Recent advances in modeling and control of epidemics using a mean field approach

AMAL ROY[1,*] , CHANDRAMANI SINGH[2] and Y NARAHARI[1]

[1]Department of Computer Science and Automation, Indian Institute of Science, Bangalore 560012, India
[2]Department of Electronic Systems Engineering, Indian Institute of Science, Bangalore 560012, India
e-mail: amalroy@iisc.ac.in; chandra@iisc.ac.in; narahari@iisc.ac.in

Abstract. The modeling and control of epidemics, such as the novel Coronavirus, have become crucial on a global scale, for effective management of epidemic situations. This paper is focused on using the mean field approach for modeling and control of epidemics. The mean field approach is an effective alternative to the classical approach of using continuous time Markov decision process (CTMDP) models, which suffer from the curse of dimensionality and entail knowledge of global state information. The mean field approach captures the collective behavior of a dynamic system consisting of numerous interacting nodes representing individuals in the population. The objectives of this paper are twofold: (a) to provide an overview of the mean field approach to epidemic modeling and control, and (b) to present recent advances in this area. Emphasizing the importance of containing and suppressing epidemic spread through non-pharmaceutical interventions, the paper highlights the need to minimize loss of lives, reduce suffering, and alleviate the burden on the public healthcare system. A potential challenge here is the presence of a section of the population who act on their free will and deviate from recommended best practices, which could lead to a potential public health crisis. Motivated by this, the paper explores two specific threads to modeling and control. The first thread assumes that individual nodes comply with a socially optimal control policy mandated by a regulatory authority. The second thread allows for independent and strategic behavior by the individual nodes, modeled as a mean field game, where the strategies of rational agents are based on mean field Nash equilibria. The paper begins by discussing the modeling of epidemics using an extended SIVR (Susceptible-Infected-Vaccinated-Recovered) compartmental model, accompanied by an illustrative example. Next, a literature review is provided, focusing on the mean field approach for socially optimal control of epidemics and how a regulatory authority can effectively contain the spread. Subsequently, the paper presents an update on the use of mean field game-based approaches in studying epidemic spread and control. Finally, future research directions in this important area are discussed.

Keywords. Mean field games; mean field control; epidemic games; optimal control; population games; mean field nash equilibrium; Stackelberg game.

1. Introduction

Throughout history, epidemics have had major effects, often catastrophic, on the lives and lifestyle of the global population. The *Bubonic Plague* also called *Black Death* ravaged Asia and Europe in several waves during the fourteenth century, and is estimated to have caused the death of as much as one third of the population of Europe between 1346 and 1350. *Spanish Flu*, also known as the *Great Influenza* epidemic, was an exceptionally deadly global influenza pandemic which started in 1918. Two years later, nearly a third of the global population, or an estimated 500 million people, had been infected in four successive waves. Estimates of deaths range from 17 to 50

million, and possibly as high as 100 million, making it one of the deadliest pandemics in human history.

More recently, the entire world has been devastated by multiple waves of a novel Corona virus and its variants since December 2019. Unsurprisingly, this has led to research and innovation efforts of an unprecedented scale towards prediction, mitigation, and management of the pandemic. Epidemiology has now moved to the centre-stage of research and policy making in public health.

1.1 *Motivation*

When an epidemic or pandemic such as Corona strikes, it is important to contain and suppress the spread of the disease to minimize the loss of life and the suffering and also to

lower the burden on the health care system. Non-pharmaceutical interventions such as reduction of social interactions, masking, social distancing, hand washing, hand-sanitizing, and disinfecting surfaces are simple yet very effective measures to limit the spread of an ongoing epidemic. However, it is observed that in spite of the threat posed by the pandemic individuals tend to go by their freewill rather than adhering to best practices such as stated above. There are numerous instances, throughout the world, where individuals have refused to comply with Covid appropriate behaviour [1]. This could dangerously push the dynamics of the entire population towards alarming or even catastrophic consequences. This lack of responsible behavior has forced regulatory authorities to seek suitable measures and incentives to improve compliance to best practices. In order for these measures to be computed scientifically, we need an appropriate modeling and control framework.

1.2 *Mean field approach*

Set in the above backdrop, modeling and control of epidemics have assumed paramount importance. A natural and powerful model that has emerged in this context is a continuous time Markov decision process (CTMDP) which encompasses the classical compartmental paradigms such as the SIR (Susceptible - Infected - Recovered) model. The use of a CTMDP based model, however, poses certain technical and computational challenges. First, the state space of any CTMDP based model grows exponentially with the population size and the computation of the optimal policy using dynamic programming principles become intractable even for moderate population sizes [2]. Second, the controller needs to know the global population state to execute the policy. These two challenges motivate the need for a more efficient approach and the *mean field approach* has emerged as an effective alternative [3]. The mean field approach computes the collective behavior of a dynamical system comprising numerous interacting nodes (individuals in the population). Our objective in this paper is to provide a bird's eye view of and a state-of-the-art update on some recent advances in the use of the mean field approach for epidemic modeling and control.

1.3 *Structure of the paper*

Our discussion in this paper proceeds along two threads. The first thread assumes that the individual nodes faithfully follow a socially optimal control policy prescribed by a regulatory authority. Here we follow a mean field approach to derive optimal control. The second thread allows the nodes to exhibit individualistic, strategic behavior. In this case, a mean field game is formulated and the individuals'

controls are governed by the mean field Nash equilibrium. This paper is structured as follows.

- **Section 2**: This section provides a description of the model used to analyze the spread of epidemics in the paper. Initially, an extension of the classical SIR model (Susceptible-Infected-Recovered model) is introduced, incorporating vaccinations, known as the SIVR (Susceptible - Infected - Vaccinated - Recovered) model. The associated costs for each node, including lockdown cost, infection cost, and vaccination cost, are outlined. Subsequently, the evolution of the population is described as a time-inhomogeneous continuous time Markov chain. The formulation of the optimal control problem aims to minimize the expected average cost for an individual over a finite time horizon. Furthermore, a stochastic game model is presented to analyze situations where the individual nodes act strategically.

- **Section 3**: This section discusses mean field modeling of epidemics. The spread of epidemics can be modeled as the mean field limit of a sequence of dynamical processes. We consider a general dynamical system and recall key results in mean field analysis. We present the standard solution to the mean field control problem. We then consider the strategic case and present the mean field game model and a solution for the same.

- **Section 4**: Here, we present an illustrative example. To keep things simple, we consider an SIR model rather than an SIVR model. For this example, we illustrate mean field control for the non-strategic case followed by a mean field game model and control for the strategic case.

- **Section 5**: This provides a state-of-the-art update of mean field optimal control of epidemics when a centralized regulatory authority prescribes a control policy and the individual nodes faithfully follow the policy. There is rich literature on applying optimal control methods to compartmental models of epidemics. We categorize the literature into (1) Non-pharmaceutical interventions (2) Vaccination strategies.

- **Section 6**: This section is devoted to a state-of-the-art update of literature on the use of the mean field game approach in the study of epidemic spread and control. We have categorized the relevant literature into the following groups:

  1. Non-pharmaceutical interventions
  2. Vaccination strategies
  3. Control and policy design.

- **Section 7**: We conclude the paper by providing several directions for future research in this section.

Table 1. List of symbols and notation.

N	Number of nodes (individuals or agents) in the population
$\mathcal{S}$	State space of a Markov chain
K	Cardinality of a discrete state space $\mathcal{S}$
T	Time horizon
M_i	Number of nodes in state i
Y_i	State of agent i. $Y_i \in \mathcal{S}$
X_i	$X_i := M_i/N$, fraction of nodes in state i
p_i	Probability of an agent being in state i
Δ_K^N	$\left\{ \frac{x}{N} \mid x \in \mathbb{Z}_+^K, \sum_{i=1}^K x_i = N \right\}$
Δ_K	$\left\{ x \in \mathbb{R}_+^K \mid \sum_{i=1}^K x_i = 1 \right\}$
X^N	State of population of N agents $:= (X_1, \ldots, X_k) \in \Delta_K^N$
$x(t)$	State of the population in mean field $:= (x_1, \ldots, x_k) \in \Delta_K$
Q	$\in \mathbb{R}^{K \times K}$ Transition rate matrix. Q_{ij} is the transition rate from state i to j
$c_L(\cdot)$	Lockdown cost per unit time per individual
$c_I(\cdot)$	Infection cost per unit time per individual
c_V	Vaccination cost per individual
$u(\cdot)$	Control variable as a function of time. $u(t) \in U \subset \mathbb{R}^m$
$g(\cdot, \cdot)$	Running cost function for optimal control problem. $g : \mathbb{R}^K \times U \to \mathbb{R}$
$h(\cdot)$	Terminal cost function for optimal control problem. $h : \mathbb{R}^K \to \mathbb{R}$
π_t	Markov decision rule $\pi_t : \Delta_K^N \to U$
π	$\pi := (\pi_t, t \in [0, T])$ is the policy
$J_\pi(x)$	Cost for the population, given the initial state x and policy π
$J_{\bar{\pi}}(i, x, \pi)$	Cost for an agent starting in state i, using policy $\bar{\pi}$, with the population starting in state x and using policy π

2. Modeling of epidemics

Modeling of spread of epidemics is a vast topic and there are numerous papers which present a variety of models. We wish to point to the paper by Guan *et al* [4] which provides a review of literature on modeling techniques and dynamic models in the context of the COVID-19 pandemic. The authors there discuss how three aspects, (a) epidemiological parameter estimation, (b) trend prediction, and (c) control measure evaluation, which are addressed in the literature. They conclude that dynamic models which are extensions of basic SIR (Susceptible - Infected - Recovered) and SEIR (Susceptible - Exposed - Infected - Recovered) models provide useful insights into these three aspects. Their conclusion is that dynamic models are useful for exploring possibilities of interventions but can fail at making strong predictions about long-term disease dynamics.

In this section, the modeling framework employed throughout the paper is presented. Firstly, an extension of the classical SIR compartmental model is described, incorporating vaccinations, referred to as the SIVR model. The compartmental model proves suitable for illustrating the computation of optimal control and mean field equilibrium. The costs incurred by each node, including lockdown cost, infection cost, and vaccination cost, are outlined. The population's evolution is then described as a time-inhomogeneous continuous time Markov chain. The formulation of the optimal control problem aims to minimize the expected average cost for an individual over a finite time horizon. Additionally, a mean field game model
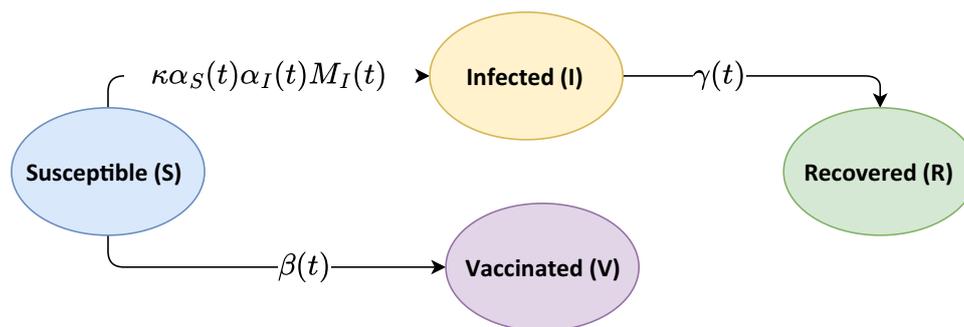
Figure 1. State transition diagram for an individual node. Values in edges are the transition rates between states.

is presented, wherein the individual nodes are strategic. Important symbols used in this paper are listed here for ready reference (table 1).

2.1 *Modeling an Individual Node: SIVR Model*

Let us consider a population of N nodes. Each node can be in one the four states, *susceptible* (S), *infected* (I), *vaccinated* (V) or *recovered* (R). The state transition diagram for an individual node is shown in figure 1 (a more detailed description of the figure appears subsequently). The nodes' states evolve with time due to their interactions with other nodes, external interventions or because of their own actions. For instance, a susceptible node becomes infected on meeting an infected node or can choose to become vaccinated. On the other hand, an infected node may recover over time either on its own or through medical treatment. It is assumed that recovered nodes will not be infected in the future and gain full immunity.

