STATE UNIVERSITY OF NEW YORK AT GENESEO

Application of the SEIRD Epidemic Model and Optimal Control to Study the Effect of Quarantine and Isolation on the Spread of COVID-19

by

Luz Melo

A thesis submitted in partial fulfillment of the
Honors in Mathematics distinction

advised by
Dr. Sedar Ngoma
Mathematics Department

May 5, 2021

STATE UNIVERSITY OF NEW YORK AT GENESEO

Abstract

Mathematical models serve as a powerful tool for visualizing and describing the dynamics of infectious diseases. In this article, the $SEIRD$ (Susceptible-Exposed-Infected-Recovered-Deceased) epidemic model consisting of a system of five non-linear differential equations is considered. By using COVID-19 data pertinent to the United States and the world situation, the use of existence and uniqueness of the disease-free equilibrium is established, the value of the basic reproductive number is determined, a stability analysis is carried out, and the parameters that fit the model to the data are estimated. An optimal control approach is performed to study the effect of quarantine and isolation on the spread of COVID-19. The goal is to explore the effectiveness of quarantine and isolation at controlling the spread of COVID-19.

Acknowledgements

First and foremost I would like to express my immense gratitude towards my research supervisor and academic advisor, Dr. Sedar Ngoma. It has been a great privilege and honor have the opportunity to work on such a captivating research while receiving guidance from a wonderful Mathematician and person. I would like to also give special thanks to professors Dr. Aaron Heap, Dr. Ahmad Almomani for believing in me. The completion of this thesis would not have been possible without the help of the great professors in charge of the Mathematics Honors Committee. This work was partially supported by funds made available under a SUNY Expanded Investment and Performance Award to SUNY Geneseo.

Contents

List of Figures

List of Tables

1 Introduction

The Coronavirus (COVID-19) is a novel disease that emerged in Wuhan, China, in December of 2019. COVID-19 is primarily transmitted from person-to-person through exposure to respiratory droplets carrying the infectious virus [3]. The symptoms can range from fever, cough, loss of taste or smell, fatigue, and muscle ache to trouble breathing, chest pain, and confusion [3]. While most individuals infected with the virus experience mild to moderate symptoms similar to the common cold, older age groups and individuals with underlying medical conditions are at risk of developing more serious, life-threatening complications. Thus, as the highly contagious virus rapidly spread in a population with absolutely no immunity, containing the virus became a global priority. On March 11 of 2020, the World Health Organization (WHO) declared COVID-19 a pandemic and, two days later, President Donald Trump declared COVID-19 a National Emergency for the United States [4]. As of April of 2021, there has been over 140 million confirmed cases and 3 million deaths across nearly 200 countries. In this study, we hope to contribute to the ongoing research on this topic and, more specifically, provide insights into the importance of quarantine and isolation as non-pharmaceutical intervention strategies for mitigating the spread of COVID-19.

1.1 Properties of the Model

In this section, the $SEIRD$ compartmental mathematical model for prediction of COVID-19 epidemic dynamics is introduced. The model is a modified version of the basic Susceptible-Infected-Removed (SIR) model proposed by W.O. Kermack and A. G. MacKendrick. The $SEIRD$ system consists of five nonlinear differential equations, in this case represented by compartments, namely S, E, I, R, and D. In this study, the data gathered represents the spread of the disease prior major intervention strategies were applied. The variables used in our model represent the number of people in each compartment at a particular time t and are described as follows:

Table 1: Variables used and their meaning.

Variable	**Meaning**
$S = S(t)$	The susceptible population at time t.
$E = E(t)$	The exposed population at time t.
$I = I(t)$	The infectious population at time t.
$R = R(t)$	The recovered population at time t.
$D = D(t)$	The dead population at time t.

More specifically, the susceptible population are those individuals that are capable of contracting the disease once in contact with an infectious individual, the exposed population are infected individuals that remain latent for some time before becoming infectious, the infectious population are the individuals in the population that spread the disease, the recovered population are the individuals who have survived through the infectious period and become immune to the disease, and the dead population are the individuals that died due to the disease.

Table 2: Parameters used and their meaning.

Parameter	Meaning
β	Disease transmitting rate.
γ	Incubation rate.
μ	Recovery rate.
ρ	Death rate due to the disease.

The disease transmitting rate is the rate at which individuals contract the disease, the incubation rate is the rate at which the exposed individuals display symptoms of the disease, the recovery rate is the rate at which infectious individuals recover from the disease, and the death rate is the rate at which infected individuals die due to the disease. The total number of fixed people in the population at time t is given by $N = S(t) + E(t) + I(t) + R(t) + D(t)$ and S_0, E_0, I_0 R_0, and D_0 are the given initial conditions. In this study, the $SEIRD$ model is defined as follows:

1.2 The SEIRD Model without Optimal Control

$$\frac{dS}{dt} = -\beta SI, \qquad S(0) = S_0 > 0, \tag{1.1}$$

$$\frac{dE}{dt} = \beta SI - \gamma E, \qquad E(0) = E_0 > 0, \tag{1.2}$$

$$\frac{dI}{dt} = \gamma E - \rho I - \mu I, \quad I(0) = I_0 > 0, \tag{1.3}$$

$$\frac{dR}{dt} = \mu I, \qquad R(0) = R_0 > 0 \tag{1.4}$$

$$\frac{dD}{dt} = \rho I \qquad D(0) = D_0 > 0 \tag{1.5}$$

The interactions between our given variables and parameters are pivotal in order to understand the model. Studies have shown that the virus is mainly transmitted through human to human contact when infected respiratory droplets are inhaled. Thus, a major interaction occurs in the first two differential equations where the susceptible population and the infected population come in contact with each other. The term βSI indicates the portion of susceptible people who were in contact with an infectious individual. Hence, the individuals who are in direct contact with an infectious individual become exposed to the virus (this is particularly true in the early stages of the COVID-19 outbreak before hygiene protocols, face masks, and quarantine are implemented), and so this portion now moves from the susceptible group to the exposed group.

As a result, in the process of disease spread for the $SEIRD$ model, the susceptible individuals S will first move to the exposed population E instead of directly moving to the infected population I. Now, βSI is the amount of individuals in the exposed population E. However, out of this population a portion of individuals will catch the virus and will undergo an incubation period before becoming infectious. That is, γE individuals will move from the exposed population

E to the infected population I. Consequently, the total number of individuals in the infected compartment will fluctuate depending on 1) how many μI individuals recover from the disease and move to the recovered class, or 2) how many ρI individuals end up dying from the disease and move to the dead class.

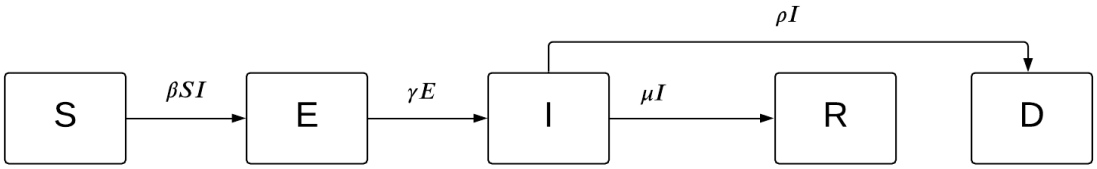


Figure 1: Flowchart diagram for the $SEIRD$ model. The five compartments represent the five distinct groups of individuals in a given population. The arrows represent the portion of individual that enters or leave a group.

1.3 Model Assumptions

Before we proceed to use the $SEIRD$ model to represent the dynamics of COVID-19 using real data, it is important to notice that this is an idealized model. As nonlinear system of equations are difficult to solve due to the high dependency of the system variables on each other, the $SEIRD$ model has been deliberately constructed as a simplified version of COVID-19 transmission dynamics. Thus, the model relies on a few but important assumptions:

1. The $SEIRD$ model is here used to model COVID-19 in a human population; hence, it is reasonable to assumed that all the variables and parameters are non-negative.
2. It is assumed that every individual in our population is susceptible to COVID-19.
3. The infected and susceptible populations are assumed to mix homogeneously. That is, each infectious individual (regardless of age, geographic location, etc.) has the same probability of coming in contact with any susceptible individual in the population.
4. The total population N remains constant over time as we ignore births, natural death, immigration, etc. Hence, no one joins the susceptible group and the only way an individual leaves the susceptible compartment is by becoming exposed to the virus. This results in the susceptible population decreasing monotonically towards zero over time as individuals become exposed to the virus.
5. All individuals in the population have the same probability to contract and die from the disease, regardless of their age or health status.
6. There is no treatment failure; a patient will either recover or die.
7. The model assumes that both Exposed-Recovered individuals and Infected-Recovered individuals become immune to the disease permanently. In other words, the individual gains immunity to the disease once recovered.

