Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(10): 7027–7045.
DOI: 10.3934/era.2026304
Received: 26 January 2026
Revised: 22 July 2026
Accepted: 11 August 2026
Published: 17 August 2026

Research article

Optimal one-round interventions for an SIR epidemic under two distinct time horizons

Jiaqi Yu, Yihong Li and Weiqiang Li*

School of Mathematics, North University of China, Taiyuan 030051, China

* **Correspondence:** Email: wqli@nuc.edu.cn.

Abstract: Non-pharmaceutical interventions play a vital role in mitigating the outbreak of infectious diseases. During the COVID-19 pandemic, numerous countries and regions imposed multiple rounds of intervention measures to curb the spread of the disease. To get insights into optimal implementation of one-round interventions, two optimal control problems were examined under distinct time horizons, based on whether the herd immunity was reached. Here, one-round interventions were defined as interventions that operated under a budget constraint. Both of optimal control policies were proved to be purely bang-bang with a single switch, but their timings of onset were different. These results may provide some theoretical support for developing intervention strategies.

Keywords: non-pharmaceutical interventions; one-round interventions; SIR model; time horizon

1. Introduction

The COVID-19 pandemic has catalyzed a major resurgence of interest in the optimal control of infectious diseases. Compared with pharmaceutical interventions [1–6], such as vaccines and drugs, non-pharmaceutical interventions (NPIs), including social distancing, school closure, and isolation, have been proved to play a vital role in mitigating the spread of disease, particularly at the early phase of the outbreak [7, 8]. Therefore, investigating the optimal implementation of NPIs is of significant practical value.

An enormous amount of attention has been devoted to the research on optimal NPIs. Among many topics, two issues have attracted a lot of interest. Considering the implementation of NPIs inevitably incurs societal and economic costs, it is unrealistic to impose interventions during the whole course of an outbreak. A portion of the literature, therefore, aims to identify optimal NPIs on a finite time interval [9, 10], during a fixed or maximal time duration, where interventions can be imposed [11–15], or under limited control resources [16, 17]. For this type of problem, the objective functional is typically to minimize the epidemic final size or the total incidence of the epidemic. It has been demonstrated that

optimal control policies consist of applying the maximal intervention level over a single time interval and no intervention over other intervals. On the other hand, it is widely recognized that an objective of implementing interventions is to ensure that the health-care system is not overwhelmed [18–21]. Accordingly, a state constraint, often named ICU constraint in the literature, is incorporated into optimal control problems: the disease prevalence cannot exceed a prescribed level, that is, the capacity of the health-care system. For this kind of problem, the main target is to minimize the total cost of control accumulated over the considered time horizon. Optimal solutions are proved to be so-called "filling the box" policies. Specifically, no action is adopted until the disease prevalence reaches the capacity of the health-care system, interventions are so imposed that the disease prevalence remains constant, and after the herd immunity has been reached interventions are completely lifted. Although there exist many differences between two classes of problems mentioned above, they share a common characteristic: the problems are discussed on long-term or infinite horizons, under which the herd immunity has been achieved.

In practice, as evidenced by the COVID-19 pandemic, at the early phase of the outbreak, due to the absence of effective vaccines, a series of intervention policies had been deployed in rolling consequence, with necessary breaks between rounds, in order to prevent the disease from spreading too rapidly. On the other hand, as the epidemic evolves, many conditions and states, such as the availability of new vaccines and the status of medical resources, may change. These facts indicate that the study of optimal interventions over a short-term horizon merits more attention.

In this paper, we attempt to find optimal one-round interventions over two distinct time horizons, depending on whether the herd immunity is reached. Here, we refer to one-round interventions as the ones under which the accumulative cost of control is equal to a given value [22]. The difference in time horizons for two optimal control problems is achieved by imposing different state constraints. Different from common methods [23], such as the Pontryagin's maximum principle and the Hamilton-Jacobi equation, inspired by the work of Angulo et al. [18], we propose an approach to solve the problems discussed in this paper. Our approach is based on the finding that if two state trajectories have identical endpoints, the costs of control along them can be compared under certain conditions. This extends the conclusion reached by Angulo et al. [18].

The remainder of this paper is organized as follows: In Section 2, we introduce an optimal control problem for an SIR model. Next, in Section 3 we derive the optimal solutions under two distinct time horizons. Finally, some conclusions and discussions are summarized in Section 4.

2. The optimal control problem

Consider optimal control problems of a classical SIR (Susceptible-Infected-Recovered) model [24,25]:

$$\begin{aligned} S' &= -(1-u)\beta S I, \\ I' &= (1-u)\beta S I - \gamma I, \\ R' &= \gamma I, \end{aligned} \tag{2.1}$$

where S, I, and R represent the proportions of susceptible, infected, and recovered individuals in a population, and β and γ are the transmission and recovery rates, respectively. In the above system, the demographic factors are not incorporated, implying that $S + I + R \equiv 1$. The effect of NPIs on epidemic dynamics is represented by the control variable u, a reduction factor in the transmission rate. In addition,

to account for the cost of control, we introduce a cost function Γ defined as

$$\Gamma(t; t_0) = \int_{t_0}^{t} u(t)\,\mathrm{d}t, \tag{2.2}$$

quantifying the accumulative cost incurred by interventions over the time interval $[t_0, t]$. Without causing ambiguities, we use the simplified notation $\Gamma(t)$ or Γ. For this form of cost function, we assume that the cost of control is proportional to the reduction factor and the length of time over which interventions are implemented. In this paper, it is assumed that initial values, S_0, I_0, and R_0, of the system (2.1) satisfy the following conditions: $0 < S_0 < 1$, $0 < I_0 < 1$, $0 \le R_0 < 1$, and $S_0 + I_0 + R_0 = 1$, and $\beta > \gamma$.

In this paper, we aim to find a control u^* and a terminal time t_T^* that maximizes the proportion of susceptible individuals at the terminal time, that is, the objective functional is $J(x, u, t_T) = S(t_T)$. The optimal control problem discussed can be written as below [26]:

$$\begin{cases} \max\limits_{u, t_T} \ J(x, u, t_T) = S(t_T) \\ \text{subject to} \\ x'(t) = f(x(t), u(t)), \ t \in [0, t_T], \\ g(x(t), u(t)) \ge 0, \ t \in [0, t_T], \\ x(0) = x_0, \ b(x(t_T), t_T) = 0, \end{cases} \tag{OC}$$

where $x = (S, I, \Gamma)^{\mathrm{T}}$ is the state vector. Here, the sign T denotes the transpose operation. Except two state variables S and I, the cost function Γ is added to the state vector x, governed by $\Gamma'(t) = u$. By means of this technique, the constraint on the accumulative cost of control can be converted into the one on the terminal state [27]. The state vector x and control function u are subject to two constraints. The mixed state constraint $g(x, u) := (u, u_{\max} - u)^{\mathrm{T}} \ge 0$ reflects that u is bounded by 0 and $u_{\max}$, where $u_{\max} \in (0, 1]$ represents the maximal intervention level. The terminal state constraint $b(x(t_T), t_T) := (I_C - I(t_T), \Gamma_C - \Gamma(t_T))^{\mathrm{T}} = 0$ indicates that the terminal proportion of infected individuals reaches a prescribed level $I_C \in (0, 1)$ and the accumulative cost of control over the interval $[0, t_T]$ is equal to a given value $\Gamma_C > 0$, which we call the cost limit. It should be emphasized that the cost function Γ is not necessarily equal to the cost limit Γ_C only at $t = t_T$; it's likely that the cost function Γ reaches the cost limit Γ_C at an intermediate time, and then maintains its value until $t = t_T$. The control function u is assumed to be piecewise continuous with finite intervals on which it takes nonzero values and remaining ones on which it is zero.

Below, we declare some facts. For an arbitrary control function, the existence and uniqueness of the solution to the augmented system (2.1) with the predetermined initial conditions can be guaranteed by the Picard-Lindelöf theorem. On each interval where the control function is continuous, the Picard-Lindelöf theorem reveals that the system (2.1) admits a unique solution. The whole solution can be formed in a splicing manner. Thus, the solution to the system (2.1) is a function with continuous components. The set $\Omega = \{(S, I, R) \in \mathbb{R}^3 \mid S \ge 0, I \ge 0, R \ge 0\}$ can be proved to be positively invariant with respect to the system (2.1). This is because the direction of the vector field $(-(1-u)\beta S I, (1-u)\beta S I - \gamma I, \gamma I)^{\mathrm{T}}$ on each coordinate plane is either tangent to the plane or pointing to the interior of Ω. Furthermore, the solution to the system (2.1) possesses the following property: $S(t) > 0$ and $I(t) > 0$ for all $t > 0$ [25]. Given a control function u, one has $\frac{\mathrm{d}S}{\mathrm{d}R} = -\frac{(1-u)\beta}{\gamma} S$, implying that $S(t) = S_0 \mathrm{e}^{-\int_0^t \frac{(1-u)\beta}{\gamma}\,\mathrm{d}R} > 0$ since $S_0 > 0$. On the other hand, $\frac{\mathrm{d}I}{\mathrm{d}t} = (1-u)\beta S I - \gamma I$ indicates that $I(t) = I_0 \mathrm{e}^{\int_0^t [(1-u)\beta S - \gamma]\,\mathrm{d}t} > 0$ due to $I_0 > 0$.

