SIMC report 2022 Korea Science Academy of KAIST
Jeewon Kim, Minwoo Kim, Jihun Park

Section 1.1: A simple ODE Model

Question 1.1.a.
We will use the following lemma to find bounds of S and I.

Lemma 0.1. *f is a differentiable function with $f(0) > 0$, and there exists a continuous function $g : [0, \infty) \to \mathbb{R}$ such that $\frac{d}{dt}f(t) = g(t)f(t)$. Then $f(t) > 0$ for all $t > 0$.*

Proof. Multiplying an integration factor, it can be shown that

$$\frac{d}{dt}(f(t)e^{-\int g(t)dt}) = (\frac{d}{dt}f(t) - f(t)g(t))e^{-\int g(t)dt} = 0 \tag{1}$$

Therefore, $f(t)e^{-\int g(t)dt}$ should equal to some constant C. Since $e^{-\int g(t)dt} \neq 0$, it is possible to divide both hand sides by that value and the result is $f(t) = Ce^{\int g(t)dt}$. $f(0) > 0$ makes it certain that $C > 0$ and thus $f > 0$. $\square$

Now, if we take $f(t) = S(t)$, $g(t) = -\beta I(t)$, we can use the lemma because $S(0) > 0$. Therefore $\underline{S(t) > 0 \text{ for all } t > 0}$.
Similarly, letting $f(t) = I(t)$, $g(t) = \beta S(t) - \gamma$ and using $I(0) > 0$ leads us to the conclusion that $\underline{I(t) > 0 \text{ for all } t > 0}$.
Note that $R(0) = 0$, R is differentiable on $[0, \infty)$, and $R'(t) = \gamma I(t) > 0$ for all t. By the Mean Value Theorem (MVT), for all $t > 0$, there is $0 < t' < t$ with $\frac{R(t)-R(0)}{t-0} = \frac{R(t)}{t} = R'(t)$. Since $R'(t) > 0$ and $t > 0$, $\underline{R(t) > 0}$.
For the sum of the three functions, we can add up the three given equations to get $\frac{d}{dt}(S(t) + I(t) + R(t)) = 0$. Therefore $\underline{S(t) + I(t) + R(t) = f(t) \text{ is constant}}$.
We can deduce a corollary that will be handy in the near future, especially in solving Question 1.1.b. and 1.1.c.

Corollary 0.2. *S is strictly decreasing. (Because $\frac{d}{dt}S(t) = -\beta I(t)S(t) < 0$ for all t.)*

Question 1.1.b. Answer: $\underline{\mathbf{S_0} > \frac{\gamma}{\beta}}$

- Case 1: $S_0 > \frac{\gamma}{\beta}$:
  Since $\frac{d}{dt}I(t) = (\beta S(t) - \gamma)I(t) > 0$ and $(\beta S(t) - \gamma)I(t)$ is continuous, there exists an interval $A = [0, \varepsilon)$ where $t \in A \Rightarrow \frac{d}{dt}I(t) > 0$. So $I(t)$ increases in interval A, which means that for all $t \in (0, \varepsilon)$, $I(t) > I_0$. So an epidemic occurs.
- Case 2: $S_0 \leq \frac{\gamma}{\beta}$:
  Because S is a decreasing function, $\beta S(t) - \gamma \leq 0$ for all $t \geq 0$ and thus $I'(t) \leq 0$ in the same domain. Therefore I is decreasing and there is no epidemic.

Question 1.1.c. Answer: $\underline{I_{max} = N - \frac{\gamma}{\beta} + \frac{\gamma}{\beta}\ln(\frac{\gamma}{\beta S})}$
Firstly, $\frac{dI}{dt} = 0$ if and only if $S = \frac{\gamma}{\beta}$, so I has its maximum value at t where $S = \frac{\gamma}{\beta}$.
Because $I > 0$, S is a strictly decreasing function. This lets one to write I as a function of S. Using the chain rule,

$$\frac{dI}{dS} = \frac{dI/dt}{dS/dt} = \frac{(\beta S - \gamma)I}{-\beta IS} = -1 + \frac{\gamma}{\beta S} \tag{2}$$

Thus $I(S) = \int_{S_0}^{S} (-1 + \frac{\gamma}{\beta S}) dS = S_0 - S + \frac{\gamma}{\beta} \ln S$. Plugging $S = \frac{\gamma}{\beta}$ in gives the result.