1.4 Existence and Uniqueness

In this section we find the equilibrium solution and briefly outline conditions under which system (5.1)–(5.5) has a solution. The result is stated in the following theorem:

Theorem 1.1 (**Existence and Uniqueness**). *Assume that the functions -$\beta S(t)I(t)$, $\beta S(t)I(t) - \gamma E(t)$, $\gamma E(t) - \mu I(t) - \rho I(t)$, $\mu I(t)$, $\rho I(t)$ appearing to the right sides of system* (5.1)–(5.5) *and their partial derivatives with respect to S, E, I, R, D are continuous in a rectangle $a < t < b$, $c < S, E, I, R, D < d$. Then for any $t_0 \in (a, b)$ and $S_0, E_0, I_0, R_0, D_0 \in (c, d)$, the initial value problem* (5.1)–(5.5) *has a unique solution valid on some open interval $a < \alpha < t < \nu < b$ containing t_0.*

Proof. Let $x(t) = (S(t), E(t), I(t), R(t), D(t))$, $t \in (a, b)$ and $S(t), E(t), I(t), R(t), D(t) \in (c, d)$. Then $x'(t) = (S', E', I', R', D')$. Define a function g as

$$g(x(t)) = (-\beta SI, \beta SI - \gamma E, \gamma E - \rho I - \mu I, \mu I, \rho I).$$

Let $t_0 \in (a, b)$ and $x(t_0) = (S(t_0), E(t_0), I(t_0), R(t_0), D(t_0)) = (S_0, E_0, I_0, R_0, D_0) = x_0 \in (c, d)$. By assumption we know that $g(x(t))$ and its partial derivative $g_x(x(t))$ are continuous in the rectangle $a < t < b$, $c < x < d$. By [10, Theorem 1.38, Page 74], the initial value problem

$$\begin{cases} x'(t) = g(x(t)) \\ x(t_0) = x_0 \end{cases}$$

has a unique solution valid on some open interval containing t_0. □

Now consider

$$-\beta SI = 0 \tag{1.6}$$
$$\beta SI - \gamma E = 0 \tag{1.7}$$
$$\gamma E - \rho I - \mu I = 0 \tag{1.8}$$
$$\mu I = 0 \tag{1.9}$$
$$\rho I = 0 \tag{1.10}$$

First, recall the vector $x'(t) = g(t, x(t))$ made from the right hand side of our system of differential equations (1.1)-(1.5). To find the equilibrium solution, we set $x'(t) = 0$. From (1.9)-(1.10), since μ and ρ are positive parameters, we must have that $I = 0$. Consequently, from (1.8) it follows that $E = 0$. Now, the term $\beta SI = 0$ since $I = 0$. This means that S could potentially be any other constant other than zero. Also, recall that, by assumption, our model does not considers births, natural death, immigration, etc. Thus, it follows that $x^* = (S_0, 0, 0, 0, 0)$ is the only equilibrium solution. This corresponds to the the disease-free equilibrium where the disease is not present in the population.

(For further explanation on the topic and a general proof of existence and uniqueness for the SI model refer to [7]).

2 Basic Reproductive Number

Definition 2.1. *The basic reproduction number, $\mathscr{R}_0$, is the average number of secondary infections produced by one infectious individual in an entirely susceptible population.*

In this section, the basic reproduction number is used as a threshold quantity to measure the transmission potential of COVID-19. To find the reproductive number, we use the approach of Next-Generation matrix given by Van den Driessche and Watmough. This method consists in a technique for the derivation of the next-generation matrix from ordinary differential equation compartmental models for disease transmission [11].

2.1 Next-Generation Matrix Approach

First, we begin by dividing the compartments into two sections: the infected compartments and the non infected compartments. The infected compartments consists of two type of individuals:

1. The individuals that have been exposed to the virus but are not yet infectious.
2. The infectious individuals that spread the virus.

The non-infected compartments are made of the remaining individuals who are not infected, that is, the susceptible population, recovered population, and dead population. Let $a = (E, I)$ be the vector of dependent variables in the infected compartments and $b = (S, R, D)$ be the vector of variables in the non infected compartments. Hence, $a \in \mathbb{R}^2$ and $b \in \mathbb{R}^3$. Now, we arrange the $SEIRD$ model such that the first 2 compartments of the system correspond to the infected compartments as follows:

$$\begin{aligned} a_i' &= f_i(a,b) && \text{for } i = 1,2. \\ b_j' &= g_j(a,b) && \text{for } j = 1,2,3. \end{aligned}$$

That is,

$$\begin{aligned} a_1' &= E' = \beta SI - \gamma E \\ a_2' &= I' = \gamma E - \mu I - \rho I \end{aligned}$$

and

$$\begin{aligned} b_1' &= S' = -\beta SI \\ b_2' &= R' = \mu I \\ b_3' &= D' = \rho I. \end{aligned}$$

In addition, we divide the right-hand side of the infected compartments such that

$$\begin{aligned} a_i' &= \mathscr{F}_i(a,b) - \mathscr{V}_i(a,b) && \text{for } i = 1,2. \\ b_j' &= g_j(a,b) && \text{for } j = 1,2,3. \end{aligned}$$

In the above equations, $\mathscr{F}_i(a,b)$ represents the rate of appearance of new infections in the i^{th} compartment and $\mathscr{V}_i(a,b)$ represents the remaining transfer of individuals into and out of the compartment i. In our case, we get that

$$\mathscr{F}_1(a,b) = \beta SI$$
$$\mathscr{F}_2(a,b) = \gamma E$$

and

$$\mathscr{V}_1(a,b) = \gamma E$$
$$\mathscr{V}_2(a,b) = \mu I + \rho I,$$

where this decomposition satisfies the properties outlined in [11, Page 105]. Now, recall that our $SEIRD$ model has a unique disease-free equilibrium, namely $x^* = (S_0, 0, 0, 0, 0)$. Thus, we determine the matrices F and V with components $F = \left[\frac{\partial \mathscr{F}_i(a,b)}{\partial x_j}\right]$ and $V = \left[\frac{\partial \mathscr{V}_i(a,b)}{\partial x_j}\right]$, for $i, j = 1, 2$, to be defined as

$$F = \begin{bmatrix} 0 & \beta S_0 \\ \gamma & 0 \end{bmatrix}$$

and

$$V = \begin{bmatrix} \gamma & 0 \\ 0 & \rho + \mu \end{bmatrix},$$

respectively. The next-generation matrix is defined to be $K = FV^{-1}$. Thus, basic matrix multiplication yields that

$$K = \begin{bmatrix} 0 & \frac{\beta S_0}{\rho+\mu} \\ 1 & 0 \end{bmatrix}.$$

The basic reproduction number, $\mathscr{R}_0$, is known to be the largest eigenvalue of K. Therefore, for the control-free SEIRD system it follows that the basic reproductive number of our model is given by

$$\mathscr{R}_0^2 = \frac{\beta S_0}{(\rho + \mu)}. \tag{2.1}$$

In order to see that formula (2.1) for $\mathscr{R}_0$ satisfies definition 2.1, notice that the number of new cases per unit of time produced by the infectious population is given by $\beta S_0 I$. When the first case of COVID-19 originated, there was only one infectious individual and so $I_0 = 1$. Thus, the number of secondary cases one infectious individual produces per unit of time will be βS_0. Since $\mu + \rho$ is the rate at which individuals leave the infected class, this means that the average time spent as an infected individual is $\frac{1}{\mu+\rho}$. Therefore, the number of secondary cases an individual will produce during the infectious period is indeed $\mathscr{R}_0^2 = \frac{\beta S_0}{(\rho+\mu)}$.

Next, the reproduction number $\mathscr{R}_0$ is calculated using data pertinent to the USA and world situation. The corresponding values of $\mathscr{R}_0$ are shown in Table 3. Since $\mathscr{R}_0 > 1$, we can estimate that COVID-19 will continue to spread for both the USA and the world.

Table 3: Basic reproductive number for the USA and world data.