The number of unrestricted meetings between any pair of nodes is assumed to constitute a Poisson process of rate κ. Furthermore, these processes are assumed to be independent across node pairs. Moreover, the number of meetings between any pair of susceptible-infected nodes constitutes a Poisson process of rate $\kappa\alpha_S(t)\alpha_I(t)$ where $\alpha(t):=(\alpha_S(t),\alpha_I(t)) \in [\alpha_{\min},1] \times [\alpha_{\min},1]=:A$ quantify the efforts of susceptible and infected individuals to contact other individuals. We also assume that the susceptible nodes' vaccination processes are independent Poisson processes with rate $\beta(t) \in B$ and infected nodes' recovery times are exponentially distributed with parameter $\gamma(t) \in C$. We refer to $u(t):=(\alpha(t),\beta(t),\gamma(t)) \in U:=A \times B \times C$ as the action at time t, and the function $u : [0,T] \to U$ as the action function.

2.2 *Evolution of the population*

Let $Y_n(t)$ denote the state of the nth node at time t and $Y(t)$ denote the state of the entire population; $Y(t):=(Y_1(t), Y_2(t),\cdots,Y_N(t))$. Let $M(t):=(M_S(t),M_I(t),M_V(t),M_R(t))$ denote the numbers of nodes in states S, I, V and R at time t, respectively;

$$M_i(t) = \sum_{n=1}^{N} \mathbb{1}_{\{Y_n(t)=i\}} \text{ for } i \in \{S,I,V,R\}.$$

From the above discussion, $Y(t), t \geq 0$ is a time inhomogeneous continuous time Markov chain (CTMC). For the nth node, its state evolution at any time t depends on the joint population state $Y(t)$ only through $Y_n(t)$ and $M(t)$. Figure 1 shows the associated state transition rate diagram. Moreover, $M(t), t \geq 0$ is also a CTMC with transition rate diagram as in figure 2. Observe that $M_S(t) + M_I(t) + M_V(t) + M_R(t) = N$ for all t, and hence, it suffices to specify three numbers, $M_S(t), M_I(t)$ and $M_V(t)$.

Further, let $X(t):=(X_S(t),X_I(t),X_V(t),X_R(t))$ denote the fractions of nodes in states S, I, V and R at time t;

$$X(t) = \left(\frac{M_S(t)}{N},\frac{M_I(t)}{N},\frac{M_V(t)}{N},\frac{M_R(t)}{N}\right) \in \Delta_4^N$$

where $\Delta_K^N:=\left\{\frac{m}{N} : m \in \mathbb{Z}_+^K, \sum_{i=1}^{K} m_i = N\right\}$. Clearly, $X(t), t \geq 0$ is also a CTMC. Both $M(t), t \geq 0$ and $X(t), t \geq 0$ offer equivalent characterisations of the same dynamical system. From the state transition diagram in figure 2, we can also write the expected conditional drift rates of $X(t)$. These are as follows:

$$\frac{d\mathbb{E}[X_S(t)|X(t)]}{dt} = \kappa\alpha_S(t)\alpha_I(t)(NX_S(t))(NX_I(t))\left(\frac{-1}{N}\right)$$
$$+ \beta(t)NX_S(t)\left(\frac{-1}{N}\right)$$
$$= -N\kappa\alpha_S(t)\alpha_I(t)X_S(t)X_I(t) - \beta(t)X_S(t).$$
$$\tag{1a}$$

$$\frac{d\mathbb{E}[X_I(t)|X(t)]}{dt} = N\kappa\alpha_S(t)\alpha_I(t)X_S(t)X_I(t) - \gamma(t)X_I(t). \tag{1b}$$

$$\frac{d\mathbb{E}[X_V(t)|X(t)]}{dt} = \beta(t)X_S(t). \tag{1c}$$
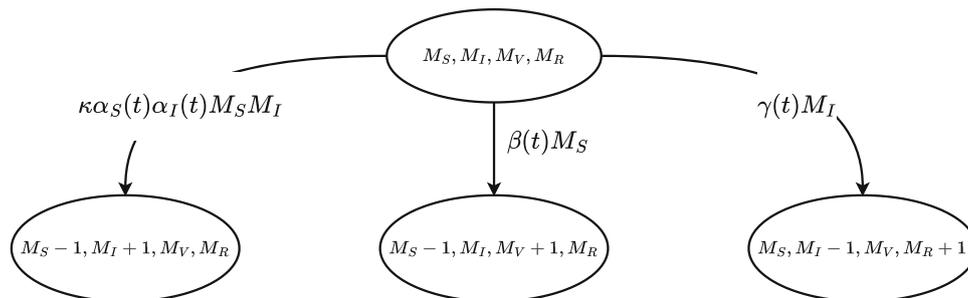
Figure 2. State transition diagram of the continuous time Markov chain $M(t), t \geq 0$ at $M(t) = (M_S, M_I, M_V, M_R)$.

We prefer to work with $X(t), t \geq 0$ for reasons that will be clear in the next section.

2.2.1 *Costs incurred by the nodes:*
We characterise the costs incurred by the nodes during the course of an epidemic. An individual can incur the following costs, according to the action at time t, $u(t) = (\alpha(t), \beta(t), \gamma(t)) \in A \times B \times C$.

1. *Lockdown cost:* This quantifies financial costs or discomfort caused to a susceptible or infected individual due to social distancing and lockdowns. These are reflected in susceptible and infected nodes' controls of meeting rates $\alpha_S(t)$ and $\alpha_I(t)$, respectively. Function $c_L : [\alpha_{\min}, 1] \to \mathbb{R}$ defines the lockdown cost where $c_L(\alpha(t))$ represents the lockdown cost per unit time for a node using action $\alpha(t)$.
2. *Infection cost:* This captures medical expenses, financial losses, etc., incurred by an infected individual. In general, the infection cost per unit time and the average recovery time $1/\gamma(t)$ are correlated. For instance, better health care incurs more cost per unit time but facilitates quicker recovery. To capture this correlation, we define function $c_I : C \to \mathbb{R}$ and $c_I(\gamma(t))$ denotes the infection cost per unit time.
3. *Vaccination cost*: This represents the per individual cost of vaccination and treatment of side effects of vaccination, if any. We assume the vaccination cost to be fixed. Let c_V represent the vaccination cost per individual.

Now we compute the average total cost, averaged over all the individuals. We express this cost as a function of initial population distribution $X(0) = x$ and actions $u(t) := (\alpha(t), \beta(t), \gamma(t)), t \geq 0$.

$$
\begin{aligned}
\frac{1}{N} \sum_{n=1}^{N} & \mathbb{E}\bigg[\int_0^T \big(c_L(\alpha_S(t)) \mathbb{1}_{Y_I(t)S} \\
& + (c_L(\alpha_I(t)) + c_I(\gamma(t))) \mathbb{1}_{Y_I(t)=I}\big) \, \mathrm{d}t \\
& + c_V \mathbb{1}_{Y_I(T)=V} \mid X(0) = x\bigg] \\
= \mathbb{E}\bigg[& \int_0^T (c_L(\alpha_S(t)) X_S(t) + (c_L(\alpha_I(t)) + c_I(\gamma(t))) X_I(t)) \, \mathrm{d}t \\
& + c_V X_V(T) \mid X(0) = x\bigg]
\end{aligned}
\tag{2}
$$

2.2.2 *The Optimal Control Problem:*
Observe that population evolution as well as the expected average cost over $[0, T]$ depend on actions $u(t), t \in [0, T]$. The optimal control problem seeks to minimize this cost via appropriate choice of actions.

Decision rule: A decision rule at time t, π_t, is function that takes the states $X(s), s \in [0, T]$ as input and output the action or decision $u(t)$. A history dependent decision rule at time t relies only on $X(s), s \in [0, t]$ for its output. Further, the output of a Markov decision rule at time t depends only

on the instantaneous state $X(t)$. More precisely, a Markov decision rule π_t is a mapping $\pi_t : \Delta_4^N \to U$.

Policy: A policy is an ordered set of decision rules for all $t \in [0, T]$:

$$
\pi := (\pi_t, t \in [0, T])
\tag{3}
$$

A policy is called history dependent (respectively, Markov) policy if it consists of history dependent (respectively, Markov) decision rules.

We have at our disposal a finite horizon continuous time Markov decision process (CTMDP) with finite state space and compact action spaces. Furthermore, the cost rates are bounded. Hence, following [3], there exists a Markov policy that is optimal within the class of all deterministic history-dependent policies. Further, observe that a Markov policy π together with the initial state $X(0)$ characterize the whole trajectory $X(t), t \in [0, T]$. So, the cost in (2) also is a function of π and $X(0) = x$, we use $J_\pi(x)$ to refer to this cost,

$$
\begin{aligned}
J_\pi(x) := \mathbb{E}\bigg[& \int_0^T c_L(\alpha_S(t)) X_S(t) \\
& + (c_L(\alpha_I(t)) + c_I(\gamma(t))) X_I(t) \, \mathrm{d}t \\
& + c_V X_V(T) \mid X(0) = x\bigg]
\end{aligned}
\tag{4}
$$

where $(\alpha(t), \beta(t), \gamma(t)) = \pi_t(X(t))$. The cost-to-go from a starting state $x \in \Delta_4^N$ is given by (5a) and the optimal control is given by (5b). The following pair of equations solves the Markov control problem.

$$
J(x) = \min_{\pi \in \Pi} J_\pi(x)
\tag{5a}
$$

$$
\pi^* \in \arg \min_{\pi \in \Pi} J_\pi(x).
\tag{5b}
$$

Here Π is the set of all Markov policies. However, this solution has two drawbacks.

1. The state space of the CTMDP grows exponentially with the population size. Consequently, the algorithm to obtain the optimal policy has a prohibitive complexity even for moderate population sizes.
2. The controller needs to know the global population state to execute the policy.

Both these problems are addressed with a mean-field approach. We work with the mean-field limit of the CTMDP as described in section 3.

2.3 *Evolution of an individual node*

We now focus on a particular node, say the nth node. Let $p(t) := (p_S(t), p_I(t), p_V(t), p_R(t))$ denote the conditional probabilities of this node being in states S, I, V and R, respectively, at time t, given $Y_n(0) = i$ and $X(0) = x$;

$$p_i(t) = \mathbb{E}\left[\mathbb{1}_{Y_n(t)=i}\big|Y_n(0)=i, X(0)=x\right] \text{ for } i \in \{S, I, V, R\}.$$

The rates of change of these probabilities are as follows.

$$\frac{dp_S(t)}{dt} = -N\kappa\alpha_S(t)\alpha_I(t)p_S(t)\,\mathbb{E}(X_I(t)) - \beta(t)p_S(t) \quad (6a)$$

$$\frac{dp_I(t)}{dt} = N\kappa\alpha_S(t)\alpha_I(t)\,\mathbb{E}(X_I(t)) - \gamma(t)p_I(t). \quad (6b)$$

$$\frac{dp_V(t)}{dt} = \beta(t)p_S(t) + \gamma(t)p_I(t). \quad (6c)$$

Suppose $Y_n(0) = i$ and $X(0) = x$. Further, suppose the nth node employs an action function $\bar{u}(t) = (\bar{\alpha}(t), \bar{\beta}(t), \bar{\gamma}(t)), t \geq 0$ whereas all other nodes use $u(t) = (\alpha(t), \beta(t), \gamma(t)), t \geq 0$. Then the expected total cost of the tagged node is given by the following expression.

$$\mathbb{E}\left[\int_0^T \left(c_L(\bar{\alpha}_S(t))\mathbb{1}_{Y_n(t)=S}\right.\right.$$
$$+ (c_L(\bar{\alpha}_I(t)) + c_I(\bar{\gamma}(t)))\mathbb{1}_{Y_n(t)=I}\right) dt$$
$$\left. + c_V\mathbb{1}_{Y_n(T)=V} \mid Y_n(0)=i, X(0)=x\right] = \int_0^T \left(c_L(\bar{\alpha}_S(t))p_S(t)\right.$$
$$\left. + (c_L(\bar{\alpha}_I(t)) + c_I(\bar{\gamma}(t)))p_I^N(t)\right) dt + c_V p_V(T). \quad (7)$$

2.4 *The case of strategic agents: stochastic game model*

Let us consider the individual nodes of the population to be strategic agents who wish to minimize their respective costs. These autonomous nodes also control their respective parameters giving rise to a stochastic game. We seek a symmetric Nash equilibrium, i.e., an equilibrium in which all the nodes employ the same policy.

