DATA-DRIVEN OPTIMAL CONTROL OF
A SEIR MODEL FOR COVID-19

HAILIANG LIU AND XUPING TIAN

ABSTRACT. Since the first case of COVID-19 in December 2019, a total of 11,357,322
confirmed cases in the US and 55,624,562 confirmed cases worldwide have been re-
ported, up to November 17, 2020, evoking fear locally and internationally. In particular,
the coronavirus is surging nationwide at this time. In this paper, we investigate a basic
Susceptible-Exposed-Infectious-Recovered (SEIR) model and calibrate the model param-
eters to the reported data. We also attempt to forecast the evolution of the outbreak over
a relatively short time period and provide scheduled optimal controls of the COVID-19
epidemic. We provide efficient numerical algorithms based on a generalized Pontryagin
Maximum Principle associated with the optimal control theory. Numerical experiments
demonstrate the effective performance of the proposed model and its numerical approx-
imations.

1. Introduction

The outbreak of COVID-19 epidemic has resulted in over millions of confirmed and
death cases, evoking fear locally and internationally. It has a huge impact on global
economy as well as everyone’s daily life. Numerous mathematical models are being pro-
duced to forecast the future of COVID-19 epidemics in the US and worldwide. These
predictions have far-reaching consequences regarding how quickly and how strongly gov-
ernments move to curb the epidemic. Mathematical models can be profoundly helpful
tools to make public health decisions and ensure optimal use of resources to reduce the
morbidity and mortality associated with the COVID-19 pandemic, but only if they are
rigorously evaluated and validated and their projections are robust and reliable.

One of the well-known and basic models in epidemiology is the SIR model proposed by
Kermack and McKendrick \cite{kermack1927} in 1927. Here, $S, I, R$ represent the number of susceptible,
infected and recovered people respectively. They use an ODE system to describe the trans-
mission dynamics of infectious diseases among the population. In the current COVID-19
pandemic, actions such as travel restrictions, physical distancing and self-quarantine are
taken to slow down the spread of the epidemic. Typically, there is a significant incubation
period during which individuals have been infected but are not yet infectious themselves.
During this period the individual is in compartment $E$ (for exposed), the resulting models
are of SEIR or SEIRS type, respectively, depending on whether the acquired immunity
is permanent or otherwise. Also such models can show how different public health in-
terventions may affect the outcome of the epidemic, and can also predict future growth
patterns.

In this paper, we integrate the optimal control with a specific SEIR model, though
the developed methods can be readily adapted to other epidemic ODE models. More

2020 Mathematics Subject Classification. Primary 34H05, 92D30; Secondary 49M05, 49M25.
precisely, we introduce a dynamic control model for monitoring the virus propagation. Here the goal is to advance our understanding of virus propagation by learning the model parameters so that the error between the reported cases and the solution to the SEIR model is minimized. In short, we formulate the following optimization problem

\[
\min_{\theta} \ J = \sum_{i=1}^{n-1} L(U(t_i)) + g(U(T)),
\]

s.t. \( \dot{U} = F(U, \theta) \quad t \in (0, T], \ U(0) = U_0, \ \theta \in \Theta. \)

Here \( \dot{U} = \frac{d}{dt}U(t) \), \( \theta \) is a time-varying vector of model parameters, \( \Theta \) is the admissible set for model parameters \( \theta \), and \( U = [S, E, I, R, D]^\top \) corresponds to the susceptible, exposed, infected, recovered and deceased population. The loss function \( J \) is composed of two parts: \( L \) measures the error between the candidate solution to the SEIR system and the reported data at intermediate observation times, and \( g \) measures the error between the candidate solution and the scheduled control data at the end time. This is in contrast to the standard approach of fitting the model to the epidemic data through the non-linear least squares approach.

### 1.1. Main contributions

In the present work we provide a fairly complete modeling discussion on parameter learning, prediction and control of epidemics spread based on the SEIR model. We begin with our discussion of basic properties of the SEIR model. Then we show how the parameters can be updated recursively by gradient-based methods for a short time period while parameters are near constants, where the loss gradient can be obtained by using both state and co-state dynamics. For an extended time period with reported data available at intermediate observation times, the parameters are typically time-varying. In this general setting, we derive necessary conditions to achieve optimal control in terms of the chosen objectives. The conditions are essentially the classical Pontryagin maximum principal (PMP)[29]. The main differences are in the way we apply the principle in each time interval, and connect them consistently by re-setting the co-state \( V \) at the end of each time interval. We thus named it the generalized PMP. We further present an algorithm to find the numerical solution to the generalized PMP in the spirit of the method of successive approximations (MSA)[9]. Our algorithm is mainly fulfilled by three parts: (1) discretization of the forward problem in such a way that solutions to \( U \) remain positive for an arbitrarily step size \( h \); (2) discretization of the co-state equation for \( V \) is made unconditionally stable; (3) Minimization of the Hamiltonian is given explicitly based on the structure of the SEIR model.

This data-driven optimal control approach can be applied to other epidemic models. In particular, the prediction and control can be combined into one framework. To this end, the cost includes terms measuring the error between confirmed cases (infection and death) and those predicted from the model during the evolution, and terms measuring the error between the scheduled numbers as desired and those predicted from the model during the control period.

### 1.2. Further related work

In the mathematical study of SIR models there is an interplay between the dynamics of the disease and that of the total population, see [2, 27, 12, 33]. We refer the reader to [25, 11] for references on SEIRS models with constant total population and to [24] for the proof of the global stability of a unique
endemic equilibrium of a SEIR model. Global stability of the endemic equilibrium for the SEIR model with non-constant population is more subtle, see [23]. Epidemic models have been treated using optimal control theory, with major control measures on medicine (vaccination), see e.g., [31, 19, 22, 15, 21, 28, 3], and for Cancer immunotherapy control [7].

For COVID-19, a data-driven model has been proposed and simulated in [26], where both the SIR model and the feed-forward network are trained jointly. Compared to this work, our data-driven model is to consider the optimal control for SEIR with rigorous derivation of optimality conditions and stable numerical approximations. On the other hand, our data-driven optimal control algorithm may be interpreted as training a deep neural network in which the SEIR model serves as the neural network with parameters to be learned. This is in contrast to the study of residual neural networks using neural ODEs (see e.g., [8]). SIR models with spatial effects have also been studied [4, 20]. A SIRT model was proposed in [4] to study the effects of the presence of a road on the spatial propagation of the COVID-19. Introduced in [20] is a mean-field game algorithm for a spatial SIR model in controlling propagation of epidemics.

1.3. Organization. In Section 2 we motivate and present the SEIR system for modelling the time evolution of epidemics, and integrate it with data into an optimal control system. Both algorithms and related numerical approximations are detailed in Section 3. Section 4 provides experimental results to show the performance of our data-driven optimal control algorithms. We end with concluding remarks in Section 5.

2. The SEIR model and optimal control

2.1. Model formulation. A population of size $N(t)$ is partitioned into subclasses of individuals who are susceptible, exposed (infected but not yet infectious), infectious, and recovered, with sizes denoted by $S(t), E(t), I(t), R(t)$, respectively. The sum $E + I$ is the total infected population.

