models the spread of a disease through a population. The three quantities are:

- \( S(t) \): The number of ‘susceptible’ people that can be infected
- \( I(t) \): The number of ‘infected’ individuals
- \( R(t) \): The number of ‘recovered’ individuals (recovered from the disease)

In addition, define the total population
\[
P = S + I + R.
\]
The system of ODEs in this model\(^1\) are
\[
\frac{dS}{dt} = -\alpha \frac{SI}{P} \\
\frac{dI}{dt} = \alpha \frac{SI}{P} - \beta I \\
\frac{dR}{dt} = \beta I
\]
Let’s assume that initially, there is some known number of susceptible and infected at the onset of an outbreak:
\[
S(0) = P_0, \quad I(0) = I_0, \quad R(0) = 0.
\]
Furthermore, suppose we have a set of data on the number of infected individuals:
\[
(t_k, I_k), \quad k = 0, 1, \cdots, N.
\]

Some (invented) data is given up to 140 days \((t = 0, 7, 14, \cdots, 140)\), taken weekly. The goal here is to estimate the infection rate \(\alpha\) and recovery rate \(\beta\) by fitting the model to the data.

a) Write a function that calculates the least-squares error \(E(\vec{r})\) (where \(\vec{r} = (\alpha, \beta)\)), taking in the relevant data. The data for \(t\) and \(I\) is provided in a text file; use the supplied ‘read’ function. This also supplies you with the initial value \(P_0\) (as \text{pop}).

You will need to solve the ODE system to get the model solution. An \texttt{rk4} routine is supplied for solving the ODE. Use a time step smaller than the spacing between the data points, e.g. if the spacing is \(\delta\) then use \(h = \delta/2^m\). The least-squares error will involve only every \(2^m\)-th value.

\(^1\)The equations vary, depending on the factors you want to include in the model.
b) An implementation of gradient descent is given. You’ll need to modify it to take in only the function $E$ and not the gradient $\nabla E$, since computing $\nabla E$ is not easy.

The fix here is to instead use the approximations

$$\frac{\partial E}{\partial \alpha} \approx \frac{E(\alpha + \delta, \beta) - E(\alpha - \delta, \beta)}{2\delta}$$

$$\frac{\partial E}{\partial \beta} \approx \frac{E(\alpha, \beta + \delta) - E(\alpha, \beta - \delta)}{2\delta}$$

in place of the derivatives. (You can put this directly in the gradient descent function).

c) Certain parameter values are unacceptable. The ODE behaves badly when $\alpha$ or $\beta$ are negative. Modify the line search so that it also requires $\alpha, \beta$ to be positive (not just a decrease in $E$). It’s important that the ODE is never solved for the bad values of $\alpha, \beta$.

iii) Introduce a tolerance parameter in the usual way so that the algorithm stops when a certain ‘accuracy’ is reached, rather than a fixed number of steps.

e) Finally, write a main function that uses gradient descent on your function in (b) to estimate the infection rate (have this printed as the output). Pick a reasonable tolerance (so you can get a good fit).

Make a plot of the points plus the model solution to demonstrate it works.

*Hint:* Both parameters are between 0 and 1.