A symmetric Nash equilibrium can be formally defined as follows. Suppose a tagged node uses a policy $\bar{\pi} = (\bar{\pi}_t, 0 \leq t \leq T)$ whereas all other nodes use a policy $\pi = (\pi_t, 0 \leq t \leq T)$. Then the occupancy measure of the other $N-1$ nodes is characterised by (6). On the other hand, the tagged node's state evolves as (6) with $u(t), t \in [0, T]$ replaced by $\bar{\pi}_t(X(t)), t \in [0, T]$, and its cost is

$$\bar{J}_{\bar{\pi}}(i, x, \pi) = \int_0^T (c_L(\bar{\alpha}_S(t))p_S(t)$$
$$+ (c_L(\bar{\alpha}_I(t)) + c_I(\bar{\gamma}(t)))p_I^N(t)) dt \quad (8)$$
$$+ c_V p_S(T)$$

where $(\bar{\alpha}(t), \bar{\beta}(t), \bar{\gamma}(t)) = \bar{\pi}_t(X(t))$. A policy $\bar{\pi}$ is called a symmetric NE if

$$\bar{\pi} \in \underset{\pi' \in \Pi}{\arg\min}\, \bar{J}_{\pi'}(i, x, \bar{\pi}) \quad (9)$$

for all $x \in \Delta_4^N$. Nash equilibrium computation and implementation are marred with computational complexity issues, much as those faced by the computation of an optimal policy. The computational complexity problem can be addressed by appealing to the mean field limit as described in section 3.

3. Mean field modeling of epidemic spread

This section discusses mean field modeling of epidemics. The spread of epidemics can be modeled as the mean field limit of a sequence of dynamical processes. To present the key results in mean field analysis, for the sake of convenience, the SIVR model presented in the previous section is moved away to a more general K-compartmental model. The standard solution to the mean field control problem is presented. Then, the strategic case is considered and the mean field game model and a solution for the same are presented.

The discussion starts by considering a sequence of systems with increasing population sizes. Under certain regularity conditions to be described below, the corresponding sequences of processes $X^N(t), t \geq 0$ and $p^N(t), t \geq 0$ converge to deterministic functions referred to as mean field limits. Below we discuss this convergence in the context of a more general dynamical system; spread of epidemics in a population will be a special case.

3.1 *Mean field limit in a general dynamical system*

Let us consider a population of N nodes, where each node can be in one of the K states, $\{1, \ldots, K\} =: S$. Let $Y^N(t) = \left(Y_1^N(t), \ldots, Y_N^N(t)\right)$ denote the states of the nodes at time t as in section 2.2.

3.1.1 *Evolution of population in the mean field case:* Let $\pi^N = \left(\pi_{i,t}^N, i \in S, t \in [0, T]\right)$ where $\pi_{i,t}^N : \Delta_K^N \to U_i$ for all $t \in [0, T]$ and U_i are compact subsets of $\mathbb{R}_+^{L_i}$, be the policy used by each of the nodes. Further, with the occupancy measure $X^N(t) \in \Delta_K^N, t \geq 0$ as defined in section 2.2, the expected conditional drift rates are, for $j = 1, \ldots, K$,

$$\frac{d\mathbb{E}\left[X_j^N(t)|X^N(t)=x\right]}{dt} = \sum_{i=1, i \neq j}^{K} x_i Q_{ij}\left(x_j, \pi_{i,t}^N(x), \pi_{j,t}^N(x)\right)$$

Here, for any i, j $(j \neq i)$, given $X^N(t) = x$, $Q_{ij}\left(x_j, \pi_{i,t}^N(x), \pi_{j,t}^N(x)\right)$ is the state transition rate of a node from state i to state j, and

$$Q_{ii}(x, \pi_{i,t}^N(x), j = 1, \ldots, K) := -\sum_{j \neq i} Q_{ij}\left(x_j, \pi_{i,t}^N(x), \pi_{j,t}^N(x)\right).$$

We can also write a general form of average cost of the population, $J_{\pi^N}^N(x)$, as follows.

$$J_{\pi^N}^N(x) = \mathbb{E}\left[\int_0^T X^N(t)^T g\left(X^N(t), \pi_{i,t}^N(X^N(t)), i \in \mathcal{S}\right) dt \right.$$
$$\left. + X^N(T)^T h\left(X^N(T)\right) \mid X^N(0) = x \right]$$

where

$$g\left(x, \pi_{i,t}^N(x), i \in \mathcal{S}\right) := \left(g_i\left(x, \pi_{i,t}^N(x)\right), i = 1, \ldots, K\right)$$

are the cost rates and $h(x) := (h_1(x), \cdots, h_K(x))$ are the terminal costs. The optimal control problem and its solution are as in (5).

3.1.2 *Evolution of an individual node in the mean field case:*

Consider a tagged node, say the nth node, which, at time $t = 0$ is in state $i \in \mathcal{S}$ and the initial population state, $X^N(0) = x$. Further, suppose that this node employs a policy $\bar{\pi}^N = (\bar{\pi}_{i,t}^N, i \in \mathcal{S}, t \in [0, T])$ whereas all other nodes use $\bar{\pi}^N = (\bar{\pi}_{i,t}^N, i \in \mathcal{S}, t \in [0, T])$. Let $p^N(t) := (p_1^N(t), \cdots, p_K^N(t))$ denote the conditional probabilities of the nth node being in states $1, \cdots, K$, respectively, at time t. The rate of change of these probabilities can be expressed as follows. For $j = 1, \cdots, K$,

$$\frac{dp_j^N(t)}{dt} = \sum_{i=1, i \neq j}^K p_i^N(t) \, \mathbb{E}\left[Q_{ij}(X_j^N(t), \bar{\pi}_{i,t}^N(X^N(t)), \bar{\pi}_{j,t}^N(X^N(t)))\right]$$

(10)

Further, the nth node's cost, $\bar{J}_{\bar{\pi}^N}(i, x, \pi^N)$, can be expressed as follows.

$$\bar{J}_{\bar{\pi}^N}^N(i, x, \pi^N) = \int_0^T p^N(t)^T g\left(X^N(t), \pi_{j,t}^N(X^N(t)), j \in \mathcal{S}\right) dt$$
$$+ p^N(T)^T h(X^N(t)).$$

Finally, the Nash equilibrium can be characterised as in (9).

Example (Epidemics): In the special case of evolution of epidemics as in section 2, $K = 4, L_1 = L_2 = 3, L_3 = L_4 = 0$ and for $u_1 = (\alpha_S, \beta)$ and $u_2 = (\alpha_I, \gamma)$,

$$Q(x, u_1, u_2) = \begin{bmatrix} -N\kappa\alpha_S\alpha_I x_I - \beta & N\kappa\alpha_S\alpha_I x_I & \beta & 0 \\ 0 & -\gamma & 0 & \gamma \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0. \end{bmatrix}.$$

Moreover,

$$g(x, u_1, u_2) = [c_L(\alpha_S) \quad c_L(\alpha_I) + c_I(\gamma) \quad 0 \quad 0]^T$$
$$h(x) = [0 \quad 0 \quad 0 \quad c_V]^T.$$

3.2 *Mean field control*

Darling [5] analyse the convergence of CTMCs to solutions of certain ODEs. Gast *et al* [3] study convergence of controlled DTMCs to continuous times controlled deterministic dynamical systems. Following these works, we make the following hypotheses on the dynamics and the costs.

3.2.1 *Mean field convergence:* Assumptions on dynamics

1. *Initial Conditions:* The initial occupancy measure $x^N(0)$ converges to $x(0) \in \Delta_K$ in probability, i.e., $\lim_{N \to \infty} P(\|X^N(0) - x(0)\| > \epsilon) = 0$ for all $\epsilon > 0$.
2. *Transition rates:* For all i, j $(j \neq i)$, the state transition rates $Q_{ij}(x_j, u_i, u_j)$ are $O(1)$. This implies that the expected number of transitions per unit time is $O(N)$.
3. *Drift rates:* The drifts are $O(N)$. Moreover, there exist bounded transition rate matrices $Q(\cdot, u) \in \mathbb{R}^{K \times K}$ $\forall u \in U := \bigotimes_{i=1}^K U_i$, such that $x^T Q(x, u)$ converges to $x^T Q(x, u)$ uniformly in (x, u).
4. *Lipschitz continuity:* $x^T Q(x, u)$ is Lipschitz continuous in (x, u).

Assumptions on the costs

1. *Cost rates:* The cost rates $g_i(\cdot, u_i)$ and the terminal costs $h_i(\cdot)$ are bounded.
2. *Lipschitz continuity:* The cost rates $g_i(\cdot, u)$ are Lipschitz continuous for all $u_i \in U_i$, and the terminal costs $h_i(\cdot)$ are also Lipschitz continuous.

Observe that the scaling conditions are satisfied in the case of epidemics if the parameters are chosen as: $\kappa^N = \kappa/N, \alpha_S^N = \alpha_S, \alpha_I^N = \alpha_I, \beta^N = \beta$ and $\gamma^N = \gamma$. We further assume that $c_L(\cdot)$ and $c_I(\cdot)$ are bounded functions. It can be easily checked that the above hypotheses are satisfied under these assumptions.

In order to state the mean field convergence results, we introduce the following optimal control problem. Consider a continuous time dynamical system

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = Q^T(x(t), u(t))x(t) \tag{11}$$

with state $x(t), t \geq 0$, initial condition $x(0)$, and action function $u(t) = (u_i(t), i = 1, \cdots, K), t \geq 0$. Let $J_u : \Delta_K \to \mathbb{R}_+$ be the associated cost function defined as follows.

$$J_u(x) = \int_0^T x(t)^T g(x(t), u(t))\, \mathrm{d}t + x(T)^T h(x(T))$$

where $x(t), t \geq 0$ is a solution to (11) given $x(0) = x$. We seek to find an action function $u^*(t), 0 \leq t \leq T$ which together with its corresponding state trajectory $x^*(t), 0 \leq t \leq T$ attains the minimum cost

$$J(x) := \min_u J_u(x).$$

We then have the following convergence results.

1. *Optimal trajectory:* The CTMC $X^N(t), 0 \leq t \leq T$ converges to the mean field limit $x(t), 0 \leq t \leq T$. More precisely, for all $\epsilon > 0$,

$$\lim_{N \to \infty} \mathbb{P}\left[\sup_{0 \leq t \leq T} \|X^N(t) - x(t)\| > \epsilon\right] = 0$$

2. *Optimal cost:* The optimal cost for the population of size N, $J^N(X^N(0))$, converges to the optimal cost of the mean field control problem, $J(x(0))$, in probability.

3. *Optimal policy:* An optimal action function for the mean field limit, say u^*, is asymptotically optimal for the system with finitely many nodes. More precisely, $J_{\pi^*}^N(X^N(0)) - J(x(0))$ converges to 0 in probability where $\pi_{i,t}^* \equiv u_i^*(t)$ for all i.

3.2.2 *Solution to the mean field control problem:* With a slight abuse of notation, we let $J_t(x)$ denote the optimal cost over $[t, T]$ given $x(t) = x$. We can write

$$J_t(x) = \min_{u \in U}\{x^T g(x, u)\mathrm{d}t + J_{t+\mathrm{d}t}(x(t + \mathrm{d}t))\}$$
$$= \min_{u \in U}\{x^T g(x, u)\mathrm{d}t + J_{t+\mathrm{d}t}(x) + x^T Q(x, u)\nabla_x J_t(x)\mathrm{d}t\}$$

On rearranging the terms, we obtain

$$-\nabla_t J_t(x) = \min_{u \in U}\{x^T g(x, u) + x^T Q(x, u)\nabla_x J_t(x)\}.$$

The above equation, referred to as the Hamilton-Jacobi-Bellman (HJB) equation, yields both the optimal action function $u^* : [0, T] \to U$ and the optimal cost $J(x) \equiv J_0(x)$ of the mean field control problem.

