MMAE 500

Discovering Dynamic Patterns from Coronavirus Data using Dynamic Mode Decomposition with Control

Julia Briden

Lab Date: April 25, 2020

Abstract

Dynamic mode decomposition was used to analyze and predict dynamic patterns from COVID-19 (coronavirus) data and an optimized method of control was generated to reduce the number of future confirmed cases.

Contents

1 Introduction **1**

2 Methods **2**

2.1 Data Pre-Processing 2

2.2 Dynamic Mode Decomposition 3

2.3 Prediction 4

2.4 Dynamic Mode Decomposition with Control 5

3 Results **6**

3.1 Identifying System Dynamics using DMD 6

3.2 Predicting Future Confirmed Cases using DMD 11

3.3 Controlling the Spread of Coronavirus using DMDc 14

4 Discussion of Results **18**

5 Conclusions **21**

6 Literature Cited **22**

List of Figures

1 Total confirmed coronavirus cases per day in the United States. 2

2 Total confirmed coronavirus cases per day in the world. 3

3 DMD reconstructed US confirmed cases data. 6

4 DMD reconstructed world confirmed cases data. 7

5 Eigenvalues of for US confirmed cases data. 8

6 Eigenvalues of for world confirmed cases data. 9

7 The modes, $\tilde{\mathbf{U}}$, and dynamics, $\tilde{\mathbf{V}}$ for US confirmed cases. 9

8 The modes, $\tilde{\mathbf{U}}$, and dynamics, $\tilde{\mathbf{V}}$ for world confirmed cases. 10

9 The mode magnitude, $\tilde{\mathbf{U}}$, for each US state. 10

10 Predicted and actual test data for the US. 11

11 Predicted and actual test data for the world. 12

12 Predicted data for the US for 100 days after April 28th, 2020. 13

13 Predicted data for the world for 100 days after April 28th, 2020. 13

14 Reduced order continuous system with no inputs for the US. 15

15 Reduced order continuous system with no inputs for the world. 15

16 Reduced order continuous system with feedback and unity proportional control for the US. 16

17 Reduced order continuous system with feedback and unity proportional control for the world. 16

18 Reduced order continuous system with optimal control for the US. 17

19 Reduced order continuous system with optimal control for the world. . . . 17

20 Functions in time domain from eigenvalue placement. 18

List of Tables

1 Eigenvalues for $\tilde{\mathbf{A}}$. 7

2 Mean error for DMD predicted test data. 12

3 Gain values and eigenvalues for $\tilde{\mathbf{A}}$ with optimal control. 15

1 Introduction

Applying quantitative methods to understanding disease dynamics and planning intervention provides an important push towards eradication of human infectious diseases. Dynamic mode decomposition (DMD) has been utilized to discover coherent spatial-temporal modes in high-dimensional data collected from complex systems with time dynamics. Recently, DMD was applied to three different infectious disease sets, including Google Flu Trends Data, pre-vaccination measles in the UK, and the paralytic poliomyelitis wild type-1 cases in Nigeria [1]. Comparing the eigenvalue spectrum and dominant modes across infectious disease data sets allowed for detection of large-scale dynamic patterns within the data. In addition, a new method which extends DMD, dynamic mode decomposition with control (DMDc), provides the ability to disambiguate between the underlying dynamics and effects of actuation, resulting in accurate input-output models [2]. DMDc was applied to the analysis of infectious disease data with mass vaccination (actuation). These finding were extended to analyze dynamic patterns from COVID-19 (coronavirus) data in this paper.

2 Methods

2.1 Data Pre-Processing

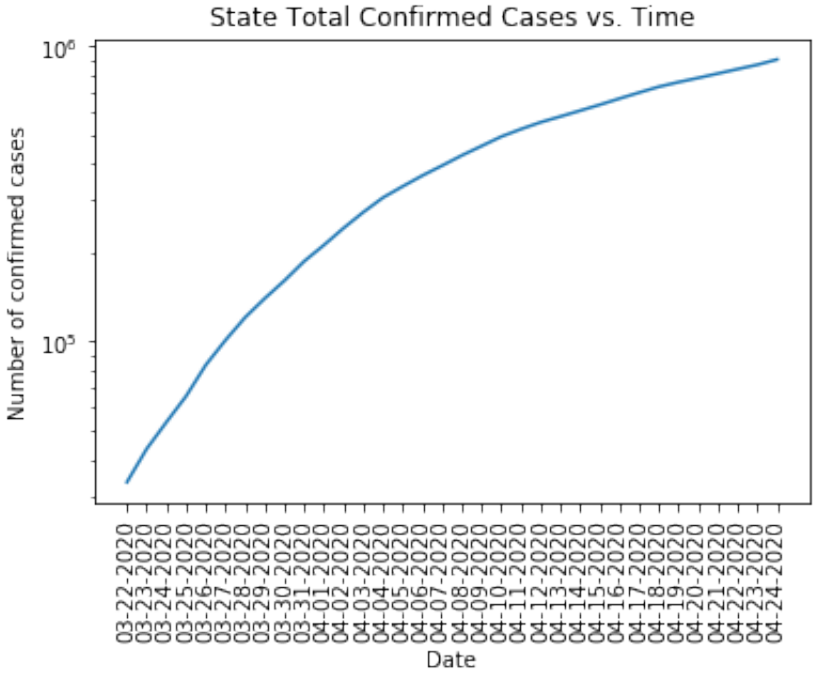

Figure 1: Total confirmed coronavirus cases per day in the United States.

The data used for determining the system dynamics of the coronavirus includes confirmed cases time series for the world and the United States (Figures 1 and 2). The data was accessed through the John Hopkins repository [3]. In order to have sufficient initial conditions for the dynamical system, the state data was spliced to start at the date where all locations had two or more confirmed cases of coronavirus. The world data was spliced so that most countries had more than one case. Each data set was arranged as a matrix with locations as rows and dates as columns (Equation 1).