Question 1.1.d. Answer: **Disease always dies out**

It is obvious that β, $\gamma > 0$

Suppose that $\lim_{t\to\infty} I(t) \neq 0$. Then for all ε, there exists $t_0 > 0$ satisfying $[I(t) > \varepsilon$ for all $t > t_0]$. Let's say $I(t) > \varepsilon_0$.
(1) Suppose that $S(t) > \frac{\gamma}{\beta}$ for all $t > t_0$. Since $I(t) > \varepsilon_0$ for all $t > t_0$, $\frac{d}{dt}S(t) = -\beta I(t)S(t) < -\beta \cdot \varepsilon_0 \cdot \frac{\gamma}{\beta} = -\varepsilon_0\gamma$. By Mean Value Theorem, $S(t) < S(t_0) - \varepsilon_0\gamma(t - t_0)$. That is, $S(\frac{S(t_0)}{\varepsilon_0\gamma} + t_0) < 0$. Contradiction.
Therefore $S(t') \leq \frac{\gamma}{\beta}$ for some $t' > t_0$. Also, since $S(t'+1) < S(t') \leq \frac{\gamma}{\beta}$, there exist $t (= t'+1)$ s.t. $S(t) < \frac{\gamma}{\beta}$.
(2) Let t^* be an arbitrary real number with $S(t^*) < \frac{\gamma}{\beta}$. Since S is a strictly decreasing function, for all $t \geq t^*$, $\beta S(t) - \gamma < \beta S(t^*) - \gamma < 0$. Let $k = \beta S(t^*) - \gamma < 0$.
Then $\frac{d}{dt}I(t) = (\beta S(t) - \gamma)I(t) < kI(t)$. Consider the function $f(t) = e^{kt} \cdot \frac{I(t^*)}{e^{kt^*}}$. Then $I(t^*) = f(t^*)$, $\frac{d}{dt}f(t) = kf(t)$, so $I(t) \leq f(t)$ for all $t > t^*$, which means $\lim_{t\to\infty} I(t) \leq \lim_{t\to\infty} f(t) = 0$. So $\lim_{t\to\infty} I(t) = 0$, which is a contradiction.

Question 1.1.e. The figure below denotes the graph of S(blue), I(orange), R(green), N(red), pred_Imax(violet) vs time, where $\beta = 0.01, \gamma = 0.03, I_0 = 1.0$.
pred_Imax is calculated with the equation given at Question 1.1.c.
The x-axis is t, which spans between 0 and 150, and y-axis is the actual value for $S(t), I(t), \ldots$
For the four graphs, We have set the value of S_0 as 0.2, 1, 5, 25, respectively.
From the figure below, we can find that

(a) S, I, R are always positive and S+I+R=N is constant.
(b) Epidemic occurs when $S_0 > 3 = \frac{\gamma}{\beta}$.
(c) The value of predicted $I_{\max}$ accords to the actual maximum values of I for all epidemic cases.
(d) The value of $I(t)$ always approaches 0.

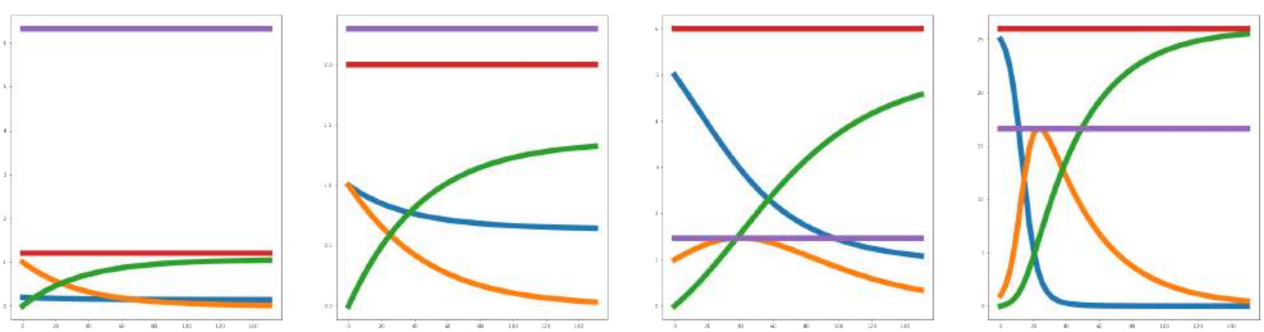

Figure 1. SimpleODE

Section 1.2: The Spatial Model

Question 1.2.a. The figure below denotes the graph of S(blue), I(orange). The x-axis is i and y-axis is the value of $S^i(t), I^i(t)$. Other parameters are set as written in the problem.
The following 6 graphs are graphs at $t = 0, 2, 4, 6, 8, 10$.