The Minimum Principle [6] : Let us introduce adjoint (also called *costate*) processes $\lambda(t) \in \mathbb{R}^K, t \in [0, T]$ and the Hamiltonian function $H : \Delta_K^N \times U \times \mathbb{R}^K \to \mathbb{R}$ given by

$$H(x, u, \lambda) = x^T g(x, u) + x^T Q(x, u)\lambda.$$

Let $u^*(t), t \in [0, T]$ be an optimal action function and $x^*(t), t \in [0, T]$ be the corresponding state trajectory, i.e., $x^*(t), t \in [0, T]$ is the solution to (11) with $u(t) = u^*(t), t \in [0, T]$ and initial condition $x^*(0) = x$. The HJB equation can be compactly written in terms of the Hamiltonian function as

$$-\nabla_t J_t(x^*(t)) = H(x^*(t), u^*(t), \nabla_x J_t(x^*(t))).$$

The following result, referred to as *the Minimum Principle*, provides a necessary condition for optimality of $u^*(t), t \in [0, T]$. Let $\lambda(t), t \in [0, T]$ be the solution to equation

$$\frac{\mathrm{d}\lambda(t)}{\mathrm{d}t} = -\nabla_x H(x^*(t), u^*(t), \lambda(t)),$$

referred to as the adjoint equation, with the boundary condition:

$$\lambda(T) = \nabla(x^*(T)^T h(x^*(T))).$$

Then, for all $t \in [0, T]$,

$$u^*(t) = \underset{u \in U}{\arg\min}\, H(x^*(t), u, \lambda(t)).$$

Furthermore, there is a constant θ such that

$$H(x^*(t), u^*(t), \lambda(t)) = \theta, \text{ for all } t \in [0, T].$$

The minimum principle can be used as the basis of a numerical solution. In the so called *two-point boundary problem method*, we use the necessary condition

$$u^*(t) = \underset{u \in U}{\arg\min}\, H(x^*(t), u, \lambda(t)).$$

to express $u^*(t)$ in terms of $x^*(t)$ and $\lambda(t)$. We then substitute the result into the system and the adjoint equations, to obtain a set of $2K$ first order differential equations in $x^*(t)$ and $\lambda(t)$. These equations can be solved using the split boundary conditions:

$$x^*(0) = x \text{ and } \lambda(T) = \nabla(x^*(T)^T h(x^*(T))).$$

3.2.3 *Solution to the finite population control problem:* We thus obtain the following procedure for solving the optimal control problem for a finite population.

- From the original system with N nodes, write the mean field limit. In particular, set $x(0) = X^N(0)$ and obtain $Q(\cdot, u)$ via appropriate scaling of parameters.
- Obtain the optimal control for the limiting problem via solving the HJB equation or via some other method, e.g., using the minimum principle [6].
- Use this control in the finite population problem. From the above discussion, this is asymptotically optimal.

3.3 *Mean field game model of epidemic spread*

Let us focus on a tagged node, say the nth node. Suppose it uses an action function $\bar{u} : [0, T] \to U$ whereas all other nodes use an action function $u : [0, T] \to U$. Following the discussion in section 3.2.1, the occupancy measures of these nodes, $X^N(t), t \in [0, T]$ asymptotically approach the mean field limit $x(t), t \in [0, T]$. Hence, in the limiting system, the probabilities of the tagged node being in various states, $p(t), t \in [0, T]$, given its initial state, say i, evolve as follows.

$$\frac{\mathrm{d}p}{\mathrm{d}t} = Q^T(x(t), \bar{u}(t))p(t) \tag{12}$$

with $p(0) = \delta_i$. Let $\bar{J}_{\bar{u}}(i, x, u)$ be the associated cost defined as follows.

$$\bar{J}_{\bar{u}}(i, x, u) = \int_0^T p(t)^T g(x(t), u(t)) \, \mathrm{d}t + p(T)^T h(x(T))$$

where $x(t), 0 \le t \le T$ and $p(t), 0 \le t \le T$ are solutions to (11) and (12), respectively, given $x(0) = x$ and $p(0) = \delta_i$. We seek to find a action function $\bar{u} : [0, T] \to U$, which, together with the corresponding trajectory $\bar{p}(t), t \in [0, T]$, minimizes the tagged node's cost when other nodes are also using the same action function $\bar{u} : [0, T] \to U$, i.e.,

$$\bar{u} \in \arg\min_{u'} \bar{J}_{u'}(i, x, \bar{u}).$$

The equilibrium trajectory, policy, and cost for the stochastic game for a finite size population are expected to be asymptotically close to those for the mean field game. This is a conjecture and we do not yet have a proof of convergence.

3.3.1 *Solution to the mean field game:*   With a slight abuse of notation, let $\bar{J}_t(i, x, u)$ denote the optimal cost of the nth node over $[t, T]$ given $Y_n(t) = i, X(0) = x$. Then,

$$\begin{aligned}
\bar{J}_t(i, x, u) &= \min_{u'_i \in U_i} \mathbb{E}\big[g_i(x(t), u'_i) \, \mathrm{d}t \\
&\quad + \bar{J}_{t+\mathrm{d}t}(Y_n(t + \mathrm{d}t), x, u) \mid Y_n(t) = i, x(0) = x\big] \\
&= \min_{u'_i \in U_i} \Big\{ g_i(x(t), u'_i) \, \mathrm{d}t \\
&\quad + \sum_{j \ne i} Q_{ij}(x_j(t), u'_i, u_j(t)) \, \mathrm{d}t \, \bar{J}_{t+\mathrm{d}t}(j, x, u) \\
&\quad + \Big(1 - \sum_{j \ne i} Q_{ij}(x_j(t), u'_i, u_j(t)) \, \mathrm{d}t\Big) \bar{J}_{t+\mathrm{d}t}(i, x, u) \Big\}
\end{aligned} \tag{13}$$

On rearranging the terms,

$$\begin{aligned}
&- \nabla_t \bar{J}_t(i, x, u) \\
&= \min_{u'_i \in U_i} \Big\{ g_i(x(t), u'_i) \\
&\quad + \sum_{j \ne i} Q_{ij}(x_j(t), u'_i, u_j(t))(\bar{J}_t(j, x, u) - \bar{J}_t(i, x, u)) \Big\} \\
&= \min_{u'_i \in U} \Big\{ g_i(x(t), u'_i) + \sum_j Q_{ij}(x_j(t), u'_i, u_j(t)) \bar{J}_t(j, x, u) \Big\}
\end{aligned} \tag{14}$$

This is the HJB equation for the mean field game. A control $\bar{u}$ is a Nash equilibrium of the game provided

$$\begin{aligned}
\bar{u}_i(t) \in \arg\min_{u'_i \in U_i} \\
\Big\{ g_i(x(t), u'_i) + \sum_j Q_{ij}(x_j(t), u'_i, \bar{u}_j(t))) \bar{J}_t(j, x, \bar{u}) \Big\}
\end{aligned}$$

for all $t \in [0, T]$. So, Nash equilibria are characterized by the system of Kolmogorov and HJB equations, (11) and (14), respectively, together with the boundary conditions:

$$x(0) = x$$
$$\text{and } \bar{J}_T(i, x, \bar{u}) = h_i(x(T)).$$

Thus we get a initial-terminal value problem (ITVP) whose fixed points yield the solutions to the mean field game. Under regularity assumptions for the cost functions $g_i(\cdot, \cdot)$ and the transition rates $Q(\cdot, \cdot)$, it can be shown that a unique minimizer exists in (14), which is the Nash equilibrium of the game [7, section 7.2.2]. Following is a fixed point iteration to obtain a Nash equilibrium given $x(0)$.

- Initialize with an action function $u : [0, T] \to U$.
- Solve Kolmogorov equations (11) to obtain trajectory $x(t), 0 \le t \le T$ corresponding to u.
- For each $j \in \{1, \cdots, K\}$, solve HJB equations (14) to obtain the best responses $u'_i : [0, T] \to U_i$ to u for all i.

- If $u'(t) = u(t)$ for all $t \in [0, T]$, set $\bar{u} = u'$ and exit. Otherwise, set $u = u'$, and continue with the next iteration (go to Step 2).

Let f_x denote the mapping from u to $\bar{u}$. When f_x is a *contractive* mapping on a complete metric space, Picard-Banach fixed point theorem states that there is a unique fixed point and that the convergence is geometric [8].

The Master Equation: [7, section 6.5] Let us introduce the adjoint processes $\bar{\lambda}(t) \in \mathbb{R}^K, t \in [0, T]$ and the Hamiltonian functions $H_i : \Delta_K^N \times U \times \mathbb{R}^K \to \mathbb{R}$ given by

$$H_i(x, u, \lambda) = g_i(x, u_i) + \sum_j Q_{ij}(x_j, u_i, u_j)\lambda_j.$$

Let $\bar{u}_i(t), i \in \mathcal{S}, t \in [0, T]$ be the best response action functions of the tagged player. The HJB equations can be compactly written in terms of the Hamiltonian functions as

$$-\nabla_t \bar{J}_t(i, x, u) = \min_{u'_i \in U_i} H_i(x(t), (u'_i, u_{-i}(t)), \bar{J}_t(\cdot, x, u)) \ \forall i \in \mathcal{S},$$

and

$$\bar{u}_i(t) \in \arg\min_{u'_i \in U_i} H_i(x(t), (u'_i, \bar{u}_{-i}(t)), \bar{J}_t(\cdot, x, \bar{u})) \ \forall i \in \mathcal{S},$$

where $\bar{J}_t(\cdot, x, u) \in \mathbb{R}^K$ has jth component $\bar{J}_t(j, x, u)$ for all $j \in \mathcal{S}$. Finally, let $\mathcal{J}_t, t \in [0, T]$ be real valued functions, $\mathcal{J}_t : [K] \times \mathbb{R}^K \to \mathbb{R}$, and let us also introduce the following equation known as the master equation.

$$
\begin{aligned}
&\nabla_t \mathcal{J}_t(i, x) + H_i(x, (v_i(x, u_{-i}(t), \mathcal{J}_t(\cdot, x)), \\
&\quad u_{-i}(t)), \mathcal{J}_t(\cdot, x)) \\
&+ \sum_j \left(\sum_l x_l Q_{lj}(x_j, v_i(x, u_{-i}(t), \mathcal{J}_t(\cdot, x)), u_j(t)) \right) \\
&\nabla_{x_j} \mathcal{J}_t(i, x)) = 0
\end{aligned}
\tag{15}
$$

where

$$v_i(x, u_{-i}, \lambda) \in \arg\min_{u'_i \in U_i} H_i(x, (u'_i, u_{-i}), \lambda)$$

where

$$q_j(x, u, z) = \sum_l x_l Q_{lj}(x_j, u_l(z), u_j).$$

Let $\mathcal{J}_t, t \in [0, T]$ be a solution to the master equation with terminal condition $\mathcal{J}_T(i, x) = h_i(x)$ for all $i \in \mathcal{S}$, and $\bar{u}(t), t \in [0, t]$ be the associated optimal action function. Let $\bar{x}(t), t \in [0, T]$ be a solution to

$$\frac{\mathrm{d}\bar{x}_i(t)}{\mathrm{d}t} = q_t(i, \bar{x}(t), \bar{u}(t), \mathcal{J}_t(\cdot, \bar{x}(t))) \tag{16}$$

with initial condition $\bar{x}(0)$. Then, setting $\bar{J}_t(i, x, u) = \mathcal{J}_t(i, \bar{x}(t)))$,

$$
\begin{aligned}
&\nabla_t \bar{J}_t(i, x, u) + H_i(x(t), \bar{u}(t), \bar{J}_t(\cdot, x(t), u(t))) \\
&= \nabla_t \mathcal{J}_t(i, x) + \sum_j q_t(j, x, \bar{u}(t), \mathcal{J}_t(\cdot, x)) \nabla_{x_j} \mathcal{J}_t(\cdot, x) \\
&\quad + H_i(x, \bar{u}(t), \mathcal{J}_t(\cdot, x)) \\
&= 0.
\end{aligned}
$$

Clearly, $\bar{J}_t(i, x, u)$ solves the HJB equation and is value function in the optimization problem (14). Also, (16) can be identified with the Kolmogorov equation (11). Consequently, $\bar{x}(t)$ is the equilibrium state trajectory. It is therefore seen that the master equation encapsulates both the Kolmogorov and the HJB equations in a single equation.

3.3.2 *Solution to the finite population stochastic game:* Finally, we can adopt a procedure similar to that in section 2.2, to obtain a solution to the stochastic game for a population of size N.

4. An illustrative example

Let us now formulate a mean field control and a mean field game problem in a compartmental model of epidemic spread. For ease of our presentation, we simplify the SIVR model introduced in section 2 to an SIR model without vaccination. We assume that the transmission rate of the disease, κ, is influenced by two factors: the disease characteristics and the baseline contact factor within the population. The recovery rate γ is taken to be a constant. Individuals can adjust their level of social interaction by choosing their *contact factor*. This is reflected by choosing a value of $\alpha(t):=(\alpha_S(t), \alpha_I(t)) \in [\alpha_{\min}, 1] \times [\alpha_{\min}, 1]=:A$. α_{min} represents the minimal transmission rate that an individual who adopts maximum protection effort will encounter.