$$\mathbf{X} = [\mathbf{x}(t1), \mathbf{x}(t2), ...\mathbf{x}(tm)] \tag{1}$$

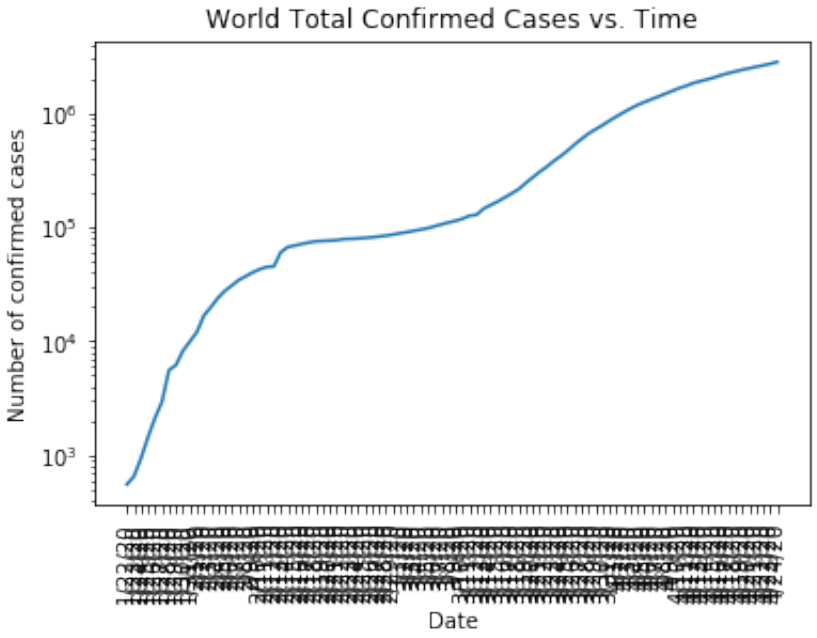

Figure 2: Total confirmed coronavirus cases per day in the world.

2.2 Dynamic Mode Decomposition

Dynamic mode decomposition (DMD) provides a modal decomposition where each mode consists of spatially correlated structures that have the same linear behavior in time [4]. The DMD algorithm finds the eigenvalues and eigenvectors of the best-fit linear operator, $\mathbf{A}$, that relates to the snapshot matrices in time (Equation 2).

$$\mathbf{X}' \approx \mathbf{A}\mathbf{X} \tag{2}$$

Where $\mathbf{X}'$ is the data matrix at a later time and $\mathbf{X}$ is the data matrix at the current time. The best-fit operator, $\mathbf{A}$, is found by multiplying the next time step of the data by the pseudo-inverse of the current data (Equation 3).

$$\mathbf{A} = \mathbf{X}'\mathbf{X}^{\dagger} \tag{3}$$

This pseudo-inverse is computed via the singular value decomposition of the matrix X

(Equation 4).

$$\mathbf{X} \approx \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^{*} \tag{4}$$

Where $\tilde{\mathbf{U}}\ \epsilon\ C^{nxr}$ are POD modes, $\tilde{\mathbf{\Sigma}}\ \epsilon\ C^{rxr}$, and $\tilde{\mathbf{V}^{*}}\ \epsilon\ C^{mxr}$ are the dynamics. The exact or approximate rank of the data matrix is r $\leq$ m. Finally, instead of computing $\mathbf{A}$ directly, the projection of A onto the leading singular vectors is computed (Equation 5).

$$\tilde{\mathbf{A}} = \tilde{\mathbf{U}}^{*}\mathbf{X}'\tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}^{-1} \tag{5}$$

Where $\tilde{\mathbf{A}}$ has the same nonzero eigenvalues as $\mathbf{A}$.

2.3 Prediction

To predict future values of $\mathbf{X}$, the dynamic modes were used to calculate the predicted data, $\mathbf{X}$'. To calculate the current reduced data matrix, Equation 6 was used.

$$\tilde{\mathbf{X}} = \tilde{\mathbf{U}}^{\dagger}\mathbf{X} \tag{6}$$

The reduced best-fit linear operator was used to iterate over to get the data matrix at a later time step (Equation 7).

$$\tilde{\mathbf{X}}' = \tilde{\mathbf{U}}\tilde{\mathbf{X}} \tag{7}$$

The high-dimensional predicted data was obtained by multiplying the reduced dynamic modes by the data matrix (Equation 8).

$$\mathbf{X} = \tilde{\mathbf{U}}\mathbf{X} \tag{8}$$

The predictions were validated by dividing the data set into test and training data. The training data was used to take the DMD and predict the test data set. Then the mean

error and standard deviation of the error between the predicted data and test data were computed.

2.4 Dynamic Mode Decomposition with Control

Linear control theory was applied to the reduced-order dynamic model, obtained through DMD, to determine the effects of feedback in the system and to determine the optimal controller for the system. Since $\tilde{\mathbf{A}}$ was found in discrete time, it must be converted to continuous-time to be used in continuous control methods. The continuous system is defined in Equations 9 and 10.

$$\dot{X} = \mathbf{A}x + \mathbf{B}u \tag{9}$$

$$y = \mathbf{C}x + \mathbf{D}u \tag{10}$$

Where u is the input vector, y is the output vector, and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are coefficient matrices. When using feedback control, the control input is dependant upon the output of the system. This allows controller to adjust based on the error from the desired output. Linear quadratic regulars (lqr) provide a method for finding a controller with guaranteed robustness. The gains and eigenvalues of the optimal controller are found by minimizing Equation 11.

$$J = \int_{t=0}^{\infty} (\vec{x}^T Q \vec{x} + \vec{u}^T R \vec{u}) dt, \tag{11}$$

Where Q governs the difference between the real and desired states and R penalizes the controller. Feedback control with unity gain and a lqr were tested with the DMD-fitted reduced order data to determine possible control strategies for the system.

3 Results

3.1 Identifying System Dynamics using DMD

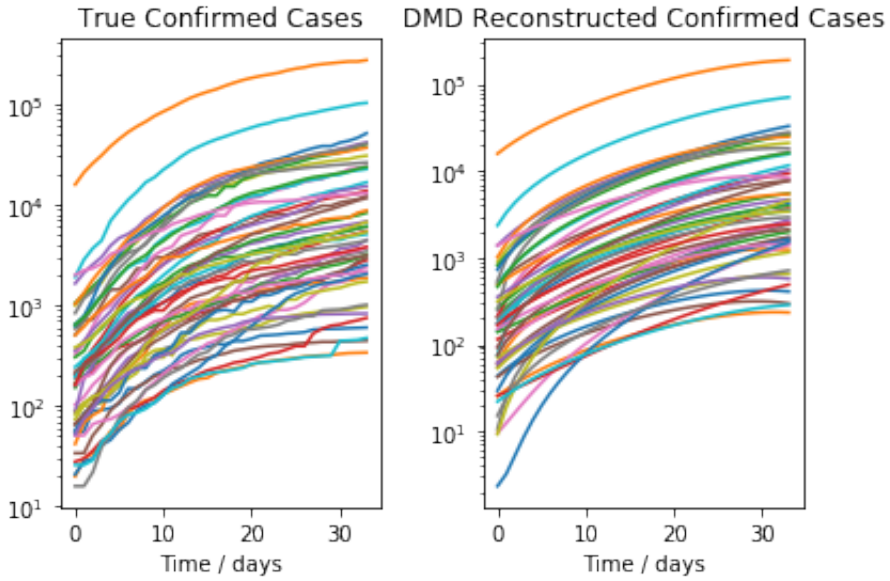

Figure 3: DMD reconstructed US confirmed cases data.

