# Modeling COVID-19 Dynamics

In this project we will fit a longer term model for COVID-19 known as the SEIR model, you can read more about it here: `https://www.idmod.org/docs/hiv/model-seir.html`
The model is given by the following differential equation:

$$\begin{aligned}
\dot{S} &= -\beta SI/(S+E+I+R) \\
\dot{E} &= \beta SI/(S+E+I+R) - \sigma E \\
\dot{I} &= \sigma E - \gamma I \\
\dot{R} &= \gamma I
\end{aligned}$$

Notice that this is a 4-dimensional system of the form $\dot{x} = f(x)$ where $x = (S, E, I, R)^\top$. The four classes are Susceptible (S), Exposed (E), Infectious (I), and Recovered (R). The exposed class have the disease but are in the incubation period and cannot pass the disease on yet. Notice that the rate of increase in the exposed class is proportional to $S * \frac{I}{S+E+I+R}$ where $\frac{I}{S+E+I+R}$ is the percentage of the population that is infectious. Some reasonable estimates for the parameters are given below:

| Parameter | Estimate | Meaning |
|---|---|---|
| $\sigma$ | $1/3$ | One over the incubation period |
| $\gamma$ | $1/4$ | One over the duration of contagiousness |
| $\beta$ | $R_0 * \gamma$ | Infectiousness rate |
| $R_0$ | 3 | Reproduction number |

COVID-19 Parameter Information:
`https://www.ncbi.nlm.nih.gov/books/NBK554776/`
`https://www.uptodate.com/contents/coronavirus-disease-2019-covid-19`

**Problem 1:** Modify the attached `pend2.m` code to simulate the SEIR model with the above parameters and rename it `SEIR.m`. You will need to extend the "ydot" function to include 4 variables and be sure to update the plot legend. Run your code as `SEIR([0 500],[330e6 10 0 0],1000);`. The peak number of exposed (E) should be $\approx 46$ million, what is the peak number of infected?

Notice that the model evolves from near the Disease-Free Equilibrium (DFE), where $(S, E, I, R) = (S, 0, 0, 0)$ and everyone is susceptible, to the Endemic Equilibrium (EE) where the number of recovered (who are assumed to be immune) is large enough to prevent the infection from spreading. The equilibria are states that don't change, so all the derivatives are equal to zero, so they can be solved by solving the following system of equations:

$$\begin{aligned}
0 &= -\beta SI/(S+E+I+R) \\
0 &= \beta SI/(S+E+I+R) - \sigma E \\
0 &= \sigma E - \gamma I \\
0 &= \gamma I
\end{aligned}$$

which can be solved with Newton's method! (not part of this project, just connecting to what we've seen before).

**Problem 2:** Do your own research on the internet (you can use the sites above) and decide on a range for each of the parameters, $\sigma, \beta$, and $R_0$. List your ranges for each parameter and pick your own best guess for each of the parameters. Rerun the simulation using your parameters and compare the plots.

Playing around with parameters is fun but if we want to connect with reality we need data! Download `full_data.csv` and `loadfulldata.m` from the Blackboard Project. You can also get the latest data set here: `https://ourworldindata.org/coronavirus-source-data`

Run `loadfulldata.m` which will extract the cumulative number of `Cases` and `Deaths` in the United States over time. Unfortunately, these observations (data) do not directly correspond to anything in our model. We first need to add two new variables to our model, $C$ for confirmed cases, and $D$ for deaths.

$$\begin{aligned}
\dot{S} &= -\beta SI/(S+E+I+R) \\
\dot{E} &= \beta SI/(S+E+I+R) - \sigma E \\
\dot{I} &= \sigma E - \gamma I - \gamma\tau I - \gamma\mu I \\
\dot{R} &= \gamma I \\
\dot{C} &= \gamma\tau I \\
\dot{D} &= \gamma\mu I
\end{aligned}$$

where $\tau$ is the percentage of infectious people who are tested and $\mu$ is the percentage of infectious people who die from illness. We will call this new model the SEIRCD model.

**Problem 3:** Create a new version of your code called `SEIRCD.m` and modify it to run the SEIRCD model. Also, modify your code to take in a vector of parameters $p = [\mu, \tau, E_0]$ (where $E_0$ is the initial number of exposed). Set `p0 = [0.015, 0.1, 10];` and run your code as `SEIRCD([0 500],[330e6 0 0 0 0],1000,p0);` then add the data for `Cases` and `Deaths` to your plot. Zoom it to see if the data matches the model.

Now we are ready to fit the model to the data. The goal is to adjust $\mu$, $\tau$, and $E_0$, to make the model fit the data. To do this we first define a function to compute the error.

**Problem 4:** Define a function `error = SEIRCDerror(Cases,Deaths,p);` that takes in the data for `Cases` and `Deaths` and the parameters `p`. Your function should simulate the SEIRCD model with the p parameters. Let `T=length(Cases);` and simulate the SEIRCD model for time $[0, T-1]$ using $n = 10 * T$ steps. Then extract the model cases `C = x(:,5);` and model deaths `D = x(:,6);` and compare `C(1:10:end)` to `Cases` and `D(1:10:end)` to `Deaths` by computing

```
error = mean((C(1:10:end-1)-Cases).^2+5*(D(1:10:end-1)-Deaths).^2);
```

Finally, we want to choose $p$ to minimize the error. We could try to use Newton's method for this but instead we will use a built in optimization tool in Matlab called `fminsearch`.

**Problem 5:** Define an inline function `f = @(p) SEIRCDerror(Cases,Deaths,p);` so that `f` only takes in `p` and returns the error. Then optimize the parameters by calling `p0 = fminsearch(f,p0);` you may need to call this a few times until the `p0` stabilizes. Use this new `p0` to rerun the SEIRCD model and plot the new simulation along with the data as in Problem 3. How does the fit compare the Problem 3? What is the peak number of infectious patients? How many deaths does your model predict? Compare to these predictions: `https://covid19.healthdata.org/projections`

This is a fairly simple model and it can be very sensitive to the parameters so we should take the results with a grain of salt. There is a lot of debate right now about how to improve the model for COVID-19.

**Problem 6:** Based on what you know about COVID-19, or based on some internet research, think of a possible way to improve the model, either by adding new variables or modifying the equations. Explain your modification and then fit parameters in your new model and compare with the SEIRCD model results.

**Problem 7:** Repeat your modeling on another country of your choosing.