4.1 *Illustrative example: Mean field control*

From a societal viewpoint, we assume that the regulating authority chooses a policy $\pi = (\pi_t, t \in [0, T])$ for the whole population. This can be thought of as a combination of non-pharmaceutical interventions like lockdowns, mandatory usage of masks, proactive testing and quarantine, etc. We

look at a finite horizon setup to identify the social optimum strategy that minimizes the total social cost.

Once the social planner commits to a policy $\pi(t)$, analogous to equation (11), the forward evolution of the state with the initial conditions $x(0) = (x_S(0), x_I(0), x_R(0))$ is given by:

$$\dot{x}_S(t) = -\kappa \alpha_I(t) \alpha_S(t) x_S(t) x_I(t) \tag{17a}$$

$$\dot{x}_I(t) = \kappa \alpha_I(t) \alpha_S(t) x_S(t) x_I(t) - \gamma x_I(t) \tag{17b}$$

$$\dot{x}_R(t) = \gamma x_I(t) \tag{17c}$$

We assume that individuals in states $i \in \{S, I\}$ choosing a contact factor $\alpha_i(t)$ incurs a running cost $c_L(\alpha_i(t)) = c_Q(1 - \alpha_i(t))$. In addition to this lockdown cost, infected individuals pay a cost c_I per unit time.

Starting from $x(0) = (x_S(0), x_I(0), x_R(0))$, policy π induces a social cost which can be expressed as:

$$J_\pi(x) = \int_0^T (c_Q(1 - \alpha_S(t)) x_S(t)$$
$$+ (c_I + c_Q(1 - \alpha_I(t))) x_I(t)) \, dt$$

Hamiltonian corresponding to the optimal control problem is defined as,

$$H(x, \pi, \lambda) = c_Q(1 - \alpha_S(t)) x_S(t)$$
$$+ (c_I + c_Q(1 - \alpha_I(t))) x_I(t)$$
$$- \lambda_2(t) \gamma x_I(t) + \kappa \alpha_I(t) \alpha_S(t) x_S(t) x_I(t)(\lambda_2(t) - \lambda_1(t))$$

Defining $\lambda(t) = (\lambda_1(t), \lambda_2(t), \lambda_3(t))$ as the co-state variable vector, with $\lambda_3(t) := 0 \quad \forall t \in [0, T]$,
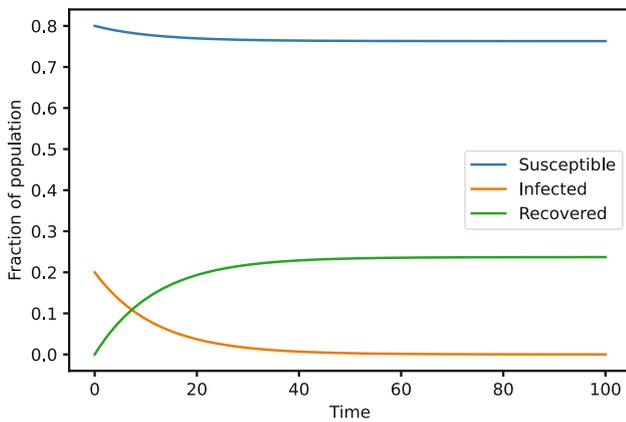
Figure 3. Time evolution of the state (x_S, x_I, x_R) under socially optimal control policy.
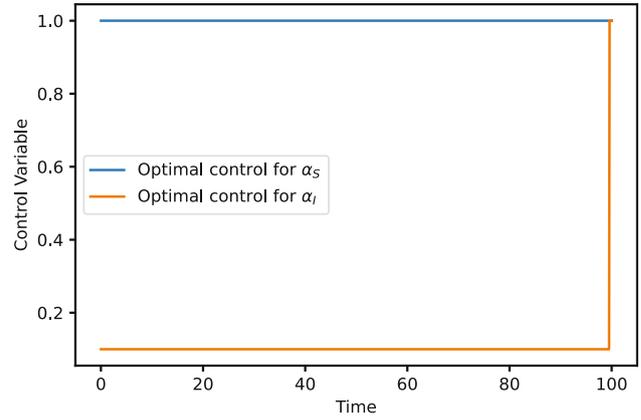
Figure 4. Socially optimal control policy.

Table 2. Set of parameters for the numerical experiments.

(S_0, I_0, R_0)	γ	κ	T	c_I	c_Q	α_{min}
(0.8, 0.2, 0)	0.1	0.2	100	100	10	0.1

$$\dot{\lambda}_1 = -\frac{\partial H}{\partial x_S} = -c_Q(1 - \alpha_S(t))$$
$$+ \kappa \alpha_I(t) \alpha_S(t) x_I(t)(\lambda_1(t) - \lambda_2(t))$$
$$\dot{\lambda}_2 = -\frac{\partial H}{\partial x_I} = -c_I - c_Q(1 - \alpha_I(t)) + \lambda_2(t)\gamma$$
$$+ \kappa \alpha_I(t) \alpha_S(t) x_S(t)(\lambda_1(t) - \lambda_2(t)) \tag{18}$$

The final constraints for the co-state variables are: $\lambda_1(T) = \lambda_2(T) = 0$.

From our discussion in section 3.2.2, functions λ_1, λ_2 exist that satisfy equation (18), and the optimal control π^* is the minimization of Hamiltonian, assuming that the co-state variables are set according to optimum control.

4.1.1 Numerical computation of optimal control:
Note that in the previous section, we had to deal with two sets of ODEs. We have the forward evolution ODEs of the compartmental model (17) along with their initial conditions. In addition to which, we had the co-state ODE (18) with terminal conditions. There are several classical procedures available to compute the optimal control.

We use the forward-backward sweep method in [9, Chapter 4]. Figure 3 shows time evolution of the epidemic in the population. Figure 4 shows the socially optimal control for the set of parameters in table 2.

4.2 Illustrative example: mean field game model

From the point of view of a strategic individual node, computing the best response to a population strategy can be seen as a mean field game. Suppose a tagged node uses a policy $\bar{u} = (\bar{\alpha}(t))_{t \in [0,T]}$ whereas all other nodes use a policy u. The cost of an individual node is

$$\bar{J}_{\bar{u}}(i,x,u) = \int_0^T \left\{ c_Q(1 - \bar{\alpha}_S(t))p_S(t) \right.$$
$$\left. + (c_I + c_Q(1 - \bar{\alpha}_I(t)))p_I(t) \right\} \mathrm{d}t.$$

We seek to find a policy $\bar{u} : [0,T] \to U$ which, together with the corresponding trajectory $\bar{p}(t), t \in [0,T]$ minimizes the tagged node's cost when other nodes are also using the same policy $\bar{u}$ (9).

Writing the corresponding Bellman equations (from section 3.3), with $J_i(t)$ denoting the cost to go at time t from state $i \in \{S, I\}$ we have:

$$-\dot{J}_S(t) = \min_{\alpha_S(t)} c_Q(1 - \alpha_S(t)) + \kappa \alpha_S(t)\alpha_I(t)p_I(t)(J_I(t) - J_S(t))$$
$$\tag{19a}$$

$$-\dot{J}_I(t) = \min_{\alpha_I(t)} c_I + c_Q(1 - \alpha_I(t)) - \gamma J_I(t) \tag{19b}$$

along with the terminal conditions $J_S(T) = J_I(T) = 0$. The cost-to-go from recovered state $J_R(t)$ is identically set to zero. From (19), we can immediately see that for an infected individual, $\alpha_I(t) = 1$ is the best response action. Defining a switching function $\phi(t) := \kappa p_I(t)(J_I(t) - J_S(t)) - c_Q$, we can find the individual best response strategy for a susceptible individual as follows: for any $t \in [0,T]$,

$$\alpha_S^*(t) = \begin{cases} \alpha_{min} & \text{if } \phi(t) \geq 0 \\ 1 & \text{if } \phi(t) < 0 \end{cases}$$

4.2.1 Numerical computation of mean field equilibrium:

Nash equilibria correspond to a fixed point of the best response function. The idea is to use an inductive sequence $u_{n+1} = f_{x_0}(u_n)$ to reach the fixed point. The Kolmogorov equations (11) are solved for a given control u_n and the best response is computed from solving the Bellman ODE (19) backward in time, giving u_{n+1}. We use the iterative scheme detailed in section 3.3.1 to compute the mean field equilibrium.

Figure 5 shows the individually optimal control for the set of parameters in table 2. We observe that the socially optimal control policy requires that the population should reduce their social interaction for a longer period compared to the mean field equilibrium policy.
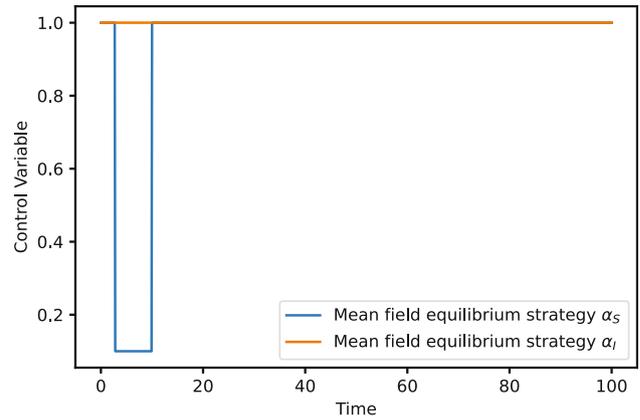
Figure 5. Mean field equilibrium control policy.

4.3 Comparison of mean field control and mean field equilibrium strategies

We can see from figure 4 that in an idealized setting where the regulating authority has full control on contact between individuals, only the infected needs to be isolated and this is reflected in the control actions α_S and α_I chosen for individuals in Susceptible and Infected states respectively. However, when we look at a mean field equilibrium strategy (figure 5), infected individuals are self interested and does not undergo mitigation efforts. When the infection is near its peak, susceptible individuals undergo protective measures to avoid getting the infection. This is consistent with the free riding phenomenon which has been observed in the context of COVID-19 pandemic.

5. Current state-of-the-art in optimal control of epidemics

Optimal control of epidemics concerns how a regulating authority can optimally contain epidemic spread in a population. Depending on the underlying epidemic model, we can classify them as (1) compartmental models (2) network based models. Tools from optimal control were applied to compartmental models as early as 1970s [10, 11]. Compartmental models are formulated as Markov chains and due to their simplicity, they will be the go-to model in our discussions. Nowzari, Preciado and Pappas [12] presents a survey on analysis and control of epidemics in complex networks. In this section, we categorize the literature into the following groups according to the control variable available to the regulating authority: (1) non-pharmaceutical interventions (2) vaccination strategies.

5.1 *Optimal control: non-pharmaceutical interventions*

The first line of defence against epidemics in the absence of preventative vaccines or in the case of vaccine hesitancy are non-pharmaceutical interventions which reduce the mixing of infected people in the community. Non-pharmaceutical interventions include (1) a full regional lockdown, where only essential services are allowed to operate (2) mandatory use of personal protective equipment (3) targeted interventions like symptomatic testing, contact tracing and quarantine, and (4) travel restrictions.

In the initial stages of the COVID-19 pandemic, the governments all around the world introduced strict regional lockdowns and travel bans to curb the disease spread. However, this led to serious economic and social disruptions as witnessed by the global economic recession following COVID-19. Early on in the disease evolution, extensive testing, contact tracing and quarantine are effective ways to prevent an exponential growth of the disease. This necessitates a timely intervention from governments in terms of healthcare infrastructure and testing capacity. Over the years, researchers have explored how to model the cost associated with these interventions and how to frame the problem faced by the regulator as an optimal control problem.

Abakuks [10, 11] first formulated stochastic and deterministic compartmental models where optimal policies are computed numerically for the stochastic model and analytically for the deterministic model. Abakuks [10] considers a compartmental SIR model of epidemic evolution. The length of infection period is assumed to have an exponential distribution. Under the assumption that a subset of infected population can be isolated instantaneously, the optimal isolation policy is calculated. Note that this is a restrictive assumption because instantaneous isolation is not practical and for many diseases, testing is required to ascertain the state of an individual agent. Optimal isolation policies are numerically computed for a stochastic model and analytically computed for a deterministic model and the two policies compared. Abakuks [11] uses a similar model to analyze optimal vaccination policies.