The DMD was taken for coronavirus confirmed cases in the United States (Figure 3) and the world (Figure 4). The rank of the singular value decomposition for the DMD was set to three. The rank was decided on by iterative techniques to determine what rank yielded the most accurate predictions, discussed in detail later.

The discrete eigenvalues of the reduced linear operators, $\tilde{\mathbf{A}}$, are plotted in Figures 5 and 6. These eigenvalues correspond to the systems' poles. Poles outside of the unit circle are unstable and poles outside of the unit circle with imaginary components show that the system will oscillate to infinity in the time domain. Poles inside of the unit circle show that the system will eventually reach a constant value and, therefore, they are stable poles. To easily determine system dynamics, the eigenvalues were converted to the continuous domain and are included in Table 1.

The modes, $\tilde{\mathbf{U}}$, and dynamics, $\tilde{\mathbf{V}}$, for the systems are plotted in Figures 7 and 8. These

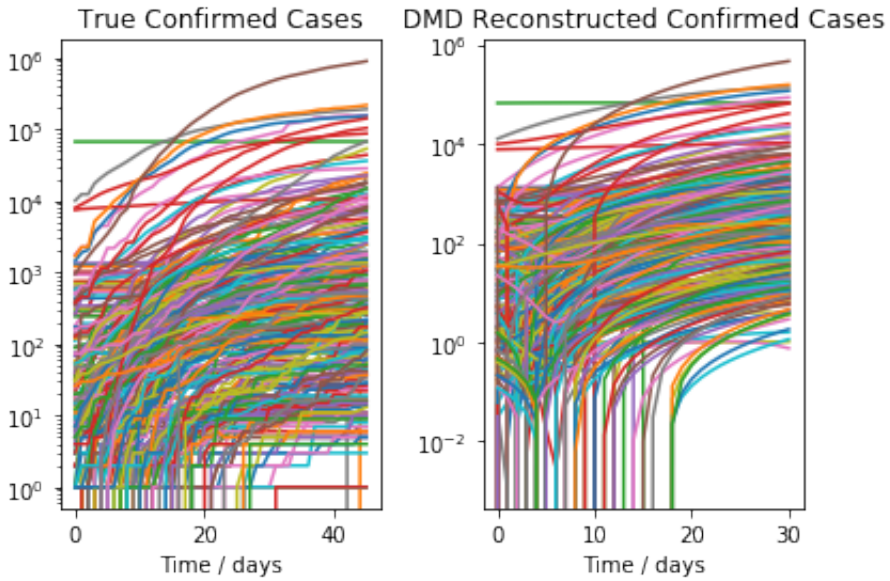

Figure 4: DMD reconstructed world confirmed cases data.

	Discrete Eigenvalues	Continuous Eigenvalues
US	1.0375+/-0.05051j, 1.0108	0.03797+/-0.04864j, 0.01073
World	1.0114+/-0.05963j, 1.0122	0.01312+/-0.05889j, 0.01209

Table 1: Eigenvalues for $\tilde{\mathbf{A}}$.

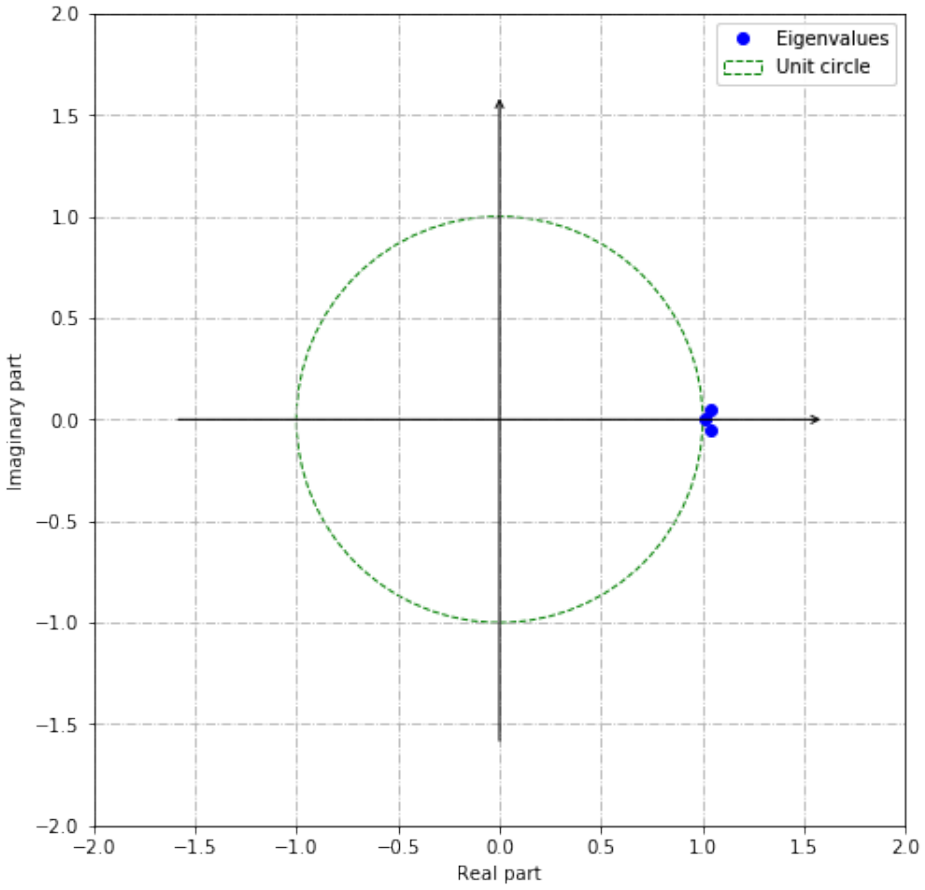

Figure 5: Eigenvalues of for US confirmed cases data.