The dynamical transfer of the population is governed by an ODE system

$$
\dot{S} = A - \beta SI/N - dS,
\dot{E} = \beta SI/N - \epsilon E - dE,
\dot{I} = \epsilon E - \mu I - \gamma I - dI,
\dot{R} = \gamma I - dR,
\dot{D} = \mu I,
$$

(2.1)

subject to initial conditions $S_0, E_0, I_0, R_0, D_0$. Here $D(t)$ denotes deaths at time $t$, $A$ denotes the recruitment rate of the population due to birth and immigration. It is assumed that all newborns are susceptible and vertical transmission can be neglected. $d$ is the natural death rate, $\mu$ is the rate for virus-related death. $\gamma$ is the rate for recovery with $1/\gamma$ being the mean infectious period, and $\epsilon$ is the rate at which the exposed individuals become infectious with $1/\epsilon$ being the mean incubation period. The recovered individuals are assumed to acquire permanent immunity (yet to be further confirmed for COVID-19); there is no transfer from the $R$ class back to the $S$ class. $\beta$ is the effective contact rate. In the limit when $\epsilon \to \infty$, the SEIR model becomes a SIR model. Note that the involved
parameters do not correspond to the actual per day recovery and mortality rates as the new cases of recovered and deaths come from infected cases several days back in time. However, one can attempt to provide some coarse estimations of the “effective/apparent” values of these epidemiological parameters based on the reported confirmed cases.

2.2. Solution properties of the SEIR model. Classical disease transmission models typically have at most one endemic equilibrium. If there is no endemic equilibrium, diseases will disappear. Otherwise, the disease will be persistent irrespective of initial positions. Large outbreaks tend to the persistence of an endemic state and small outbreaks tend to the extinction of the diseases.

To understand solution properties of the SEIR system, we simply take $A = bN$ with $b$ the natural birth rate, and focus on the following sub-system:

\[
\begin{align*}
\dot{S} &= bN - \beta SI/N - dS, \\
\dot{E} &= \beta SI/N - \epsilon E - dE, \\
\dot{I} &= \epsilon E - \mu I - \gamma I - dI, \\
\dot{R} &= \gamma I - dR,
\end{align*}
\]

with the total population size $N = S + E + I + R$. By adding the equations above we obtain

\[
\dot{N} = (b - d)N - \mu I.
\]

Let $s = S/N, e = E/N, i = I/N, r = R/N$ denote the fractions of the classes $S, E, I, R$ in the population, respectively, then one can verify that

\[
\begin{align*}
\dot{s} &= b - bs - \beta is + \mu is, \\
\dot{e} &= \beta is - (\epsilon + b)e + \mu ie, \\
\dot{i} &= \epsilon e - (\mu + \gamma + b)i + \mu i^2, \\
\dot{r} &= \gamma i - br + \mu ir.
\end{align*}
\]

From biological considerations, we study system (2.2) in a feasible region

\[
\Sigma = \{(s, e, i) \in \mathbb{R}^3_+, 0 \leq s + e + i \leq 1\}.
\]

It can be verified that $\Sigma$ is positively invariant with respect to the underlying dynamic system. We denote $\partial \Sigma$ and $\Sigma^0$ the boundary and the interior of $\Sigma$ in $\mathbb{R}^3$, respectively. A special solution of form $P^0 = (1, 0, 0)$ on the boundary of $\Sigma$ is the disease-free equilibrium of system (2.2) and it exists for all non-negative values of its parameters. Any equilibrium in the interior $\Sigma^0$ of $\Sigma$ corresponds to the disease being endemic and is named an endemic equilibrium. To identify the critical threshold for the parameters, we take $\epsilon \times (2.2a) + (\epsilon + b) \times (2.2c)$ to obtain

\[
\frac{d}{dt} [\epsilon e + (\epsilon + b)i] = i [\epsilon \beta s + \epsilon \mu e + \mu (\epsilon + b)i - (\epsilon + b)(\gamma + \mu + b)] \\
\leq i [\max\{\epsilon \beta, \epsilon \mu, \mu (\epsilon + b)\} - (\epsilon + b)(\gamma + \mu + b)] \leq 0,
\]

where $r$ can be obtained from $r = 1 - s - e - i$ or $\dot{r} = \gamma i - br + \mu ir$. 

2.2. Solution properties of the SEIR model. Classical disease transmission models typically have at most one endemic equilibrium. If there is no endemic equilibrium, diseases will disappear. Otherwise, the disease will be persistent irrespective of initial positions. Large outbreaks tend to the persistence of an endemic state and small outbreaks tend to the extinction of the diseases.

To understand solution properties of the SEIR system, we simply take $A = bN$ with $b$ the natural birth rate, and focus on the following sub-system:

\[
\begin{align*}
\dot{S} &= bN - \beta SI/N - dS, \\
\dot{E} &= \beta SI/N - \epsilon E - dE, \\
\dot{I} &= \epsilon E - \mu I - \gamma I - dI, \\
\dot{R} &= \gamma I - dR,
\end{align*}
\]

with the total population size $N = S + E + I + R$. By adding the equations above we obtain

\[
\dot{N} = (b - d)N - \mu I.
\]

Let $s = S/N, e = E/N, i = I/N, r = R/N$ denote the fractions of the classes $S, E, I, R$ in the population, respectively, then one can verify that

\[
\begin{align*}
\dot{s} &= b - bs - \beta is + \mu is, \\
\dot{e} &= \beta is - (\epsilon + b)e + \mu ie, \\
\dot{i} &= \epsilon e - (\mu + \gamma + b)i + \mu i^2, \\
\dot{r} &= \gamma i - br + \mu ir.
\end{align*}
\]

From biological considerations, we study system (2.2) in a feasible region

\[
\Sigma = \{(s, e, i) \in \mathbb{R}^3_+, 0 \leq s + e + i \leq 1\}.
\]

It can be verified that $\Sigma$ is positively invariant with respect to the underlying dynamic system. We denote $\partial \Sigma$ and $\Sigma^0$ the boundary and the interior of $\Sigma$ in $\mathbb{R}^3$, respectively. A special solution of form $P^0 = (1, 0, 0)$ on the boundary of $\Sigma$ is the disease-free equilibrium of system (2.2) and it exists for all non-negative values of its parameters. Any equilibrium in the interior $\Sigma^0$ of $\Sigma$ corresponds to the disease being endemic and is named an endemic equilibrium. To identify the critical threshold for the parameters, we take $\epsilon \times (2.2a) + (\epsilon + b) \times (2.2c)$ to obtain

\[
\frac{d}{dt} [\epsilon e + (\epsilon + b)i] = i [\epsilon \beta s + \epsilon \mu e + \mu (\epsilon + b)i - (\epsilon + b)(\gamma + \mu + b)] \\
\leq i [\max\{\epsilon \beta, \epsilon \mu, \mu (\epsilon + b)\} - (\epsilon + b)(\gamma + \mu + b)] \leq 0,
\]

where $r$ can be obtained from $r = 1 - s - e - i$ or $\dot{r} = \gamma i - br + \mu ir$. 