- There is a pandemic occuring. The waves of S, I travel in the positive x direction together.
- It seems that S has an inflection point where I has the local maximum.
- The maximum value of I is in the range of $35 \sim 40$.
- There are people who survive the pandemic, that is, S_{-} is in the range of $1 \sim 5$.
- The graph seems to move at about the speed of $10 units/sec$.

Note that this is similar with the case where x is continuous, so We used this program for different initial values, and tried to figure out the relations between the parameters and the properties of the graph.

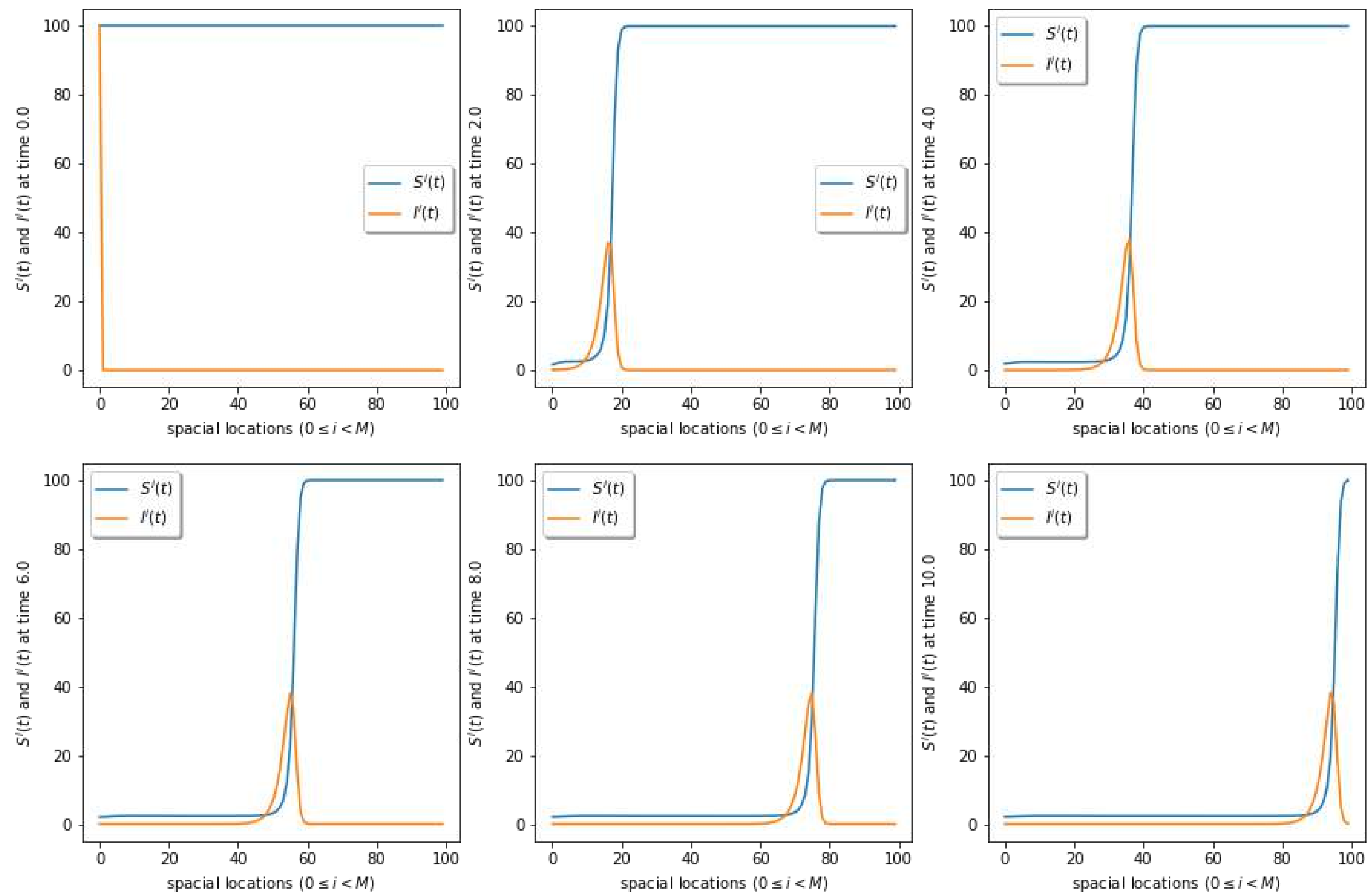

Figure 2. Result of the notebook code

Question 1.2.b. Answer: $d = \delta$

We can use the following approximation to check the two terms are similar.

$$S^{i-1}(t) - 2S^i(t) + S^{i+1}(t) \approx \frac{S(t,x-1) - 2S(t,x) + S(t,x+1)}{1^2} \approx \frac{S(t,x+\triangle x) - 2s(t,x) + S(t,x-\triangle x)}{(\triangle x)^2}$$

$$\approx \frac{\frac{S(t,x+\triangle x)-s(t,x)}{\triangle x} - \frac{S(t,x)-S(t,x-\triangle x)}{\triangle x}}{\triangle x} \approx \frac{\partial}{\partial x} S(t,x)$$

Also, let's note that $v = d/2$...? Are you sure?