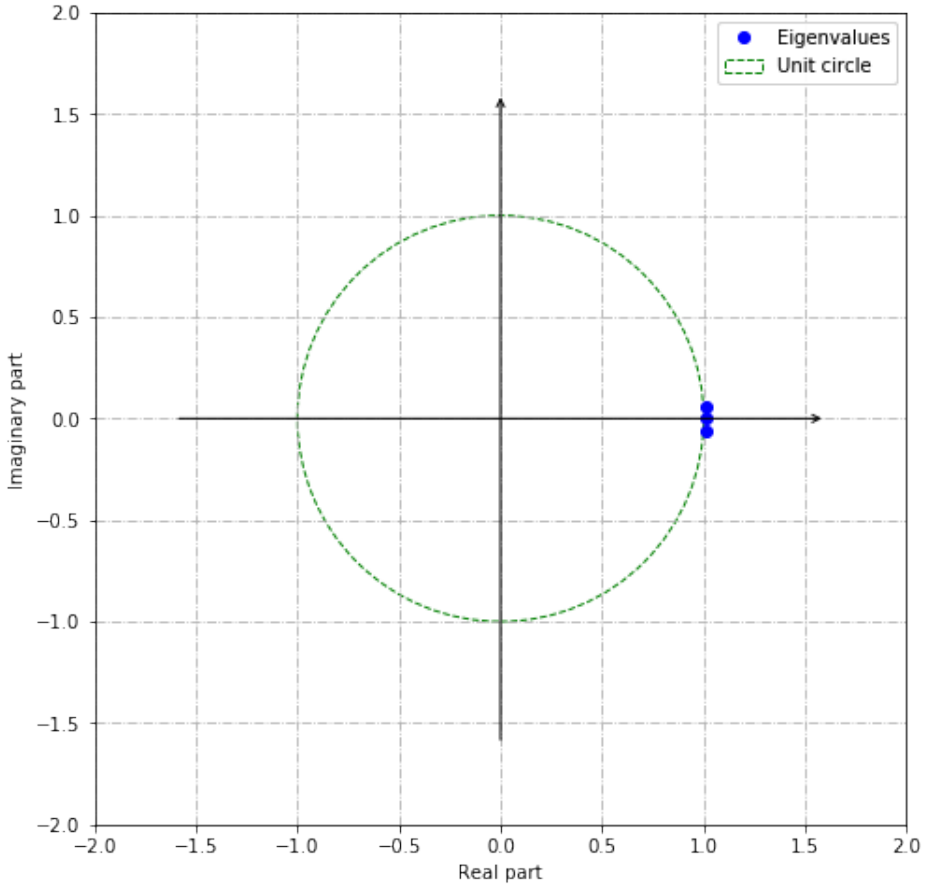

Figure 6: Eigenvalues of for world confirmed cases data.

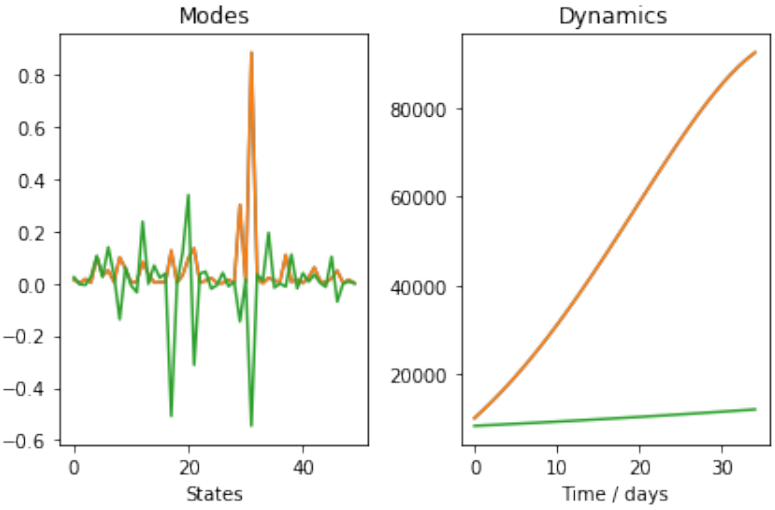

Figure 7: The modes, $\tilde{\mathbf{U}}$, and dynamics, $\tilde{\mathbf{V}}$ for US confirmed cases.

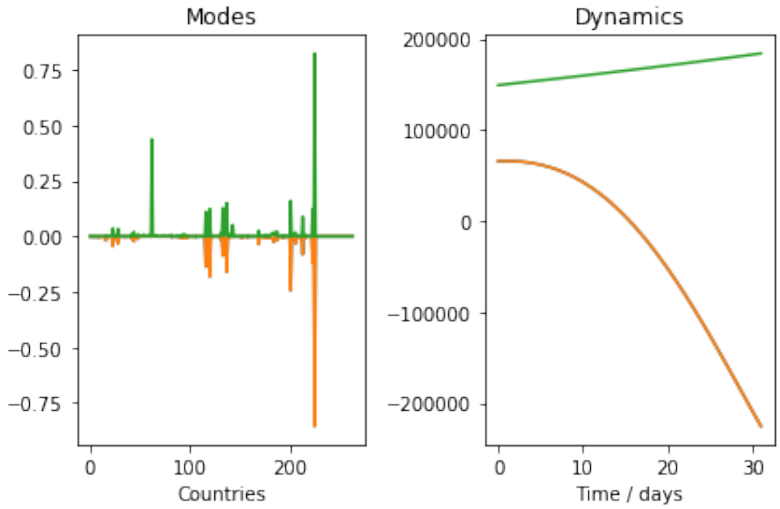

Figure 8: The modes, $\tilde{\mathbf{U}}$, and dynamics, $\tilde{\mathbf{V}}$ for world confirmed cases.