Wickwire [13] extended the previous works of Abakuks, [10, 11] on isolation and vaccination to analyze Kermack-McKendrick type compartmental models by means of value function and the Bellman equation. Optimal isolation policies for deterministic and stochastic epidemics were determined by [13]. Wickwire *et al* also relaxed the unrealistic assumption that an arbitrary number of individuals can be isolated instantaneously, and instead assumed that there is a hard constraint on the isolation rate. See [14] for a survey.

Behncke [15] further adapts Wickwire's model [14] to a more general control and cost setup and deals with both finite and infinite time horizon problems. They use the Pontryagin's maximum principle [6] throughout instead of

Bellman equations and study the control problem qualitatively, thus avoiding differentiability assumptions on the value function.

In the context of the COVID-19 pandemic, optimal control has been applied to compartmental and network based models. Tsay *et al* [16] use a compartmental SEAIR (susceptible - exposed - asymptomatic - infectious - recovered) model with an asymptomatic compartment, provide methods for estimation of parameters, and study the effect of isolation measures. The work by Perkins *et al* [17] performs optimal control analysis of a compartmental SEAIHV (susceptible - exposed - asymptomatic - infected - hospitalised - vaccinated) model and concludes that (a) heightened control early on in the pandemic is important for achieving long-term success, (b) preventing a large wave that overwhelms the public health system may not even be possible under some parameter combinations and (c) prioritizing the minimization of deaths versus days under control leads to vastly different outcomes. Kohler *et al* [18] use a compartmental model with 8 states and employ a robust model predictive control (MPC) based feedback policy. This policy adapts the social distancing measures cautiously and safely, thus leading to a minimum number of fatalities even if measurements are inaccurate and the infection rates cannot be precisely specified. Silva *et al* [19] considers a social opinion biased SAIRP model to provide forecasting mathematical models to anticipate the consequences of political decisions.

Kantner and Koprucki [20] compute the optimal non-pharmaceutical intervention strategy based on an extended SEIR (Susceptible-Exposed-Infected-Recovered) model and continuous time optimal control theory. The optimal control must satisfy the the following requirements: (1) minimize disease related deaths; (2) establish a sufficient degree of natural immunity at the end of horizon, to exclude a second wave; and (3) keep the socio-economic costs of interventions minimum. This model was then calibrated to reproduce the initial exponential growth phase of COVID-19 pandemic in Germany. The optimal intervention strategy can be structured into 3 phases. In the first phase, intervention begins with a strict initial lockdown to hold the effective reproduction number, $\mathcal{R}_{eff} < 1$. In the second phase, there is a *critical period* where the number of simultaneous cases is approximately held constant ($\mathcal{R}_{eff} \approx 1$). During this period, the non-pharmaceutical interventions are relaxed on a gradually increasing rate. Phase 3 commences after the critical period is over and the number of active cases start to decay. In Phase 3, a final moderate tightening of measures is required.

Kruse and Strack [21] derive optimal policies for social distancing in an SIR model. They find that the optimal policy has the following features: (1) if the death rate is not too sensitive to the number of infected, the optimal policy has two phases. A first phase of strong interventions, followed by a second phase with weaker interventions. (2) If

the cost of reducing transmission rate is linear, the optimal policy is always extreme (bang-bang).

Richard and co-authors [22] identify optimal age-stratified non-pharmaceutical interventions to implement as a function of time since the onset of epidemic. By applying optimal control theory, they arrive at a solution which minimizes deaths and control costs. This strategy is implemented for three countries with contrasted age distributions. They also show that this age-stratified policy strongly outperforms a constant uniform control over the whole population or over the younger population.

Bliman *et al* [23] aims to study how partial or total containment can be applied to an SIR epidemic model to minimize the epidemic final size (cumulative number of infected cases during the complete course of an epidemic). Theoretical and numerical results demonstrate that this approach can lead to a significant decrease in epidemic final size. It is shown that optimal intervention has to begin before the number of susceptible individuals has crossed the herd immunity level.

Dimarco *et al* [24] use an SEIAR type compartmental model to study the optimal control strategy when agents reduce their mean number of contacts. The novelty lies in using a kinetic-type model to account for heterogeneity in contact distribution of the population. Using a data driven approach to determine the relevant epidemiological parameters, they show that different types of control can lead to very different mitigation effects according to the level of heterogeneity in contact distribution of agents.

Arruda *et al* [25] incorporates multiple viral strains and reinfection to an SEIR model and studies optimal control policies. The model is validated from epidemiology data from COVID-19 in England and Brazil. They consider the cost of mitigation efforts to grow exponentially with the mitigation effort and solve an optimal control problem to to determine optimal mitigation measures. Their results point to the importance of controlling an epidemic from outset in hindering the emergence of new strains and avoiding the effects of a prolonged epidemic.

Morris *et al* [26] studies the role of non-pharmaceutical interventions in reducing or delaying the peak number of infected individuals. A classical SIR model is used to derive the theoretically optimum strategy and show that easier to implement strategies without perfect information about current state can perform near optimally. However, neither the optimal strategy nor the near-optimal strategies are robust strategies. Small deviations in intervention timing can cause large increases in the infection peak. Robust controls should therefore aim at strong, early interventions which are sustained in an ideal scenario.

The general conclusion which we can draw from the above literature is that for linear cost of control, the optimal policy switches from exerting maximum control effort until some point in time and then switches off the control efforts after that time point. This *bang-bang* solution, with at most one switch, is common in similar problems. Although the bang-bang solution is common, different types of solutions can be obtained for various formulations of the optimal control problem [27, 28]. For an SIR model with quadratic control cost, [28] shows that the optimal solution is not a bang-bang controller.

5.2 *Optimal control with vaccinations*

Abakuks [11] started the work on optimal vaccination strategies for epidemics. However, this model works under the assumption that the whole of susceptible population is instantaneously vaccinated once a vaccine is available. This work computes the optimal control for a stochastic model as well as a deterministic model and provides a comparative study.

Bauch and Earn [29] models spread of childhood diseases and analyze voluntary vaccination policies using a compartmental SIR model. Morton and Wickwire [30] consider an immunisation model for susceptible individuals where the vaccination control is bounded. This boundedness assumption on vaccination control is needed because in practical scenarios, there will be a bound on the maximum number of vaccines available. Kuga and co-authors [31] combine evolutionary game theory and mathematical epidemiology to evaluate the performance of vaccination subsidizing policies for a seasonal epidemic. Multi-agent simulations are used to find how the topology of network structure affects the vaccination behavior. Mean field approximations are used in this paper to confirm the simulation results and to see the change in social behavior when the vaccine is imperfect. The authors are able to point out instances where vaccine subsidizing policies could be counterproductive.

Arefin and co-authors [32] build a mean-field vaccination game to analyse the effect of an imperfect vaccine on a two-strain epidemic. The vaccination-decision takes place at the beginning of an epidemic season and depends upon the vaccine-effectiveness along with the cost. An additional situation where the original strain continuously converts to a strain by mutation is also considered.

Zaman *et al* [33] propose an SIR epidemic model where a percentage of susceptible population is vaccinated. They show that an optimal control exists for the optimal vaccination problem and describe numerical simulations using a Runge-Kutta fourth order procedure. Furthermore, a real-world example is constructed where smoking is modelled as an epidemic, demonstrating the efficiency of optimal control. Kar and Batabyal [34] study an optimal control problem with vaccination coverage as a control variable on an SIR epidemic model. With the help of the Pontryagin maximum principle and and an iterative method, it is shown that there are two equilibria, one a disease free equilibrium and the other an endemic equilibrium. The existence and stability of these equilibria are studied and the optimality system is then solved numerically.

Tchuenche and co-authors [35] analyze the dynamics of an influenza pandemic model with vaccination and treatment using two preventive scenarios: increase in vaccine uptake and decrease in vaccine uptake. The optimal control is computed using Potryagin's maximum principle and sensitivity analysis and simulations are performed to determine the relative importance of transmission parameters.

Acuna *et al* [36] formulate an optimal control problem where vaccination coverage (covering a certain percentage of the population in a given period) and hospital occupancy are constrained and identifies vaccination policies that minimizes the number of disability-adjusted years of life lost. A compartmental model is used in the analysis and the burden of COVID-19 is studied with respect to different scenarios such as optimal vs constant vaccination policies, vaccine efficacy, induced vaccine immunity and natural immunity.

5.2.1 *Note:*

It should be noted that although this section focused on optimal control in the context of epidemics, the same models and tools are directly applied to general spreading processes in complex networks. Examples of optimal control applied to these models include: malware propagation in computer networks [27, 37], opinion dynamics in social networks [38], and adoption of a new product in a marketplace. Nowzari *et al* [12] provides a survey on analysis and control of epidemics in complex networks.

6. Current state-of-the-art in mean field game modeling of epidemics

This section is devoted to the use of the mean field game approach in the study of epidemic spread and control. We have categorized the relevant literature into the following groups: (1) Analysis of epidemic spread and non-pharmaceutical interventions (2) Study of vaccination effects (3) Control and policy design for Epidemics.

The paper by Huang and Zhu [39] is a recent survey of the use of game theoretic models (including mean field games) for epidemic spread and control. The paper starts with a review of various models (such as the SIR model) for epidemic spread. The focus of the survey is on the use of the models in answering important questions such as what interventions, when to intervene, etc. The review also provides a taxonomy of the literature based on (1) types of games, such as static games, differential games, stochastic games, evolutionary games, and mean field games (2) types of interventions, such as social distancing, vaccination, quarantine, and antidotes; and (3) types of decision-makers, such as individual nodes, adversaries, and central authorities.

6.1 *Mean field game modeling and analysis of epidemic spread and non-pharmaceutical interventions*

Elie *et al* [40] formulate a model of COVID-19 spread and control using an SIR model with an embedded mean field game. The control in the SIR model is induced by the degree of contact among the individual nodes. An individual node can decrease the contact rate during the epidemic and this intervention has a social cost and an effort cost. The mean field game model is shown to have an equilibrium in which the transmission rate of the epidemic is reduced. The transmission rate achieved in the equilibrium is however higher compared to a socially optimal solution. Similar results are derived in the case of an SEIR model where an additional *exposed* state is introduced. The divergence between autonomous behavior and socially optimal behaviors are shown to be more prominent immediately before and immediately after the peak of the epidemic.

The paper by Petrakova and Krivorotko [41] presents a model for spread of an epidemic like COVID-19 by considering three separate groups of population, namely, suspectable (S), infected (I), removed (R) and cross-immune (C) ones. The model is based on the mean-field control inside these three groups of population. This model takes into account population heterogeneity and is therefore superior to a traditional SIR model. The numerical experiments are able to produce accurate estimates for COVID-19 spread in Novosibirsk, Russia for two 100-day periods.

Cho [42] presents a mean field game model of individual nodes in a population affected by an epidemic, where each node chooses a dynamic strategy of interactions, given the benefits of the interactions as well as the risk involved in getting infected. The mean field equilibrium that results from the non-cooperative game model is computed and the outcome is compared to the socially optimal outcome which maximizes the total utility of the population. It is shown that the mean field equilibrium strategy is to make more contacts than the level at which it would be socially optimal, in the absence of any public policy or incentives. If incentives are offered and the cost of incentivizing people is included, then it is shown that policies reducing contacts of the infected should continued to be enforced even after the peak of epidemic has passed. The paper also computes the price of anarchy to get an idea of the conditions under which the discrepancies between the mean field behaviour and socially optimal behaviour warrant public policy interventions.

Olmez *et al* [43] develop mean field game models for the evolution of epidemics. The specific problem that is modelled as a mean field game is the decision facing an individual node regarding the degree of social activity. This is modelled as a mean field game, taking into accounts healthcare related cost and benefits accruing from social interactions. The authors investigated the fully observed

setting as well as a partially observed setting. The paper presents a complete analysis of the fully observed case and some analytical results for the partially observed case. In the fully observed case, each individual node knows its epidemiological status perfectly. It turns out that a susceptible node will engage in a social interaction if any only the reward outweighs the risk whereas an infected individual will choose to quarantine. In the partially observed case, the nodes do not know their epidemiological status perfectly and it turns out that an infected node behaves like a susceptible node. This could make the epidemic spread faster. The paper [44] is a follow-up of the work in [43] and studies the behavior of self-interested agents in a large heterogeneous population and how pre-symptomatic agents can drive the growth in infection. A mean field type optimal control model is used to investigate the effect of partial observation on individual decision making.