2.2. Solution properties of the SEIR model. Classical disease transmission models typically have at most one endemic equilibrium. If there is no endemic equilibrium, diseases will disappear. Otherwise, the disease will be persistent irrespective of initial positions. Large outbreaks tend to the persistence of an endemic state and small outbreaks tend to the extinction of the diseases.

To understand solution properties of the SEIR system, we simply take $A = bN$ with $b$ the natural birth rate, and focus on the following sub-system:

\[
\begin{align*}
\dot{S} &= bN - \beta SI/N - dS, \\
\dot{E} &= \beta SI/N - \epsilon E - dE, \\
\dot{I} &= \epsilon E - \mu I - \gamma I - dI, \\
\dot{R} &= \gamma I - dR,
\end{align*}
\]

with the total population size $N = S + E + I + R$. By adding the equations above we obtain

\[
\dot{N} = (b - d)N - \mu I.
\]

Let $s = S/N, e = E/N, i = I/N, r = R/N$ denote the fractions of the classes $S, E, I, R$ in the population, respectively, then one can verify that

\[
\begin{align*}
\dot{s} &= b - bs - \beta is + \mu is, \\
\dot{e} &= \beta is - (\epsilon + b)e + \mu ie, \\
\dot{i} &= \epsilon e - (\mu + \gamma + b)i + \mu i^2, \\
\dot{r} &= \gamma i - br + \mu ir.
\end{align*}
\]

From biological considerations, we study system (2.2) in a feasible region

\[
\Sigma = \{(s, e, i) \in \mathbb{R}^3_+, 0 \leq s + e + i \leq 1\}.
\]

It can be verified that $\Sigma$ is positively invariant with respect to the underlying dynamic system. We denote $\partial \Sigma$ and $\Sigma^0$ the boundary and the interior of $\Sigma$ in $\mathbb{R}^3$, respectively. A special solution of form $P^0 = (1, 0, 0)$ on the boundary of $\Sigma$ is the disease-free equilibrium of system (2.2) and it exists for all non-negative values of its parameters. Any equilibrium in the interior $\Sigma^0$ of $\Sigma$ corresponds to the disease being endemic and is named an endemic equilibrium. To identify the critical threshold for the parameters, we take $\epsilon \times (2.2a) + (\epsilon + b) \times (2.2c)$ to obtain

\[
\frac{d}{dt} [\epsilon e + (\epsilon + b)i] = i [\epsilon \beta s + \epsilon \mu e + \mu (\epsilon + b)i - (\epsilon + b)(\gamma + \mu + b)] \\
\leq i [\max\{\epsilon \beta, \epsilon \mu, \mu (\epsilon + b)\} - (\epsilon + b)(\gamma + \mu + b)] \leq 0,
\]
provided a key quantity $\sigma \leq 1$, where

$$\sigma := \frac{\beta \epsilon}{(\epsilon + b)(\gamma + \mu + b)}$$

is called the modified contact number [23]. The global stability of $P_0$ when $\sigma \leq 1$ follows from LaSalle’s invariance principle. The following theorem (see [23]), a standard type in mathematical epidemiology, shows that the basic number $\sigma$ determines the long-term outcome of the epidemic outbreak.

**Theorem 2.1.** 1. If $\sigma \leq 1$, then $P_0$ is the unique equilibrium and globally stable in $\Sigma$. 2. If $\sigma > 1$, then $P_0$ is unstable and the system is uniformly persistent in $\Sigma^0$.

By this theorem, the disease-free equilibrium $P_0$ is globally stable in $\Sigma$ if and only if $\sigma \leq 1$. Its epidemiological implication is that the infected fraction of the population vanishes in time so the disease dies out. In addition, the number $\sigma$ serves as a sharp threshold parameter; if $\sigma > 1$, the disease remains endemic.

In the epidemic literature, another threshold quantity is the basic reproduction number

$$R_0 = \frac{\beta}{\gamma + \mu},$$

which is the product of the contact rate $\beta$ and the average infectious period $\frac{1}{\gamma + \mu}$. It is a parameter well known for quantifying the epidemic spread [10, 35]. On the other hand, the contact number $\sigma$ is defined at all times. In general, we have

$$R_0 \geq \sigma.$$  

Note that for most models, $\sigma = R_0$, both quantities can be used interchangeably. This is the case in our experiments when taking $b = 0$.

**Remark 2.2.** Extensive evidence shows that the disease spread rate is sensitive to at least three factors: (1) daily interactions, (2) probability of infection, and (3) duration of illness. The above assertion shows that making efforts to decrease $R_0$ is essential for controlling propagation of epidemics. Hence measures such as social distancing (self-quarantine, physical separation), washing hands and wearing face coverings, as well as testing / timely-hospitalization can collectively decrease $R_0$.

**Remark 2.3.** The above model if further simplified will reduce to the SIR model or SIS model, which is easier to analyze [13, 5, 34]. It can also be enriched by dividing into different groups [18, 32], or by considering the spatial movement effects [14, 16]. In this work, we use the SEIR model as a base for our analysis, prediction, and control.

### 2.3. A simple control.

In compact form, the SEIR system may be written as

$$\dot{U} = F(U; \theta), \quad U(0) = U_0,$$

where

$$U = [S, E, I, R, D]^\top, \quad \theta = [\beta, \epsilon, \gamma, \mu]^\top$$

are the column vectors of the state and parameters, respectively. Let $g(\cdot)$ be a loss function (to be specified later) at the final time $T$, then the problem of determining the model parameters based on this final cost can be cast as a dynamic control problem:

$$\min_{\theta} \{g(U(T)) \mid \text{subject to } \dot{U} = F(U, \theta), \quad 0 < t \leq T; \quad U(0) = U_0, \quad \theta \in \Theta\}.$$
Here the dependence of $g$ on $\theta$ is through $U(T)$. The set of admissible parameters $\Theta$ may be estimated from other sources.

For a short time period, the model parameters are near constant, then we can use gradient-based methods to directly update $\theta$, for which $\nabla_\theta g$ needs to be evaluated. We define the Lagrangian functional
\[
\mathcal{L}(\theta) = g(U(T)) - \int_0^T (\dot{U} - F(U, \theta))^\top V dt,
\]
where $V = V(t)$ is the Lagrangian multiplier depending on time and can be chosen freely, and $U$ depends on $\theta$ through the ODE. A formal calculation gives
\[
\nabla_\theta g = \nabla_\theta \mathcal{L}
\]
\[
= (\nabla_\theta U(T))^\top \nabla_U g + \int_0^T \left((\nabla_U F)(\nabla_\theta U(t)) + \nabla_\theta F - \nabla_\theta \dot{U}(t)\right)^\top V(t) dt
\]
\[
= (\nabla_\theta U(T))^\top \left((\nabla_U g - V(T)) + (\nabla_\theta U(0))^\top V(0)\right)
\]
\[
+ \int_0^T (\nabla_\theta U(t))^\top \left((\nabla_U F)^\top V(t) + \dot{V}(t)\right) + (\nabla_\theta F)^\top V(t) dt.
\]
Thus $\nabla_\theta g$ can be determined in the following steps:

- Solve the forward problem for state $U$,
  \[
  \dot{U}(t) = F(U(t), \theta), \quad U(0) = U_0.
  \]
- Solve the backward problem for co-state $V$,
  \[
  \dot{V}(t) = -((\nabla_U F)^\top V(t), \quad V(T) = \nabla_U g(U(T)).
  \]
- Evaluate the gradient of $g$ by
  \[
  \nabla_\theta g = \int_0^T (\nabla_\theta F)^\top V(t) dt.
  \]

In the context of the optimal control theory [29], such $V$ exists and is called the co-state function. We note that the $(U, V)$ dynamical system models the optimal strategies for $S, E, I, R$ populations.