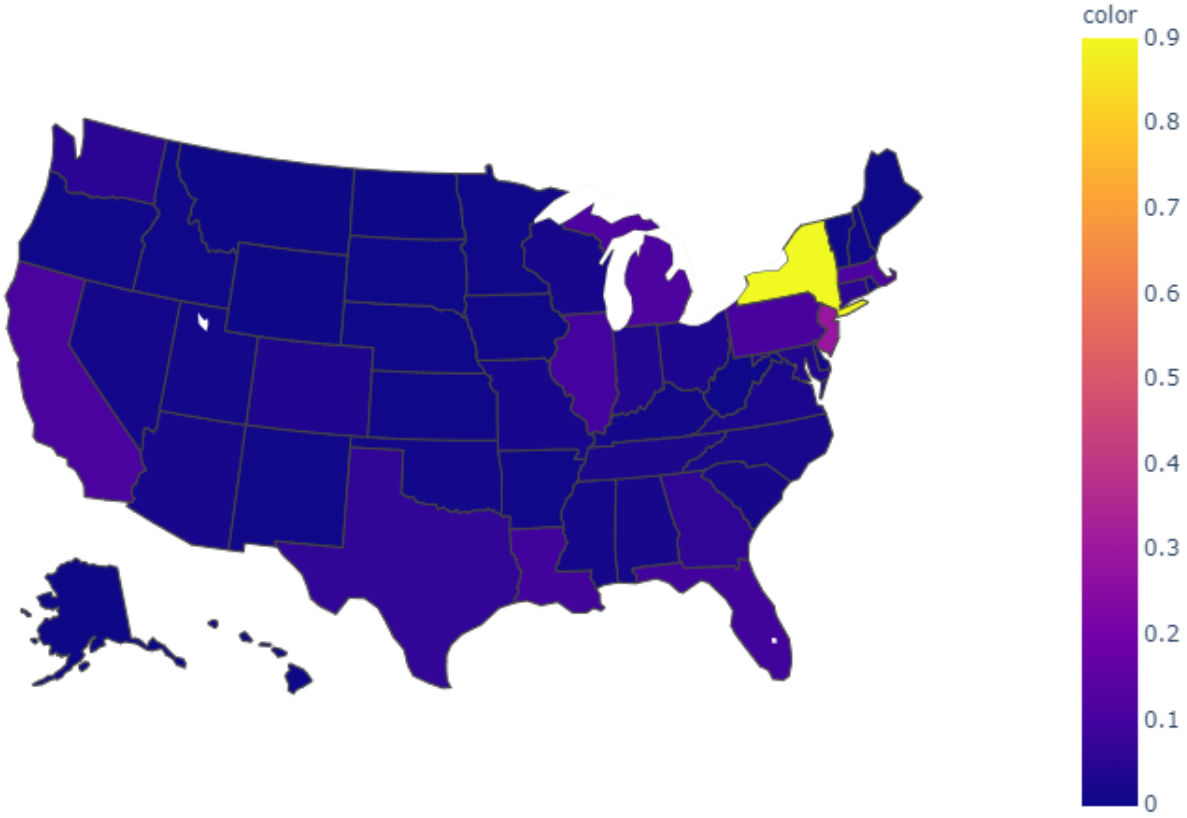

Figure 9: The mode magnitude, $\tilde{\mathbf{U}}$, for each US state.

values represent system variations across locations and time. Figure 9 shows the magnitude of the dynamic modes for each state when taking an SVD of rank one.

3.2 Predicting Future Confirmed Cases using DMD

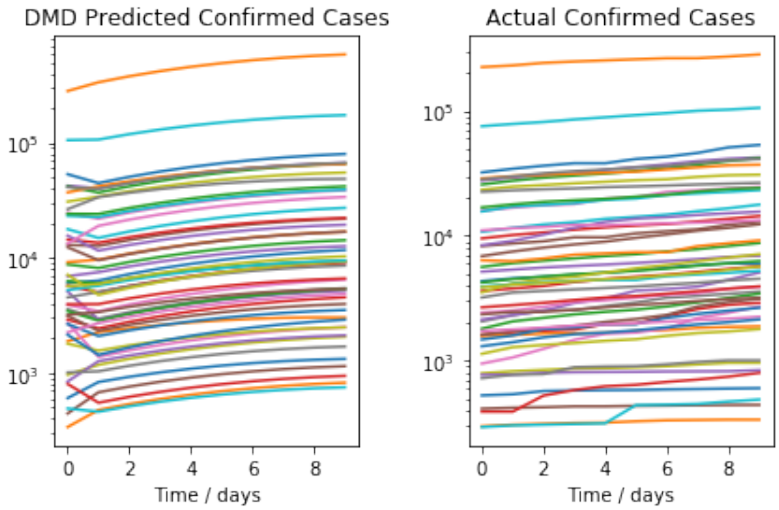

Figure 10: Predicted and actual test data for the US.

DMD was used to predict future trends for coronavirus confirmed cases data. The accuracy of DMD predictions was calculated by dividing the data matrix, $\mathbf{X}$, into testing and training data. The training data was used for taking the DMD and future data predictions were made for the duration of the test data. The optimal number of days to use for testing data was determined to be 10 days, using iterative methods. In addition, the optimal SVD rank was iteratively determined to be three, since it consistently yielded the best accuracy when predicting test data. The DMD predicted confirmed cases and actual confirmed cases are plotted in Figures 10 and 11. Table 2 shows the mean error for the DMD predicted test data for the US and the world for a 10 day span.

The number of future confirmed coronavirus cases were predicted using the DMD linear operator. Figures 12 and 13 show the predicted future coronavirus confirmed cases for

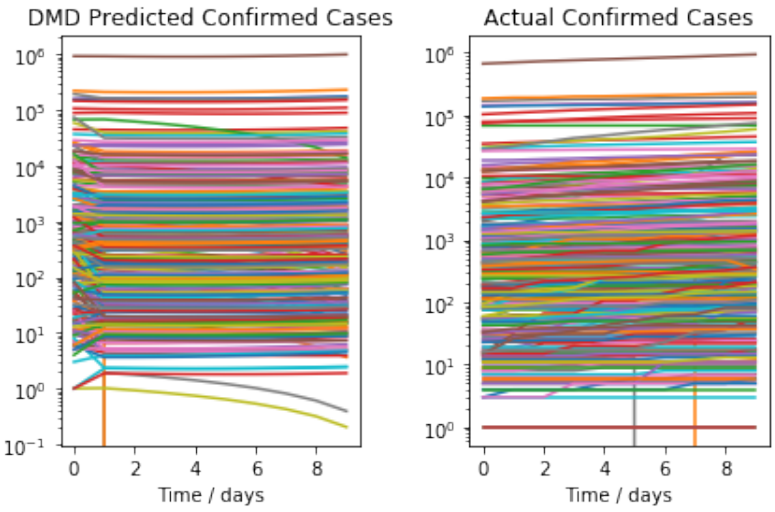

Figure 11: Predicted and actual test data for the world.

Day	US	World
1	54+/-33%	35+/-38%
2	38+/-11%	21+/-31%
3	48+/-15%	17+/-24%
4	56+/-18%	14+/-19%
5	63+/-22%	13+/-16%
6	69+/-25%	12+/-13%
7	71+/-27%	12+/-12%
8	72+/-31%	12+/-11%
9	70+/-33%	13+/-10%
10	69+/-35%	14+/-10%

Table 2: Mean error for DMD predicted test data.