3. Results

Now, we state the main results in this paper. Before proceeding, we derive a lemma, which is related to two different phase trajectories linking the same beginning and ending points.

Lemma 3.1. *For two phase trajectories, L_1 and L_2, connecting the same beginning and ending points in the SI-plane, corresponding to two processes (x_1, u_1) and (x_2, u_2) defined on intervals $[t_0^L, t_1^L]$ and $[t_0^L, t_2^L]$, respectively, when $S_1(t) > \frac{\gamma}{\beta}$ for $t \in [t_0^L, t_1^L]$ and $S_2(t) > \frac{\gamma}{\beta}$ for $t \in [t_0^L, t_2^L]$, if L_1 is located above L_2, in other words, for two arbitrary times $t_1 \in (t_0^L, t_1^L)$ and $t_2 \in (t_0^L, t_2^L)$ where $S_1(t_1) = S_2(t_2)$, $I_1(t_1) > I_2(t_2)$ holds, then one has*

$$\Gamma_1(t_1^L) = \int_{t_0^L}^{t_1^L} u_1(t)\,\mathrm{d}t < \Gamma_2(t_2^L) = \int_{t_0^L}^{t_2^L} u_2(t)\,\mathrm{d}t;$$

when $S_1(t) < \frac{\gamma}{\beta}$ for $t \in [t_0^L, t_1^L]$ and $S_2(t) < \frac{\gamma}{\beta}$ for $t \in [t_0^L, t_2^L]$, if L_1 is located above L_2, then one has $\Gamma_1(t_1^L) > \Gamma_2(t_2^L)$. Here, $x_1 = (S_1, I_1, \Gamma_1)^{\mathrm{T}}$ and $x_2 = (S_2, I_2, \Gamma_2)^{\mathrm{T}}$ are the corresponding states.

Proof. By the system (2.1), one has

$$\begin{aligned}\Gamma_i(t_i^L) &= \int_{t_0^L}^{t_i^L} u_i(t)\,\mathrm{d}t = \int_{t_0^L}^{t_i^L} \left(1 + \frac{S_i'(t)}{\beta S_i(t) I_i(t)}\right) \mathrm{d}t \\ &= \int_{t_0^L}^{t_i^L} \left(-\frac{S_i'(t) + I_i'(t)}{\gamma I} + \frac{S_i'(t)}{\beta S_i(t) I_i(t)}\right) \mathrm{d}t\end{aligned} \tag{3.1}$$

for $i = 1, 2$. Then, the difference between $\Gamma_1(t_1^L)$ and $\Gamma_2(t_2^L)$ can be expressed as

$$\begin{aligned}\Gamma_1(t_1^L) - \Gamma_2(t_2^L) &= \int_{L_1} \left(-\frac{1}{\gamma I} + \frac{1}{\beta S I}\right) \mathrm{d}S - \frac{1}{\gamma I}\,\mathrm{d}I - \int_{L_2} \left(-\frac{1}{\gamma I} + \frac{1}{\beta S I}\right) \mathrm{d}S - \frac{1}{\gamma I}\,\mathrm{d}I \\ &= \oint_L \left(-\frac{1}{\gamma I} + \frac{1}{\beta S I}\right) \mathrm{d}S - \frac{1}{\gamma I}\,\mathrm{d}I = -\iint_D \left(\frac{\beta S - \gamma}{\beta \gamma S I^2}\right) \mathrm{d}S\,\mathrm{d}I,\end{aligned} \tag{3.2}$$

where L is the closed curve composed of L_1 and L_2 in the counterclockwise direction, and D is the region bounded by L. In the above process, we have used Green's theorem. From Eq (3.2), we conclude that if $S_1 > \frac{\gamma}{\beta}$ and $S_2 > \frac{\gamma}{\beta}$, then one has $\Gamma_1(t_1^L) < \Gamma_2(t_2^L)$; however, if $S_1 < \frac{\gamma}{\beta}$ and $S_2 < \frac{\gamma}{\beta}$, then $\Gamma_1(t_1^L) > \Gamma_2(t_2^L)$ holds. □

This lemma declares that under certain conditions the accumulative costs of control along two phase trajectories can be directly compared by examining their relative locations. It should be noted that Lemma 3.1 holds, even though $S_1(t) = \frac{\gamma}{\beta}$ on some interval $[t_1, t_2] \subset [t_0^L, t_1^L]$ or $S_2(t) = \frac{\gamma}{\beta}$ on some interval $[t_1, t_2] \subset [t_0^L, t_2^L]$.

3.1. Long-term horizon

To begin with, we consider the problem (**OC**) over the long-term horizon. In this part, it is assumed that $u_{\max} = 1$.

Theorem 3.2. *For the initial condition $x_0 = (S_0, I_0, \Gamma_0)^{\mathrm{T}}$, if $S_0 > \frac{\gamma}{\beta}$, $I_0 > I_C$, and $\Gamma_C < -\frac{1}{\gamma}\ln\frac{I_C}{I_0}$, then the optimal control for the problem (***OC***) has the following form:*

$$u^*(t) = \begin{cases} 0, & t \in [0, t_p); \\ 1, & t \in [t_p, t_p + \Gamma_C); \\ 0, & t \in [t_p + \Gamma_C, t_T^*]. \end{cases} \tag{3.3}$$

Here, t_p is the time instant at which the disease reaches the peak under no control, that is, $I'(t_p)\big|_{u=0} = 0$, or equivalently, $S(t_p)\big|_{u=0} = \frac{\gamma}{\beta}$, and t_T^* is the optimal terminal time.

Remark 3.3. *The condition of $\Gamma_C < -\frac{1}{\gamma}\ln\frac{I_C}{I_0}$ guarantees that under an arbitrary admissible control function, the herd immunity can be achieved at the terminal time, in other words, $S(t_T) < \frac{\gamma}{\beta}$. This condition reflects the requirement that the cost limit Γ_C is less than the accumulative cost of control along the phase trajectory in the SI-plane, which begins at the initial point (S_0, I_0), ends at the line $I = I_C$, and is generated by the control function $u = u_{\max}$.*

Remark 3.4. *The terminal equality constraint $\Gamma(t_T) = \Gamma_C$ can, in fact, be replaced with the inequality one $\Gamma(t_T) \le \Gamma_C$ in the problem (***OC***).*

To prove Theorem 3.2, it's sufficient to verify that for any admissible control, named u^r, that is the control satisfying the constraints of the problem (**OC**), we have $J(x^r, u^r, t_T^r) \le J(x^*, u^*, t_T^*)$, or more specifically, $S^r(t_T^r) \le S^*(t_T^*)$. Here, $x^r = (S^r, I^r, \Gamma^r)^{\mathrm{T}}$ and t_T^r are the state trajectory and the terminal time corresponding to u^r, respectively, and $x^* = (S^*, I^*, \Gamma^*)^{\mathrm{T}}$ is the state trajectory corresponding to u^*. Obviously, either u^r is purely bang-bang or u^r contains one or more singular components.

Let us begin by the case where there exist singular intervals within $[0, t_T^r]$. Without loss of generality, it is assumed that u^r contains only one singular component. Meanwhile, u^r may take its upper bound $u_{\max} = 1$ on other intervals. Therefore, the following arguments are split into two scenarios.

Scenario 1: there exists no interval on which $u^r = 1$. It implies that u^r is singular on an interval, named $[t_1, t_2)$, while $u^r = 0$ on the other interval(s). This can be further discussed in three cases: $0 < t_1 < t_2 < t_T^r$, $0 = t_1 < t_2 < t_T^r$, and $0 < t_1 < t_2 = t_T^r$, which we refer to as Cases 1–3, respectively.

We first consider Case 1. Below, it is assumed that $S^r(t_1) > \frac{\gamma}{\beta}$ and $S^r(t_2) < \frac{\gamma}{\beta}$. The non-optimality of u^r can be proved in two steps.

Step 1: Construct an auxiliary control function u^n. Define u^n as

$$u^n(t) = \begin{cases} 0, & t \in [t_1, t_p); \\ 1, & t \in [t_p, t_c^n); \\ 0, & \text{otherwise}, \end{cases} \tag{3.4}$$

where t_c^n is chosen under the condition that there exists a time $t_2^n > t_c^n$ at which $S^n(t_2^n) = S^r(t_2)$ and $I^n(t_2^n) = I^r(t_2)$. Here, $x^n = (S^n, I^n, \Gamma^n)^{\mathrm{T}}$ is the state trajectory corresponding to u^n. As shown in Figure 1(a), by comparing phase trajectories, denoted by L^n and L^r, of x^n and x^r in the SI-plane, we find that all components of L^n and L^r are the same except for two ones: L^n on $[t_1, t_i^n) \cup [t_i^n, t_2^n)$, named L_1^n and L_2^n, and L^r on $[t_1, t_i^r) \cup [t_i^r, t_2)$, called L_1^r and L_2^r. Here, t_i^n and t_i^r are the corresponding times at which L^n and L^r intersect. Obviously, L_1^n is located above L_1^r, while L_2^n is located below L_2^r. It should be noted that u^n is not admissible, which can be confirmed later.