Remark 2.4. It is remarkable that computing gradients of a scalar-valued loss with respect to all parameters is done using only state and co-state functions, without backpropagating through the operations of the solver. In addition, modern ODE solvers may be used to compute both $U$ and $V$.

2.4. Data-driven optimal control. Now we consider an extended time period with reported data available at intermediate observation times. The data are taken from the reported cumulative infection and death cases. In such general case the parameters are typically time-varying, we need to derive a more refined data-driven optimal control. Arranging the data in a vector $U_c = [I_c, D_c]^\top$ at times $0 = t_0 < t_1 < t_2 < ... < t_n = T$, we aim to

*Data used here is publicly available in the CSSEGISandData/COVID-19 GitHub repository, which collects data from official sources and organizations.*
• find optimal parameter $\theta(t)$ for $0 \leq t \leq t_{n-1}$ such that the solution to the SEIR system fits the reported data at the grid points $\{t_i\}_{i=1}^{n-1}$ as close as possible, and
• find desired parameter $\theta(t)$ for $t_{n-1} \leq t \leq T$ that is able to control the epidemic spreading at time $T$ at desired values.

To achieve these goals, we first define a loss function by

$$J = \sum_{i=1}^{n-1} L(U(t_i)) + g(U(T)),$$

where

$$L(U(t_i)) = \lambda_1 |I(t_i) - I_c(t_i)|^2 + \lambda_2 |D(t_i) - D_c(t_i)|^2, \quad 1 \leq i \leq n - 1,$$

$$g(U(T)) = \lambda_1 |I(T) - I_d(T)|^2 + \lambda_2 |D(T) - D_d(T)|^2.$$

Here $\lambda_1, \lambda_2$ are user-defined normalization factors. This loss function is composed of two parts: the running cost $L$ which measures the error between the candidate solution $(I, D)$ to the SEIR system and the reported data $(I_c, D_c)$ at intermediate times; and the final cost $g$, which measures the error between the candidate solution $(I, D)$ and the desired data $(I_d, D_d)$ at the end time.

We then formulate the task as the following optimal control problem

$$\min J = \sum_{i=1}^{n-1} L(U(t_i)) + g(U(T)),$$

s.t. $U = F(U, \theta) \quad t \in (0, T], \quad U(0) = U_0, \quad \theta \in \Theta$.

Motivated by the classical optimal control theory, we derive the necessary conditions for the optimal solution to problem (2.4), stated in the following theorem.

**Theorem 2.5.** Let $\theta^*$ be the optimal solution to problem (2.4) and $U^*$ be the corresponding state function, then there exists a piece-wise smooth function $V^*$ and a Hamiltonian $H$ defined by

$$H(U, V, \theta) = V^\top F(U, \theta)$$

such that

$$\dot{U}^* = F(U^*, \theta^*), \quad t \in (0, T], \quad U^*(0) = U_0,$$

$$\dot{V}^* = -(\nabla_U F(U^*, \theta^*))^\top V, \quad t_{i-1} \leq t < t_i, \quad i = n, ..., 1,$$

$$V^*(T) = \nabla_U g(U^*(T)),$$

$$V^*(t^-_i) = V^*(t^+_i) + \nabla_U L(U^*(t_i)), \quad i = n - 1, ..., 1$$

are satisfied. Moreover, the Hamiltonian minimization condition

$$H(U^*, V^*, \theta^*) \leq H(U^*, V^*, a) \quad \forall a \in \Theta$$

holds for all time $t \in [0, T]$ but $\{t_i\}_{i=1}^{n-1}$.

**Proof.** Recall the classical Bolza optimal control problem

$$\min J = \int_0^T L(U(t), \theta(t)) dt + g(U(T))$$
s.t. \( \dot{U} = F(U, \theta), \quad t \in [0, T], \quad U(0) = U_0, \quad \theta(t) \in \Theta. \)

The Pontryagin’s maximum principle states the necessary conditions for optimality: assume \( \theta^* \) and \( U^* \) are the optimal control function and corresponding state of the system, then there exists a function \( V^* \) and a Hamiltonian \( H \) defined for all \( t \in [0, T] \) by

\[
H(U, V, \theta) = V^T F(U, \theta) + L(U, \theta)
\]

such that

\[
\dot{U}^* = \nabla_V H(U^*, V^*, \theta^*), \quad U^*(0) = U_0,
\]

\[
\dot{V}^* = -\nabla_U H(U^*, V^*, \theta^*), \quad V^*(T) = \nabla_U g(U^*(T)),
\]

are satisfied. Moreover, the Hamiltonian minimization condition

\[
H(U^*, V^*, \theta^*) \leq H(U^*, V^*, a)
\]

holds for all time \( t \in [0, T] \) and for all permissible control function \( a \in \Theta. \)

Notice that the loss function in (2.4) can be rewritten as

\[
J = \int_0^T \sum_{i=1}^{n-1} L(U(t))\delta(t - t_i)dt + g(U(T)),
\]

where \( \delta(x) \) is the Dirac delta function, then the corresponding Hamiltonian reads as

\[
H(U, V, \theta) = V^T F(U, \theta) + \sum_{i=1}^{n-1} L(U(t))\delta(t - t_i).
\]

Hence (2.8) can be expressed as (2.6), (2.9) has the form of

\[
\dot{V} = -\sum_{i=1}^{n-1} \nabla_U L(U(t))\delta(t - t_i) - (\nabla_U F(U, \theta))^\top V, \quad V(T) = \nabla_U g(U(T)).
\]

Integrate (2.11) from \( t_i^- - \epsilon \) to \( t_i^+ + \epsilon \) and let \( \epsilon \) goes to 0, then for \( i = n - 1, \ldots, 1 \)

\[
V(t_i^+) - V(t_i^-) = \nabla_U L(U(t_i)).
\]

Thus in each interval, (2.11) reduces to

\[
\dot{V} = -(\nabla_U F(U, \theta))^\top V, \quad t_{i-1} \leq t < t_i, \quad i = n, \ldots, 1,
\]

which is exactly (2.7). Correspondingly, the Hamiltonian (2.10) reduces to (2.5) for all \( t \in [0, T] \) but \( \{t_i\}_{i=1}^{n-1} \).