Region	**Basic reproductive number**
USA	1.82
World	1.43

3 Local Stability of Disease-Free Equilibrium

In this section we determine the stability of the disease-free equilibrium. To determine the stability of the $SEIRD$ system, will perform linearization to obtain quantitative features of the solution in the close neighborhood of the disease-free equilibrium. Recall the Hartman-Grobman Theorem which says that the stability type of an equilibrium of a non-linear system is the same as that of the linearized system.

Theorem 3.1 (**Hartman–Grobman Theorem).** *The solutions of an $n \times n$ autonomous system of ordinary nonlinear differential equations in a neighborhood of a steady state look "qualitatively" just like the solutions of the linearized system near the origin [11].*

Theorem 3.2 (**Local Stability).** *An isolated equilibrium state x^* is locally stable if there in an open interval I containing x^* such that $\lim_{t\to+\infty} x(t) = x^*$ for any solution $x = x(t)$ with $x(0)$ in I [10].*

Definition 3.1 (**Jacobian Matrix).** *The Jacobian matrix is defined as the matrix of all first-order partial derivatives of a vector-valued function.*

The general form of the Jacobian matrix for the $SEIRD$ system is denoted as

$$J = \begin{bmatrix} -\beta I & 0 & -\beta S_0 & 0 & 0 \\ \beta I & -\gamma & \beta S_0 & 0 & 0 \\ 0 & \gamma & -\rho - \mu & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & \rho & 0 & 0 \end{bmatrix}.$$

Subsequently, the Jacobian matrix evaluated at the disease-free equilibrium $x^* = (S_0, 0, 0, 0, 0)$, in particular where $I = 0$, is given by

$$J(x^*) = \begin{bmatrix} 0 & 0 & -\beta S_0 & 0 & 0 \\ 0 & -\gamma & \beta S_0 & 0 & 0 \\ 0 & \gamma & -\rho - \mu & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & \rho & 0 & 0 \end{bmatrix}.$$

Theorem 3.3 (**Unstable equilibrium).** *The disease-free equilibrium, x^*, is unstable if and only if the basic reproductive number $\mathscr{R}_0 > 1$.*

Now, we wish to further expand on the stability of the $SEIRD$ model. To do this, we will prove that the disease-free equilibrium is locally unstable by looking at the eigenvalues of the Jacobian matrix and the basic reproductive number. Recall that the basic reproductive number of the $SEIRD$ system is computed to be $\mathscr{R}_0^2 = \frac{\beta S_0}{(\mu+\rho)}$. Since the basic reproductive number yield to be $\mathscr{R}_0 > 1$ for both the USA and world data, by Theorem 3.3 it follows that the $SEIRD$ system is locally unstable at the disease-free equilibrium. The following proof of Theorem 3.3 is meant to further validate our conclusion.

Proof. The disease-free equilibrium, x^*, is unstable if and only if $\lambda_1 < 0$ and $\lambda_2 > 0$. Assume that the disease-free equilibrium, x^*, is unstable. The characteristic equation of the Jacobian $|J - \lambda I| = 0$ is a quadratic polynomial in λ given by

$$A\lambda^2 + B\lambda + C = 0$$

where

$$\begin{aligned} A &= 1 \\ B &= (\gamma + \rho + \mu) \\ C &= \gamma(\rho + \mu - \beta S_0). \end{aligned}$$

Thus, using the quadratic equation, we obtain that

$$\lambda_1 = \frac{-(\gamma + \rho + \mu) - \sqrt{(\gamma + \rho + \mu)^2 - 4\gamma(\rho + \mu - \beta S_0)}}{2}. \tag{3.1}$$

$$\lambda_2 = \frac{-(\gamma + \rho + \mu) + \sqrt{(\gamma + \rho + \mu)^2 - 4\gamma(\rho + \mu - \beta S_0)}}{2}. \tag{3.2}$$

Notice that $\lambda_1 < 0$ and $\lambda_2 > 0$ can only happen when $\rho + \mu - \beta S_0 < 0$ so that the discriminant $(\gamma + \rho + \mu)^2 - 4\gamma(\rho + \mu - \beta S_0) > 0$. If $\rho + \mu - \beta S_0 < 0$, then $\rho + \mu < \beta S_0$, $\frac{\rho+\mu}{\rho+\mu} < \frac{\beta S_0}{\rho+\mu}$, $\mathscr{R}_0^2 > 1$, and so $\mathscr{R}_0 > 1$, as desired.

Conversely, assume that $\mathscr{R}_0 > 1$. Then, $\mathscr{R}_0^2 > 1$, $\frac{\beta S_0}{\rho+\mu} > 1$, $\beta S_0 > \rho + \mu$, $\beta S_0 - (\rho + \mu) > 0$, or $(\rho + \mu) - \beta S_0 < 0$. This implies that the discriminant in (3.1)-(3.2) is positive and consequently $\lambda_1 < 0$ and $\lambda_2 > 0$. Hence, since $\lambda_1 < 0$ and $\lambda_2 > 0$, the disease-free equilibrium, x^*, is unstable. Therefore, the disease-free equilibrium, x^*, is unstable if and only if the basic reproductive number is $\mathscr{R}_0 > 1$. □

Theorem 3.4 (**Stable equilibrium).** *The disease-free equilibrium, x^*, is locally asymptotically stable if and only if the basic reproductive number $\mathscr{R}_0 < 1$.*

Proof. For the sake of curiosity, assume that the disease-free equilibrium, x^*, is stable. Then we must have $\rho + \mu - \beta S_0 > 0$ so that the discriminant is negative. This results in two complex eigenvalues with negative real part. Thus, the disease-free equilibrium, x^*, would be asymptotically stable [10, Page 251]. The converse would hold in a similar argument. □

4 Fitting the Model to the Data

In this model, we want to verify that the $SEIRD$ model accurately depicts the transmission dynamics of COVID-19. To do this, we use real world data provided by the World Health Organization (WHO)[2] and fit the model to the data using the least squares approach. We fit the model to incidence weekly data of the United States and the world situation within a fifteen-week time span. The goal is to determine from the fitting curve the parameters β, ρ, and μ that best represent the observed data.

Notice that different time spans are used for the USA data and the world data. Distinct time intervals are used to correspond the time periods where the region experienced a constant weekly increase of COVID-19 cases. For the purpose of our study, it is crucial to collect data prior to major intervention strategies are implemented. Thus, the data has been deliberately selected prior the peak of the infectious period.

Table 4: Weekly COVID-19 Cases in the United States

Weeks	**Cases**
09/07/20	242694
09/14/20	275171
09/21/20	298149
09/28/20	296082
10/05/20	327514
10/12/20	382981
10/19/20	436292
10/26/20	548965
11/02/20	684493
11/09/20	1004852
11/16/20	1147581
11/23/20	1150654
11/30/20	1251632
12/07/20	1456800
12/14/20	1666736

Table 5: Weekly COVID-19 Cases Globally

Weeks	**Cases**
04/13/20	533134
04/20/20	553798
04/27/20	547077
05/04/20	582996
05/11/20	597373
Continued on next page	

Table 5 – continued from previous page

First column	Second column
05/18/20	679659
05/25/20	719524
06/01/20	842755
06/08/20	888759
06/15/20	1019230
06/22/20	1139197
06/29/20	1305436
07/06/20	1431541
07/13/20	1560279
07/20/20	1743320

Definition 4.1 (**Curve-fitting**). *The process of identifying the parameters of the model so that the solution best fits a given set of data [11].*

4.1 The Least Squares Approach

Definition 4.2 (**The Method of Least Squares**). *The least squares method is a statistical procedure to find the best fit for a set of data points by minimizing the sum of the squared residuals of data points from the fitted curve.*

Our objective consists on applying the least squares approach to estimate the parameters of the $SEIRD$ model that best fit the collected real data. The parameters to be estimated are β, ρ, μ, where γ is excluded from the fitting procedure. We assume that the x-coordinates (time in weeks) of the data are exact and represent the time independent variable, while the y-coordinates (number of infected individuals taken from WHO data) may be noisy or distorted and represent the time dependent variable. Particularly, our data is given in the form $\{(t_0, Y_0), \ldots, (t_{14}, Y_{14})\}$.