Step 2: Compare u^n with u^*. It has been established in [18, 25] (refer to Section 2.1.2 in [25]) that if $u^n = 0$ on $[0, t_p)$, then one has

$$I^n(t) + S^n(t) - \frac{\gamma}{\beta}\ln S^n(t) = I^n(0) + S^n(0) - \frac{\gamma}{\beta}\ln S^n(0)$$

for $t \in (0, t_p)$, implying that

$$I^n(t_p) + S^n(t_p) - \frac{\gamma}{\beta}\ln S^n(t_p) = C_0^n. \tag{3.5}$$

Here, $C_0^n := I^n(0)+S^n(0)-\frac{\gamma}{\beta}\ln S^n(0)$ is a constant, determined by $S^n(0)$ and $I^n(0)$. Next, when $t \in (t_p, t_c^n)$, we have

$$S^n(t) = S^n(t_p),\ I^n(t) = I^n(t_p)\mathrm{e}^{-\gamma(t-t_p)}$$

since $(S^n)' = 0$ and $(I^n)' = -\gamma I^n$. Thus, it follows that

$$S^n(t_c^n) = S^n(t_p),\ I^n(t_c^n) = I^n(t_p)\mathrm{e}^{-\gamma\Delta t^n}. \tag{3.6}$$

Here, $\Delta t^n := t_c^n - t_p$. Finally, by the similar arguments, the following equality holds:

$$I^n(t_T^n) + S^n(t_T^n) - \frac{\gamma}{\beta}\ln S^n(t_T^n) = I^n(t_c^n) + S^n(t_c^n) - \frac{\gamma}{\beta}\ln S^n(t_c^n), \tag{3.7}$$

where t_T^n represents the terminal time corresponding to u^n. To sum up, from Eqs (3.5)–(3.7), we obtain

$$\begin{aligned} I^n(t_T^n) + S^n(t_T^n) - \frac{\gamma}{\beta}\ln S^n(t_T^n) &= I^n(t_c^n) + S^n(t_c^n) - \frac{\gamma}{\beta}\ln S^n(t_c^n) \\ &= I^n(t_p)\mathrm{e}^{-\gamma\Delta t^n} + S^n(t_p) - \frac{\gamma}{\beta}\ln S^n(t_p) \\ &= I^n(t_p)(\mathrm{e}^{-\gamma\Delta t^n} - 1) + C_0^n. \end{aligned}$$

Then, we further have

$$S^n(t_T^n) - \frac{\gamma}{\beta}\ln S^n(t_T^n) = I^n(t_p)(\mathrm{e}^{-\gamma\Delta t^n} - 1) + C_0^n - I^n(t_T^n). \tag{3.8}$$

On the other hand, under the control function u^*, $S^*(t_T^*)$ satisfies the following equality:

$$S^*(t_T^*) - \frac{\gamma}{\beta}\ln S^*(t_T^*) = I^*(t_p)(\mathrm{e}^{-\gamma\Gamma_C} - 1) + C_0^* - I^*(t_T^*). \tag{3.9}$$

Here, $C_0^* := I^*(0) + S^*(0) - \frac{\gamma}{\beta}\ln S^*(0)$.

By Lemma 3.1, we have $\Gamma^n(t_T^n) < \Gamma^r(t_T^r) = \Gamma^*(t_T^*) = \Gamma_C$, implying that u^n is not admissible. Recall that x^r and x^* are admissible state trajectories. Thus, it follows that $\Delta t^n < \Gamma_C$. From Eqs (3.8) and (3.9), we obtain $S^r(t_T^r) = S^n(t_T^n) < S^*(t_T^*)$. Note that $y - \frac{\gamma}{\beta}\ln y$ is a decreasing function of y when $y < \frac{\gamma}{\beta}$. Note also that $I^n(t_p) = I^*(t_p) = I(t_p)|_{u=0}$, $I^n(t_T^n) = I^*(t_T^*) = I_C$, and $C_0^n = C_0^*$. This is because both L^n and L^* begin at the same initial point, i.e., $S^n(0) = S^*(0)$ and $I^n(0) = I^*(0)$, and end at the line $I = I_C$.

For the case of $S^r(t_2) > \frac{\gamma}{\beta}$ and $S^r(t_1) < \frac{\gamma}{\beta}$, the similar results can be obtained based on Cases 2 and 3.

Next, we turn to Case 2. Now, we assume that $S^r(t_2) \geq \frac{\gamma}{\beta}$. If $S^r(t_2) < \frac{\gamma}{\beta}$, then Case 2 can be similarly analyzed as Case 1. To prove that u^r is not optimal, we introduce an auxiliary control function u^m. Define u^m as

$$u^m(t) = \begin{cases} 0, & t \in [0, t_c^m); \\ 1, & t \in [t_c^m, t_2^m); \\ 0, & \text{otherwise}, \end{cases} \tag{3.10}$$

where t_c^m and t_2^m are set such that $S^m(t_c^m) = S^r(t_2)$ and $I^m(t_2^m) = I^r(t_2)$. Here, $x^m = (S^m, I^m, \Gamma^m)^{\mathrm{T}}$ is the state trajectory corresponding to u^m. Let L^m be the phase trajectory of x^m in the SI-plane. As illustrated in Figure 1(b), there exists one distinct component between L^m and L^r: L^m on $[t_1, t_2^m)$ and L^r on $[t_1, t_2]$, named L_d^m and L_d^r. Obviously, L_d^m is located above L_d^r. By the similar arguments as Case 1, one has

$$\begin{aligned} I^m(t_T^m) + S^m(t_T^m) - \frac{\gamma}{\beta} \ln S^m(t_T^m) &= I^m(t_2^m) + S^m(t_2^m) - \frac{\gamma}{\beta} \ln S^m(t_2^m) \\ &= I^m(t_c^m) \mathrm{e}^{-\gamma \Delta t^m} + S^m(t_c^m) - \frac{\gamma}{\beta} \ln S^m(t_c^m) \\ &= I^m(t_c^m)(\mathrm{e}^{-\gamma \Delta t^m} - 1) + C_0^m, \end{aligned}$$

where t_T^m denotes the terminal time corresponding to u^m, $\Delta t^m := t_2^m - t_c^m$, and $C_0^m := I^m(0) + S^m(0) - \frac{\gamma}{\beta} \ln S^m(0)$. It implies that

$$S^m(t_T^m) - \frac{\gamma}{\beta} \ln S^m(t_T^m) = I^m(t_c^m)(\mathrm{e}^{-\gamma \Delta t^m} - 1) + C_0^m - I^m(t_T^m). \tag{3.11}$$

According to Lemma 3.1, we obtain $\Delta t^m < \Gamma_C$. Furthermore, we have $I^m(t_c^m) \leq I^*(t_p) = I(t_p)|_{u=0}$ since $S^r(t_2) \geq \frac{\gamma}{\beta}$. Thus, by comparing Eqs (3.13) and (3.9), we conclude that $S^r(t_T^r) = S^m(t_T^m) < S^*(t_T^*)$. Note that $I^m(t_T^m) = I^*(t_T^*) = I_C$, and $C_0^m = C_0^*$.

Finally, we focus on Case 3. In this part, we discuss the situation where $S^r(t_1) \leq \frac{\gamma}{\beta}$. For the case of $S^r(t_1) > \frac{\gamma}{\beta}$, Case 3 can be similarly treated as Case 1. An auxiliary control function u^w is introduced, defined as

$$u^w(t) = \begin{cases} 1, & t \in [t_1, t_c^w); \\ 0, & t \in [t_c^w, t_T^w]; \\ 0, & \text{otherwise}, \end{cases} \tag{3.12}$$

where t_c^w is chosen subject to the condition that there exists a time $t_T^w > t_c^w$ at which $S^w(t_T^w) = S^r(t_T^r)$ and $I^w(t_T^w) = I^r(t_T^r)$. Here, $x^w = (S^w, I^w, \Gamma^w)^{\mathrm{T}}$ and t_T^w are the state trajectory and the terminal time, corresponding to u^w, respectively. Let L^w be the phase trajectory of x^w in the SI-plane. Figure 1(c) shows that L^w and L^r coincide on $[0, t_1)$, but L^w defined on $[t_1, t_T^w]$ is located below L^r defined on $[t_1, t_T^r]$. Following the similar procedures as Case 1, one has

$$S^w(t_T^w) - \frac{\gamma}{\beta} \ln S^w(t_T^w) = I^w(t_1)(\mathrm{e}^{-\gamma \Delta t^w} - 1) + C_0^w - I^w(t_T^w), \tag{3.13}$$

where $\Delta t^w := t_c^w - t_1$, and $C_0^w := I^w(0) + S^w(0) - \frac{\gamma}{\beta} \ln S^w(0)$. Thus, we obtain $S^r(t_T^r) = S^w(t_T^w) < S^*(t_T^*)$, based on two facts that $\Delta t^w < \Gamma_C$ by Lemma 3.1 and $I^w(t_1) \leq I^*(t_p) = I(t_p)|_{u=0}$ due to $S^r(t_1) \leq \frac{\gamma}{\beta}$. Note that $I^w(t_T^w) = I^*(t_T^*) = I_C$ and $C_0^w = C_0^*$.