3. Numerical discretization and algorithms

In this section, we present implementation details to solve problem (2.4) via solving the generalized PMP by iteration. We proceed in the following manner. First we make an initial guess \( \theta_0 \in \Theta. \) From the control function \( \theta_l(t) \) in the \( l \)-th iteration for \( l = 0, 1, 2, \ldots \), we obtain \( \theta_{l+1}(t) \) in three steps:

**Step 1:** Solve the forward problem (2.6) to obtain \( U_l. \)

**Step 2:** Solve the sequence of backward problems (2.7) to obtain \( V_l. \)

**Step 3:** \( \theta_{l+1}(t) = \arg\min_{\theta \in \Theta} H(U_l, V_l, \theta, t) \) for each \( t \in [0, T]. \)
This is essentially the method of successive approximations (MSA) \[9\]. An important feature of MSA is that the Hamiltonian minimization step is decoupled for each \(t\). However, in general MSA tends to diverge, especially if the bad initial guess is taken \[9\].

Our goal is to adopt a careful discretization at each step to control possible divergent behavior, instead of simply call an existing ODE solver for Steps 1-2, and an optimization algorithm for Step 3.

Within each interval \((t_{i-1}, t_i)\), we approximate both \(U\) and \(V\) at the same \(m\) points with step size \(h = (t_i - t_{i-1})/m\), so that the value of \(U\) required in Step 2 are already calculated in Step 1. We use \(U^{i,k}, V^{i,k}, \theta^{i,k}\), where \(i = 1, ..., n\) and \(k = 0, ..., m\) to denote solution values at \(k\)-th point in the \(i\)-th interval.

Note that population dynamics (births or deaths) may be neglected at a very crude level on the grounds that epidemic dynamics often occur on a faster time scale than host demography, or we can say heuristically that death of an infected individual and subsequent replacement by a susceptible (in the absence of vertical transmission) is equivalent to a recovery event. Hence, in the numerical study, both \(b\) and \(d\) are taken to be zero.

Below we discuss the discretization of the three iteration steps for the case \(b = d = 0\).

### 3.1. Forward discretization

We focus on the time interval \((t_{i-1}, t_i)\). For notation simplicity, we use \(S^k, E^k, I^k\) to represent corresponding values at the \(k\)-th point. We discretize the forward problem (2.6) by an explicit-implicit method with the Gauss-Seidel type update:

\[
\begin{align*}
\frac{S^{k+1} - S^k}{h} &= -\beta^k S^{k+1} I^k / N^k, \\
\frac{E^{k+1} - E^k}{h} &= \beta^k S^{k+1} I^k / N^k - \epsilon^k E^{k+1}, \\
\frac{I^{k+1} - I^k}{h} &= \epsilon^k E^{k+1} - (\gamma^k + \mu^k) I^{k+1}, \\
\frac{R^{k+1} - R^k}{h} &= \gamma^k I^{k+1}, \\
\frac{D^{k+1} - D^k}{h} &= \mu^k I^{k+1}.
\end{align*}
\]

This gives

\[
\begin{align*}
S^{k+1} &= \frac{S^k}{1 + h\beta^k I^k / N^k}, \\
E^{k+1} &= \frac{E^k + h\beta^k S^k I^k / N^k}{1 + h\epsilon^k}, \\
I^{k+1} &= \frac{I^k + h\epsilon^k E^{k+1}}{1 + h(\gamma^k + \mu^k)}, \\
R^{k+1} &= R^k + h\gamma^k I^{k+1}, \\
D^{k+1} &= D^k + h\mu^k I^{k+1},
\end{align*}
\]
where \( k = 0, 1, ..., m-1 \), \( h = (t_i - t_{i-1})/m \) is the step size. The most important property of this update is its unconditional positivity-preserving property, i.e.,
\[
U^k \geq 0 \Rightarrow U^{k+1} > 0, \quad k = 0, \cdots
\]
irrespective of the size of the step size \( h \). For the starting value in each interval, we have
\[
U^{1,0} = U(0), \quad U^{i,0} = U^{i-1,m}, \quad i = 2, \cdots, n-1.
\]

3.2. **Backward discretization.** Let \( V = [V_S, V_E, V_I, V_R, V_D]^T \) and notice that
\[
\nabla_U F = \begin{bmatrix}
-\beta I(N - S)/N^2 & 0 & -\beta S(N - I)/N^2 & 0 & 0 \\
\beta I(N - S)/N^2 & -\epsilon & \beta S(N - I)/N^2 & 0 & 0 \\
0 & 0 & -\gamma & 0 & 0 \\
0 & 0 & 0 & \mu & 0 \\
\end{bmatrix},
\]
then (2.7) takes the following form
\[
\begin{aligned}
\dot{V}_S &= \beta I(N - S)(V_S - V_E)/N^2, \\
\dot{V}_E &= \epsilon(V_E - V_I), \\
\dot{V}_I &= \beta S(N - I)(V_S - V_E)/N^2 + \gamma(V_I - V_E) + \mu(V_I - V_D), \\
\dot{V}_R &= \dot{V}_D = 0.
\end{aligned}
\]
In a similar fashion to the discretization of \( U \), we discretize this system by an explicit-implicit method by
\[
\begin{aligned}
\frac{V_{S}^{k+1} - V_{S}^{k}}{h} &= \beta k I^k (N^k - S^k)(V_S^k - V_{E}^{k+1})/(N^k)^2, \\
\frac{V_{E}^{k+1} - V_{E}^{k}}{h} &= \epsilon^k (V_E^k - V_{I}^{k+1}), \\
\frac{V_{I}^{k+1} - V_{I}^{k}}{h} &= \beta k S^k (N^k - I^k)(V_S^k - V_{E}^{k})/(N^k)^2 + \gamma^k (V_I^k - V_{E}^{k+1}) + \mu^k (V_I^k - V_{D}^{k+1}), \\
V_{R}^{k} &= V_{R}^{k+1}, \quad V_{D}^{k} = V_{D}^{k+1}.
\end{aligned}
\]
Then the scheme is given by
\[
\begin{aligned}
V_{S}^{k} &= \frac{V_{S}^{k+1}(N^k)^2 + h\beta k I^k (N^k - S^k) V_{E}^{k+1}}{(N^k)^2 + h\beta k I^k (N^k - S^k)}, \\
V_{E}^{k} &= \frac{V_{E}^{k+1} + h\epsilon^k V_{I}^{k+1}}{1 + h\epsilon^k}, \\
V_{I}^{k} &= \frac{V_{I}^{k+1} + h(\gamma^k V_R + \mu^k V_D - \beta^k S^k (N^k - I^k)(V_S^k - V_{E}^{k})/(N^k)^2)}{1 + h(\gamma^k + \mu^k)}, \\
V_{R}^{k} &= V_{R}^{k+1}, \quad V_{D}^{k} = V_{D}^{k+1},
\end{aligned}
\]
where \( k = m - 1, ..., 0 \), \( h = (t_i - t_{i-1})/m \) is the step size. For the starting value in each interval, we have
\[
V^{n,m} = \nabla_U g(U(T)), \quad V^{i,m} = V^{i+1,0} + \nabla_U L(U(t_i)), \quad i = n - 1, \cdots, 1.
\]
Note that the above discretization is well-defined for any $h > 0$, hence unconditionally stable since the system is linear in $V$.