Definition 0.3. Whenever there is occurs a pandemic, $S(t,x)$ and $I(t,x)$ satisfies the transport equation,
so we denote $S(t,x) = f(x-ct)$ and $I(t,x) = g(x-ct)$ where c is the velocity of the pandemic wave. Let $u = x - ct$.

Question 1.2.c. Answer: $S_0 < \frac{\gamma}{\beta}$
$S(-\infty, x) = S_0$ for all $x \in \mathbb{R}$

- $S_0 < \frac{\gamma}{\beta}$
  Assume that a pandemic occurs. Use the Definition 2.1. Then $\int_{-\infty}^{\infty} g(x-ct)dx$ is a constant for all t. Then $0 = \frac{\partial}{\partial t}\int_{-\infty}^{\infty} g(x-ct)dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial t} g(x-ct)dx = \int_{-\infty}^{\infty} \beta f(x-ct)g(x-ct) - \gamma g(x-ct) + \delta g''(x-ct)dx = \int_{-\infty}^{\infty}[\beta f(x-ct)-\gamma]g(x-ct)dx + [g'(x-ct)]_{-\infty}^{\infty} = \int_{-\infty}^{\infty}[\beta f(x-ct)-\gamma]g(x-ct)dx < \int_{-\infty}^{\infty}[\beta S_0 - \gamma]g(x-ct)dx < 0$, which is a contradiction.

Question 1.2.d. The figures denote 2 graphs over time, S, I resepectively.