Definition 4.3 (**Residual**). *The difference between the actual value of the time dependent variable and the value estimated by the model.*

Thus, since we are fitting the number of COVID-19 cases for the infected variable $I(t)$ in the $SEIRD$ model (as shown in Table 4-5), we have that the residual is given by

$$r_i = Y_i - I(t_i) \qquad \text{for } i = 0, 2, \ldots, 14.$$

and the sum-of-squares error is given by

$$SSE = \sum_{i=0}^{15} r_i^2.$$

The sum-of-squares error (SSE) is a function of the parameters of the model [11]. Hence, in order to determine the best fit for the data, the values of β, ρ, μ that yield the fitted curve with the smallest SSE is desired. This becomes an optimization problem where we seek to minimize the SSE as much as possible. To do this, we use MATLAB's built-in function, *fminsearch*, to perform a minimization routine. The fitted curve for the USA data and world situation is shown below in Figure 2-3.

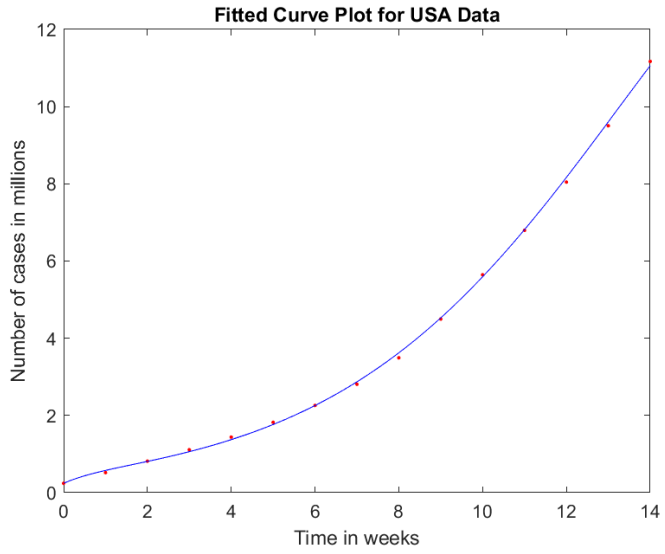


Figure 2: The graph display time in weeks versus the incidence number of cases in a weekly basis of the United States between September 7, 2020 and December 14, 2020. The red dots represent weekly amounts of cases obtained by the WHO. The initial conditions are as follow: $N = 330660000, S_0 = 325595940, E_0 = 4497300, I_0 = 242694, R_0 = 318920$, and $D_0 = 5146$. Convergence was reached after 52 iterations with an error of 0.070624.

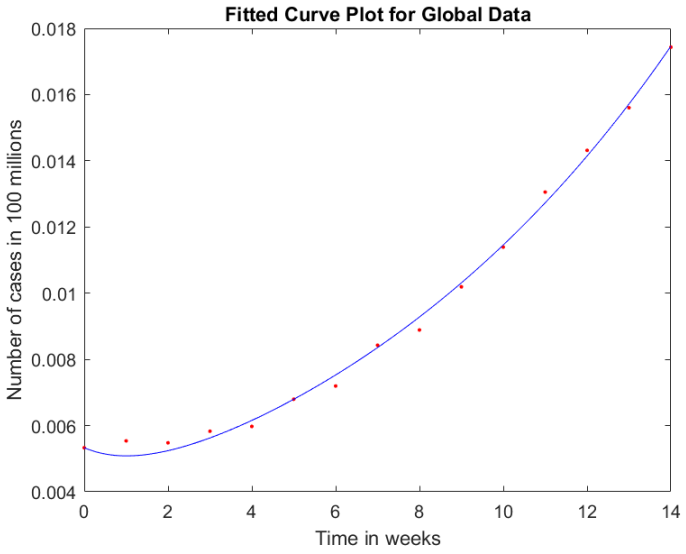


Figure 3: The graph display time in weeks versus the incidence number of cases in a weekly basis of the global situation between April 13, 2020 and July 20, 2020. The red dots represent weekly amounts of cases obtained by the WHO. The initial conditions are as follow: $N = 7800000000$, $S_0 = 7791272494, E_0 = 1171224, I_0 = 533134, R_0 = 7023148$, and $D_0 = 51585$. Convergence was reached after 131 iterations with an error of 0.000001.

4.2 Estimation of Parameters

The specified value for the initial conditions and initial guesses for the parameters have a pivotal role on the fit. In this fitting, we do not fit the initial conditions. Instead, we determine their value based on the corresponding collected data. The total number of fixed people in the population at time t_0, $N = S_0+E_0+I_0+R_0+D_0$, is taken to be the total population of the United States on September 7, 2020 and the total global population on April 13, 2020, respectively. We let I_0 denote the number of infectious individuals at time t_0 and D_0 to be number of dead individuals at time t_0. Based on I_0 and D_0, we adjust the values of E_0 and R_0 as necessarily. Thus, with N, E_0, I_0, and R_0 known, the susceptible population at t_0 follows to be $S_0 = N-(E_0+I_0+R_0+D_0)$.

A classical approach to the estimation of parameters is to identify features of the set of data and then choose parameters of our model to match those features. Based on existing research, the incubation period of COVID-19 and other coronaviruses ranges from 2–14 days [2]. Thus, we fix the incubation rate to be $\gamma = 0.16$, which is 6.25 days. Early estimates predict that the overall COVID-19 recovery rate to be between 97% and 99% and death rate to be between $1-2\%$, so we pre-estimate ρ and γ to be $\rho = 0.01818$ and $\mu = 0.9818$. Lastly, we fix the obtained estimated parameters, γ, ρ and μ, to manipulate the value of β until a good agreement with the data is attained. Ultimately, the desirable value yields to be $\beta = 0.01$.

Table 6: Fitted parameters for the USA and world data.

Region	**USA**	**Global**
β	0.0101	0.0119.
μ	0.9780	0.4311.
ρ	0.0185	0.0223.

Next, we examine the residual graphs of the USA data and the world data in order to verify that fitting curve is appropriate for the data. Notice that positive values for the residual mean the prediction was too low and negative values mean the prediction was too high. If the residual is equal to zero, this means that our fitted value was exactly correct. Consequently, in order for the fit to be good, the residuals should be randomly distributed, tending to cluster towards the middle of the plot.
By examining the residuals in Figure 4, we notice that for both the USA data and the world data, three out of the fifteen red dots almost perfectly fall on the blue line. This is a great indication of a good fit since it implies that at those fitted values the error is negligible. Also, the residual plots display a fairly random pattern, with approximately one half of the residuals positive and the other half negative. Therefore, since the fitting curve effectively reflects the data, it is safe to conclude that the $SEIRD$ model accurately depicts the transmission dynamics of COVID-19.

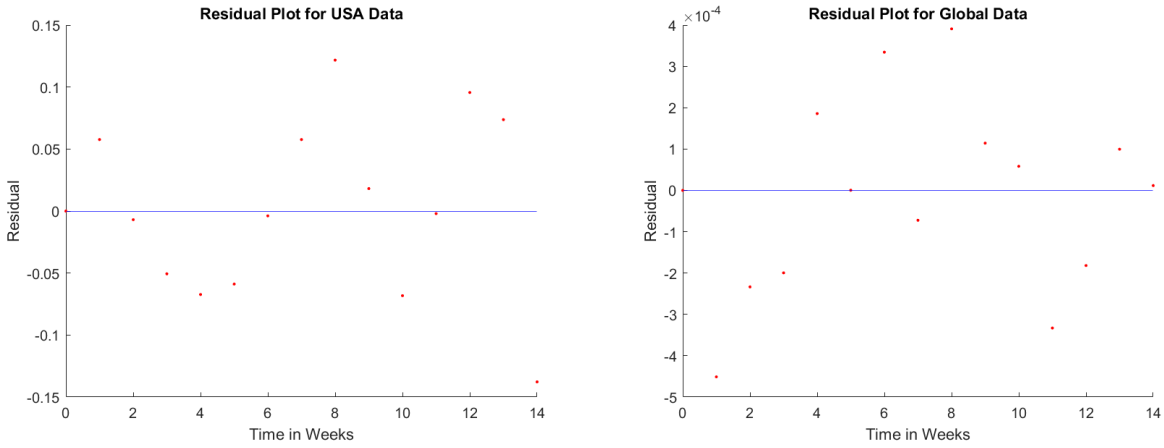


Figure 4: The graphs display the least squares residuals of the USA (*left*) and global situation (*right*). The distance from the red dots to the blue line at 0 illustrates how far the fitted value is from the actual value.