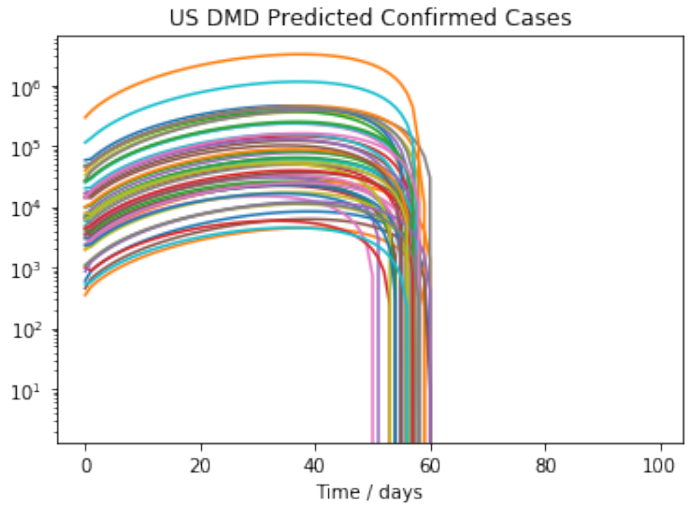

Figure 12: Predicted data for the US for 100 days after April 28th, 2020.

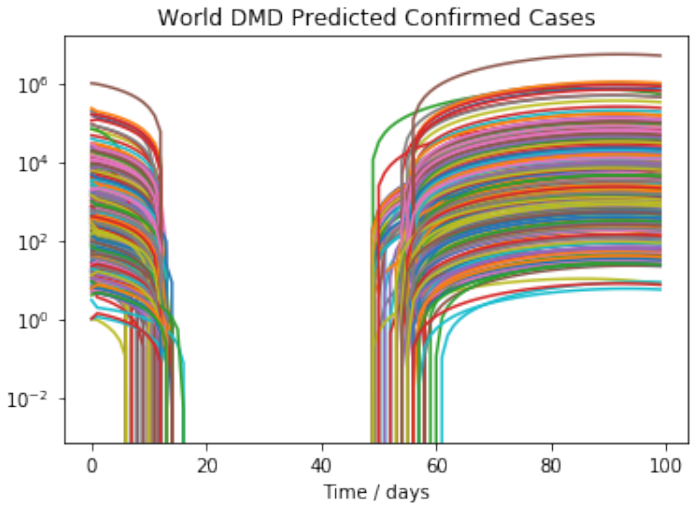

Figure 13: Predicted data for the world for 100 days after April 28th, 2020.

100 days after April 28th, 2020. Since there is one pair of eigenvalues with imaginary components, the predicted data exhibits oscillatory behavior.

3.3 Controlling the Spread of Coronavirus using DMDc

DMDc was used to determine the effects of control strategies on the number of coronavirus confirmed cases. The coefficients from Equations 9 and 10 were determined the be the following: **A** was determined from converting the reduced order $\tilde{\mathbf{A}}$ to continuous-time. **B** was determined to be a vector of ones because ideally all states can be actuated by the input. **C** was also chosen to be a vector of ones because all states contribute to the output. The system parameters were that the initial condition is the reduced-order data from the last available day and the time was for 50 days for the US and 365 days for the world. The continuous transfer functions were calculated to be Equation 12 for the US and Equation 13 for the world.

$$\frac{3s^2 - 0.247s + 0.0008641}{s^3 - 0.05268s^2 + 0.002976s + 1.925x10^{-}5} \tag{12}$$

$$\frac{3s^2 - 0.2128s - 0.00108}{s^3 - 0.03731s^2 + 0.003966s - 4.878x10^{-}5} \tag{13}$$

The reduced order model was plotted with no input in Figures 14 and 15. The addition of feedback with unity proportional control is shown in Figures 16 and 17. Feedback began at day 10 in the US and at day 100 in the world for the simulation. Lastly, to determine the optimal gains for the proportional controller, a lqr was used. The coefficient, R, which yielded the smoothest control was determined to be 1 for the US and 10^5 for the world. The Reduced order models with optimal control implemented are shown in Figures 18 and 19. The gain values found by the lqr and the resulting continuous eigenvalues are shown in Table 3. Physical meanings for the gain values can be drawn for optimal social distancing and vaccination practices.

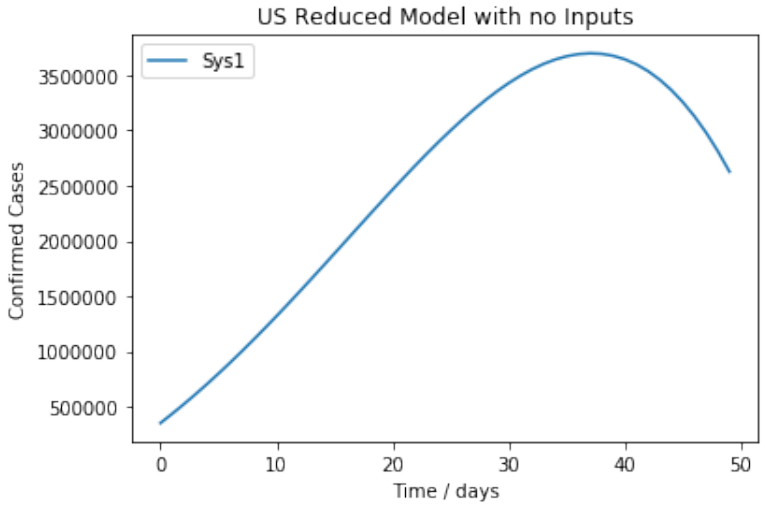

Figure 14: Reduced order continuous system with no inputs for the US.

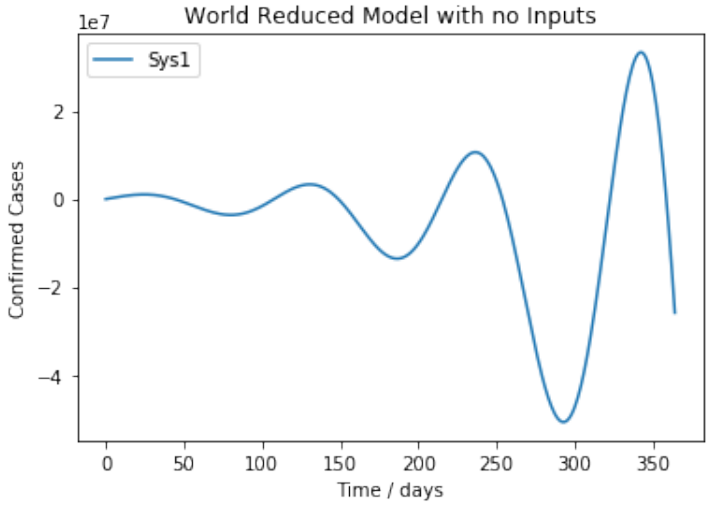

Figure 15: Reduced order continuous system with no inputs for the world.