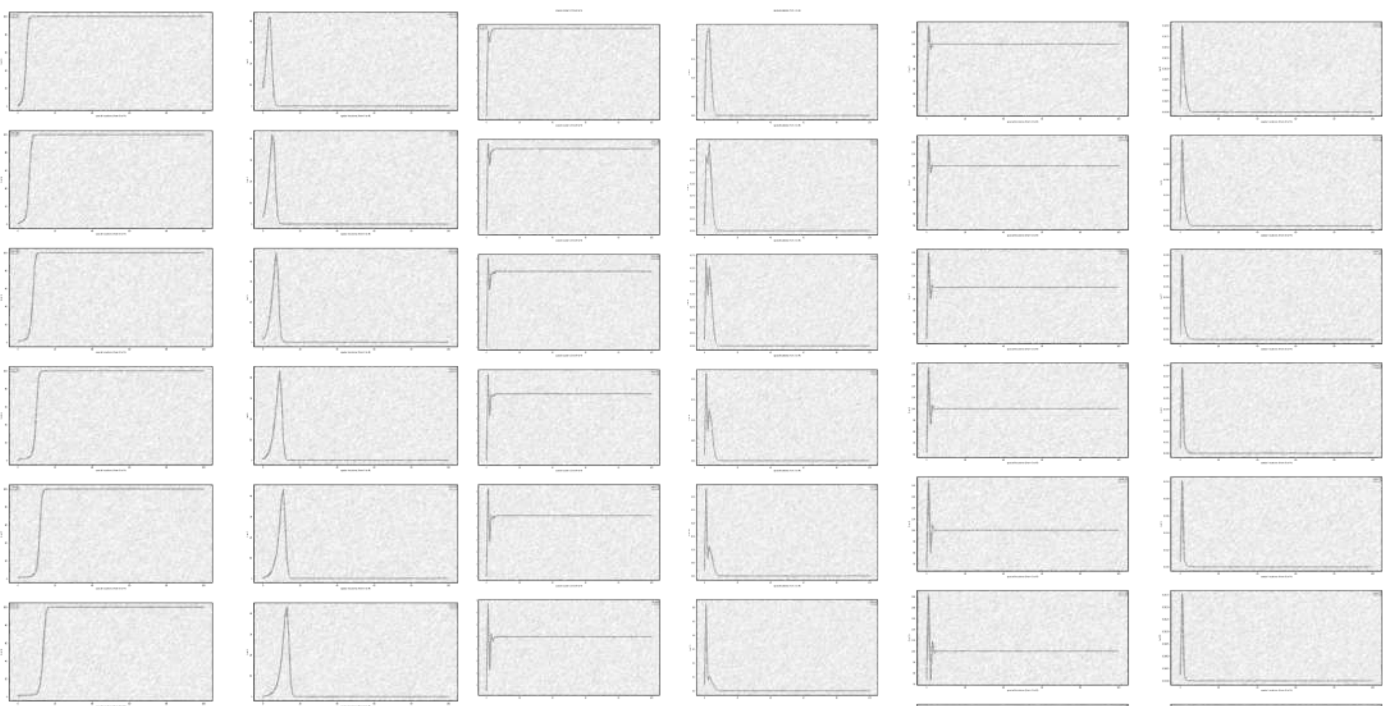

Figure 3. case 1: $S_0 > \gamma/\beta$; case 2: $S_0 = \gamma/\beta$; case 3: $S_0 < \gamma/\beta$

(1) If $S_0 > \gamma/\beta$, the graph looks like it is a pandemic.
(2) If $S_0 = \gamma/\beta$, the graph slowly becomes a sharp wave, but we think that this is a error due to the python package we used to program: scipy.integrate.solveivp
(3) If $S_0 < \gamma/\beta$, the graph is clearly not a pandemic. Both graphs quickly become very sharpy oscillating waves.

Question 1.2.e.

$$\frac{\partial}{\partial t}S(t,x) = -\beta I(t,x)S(t,x) - v\frac{\partial}{\partial x}S(t,x) \tag{3}$$

$$\frac{\partial}{\partial t} I(t,x) = \beta I(t,x)S(t,x) - \gamma I(t,x) + \delta \frac{\partial^2}{\partial x^2} I(t,x) \tag{4}$$

For any t and x, let's only consider the population of S that moves. Denote that function as S^*. And let v equal to the speed that people of S move. Then $v = \frac{\triangle x}{\triangle t}$.

$$\frac{\partial}{\partial t} S^*(t,x) = \lim_{\triangle t \to 0} \frac{S(t+\triangle t, x) - S(t,x)}{\triangle t} = \lim_{\triangle t \to 0} \frac{S(t, x-\triangle x) - S(t,x)}{\triangle t}$$
$$= \lim_{\triangle x \to 0} \frac{S(t, x-\triangle x) - S(t,x)}{\triangle x} \cdot v = -v \frac{\partial}{\partial x} S(t,x)$$

So instead of $\delta \frac{\partial^2}{\partial x^2} S(t,x)$, we use $-v \frac{\partial}{\partial x} S(t,x)$

Question 1.2.f. Answer: $(v \geq c)$

Assume that there is a pandemic. Use the Definition 2.1. Put it in equation (2), then we have $(c - v)f'(u) = \beta f(u)g(u)$. Since the RHS is positive, $v \geq c$ implies that there is no solution, so contradiction.
If $v < c$, there exists a solution of $(c - v)f'(u) = \beta f(u)g(u)$, so such f, g will be S, I, and there will be a pandemic.

Question 1.2.g. Answer: **If** $S_0 \geq \frac{\gamma}{\beta}$, $\rho = -xW(-\frac{e^{-1/x}}{x})$ **where** $x = \frac{\gamma}{\beta S_0}$**. Else:** $\rho = 1$

Since there is a pandemic, use the Definition 2.1. Then the differential equations (2), (3) becomes

$$(c - v)f'(u) = \beta f(u)g(u) \tag{5}$$
$$-cg'(u) = \beta f(u)g(u) - \gamma g(u) + \delta g''(u) \tag{6}$$

where $u = x - ct$

- Solve (4):

$$\frac{\beta}{c-v} g(u) = \frac{f'(u)}{f(u)}$$
$$\frac{\beta}{c-v} \int_{-\infty}^{u} g(z)dz = \ln f(u) + C$$
$$u \to -\infty: \; 0 = \ln S_- + C, \; C = -\ln S_-$$
$$u \to \infty: \; I_0 = \int_{-\infty}^{\infty} g(z)dz, \;\text{ then }\; \frac{\beta}{c-v} I_0 = \ln S_0 - \ln S_-$$

- Solve (4)-(5):

$$(c - v)f'(u) + cg'(u) = \gamma g(u) - \delta g''(u)$$
$$(c - v)f(u) + cg(u) = \gamma \int_{-\infty}^{u} g(z)dz - \delta g'(u) + C$$
$$u \to -\infty: \; (c - v)S_- = 0 + 0 + C$$
$$u \to \infty: \; (c - v)S_0 = \gamma I_0 + C = \gamma I_0 + (c - v)S_-$$
$$(c - v) = \frac{\gamma I_0}{S_0 - S_-}$$