Scenario 2: there exist one or more intervals on which $u^r = 1$. In particular, we consider that $u^r = 1$ on only one interval. Thus, it is assumed that u^r is singular on $[t_1, t_2]$, $u^r = 1$ on $[t_3, t_4]$, and $u^r = 0$ on other intervals, where $0 \le t_1 < t_2 < t_3 < t_4 < t_T^r$. Below, we discuss the special case where $S^r(t_2) > \frac{\gamma}{\beta}$ and $S^r(t_3) = \frac{\gamma}{\beta}$. The whole proof that u^r is not optimal consists of two steps.

Step 1: Introduce an auxiliary control function u^v. Define u^v as

$$u^v(t) = \begin{cases} 0, & t \in [t_1, t_c^v); \\ 1, & t \in [t_c^v, t_2^v); \\ 1, & t \in [t_2^v + t_3 - t_2, t_2^v + t_4 - t_2); \\ 0, & \text{otherwise}, \end{cases} \tag{3.14}$$

where t_c^v and t_2^v are set such that $S^v(t_c^v) = S^r(t_2)$ and $I^v(t_2^v) = I^r(t_2)$. Here, $x^v = (S^v, I^v, I^v)^{\mathrm{T}}$ is the state trajectory corresponding to u^v. Let L^v be the phase trajectory of x^v in the SI-plane. Then, as revealed in Figure 1(d), we observe that L^v is different from L^r in one component: L^v on $[t_1, t_2^v)$ and L^r on $[t_1, t_2)$.

Step 2: Contrast u^v with u^*. Based on the similar analysis, one has

$$\begin{aligned} I^v(t_3^v) + S^v(t_3^v) - \frac{\gamma}{\beta}\ln S^v(t_3^v) &= I^v(t_2^v) + S^v(t_2^v) - \frac{\gamma}{\beta}\ln S^v(t_2^v) \\ &= I^v(t_c^v)\mathrm{e}^{-\gamma\Delta t_1^v} + S^v(t_c^v) - \frac{\gamma}{\beta}\ln S^v(t_c^v) \\ &= I^v(t_c^v)(\mathrm{e}^{-\gamma\Delta t_1^v} - 1) + C_0^v, \end{aligned}$$

where $t_3^v := t_2^v + t_3 - t_2$, $\Delta t_1^v := t_2^v - t_c^v$, and $C_0^v := I^v(0) + S^v(0) - \frac{\gamma}{\beta}\ln S^v(0)$. From the above result, we further obtain

$$I^v(t_3^v) = I^v(t_c^v)(\mathrm{e}^{-\gamma\Delta t_1^v} - 1) + I(C_0^v) > I(C_0^v)\mathrm{e}^{-\gamma\Delta t_1^v}.$$

Here, we have used two facts: the first one is that there exists a value $I(C_0^v)$ such that $I(C_0^v) + S^v(t_3^v) - \frac{\gamma}{\beta}\ln S^v(t_3^v) = C_0^v$; the second one is that $I^v(t_c^v) < I(C_0^v)$. This is because $S^v(t_c^v) > \frac{\gamma}{\beta}$ and $S^v(t_3^v) = \frac{\gamma}{\beta}$. The value $I(C_0^v)$ indicates the value of $I(t)$ at the time t, where $S(t) = \frac{\gamma}{\beta}$, based on the system (2.1) with $u = 0$. Thus, we further have

$$I^v(t_3^v)\mathrm{e}^{-\gamma\Delta t_2} > I(C_0^v)\mathrm{e}^{-\gamma\Delta t^v},$$

where $\Delta t_2 := t_4 - t_3$ and $\Delta t^v := \Delta t_1^v + \Delta t_2$. On the other hand, one has

$$\begin{aligned} I^v(t_T^v) + S^v(t_T^v) - \frac{\gamma}{\beta}\ln S^v(t_T^v) &= I^v(t_4^v) + S^v(t_4^v) - \frac{\gamma}{\beta}\ln S^v(t_4^v) \\ &= I^v(t_3^v)\mathrm{e}^{-\gamma\Delta t_2} + S^v(t_3^v) - \frac{\gamma}{\beta}\ln S^v(t_3^v) \\ &= I^v(t_3^v)\mathrm{e}^{-\gamma\Delta t_2} - I(C_0^v) + C_0^v, \end{aligned}$$

where $t_4^v := t_2^v + t_4 - t_2$.

By the foregoing results, we obtain

$$\begin{aligned} S^v(t_T^v) - \frac{\gamma}{\beta}\ln S^v(t_T^v) &= I^v(t_3^v)\mathrm{e}^{-\gamma\Delta t_2} - I(C_0^v) + C_0^v - I^v(t_T^v) \\ &> I(C_0^v)(\mathrm{e}^{-\gamma\Delta t^v} - 1) + C_0^v - I^v(t_T^v), \end{aligned} \tag{3.15}$$

Thus, by combining Eq (3.15) with Eq (3.9), we have $S^r(t_T^v) = S^v(t_T^v) < S^*(t_T^*)$ since $\Delta t^v < \Gamma_C$ and $I(C_0^v) = I^*(t_p) = I(t_p)|_{u=0}$. Note that $I^v(t_T^v) = I^*(t_T^*) = I_C$ and $C_0^v = C_0^*$.

Now, let us consider the case where u^r is purely bang-bang. We first assume that $u^r = 1$ on only one interval, named $[t_1, t_2)$. This can be further divided into three cases: $0 < t_1 < t_2 < t_T$, $0 = t_1 < t_2 < t_T$, and $0 < t_1 < t_2 = t_T$, which we call Cases 1–3, respectively.

Let us begin with Case 1. After the similar analysis, we have

$$\begin{aligned} I^r(t_T^r) + S^r(t_T^r) - \frac{\gamma}{\beta}\ln S^r(t_T^r) &= I^r(t_2) + S^r(t_2) - \frac{\gamma}{\beta}\ln S^r(t_2) \\ &= I^r(t_1)\mathrm{e}^{-\gamma\Gamma_C} + S^r(t_1) - \frac{\gamma}{\beta}\ln S^r(t_1) \\ &= I^r(t_1)(\mathrm{e}^{-\gamma\Gamma_C} - 1) + C_0^r, \end{aligned}$$

where $C_0^r := I^r(0) + S^r(0) - \frac{\gamma}{\beta}\ln S^r(0)$. Then, it follows that

$$S^r(t_T^r) - \frac{\gamma}{\beta}\ln S^r(t_T^r) = I^r(t_1)(\mathrm{e}^{-\gamma\Gamma_C} - 1) + C_0^r - I^r(t_T^r). \tag{3.16}$$

Thus, from Eqs (3.16) and (3.9), we conclude that $S^r(t_T^r) \le S^*(t_T^*)$. Note that $I^r(t_1) \le I^*(t_p) = I(t_p)|_{u=0}$, $I^r(t_T^r) = I^*(t_T^*) = I_C$, and $C_0^r = C_0^*$.

Next, let us consider Case 2. By the similar arguments, we obtain

$$S^r(t_T^r) - \frac{\gamma}{\beta}\ln S^r(t_T^r) = I^r(0)(\mathrm{e}^{-\gamma\Gamma_C} - 1) + C_0^r - I^r(t_T^r).$$

Thus, we have $S^r(t_T^r) \le S^*(t_T^*)$ due to $I^r(0) \le I^*(t_p) = I(t_p)|_{u=0}$.

Finally, we discuss Case 3. Under the similar procedures, the following equality holds:

$$\begin{aligned} S^r(t_T^r) - \frac{\gamma}{\beta}\ln S^r(t_T^r) &= S^r(t_1) - \frac{\gamma}{\beta}\ln S^r(t_1) \\ &= C_0^r - I^r(t_1) \\ &= C_0^r - I^r(t_T^r)\mathrm{e}^{\gamma\Gamma_C}. \end{aligned} \tag{3.17}$$

Let D be the difference between the right-hand sides of Eqs (3.17) and (3.9). Then, one has

$$\begin{aligned} D &= [I^*(t_p)(\mathrm{e}^{-\gamma\Gamma_C} - 1) + C_0^* - I^*(t_T^*)] - (C_0^r - I^r(t_T^r)\mathrm{e}^{\gamma\Gamma_C}) \\ &= I^*(t_p)(\mathrm{e}^{-\gamma\Gamma_C} - 1) + I_C(\mathrm{e}^{\gamma\Gamma_C} - 1) \\ &= (\mathrm{e}^{\gamma\Gamma_C} - 1)(I_C - I^*(t_p)\mathrm{e}^{-\gamma\Gamma_C}) < 0. \end{aligned}$$

This is because $\Gamma_C < -\frac{1}{\gamma}\ln\frac{I_C}{I_0}$, implying that $I_0\mathrm{e}^{-\gamma\Gamma_C} > I_C$. Thus, we obtain $S^r(t_T^r) \le S^*(t_T^*)$.