	Gain Values	Continuous Eigenvalues
US	-4.985, -39.18, 49.82	-5.476, -0.06444+/-0.03151
World	0.03507, 0.1324, -0.08660	-0.01295, -0.01532+/-0.05978

Table 3: Gain values and eigenvalues for $\tilde{\mathbf{A}}$ with optimal control.

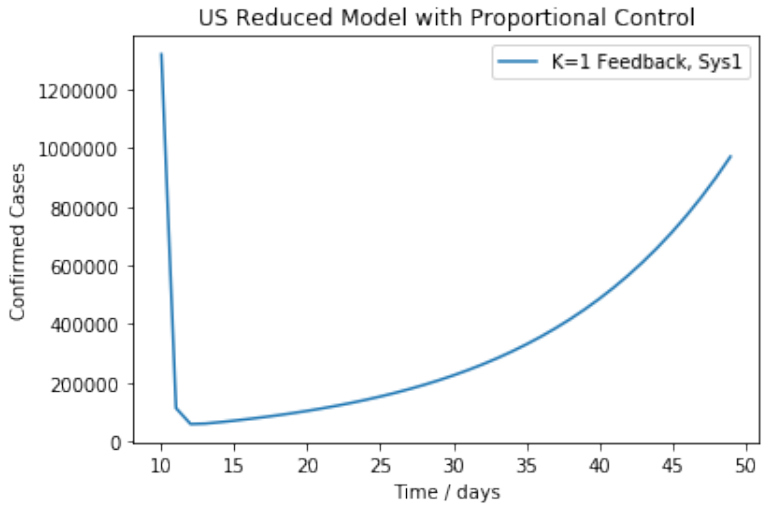

Figure 16: Reduced order continuous system with feedback and unity proportional control for the US.

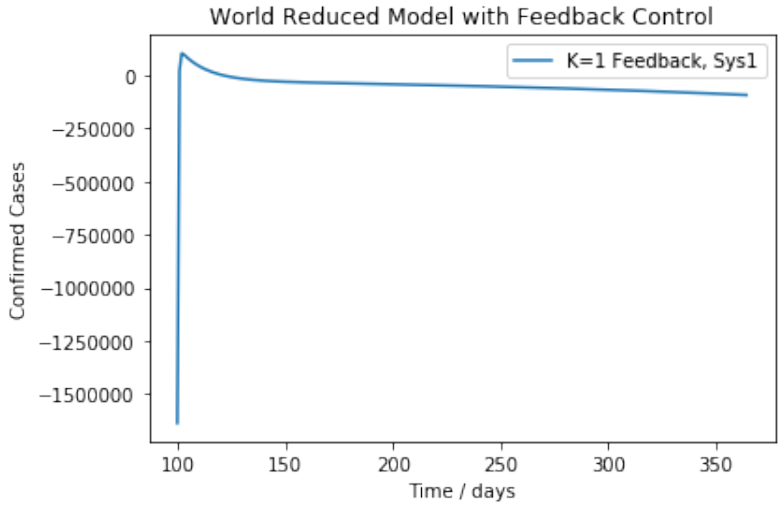

Figure 17: Reduced order continuous system with feedback and unity proportional control for the world.

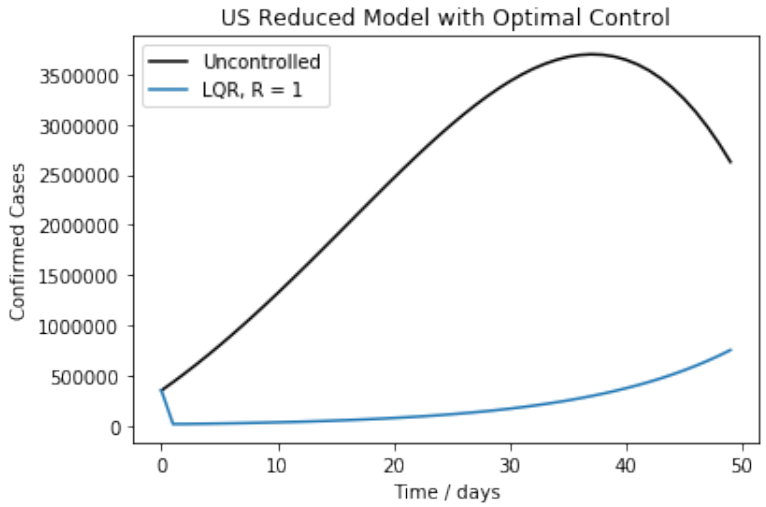

Figure 18: Reduced order continuous system with optimal control for the US.

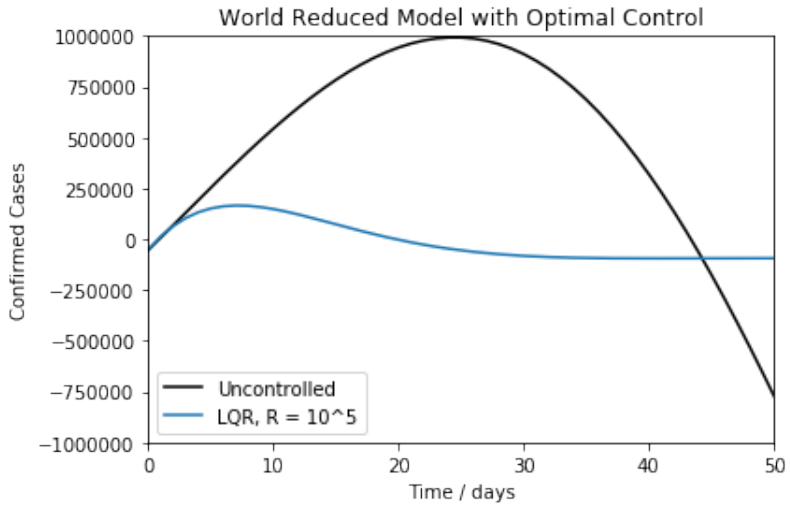

Figure 19: Reduced order continuous system with optimal control for the world.