- Combine the two results and we get $\frac{\beta}{\gamma}(S_0 - S_-) = \ln S_0 - \ln S_-$. Then we have $\frac{\gamma}{\beta} = \frac{S_0(1-\rho)}{\ln(1/\rho)}$.
  Using Lambert W function, we can show that $\rho = -xW(-\frac{e^{-1/x}}{x})$ where $x = \frac{\gamma}{\beta S_0}$.
- If $S_0 < \frac{\gamma}{\beta}$, by Mean Value Theorem, there exists some $S^* \in (S_-, S_0)$ such that $\frac{\beta}{\gamma} = \frac{1}{S^*} > \frac{1}{S_0} > \frac{\beta}{\gamma}$, so it is impossible.
- In the other case, there is no solution, and notice that we divided by $S_0 - S_-$, so we can know that $\rho = S_-/S_0 = 1$ because S_- has to converge to some value. In this case, pandemic wouldn't occur.

Corollary 0.4. *In our new model for modified disease spread, if pandemic occurs, then* $S_0 > \frac{\gamma}{\beta}$.

Section 1.3: Prediction of Disease Spreading Dynamics from Data

Observations.
Observation 1) The value of $S(t,x)+I(t,x)+R(t,x)$ is almost constant. Therefore we can assume that all people move all direction with same probability. Therefore, we decided to refer to the *2-dimensional diffusion model to predict the movement of people.*

Observation 2) For each year, I starts changing at the specific day of the year. Also, note that the disease starts spreading from one spot. *We assume that the infection starts on January 21st from a unique patient zero.*

Observation 3) Unlike the preexisting SIR model above, the number of the recovered(/removed) individual decreases from June, and it eventually reaches zero. We can say that the recovered people lose their immune to the disease. *We assume that small proportion of the recovered individuals lose immunity and become susceptible.*

Trial 1.

Model 1. From the observations above, we get the following model.

$$\begin{cases} \frac{\partial}{\partial t}S(t,x,y) = \alpha R(t,x,y) - \beta I(t,x,y)S(t,x,y) + \delta(\frac{\partial^2}{\partial x^2}S(t,x,y) + \frac{\partial^2}{\partial y^2}S(t,x,y)) \\ \frac{\partial}{\partial t}I(t,x,y) = \beta I(t,x,y)S(t,x,y) - \gamma I(t,x,y) + \delta(\frac{\partial^2}{\partial x^2}I(t,x,y) + \frac{\partial^2}{\partial y^2}I(t,x,y)) \\ \frac{\partial}{\partial t}R(t,x,y) = \gamma I(t,x,y) - \alpha R(t,x,y) + \delta(\frac{\partial^2}{\partial x^2}R(t,x,y) + \frac{\partial^2}{\partial y^2}R(t,x,y)) \end{cases}$$

(Initial value: $I(t_0 x_0, y_0) > 0$ for the first time.)
where $S(t,x,y), I(t,x,y), R(t,x,y)$ denote the number of susceptible / infected recovered individuals at time t and position (x,y). (Time 0: January 1st, Time 1: January 2nd, ..., Time 364: December 31st)
Each of the variable represents follows.

- α is a coefficient denoting how much of recovered individuals switch to susceptible ones.
- β, analogous to the previous examples, is a coefficient of infection.
- γ is used as recovery coefficient also in this context.
- δ denotes rate of diffusion among the three groups. It will be almost same among three groups S, I, R, as seen in Observation 1.

Especially, the terms in the form of $\frac{\partial^2}{\partial x^2}f(t,x,y)+\frac{\partial^2}{\partial y^2}f(t,x,y)$ is taken from the 2-dimensional diffusion model, just as how transport model and 1-dimensional diffusion model was used in previous cases.

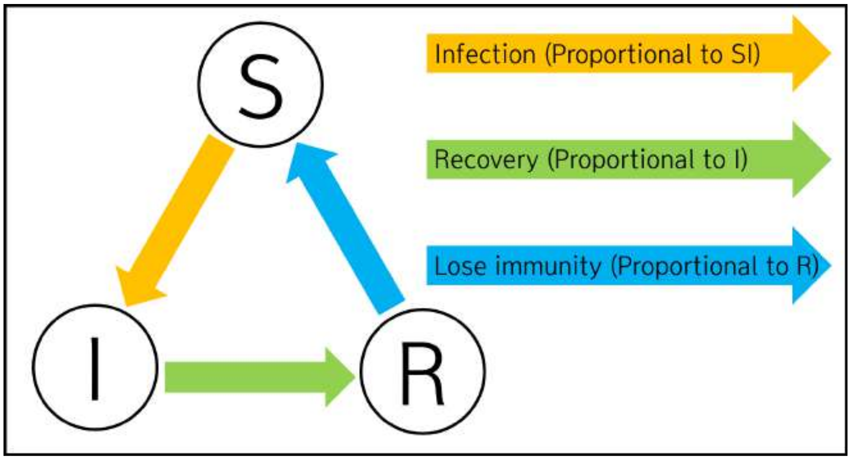

Figure 4. lose immunity

Prediction Results. Given the coefficients, we change them manually by scoring on how well they behave. Our first goal was to set the parameters so that I, R descend to 0 naturally at winter. The best set of parameters were as following.