3.3. Hamiltonian minimization. The Hamiltonian (2.5) is given by

$$H(U, V, \theta) = -V_S \beta SI/N + V_E(\beta SI/N - \epsilon E) + V_I(\epsilon E - (\gamma + \mu)I) + V_R\gamma I + V_D\mu I.$$  

After plugging in $U(t), V(t)$ that is obtained by solving the above forward and backward problems with given $\theta(t)$, $H$ may be seen as a functional only with respect to $\theta$. We solve it by the proximal point algorithm (PPA) [30], that is for any fixed time $t \in (t_{i-1}, t_i)$, $i = 1, 2, ..., n$,

$$\theta_{i+1}(t) = \arg\min_{\theta(t) \in \Theta} \left\{ H(U_i(t), V_i(t), \theta(t)) + \frac{1}{2\tau} \|\theta(t) - \theta_i(t)\|^2 \right\},$$

where $l$ is the index for iteration, $\tau$ is the step size.

**Remark 3.1.** The use of PPA brings two benefits: (1) the objective function in (3.3) is a convex function, which ensures the existence of the solution to the minimization problem; (2) it is numerically stable: PPA has the advantage of being monotonically decreasing, which is guaranteed for any step size $\tau > 0$. In this way, the convergence of minimizing the original loss in (2.4) with a regularization term is ensured, hence address the convergence issue that MSA generally has [1].

Given discretized $U^\alpha$ and $V^\alpha$ with $\alpha = \{i, k\}$, we solve (3.3) on grid points, that is to solve

$$\theta_{i+1}^\alpha = \arg\min_{\theta^\alpha} \left\{ H(U_i^\alpha, V_i^\alpha, \theta^\alpha) + \frac{1}{2\tau} \|\theta - \theta_i^\alpha\|^2 \right\}.$$

Since $H$ is smooth, the above formulation when the constraint is not imposed is equivalent to the following

$$\theta_{i+1}^\alpha = \theta_i^\alpha - \tau \nabla_{\theta} H(U_i^\alpha, V_i^\alpha, \theta_i^\alpha).$$

The special form of $H$ allows us to obtain a closed form solution:

$$\begin{align*}
\beta_{i+1}^\alpha &= \beta_i^\alpha + \tau S_i^\alpha I_i^\alpha ((V_S)_i^\alpha - (V_E)_i^\alpha)/N_i^\alpha, \\
\epsilon_{i+1}^\alpha &= \epsilon_i^\alpha + \tau E_i^\alpha ((V_E)_i^\alpha - (V_I)_i^\alpha), \\
\gamma_{i+1}^\alpha &= \gamma_i^\alpha + \tau I_i^\alpha ((V_I)_i^\alpha - (V_R)_i^\alpha), \\
\mu_{i+1}^\alpha &= \mu_i^\alpha + \tau I_i^\alpha ((V_I)_i^\alpha - (V_D)_i^\alpha).
\end{align*}$$

(3.5)

Now taking the constraint $\theta \in \Theta$ into consideration, we simply project the solution (3.5) back into the feasible region after each iteration, that is

$$\theta_{i+1}^\alpha = \text{clip}(\theta_{i+1}^\alpha, \Theta, \bar{\Theta})$$

where $\Theta, \bar{\Theta}$ are element-wise lower bound and upper bound of $\Theta$, respectively. That guarantees the output is constrained to be in $\Theta$.

**Remark 3.2.** The step size for each parameter may be set differently according to their magnitude scale, and this can indeed improve the training performance, as observed in our experiments.
3.4. Algorithms. The above computing process is wrapped up in Algorithm 1. Since we can update \( \{ \theta^{i,k} \} \) simultaneously, we consider \( \{ \theta^{i,k} \} \) as a vector and still denote it as \( \theta \) for notation simplicity. We use \( \| \cdot \| \) to stand for the \( L_2 \) norm.

Algorithm 1

Require: \( \{ U_c(t_i) \}_{i=1}^n \): data, \( t_n = T \): final time, \( U_0 \): initial data at 0, \( \theta_0 \): initial guess, \( \tau \): step size for the minimization problem

1: while \( \| \theta^l - \theta^{l-1} \| / \| \theta^{l-1} \| > \text{Tol} \) do
2: \hspace{1em} for \( i = 1 \) to \( n \) do
3: \hspace{2em} for \( k = 0 \) to \( m - 1 \) do
4: \hspace{3em} \( U^{i,k+1} \leftarrow U^{i,k} \) (solve the forward problem, refer to (3.1))
5: \hspace{3em} \( U^{i,0} \leftarrow U^{i-1,m} \) (update the initial condition for ODE solver)
6: \hspace{2em} \( V^{n,m} \leftarrow \partial U g(U(T)) \) (set the initial data for the backward problem)
7: \hspace{1em} for \( i = n \) to 1 do
8: \hspace{2em} for \( k = m - 1 \) to 0 do
9: \hspace{3em} \( V^{i,k} \leftarrow V^{i+1,0} + \partial U L(U(t_i)) \) (update the initial condition for ODE solver)
10: \hspace{2em} \( \theta^{l+1} \leftarrow \theta^l \) (solve the minimization problem, refer to (3.5))
11: \hspace{1em} \( \theta^{l+1} \leftarrow \text{clip}(\theta^{l+1}, \Theta, \Theta) \) (ensure \( \theta \in \Theta \))
12: return \( \theta, U \)

This data-driven optimal control algorithm provides an optimal fitting to both the data and the SEIR model, it can help to reduce the uncertainty in conventional model predictions with standard data fitting.

Initialization: We now discuss how to initialize the control function \( \theta \). A simple choice is to take \( \theta^{0,k}_i \) to be a constant vector for all \( 1 \leq i \leq n, 1 \leq k \leq m \), where the value of each component relies on a priori epidemiological and clinical information about the relative parameter magnitude. They vary with the area from where the data were sampled.

We can also use the data \( \{ D_c(t_i) \}_{i=0}^n, \{ I_c(t_i) \}_{i=0}^n \) to obtain a better initial guess for \( \mu \). More precisely, from (2.1), i.e., \( \mu = D/I \) we take

\[
\mu^{i,k}_0 = \frac{D_c(t_{i+1}) - D_c(t_i)}{(t_{i+1} - t_i)I_c(t_{i+1})} \quad \forall k = 0, 1, \ldots, m.
\]

The initial data for the forward problem is set as

\[
U_0 = [N(0) - I_c(0), 0, I_c(0), 0, D_c(0)]^T,
\]

where \( N(0) \) is the initial population of the area analyzed, \( I_c(0), D_c(0) \) are the confirmed infections and deaths on the day of the first confirmed cases, respectively.

The initial condition of the backward problem is given in (2.7). In the present setup, the loss function \( g \) only depends on \( I \) and \( D \), hence the initial condition is given by

\[
V^{n,m} = [0, 0, \partial_I g(U(T)), 0, \partial_D g(U(T))]^T.
\]

Finally, we should point out that Algorithm 1 with a rough initial guess \( \theta_0 \) can be rather inefficient when \( T \) is large. In such case, we divide the whole interval as \( 0 = t_{n_0} < \)}
\[ t_{n_1} < ... < t_{n_s} < t_{n_{s+1}} = T \text{ with } t_{n_j} \in \{t_i\}_{i=1}^{n-1} \text{ for } 1 \leq j \leq s \text{ and apply Algorithm 1 to each subinterval consecutively. This treatment is summarized in Algorithm 2.} \]

Algorithm 2

Require: \( \{U_c(t_i)\}_{i=1}^{n_j}, U_0 \): initial data, \( \tau \): step size for the minimization problem.