As for the case where $u^r = 1$ on two or more intervals, the arguments mentioned in Scenario 2 are applicable under some modifications. One main change is that the auxiliary control function is not needed any more.

To summarize, compared with any admissible control u^r, the control u^* given by Eq (3.3) brings about the maximal terminal proportion of susceptible individuals. Therefore, the pair (u^*, t_T^*) is the optimal solution to the problem (**OC**).

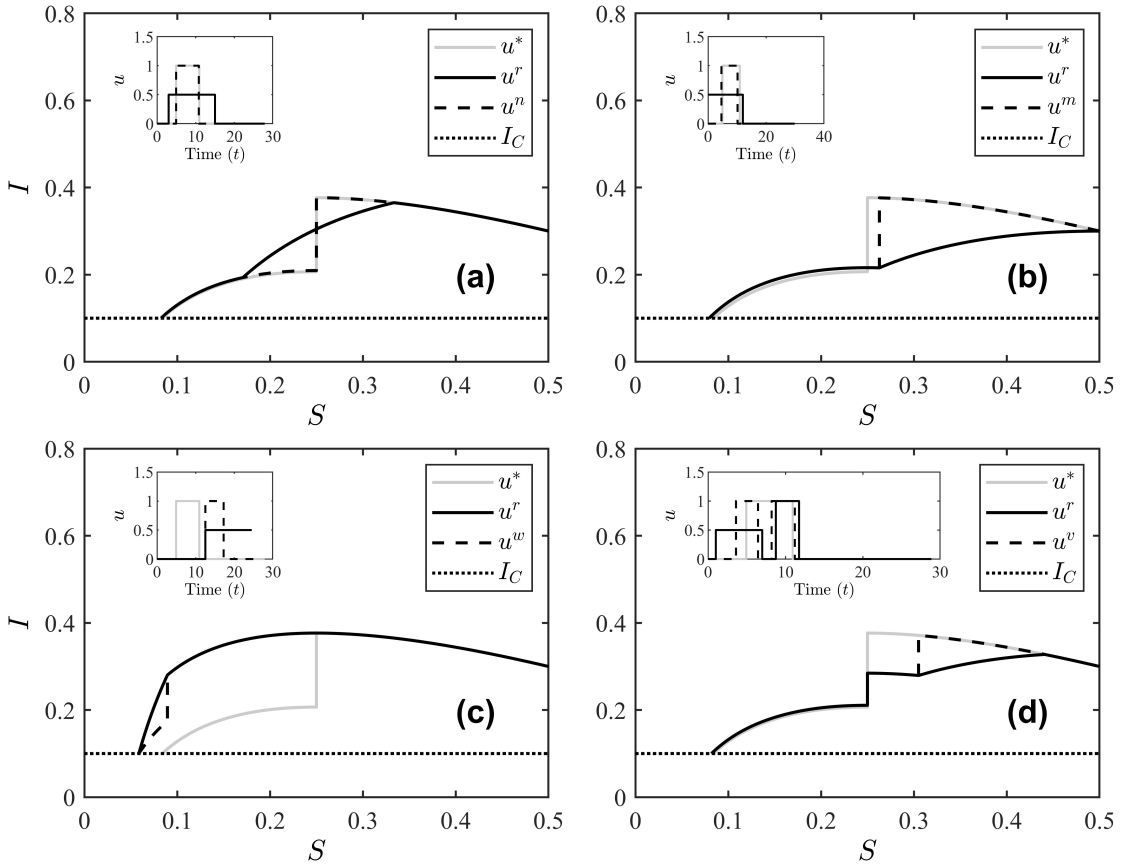

Figure 1. Comparison of two phase trajectories corresponding to an admissible control (black solid line), u^r, and the optimal control (gray solid line), u^*, for the problem (**OC**) in four cases. In the first three cases (a)–(c), u^r is singular on an interval $[t_1, t_2]$ and $u^r = 0$ on the other interval(s), where (a) $0 < t_1 < t_2 < t_T^r$, (b) $0 = t_1 < t_2 < t_T^r$, and (c) $0 < t_1 < t_2 = t_T^r$, while in the last case (d), u^r is singular on an interval $[t_1, t_2]$, $u^r = 1$ on another interval $[t_3, t_4]$, and $u^r = 0$ on other intervals, where $0 \le t_1 < t_2 < t_3 < t_4 < t_T^r$. Consider the case of $S^r(t_1) > \frac{\gamma}{\beta}$, and $S^r(t_3) = \frac{\gamma}{\beta}$. Here, $S^r(t)$ represents the proportion of susceptible individuals at the time t. In addition, four auxiliary controls (black dotted lines) associated with u^r are introduced: u^n, u^m, u^w, and u^v. The example phase trajectories are generated by three classes of control functions illustrated in the insets based on the problem (**OC**). Here, $\beta = 0.4$, $\gamma = 0.1$, $(S_0, I_0)^{\mathrm{T}} = (0.5, 0.3)^{\mathrm{T}}$, $I_C = 0.1$, and $\Gamma_C = 6$.

Remark 3.5. *It can be easily demonstrated that if $S_0 \le \frac{\gamma}{\beta}$ and $I_0 > I_C$, then the optimal control u^* for the problem (***OC***) has the following form: $u^* = 1$ on $[0, \Gamma_C)$, while $u^* = 0$ on $[\Gamma_C, t_T^*]$.*

Remark 3.6. *Scenario 2 can, in fact, be split into two categories: $S^r(t_2) > \frac{\gamma}{\beta}$ and $S^r(t_2) \le \frac{\gamma}{\beta}$. For the former category, once u^r is given on $[t_1, t_2)$, the optimal setting of u^r on $[t_2, t_T^r]$ is that $u^r = 1$ on an interval $[t_3, t_4)$ and $u^r = 0$ on other intervals, where $S^r(t_3) = \frac{\gamma}{\beta}$, and $t_4 - t_3 = \Gamma_C - \Gamma(t_2; t_1)$. This can be explained by Scenario 1, in which $x^r(t_2)$ is regarded as the new initial condition. For the latter category, when u^r is fixed on $[t_1, t_2)$, Remark 3.5 reveals that the following setting of u^r on $[t_2, t_T^r]$ is optimal: $u^r = 1$ on $[t_3, t_4)$ and $u^r = 0$ on the remaining interval, where $t_3 = t_2$ and $t_4 - t_3 = \Gamma_C - \Gamma(t_2; t_1)$. This form of u^r defined on $[0, t_T^r]$ has been proved to be not optimal.*

Remark 3.7. *For a general case where u^r contains three or more intervals on which u^r is singular or $u^r = 1$, a series of auxiliary control functions can be provided, each featuring one fewer interval of $u^r = 1$ than its predecessor. This process can be terminated since there exist a limited number of intervals on which u^r is singular or u^r=1. As an example, it is assumed that there exist three intervals, denoted by $[t_1, t_2)$, $[t_3, t_4)$, and $[t_5, t_6)$, on which u^r is singular or $u^r = 1$. The first auxiliary control function, named u^f, can be constructed as one of the following three alternatives according to the comparison of $u^r(t_3)$ and $u^r(t_6)$ with $\frac{\gamma}{\beta}$. The first option is that $u^f(t) = 0$, $t \in [t_3, t_p^f)$; $u^f(t) = 1$, $t \in [t_p^f, t_c^f)$; $u^f(t) = 0$, $t \in [t_c^f, t_6^f)$; $u^f(t) = 0$ on other intervals, where t_c^f is chosen under the condition that there exists a time $t_6^f > t_c^f$ at which $S^f(t_6^f) = S^r(t_6)$ and $I^f(t_6^f) = I^r(t_6)$. It should be noted that t_p^f is the time at which $S(t_p^f)|_{u=0} = \frac{\gamma}{\beta}$ based on the condition $(S(t_2), I(t_2))^{\mathrm{T}} = (S^r(t_2), I^r(t_2))^{\mathrm{T}}$. The second option is that $u^f(t) = 0$, $t \in [t_3, t_c^f)$; $u^f(t) = 1$, $t \in [t_c^f, t_6^f)$; $u^f(t) = 0$ on other intervals, where t_c^f and t_6^f are set such that $S^f(t_c^f) = S^r(t_6)$ and $I^f(t_6^f) = I^r(t_6)$. The final option is that $u^f(t) = 1$, $t \in [t_3, t_c^f)$; $u^f(t) = 0$, $t \in [t_c^f, t_6^f)$; $u^f(t) = 0$ on other intervals, where t_c^f is chosen subject to the condition that there exists a time $t_6^f > t_c^f$ at which $S^f(t_6^f) = S^r(t_6)$ and $I^f(t_6^f) = I^r(t_6)$. Thus, u^f can be regarded as a replacement of u^r with two intervals on which u^f is singular or $u^f = 1$. Then, the second auxiliary control function, called u^d, can be similarly formed by replacing u^f defined on $[t_1, t_6^f)$ with u^d defined on $[t_1, t_6^d)$ under the condition of $S^d(t_6^d) = S^f(t_6^f)$ and $I^d(t_6^d) = I^f(t_6^f)$. Therefore, u^d contains only one interval on which $u^d = 1$. On the other hand, based on the above process, we have $\Gamma^d(t_T^d) < \Gamma^f(t_T^f) < \Gamma^r(t_T^r)$ by Lemma 3.1. Obviously, u^d is not optimal compared with u^*. Here, $x^f = (S^f, I^f, \Gamma^f)^{\mathrm{T}}$ and $x^d = (S^d, I^d, \Gamma^d)^{\mathrm{T}}$ are state trajectories corresponding to u^f and u^d, and t_T^f and t_T^d are the corresponding terminal times, respectively.*

3.2. Short-term horizon

Next, we turn our attention to the short-term issue. Another state constraint is introduced to the problem (**OC**):

$$h(x(t)) = I_{\max} - I(t) \ge 0 \tag{H}$$

for $t \in [0, t_T]$, describing that the disease prevalence cannot exceed the capacity $I_{\max}$ of the health-care system during the whole horizon $[0, t_T]$. We call the problem (**OC**) coupled with the constraint (H) the problem (**OC'**) hereafter.