$$\alpha = 10, \beta = 0.05, \gamma = 10, \delta = 0.002$$

β had to be small because β has huge impact since both I, S has effect on it, But regardless to the values of the parameters, they always tended to converge to some positive constant. So we made a change to the model, and we had our second try.

Trial2.

Model 2. We noticed that S explodes rapidly in summer in the raw data. So we make a hypothesis that this disease is a seasonal disease, which only has high activity in summer. Then β will have to change throughout time because it will be more infectious at summer. So instead of using the constant β, we use $\beta sin(\pi t/365)$. Then in summer where t is about 365/2, the infection rate would be the highest.

$$\begin{cases} \frac{\partial}{\partial t}S(t,x,y) = \alpha R(t,x,y) - \beta sin(\pi t/365) I(t,x,y)S(t,x,y) + \delta(\frac{\partial^2}{\partial x^2}S(t,x,y) + \frac{\partial^2}{\partial y^2}S(t,x,y)) \\ \frac{\partial}{\partial t}I(t,x,y) = \beta sin(\pi t/365) I(t,x,y)S(t,x,y) - \gamma I(t,x,y) + \delta(\frac{\partial^2}{\partial x^2}I(t,x,y) + \frac{\partial^2}{\partial y^2}I(t,x,y)) \\ \frac{\partial}{\partial t}R(t,x,y) = \gamma I(t,x,y) - \alpha R(t,x,y) + \delta(\frac{\partial^2}{\partial x^2}R(t,x,y) + \frac{\partial^2}{\partial y^2}R(t,x,y)) \end{cases}$$

(Initial value: $I(t_0 x_0, y_0) > 0$ for the first time.)

Prediction Results. Our goal again was to set the parameters so that I, R descend to 0 naturally as time goes. The best set of parameters were as following.

$$\alpha = 2, \beta = 0.04, \gamma = 2, \delta = 0.001$$

We found out that this model fits pretty well to the data, but we observed one more thing. In our second model, the population of I grew too slow and lasted too long, the given data's I was very big at the beginning, and it shouldn't last too long.

Trial3.

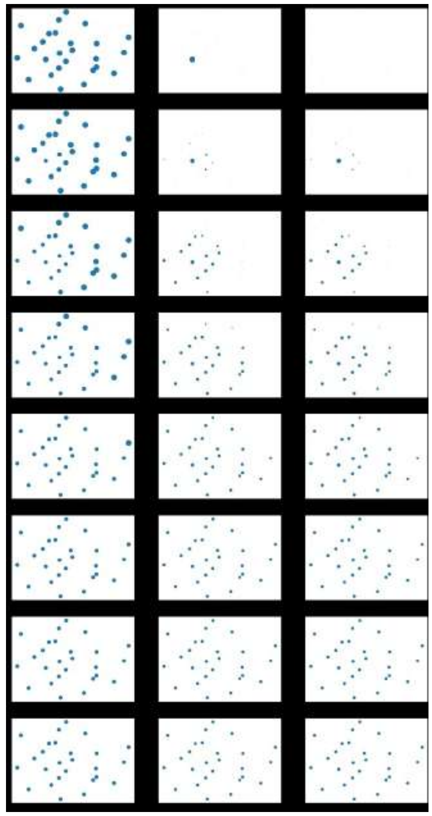

Figure 5. Trial 1: (S, I, R, respectively)

Model 3. So instead of using the sine function for β, we use $\beta(1-t/365)$. Then I would grow very fast at the beginning, and decrease rather quickly than the sine function.

$$\begin{cases} \frac{\partial}{\partial t}S(t,x,y) = \alpha R(t,x,y) - \beta(1-t/365)I(t,x,y)S(t,x,y) + \delta(\frac{\partial^2}{\partial x^2}S(t,x,y) + \frac{\partial^2}{\partial y^2}S(t,x,y)) \\ \frac{\partial}{\partial t}I(t,x,y) = \beta(1-t/365)I(t,x,y)S(t,x,y) - \gamma I(t,x,y) + \delta(\frac{\partial^2}{\partial x^2}I(t,x,y) + \frac{\partial^2}{\partial y^2}I(t,x,y)) \\ \frac{\partial}{\partial t}R(t,x,y) = \gamma I(t,x,y) - \alpha R(t,x,y) + \delta(\frac{\partial^2}{\partial x^2}R(t,x,y) + \frac{\partial^2}{\partial y^2}R(t,x,y)) \end{cases}$$

(Initial value: $I(t_0 x_0, y_0) > 0$ for the first time.)