Require: \( \{t_{n_j}\}_{j=0}^{s+1}, \theta_0 \): initial guess of \( \theta \) on \([t_0, t_1]\)

1: for \( j = 0 \) to \( s \) do
2: \( \theta, U = \text{Algorithm 1}(\{U_c(t_i)\}_{i=n_j}^{n_{j+1}}, [t_{n_j}, t_{n_{j+1}}], U_0, \theta_0, \tau) \)
3: \( \theta_0, U_0 \leftarrow \theta(t_{n_{j+1}}), U(t_{n_{j+1}}) \)
4: return \( \theta, U \)

Remark 3.3. If parameters vary dramatically, \( \theta(t_{n_j}) \) may not be a good initialization for \( \theta \) on \([t_{n_j}, t_{n_{j+1}}]\). When this happens we switch to the rough initial guess and test by trial and error.

Figure 1. (a) Reported and fitted cumulative infection and death cases in the US (b) Estimated SEIR parameters and the basic reproduction number. \( \beta (\mu) \) corresponds to the left (right) vertical axis, \( \epsilon = 0.2 \) and \( \gamma = 0.1 \) are almost constant. The dashed line in \( R_0 \) is a zoomed-in version on the tail of the solid line.

4. Experiments

We now present experimental results to demonstrate the good performance of our algorithms. In all the experiments, the normalization factors \( \lambda_1, \lambda_2 \) in (2.3) are chosen such that the loss with respect to the infection cases is at the same scale as the loss with
respect to the death cases. According to the Centers for Disease Control and Prevention (CDC) in the US, the median incubation period is 4-5 days from exposure to symptoms onset, persons with mild to moderate COVID-19 remain infectious no longer than 10 days after symptom onset, therefore, we set

$$\Theta = \{ \theta \mid 0 \leq \beta \leq 5, \ 0.2 \leq \epsilon \leq 0.25, \ 0.1 \leq \gamma \leq 0.2, \ 0 \leq \mu \leq 0.01 \}.$$ 

For the step size, considering that $\beta, \epsilon$ and $\gamma$ are almost at the same scale, which is 100 times greater than $\mu$, and the permissible range of $\beta$ is much large than $\epsilon$ and $\gamma$, we set $\tau_\epsilon = \tau_\gamma = \tau$, $\tau_\beta = 100\tau$ and $\tau_\mu = \tau/100$.

4.1. Covid-19 epidemic in the US. As of today [November 17, 2020], it has been about 300 days since the first infection case of COVID-19 was reported in the US. We thus consider the time period $[0, 300]$ and sample the data $[I_c, D_c]$ at $t_i = 2i$ for $0 < i \leq 150$. To apply Algorithm 2 we take $\{t_{n_j}\}_{j=0}^{s+1}$ as $\{0, 30, 60, 90, 150, 210, 270, 300\}$.

Data fitting via optimal control with SEIR model. Figure 1 (a) shows that our data-driven optimal control algorithm learns the data very well. In Figure 1 (b), there is a noticeable peak over the second month, where the value of reproduction number $R_0$ is very large due to a dramatic increase in infection rate $\beta$. After that, $R_0$ goes down to a lower range, with a slight rise around the 6-th month. Over the last two months we observed another increase in $R_0$. Overall, the value of $R_0$ stays above 1 and the pattern of $R_0$ is consistent with the increasing trend of the confirmed cases.

**Figure 2.** Prediction for the cumulative number of infection and death cases in the US for the next 14 days by SEIR model

**Short-time prediction.** For a short time period, one may expect the transmission to continue the same way; hence the obtained model parameters could be used for prediction over a short coming period. Here, we use the parameters at the 300th day ($\beta = 0.195$, $\epsilon = 0.2$, $\gamma = 0.1$, $\mu = 1.09 \times 10^{-4}$) to make prediction for the cumulative number of infection and death cases in the next 14 days. For this prediction, we simply exploit the 4th-order Runge-Kutta solver. Figure 2 shows the prediction result for 300 – 314 days.

**Scheduled control.** From the above prediction result, we see that without interventions, the amount of confirmed and death cases will increase rapidly. We would like to
see a slow down of the epidemic spreading as the outcome of various public health interventions. With the present approach, this can be formulated as a scheduled control. Specifically, with a desired level of infection and death cases at the final time $T$, we schedule a sequence of values at intermediate times, from which we apply our algorithm to learn the optimal parameters (control function) such that the state function reaches the desired value at $T$ along the scheduled path.

For instance, starting from the 270-th day, with the goal of controlling the cumulative number of infection and death cases at 9,746,063 and 234,390 on the 300-th day, respectively, we set a pair of values for each day in the 30 days as shown in Figure 3 (a), then learn the parameters from the scheduled data. The results are presented in Figure 3 (b). Figure 3 (a) shows that the goal can be achieved by setting the parameters as what have been learned.

To compare the situations with or without a scheduled control, we also present the reported data and corresponding training results in Figure 3. In fact, the scheduled intermediate values are obtained by assuming the daily increases were half of the reported daily increases. From Figure 3 (b), we see that the most significant difference occurs in $\beta$. This can be roughly interpreted as: if the contact rate $\beta$ could be reduced by 0.02, the number of confirmed cases over the last 30 days could have been reduced by 50%, though the corresponding $R_0$ is still greater than 1. For virus propagation to eventually stop, $R_0$ needs to be less than 1, for which $\beta$ must be less than $\gamma + \mu$.

4.2. Experimental results for other countries. The coronavirus pandemic continues to affect every region of the world, but some countries are experiencing higher rates of infection, while others appear to have mostly controlled the virus. In order to see the virus dynamics in other regions, we also provide results for some other selected countries.
Figure 4. (a) Reported and fitted cumulative infection and death cases in the UK (b) Estimated SEIR parameters and the basic reproduction number. \( \beta \) (\( \mu \)) corresponds to the left (right) vertical axis, \( \epsilon = 0.2 \) and \( \gamma = 0.1 \) are almost constant. The dashed line in \( R_0 \) is a zoomed-in version on the tail of the solid line.

From Figure 4 and 5, we see that the confirmed cases in the UK and France display similar patterns. Figure 6 shows that China was hit hard early on, but the number of new cases has largely been under control for months.

5. Discussion

In this paper, we introduced a data-driven optimal control model for learning the time-varying parameters of the SEIR model, which reveals the virus spreading process among several population groups. Here the state variables represent the population status, such as \( S, E, I, R, D \) while the control variables are rate parameters of transmission between population groups. The running cost is of discrete form fitting the reported data of infection and death cases at the observation times. The terminal cost can be used to quantify the desired level of the total infection and death cases at scheduled times. Numerical algorithms are derived to solve the proposed model efficiently. Experimental results show that our approach can effectively fit, predict and control the infected and deceased populations.