4 Discussion of Results

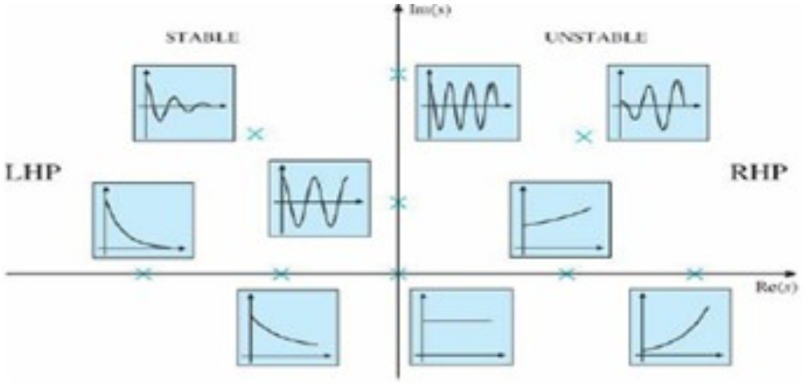

Figure 20: Functions in time domain from eigenvalue placement.

Using DMD to identify system dynamics and to predict future trends for coronavirus data proves a useful method for determining sufficient self-isolation methods. Using DMD with an SVD rank of three results in a very accurate reconstruction of confirmed cases trends (Figures 3 and 4). The eigenvalues for the linear operator of the US confirmed cases and the world are very close in value (Table 1). Both data sets have one continuous real eigenvalue on the right hand plane (RHP) and two continuous eigenvalues with imaginary components on the left hand plane (LHP). The general rules for determining the dynamics of a function based on its continuous roots are shown in Figure 20. From the DMD-calculated eigenvalues, the real eigenvalue on the RHP is very close to the origin, indicating a slow time response. The two eigenvalues with imaginary components indicate oscillation. Since the eigenvalues are on the RHP, the system will grow over time without bounds. This means that the system is unstable if no control methods are implemented. Compared to previous DMD-defined system poles of disease data, the coronavirus poles are close in value, but exist in the RHP [1]. This means that disease dynamics often have slow oscillatory frequencies. This makes sense because disease trends tend to follow logarithmic curves which oscillate when a certain portion of the population can no longer be infected or when

social distancing measures are enforced.

The modes and dynamics for the US and world confirmed cases data show dynamics which seem logarithmic and a small number of locations with modes with large magnitudes. The mode graph of the US shows NY as an obvious outlier in mode magnitude (Figure 9). Most states have very low magnitude modes, shown by dark blue on the map.

Use of DMD for predicting coronavirus confirmed cases is accurate for about a ten day time span (Table 2). Using world data provides about 5x more accurate predictions than the US data does. This makes sense because the is a larger distribution of states for each time step. The ability to use different locations at very different stages of infection for fitting the DMD model would provide more information for future predictions. The predictions for the next 100 days of confirmed cases shows a predicted oscillation (Figures 12 and 13). This oscillation does not seem likely because the number of confirmed cases cannot physically drop from 100,000 to 0 in a period of five days. This information is useful in determining that the current self-isolation mandates in place are causing a decrease in the rate of change of the number of confirmed coronavirus cases. Predictions for US data were also calculated without NY, the outlier, to test for high prediction accuracy. When NY was excluded, prediction accuracy did increase modestly, but the prediction oscillatory patterns remained the same.

Proportional feedback control was tested on the confirmed cases data set for the world and US to determine if an optimal control strategy could be developed to mitigate the number of confirmed coronavirus cases. A continuous control system was developed from using the equivalent continuous-time linear operator from the DMD, with an SVD rank of three. The low rank approximation was used to generalize system dynamics to a low-dimensional sub-space. The resulting transfer functions for the systems have poles and zeros which are close in value (Equations 12 and 13). This is expected, since the eigenvalues of the systems were shown to be close in value. Since the transfer functions were roughly the same, the US data set was used for determining a short-term control strategy and the world

data set was used for determining a long-term control strategy. Figures 14 and 15 show that both systems will grow without bound when there are no external inputs. When unity feedback control is introduced, it quickly stabilizes the system for the world but causes the system for the US to become unstable (Figures 16 and 17). Optimal proportional control provides an efficient means of flattening the curve of confirmed coronavirus cases (Figures 18 and 19). With an R value of 1, the controller works quickly to lower the number of confirmed cases quickly for the US. In contrast, when R was set to 10^5 the model grew at the same rate for the first few days, but the peak of the curve was reached much earlier and then stabilized at zero. From Table 3, the effects of changing the R values for the lqr are shown; the gain values for the US and world differ by three orders of magnitude. The gain values are opposite in sign most likely because the reduced order models at different phases of the system response. The continuous eigenvalues show that it is possible to use proportional control to move the eigenvalues to the LHP, obtaining stability. Furthermore, it is possible to significantly reduce overshoot with small actuations (social distancing measures or vaccinations), in response to the output.

While DMD and DMDc provided an effective means for categorizing the coronavirus system, further analysis of the system could be completed. Determining the full-state discrete controller for the system would garner additional information on the specific magnitudes of input needed to stabilize each region. In addition, further analysis on the physical implication of eigenvalue locations, mode magnitudes, and dynamics for real epidemiological systems would likely provide more insight into predictions and control measures developed. Lastly, research in DMD methods has led to new variations in DMD, including multi-resolution DMD, compressed DMD, higher order DMD, forward/backward DMD, and optimal closed-form DMD. Applying these methods to epidemiological problems have the potential to increase system modelling and prediction accuracy.

5 Conclusions

DMD is an effective method for identifying system dynamics and predicting future trends for time-dependent data sets. In applying this method to the confirmed coronavirus cases in the US and the world, future behaviors of the systems were predicted and optimal control strategies for disease containment were determined. Both the world and US systems were determined to be slightly unstable, with low frequency oscillatory behaviors. Therefore, a control strategy with small corrections in input magnitude, such as social distancing measures or distribution of vaccines in high-impact regions, would be effective in stabilizing the outbreak.

6 Literature Cited

References

[1] Joshua L. Proctor and Philip A. Eckhoff. *Discovering Dynamic Patterns from Infectious Disease Data using Dynamic Mode Decomposition.* International Health, 2015.

[2] Steven L. Brunton Joshua L. Proctor and Nathan Kutz. *Dynamic Mode Decomposition with Control.* SIAM Journal on Applied Dynamical Systems, 2016.

[3] *https://github.com/CSSEGISandData/.* John Hopkins University CSSE, 2020.

[4] J. Nathan Kutz Steven L. Brunton. *Data-Driven Science and Engineering.* Cambridge University Press.