Prediction Results. Our goal was to set the parameters so that population of I behaved as in the given datas.
The best set of parameters were as following.

$$\alpha = 10, \beta = 0.02, \gamma = 4, \delta = 0.0001$$

We found out that this model fits very well to the data. It fit well when we used $(1 - t/365)$, so we can think of situations where people will wear masks more or get vaccinated more as time goes by, so the infection rate decreases over time. In the following figure, we can see that I rapidly increases at the beginning.

In conclusion, our model has 2 main differences with the models before. First, we considered about the rate people losing immunity (α). Second, we considered β decreasing as time goes, especially with function $(1 - t/365)$. We found out that this model showed very good approximations of the data given in 2021, 2022.

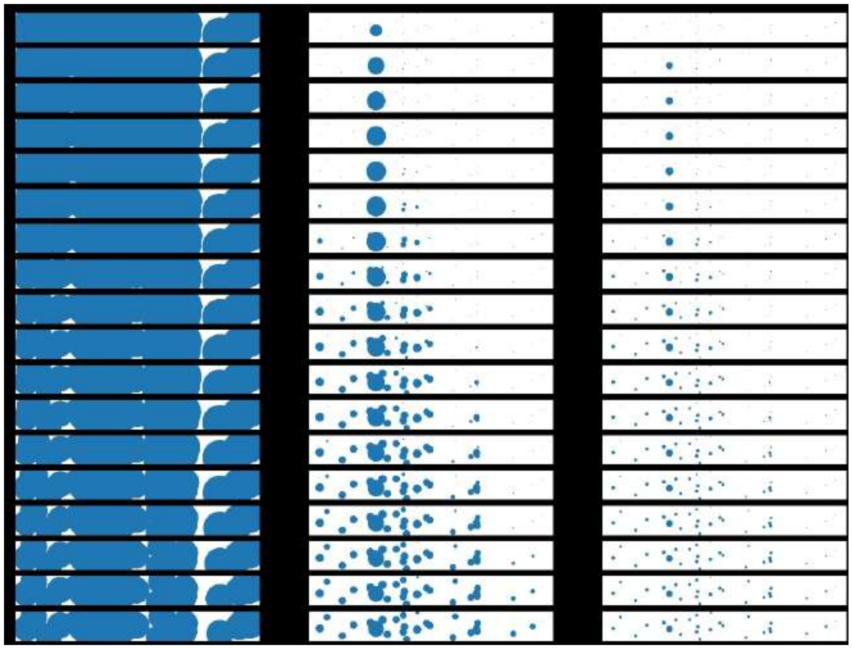

Figure 6. Trial 3: (S, I, R, respectively

Further Research.

Ways to Change the Model.

- In reality, **fatality rate** would be important. However, in this data, the total population was constant so this was not considered in this model.
- **Relapse rate**: this is actually very similar to rate of losing immunity.
- **Urban Cities**: Certain locations have more people coming in and coming out, increasing the δ.
- **SEIR model**: We can add the population of exposed but not infected people.
- Instead of using just δ for all S, I, R, use δ_1, δ_2, δ_3 as it might differ among them. Also when $\delta = \delta_1$, δ_2, δ_3, let $N(t, x, y) = S + I + R$, then $\frac{\partial}{\partial t} N(t, x, y) = \delta \nabla^2 N$. That is a heat equation, so N would become smooth after a long time, and have an equal distribution, but in the data, they don't. So using δ_1, δ_2, δ_3 would be more reasonable.

Different way to calculate. It is important to note that this score function is purely a function of the coefficients. Therefore, calculating the gradient of the score function gives a vector in which we can modify coefficients in that way to reduce the score. This method is known as gradient descent.
Using the method of **Gradient Descent**, it seems that we can easily find all the parameters in the equation.
Then the score function of a model should be defined using MSE(mean squared error) between the predicted number of susceptible/infected/recovered individuals and the actual value.
Although we partially wrote the code for Gradient Descent (refer to the Jupyter file), we could not run the method because the estimated run-time is too long. (It takes about 1 minute to execute the score function once.) Therefore for now, we didn't use the Gradient Descent to get all the parameters.