The data-driven modeling approach presented in this work is applicable to more advanced models such as with spatial movement effects, interaction of different class of populations, for which we need to formulate mean-field game type models over a spatial domain, see [20] for the mean-field game formulation of an epidemic control problem.
Figure 5. (a) Reported and fitted cumulative infection and death cases in France (b) Estimated SEIR parameters and the basic reproduction number. $\beta$ ($\mu$) corresponds to the left (right) vertical axis, $\epsilon = 0.2$ and $\gamma = 0.1$ are almost constant. The dashed line in $R_0$ is a zoomed-in version on the tail of the solid line.

the computational side, our approach involves a non-convex optimization problem, which comes from the multiplicative terms of the SEIR model itself. In future work, we intend to extend our algorithm to more advanced models.

References

1. Vladimir V Aleksandrov, *On the accumulation of perturbations in the linear systems with two coordinates*, Vestnik MGU 3 (1968).
2. Roy M. Anderson and Robert M. May, *Population biology of infectious diseases: Part I*, Nature 280 (1979), 361–367.
3. H. Behncke, *Optimal control of deterministic epidemics*, Optimal control applications and methods 21(6) (2000), 269–285.
4. Henri Berestycki, Jean-Michel Roquejoffre, and Luca Rossi, *Propagation of epidemics along lines with fast diffusion*, ArXiv abs/2005.01859 (2020).
5. Fred Brauer and Carlos Castillo-Chávez, *Mathematical models in population biology and epidemiology*, Texts in Applied Mathematics, vol. 40, Springer-Verlag, New York, 2001. MR 1822695
6. Alberto Bressan and Benedetto Piccoli, *Introduction to the mathematical theory of control*, AIMS Series on Applied Mathematics, vol. 2, American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2007. MR 2347697
7. F. Castiglione and B. Piccoli, *Cancer immunotherapy, mathematical modeling and optimal control*, Journal of Theoretical Biology 247(4) (2007), 723–732.
8. Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud, *Neural ordinary differential equations*, Conference on Neural Information Processing Systems (NIPS), 2018.
Figure 6. (a) Reported and fitted cumulative infection and death cases in China (b) Estimated SEIR parameters and the basic reproduction number. $\beta$ ($\mu$) corresponds to the left (right) vertical axis, $\epsilon = 0.2$ and $\gamma = 0.2$ are almost constant. The dashed line in $R_0$ is a zoomed-in version on the tail of the solid line.

9. F. L. Chernous and A. A. Lyubushin, Method of successive approximations for solution of optimal control problems, Optimal Control Appl. Methods 3 (1982), no. 2, 101–114.
10. Odo Diekmann and J. A. P. Heesterbeek, Mathematical epidemiology of infectious diseases, Wiley Series in Mathematical and Computational Biology, John Wiley & Sons, Ltd., Chichester, 2000, Model building, analysis and interpretation. MR 1882991
11. D. Greenhalgh, Hopf bifurcation in epidemic models with a latent period and nonpermanent immunity, Mathematical and Computer Modelling 25(2) (1997), 85–107.
12. D. Greenhalgh and R. Das, Modeling epidemics with variable contact rates, Theoretical Population Biology 47(2) (1995), 129–179.
13. Herbert W. Hethcote, The mathematics of infectious diseases, SIAM Review 42(4) (2000), 599–653.
14. Yuzo Hosono and Bilal Ilyas, Traveling waves for a simple diffusive epidemic model, Math. Models Methods Appl. Sci. 5 (1995), no. 7, 935–966. MR 1359215
15. Junyoung Jang, Hee Dae Kwon, and Jeehyun Lee, Optimal control problem of an SIR reaction–diffusion model with inequality constraints, Mathematics and Computers in Simulation 171 (2020), 136–151.
16. David G Kendall, Mathematical models of the spread of infection, Mathematics and computer science in biology and medicine 171 (1965), 213–225.
17. W. O. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, Proceedings of the Royal Society of London Series A 115(772) (1927), 700–721.
18. Andrei Korobeinikov, Global properties of SIR and SEIR epidemic models with multiple parallel infectious stages, Bull. Math. Biol. 71 (2009), no. 1, 75–83. MR 2469618
19. A. Lahrouz, H. El Mahjour, A. Settati, and A. Bernoussi, Dynamics and optimal control of a non-linear epidemic model with relapse and cure, Physica A: Statistical Mechanics and its Applications 496 (2018), 299–317.
20. Wonjun Lee, Siting Liu, Hamidou Tembine, Wuchen Li, and Stanley Osher, *Controlling propagation of epidemics via mean-field games*, ArXiv abs/2006.01249 (2020).
21. S. Lenhart and J. Workman, *Optimal control applied to biological models*, CRC Press, 2007.
22. Kezan Li, Guanghu Zhu, Zhongjun Ma, and Lijuan Chen, *Dynamic stability of an SIQS epidemic network and its optimal control*, Communications in Nonlinear Science and Numerical Simulation 66 (2019), 84–95.
23. Michael Y. Li, John R. Graef, Liancheng Wang, and János Karsai, *Global dynamics of a SEIR model with varying total population size*, Mathematical Biosciences 160(2) (1999), 191–213.
24. Michael Y. Li and James S. Muldowney, *Global stability for the SEIR model in epidemiology*, Mathematical Biosciences 125(2) (1995), 155–164.
25. Weimin Liu, Herbert W. Hethcote, and Simon A. Levin, *Dynamical behavior of epidemiological models with nonlinear incidence rates*, Journal of Mathematical Biology 25 (1987), 359–380.
26. Luca Magri and Nguyen Anh Khoa Doan, *First-principles machine learning modelling of COVID-19*, ArXiv abs/2004.09478 (2020).
27. Jaime Mena-Lorcat and Herbert W. Hethcote, *Dynamic models of infectious diseases as regulators of population sizes*, Journal of Mathematical Biology 30 (1992), 693–716.
28. R. Morton and K. H. Wickwire, *On the optimal control of a deterministic epidemic*, Advances in Applied Probability 6(4) (1974), 622–635.
29. L.S. Pontryagin, V.G Boltyanskii, R.V Gamkrelidze, and E.F. Mishchenko, *The mathematical theory of optimal processes*, CRC Press, 1962.
30. R. Tyrrell Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization 14(5) (1976), 877–898.
31. Suresh P. Sethi and Preston W. Staats, *Optimal control of some simple deterministic epidemic models*, The Journal of the Operational Research Society 29(2) (1978), 129–136.
32. Ruoyan Sun, *Global stability of the endemic equilibrium of multigroup SIR models with nonlinear incidence*, Comput. Math. Appl. 60 (2010), no. 8, 2286–2291. MR 2725319
33. Horst R. Thieme, *Epidemic and demographic interaction in the spread of potentially fatal diseases in growing populations*, Mathematical Biosciences 111(1) (1992), 99–130.
34. M.K. Makc Tiberiu Harko, Francisco S.N. Lobo, *Exact analytical solutions of the susceptible-infected-recovered (SIR) epidemic model and of the SIR model with equal death and birth rates*, Applied Mathematics and Computation 236(1) (2014), 184–194.
35. P. van den Driessche and James Watmough, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, vol. 180, 2002, John A. Jacquez memorial volume, pp. 29–48. MR 1950747

IOWA STATE UNIVERSITY, MATHEMATICS DEPARTMENT, AMES, IA 50011
*Email address: hliu@iastate.edu*