5 Optimal Control

Now, we introduce the finite-dimensional optimal control problem to determine how effective quarantine and isolation are at minimizing the number of infected individuals over time. Hence, we consider two controls which serve as prevention methods for COVID-19: quarantine, $q(t)$, and isolation, $z(t)$. In general, these two controls are widely used by health officials to combat the spread of an epidemic.

Definition 5.1 (**Quarantine**). *Separates and restricts the movement of people who were exposed to a contagious disease to see if they become sick.*

Definition 5.2 (**Isolation**). *Separates sick people with a contagious disease from people who are not sick [1].*

The goal of this study is to determine the effectiveness of implementing quarantine and isolation together to control the spread of COVID-19 over time. In our study, we consider the previous studies for COVID-19 dynamics that have explored on the topic on quarantine and isolation as optimal controls [13, 14, 8, 12]. For instance, we see that many previous research [14] consider an optimal control approach for the $SEIR$ system for vaccination and quarantine, but excludes isolation as possible preventive strategy. In fact, various research in the topic explore quarantine as its own compartment but considers the patients in the infectious period to be partially isolated so that these can not participate in the transmission of the virus [8]. On the other hand, some studies have found that isolation of the infected is not actively controlled in the SIR model, and therefore isolation is prioritized over quarantine [12]. Despite the fact that the study of optimal control for both quarantine and isolation has been also considered [13], the model could contains major constraints such as having the assumption that only limited quarantine control is available. In our study, we wish to study the impact on the transmission of COVID-19 dynamics when quarantine and isolation are applied both separately and simultaneously and contribute to the extensive archive of research on the topic.

5.1 The SEIRD Model with Optimal Control

In this section, we modify the SEIRD model by introducing the controls $q(t)$ and $z(t)$ as follows

$$\frac{dS}{dt} = -\beta SI \qquad S(0) = S_0 > 0, \tag{5.1}$$

$$\frac{dE}{dt} = \beta SI - \gamma E - q(t)E, \qquad E(0) = E_0 > 0, \tag{5.2}$$

$$\frac{dI}{dt} = \gamma E - z(t)I - \rho I - \mu I, \qquad I(0) = I_0 > 0, \tag{5.3}$$

$$\frac{dR}{dt} = q(t)E + z(t)I + \mu I, \qquad R(0) = R_0 > 0 \tag{5.4}$$

$$\frac{dD}{dt} = \rho I \qquad D(0) = D_0 > 0 \tag{5.5}$$

In this case, a portion of the exposed population to the virus will be quarantined for a certain amount of time t. The expression $q(t)E$ represents the portion of exposed individuals who went into quarantine. Similarly, $z(t)I$ is the portion of infectious individuals who went into isolation. It is assumed that after being in quarantine or isolation for t amount of time the person recovers from the virus. Thus, $q(t)E$ and $z(t)I$ individuals will automatically move to the recovered group once the period of quarantine/isolation is over.

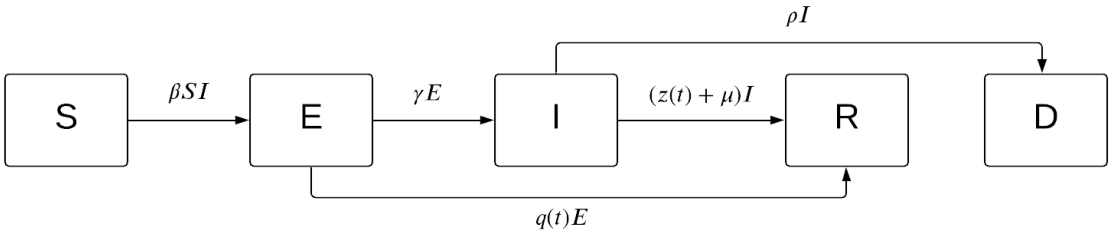


Figure 5: Flowchart diagram for the SEIRD model with controls $q(t)$ and $z(t)$.

In our optimal control problem for ordinary differential equations, the functions $q(t)$ and $z(t)$ represent the controls and $x(t)$ represents the state variable. The state variable satisfies a differential equation which depends on the control variables:

$$x'(t) = g(t, x(t), q(t), z(t)).$$

The principle technique for such an optimal control problem is to solve a set of necessary conditions that an optimal control and corresponding state must satisfy. In particular, we seek for piece-wise continuous controls $q^*(t)$ and $z^*(t)$ and the associated state $x(t)$ to minimize the given objective functional:

$$\min_{q,z} J(q, z) = \min_{q,z} \int_{t_0}^{t_1} f(t, x(t), q(t), z(t))dt$$

subject to

$$x'(t) = g(t, x(t), q(t), z(t))$$

$$x(t_0) = x_0 \text{ and } x(t_1) \text{ free}$$

where f is a continuously differentiable function defined as

$$f(t, x(t), q(t), z(t)) = I(t) + \frac{A}{2}q^2(t) + \frac{B}{2}z^2(t)$$

with A and B as positive constants. In order to find an optimal solution, we define the Hamiltonian and use Pontryangin's Maximun Principle to derive the adjoint equations and optimality conditions.

Theorem 5.1 (**Pontryangin's Maximun Principle**). *Suppose that $\overline{q}(t)$ and $\overline{z}(t)$ are optimal for the given objective functional and state equation $\overline{x}$. Then, there exists a piecewise differential adjoint variable $\lambda(t) = (\lambda_S, \lambda_E, \lambda_I, \lambda_R, \lambda_D)$ such that*

$$H(t, \overline{x}(t), q(t), z(t), \lambda(t)) \leq H(t, \overline{x}(t), \overline{q}(t), \overline{z}(t), \lambda(t))$$

for all controls $q(t)$ and $z(t)$ at each time t, where the Hamiltonian H is defined as

$$H = f(t, x(t), q(t), z(t)) + \lambda(t) \cdot g(t, x(t), q(t), z(t))$$

and the adjoint system is given by

$$\begin{aligned} \lambda'(t) &= -\frac{\partial H(t, \overline{x}(t), \overline{q}(t), \overline{z}(t), \lambda(t))}{\partial x}, \\ \lambda(t_1) &= 0. \end{aligned}$$

In our case, the Hamiltonian is equivalent to the addition of the integrand of the objective function and the dot product of the vectors $\lambda(t)$ and $x'(t)$. Thus, the Hamiltonian is defined as

$$\begin{aligned} H = &\left(I(t) + \frac{A}{2}q^2(t) + \frac{B}{2}z^2(t)\right) + \lambda_S\left(-\beta SI\right) \\ &+ \lambda_E\left(\beta SI - (\gamma + q(t))E\right) + \lambda_I\left(\gamma E - (z(t) + \rho + \mu)I\right) \\ &+ \lambda_R\left(q(t)E + (z(t) + \mu)I\right) + \lambda_D(\rho I)\,. \end{aligned}$$

Next, we use the Hamiltonian to derive the adjoint equations. Notice that, since our model consists of five differential equations, there will be an adjoint equation corresponding to each of the S, E, I, R and D compartment. Hence, the adjoint equations are defined as

$$\begin{aligned} \lambda_S' &= \lambda_S\beta I - \lambda_E\beta I \\ \lambda_E' &= \lambda_E(\gamma + q(t)) - \lambda_I\gamma - \lambda_R q(t). \\ \lambda_I' &= -1 + \lambda_S\beta S - \lambda_E\beta S + \lambda_I(z(t) + \rho + \mu) - \lambda_R(z(t) + \mu) - \lambda_D\rho. \\ \lambda_R' &= 0. \\ \lambda_D' &= 0. \end{aligned}$$

Lastly, posing the necessary conditions from Pontryagin's principle, we have that the Hamiltonian must have a critical point in the variable $q(t)$ and $z(t)$ at $\overline{q}(t)$ and $\overline{z}(t)$, respectively. Thus, the optimal conditions are obtained by setting the partial derivative of the Hamiltonian with respect to $q(t)$ and $z(t)$ to zero. Solving for $q(t)$ yields

$$\overline{q}(t) = \frac{\lambda_E E - \lambda_R E}{A}.$$

Similarly,

$$\overline{z}(t) = \frac{\lambda_I I - \lambda_R I}{B}.$$

To provide numerical simulations of the control problem, we use a numerical approach known as the Forward-backward sweep method. This method will be used to solve the state $x(t)$ forward in time from t_0 to t_1 and solve the adjoint equations $\lambda(t)$ backward in time from t_1 to t_0. To do this, we will use a modified version of the MATLAB's code provided in [9, Page 50].