Theorem 3.8. *For the initial condition $x_0 = (S_0, I_0, \Gamma_0)^{\mathrm{T}}$, if $S_0 > \frac{\gamma}{\beta}$, $I_0 = I_C = I_{\max}$, and $\Gamma_C \le$*

$\int_0^T \left(1 - \frac{\gamma}{\beta}\frac{1}{S_0 - \gamma I_{\max} t}\right) \mathrm{d}t$, *then the optimal control for the problem (***OC'***) is given by*

$$u^*(t) = \begin{cases} u_{\max}, & t \in [0, \Gamma_C/u_{\max}); \\ 0, & t \in [\Gamma_C/u_{\max}, t_T^*]. \end{cases} \tag{3.18}$$

Here, $T = \frac{1}{\gamma I_{\max}}(S_0 - \frac{\gamma}{\beta})$, *and* t_T^* *is the optimal terminal time.*

Remark 3.9. *The restriction on* Γ_C *guarantees that for any admissible control function the terminal proportion of susceptible individuals is not less than the threshold parameter* $\frac{\gamma}{\beta}$, *that is,* $S(t_T) \geq \frac{\gamma}{\beta}$. *This restriction reflects the requirement that the cost limit* Γ_C *is less than the accumulative cost of control along the phase trajectory in the* SI*-plane, which begins at the initial point* $(S_0, I_{\max})$, *ends at the point* $(\frac{\gamma}{\beta}, I_{\max})$, *and stays at the line* $I = I_{\max}$.

For any admissible state trajectory of the problem (**OC'**), its phase trajectory in the SI-plane is, obviously, made up of one or more arc components, all of which begin and end at the line $I = I_{\max}$. The proof of Theorem 3.8 relies on the following propositions. Before stating them, we need to introduce some lemmas.

Let us define two kinds of arcs: boundary arcs and interior arcs. Given two points in the SI-plane, for example, the beginning point $m_b = (S_b, I_b)$ and the ending point $m_d = (S_d, I_d)$, both of which lie on $I = I_{\max}$, implying $I_b = I_d = I_{\max}$, consider all possible arcs connecting them. Here, the term 'arc' refers to the phase trajectory of state trajectory of the system (2.1) in the SI-plane. If $S_b > \frac{\gamma}{\beta}$ and $S_d > \frac{\gamma}{\beta}$, then Lemma 3.1 reveals that the most "costly" arc, i.e., the one along which the cost function is maximal, is generated by the following control function $\hat{u}$:

$$\hat{u}(t) = \begin{cases} u_{\max}, & t \in [t_b, t_c); \\ 0, & t \in [t_c, t_d], \end{cases} \tag{3.19}$$

where t_b, t_d, and t_c stand for the initial time, the terminal time, and the switching time, respectively. It should be stressed that once t_b and m_b are given, t_c and t_d can be uniquely determined by m_d, or, more precisely, S_d. To distinguish this arc from the other arcs, we call it the boundary arc and collectively the others interior arcs. Unless otherwise specified, both of arcs satisfy $S_b > \frac{\gamma}{\beta}$ and $S_d > \frac{\gamma}{\beta}$ hereafter.

Two conclusions about boundary arcs are drawn as below.

Lemma 3.10. *For two boundary arcs,* L^1 *and* L^2, *linking* $m_b^i = (S_b^i, I_b^i)$ *and* $m_d^i = (S_d^i, I_d^i)$ *for* L^i, $i = 1, 2$, *where* $S_b^1 = S_b^2$, *if* $S_d^1 < S_d^2$, *then* $\Gamma^1 > \Gamma^2$, *and vice versa. Here,* Γ^1 *and* Γ^2 *are the cost functions corresponding to* L^1 *and* L^2, *respectively.*

Obviously, Lemma 3.10 holds by Lemma 3.1.

Lemma 3.11. *When an initial time* t_b, *a beginning point* $m_b = (S_b, I_b)$, *and a parameter* Γ_0 *are given, there exist uniquely a switching time* t_c, *a terminal time* t_d, *and an ending point* m_d *such that a boundary arc linking* m_b *and* m_d *can be generated by the control function* $\hat{u}$ *of the form* (3.19), *along which the cost function is equal to* Γ_0.

Proof. Let us first deduce a result. By the system (2.1), we know that if $u \equiv u_0 \neq 1$ on an interval $[t_1, t_2]$, then the following equality holds [18, 25]:

$$I(t) + S(t) - \frac{\gamma}{(1 - u_0)\beta} \ln S(t) = I(t_1) + S(t_1) - \frac{\gamma}{(1 - u_0)\beta} \ln S(t_1) \tag{3.20}$$

for $t \in (t_1, t_2]$. From Eq (3.20), it can be concluded that once $S(t_1)$ and $I(t_1)$ are given, $I(t)$ can be uniquely expressed by $S(t)$. We use a function $I(S; S(t_1), I(t_1), u_0)$ to describe this relation. Note that this function is related to the control constant u_0. Thus, by the system (2.1), one has

$$t_2 - t_1 = -\frac{1}{(1-u_0)\beta} \int_{S(t_1)}^{S(t_2)} \frac{\mathrm{d}S}{S\, I(S; S(t_1), I(t_1), u_0)}. \tag{3.21}$$

In particular, when $u \equiv 1$ on $[t_1, t_2]$, one has

$$t_2 - t_1 = -\frac{1}{\gamma} \ln\left(\frac{I(t_2)}{I(t_1)}\right).$$

Now, if there exists a boundary arc, say L_{bd}, linking m_b and m_d under the control function $\hat{u}$ of the form (3.19), then one has

$$t_c - t_b = \begin{cases} -\frac{1}{(1-u_{\max})\beta} \int_{S_b}^{S_c} \frac{\mathrm{d}S}{S\, I(S; S_b, I_b, u_{\max})}, & u_{\max} < 1; \\ -\frac{1}{\gamma} \ln\left(\frac{I_c}{I_b}\right), & u_{\max} = 1, \end{cases} \tag{3.22}$$

and

$$I_d + S_d - \frac{\gamma}{\beta} \ln S_d = I_c + S_c - \frac{\gamma}{\beta} \ln S_c. \tag{3.23}$$

Here, $m_d := (S_d, I_d)$, $S_c := S(t_c)$, and $I_c := I(t_c)$. In addition, the cost function along L_{bd} is equal to $(t_c - t_b)u_{\max}$, i.e., $\Gamma_0 = (t_c - t_b)u_{\max}$, by which t_c can be uniquely determined. Next, Eq (3.22) reveals that its right-hand side is continuous with respect to S_c or I_c. Furthermore, it is also monotonic. This is because for the case of $u_{\max} < 1$, $S(t) > 0$ and $I(t) > 0$ for all t and S_c only serves as the upper bound of the integral. Thus, we can solve for S_c or I_c uniquely from Eq (3.22). Finally, S_d can be uniquely obtained from Eq (3.23) since $I_d = I_{\max}$, and thus t_d can be uniquely determined by Eq (3.21), where $u_0 = 0$, $t_1 = t_c$, and $t_2 = t_d$. □

Based on the above lemmas, we give the first proposition. In the following arguments, we make no distinction between state trajectories and their phase trajectories. The phrase 'state trajectory' is used to refer to either the state trajectory itself or its phase trajectory.

Proposition 3.12. *For the problem (***OC'***), the optimal state trajectory is composed of one or more boundary arc components.*

Proof. Suppose that the optimal state trajectory x^* contains an interior arc L^*. Now, we construct a new admissible state trajectory, which is made up of the boundary arc component(s).