Tembine [45] considers a class of mean-field-type games with discrete-continuous state spaces and presents Bellman systems that provide sufficiency conditions for the existence of mean-field-type equilibria in state-and-mean-field-type feedback form. The author derives unnormalized master adjoint systems (MASS) which provide a methodology powerful enough to model the propagation of the COVID-19 virus in the globe. Based on MASS, the author presents data-driven modelling and analytics for mitigating COVID-19. The model is very versatile and captures many aspects: untested cases, age-structure, decision-making, gender, pre-existing health conditions, location, testing capacity, hospital capacity, and a mobility map of local areas, including in-cities, inter-cities, and international aspects. The author shows that this data-driven model can capture the trends of the reported data on COVID-19.

The report by Bremaud [8] models the propagation of epidemics in which individual nodes have control on some parameters such as vaccination rate or social interactions. The report focuses on two models based on the standard SIR model. The first model focuses on vaccination control. The existence of a unique mean field equilibrium is shown and a numerical method is presented for computing the equilibrium. The second model captures the effect of social contacts. Here again, a mean field equilibrium is computed numerically.

Bremaud and Ullmo [46] consider a SIR compartmental model with social structure, where individuals are grouped by age and interact in different settings where individuals have contact with the other (schools, households, community, etc.). The mean field Nash equilibrium is computed in this setting and is compared with the social optimum. They also investigate how an approximation of socially optimum policy can be obtained with social policies like lockdown.

Aurell et al [1] discuss the modeling of epidemics using graphon games. A crucial assumption in mean field game theory is that nodes are indistinguishable and interact identically regardless of with whom they interact. In the context of modeling of epidemics, there is often a need to model the diversity of individual nodes and the variation of their interactions (for example, travel restrictions, multiple age groups with distinct social behavior and risk profiles, and a wide range of co-morbid conditions, etc.). Games with a large number of non-identical players can be analyzed with so-called graphon games whenever the network specifying the interactions is dense. A graphon can be viewed as the limit of a dense random graph. The authors develop a framework for epidemic modeling using graphon games and analyze their Nash equilibria. They provide a sufficient condition for the existence of a Nash equilibrium and propose a numerical approach based on machine learning tools to compute the equilibrium. The paper also presents numerical results on several applications of compartmental models for epidemics.

In the paper by Kordonis et al [47], the authors study a dynamic game model that captures social distancing behaviors during an epidemic, assuming a continuum of players and infection dynamics caused by individual choice. The authors assume a slight variant of the SIR model. The players have incomplete information about their infection state, and their choice of actions is determined by the individual beliefs on the probabilities of being susceptible, infected, or recovered. The cost of each player is determined by the infection and the contact factor. The authors show that a Nash equilibrium is guaranteed to exist and develop an efficient computational procedure for the same. Even when the players have the same parameters, they could exhibit different behaviors. The work studies the effect of various parameters like the vulnerability (co-morbidity for example) of players, the time horizon, and the various interventions on the optimal policies and the costs of the nodes.

The paper by Gao et al [48] is concerned with accurate modeling of COVID-19 evolution with mean field evolutionary dynamics (MFEDs) by invoking optimal transport theory and mean field games on graphs. The authors compute the payoff functions for different individual states from the commonly used replicator dynamics (RDs) and employ them to govern the evolution of epidemics. The authors compare epidemic modeling based on MFEDs with that based on RDs through numerical experiments. The efficacy of MFEDs is demonstrated by fitting the model to the COVID-19 statistics of Wuhan, China. The authors also analyze the effects of one-time social distancing as well as the seasonality of COVID-19 through the post-pandemic period.

6.2 *Mean field game modeling of the effect of vaccinations*

Doncel et al [49, 50] present a mean field game model under the SIR model when the individual nodes in the population choose when to get vaccinated. The authors

prove the existence of a unique mean field equilibrium which shows a bang-bang control behavior. Specifically, there is a threshold time until which the individuals will get vaccinated at maximal rate and beyond which they do not get vaccinated at all. Interestingly, the vaccination strategy that minimizes the total cost has the same structure as the mean field equilibrium. However, the vaccination period of the mean field equilibrium is always smaller than the one that minimizes the total cost. This essentially means that vaccination should be subsidized appropriately in order to nudge the people to exhibit optimal vaccination behavior.

Hubert and Turinici [51] consider an SIR model with vaccination where the vaccination is not mandatory. In particular, they study newborn vaccination. The evolution of each individual node is modelled as a Markov chain and the dynamics of the population is modelled using a mean field approach. The vaccination decision optimizes a criterion depending on the time-dependent societal vaccination rate and the future epidemic dynamics. It is shown that a Nash mean field equilibrium exists in the proposed model. The paper presents a method for numerical computation of the equilibrium.

Laguzet *et al* [52] analyzes individual vaccination strategies for a Susceptible-Infected-Recovered (SIR) model. The decision on whether or not to get vaccinated is taken by individual nodes and is based on the risk of infection, the possible side effects of the vaccine and the overall severity of the epidemic course. An important realistic condition is captured, namely limited capacity of vaccination. It is shown that a mean field equilibrium exists between the individual decisions and the epidemic evolution. This enables one to compute an optimal vaccination policy.

Salvarani and Turinici [53] model the behavior of individual nodes in protecting themselves against an epidemic when the vaccination is voluntary. They capture two aspects of vaccines namely imperfect efficacy and limited duration of protection. They show the existence of a Nash equilibrium, assuming non-persistent immunity. They also provide a numerical method for computing the equilibrium. Numerical experiments reveal interesting insights which are useful in planning a vaccination campaign (timing, etc.) in the realistic scenario of imperfect vaccines.

Lee *et al* [54] deals with the problem of transporting and distributing the COVID-19 vaccine(s), to achieve optimal control of the pandemic. The methodology is based on a mean-field variational problem in a spatial domain, which controls the propagation of pandemic by the optimal transportation strategy of vaccine distribution. The authors integrate the vaccine distribution into the mean-field SIR model designed by the authors in [55]. Experimental results show that the proposed model indeed provides effective strategies in vaccine distribution on a spatial domain.

6.3 *Mean field game approach to control and policy design for epidemics*

Aurell *et al* [56] model the spread and regulation of an epidemic as a Stackelberg game between the federal Government and the population. The Government which formulates mitigation policies through incentives is the leader in the game while the mean field of nodes (citizens) representing the population is the follower. A standard SIR model is considered to represent the spread of the epidemic. The paper considers for the first time, a compartmental model of epidemics capturing the interplay between independent autonomous nodes and a regulator. The individual nodes interact via a noncooperative game where utility maximization is equivalent to minimizing the individual cost by controlling the degree of interactions with other nodes (rates of transitions between the states). The Nash equilibrium of this non-cooperative game is determined. The leader, guided by a social objective, applies incentive policies and non-pharmaceutical interventions which determine the Nash equilibrium of the mean field game among the individual nodes.

The mean field game proposed is an extended MFG in which dependency on the joint distribution of action and state is explicitly taken into account (in earlier models of extended MFG, this dependency involved only the marginal distributions). In terms of the technical novelty of the model, this paper formulates the leader's problem under the constraint of the mean field Nash equilibrium of the individual nodes as an optimal control problem with two forward stochastic equations. The numerical solution of this problem uses an approximation of the population by an interacting particle system and the approximation of controls by neural networks, including the leader's policy. The optimization of the leader's cost is carried out using a variant of stochastic gradient descent to update the neural networks' parameters (see [56] for details).

The numerical experiments reveal two interesting insights: (a) In the first case, when the leader applies a containment policy, it is found that the nodes are more cautious about their interactions in the resulting Nash equilibrium when compared to a free spread scenario without any containment policy in place. An early lockdown has a higher impact than any action taken at a later point in time. (b) When the leader optimizes over its policies to minimize its own cost, the Nash equilibrium of the nodes shows interesting outcomes. For example, when the SIR model is expanded to include two additional states namely Exposed (E) and Deceased (D), the nodes are found to choose lower contact levels in their Nash equilibrium than recommended by the regulator.

Hubert *et al* [57] investigate the optimal control of an epidemic by offering incentives to lockdown and testing. The interaction between the Government and the individual nodes in the population is modelled as a principal-

node problem with moral hazard. This results in a Stackelberg game model. The spread of the epidemic is modelled by stochastic model with SIS or SIR compartments. The transmission rate of the epidemic is proposed to be controlled by decreasing the contact rate between individuals (through various means such as reducing physical interactions). This causes a social cost as well as monetary cost to the individual nodes. The paper proposes that the government can aid this through a tax or subsidy as well as implement a testing policy. The testing policy will enable to determine accurately the spread of the epidemic, facilitating isolation of infected individuals. The work derives an optimal form of the tax which is indexed on the (a) proportion of infected individuals and (b) the optimal effort of the population, namely the transmission rate chosen in response to this tax. This yields an optimization problem to be solved by the government, namely solving an appropriate Hamilton-Jacobi-Bellman equation. Experimental results show that the imposition of a tax policy will induce the individual nodes to cut down on their interactions.

The paper also studies the influence a testing policy can have on limiting the spread of the epidemic. If the testing is done aggressively, then individuals who have tested positive can be isolated. This will enable the individual nodes to interact more freely than if there were no testing and no isolation of individuals who tested positive.

Charpentier et al [58] consider an extended SIR model with several realistic features of COVID-19 pandemic incorporated. The authors derive an optimal policy for controlling the spread of the epidemic using and taking into account: (a) lockdown intervention as well as detection and isolation intervention, (b) the trade-off between the sanitary and the socio-economic cost of the pandemic, and (c) limited capacity of intensive care units (ICUs). A detailed sensitivity analysis is carried out with parameters chosen from the COVID-19 literature. It is shown that the optimal lockdown policy is structured into 4 phases: (1) A quick and strong lockdown intervention to stop the exponential growth of the contagion; (2) A short transition to reduce the prevalence of the virus; (3) a long period with full ICU capacity and stable virus prevalence; (4) a return to normal social interactions with disappearance of the virus. This optimal scenario avoids the second wave of infection, provided the lockdown is released sufficiently slowly. It is also shown that with aggressive testing followed by isolation of infected individuals, social distancing norms can be relaxed.

Lee et al [55] presents a mean field game model to control the spread of epidemics in the spatial domain. A standard SIR model is considered and the spatial velocities in the three states S,I, and R are chosen as the control variables. The authors consider three crowds, Susceptible, Infected, and Recovered, which evolve spatially. This is due to mobility, interactions, etc. The central planner seeks to mitigate the risk of infection by controlling the spatial velocity of the nodes in the three states. The paper provides efficient algorithms based on proximal primal-dual methods for obtaining the solutions. It is shown that the proposed model can be effectively used for identifying the infected and susceptible populations in a spatial domain.

Xu et al [59] discusses three specific COVID-19 epidemic control models: (1) the susceptible, exposed, infectious, recovered (SEIR) model with vaccination control; (2) the SEIR model with shield immunity control; and (3) the susceptible, un-quarantined infected, quarantined infected, confirmed infected (SUQC) model with quarantine control. The paper expresses control outcomes using metric temporal logic (MTL) which is a formal specification language. An example of a control outcome would be: the population immune from the disease should exceed 200 thousand in the next 100 to 200 days [59]. The paper presents methods to synthesize control strategies with MTL specifications and presents simulation results for three scenarios: (a) vaccination control for the COVID-19 epidemic; (b) shield immunity control for the COVID-19; and (c) quarantine control for the COVID-19.

7. Future directions

A prominent direction that has been pursued by many researchers is to investigate how socially optimal control and mean field game based control differ in an epidemic model. The control policies influence various costs including infection cost, vaccination cost, quarantine cost, etc. Assuming that these costs can be suitably chosen (with the support of a regulatory authority), a few researchers have looked into the problem of nudging the individuals of the population in a way that mean field game control approaches the performance of socially optimal control. Typically, researchers conduct mean field game analyses and subsequently suggest certain incentives or deterrents based on the associated equilibria. However, the analyses do not take into account how the individuals react to the policies, except for the work by Aurell et al [56] who arrive at an optimal contract by explicitly taking into account the reaction of the individuals through a Stackelberg game type of model. Here, the regulator announces a contract with recommends levels of social interaction for individuals in each compartment. Agents in the population incur a running cost for not adhering to the regulator recommendation. This two level optimization of Stackelberg type models mimics the real world scenario of government announcing a policy and a population of rational agents responding to the policy. One specific problem here is to extend the work of [56] by finding optimal levels of pharmaceutical interventions like vaccination and testing.