The basic algorithm of the Forward-Backward Sweep Method is shown below:

5.2 Forward-Backward Sweep Method

1. **Step 1.** Make an initial guess for q and z over the interval.
2. **Step 2.** Using the initial condition $x(t_0) = x_0$ and the values for q and z, solve for x forward in time according to its differential equation in the optimality system.
3. **Step 3.** Use the transversality condition $\lambda_{N+1} = \lambda(t_1) = 0$ and the values for q, z and x to solve λ backward in time.
4. **Step 4.** Update q and z in each iteration by entering the new x and λ values into the characterization of the optimal control.
5. **Step 5.** Check for convergence. If values of the variables in the current iteration and the last iteration are negligibly close, output the current values as solutions. If values are not close, return to **Step 2**.

Even though there exists many methods for solving initial value problems, the Runge-Kutta 4th order (RK4) method is used in our study. The procedure used to implement RK4 to the $SEIRD$ model is shown below:

5.3 Runge-Kutta 4th Order Method

For the step size h and the state $x'(t) = g(t, x(t))$, the approximation of $x(t+h)$ is defined as

$$x(t+h) \approx x(t) + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

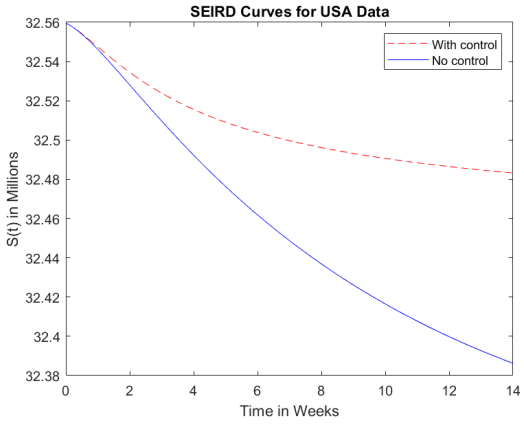
where

$$\begin{aligned}
k_1 &= g(t, x(t)) \\
k_2 &= g\left(t + \frac{h}{2}, x(t) + \frac{h}{2}k_1\right) \\
k_3 &= g\left(t + \frac{h}{2}, x(t) + \frac{h}{2}k_2\right) \\
k_4 &= g\left(t + \frac{h}{2}, x(t) + \frac{h}{2}k_3\right).
\end{aligned}$$

The initial conditions are the same as those used for our curve fitting and the initial guesses for the optimal controls are $q = 0$ and $z = 0$. The error for Runge-Kutta 4th is, as the name suggest, is $\mathscr{O}(h^4)$. Additional information on the accuracy of this method can be found in numerous texts [6, 5]. In the numerical experiment we use $h = \frac{T-t_0}{M}$ where T is the final time, t_0 is the initial time, and $M = 1000$ is the total number of subintervals.

6 Discussion

The $SEIRD$ model constructed considers the model with and without quarantine and isolation as optimal controls. We use MATLAB to execute the forward-backward sweep on the optimal $SEIRD$ model. The simulation results for the $SEIRD$ curves given by [a]-[e] in Figure 6 compares the trajectory of the disease as time progresses when a weight of $A = B = 0.5$ is applied on the optimal conditions $\overline{q}(t)$ and $\overline{z}(t)$. The solid and dashed lines represent the model without and with control, respectively. More specifically, the solid lines represent the model where no intervention method is applied to stop the spread of the disease whereas the dashed lines represent the model with the condition that each of the compartments must consider $q(t)$ and $z(t)$ as intervention strategies. The behaviour of the curves adheres to the same intuitive pattern we would expect in the real world. First, notice that the dashed curve for the susceptible class (Figure 6 [a]) decreases monotonically at a slower rate than its solid counterpart. This is due to the fact that more individuals become exposed to the virus in the $SEIRD$ model without control than with control, so the solid curve will decrease faster as individuals are contracting the virus and becoming part of the exposed compartment at a faster pace. Also, notice that within the first week of exposure to the virus, there is no significant distinction between the curves of susceptible individuals with control and without control. This implies that the time spend in quarantine by the majority of susceptible individuals is approximately a week. Thus, it is expected for the susceptible population to being to decrease after the first week as individuals are assumed to move automatically to the recovered compartment after being in quarantine, isolation, or both.



[a]

The remaining of the curves also coincide with the data. First, notice that from our system of equations, the portion of individuals being quarantined are extracted from the exposed compartment. Thus, the curve for the exposed population (Figure 6, [b]) decreases abruptly as it is directly linked to the portion of quarantined individuals in subject . The infected curve (Figure

6 [c]) shows that between weeks 0 to 1.5 the rate at which individuals contract the disease increased. This correlates with what is expected since in the early stages there is a greater portion of susceptible individuals that are not immune to the disease. Similarly, the curve for the recovered population (Figure 6, [d]) and death population (Figure 6, [e]) is consistent with what is predicted since the use of quarantine and isolation results in fewer fatalities since less individuals are becoming infected in the first place.

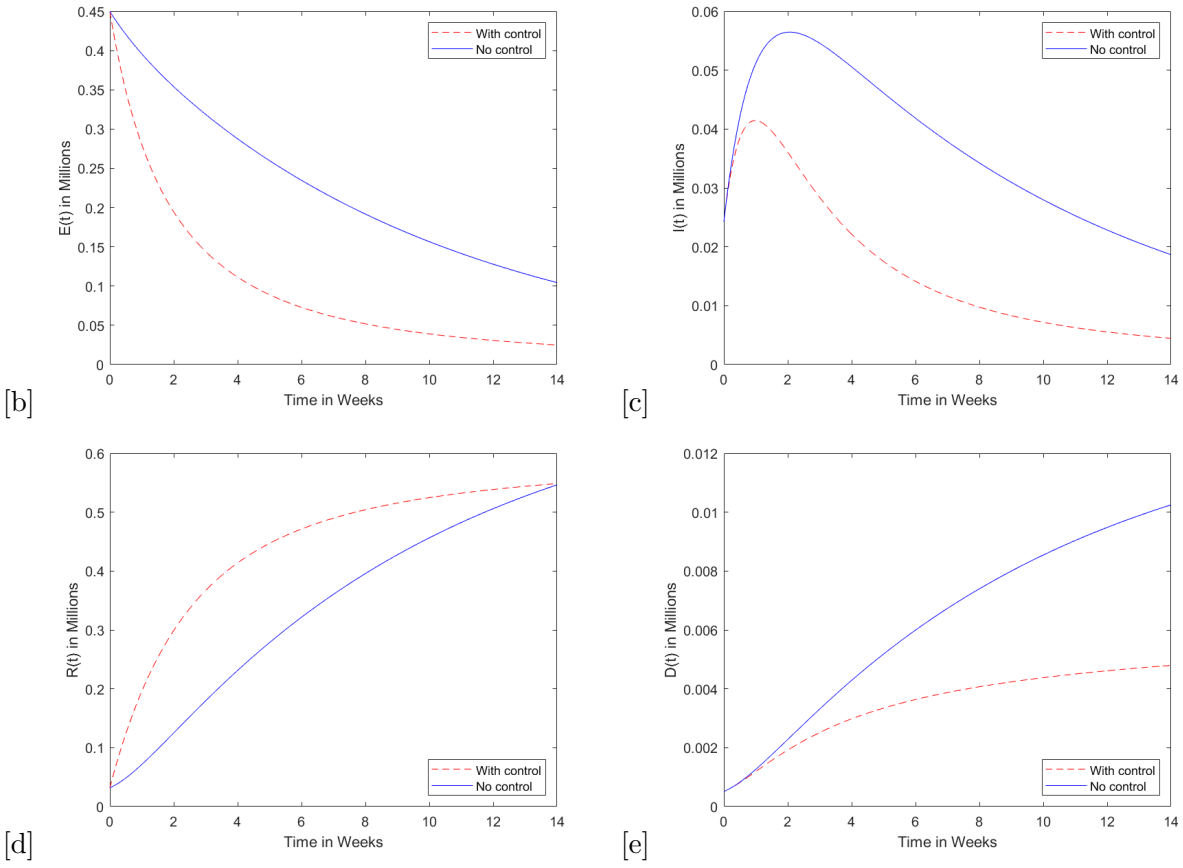


Figure 6: The five plots display each of the curves for the $SEIRD$ model in millions for the USA data. The solid curves are the disease running its course through the population. The dashed curves display the effects of quarantine and isolation on containing the spread of the disease. The initial conditions are as follow: $N = 330660000, S_0 = 325595940, E_0 = 4497300, I_0 = 242694, R_0 = 318920,\ D_0 = 5146$, with $A = 0.5$, and $B = 0.5$

The following curves for the optimal $SEIRD$ model using world data (Figure 7) yield a similar result.