Consider first the case where L^* is the last component of x^*. We introduce two boundary arcs, named L^1 and L^2, as shown in Figure 2. For the sake of brevity, we keep up the previous notations. On the one hand, both L^1 and L^2 begin from the same point as L^*. On the other hand, L^1 has the same ending point as L^*, while along L^2 and L^*, $\Gamma^2 = \Gamma^*$. Here, Γ^* is the cost function along L^*. Thus, according to Lemma 3.1, we have

$$\Gamma^1 > \Gamma^* = \Gamma^2.$$

Then, by Lemma 3.10, we further obtain

$$S_d^1 = S_d^* < S_d^2, \tag{3.24}$$

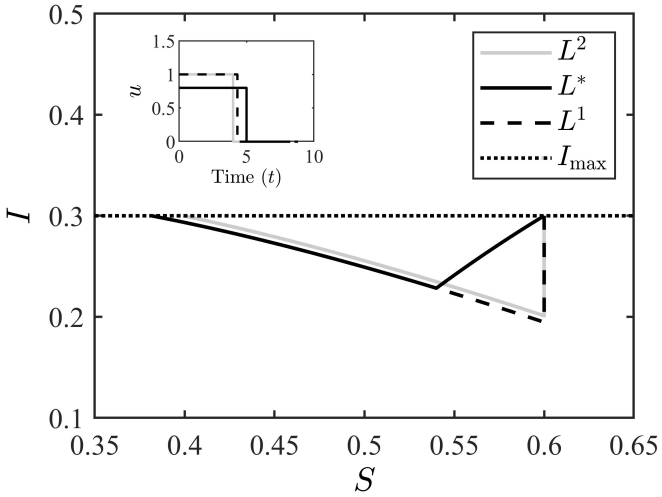

Figure 2. Comparison of three arcs: an interior arc (black solid line), L^*, and two boundary arcs (black dotted and gray solid lines), L^1 and L^2. Both L^1 and L^2 begin from the same point as L^*. In addition, L^1 has the same ending point as L^*, while along L^2 and L^*, $\Gamma^2 = \Gamma^*$. The terminal proportion of susceptible individuals is large for L^2 compared with L^1 and L^*. Three example arcs are formed by the control functions illustrated in the inset according to the system (2.1). Here, $\beta = 0.4$, $\gamma = 0.1$, $(S_0, I_0)^{\mathrm{T}} = (0.6, 0.3)^{\mathrm{T}}$, $I_{\max} = 0.3$, and $\Gamma^* = 4$.

where S_d^* stands for the terminal proportion of susceptible individuals corresponding to L^*. By replacing L^* with L^2 and keeping the other components unchanged, an admissible state trajectory x^{new} is formed. From Eq (3.24), we know

$$J(x^{new}, u^{new}, t_T^{new}) > J(x^*, u^*, t_T^*), \tag{3.25}$$

since L^* and L^2 are the last components of x^* and x^{new}, respectively. Here, u^{new} and t_T^{new} are the control function and the terminal time corresponding to x^{new}, respectively. This leads to a contradiction.

Next, let us turn to the other case. In other words, L^* lies in the middle part of x^*. Different from the previous case, we introduce an admissible state trajectory by two steps. First, we change L^* to a boundary arc with the same endpoints as L^*. As demonstrated before, this process gives rise to an increase in the cost function along the modified state trajectory, say x^{inter}, compared with the original one x^*. Second, we alter the last boundary arc component(s) of x^{inter} with a boundary arc such that the cost function along the new state trajectory, named x^{new}, is equal to the cost limit Γ_C. Thus, the state trajectory x^{new} is admissible. This step is feasible based on Lemma 3.11. It should be pointed out that in the second step the last two or more boundary arcs will involve in the alteration process, provided that the last one cannot compensate for the increase in the cost function during the first step. Since the cost function along the last component of x^{new} is less than the one along the last component(s) of x^{inter}, Eq (3.25) can be obtained similarly. This result contradicts the optimality assumption of x^*. □

The following proposition answers the question about the number of boundary arc components the optimal state trajectory contains for the problem (**OC'**).

Proposition 3.13. *For the problem (***OC'***), if two admissible state trajectories, x^n and x^m, consist of one*

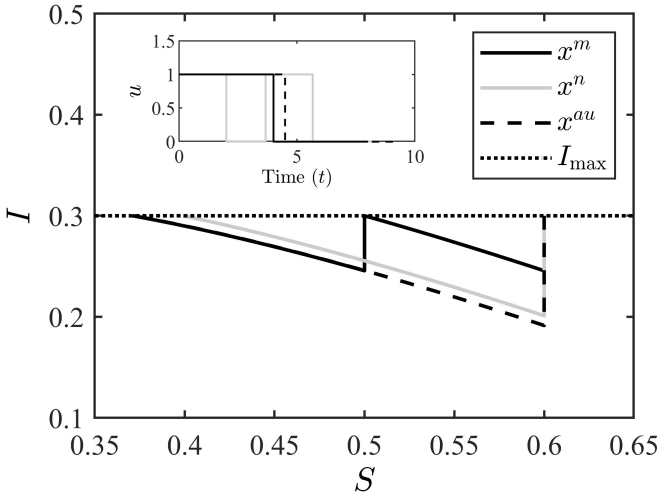

Figure 3. Comparison of three state trajectories: two admissible state trajectories with one boundary arc (gray solid line), x^n, and two boundary arcs (black solid line), x^m, and a state trajectory with one boundary arc (black dotted line), x^{au}. Both x^{au} and x^m have the same endpoints. The terminal proportion of susceptible individuals is larger for x^n than for x^m. Three example state trajectories are formed by the control function illustrated in the inset according to the problem (**OC'**). Here, $\beta = 0.4$, $\gamma = 0.1$, $(S_0, I_0)^{\mathrm{T}} = (0.6, 0.3)^{\mathrm{T}}$, $I_{\max} = 0.3$, and $\Gamma_C = 4$.

boundary arc and two boundary arcs, respectively, then one has

$$J(x^n, u^n, t_T^n) > J(x^m, u^m, t_T^m),$$

where u^n and t_T^n are the control function and the terminal time corresponding to x^n, and u^m and t_T^m are the ones corresponding to x^m, respectively.

Proof. Let us introduce an auxiliary state trajectory, say x^{au}, which contains only one boundary arc and has the same endpoints as x^m, as shown in Figure 3. Thus, we have

$$\Gamma^n = \Gamma^m < \Gamma^{au},$$

where Γ^n, Γ^m, and Γ^{au} are the corresponding cost functions. Then, by Lemma 3.10, we obtain

$$S_T^n > S_T^m = S_T^{au}.$$

Here, S_T^n, S_T^m, and S_T^{au} are the terminal proportions of susceptible individuals corresponding to three state trajectories. From the above result, it is concluded that

$$J(x^n, u^n, t_T^n) > J(x^m, u^m, t_T^m).$$

□

Proposition 3.13 shows that the performance of the admissible state trajectory with only one boundary arc is better than the one of the admissible trajectory with two boundary arcs. It can be easily found that

this result can be extended to a more general case, specifically speaking, the admissible state trajectory with three or more boundary arcs performs poorer than the one with one boundary arc. In fact, by constructing an auxiliary state trajectory, which is composed by one boundary arc and connects the same endpoints as the state trajectory with more than two boundary arcs, the above result can be similarly derived.

Based on above propositions, we conclude that the optimal state trajectory consists of one boundary arc, which implies that the control function of the form (3.18) is optimal.

Next, we relax the limitations on the initial condition and give the following corollary.

Corollary 3.14. *For the problem (**OC'**), if* $x_0 = (S_0, I_0, \Gamma_0)^{\mathrm{T}}$ *satisfies* $S_0 > \frac{\gamma}{\beta}$ *and* $I_0 < I_{\max}$, $I_C = I_{\max}$, *and* $\Gamma_C \leq \int_0^T \left(1 - \frac{\gamma}{\beta}\frac{1}{S_c - \gamma I_{\max} t}\right) \mathrm{d}t$, *then the optimal control has the form* (3.18). *Here,* $T = \frac{1}{\gamma I_{\max}}(S_c - \frac{\gamma}{\beta})$, *where* S_c *satisfies the equality* $I_{\max} + S_c - \frac{\gamma}{\beta}\ln S_c = I_0 + S_0 - \frac{\gamma}{\beta}\ln S_0$.

Proof. Under the hypotheses in the corollary, all admissible state trajectories can be classified into three categories according to the proportion of infected individuals at the time instant, say t_e, when interventions first come into effect: $I(t_e) = I_0$, $I(t_e) \in (I_0, I_{\max})$, and $I(t_e) = I_{\max}$.

Before proceeding, we need to introduce several types of arcs. We first extend the meaning of boundary arcs. Given two arbitrary points, among all arcs connecting these two points, the boundary arc is defined as the one that is generated by the control function of the form (3.19). It should be noted that the endpoints of boundary arcs are not required to lie on the line $I = I_{\max}$ anymore. Next, we call a special arc the free arc, named L^f, which begins at the initial point (S_0, I_0) and ends at the line $I = I_{\max}$, and along which no intervention is imposed. Related to the free arc L^f, we introduce three classes of boundary arcs: they begin at the beginning point, an intermediate point, and the ending point of L^f, and all end at the line $I = I_{\max}$, which we refer to as L^b, L^t, and L^d, respectively. Thus, we deduce that the optimal state trajectory is among three kinds of admissible state trajectories, named x^b, x^t, and x^d, which are composed of L^b, a part of L^f and L^t, and L^f plus L^d, respectively. This deduction is based on a fact that any admissible state trajectory containing one of L^b, L^t, and L^d, and another subsequent boundary arc, is not optimal.