In the existing literature, the population is treated as consisting of a single group of individuals. In reality, however, there are multiple logical groups such that the weights on the compartmental model would be different for

different groups: (a) healthy, robust individuals; (b) individuals with co-morbidities; (c) elderly individuals (which are more vulnerable), etc. Although such models exist in the literature for modeling of epidemics, to the best of our knowledge, mean field control and mean field games have not been analysed for more complex models. While analysis of complex models can lead into difficulties, numerical methods can help circumvent these difficulties.

The ongoing pandemic is continuously evolving. Research and innovation have led to multiple vaccines and vaccinated people are well protected. However, compliance to vaccination and pandemic appropriate behaviour is not 100 percent. It would be valuable and interesting to capture some of these phenomena in the epidemic models. Data about the ongoing epidemic will be very useful for these studies. The models developed will be much more credible if available data is incorporated appropriately.

Testing plays an important role in controlling the spread of the epidemic and also in better management of the epidemic effects. For infected individuals, early detection is important for medical intervention and prevention of permanent harm to their bodies. It is important to study this problem as a social problem and compute the optimal testing strategy for the community as a whole through an appropriate model.

Analytical techniques are hard to develop for the problems mentioned in this paper. This offers a formidable technical challenge and will be an interesting research direction. Most existing solutions are numerical. Hu *et al* [60] gives an overview of state-of-the-art machine learning methods applied to stochastic control problems and games. Deep learning methods have enabled the possibility to solve problems even when the underlying MDP is complex or the dimension is high. This paper builds on recent advancements in machine learning methods for high dimensional PDEs and backward stochastic differential equations. Dayanikli *et al* [61] propose a single level numerical approach that uses machine learning techniques to solve two level Stackelberg problems between a regulator and a mean field population of agents. Therefore, another promising direction will be to develop improved numerical methods leveraging on advances in machine learning methods.

References

[1] Aurell A, Carmona R, Dayanikli G, and Lauriere M 2021 *Finite state graphon games with applications to epidemics.* arXiv preprint arXiv:2106.07859

[2] Sutton R S and Barto A 2018 *Reinforcement learning: An introduction.* MIT press, Cambridge

[3] Gast N, Gaujal B and Le Boudec J-Y 2012 Mean field for Markov decision processes: From discrete to continuous optimization. *IEEE Trans. Automat. Control* 57: 2266–2280

[4] Guan J, Wei Y, Zhao Y and Chen F 2020 Modeling the transmission dynamics of covid-19 epidemic: A systematic review. *J. Biomed. Res.* 34: 422

[5] Darling R 2002 *Fluid limits of pure jump Markov processes: A practical guide.* arXiv preprint arXiv:math/0210109

[6] Bertsekas D 2012 *Dynamic programming and optimal control: Volume I.* Athena Scientific, Nashua

[7] Carmona R and Delarue F *et al*. 2018 *Probabilistic theory of mean field games with applications I–II.* Springer

[8] Brémaud L and Restagno F 2021 Mean field game modeling of epidemic propagation

[9] Lenhart S and Workman J T 2007 *Optimal control applied to biological models.* Chapman and Hall/CRC, London

[10] Abakuks A 1973 An optimal isolation policy for an epidemic. *J. Appl. Prob.* 10: 247–262

[11] Abakuks A 1974 Optimal immunisation policies for epidemics. *Adv. Appl. Prob.* 6: 494–511

[12] Nowzari C, Preciado V M and Pappas G J 2016 Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Syst. Mag.* 36: 26–46

[13] Wickwire K H 1975 Optimal isolation policies for deterministic and stochastic epidemics. *Math. Biosci.* 26: 325–346

[14] Wickwire K 1977 Mathematical models for the control of pests and infectious diseases: A survey. *Theor. Popul. Biol.* 11: 182–238

[15] Behncke H 2000 Optimal control of deterministic epidemics. *Opt. Control Appl. Methods* 21: 269–285

[16] Tsay C, Lejarza F, Stadtherr M A and Baldea M 2020 Modeling, state estimation, and optimal control for the us covid-19 outbreak. *Sci. Rep.* 10: 10711

[17] Perkins T A and España G 2020 Optimal control of the covid-19 pandemic with non-pharmaceutical interventions. *Bull. Math. Biol.* 82: 118

[18] Köhler J, Schwenkel L, Koch A, Berberich J, Pauli P and Allgöwer F 2021 Robust and optimal predictive control of the covid-19 outbreak. *Annu. Rev. Control* 51: 525–539

[19] Silva C J *et al*. 2021 Optimal control of the covid-19 pandemic: Controlled sanitary deconfinement in Portugal. *Sci. Rep.* 11: 3451

[20] Kantner M and Koprucki T 2020 Beyond just "flattening the curve": Optimal control of epidemics with purely non-pharmaceutical interventions. *J. Math. Ind.* 10: 1–23

[21] Kruse T and Strack P 2020 Optimal control of an epidemic through social distancing. In: *Cowles Foundation Discussion Papers 2229R, Cowles Foundation for Research in Economics*, Yale University

[22] Richard Q, Alizon S, Choisy M, Sofonea M T and Djidjou-Demasse R 2021 Age-structured non-pharmaceutical interventions for optimal control of covid-19 epidemic. *PLoS Comput. Biol.* 17: e1008776

[23] Bliman P-A, Duprez M, Privat Y and Vauchelet N 2021 Optimal immunity control and final size minimization by social distancing for the sir epidemic model. *J. Optim. Theory Appl.* 189: 408

[24] Dimarco G, Toscani G and Zanella M 2022 Optimal control of epidemic spreading in the presence of social heterogeneity. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 380: 20210160

[25] Arruda E F, Das S S, Dias C M and Pastore D H 2021 Modelling and optimal control of multi strain epidemics, with application to covid-19. *PLoS ONE* 16: e0257512

[26] Morris D H, Rossine F W, Plotkin J B and Levin S A 2021 Optimal, near-optimal, and robust epidemic control. *Commun. Phys.* 4: 78

[27] Khouzani M. R, Sarkar S, and Altman E 2010 Dispatch then stop: Optimal dissemination of security patches in mobile wireless networks. In: *49th IEEE Conference on Decision and Control (CDC)*, IEEE, pp. 2354–2359

[28] Khouzani M, Sarkar S, and Altman E 2011 Optimal control of epidemic evolution. In: *2011 Proceedings IEEE INFOCOM*, IEEE, pp. 1683–1691

[29] Bauch C T and Earn D J 2004 Vaccination and the theory of games. *Proc. Natl. Acad. Sci.* 101: 13391–13394

[30] Morton R and Wickwire K H 1974 On the optimal control of a deterministic epidemic. *Adv. Appl. Prob.* 6: 622–635

[31] Kuga K, Tanimoto J and Jusup M 2019 To vaccinate or not to vaccinate: A comprehensive study of vaccination-subsidizing policies with multi-agent simulations and mean-field modeling. *J. Theor. Biol.* 469: 107–126

[32] Arefin M R, Kabir K A and Tanimoto J 2020 A mean-field vaccination game scheme to analyze the effect of a single vaccination strategy on a two-strain epidemic spreading. *J. Stat. Mech. Theory Exp.* 2020: 033501

[33] Zaman G, Han Kang Y and Jung I H 2008 Stability analysis and optimal vaccination of an sir epidemic model. *Biosystems* 93: 240–249

[34] Kar T and Batabyal A 2011 Stability analysis and optimal control of an sir epidemic model with vaccination. *Biosystems* 104: 127–135

[35] Tchuenche J, Khamis S, Agusto F and Mpeshe S 2011 Optimal control and sensitivity analysis of an influenza model with treatment and vaccination. *Acta Biotheoret.* 59: 1–28

[36] Acuña-Zegarra M A, Díaz-Infante S, Baca-Carrasco D and Olmos-Liceaga D 2021 Covid-19 optimal vaccination policies: A modeling study on efficacy, natural and vaccine-induced immunity responses. *Math. Biosci.* 337: 108614

[37] Khouzani M, Altman E, and Sarkar S 2009 Optimal quarantining of wireless malware through power control. *2009 Information Theory and Applications Workshop*, IEEE, pp. 301–310

[38] Albi G, Pareschi L, and Zanella M 2015 On the optimal control of opinion dynamics on evolving networks. *IFIP Conference on System Modeling and Optimization*, Springer, pp. 58–67

[39] Huang Y and Zhu Q 2022 Game-theoretic frameworks for epidemic spreading and human decision-making: A review. *Dyn. Games Appl.* 12: 7–48

[40] Elie R, Hubert E and Turinici G 2020 Contact rate epidemic control of covid-19: An equilibrium view. *Math. Modell. Nat. Phenom.* 15: 35

[41] Petrakova V and Krivorotko O 2022 Mean field game for modeling of covid-19 spread. *J. Math. Anal. Appl.* 154: 126271

[42] Cho S 2020 *Mean-field game analysis of sir model with social distancing*. arXiv preprint arXiv:2005.06758

[43] Olmez S Y, Aggarwal S, Kim J W, Miehling E, Başar T, West M and Mehta P G 2022 Modeling presymptomatic spread in epidemics via mean-field games. *2022 American Control Conference (ACC)*, IEEE, pp. 3648–3655

[44] Olmez S Y, Aggarwal S, Kim J W, Miehling E, Başar T, West M and Mehta P G 2022 How does a rational agent act in an epidemic? *2022 IEEE 61st Conference on Decision and Control (CDC)*, IEEE, pp. 5536–5543

[45] Tembine H 2020 Covid-19: Data-driven mean-field-type game perspective. *Games* 11: 51

[46] Bremaud L and Ullmo D 2022 Social structure description of epidemic propagation with a mean-field game paradigm. *Phys. Rev. E* 106: L062301

[47] Kordonis I, Lagos A -R and Papavassilopoulos G P 2022 Dynamic games of social distancing during an epidemic: Analysis of asymmetric solutions. *Dyn. Games Appl.* 12

[48] Gao H, Li W, Pan M, Han Z and Poor H V 2021 Modeling covid-19 with mean field evolutionary dynamics: Social distancing and seasonality. *J. Commun. Netw.* 23: 314–325

[49] Doncel J, Gast N and Gaujal B 2020 A mean field game analysis of sir dynamics with vaccination. *Prob. Eng. Inform. Sci.* 36: 482–499

[50] Gaujal B, Doncel J and Gast N 2021 Vaccination in a large population: Mean field equilibrium versus social optimum. In: *NETGCOOP 2020-10th International Conference on Network Games, Control and OPtimization*, pp. 1–9

[51] Hubert E and Turinici G 2018 Nash-mfg equilibrium in a sir model with time dependent newborn vaccination. *Ricerche Matemat.* 67: 227–246

[52] Laguzet L, Turinici G and Yahiaoui G 2016 Equilibrium in an individual-societal sir vaccination model in presence of discounting and finite vaccination capacity. In: *New trends in differential equations, control theory and optimization: Proceedings of the 8th congress of Romanian mathematicians*. World Scientific, pp. 201–214

[53] Salvarani F and Turinici G 2018 Optimal individual strategies for influenza vaccines with imperfect efficacy and durability of protection. *Math. Biosci. Eng.* 15

[54] Lee W, Liu S, Li W and Osher S 2022 Mean field control problems for vaccine distribution. *Res. Math. Sci.* 9: 51

[55] Lee W, Liu S, Tembine H, Li W and Osher S 2021 Controlling propagation of epidemics via mean-field control. *SIAM J. Appl. Math.* 81: 190–207

[56] Aurell A, Carmona R, Dayanikli G and Lauriere M 2022 Optimal incentives to mitigate epidemics: A Stackelberg mean field game approach. *SIAM J. Control Optim.* 60: S294–S322

[57] Hubert E, Mastrolia T, Possamaï D and Warin X 2022 Incentives, lockdown, and testing: From thucydides' analysis to the covid-19 pandemic. *J. Math. Biol.* 84: 37

[58] Charpentier A, Elie R, Laurière M and Tran V C 2020 Covid-19 pandemic control: Balancing detection policy and lockdown intervention under ICU sustainability. *Math. Modell. Nat. Phenom.* 15: 57

[59] Xu Z, Wu B and Topcu U 2021 Control strategies for covid-19 epidemic with vaccination, shield immunity and quarantine: A metric temporal logic approach. *PLoS ONE* 16: e0247660

[60] Hu R and Lauriere M 2023 *Recent developments in machine learning methods for stochastic control and games*. arXiv preprint arXiv:2303.10257

[61] Dayanikli G and Lauriere M 2023 *A machine learning method for Stackelberg mean field games*. arXiv preprint arXiv:2302.10440