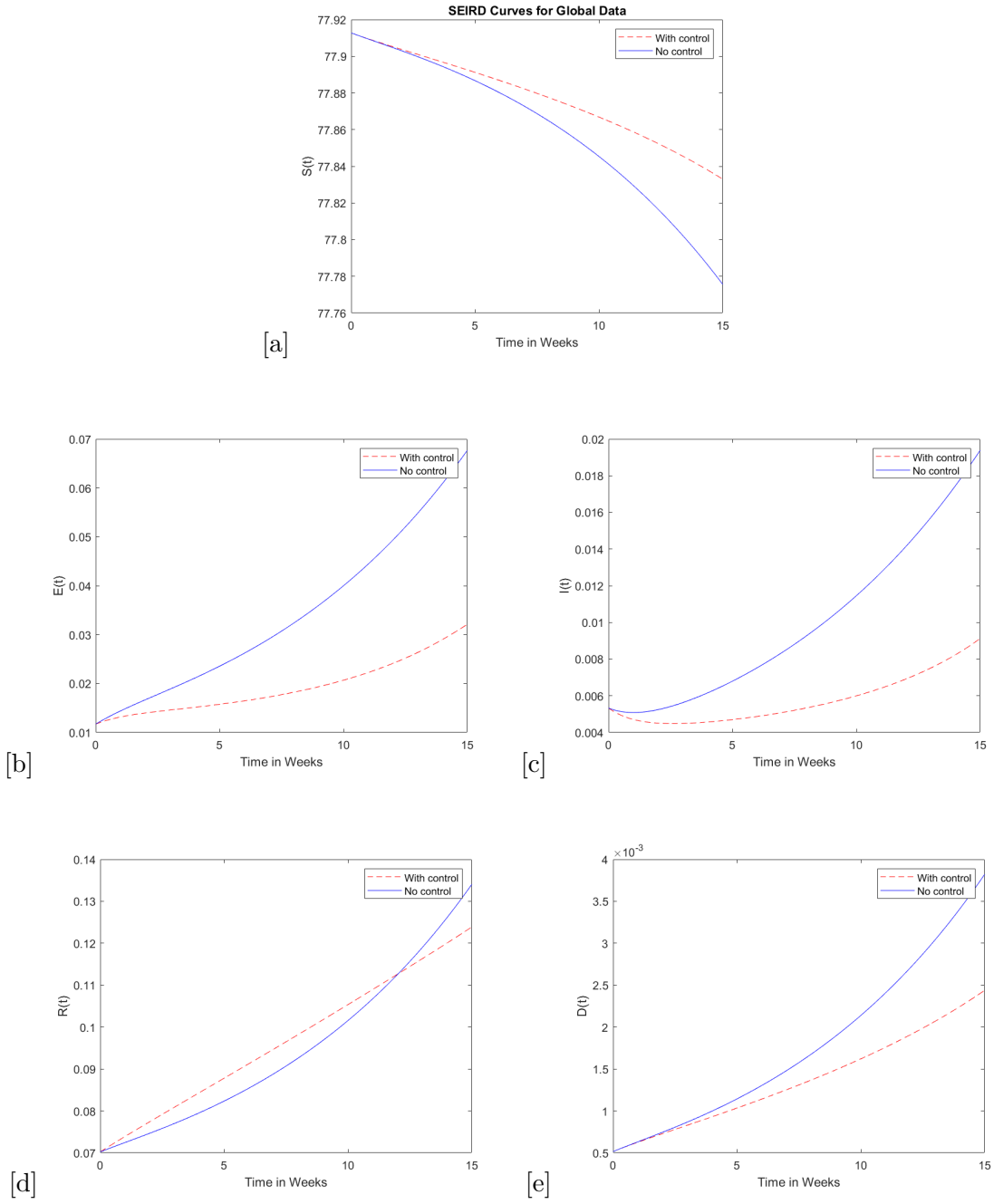
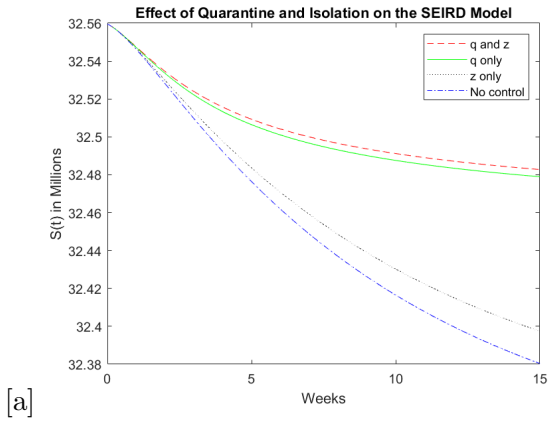
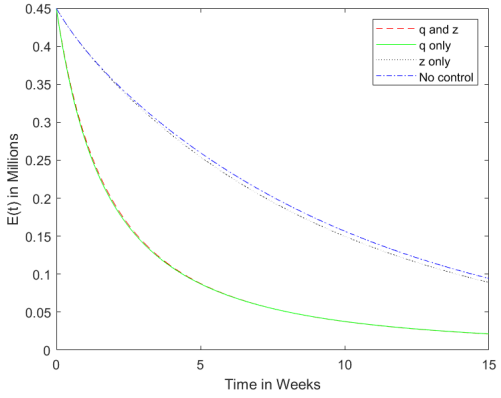


Figure 7: The five plots display each of the curves for the $SEIRD$ model in 100 millions. The solid curves are the disease running its course through the population. The dashed curves display the effects of quarantine and isolation on containing the spread of the disease. The initial conditions are as follow:$N = 7800000000$, $S_0 = 7791272494, E_0 = 1171224, I_0 = 533134, R_0 = 7023148$, and $D_0 = 51585$, with $A = 0.5$, and $B = 0.5$

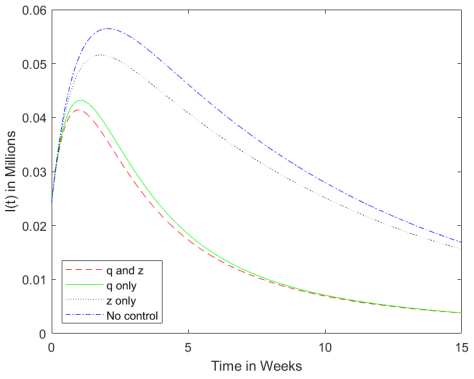
Next, we examine the effect of quarantine and isolation separately using USA data. The plot of Figure 8 considers the curves for each of the compartments when quarantine and isolation are examined separately as control strategies. First, from Figure 8 [a]-[b] we notice an eye catching pattern. Notice that the curve for the model with both control (red dashed curve) and the curve with only quarantine as a control (green solid curve) have very similar trajectories. This phenomenon is also present between the curve without control (blue dotted-dashed curve) and the curve with only isolation as a control (black dotted curve). This could be due to the fact that quarantine has greater impact than isolation on stopping the spread of the disease. This pivotal because it shows that quarantine must be prioritized when it comes to implementing early intervention strategies to control the disease. Also, notice that as the value of A and B grow in size, the curves will get closer to each other. As a result, based on the curves we determined that preventive methods such as quarantine and isolation, in particular quarantine, are key in order to stop the spread of COVID-19.