Now, let us define two kinds of auxiliary state trajectories, say x^t_{au} and x^d_{au}, which have the same endpoints as x^t and x^d, respectively, and is made up of one boundary arc. Thus, according to Lemma 3.1, one has

$$\Gamma^t_{au} > \Gamma^t = \Gamma^b \quad \text{and} \quad \Gamma^d_{au} > \Gamma^d = \Gamma^b,$$

where Γ^b, Γ^t, Γ^d, Γ^t_{au}, and Γ^d_{au} are the corresponding costs functions. Then, we further obtain

$$S^t_T < S^b_T \quad \text{and} \quad S^d_T < S^b_T,$$

where S^b_T, S^t_T, and S^d_T are the terminal proportions of susceptible individuals corresponding to x^b, x^t, and x^d, respectively. To apply Lemma 3.10, we supplement x^t_{au}, x^d_{au}, and x^b with a same imaginary arc, which begins at the line $I = I_{\max}$, ends at the initial point (S_0, I_0), and is generated by $u = u_{\max}$. Based on this result, it is proved that x^b is the optimal state trajectory. □

4. Conclusions

In this paper, we investigate optimal one-round non-pharmaceutical interventions for an SIR epidemic under two distinct time horizons. The results show that both of the optimal solutions to the problems are

purely bang-bang with only a switch. However, their timings of onset are different: for the first problem, the optimal control policies come into effect at the time where the disease prevalence reaches the peak; for the second problem, the optimal intervention strategies are imposed at the initial time.

There exist some shortcomings in this work. First, the optimal control problems are developed under several hypotheses. For example, the cost function is assumed to be linear in the reduction strength, and the terminal proportion of infected individuals is set to equal to a predetermined constant. These assumptions are introduced mainly out of mathematical tractability, but not based on the realistic consideration. Thus, it hinders the obtained results from being applied to real-world scenarios. However, these researches may provide some guidelines on the formulation of intervention measures [16, 18]. Second, for the problem (**OC**), the assumption of $u_{\max} = 1$ is quite restrictive. Once $u_{\max} < 1$, the approach we develop is not valid anymore. To deal with this trouble, we need to introduce a new method or technique [16].

Some improvements will be made in future work. On the one hand, a more general terminal state constraint will be examined since the constraint $I = I_C$ lacks practical significance. On the other hand, an attempt will be made to incorporate the network structure into the models to reflect the heterogeneity of population contacts.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 12301640) and the Fundamental Research Program of Shanxi Province (Grant No. 202203021222078).

Conflict of interest

The authors declare there is no conflicts of interest.

References

1. L. Bolzoni, E. Bonacini, R. D. Marca, M. Groppi, Optimal control of epidemic size and duration with limited resources, *Math. Biosci.*, **315** (2019), 108232. https://doi.org/10.1016/j.mbs.2019.108232
2. L. Bolzoni, R. D. Marca, On the optimal vaccination control of SIR model with Erlang-distributed infectious period, *J. Optim. Theory Appl.*, **205** (2025), 39. https://doi.org/10.1007/s10957-025-02645-0
3. S. Dong, L. Xu, Y. A, Z. Z. Lan, D. Xiao, B. Gao, Application of a time-delay SIR model with vaccination in COVID-19 prediction and its optimal control strategy, *Nonlinear Dyn.*, **111** (2023), 10677–10692. https://doi.org/10.1007/s11071-023-08308-x
4. L. M. Moschen, M. S. Aronna, Bang-bang optimal control of vaccination in metapopulation epidemics with linear cost structures, *IEEE Control Syst. Lett.*, **9** (2025), 492–497. https://doi.org/10.1109/LCSYS.2025.3575227

5. F. Saldaña, Vaccination strategies in a pair formation model for human papillomavirus infection: an optimal control approach, *J. Theor. Biol.*, **597** (2025), 111994. https://doi.org/10.1016/j.jtbi.2024.111994
6. M. D. Rossa, L. Freddi, D. Goreac, Optimality of vaccination for an SIR epidemic with an ICU constraint, *J. Optim. Theory Appl.*, **204** (2025), 8. https://doi.org/10.1007/s10957-024-02598-w
7. S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, et al., Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, *Nature*, **584** (2020), 257–261. https://doi.org/10.1038/s41586-020-2405-7
8. S. Lai, N. W. Ruktanonchai, L. Zhou, O. Prosper, W. Luo, J. R. Floyd, et al., Effect of non-pharmaceutical interventions to contain COVID-19 in China, *Nature*, **585** (2020), 410–413. https://doi.org/10.1038/s41586-020-2293-x
9. P. Bliman, M. Duprez, Y. Privat, N. Vauchelet, Optimal immunity control and final size minimization by social distancing for the SIR epidemic model, *J. Optim. Theory Appl.*, **189** (2021), 408–436. https://doi.org/10.1007/s10957-021-01830-1
10. D. I. Ketcheson, Optimal control of an SIR epidemic through finite-time non-pharmaceutical intervention, *J. Math. Biol.*, **83** (2021), 7. https://doi.org/10.1007/s00285-021-01628-9
11. P. Bliman, M. Duprez, How best can finite-time social distancing reduce epidemic final size?, *J. Theor. Biol.*, **511** (2021), 110557. https://doi.org/10.1016/j.jtbi.2020.110557
12. Y. Feng, G. Iyer, L. Li, Scheduling fixed length quarantines to minimize the total number of fatalities during an epidemic, *J. Math. Biol.*, **82** (2021), 69. https://doi.org/10.1007/s00285-021-01615-0
13. P. Bliman, A. Rapaport, On the problem of minimizing the epidemic final size for SIR model by social distancing, *IFAC-PapersOnLine*, **56** (2023), 4049–4054. https://doi.org/10.1016/j.ifacol.2023.10.1726
14. R. Balderrama, M. I. Prieto, C. S. de la Vega, F. Vázquez, Optimal control for an SIR model with limited hospitalised patients, *Math. Biosci.*, **378** (2024), 109317. https://doi.org/10.1016/j.mbs.2024.109317
15. J. M. Greene, E. D. Sontag, Minimizing the infected peak utilizing a single lockdown: a technical result regarding equal peaks, in *2022 American Control Conference (ACC)*, IEEE, Atlanta, GA, USA, (2022), 3640–3647. https://doi.org/10.23919/ACC53348.2022.9867521
16. T. Britton, L. Leskelä, Optimal intervention strategies for minimizing total incidence during an epidemic, *SIAM J. Appl. Math.*, **83** (2023), 354–373. https://doi.org/10.1137/22M1504433
17. E. Molina, A. Rapaport, An optimal feedback control that minimizes the epidemic peak in the SIR model under a budget constraint, *Automatica*, **146** (2022), 110596. https://doi.org/10.1016/j.automatica.2022.110596
18. M. T. Angulo, F. Castaños, R. Moreno-Morton, J. X. Velasco-Hernández, J. A. Moreno, A simple criterion to design optimal non-pharmaceutical interventions for mitigating epidemic outbreaks, *J. R. Soc. Interface*, **18** (2021), 20200803. https://doi.org/10.1098/rsif.2020.0803
19. L. Miclo, D. Spiro, J. Weibull, Optimal epidemic suppression under an ICU constraint: an analytical solution, *J. Math. Econ.*, **101** (2022), 102669. https://doi.org/10.1016/j.jmateco.2022.102669

20. F. Avram, L. Freddi, D. Goreac, Optimal control of a SIR epidemic with ICU constraints and target objectives, *Appl. Math. Comput.*, **418** (2022), 126816. https://doi.org/10.1016/j.amc.2021.126816
21. D. Breda, M. D. Rossa, L. Freddi, Viability and control of a delayed SIR epidemic with an ICU state constraint, *ESAIM: COCV*, **31** (2025), 9. https://doi.org/10.1051/cocv/2024083
22. F. D. Lauro, I. Z. Kiss, J. C. Miller, Optimal timing of one-shot interventions for epidemic control, *PLoS Comput. Biol.*, **17** (2021), e1008763. https://doi.org/10.1371/journal.pcbi.1008763
23. F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*, Springer London, (2013), 435–501. https://doi.org/10.1007/978-1-4471-4820-3
24. W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. London, Ser. A*, **115** (1927), 700–721. https://doi.org/10.1098/rspa.1927.0118
25. M. Y. Li, *An Introduction to Mathematical Modeling of Infectious Diseases*, Springer Cham, (2018), 37–41. https://doi.org/10.1007/978-3-319-72122-4
26. S. P. Sethi, *Optimal Control Theory: Applications to Management Science and Economics*, Springer Cham, (2021), 119–121. https://doi.org/10.1007/978-3-030-91745-6
27. E. Hansen, T. Day, Optimal control of epidemics with limited resources, *J. Math. Biol.*, **62** (2011), 423–451. https://doi.org/10.1007/s00285-010-0341-0

AIMS Press