[a]



[b]



[c]

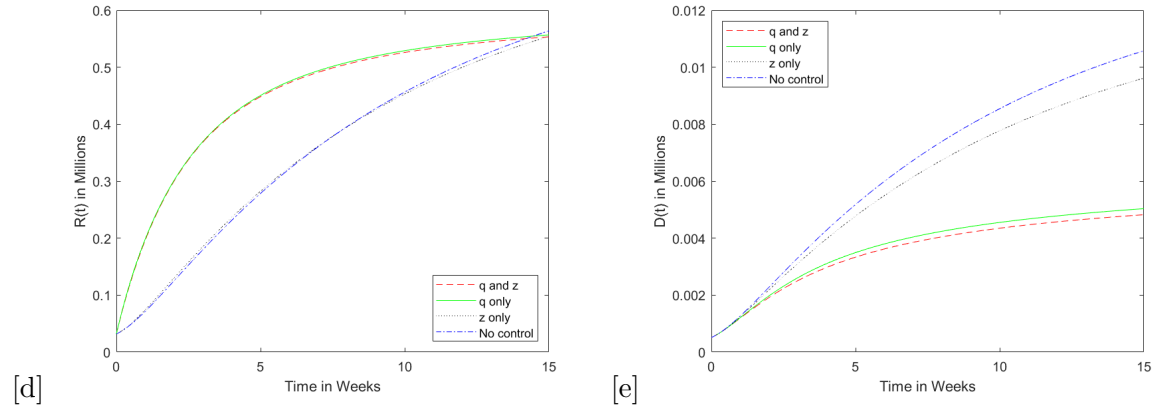


Figure 8: The five plots display each of the curves for the $SEIRD$ model in millions for the USA data with no control, with q and z independently, and both q and z simultaneously. The initial conditions are as follow: $N = 330660000, S_0 = 325595940, E_0 = 4497300, I_0 = 242694, R_0 = 318920$, $D_0 = 5146$, $A = 0.5$, and $B = 0.5$

As previously mentioned, one of the main goals of this study was to determine how effective quarantine and isolation are at controlling the spread of COVID-19. To archive this, the optimal conditions for quarantine and isolation were determined through an optimal control approach using real data. However, the $SEIRD$ system is extremely sensitive to the initial conditions and initial guesses of the parameters; slight deviations can lead to completely different outcomes. Thus, we take a closer look at the parameters and variables to validate our model. The goal is to decide qualitatively how much influence the quarantine rate and isolation rate have in the model's output [11]. From Figures 6 and 7, it is evident that introducing an optimal control is an effective preventive measure in the population. However, when fitting the model to our data, the estimated fitted parameters vary drastically depending on the scaling of the initial guesses for our parameters, the initial conditions, and the data set. Recall that fitted value for μ is $\mu = 0.97796$ for the USA data and $\mu = 0.43108$ for the world data. The difference in recovery rates relies on the idea that COVID-19 was more fatal globally than locally in the USA. Also, we focus on the rates of the optimal controls, $\overline{q}(t)$ and $\overline{z}(t)$ when the value of A and B ranges from 0.1 to 0.5 (Figure 9). Notice that as the weight of A and B increase, the weight of $\overline{q}$ and $\overline{z}$ decrease. From this we can conclude that when the value of A and B becomes sufficiently large, the values of $\overline{q}$ and $\overline{z}$ become ineligible. Hence, it is safe to conclude that the values of A and B are sensitive constants since the rate at which exposed individuals are being quarantined and the rate at which infected individuals are being isolated are directly related to the values of A and B.

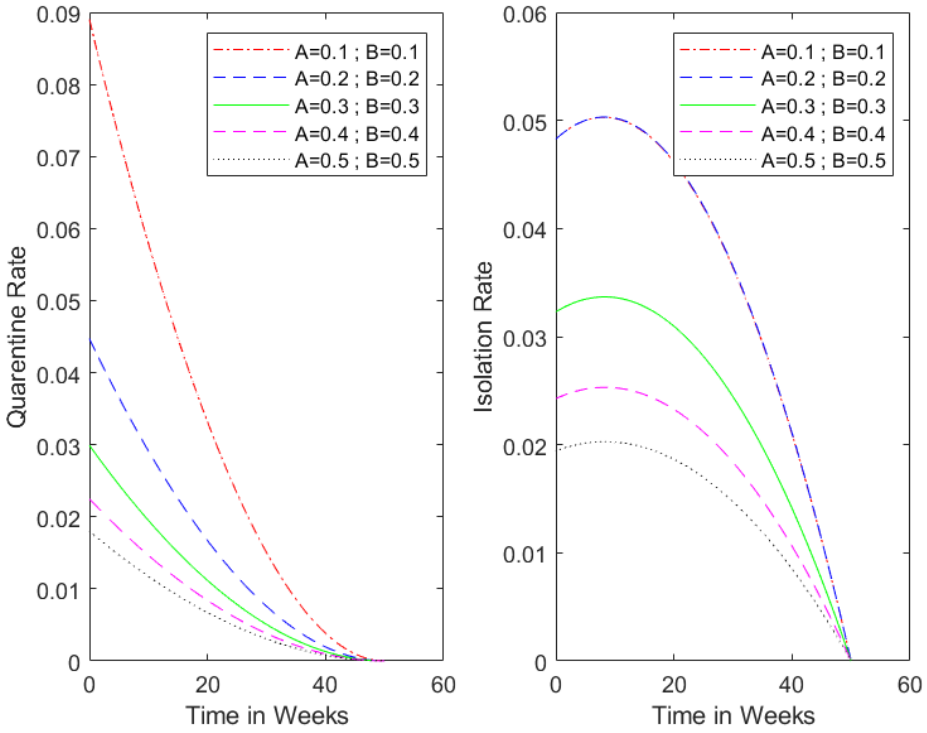


Figure 9: The graph displays the time in weeks vs the quarantine rate (*left*) and isolation rate (*right*) for the USA. The values for A and B range between 0.1 and 0.5 with a 0.1 increase.

In addition, notice that over a time interval of sixty weeks, the quarantine rate decreases in an exponential decay. More importantly, observe that approximately, during the first fifty weeks, the quarantine rate decreases entirely. This could imply that as more individuals are being quarantined over time, the more people will recover and gain immunity in the long run. Furthermore, notice an interesting pattern present in the curves given by the isolation rate. We observe that the isolation rate increases for approximately the first ten weeks and then decreases. Thus, we see that the isolation rate behaves similar to the curves of the infected population given in Figure 6 [c]. This could be because as the infected population increases, the isolation rate also increases to fight the disease. In general, the fact that the quarantine and isolation rate is zero after fifty weeks of being introduced as controls implies that after fifty weeks (or about a year) we can expect quarantine and isolation to no longer necessary. This is critical information because it provides insight on the portion of individuals that should be quarantined and isolated in order to stop the spread of the disease in a desired time interval.

7 Conclusion

In this paper we implement the proposed control approach in the COVID-19 $SEIRD$ model with real data collected from WHO [2]. The region from which the data was collected was for the United States and the world. The use of numerical and analytical methods were used to quantify and interpret the dynamics by which COVID-19 spreads. To measure the transmission potential of COVID-19, the basic reproductive number was calculated and determined to be $\mathscr{R}_0 > 1$ for both the USA and world data. We conclude that the $SEIRD$ system is locally unstable at the disease-free equilibrium. The curves for the control provide critical and useful information that indicates that COVID-19 will continue to spread if no further action is taken by heath officials.

References

[1] Quarantine and isolation. Centers for Disease Control and Prevention.

[2] Who coronavirus disease (covid-19) dashboard. World Health Organization, covid19.who.int/.

[3] Covid-19 overview and infection prevention and control priorities in non-us healthcare settings. 2021.

[4] A timeline of covid-19 developments in 2020. 2021.

[5] John Butcher. A multistep generalization of runge-kutta methods with four or five stages. Journal of the ACM.

[6] John Butcher. On the convergence of numerical solutions of ordinary differential equations. Math. Comp, 1966.

[7] ADNAOUI Khalid EL BERRAI Imane and BOUYAGHROUMNI Jamal. Proof of existence and uniqueness of solutions in epidemiology: Case of si model. 2017.

[8] Zhu Y. Gao Y Huang, B. The analysis of isolation measures for epidemic control of covid-19. Appl Intell.

[9] S. Lenhart and J.Workman. Optimal control applied to biological models.

[10] J. David Logan. A First Course in Differential Equations.

[11] M. Martcheva. An introduction to mathematical epidemiology. Text in Applied Mathematics, Springer.

[12] Osamu Mitarai and Nagato Yanagi. Suppression of covid-19 infection by isolation time control based on the sir model and an analogy from nuclear fusion research. 2020.

[13] Vladislav Nenchev. Optimal quarantine control of an infectious outbreak. Chaos, solitons, and fractals.

[14] Topcu U Xu Z, Wu B. Control strategies for covid-19 epidemic with vaccination, shield immunity and quarantine: A metric temporal logic approach. Chaos, solitons, and fractals.